A DECISION PROBLEM FOR ULTIMATELY PERIODIC SETS IN NON-STANDARD NUMERATION SYSTEMS

EMILIE CHARLIER AND MICHEL RIGO

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 *numeration system* is given by a (strictly) increasing sequence $U=(U_i)_{i\ge 0}$ of integers such that $U_0=1$ and $C_U:=\sup_{i\ge 0}\lceil U_{i+1}/U_i\rceil$ 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 finite automaton). 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 2. 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 $\mathrm{rep}_U(x)<_{gen}\mathrm{rep}_U(y)$ where $<_{gen}$ is the genealogical ordering over A_U^*: words are ordered by increasing length and for words of same 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 $a<b$.

Definition 3. A numeration system $U=(U_i)_{i\ge 0}$ is said to be *linear*, if the sequence U satisfies a homogenous linear recurrence relation. For all $i\ge 0$, we have

$$U_{i+k}=a_1U_{i+k-1}+\cdots+a_kU_i \tag{1}$$

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

Example 4. Consider the sequence defined by $F_0=1$, $F_1=2$ and for all $n\ge 0$, $F_{n+2}=F_{n+1}+F_n$. The *Fibonacci (linear numeration) system* is given by $F=(F_i)_{i\ge 0}=(1,2,3,5,8,13,\ldots)$. 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 [2]. 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 [3].

J. Honkala showed in [6] 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 explicitely.

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 [7].

The structure of this paper is the same as [6]. First we give an upper bound on the admissible periods of a U-recognizable set when it is assumed to be ultimately periodic, then an upper bound on the admissible preperiods is obtained. 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 [6]. Actually they cannot be applied to integer base systems (see Remark **??**).

In the next section, Theorem **??** 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 [7]. 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 **??**. 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 numeration systems with $a_k = \pm 1$

Lemma 5. *Let $U = (U_i)_{i\geq 0}$ be a numeration system such that*

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

Then for all j, there exists $L \geq j$ such that for all $n \geq L - j$,

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

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.

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

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

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

Remark 6. 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 [?, 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 numeration system $U_\beta = (U_i)_{i\geq 0}$ associated with β (see [?]). If $d^*_\beta(1) = (t_i)_{i\geq 1}$, then

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

If β is a Parry number (i.e., $d_\beta(1)$ is finite or ultimately periodic) then the sequence U_β satisfies obviously a linear recurrence relation and as a consequence of Bertrand's theorem 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 [?] 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 7. 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 [?], then for $x, y \in \mathrm{rep}_F(\mathbb{N})$, we have $x0y \in \mathrm{rep}_F(\mathbb{N})$.

Definition 8. 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 $|u|$ (resp. $|v|$) 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 9. *Let $X \subseteq \mathbb{N}$ be an ultimately periodic set of period $|v|$ and preperiod $|u|$. Let $i, j \geq |u|$. If $i \not\equiv j \mod |v|$ then there exists $t < |v|$ 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 [?]) 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 10. *Let $U = (U_i)_{i\geq 0}$ be a numeration system satisfying condition* (2) *of Lemma 1. If $X \subseteq \mathbb{N}$ is an ultimately periodic U-recognizable set of period $|v|$ and preperiod $|u|$, then any deterministic finite automaton accepting $\mathrm{rep}_U(X)$ has at least $N_U(|v|)$ states.*

Proof. By Lemma 1, there exists L such that for any $h \geq L$, the words

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

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

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

and $h_1, \ldots, h_N$ can be chosen such that $U_{h_i} > |u|$ 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 $t_{i,j} < |v|$ 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,

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

is a word such that either

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

or

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

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

The previous proposition has an immediate consequence.

Corollary 11. *Let $U = (U_i)_{i\geq 0}$ be a numeration system satisfying condition* (2) *of Lemma 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 $m \geq s_0$, $N_U(m) > d$.

For a sequence $(U_i)_{i\geq 0}$ of integers, if $(U_i \mod m)_{i\geq 0}$ 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 2 is satisfied.

Lemma 12. *If $U = (U_i)_{i\geq 0}$ 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 $m \geq 2$, the sequence $(U_i \mod m)_{i\geq 0}$ is purely periodic. Indeed, for all $i \geq 0$, U_{i+k} is determined by the k previous terms $U_{i+k-1}, \dots, U_i$. But since $a_k = \pm 1$, for all $i \geq 0$, U_i is also determined by the k following terms $U_{i+1}, \dots, U_{i+k}$. So, by definition of $N_U(m)$, the sequence $(U_i \mod m)_{i\geq 0}$ 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 $U_{\alpha(m)} \geq m$ is nondecreasing and $\lim_{m\to+\infty} \alpha(m) = +\infty$. The conclusion follows, as $N_U(m) \geq \alpha(m)$. □

Remark 13. If the sequence $(U_i \mod m)_{i\geq 0}$ is not purely periodic as in the previous proof but only ultimately periodic, then a similar argument can be applied if it is assumed that $\lim_{m\to\infty} \pi_U(m) = +\infty$. By minimality of $\pi_U(m)$, for any $i, j \in \{\iota_U(m), \dots, \iota_U(m) + \pi_U(m) - 1\}$ such that $i \neq j$, the k-tuples

$$(U_i \mod m, \dots, U_{i+k-1} \mod m) \text{ and } (U_j \mod m, \dots, U_{j+k-1} \mod m)$$

are different. By definition of $N_U(m)$, the sequence $(U_i \mod m)_{i\geq \iota_U(m)}$ takes exactly $N_U(m)$ different values, and the maximal number of different k-tuples is $(N_U(m))^k$. Therefore, $N_U(m) \geq \sqrt[k]{\pi_U(m)}$ tends to infinity if $m \to +\infty$.

Remark 14. Let $U = (U_i)_{i\geq 0}$ be a numeration system satisfying hypothesis of Lemma **??** 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 2 can be estimated as follows.

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

Moreover, if U satisfies condition (2) of Lemma 1 and if X is an ultimately periodic set, then, by Corollary 2, 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 $(U_i \mod m)_{i\geq 0}$ for all $m \geq 2$.

Proposition 15. *Let $U = (U_i)_{i\geq 0}$ be a numeration system satisfying condition* (2) *of Lemma 1 and $X \subseteq \mathbb{N}$ be an ultimately periodic U-recognizable set of period $|v|$ and preperiod $|u|$. 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.*

Proof. Applying several times Lemma 1, there exist $n_1, \dots, n_{|v|}$ such that

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

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

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

and

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

For $i, j \in \{1, \dots, |v|\}$, $i \neq j$, 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 $0^{|\mathrm{rep}_U(|v|-1)| - |\mathrm{rep}_U(t)|} \mathrm{rep}_U(t)$ with $t < |v|$, 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 16. *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.*

The arguments of the following proof are similar to the one found in [6].

Proof. The sequence $(U_i \mod |v|)_{i\geq 0}$ is ultimately periodic with preperiod $\iota_U(|v|)$ and period $\pi_U(|v|)$. Proceed by contradiction and assume that $\mathcal{A}$ is a deterministic finite automaton with less than $|\operatorname{rep}_U(|u|-1)| - \iota_U(|v|)$ states accepting $\operatorname{rep}_U(X)$. The greedy U-representation of $|u|-1$ can be factorized as

$$\operatorname{rep}_U(|u|-1) = ww_4$$

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

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

By minimality of $|u|$ and $|v|$, either $|u|-1 \in X$ and for all $n \geq 1$, $|u|+n|v|-1 \notin X$, or $|u|-1 \notin X$ and for all $n \geq 1$, $|u|+n|v|-1 \in X$. But notice that

$$\operatorname{val}_U(w_1w_2^{|v|\pi_U(|v|)}w_2w_3w_4) \equiv \operatorname{val}_U(w_1w_2w_3w_4) \mod |v|,$$

leading to a contradiction. □

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

Lemma 17. *Let a, b be nonnegative integers and $U = (U_i)_{i\geq 0}$ be a linear numeration system. The language*

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

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

Before giving the proof, notice that for any integer $n \geq 0$, $\operatorname{val}_U^{-1}(n)$ is a finite set of words $\{x_1, \ldots, x_{t_n}\}$ over A_U such that $\operatorname{val}_U(x_i) = n$ for all $i = 1, \ldots, t_n$. This set contains in particular $\operatorname{rep}_U(n)$.

Proof. Since regular sets are stable under finite modification, we can assume that $0 \leq b < a$. The sequence $(U_i \mod a)_{i\geq 0}$ is ultimately periodic with preperiod $\ell = \iota_U(a)$ and period $p = \pi_U(a)$. It is an easy exercise to build a deterministic finite automaton $\mathcal{A}$ accepting reversal of the words in $\{w \in A_U^* \mid \operatorname{val}_U(w) \in a\mathbb{N}+b\}$. The alphabet of the automaton is A_U. States are pairs (r, s) where $0 \leq r < a$ and $0 \leq s < \ell + p$. The initial states is $(0, 0)$. Final states are the ones with the first component equal to b. Transitions are defined as follows

$$\forall s < \ell + p - 1 : \ (r, s) \xrightarrow{j} (jU_s + r \mod a,\ s+1)$$

$$(r, \ell + p - 1) \xrightarrow{j} (jU_s + r \mod a,\ \ell),$$

for all $j \in A_U$. Notice that $\mathcal{A}$ does not check the greediness of the accepted words, the construction only relies on the U-numerical value of the words modulo a.

For the particular case, one has to consider the intersection of two regular languages $\operatorname{rep}_U(\mathbb{N}) \cap \operatorname{val}_U^{-1}(a\mathbb{N}+b)$. □

Remark 18. 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 [8], 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 [5].

Theorem 19. *Let $U = (U_i)_{i\ge 0}$ 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) *of Lemma 1. 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 $\operatorname{rep}_U(X)$.

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 **??**). Then, using Proposition 5, the admissible preperiods are also bounded by a constant. Indeed, assume that X is ultimately periodic with period $|v| \le t_0$ and preperiod $|u|$. We have $\iota_U(|v|) = 0$ and any DFA accepting $\operatorname{rep}_U(X)$ must have at least $|\operatorname{rep}_U(|u|-1)|$ states. Therefore, the only values that $|u|$ can take satisfy $|\operatorname{rep}_U(|u|-1)| \le d$.

Consequently the sets of admissible preperiods and periods that we have to check are finite. Thanks to Lemma 3, one can build an automaton for each pair (i,p) of admissible preperiods and periods and then compare the language $L_{i,p}$ accepted by this automaton with $\operatorname{rep}_U(X)$. (Recall that testing whether $L_{i,p} \setminus \operatorname{rep}_U(X) = \emptyset$ and $\operatorname{rep}_U(X) \setminus L_{i,p} = \emptyset$ is decidable algorithmically). □

Remark 20. 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 $\gcd(a_1,\dots,a_k) = g \ge 2$, for all $n \ge 1$ and for all i large enough, we have $U_i \equiv 0 \mod g^n$ and assumption about $N_U(m)$ in Corollary 2 does not hold [4]. Indeed, the only value taken infinitely often by the sequence $(U_i \mod g^n)_{i\ge 0}$ 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 $b \ge 2$ numeration system where the only value taken infinitely often by the sequence $(b^i \mod b^n)_{i\ge 0}$ is 0, for all $n \ge 1$.

3. A decision procedure for a class of abstract numeration systems

Let $S = (L, \Sigma, <)$ be an *abstract numeration system* [7] 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. $\le j$) accepted from $q \in Q_L$ in $\mathcal{M}_L$. By classical arguments, the sequences $(\mathbf{u}_j(q))_{j\ge 0}$ (resp. $(\mathbf{v}_j(q))_{j\ge 0}$) satisfy the same homogenous linear recurrence relation for all $q \in Q_L$ (for details, see Remark **??**).

To define an abstract numeration system, L is genealogically ordered (words are ordered by increasing length and for words of same length, one uses the lexicographical ordering induced by the total ordering $<$ on the alphabet Σ), then 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 (the first word in L having index 0). A set $X \subseteq \mathbb{N}$ of integers is *S-recognizable* if the language $\operatorname{rep}_S(X)$ over Σ is regular (i.e., accepted by a finite automaton).

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 21. 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 22. 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. For all $n \geq 1$, we have $\mathbf{u}_n(q_{0,L}) = 2F_n$

FIGURE 1. A DFA accepting L.

and $\mathbf{u}_0(q_{0,L}) = 1$. Consequently, for $n \geq 1$,

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

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

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

Remark 23. 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 **??**.

Lemma 24. [7] *Let* $w = \sigma_1 \cdots \sigma_n \in L$. *We have*

$$\mathrm{val}_S(w) = \sum_{q \in Q_L} \sum_{i=1}^{|w|} \beta_{q,i}(w)\, \mathbf{u}_{|w|-i}(q) \tag{3}$$

where

$$\beta_{q,i}(w) := \#\{\sigma < \sigma_i \mid \delta_L(q_{0,L}, \sigma_1 \cdots \sigma_{i-1}\sigma) = q\} + \mathbf{1}_{q,q_{0,L}} \tag{4}$$

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 25. [7] *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.*

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 26. *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* $|v|$ *and preperiod* $|u|$, *then any deterministic finite automaton accepting* $\mathrm{rep}_S(X)$ *has at least* $\lceil N_{\mathbf{v}}(|v|)/\#Q_L \rceil$ *states where* $\mathbf{v} = (\mathbf{v}_j(q_{0,L}))_{j \geq 0}$.

Proof. Since for all states q of $\mathcal{M}_L$, we have $\lim_{j \to +\infty} \mathbf{u}_j(q) = +\infty$, there exists a minimal constant $J > 0$ such that $\mathbf{u}_J(q) \geq |v|$ for all $q \in Q_L$. Consider for any $j \geq 0$, the word

$$w_j = \mathrm{rep}_S(\mathbf{v}_j(q_{0,L})),$$

corresponding to the first word of length $j+1$ in the genealogically ordered language L. Consequently, for $j \geq J - 1$, w_j is factorized as $w_j = a_j b_j$ with $|b_j| = J$ and we define $q_j := \delta_L(q_{0,L}, a_j)$. Notice that b_j is the smallest word of length J accepted from q_j. By definition of J, from each q_j, there are at least $|v|$ words of length J leading to a final state. If we order them by genealogical ordering, we denote the $|v|$ first of them by

$$b_j = b_{j,0} < b_{j,1} < \cdots < b_{j,|v|-1}.$$

Notice that for $i \in \{0, \ldots, |v| - 1\}$, we have

$$\mathrm{val}_S(a_j b_{j,i}) = \mathrm{val}_S(a_j b_j) + i = \mathbf{v}_j(q_{0,L}) + i.$$

The sequence $(\mathbf{v}_j(q_{0,L}) \mod |v|)_{j \geq 0}$ is ultimately periodic and takes infinitely often $N_{\mathbf{v}}(|v|) =: N$ different values. Let $h_1, \ldots, h_N \geq J - 1$ such that

$$i \neq j \Rightarrow \mathbf{v}_{h_i}(q_{0,L}) \not\equiv \mathbf{v}_{h_j}(q_{0,L}) \mod |v|$$

and for all $j \in \{1, \ldots, N\}$, $\mathbf{v}_{h_j}(q_{0,L}) \geq |u|$. We have

$$\mathrm{rep}_S(\mathbf{v}_{h_j}(q_{0,L})) = w_{h_j} = a_{h_j} b_{h_j} \text{ and } q_{h_j} = \delta_L(q_{0,L}, a_{h_j}).$$

The elements in the set $\{q_{h_1}, \ldots, q_{h_N}\}$ can take only $\#Q_L$ different values. So at least $\sigma := \lceil N/\#Q_L \rceil$ of them are the same. For the sake of simplicity, assume that they are $q_{h_1}, \ldots, q_{h_\sigma}$. Consequently, for $i, j \in \{1, \ldots, \sigma\}$ and for all $k = 0, \ldots, |v|-1$, we have $b_{h_i,k} = b_{h_j,k}$. For all $i, j \in \{1, \ldots, \sigma\}$ such that $i \neq j$, by Lemma 2, there exists $t_{i,j} < |v|$ such that either $\mathbf{v}_{h_i}(q_{0,L}) + t_{i,j} \in X$ and $\mathbf{v}_{h_j}(q_{0,L}) + t_{i,j} \notin X$ or, $\mathbf{v}_{h_i}(q_{0,L}) + t_{i,j} \notin X$ and $\mathbf{v}_{h_j}(q_{0,L}) + t_{i,j} \in X$. Therefore, the words a_{h_i} and a_{h_j} do not belong to the same equivalence class for the relation $\sim_{\mathrm{rep}_S(X)}$. This can be shown by concatenating the word $b_{h_i,t_{i,j}} = b_{h_j,t_{i,j}}$. Hence the minimal automaton of $\mathrm{rep}_S(X)$ has at least σ states. □

Corollary 27. *Let $S = (L, \Sigma, <)$ be an abstract numeration system having the same properties as in Proposition **??**. 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 28. *Let $S = (L, \Sigma, <)$ be an abstract numeration system. If $X \subseteq \mathbb{N}$ is an ultimately periodic set of period $|v|$ such that $\mathrm{rep}_S(X)$ is accepted by a DFA with d states, then the preperiod $|u|$ of X is bounded by a constant C depending only on d and $|v|$.*

Proof. Let $\mathcal{A} = (Q, q_0, \Sigma, \delta, F)$ be a DFA with d states accepting $\mathrm{rep}_S(X)$. As usual, $\mathcal{M}_L = (Q_L, q_{0,L}, \Sigma, \delta_L, F_L)$ is the minimal automaton of L and for any state $q \in Q_L$, $\mathbf{u}_j(q)$ is the number of words of length j accepted from q in $\mathcal{M}_L$. Since $(\mathbf{u}_j(q))_{j\geq 0}$ satisfies a linear recurrence relation, the sequences $(\mathbf{u}_j(q) \mod |v|)_{j\geq 0}$ are ultimately periodic for all $q \in Q_L$. As usual, we denote by $\iota_{\mathbf{u}(q)}(|v|)$ (resp. $\pi_{\mathbf{u}(q)}(|v|)$) the preperiod (resp. the period) of $(\mathbf{u}_j(q) \mod |v|)_{j\geq 0}$. We set

$$I(|v|) := \max_{q\in Q_L} \iota_{\mathbf{u}(q)}(|v|)$$

and

$$P(|v|) := \mathrm{lcm}_{q\in Q_L}\, \pi_{\mathbf{u}(q)}(|v|).$$

For $|u|$ large enough, we have $|\mathrm{rep}_S(|u| - 1)| > d\,\#Q_L$. By the pumping lemma applied to the product automaton[1] $\mathcal{A} \times \mathcal{M}_L$, there exist x, y, z with $y \neq \varepsilon$, $|xy| \leq d\,\#Q_L$, $\delta(q_0, x) = \delta(q_0, xy)$, $\delta_L(q_{0,L}, x) = \delta_L(q_{0,L}, xy)$ and such that

$$\mathrm{rep}_S(|u| - 1) = xyz$$

and for all $n \geq 0$,

$$xy^n z \in \mathrm{rep}_S(X). \tag{5}$$

Since $|xy|$ is bounded by a constant, we also have $|z| > I(|v|)$ if $|u|$ is chosen large enough.

Since $|z| > I(|v|)$, using (**??**), (**??**) and for all $q \in Q_L$ the periodicity of the sequences $(\mathbf{u}_j(q) \mod |v|)_{j\geq 0}$, we have for all $\ell \geq 0$ that

$$\mathrm{val}_S(xy^{\ell|v|P(|v|)}yz) \equiv \mathrm{val}_S(xyz) \mod |v|. \tag{6}$$

Let us give some extra details on how we derive identity (**??**). Assume $x = x_1 \cdots x_r$, $y = y_1 \cdots y_s$ and $z = z_1 \cdots z_t$. For all $n \geq 1$, using (**??**) for $w = xy^n z$, we get

[1]The automaton $\mathcal{A} \times \mathcal{M}_L$ is defined as follows. For any state (q, q') in the set of states $Q \times Q_L$, when reading $a \in \Sigma$, one reaches in $\mathcal{A} \times \mathcal{M}_L$ the state $(\delta(q, a), \delta_L(q', a))$. The initial state is $(q_0, q_{0,L})$ and the set of final states is $F \times F_L$. Roughly speaking, the product automaton mimics the behavior of both automata $\mathcal{A}$ and $\mathcal{M}_L$.

$|w| = r + ns + t$ and

$$\begin{aligned}\mathrm{val}_S(xy^nz) &= \sum_{q\in Q_L}\left(\sum_{i=1}^{r}\beta_{q,i}(w)\mathbf{u}_{|w|-i}(q)\right.\\ &\quad+\sum_{i=r+1}^{r+s}\beta_{q,i}(w)\mathbf{u}_{|w|-i}(q)+\cdots+\sum_{i=r+(n-1)s+1}^{r+ns}\beta_{q,i}(w)\mathbf{u}_{|w|-i}(q)\\ &\quad\left.+\sum_{i=r+ns+1}^{r+ns+t}\beta_{q,i}(w)\mathbf{u}_{|w|-i}(q)\right),\end{aligned}$$

where the first (resp. second, third) line corresponds, as explained below, to the contribution of x (resp. y^n, z). By definition (**??**) of the coefficients $\beta_{q,i}(w)$, we know that $\beta_{q,1}(w)$ depends only on x_1 but $\beta_{q,2}(w)$ depends only on x_2 and on $\delta_L(q_{0,L}, x_1)$. Continuing this way, $\beta_{q,r}(w)$ depends only on x_r and on $\delta_L(q_{0,L}, x_1\cdots x_{r-1})$ and for $1 \le j \le s$, $\beta_{q,r+j}(w)$ depends on y_j and on $\delta_L(q_{0,L}, xy_1\cdots y_{j-1})$. Now $\beta_{q,r+s+1}(w)$ depends only on y_1 and on $\delta_L(q_{0,L}, xy_1\cdots y_s) = \delta_L(q_{0,L}, xy) = \delta_L(q_{0,L}, x)$. This implies that $\beta_{q,r+s+j}(w) = \beta_{q,r+j}(w)$ for all $q \in Q_L$ and all $j \in \{1,\ldots,s\}$. This argument can be repeated with every copy of y appearing in w. Consequently, the previous expansion become

$$\begin{aligned}\mathrm{val}_S(xy^nz) &= \sum_{q\in Q_L}\left(\sum_{i=1}^{r}\beta_{q,i}(w)\mathbf{u}_{|w|-i}(q)+\sum_{i=r+1}^{r+s}\beta_{q,i}(w)\underbrace{\sum_{j=0}^{n-1}\mathbf{u}_{|w|-i-js}(q)}_{(*)}\right.\\ &\quad\left.+\sum_{i=r+ns+1}^{r+ns+t}\beta_{q,i}(w)\mathbf{u}_{|w|-i}(q)\right).\end{aligned}$$

Assume now that $n = 1+\ell|v|P(|v|)$, with $\ell \ge 0$. For $q \in Q_L$ and $i = r+1,\ldots,r+s$, we have

$$(*) = \sum_{j=0}^{n-1}\mathbf{u}_{|w|-i-js}(q) = \mathbf{u}_{|w|-i}(q) + \sum_{j=1}^{\ell|v|P(|v|)}\mathbf{u}_{|w|-i-js}(q)$$

and the second term is congruent to 0 modulo $|v|$ due to the periodicity of the sequences $(\mathbf{u}_j(q) \mod |v|)_{j\ge 0}$ (recall that in the case we are considering, $|z| = t > I(|v|)$). Consequently, for $n = 1+\ell|v|P(|v|)$, we have

$$\begin{aligned}\mathrm{val}_S(xy^nz) &\equiv \sum_{q\in Q_L}\left(\sum_{i=1}^{r}\beta_{q,i}(w)\mathbf{u}_{|w|-i}(q)+\sum_{i=r+1}^{r+s}\beta_{q,i}(w)\mathbf{u}_{|w|-i}(q)\right.\\ &\quad\left.+\sum_{i=r+ns+1}^{r+ns+t}\beta_{q,i}(w)\mathbf{u}_{|w|-i}(q)\right) \mod |v|.\end{aligned}$$

It is then easy to derive (**??**).

We now use the minimality of $|u|$ to get a contradiction. Assume that $|u|-1$ is in X (the case not in X is similar). Therefore for all $n \ge 1$, $|u|+n|v|-1$ is not in X. From (**??**), for $\ell > 0$ we get $xy^{\ell|v|P(|v|)}yz \in \mathrm{rep}_S(X)$, but from (**??**) this word represents a number of the kind $|u|+n|v|-1$ with $n > 0$ which cannot belong to X. □

Remark 29. 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 $|u| > C$ implies $|\mathrm{rep}_S(|u|-1)| - d\,\#Q_L > I(|v|)$. Since the abstract numeration system S, the period $|v|$ and the number d of states are given, $I(|v|)$ and $\mathrm{rep}_S(n)$ for all $n \geq 0$ can be effectively computed.

Theorem 30. *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 **??**. 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 **??**, $\lim_{m \to +\infty} N_{\mathbf{v}}(m) = +\infty$. 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 $|v| \leq t_0$, then using proposition **??**, its preperiod is bounded by a constant (which can also be computed effectively thanks to Remark **??**).

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

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

[13] E. Zeckendorf, Représentation des nombres naturels par une somme des nombres de Fibonacci ou de nombres de Lucas, *Bull. Soc. Roy. Sci. Liège* **41** (1972), 179–182.

Institute of Mathematics, University of Liège, Grande Traverse 12 (B 37), B-4000 Liège, Belgium

E-mail address: {echarlier,M.Rigo}@ulg.ac.be