STRUCTURE OF THE MINIMAL AUTOMATON OF A NUMERATION LANGUAGE

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.

1. Introduction

In this paper, when i, j are integers satisfying $i \leq j$, we use the notation $[\![i, j]\!]$ to designate the interval of integers $\{i, i+1, \ldots, j-1, j\}$. Moreover, 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 $i \in [\![0, n-1]\!]$.

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 $[\![0, C_U - 1]\!]$. A greedy representation of a non-negative integer n is a word $w = w_{\ell-1} \cdots w_0$ over A_U satisfying

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

We denote by $\operatorname{rep}_U(n)$ the greedy representation of $n > 0$ satisfying $w_{\ell-1} \neq 0$. 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. 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 $\sum_{i=0}^{\ell-1} d_i U_i$.

From the point of view of Chomsky hierarchy, a U-recognizable set X of integers can be considered as having a low computational complexity: the elements belonging to $\operatorname{rep}_U(X)$ have simple syntactical properties recognized by some finite automaton. Since the seminal work of Alan Cobham [9] showing that the recognizability of a set depends on the numeration system under consideration, many properties of U-recognizable sets have been investigated, e.g., algebraic, logic or automatic characterizations of U-recognizable sets for integer base numeration systems [6], extensions of these characterizations to systems based on a Pisot number [5], study of the normalization map [11], introduction of abstract numeration systems [16],. . .

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. Conditions on a linear numeration system U for $\mathbb{N}$ to be U-recognizable are considered in [13].

Among linear numeration systems, the class of systems whose characteristic polynomial is the minimal polynomial of a Pisot number has been widely studied [5]. An example of such a system is given by the famous 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) [4, 12].

Currently little is known about the automata accepting other kind of numeration languages. The aim of this paper is to study the structure of these automata for a wide class of numeration systems. Our primary motivation was to understand the state complexity of languages of the form $0^* \operatorname{rep}_U(m\mathbb{N})$, that is, the language made up of the representations of the multiples of m in a given numeration system (see [1, 15]), in connection with the following decidability problem. Let U be a numeration system such that $\mathbb{N}$ is U-recognizable and let $X \subseteq \mathbb{N}$ be a U-recognizable set of integers given by some deterministic finite automaton recognizing $\operatorname{rep}_U(X)$. For integer base systems, Honkala has proved that one can decide whether or not X is ultimately periodic [14]. A shorter proof of this result was given in [2]. The same decidability question was answered positively in [8, 3] for a wide class of linear numeration systems containing the Fibonacci numeration system. Furthermore, in [7], as an application, for the Fibonacci numeration system F, we show that the number of states of the trim minimal automaton accepting $0^* \operatorname{rep}_F(m\mathbb{N})$ is $2m^2$.

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.

2. Background on Numeration Systems

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 $\operatorname{rep}_U(m)$ is genealogically less than $\operatorname{rep}_U(n)$.

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

Let us recall the theorems of Bertrand [4] (also see [18, Thm. 7.3.8]) and Parry [19] (also see [18, 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* $A_\beta = [\![0, \lfloor\beta\rfloor]\!]$. Also observe that, for all $x, y \in [0, 1]$, we have $x < y \Leftrightarrow d_\beta(x) < d_\beta(y)$. The set $\operatorname{Fact}(D_\beta)$ 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*.

Lemma 1. *Let $x = x_{k-1} \cdots x_0$ be a word over $\mathbb{N}$. We have*

$$\forall \ell \in [\![1, k]\!],\ x_{\ell-1} \cdots x_0 0^\omega \left\{ \begin{array}{c} < \\ \leq \end{array} \right. d_\beta(1) \quad \Leftrightarrow \quad \forall \ell \in [\![1, k]\!],\ \sum_{i=0}^{\ell-1} x_i \beta^{i-\ell} \left\{ \begin{array}{c} < \\ \leq \end{array} \right. 1.$$

Theorem 1 (Bertrand [4])**.** *Let $U = (U_n)_{n\geq 0}$ be a numeration system. There exists a real number $\beta > 1$ such that $0^* \operatorname{rep}_U(\mathbb{N}) = \operatorname{Fact}(D_\beta)$ if and only if U is a Bertrand numeration system. In that case, if $d^*_\beta(1) = (t_i)_{i\geq 1}$, then*

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

Note that if β is a Parry number, then (1) defines a linear recurrence sequence and β is a root of its characteristic polynomial.

Theorem 2 (Parry [19])**.** *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 $n \in \mathbb{N}$.*

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 $\operatorname{Fact}(D_\beta)$. 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 $Q_\beta = \{q_{\beta,0}, \ldots, q_{\beta,i+p-1}\}$. All states are final. For every $j \in [\![1, i + p]\!]$, we have t_j edges $q_{\beta,j-1} \to q_{\beta,0}$ labeled by $0, \ldots, t_j - 1$ 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, [12, 17]. Note that in [18, Thm. 7.2.13], $\mathcal{A}_\beta$ is shown to be the trim minimal automaton of $\operatorname{Fact}(D_\beta)$. 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.

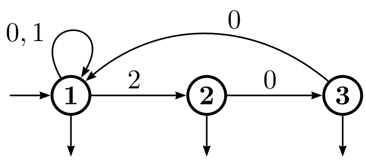

FIGURE 1. The automaton $\mathcal{A}_\beta$ for $d_\beta(1) = 2010^\omega$.

Definition 2. 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 $\operatorname{rep}_U(\mathbb{N})$ is regular, then the dominant root β is a Parry number [13].

Let $\mathcal{A}_U = (Q_U, q_{U,0}, F_U, A_U, \delta_U)$ be the trim minimal automaton of the language $0^* \operatorname{rep}_U(\mathbb{N})$ having $\#\mathcal{A}_U$ states. 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 [12, 17, 18]. Our aim in this paper is to provide a more comprehensive analysis of the relationship between these two automata.

Recall 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.

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

Example 2 presents the well-known Fibonacci numeration system. Note that in Examples 2 and 3, 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 $d^*_\beta(1) = (10)^\omega$.

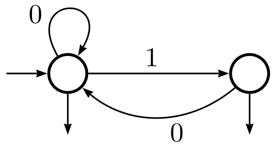

FIGURE 2. The automaton $\mathcal{A}_U$ for the Fibonacci numeration system.

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

Example 3 (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), which results in a sofic system (the set of forbidden factors is an infinite regular language). We have $A_U = \{0, 1, 2\}$ and 21^*2 is the set of forbidden factors. Moreover $d_\beta(1) = d^*_\beta(1) = 21^\omega$.

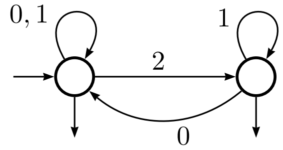

FIGURE 3. The automaton $\mathcal{A}_U$ for the Bertrand system associated with $(3 + \sqrt{5})/2$.

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 4 has one more state, from which only finitely many words may be accepted.

Example 4 (Modified Fibonacci system)**.** Consider the sequence $U = (U_n)_{n\geq 0}$ 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 $(U_n)_{n\geq 0} = 1, 3, 4, 7, 11, 18, 29, 47, \ldots$ 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 4.

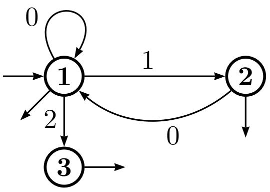

FIGURE 4. The automaton $\mathcal{A}_U$ for the modified Fibonacci system.

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

Example 5. Consider the numeration system $U = (U_n)_{n\geq 0}$ 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 5.

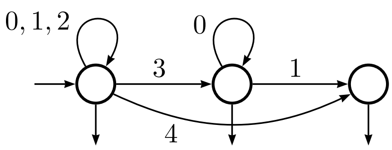

FIGURE 5. The automaton $\mathcal{A}_U$ for $U_{n+1} = 3U_n + 2$ and $U_0 = 1$.

As a prelude to Theorem 3, 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 6. 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 3. 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 4 is not a strongly connected component.

Theorem 3. *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 $n \geq N$. 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 $\sigma \neq 0$.*

(iii) *If $\mathcal{C}_U$ is the only non-trivial strongly connected component of $\mathcal{A}_U$, then we have $\lim\limits_{n\to+\infty} U_{n+1} - U_n = +\infty$.*

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

Example 7. This example illustrates property (iii) of Theorem 3. Let U be the Bertrand numeration system associated with a Parry number $\beta > 1$. From Theorems 1 and 2, we see that $\mathbb{N}$ is U-recognizable and $\mathcal{A}_U$ has only one non-trivial strongly connected component.

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

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

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, $uv \sim_{0^* \mathrm{rep}_U(\mathbb{N})} u0^{|v|}$ 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 $0^* \mathrm{rep}_U(\mathbb{N})$. For each word x (resp. y) in $0^* \mathrm{rep}_U(\mathbb{N})$ of length n (resp. $n - t$), the word $0x$ (resp. $10^t y$) has length $n + 1$ and belongs to $0^* \mathrm{rep}_U(\mathbb{N})$. Therefore, we obtain $U_{n+1} \geq U_n + U_{n-t}$ for all n.

(iv) Assume that $\lim\limits_{n\to+\infty} U_{n+1} - U_n = +\infty$. 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

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

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

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

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.

□

Theorem 4. *Let U be a linear numeration system having a dominant root $\beta > 1$ such that $\mathrm{rep}_U(\mathbb{N})$ is regular. Then we have the following.*

(i) *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 $[\![0, \lceil\beta\rceil - 1]\!]$.*

(ii) *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 $\alpha \leq \sigma$.*

(iii) *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 leading to such a component, then there exists a word y over $[\![0, \lceil\beta\rceil - 1]\!]$ 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 $\#\Phi^{-1}(q_{\beta,|s|-1})$.*

(iv) *If $U_{n+1}/U_n \to \beta^-$ as n tends to infinity, then the only non-trivial strongly connected component is $\mathcal{C}_U$.*

(v) *If the following conditions hold:*
   (v.1) *$U_{n+1}/U_n \to \beta^+$, as n tends to infinity,*
   (v.2) *there exists infinitely many n such that $U_{n+1}/U_n \neq \beta$, and*
   (v.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 (v.2) holds true.

Example 8. This example illustrates property (ii) of Theorem 4. Consider the same recurrence relation as in Example 6 but with $(U_0, U_1, U_2) = (1, 5, 6)$. In Example 6, 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 6. 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.

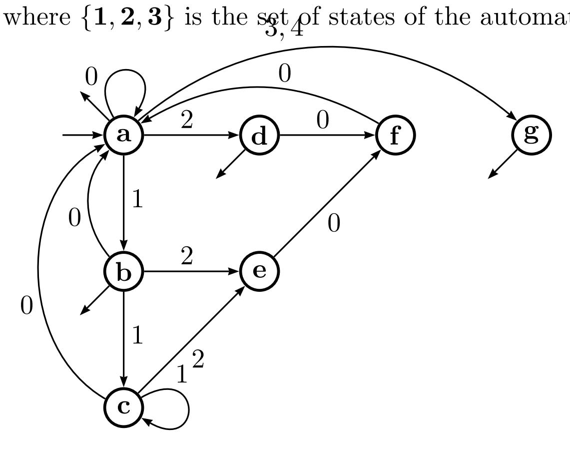

FIGURE 6. The automaton $\mathcal{A}_U$ for $(U_0, U_1, U_2) = (1, 5, 6)$.

Example 9. 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. This illustrates, in particular, properties (iii) and (v) of Theorem 4. 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 $0^* \operatorname{rep}_U(\mathbb{N}) = \{0,1\}^* \cup \{0,1\}^* 2(0^t)^*$.

Remark 2. 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 $q_{U,0}$.

Proof. (i) Let $x = x_{k-1} \cdots x_0$ be a word over A_U such that infinitely many words are accepted from $\delta_U(q_{U,0}, x)$. Due to the greediness of the representations, there exist infinitely many n such that $x0^n$ is a greedy representation. We obtain

$$\forall \ell \in [\![1, k]\!], \ \sum_{i=0}^{\ell-1} x_i U_{i+n} < U_{\ell+n}$$

for infinitely many n. Dividing by $U_{\ell+n}$ and letting n tend to infinity, we get

$$\forall \ell \in [\![1, k]\!], \quad \sum_{i=0}^{\ell-1} x_i \beta^{i-\ell} \leq 1.$$

Now assume that $\delta_U(q_{U,0}, x)$ belongs to $\mathcal{C}_U$. From (ii) and (iv), there exist $m, N \in \mathbb{N}$ such that for all $n \geq N$, we have $\delta_U(q_{U,0}, x0^m10^n) = q_{U,0}$, which is a final state. By the same reasoning as before, we obtain that

$$\forall \ell \in [\![1, k]\!], \quad \sum_{i=0}^{\ell-1} x_i \beta^{i-\ell} + \beta^{-\ell-m-1} \leq 1.$$

This implies that

$$\forall \ell \in [\![1, k]\!], \quad \sum_{i=0}^{\ell-1} x_i \beta^{i-\ell} < 1.$$

To show the other direction, now assume that $\delta_U(q_{U,0}, x)$ does not belong to $\mathcal{C}_U$. For all $n \in \mathbb{N}$, we have $\delta_U(q_{U,0}, x0^n) \neq q_{U,0}$. Therefore, by Remark 2, for all $n \in \mathbb{N}$, there exists a greedy representation $w^{(n)}$ of length at most $(\#\mathcal{A}_U)^2$ such that $x0^n w^{(n)}$ is not a greedy representation. Hence, by the pigeonhole principle, there exists a greedy representation w of length at most $(\#\mathcal{A}_U)^2$ such that for infinitely many n, the word $x0^n w$ is not a greedy representation. Therefore

$$\exists \ell \in [\![1, k]\!], \ \sum_{i=0}^{\ell-1} x_i U_{i+n+|w|} + \operatorname{val}_U(w) \geq U_{\ell+n+|w|}$$

for infinitely many n. We conclude that

$$\exists \ell \in [\![1, k]\!], \ \sum_{i=0}^{\ell-1} x_i \beta^{i-\ell} \geq 1.$$

Using Lemma 1, we obtain the desired result.

(ii) Consider the automaton whose transition diagram is the subgraph induced by $\mathcal{C}_U$ and where all states are assumed to be final. From (i) and Theorem 2, 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 [18, Theorem 7.2.13]. From a classical result in automata theory (see, for instance, [10, Chap. 3, Thm. 5.2]), such a map Φ exists.

(iii) 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 $x \in A_\beta^*$ 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 (i), 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 (i) 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 (ii), $\delta_U(q_{\beta,0}, y) = q_{\beta,0}$. This is immediate by the definitions of $\mathcal{A}_\beta$ and $d_\beta(1)$.

(iv) Suppose that $U_{n+1}/U_n \to \beta^-$ but $\mathcal{A}_U$ has more than one non-trivial strongly connected component. Let $x = x_{k-1} \cdots x_0$ be a word such that $\delta(q_{U,0}, x)$ is not in C_U and such that there exists an infinite sequence $j_1 < j_2 < \cdots$ such that for all $n \geq 1$, the word $x0^{j_n}$ is a greedy representation. Thus for all $\ell \in [\![1, k]\!]$,

$$\forall n \geq 1, \ \sum_{i=0}^{\ell-1} x_i \frac{U_{i+j_n}}{U_{\ell+j_n}} < 1. \tag{2}$$

Since $U_{n+1}/U_n \to \beta^-$ and by (i), we see that

$$\sum_{i=0}^{\ell-1} x_i \frac{U_{i+j_n}}{U_{\ell+j_n}} \longrightarrow \left(\sum_{i=0}^{\ell-1} x_i \beta^{i-\ell}\right)^+ = 1^+ \quad \text{as } n \to +\infty,$$

which is not possible in view of (2).

(v) Let $d_\beta(1) = s0^\omega$, where $s = s_{k-1} \cdots s_0$ is a word over A_β. In view of (iii), to show that there is a second strongly connected component, it suffices to show that for infinitely many n the words $s0^n$ are greedy representations. Equivalently, it suffices to show that for infinitely many n, we have

$$\forall \ell \in [\![1, k]\!], \sum_{i=0}^{\ell-1} s_i \frac{U_{i+n}}{U_{\ell+n}} < 1. \tag{3}$$

Let $\ell \in [\![1, k]\!]$. We have

$$\beta^{\ell-k} \sum_{i=0}^{\ell-1} s_i \beta^{i-\ell} = \sum_{i=0}^{\ell-1} s_i \beta^{i-k} = \sum_{i=0}^{k-1} s_i \beta^{i-k} - \sum_{i=\ell}^{k-1} s_i \beta^{i-k} \leq \beta^{\ell-k}.$$

Applying the hypotheses (v.1) and (v.2), we obtain (3), as required. □

5. Perspectives and Conjectures

(1) We use the same notations as in Theorem 3. In the case where the numeration system U has a dominant root $\beta > 1$, if $d_\beta(1)$ is finite, then $d_\beta^*(1) = (t_1 \cdots t_{m-1}(t_m - 1))^\omega$ where $t_m \neq 0$ and then we clearly have $\#\Phi^{-1}(q_{\beta,i}) \geq \#\Phi^{-1}(q_{\beta,i+1})$ for all $i \in [\![0, m-2]\!]$. We conjecture that, in this case, $\#\Phi^{-1}(q_{\beta,m-1}) = 1$. In other words, we conjecture that, in this case, $\mathcal{A}_U$ has at most two non-trivial strongly connected components.

(2) When the numeration system U does not satisfy the dominant root condition, we have not provided a precise description of $\mathcal{A}_U$. In this case, new kinds of phenomena may appear. For instance, in the following two examples, there exist more than one non-trivial strongly connected components containing transitions not labeled by 0. Furthermore, thanks to the first example, we see that $\mathcal{A}_U$ may have more than two non-trivial strongly connected components.

Example 10. Consider the numeration system $(U_n)_{n\geq 0}$ defined by $U_{n+3} = 24U_n$ and $(U_0, U_1, U_2) = (1, 2, 6)$. The corresponding trim minimal automaton is depicted in Figure 7. States in the same strongly connected component have the same label: **1**, **2** and **3**, respectively.

Example 11. Consider the numeration system $(U_n)_{n\geq 0}$ defined by $U_{n+4} = 3U_{n+2} + U_n$ and $(U_0, U_1, U_2, U_3) = (1, 2, 3, 7)$. The corresponding trim minimal automaton is depicted in Figure 8. Again, states in the same strongly connected component have the same label: **1** and **2**, respectively. Even if the sequence U_{n+1}/U_n does not converge, we have $\lim_{n\to+\infty} U_{2n+2}/U_{2n} = \lim_{n\to+\infty} U_{2n+3}/U_{2n+1} = (3 + \sqrt{13})/2$. Note that the latter observation is consistent with Hollander's conjecture [13].

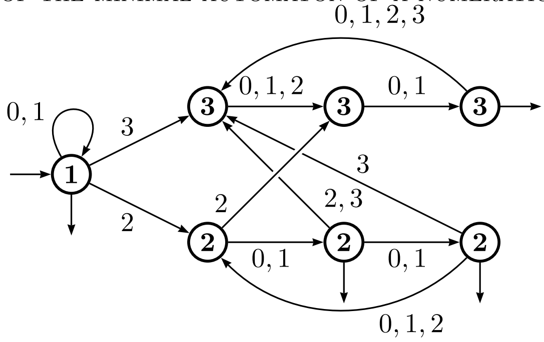

FIGURE 7. An automaton $\mathcal{A}_U$ for a numeration system $U = (U_n)_{n\geq 0}$ not satisfying the dominant root condition.

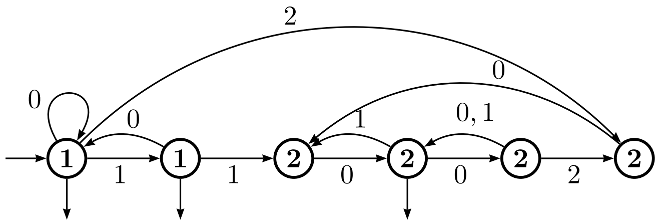

FIGURE 8. An automaton $\mathcal{A}_U$ for a numeration system $U = (U_n)_{n\geq 0}$ not satisfying the dominant root condition.

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

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

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

[7] E. Charlier, N. Rampersad, M. Rigo, L. Waxweiler, State complexity of testing divisibility. Work in progress.

[8] 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.

[9] A. Cobham, On the base-dependence of sets of numbers recognizable by finite automata, *Math. Systems Theory* **3** (1969), 186–192.

[10] S. Eilenberg, *Automata, languages, and machines*, Vol. A, Pure and Applied Mathematics, Vol. 58, Academic Press , New York (1974).

[11] Ch. Frougny, Representations of numbers and finite automata, *Math. Systems Theory* **25** (1992), 37–60.

[12] 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).

[13] M. Hollander, Greedy numeration systems and regularity, *Theory Comput. Systems* **31** (1998), 111–133.

[14] J. Honkala, A decision method for the recognizability of sets defined by number systems, *Theor. Inform. Appl.* **20** (1986), 395–403.

[15] D. Krieger, A. Miller, N. Rampersad, B. Ravikumar, J. Shallit, Decimations of languages and state complexity, *Theoret. Comput. Sci.* **410** (2009), 2401–2409.

[16] P. Lecomte and M. Rigo, *Numeration systems on a regular language*, Theory Comput. Syst. **34** (2001), 27–44.

[17] P. Lecomte, M. Rigo, Real numbers having ultimately periodic representations in abstract numeration systems, *Inform. and Comput.* **192** (2004), 57–83.

[18] M. Lothaire, Algebraic Combinatorics on Words, Encyclopedia of Math. and its Applications, vol. **90**, Cambridge University Press (2002).

[19] W. Parry, On the β-expansions of real numbers, *Acta Math. Acad. Sci. Hungar.* **11** (1960), 401–416.

INSTITUTE OF MATHEMATICS, UNIVERSITY OF LIÈGE,, GRANDE TRAVERSE 12 (B 37), B-4000 LIÈGE, BELGIUM, `{echarlier,nrampersad,M.Rigo,L.Waxweiler}@ulg.ac.be`