STRUCTURE OF THE MINIMAL AUTOMATON OF A NUMERATION LANGUAGE AND APPLICATIONS TO STATE COMPLEXITY

ÉMILIE CHARLIER, NARAD RAMPERSAD, MICHEL RIGO, AND LAURENT WAXWEILER

Abstract. We study the structure of automata accepting the greedy representations of $\mathbb{N}$ in a wide class of numeration systems. We describe the conditions under which such automata can have more than one strongly connected component and the form of any such additional components. Our characterization applies, in particular, to any automaton arising from a Bertrand numeration system. Furthermore, we show that for any automaton $\mathcal{A}$ arising from a system with a dominant root $\beta > 1$, there is a morphism mapping $\mathcal{A}$ onto the automaton arising from the Bertrand system associated with the number β. Under some mild assumptions, we also study the state complexity of the trim minimal automaton accepting the greedy representations of the multiples of $m \geq 2$ for a wide class of linear numeration systems. As an example, the number of states of the trim minimal automaton accepting the greedy representations of $m\,\mathbb{N}$ in the Fibonacci system is exactly $2m^2$.

1. Introduction

Cobham [11] showed that ultimately periodic sets of non-negative integers are the only sets that are recognized by a finite automaton in every integer base numeration system. The ultimately periodic sets are also exactly the sets definable by first order formulas in the Presburger arithmetic $\langle \mathbb{N}, + \rangle$. In the context of a non-standard numeration system U, if $\mathbb{N}$ is U-recognizable, then U is easily seen to be a linear numeration system, that is, U satisfies a linear recurrence with integer coefficients [24]. For linear numeration systems, ultimately periodic sets are all recognized by finite automata if and only if $\mathbb{N}$ is (see Theorem 1 below). Conditions on a linear numeration system U for $\mathbb{N}$ to be U-recognizable are considered in [16, 21]. From the point of view of the Chomsky hierarchy, a U-recognizable set X of integers can be considered as having a low computational complexity: the greedy representations of the elements in X in the numeration system U have simple syntactical properties recognized by some finite automaton, i.e., $\mathrm{rep}_U(X)$ is a regular language. Since the seminal work of Alan Cobham [11] many properties of U-recognizable sets have been investigated, e.g., algebraic, logical or automatic characterizations of U-recognizable sets for integer base numeration systems [7], extensions of these characterizations to systems based on a Pisot number [6], study of the normalization map [14], introduction of abstract numeration systems [19], ... Among linear numeration systems for which $\mathbb{N}$ is U-recognizable, the class of systems whose characteristic polynomial is the minimal polynomial of a Pisot number has been widely studied [6]. An example of such a system is given by the Fibonacci numeration system (see Example 2). In particular, the automata accepting these numeration languages are well-known. Another well-known class of numeration languages, which has given rise to many successful applications concerning β-numerations, consists of the languages arising from Bertrand systems associated with a Parry number (see Section 2) [5, 15].

Currently little is known about the automata accepting other kind of numeration languages. In the first part of this paper we study the structure of these automata for a wide class of numeration systems. In Section 2 we review the needed background concerning numeration systems. Then in Section 3 we provide several examples in order to illustrate the different types of automata that can arise from these numeration systems. In Section 4 we describe the conditions under which such automata can have more than one strongly connected component and the form of any such additional strongly connected component. In the case where the numeration system has a dominant root $\beta > 1$ (see the next section for the definition), we are able to provide a more specific description of the structure. For instance, we show that for any automaton $\mathcal{A}$ arising from a numeration system with a dominant root $\beta > 1$, there is a morphism mapping $\mathcal{A}$ onto the automaton arising from the Bertrand system associated with the number β.

Our primary motivation is to understand the state complexity of languages of the form $0^* \operatorname{rep}_U(m\mathbb{N})$, that is, the language of the representations of the multiples of m in a given numeration system U (see [1, 18]), in connection with the following decidability problem. Let U be a linear numeration system and X be a U-recognizable set of non-negative integers given by some deterministic finite automaton recognizing the greedy representations of elements of X. For integer base systems, Honkala proved that one can decide whether or not X is ultimately periodic [17]. Another, shorter proof of this result can be found in [2]. For a wide class of linear numeration systems containing the Fibonacci numeration system, the same decidability question is answered positively in [10, 3]. For all the above mentioned reasons ultimately periodic sets of integers and, in particular, the recognizability of a given divisibility criterion by finite automata deserve special interest.

Lecomte and Rigo [19] showed the following: given a regular language $L = \{w_0 < w_1 < \cdots\}$ genealogically ordered, extracting from L words whose indices belong to an ultimately periodic set $I \subset \mathbb{N}$ is a regularity-preserving operation defining a language L_I. Krieger *et al.* [18] considered the state complexity of this operation. If the minimal automaton of L has n states, it is natural to give bounds or try to estimate the number of states of the minimal automaton of L_I as a function of n, the preperiod and period of I. Such results could be useful in solving the decidability question mentioned in the last paragraph. For example, Alexeev [1] recently gave the following formula for the number of states of the minimal automaton of the language $0^* \operatorname{rep}_b(m\mathbb{N})$, that is, the set of b-ary representations of the multiples of $m \geq 1$. Let N, M be such that $b^N < m \leq b^{N+1}$ and $(m, 1) < (m, b) < \cdots < (m, b^M) = (m, b^{M+1}) = (m, b^{M+2}) = \cdots$. The minimal automaton of $0^* \operatorname{rep}_b(m\mathbb{N})$ has exactly

$$\frac{m}{(m, b^{N+1})} + \sum_{t=0}^{\inf\{N, M-1\}} \frac{b^t}{(m, b^t)} \tag{1}$$

states.

In the second part of this paper, we study the state complexity for the divisibility criterion by $m \geq 2$ in the framework of linear numeration systems. Under some mild assumptions, Theorem 10 gives the number of states of the trim minimal automaton of $0^* \operatorname{rep}_U(m\mathbb{N})$ from which infinitely many words are accepted. As a corollary, we show that, for a certain class of numeration systems, we can give the precise number of states of this automaton. For instance, for the Fibonacci numeration system, the corresponding number of states is $2m^2$, see Corollary 2. Finally we are able to give a lower bound for the state complexity of $0^* \operatorname{rep}_U(m\,\mathbb{N})$ for any numeration system.

Note that the study of state complexity could possibly be related to the length of the formulas describing such sets in a given numeration system. It is noteworthy that for linear numeration systems whose characteristic polynomial is the minimal polynomial of a Pisot number, U-recognizable sets can be characterized by first order formulas of a convenient extension of $\langle \mathbb{N}, + \rangle$, see [6].

2. Background on Numeration Systems

In this paper, when we write $x = x_{n-1} \cdots x_0$ where x is a word, we mean that x_i is a letter for all $i \in \{0, \ldots, n-1\}$.

An increasing sequence $U = (U_n)_{n \geq 0}$ of integers is a *numeration system*, or a *numeration basis*, if $U_0 = 1$ and $C_U := \sup_{n \geq 0} \lceil \frac{U_{n+1}}{U_n} \rceil < +\infty$. We let A_U be the alphabet $\{0, \ldots, C_U - 1\}$. A greedy representation of a non-negative integer n is a word $w = w_{\ell-1} \cdots w_0$ over A_U satisfying

$$\sum_{i=0}^{\ell-1} w_i U_i = n \quad \text{and} \quad \forall j \in \{1, \ldots, \ell\}, \quad \sum_{i=0}^{j-1} w_i U_i < U_j.$$

We denote the greedy representation of $n > 0$ satisfying $w_{\ell-1} \neq 0$ by $\operatorname{rep}_U(n)$. By convention, $\operatorname{rep}_U(0)$ is the empty word ε. The language $\operatorname{rep}_U(\mathbb{N})$ is called the *numeration language*. A set X of integers is *U-recognizable* if $\operatorname{rep}_U(X)$ is regular, i.e., accepted by a finite automaton. If $\mathbb{N}$ is U-recognizable, then we let $\mathcal{A}_U = (Q_U, q_{U,0}, F_U, A_U, \delta_U)$ denote the trim minimal automaton of the language $0^* \operatorname{rep}_U(\mathbb{N})$ having $\#\mathcal{A}_U$ states. The *numerical value map* $\operatorname{val}_U : A_U^* \to \mathbb{N}$ maps any word $d_{\ell-1}\cdots d_0$ over A_U to $\sum_{i=0}^{\ell-1} d_i U_i$. For example, if $(U_0, U_1, U_2) = (1, 2, 3)$ and $A_U = \{0, 1\}$, then $\mathrm{val}_U(100) = 3$ and $\mathrm{val}_U^{-1}(3) = \{11, 100\}$.

Definition 1. A numeration system $U = (U_n)_{n\geq 0}$ is said to be *linear*, if there exist $k \geq 1$ and $a_0, \ldots, a_{k-1} \in \mathbb{Z}$ such that

$$\forall n \in \mathbb{N},\ U_{n+k} = a_{k-1}U_{n+k-1} + \cdots + a_0 U_n. \tag{2}$$

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

Theorem 1. [4, Proposition 3.1.9] *Let $p, r \geq 0$. If $U = (U_n)_{n\geq 0}$ is a linear numeration system, then*

$$\mathrm{val}_U^{-1}(p\,\mathbb{N} + r) = \{w \in A_U^* \mid \mathrm{val}_U(w) \in p\,\mathbb{N} + r\}$$

is accepted by a deterministic finite automaton that can be effectively constructed. In particular, if $\mathbb{N}$ is U-recognizable, then any eventually periodic set is U-recognizable.

Let u, v be two finite words of the same length (resp. two infinite words) over an alphabet $A \subset \mathbb{N}$. We say that u is *lexicographically less* than v and we write $u < v$, if there exist $p \in A^*$, $a, b \in A$ with $a < b$ and words u', v' over A such that $u = pau'$, $v = pbv'$. If u and v are two finite words (not necessarily of the same length), then we say that u is *genealogically less* than v if either $|u| < |v|$, or $|u| = |v|$ and $u < v$ (with respect to the lexicographic order). We also write $u < v$ to denote the genealogical order. Note that if U is a numeration system, then for all $m, n \in \mathbb{N}$, we have $m < n$ if and only if $\mathrm{rep}_U(m)$ is genealogically less than $\mathrm{rep}_U(n)$.

Observe that if uv is a greedy representation, then so is v. However, if u is a greedy representation, there is no reason for $u0$ to still be greedy. As an example, if $U_0 = 1$, $U_1 = 3$ and $U_2 = 5$, then 2 is a greedy representation but 20 is not.

Definition 2. A numeration system $U = (U_n)_{n\geq 0}$ is a *Bertrand numeration system* if, for all $w \in A_U^+$, $w \in \mathrm{rep}_U(\mathbb{N}) \Leftrightarrow w0 \in \mathrm{rep}_U(\mathbb{N})$.

Let us recall the theorems of Bertrand [5] (also see [22, Thm. 7.3.8]) and Parry [23] (also see [22, Thm. 7.2.9]). Let $\beta > 1$ be a real number. The *β-expansion* of a real number $x \in [0, 1]$ is the sequence $d_\beta(x) = (x_i)_{i\geq 1} \in \mathbb{N}^\omega$ satisfying

$$x = \sum_{i=1}^{+\infty} x_i \beta^{-i}$$

and which is the maximal element in $\mathbb{N}^\omega$ having this property with respect to the lexicographic order over $\mathbb{N}$. Note that the β-expansion is also obtained by using the greedy algorithm and that it only contains letters in the *canonical alphabet* $A_\beta = \{0, \ldots, \lfloor\beta\rfloor\}$. Also observe that, for all $x, y \in [0, 1]$, we have $x < y \Leftrightarrow d_\beta(x) < d_\beta(y)$. The set $\mathrm{Fact}(D_\beta)$ is the set of factors occurring in the β-expansions of the real numbers in $[0, 1)$. If $d_\beta(1) = t_1 \cdots t_m 0^\omega$, with $t_1, \ldots, t_m \in A_\beta$ and $t_m \neq 0$, then we say that $d_\beta(1)$ is *finite* and we set $d_\beta^*(1) = (t_1 \cdots t_{m-1}(t_m - 1))^\omega$. Otherwise, we set $d_\beta^*(1) = d_\beta(1)$. If $d_\beta^*(1)$ is ultimately periodic, then β is said to be a *Parry number*.

Theorem 2 (Bertrand [5])**.** *Let $U = (U_n)_{n\geq 0}$ be a numeration system. There exists a real number $\beta > 1$ such that $0^* \mathrm{rep}_U(\mathbb{N}) = \mathrm{Fact}(D_\beta)$ if and only if U is a Bertrand numeration system. In that case, if $d_\beta^*(1) = (t_i)_{i\geq 1}$, then*

$$U_n = t_1 U_{n-1} + \cdots + t_n U_0 + 1. \tag{3}$$

Note that if β is a Parry number, then (3) defines a linear recurrence sequence and β is a root of its characteristic polynomial.

Theorem 3 (Parry [23])**.** *A sequence $s = (s_i)_{i\geq 1}$ over $\mathbb{N}$ is the β-expansion of a real number in $[0, 1)$ if and only if $(s_{n+i})_{i\geq 1}$ is lexicographically less than $d_\beta^*(1)$ for all $n \in \mathbb{N}$.*

As a consequence of the previous two theorems, with any Parry number β is canonically associated a deterministic finite automaton $\mathcal{A}_\beta = (Q_\beta, q_{\beta,0}, F_\beta, A_\beta, \delta_\beta)$ accepting the language $\mathrm{Fact}(D_\beta)$. Let $d_\beta^*(1) = t_1 \cdots t_i (t_{i+1} \cdots t_{i+p})^\omega$ where $i \geq 0$ and $p \geq 1$ are the minimal preperiod and period respectively. The set of states of $\mathcal{A}_\beta$ is $Q_\beta = \{q_{\beta,0}, \dots, q_{\beta,i+p-1}\}$. All states are final. For every $j \in \{1, \dots, i+p\}$, we have t_j edges $q_{\beta,j-1} \to q_{\beta,0}$ labeled by $0, \dots, t_j - 1$ and, for $j < i+p$, one edge $q_{\beta,j-1} \to q_{\beta,j}$ labeled by t_j. There is also an edge $q_{\beta,i+p-1} \to q_{\beta,i}$ labeled by t_{i+p}. See, for instance, [13, 15, 20]. Note that in [22, Thm. 7.2.13], $\mathcal{A}_\beta$ is shown to be the trim minimal automaton of $\mathrm{Fact}(D_\beta)$. A deterministic finite automaton is *trim* if it is accessible and coaccessible, i.e., any state can be reached from the initial state and from any state, a final state can be reached.

Example 1. Let β be the dominant root of the polynomial $X^3 - 2X^2 - 1$. We have $d_\beta(1) = 2010^\omega$ and $d^*_\beta(1) = (200)^\omega$. The automaton $\mathcal{A}_\beta$ is depicted in Figure 1.

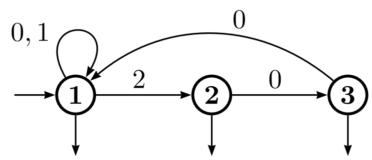

FIGURE 1. The automaton $\mathcal{A}_\beta$ for $d^*_\beta(1) = (200)^\omega$.

Definition 3. Let U be a linear numeration system. If $\lim_{n\to+\infty} U_{n+1}/U_n = \beta$ for some real $\beta > 1$, then U is said to *satisfy the dominant root condition* and β is called the *dominant root* of the recurrence.

Remark 1. If U is a linear numeration system satisfying the dominant root condition and if $\mathrm{rep}_U(\mathbb{N})$ is regular, then the dominant root β is a Parry number [16].

In the case where U has a dominant root $\beta > 1$, some connections between $\mathcal{A}_U$ and $\mathcal{A}_\beta$ have been previously explored by several authors [15, 20, 22]. Our aim in this paper is to provide a more comprehensive analysis of the relationship between these two automata.

Recall [12] that the states of the minimal automaton of an arbitrary language L over an alphabet A are given by the equivalence classes of the Myhill-Nerode congruence $\sim_L$, which is defined by

$$\forall w, z \in A^*,\ w \sim_L z \Leftrightarrow \{x \in A^* \mid wx \in L\} = \{x \in A^* \mid zx \in L\}.$$

Equivalently, the states of the minimal automaton of L correspond to the sets $w^{-1}L = \{x \in A^* \mid wx \in L\}$. In this paper the symbol $\sim$ will be used to denote Myhill-Nerode congruences.

Remark 2. In Theorem 6 we will describe a map between a restriction of $\mathcal{A}_U$ and $\mathcal{A}_\beta$. Note that similar obervations have been considered in other contexts [13, 6]. For example, if U is the Bertrand numeration system associated with a Pisot number β, then for any U-recognizable set X of integers, there exist an automaton recognizing X and a morphism mapping this automaton onto $\mathcal{A}_U = \mathcal{A}_\beta$ [6].

3. EXAMPLES OF AUTOMATA $\mathcal{A}_U$

The first two examples present the well-known Fibonacci numeration system and its generalization to an ℓ-order recurrence relation. Note that in the first four examples, Examples 2 to 5, the automaton $\mathcal{A}_U$ is exactly an automaton of the kind $\mathcal{A}_\beta$.

Example 2 (Fibonacci numeration system). With $U_{n+2} = U_{n+1} + U_n$ and $U_0 = 1$, $U_1 = 2$, we get the usual Fibonacci numeration system associated with the Golden Ratio. The dominant root is $\beta = (1+\sqrt{5})/2$. For this system, $A_U = \{0, 1\}$ and $\mathcal{A}_U$ accepts all words over A_U except those containing the factor 11. Moreover, we have $d_\beta(1) = 110^\omega$ and $d^*_\beta(1) = (10)^\omega$.

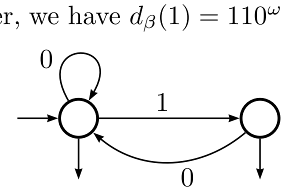

FIGURE 2. The automaton $\mathcal{A}_U$ for the Fibonacci numeration system.

Example 3 (ℓ-bonacci numeration system)**.** Let $\ell \geq 2$. Consider the linear recurrence sequence defined by

$$\forall n \in \mathbb{N},\ U_{n+\ell} = \sum_{i=0}^{\ell-1} U_{n+i}$$

and for $i \in \{0, \dots, \ell-1\}$, $U_i = 2^i$. For this system, $A_U = \{0, 1\}$ and $\mathcal{A}_U$ accepts all words over A_U except those containing the factor 1^ℓ. We have $d_\beta(1) = 1^\ell 0^\omega$ and $d^*_\beta(1) = (1^{\ell-1}0)^\omega$.

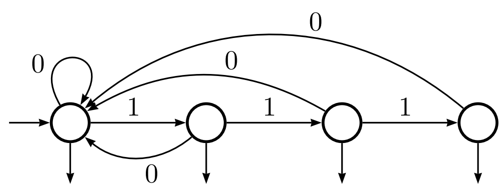

FIGURE 3. The automaton $\mathcal{A}_U$ for the 4-bonacci numeration system.

The third example is also classical. Compared to the previous examples where the β-expansions of the real numbers in $[0, 1)$ avoid a single factor, here the β-expansions avoid factors in an infinite regular language.

Example 4 (Square of the Golden Ratio)**.** With $U_{n+2} = 3U_{n+1} - U_n$, $U_0 = 1$ and $U_1 = 3$, we get the Bertrand numeration system associated with $\beta = (3 + \sqrt{5})/2$ (the square of the Golden Ratio). We have $A_U = \{0, 1, 2\}$ and 21^*2 is the set of minimal forbidden factors. Moreover $d_\beta(1) = d^*_\beta(1) = 21^\omega$.

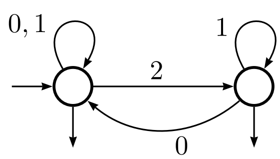

FIGURE 4. The automaton $\mathcal{A}_U$ for the Bertrand system associated with $(3 + \sqrt{5})/2$.

The recurrence involved in the following example will show some interesting properties and is related to Example 12.

Example 5. With $U_{n+2} = 2U_{n+1} + U_n$, $U_0 = 1$, $U_1 = 3$, we have the Bertrand numeration system

$$(U_n)_{n \geq 0} = 1, 3, 7, 17, 41, 99, 239, \dots$$

associated with $\beta = 1 + \sqrt{2}$. We have $d_\beta(1) = 210^\omega$ and $d^*_\beta(1) = (20)^\omega$. The corresponding automaton $\mathcal{A}_U$ is depicted in Figure 5.

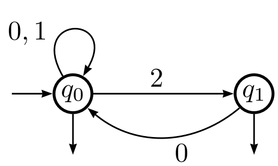

FIGURE 5. The automaton $\mathcal{A}_U$ for the Bertrand system associated with $1 + \sqrt{2}$.

The next example reveals some interesting properties and should be compared with the usual Fibonacci system. Observe that we have the same strongly connected component as for the Fibonacci system but the automaton in Figure 6 has one more state, from which only finitely many words may be accepted.

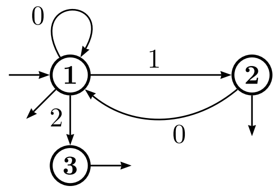

FIGURE 6. The automaton $\mathcal{A}_U$ for the modified Fibonacci system.

Example 6 (Modified Fibonacci system)**.** Consider the sequence $U = (U_n)_{n\ge 0}$ defined by the recurrence $U_{n+2} = U_{n+1} + U_n$ of Example 2 but with the initial conditions $U_0 = 1$, $U_1 = 3$. We get a numeration system $(U_n)_{n\ge 0} = 1, 3, 4, 7, 11, 18, 29, 47, \ldots$ which is no longer Bertrand. Indeed, 2 is a greedy representation but 20 is not because $\mathrm{rep}_U(\mathrm{val}_U(20)) = 102$. For this system, $A_U = \{0, 1, 2\}$ and $\mathcal{A}_U$ is depicted in Figure 6.

The following example illustrates the case where β is an integer.

Example 7. Consider the numeration system $U = (U_n)_{n\ge 0}$ defined by $U_{n+1} = 3U_n + 2$ and $U_0 = 1$. We have $A_U = \{0, 1, 2, 3, 4\}$. This system is linear and has the dominant root $\beta = 3$. We have $d_\beta(1) = 30^\omega$ and $d^*_\beta(1) = 2^\omega$. The automaton $\mathcal{A}_U$ is depicted in Figure 7.

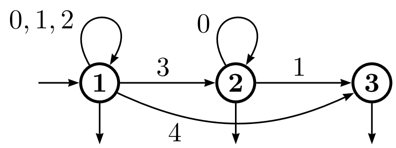

FIGURE 7. The automaton $\mathcal{A}_U$ for $U_{n+1} = 3U_n + 2$ and $U_0 = 1$.

As a prelude to Theorem 4, the next example shows that when the initial conditions are changed, the automaton $\mathcal{A}_U$ may have the same transition graph as the canonical automaton $\mathcal{A}_\beta$, but the set of final states may change.

Example 8. Consider the recurrence relation $U_{n+3} = 2U_{n+2} + U_n$. If we choose $(U_0, U_1, U_2) = (1, 3, 7)$, we get the Bertrand numeration system U such that $\mathcal{A}_U$ is exactly the automaton $\mathcal{A}_\beta$ from Example 1 depicted in Figure 1. If $(U_0, U_1, U_2) = (1, 2, 4)$, we get the same graph but only state **1** is final. If $(U_0, U_1, U_2) = (1, 2, 5)$, we get the same graph but only states **1** and **3** are final. Finally, with $(U_0, U_1, U_2) = (1, 3, 6)$, states **1** and **2** are final.

4. Structure of the Automaton $\mathcal{A}_U$

In this section we give a precise description of the automaton $\mathcal{A}_U$ when U is a linear numeration system satisfying the dominant root condition and such that $\mathrm{rep}_U(\mathbb{N})$ is regular.

Definition 4. A directed graph is *strongly connected* if for all pairs of vertices (s, t), there is a directed path from s to t. A *strongly connected component* of a directed graph is a maximal strongly connected subgraph. Such a component is said to be *non-trivial* if it does not consist of a single vertex with no loop.

For instance, state **3** in Figure 6 is not a non-trivial strongly connected component and state **2** in Figure 7 is a non-trivial strongly connected component.

Theorem 4. *Let U be a linear numeration system such that $\mathrm{rep}_U(\mathbb{N})$ is regular.*

(i) *The automaton $\mathcal{A}_U$ has a non-trivial strongly connected component $\mathcal{C}_U$ containing the initial state.*
(ii) *If p is a state in $\mathcal{C}_U$, then there exists $N \in \mathbb{N}$ such that $\delta_U(p, 0^n) = q_{U,0}$ for all $n \ge N$. In particular, if q (resp. r) is a state in $\mathcal{C}_U$ (resp. not in $\mathcal{C}_U$) and if $\delta_U(q, \sigma) = r$, then $\sigma \ne 0$.*
(iii) *If $\mathcal{C}_U$ is the only non-trivial strongly connected component of $\mathcal{A}_U$, then we have $\lim_{n\to+\infty} U_{n+1} - U_n = +\infty$.*

(iv) *If* $\lim_{n\to+\infty} U_{n+1} - U_n = +\infty$*, then the state* $\delta_U(q_{U,0}, 1)$ *belongs to* $\mathcal{C}_U$*.*

Proof. (i) The initial state $q_{U,0}$ has a loop with label 0 and therefore $\mathcal{A}_U$ has a non-trivial strongly connected component $\mathcal{C}_U$ containing $q_{U,0}$.

(ii) Let p be a state in $\mathcal{C}_U$. There exist $u, v \in A_U^*$ such that $\delta_U(q_{U,0}, u) = p$ and $\delta_U(p, v) = q_{U,0}$. We have

$$\forall x \in A_U^*,\ uvx \in 0^* \operatorname{rep}_U(\mathbb{N}) \Leftrightarrow u0^{|v|}x \in 0^* \operatorname{rep}_U(\mathbb{N}).$$

Indeed, if uvx is a greedy representation, so is $u0^{|v|}x$. Furthermore, if $u0^{|v|}x$ is a greedy representation, so is x, which must be accepted from $q_{U,0} = \delta_U(q_{U,0}, uv)$. Hence, uvx is a greedy representation. In other words, $uv \sim_{0^* \operatorname{rep}_U(\mathbb{N})} u0^{|v|}$ and $\delta_U(p, 0^{|v|}) = q_{U,0}$. Since $q_{U,0}$ has a loop labeled by 0, we obtain the desired result.

(iii) Assume that $\mathcal{A}_U$ has only one non-trivial strongly connected component $\mathcal{C}_U$. Since 10^n is a greedy representation for all n, infinitely many words are accepted from $\delta_U(q_{U,0}, 1)$, and so $\delta_U(q_{U,0}, 1)$ belongs to $\mathcal{C}_U$. From (ii), there exists a minimal $t \in \mathbb{N}$ such that $\delta_U(q_{U,0}, 10^t) = q_{U,0}$. Observe that U_n is the number of words of length n in $0^* \operatorname{rep}_U(\mathbb{N})$. For each word x (resp. y) in $0^* \operatorname{rep}_U(\mathbb{N})$ of length n (resp. $n - t$), the word $0x$ (resp. $10^t y$) has length $n+1$ and belongs to $0^* \operatorname{rep}_U(\mathbb{N})$. Therefore, we obtain $U_{n+1} \geq U_n + U_{n-t}$ for all $n \geq t$.

(iv) Assume that $\lim_{n\to+\infty} U_{n+1} - U_n = +\infty$. It is enough to show that there exists ℓ such that $\delta_U(q_{U,0}, 10^\ell) = q_{U,0}$. That is, we have to show that

$$\exists \ell \in \mathbb{N},\ \forall x \in A_U^*,\ 10^\ell x \in 0^* \operatorname{rep}_U(\mathbb{N}) \Leftrightarrow x \in 0^* \operatorname{rep}_U(\mathbb{N}).$$

Since we can always distinguish two states by a word of length at most $g = (\#\mathcal{A}_U)^2$, it is equivalent to show that

$$\exists \ell \in \mathbb{N},\ \forall x \in A_U^{\leq g},\ 10^\ell x \in 0^* \operatorname{rep}_U(\mathbb{N}) \Leftrightarrow x \in 0^* \operatorname{rep}_U(\mathbb{N}),$$

where $A_U^{\leq g}$ denotes the set of the words of length at most g over A_U. Since $U_{n+1} - U_n$ tends to $+\infty$, there exists ℓ such that for all $n \geq \ell$, we have $U_{n+1} - U_n > U_g - 1$, which shows that $10^\ell x$ is a greedy representation for any greedy representation x of length less than or equal to g. The other direction is immediate. □

The proof of the next result is mainly a consequence of the greediness of the involved representations.

Theorem 5. *Let U be a linear numeration system, having a dominant root $\beta > 1$, such that* $\operatorname{rep}_U(\mathbb{N})$ *is regular. Let x be a word over A_U such that infinitely many words are accepted from* $\delta_U(q_{U,0}, x)$*. Then $y0^\omega \leq d_\beta(1)$ for all suffixes y of x. Furthermore, the state $\delta_U(q_{U,0}, x)$ belongs to $\mathcal{C}_U$ if and only if $y0^\omega < d_\beta(1)$ for all suffixes y of x. In particular, in this case, the word x only contains letters in* $\{0, \ldots, \lceil\beta\rceil - 1\}$*.*

Remark 3. Let q be a state of $\mathcal{A}_U$ distinct from $q_{U,0}$. Since $\mathcal{A}_U$ is minimal, there exists a word w_q that *distinguishes* $q_{U,0}$ and q: that is, either w_q is accepted from $q_{U,0}$ and not from q, or w_q is accepted from q and not from $q_{U,0}$. Let us show that in the setting of numeration languages the second situation never occurs. Let x be such that $\delta_U(q_{U,0}, x) = q$. Assume that xw_q is accepted by $\mathcal{A}_U$. Then w_q is a greedy representation which must be accepted from $q_{U,0}$.

Theorem 6. *Let U be a linear numeration system, having a dominant root $\beta > 1$, such that* $\operatorname{rep}_U(\mathbb{N})$ *is regular. There exists a map $\Phi\colon \mathcal{C}_U \to Q_\beta$ such that $\Phi(q_{U,0}) = q_{\beta,0}$, and for all states q and all letters σ such that q and $\delta_U(q, \sigma)$ are states in $\mathcal{C}_U$, we have $\Phi(\delta_U(q, \sigma)) = \delta_\beta(\Phi(q), \sigma)$. Furthermore, if q is a state in $\mathcal{C}_U$ and σ is the maximal letter that can be read from $\Phi(q)$ in $\mathcal{A}_\beta$, then for any letter α in A_U, the state $\delta_U(q, \alpha)$ is in $\mathcal{C}_U$ if and only if $\alpha \leq \sigma$.*

Proof. Consider the automaton whose transition diagram is the subgraph induced by $\mathcal{C}_U$ and where all states are assumed to be final. From Theorems 2, 3 and 5, the language accepted by this automaton is exactly the same as the one accepted by $\mathcal{A}_\beta$. Note that $\mathcal{A}_\beta$ is a trim minimal automaton [22, Theorem 7.2.13]. From a classical result in automata theory (see, for instance, [12, Chap. 3, Thm. 5.2]), such a map Φ exists. □

Example 9. Consider the same recurrence relation as in Example 8 but with $(U_0, U_1, U_2) = (1, 5, 6)$. In Example 1 (see also Example 8), the automaton $\mathcal{A}_\beta$ with $d_\beta(1) = 2010^\omega$ and $\mathcal{A}_U$ had the same transition graph. Here we get a more complex situation described in Figure 8. The non-trivial strongly connected component $\mathcal{C}_U$ consists of the states $Q_U \setminus \{\mathbf{g}\}$. The map Φ is the map that sends the states $\mathbf{a}, \mathbf{b}, \mathbf{c}$ onto $\mathbf{1}$; the states $\mathbf{d}, \mathbf{e}$ onto $\mathbf{2}$; and the states $\mathbf{f}$ onto $\mathbf{3}$; where $\{\mathbf{1}, \mathbf{2}, \mathbf{3}\}$ is the set of states of the automaton $\mathcal{A}_\beta$ given in Figure 1.

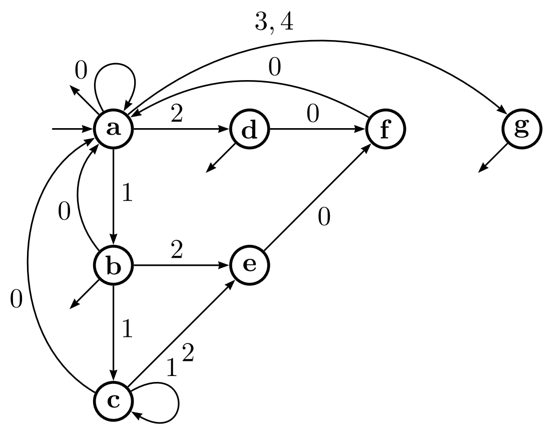

FIGURE 8. The automaton $\mathcal{A}_U$ for $(U_0, U_1, U_2) = (1, 5, 6)$.

Theorem 7. *Let U be a linear numeration system, having a dominant root $\beta > 1$, such that $\mathrm{rep}_U(\mathbb{N})$ is regular. If there exists a non-trivial strongly connected component distinct from $\mathcal{C}_U$, then $d_\beta(1)$ is finite. In this case, if s denotes the longest prefix of $d_\beta(1)$ which does not end with 0, then $\delta_U(q_{U,0}, u) \in \mathcal{C}_U$ for all proper prefixes u of s and $\delta_U(q_{U,0}, s) \notin \mathcal{C}_U$. In addition, if x is a word over A_U such that $\delta_U(q_{U,0}, x)$ is a state not in $\mathcal{C}_U$ leading to such a component, then there exists a word y over $\{0, \ldots, \lceil\beta\rceil - 1\}$ such that $\delta_U(q_{U,0}, y) \in \Phi^{-1}(q_{\beta,0})$ and $x = ys0^n$ for some n. In particular, the number of non-trivial strongly connected components distinct from $\mathcal{C}_U$ is bounded by $\#\Phi^{-1}(q_{\beta,|s|-1})$.*

Proof. Assume that there exists a non-trivial strongly connected component distinct from $\mathcal{C}_U$. Consider a state q not in $\mathcal{C}_U$ leading to such a component and a word u over A_U such that $\delta_U(q_{U,0}, u) = q$. Take the longest prefix x of u such that $\delta_U(q_{U,0}, x) \in \mathcal{C}_U$. Hence from Theorem 5 $x \in A_\beta^*$ and if $\sigma \in A_U$ and $v \in A_U^*$ are such that $u = x\sigma v$, then $\delta_U(q_{U,0}, x\sigma) \notin \mathcal{C}_U$. Using Theorem 5, there exists a suffix z of x such that $d_\beta(1) = z\sigma 0^\omega$, and so $d_\beta(1)$ is finite. The longest prefix of $d_\beta(1)$ which does not end with 0 is $s = z\sigma$. Furthermore, by Theorem 5 again, we see that v belongs to 0^*.

We still have to show that if $x = yz$, then $\delta_U(q_{U,0}, y)$ belongs to $\Phi^{-1}(q_{\beta,0})$, or equivalently in view of Theorem 6, $\delta_\beta(q_{\beta,0}, y) = q_{\beta,0}$. This is immediate by the definitions of $\mathcal{A}_\beta$ and $d_\beta(1)$. $\square$

Example 10. We give an illustration of the fact that if $\mathcal{A}_U$ contains more than one strongly connected component, then all components other than $\mathcal{C}_U$ consist of cycles labeled by 0. Here we are able to build a cycle with label 0^t for all $t \in \mathbb{N}$. Consider the sequence defined by $U_0 = 1$, $U_{tn+1} = 2U_{tn} + 1$ and $U_{tn+r} = 2U_{tn+r-1}$, for $1 < r \leq t$. This is a linear recurrence sequence and we get $0^* \mathrm{rep}_U(\mathbb{N}) = \{0, 1\}^* \cup \{0, 1\}^* 2(0^t)^*$.

Uniqueness of the non-trivial strongly connected component is discussed in the next two results.

Theorem 8. *Let U be a linear numeration system, having a dominant root $\beta > 1$, such that $\mathrm{rep}_U(\mathbb{N})$ is regular. If $U_{n+1}/U_n \to \beta^-$ as n tends to infinity, then the only non-trivial strongly connected component is $\mathcal{C}_U$.*

Theorem 9. *Let U be a linear numeration system, having a dominant root $\beta > 1$, such that $\mathrm{rep}_U(\mathbb{N})$ is regular. If the following conditions hold:*

(1) *$U_{n+1}/U_n \to \beta^+$, as n tends to infinity,*
(2) *there exists infinitely many n such that $U_{n+1}/U_n \neq \beta$, and*
(3) *$d_\beta(1)$ is finite,*

then $\mathcal{A}_U$ has more than one non-trivial strongly connected component. Note that, if $\beta \notin \mathbb{N}$, then (2) holds true.

Example 11. The numeration systems of Example 10 satisfy the hypotheses of the previous theorem and we have already shown that the corresponding automata have more than one non-trivial strongly connected component.

5. State complexity for divisibility criterion

We now turm to second issue of this paper. Namely we will study the state complexity of $0^* \operatorname{rep}_U(m\mathbb{N})$.

Definition 5. Let $U = (U_n)_{n\geq 0}$ be a numeration system and $m \geq 2$ be an integer. The sequence $(U_n \bmod m)_{n\geq 0}$ satisfies a linear recurrence relation of minimal length. This integer is denoted by $k_{U,m}$ or simply by k if the context is clear. This quantity is given by the largest t such that

$$\det H_t \not\equiv 0 \pmod m, \text{ where } H_t = \begin{pmatrix} U_0 & U_1 & \cdots & U_{t-1} \\ U_1 & U_2 & \cdots & U_t \\ \vdots & \vdots & \ddots & \vdots \\ U_{t-1} & U_t & \cdots & U_{2t-2} \end{pmatrix}.$$

Example 12. Let $m = 2$ and consider the sequence introduced in Example 5. The sequence $(U_n \bmod 2)_{n\geq 0}$ is constant and trivially satisfies the recurrence relation $U_{n+1} = U_n$ with $U_0 = 1$. Therefore, we get $k_{U,2} = 1$. For $m = 4$, one can check that $k_{U,4} = 2$.

Definition 6. Let $U = (U_n)_{n\geq 0}$ be a numeration system and $m \geq 2$ be an integer. Let $k = k_{U,m}$. Consider the system of linear equations

$$H_k\, \mathbf{x} \equiv \mathbf{b} \pmod m$$

where H_k is the $k \times k$ matrix given in Definition 5. We let $S_{U,m}$ denote the number of k-tuples $\mathbf{b}$ in $\{0, \ldots, m-1\}^k$ such that the system $H_k\, \mathbf{x} \equiv \mathbf{b} \pmod m$ has at least one solution $\mathbf{x}$.

Example 13. Again take the same recurrence relation as in Example 5 and $m = 4$. Consider the system

$$\begin{cases} 1\,x_1 + 3\,x_2 & \equiv & b_1 \pmod 4 \\ 3\,x_1 + 7\,x_2 & \equiv & b_2 \pmod 4 \end{cases}$$

We have $2x_1 \equiv b_2 - b_1 \pmod 4$. Hence for each value of b_1 in $\{0, \ldots, 3\}$, b_2 can take at most 2 values. One can therefore check that $S_{U,4} = 8$.

Remark 4. Let $\ell \geq k = k_{U,m}$. Then the number of ℓ-tuples $\mathbf{b}$ in $\{0, \ldots, m-1\}^\ell$ such that the system $H_\ell\, \mathbf{x} \equiv \mathbf{b} \pmod m$ has at least one solution equals $S_{U,m}$. Let us show this assertion for $\ell = k + 1$. Let H'_ℓ denote the $\ell \times k$ matrix obtained by deleting the last column of H_ℓ and let $\mathbf{x}'$ denote the k-tuple obtained by deleting the last element of $\mathbf{x}$. Observe that the ℓ-th column of H_ℓ is a linear combination of the other columns of H_ℓ. It follows that if $\mathbf{b} = (b_0, \ldots, b_{k-1}, b)^T \in \{0, \ldots, m-1\}^\ell$ is an ℓ-tuple for which the system $H'_\ell \mathbf{x}' \equiv \mathbf{b} \pmod m$ has a solution, then $\mathbf{b}' = (b_0, \ldots, b_{k-1})^T \in \{0, \ldots, m-1\}^k$ is a k-tuple for which the system $H_k \mathbf{x}' \equiv \mathbf{b}' \pmod m$ also has a solution. Furthermore, the ℓ-th row of H'_ℓ is a linear combination of the other rows of H'_ℓ, so for every such $\mathbf{b}'$, there is exactly one $\mathbf{b}$ such that $H'_\ell \mathbf{x}' \equiv \mathbf{b} \pmod m$ has a solution. This establishes the claim.

We define two properties that $\mathcal{A}_U$ may satisfy in order to get our results:

(H.1) $\mathcal{A}_U$ has a single strongly connected component denoted by $\mathcal{C}_U$,
(H.2) for all states p, q in $\mathcal{C}_U$, with $p \neq q$, there exists a word x_{pq} such that $\delta_U(p, x_{pq}) \in \mathcal{C}_U$ and $\delta_U(q, x_{pq}) \notin \mathcal{C}_U$, or, $\delta_U(p, x_{pq}) \notin \mathcal{C}_U$ and $\delta_U(q, x_{pq}) \in \mathcal{C}_U$.

Theorem 10. *Let $m \geq 2$ be an integer. Let $U = (U_n)_{n\geq 0}$ be a linear numeration system satisfying the recurrence relation* (2) *such that*

(a) $\mathbb{N}$ *is U-recognizable and $\mathcal{A}_U$ satisfies the assumptions* (H.1) *and* (H.2),

(b) $(U_n \bmod m)_{n\geq 0}$ *is purely periodic.*

Then the number of states of the trim minimal automaton $\mathcal{A}_{U,m}$ of the language

$$0^* \operatorname{rep}_U(m\mathbb{N})$$

from which infinitely many words are accepted is

$$(\#\mathcal{C}_U)S_{U,m}.$$

From now on we fix an integer $m \geq 2$ and a numeration system $U = (U_n)_{n\geq 0}$ satisfying the recurrence relation (2) and such that $\mathbb{N}$ is U-recognizable. Let $k = k_{U,m}$.

Definition 7. We define a relation $\equiv_{U,m}$ over A_U^*. For all $u, v \in A_U^*$,

$$u \equiv_{U,m} v \Leftrightarrow \begin{cases} u \sim_{0^* \operatorname{rep}_U(\mathbb{N})} v \quad \text{and} \\ \forall i \in \{0, \ldots, k-1\},\ \operatorname{val}_U(u0^i) \equiv \operatorname{val}_U(v0^i) \pmod m \end{cases}$$

where $\sim_{0^* \operatorname{rep}_U(\mathbb{N})}$ is the Myhill-Nerode equivalence for the language $0^* \operatorname{rep}_U(\mathbb{N})$ accepted by $\mathcal{A}_U$.

The proof of the main result of this part relies on the following technical proposition.

Proposition 1. *Assume that the numeration system U satisfies the assumptions of Theorem 10. Let $u, v \in A_U^*$ be such that $\delta_U(q_{U,0}, u)$ and $\delta_U(q_{U,0}, v)$ belong to $\mathcal{C}_U$. We have $u \equiv_{U,m} v$ if and only if $u \sim_{0^* \operatorname{rep}_U(m\mathbb{N})} v$.*

Proof of Theorem 10. If u is a word such that $\delta_U(q_{U,0}, u)$ belongs to $\mathcal{C}_U$, then there exist infinitely many words x such that $ux \in 0^* \operatorname{rep}_U(m\mathbb{N})$. On the other hand, by (H.1), if v is a word such that $\delta_U(q_{U,0}, v)$ does not belong to $\mathcal{C}_U$, there exist finitely many words x such that $vx \in 0^* \operatorname{rep}_U(m\mathbb{N})$. Therefore, the number of states of the trim minimal automaton of the language $0^* \operatorname{rep}_U(m\mathbb{N})$ from which infinitely many words are accepted is the number of sets $u^{-1}0^* \operatorname{rep}_U(m\,\mathbb{N})$ where u is a word over A_U such that $\delta_U(q_{U,0}, u)$ belongs to $\mathcal{C}_U$. Hence, as a consequence of Proposition 1, this number is also the number of equivalence classes $[u]_{\equiv_{U,m}}$ with u being such that $\delta_U(q_{U,0}, u) \in \mathcal{C}_U$. What we have to do to conclude the proof is therefore to count the number of such equivalence classes.

First we show that there are at most $\#\mathcal{C}_U S_{U,m}$ such classes. By definition, if $u, v \in A_U^*$ are such that $\delta_U(q_{U,0}, u) \neq \delta_U(q_{U,0}, v)$, then $u \not\equiv_{U,m} v$. Otherwise, $u \not\equiv_{U,m} v$ if and only if there exists $\ell < k$ such that $\operatorname{val}_U(u0^\ell) \not\equiv \operatorname{val}_U(v0^\ell) \pmod m$.

Let $u = u_{r-1} \cdots u_0 \in A_U^*$. We let $\mathbf{b}_u$ denote the k-tuple $(b_0, \ldots, b_{k-1})^T \in \{0, \ldots, m-1\}^k$ defined by

$$\forall s \in \{0, \ldots, k-1\},\ \operatorname{val}_U(u0^s) \equiv b_s \pmod m. \tag{4}$$

Using the fact that the sequence $(U_n)_{n\geq 0}$ satisfies (2), there exist $\alpha_0, \ldots, \alpha_{k-1}$ such that

$$\forall s \in \{0, \ldots, k-1\},\ \operatorname{val}_U(u0^s) = \sum_{i=0}^{r-1} u_i U_{i+s} = \sum_{i=0}^{k-1} \alpha_i U_{i+s}. \tag{5}$$

Using (4) and (5), we see that the system $H_k\mathbf{x} \equiv \mathbf{b}_u \pmod m$ has a solution $\mathbf{x} = (\alpha_0, \ldots, \alpha_{k-1})^T$.

If $u, v \in A_U^*$ are such that $\delta_U(q_{U,0}, u) = \delta_U(q_{U,0}, v)$ but $u \not\equiv_{U,m} v$, then $\mathbf{b}_u \neq \mathbf{b}_v$. From the previous paragraph the systems $H_k\mathbf{x} \equiv \mathbf{b}_u \pmod m$ and $H_k\mathbf{x} \equiv \mathbf{b}_v \pmod m$ both have a solution. Therefore, there are at most $\#\mathcal{C}_U S_{U,m}$ infinite equivalence classes.

Second we show that there are at least $\#\mathcal{C}_U S_{U,m}$ such classes. Let $\mathbf{c} = (c_0, \ldots, c_{k-1})^T \in \{0, \ldots, m-1\}^k$ be such that the system $H_k\mathbf{x} \equiv \mathbf{c} \pmod m$ has a solution $\mathbf{x_c} = (\alpha_0, \ldots, \alpha_{k-1})^T$. Let q be any state in $\mathcal{C}_U$. Our aim is to build a word y over A_U such that

$$\delta_U(q_{U,0}, y) = q \text{ and } \forall s \in \{0, \ldots, k-1\},\ \operatorname{val}_U(y0^s) \equiv c_s \pmod m.$$

Since $\mathcal{A}_U$ is accessible, there exists a word $u \in A_U^*$ such that $\delta_U(q_{U,0}, u) = q$. With this word u is associated a unique $\mathbf{b}_u = (b_0, \ldots, b_{k-1})^T \in \{0, \ldots, m-1\}^k$ given by (4). The system $H_k\mathbf{x} \equiv \mathbf{b}_u \pmod m$ has a solution denoted by $\mathbf{x}_u$.

Define $\gamma_0, \ldots, \gamma_{k-1} \in \{0, \ldots, m-1\}$ by $\mathbf{x_c} - \mathbf{x}_u \equiv (\gamma_0, \ldots, \gamma_{k-1})^T \pmod m$. Thus

$$H_k(\mathbf{x}_c - \mathbf{x}_u) \equiv \mathbf{c} - \mathbf{b}_u \pmod m. \tag{6}$$

Using properties (ii)–(iv) from Theorem 4 from the initial state $q_{U,0}$, there exist $t_{1,1}, \ldots, t_{1,\gamma_0}$ such that the word

$$w_1 = (0^{pt_{1,1}-1}1) \cdots (0^{pt_{1,\gamma_0}-1}1)$$

satisfies $\delta_U(q_{U,0}, w_1) \in \mathcal{C}_U \cap F_U$ and $\mathrm{val}_U(w_1) \equiv \gamma_0 U_0 \pmod m$. We can iterate this construction. For $j \in \{2, \ldots, k\}$, there exist $t_{j,1}, \ldots, t_{j,\gamma_j}$ such that the word

$$w_j = w_{j-1}(0^{pt_{j,1}-j}10^{j-1}) \cdots (0^{pt_{j,\gamma_j}-j}10^{j-1})$$

satisfies $\delta_U(q_{U,0}, w_j) \in \mathcal{C}_U \cap F_U$ and $\mathrm{val}_U(w_j) \equiv \mathrm{val}_U(w_{j-1}) + \gamma_{j-1}U_{j-1} \pmod m$. Consequently, we have

$$\mathrm{val}_U(w_k) \equiv \gamma_{k-1}U_{k-1} + \cdots + \gamma_0 U_0 \pmod m.$$

Now take r and r' large enough such that $\delta_U(q_{U,0}, w_k 0^{rp}) = q_{U,0}$ and $r'p \geq |u|$. Such an r exists by (ii) in Theorem 4. The word

$$y = w_k 0^{(r+r')p-|u|}u$$

is such that $\delta_U(q_{U,0}, y) = \delta_U(q_{U,0}, u) = q$ and taking into account the periodicity of $(U_n \bmod m)_{n\geq 0}$, we get

$$\mathrm{val}_U(y) \equiv \mathrm{val}_U(w_k) + \mathrm{val}_U(u) \pmod m.$$

In view of (6), we obtain

$$\forall s \in \{0, \ldots, k-1\},\ \mathrm{val}_U(y0^s) \equiv \sum_{i=0}^{k-1} \gamma_i U_{i+s} + b_s \equiv c_s - b_s + b_s = c_s \pmod m.$$

□

Corollary 1. *Assume that the numeration system U satisfies the assumptions of Theorem 10. Assume moreover that $\mathcal{A}_U$ is strongly connected (i.e. $\mathcal{A}_U = \mathcal{C}_U$). Then the number of states of the trim minimal automaton of the language $0^* \mathrm{rep}_U(m\mathbb{N})$ is $(\#\mathcal{C}_U)S_{U,m}$.*

Corollary 2. *Let $\ell \geq 2$. For the ℓ-bonacci numeration system $U = (U_n)_{n\geq 0}$ defined by $U_{n+\ell} = U_{n+\ell-1} + \cdots + U_n$ and $U_i = 2^i$ for all $i < \ell$, the number of states of the trim minimal automaton of the language $0^* \mathrm{rep}_U(m\mathbb{N})$ is ℓm^ℓ.*

To build the minimal automaton of $\mathrm{rep}_U(m\mathbb{N})$, one can use Theorem 1 to first have an automaton accepting the reversal of the words over A_U whose numerical value is divisible by m. We consider the reversal representations, that is least significant digit first, to be able to handle the period[1] of $(U_n \bmod m)_{n\geq 0}$. Such an automaton has m times the length of the period of $(U_n \bmod m)_{n\geq 0}$ states. Then minimizing the intersection of the reversal of this automaton with the automaton $\mathcal{A}_U$, we get the expected minimal automaton of $0^* \mathrm{rep}_U(m\mathbb{N})$.

Taking advantage of Proposition 1, we get an automatic procedure to obtain directly the minimal automaton $\mathcal{A}_{U,m}$ of $0^* \mathrm{rep}_U(m\mathbb{N})$. States of $\mathcal{A}_{U,m}$ are given by $(k+1)$-tuples. The state reached by reading w has as first component the state of $\mathcal{A}_U$ reached when reading w and the other components are $\mathrm{val}_U(w) \bmod m, \ldots, \mathrm{val}_U(w0^{k-1}) \bmod m$.

Example 14. Consider the Fibonacci numeration system and $m = 3$. The states of $\mathcal{A}_U$ depicted in Figure 2 are denoted by q_0 and q_1. The states of $\mathcal{A}_{U,3}$ are $r_0, \ldots, r_{17}$. The transition function of $\mathcal{A}_{U,3}$ is denoted by τ and is described in Table 1.

All the systems presented in Examples 2, 3 and 5 are Bertrand numeration systems. As a consequence of Parry's theorem [23, 22] and Bertrand's theorem [5, 22], the canonical automaton $\mathcal{A}_\beta$ associated with β-expansions is a trim minimal automaton (therefore, any two distinct states are distinguished) which is moreover strongly connected. The following result is therefore obvious.

[1] Another option is to consider a non-deterministic finite automaton reading most significant digits first.

w	$r = (\delta_U(q_0,w), \mathrm{val}_U(w), \mathrm{val}_U(w0))$	$\tau(r,0)$	$\tau(r,1)$
$\varepsilon, 0, 10^3 10$	$r_0 = (q_0, 0, 0)$	r_0	r_1
1	$r_1 = (q_1, 1, 2)$	r_2	
$10, 10100$	$r_2 = (q_0, 2, 0)$	r_3	r_4
100	$r_3 = (q_0, 0, 2)$	r_5	r_6
101	$r_4 = (q_1, 1, 1)$	r_7	
$1000, (10)^3$	$r_5 = (q_0, 2, 2)$	r_8	r_9
1001	$r_6 = (q_1, 0, 1)$	r_{10}	
$1010, (100)^2$	$r_7 = (q_0, 1, 2)$	r_2	r_{11}
$10^4, 10^4 10$	$r_8 = (q_0, 2, 1)$	r_{12}	r_{13}
$10^3 1$	$r_9 = (q_1, 0, 0)$	r_0	
$10010, 10^7$	$r_{10} = (q_0, 1, 1)$	r_7	r_{14}
10101	$r_{11} = (q_1, 0, 2)$	r_5	
10^5	$r_{12} = (q_0, 1, 0)$	r_{15}	r_{16}
$10^4 1$	$r_{13} = (q_1, 2, 2)$	r_8	
100101	$r_{14} = (q_1, 2, 1)$	r_{12}	
10^6	$r_{15} = (q_0, 0, 1)$	r_{10}	r_{17}
$10^5 1$	$r_{16} = (q_1, 1, 0)$	r_{15}	
$10^6 1$	$r_{17} = (q_1, 2, 0)$	r_3	

TABLE 1. The transition function of $\mathcal{A}_{U,3}$.

Proposition 2. *Let U be the Bertrand numeration system associated with a non-integer Parry number $\beta > 1$. The set $\mathbb{N}$ is U-recognizable and the trim minimal automaton $\mathcal{A}_U$ of $0^* \mathrm{rep}_U(\mathbb{N})$ fulfills properties* (H.1) *and* (H.2).

We can therefore apply Theorem 10 to the class of Bertrand numeration systems.

Finally, we give a lower bound when the numeration system satisfies weaker hypotheses than those of Theorem 10.

Proposition 3. *Let U be any numeration system (not necessarily linear). The number of states of $\mathcal{A}_{U,m}$ is at least $|\mathrm{rep}_U(m)|$.*

Proof. Let $n = |\mathrm{rep}_U(m)|$. For each $i \in \{1, \ldots, n\}$, we define p_i (resp. s_i) to be the prefix (resp. suffix) of length i (resp. $n-i$) of $\mathrm{rep}_U(m)$. We are going to prove that for all $i, j \in \{1, \ldots, n\}$, we have $p_i \not\sim_{0^* \mathrm{rep}_U(m\,\mathbb{N})} p_j$. Let $i, j \in \{1, \ldots, n\}$. We may assume that $i < j$. Obviously, the word $p_j s_j$ belongs to $0^* \mathrm{rep}_U(m\,\mathbb{N})$. On the other hand, observe that $|p_i s_j| \in \{1, \ldots, n-1\}$. Therefore the word $p_i s_j$ does not belong to $0^* \mathrm{rep}_U(m\,\mathbb{N})$ since it cannot simultaneously be greedy and satisfy $\mathrm{val}_U(p_i s_j) \equiv 0 \pmod{m}$. Hence, the word s_j distinguishes p_i and p_j. □

References

[1] B. Alexeev, Minimal DFA for testing divisibility, *J. Comput. Syst. Sci.* **69** (2004), 235–243.

[2] J.-P. Allouche, N. Rampersad, J. Shallit, Periodicity, repetitions, and orbits of an automatic sequence, *Theoret. Comput. Sci.* **410** (2009), 2795–2803.

[3] J. P. Bell, E. Charlier, A. S. Fraenkel, M. Rigo, A decision problem for ultimately periodic sets in non-standard numeration systems, *Int. J. Algebra and Computation* **19** (2009), 809–839.

[4] V. Berthé, M. Rigo, Eds., *Combinatorics, Automata and Number Theory*, Encyclopedia of Math. and its Applications, vol. **135**, Cambridge University Press (2010).

[5] A. Bertrand, Comment écrire les nombres entiers dans une base qui n'est pas entière, *Acta Math. Hungar.* **54** (1989), 237–241.

[6] V. Bruyère, G. Hansel, Bertrand numeration systems and recognizability, *Theoret. Comput. Sci.* **181** (1997), 17–43.

[7] V. Bruyère, G. Hansel, C. Michaux, R. Villemaire, Logic and p-recognizable sets of integers, *Bull. Belg. Math. Soc.* **1** (1994), 191–238.

[8] E. Charlier, N. Rampersad, M. Rigo, L. Waxweiler, Structure of the minimal automaton of a numeration language, *to appear in the proceedings of the 12th Descriptional Complexity of Formal Systems.*

[9] E. Charlier, N. Rampersad, M. Rigo, L. Waxweiler, State complexity of testing divisibility, *to appear in the proceedings of the 13th Mons Theoretical Computer Science Days*.
[10] E. Charlier, M. Rigo, A decision problem for ultimately periodic sets in non-standard numeration systems, *Lect. Notes in Comput. Sci.* **5162** (2008), Mathematical Foundations of Computer Science 2008, 241–252.
[11] A. Cobham, On the base-dependence of sets of numbers recognizable by finite automata, *Math. Systems Theory* **3** (1969) 186–192.
[12] S. Eilenberg, *Automata, languages, and machines*, Vol. A, Pure and Applied Mathematics, Vol. 58, Academic Press , New York (1974).
[13] S. Fabre, Substitutions et β-systèmes de numération, *Theoret. Comput. Sci.* **137** (1995), 219–236.
[14] Ch. Frougny, Representations of numbers and finite automata, *Math. Systems Theory* **25** (1992), 37–60.
[15] Ch. Frougny, B. Solomyak, On 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).
[16] M. Hollander, Greedy numeration systems and regularity, *Theory Comput. Systems* **31** (1998), 111–133.
[17] J. Honkala, A decision method for the recognizability of sets defined by number systems, *Theor. Inform. Appl.* **20** (1986), 395–403.
[18] D. Krieger, A. Miller, N. Rampersad, B. Ravikumar, J. Shallit, Decimations of languages and state complexity, *Theoret. Comput. Sci.* **410** (2009), 2401–2409.
[19] P. Lecomte, M. Rigo, Numerations systems on a regular language, *Theory Comput. Syst.* **34** (2001), 27–44.
[20] P. Lecomte, M. Rigo, Real numbers having ultimately periodic representations in abstract numeration systems, *Inform. and Comput.* **192** (2004), 57–83.
[21] N. Loraud, β-shift, systèmes de numération et automates. *J. Théor. Nombres Bordeaux* **7** (1995), 473–498.
[22] M. Lothaire, Algebraic Combinatorics on Words, Encyclopedia of Math. and its Applications, vol. **90**, Cambridge University Press (2002).
[23] W. Parry, On the β-expansions of real numbers, *Acta Math. Acad. Sci. Hungar.* **11** (1960), 401–416.
[24] J. Shallit, Numeration systems, linear recurrences, and regular sets, *Inform. and Comput.* **113** (1994), 331–347.

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