DCFS 2010 (Draft)

Descriptional Complexity of Formal Systems
Deadline for submissions: April 25, 2010
Final versions: July 8, 2010

State complexity of testing divisibility

Emilie Charlier[(A)] Narad Rampersad[(A)] Michel Rigo[(A)]
Laurent Waxweiler[(A)]

[(A)]Department of Mathematics – University of Liège
Grande traverse 12 (B37) – B-4000 Liège – Belgium
echarlier@ulg.ac.be (E. Charlier) nrampersad@ulg.ac.be (N. Rampersad)
M.Rigo@ulg.ac.be (M. Rigo) L.Waxweiler@ulg.ac.be (L. Waxweiler)

Abstract. Under some mild assumptions, we 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 $2m^2$.

Keywords: finite automata, state complexity, divisibility testing, linear recurrence relation.

1 Introduction

Cobham [9] 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 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 [19]. For linear numeration systems, ultimately periodic sets are all recognized by finite automata if and only if $\mathbb{N}$ is (see Theorem 2 below). Conditions on a linear numeration system U for $\mathbb{N}$ to be U-recognizable are considered in [12]. 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 4).

Let U be a linear numeration system and X be a U-recognizable set of non-negative integers given by some DFA recognizing the greedy representations of elements of X. For integer base systems, Honkala has proved that one can decide whether or not X is ultimately periodic [13]. Another, shorter proof of this result can be found in [2]. For a wide class of linear numeration systems, the same decidability question is answered positively in [8, 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 [15] 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.* [14] 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 an exact 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 $m \geq 1$.

In this paper, we study the state complexity for the divisibility criterion by $m \geq 2$ in the framework of linear numeration systems. Let $0^* \operatorname{rep}_U(m\mathbb{N})$ be the language of greedy representations of the multiples of $m \geq 1$ in the numeration system U. Under some mild assumptions, Theorem 14 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 19. 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].

Our result can only be fully understood when one has a clear picture of $\mathcal{A}_U$, the trim minimal automaton recognizing the set $0^* \operatorname{rep}_U(\mathbb{N})$ of all greedy representations. Such a description for a linear numeration system satisfying the dominant root condition (see below) is partially recalled in Theorem 8 [7].

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

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

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. 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 $\sum_{i=0}^{\ell-1} d_i U_i$.

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

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

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

Theorem 2 [4]. *Let $p, r \geq 0$. If $U = (U_n)_{n\geq 0}$ is a linear numeration system, then*

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

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

Definition 3. If $U = (U_n)_{n\geq 0}$ is a linear numeration system satisfying

$$\lim_{n\to+\infty} \frac{U_{n+1}}{U_n} = \beta$$

for some real $\beta > 1$, then it is said to *satisfy the dominant root condition* and β is called the *dominant root* of the recurrence.

Example 4 (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. The Golden Ratio $(1 + \sqrt{5})/2$ is the dominant root. For this system, $A_U = \{0, 1\}$ and $\mathcal{A}_U$ accepts all words over A_U except those containing the factor 11.

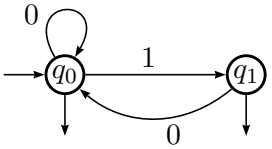

Figure 1. The automaton $\mathcal{A}_U$ for the Fibonacci numeration system.

Example 5 (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 $i \in \{0, \ldots, \ell - 1\}$, $U_i = 2^i$. For this system, $A_U = \{0, 1\}$ and $\mathcal{A}_U$ accepts all words over A_U except those containing the factor 1^ℓ.

Example 6. With $U_{n+2} = 2U_{n+1} + U_n$, $U_0 = 1$, $U_1 = 3$, we have a numeration system

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

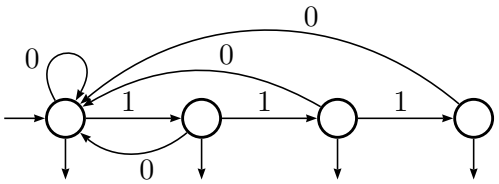

Figure 2. The automaton $\mathcal{A}_U$ for the 4-bonacci numeration system.

having a dominant root $\beta = 1 + \sqrt{2}$ and where the corresponding automaton $\mathcal{A}_U$ is depicted in Figure 3.

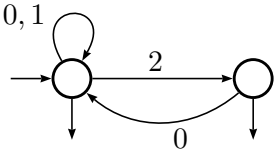

Figure 3. The automaton $\mathcal{A}_U$ for the system having dominant root $1 + \sqrt{2}$.

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.

Definition 7. A directed multi-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 multi-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.

Theorem 8. *[7] 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 $\lim_{n \to +\infty} U_{n+1} - U_n = +\infty$.*

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

In the case where the numeration system U has a dominant root $\beta > 1$, if the automaton $\mathcal{A}_U$ has more than one non-trivial strongly connected component, then any such component distinct from $\mathcal{C}_U$ is restricted to a cycle all of whose edges are labelled 0.

3 State complexity for divisibility criterion

Definition 9. 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 10. Let $m = 2$ and consider the sequence introduced in Example 6. 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 11. 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 9. We let $S_{U,m}$ denote the number of k-tuples $\mathbf{b}$ in $\{0, \ldots, m-1\}^k$ such that the system $H_k \mathbf{x} \equiv \mathbf{b} \bmod m$ has at least one solution.

Example 12. Again take the same recurrence relation as in Example 6 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 13. 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 $\mathbf{b} = (b_0, \ldots, b_{k-1}, b) \in \{0, \ldots, m-1\}^\ell$ is an ℓ-tuple for which the system $H'_\ell \mathbf{x}' \equiv \mathbf{b} \pmod m$ has a solution, then $\mathbf{b}' = (b_0, \ldots, b_{k-1}) \in \{0, \ldots, m-1\}^k$ 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:

(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 14. *Let $m \geq 2$ be an integer. Let $U = (U_n)_{n\geq 0}$ be a linear numeration system satisfying the recurrence relation* (1) *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 (1) and such that $\mathbb{N}$ is U-recognizable.

Definition 15. We define a relation $\equiv_{U,m}$ over A_U^*. For all $u, v \in A_U^*$,

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

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

Lemma 16. *Let $u, v, x \in A_U^*$. If $u \equiv_{U,m} v$ and $ux, vx \in \operatorname{rep}_U(\mathbb{N})$, then $ux \equiv_{U,m} vx$ and in particular, $\operatorname{val}_U(ux) \equiv \operatorname{val}_U(vx) \pmod m$.*

Proof: By assumption, for all $i \in \{0, \ldots, k-1\}$, $\operatorname{val}_U(u0^i) \equiv \operatorname{val}_U(v0^i) \pmod m$. Hence, for all $i \in \{0, \ldots, k-1\}$, $a_i \operatorname{val}_U(u0^i) \equiv a_i \operatorname{val}_U(v0^i) \pmod m$. Assume that $u = u_{\ell-1} \cdots u_0$. Note that

$$\sum_{i=0}^{k-1} a_i \operatorname{val}_U(u0^i) = \sum_{j=0}^{\ell-1} u_j \sum_{i=0}^{k-1} a_i U_{j+i} = \sum_{j=0}^{\ell-1} u_j U_{j+k} = \operatorname{val}_U(u0^k).$$

Therefore, we can conclude that $\operatorname{val}_U(u0^k) \equiv \operatorname{val}_U(v0^k) \pmod m$. Iterating this argument, we have, for all $n \geq 0$,

$$\operatorname{val}_U(u0^n) \equiv \operatorname{val}_U(v0^n) \pmod m. \tag{2}$$

Since the Myhill-Nerode relation is a right congruence, we have that

$$ux \sim_{0^* \operatorname{rep}_U(\mathbb{N})} vx.$$

Let $i \in \{0, \ldots, k-1\}$. From (2), we deduce that

$$\mathrm{val}_U(u0^{|x|+i}) + \mathrm{val}_U(x0^i) \equiv \mathrm{val}_U(v0^{|x|+i}) + \mathrm{val}_U(x0^i) \pmod{m}$$

and therefore $\mathrm{val}_U(ux0^i) \equiv \mathrm{val}_U(vx0^i) \pmod{m}$. □

Proposition 17. *Assume that the numeration system U satisfies the assumptions of Theorem 14. 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 $u \sim_{0^* \mathrm{rep}_U(m\mathbb{N})} v$.*

Proof: From (b) the sequence $(U_n \bmod m)_{n\geq 0}$ is purely periodic, say of period p.

Assume that $u \not\equiv_{U,m} v$. Our aim is to show that there exists a word $y \in A_U^*$ that distinguishes u and v in the minimal automaton of $0^* \mathrm{rep}_U(m\mathbb{N})$, i.e., either $uy \in 0^* \mathrm{rep}_U(m\mathbb{N})$ and $vy \notin 0^* \mathrm{rep}_U(m\mathbb{N})$, or $uy \notin 0^* \mathrm{rep}_U(m\mathbb{N})$ and $vy \in 0^* \mathrm{rep}_U(m\mathbb{N})$.

As a first case, assume $u \not\sim_{0^* \mathrm{rep}_U(\mathbb{N})} v$. Since $\delta_U(q_{U,0}, u)$ and $\delta_U(q_{U,0}, v)$ both belong to $\mathcal{C}_U$, this means that $\delta_U(q_{U,0}, u)$ and $\delta_U(q_{U,0}, v)$ are two different states in $\mathcal{C}_U$. By (H.2), without loss of generality, we may assume that there exists a word x such that

$$\delta_U(q_{U,0}, ux) \in \mathcal{C}_U \quad \text{and} \quad \delta_U(q_{U,0}, vx) \notin \mathcal{C}_U.$$

Since $\mathcal{A}_U$ contains only one strongly connected component, only finitely many words may be accepted from $\delta_U(q_{U,0}, vx)$. Let T be the length of the longest word accepted from $\delta_U(q_{U,0}, vx)$. Let $i \in \{1, \ldots, m\}$ be such that $\mathrm{val}_U(ux) + i \equiv 0 \pmod{m}$. Using properties (ii)–(iv) from Theorem 8 i times and the fact that $\delta_U(q_{U,0}, 1)$ is finite, there exist $r_1, \ldots, r_i \geq 0$ such that the word

$$y = x(0^{r_1 p}0^{p-1}1)(0^{r_2 p}0^{p-1}1)\cdots(0^{r_i p}0^{p-1}1)$$

has a length larger than $T + |x|$ and is such that uy is a greedy representation. Moreover, due to the periodicity of $(U_n \bmod m)_{n\geq 0}$, we have $\mathrm{val}_U(uy) \equiv 0 \pmod{m}$ and therefore uy belongs to $0^* \mathrm{rep}_U(m\mathbb{N})$. Hence, the word y distinguishes u and v for the language $0^* \mathrm{rep}_U(m\mathbb{N})$.

Now assume that $u \sim_{0^* \mathrm{rep}_U(\mathbb{N})} v$ and there exists $j \in \{0, \ldots, k-1\}$ such that $\mathrm{val}_U(u0^j) \not\equiv \mathrm{val}_U(v0^j) \pmod{m}$. There exists $i < m$ such that $\mathrm{val}_U(u0^j) + i \equiv 0 \pmod{m}$ and $\mathrm{val}_U(v0^j) + i \not\equiv 0 \pmod{m}$. Using properties (ii)-(iv) from Theorem 8 there exist $s_1, \ldots, s_i \geq 0$ such that the word

$$y = (0^{s_1 p}0^{p-1}1)(0^{s_2 p}0^{p-1}1)\cdots(0^{s_i p}0^{p-1}1)$$

distinguishes u and v.

Consider the other implication and assume that $u \equiv_{U,m} v$. Let x be a word such that $ux \in 0^* \mathrm{rep}_U(m\mathbb{N})$. From Lemma 16, we only have to show that vx is a greedy representation. Since v is a greedy representation and $u \sim_{0^* \mathrm{rep}_U(\mathbb{N})} v$, we can conclude that vx is a greedy representation. Hence the conclusion follows. □

Proof: [Proof of Theorem 14] If u is a word such that $\delta_U(q_{U,0}, u)$ belongs to $\mathcal{C}_U$, then with the same reasoning as in the proof of Proposition 17, 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 17, 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

$$\forall s \in \{0, \ldots, k-1\},\ \operatorname{val}_U(u0^s) \equiv b_s \pmod m. \tag{3}$$

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

$$\forall s \in \{0, \ldots, k-1\},\ \operatorname{val}_U(u0^s) = \sum_{i=0}^{r-1} u_i U_{i+s} = \sum_{i=0}^{k-1} \alpha_i U_{i+s}. \tag{4}$$

Using (3) and (4), 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 (3). 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

$$H_k(\mathbf{x}_c - \mathbf{x}_u) \equiv \mathbf{c} - \mathbf{b_u} \pmod m. \tag{5}$$

Using properties (ii)–(iv) from Theorem 8 from the initial state $q_{U,0}$, there exist $t_{1,1},\dots,t_{1,\gamma_0}$ such that the word

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

satisfies $\delta_U(q_{U,0},w_1)\in\mathcal{C}_U\cap F_U$ and $\mathrm{val}_U(w_1)\equiv\gamma_0U_0 \pmod m$. We can iterate this construction. For $j\in\{2,\dots,k\}$, there exist $t_{j,1},\dots,t_{j,\gamma_j}$ such that the word

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

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

$$\mathrm{val}_U(w_k)\equiv\gamma_{k-1}U_{k-1}+\cdots+\gamma_0U_0 \pmod m.$$

Now take r and r' large enough such that $\delta_U(q_{U,0},w_k0^{rp})=q_{U,0}$ and $r'p\geq|u|$. Such an r exists by (ii) in Theorem 8. The word

$$y = w_k0^{(r+r')p-|u|}u$$

is such that $\delta_U(q_{U,0},y)=\delta_U(q_{U,0},u)=q$ and taking into account the periodicity of $(U_n \bmod m)_{n\geq0}$, we get

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

In view of (5), we obtain

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

□

Corollary 18. *Assume that the numeration system U satisfies the assumptions of Theorem 14. 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 $(\#\mathcal{C}_U)S_{U,m}$.*

Proof: We use the same argument as in the beginning of the proof of Theorem 14. Since $\mathcal{A}_U=\mathcal{C}_U$, all of the sets $u^{-1}0^*\,\mathrm{rep}_U(m\,\mathbb{N})$ are infinite. Hence, infinitely many words are accepted from any state of $\mathcal{A}_{U,m}$. □

Corollary 19. *Let $\ell\geq2$. For the ℓ-bonacci numeration system $U=(U_n)_{n\geq0}$ 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 $\ell.m^\ell$.*

Proof: First note that the trim minimal automaton of $0^*\,\mathrm{rep}_U(\mathbb{N})$ consists of a unique strongly connected component made of ℓ states (see Figure 1) and $\mathcal{A}_U$ satisfies all the required assumptions. The matrix $\mathbf{H}_\ell$ has a determinant equal to ±1. Therefore, for all $\mathbf{b}\in\{0,\dots,m-1\}^\ell$, the system $\mathbf{H}_\ell\mathbf{x}\equiv\mathbf{b} \pmod m$ has a solution. There are m^ℓ such vectors $\mathbf{b}$. We conclude by using Corollary 18. □

To build the minimal automaton of $\mathrm{rep}_U(m\mathbb{N})$, one can use Theorem 2 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 representation, 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 17, 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 20. Consider the Fibonacci numeration system and $m = 3$. The states of $\mathcal{A}_U$ depicted in Figure 1 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.

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	

Definition 21. A numeration system $U = (U_n)_{n\geq 0}$ is a *Bertrand numeration system* if, for all $w \in A_U^+$, $w \in \mathrm{rep}_U(\mathbb{N}) \Leftrightarrow w0 \in \mathrm{rep}_U(\mathbb{N})$.

All the systems presented in Examples 4, 5 and 6 are Bertrand numeration systems. As a consequence of Parry's Theorem [18, 17] and Bertrand's theorem [5, 17], the canonical automaton $\mathcal{A}_\beta$ associated with β-expansions is a trim minimal

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

automaton (therefore, any two distinct states are distinguished) which is moreover strongly connected. The following result is therefore obvious.

Proposition 22. *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 $0^* \operatorname{rep}_U(\mathbb{N})$ fulfills properties* (H.1) *and* (H.2).

We can therefore apply Theorem 14 to this class of Bertrand numeration systems.

Finally, we give a lower bound when the numeration system satisfies weaker hypotheses than those of Theorem 14.

Proposition 23. *Let U be any numeration system (not necessarily linear). The number of state of $\mathcal{A}_{U,m}$ is at least $|\operatorname{rep}_U(m)|$.*

Proof: Let $n = |\operatorname{rep}_U(m)|$. 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 $\operatorname{rep}_U(m)$. We are going to prove that for all $i, j \in \{1, \ldots, n\}$, we have $p_i \not\sim_{0^* \operatorname{rep}_U(m\,\mathbb{N})} p_j$. Let $i, j \in \{1, \ldots, n\}$. We may assume that $i < j$. Obviously, the word $p_j s_j$ belongs to $0^* \operatorname{rep}_U(m\,\mathbb{N})$. 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 $0^* \operatorname{rep}_U(m\,\mathbb{N})$ since it cannot simultaneously be greedy and satisfy $\operatorname{val}_U(p_i s_j) \equiv 0 \pmod{m}$. Hence, the word s_j distinguishes p_i and p_j. □

4 Perspectives

- With the same assumptions as in Theorem 14, can we count the number of states from which only finitely many words are accepted?
- Can we weaken the assumptions of Theorem 14?
- If X is a finite union of arithmetic progressions, can we give bounds for the number of states of the trim minimal automaton accepting $0^* \operatorname{rep}_U(X)$?

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] E. Charlier, N. Rampersad, M. Rigo, L. Waxweiler, Structure of the minimal automaton of a numeration language, *submitted for publication*.

[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, 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).

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

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

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

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

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

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

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

[19] J. Shallit, Numeration systems, linear recurrences, and regular sets, *Inform. and Comput.* **113** (1994), 331–347.