An efficient algorithm to decide periodicity of b-recognisable sets using MSDF convention

Bernard Boigelot, Isabelle Mainz, Victor Marsault*, and Michel Rigo

Montefiore Institute & Department of Mathematics, Université de Liège, Belgium
{bernard.boigelot, isabelle.mainz, victor.marsault, m.rigo}@ulg.ac.be

Abstract

Given an integer base $b > 1$, a set of integers is represented in base b by a language over $\{0, 1, ..., b-1\}$. The set is said to be b-recognisable if its representation is a regular language. It is known that eventually periodic sets are b-recognisable in every base b, and Cobham's theorem implies the converse: no other set is b-recognisable in every base b.

We are interested in deciding whether a b-recognisable set of integers (given as a finite automaton) is eventually periodic. Honkala showed that this problem decidable in 1986 and recent developments give efficient decision algorithms. However, they only work when the integers are written with the least significant digit first.

In this work, we consider the natural order of digits (Most Significant Digit First) and give a quasi-linear algorithm to solve the problem in this case.

Keywords and phrases integer-base systems; automata; recognisable sets; periodic sets.

Introduction

Let $b > 1$ be an integer base. We let $[\![b]\!] = \{0, 1, \ldots, b-1\}$ denote the canonical alphabet of base-b digits. If $u = u_\ell \cdots u_0$ belongs to $[\![b]\!]^*$, we let $\overline{u}$ denote the *value* of u in base b, *i.e.*, $\overline{u} = \sum_{i=0}^{\ell} u_i \, b^i$. Note that the leftmost digit is the most significant one. We let $\langle n \rangle$ denote the (shortest) *base-b representation* of n. We set $\langle 0 \rangle$ to be the empty word ε. If reference to the base b is needed, we write $\langle n \rangle_b$. Thus $\langle n \rangle$ is the unique word u over $[\![b]\!]$ not starting with 0 and such that $\overline{u} = n$. Moreover, for every $u \in [\![b]\!]^*$ such that $\overline{u} = n$, there exists $i \geq 0$ such that $u = 0^i \langle n \rangle$.

Our contribution

Let $b > 1$ be an integer base. In this paper, we develop an algorithm to decide whether a given deterministic automaton $\mathcal{A}$ over the alphabet $[\![b]\!]$ accepts, by value, an (eventually) periodic set of integers. More precisely, the question is to decide whether there exist integers $p \geq 1$ and $N \geq 0$ such that, for all words $u \in [\![b]\!]^*$, if $\overline{u} \geq N$, then u is accepted by $\mathcal{A}$ if and only if $\langle \overline{u} + p \rangle$ is accepted as well. *Acceptance by value* means that words sharing the same value are either all accepted or all rejected. Stated otherwise, a word u is accepted by $\mathcal{A}$ if and if only if $0u$ is accepted. The main result of this paper is the following one.

► **Theorem 1.** *Given an integer base $b > 1$ and a n-state deterministic automaton $\mathcal{A}$ over the alphabet $[\![b]\!]$, it is decidable in $O(bn \log n)$ time whether or not $\mathcal{A}$ accepts, by value, some eventually periodic set of integers.*

[*] Supported by a Marie Skłodowska-Curie fellowship, co-funded by the European Union.

Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

We stress the fact that the input automaton $\mathcal{A}$ reads words most significant digit first (MSDF). This is an important difference with other results discussed in the literature. For instance, an efficient algorithm to solve this decision problem is provided for automata reading least significant digits first (LSDF) [15]. One can therefore think that it is enough to take the reversal of $\mathcal{A}$ and thus consider entries LSDF. Nevertheless, the reversal of $\mathcal{A}$ has first to be determinised. This potentially leads to an exponential blow-up in the number of states and thus to an inefficient procedure.

Motivations and related results

We say that a set $X \subseteq \mathbb{N}$ is *b-recognisable* if $\langle X\rangle_b$ is accepted by some finite automaton. One reason why eventually periodic sets of integers play a special role comes from the celebrated theorem of Cobham about the dependence to the base of b-recognisability.

▶ **Theorem** (Cobham, [11]). *Let $b, c > 1$ be two multiplicatively independent integers. A set X of integers is such that the languages $\langle X\rangle_b$ and $\langle X\rangle_c$ are both accepted by finite automata if and only if X is eventually periodic.*

In combinatorics on words, when studying morphic words (for details and definitions, for instance, see [2, 5]), Cobham's theorem can be reformulated as follows. Let $b, c > 1$ be two multiplicatively independent integers. An infinite word $\mathbf{x}$ is both b-automatic and c-automatic if and only if $\mathbf{x}$ is of the form uv^ω where u, v are finite words. Indeed, a set of integers is b-recognisable if and only if its characteristic sequence is b-automatic. The decision problem considered in our Theorem 1 is well known to be decidable.

▶ **Theorem** (Honkala, [13]). *It is decidable whether or not a given b-automatic word is ultimately periodic.*

Complexity issues are however not all considered in Honkala's paper. The decidability of our problem of interest can also be obtained using a first-order logic characterization of b-recognisable sets given by Büchi's theorem, and the fact that Presburger arithmetic is decidable [9, 1]. These independent approaches all lead to decision procedures with exponential complexity.

Using LSDF convention, efficient decision procedures are known. First, Leroux obtained a quadratic decision procedure [14] for utimately-periodic b-recognisable sets of integers. Then, the result was improved as follows.

▶ **Theorem** (Marsault, Sakarovitch, [15]). *Given an integer base $b > 1$ and a n-state deterministic automaton $\mathcal{A}$ over the alphabet $[\![b]\!]$, it is decidable in $O(b\,n \log n)$ time whether or not $\mathcal{A}$ accepts, with LSDF convention, some eventually periodic set of integers.*

Leroux's result is stated in a multi-dimensional setting, *i.e.*, the problem is to decide whether or not a b-recognisable subset of $\mathbb{N}^d$ is semi-linear. In that direction, see [19, 17, 14].

Generalisation to real numbers

Real numbers can be encoded in a base $b > 1$ by extending positional encoding to infinite words: A word encoding a real is composed of a finite prefix corresponding to an integer part, followed by a single occurrence of a distinguished symbol acting as a separator, and an infinite suffix representing a fractional part. Infinite-word automata are then able to recognise sets of reals. It has been established that *weak deterministic automata*, a restricted class of infinite-word automata, are sufficiently expressive for recognising all sets definable in mixed integer and real first-order additive arithmetic [7].

The properties of sets of real numbers that can be recognised by weak deterministic automata in all bases $b > 1$ have been investigated [6]. Such sets generalise to the real domain the notion of eventual periodicity; they precisely correspond to finite combinations of eventually periodic sets of integers, and intervals of $[0, 1]$. Checking whether an automaton recognises such a set can be done by first splitting this automaton into finite-state machines operating on the integer and fractional parts of encodings. The former are then checked in the same way as for MSDF integer encodings, and the latter by verifying that they obey the simple structure documented in [6], which is a simple operation. As a consequence, the algorithm developed in this paper also leads to an efficient procedure for checking that a weak deterministic automaton recognises an eventually periodic set of reals.

Generalisation to other numeration systems

Automatic words form a particular class of morphic words. Similarly, integer-base systems are special cases of more general numeration systems such as those built on a linear recurrent sequence. One can define a *numeration system* as a one-to-one map s from $\mathbb{N}$ to a language L over a finite alphabet. The integer n is mapped to its representation $s(n)$ within the considered system. Hence, it is natural to ask, for given a numeration system s and a subset M of L accepted by a finite automaton $\mathcal{A}$, whether or not the s-recognisable set $s^{-1}(M) \subseteq \mathbb{N}$ is eventually periodic.

On the one hand, Honkala's result is extended as follows. It is decidable whether or not a given morphic word is ultimately periodic [12, 16]. On the other hand, Büchi's theorem can be extended to linear numeration systems whose characteristic polynomial is the minimal polynomial of a Pisot number. See, for details, [8]. In that setting, several decision problems in combinatorics on words, including the ultimate periodicity problem, are decidable [10]. Using Honkala's techniques, the decision problem considered in our Theorem 1 is generalized to a large class of numeration systems in [4]. In particular, there are systems in this class for which the logical setting may not be applied. For all these decidability results presented in a wider context, no efficient procedure is known.

1 Preliminaries

In this paper, we only consider deterministic accessible finite automata with an input alphabet of the form $[\![b]\!]$. We use the acceptance-by-value convention. Thus, we may assume that the initial state bears a loop with label 0. In particular, this will always be the case after minimisation. Let $\mathcal{A}$ be an automaton. Its set of states (resp. its initial state, its set of final states) is denoted by $Q_{\mathcal{A}}$ (resp. $i_{\mathcal{A}}$, $F_{\mathcal{A}}$). If the considered automaton is clear from the context, $(s \cdot u)$ is the state s' such that $s \xrightarrow{u} s'$. The language accepted by $\mathcal{A}$ is denoted by $L(\mathcal{A})$. In this section, we recap basic results about automata.

1.1 Automaton morphisms and pseudo-morphisms

▶ **Definition 2.** Given two (accessible) automata $\mathcal{A}$ and $\mathcal{M}$ over $[\![b]\!]$, an *automaton morphism* $\mathcal{A} \to \mathcal{M}$ is a function $\phi : Q_{\mathcal{A}} \to Q_{\mathcal{M}}$ that satisfies:

$$\phi(i_{\mathcal{A}}) = i_{\mathcal{M}} \tag{1}$$

$$\forall s \in Q_{\mathcal{A}},\ \forall a \in [\![b]\!] \quad (s \cdot a) \text{ exists in } \mathcal{A} \iff (\phi(s) \cdot a) \text{ exists in } \mathcal{M} \tag{2}$$

$$\forall s, s' \in Q_{\mathcal{A}},\ \forall a \in [\![b]\!] \quad s \xrightarrow{a} s' \text{ in } \mathcal{A} \implies \phi(s) \xrightarrow{a} \phi(s') \text{ in } \mathcal{M} \tag{3}$$

$$F_{\mathcal{A}} = \phi^{-1}(F_{\mathcal{M}}) \tag{4}$$

▶ **Definition 3.** If a function ϕ satisfies (1), (2) and (3) but not necessarily (4), then we say that we have an *automaton pseudo-morphism.*

▶ **Definition 4.** Two states s, s' of an automaton $\mathcal{A}$ are *Nerode-equivalent* if, for every word u, $(s \cdot u)$ exists and is final if and only if $(s' \cdot u)$ exists and is final.

The next result is classical. See, for instance, [18].

▶ **Theorem 5** (Myhill–Nerode). *Let $\mathcal{A}$ be a complete automaton. Among all the complete automata accepting $L(\mathcal{A})$, up to isomorphism, there exists a unique one with a minimal number of states, called the* minimisation of $\mathcal{A}$*. Moreover, if $\mathcal{M}$ denotes the minimisation of $\mathcal{A}$, then there exists an automaton morphism $\phi : \mathcal{A} \to \mathcal{M}$ (called the* minimisation morphism*) such that*

$$\forall s, s' \in \mathcal{A} \quad \phi(s) = \phi(s') \iff s \text{ and } s' \text{ are Nerode-equivalent.} \tag{5}$$

If $\mathcal{A}$ is an automaton and u is a word, we write $(\mathcal{A} \cdot u)$ as a shorthand for $(i_{\mathcal{A}} \cdot u)$, i.e., the state reached by the run of u in $\mathcal{A}$.

▶ **Lemma 6.** *Let $\mathcal{A}$ and $\mathcal{M}$ be two complete (and accessible) automata. There exists a pseudo-morphism $\mathcal{A} \to \mathcal{M}$ if and only if every pair of words u, u' such that $(\mathcal{M} \cdot u) \neq (\mathcal{M} \cdot u')$ also satisfies $(\mathcal{A} \cdot u) \neq (\mathcal{A} \cdot u')$.*

Proof. Forward direction. Since a pseudo-morphism ϕ respects transitions and the initial state, it follows that, for every word u, $(\mathcal{M} \cdot u) = \phi(\mathcal{A} \cdot u)$. The statement follows immediately.

Backward direction. For every state s, we choose a word u_s such that $(\mathcal{A} \cdot u_s) = s$ (such a word exists because $\mathcal{A}$ is accessible). We define a function $\phi : Q_{\mathcal{A}} \to Q_{\mathcal{M}}$ as follows. For every state $s \in Q_{\mathcal{A}}$, $\phi(s) = (\mathcal{M} \cdot u_s)$. Let us show that ϕ is an automaton pseudo-morphism.

Let s be a state of $\mathcal{A}$ and let u be a word such that $(\mathcal{A} \cdot u) = s$. Since $(\mathcal{A} \cdot u) = (\mathcal{A} \cdot u_s)$, the hypothesis implies $(\mathcal{M} \cdot u) = (\mathcal{M} \cdot u_s)$. The definition of ϕ is therefore independent of the choice of the words u_s.

In particular, $\phi(i_{\mathcal{A}}) = (\mathcal{M} \cdot u_{i_{\mathcal{A}}}) = (\mathcal{M} \cdot \varepsilon) = i_{\mathcal{M}}$ hence ϕ satisfies (1). Moreover, since both $\mathcal{A}$ and $\mathcal{M}$ are complete, and since ϕ is a total function, ϕ also satisfies (2). Let $t \xrightarrow{a} t'$ be a transition of $\mathcal{A}$. By definition $\phi(t) = (\mathcal{M} \cdot u_t)$ and since the definition of ϕ does not depend on the choice of the words u_s, we may assume that $u_{t'} = u_t a$. It then follows that

$$\phi(t') = (\mathcal{M} \cdot (u_t a)) = ((\mathcal{M} \cdot u_t) \cdot a) = \phi(t) \cdot a\ .$$

In other words, $\phi(t) \xrightarrow{a} \phi(t')$ is a transition of $\mathcal{M}$. ◀

1.2 Ultimately-equivalent states

Our decision procedure involves the determination of ultimately-equivalent states defined as follows.

▶ **Definition 7.** Let $\mathcal{A}$ be an automaton over $[\![b]\!]$. Let $m \geq 1$ be an integer. Two states s, s' of $\mathcal{A}$ are *m-ultimately-equivalent* if

$$\forall u \in [\![b]\!]^* \quad |u| \geq m \implies (s \cdot u) = (s' \cdot u) \ .$$

Two states are *ultimately-equivalent* if they are m-ultimately-equivalent for some $m \geq 1$.

▶ **Remark 8.** *Note that ultimate-equivalence is indeed an equivalence relation: if s and s' are m-ultimately-equivalent while s' and s'' are m'-ultimately-equivalent, then s and s'' are* $\max(m, m')$*-ultimately-equivalent.*

Considering an automaton $\mathcal{A}$ over $[\![b]\!]$, the computation of this relation is easy. Let us build a directed graph $\mathcal{G} = (V, E)$ as follows. The vertex-set is $V = Q_\mathcal{A} \times Q_\mathcal{A}$ and the edge set is:

$$\forall (s,t), (s',t') \in V,\ s \neq t$$
$$(s,t) \to (s',t') \text{ in } \mathcal{G} \quad \iff \exists a \in [\![b]\!] \text{ such that } \mathcal{A} \text{ features } \left\{ \begin{array}{l} s \xrightarrow{a} s' \\ t \xrightarrow{a} t' \end{array} \right. \ . \tag{6}$$

In particular, vertices of the form (s,s) never qualify for the above condition and thus never have outgoing edges. Observe that two distinct states s, t of $\mathcal{A}$ are ultimately-equivalent if and only if (s,t) may not reach in $\mathcal{G}$ a strongly connected component.

Computing the strongly connected components of a graph is done in linear time (see, for instance, Tarjan's algorithm [20]). Hence, the set of the pairs of states of $\mathcal{A}$ that are ultimately-equivalent may be decided in time $O(bn^2)$. This complexity can be improved as follows.

▶ **Proposition 9** (Béal, Crochemore, [3]). *Let $\mathcal{A}$ be an automaton over $[\![b]\!]$. We write n the number of states of $\mathcal{A}$. The set of the pairs of states of $\mathcal{A}$ that are ultimately-equivalent may be decided in time $O(bn \log n)$.*

Sketch. We take verbatim the algorithm in [3]. Start from the trivial partition and iteratively merge states. Each step of the algorithm consists in merging two states that are 1-ultimately-equivalent. The purpose of Béal and Crochemore was to show that starting with a so-called AFT automaton $\mathcal{A}$, the result is the minimisation of $\mathcal{A}$. Starting with any automaton $\mathcal{A}$, the resulting automaton is not necessarily minimal. However, one can observe that its states are precisely the ultimate-equivalence classes of $\mathcal{A}$. ◀

As a direct consequence of the definition of an automaton morphism, ultimate-equivalence commutes with automaton morphisms.

▶ **Lemma 10.** *Let $\mathcal{A}$ and $\mathcal{M}$ be two automata such that there is an automaton morphism $\phi : \mathcal{A} \to \mathcal{M}$. Let s and s' be two states of $\mathcal{A}$ that are ultimately-equivalent (w.r.t. $\mathcal{A}$), then $\phi(s)$ and $\phi(s')$ are also ultimately-equivalent (w.r.t. $\mathcal{M}$).*

2 Purely periodic b-recognisable sets

▶ **Notation 11.** *Let $p > 0$ and $b > 1$ be two integers. Throughout this section, the quantities k, d, j, ψ are fixed as follows.*

- *Let k, d be the unique integers such that $p = k\,d$ where k is the greatest divisor of p coprime with b. In particular, the prime factors occurring in the prime decomposition of d all appear in the prime decomposition of b. Moreover, $(k, d) = 1$.*
- *Let j be the least integer such that d is a divisor of b^j.*
- *Since $(k, b) = 1$, the order of b in $\mathbb{Z}/k\mathbb{Z}$ is well defined and denoted by ψ, i.e., $b^\psi \equiv 1\ [k]$.*

Let $s < k$ and $t < d$ be two integers. We let $\langle s, t\rangle$ denote the (unique) integer of $\mathbb{Z}/p\mathbb{Z}$ congruent to s modulo k and t modulo d. Note that if n is an integer less than p, then $n = \langle n\%k, n\%d\rangle$ where $n\%k$ denote the remainder of the division of n by k.

2.1 The automaton $\mathcal{A}_{(p,R)}$ and its minimisation

▶ **Definition 12.** A subset P of integers is *purely periodic*, if there exist $p \geq 1$ and a subset $R \subseteq \{0, \ldots, p-1\}$ such that $P = R + p\mathbb{N}$.

For instance, $\{0, 1\} + 4\mathbb{N}$ is purely periodic but $\{4, 5\} + 4\mathbb{N}$ is not. Let $p \geq 1$ be an integer and R be a subset of $\{0, \ldots, p-1\}$. We say that the parameter (p, R) is *proper*, if p is the smallest period of the purely periodic set $R + p\mathbb{N}$. For instance, $(4, \{0, 1\})$ is proper but $(4, \{0, 2\})$ is not because $\{0, 2\} + 4\mathbb{N} = \{0\} + 2\mathbb{N}$.

The following definition is ubiquitous when dealing with periodic sets of integers. It is an easy exercise to show that this automaton accepts base-b representations of integers whose remainder modulo p belongs to R.

▶ **Definition 13.** We let $\mathcal{A}_{(p,R)}$ denote the automaton $\mathcal{A}_{(p,R)} = \langle [\![b]\!], \mathbb{Z}/p\mathbb{Z}, \delta, 0, R\rangle$ where δ is defined as

$$\forall n \in \mathbb{Z}/p\mathbb{Z},\ \forall a \in [\![b]\!] \quad n \xrightarrow{a} nb + a\ .$$

When we are only interested in the transitions of the automaton $\mathcal{A}_{(p,R)}$, it is sometimes convenient to leave the set of final states unspecified. In that case, we write $\mathcal{A}_{(p,?)}$ for the automaton where the final/non-final status of the states is not set.

▶ **Example 14.** Figure 1c shows $\mathcal{A}_{(12,\{5,7\})}$ in base 2. Transitions with label 1 (resp. 0) are represented with bold (resp. thin) edges.

As can be seen, for instance, in Figure 2, the automaton $\mathcal{A}_{(p,R)}$ is not necessarily minimal.

▶ **Lemma 15.** *For every word $u \in [\![b]\!]^*$, $(\mathcal{A}_{(p,R)} \cdot u) = (\overline{u}\%p) = \langle \overline{u}\%k, \overline{u}\%d\rangle$.*

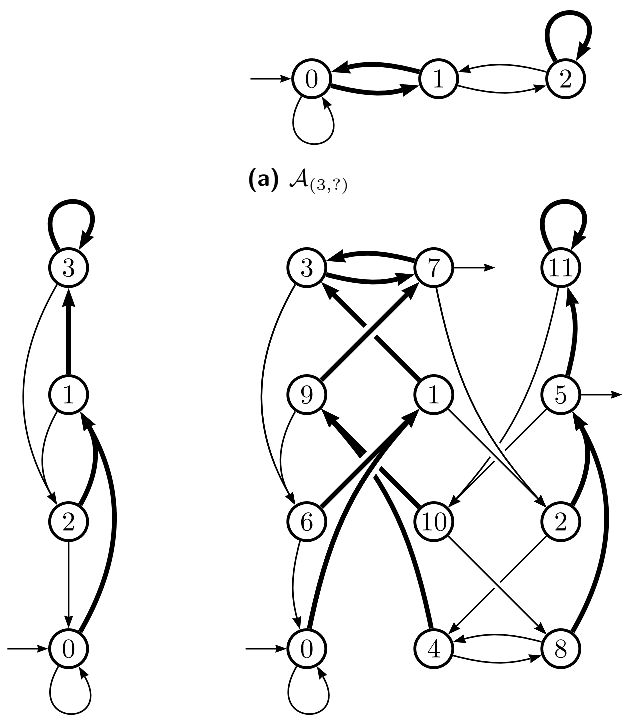
Proof. This follows directly from the definition of the transition function of $\mathcal{A}_{(p,?)}$. ◀

▶ **Property 16.** *The automaton $\mathcal{A}_{(p,R)}$ is strongly connected.*

Proof. Let n, m be two states. The state n is of the form $\langle i, i'\rangle$. Let u be a word satisfying

$$\overline{u} \equiv \langle k-i, 0\rangle [p]\ , \quad |u| \geq j \quad \text{and} \quad |u| \equiv 0[\psi]\ .$$

The last two conditions are easily satisfied by adding a suitable number of leading zeroes. Reading u from n leads to the initial state 0. Obviously, reading $\langle m\rangle$ from 0 leads to m. ◀



(a) $\mathcal{A}_{(3,?)}$

(b) $\mathcal{A}_{(4,?)}$

(c) $\mathcal{A}_{(12,\{5,7\})}$

Figure 1 The automaton $\mathcal{A}_{(12,\{5,7\})}$, as the product automaton of $\mathcal{A}_{(4,?)}$ by $\mathcal{A}_{(3,?)}$

The next lemma states that the automaton $\mathcal{A}_{(p,?)}$ is the product automaton $\mathcal{A}_{(k,?)} \times \mathcal{A}_{(d,?)}$. This easily follows from the Chinese remainder theorem and Lemma 15.

▶ **Lemma 17.** *For all integers $s, s' \in \mathbb{Z}/k\mathbb{Z}$, $t, t' \in \mathbb{Z}/d\mathbb{Z}$ and every word $u \in [\![b]\!]^*$,*

$$\langle s,t\rangle \xrightarrow{u} \langle s',t'\rangle \text{ in } \mathcal{A}_{(p,?)} \iff \begin{cases} s \xrightarrow{u} s' \text{ in } \mathcal{A}_{(k,?)} \\ t \xrightarrow{u} t' \text{ in } \mathcal{A}_{(d,?)} \end{cases}$$

The fact that k is coprime with b implies the following result.

▶ **Lemma 18.** *With the definition introduced in Notation 11, the automaton $\mathcal{A}_{(k,?)}$ is a group automaton: each letter induces a permutation on the set of states.*

Proof. Since k is coprime with b, the function $f_0 : \mathbb{Z}/k\mathbb{Z} \to \mathbb{Z}/k\mathbb{Z}$ defined by $s \mapsto sb$ is a permutation of $\mathbb{Z}/k\mathbb{Z}$. Hence, so is the function f_a defined by $s \mapsto (sb + a)$, for every letter $a \in [\![b]\!]$. The action of a in $\mathcal{A}_{(k,?)}$ is exactly f_a, a permutation of the states. ◀

2.2 Nerode-equivalence and ultimate-equivalence in $\mathcal{A}_{(p,R)}$

Within the setting of Example 14 where rows (resp. columns) of the product automaton $\mathcal{A}_{(p,R)} \approx \mathcal{A}_{(d,?)} \times \mathcal{A}_{(k,?)}$ correspond to the equivalence classes modulo d (resp. modulo k), the forthcoming Proposition 20 shows that Nerode-equivalent states in $\mathcal{A}_{(p,R)}$ must belong to the same column. See, for instance, Figure 2. Then, we show that all states belonging to the same column are ultimately-equivalent.

▶ **Lemma 19.** *If (p, R) is proper, then for all distinct integers i and i', $0 \le i, i' < k$, the states id and $i'd$ are not Nerode-equivalent.*

Proof. Since (p, R) is proper and $id \neq i'd$, there exists an integer m such that $(id + m) \in R + p\mathbb{N}$ and $(i'd + m) \notin R + p\mathbb{N}$.

We let u denote a word such that $\overline{u} = m$ and $|u| \equiv 0[\psi]$ (in other words, u is the word $\langle m \rangle$ padded with an appropriate number of 0's); it thus holds that $b^{|u|} \equiv 1\ [k]$. Reading the word u respectively from the states id and $i'd$ leads to the states:

$$id \cdot u = idb^{|u|} + m \quad \text{and} \quad i'd \cdot u = i'db^{|u|} + m\ .$$

The integer $(idb^{|u|} + m)$ is congruent to $(id + m)$ modulo k (since $b^{|u|} \equiv 1\ [k]$) as well as modulo d (since both are obviously congruent to m) hence modulo p. The same reasoning also applies to the second state, finally yielding:

$$id \cdot u = id + m \quad \text{and} \quad i'd \cdot u = i'd + m\ .$$

The first state belongs to R and is thus final while the second does not belong to R and thus is not final. The word u is then a witness of the fact that id and $i'd$ are not Nerode-equivalent. ◀

▶ **Proposition 20.** *Let (p, R) be proper. If i and i' are Nerode-equivalent states, then they are congruent modulo k.*

Proof. Proof by contrapositive. Let i and i' be two states that are not congruent modulo k. By definition of j, see Notation 11, the states $(i \cdot 0^j)$ and $(i' \cdot 0^j)$ are both congruent to 0 modulo d. However the operation $i \mapsto ib$ is a permutation of $\mathbb{Z}/k\mathbb{Z}$, hence $(i \cdot 0^j)$ and $(i' \cdot 0^j)$ are not congruent modulo k. It follows that $(i \cdot 0^j) = ld$ and $(i' \cdot 0^j) = l'd$ for some distinct $l, l' \in \mathbb{Z}/k\mathbb{Z}$. Lemma 19 then yields that these states are not Nerode-equivalent, hence that i and i' are not either. ◀

▶ **Lemma 21.** *Let s and s' be two states of $\mathcal{A}_{(p,R)}$. With the definition introduced in Notation 11, if $s \equiv s'[k]$, then s and s' are j-ultimately-equivalent.*

Proof. Let u be any word of length j. Since s and s' are congruent modulo k, there exists $i \in \mathbb{Z}/k\mathbb{Z}$ and $l, l' \in \mathbb{Z}/d\mathbb{Z}$ such that $s = \langle i, l \rangle$ and $s' = \langle i, l' \rangle$. Then, from Lemma 17 and using the fact that $lb^j \equiv 0\ [d]$, we get

$$(s \cdot u) = \langle ib^j + \overline{u}\,,\, lb^j + \overline{u} \rangle = \langle ib^j + \overline{u}\,,\, \overline{u} \rangle\ .$$

Similarly $(s' \cdot u) = \langle ib^j + \overline{u}\,,\, \overline{u} \rangle = (s \cdot u)$. ◀

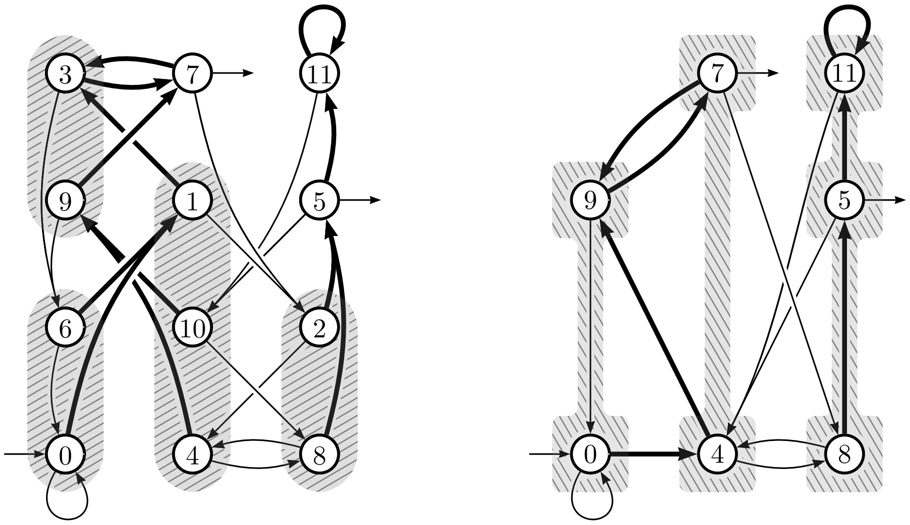
2.3 Circuits labelled by the digit 0

A circuit whose every arc is labelled by the digit 0 is called for short a *0-circuit*. For instance, the automaton $\mathcal{A}_{(12,\{5,7\})}$ depicted in Figure 1 has two such circuits: one reduced to the state 0 and one made of the states 4 and 8. We will see that the number of states belonging to 0-circuits has a special meaning.

▶ **Lemma 22.** *A state of $\mathcal{A}_{(p,R)}$ is a multiple of d if and only if it belongs to a 0-circuit.*

Proof. Forward direction. It is enough to show that every state of the form id, for $i \in \mathbb{Z}/k\mathbb{Z}$, has a predecessor by 0 of the form $i'd$, $i' \in \mathbb{Z}/k\mathbb{Z}$. Simple arithmetic yields that $(b^{-1}i)d$ is suitable, where b^{-1} is the inverse of b in $\mathbb{Z}/k\mathbb{Z}$.

Backward direction. Proof by contrapositive. Let s be a state which is not a multiple of d. The state $(s \cdot 0^j)$ is a multiple of d. Therefore, for every integer $i \geq j$, the state $(s \cdot 0^i)$ is a multiple of d, hence is not equal to s. Since $\mathcal{A}_{(p,R)}$ is deterministic, $(s \cdot 0^i)$ cannot be equal to s for $i < j$ either. ◀



(a) Nerode-equivalence classes of $\mathcal{A}_{(12,\{5,7\})}$

(b) Pseudo-morphism equivalence classes in the minimisation of $\mathcal{A}_{(12,\{5,7\})}$

Figure 2 Minimisation morphism of $\mathcal{A}_{(12,\{5,7\})}$ and pseudo-morphism of its minimisation

The next proposition follows from Lemmas 22 and 19. Recall that k is the largest integer coprime with b such that $p = k\,d$ and $d \geq 1$ (see Notation 11).

▶ **Proposition 23.** *If (p, R) is proper, the minimisation of $\mathcal{A}_{(p,R)}$ possesses exactly k states that belong to 0-circuits.*

3 Characterisation of automata accepting purely periodic sets

The next result will allow us to decide whether a deterministic automaton $\mathcal{A}$ over $[\![b]\!]$, given as input, is such that $\overline{L(\mathcal{A})}$ is a purely periodic set of integers, i.e., whether or not it is of the form $R + p\mathbb{N}$ for some R and p.

▶ **Theorem 24.** *Let $b > 1$ be a base and $\mathcal{A}$ be a minimal automaton over $[\![b]\!]$ bearing a self-loop labelled by 0 on the initial state. Let ℓ be the number of states in $\mathcal{A}$ that belong to 0-circuits. The automaton $\mathcal{A}$ accepts by value a purely periodic set of integers if and only if the following two conditions are fulfilled.*

a. *There exists a pseudo-morphism $\phi : \mathcal{A} \to \mathcal{A}_{(\ell,?)}$.*

b. *The equivalence relation induced by ϕ is a refinement of the ultimate-equivalence relation.*

Proof of forward direction. Since $\mathcal{A}$ accepts by value a purely periodic set of integers, there exists a smallest period p and a remainder-set $R \subseteq \{0, \ldots, p-1\}$ such that $L(\mathcal{A}) = 0^*\langle R+p\mathbb{N}\rangle$. Note that (p, R) is proper by choice of p. We make use of Notation 11. In particular, k is the greatest divisor of p that is coprime with b.

Since $\mathcal{A}$ is minimal, it is isomorphic to the minimisation of any automaton accepting $L(\mathcal{A})$, in particular, to the minimisation of $\mathcal{A}_{(p,R)}$. It then follows from Proposition 23 that $\ell = k$.

To prove that there exists a pseudo-morphism $\phi : \mathcal{A} \to \mathcal{A}_{(k,?)}$, we will apply Lemma 6. Let u,u' be two words such that $(\mathcal{A}_{(k,?)} \cdot u) \neq (\mathcal{A}_{(k,?)} \cdot u')$. Let us show that $(\mathcal{A} \cdot u) \neq (\mathcal{A} \cdot u')$. Since $(\mathcal{A}_{(k,?)} \cdot u) \neq (\mathcal{A}_{(k,?)} \cdot u')$, we have that $\overline{u} \not\equiv \overline{u'}\ [k]$. Due to Lemma 15, $(\mathcal{A}_{(p,R)} \cdot u)$ and $(\mathcal{A}_{(p,R)} \cdot u')$ are not congruent modulo k. It then follows from Proposition 20 that the states $(\mathcal{A}_{(p,R)} \cdot u)$ and $(\mathcal{A}_{(p,R)} \cdot u')$ are not Nerode-equivalent, which implies that $(\mathcal{A} \cdot u) \neq (\mathcal{A} \cdot u')$ because $\mathcal{A}$ is the minimisation of $\mathcal{A}_{(p,R)}$.

Let s and s' be two states of $\mathcal{A}$ such that $\phi(s) = \phi(s')$. We have to show that s and s' are ultimately-equivalent. Let u and u' be two words such $(\mathcal{A} \cdot u) = s$ and $(\mathcal{A} \cdot u') = s'$. Since ϕ is a pseudo-morphism, we get that

$$(\mathcal{A}_{(k,?)} \cdot u) = \phi(s) = \phi(s') = (\mathcal{A}_{(k,?)} \cdot u')$$

and so $\overline{u} \equiv \overline{u'}\ [k]$. Applying Lemma 15 yields that the states $(\mathcal{A}_{(p,R)} \cdot u)$ and $(\mathcal{A}_{(p,R)} \cdot u')$ are congruent modulo k, and by Lemma 21, these states are ultimately-equivalent. Since $\mathcal{A}$ is the minimisation of $\mathcal{A}_{(p,R)}$, we have an automaton morphism $\mathcal{A}_{(p,R)} \to \mathcal{A}$. Finally, since ultimate-equivalence commutes with automaton morphism (Lemma 10), $(\mathcal{A} \cdot u) = s$ and $(\mathcal{A} \cdot u') = s'$ are ultimately-equivalent. ◀

Proof of backward direction. By assumption, for all $i \in \mathbb{Z}/\ell\mathbb{Z}$, the states in $\phi^{-1}(i)$ are ultimately-equivalent. For every integer $i \in \mathbb{Z}/\ell\mathbb{Z}$, we let m_i denote the least integer such that, for all s, s' in $\phi^{-1}(i)$, $(s \cdot u) = (s' \cdot u)$ whenever $|u| \geq m_i$. Let $m = \max\{m_i \mid i \in \mathbb{Z}/\ell\mathbb{Z}\}$.

Let u, u' be two words with respective values that are congruent modulo ℓb^m. Note that, in particular, $\overline{u}$ and $\overline{u'}$ are thus congruent modulo b^m. Let us show that u and u' reach the same state in $\mathcal{A}$.

Since $\mathcal{A}$ bears a self-loop labelled by 0 on the initial state, the word $0^m u$ is such that $\overline{0^m u} = \overline{u}$ and $\mathcal{A} \cdot 0^m u = \mathcal{A} \cdot u$. We may thus assume that u and u' are longer than m. There exist factorisations $u = vw$ and $u' = v'w'$ such that the lengths of w and w' are both equal to m. Since $\overline{u}$ and $\overline{u'}$ are congruent modulo b^m, w and w' are equal: $u = vw$, $u' = v'w$.

Assume without loss of generality that $\overline{u} \geq \overline{u'}$. Hence $\overline{u} - \overline{u'} = (\overline{v} - \overline{v'})b^m$ is congruent to 0 modulo ℓb^m. We deduce that $\overline{v}$ and $\overline{v'}$ are congruent modulo ℓ. By Lemma 15, the respective runs of v and v' in $\mathcal{A}_{(\ell,?)}$ reach the same state: $(\mathcal{A}_{(\ell,?)} \cdot v) = (\mathcal{A}_{(\ell,?)} \cdot v')$. From assumption (a), we get $\phi(\mathcal{A} \cdot v) = \phi(\mathcal{A} \cdot v')$. In other words, the states $(\mathcal{A} \cdot v)$ and $(\mathcal{A} \cdot v')$ are ϕ-equivalent. Hence, by assumption (b), they are m_i-ultimately-equivalent. Since $|w| = m \geq m_i$ (by choice of m), we get that $(\mathcal{A} \cdot v \cdot w) = (\mathcal{A} \cdot v' \cdot w)$: the run in $\mathcal{A}$ of the words $u = vw$ and $u' = v'w$ indeed reach the same state.

We have just shown that words whose values are congruent modulo ℓb^m have runs in $\mathcal{A}$ reaching the same states, hence either all are accepted by $\mathcal{A}$ or none of them are. The run of a word u is then accepted by $\mathcal{A}$ if and only if $\langle \overline{u} \% (\ell b^m) \rangle$ is. Finally, a word u is accepted by $\mathcal{A}$ if and only if $\overline{u} \% (\ell b^m)$ belongs to the set $R \subseteq \{0, \ldots, \ell b^m - 1\}$, defined by

$$R = \{\ i \in \mathbb{Z}/\ell b^m \mathbb{Z} \mid (\mathcal{A} \cdot \langle i \rangle) \text{ is final }\} \ . \quad ◀$$

▶ **Remark 25.** *In the proof of the forward direction, it was stated that $\ell = k$ (where k is the greatest divisor of the period which is coprime with the base). It is also the case in the backward direction. Indeed, the automaton $\mathcal{A}$ is shown to accept a purely periodic set of integers. Let (p, R) denotes the **proper** parameter of this set (it is not necessarily the one given in the proof). Since $\mathcal{A}$ is minimal, it is the quotient of $\mathcal{A}_{(p,R)}$. It then follows from Proposition 23 that, ℓ, the number of states belonging to 0-circuits, is equal to k, the greatest divisor of the period which is coprime with the base.*

3.1 Complexity and algorithmic issues

Theorem 24 yields an algorithm to decide whether a given deterministic automaton $\mathcal{A}$ accepts by value a purely periodic set of integers:

0. if necessary, minimise $\mathcal{A}$ and make it complete;
1. count the number ℓ of states of $\mathcal{A}$ that belong to 0-circuits;
2. build the automaton $\mathcal{A}_{(\ell,?)}$;
3. construct, if it exists, the pseudo morphism $\phi : \mathcal{A} \to \mathcal{A}_{(\ell,?)}$;
4. check whether, for all $x \in \mathbb{Z}/\ell\mathbb{Z}$, the states of $\phi^{-1}(x)$ are ultimately-equivalent.

Let us denote by n the number of states of $\mathcal{A}$. Step (0) can be carried out in $O(bn \log n)$ time. Steps (1), (2) can obviously be performed in $O(bn)$ time. A morphism between deterministic automata, if it exists, can be computed by a single traversal of the bigger automaton; the same algorithm also works for pseudo-morphisms: Step (3) also runs in $O(bn)$ time. The ultimate-equivalence classes of $\mathcal{A}$ can be computed in time $O(bn \log n)$ from Proposition 9, hence so is the execution of Step (4).

▶ **Corollary 26.** *Let $b > 1$ be a base and $\mathcal{A}$ be a n-state deterministic automaton over $[\![b]\!]$. It is decidable in $O(bn \log n)$ time whether $\mathcal{A}$ accepts by value a purely periodic set of integers.*

▶ **Remark 27.** *Remark 25 gives a very fast rejection test. Indeed, before Step (2) we may check whether the integer ℓ (computed by Step (1)) is coprime with b. If it is not the case, $\mathcal{A}$ may be rejected already.*

▶ **Example 28.** We start with the minimal automaton $\mathcal{A}$ depicted in Figure 3. Step (1) is shown in Figure 4: $\mathcal{A}$ has five states belonging 0-circuits and thus, $\ell = 5$. Step (2) then consists in constructing $\mathcal{A}_{(5,?)}$, shown in 5. There is a pseudo-morphism $\mathcal{A} \to \mathcal{A}_{(5,?)}$, whose equivalence classes are represented in Figure 6. Finally, one could check that Step (4) holds: all states belonging to the same class are 3-ultimately-equivalent. Hence $\mathcal{A}$ accepts an eventually periodic set of period $2^3 \times 5$. It is indeed the minimisation of $\mathcal{A}_{(40,\{0,3\})}$.

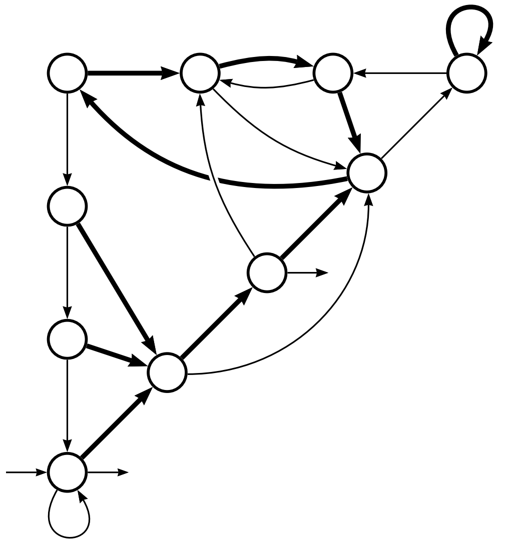


Figure 3 An automaton $\mathcal{A}$

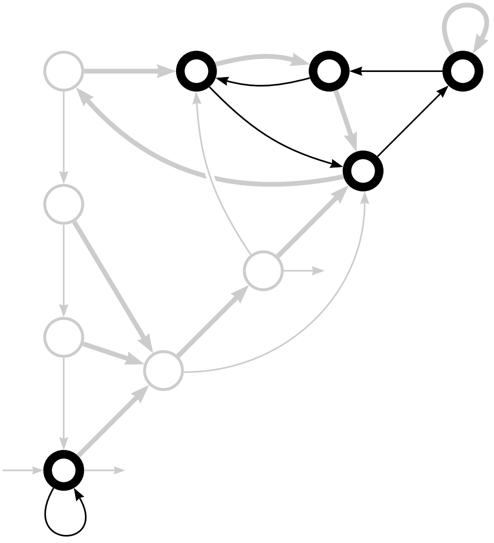


Figure 4 The 0-circuits of $\mathcal{A}$ have 5 states in total

4 Impurely periodic b-recognisable sets

In this section, we will study the eventually periodic sets of integers that are not purely periodic (see Definition 12). We say that such sets are *impurely periodic*. For denotational

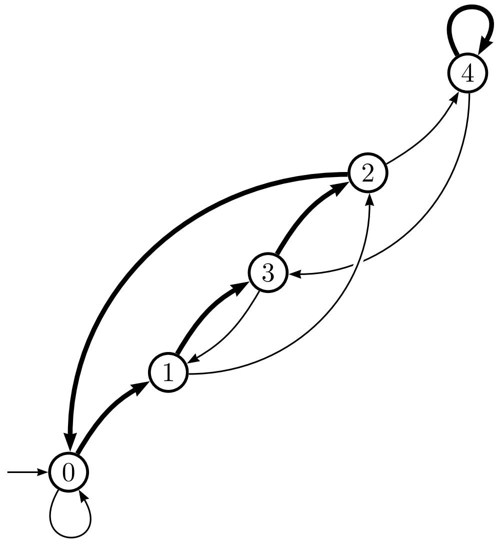


Figure 5 The automaton $\mathcal{A}_{(5,?)}$

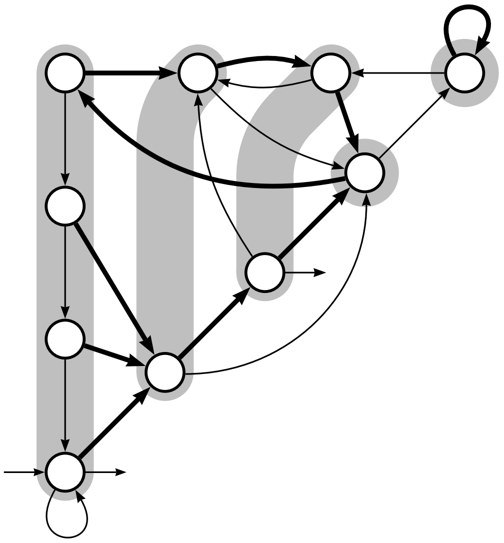


Figure 6 Equivalence classes of the relation induced by the pseudo-morphism $\mathcal{A} \to \mathcal{A}_{(5,?)}$

reasons, we will describe eventually periodic sets S with three parameters: a period p, a remainder-set $R \subseteq \{0, \dots, p-1\}$ and a finite set $I \subseteq \mathbb{N}$ of "mismatches" with a purely periodic set. Such a triplet (p, R, I) is a *parameter* of S if

$$S = (R + p\mathbb{N}) \oplus I \ ,$$

where $\oplus$ is the *exclusive disjunction* operation: an integer belongs to S if it belongs either to $(R + p\mathbb{N})$ or to I, but not both. One can also find the terminology *symmetric difference* or *disjunctive union* (and the notation Δ).

▶ **Example 29.** The set $S = \{0, 6\} \cup (\{4, 5\} + 4\mathbb{N})$ can be described by the parameter $(4, \{0, 1\}, \{1, 6\})$. Indeed, the purely periodic set $P = \{0, 1\} + 4\mathbb{N}$ and the set S differ only by the fact that $1 \in P \setminus S$ and $6 \in S \setminus P\}$.

This way of describing eventually periodic sets has several advantages: the parameter has only three components, the set I is uniquely defined and it allows to determine if S is purely periodic (Lemmas 30 and 31),

▶ **Lemma 30.** *Let S be an eventually periodic set of integers. There is a unique purely periodic set $P \subseteq \mathbb{N}$ and a unique finite set $I \subseteq \mathbb{N}$ of mismatches such that $S = I \oplus P$.*

We then say that the triplet (p, R, I) is *the proper parameter* of an eventually periodic set S if $S = (R + p\mathbb{N}) \oplus I$ and p is the smallest positive period for which such R, I exist. We take the convention that the proper parameter of a finite set S is $p, R, I = (1, \emptyset, S)$, (instead of considering that the period equals 0); this is why the smallest period is assumed to be positive in the previous sentence.

▶ **Lemma 31.** *An eventually periodic set S of parameter (p, R, I) is purely periodic if and only if I is empty.*

▶ **Notation 32.** *In what follows, we consider impurely periodic sets of integers, hence a finite* ***non-empty*** *set $I \subseteq \mathbb{N}$ of mismatches is given. Moreover, we still follow the convention of*

Notation 11 recapped hereafter. A period p and a remainder-set $R \subseteq \{0, \dots, p-1\}$ are given. We let k denote the greatest divisor of p that is coprime with the base b, d is the integer such that $kd = p$ and j is the smallest integer such that d divides b^j.

4.1 The automata $\mathcal{B}_I$ and $\mathcal{C}_{(p,R,I)} = \mathcal{A}_{(p,R)} \oplus \mathcal{B}_I$

We will describe an automaton accepting, by value, an eventually periodic set with parameter (p, R, I). We first have to deal with the set I of mismatches.

▶ **Definition 33.** We denote by m the greatest element of I. We denote by $\mathcal{B}_I$ the automaton:

$$\mathcal{B}_I = \langle [\![b]\!],\ \{0, \dots, m\} \cup \{\bot\},\ \delta,\ 0,\ I \rangle \ ,$$

where δ if defined as follows.

$$\forall i \in \{0, \dots, m\},\ \forall a \in [\![b]\!] \quad \begin{cases} i \xrightarrow{a} (ib+a) & \text{if } ib + a \leq m \\ i \xrightarrow{a} \bot & \text{otherwise} \end{cases} \tag{7}$$

$$\forall a \in [\![b]\!] \quad \bot \xrightarrow{a} \bot \tag{8}$$

Simple and formal verification yields the following properties of $\mathcal{B}_I$. We write *scc* for strongly connected component. A *trivial* scc is a state belonging to no circuit.

▶ **Lemma 34.** *The automaton $\mathcal{B}_I$*

a. *is deterministic, complete and trim;*
b. *has exactly two non-trivial sccs: $\{0\}$ and $\{\bot\}$;*
c. *has exactly two 0-circuits, the respective self-loops on 0 and $\bot$;*
d. *accepts a word $u \in [\![b]\!]^*$ if and only if $\overline{u} \in I$.*

In the next definition, the exclusive disjunction $\oplus$ is extended to sets of pairs of states.

▶ **Definition 35.** Given two complete automata $\mathcal{A}$ and $\mathcal{B}$ over the alphabet $[\![b]\!]$. We define the *exclusive disjunction* $\mathcal{A} \oplus \mathcal{B}$ as usual:

$$\mathcal{A} \oplus \mathcal{B} = \langle [\![b]\!],\ Q_{\mathcal{A}} \times Q_{\mathcal{B}},\ \delta,\ (i_{\mathcal{A}}, i_{\mathcal{B}}),\ (F_{\mathcal{A}} \times Q_{\mathcal{B}}) \oplus (Q_{\mathcal{A}} \times F_{\mathcal{B}}) \rangle \ ,$$

where δ is defined as follows.

$$\forall s, s' \in Q_{\mathcal{A}},\ \forall t, t' \in Q_{\mathcal{B}},\ \forall a \in [\![b]\!] \quad (s,t) \xrightarrow{a} (s', t') \iff \begin{cases} s \xrightarrow{a} s' \text{ in } \mathcal{A} \\ t \xrightarrow{a} t' \text{ in } \mathcal{B} \end{cases} .$$

It is quite obvious that a word u is accepted by $\mathcal{A} \oplus \mathcal{B}$ if and only if it is accepted by $\mathcal{A}$ or $\mathcal{B}$, but not by both of them.

▶ **Notation 36.** *We let $\mathcal{C}_{(p,R,I)}$ denote the exclusive disjunction $\mathcal{C}_{(p,R,I)} = \mathcal{A}_{(p,R)} \oplus \mathcal{B}_I$.*

The next lemma gives properties of $\mathcal{C}_{(p,R,I)} = \mathcal{A}_{(p,R)} \oplus \mathcal{B}_I$ that follow from Lemma 34 and Definition 35. Recall that we have seen in Property 16 that $\mathcal{A}_{(p,R)}$ is strongly connected.

▶ **Lemma 37.** *The following properties hold.*

a. *$\mathcal{C}_{(p,R,I)}$ is deterministic, complete and trim.*
b. *$\mathcal{C}_{(p,R,I)}$ possesses exactly two non-trivial sccs:*
  - *the singleton made of the initial state, $\{(0,0)\}$,*
  - *and $\{ (s, \bot) \mid s \text{ is a state of } \mathcal{A}_{(p,R)} \}$.*

  Moreover this second scc is isomorphic to $\mathcal{A}_{(p,R)}$ by projecting to the first component, hence complete.

c. *$\mathcal{C}_{(p,R,I)}$ accepts a word $u \in [\![b]\!]^*$ if and only if $\overline{u} \in ((R + p\mathbb{N}) \oplus I)$.*

4.2 The 0-circuits of $\mathcal{C}_{(p,R,I)}$ and of its minimisation

The next statement gives a description of the 0-circuits of $\mathcal{C}_{(p,R,I)}$ and follows from Lemmas 22 and 37.

▶ **Lemma 38.** *A state s belongs to a 0-circuit of $\mathcal{C}_{(p,R,I)}$ if and only if either*

a. *the state s is initial, or*

b. *there exists $x \in \mathbb{Z}/k\mathbb{Z}$ such that $s = (xd, \perp)$.*

The relationship of $\mathcal{C}_{(p,R,I)}$ with its minimisation is stated by the next lemma. It is similar to the one of $\mathcal{A}_{(p,R)}$ with its minimisation.

▶ **Lemma 39.** *If (p, R, I) is proper, the following statements hold.*

a. *For every distinct integers $x, x' \in \mathbb{Z}/k\mathbb{Z}$, the states $(xd, \perp)$ and $(x'd, \perp)$ of $\mathcal{C}_{(p,R,I)}$ are not Nerode-equivalent.*

b. *The states of $\mathcal{C}_{(p,R,I)}$ that belong to 0-circuits are pairwise Nerode-inequivalent.*

c. *The initial state of $\mathcal{C}_{(p,R,I)}$ is not Nerode-equivalent to any other state.*

Proof. Item (a) follows directly from Lemmas 19 and 37(b).

(b). From item (a) and Lemma 38, it suffices to show that there is no state $s = (xd, \perp)$ which is Nerode-equivalent to the initial state. For the sake of contradiction let us assume that such a state exists.

We denote by i the initial state of $\mathcal{C}_{(p,R,I)}$. Let v be any word whose run reaches s, hence satisfying $i \cdot v = s$. Since i bears a loop labelled by 0, the run of $0v$ reaches s, hence, without loss of generality, we may assume that $|v| \equiv 0\ [\psi]$.

Since s and i are Nerode-equivalent, so are $s \cdot v$ and $(i \cdot v) = s$. By iterating this reasoning, we obtain that i is Nerode-equivalent to $(i \cdot v^k)$. Similarly, i is Nerode-equivalent to the state $(i \cdot v^k 0^j)$, that we denote by s'. Moreover, since s' is reachable from s and since s belongs to the $\perp$-scc (complete from Lemma 37(b)), s' belongs to the $\perp$-scc as well.

Since $|v| \equiv 0\ [\psi]$, $\overline{v^k} \equiv k\overline{v} \equiv 0\ [k]$, hence $\overline{v^k 0^j} \equiv 0\ [k]$. Since $\overline{v^k 0^j}$ is obviously a multiple of d, it is also a multiple of $p = kd$. In other, words $s' = (0, \perp)$ and the initial state is Nerode-equivalent to $(0, \perp)$. This contradicts the fact that I is non-empty.

(c). Let us denote by X the Nerode-equivalence class of the initial state. Since the initial state bears a loop labeled by 0, the set X is stable by reading the digit 0. Therefore, if X were containing a non-initial state, then it would contain a whole 0-circuit (distinct from the initial state), contradicting item (b). ◀

The next statement can then be established using Lemma 39(b) much like Proposition 20 was shown using Lemma 19.

▶ **Proposition 40.** *If (p, R, I) is proper, then two Nerode-equivalent states (s', t') and (s', t') are necessarily such that s and s' are congruent modulo k.*

It follows from Lemma 38 that $\mathcal{C}_{(p,R,I)}$ has $(k + 1)$ states that belong to 0-circuits and from Lemma 39(b) that such states are not merged by the minimisation process, hence the next proposition holds.

▶ **Proposition 41.** *If (p, R, I) is proper, the 0-circuits of the minimisation of $\mathcal{C}_{(p,R,I)}$ have a total of $(k + 1)$ states*

4.3 Ultimate equivalence class of $\mathcal{C}_{(p,R,I)}$

▶ **Lemma 42.** *Let (s,t) and (s',t') be two states of $\mathcal{C}_{(p,R,I)}$ such that $s \equiv s'[k]$. If neither (s,t) nor (s',t') is the initial state, then (s,t) and (s',t') are ultimately-equivalent.*

Proof. Since by hypothesis, (s,t) and (s',t') are not initial, there exists a bound m such that for every word u longer that m, the states $(s,t)\cdot u$ and $(s,t')\cdot u$ belong to the $\bot$-scc. Without loss of generality, we may assume that $m \geq j$.

Let u be a word longer than m. Then,

$$(s,t)\cdot u = (sb^{|u|} + \overline{u}, \bot) \text{ and } (s',t')\cdot u = (s'b^{|u|} + \overline{u}, \bot)\ .$$

Since s and s' are congruent modulo k, and since k is coprime with b, it holds:

$$sb^{|u|} + \overline{u} \equiv s'b^{|u|} + \overline{u}\ [k]\ . \tag{9}$$

Moreover, since $|u| \geq j$, and since d divises b^j,

$$sb^{|u|} + \overline{u} \equiv \overline{u} \equiv s'b^{|u|} + \overline{u}\ [d]\ . \tag{10}$$

Finally, since d and k are coprime, (9) and (10) yield

$$sb^{|u|} + \overline{u} \equiv s'b^{|u|} + \overline{u}\ [p]\ ,$$

hence $(s,t)\cdot u = (s',t')\cdot u$. ◀

▶ **Lemma 43.** *The initial state of $\mathcal{C}_{(p,R,I)}$ is not ultimately equivalent to any other state.*

Proof. The only state from whom the initial state may be reached is the initial state itself. Moreover, as the initial state bears a loop labelled by 0, the words of 0^* are witnesses of the fact that no state is ultimately equivalent to the initial state. ◀

5 Characterisation of automata accepting impurely periodic sets

▶ **Theorem 44.** *Let $b > 1$ be a base and $\mathcal{A}$ be a minimal automaton over $[\![b]\!]$. We write $(\ell+1)$ for the number of states in $\mathcal{A}$ that belong to 0-circuits. The automaton $\mathcal{A}$ accepts by value an **impurely** periodic set of integers if and only if the following conditions are met.*

a. *There exists a pseudo-morphism $\phi : \mathcal{A} \to \mathcal{A}_{(\ell,?)}$.*

b. *The initial state excluded, the equivalence relation induced by ϕ is a refinement of the ultimate-equivalence relation.*

c. *The initial state bears a self-loop labelled by the digit 0 and features no other incoming transitions.*

Proof of forward direction. Conditions (a) and (b) are obtained much like it was done in Theorem 24. We simply apply Propositions 41 and 40 instead of 23 and 20.

Since $\mathcal{A}$ is minimal and accepts by value an impurely periodic set, there exists a parameter (p,R,I) such that $\mathcal{A}$ is the minimisation of $\mathcal{C}_{(p,R,I)}$. A simple verification yields that Condition (c) is satisfied by $\mathcal{C}_{(p,R,I)}$. Besides, it follows from Lemma 39(c) that the minimisation process does not merge any state of $\mathcal{C}_{(p,R,I)}$ with the initial state. As a result, the incoming transitions to the initial state are the same in $\mathcal{A}$ and $\mathcal{C}_{(p,R,I)}$. ◀

Proof of backward direction. There are finitely many ultimate-equivalence classes. Hence there exists an integer m such that, if two states s and s' are ultimately equivalent, then they are m-ultimately-equivalent.

Note also that since $\mathcal{A}$ is complete, Condition (c) implies that $\ell \geq 1$.

Let u, u' be two words whose respective values are congruent modulo ℓb^m and greater than b^m. Thus, there are words v, v', w, w', $|w| = |w'| = m$, satisfying $u = vw$, $u' = v'w'$ and such that v, v' both possess a non-zero digit. In particular, neither $\mathcal{A} \cdot v$ nor $\mathcal{A} \cdot v'$ is the initial state. With exactly the same proof as was given in Theorem 24, it may then be shown that $\mathcal{A} \cdot u = \mathcal{A} \cdot u'$.

In other words, $\mathcal{A}$ accepts an ultimately periodic set of integers S of period ℓb^m. (In general, this period is not the smallest one, which would be ℓd for some d dividing b^m.) We moreover write I the set of mismatches (existence and unicity ensured by Lemma 30). Let us show that it is not purely periodic, or equivalently that I is not empty (Lemma 31).

We denote by s the state reached by the run of the word $\langle \ell b^m \rangle$, *i.e.*, $s = \mathcal{A} \cdot \langle \ell b^m \rangle$. Since $\ell \geq 1$, this word possesses a non-zero digit, hence s is not the initial state of $\mathcal{A}$ (Condition (c)). Since $\mathcal{A}$ is minimal, s and $i_{\mathcal{A}}$ are not Nerode-equivalent. Hence there exists a word w such that exactly one of the states in $\{s \cdot w,\ i_{\mathcal{A}} \cdot w\}$ is final. Since $\overline{w}$ and $\overline{\langle \ell b^m \rangle w}$ are obviously congruent modulo ℓb^m, $\overline{w}$ is a mismatch: it belongs to I. ◀

As stated below, Theorem 44 gives an algorithm to decide whether an automaton accepts an ultimately periodic set of integers. It is the same as the one from Section 3.1 with an additional Step (5) at the end. It consists in verifying that Condition 44(c) holds.

▶ **Corollary 45.** *Let b be a base and $\mathcal{A}$ be a n-state deterministic automaton over $[\![b]\!]$. It is decidable in $O(bn \log n)$ time whether $\mathcal{A}$ accepts by value an impurely periodic set of integers.*

Since an eventually periodic set is either purely or impurely periodic, Theorem 1 is a direct consequence of Corollaries 26 and 45.

References

1 Jean-Paul Allouche, Narad Rampersad, and Jeffrey Shallit. Periodicity, repetitions, and orbits of an automatic sequence. *Theoret. Comput. Sci*, 410:2795–2803, 2009.

2 Jean-Paul Allouche and Jeffrey Shallit. *Automatic Sequences: Theory, Applications, Generalizations.* Cambridge University Press, 2003.

3 Marie-Pierre Béal and Maxime Crochemore. Minimizing local automata. In M. Fossorier G. Caire, editor, *IEEE International Symposium on Information Theory*, pages 1376–1380, 2007.

4 Jason Bell, Emilie Charlier, Aviezri S. Fraenkel, and Michel Rigo. A decision problem for ultimately periodic sets in nonstandard numeration systems. *IJAC*, 19(6):809–839, 2009.

5 Valérie Berthé and Michel Rigo, editors. *Combinatorics, Automata and Number Theory.* Number 135 in Encyclopedia Math. Appl. Cambridge University Press, 2010.

6 Bernard Boigelot and Julien Brusten. A generalization of Cobham's theorem to automata over real numbers. *Theor. Comput. Sci.*, 410(18):1694–1703, 2009.

7 Bernard Boigelot, Sébastien Jodogne, and Pierre Wolper. An effective decision procedure for linear arithmetic over the integers and reals. *ACM Trans. Comput. Log.*, 6(3):614–633, 2005.

8 V. Bruyère and G. Hansel. Recognizable sets of numbers in nonstandard bases. In R. Baeza-Yates, E. Goles, and P. V. Poblete, editors, *LATIN '95: Theoretical Informatics*, volume 911 of *Lect. Notes Comput. Sci.*, pages 167–179. Springer, 1995.

9 V. Bruyère, G. Hansel, C. Michaux, and R. Villemaire. Logic and p-recognizable sets of integers. *Bull. Belg. Math. Soc.*, 1:191–238, 1994. Corrigendum, *Bull. Belg. Math. Soc.* **1** (1994), 577.

10 Emilie Charlier, Narad Rampersad, and Jeffrey Shallit. Enumeration and decidable properties of automatic sequences. *Int. J. Found. Comput. Sci.*, 23(5):1035–1066, 2012.

11 Alan Cobham. On the base-dependence of sets of numbers recognizable by finite automata. *Mathematical Systems Theory*, 3(2):186–192, 1969.

12 Fabien Durand. Decidability of the HD0L ultimate periodicity problem. *RAIRO - Theor. Inf. and Applic.*, 47(2):201–214, 2013.

13 Juha Honkala. A decision method for the recognizability of sets defined by number systems. *ITA*, 20(4):395–403, 1986.

14 Jérôme Leroux. A polynomial time Presburger criterion and synthesis for number decision diagrams. In *Logic in Computer Science 2005 (LICS 2005)*, pages 147–156. IEEE Comp. Soc. Press, 2005.

15 Victor Marsault and Jacques Sakarovitch. Ultimate Periodicity of b-Recognisable Sets: A Quasilinear Procedure. In Marie-Pierre Béal and Olivier Carton, editors, *Developments in Language Theory - 17th International Conference (DLT 2013)*, number 7907 in Lect. Notes Comput. Sci., pages 362–373. Springer, 2013.

16 Ivan Mitrofanov. A proof for the decidability of HD0L ultimate periodicity (in russian). Preprint arXiv:1110.4780, 2011.

17 Andrei A. Muchnik. The definable criterion for definability in Presburger arithmetic and its applications. *Theor. Comput. Sci.*, 290(3):1433–1444, 2003. English translation of a prior article with the same name in Russian, Moscow's Institute of New Technologies, 1991.

18 Jacques Sakarovitch. *Elements of Automata Theory.* Cambridge University Press, 2009. Corrected English translation of *Éléments de théorie des automates*, Vuibert, 2003.

19 Alexei L. Semenov. Presburgerness of predicates regular in two number systems. *Siberian Mathematical Journal*, 18(2):289–300, 1977. English translation from Russian Translated from Sibirskii Matematicheskii Zhurnal, 18(2), pp. 403–418, 1977.

20 Robert E. Tarjan. Depth-first search and linear graph algorithms. *SIAM J. Comput.*, 1(2):146–160, 1972.