ELSEVIER

Theoretical Computer Science 181 (1997) 17–43

Theoretical Computer Science

Bertrand numeration systems and recognizability [1]

Véronique Bruyère [a,*,2], Georges Hansel [b]

[a] *Université de Mons-Hainaut, 15 Avenue Maistriau, B-7000 Mons, Belgium*
[b] *Université de Rouen, Laboratoire d'Informatique de Rouen, Place Blondel, F-76134 Mont Saint-Aignan, France*

Abstract

There exist various well-known characterizations of sets of numbers recognizable by a finite automaton, when they are represented in some integer base $p \geqslant 2$. We show how to modify these characterizations, when integer bases p are replaced by linear numeration systems whose characteristic polynomial is the minimal polynomial of a Pisot number. We also prove some related interesting properties.

1. Introduction

Since the work of [9], sets of integers recognizable by finite automata have been studied in numerous papers. One of the jewels in this topic is the famous Cobham's theorem [11]: the only sets of numbers recognizable by finite automata, independently of the base of representation, are those which are ultimately periodic. Other studies are concerned with computation models equivalent to finite automata in the recognition of sets of integers. The proposed models use first-order logical formulæ [9], uniform substitutions [12], algebraic formal series [10]. We refer the reader to the surveys [13, Chapter 5; 23, Section 8; 8].

During the last years, many researchers have investigated representations of numbers in nonstandard bases (like the Fibonacci numeration system) and their relationship with finite automata [22, 4, 25, 14–17, 21, 2]. Given a nonstandard numeration system U, an integer can be represented by more than one U-representation. One representation is distinguished, it is the one computed by the Euclidean algorithm. The normalization ν_U is the map transforming any U-representation into the normalized one. The use of nonstandard numeration systems U, instead of usual bases, raises several problems. The normalization ν_U and the set $\mathcal{N}_U$ of all the normalized U-representations, are generally

* Corresponding author. E-mail: veronique.bruyere@umh.ac.be.
[1] Work partially supported by ESPRIT-BRA Working Group 6317 ASMICS.
[2] Work done in part while visiting the University of Rouen.

PII S0304-3975(96)00260-5

not recognizable by a finite automaton [25, 17]. The characterization of numeration systems U with recognizable ν_U and $\mathcal{N}_U$ is not completely settled yet.

This paper deals with numeration systems U defined by a linear recurrence relation whose characteristic polynomial is the minimal polynomial of a Pisot number. The Fibonacci system belongs to this class. Other examples are Bertrand numeration systems studied in [4]. Under this hypothesis, the normalization ν_U and the set $\mathcal{N}_U$ are both recognizable by finite automata [17]. We give new simple proofs of these results.

We are mainly interested in U-recognizable sets of integers and the possible computation models associated with them. A subset of $\mathbb{N}$ is called U-recognizable if the normalized U-representations of its elements are recognizable by a finite automaton. We here propose three models: U-automata, non-uniform U-substitutions and first-order U-formulæ. They are all equivalent (Theorems 16, 20 and 21) and they are strongly dependent on the recognizability of ν_U and $\mathcal{N}_U$. Moreover, these characterizations hold in higher dimensions. Our models naturally generalize the models known for usual bases [9, 12], as well as the ones proposed in [14, 7] for the Bertrand numeration systems.

We also prove that a set is U-recognizable if and only if it is U'-recognizable, for any numeration systems U' satisfying the same recurrence relation as the system U. Such an equivalence was only known for the set $\mathcal{N}_U$ and the normalization ν_U.

The paper is organized as follows. In Section 2, we recall the definitions of numeration systems and Bertrand numeration systems. We also indicate our work's hypotheses. In Section 3, we introduce the notion of U-recognizable set and we give the new proofs that ν_U and $\mathcal{N}_U$ are recognizable by a finite automaton. We give in Section 4 the first computation model based on first-order logical formulæ. The second and the third models, using U-automata and U-substitutions respectively, are developed in the next section. In Section 6, we prove that the U-recognizability only depends on the recurrence relation that the system U satisfies. The last section contains the conclusions.

The reader is supposed to be familiar with the theory of finite automata.

2. Linear numeration systems

2.1. Representations of integers

Let $U=(U_n)_{n\in\mathbb{N}}$ be a *numeration system*, i.e., a strictly increasing sequence of integers such that $U_0 = 1$ and U_{n+1}/U_n is bounded. Using the Euclidean algorithm, any integer $x > 0$ is represented by

$$x = a_iU_i + a_{i-1}U_{i-1} + \cdots + a_0U_0$$

in the following way. Let i be such that $U_i \leqslant x < U_{i+1}$, let $x_i = x$. Compute the Euclidean division $x_j = a_jU_j + x_{j-1}$ for $j \in \{0,\ldots,i\}$ [15]. The word $a_i \cdots a_0$ is over the *canonical alphabet*

$$A_U = \{0,\ldots,c\}$$

with c the greatest integer less than $\sup U_{n+1}/U_n$. By convention, the integer 0 is represented by the empty word.

An integer x can have several representations $w = a_i \cdots a_0 \in B^*$, with $B \subset \mathbb{Z}$ a finite alphabet, such that $x = \sum_{j=0}^{i} a_j U_j$. Any such representation is called a *U-representation*. We say that

$$\pi_U(w) = \sum_{j=0}^{i} a_j U_j$$

is the *numerical value* of w.

The representation computed by the Euclidean algorithm is called the *normalized U-representation*. It is denoted by $\rho_U(x)$. We denote by $\mathcal{N}_U$ the set of all the normalized U-representations, allowing leading zeros

$$\mathcal{N}_U = 0^*\{\rho_U(x) \mid x \in \mathbb{N}\}.$$

Normalized U-representations are characterized as follows.

Proposition 1. *Let U be a numeration system. Then*

$$w = a_i \cdots a_0 \in \mathcal{N}_U \iff \forall j \in \{0, \ldots, i\}, \quad \sum_{k=0}^{j} a_k U_k < U_{j+1}.$$

Elements of $\mathcal{N}_U$ satisfy the following interesting properties.

Proposition 2. *Let U be a numeration system.*

(1) *Let $u, v \in \mathcal{N}_U$ such that $|u| = |v|$. Let $x = \pi_U(u)$, $y = \pi_U(v)$ their numerical values. Then*

$$x < y \iff u < v,$$

where the ordering $u < v$ is the lexicographic ordering.

(2) *If $uv \in \mathcal{N}_U$, then $v \in \mathcal{N}_U$.*

(3) *If $uv \in \mathcal{N}_U$, then $uv' \in \mathcal{N}_U$, for any $v' \in \mathcal{N}_U$ such that $|v'| = |v|$ and $v' < v$. In particular, if $uv \in \mathcal{N}_U$, then $u0^{|v|} \in \mathcal{N}_U$.*

Remark 1. Notice that, in the general case, $uv \in \mathcal{N}_U$ does not imply that $u \in \mathcal{N}_U$.

Given $B \subset \mathbb{Z}$ a finite alphabet, we define a partial function [16]

$$\nu_{B,U} : B^* \to \mathcal{N}_U$$

called *normalization*, as follows. Let $w \in B^*$ such that $\pi_U(w) \in \mathbb{N}$, then $\nu_{B,U}(w) = \rho_U(\pi_U(w))$.

Linear numeration systems are systems U defined by a linear recurrence relation

$$U_n = d_{k-1} U_{n-1} + \cdots + d_0 U_{n-k}, \quad d_l \in \mathbb{Z}, \ l \in \{0, \ldots, k-1\}, \ d_0 \neq 0$$

for all $n \geqslant k$. The polynomial

$$P_U(X) = X^k - d_{k-1}X^{k-1} - \cdots - d_0$$

is called the *characteristic polynomial* of the system U.

Example 1. Let U be the *Fibonacci numeration system* defined by $U_0 = 1$, $U_1 = 2$ and $U_n = U_{n-1} + U_{n-2}$ for $n \geqslant 2$. The normalized U-representations are those with no two consecutive 1's. Then $\mathcal{N}_U = A_U^* \backslash A_U^* 11 A_U^*$, with $A_U = \{0, 1\}$. The characteristic polynomial of U is $P_U = X^2 - X - 1$.

2.2. Bertrand numeration systems

Definition 1. *Bertrand numeration systems* are systems U such that

$$w \in \mathcal{N}_U \iff w0^n \in \mathcal{N}_U, \quad \forall n \in \mathbb{N}.$$

Example 2. The usual decimal system $U = (10^n)_{n \in \mathbb{N}}$ is a Bertrand system. This is no longer true for the numeration system $U_0 = 1$, $U_1 = 4$, $U_n = U_{n-1} + U_{n-2}$, $n \geqslant 2$. For instance, $3 \in \mathcal{N}_U$ but $30 \notin \mathcal{N}_U$.

In [4], Bertrand numeration systems are characterized thanks to the θ-shift S_θ of some real number $\theta > 1$, as explained below.

Let $\theta > 1$ be a real number. For any $x \in \mathbb{R}$, let $[x]$ be its integer part and $\{x\}$ its fractional part. Any real number $x \in [0, 1]$ is uniquely written as [22]

$$x = \sum_{n \geqslant 1} a_n \theta^{-n} \tag{1}$$

such that $x_1 = x$ and for any $n \geqslant 1$, $a_n = [\theta x_n]$, $x_{n+1} = \{\theta x_n\}$. The infinite sequence

$$e_\theta(x) = (a_n)_{n \geqslant 1} = a_1 \ldots a_n \ldots$$

is called the *θ-expansion* of x.

A particular case is the θ-expansion $e_\theta(1)$ of the number 1. In the case it ends with an infinite sequence of 0's, i.e., $e_\theta(1) = d_1 \ldots d_{n-1} d_n 0^\omega$, $d_n \neq 0$, then we put instead

$$e_\theta(1) = (d_1 \ldots d_{n-1}(d_n - 1))^\omega.$$

This new sequence also satisfies equality (1).

With this convention, the next property holds [4].

Proposition 3. *Let $\theta > 1$ be a real number. A sequence $(a_n)_{n \geqslant 1}$ is the θ-expansion $e_\theta(x)$ of a number $x \in [0, 1[$ if and only if for all $i \in \mathbb{N}$, the shifted sequence $(a_{n+i})_{n \geqslant 1}$ is lexicographically less than the sequence $e_\theta(1)$.*

Moreover, for all $n \in \mathbb{N} \backslash \{0\}$, $a_n < \theta$.

The *alphabet* A_θ associated with θ is then defined as $\{0, 1, \ldots, [\theta]\}$ if $\theta \in \mathbb{R} \backslash \mathbb{N}$, and $\{0, 1, \ldots, \theta - 1\}$ if $\theta \in \mathbb{N}$.

The *θ-shift* S_θ is the set of all the θ-expansions $e_\theta(x)$, $x \in [0,1]$. We denote by $L(\theta)$ the set of all the finite factors of the sequences in S_θ.

Theorem 4 (Bertrand-Mathis [4]). *Let U be a numeration system. Then U is a Bertrand numeration system if and only if there exists a real number $\theta > 1$ such that $\mathcal{N}_U = L(\theta)$.*

In this case, $A_U = A_\theta$ and if $e_\theta(1) = (d_n)_{n \geqslant 1}$, then

$$U_0 = 1, \quad U_n = d_1 U_{n-1} + d_2 U_{n-2} + \cdots + d_n U_0 + 1, \quad n \geqslant 1.$$

Corollary 5. *Let U be a Bertrand numeration system and $\theta > 1$ a real number such that $\mathcal{N}_U = L(\theta)$. Then U is linear if and only if the sequence $e_\theta(1)$ is ultimately periodic. In this case, the minimal polynomial P_θ of θ divides the characteristic polynomial P_U of U.*

Proposition 6. *Let U be a Bertrand numeration system. Then*

$$w = a_i \cdots a_0 \in \mathcal{N}_U \iff \forall j \in \{0, \ldots, i\}, \quad \sum_{k=0}^{j} a_k \theta^k < \theta^{j+1}. \tag{2}$$

Moreover, if U is linear, then there exists a constant $\alpha > 0$ such that for any $w = a_i \cdots a_0 \in \mathcal{N}_U$

$$\forall j \in \{0, \ldots, i\}, \quad \sum_{k=0}^{j} a_k \theta^k < \theta^{j+1} - \alpha. \tag{3}$$

Proof. By Theorem 4, $\mathcal{N}_U = L(\theta)$ for some $\theta > 1$. By definition of the θ-expansions, we have (2) [22].

Let $e_\theta(1) = d_1 \cdots d_n \cdots$, or equivalently, $1 = (d_1/\theta) + \cdots + (d_n/\theta^n) + \cdots$. By Proposition 3, $a_i \cdots a_0 \leqslant d_1 \cdots d_{i+1}$. Hence by (2)

$$\begin{aligned}\sum_{k=0}^{j} a_k \theta^k &\leqslant d_1 \theta^j + d_2 \theta^{j-1} + \cdots + d_{j+1} \theta^0 = \theta^{j+1} \left(\frac{d_1}{\theta} + \frac{d_2}{\theta^2} + \cdots + \frac{d_{j+1}}{\theta^{j+1}} \right) \\ &= \theta^{j+1} \left(1 - \frac{d_{j+2}}{\theta^{j+2}} - \frac{d_{j+3}}{\theta^{j+3}} - \cdots - \frac{d_{j+n}}{\theta^{j+n}} \cdots \right).\end{aligned}$$

If U is linear, then $e_\theta(1)$ is ultimately periodic by Corollary 5. Therefore (3) holds because the possible values of the last factor are in finite number. $\square$

2.3. The hypothesis

This paper is concerned with linear numeration systems U whose characteristic polynomial P_U is the minimal polynomial of a Pisot number θ. A *Pisot number* is an algebraic integer $\theta > 1$ such that the roots of its minimal polynomial, distinct from θ, have modulus less than 1. Notice that for θ a Pisot number, $e_\theta(1)$ is always ultimately periodic. Hence, the systems related to the *same* Pisot number θ only differ by the initial values $U_0, U_1, \ldots, U_{k-1}$, where k is the degree of the minimal polynomial of θ. The values U_n, with $n \geqslant k$, are computed with respect to the same recurrence relation. Among these systems, only one is a Bertrand numeration system, whose the initial values are in Theorem 4.

Hypothesis 1. In the sequel of the paper, we work only with linear numeration systems U whose characteristic polynomial P_U is the minimal polynomial P_θ of some Pisot number θ. We denote by $\mathscr{U}_\theta$ the class of all the numeration systems U such that $P_U = P_\theta$. They satisfy the same linear recurrence relation but may differ on the initial values of the sequence U. The unique Bertrand numeration system of the class $\mathscr{U}_\theta$ is denoted by U_θ.

Let us recall the following well-known fact. As the roots $\theta_1 = \theta$, θ_2, ..., θ_k of the polynomial $P_U = P_\theta$ are simple, there exist complex constants γ_l such that for all $n \in \mathbb{N}$

$$U_n = \sum_{l=1}^{k} \gamma_l \theta_l^n. \tag{4}$$

Let $w = a_i \cdots a_0 \in A_U^*$. We denote

$$\pi_\theta(w) = \gamma_1 \sum_{j=0}^{i} a_j \theta^j.$$

The proof of the following proposition is not difficult. It uses Hypothesis 1 and (4).

Proposition 7. *Let* $U \in \mathscr{U}_\theta$.

(1) *There exists a constant* e *such that*

$$\forall w \in A_U^*, \quad |\pi_U(w) - \pi_\theta(w)| < e. \tag{5}$$

(2) *For any* $\varepsilon > 0$, *there exists a constant* $M_\varepsilon \in \mathbb{N}$ *such that*

$$\forall w \in A_U^*, \ \forall n \geqslant M_\varepsilon, \ |\pi_U(w0^n) - \pi_\theta(w0^n)| < \varepsilon. \tag{6}$$

Example 3. (1) Let $p \in \mathbb{N}$ such that $p \geqslant 2$. Then p is a Pisot number with minimal polynomial $X - p$. The class $\mathscr{U}_p$ only contains the Bertrand numeration system $(p^n)_{n \in \mathbb{N}}$.

(2) The golden number $\phi = \frac{1}{2}(1 + \sqrt{5})$ is a Pisot number with $P_\phi = X^2 - X - 1$ and $e_\phi(1) = (10)^\omega$. The Bertrand numeration system U_ϕ of the class $\mathscr{U}_\phi$ is the Fibonacci system. We already mentioned in Example 2 another numeration system of the class $\mathscr{U}_\phi$: the system $U_0 = 1$, $U_1 = 4$, $U_n = U_{n-1} + U_{n-2}$, $n \geqslant 2$.

(3) Let θ be the Pisot number $\phi^2 = \frac{1}{2}(3 + \sqrt{5})$. Then $P_\theta = X^2 - 3X + 1$ and $e_\theta(1) = 21^\omega$. The system U_θ is defined by $U_0 = 1$, $U_1 = 3$, $U_n = 3U_{n-1} - U_{n-2}$, $n \geqslant 2$.

3. *U*-recognizable sets

The aim of this paper is to give three characterizations of sets of positive integers whose normalized U-representations are recognizable by a finite automaton. In this study, the numeration systems all belong to a class $\mathscr{U}_\theta$, with θ a Pisot number. In this section, we introduce the concept of U-recognizable set. The characterizations are given in Sections 4 and 5.

Definition 2. Let $m \geqslant 1$ and $X \subseteq \mathbb{N}^m$. Let U be a numeration system. We say that X is *U-recognizable* if the set $L = 0^*\rho_U(X)$ is recognizable by a finite automaton.

This notion naturally generalizes the classical concept of *p-recognizable* sets, with p an integer greater than 1 (see [13, 8]).

In the previous definition, if $m > 1$, notation 0 means the m-tuple $(0,\dots,0)$ and $\rho_U(x)$ is defined as

$$(0^{l-l_1}\rho_U(x_1), 0^{l-l_2}\rho_U(x_2), \dots, 0^{l-l_m}\rho_U(x_m))$$

where $x = (x_1,\dots,x_m)$, $l_i = |\rho_U(x_i)|$ and $l = \max l_i$. Hence, elements of L are m-tuples of words of $\mathscr{N}_U$ with the same length. In other words, the finite automaton recognizing L has its edges labelled by m-tuples of letters of the alphabet A_U. Such automata are called *letter-to-letter automata*.

Example 4. (1) It is well known that the set $\{(x,y,z) \in \mathbb{N}^3 \mid x+y=z\}$ is 2-recognizable, by an automaton with three states (see Fig. 1).

(2) Consider the Fibonacci numeration system U_ϕ. The set $\mathbb{N}$ is U_ϕ-recognizable by the finite automaton given on Fig. 2 (see Example 1). The automaton of Fig. 3 recognizes the set of words $w \in \mathscr{N}_{U_\phi}$ with an even number of 1's. The corresponding set $X \subseteq \mathbb{N}$ is then U_ϕ-recognizable.

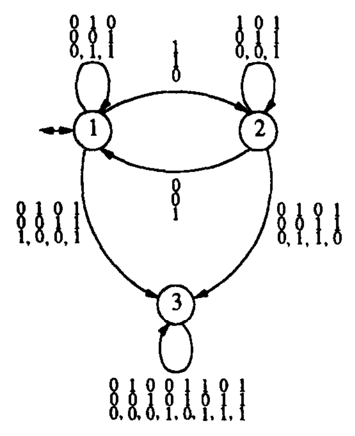

Fig. 1. The addition is 2-recognizable.

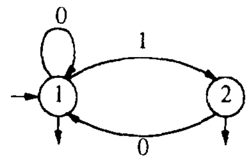

Fig. 2. The minimal automaton for the set of normalized U_ϕ-representations.

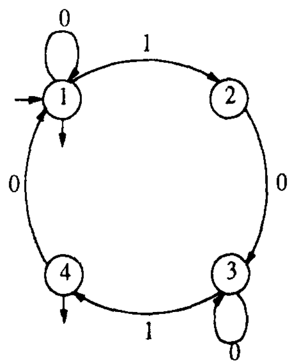

Fig. 3. The Thue–Morse words in the Fibonacci numeration system.

As we will see in Section 4, the recognizability of $\mathcal{N}_U$ and the recognizability of the normalization $v_{B,U}$ are crucial in the study of U-recognizable sets.

3.1. Recognizability of $\mathcal{N}_U$

The recognizability of $\mathcal{N}_U$ is studied in several papers. If U is a numeration system such that $\mathbb{N}$ is U-recognizable, then U is necessarily linear, however the converse is false [25, 21]. If U is a Bertrand numeration system, $\mathbb{N}$ is U-recognizable if and only if U is linear [5]. For the numeration systems $U \in U_\theta$ studied in this paper, $\mathbb{N}$ is always U-recognizable [17]. Recently [20], Hollander has completely characterized which linear numeration systems have a recognizable set $\mathcal{N}_U$.

Theorem 8. *Let $U \in \mathcal{U}_\theta$. Then $\mathbb{N}$ is U-recognizable.*

We give below a simple proof of Theorem 8 which is different from the proof given in [17]. It is mainly based on Proposition 9 which states that for $U \in \mathcal{U}_\theta$, normalized U-representations coincide with normalized U_θ-representations, provided they end with enough zeros.

Proposition 9. *Let $U \in \mathcal{U}_\theta$. There exists $M \in \mathbb{N}$ such that for any $n \geq M$ and any word w,*

$$w0^n \in \mathcal{N}_U \iff w0^n \in \mathcal{N}_{U_\theta}.$$

Proof. We use the notations of equality (4) and Propositions 6, 7. We also use the characterizations of normalized U-representations and θ-expansions given in Propositions 1 and 6, respectively.

Let $\varepsilon = \min\{\frac{1}{2}, \frac{1}{2}\gamma_1\alpha\}$ and $M = M_\varepsilon$.

(1) Let $w = a_i \cdots a_0$ such that $w0^n \in \mathcal{N}_U$. Let $j \in \{0, \ldots, i\}$. By (6) and Proposition 1

$$\pi_\theta(a_j \cdots a_0 0^n) < \pi_U(a_j \cdots a_0 0^n) + \varepsilon \leqslant U_{n+j+1} - 1 + \varepsilon$$
$$< \gamma_1 \theta^{n+j+1} - 1 + 2\varepsilon \leqslant \gamma_1 \theta^{n+j+1}.$$

Therefore $w0^n \in \mathcal{N}_{U_\theta}$ by Proposition 6.

(2) Let $w = a_i \cdots a_0$ such that $w0^n \in \mathcal{N}_{U_\theta}$. Let $j \in \{0, \ldots, i\}$. By (6) and (3)

$$\pi_U(a_j \cdots a_0 0^n) < \pi_\theta(a_j \cdots a_0 0^n) + \varepsilon \leqslant \gamma_1(\theta^{n+j+1} - \alpha) + \varepsilon$$
$$< U_{n+j+1} - \gamma_1\alpha + 2\varepsilon \leqslant U_{n+j+1}.$$

It follows that $w0^n \in \mathcal{N}_U$ by Proposition 1. □

Corollary 10. *Let M as in the previous proposition. Let $w \in A_U^*$, let $n \geqslant M$. Then $w0^M \in \mathcal{N}_U$ if and only if $w0^n \in \mathcal{N}_U$.*

Proof. Remember that U_θ is a Bertrand numeration system. □

The next three lemmas will be used in the proof of Theorem 8.

Lemma 11. *Let $u \in \mathcal{N}_U$ and $n \in \mathbb{N}$. Denote*

$$\Delta_n(u) = \{w \in \mathcal{N}_U \mid uw \in \mathcal{N}_U, |w| = n\}$$

and $\delta_n(u) = \operatorname{card} \Delta_n(u)$. Then function π_U defines a bijection between $\Delta_n(u)$ and $\{x \in \mathbb{N} \mid 0 \leqslant x < \delta_n(u)\}$.

Proof. Take v the greatest word in $\Delta_n(u)$. By Proposition 2(3), $\Delta_n(u) = \{w \in \mathcal{N}_U \mid |w| = n, w \leqslant v\}$. By (1) of the same proposition, we get $\pi_U(\Delta_n(u)) = \{0, 1, \ldots, \pi_U(v) = \delta_n(u) - 1\}$. □

Let M as in Proposition 9. Denote by A_M the set

$$A_M = \{w \in A_U^* \mid w0^M \in \mathcal{N}_U\}.$$

By Corollary 10, we have also $A_M = \{w \in A_U^* \mid w0^n \in \mathcal{N}_U, \ \forall n \geqslant M\}$. We enumerate the set

$$A_M \backslash 0^+ A_M = \{w_0 < w_1 < \cdots < w_n < \cdots\}$$

with respect to the following ordering. Let $u, v \in A_M \backslash 0^+ A_M$, then $u < v$ if and only if either $|u| < |v|$, or $|u| = |v|$ and u is lexicographically less that v.

Given $u \in A_M$, we denote by $\overline{u}$ its *successor* in A_M, such that $u \in 0^* w_n$, $\overline{u} = w_{n+1}$ for some $n \in \mathbb{N}$.

Lemma 12. *Let* $u \in A_M$ *and* $n \geqslant M$. *Then*

$$\delta_n(u) = \pi_U(\overline{u}0^n) - \pi_U(u0^n).$$

Proof. Let $t \in 0^*u$, $t' \in 0^*\overline{u}$ such that $|t| = |t'|$. Let v be the greatest word in $\Delta_n(t)$. Then $\pi_U(tv) < \pi_U(t'0^n)$ by Proposition 2(1). Assume that there exists $t''v'' \in \mathcal{N}_U$ such that $|t''| = |t|$, $|v''| = |v|$ and $\pi_U(tv) < \pi_U(t''v'') < \pi_U(t'0^n)$. By Proposition 2(1) and (3), $tv < t''v'' < t'0^n$ and $t''0^n \in \mathcal{N}_U$. Then $t < t'' < t'$ with $t'' \in A_M$, in contradiction with the definition of $\overline{u}$. Hence

$$\pi_U(tv) + 1 = \pi_U(t'0^n).$$

By Lemma 11, we get

$$\begin{aligned}\delta_n(t) &= \pi_U(v) + 1 = \pi_U(tv) - \pi_U(t0^n) + 1\\ &= \pi_U(t'0^n) - \pi_U(t0^n).\end{aligned}$$

It follows that $\delta_n(u) = \pi_U(\overline{u}0^n) - \pi_U(u0^n)$. $\square$

Lemma 13. *Let* $V_n = \sum_{j=0}^{i} g_j U_{j+n}$ *where* $i \in \mathbb{N}$ *and* $g_j \in \mathbb{R}$, $j \in \{0,\dots,i\}$, *are constants. Then the sequence* $(V_n)_{n\in\mathbb{N}}$ *satisfies the same recurrence relation as* $(U_n)_{n\in\mathbb{N}}$.

Proof of Theorem 8. To show that $\mathbb{N}$ is U-recognizable, let us prove that the syntactic congruence $\sim_{\mathcal{N}_U}$ defined on A_U^* by

$$u \sim_{\mathcal{N}_U} v \;\Leftrightarrow\; \{\forall w \in A_U^*, uw \in \mathcal{N}_U \Leftrightarrow vw \in \mathcal{N}_U\}$$

has finite index. By Lemma 11, we have

$$u \sim_{\mathcal{N}_U} v \;\Leftrightarrow\; \{\forall n \in \mathbb{N}, \delta_n(u) = \delta_n(v)\}.$$

Let M as in Proposition 9. Let k be the degree of the polynomial P_U. We define an equivalence relation R on A_U^* as follows:

$$u R v \;\Leftrightarrow\; \{\forall n \in \{0,\dots,M+k-1\}, \delta_n(u) = \delta_n(v)\}.$$

The relation R has finite index because $\delta_n(u)$ is bounded by $(card\, A_U)^n$. Let us prove that $u R v \Rightarrow u \sim_{\mathcal{N}_U} v$.

Let $u R v$. Either $\delta_M(u) = \delta_M(v) = 0$ or $\delta_M(u) = \delta_M(v) \neq 0$. In the first case, by Corollary 10, $\delta_n(u) = \delta_n(v) = 0$ for all $n \geqslant M$. Consequently, $u \sim_{\mathcal{N}_U} v$. The second case leads to the same conclusion in the following way. By Lemma 12, as $u, v \in A_M$,

$$\delta_n(u) = \pi_U(\overline{u}0^n) - \pi_U(u0^n)$$

$$\delta_n(v) = \pi_U(\overline{v}0^n) - \pi_U(v0^n)$$

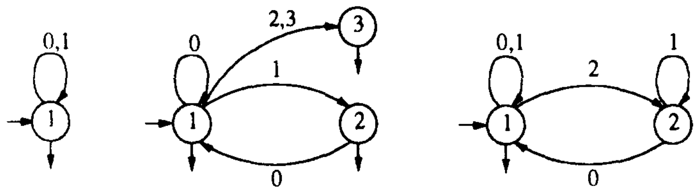

Fig. 4. Three canonical automata.

for all $n \geq M$. Hence by Lemma 13, the sequences $(\delta_{M+n}(u))_{n\in\mathbb{N}}$ and $(\delta_{M+n}(v))_{n\in\mathbb{N}}$ satisfy the same recurrence relation as $(U_n)_{n\in\mathbb{N}}$. Since they coincide on the first k values, they are equal. Conseqently, $u \sim_{\mathcal{N}_U} v$. □

3.2. Automaton for $\mathcal{N}_U$

Thanks to Theorem 8, we associate with $\mathcal{N}_U$ a canonical automaton $\mathcal{A}_U$.

Definition 3. Let $U \in \mathcal{U}_\theta$. We denote by $\mathcal{A}_U$ the trim minimal automaton which recognizes $\mathcal{N}_U$. We say that $\mathcal{A}_U$ is the *canonical automaton* of U.

Remark 2. There exists in $\mathcal{A}_U$ a loop with label 0 on the initial state.

Remark 3. All states of $\mathcal{A}_U$ are final if and only if for any $u, v \in A_U^*$, $uv \in \mathcal{N}_U \Rightarrow u \in \mathcal{N}_U$. By Proposition 2(3), this condition is equivalent to $u0^n \in \mathcal{N}_U \Rightarrow u \in \mathcal{N}_U$ for all $u \in A_U^*$, $n \in \mathbb{N}$. In particular, as U_θ is a Bertrand numeration system, all states of $\mathcal{A}_{U_\theta}$ are final.

Remark 4. The canonical automaton of the numeration system U_θ has the following particular form (see [14, 17]). By Corollary 5, $e_\theta(1) = uv^\omega$. Draw the "frying-pan" automaton corresponding to uv^ω (a stick labelled by u followed by a circle labelled by v). For any state q, let a be the label of the outgoing transition. Then, for any $b \in A_{U_\theta}$, $b < a$, draw a transition from state q to the initial state.

Example 5. Fig. 4 indicates the canonical automaton for the three following numeration systems. The first one is $(2^n)_{n\in\mathbb{N}}$. The second is defined by $U_0 = 1$, $U_1 = 4$, $U_n = U_{n-1} + U_{n-2}$. The last one is the system $U_0 = 1$, $U_1 = 2$, $U_n = 3U_{n-1} - U_{n-2}$. See also Fig. 2.

3.3. Recognizability of $\nu_{B,U}$

The recognizability of the normalization is an important property. This implies that the addition and the substraction in $\mathbb{N}$ are U-recognizable, i.e., the sets $\{(x, y, z) \subseteq \mathbb{N}^3 \mid x + y = z\}$ and $\{(x, y, z) \subseteq \mathbb{N}^3 \mid x - y = z\}$ are U-recognizable. Indeed, to add two integers $x, y \in \mathbb{N}$, perform the addition on their normalized U-representations, letter by letter, without any carry. After, apply the normalization $\nu_{B,U}$ to the result, with $B = \{0, \ldots, 2c\}$ if $A_U = \{0, \ldots, c\}$.

Theorem 14 (Frougny and Solomyak [17]). *Let $U \in \mathscr{U}_\theta$. Let $B \subset \mathbb{Z}$ be a finite alphabet. Then the set $\{(u,v) \in B^* \times \mathscr{N}_U \mid \nu_{B,U}(u) = v\}$ is recognizable by a finite letter-to-letter automaton.*

Corollary 15. *Let $U \in \mathscr{U}_\theta$. Then the set $\{(x,y,z) \in \mathbb{N}^3 \mid x + y = z\}$ is U-recognizable.*

A direct proof of Corollary 15 is given in [7] for the particular numeration system U_θ of $\mathscr{U}_\theta$. We now give a proof of Theorem 14 which is simpler than the proof given in [17].

Proof of Theorem 14. Let $A_U = \{0, \ldots, c\}$, let $A_c = \{-c-1, \ldots, c+1\}$. To prove that $\{(u,v) \in B^* \times \mathscr{N}_U \mid \nu_{B,U}(u) = v\}$ is recognizable by a finite letter-to-letter automaton, we must prove that the sets $\mathscr{N}_U$ and

$$Z_{U,c} = \{a_i \cdots a_0 \in A_c^* \mid \pi_U(a_i \cdots a_0) = 0\}$$

are both recognizable (see [17]). We already know that $\mathscr{N}_U$ is recognizable (Theorem 8). To prove that $Z_{U,c}$ is recognizable, we will use Proposition 7 which remains true if A_U is replaced by A_c.

First recall that $Z_{U,c}$ is recognizable if and only if the syntactic congruence on A_c^*

$$u \sim_{Z_{U,c}} v \iff \{\forall w \in A_c^*, uw \in Z_{U,c} \Leftrightarrow vw \in Z_{U,c}\}$$

has finite index. Let $P(Z_{U,c})$ be the set of prefixes of the words of $Z_{U,c}$. Notice that $A_U^* \setminus P(Z_{U,c})$ is a particular class of $\sim_{Z_{U,c}}$.

Let k be the degree of the polynomial P_U. We define on $P(Z_{U,c})$ the equivalence relation R as follows

$$uRv \iff \{\forall n \in \{0, \ldots, k-1\},\ \pi_U(u0^n) = \pi_U(v0^n)\}.$$

Let uRv. Let us show that $u \sim_{Z_{U,c}} v$. By Lemma 13, the sequences $(\pi_U(u0^n))_{n \in \mathbb{N}}$ and $(\pi_U(v0^n))_{n \in \mathbb{N}}$ satisfy the same recurrence relation as $(U_n)_{n \in \mathbb{N}}$. Since they coincide on the first k values, they are equal. Consequently, for any $w \in A_c^*$

$$\begin{aligned} uw \in Z_{U,c} &\iff \pi_U(u0^{|w|}) + \pi_U(w) = 0 \\ &\iff \pi_U(v0^{|w|}) + \pi_U(w) = 0 \\ &\iff vw \in Z_{U,c}. \end{aligned}$$

These equivalences show that $u \sim_{Z_{U,c}} v$.

It remains to prove that R has finite index. This needs two steps.

(a) $\exists \beta,\ \forall u \in P(Z_{U,c}), |\pi_\theta(u)| < \beta$.

Since $u \in P(Z_{U,c})$, there exists $w \in A_c^*$ such that $uw \in Z_{U,c}$. By (5), we get

$$\begin{aligned}0 = \pi_U(u0^{|w|}) + \pi_U(w) &< \pi_\theta(u0^{|w|}) + \pi_\theta(w) + 2e \\ &< \pi_\theta(u)\theta^{|w|} + (c+1)\theta^{|w|} + 2e\theta^{|w|}.\end{aligned}$$

Hence,

$$\pi_\theta(u) > -c - 1 - 2e.$$

In the same way

$$\pi_\theta(u) < c + 1 + 2e.$$

We get the result with $\beta = c + 1 + 2e$.

(b) $\exists \beta'$, $\forall u \in P(Z_{U,c}), \forall n \in \{0,\dots,k-1\}, |\pi_U(u0^n)| < \beta'$.

By (5) and step (a), we have

$$\begin{aligned}|\pi_U(u0^n)| &\leqslant |\pi_U(u0^n) - \pi_\theta(u0^n)| + |\pi_\theta(u0^n)| \\ &< e + |\pi_\theta(u)|\theta^n < e + \beta\theta^n .\end{aligned}$$

Therefore $\beta' = e + \beta\theta^{k-1}$.

By step (b), the possible values of $\pi_U(u0^n)$, $u \in P(Z_{U,c})$, $n \in \{0,\dots,k-1\}$ are in finite number. Hence R has finite index. □

4. U-definable sets

In this section, we show that first-order formulæ are a useful and simple method to describe U-recognizable sets. This is our first characterization of U-recognizable sets.

Definition 4. Let U be a numeration system. We consider the *structure*

$$\langle \mathbb{N}, +, V_U \rangle,$$

where $+$ denotes the relation $\{(x,y,z) \in \mathbb{N}^3 \mid x + y = z\}$. The binary relation $\{(x,y) \in \mathbb{N}^2 \mid V_U(x) = y\}$ is defined as follows. Either $x = 0$ and $y = U_0 = 1$, or $x \neq 0$, $\rho_U(x) = a_i \cdots a_j 0^j$ with $a_j \neq 0$ and $y = U_j$. In other words, y is the smallest U_n appearing in the normalized U-representation of x with a non null coefficient.

Formulæ are inductively constructed from variables $x, y, \dots$ describing elements of $\mathbb{N}$, the equality $=$, the relations $+$, V_U, the connectives $\vee$, $\wedge$, $\neg$, $\rightarrow$, $\leftrightarrow$, and the quantifiers $\forall$, $\exists$ on variables.

Definition 5. Let U be a numeration system. Let $m \geqslant 1$ and $X \subseteq \mathbb{N}^m$. We say that X is *U-definable* if there exists a formula $\psi(x_1,\dots,x_m)$ of $\langle \mathbb{N}, +, V_U \rangle$ such that X is the set of m-tuples of $\mathbb{N}^m$ for which the formula ψ is true.

Example 6. The set $\{U_n \mid n \geqslant 0\}$ is U-definable in $\langle \mathbb{N}, +, V_U \rangle$ by the formula $\psi(x)$ equal to $V_U(x) = x$. The set $\{(U_n, U_{n+1}) \mid n \geqslant 0\}$ is definable by the following formula $\varphi(x, y)$

$$\psi(x) \wedge \psi(y) \wedge (x < y) \wedge (\forall z)(\psi(z) \wedge x < z \rightarrow y \leqslant z).$$

Notice that the order $x < y$ is definable by $(\exists z)(x + z = y) \wedge \neg(x = y)$.

Theorem 16 states that, for any numeration system $U \in \mathscr{U}_\theta$ with θ a Pisot number, a subset X of $\mathbb{N}^m$ is U-recognizable if and only if it is U-definable. The equivalence was already proved by Büchi [9] for numeration systems $(p^n)_{n \in \mathbb{N}}$ such that p is an integer greater than 1 (see also [8]). It has also been proved for the Bertrand numeration system U_θ of $\mathscr{U}_\theta$ [7].

Theorem 16. *Let $U \in \mathscr{U}_\theta$. Let $m \geqslant 1$ and $X \subseteq \mathbb{N}^m$. Then X is U-recognizable if and only if X is U-definable.*

The proof of this result is very close to the one given in [8, p. 207] for p-recognizable sets. It is strongly based on Theorem 8 and Corollary 15.

Proof. Suppose first that X is U-definable by some formula ψ of $\langle \mathbb{N}, +, V_U \rangle$. It is proved in [19] (see also [8]) that, to show that X is U-recognizable, it is enough to prove that the set $\mathbb{N}$ is U-recognizable and the relations $\{(x, y) \mid x = y\}$, $\{(x, y, z) \mid x + y = z\}$ and $\{(x, y) \mid V_U(x) = y\}$ are U-recognizable. Indeed, an automaton for X is then constructed by induction on the complexity of the formula ψ defining X.

By Theorem 8 and Corollary 15, $\mathbb{N}$ and $\{(x, y, z) \mid x + y = z\}$ are both U-recognizable. It is easy to construct a finite automaton for $\{(x, y) \mid x = y\}$: replace any label a by the label $\binom{a}{a}$ in the canonical automaton $\mathscr{A}_U$. It is also easy to find an automaton for $\{(x, y) \mid V_U(x) = y\}$. This is an automaton recognizing the following rational set (component x is up and component y is down)

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix}^* \begin{pmatrix} 0 \\ 1 \end{pmatrix} \cup \left\{ \begin{pmatrix} A_U \\ 0 \end{pmatrix}^* \begin{pmatrix} A_U \backslash 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix}^* \cap L \right\}$$

with $L = \{(u, v) \in \mathscr{N}_U \times A_U^* \mid |u| = |v|\}$. The set L is recognized by the canonical automaton $\mathscr{A}_U$ where any label a has been replaced by the labels $\binom{a}{b}$, $b \in A_U$.

Suppose now that X is U-recognizable. To show that X is U-definable, the idea is to describe by a formula of $\langle \mathbb{N}, +, V_U \rangle$, the behavior of a finite deterministic automaton $\mathscr{A}$ associated with X. This formula says that a word w is recognized by $\mathscr{A}$ if and only if there is a sequence $q, q', \ldots, q''$ of states, beginning with the initial state q, ending with some final state q'', and respecting the transitions imposed by w. The formula proposed in [26] (see also [8]) for numeration systems $(p^n)_{n \in \mathbb{N}}$, uses a particular relation $e(x, y)$ definable in $\langle \mathbb{N}, +, V_p \rangle$ and a coding of the sequence $q, q', \ldots, q''$.

For the numeration system U, the formula we use is exactly the formula given in [8, p. 210]. The relation $e(x, y)$ is still definable in $\langle \mathbb{N}, +, V_U \rangle$ [24]. However a difficulty appears with the coding of the sequence $q, q', \ldots, q''$. In [8], the proposed coding of the k states q_l of $\mathscr{A}$ was done by k-tuples of letters of A_U, respectively $(1,0,\ldots,0)$ for q_1, $(0,1,\ldots,0)$ for $q_2,\ldots$, and $(0,\ldots,0,1)$ for q_k. The sequence $q, q', \ldots, q''$ was then coded by a k-tuple $(u_1,\ldots,u_k)$ of words $u_l \in A_U^*$. This coding is here no longer convenient because any u_l must belong to $\mathscr{N}_U$. A correct coding is based on the following lemma.

Lemma 17. *Let $U \in \mathscr{U}_\theta$ and $w' \in \mathscr{N}_U \cap \mathscr{N}_{U_\theta}$. There exists $K \in \mathbb{N}$ such that for any $m \geqslant K$,*

$$w \in \mathscr{N}_U \cap \mathscr{N}_{U_\theta} \;\Rightarrow\; w0^m w' \in \mathscr{N}_U \cap \mathscr{N}_{U_\theta}.$$

Proof of the lemma. We first give a property of the numeration system U_θ. Let $e_\theta(1) = uu'^\omega$ be the θ-expansion of 1 (see Corollary 5). As $\theta > 1$, u begins with a letter distinct from 0. By definition of $e_\theta(1)$, $u' \notin 0^*$. As $\mathscr{N}_{U_\theta} = L(\theta)$, it follows by Proposition 3 that

$$v, v' \in \mathscr{N}_{U_\theta} \;\Rightarrow\; v0^n v' \in \mathscr{N}_{U_\theta} \tag{7}$$

for any $n \geqslant N = |uu'|$.

Next, let $t \in \mathscr{N}_{U_\theta}$ such that

$$\pi_\theta(t) \geqslant \pi_U(w') + 2\varepsilon \tag{8}$$

with ε as in Proposition 7. Define $K = N + |t|$.

Now, by (7), we have $w0^m w' \in \mathscr{N}_{U_\theta}$. To prove that $w0^m w' \in \mathscr{N}_U$, we proceed as in the proof of Proposition 9. Let $w = a_i \cdots a_0$, then for any $j, 0 \leqslant j \leqslant i$,

$$\begin{aligned}
\pi_U(a_j \cdots a_0 0^m w') &= \pi_U(a_j \cdots a_0 0^{m+|w'|}) + \pi_U(w') \\
&< \pi_\theta(a_j \cdots a_0 0^{m+|w'|}) + \varepsilon + \pi_U(w') && \text{(Proposition 7)} \\
&\leqslant \pi_\theta(a_j \cdots a_0 0^{m+|w'|}) + \pi_\theta(t) - \varepsilon && \text{(by (8))} \\
&= \pi_\theta(a_j \cdots a_0 0^{m+|w'|-|t|} t) - \varepsilon \\
&< \gamma_1 \theta^{j+1+m+|w'|} - \varepsilon && \text{((7) and Proposition 6)} \\
&< U_{j+1+m+|w'|}. && \text{(Proposition 7)}
\end{aligned}$$

It follows that $w0^m w' \in \mathscr{N}_U$ by Proposition 1. □

Using the previous lemma with $w' = 1$, it is clear that each component u_l of the k-tuple $(u_1,\ldots,u_k)$ belongs to $\mathscr{N}_U$ if the cycles of the automaton $\mathscr{A}$ have length greater than or equal to K.

Suppose that $\mathscr{A}$ has a cycle with length $L < K$. We are going to split the states of this cycle in a way to replace it by a cycle of length $2L$, without modifying the language recognized by $\mathscr{A}$. The splitting procedure is as follows. Denote the cycle by L states $q_0, q_1, \ldots, q_L = q_0$ and L transitions $T(q_0, a_1) = q_1, T(q_1, a_2) = q_2, \ldots, T(q_{L-1}, a_L) = q_L$. Any state $q_k, 0 \leqslant k \leqslant L$, is split into q'_k, q''_k with $q'_L = q''_0$ and $q''_L = q'_0$. If q_k is initial, then q'_k is initial. If q_k is final, thenn q'_k, q''_k are final. Any transition $T(p,b) = q_k$, with $p \neq q_{k-1}$, is replaced by $T(p,b) = q'_k$. Any transition $T(q_k,b) = p$, with $p \neq q_{k+1}$, is replaced by $T(q'_k,b) = p$ and $T(q''_k,b) = p$. Finally any transition $T(q_k,b,q_{k+1})$ is replaced by $T(q'_k,b,q'_{k+1})$ and $T(q''_k,b,q''_{k+1})$. □

5. U-automata and U-substitutions

In this section, we give the second and the third characterizations of U-recognizable sets of integers. We first associate with $\mathcal{N}_U$ a canonical substitution S_U which simulates the canonical automaton $\mathcal{A}_U$. Next, we introduce the concept of U-automaton. After that, we define the U-substitutions which simulate the U-automata. We also show how these new notions are connected to U-recognizability.

5.1. *Substitution for $\mathcal{N}_U$*

Definition 6. A *substitution S* is triple (f,g,b) such that

1. $f : B^* \to B^*$ and $g : B^* \to B^*$ are morphisms on some finite alphabet B,
2. $b \in B$ and $f(b) = bw$ for some $w \in B^*$,
3. $g(c) = c$ or $g(c) = \lambda$ for any $c \in B$ (λ is the empty word).

Any substitution generates a word ω_S equal to $g(f^\omega(b))$. In the sequel, ω_S is always an infinite word.

This notion of substitution is new. Classically, substitutions which generate infinite words, are pairs (f,b) (g is supposed to be the identity on B).

Definition 7. Let $U \in \mathcal{U}_0$. Let $\mathcal{A}_U = (Q,i,F,\tau)$ be the canonical automaton of U, with Q its set of states, i its initial state, F its final states and τ its transition function. We associate with $\mathcal{A}_U$ the *canonical substitution* $S_U = (f,g,i)$ defined by

$$\begin{aligned} f : Q &\to Q^* \\ q &\to \tau(q,0)\tau(q,1)\cdots\tau(q,a_q) \end{aligned}$$

with $a_q = \max\{a \in A_U \mid \tau(q,a) \text{ is defined}\}$ and

$$\begin{aligned} g : Q &\to Q^* \\ q &\to q \text{ if } q \in F, \quad q \to \lambda \text{ if } q \notin F. \end{aligned}$$

Such a substitution is defined in [14, 7] for the numeration system U_θ. In this case, g is always the identity.

Example 7. (1) Let $\theta = \frac{1}{2}(3+\sqrt{5})$ and $U \in \mathcal{U}_\theta$ defined by $U_0 = 1$, $U_1 = 2$ and $U_n = 3U_{n-1} - U_{n-2}$, $n \geqslant 2$. The canonical automaton $\mathcal{A}_U$ is the third automaton of Fig. 4. The corresponding canonical substitution S_U is defined by $f(1) = 112$, $f(2) = 12$ and $g(1) = 1$, $g(2) = \lambda$. The word generated by S_U is equal to $g(1121121211211212\ldots) = 1^\omega$.

(2) Let $U=(p^n)_{n\in\mathbb{N}}$ with p an integer greater than 1. The automaton $\mathscr{A}_U$ is reduced to one initial and final state q, and a loop labelled by $0,\dots,p-1$ on state q. Therefore, S_U is the triple (f,g,q) with $f(q)=q^p$ and $g(q)=q$.

Remark 5. In Definition 7, notice that f is well-defined. Indeed, Q is finite by Theorem 8, transition $\tau(q,b)$ exists for any $b\leqslant a_q$ by Proposition 2(3), and $f(i)=iw$ for some $w\in Q^*$ by Remark 2. Notice also that as $\mathscr{A}_U$ is trim, $f(q)$ is equal to λ for all states q without outgoing transition.

We have mentioned in Remarks 3 and 4 the particular form of the canonical automaton $\mathscr{A}_{U_\theta}$ of the numeration system U_θ: all states are final and have at least one outgoing transition. This means that for S_{U_θ}, $f(Q)\subset Q^+$ and g is the identity.

The next proposition shows that the canonical substitution S_U mimics the canonical automaton $\mathscr{A}_U$.

Proposition 18. *Let $U\in\mathscr{U}_\theta$. Let $\mathscr{A}_U=(Q,i,F,\tau)$ be the canonical automaton of U. Let $S_U=(f,g,i)$ be the canonical substitution and $\omega_{S_U}=\omega_0\omega_1\cdots\omega_n\cdots$ the word generated by S_U. Then for any $w\in\mathscr{N}_U$ and $x\in\mathbb{N}$ such that $x=\pi_U(w)$*

$$\tau(i,w)=\omega_x.$$

Proof. Let $n\in\mathbb{N}$. Let $A_n=\{w\in A_U^*\mid |w|=n \text{ and } \tau(i,w) \text{ exists}\}$. Enumerate A_n by increasing lexicographic ordering $\{w_0=0^n<w_1<\cdots<w_{k_n}\}$. We denote

$$\tau^{(n)}=\tau(i,w_0)\tau(i,w_1)\cdots\tau(i,w_{k_n})\in Q^+.$$

Let us prove by induction on n that

$$f^n(i)=\tau^{(n)}.$$

If $n=0$, then $f^0(i)=i=\tau(i,\lambda)=\tau^{(0)}$. Suppose that $n\geqslant 0$ and $f^n(i)=\tau^{(n)}$. Then $f^{n+1}(i)=f(\tau(i,w_0))f(\tau(i,w_1))\cdots f(\tau(i,w_{k_n}))$. By definition of f, for any $l\in\{0,\dots,k_n\}$, we get for $q_l=\tau(i,w_l)$

$$f(\tau(i,w_l))=f(q_l)=\tau(i,w_l0)\tau(i,w_l1)\cdots\tau(i,w_la_{q_l}).$$

As $w_l0<w_l1<\cdots<w_la_{q_l}$ and $w_la_{q_l}<w_{l+1}0$, it follows that $f^{n+1}(i)=\tau^{(n+1)}$.

Recall that the words w and $0w$ introduced in the initial state i lead to the same state of $\mathscr{A}_U$, if it exists (see Remark 2). Therefore, $f^n(i)$ is prefix of $f^{n+1}(i)$ for any $n\in\mathbb{N}$.

The same kind of property still holds after applying g. Let $n\in\mathbb{N}$ and $B_n=A_n\cap\mathscr{N}_U=\{v_0<v_1<\cdots<v_{l_n}\}$. Let

$$\tau_F^{(n)}=\tau(i,v_0)\tau(i,v_1)\cdots\tau(i,v_{l_n}).$$

Then

$$g(f^n(i))=g(\tau^{(n)})=\tau_F^{(n)},$$

because the automaton $\mathscr{A}_U$ recognizes $\mathscr{N}_U$. Moreover, $g(f^n(i))$ is prefix of $g(f^{n+1}(i))$.

Now, let $w \in \mathcal{N}_U$ and $x = \pi_U(w) \in \mathbb{N}$. If $n = |w|$, then

$$\omega_x = g(f^{\omega}(i))_x = g(f^n(i))_x = (\tau_F^{(n)})_x = \tau(i,w). \qquad \square$$

5.2. U-automata

Definition 8. Let $U \in \mathscr{U}_\theta$ and $\mathscr{A}_U$ its canonical automaton. We say that a finite deterministic automaton $\mathscr{A}$ is a *U-automaton* if there exists a morphism $\varphi : \mathscr{A} \to \mathscr{A}_U$.

Given two deterministic automata $\mathscr{A} = (Q,i,F,\tau)$ and $\mathscr{A}' = (Q',i',F',\tau')$ working on the same alphabet A, a *morphism* $\varphi : \mathscr{A} \to \mathscr{A}'$ is an onto map $\varphi : Q \to Q'$ such that

1. $\varphi(i) = i'$,
2. $\varphi(\tau(q,a)) = \tau'(\varphi(q),a)$, for all $q \in Q$ and $a \in A$,
3. there exists $\mathscr{F}$, $F \subseteq \mathscr{F} \subseteq Q$, such that $\varphi(\mathscr{F}) = F'$ and $\varphi^{-1}(F') = \mathscr{F}$.

In the second condition, if $\tau(q,a)$ is not defined, then $\tau'(\varphi(q),a)$ is not defined, and conversely. The notion of morphism given in [13, p. 38] is close to our definition. In [13], $\mathscr{F}$ is always equal to F.

Example 8. Let $U = (p^n)_{n \in \mathbb{N}}$ and $\mathscr{A}_U$ its canonical automaton. Any U-automaton is a finite deterministic automaton over the alphabet $\{0,\dots,p-1\}$ which is complete.

Proposition 19. *Let $\mathscr{A}_U$ be the canonical automaton of $U \in \mathscr{U}_\theta$. Then $\mathscr{A} = (Q,i,F,\tau)$ is a U-automaton if and only if there exists $\mathscr{F}$, $F \subseteq \mathscr{F} \subseteq Q$ such that $\mathscr{B} = (Q,i,\mathscr{F},\tau)$ recognizes $\mathcal{N}_U$.*

Proof. The proof uses standard properties of finite automata [13].

If there exists a morphism $\varphi : \mathscr{A} \to \mathscr{A}_U$, then the automaton $\mathscr{B}$ defined by φ and the canonical automaton $\mathscr{A}_U$ recognize the same language $\mathcal{N}_U$.

Conversely, suppose that the automaton $\mathscr{B}$ recognize $\mathcal{N}_U$, then there exists a morphism $\varphi : \mathscr{A} \to \mathscr{A}_U$ because $\mathscr{A}_U$ is the minimal automaton recognizing $\mathcal{N}_U$. $\square$

Theorem 20. *Let $U \in \mathscr{U}_\theta$. Let $X \subseteq \mathbb{N}$. Then X is U-recognizable if and only if $L = 0^*\rho_U(X)$ is recognizable by a U-automaton.*

Proof. Indeed, a U-automaton for L can be constructed from the congruence equal to the intersection of the two syntactic congruences $\sim_{\mathcal{N}_U}$ and $\sim_L$. $\square$

5.3. U-substitutions

Definition 9. Let $U \in \mathscr{U}_\theta$. Let S_U be the canonical substitution. We say that a substitution $S = (f,g,b)$ is a *U-substitution* if there exists a morphism $\varphi : S \to S_U$.

Given two substitutions $S = (f,g,b)$ and $S' = (f',g',b')$ defined on the alphabets B and B' respectively, a *morphism* $\varphi : S \to S'$ is a surjective morphism $\varphi : B \to B'$ such that

1. $\varphi(b) = b'$,

2. $\varphi(f(c)) = f'(\varphi(c))$, for all $c \in B$,
3. $\varphi(g(c)) = g'(\varphi(c))$, for all $c \in B$.

The notion of U-substitution is defined in [14, 7] for classical substitutions (such that g is the identity).

Example 9. In Example 7, the canonical substitution S_U of the numeration system $U = (p^n)_{n\in\mathbb{N}}$ is described. In this case, any U-substitution is a triple (f, g, b) such that $f(B) \subset B^p$ and g is the identity on B. These particular substitutions are the *uniform tag systems* of [12] or the *p-substitutions* of [8].

We are now able to define sets of integers generated by U-substitution. This naturally generalizes sets generated by uniform tag systems [12, 8]. We will prove that for any $U \in \mathscr{U}_\theta$, the sets generated by U-substitutions are exactly the U-recognizable sets. This equivalence has been proved in [14, 7] for the numeration system U_θ.

Definition 10. Let $U \in \mathscr{U}_\theta$. Let $X \subseteq \mathbb{N}$. We say that X is *generated by U-substitution* if there exist a U-substitution $S = (f, g, b)$ on the alphabet B and a map $h : B \to \{0, 1\}$ such that

$$x \in X \Leftrightarrow (h(\omega_S))_x = 1.$$

Theorem 21. *Let $U \in \mathscr{U}_\theta$. Let $X \subseteq \mathbb{N}$. Then X is U-recognizable if and only if X is generated by U-substitution.*

Proof. By Theorem 20 we have to prove that $L = 0^*\rho_U(X)$ is recognizable by a U-automaton if and only if X is generated by U-substitution.

(a) Let $\mathscr{A} = (Q, i, F, \tau)$ be a U-automaton for L and $\varphi : \mathscr{A} \to \mathscr{A}_U$ the related morphism. By Proposition 19, there exists $\mathscr{F}$, $F \subseteq \mathscr{F} \subseteq Q$, such that $\mathscr{B} = (Q, i, \mathscr{F}, \tau)$ recognizes $\mathscr{N}_U$. Let $\mathscr{A}_U = (Q', i', F', \tau')$ be the canonical automaton and $S_U = (f', g', i')$ be the canonical substitution.

A substitution $S = (f, g, i)$ simulating $\mathscr{B}$ is constructed as in Definition 7: we define

$$\begin{aligned} f : Q &\to Q^* \\ q &\to \tau(q,0)\tau(q,1)\cdots\tau(q,a_q) \end{aligned}$$

with $a_q = \max\{a \in A_U \mid \tau(q, a) \text{ is defined}\}$ and

$$\begin{aligned} g : Q &\to Q^* \\ q &\to q \text{ if } q \in \mathscr{F}, \quad q \to \lambda \text{ if } q \notin \mathscr{F}. \end{aligned}$$

This substitution S is a U-substitution since the morphism $\varphi : \mathscr{A} \to \mathscr{A}_U$ can also be considered as a morphism $\varphi : S \to S_U$. Let $\omega_S = \omega_0\omega_1 \ldots \omega_n \ldots$ be the word generated by S. We get the property of Proposition 18:

$$\forall w \in \mathscr{N}_U, \quad x = \pi_U(w) \in \mathbb{N}, \qquad \tau(i, w) = \omega_x. \tag{9}$$

The proof is identical, due to Proposition 19.

It remains to show that X is generated by U-substitution. Let S as above, let

$$h : Q \to \{0,1\}$$
$$q \to 1 \text{ if } q \in F, \quad q \to 0 \text{ if } q \notin F.$$

Then, by (9)

$$x \in X \Leftrightarrow \tau(i, \rho_U(x)) \in F \Leftrightarrow \omega_x \in F \Leftrightarrow h(\omega_x) = 1.$$

(b) Conversely, suppose that X is generated by U-substitution. One constructs a U-automaton for L, by following step (b) backwards. □

Example 10. Consider the Fibonacci numeration system U_ϕ and the U_ϕ-recognizable set $X = \{x \in \mathbb{N} \mid \rho_{U_\phi}(x)$ has an even number of 1's$\}$. The canonical automaton $\mathscr{A}_{U_\phi}$ is depicted on Fig. 2, an automaton $\mathscr{A}$ for X is given on Fig. 3. One easily verifies that $\mathscr{A}$ is a U_ϕ-automaton. The canonical substitution $S_{U_\phi} = (f', g', 1)$ is defined by $f'(1) = 12$, $f'(2) = 1$ and $g'(1) = 1$, $g'(2) = 2$. The substitution $S = (f, g, 1)$ associated with $\mathscr{A}$ is such that $f(1) = 12$, $f(2) = 3$, $f(3) = 34$, $f(4) = 1$ and $g(1) = 1$, $g(2) = 2$, $g(3) = 3$, $g(4) = 4$. Clearly S is a U_ϕ-substitution. One see that X is generated by U_ϕ-substitution, using S and h such that $h(1) = 1$, $h(2) = 0$, $h(3) = 0$, $h(4) = 1$:

$$h(\omega_S) = h(1233434134112\ldots) = 1000101101110\ldots$$

5.4. Higher dimensions

Theorems 20 and 21 only concern sets $X \subseteq \mathbb{N}$ of integers. It remains true for sets $X \subseteq \mathbb{N}^m$, $m \geqslant 2$, because U-automata and U-substitutions exist in higher dimensions. We will briefly sketch the ideas, skipping the details and proofs. See [8] for U-substitutions related to the numeration systems $U = (p^n)_{n \in \mathbb{N}}$, in any dimension $m \geqslant 1$.

Definition 11. Let $U \in \mathscr{U}_\theta$. Let $m \geqslant 1$. By Theorem 8, the set $\mathbb{N}^m$ is U-recognizable. We denote by $\mathscr{A}_U^m$ the trim minimal automaton recognizing $0^* \rho_U(\mathbb{N}^m)$. We denote by S_U^m the substitution which simuates $\mathscr{A}_U^m$. This substitution is constructed exactly as in Definition 7 (the alphabet of $\mathscr{A}_U^m$ is simply m-tuples of letters of A_U).

Example 11. Let $m = 2$ and U_ϕ the Fibonacci numeration system. The automaton $\mathscr{A}_{U_\phi}^2$ is given on Fig. 5.

The substitution $S_{U_\phi}^2 = (f, g, 1)$ is defined by

$$f(1) = \begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix}, \quad f(2) = (1), \quad f(3) = (1\ 4), \quad f(4) = \begin{pmatrix} 3 \\ 1 \end{pmatrix},$$

$$g(1) = 1, \quad g(2) = 2, \quad g(3) = 3, \quad g(4) = 4.$$

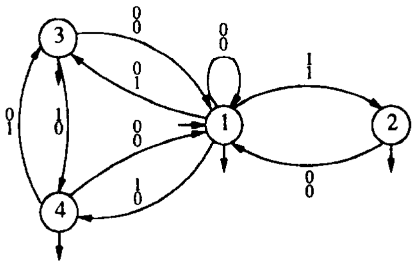

Fig. 5. The canonical automaton $\mathscr{A}^2_{U_\Phi}$ of the Fibonacci numeration system.

The array $\left(\begin{smallmatrix} 3 & 2 \\ 1 & 4 \end{smallmatrix}\right)$ indicates the states reached by the array of letters

$$\begin{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} & \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ \begin{pmatrix} 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 1 \\ 0 \end{pmatrix} \end{pmatrix}.$$

The iteration of f on 1 gives rise to larger and larger arrays

$$(1) \xrightarrow{f} \begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix} \xrightarrow{f} \begin{pmatrix} 1 & 4 & 1 \\ 3 & 2 & 3 \\ 1 & 4 & 1 \end{pmatrix} \xrightarrow{f} \begin{pmatrix} 3 & 2 & 3 & 3 & 2 \\ 1 & 4 & 1 & 1 & 4 \\ 1 & 4 & 1 & 1 & 4 \\ 3 & 2 & 3 & 3 & 2 \\ 1 & 4 & 1 & 1 & 4 \end{pmatrix} \xrightarrow{f} \cdots$$

Let $U \in \mathscr{U}_\theta$ and $m \geqslant 1$. Thanks to the example above, the reader can imagine what are U-automata and U-substitutions in dimension m and how U-substitutions simulate U-automata. Such definitions are based on morphisms onto the automaton $\mathscr{A}^m_U$ and the substitution S^m_U, respectively.

Theorem 22. *Let $m \geqslant 1$ and $U \in \mathscr{U}_\theta$. Let $X \subseteq \mathbb{N}^m$. Then X is U-recognizable if and only if $L = 0^*\rho_U(X)$ is recognizable by a U-automaton if and only if X is generated by U-substitution.*

6. Independence on the initial values

Theorem 8 and Corollary 15 state that the sets $\mathbb{N}$ and $\{(x,y,z) \in \mathbb{N}^3 \mid x + y = z\}$ are U-recognizable for any $U \in \mathscr{U}_\theta$. In this section we will prove more: a set $X \subseteq \mathbb{N}^m$ is U-recognizable if and only if X is U_θ-recognizable. In other words, the existence of an automaton for X only depends on the Pisot number θ, i.e., the recurrence relation given by its minimal polynomial P_θ. This is independent on the initial values $U_0, \ldots, U_{k-1}$ which define a particular numeration system U of the class $\mathscr{U}_\theta$ (k is the degree of P_θ). Hence, we could say that X is θ-recognizable instead of U-recognizable, with $U \in \mathscr{U}_\theta$.

Theorem 23. *Let $\mathscr{U}_\theta$. Let $m \geqslant 1$ and $X \subseteq \mathbb{N}^m$. Then X is U-recognizable for some $U \in \mathscr{U}_\theta$ if and only if X is U-recognizable for all $U \in \mathscr{U}_\theta$.*

Let us begin with a lemma.

Lemma 24. *Let $U, U' \in \mathscr{U}_\theta$. Let k be the degree of P_θ. Then there exist $\beta \in \mathbb{Z}$, $\beta \neq 0$ and $\beta_0, \dots \beta_{k-1} \in \mathbb{Z}$ such that*

$$\beta U'_n = \beta_0 U_n + \cdots + \beta_{k-1} U_{n+k-1} \quad \forall n \geqslant 0.$$

Proof. (a) Let $P_\theta(X) = X^k - d_{k-1}X^{k-1} - \cdots - d_0$. Then for any $n \geqslant k$

$$U_n = d_{k-1}U_{n-1} + \cdots + d_0 U_{n-k}, \tag{10}$$

$$U'_n = d_{k-1}U'_{n-1} + \cdots + d_0 U'_{n-k}. \tag{11}$$

With (10) and (11) we associate the $k \times k$ recurrence matrix

$$\mathscr{R} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ d_0 & d_1 & & \dots & d_{k-1} \end{pmatrix}.$$

(b) Consider the following $k \times k$ matrix

$$\mathscr{M} = \begin{pmatrix} U_0 & U_1 & \dots & U_{k-1} \\ U_1 & U_2 & \dots & U_k \\ \vdots & & & \vdots \\ U_{k-1} & U_k & \dots & U_{2k-2} \end{pmatrix}.$$

Let us prove that

$$\det(\mathscr{M}) \neq 0.$$

Assume the contrary: there exists $l \in \{1, \dots, k-1\}$ such that the lth column is linearly dependent of the previous ones. Hence there exist $\eta_0, \dots, \eta_{l-1} \in \mathbb{Q}$ such that

$$\begin{cases} U_l & = \eta_{l-1}U_{l-1} & + & \cdots & + \eta_0 U_0 \\ U_{l+1} & = \eta_{l-1}U_l & + & \cdots & + \eta_0 U_1 \\ \vdots & & & & \vdots \\ U_{l+k-1} & = \eta_{l-1}U_{l+k-2} & + & \cdots & + \eta_0 U_{k-1} \end{cases}$$

Applying (10) on each column of the previous system, we get

$$U_n = \eta_{l-1}U_{n-1} + \cdots + \eta_0 U_{n-l} \quad \forall n \geqslant l. \tag{12}$$

We denote by $q(X)$ the polynomial $X^l - \eta_{l-1}X^{l-1} - \cdots - \eta_0 \in \mathbb{Q}[X]$ with degree $l < k$.

Let us show that $q(\theta) = 0$. We will get the contradiction because the minimal polynomial P_θ of θ has degree k. Let $\varepsilon > 0$. By Proposition 7, for any $n \geqslant M_\varepsilon$

$$|U_n - \gamma_1\theta^n| < \varepsilon.$$

Hence, for $n \geqslant M_\varepsilon + l$, by (12)

$$|\gamma_1\theta^{n-l}q(\theta)| = |\gamma_1(\theta^n - \eta_{l-1}\theta^{n-1} - \cdots - \eta_0\theta^{n-l})| < \varepsilon\left(1 + \sum_{i=0}^{l-1}|\eta_i|\right).$$

So $\gamma_1\theta^{n-l}q(\theta)$ is bounded independently of n, i.e., $q(\theta) = 0$.

(c) Consider the following system of linear equations

$$\begin{pmatrix} U'_0 \\ U'_1 \\ \vdots \\ U'_{k-1} \end{pmatrix} = \mathscr{M} \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{k-1} \end{pmatrix}$$

Since $\det \mathscr{M} \neq 0$, the system has one solution $(\alpha_0, \ldots, \alpha_{k-1}) \in \mathbb{Q}^k$. By (10) and (11) we get for any $n \in \mathbb{N}$

$$\begin{pmatrix} U'_n \\ \vdots \\ U'_{n+k-1} \end{pmatrix} = \mathscr{R}^n \begin{pmatrix} U'_0 \\ \vdots \\ U'_{k-1} \end{pmatrix} = \mathscr{R}^n \mathscr{M} \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_{k-1} \end{pmatrix}$$

$$= \begin{pmatrix} U_n & \cdots & U_{n+k-1} \\ \vdots & & \vdots \\ U_{n+k-1} & \cdots & U_{n+2k-2} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_{k-1} \end{pmatrix}.$$

Then $U'_n = \alpha_0 U_n + \cdots + \alpha_{k-1}U_{n+k-1}$, for any $n \geqslant 0$. The thesis follows with $\beta \in \mathbb{Z}\backslash\{0\}$ the lcm of $\alpha_0, \ldots, \alpha_{k-1}$ and $\beta_l = \beta\alpha_l \in \mathbb{Z}$, $l \in \{0, \ldots, k-1\}$. $\square$

Here is the first proof of Theorem 23. It uses local automata [3]. For readers who are fond of purely logical proofs, we give afterwards a second proof.

We first prove the next lemma.

Lemma 25. *Let $U, U' \in \mathscr{U}_\theta$. Let k, β as in Lemma 24. Then there exist a finite alphabet $B \subset \mathbb{Z}$ and a function*

$$\phi : A^*_{U'} \to B^*, \qquad u \to v = \phi(u)$$

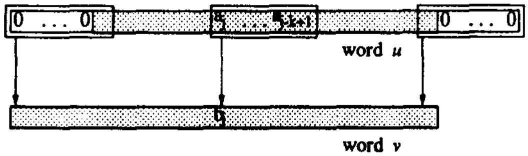

Fig. 6. The k-local map ϕ.

such that $\beta\pi_{U'}(u) = \pi_U(v)$ *and* $|v| = |u| + k - 1$. *Moreover, the set*

$$\{(u0^{k-1}, v) \in A_{U'}^* \times B^* \mid v = \phi(u)\}$$

is recognizable by a letter-to-letter automaton.

Proof. Let $u = a_i \cdots a_0 \in A_{U'}^* = \{0, \cdots, c'\}^*$. Add coefficients $a_{i+k-1} = \cdots = a_{i+1} = 0$ and $a_{-1} = \cdots = a_{-k+1} = 0$. Then, by Lemma 24, ($U'_{-1}, \ldots, U'_{-k+1}$ are defined arbitrarily)

$$\beta\pi_{U'}(u) = \beta \sum_{j=0}^{i} a_j U'_j = \beta \sum_{j=-k+1}^{i+k-1} a_j U'_j = \sum_{j=-k+1}^{i+k-1} a_j \left(\sum_{l=0}^{k-1} \beta_l U_{j+l} \right).$$

In the latter summation, the coefficients of $U_{i+2k-2}, \ldots, U_{i+k}$ and $U_{-1}, \ldots, U_{-k+1}$ are null. It follows that

$$\beta\pi_{U'}(u) = \sum_{j=0}^{i+k-1} \left(\sum_{l=0}^{k-1} a_{j-l}\beta_l \right) U_j = \sum_{j=0}^{i+k-1} b_j U_j = \pi_U(v) \tag{13}$$

with $v = b_{i+k-1} \cdots b_0 \in B^*$. The alphabet $B \subset \mathbb{Z}$ is defined by $B = \{-b, \ldots, b\} \subset \mathbb{Z}$ such that $b = c' \sum_{l=0}^{k-1} |\beta_l|$. Then there exists a map $\phi : A_{U'}^* \to B^*$ such that $\phi(u) = v$, $\beta\pi_{U'}(u) = \pi_U(v)$ and $|v| = |u| + k - 1$.

The previous computation shows that ϕ is a k-local map [3, p. 95], because letter b_j of $v = \phi(u)$ is computed as a function of the k letters $a_j, \ldots, a_{j-k+1}$ of u (see (13) and Fig. 6). Hence [3, p. 98], the set

$$\{(u0^{k-1}, v) \in A_{U'}^* \times B^* \mid v = \phi(u)\}$$

is recognizable by the universal k-local automaton $\mathscr{C} = (Q, i, F, \tau)$ in the following way:

1. Q is the set of all the k-tuples of letters of $A_{U'}$,
2. for any $q = (c_k, \ldots, c_1) \in Q$ and $c_0 \in A_{U'}$, $\tau(q, (c_0, d)) = q'$ such that $q' = (c_{k-1}, \ldots, c_1, c_0)$ and $d = \sum_{l=0}^{k-1} c_{k-1-l}\beta_l \in B$,
3. the initial state i is $(0, \ldots, 0)$,
4. F is the set $\{(c, 0, \ldots, 0) \mid c \in A_{U'}\}$. $\square$

Proof of Theorem 23. Now let $U, U' \in \mathscr{U}_\theta$. Let $X \subseteq \mathbb{N}$ be a U'-recognizable set. We are going to prove that X is also U-recognizable. We use the notations of Lemmas 24 and 25.

Since X is U'-recognizable, the set $L' = 0^*\rho_{U'}(X)$ is recognizable. Using the automaton of Lemma 25 on the recognizable set $L'0^{k-1} \subseteq A_{U'}^*$, there exists a recognizable set $L \subseteq B^*$ such that $\pi_U(L) = \beta\pi_{U'}(L') = \beta X$. Using Theorem 14, the set $L'' = \nu_{B,U}(L) \subseteq \mathcal{N}_U$ is recognizable. Again $\pi_U(L'') = \beta X$. It follows that βX is U-recognizable. Let $\psi(x_1, \ldots, x_m)$ be a formula of $\langle \mathbb{N}, +, V_U \rangle$ defining βX (see Theorem 16), then the formula $\varphi(y_1, \ldots, y_m)$

$$(\exists x_1)\ldots(\exists x_m)(\beta y_1 = x_1) \wedge \ldots \wedge (\beta y_m = x_m) \wedge \psi(x_1, \ldots, x_m)$$

defines X. This shows that X is U-recognizable.

The proof is easily adapted to subsets X of $\mathbb{N}^m$ with $m \geqslant 2$. Lemma 25 clearly remains true if u and v are m-tuples of words with the same length respectively. □

Second proof of Theorem 23. We give the proof for subsets X of $\mathbb{N}$ only. It is not difficult to generalize it to sets $X \subseteq \mathbb{N}^m$ such that $m \geqslant 2$.

Let $U, U' \in \mathscr{U}_\theta$. We are going to prove that if $X \subseteq \mathbb{N}$ is U'-recognizable, then it is U-recognizable. We use the notations of Lemma 24. We also use the constant M of Proposition 9 such that it works for both numeration systems U and U'.

Any $x \in X$ is uniquely written as

$$x = v(x) + w(x)$$

in the following way. Let $vw \in \mathcal{N}_{U'}$ such that $\pi_{U'}(vw) = x$ and $|w| = M$. Then

$$x = \pi_{U'}(v0^M) + \pi_{U'}(w) = v(x) + w(x).$$

As $|w| = M$, $w(x) \in \{0, \ldots, U'_M - 1\}$. It follows that X is equal to the disjoint union

$$X = \sum_{r=0}^{U'_M - 1} (X_r + r).$$

where $X_r = \{v(x) \mid x \in X, w(x) = r\}$.

Let us show that any X_r is U-recognizable. Denote by $L_r \subseteq \mathcal{N}_{U'}$ the set $0^*\rho_{U'}(X_r)$. By Definition of X_r, L_r is a recognizable set over the alphabet $A_{U'}$ and all its elements are of the form $v0^n$ with $n \geqslant M$. By Proposition 9, L_r is also a recognizable set over the alphabet A_U. Let $Y_r = \pi_U(L_r)$ be the corresponding U-recognizable set.

Let $v0^n = a_i \cdots a_n 0^n \in L_r$. Hence by Lemma 24

$$\beta \sum_{j=n}^{i} a_j U'_j = \sum_{l=0}^{k-1} \beta_l \left(\sum_{j=n}^{i} a_j U_{j+l} \right) = \sum_{l=0}^{k-1} \beta_l \mathscr{S}^l \left(\sum_{j=n}^{i} a_j U_j \right),$$

where $\mathscr{S}$ is the successor function such that $\mathscr{S}(U_j) = U_{j+1}$. In other words, for $x = \pi_{U'}(v0^n) \in X_r$ and $y = \pi_U(v0^n) \in Y_r$, we get

$$\beta x = \sum_{l=0}^{k-1} \beta_l \mathscr{S}^l(y). \tag{14}$$

Consequently, since Y_r is U-definable (see Theorem 16), since function $\mathscr{S}$ (restricted to $y \in \mathbb{N}$ such that $V_U(y) \geqslant M$) is definable in $\langle \mathbb{N}, +, V_U \rangle$, formula (14) shows that X_r is U-definable.

This concludes the proof because $X = \sum_{r=0}^{U'_M - 1}(X_r + r)$ is also U-definable. □

7. Conclusions

In this paper, we have studied classes $\mathscr{U}_\theta$ of linear numeration systems U whose characteristic polynomial is the minimal polynomial of a Pisot number θ. Under this condition, the set $\mathscr{N}_U$ and the normalization $v_{B,U}$ are recognizable by a finite automaton.

With any system $U \in \mathscr{U}_\theta$, we have associated a canonical automaton $\mathscr{A}_U$ equal to the minimal automaton of $\mathscr{N}_U$, and a canonical substitution S_U which mimics the behavior of $\mathscr{A}_U$. We have introduced the concept of U-automaton and U-substitution, which are "splittings" of $\mathscr{A}_U$ and S_U, respectively.

Theorem 21 states the equivalence between U-automata and U-substitutions for all numeration systems $U \in \mathscr{U}_\theta$. The proof only needs Proposition 2 and Theorem 8. This means that Theorem 21 still holds for any numeration system U such that $\mathscr{N}_U$ is recognizable. This also points out that U-substitutions are "disguised" U-automata.

Theorem 16 is in a certain way more interesting. Through formulæ of $\langle \mathbb{N}, +, V_U \rangle$, it shows that U-recognizability is built on the recognizability of both $\mathscr{N}_U$ and $v_{B,U}$. Notice the usefulness of Lemma 17 whose proof heavily depends on Hypothesis 1. Therefore, the logical characterization of U-recognizable sets remains true for numeration systems U such that $\mathscr{N}_U$ and $v_{B,U}$ are recognizable, and Lemma 17 holds.

Theorem 23 allows to transfer certain properties of the Bertrand numeration system U_θ to all the numeration systems $U \in U_\theta$. For instance, in [24], Cobham's theorem is generalized to two numeration systems, one of which is a usual base $p \geqslant 2$ and the other is a Bertrand numeration system U_θ. This is still true for two Bertrand numeration systems U_θ, U_ϕ with θ and ϕ two multiplicatively independent Pisot numbers [18, 6]. As a matter of fact, this generalization holds for any pair (U, U') of numeration systems such that $U \in U_\theta$ and $U' \in U_\phi$.

References

[1] J.-P. Allouche, q-regular sequences and other generalizations of q-automatic sequences, Proc. Latin '92, Lecture Notes in Computer Science, Vol. 583 (1992) 15–23.

[2] J.-P. Allouche, E. Cateland, W.J. Gilbert, H.-O. Peitgen, J.O. Shallit and G. Skordev, Automatic maps in exotic numeration systems, *Math. Systems Theory* (1996) to appear.

[3] M.P. Béal, *Codage symbolique*, Masson (1993).

[4] A. Bertrand-Mathis, Comment écrire les nombres entiers dans une base qui n'est pas entière, *Acta Math. Acad. Sci. Hungary* **54** (1989) 237–241.

[5] A. Bertrand-Mathis, Développement en base θ, répartition modulo un de la suite $(x\theta^n)_{n \geqslant 0}$, langages codés et θ-shift, *Bull. Soc. Math. France* **114** (1986) 271–323.

[6] A. Bès, An extension of the Cobham-Semenov theorem, preprint (1996).

[7] V. Bruyère and G. Hansel, Recognizable sets of numbers in nonstandard bases, Lecture Notes in Computer Science, Vol. 911 (1995) 167–179.
[8] V. Bruyère, G. Hansel, C. Michaux and R. Villemaire, Logic and p-recognizable sets of integers, *Bull. Belg. Math. Soc.* **1** (1994) 191–238.
[9] J.R. Büchi, Weak second-order arithmetic and finite automata, *Z. Math. Logik Grundlag. Math.* **6** (1960) 66–92.
[10] G. Christol, T. Kamae, M. Mendès France and G. Rauzy, Suites algébriques, automates et substitutions, *Bull. Soc. Math. France* **108** (1980) 401–419.
[11] A. Cobham, On the base-dependence of sets of numbers recognizable by finite automata, *Math. Systems Theory* **3** (1969) 186–192.
[12] A. Cobham, Uniform tag systems, *Math. Systems Theory* **6** (1972) 164–192.
[13] S. Eilenberg, *Automata, Languages and Machines*, Vol. A (Academic Press, New York, 1974).
[14] S. Fabre, Substitutions et indépendence des systèmes de numération, *Thèse*, Université Aix-Marseille II (1992).
[15] A.S. Fraenkel, Systems of numeration, *Amer. Math. Monthly* **92** (1985) 105–114.
[16] C. Frougny, Representations of numbers and finite automata, *Math. Systems Theory* **25** (1992) 37–60.
[17] C. Frougny and B. Solomyak, On representations of integers in linear numeration systems, in: M. Pollicot and K. Schmidt, eds., *Ergodic Theory of $\mathbb{Z}^d$-Actions*, London Math. Soc. Lecture Note, Vol. 228 (1996) 345–368.
[18] G. Hansel, Sur la syndéticité d'une partie (θ, θ')-reconnaissable de $\mathbb{N}$, manuscript (1995).
[19] B.R. Hodgson, Décidabilité par automate fini, *Ann. Sci. Math. Québec* **7** (1983) 39–57.
[20] M. Hollander, Greedy numeration systems and regularity, preprint (1995)
[21] N. Loraud, β-shift, systèmes de numération et automates, preprint (1995).
[22] W. Parry, On the β-expansions of real numbers, *Acta Math. Acad. Sci. Hungar.* **11** (1960) 401–416.
[23] D. Perrin, Finite automata, in: J. Van Leeuwen, ed., *Handbook of Theoretical Computer Science*, Vol. B (Elsevier, Amsterdam, 1990) 2–57.
[24] F. Point and V. Bruyère, On the Cobham–Semenov theorem, *Theory Comput. Systems* **30** (1997) 197–220.
[25] J. Shallit, Numeration systems, linear recurrences and regular sets, Lecture Notes in Computer Science, Vol. 623 (1992) 89–100.
[26] R. Villemaire, The theory of $\langle \mathbb{N}, +, V_k, V_l \rangle$ is undecidable, *Theoret. Comput. Sci.* **106** (1992) 337–349.