An Efficient Algorithm to Decide Periodicity of b-Recognisable Sets Using MSDF Convention*

Bernard Boigelot[1], **Isabelle Mainz**[2], **Victor Marsault**[†3], **and Michel Rigo**[4]

1 **Montefiore Institute & Department of Mathematics, Université de Liège, Liège, Belgium**
`bernard.boigelot@ulg.ac.be`

2 **Montefiore Institute & Department of Mathematics, Université de Liège, Liège, Belgium**
`isabelle.mainz@ulg.ac.be`

3 **Montefiore Institute & Department of Mathematics, Université de Liège, Liège, Belgium**
`victor.marsault@ulg.ac.be`

4 **Montefiore Institute & Department of Mathematics, Université de Liège, Liège, Belgium**
`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 is 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.

1998 ACM Subject Classification G.2.2 Graph Theory

Keywords and phrases integer-base systems, automata, recognisable sets, periodic sets

Digital Object Identifier 10.4230/LIPIcs.ICALP.2017.118

1 Introduction

Let $b > 1$ be an integer base. We let $\llbracket b \rrbracket = \{0, 1, \ldots, b-1\}$ denote the canonical alphabet of base-b digits. If $u = u_\ell \cdots u_0$ belongs to $\llbracket b \rrbracket^*$, 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 $\llbracket b \rrbracket$ not starting with

[*] An extended version of this paper is available in arXiv:1702.03715 [8], `https://arxiv.org/abs/1702.03715`.

[†] Corresponding author. Supported by a Marie Skłodowska-Curie fellowship, co-funded by the European Union and the University of Liège, Belgium.

44th International Colloquium on Automata, Languages, and Programming (ICALP 2017).
Editors: Ioannis Chatzigiannakis, Piotr Indyk, Fabian Kuhn, and Anca Muscholl;
Article No. 118; pp. 118:1–118:14
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

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 $u = 0^i \langle n \rangle$.

1.1 Our contribution

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. In the literature, one also finds the term "saturated language". The main result of this paper is the following.

▶ **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.*

Due to space constraints, this paper deals with the purely periodic case only. The general case is treated similarly [8].

Theorem 1 relies on a characterisation of the minimal automata accepting by value purely periodic sets. This characterisation is given Section 4. It relies on the notion of ultimately-equivalent states which is previously introduced in Section 2.2. In Section 3, we study the structure of the naive automaton accepting an arbitrary purely periodic sets, and its minimisation.

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 digit first (LSDF) [15, 16]. 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. For instance, this event occurs for the language $L_n = 0^* \, 1 \, (0+1)^n \, 1 \, (0+1+\varepsilon)^n$ and its mirror K_n: the number of states in the minimal automaton accepting L_n (resp. K_n) grows linearly (resp. exponentially) with n. Evaluating L_n as MSDF encodings or K_n as LSDF encodings yields the same finite (thus eventually periodic) set of integers.

1.2 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, [12]). *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, [14]). *It is decidable whether or not a given b-automatic word is eventually periodic.*

Complexity issues are however not considered at all in Honkala's paper. The decidability of our problem of interest can also be obtained using a first-order logic characterisation of b-recognisable sets given by Büchi's theorem, and the fact that Presburger arithmetic is decidable [10, 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 [15] for eventually-periodic b-recognisable sets of integers, that relies on intricate geometrical constructions. (Leroux's result is indeed 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.) Still using LSDF convention, the third author and Sakarovitch designed a quasilinear algorithm [16]. The general idea is similar to the one we use here: finding characteristic properties that are preserved by minimisation. However, the criterion used in the present work is mostly unrelated to [16] and yields a very different decision algorithm.

1.3 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 structure documented in [6].

1.4 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 eventually periodic [13, 17]. 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, [9]. In that setting, several decision problems in combinatorics on words, including the ultimate periodicity problem, are decidable [11]. Using Honkala's techniques, the decision problem considered in our Theorem 1 is generalised 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.

2 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.

2.1 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.

We recall the following classical result. 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 $\mathcal{M}$. ◀

2.2 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. 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.

Given 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:

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

In particular, vertices of the form (s, s) never qualify for the above condition and thus have no 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 (e.g., with Tarjan's algorithm [19]). Hence, the set of the pairs of states of $\mathcal{A}$ that are ultimately-equivalent may be computed in time $O(bn^2)$. This complexity can be improved as follows.

▶ **Proposition 8** (Béal, Crochemore, [3]). *Let $\mathcal{A}$ be an automaton over $[\![b]\!]$ and n the number of states in $\mathcal{A}$. The ultimate-equivalence classes of $\mathcal{A}$ may be computed in time $O(bn \log n)$.*

Proof Sketch. We take verbatim the algorithm in [3]. One starts from the trivial partition and iteratively merges 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 9.** *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. $\mathcal{M}$).*

3 Purely periodic b-recognisable sets

The content of this section is the following. Section 3.1 gives the definition and main properties of the "naive" automaton $\mathcal{A}_{(p,R)}$ accepting the purely periodic set $R + p\mathbb{N}$. Then, in Section 3.2 we study the relationship between ultimate equivalence and Nerode equivalence in $\mathcal{A}_{(p,R)}$. In Section 3.3, we show how to extract relevant information on the period p from the minimisation of $\mathcal{A}_{(p,R)}$.

▶ **Notation 10.** *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, $(k, d) = 1$.*
- *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 j be the least integer such that d is a divisor of b^j.*

Let $s < k$ and $t < d$ be two integers. Let $\langle s, t\rangle$ denote the integer of $\mathbb{Z}/p\mathbb{Z}$ congruent to s modulo k and t modulo d. This integer is unique by the Chinese remainder theorem. 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.

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

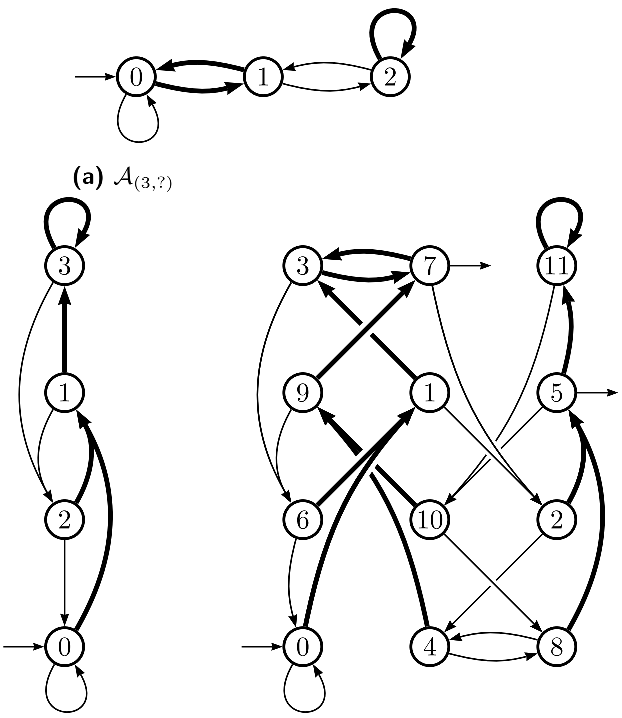
▶ **Definition 11.** A subset P of integers is *purely periodic*, if there exist $p \geq 1$ and a subset $R \subseteq \{0, \ldots, p-1\}$ such that $P = R + p\mathbb{N}$.

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 12.** 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

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



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

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 13.** 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 14.** *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 15.** *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. ◀

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 14.

▶ **Lemma 16.** *For all integers $s, s' \in \mathbb{Z}/k\mathbb{Z}$, $t, t' \in \mathbb{Z}/d\mathbb{Z}$ and every word $u \in [\![b]\!]^*$,*

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

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

▶ **Lemma 17.** *With the definition introduced in Notation 10, 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. ◀

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

Within the setting of Example 13 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 19 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 18.** *If (p, R) is proper, then for all distinct integers i and i', $0 \leq i, i' < k$, 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:

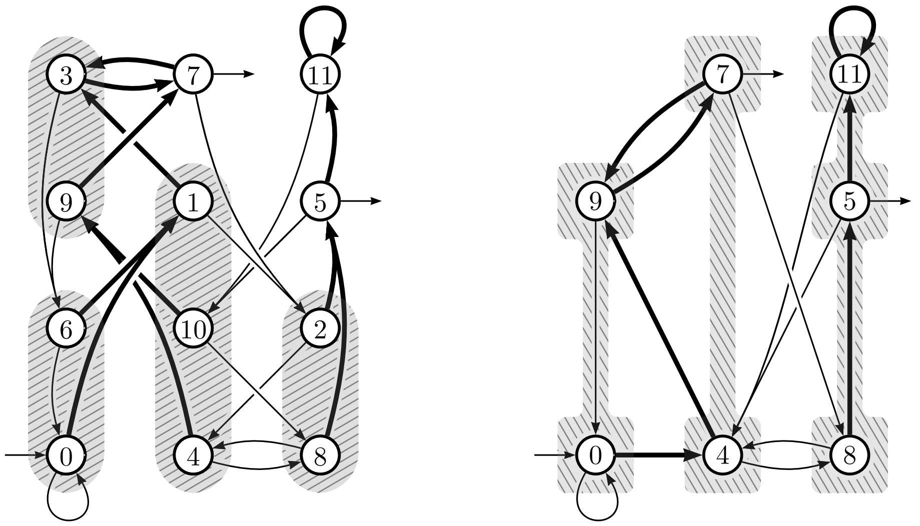
$$id \cdot u = id + m \quad \text{and} \quad i'd \cdot u = i'd + m \ .$$

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 19.** *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 10, 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 18 then yields that these states are not Nerode-equivalent, hence that i and i' are not either. ◀

▶ **Lemma 20.** *Let s and s' be two states of $\mathcal{A}_{(p,R)}$. With the definition introduced in Notation 10, if $s \equiv s'[k]$, then s and s' are j-ultimately-equivalent.*



(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.

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

3.3 Circuits labelled by the digit 0

A circuit in which 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: $0 \xrightarrow{0} 0$ and $4 \xrightarrow{0} 8 \xrightarrow{0} 4$. We will see that the number of states belonging to 0-circuits has a special meaning.

▶ **Lemma 21.** *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. ◀

The next proposition follows from Lemmas 21 and 18. Recall that k is the largest integer coprime with b such that $p = k\,d$ and $d \geq 1$ (see Notation 10).

▶ **Proposition 22.** *If (p, R) is proper, the minimisation of $\mathcal{A}_{(p,R)}$ possesses exactly k states that belong to 0-circuits.*

4 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. We say that an equivalence relation $\sim_1$ is a *refinement* of another equivalence relation $\sim_2$ if for every x and y, $x \sim_1 y \implies x \sim_2 y$.

▶ **Theorem 23.** *Let $b > 1$ be a base and $\mathcal{A}$ a minimal automaton over $[\![b]\!]$. 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 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.*

(c) *The initial state of $\mathcal{A}$ bears a self-loop labelled by the digit* 0

Proof of forward direction. Since the automaton $\mathcal{A}$ is minimal and accepts by value, the initial state of $\mathcal{A}$ necessarily bears a loop labelled by the digit 0; in other words, item (3) holds.

More precisely, 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 10. 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 22 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 14, $(\mathcal{A}_{(p,R)} \cdot u)$ and $(\mathcal{A}_{(p,R)} \cdot u')$ are not congruent modulo k. It then follows from Proposition 19 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 14 yields that the states $(\mathcal{A}_{(p,R)} \cdot u)$ and $(\mathcal{A}_{(p,R)} \cdot u')$ are congruent modulo k, and by Lemma 20, 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 9), $(\mathcal{A} \cdot u) = s$ 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 14, 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 (1), 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 (2), 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 $R \subseteq \{0, \dots, \ell b^m - 1\}$, defined by

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

▶ **Remark 24.** *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 a quotient of $\mathcal{A}_{(p,R)}$. It then follows from Proposition 22 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.*

4.1 Complexity and algorithmic issues

Theorem 23 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) and (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 8, hence so is the execution of Step (4).

▶ **Corollary 25.** *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 26.** *Remark 24 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 27.** 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

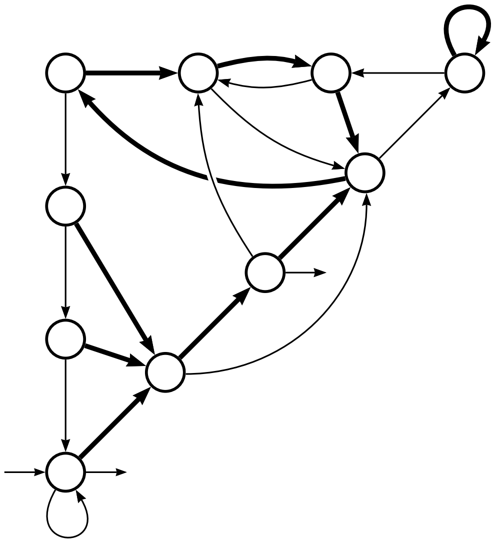


Figure 3 An automaton $\mathcal{A}$.

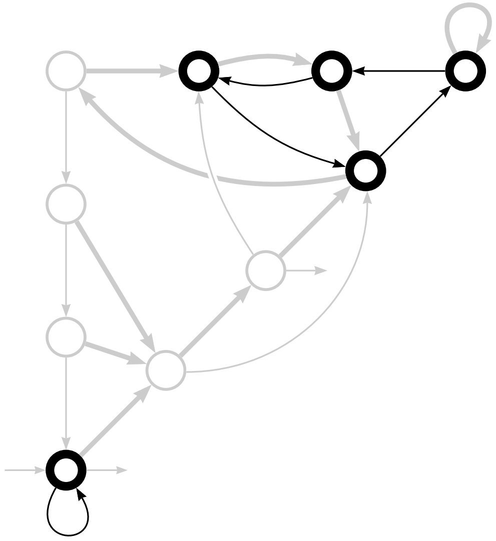


Figure 4 The 0-circuits have 5 states in total.

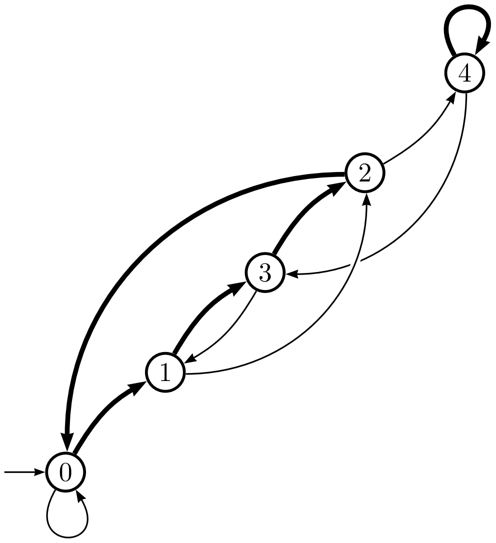


Figure 5 The automaton $\mathcal{A}_{(5,?)}$.

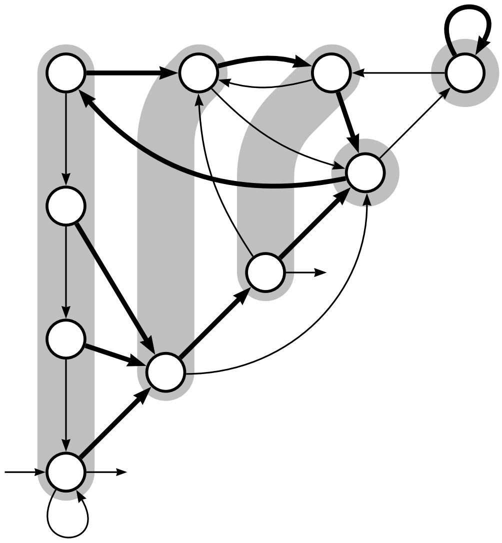


Figure 6 Equivalence classes of the relation induced by the pseudo-morphism $\mathcal{A} \to \mathcal{A}_{(5,?)}$.

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 $\mathcal{A}_{(40,\{0,3\})}$.

5 Generalisation to eventually periodic sets

Let us now consider the eventually periodic sets of integers that are not purely periodic (see Definition 11). We say that such sets are *impurely periodic* and Theorem 28 below gives a characterisation of the minimal automata that accept them.

▶ **Theorem 28.** *Let $b > 1$ be a base and let $\mathcal{A}$ 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* ***im****purely 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.*

Due to space constraints we do not detail the proof of Theorem 28. Although it is not immediate, it is much similar to the proof of Theorem 23 and may be found in arXiv [8].

As stated by the next corollary, Theorem 28 gives an algorithm to decide whether an automaton accepts an impurely periodic set of integers. It is the same as the one from Section 4.1 with the following modification and addition:

0 to **3.** same tests as in Section 4.1.

4. check whether, for every $x \in \mathbb{Z}/\ell\mathbb{Z}$, the **non-initial** states of $\phi^{-1}(x)$ are ultimately-equivalent;

5. check whether the initial state has no incoming transition.

▶ **Corollary 29.** *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 25 and 29.

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 Int. Symp. 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 Bernard Boigelot, Isabelle Mainz, Victor Marsault, and Michel Rigo. An efficient algorithm to decide periodicity of b-recognisable sets using MSDF convention, 2017. Preprint arXiv:1702.03715.

9 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.

10 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.

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

12 Alan Cobham. On the base-dependence of sets of numbers recognizable by finite automata. *Mathematical Systems Theory*, 3(2):186–192, 1969. doi:10.1007/BF01746527.

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

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

15 Jérôme Leroux. A polynomial time Presburger criterion and synthesis for number decision diagrams. In *LICS 2005*, pages 147–156. IEEE Comp. Soc. Press, 2005.

16 Victor Marsault and Jacques Sakarovitch. Ultimate Periodicity of b-Recognisable Sets: A Quasilinear Procedure. In *DLT 2013*, number 7907 in Lect. Notes Comput. Sci., pages 362–373. Springer, 2013.

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

18 Jacques Sakarovitch. *Elements of Automata Theory.* Cambridge University Press, 2009.

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