LOGIC AND p-RECOGNIZABLE SETS OF INTEGERS

Véronique Bruyère* Georges Hansel Christian Michaux*†
Roger Villemaire‡

Abstract

We survey the properties of sets of integers recognizable by automata when they are written in p-ary expansions. We focus on Cobham's theorem which characterizes the sets recognizable in different bases p and on its generalization to $\mathbb{N}^m$ due to Semenov. We detail the remarkable proof recently given by Muchnik for the theorem of Cobham-Semenov, the original proof being published in Russian.

1 Introduction

This paper is a survey on the remarkable theorem of A. Cobham stating that the only sets of numbers recognizable by automata, independently of the base of representation, are those which are ultimately periodic. The proof given by Cobham, even if it is elementary, is rather difficult [15]. In his book [24], S. Eilenberg proposed as a challenge to find a more reasonable proof. Since this date, some researchers found more comprehensible proofs for subsets of $\mathbb{N}$, and more generally of $\mathbb{N}^m$. The more recent works demonstrate the power of first-order logic in the study of recognizable sets of numbers [54, 49, 50].

One aim of this paper is to collect, from Büchi to Muchnik's works [9, 54], all the base-dependence properties of sets of numbers recognizable by finite automata, with some emphasis on logical arguments. It contains several examples and some logical proofs. In particular, the fascinating proof recently given by A. Muchnik is detailed

*This work was partially supported by ESPRIT-BRA Working Group 6317 *ASMICS* and Co-operation Project C.G.R.I.-C.N.R.S. *Théorie des Automates et Applications*.

†Partially supported by a F.N.R.S. travel grant. The author thanks for its hospitality the Department of Mathematics and Computer Science of UQAM and the Centre Ricerca Matematica at Barcelona.

‡The author thanks for its hospitality the team of Mathematical Logic at Paris 7 and also the Department of Mathematics and Computer Science of UQAM for financial support.

Received by the editors November 1993, revised February 1994.
Communicated by M. Boffa.
AMS Mathematics Subject Classification : 11B85, 03D05, 68R15, 68Q45.
Keywords : infinite words, p-recognizable sequences, finite automata, first-order definability, formal power series.

and simplified. It states Cobham's theorem and the generalization of Semenov to $\mathbb{N}^m$. The paper also contains bibliographic notes on the history of the results.

This survey is born of several lectures given by the authors, respectively V. Bruyère, 29 April 93 - Université Libre de Bruxelles, G. Hansel, 26 April and 3 May 93 - Université Paris 6, C. Michaux, 2 September 92 - Université de Mons-Hainaut and R. Villemaire, 23 November 92 - University McGill, 3 December 92 - Université UQAM. It addresses readers accustomed with automata, but less familiar with first-order logic. It may be read in different ways, depending on the interests of the reader.

The paper is arranged in the following way. It begins with a brief section on automata and a lesson of logic to make the reader more familiar with first-order structures, definable sets and decidable theories.

Next, Section 4 deals with p-recognizable sets of numbers, i.e., sets of numbers whose p-ary expansions are recognizable by a finite automaton. Various characterizations of p-recognizability are related in Theorem 4.1 : iterated uniform morphisms, algebraic formal power series, and definability by first-order formulae. Section 4 is centered around these four models of p-recognizability. It begins and ends with two generic examples. It also contains some bibliographic notes related to Theorem 4.1, as well as notes about the automata associated with p-recognizable sets.

Section 5 is in the same spirit but for p-recognizable subsets of $\mathbb{N}^m$. The four models still hold and are again equivalent (Theorem 5.1).

Among the different characterizations of p-recognizability, Section 6 emphasizes the logical one. The currently simplest proof of this equivalence is given, following the references [40, 74]. At the origin, it was proved by R. Büchi in his paper [9]. Some corollaries of decidability and non p-recognizability are easily derived, together with a powerful tool for operations preserving p-recognizability (Corollaries 6.2, 6.3 and 6.4).

Next, Section 7 studies the dependence of p-recognizability on the base of representation. In particular it contains Cobham's theorem (Theorem 7.7). It shows that there are essentially three kinds of subsets of $\mathbb{N}^m$: the sets recognizable in every base p, the sets recognizable in certain bases only, and the sets recognizable in no base. The first class is quite restricted, as it is limited to the rational sets of the monoid $(\mathbb{N}^m, +)$. When $m = 1$, it is precisely the ultimately periodic sets. It is rather easy to see that ultimately periodic sets are p-recognizable for every integer $p \geqslant 2$, but the converse relies on the deep theorem of Cobham. As for p-recognizable sets, sets recognizable in every base are characterized in various ways (Theorems 7.3 and 7.4), including a logical characterization and a fine definability criterion found recently by A. Muchnik [54]. Muchnik's criterion is at the heart of his remarkable proof of Cobham's theorem over $\mathbb{N}^m$, it also allows one to decide whether a p-recognizable set of $\mathbb{N}^m$ is recognizable in every base (Proposition 7.6).

Section 8 is devoted to Muchnik's proofs of the definability criterion and of the theorem of Cobham-Semenov over $\mathbb{N}^m$. The original proof is in Russian. We follow it, simplifying some parts and detailing others.

The last section is a brief introduction to related works and references. The list is certainly not complete; however it may give a flavour of connected research works.

2 Preliminaries

We briefly recall some definitions about automata, automata with output and rational operations. These notions are well detailed in [24, 43, 58].

The sets A and B are finite *alphabets*. We denote by B^* the set of all *words* written with *letters* of B, including the *empty word* λ. Set B^* is a monoid called the *free monoid generated by* B, with concatenation as product operation and λ as neutral element. The symbol $|w|$ denotes the *length* of the word $w \in B^*$ and w^{R} the *reverse* of w. In this paper, A and B are often finite subsets of the set $\mathbb{N} = \{0, 1, 2, \ldots\}$ of natural numbers.

An *automaton* $\mathcal{A} = (Q, I, F, T)$ is a graph with a set Q of vertices or *states* and a set $T \subseteq Q \times B \times Q$ of edges or *transitions* labelled by an alphabet B. Set $I \subseteq Q$ is the set of *initial* states and $F \subseteq Q$ the set of *final* states. A word $w \in B^*$ is *accepted* (or computed) by $\mathcal{A}$ if it is the label of some path in $\mathcal{A}$ beginning with an initial state and ending with a final state. We say that $L \subseteq B^*$ is *recognizable* if it is the set of words computed by some *finite* automaton, i.e., with Q finite. An automaton $\mathcal{A}$ is *deterministic* if it has a unique initial state and if T is a (partial) function $T\ : Q \times B \to Q$. If T is a total function, $\mathcal{A}$ is moreover called *complete*.

Automata with output generalize the concept of automaton. Instead of a set F of final states, they have an *output function* labelling each state by a letter of some alphabet A. For classical automata, $A = \{0, 1\}$, and a state is labelled by 1 if it is final; otherwise it is labelled by 0. An automaton with output *computes* a relation $R \subseteq B^* \times A$ in the following way : (w, a) is in R if and only if there is a path labelled by w from some initial state to some state whose output is a. Any complete deterministic automaton with output computes a function $s\ : B^* \to A$.

Several proofs in this paper use well-known properties of automata and recognizable sets, for instance the equivalence between automata and complete deterministic automata, the closure under Boolean operations of the family of recognizable sets, and the equivalence between automata reading words from left to right and those reading them from right to left.

In particular, we frequently use properties of right-congruences associated with recognizable sets. Let $L \subseteq B^*$; recall that the following relation $\sim_L$ over B^*

$$u \sim_L v \quad \Leftrightarrow \quad [\ \forall w \in B^*, \quad uw \in L \Leftrightarrow vw \in L\ ]$$

is a *right-congruence*, i.e., an equivalence relation which is *right-stable* :

$$u \sim_L v \quad \Rightarrow \quad uw \sim_L vw\ .$$

Set L is the union of some equivalence classes of $\sim_L$. Moreover, L is recognizable if and only if $\sim_L$ has finite index. All these porperties are known as the Myhill-Nerode theorem.

The *minimal automaton* $\mathcal{A}(L)$ of a recognizable set $L \subseteq B^*$ is the smallest (in number of states) complete deterministic automaton computing it. It is unique up to isomorphism and can be constructed in the following way. Its states are the classes of $\sim_L$, the initial state is the class of the empty word and the final states are the classes containing the words of L. A transition labelled by $b \in B$ goes from the class of w to the class of wb.

The automaton $\mathcal{A}(L)$ has two nice properties. For any pair of distinct states q, q', there exists $w \in B^*$ such that $T(q, w)$ is final and $T(q', w)$ is not, or vice versa. Any class C of $\sim_L$ is a state of $\mathcal{A}(L)$, but it is also the set of words labelling paths from the initial state to state C.

As usual, the *rational* operations are $\cup, \cdot$ and *. They operate on sets of words $L, L' \subseteq B^*$, $L \cup L'$ is their union, $L \cdot L' = \{ww' \mid w \in L, w' \in L'\}$ their concatenation and L^* the submonoid of B^* generated by L. The set $L \subseteq B^*$ is then said to be *rational* if it can be constructed from finite subsets of B^* using a finite number of rational operations. *Kleene's theorem* states that $L \subseteq B^*$ is rational if and only if it is recognizable.

The set $\mathbb{N} = \{0, 1, 2, \ldots\}$ is also a monoid, with operation $+$ and neutral element 0. The same holds for $\mathbb{N}^m, m \geqslant 2$, with the vectorial sum $+$ and neutral element $\overline{0}$. We use the notation $\overline{n}$ for the m-tuple $(n_1, \ldots, n_m)$. Rational operations and rational subsets can also be defined in these monoids. In $\mathbb{N}^m$, the rational operations $\cup, \cdot$ and * are interpreted as the operations $\cup, +$ and the closure under addition (L^* is the submonoid of $\mathbb{N}^m$ generated by L). On the other hand, the notion of recognizable set is extended to $\mathbb{N}^m$ using the right-congruence defined above : $L \subseteq \mathbb{N}^m$ is said to be *recognizable* if the right-congruence $\sim_L$ has finite index. Kleene's theorem holds in $\mathbb{N}$ which is the free monoid generated by 1, but it is no longer true in $\mathbb{N}^m, m \geqslant 2$. Indeed, any recognizable subset is rational, but the converse is false. For instance the *diagonal* $L = \{(n, n) \mid n \in \mathbb{N}\}$ is the rational set $(1, 1)^*$ of $\mathbb{N}^2$ but it is not recognizable (since any two points $(n, 0)$ and $(m, 0)$ with $n \neq m$ are not equivalent for $\sim_L$) (see for example [58, page 16]).

3 Some Notions of Logic

In this section, we give a short lesson in first-order logic. The reader is referred to [3] for more details about first-order logic.

3.1 Structures and Formulae

In the sequel, we will often meet the structures $\langle \mathbb{N}, + \rangle$ and $\langle \mathbb{N}, +, V_p \rangle$. In first-order logic, a *structure*

$$\mathcal{S} = \langle D, (R_i)_{i \in I}, (f_j)_{j \in J}, (c_k)_{k \in K} \rangle$$

consists of a *domain* D (some set), a family of *relations* $(R_i)_{i \in I}$ on D, a family of *functions* $(f_j)_{j \in J}$ on D and a family of *constants* $(c_k)_{k \in K}$ of D. The relation R_i is a subset of D^{n_i} and f_j is a function from D^{n_j} to D. The set $\{(R_i)_{i \in I}, (f_j)_{j \in J}, (c_k)_{k \in K}\}$ is called the *language* of the structure $\mathcal{S}$.

For instance, the structure $\langle \mathbb{N}, + \rangle$ has the set $\mathbb{N}$ of natural numbers as domain and the usual function $+$. It has no relation and no constant.

First-order formulae of the structure $\mathcal{S}$ (or in the language of $\mathcal{S}$) are constructed by certain rules. But first we need to list the *symbols* used in the formulae and to define the *terms*.

In addition to the symbols of relations R_i, functions f_j and constants c_k, there are also a countable set of *variables* $x, y, z, \ldots$, the usual *connectives* $\vee$ (or), $\wedge$ (and),

$\neg$ (not), $\rightarrow$ (if then), $\leftrightarrow$ (if and only if), the *quantifiers* $\forall$ (for all), $\exists$ (there exists) and the symbol $=$ (equal).
The *terms* are defined by induction following two rules :

1. any variable and constant is a term,

2. if f_j is a n-ary function and $t_1, \ldots, t_n$ are terms, then $f_j(t_1, \ldots, t_n)$ is a term.

The *formulae* are generated by four rules :

1. if t_1, t_2 are terms, then $t_1 = t_2$ is a formula,

2. if R_i is a n-ary relation and $t_1, \ldots, t_n$ are terms, then $R_i(t_1, \ldots, t_n)$ is a formula,

3. if φ, ϕ are formulae, then $\varphi \vee \phi, \varphi \wedge \phi, \neg\varphi, \varphi \rightarrow \phi, \varphi \leftrightarrow \phi$ are formulae,

4. if φ is a formula and x is a variable, then $\forall x\varphi, \exists x\varphi$ are formulae.

For clarity, parentheses (,) are necessary during the construction of formulae. Formulae generated only by the first two rules are called *atomic formulae*. *Sentences* are formulae with no *free* variables, i.e., variables which are not under the scope of a quantifier. We sometimes write $\varphi(x_1, \ldots, x_n)$ to explicitly mention the free variables of the formula φ.

We point out that the presentation is here a bit different from the one usually given in logic textbooks.

3.2 Examples

The formula

$$(\exists z)(x + z = y)$$

of the structure $\langle \mathbb{N}, + \rangle$ means "there exists $z \in \mathbb{N}$ such that $x + z = y$"; hence it defines the relation $x \leqslant y$. The constant $x = 0$ can also be defined in $\langle \mathbb{N}, + \rangle$ by the formula "$x \leqslant y$ for all $y \in \mathbb{N}$", i.e.

$$(\forall y)(x \leqslant y)$$

where $x \leqslant y$ is the formula above. More generally, any element of $\mathbb{N}$ can be defined in the language of $\langle \mathbb{N}, + \rangle$. For $x = 1$, it is the following formula

$$(\neg(x = 0)) \ \wedge \ ((\forall y)(\neg(y = 0)) \rightarrow (x \leqslant y)) \ .$$

which means that x is not 0 and $x \leqslant y$ for any $y \in \mathbb{N}\backslash\{0\}$. We use $x = 0$ as a short-hand for the formula $(\forall y)(x \leqslant y)$ defining it.

It is also easy to see that the multiplication by a constant can be defined in $\langle \mathbb{N}, + \rangle$. Multiplication of x by 3, $y = 3x$, is simply the formula $y = x + x + x$.

Finally, the property of commutativity of addition is described by the sentence

$$(\forall x)(\forall y)(x + y = y + x) \ .$$

3.3 Definable Sets and Functions

Let $\varphi(x_1, \ldots, x_n)$ be a formula of the structure $\mathcal{S}$ and $d_1, \ldots, d_n$ elements of the domain D. We use the notation

$$\mathcal{S} \models \varphi(d_1, \ldots, d_n)$$

to state that $\varphi(x_1, \ldots, x_n)$ is true when the free variables x_i are replaced by the elements d_i of D, $1 \leqslant i \leqslant n$. Therefore

$$\{\ (d_1, \ldots, d_n) \in D^n \quad | \quad \mathcal{S} \models \varphi(d_1, \ldots, d_n)\ \}$$

is the set of n-tuples of elements of D for which φ is true. We say that this set is *first-order definable* (or in short *definable*) in the structure $\mathcal{S}$ by the formula φ.

We say that a constant c is *definable* in the structure $\mathcal{S}$ if the singleton $\{c\}$ is definable in $\mathcal{S}$. A relation R is *definable* in $\mathcal{S}$ if the subset associated with R is definable in $\mathcal{S}$. In the same way, we say that a function $f\ : D^n \rightarrow D$ is *definable* in the structure $\mathcal{S}$ if its *graph*

$$\{\ (d_1, \ldots, d_n, d) \in D^{n+1} \quad | \quad f(d_1, \ldots, d_n) = d\ \}$$

is definable in $\mathcal{S}$.

In this paper, we mainly study sequences $(s_n)_{n \geqslant 0}$ and more generally multi-dimensional sequences, whose values belong to a finite subset A of $\mathbb{N}$. These sequences are simply functions $\mathbf{s}\ : \mathbb{N}^m \rightarrow A, m \geqslant 1$. Thus, we can speak of *first-order definable* sequences.

For instance the sequence $\mathbf{s} : \mathbb{N} \rightarrow A$ defined by $s_n = n \bmod 3$ is definable in the structure $\langle \mathbb{N}, + \rangle$. Indeed, $\mathbf{s}^{-1}(0)$ is the set of multiples of 3 which is defined by the formula $\varphi_0(x)$

$$(\exists z)(x = 3z)\ .$$

The set $\mathbf{s}^{-1}(1)$ is defined by $\varphi_1(x)$ equal to $(\exists z)(x = 3z + 1)$, and $\mathbf{s}^{-1}(2)$ by $\varphi_2(x)$ equal to $(\exists z)(x = 3z + 2)$. Therefore, the sequence $\mathbf{s} : \mathbb{N} \rightarrow \{0, 1, 2\}$, or equivalently its graph, is first-order definable by the following formula $\varphi(x, y)$

$$(\varphi_0(x) \wedge y = 0) \vee (\varphi_1(x) \wedge y = 1) \vee (\varphi_2(x) \wedge y = 2)\ .$$

This example also shows that a sequence $\mathbf{s} : \mathbb{N}^m \rightarrow A$, with A a finite subset of $\mathbb{N}$, is definable if and only if all the sets $\mathbf{s}^{-1}(a), a \in A$, are definable.

3.4 Equivalent Structures

Consider two structures $\mathcal{S}, \mathcal{S}'$ with the same domain D. Any formula $\varphi(x_1, \ldots, x_n)$ of $\mathcal{S}$ defines the set

$$M_\varphi = \{\ (d_1, \ldots, d_n) \in D^n\ \ |\ \ \mathcal{S} \models \varphi(d_1, \ldots, d_n)\ \}\ .$$

In the same way, any formula φ' of $\mathcal{S}'$ defines the set $M_{\varphi'}$.

We say that $\mathcal{S}$ and $\mathcal{S}'$ are *equivalent* if for every formula $\varphi(x_1, \ldots, x_n)$ of $\mathcal{S}$, there exists a formula $\varphi'(x_1, \ldots, x_n)$ of $\mathcal{S}'$ such that

$$M_\varphi = M_{\varphi'}$$

and conversely. In other words, the sets definable in $\mathcal{S}$ are the same as in $\mathcal{S}'$.

It is easy to check that two structures are equivalent. This holds if the relations, the functions and the constants of the first structure are definable in the second structure, and conversely.

For instance, $\langle \mathbb{N}, + \rangle$ and $\langle \mathbb{N}, +, \leqslant \rangle$ are equivalent structures, because $\leqslant$ is definable in $\langle \mathbb{N}, + \rangle$ (see above). Another example of equivalent structures is $\langle \mathbb{N}, +, \cdot \rangle$ and $\langle \mathbb{N}, +, x^2 \rangle$, where $\cdot$ is the usual product and $y = x^2$ the square function. The function $y = x^2$ is of course definable in $\langle \mathbb{N}, +, \cdot \rangle$ and conversely the product $x \cdot y = z$ is definable in $\langle \mathbb{N}, +, x^2 \rangle$ by the formula

$$(x+y)^2 = x^2 + 2 \cdot z + y^2 \ .$$

Notice that any structure $\langle D, (R_i)_{i \in I}, (f_j)_{j \in J}, (c_k)_{k \in K} \rangle$ is equivalent to a structure with relations only. Functions f_j and constants c_k are easily replaced by relations (its graph for the function f_j and $\{c_k\}$ for the constant c_k).

3.5 Decidable Theories

Given a structure $\mathcal{S}$, the set of the sentences true for $\mathcal{S}$ is the *theory* of $\mathcal{S}$, denoted by $Th(\mathcal{S})$. The theory $Th(\mathcal{S})$ is *decidable* if there exists an algorithm which decides if any sentence of $\mathcal{S}$ is true or false for $\mathcal{S}$, i.e., if it belongs to $Th(\mathcal{S})$ or not. There exist various techniques to prove the decidability of a theory : quantifier elimination, axiomatisation of the theory, finite automata [60].

A classical example of a decidable theory is $Th(\langle \mathbb{N}, + \rangle)$ [59, 28], and an undecidable theory is $Th(\langle \mathbb{N}, +, \cdot \rangle)$ [14, 28].

4 Recognizability over $\mathbb{N}$

4.1 An Appetizing Example

In this section we intuitively introduce four different methods to generate the characteristic sequence of the powers of 2. The definitions will be more precise in the next section.

Let $\mathbf{p} : \mathbb{N} \to \{0, 1\}$ be the characteristic sequence of the powers of 2 :

$$011010001000000010\ldots \ .$$

The alphabet is $A = \{0, 1\}$ and $p_n = 1$ if n is a power of 2, $p_n = 0$ otherwise. This sequence has remarkable properties.

1. *It is generated by a 2-substitution*

   Let

$$f \ : \ \{a,b,c\} \to \{a,b,c\}^2 \ : \ \begin{array}{lll} a & \to & ab \\ b & \to & bc \\ c & \to & cc \end{array}$$

$$g \ : \ \{a,b,c\} \to \{0,1\} \ : \ \begin{array}{lll} a & \to & 0 \\ b & \to & 1 \\ c & \to & 0 \end{array}$$

The iteration of f on the letter a gives rise to a sequence on the alphabet $\{a, b, c\}$, whose image by g is the sequence $\mathbf{p}$.

$$\begin{array}{ccccccccccc}
a & \xrightarrow{f} & ab & \xrightarrow{f} & abbc & \xrightarrow{f} & abbcbccc & \xrightarrow{f} & abbcbcccbccccccc & \ldots \\
\downarrow g & & \downarrow g & & \downarrow g & & \downarrow g & & \downarrow g & \\
0 & \longrightarrow & 01 & \longrightarrow & 0110 & \longrightarrow & 01101000 & \longrightarrow & 0110100010000000 & \ldots
\end{array}$$

Figure 1. Iteration of the 2-substitution f

2. *It is 2-recognizable*

The sequence $\mathbf{p}$ is computed by the following finite automaton with output (the output of each state is indicated under the state).

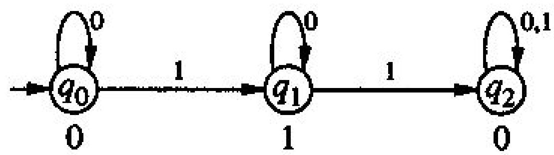

Figure 2. An automaton computing $\mathbf{p}$ in base 2

The automaton computes the symbol p_n from the binary expansion $(n)_2$ of n. More precisely the automaton reads the word $(n)_2$ (the most significant digit of $(n)_2$ is read first) from the initial state q_0 to some state q whose output symbol is p_n. For instance, the binary expansion $(8)_2$ of 8 is 1000, the state reached after reading 1000 is q_1 with output 1, so $p_8 = 1$.

3. *It is 2-definable*

Consider the structure $\langle \mathbb{N}, +, V_2 \rangle$ where V_2 is the function defined by

$$\begin{array}{rcll}
V_2(x) & = & y & \text{where } y \text{ is the greatest power of 2 dividing } x \ (x \neq 0), \\
V_2(0) & = & 1. &
\end{array}$$

The sequence $\mathbf{p}$ is first-order definable in $\langle \mathbb{N}, +, V_2 \rangle$ since the subsets of integers $\mathbf{p}^{-1}(0)$ and $\mathbf{p}^{-1}(1)$ are both definable by a formula of $\langle \mathbb{N}, +, V_2 \rangle$ (see Section 3.3). Indeed $p_n = 1$ if and only if n is a power of 2 if and only if $V_2(n) = n$. Let $P_2(x)$ be the formula $V_2(x) = x$; then

$$\begin{array}{rcl}
\mathbf{p}^{-1}(1) & = & \{ \, n \in \mathbb{N} \mid \langle \mathbb{N}, +, V_2 \rangle \models P_2(n) \, \} \ , \\
\mathbf{p}^{-1}(0) & = & \{ \, n \in \mathbb{N} \mid \langle \mathbb{N}, +, V_2 \rangle \models \neg P_2(n) \, \} \ .
\end{array}$$

4. *It is 2-algebraic*

Consider the finite field $\mathbb{F}_2 = \{0, 1\}$. The formal power series $P(x) \in \mathbb{F}_2[[x]]$

$$P(x) = \sum_{n \geqslant 0} p_n x^n = \sum_{n \geqslant 0} x^{2^n}$$

is naturally associated with the sequence $\mathbf{p}$. One verifies that $P(x)$ is algebraic over the ring $\mathbb{F}_2[x]$, i.e., $P(x)$ is a root of the following polynomial $Q(t)$ with coefficients in $\mathbb{F}_2[x]$:

$$Q(t) = t^2 + t + x \ .$$

Indeed, $P(x)^2 = P(x^2) = P(x) - x$ (remember that $-1 = 1$ and $(a+b)^2 = a^2 + b^2$ in $\mathbb{F}_2$).

4.2 Four Modes of Computing

We now give precise definitions of the four methods intuitively described in the previous section. Theorem 4.1 states that they all generate the same sequences.

Let $p \geqslant 2$ be an integer and $\mathbf{s} : \mathbb{N} \to A$ a *sequence* with values in a finite alphabet $A \subset N$.

1. *p-substitution*

Let B be a finite alphabet. Let $f : B \to B^p$ be a function called a *p-substitution*, which replaces each letter of B by some word of B^* of length p. The function f can be extended to a morphism on B^*. If $f(b)$ begins with the letter b for some $b \in B$, then the sequence $(f^n(b))_{n \geqslant 0}$ converges towards a fixed point $f^\omega(b)$ of f. Let $g : B \to A$ be a function. Then, the image by g of the fixed point $f^\omega(b)$ yields a sequence $\mathbf{s} : \mathbb{N} \to A$.

A sequence $\mathbf{s}$ generated by this kind of process is said to be *generated by p-substitution*.

2. *p-automaton*

A *p-automaton* is a complete deterministic finite automaton with output, whose transitions are labelled by $\{0, 1, \ldots, p-1\}$ and whose states are labelled by A (the output).

A sequence $\mathbf{s}$ is called *p-recognizable* if it is computed by some p-automaton in the following way. The p-ary expansion $(n)_p$ of $n \in \mathbb{N}$ is a word of $\{0, 1, \ldots, p-1\}^*$. Starting in the initial state and using the transitions labelled by the letters of $(n)_p$, one reaches some state q. Then s_n is equal to the output of q. The way a p-automaton reads $(n)_p$ is from the most significant digit to the least one; this choice is arbitrary.

3. *p-definability*

We consider the structure $\langle \mathbb{N}, +, V_p \rangle$, where the function V_p is defined as

$$\begin{array}{lcll} V_p(x) & = & y & \text{where } y \text{ is the greatest power of } p \text{ dividing } x \ (x \neq 0), \\ V_p(0) & = & 1. & \end{array}$$

A sequence $\mathbf{s}$ is *p-definable* if for each letter $a \in A$, there exists a first-order formula φ_a of $\langle \mathbb{N}, +, V_p \rangle$ such that

$$\mathbf{s}^{-1}(a) = \{ \ n \in \mathbb{N} \ \mid \ \langle \mathbb{N}, +, V_p \rangle \models \varphi_a(n) \ \} \ .$$

4. *p-algebraicity*

We assume that p is a *prime number*.

Let K be a finite field with characteristic p such that A is embedded into K (for instance $K = \mathbb{F}_p = \{0, 1, \ldots, p-1\}$ if the cardinality of A is less than or equal to p). With the sequence $\mathbf{s}$ is associated the formal power series

$$S(x) = \sum_{n \geqslant 0} s_n x^n \in K[[x]] \ .$$

We say that $\mathbf{s}$ is *p-algebraic* if $S(x)$ is algebraic over $K[x]$, i.e., if there exist polynomials $q_i(x) \in K[x]$ such that $S(x)$ is a root of

$$Q(t) = q_j(x)t^j + q_{j-1}(x)t^{j-1} + \cdots + q_0(x) \in K[x][t] \backslash \{0\} \ .$$

Theorem 4.1 *Let $p \geqslant 2$ be an integer and $\mathbf{s} : \mathbb{N} \to A$ a sequence with values in a finite alphabet $A \subset \mathbb{N}$. The following are equivalent :*
(1) $\mathbf{s}$ is generated by p-substitution,
(2) $\mathbf{s}$ is p-recognizable,
(3) $\mathbf{s}$ is p-definable,
(4) $\mathbf{s}$ is p-algebraic (under the additional assumption that p is prime).

Some hints on the proof are given in Section 4.4.

There exist sequences which are not p-recognizable, for any $p \geqslant 2$; for instance the characteristic sequence of the squares, or the prime numbers (see [9, 63, 51]; see also Corollary 6.3).

4.3 Notes on p-Automata

Given an integer n, its *p-ary expansion* is the word $(n)_p = w_0 w_1 \dots w_k$ of $\{0, 1, \dots, p-1\}^*$ such that $w_0 \neq 0$ and

$$n = w_0 p^k + w_1 p^{k-1} + \cdots + w_k p^0 \quad .$$

By convention, $(0)_p$ is the empty word λ. Conversely, to any word $w = w_0 w_1 \dots w_k \in \{0, \dots, p-1\}^*$ corresponds its *value* $[w]_p \in \mathbb{N}$ equal to $w_0 p^k + w_1 p^{k-1} + \cdots + w_k p^0$. Different words can have the same integer as value, due to the leading zeros. In fact, any $n \in \mathbb{N}$ has an infinite number of representations w such that $[w]_p = n$; it is the infinite set $0^*(n)_p$.

By definition, p-automata only treat the p-ary expansion of each integer n. We can always suppose that a transition labelled by 0 exists, which loops on the initial state q_0. In this way, the p-automaton identically treats all the words w such that $[w]_p = n$ (see Figure 2). From now on, we will always assume that any p-automaton has *a loop labelled by 0 on its initial state*.

The family of p-recognizable subsets of $\mathbb{N}$ are also much studied. We say that $M \subseteq \mathbb{N}$ is *p-recognizable* if its characteristic sequence $\mathbf{m} : \mathbb{N} \to \{0, 1\}$ defined by

$$m_n = 1 \quad \Leftrightarrow \quad n \in M$$

is p-recognizable.

Equivalently M is p-recognizable if and only if there is a finite automaton accepting the set $\{ w \in \{0, \dots, p-1\}^* \mid [w]_p \in M \}$. This automaton is, for instance, some p-automaton computing the sequence $\mathbf{m}$ whose states with output 1 are considered as final states. Actually we have the following more general result, stating that p-recognizability is independent of leading zeros (see [24, page 106]).

Proposition 4.2 *A set $M \subseteq \mathbb{N}$ is p-recognizable if and only if there exists a finite (deterministic or not) automaton accepting $L \subseteq \{0, \dots, p-1\}^*$ such that*

$$M = \{[w]_p \mid w \in L\} \quad .$$

It is also possible to characterize p-recognizable sets M of integers by an equivalence relation of finite index (see [24, page 107]. This relation is just the translation to

$\mathbb{N}$ of the right-congruence $\sim_L$ of a finite automaton computing $L = \{\ w \mid [w]_p \in M\ \}$ (see Section 2). More precisely, the relation $\sim_{p,M}$ is defined as follows. Let $n, m \in \mathbb{N}$,

$$n \sim_{p,M} m \quad \Leftrightarrow \quad [\ np^k + r \in M \Leftrightarrow mp^k + r \in M \quad \forall k \geqslant 0, \forall r\ 0 \leqslant r < p^k\ ]\ .$$

Consequently, we have

Proposition 4.3 *A set $M \subseteq \mathbb{N}$ is p-recognizable if and only if the equivalence relation $\sim_{p,M}$ has finite index.*

In the sequel, we will need automata reading p-ary expansions of integers from *right to left* instead of left to right. In that case, the equivalence relation $\sim_{p,M^{\mathrm{R}}}$ is slightly different :

$$n \sim_{p,M^{\mathrm{R}}} m \quad \Leftrightarrow \quad [\ n + rp^k \in M \Leftrightarrow m + rp^l \in M \quad \forall r,\ \forall p^k > n,\ p^l > m\ ]\ .$$

We have the analogues of Proposition 4.2 and Proposition 4.3 when reading words from right to left. Since reading from left to right does not change the concept, the choice is a matter of convenience. Later we will see that reading from right to left is a good choice for generalization to higher dimensions.

We have seen that p-recognizable sets of integers coincide with p-recognizable characteristic sequences. Conversely any p-recognizable sequence $\mathbf{s} : \mathbb{N} \to A$ with values in a finite alphabet A is associated with the p-recognizable sets $\mathbf{s}^{-1}(a) \subseteq \mathbb{N}$. Indeed, if $\mathcal{A}$ is a p-automaton for $\mathbf{s}$, then $\mathbf{s}^{-1}(a)$ is computed by $\mathcal{A}$ where the states with output a are considered as the final states [24, Chapter 15].

Proposition 4.4 *Let $A \subset \mathbb{N}$ be a finite alphabet and $\mathbf{s} : \mathbb{N} \to A$ a sequence. Then $\mathbf{s}$ is p-recognizable if and only if each set $\mathbf{s}^{-1}(a)$, $a \in A$, is p-recognizable.*

This proposition allows to transfer theorems on p-recognizable sets into theorems on p-recognizable sequences. We will often use this principle in the sequel.

As for p-recognizable sets $M \subseteq \mathbb{N}$, any p-recognizable sequence $\mathbf{s}$ is characterized by the finite index of the equivalence $\sim_{p,\mathbf{s}}$ (or $\sim_{p,\mathbf{s}^{\mathrm{R}}}$) defined by

$$n \sim_{p,\mathbf{s}} m \quad \Leftrightarrow \quad s_{np^k+r} = s_{mp^k+r} \quad \forall k, \forall r < p^k\ .$$

4.4 Bibliographic Notes

Theorem 4.1 results from several independent works.

The equivalence (1) $\Leftrightarrow$ (2) is proved in [16] (see also [24, Chapter 15]). The idea of the proof is the following. Let $\mathcal{A}$ be a p-automaton computing the sequence $\mathbf{s}$. Let Q be the set of states, q_0 the initial state and $T\ : Q \times \{0, \ldots, p-1\} \to Q$ the transition function. We can suppose that $T(q_0, 0) = q_0$ (see Section 4.3). We define the p-substitution $f\ : Q \to Q^p$ by

$$f(q) = T(q,0)T(q,1)\cdots T(q,p-1)$$

for each $q \in Q$. The function f has a fixed point $f^\omega(q_0)$ because $f(q_0)$ begins with q_0. Let g be the function from Q to A defined by $g(q)$ equal to the output of q. Then the image by g of the fixed point of f is the sequence $\mathbf{s}$ (this is proved by induction on the length of the p-ary expansion of n). The proof of the reversed implication uses the same construction backwards.

The equivalence (2) $\Leftrightarrow$ (4) is proved in [13] (see also [12]). Here p-automata for the sequence $\mathbf{s}$ read p-ary expansions of n from right to left (see Section 4.3). The proof is not easy; it is based on the finiteness of the p-kernel of the sequence $\mathbf{s}$. The *p-kernel* is the set of subsequences of $\mathbf{s}$ equal to

$$\{(s_{np^k+r})_{n\geqslant 0} \mid k \geqslant 0, r < p^k\} \ .$$

The p-kernel is finite if and only if the equivalence relation $\sim_{p,\mathbf{s}^{\mathrm{R}}}$ has finite index.

The equivalence (2) $\Leftrightarrow$ (3) is proved in detail in Section 6, following [40, 74].

The first proof of this equivalence was given by J. R. Büchi in 1960 [9]. It was well detailed for $p = 2$ but just sketched for $p > 2$. Büchi proved that sequences are 2-recognizable if and only if they are defined by weak-monadic second-order formulae of the structure $\langle \mathbb{N}, S\rangle$ where S is the successor function[1]. Roughly the formulae describe how 2-automata compute 2-recognizable sequences. Büchi then stated that these formulae are equivalent to first-order formulae of the structure $\langle \mathbb{N}, +, P_2\rangle$, where $P_2(x)$ is the unary relation "x is a power of 2".

In 1963, R. MacNaughton reviewed Büchi's paper [46]. He noticed that this equivalence with the structure $\langle \mathbb{N}, +, P_2\rangle$ was not correctly proved. He suggested replacing it with the structure $\langle \mathbb{N}, +, \in_2\rangle$, where $\in_2(x, y)$ is the binary relation "y is a power of 2 occurring in the binary expansion of x" (here "occurring" means that the coefficient of y is 1 in the binary expansion of x, i.e., $x = \sum\limits_{\in_2(x,y)} y$).

Referring to the works of [46, 72], M. Boffa suggested the use of the structure $\langle \mathbb{N}, +, V_p\rangle$ instead of $\langle \mathbb{N}, +, P_p\rangle$ [7]. This led to the work [8] where Büchi's proof was detailed and corrected for any $p \geqslant 2$. The implication (3) $\Rightarrow$ (2) is proved directly without any use of weak-monadic second-order formulae, based on the reference [40]. The other implication is in the same spirit as in [9].

C. Michaux and F. Point gave in 1986 another proof of (2) $\Leftrightarrow$ (3) [48]. The proof of the implication (3) $\Rightarrow$ (2) was the same as in [8]. For the converse, they used an induction on rational expressions over the alphabet $\{0, \dots, p-1\}$.

Recently, R. Villemaire gave a short proof for the implication (2) $\Rightarrow$ (3) directly using first-order formulae describing sets of integers computed by p-automata [73, 74].

Let us come back to the structures $\langle \mathbb{N}, +, P_2\rangle$, $\langle \mathbb{N}, +, \in_2\rangle$ and $\langle \mathbb{N}, +, V_2\rangle$. It is easy to see that $\langle \mathbb{N}, +, \in_2\rangle$ and $\langle \mathbb{N}, +, V_2\rangle$ are equivalent structures [48]. The predicate $P_2(x)$ is definable in $\langle \mathbb{N}, +, V_2\rangle$ by the formula $V_2(x) = x$. However A. Semenov proved in [67] that the function V_2 is not definable in $\langle \mathbb{N}, +, P_2\rangle$. This shows that $\langle \mathbb{N}, +, P_2\rangle$ and $\langle \mathbb{N}, +, V_2\rangle$ are not equivalent structures, as conjectured by MacNaughton [46].

More generally, the structures $\langle \mathbb{N}, +, P_p\rangle$ and $\langle \mathbb{N}, +, V_p\rangle$ are not equivalent. This property is a corollary of several decidability results; this has been first noticed by

[1]Weak-monadic second-order formulae of $\langle \mathbb{N}, S\rangle$ are generalizations of first-order ones by allowing additional variables describing finite subsets of $\mathbb{N}$ and quantification over them.

F. Delon [20] (see Figure 3). The theories of the following first-order structures are decidable : $\langle \mathbb{N}, +, V_p \rangle$ [8], $\langle \mathbb{N}, +, p^x \rangle$ where p^x is the exponential function [68, 11]. However, one can show that the theory of $\langle \mathbb{N}, +, V_p, p^x \rangle$ is undecidable (see [11] or [73]). On the other hand, P_p is definable in $\langle \mathbb{N}, +, V_p \rangle$ and $\langle \mathbb{N}, +, p^x \rangle$.

$$\begin{array}{ccccc} & & \langle \mathbb{N}, +, V_p \rangle \ \text{decidable} & & \\ & \nearrow & & \searrow & \\ \langle \mathbb{N}, +, P_p \rangle \ \text{decidable} & & & & \langle \mathbb{N}, +, V_p, p^x \rangle \ \text{undecidable} \\ & \searrow & & \nearrow & \\ & & \langle \mathbb{N}, +, p^x \rangle \ \text{decidable} & & \end{array}$$

Figure 3. Definability relations between four structures

Assume now that $\langle \mathbb{N}, +, P_p \rangle$ and $\langle \mathbb{N}, +, V_p \rangle$ are equivalent. Then the undecidable theory $\langle \mathbb{N}, +, V_p, p^x \rangle$ is equivalent to $\langle \mathbb{N}, +, P_p, p^x \rangle$, itself equivalent to the decidable theory $\langle \mathbb{N}, +, p^x \rangle$. This is impossible.

4.5 A Dessert Example

We end Section 4 by considering the remarkable Thue-Morse sequence $\mathbf{t} : \mathbb{N} \to \{0, 1\}$

$$1001011001101001 \ldots \quad .$$

The alphabet is $A = \{0, 1\}$ and $t_n = 1$ if $(n)_2$ has an even number of 1, $t_n = 0$ otherwise. This sequence has all the properties described in Theorem 4.1.

It is easy to find a 2-automaton computing it. This automaton counts the symbols 1 inside the words $w \in \{0, 1\}^*$.

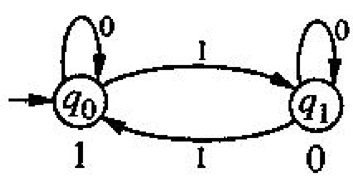

Figure 4. A 2-automaton for the Thue-Morse sequence

From this automaton, we construct the following 2-substitution (see Section 4.4) :

$$f \ : \ \begin{array}{ccc} 0 & \to & 01 \\ 1 & \to & 10 \end{array} \qquad g \ : \text{identity}$$

One of the two fixed points of f is the sequence $\mathbf{t}$, the other is the sequence $1 - \mathbf{t}$.

The sequence $\mathbf{t}$ is also 2-algebraic. Indeed its 2-kernel

$$\{ \ (t_{n2^k+r})_{n \geqslant 0} \ \mid \ k \geqslant 0, r < 2^k \ \}$$

has two elements : the sequences $\mathbf{t}$ (for $k = 1, r = 0$) and $1 - \mathbf{t}$ (for $k = 1, r = 1$). Then

$$\begin{aligned} T(x) &= \sum t_{2n} x^{2n} + \sum t_{2n+1} x^{2n+1} \\ &= \sum t_n x^{2n} + \sum (1 - t_n) x^{2n+1} \\ &= T(x^2) + \frac{x}{1 + x^2} - x T(x^2) \ . \end{aligned}$$

The series $T(x)$ is a root of the polynomial

$$Q(t) = (1+x)^3t^2 + (1+x)^2t + x \in \mathbb{F}_2[x][t] \ .$$

Finally $\mathbf{t}$ is 2-definable. We just give an idea of a formula of $\langle \mathbb{N}, +, V_2 \rangle$ defining it. It will be made more precise in Section 6. The required formula should say that $t_n = 1$ if and only if the binary expansion $(n)_2$ of n contains an even number of 1's. Equivalently, there exists an integer m such that $(m)_2$ "counts" the even number of 1's in $(n)_2$. Roughly, $(m)_2$ is constructed from $(n)_2$ by keeping one 1 among two consecutive 1's of $(n)_2$ and replacing the other by 0, as shown on the following example (for simplicity the case $n = 0$ is treated separately).

$$\begin{aligned}(n)_2 &= 100110100101000\\ (m)_2 &= 000100100001000\end{aligned}$$

More precisely, $t_n = 1, n \geqslant 1$, if and only if $\langle \mathbb{N}, +, V_2 \rangle \models \varphi(n)$ where the formula $\varphi(x)$ says that there exists y such that

1. the first power of 2 occurring in the 2-expansion of x is the same than the one occurring in y ($V_2(x) = V_2(y)$),

2. the last power of 2 (denoted by $\lambda_2(x)$) occurring in the 2-expansion of x does not occur in y ($\neg\in_2(y, \lambda_2(x))$),

3. for any two consecutive powers of 2 occurring in x, one occurs in y if and only if the other one does not.

This is a formula of $\langle \mathbb{N}, +, V_2 \rangle$, because $\in_2(x, y)$ and $\lambda_2(x)$ are definable in $\langle \mathbb{N}, +, V_2 \rangle$.

5 Recognizability over $\mathbb{N}^m$

5.1 Four Modes of Computing

The four modes of computing p-recognizable sequences $\mathbf{s} : \mathbb{N} \to A$ remain applicable for functions $\mathbf{s} : \mathbb{N}^m \to A$, for every $m \geqslant 2$. These functions $\mathbf{s}$ are called again *sequences*. Theorem 4.1 is still valid in this general context. For simplicity, we only consider sequences $\mathbf{s} : \mathbb{N}^2 \to A$. Each of the following definitions is easily generalized to sequences $\mathbf{s} : \mathbb{N}^m \to A$, for any $m > 2$.

Let $p \geqslant 2$ be an integer and $\mathbf{s} : \mathbb{N}^2 \to A$ a sequence. We adapt to $\mathbb{N}^2$ the four definitions of p-substitution, p-automaton, p-definability and p-algebraicity. We illustrate each of them with a particular sequence $\mathbf{t} : \mathbb{N}^2 \to \{0, 1\}$, essentially Pascal's triangle modulo 2. It is defined by $t_{n,m} = 0$ if, for some $k \geqslant 0$, the same power 2^k of 2 occurs in the binary expansions of n and m, otherwise $t_{n,m} = 1$.

$$
\begin{array}{|l}
\uparrow m \qquad \vdots \\
\\
1\,0\,0\,0\,0\,0\,0\,0 \\
1\,1\,0\,0\,0\,0\,0\,0 \\
1\,0\,1\,0\,0\,0\,0\,0 \\
1\,1\,1\,1\,0\,0\,0\,0 \\
1\,0\,0\,0\,1\,0\,0\,0 \quad \cdots \\
1\,1\,0\,0\,1\,1\,0\,0 \\
1\,0\,1\,0\,1\,0\,1\,0 \\
1\,1\,1\,1\,1\,1\,1\,1 \quad \xrightarrow{n} \\
\hline
\end{array}
$$

Figure 5. A sequence similar to Pascal's triangle modulo 2

1. *p-definability*

The sequence $\mathbf{s}$ is called *p-definable* if for any $a \in A$, the set $\mathbf{s}^{-1}(a) \subseteq \mathbb{N}^2$ is definable by a first-order formula $\varphi_a(x,y)$ of $\langle \mathbb{N}, +, V_p \rangle$.

The example of sequence $\mathbf{t}$ is definable in $\langle \mathbb{N}, +, V_2 \rangle$. Indeed the formula $\varphi(x,y)$

$$(\exists z)(\ \in_2(x,z)\ \wedge\ \in_2(y,z)\ )$$

defines the set $\mathbf{t}^{-1}(0) \subseteq \mathbb{N}^2$ and its negation defines the set $\mathbf{t}^{-1}(1)$.

2. *p-automaton*

A *p-automaton* is a complete deterministic finite automaton with output (in the alphabet A). Its transitions are labelled by the alphabet $\{0, \ldots, p-1\}^2$ in a way to read pairs of integers.

More precisely, any word (u,v) over the alphabet $\{0,\ldots,p-1\}^2$ has components u, v with the same length. Its value $([u]_p, [v]_p)$ is a pair (n,m) of integers. It may happen that u has leading zeros and v not, as $|u| = |v|$. Conversely, given a pair (n,m) of integers (suppose for instance that $n \geqslant m$), let $u = (n)_p, v = (m)_p$ and $i = |u| - |v|$, then $(0,0)^*(u, 0^i v)$ is the set of all pairs (u', v') over the alphabet $\{0,\ldots,p-1\}^2$ such that $[u']_p = n, [v']_p = m$.

This *p*-automaton computes a *p-recognizable* sequence $\mathbf{s} : \mathbb{N}^2 \to A$ in the following way. Let $(u,v) \in [\{0,\ldots,p-1\}^2]^*$ such that $n = [u]_p, m = [v]_p$. Starting with the initial state, the reading of (u,v) leads to some state whose ouput defines $s_{n,m}$.

It is easy to construct a 2-automaton computing the particular sequence $\mathbf{t}$. The alphabet labelling the edges is $\{0,1\}^2 = \{\binom{0}{0}, \binom{1}{0}, \binom{0}{1}, \binom{1}{1}\}$.

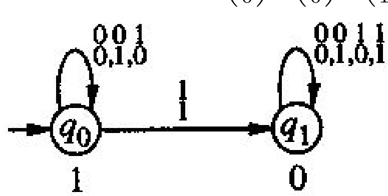

Figure 6. A 2-automaton for sequence $\mathbf{t}$

3. *p-substitution*

Let B be a finite alphabet. Let $f\ : B \to B^{p \times p}$ be a *p-substitution*, it replaces each letter of B by some square of $B^{p\times p}$ with side p. The substitution f extends into a function operating over the set of squares (see the example below). It has a fixed point if the bottom-left corner of $f(b)$ is equal to b for some letter $b \in B$. Let $g : B \to A$ be a function; the image by g of this fixed point yields a sequence $\mathbf{s} : \mathbb{N}^2 \to A$.

A sequence $\mathbf{s}$ generated by this kind of process is said to be *generated by p-substitution.*

The example $\mathbf{t}$ is generated by 2-substitution, in the following way. We construct it by looking at the 2-automaton of Figure 6 (see also Section 4.4). Let $A = B = \{0, 1\}$. Then $f : B \to B^{2\times 2}$ is defined as

$$f(1) = \begin{matrix} 1 & 0 \\ 1 & 1 \end{matrix} \quad , \quad f(0) = \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix}$$

The function $g : B \to A$ is here the identity. The iteration of f on 1 gives a fixed point whose image by g is the sequence $\mathbf{t}$.

$$\begin{matrix} & & & & & & 1\,0\,0\,0\,0\,0\,0\,0 & & \\ & & & & & & 1\,1\,0\,0\,0\,0\,0\,0 & & \\ & & & & & & 1\,0\,1\,0\,0\,0\,0\,0 & & \\ & & & & & & 1\,1\,1\,1\,0\,0\,0\,0 & & \\ & & & & 1\,0\,0\,0 & & 1\,0\,0\,0\,1\,0\,0\,0 & & \\ & & & & 1\,1\,0\,0 & & 1\,1\,0\,0\,1\,1\,0\,0 & & \\ & & 1\,0 & & 1\,0\,1\,0 & & 1\,0\,1\,0\,1\,0\,1\,0 & & \\ 1 & \to & 1\,1 & \to & 1\,1\,1\,1 & \to & 1\,1\,1\,1\,1\,1\,1\,1 & \to & \dots \end{matrix}$$

Figure 7. Iteration of the 2-substitution f

4. *p-algebraicity*

We assume that p is a prime number.

Let K be a finite field with characteristic p such that A is embedded into K. The sequence $\mathbf{s} : \mathbb{N}^2 \to A$ is *p-algebraic* if the formal power series

$$S(x, y) = \sum_{n,m \geqslant 0} s_{n,m} x^n y^m \in K[[x, y]]$$

is algebraic over $K[x, y]$, i.e., there exist polynomials $q_i(x, y) \in K[x, y]$ such that $S(x, y)$ is a root of the polynomial

$$Q(t) = q_j(x, y)t^j + q_{j-1}(x, y)t^{j-1} + \cdots + q_0(x, y) \in K[x, y][t] \backslash \{0\} \ .$$

The sequence $\mathbf{t}$ is 2-algebraic because the series $T(x, y)$ is algebraic over $\mathbb{F}_2[x, y]$:

$$(1 + x + y)T(x, y) + 1 = 0 \ .$$

Indeed, considering the 2-kernel of $\mathbf{t}$, we see that $t_{n,m} = t_{2n,2m} = t_{2n+1,2m} = t_{2n,2m+1}$ and that $t_{2n+1,2m+1} = 0$ for all $n, m \geqslant 0$. So

$$\begin{aligned} T(x, y) &= \sum t_{2n,2m} x^{2n} y^{2m} + \sum t_{2n+1,2m} x^{2n+1} y^{2m} \\ &\quad + \sum t_{2n,2m+1} x^{2n} y^{2m+1} + \sum t_{2n+1,2m+1} x^{2n+1} y^{2m+1} \\ &= (1 + x + y) \sum t_{n,m} x^{2n} y^{2m} + 0 \\ &= (1 + x + y) T(x^2, y^2) \ . \end{aligned}$$

Theorem 5.1 *Let $p \geqslant 2$ and $m \geqslant 1$. Let $\mathbf{s} : \mathbb{N}^m \to A$ be a sequence. The following are equivalent :*
(1) $\mathbf{s}$ is generated by p-substitution,
(2) $\mathbf{s}$ is p-recognizable,
(3) $\mathbf{s}$ is p-definable,
(4) $\mathbf{s}$ is p-algebraic (under the additional assumption that p is a prime number).

In each of the four modes, m is respectively the dimension of $f(b), b \in B$ where f is a p-substitution for $\mathbf{s}$, the number of components of letters labelling the transitions of a p-automaton computing $\mathbf{s}$, the number of variables of a formula defining $\mathbf{s}$ in $\langle \mathbb{N}, +, V_p \rangle$, or the number of variables of the formal power series associated with $\mathbf{s}$.

5.2 Notes

Several authors have independently contributed to Theorem 5.1. They all observed that in $\mathbb{N}^m, m \geqslant 2$, the "good" p-automata are those reading m-tuples $(w_1, \ldots, w_m)$ with components w_i of equal length. In other words, the "good" monoid is $[\{0, \ldots, p-1\}^m]^*$ rather than $[\{0, \ldots, p-1\}^*]^m$. The monoid $[\{0, \ldots, p-1\}^m]^*$ is free, so Kleene's theorem holds.

The proof of equivalence (1) $\Leftrightarrow$ (2) is in the same spirit as for Theorem 4.1 (see [10] where more general substitutions and automata are also considered). We followed this idea for the example $\mathbf{t}$. The equivalence (2) $\Leftrightarrow$ (3) was already included in the one-dimensional case (see Section 4.4). It is proved in details in the next section. The equivalence (2) $\Leftrightarrow$ (4) is established in [23]. See also [64, 65] for another proof of equivalences (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (4).

All the notes we gave for p-automata labelled by $\{0, \ldots, p-1\}$ still hold for the labelling by $\{0, \ldots, p-1\}^m$, $m \geqslant 2$. By definition, *p-recognizable* sets $M \subseteq \mathbb{N}^m$ are those sets whose characteristic sequence $\mathbf{m} : \mathbb{N}^m \to \{0, 1\}$ is p-recognizable. Conversely, any p-recognizable sequence $\mathbf{s} : \mathbb{N}^m \to A$ gives the p-recognizable sets $\mathbf{s}^{-1}(a) \subseteq \mathbb{N}^m$, $a \in A$.

Proposition 5.2 *Let $m \geqslant 1$ and $\mathbf{s} : \mathbb{N}^m \to A$ be a sequence. Then $\mathbf{s}$ is p-recognizable if and only if each set $\mathbf{s}^{-1}(a)$, $a \in A$, is p-recognizable.*

6 Logic and Automata

We give here a simple proof of the equivalence (2) $\Leftrightarrow$ (3) of Theorem 5.1. We consider sets $M \subseteq \mathbb{N}^m$ instead of sequences $\mathbf{s} : \mathbb{N}^m \to A$ (see Proposition 5.2). The proof follows the ideas of references [40, 74].

Theorem 6.1 *Let $m \geqslant 1$ and $M \subseteq \mathbb{N}^m$. Let $p \geqslant 2$. Then M is p-recognizable if and only if M is p-definable.*

Proof. (1) First we construct a finite automaton $\mathcal{A}_\varphi$ for any formula $\varphi(x_1, \ldots, x_m)$ of $\langle \mathbb{N}, +, V_p \rangle$ defining the set

$$M_\varphi = \{ \ (n_1, \ldots, n_m) \in \mathbb{N}^m \ \mid \ \langle \mathbb{N}, +, V_p \rangle \models \varphi(n_1, \ldots, n_m) \ \} \ .$$

This automaton $\mathcal{A}_\varphi$ computes the set of all words $(w_1, \ldots, w_m)$ over the alphabet $\{0, \ldots, p-1\}^m$ such that

$$([w_1]_p, \ldots, [w_m]_p) \in M_\varphi$$

(all the possible leading zeros are considered), it reads words from right to left, it is complete and deterministic (see Sections 4.3 and 5.2).

The proof is by induction on the formulae. For simplicity of the proof, we work with the structure $\langle \mathbb{N}, R_+, R_{V_p} \rangle$, where $R_+(x, y, z)$ is the relation $x + y = z$ and $R_{V_p}(x, y)$ is the relation $V_p(x) = y$. This structure is equivalent to $\langle \mathbb{N}, +, V_p \rangle$ (see Section 3.4)

The atomic formulae of $\langle \mathbb{N}, R_+, R_{V_p} \rangle$ are the equality $x = y$ and the two relations $R_+(x, y, z)$, $R_{V_p}(x, y)$. The corresponding sets $M_=$, M_+, M_{V_p} are p-recognizable. Indeed, for $p = 2$, the automata $\mathcal{A}_=, \mathcal{A}_+, \mathcal{A}_{V_p}$ are the following ones (each final state is denoted by an outgoing small arrow). The addition realized by $\mathcal{A}_+$ is the usual addition with carry.

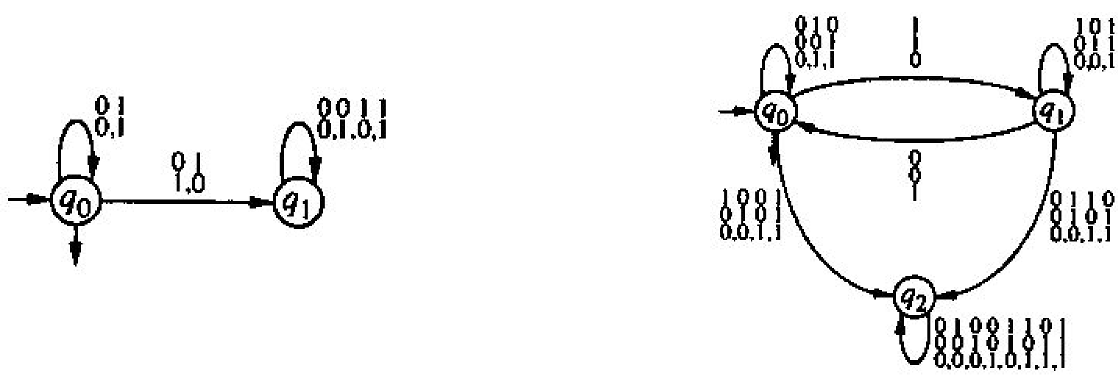

Figure 8.1 Automata $\mathcal{A}_=$ and $\mathcal{A}_+$ in base 2

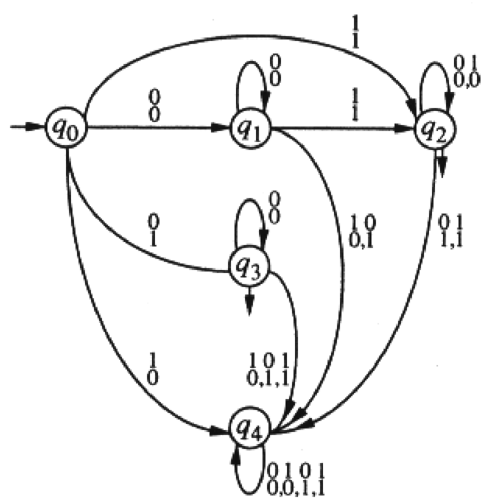

Figure 8.2 Automaton $\mathcal{A}_{V_2}$ in base 2

Now, by induction, assume that automata $\mathcal{A}_\varphi$ and $\mathcal{A}_\psi$ are constructed, for formulae φ and ψ respectively. We show how to obtain automata $\mathcal{A}_{\varphi\vee\psi}$, $\mathcal{A}_{\neg\varphi}$ and $\mathcal{A}_{\exists x \varphi}$.

First, consider the formula $\phi(x_1, \ldots, x_k, y_1, \ldots, y_l, z_1, \ldots, z_m)$ defined as

$$\varphi(x_1, \ldots, x_k, y_1, \ldots, y_l) \ \vee \ \psi(y_1, \ldots, y_l, z_1, \ldots, z_m) \ .$$

Any edge of $\mathcal{A}_\varphi$ is labelled by some letter $(a_1,\ldots,a_k,b_1,\ldots,b_l)$ of the alphabet $\{0,\ldots,p-1\}^{k+l}$. We replace this letter by the set of letters

$$\{(a_1,\ldots,a_k,b_1,\ldots,b_l)\} \times \{0,\ldots,p-1\}^m \ .$$

In the same way, for any edge of $\mathcal{A}_\psi$, we replace its labelling $(b_1,\ldots,b_l,c_1,\ldots,c_m) \in \{0,\ldots,p-1\}^{l+m}$ by the set of letters

$$\{0,\ldots,p-1\}^k \times \{(b_1,\ldots,b_l,c_1,\ldots,c_m)\} \ .$$

The two new automata have now their labelling in the same alphabet $\{0,\ldots,p-1\}^{k+l+m}$. Their union gives the automaton $\mathcal{A}_{\phi\vee\psi}$.

Secondly, consider the formula $\varphi(x_1,\ldots,x_k)$ and the automaton $\mathcal{A}_\varphi$. The set $M_{\neg\varphi}$ is equal to $\mathbb{N}^k \backslash M_\varphi$. The automaton $\mathcal{A}_{\neg\varphi}$ is simply the complement of $\mathcal{A}_\varphi$.

Finally, given the formula $\varphi(x,x_1,\ldots,x_k)$ and the automaton $\mathcal{A}_\varphi$, it remains to construct the automaton $\mathcal{A}_{\exists x\varphi}$ associated with the formula $\exists x\varphi(x,x_1,\ldots,x_k)$. The alphabet labelling the transitions of $\mathcal{A}_\varphi$ is $\{0,\ldots,p-1\}^{k+1}$. Each letter $(a,a_1,\ldots,a_k)$ is then replaced by the letter $(a_1,\ldots,a_k)$ where a is suppressed. The new automaton is generally no longer in the suitable form : it may be not deterministic, and a problem with the lack of leading zeros may happen whenever the label $(a,0,\ldots,0)$ of a transition in $\mathcal{A}_\varphi$ going to a final state has been replaced by $(0,\ldots,0)$. To solve the last problem, use Proposition 4.2 in a way to have again all possible leading zeros. Now the non deterministic automaton can be transformed to a deterministic one in the usual way.

(2) For the converse, we show how to encode any automaton in $\langle\mathbb{N},+,V_p\rangle$.

First we introduce p new relations $\in_{0,p}(x,y), \in_{1,p}(x,y),\ldots,\in_{p-1,p}(x,y)$ and a new function $\lambda_p(x)$, generalizing $\in_2(x,y)$ and $\lambda_2(x)$ introduced in Sections 4.4 and 4.5. The relation $\in_{j,p}(x,y)$, for $0 \leqslant j < p$, means that y is a power of p, and the coefficient of y in the p-ary expansion of x is equal to j, i.e., $x = \sum_{\in_{j,p}(x,y)} j\cdot y$. For powers y strictly greater than x, we consider $\in_{0,p}(x,y)$ to be satisfied (leading zeros). The function $\lambda_p(x)$ denotes the greatest power of p occurring with a nonzero coefficient in the p-ary expansion of x. By convention, $\lambda_p(0)=1$.

The relation $\in_{j,p}(x,y)$, $0 \leqslant j < p$, is definable in $\langle\mathbb{N},+,V_p\rangle$ by the formula

$$P_p(y) \ \wedge \ [\ (\exists z)(\exists t)(x = z + j\cdot y + t) \ \wedge \ (z < y) \ \wedge \ (\ (y < V_p(t)) \vee (t=0)\ )\ ] \ .$$

Roughly this formula says that the powers of p of the p-ary expansion of x are shared into three groups : one group is y only (or equivalently the integer $j\cdot y$), the powers less than y are the second group (the integer z) and the powers greater than y are the third group (the integer t). So, it is possible to express in $\langle\mathbb{N},+,V_p\rangle$ the different letters $w_0,\ldots,w_k$ of the p-ary expansion $(n)_p = w_0\ldots w_k$ of any integer n, as well as leading zeros.

There is also a formula in $\langle\mathbb{N},+,V_p\rangle$ for $\lambda_p(x) = y$:

$$\begin{array}{ll} & [\ P_p(y) \quad \wedge \quad y \leqslant x \quad \wedge \quad ((\forall z)(P_p(z) \wedge y < z) \rightarrow (x < z))\ ] \\ \vee & [\ (x=0) \quad \wedge \quad (y=1)\ ] \end{array}$$

which means that y is a power of p less than or equal to x, such that any power of p greater than y must be greater than x.

We now complete the proof of the theorem. The set $M \subseteq \mathbb{N}^m$ is p-recognizable by hypothesis. Let $\mathcal{A}$ be a complete deterministic finite automaton with set of states Q, initial state q_0, set of final states F and transition function $T : Q \times \{0, \ldots, p-1\}^m \to Q$. For simplicity we suppose that $\mathcal{A}$ reads words from right to left. The m-tuple $(n_1, \ldots, n_m)$ belongs to M if and only if there exists a word $(w_1, \ldots, w_m)$ over the alphabet $\{0, \ldots, p-1\}^m$ such that $[w_i]_p = n_i, 1 \leqslant i \leqslant m$, and if the m-tuple of reversed words $(w_1^{\mathrm{R}}, \ldots, w_m^{\mathrm{R}})$ labels a path in $\mathcal{A}$ from q_0 to some state $q \in F$. This defines a finite sequence $q_0 \ldots q$ of states, beginning with q_0, ending with q, and respecting the transitions.

Without loss of generality, we can replace any of the l states of $\mathcal{A}$ by a l-tuple of letters of $\{0, \ldots, p-1\}$, respectively q_0 by $(1, 0, \ldots, 0)$, q_1 by $(0, 1, 0, \ldots, 0), \ldots$ and q_{l-1} by $(0, \ldots, 0, 1)$. The finite sequence $q_0 \ldots q$ is now a word $(u_1, \ldots, u_l)$ over the alphabet $\{0, \ldots, p-1\}^l$. This sequence can also be considered as a particular l-tuple of integers $(y_1, \ldots, y_l)$ equal to $([u_1]_p, \ldots, [u_l]_p)$. The formula we want to construct describes such a l-tuple.

For any integer n, we denote by $n(i), i \geqslant 0$, the digit j such that $\in_{j,p} (n, p^i)$ is true. So $n = \sum_{i=0}^{+\infty} n(i) p^i$. For any state q, we denote by $q(i)$ its ith component, $1 \leqslant i \leqslant l$.

Now $(x_1, \ldots, x_m)$ belongs to M if and only if there exists a l-tuple of integers $y_1, \ldots, y_l$ such that

1. $(y_1(0), \ldots, y_l(0))$ is the initial state $q_0 = (1, 0, \ldots, 0)$,
2. $(y_1(k), \ldots, y_l(k))$ is some final state of F, with $p^k \geqslant \max_{1 \leqslant j \leqslant m} \lambda_p(x_j)$,
3. for all $0 \leqslant i < k$, if $(y_1(i), \ldots, y_l(i))$ is the state q, then $(y_1(i+1), \ldots, y_l(i+1))$ is the state $T(q, (x_1(i), \ldots, x_m(i)))$.

These three conditions can be expressed by a formula $\varphi(x_1, \ldots, x_m)$, precisely :

$$\begin{aligned}(\exists y_1) \ldots (\exists y_l)(\exists z) \quad & P_p(z) \\ \wedge \quad & (z \geqslant \max_{1 \leqslant j \leqslant m} \lambda_p(x_j)) \\ \wedge \quad & \varphi_1(y_1, \ldots, y_l) \\ \wedge \quad & \varphi_2(y_1, \ldots, y_l, z) \\ \wedge \quad & \varphi_3(x_1, \ldots, x_m, y_1, \ldots, y_l, z)\end{aligned}$$

with

$$\begin{aligned}\varphi_1 \;:\; & \textstyle\bigwedge_{j=1}^{l} \in_{q_0(j),p} (y_j, 1) \\ \varphi_2 \;:\; & \bigvee_{q \in F} \textstyle\bigwedge_{j=1}^{l} \in_{q(j),p} (y_j, z) \\ \varphi_3 \;:\; & (\forall t)(\; P_p(t) \quad \wedge \quad (t < z) \quad \wedge \\ & \bigwedge_{T(q,(a_1,\ldots,a_m))=q'} \quad [\, \textstyle\bigwedge_{j=1}^{l} \in_{q(j),p} (y_j, t) \wedge \bigwedge_{j=1}^{m} \in_{a_j,p} (x_j, t) \\ & \qquad\qquad \to \textstyle\bigwedge_{j=1}^{l} \in_{q'(j),p} (y_j, p \cdot t) \;] \; .\end{aligned}$$

Do not forget that the automaton $\mathcal{A}$ is given and therefore is considered as a constant in the previous formula.

∎

Another simple proof of Büchi's theorem is given in [71]. It has the same structure as ours, but it uses second-order formulae applied to words (as in [9]) instead of first-order formulae applied to integers. So the proof given in [71] together with a standard (for logicians) translation from second-order to first-order logic, leads to another way of proving Theorem 6.1.

The first part of our proof follows ideas given by Hodgson in [40]. He showed in this paper how automata can be used to prove that some theories are decidable. In particular, the theory $Th(\langle \mathbb{N}, +, V_p \rangle)$ is decidable, as a corollary of the previous theorem (see also [9]).

Corollary 6.2 *$Th(\langle \mathbb{N}, + \rangle)$ and $Th(\langle \mathbb{N}, +, V_p \rangle)$ are decidable theories.*

Proof. It is enough to give the proof for $\langle \mathbb{N}, +, V_p \rangle$. Let φ be a sentence in $\langle \mathbb{N}, +, V_p \rangle$. We can assume that φ is the formula $(\exists x)\psi(x)$ or $\neg(\exists x)\psi(x)$ (by making some manipulations of formulae if necessary). The proof above shows how to construct a p-automaton for the p-recognizable set $M_\psi = \{ \ n \in \mathbb{N} \ \mid \ \langle \mathbb{N}, +, V_p \rangle \models \psi(n) \ \}$. By classical results of automata theory, the emptiness of the set M_ψ is decidable. It follows that it is decidable whether the sentence φ is true or not.

■

In Section 8.1, we will prove the interesting Proposition 7.6, as an easy consequence of the previous corollary. We have also the following corollary [9].

Corollary 6.3 *The characteristic sequence of the set of squares $\{n^2 \mid n \in \mathbb{N}\}$ is not p-recognizable, for any $p \geqslant 2$.*

Proof. We first prove that the square fonction $y = x^2$ is definable in $\langle \mathbb{N}, +, R_S \rangle$ where $R_S(y)$ is the relation "y is a square". The function $y = x^2$ is defined by a formula saying that "y is a square and the next square z after y has the property that $y + 2x + 1 = z$". This formula exists in $\langle \mathbb{N}, +, R_S \rangle$.

Ab absurdo, assume that the characteristic sequence of the squares is p-definable, i.e., the relation $R_S(y)$ is p-definable by some formula $\varphi(y)$. Thus, the function $y = x^2$ is also p-definable. Indeed, in the formula defining $y = x^2$ in $\langle \mathbb{N}, +, R_S \rangle$, replace each occurrence of R_S by the formula φ of $\langle \mathbb{N}, +, V_p \rangle$. More generally, any formula of $\langle \mathbb{N}, +, x^2 \rangle$ is a formula of $\langle \mathbb{N}, +, V_p \rangle$.

This means that $\langle \mathbb{N}, +, x^2 \rangle$ is decidable, since $\langle \mathbb{N}, +, V_p \rangle$ is. But $\langle \mathbb{N}, +, x^2 \rangle$ and $\langle \mathbb{N}, +, \cdot \rangle$ are equivalent structures and $Th(\langle \mathbb{N}, +, \cdot \rangle)$ is undecidable (see Section 3). This yields the contradiction.

■

We have seen in the proof of Theorem 6.1 that p-recognizability is preserved by the Boolean operations and also by projection. Another interesting corollary of this theorem is that some operations over integers preserve p-recognizability.

For instance, if $M \subseteq \mathbb{N}$ is p-recognizable, then $c \cdot M$ is still p-recognizable where c is any constant. Indeed, if $\varphi(x)$ is a formula of $\langle \mathbb{N}, +, V_p \rangle$ defining M, then the formula $(\exists y)(\ (x = c \cdot y) \wedge \varphi(y)\ )$ defines the set $c \cdot M$. Also addition, substraction, multiplication or division by a constant are operations which preserve p-recognizability.

The diagonal of any p-recognizable subset M of $\mathbb{N}^2$, defined by $\{n \in \mathbb{N} \mid (n, n) \in M\}$, is also a p-recognizable subset of $\mathbb{N}$. If $\varphi(x, y)$ is a formula for M, then $\varphi(x, x)$ is a formula for its diagonal.

Generally, any operation definable in $\langle \mathbb{N}, +, V_p \rangle$, preserves p-recognizability. The proof of this property is straightforward and uniform in $\langle \mathbb{N}, +, V_p \rangle$. There also exist proofs using automata for the operations given previously as examples (see [58] or [65]). However, each operation needs its own proof and it soon becomes difficult to find a proof for more complex operations.

Corollary 6.4 *Let $f : \mathbb{N}^m \rightarrow \mathbb{N}^n$ be an operation over integers. If f is definable in $\langle \mathbb{N}, +, V_p \rangle$ and if $M \subseteq \mathbb{N}^m$ is p-recognizable, then $f(M)$ is p-recognizable.*

Proof. The proof is very easy. M is defined by a formula $\varphi(y_1, \ldots, y_m)$. The graph of $f : \mathbb{N}^m \rightarrow \mathbb{N}^n$ is defined by a formula $\phi(y_1, \ldots, y_m, x_1, \ldots, x_n)$. Then $f(M)$ is defined by the formula $(\exists y_1) \ldots (\exists y_m) \varphi(y_1, \ldots, y_m) \wedge \phi(y_1, \ldots, y_m, x_1, \ldots, x_n)$. ■

7 Base-Dependence

Four equivalent modes characterize p-recognizable sequences $\mathbf{s} : \mathbb{N}^m \rightarrow A$: p-substitutions, p-automata, p-definability, p-algebraicity (Theorems 4.1, 5.1). They heavily depend on the base p. We are going to see that there are three kinds of sequences $\mathbf{s}$: the sequences recognizable in every base $p \geqslant 2$, the sequences recognizable in certain bases p only and the sequences recognizable in no base p.

7.1 Base p^k

Let us come back to the characteristic sequence $\mathbf{p}$ of the powers of two (see Section 4.1). It is generated by the 2-substitution

$$f : \{a,b,c\} \rightarrow \{a,b,c\}^2 : \begin{array}{lll} a & \rightarrow & ab \\ b & \rightarrow & bc \\ c & \rightarrow & cc \end{array}$$

$$g : \{a,b,c\} \rightarrow \{0,1\} : \begin{array}{lll} a & \rightarrow & 0 \\ b & \rightarrow & 1 \\ c & \rightarrow & 0 \end{array}$$

It is also generated by some 2^k-substitution, for all $k \geqslant 1$. To see this, simply replace f by the iteration f^k. For instance, for $k = 2$, f^2 is the 4-substitution

$$f^2 : \{a,b,c\} \rightarrow \{a,b,c\}^4 : \begin{array}{lll} a & \rightarrow & abbc \\ b & \rightarrow & bccc \\ c & \rightarrow & cccc \end{array}$$

This property is also verified on automata. The 4-recognizable sequence $\mathbf{p}$ is computed by the following 4-automaton.

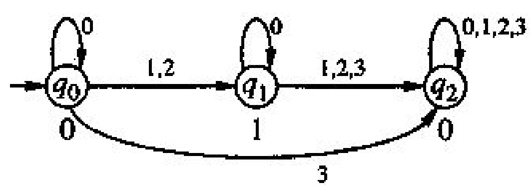

Figure 9. A 4-automaton computing $\mathbf{p}$

This automaton is simply the 2-automaton of Figure 2 with the transitions modified in the following way. The edges are the paths of length 2 of the 2-automaton, with the labelling $0, 1, 2, 3$ instead of $00, 01, 10, 11$ respectively.

From the logical point of view, the argument is simple too. The function V_2 is definable in $\langle \mathbb{N}, +, V_4 \rangle$. Indeed, if $V_4(x) = V_4(2 \cdot x)$, then $V_2(x)$ is equal to $V_4(x)$, otherwise $V_2(x)$ is equal to $2 \cdot V_4(x)$. The formula φ of $\langle \mathbb{N}, +, V_4 \rangle$ defining $V_2(x) = y$ is then the following

$$((V_4(x) = V_4(2 \cdot x)) \wedge (y = V_4(x))) \ \vee \ ((V_4(x) \neq V_4(2 \cdot x)) \wedge (y = 2 \cdot V_4(x))) \ .$$

The sequence $\mathbf{p}$ is then 4-definable. Indeed, in the formulae of $\langle \mathbb{N}, +, V_2 \rangle$ defining $\mathbf{p}^{-1}(0)$ and $\mathbf{p}^{-1}(1)$, replace each occurrence of V_2 by φ. The two new formulae are formulae of $\langle \mathbb{N}, +, V_4 \rangle$ showing the 4-definability of $\mathbf{p}$.

The property observed on the example $\mathbf{p}$ holds for any p-recognizable sequences : they are also p^k-recognizable. More generally, they are q-recognizable as soon as p, q are *multiplicatively dependent* integers, i.e., there exist $k, l \geqslant 1$ such that $p^k = q^l$, or equivalently $p = r^k, q = r^l$, for some $r \geqslant 2$ and $k, l \geqslant 1$.

Proposition 7.1 *Let $p, q \geqslant 2$ be multiplicatively dependent integers. Let $m \geqslant 1$ and $\mathbf{s} : \mathbb{N}^m \to A$ be a sequence. Then $\mathbf{s}$ is p-recognizable if and only if $\mathbf{s}$ is q-recognizable.*

Proof.

We prove that the structures $\langle \mathbb{N}, +, V_p \rangle$ and $\langle \mathbb{N}, +, V_{p^k} \rangle$ are equivalent. Function V_{p^k} is definable in $\langle \mathbb{N}, +, V_p \rangle$:

$V_{p^k}(x) = y$ if and only if "y is the greatest power of p^k less than or equal to $V_p(x)$"

The predicate $P_{p^k}(y)$ is definable in $\langle \mathbb{N}, +, V_p \rangle$, observing that "$y$ is a power of p^k if and only if y is a power of p and $p^k - 1$ divides $y - 1$". Indeed, assume that $y - 1 = (p^k - 1)z$ for some $z \neq 0$ and write y as p^{ak+b}, with $0 \leqslant b < k$. Then

$$y - 1 = p^b \cdot (p^{ak} - 1) + (p^b - 1) \ .$$

As $p^k - 1$ divides $y - 1$ and $p^{ak} - 1$, it also divides $p^b - 1$. Hence $b = 0$. The other implication is trivial.

Conversely, V_p is definable in $\langle \mathbb{N}, +, V_{p^k} \rangle$ (this is similar to the case V_2 and V_4 we have just explained) :

$$\begin{array}{llll}
\text{“If} & V_{p^k}(x) = V_{p^k}(p^{k-1} \cdot x), & \text{then} & V_p(x) = V_{p^k}(x) \quad , \\
\text{else if} & V_{p^k}(x) = V_{p^k}(p^{k-2} \cdot x), & \text{then} & V_p(x) = p \cdot V_{p^k}(x) \quad , \\
\ldots & & & \\
\text{else if} & V_{p^k}(x) = V_{p^k}(p \cdot x), & \text{then} & V_p(x) = p^{k-2} \cdot V_{p^k}(x) \quad , \\
\text{else} & & & V_p(x) = p^{k-1} \cdot V_{p^k}(x)\text{”} \quad .
\end{array}$$

Then we have shown that the structures $\langle \mathbb{N}, +, V_p \rangle$, $\langle \mathbb{N}, +, V_{p^k} \rangle$ are equivalent and that the structures $\langle \mathbb{N}, +, V_{q^l} \rangle$, $\langle \mathbb{N}, +, V_q \rangle$ are also equivalent. By hypothesis p and q are multiplicatively dependent. Let $k, l \geqslant 1$ be such that $p^k = q^l$. It follows that $\langle \mathbb{N}, +, V_p \rangle$ and $\langle \mathbb{N}, +, V_q \rangle$ are equivalent.

■

7.2 Base 1 over $\mathbb{N}$

A sequence $\mathbf{s} : \mathbb{N} \to A$ is said to be *ultimately periodic* if there exists $v \geqslant 1$ such that

$$\exists n_0, \forall n \geqslant n_0, s_n = s_{n+v} \ .$$

The integer v is called a *period* of the ultimately periodic sequence. Ultimately periodic sequences are a family of interesting sequences, as they are p-recognizable for any $p \geqslant 2$.

For instance, the sequence $\mathbf{u} : \mathbb{N} \to \{0,1\}$ equal to $00001001001001\ldots$ is ultimately periodic (with $v = 3, n_0 = 2$). It is p-recognizable for any $p \geqslant 2$. Indeed, u_n equals 1 if and only if 3 divides $n - 4$. The set $\mathbf{u}^{-1}(1)$ is then defined by the following formula $\varphi(x)$ of $\langle \mathbb{N}, + \rangle$

$$(\exists y)\ (x = 3y + 4) \ .$$

The set $\mathbf{u}^{-1}(0)$ is defined by $\neg\varphi(x)$. As any formula of $\langle \mathbb{N}, + \rangle$ is a formula of $\langle \mathbb{N}, +, V_p \rangle$, the sequence $\mathbf{u}$ is p-definable for any $p \geqslant 2$.

This property can also be proved with formal power series or automata. The formal power series

$$U(x) = \sum x^{3m+4} = \frac{x^4}{1 - x^3}$$

associated with $\mathbf{u}$, is rational. It is a root of the polynomial $(1-x^3)t - x^4 \in \mathbb{F}_p[x][t]$. Hence $\mathbf{u}$ is p-algebraic for all prime numbers p.

Let $p \geqslant 2$. Let us define $\mathbf{v} : \mathbb{N} \to \{0,1\}$ by $v_n = u_{n+4}$, for all n. A p-automaton $\mathcal{A}$ for $\mathbf{v}$ has 3 states $\{q_0, q_1, q_2\}$ and transitions $T(q_i, b) = q_j$, for $b \in \{0, \ldots, p-1\}$, such that

$$j = p \cdot i + b \bmod 3 \ .$$

The initial state is q_0. The output of q_0 is 1 and the ouput of q_1, q_2 is 0. It is easy to modify $\mathcal{A}$ into a p-automaton computing $\mathbf{u}$.

Proposition 7.2 *Let* $\mathbf{s} : \mathbb{N} \to A$ *be a sequence. If* $\mathbf{s}$ *is ultimately periodic, then* $\mathbf{s}$ *is p-recognizable for all* $p \geqslant 2$.

Proof. Roughly, $\mathbf{s}^{-1}(a), a \in A$, is a finite union of arithmetic progressions. As in the example, one shows that any arithmetic progression is definable in the structure $\langle \mathbb{N}, + \rangle$ and therefore in $\langle \mathbb{N}, +, V_p \rangle$.

■

The previous example suggests intrinsic properties of ultimately periodic sequences. The next theorem characterizes ultimately periodic sequences via "automatic", logical and algebraic arguments.

Theorem 7.3 *Let* $\mathbf{s} : \mathbb{N} \to A$ *be a sequence. The following are equivalent :*
(1) $\mathbf{s}$ *is ultimately periodic,*
(2) $\mathbf{s}$ *is definable in* $\langle \mathbb{N}, + \rangle$,
(3) The series $S(x) = \sum_{n \geqslant 0} s_n x^n$ *is rational :*

$$S(x) = \frac{p(x)}{q(x)} \quad \text{with } p(x) \in \mathbb{Z}[x], q(x) \in \mathbb{Z}[x] \backslash \{0\} \ ,$$

(4) $\mathbf{s}$ *is 1-recognizable (by a 1-automaton),*
(5) The sets $\mathbf{s}^{-1}(a)$, $a \in A$, *are rational subsets of the monoid* $\mathbb{N}$.

The example of the sequence $\mathbf{u}$ can help to imagine a proof for Theorem 7.3.

It is proved in [59] (see also [28]) that any formula $\varphi(x_1, \ldots, x_m)$ of $\langle \mathbb{N}, + \rangle$ can be written as a finite combination of disjunctions, conjunctions and negations of the formulae

$$\begin{array}{rclll} t_i(x_1, \ldots, x_m) & \geqslant & c_i & & 1 \leqslant i \leqslant r \ , \\ t_i(x_1, \ldots, x_m) & = & c_i & \bmod d & r < i \leqslant s \ , \end{array}$$

where $c_i \in \mathbb{Z}, d, m \in \mathbb{N}$ are constants and $t_i(x_1, \ldots, x_m)$ is equal to $\sum u_{i,j} x_j$, with $u_{i,j} \in \mathbb{Z}$. This explains why implication (2) $\Rightarrow$ (1) holds.

The equivalence (1) $\Leftrightarrow$ (3) is analoguous to the fact that a real number has periodic expansion if and only if it is rational.

A 1-automaton looks like a "frying pan", it is something quite special and has little relationship to p-automata. The integer n is represented by the "1-ary expansion" 0^n, here 0 could be replaced by any other symbol (see references [24, 8] for more details). It is easy to prove the equivalence (1) $\Leftrightarrow$ (4). For the preceding example $\mathbf{u}$, a 1-automaton looks like

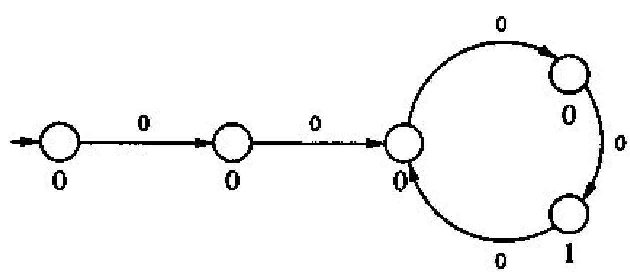

Figure 10. A frying pan automaton

Saying that a sequence $\mathbf{s}$ is ultimately periodic is the same as saying that the sets $\mathbf{s}^{-1}(a)$, $a \in A$, are rational (or equivalently recognizable) subsets of the free monoid $(\mathbb{N}, +)$ generated by 1. Indeed, rational subsets of $\mathbb{N}$ are exactly finite unions of integers and linear progressions. For instance, the set $\mathbf{u}^{-1}(1)$ of the sequence $\mathbf{u}$ is the rational subset $1^4.(1^3)^*$ of $\mathbb{N}$ (where the product operation is here interpreted as the addition in $\mathbb{N}$).

7.3 Base 1 over $\mathbb{N}^m$

First we must define a convenient generalization to $\mathbb{N}^m$ of ultimately periodic sequences. In order to keep an analog of Theorem 7.3 and Proposition 7.2 in all dimensions, the logical characterization (2) in Theorem 7.3 is clearly a good candidate, since definability in $\langle \mathbb{N}, + \rangle$ is a notion independent of the dimension.

S. Ginsburg and E. Spanier showed in [35] that $M \subseteq \mathbb{N}^m$ is definable in $\langle \mathbb{N}, + \rangle$ if and only if it is *semilinear*, which means that M is defined by a finite disjunction of formulae $\varphi(\overline{x})$ of the following form :

$$\begin{array}{llrcl} \text{either} & & (\overline{x} & = & \overline{a}) \\ \text{or} & (\exists y_1) \ldots (\exists y_j) & (\overline{x} & = & \overline{a}_0 + \overline{a}_1 y_1 + \cdots + \overline{a}_j y_j) \end{array}$$

where $\overline{x}$ is the m-tuple $(x_1, \ldots, x_m)$, $\overline{a}, \overline{a}_0, \ldots, \overline{a}_j \in \mathbb{N}^m$ are constants and $\overline{a} \cdot y$ is intended as the product $(a_1 y, \ldots, a_m y)$. Hence M is semilinear if and only if it is a finite union of *points* (formula $\overline{x} = \overline{a}$) and of *cones* (formula $(\exists y_1) \ldots (\exists y_j)(\overline{x} = \overline{a}_0 + \overline{a}_1 y_1 + \cdots + \overline{a}_j y_j))$.

Semilinearity is equivalent to rationality over the monoid $\mathbb{N}^m$. Indeed the two previous formulae define the rational sets $\{\overline{a}\}$ and $\overline{a}_0 \cdot \{\overline{a}_1, \cdots, \overline{a}_j\}^*$ of $\mathbb{N}^m$. Conversely one can show that any rational subset of $\mathbb{N}^m$ is semilinear by induction on the rational operations.

Recently A. Muchnik gave an impressive characterization of semilinear sets in terms of "local periodicity" [54]. It is the last characterization in the next theorem. Muchnik mentioned it as the *definability criterion*. We give the proof of this criterion in Section 8.1.

Theorem 7.4 *Let $m \geqslant 1$ and $\mathbf{s} : \mathbb{N}^m \to A$ be a sequence. Let $M_a = \mathbf{s}^{-1}(a)$, for each $a \in A$. The following are equivalent :*
(1) $\mathbf{s}$ is definable in $\langle \mathbb{N}, + \rangle$,
(2) each M_a is semilinear,
(3) each M_a is a rational subset of the monoid $\mathbb{N}^m$,
(4) each M_a is locally periodic and every $(m-1)$-dimensional sections of M_a is definable in $\langle \mathbb{N}, + \rangle$.

The last characterization needs some explanations. Let $M \subseteq \mathbb{N}^m$. The *section* $M_{i,c}$ of M is obtained by fixing the ith component to the constant c :

$$M_{i,c} = \{\overline{n} \in M \mid n_i = c\} \ .$$

Then $M_{i,c} \subseteq \mathbb{N}^{i-1} \times \{c\} \times \mathbb{N}^{m-i}$ can be considered as a subset of $\mathbb{N}^{m-1}$.

We say that M is *locally periodic* if there exists a finite set V of vectors $\overline{v} \in \mathbb{N}^m$ different from $\overline{0}$ such that for some $K > |V|$ and $L \geqslant 0$, one has :

$$(\forall \overline{n} \in \mathbb{N}^m, |\overline{n}| \geqslant L)(\exists \overline{v} \in V)(M \text{ is } \overline{v}\text{-periodic inside } \mathcal{N}(\overline{n}, K)) \ .$$

Let $X \subseteq \mathbb{N}^m$, set M is *$\overline{v}$-periodic* inside X if for any $\overline{m}, \overline{m} + \overline{v} \in X$

$$\overline{m} \in M \quad \Leftrightarrow \quad \overline{m} + \overline{v} \in M \ .$$

The vector $\overline{v}$ is called a *period* for M. In $\mathbb{N}^2$, this means that M is periodic in the direction of $\overline{v}$, when looking at M through the "window" X.
The set $\mathcal{N}(\overline{n}, K)$ is the *K-neighbourhood* of $\overline{n}$, it is the set

$$\mathcal{N}(\overline{n}, K) = \{\overline{n} + \overline{r} \mid \overline{r} \in \mathbb{N}^m, |\overline{r}| < K\} \ ,$$

where the norm $|\overline{r}|$ of $\overline{r}$ is equal to $\max\{r_1, \ldots, r_m\}$. For instance, in $\mathbb{N}^2$, this is a square with size K and bottom-left corner $\overline{n}$. Finally, notation $|V|$ means $\sum_{\overline{v} \in V} |\overline{v}|$. Therefore M is locally periodic if there exists a finite number of periods $\overline{v}$ for M such that for some large enough K, for any K-neighbourhood $\mathcal{N}(\overline{n}, K)$ far enough from the origin $\overline{0}$, M seen through $\mathcal{N}(\overline{n}, K)$ is periodic with one of the periods $\overline{v}$.

Let us now look at an example. The following sequence $\mathbf{c} : \mathbb{N}^2 \to \{0, 1\}$ is the characteristic sequence of a semilinear set. Figure 11 shows two points $(0, 1), (2, 4)$ and two cones one of which is degenerated into the diagonal.

$$
\begin{array}{|ccccccccccccccccccc l}
\uparrow & x_2 & & & & \vdots & & & & & & & & & & & & & & \\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0 & \\
0&0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1&1&1 & \\
0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0 & \\
0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1&1&1&1 & \\
0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0 & \\
0&0&0&0&0&0&0&1&1&1&1&1&1&1&1&1&1&1&1 & \\
0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0 & \\
0&0&0&0&0&0&1&1&1&1&1&1&1&1&1&1&1&1&1 & \\
0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0 & \ldots \\
0&0&0&0&0&1&1&1&1&1&1&1&1&1&1&1&1&1&1 & \\
0&0&1&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0 & \\
0&0&0&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 & \\
0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 & \\
1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 & x_1 \\
1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 & \longrightarrow \\
\hline
\end{array}
$$

Figure 11. The characteristic sequence of a semilinear set

The set $\mathbf{c}^{-1}(1)$ is defined by the formula $\varphi(x_1, x_2)$ of $\langle \mathbb{N}, + \rangle$

$$
\begin{array}{ll}
 & (x_1, x_2) = (0, 1) \\
\vee & (x_1, x_2) = (2, 4) \\
\vee & (\exists y)(\ (x_1, x_2) = (1, 1)y\ ) \\
\vee & (\exists y_1)(\exists y_2)(\ (x_1, x_2) = (4, 3) + (1, 2)y_1 + (1, 0)y_2\ )\ .
\end{array}
$$

It is easy to find, from this formula, a rational expression for $\mathbf{c}^{-1}(1)$:

$$\mathbf{c}^{-1}(1) = (0, 1) \cup (2, 4) \cup (1, 1)^* \cup (4, 3)\cdot \{(1, 2), (1, 0)\}^*\ .$$

Furthermore, $\mathbf{c}^{-1}(1)$ is locally periodic with $V = \{(2, 2), (1, 2), (1, 0)\}$ (see Figure 11). Indeed, far enough from the origin, a K-neighbourhood is either completely inside the cone defined by $(\exists y_1)(\exists y_2)$ ($(x_1, x_2) = (4, 3) + (1, 2)y_1 + (1, 0)y_2$), either completely outside this cone, or it overlaps one of its borders. In the last case, we choose period (1,2) for the border defined by (1,2) and period (1,0) for the other border. When the K-neighbourhood is inside the cone, it can meet the diagonal defined by $(\exists y)(\ (x_1, x_2) = (1, 1)y\ )$ or not. For both situations, period (2,2) is convenient. Finally, outside the cone, any period works.

As any formula of $\langle \mathbb{N}, + \rangle$ is trivially a formula of $\langle \mathbb{N}, +, V_p \rangle$, for any $p \geqslant 2$, we have the analog of Proposition 7.2.

Proposition 7.5 *Any sequence* $\mathbf{s} : \mathbb{N}^m \to A$ *definable in* $\langle \mathbb{N}, + \rangle$ *is p-recognizable, for all* $p \geqslant 2$.

■

Of course, there exist p-recognizable sequences which are not definable in $\langle \mathbb{N}, + \rangle$. The characteristic sequence $\mathbf{p}$ of the powers of 2 (Section 4.1) and the sequence $\mathbf{t}$ describing Pascal triangle modulo 2 (Section 5.1) are such examples. Muchnik's definability criterion allows to decide whether a p-recognizable sequence $\mathbf{s} : \mathbb{N}^m \to A$ is definable in $\langle \mathbb{N}, + \rangle$ [54]. This result was already proved for the one-dimensional case [39, 42, 55]. The proof in the general case is easy [54], using the decidability of $Th(\langle \mathbb{N}, +, V_p \rangle)$ (Corollary 6.2); it is given in Section 8.1.

Proposition 7.6 *Let $m \geqslant 1$ and $p \geqslant 2$. Let $\mathbf{s} : \mathbb{N}^m \to A$ be a p-recognizable sequence. It is decidable whether $\mathbf{s}$ is definable in $\langle \mathbb{N}, + \rangle$.*

To conclude this section, let us remark that in $\mathbb{N}^m$, $m \geqslant 2$, Kleene's theorem is no longer true : the family of recognizable subsets is strictly included in the family of rational subsets (the diagonal is rational but not recognizable). Recognizable sets may be understood as a certain "tiling" of $\mathbb{N}^m$ by a finite number of parallelepipeds. Theorem 7.4 shows that this concept does not give a sufficient generalization of ultimately periodic sequences in $\mathbb{N}^m$.

This is emphasized by the theorem of Cobham-Semenov in the next section.

7.4 Theorem of Cobham-Semenov

Any ultimately periodic sequence $\mathbf{s} : \mathbb{N} \to A$, and more generally any sequence $\mathbf{s} : \mathbb{N}^m \to A$ definable in $\langle \mathbb{N}, + \rangle$, is p-recognizable for any $p \geqslant 2$ (Propositions 7.2, 7.5). In 1969, A. Cobham proved the converse in the case of $\mathbb{N}$ [15]. Later in 1977, A. Semenov generalized this result to $\mathbb{N}^m$ [66]. As a matter of fact, Cobham and Semenov proved a stronger property : as soon as a sequence $\mathbf{s} : \mathbb{N}^m \to A$ is p- and q-recognizable, for some multiplicatively independent integers $p, q \geqslant 2$, then $\mathbf{s}$ is definable in $\langle \mathbb{N}, + \rangle$. We recall that $p, q \geqslant 2$ are *multiplicatively independent* if and only if the equation $p^k = q^l$ has the solution $k = l = 0$ only. Using the logical characterizations (see Theorems 5.1 and 7.4), this is reformulated in the following way : let $p, q \geqslant 2$ be multiplicatively independent integers, then $\mathbf{s}$ is both p- and q-definable if and only if $\mathbf{s}$ is definable in $\langle \mathbb{N}, + \rangle$. This result says that if $\mathbf{s}$ can be defined by using $+, V_p$ or $+, V_q$, then it can be defined by using $+$ only. This is clearly not obvious.

The theorem of Cobham-Semenov theorem is one of the most beautiful results in the theory of recognizability of natural numbers. We give in Section 8.2 an elegant proof of this result, following the reference [54].

Theorem 7.7 (Cobham-Semenov) *Let $m \geqslant 1$, let $p, q \geqslant 2$ be multiplicatively independent integers. Let $\mathbf{s} : \mathbb{N}^m \to A$ be a sequence. If $\mathbf{s}$ is p-recognizable and q-recognizable, then $\mathbf{s}$ is definable in $\langle \mathbb{N}, + \rangle$.*

The characteristic sequence of powers of 2 is certainly not ultimately periodic. It is 2-recognizable and also 2^k-recognizable, for all $k \geqslant 1$. As 3 and 2 are multiplicatively independent, this sequence is not 3-recognizable and thus not 3^k-recognizable, $k \geqslant 1$. More generally, it is not p-recognizable, for each $p \geqslant 2$ which is not a power of 2.

The relation "being multiplicatively dependent" is an equivalence relation on $\mathbb{N}\backslash\{0,1\}$. In each equivalence class, there exists a smallest integer p which we call *simple*. Any other element of this class is a power p^k of p, $k \geqslant 1$. An integer p is simple if and only if $\gcd(k_1, \ldots, k_l) = 1$ where $p = p_1^{k_1} \ldots p_l^{k_l}$ is the prime factorization of p. The first simple integers are $2, 3, 5, 6, 7, 10, 11, \ldots$.

Assume the sequence $\mathbf{s} : \mathbb{N}^m \to A$ is p-recognizable, but not definable in $\langle \mathbb{N}, + \rangle$. We can suppose that p is simple. The theorem of Cobham-Semenov states that the only bases q for which $\mathbf{s}$ is q-recognizable are $q = p^k, k \geqslant 1$.

To summarize, sequences $\mathbf{s}$ are classified into 3 distinct groups :

1. *The sequences p-recognizable in every base $p \geqslant 2$*

   These sequences are exactly the sequences definable in $\langle \mathbb{N}, + \rangle$. They are also p-definable for each $p \geqslant 2$.

2. *The sequences p-recognizable for certain bases p only*

   These sequences are not definable in $\langle \mathbb{N}, + \rangle$. They are only p^k-recognizable for $k \geqslant 1$ (p being simple).

3. *The sequences not p-recognizable for any $p \geqslant 2$*

   The characteristic sequence of squares belongs to this family (see Section 4.2 and Corollary 6.3).

7.5 Bibliographic Notes

The first step towards the theorem of Cobham-Semenov is in Büchi's paper [9]. He proved that P_p is definable in $\langle \mathbb{N}, +, P_q \rangle$ if and only if p, q are multiplicatively dependent. In this paper, he also showed that ultimately periodic sequences are p-recognizable for all $p \geqslant 2$. He also proved that the p^k-recognizable sequences are exactly the p-recognizable sequences.

In 1969, A. Cobham proved Theorem 7.7 in the case of subsets M of $\mathbb{N}$. In a footnote, he mentioned with no reference, the weaker theorem established independently by J. Nievergelt, that "a set which is recognizable in n-ary notation for all $n \geqslant 2$ is necessarily ultimately periodic". The property proved by Büchi, that P_p is definable in $\langle \mathbb{N}, +, P_q \rangle$ if and only if p, q are multiplicatively dependent, helped Cobham to formulate his theorem in the more general situation of sets $M \subseteq \mathbb{N}$ both p- and q-recognizable, with p, q multiplicatively independent integers.

Cobham's proof is difficult. It is based on a deep analysis of p- and q-automata computing the characteristic sequence $\mathbf{m}$ of M. The proof is divided into two parts.

The first part states that if $M \subseteq \mathbb{N}$ is p- and q-recognizable, with p, q being multiplicatively independent, then $\mathbf{m}$ is *syndetic*, that is there exists $v \geqslant 1$ such that

$$\exists n_0, \forall n \geqslant n_0, \exists v', 0 < v' \leqslant v, m_n = m_{n+v'} \ .$$

In other words, the distance between two consecutive occurrences of a given letter in the sequence is bounded by v. The property of syndeticity is weaker than the property of ultimate periodicity. Indeed the Thue-Morse sequence is syndetic but not ultimately periodic (see Section 4.5).

The second part of the proof is more technical. Cobham proves that the sequence $\mathbf{m}$ is ultimately periodic, by extracting the period of the sequence from the automata.

In his book [24], S. Eilenberg devotes Chapter 5 to the study of integers and the different notions of recognizability. He mentioned Cobham's theorem without proof but with the comment "It is a challenge to find a more reasonable proof of this fine theorem" (p. 118).

In 1977, A. Semenov generalized Cobham's theorem to subsets M of $\mathbb{N}^m$, $m \geqslant 1$ [66]. His proof is by induction on m. The first step, $m = 1$, is Cobham's proof. Twenty-eight lemmas lead to the theorem. So, Cobham's proof together with Semenov's proof establish Theorem 7.7.

Stimulated by Eilenberg's challenge, several researchers have attempted to find a simpler proof of Theorem 7.7 in the case $m = 1$. In 1982, G. Hansel succeeded in proving Cobham's theorem in a more reasonable way [36] (see also [62, 58]). The proof is combinatorial. The first part is the same than in [15] where the characteristic sequence $\mathbf{m}$ of $M \subseteq \mathbb{N}$ is proved to be syndetic. The second part is simpler and more comprehensible. To prove that $\mathbf{m}$ is ultimately periodic, instead of directly constructing a period for the sequence as done by Cobham, Hansel used the following characterization [52, 53] : *a sequence* $\mathbf{s}$ *is ultimately periodic if and only if there exists* $l \geqslant 1$ *such that the number of recurrent factors of length* l *of the sequence, is bounded by* l (a factor is a word $s_n s_{n+1} \cdots s_m$ of consecutive letters of $\mathbf{s}$; and a factor is *recurrent* if it occurs infinitely often inside $\mathbf{s}$).

Very recently, C. Michaux and R. Villemaire gave another simpler proof [49]. Their proof has also two parts. The first part is the syndeticity of the sequence as in [15]. The second one is a proof ab absurdo. It uses all the expressive power of the structure $\langle \mathbb{N}, +, V_p \rangle$, as pointed out by Corollary 6.4. The combinatorial part of the proof is very reduced.

Up to now, there does not exist a counterpart to the notion of syndeticity for multi-dimensional sequences. In 1991, A. Muchnik gave a comprehensible proof of the theorem of Cobham-Semenov, for any $m \geqslant 1$. The proof is based on his powerful definability criterion [54] and has plenty of new ideas. It is much simpler than Semenov's proof and also gives a new proof of Cobham's theorem. The most impressive thing about Muchnik's proof is that he attacks the problem in a direct fashion but still succeeds in solving it. Actually he takes a natural number $n \in M$, transforms it from base p to base q and conversely, using the relations $\sim_{p,M}$ and $\sim_{q,M}$, in order to obtain a $n' \in M$ which is close to n. He shows in this way that M is ultimately periodic. Many researchers have tried to attack the question in this way, but there are many difficulties that only Muchnik succeeded to overcome.

The main result of [49] is extended to the multi-dimensional case in [50], using Muchnik's definability criterion. They show that if $M \subseteq \mathbb{N}^m$ is not definable in $\langle \mathbb{N}, + \rangle$, then there exists a non-syndetic 1-dimensional set L which is first-order definable in the structure $\langle \mathbb{N}, +, R_M \rangle$ (R_M is the relation of membership to M). This yields a new proof of the theorem of Cobham-Semenov.

8 Muchnik's proof

We give in this section an account as precise as possible of the work of Muchnik. This is in order to make his result available to people who do not read Russian.

8.1 Definability Criterion

We here prove in details the definability criterion stated by Muchnik in [54]. We recall its statement for sets $M \subseteq \mathbb{N}^m$:

Theorem 8.1 *Let $M \subseteq \mathbb{N}^m$. Then M is definable in $\langle \mathbb{N}, + \rangle$ if and only if M is locally periodic and if any of the sections of M is definable in $\langle \mathbb{N}, + \rangle$.*

Proof. The proof of the first implication is not given here, but in the Appendix. This implication is not necessary to the proof of the theorem of Cobham-Semenov. Only the other one is used.

For the converse, assume that all the sections of M are definable in $\langle \mathbb{N}, + \rangle$ and there exists a finite set V of periods $\overline{v} \in \mathbb{N}^m \backslash \{\overline{0}\}$ such that for some $K > |V|$ and $L \geqslant 0$, we have

$$\forall \overline{n} \in \mathbb{N}^m, |\overline{n}| \geqslant L, \exists \overline{v} \in V,\ M \text{ is } \overline{v}\text{-periodic inside } \mathcal{N}(\overline{n}, K)\ \ .$$

We prove by induction on Card V that M is definable in $\langle \mathbb{N}, + \rangle$.

(1) First we suppose that $V = \{\overline{v}\}$. Therefore, any neighbourhood has the same period $\overline{v}$. This means that M is $\overline{v}$-periodic as soon as we are far enough from the origin.

The set M is easily definable using its sections. Indeed, let $\overline{v} = (v_1, \ldots, v_m)$, we denote for any i, $1 \leqslant i \leqslant m$,

$$M_i = M_{i,L} \cup M_{i,L+1} \cup \cdots \cup M_{i,L+v_i-1}\ \ .$$

M_i is the union of the v_i consecutive sections of M where the ith component has been fixed to $L, L+1, \ldots, L+v_i-1$ respectively. We also denote for any i, $1 \leqslant i \leqslant m$, the sets

$$N_i = M_{i,0} \cup M_{i,1} \cup \cdots \cup M_{i,L-1}\ \ .$$

The sets M_i, N_i and $\bigcup_i M_i, \bigcup_i N_i$ are definable in $\langle \mathbb{N}, + \rangle$ as all the sections of M are definable by assumption.

We have

$$M = \bigcup_i N_i\ \cup\ \bigcup_i M_i {\cdot}\, \overline{v}^*$$

where $\cup, \cdot, ^*$ are the rational operations in the monoid $\mathbb{N}^m$. The set M is definable in $\langle \mathbb{N}, + \rangle$ because the set $\bigcup_i M_i {\cdot}\, \overline{v}^*$ is definable by the following formula $\varphi(\overline{x})$ (recall that $\overline{v}$ is a constant)

$$(\exists \overline{y} \in \bigcup_i M_i)(\exists j \geqslant 0)\ (\overline{x} = \overline{y} + j.\overline{v})\ \ .$$

(2) Suppose now that Card $V \geqslant 2$. Let $\overline{v} \in V$.

We show that M is definable in $\langle \mathbb{N}, + \rangle$ via the sets $\mathcal{B}(M, \overline{v})$ and $\mathcal{B}(M, -\overline{v})$. These sets will be definable in $\langle \mathbb{N}, + \rangle$ by the induction hypothesis.
The set $\mathcal{B}(M, \overline{v})$ is the "border of M in the direction $\overline{v}$" and is defined as

$$\mathcal{B}(M, \overline{v}) = \{\overline{n} \in M \mid \overline{n} + \overline{v} \notin M\} \ .$$

In the same way,

$$\mathcal{B}(M, -\overline{v}) = \{\overline{n} \in M \mid \overline{n} - \overline{v} \notin M\} \ .$$

The set $\mathcal{B}(M, -\overline{v})$ is never empty since $M \subseteq \mathbb{N}^m$.

The set M is definable in $\langle \mathbb{N}, + \rangle$ using the borders $\mathcal{B}(M, \overline{v}), \mathcal{B}(M, -\overline{v})$. Indeed, given $\overline{n} \in M$, consider the "line"

$$L = \{\overline{n} + j \cdot \overline{v} \mid j \in \mathbb{Z}\} \ .$$

This line always intersects $\mathcal{B}(M, -\overline{v})$. We have to consider two cases : this line intersects $\mathcal{B}(M, \overline{v})$ or it does not. The second case is described by the following formula $\varphi(\overline{x})$

$$(\exists \overline{y} \in \mathcal{B}(M, -\overline{v})) \ ((\exists j_1 \geqslant 0)(\overline{x} = \overline{y} + j_1.\overline{v})) \wedge ((\forall j_2 \geqslant 0)(\overline{y} + j_2.\overline{v} \notin \mathcal{B}(M, \overline{v}))) \ .$$

In the first case, $\overline{n}$ is placed between a point of $\mathcal{B}(M, -\overline{v})$ and the point of $\mathcal{B}(M, \overline{v})$ met just after, in the direction of $\overline{v}$. The corresponding formula $\phi(\overline{x})$ is

$$\begin{array}{l}(\exists \overline{y}_1 \in \mathcal{B}(M, -\overline{v}))(\exists \overline{y}_2 \in \mathcal{B}(M, \overline{v}))(\exists j_1, j_2 \geqslant 0) \\ (\overline{x} = \overline{y}_1 + j_1.\overline{v}) \wedge (\overline{y}_2 = \overline{x} + j_2.\overline{v}) \\ \wedge (\forall j)[\ (0 < j < j_1 + j_2) \rightarrow (\overline{y}_1 + j.\overline{v} \notin \mathcal{B}(M, -\overline{v}) \wedge \overline{y}_1 + j.\overline{v} \notin \mathcal{B}(M, \overline{v}))\ ] \ .\end{array}$$

Hence M is defined by the formula $\varphi(\overline{x}) \vee \phi(\overline{x})$.

It remains to show that $\mathcal{B}(M, \overline{v})$ and $\mathcal{B}(M, -\overline{v})$ satisfy the induction hypothesis. Their sections are definable in $\langle \mathbb{N}, + \rangle$ because they can be defined from the sections of M. For instance, $\overline{n}$ belongs to section $\mathcal{B}(M, \overline{v})_{i,c}$ where the ith component is fixed to c if and only if $\overline{n} \in M_{i,c}$ and $\overline{n} + \overline{v} \notin M_{i,c+v_i}$.
We now prove that $\mathcal{B}(M, \overline{v})$ is $V \backslash \{\overline{v}\}$-periodic. Let

$$K' = K - |\overline{v}| > |V \backslash \{\overline{v}\}| \text{ and } L' = L \ .$$

Let $\overline{n} \in \mathbb{N}^m, |\overline{n}| \geqslant L'$. By hypothesis, M is $\overline{w}$-periodic in $\mathcal{N}(\overline{n}, K)$, for some $\overline{w} \in V$. Consider $\mathcal{N}(\overline{n}, K')$. If $\overline{v} \neq \overline{w}$, take $\overline{m}$ and $\overline{m} + \overline{w}$ in $\mathcal{N}(\overline{n}, K')$. Then

$$\overline{m}, \overline{m} + \overline{w} \quad \text{and} \quad \overline{m} + \overline{v}, \overline{m} + \overline{w} + \overline{v}$$

all belong to $\mathcal{N}(\overline{n}, K)$ inside which M is $\overline{w}$-periodic. It follows that $\mathcal{B}(M, \overline{v})$ is $\overline{w}$-periodic inside $\mathcal{N}(\overline{n}, K')$. If $\overline{v} = \overline{w}$, then $\mathcal{B}(M, \overline{v})$ is empty inside $\mathcal{N}(\overline{n}, K')$ and then $\overline{w}$-periodic in $\mathcal{N}(\overline{n}, K')$ for any $\overline{w} \in V \backslash \{\overline{v}\}$.
In the same way, one proves that $\mathcal{B}(M, -\overline{v})$ is $V \backslash \{\overline{v}\}$-periodic with $K' = K - |\overline{v}|$ and $L' = L + |\overline{v}|$.

∎

We now give the proof of Proposition 7.6 mentioned in Section 7.3, stating that it is decidable whether a p-definable sequence is definable in $\langle \mathbb{N}, + \rangle$.

Proposition 8.2 *Let $m \geqslant 1$ and $p \geqslant 2$. Let $\mathbf{s} : \mathbb{N}^m \to A$ be a p-recognizable sequence. Then it is decidable whether $\mathbf{s}$ is definable in $\langle \mathbb{N}, + \rangle$.*

Proof. By Proposition 5.2, it is sufficient to give the proof for p-recognizable sets $M \subseteq \mathbb{N}^m$.

We first consider the one-dimensional case $m = 1$. Let $\varphi_M(x)$ be a formula of $\langle \mathbb{N}, +, V_p \rangle$ defining M. The characteristic sequence $\mathbf{m}$ of M is ultimately periodic if and only if

$$(\exists t)(\exists z)(\forall x)[\ (x \geqslant z \wedge \varphi_M(x)) \to \varphi_M(x+t)\ ]\ .$$

This is a sentence of $\langle \mathbb{N}, +, V_p \rangle$. As $Th(\langle \mathbb{N}, +, V_p \rangle)$ is a decidable theory, it is decidable whether this sentence is true, i.e., whether $\mathbf{m}$ is ultimately periodic.

We now treat the two-dimensional case. Let $M \subseteq \mathbb{N}^2$ be a p-recognizable set and $\varphi_M(x, y)$ a formula defining M in $\langle \mathbb{N}, +, V_p \rangle$. The proof is the same : again we want to find a formula in $\langle \mathbb{N}, +, V_p \rangle$ which expresses that M is definable in $\langle \mathbb{N}, + \rangle$. By the definability criterion, M is definable in $\langle \mathbb{N}, + \rangle$ if and only if M is locally periodic and all its sections are definable in $\langle \mathbb{N}, + \rangle$.

Consider the second condition. Sections of M have dimension 1. So they are definable in $\langle \mathbb{N}, + \rangle$ if and only if their characteristic sequences are ultimately periodic. As we did before, the following formula of $\langle \mathbb{N}, +, V_p \rangle$

$$(\forall y)(\exists t)(\exists z)(\forall x)[\ (x \geqslant z \wedge \varphi_M(x, y)) \to \varphi_M(x+t, y)\ ]$$

states that all sections of M, when fixing the second component y, are definable in $\langle \mathbb{N}, + \rangle$. The same kind of formula exists for sections of M, when fixing the first component.

The local periodicity of M is also definable in $\langle \mathbb{N}, +, V_p \rangle$. In the definition of local periodicity, the finite set V of periods and the integer $K > |V|$ cannot be directly expressed in $\langle \mathbb{N}, +, V_p \rangle$. However, we first notice that the condition $\exists K > |V|$ can be replaced by the stronger condition $\forall K$ (by checking the proof of the definability criterion). Secondly, the set V can be replaced by the finite set of $\overline{v} \in \mathbb{N}^m$ such that $|\overline{v}| \leqslant d$ for some constant d. Hence M is locally periodic if and only if

$$(\exists d)(\forall K)(\exists L)(\forall \overline{n} \in \mathbb{N}^m, |\overline{n}| \geqslant L)(\exists \overline{v}, |\overline{v}| \leqslant d) \qquad M \text{ is } \overline{v}\text{-periodic in } \mathcal{N}(\overline{n}, K)\quad .$$

This new condition is definable in $\langle \mathbb{N}, +, V_k \rangle$: the norm $|\ |$ and the relation $\in \mathcal{N}(\overline{n}, K)$ are both definable, the $\overline{v}$-periodicity is also definable as follows

$$(\forall \overline{x})[\ (\ \overline{x} \in \mathcal{N}(\overline{n}, K)\ \wedge\ \overline{x} + \overline{v} \in \mathcal{N}(\overline{n}, K)\ )\ \to\ (\ \varphi_M(\overline{x}) \leftrightarrow \varphi_M(\overline{x} + \overline{v})\ )\ ]\ .$$

So it is possible to say with a sentence of $\langle \mathbb{N}, +, V_p \rangle$ that M is definable in $\langle \mathbb{N}, + \rangle$. We can decide if this sentence is true, since $Th\langle \mathbb{N}, +, V_p \rangle$ is decidable.

The previous cases $m = 1$ and $m = 2$ show how the general case works. The proof is by induction on m. The basis of the induction is proved above. Let $m \geqslant 2$ be a fixed integer. Let M be a p-recognizable subset of $\mathbb{N}^m$ defined by some formula $\varphi_M(\overline{x})$ of $\langle \mathbb{N}, +, V_p \rangle$. We express as before the local periodicity property. A formula stating that the sections of M are definable exists by induction hypothesis, as the sections have dimension $m-1$. Therefore, there exists a sentence of $\langle \mathbb{N}, +, V_p \rangle$ which says that M satisfies the definability criterion. This sentence is certainly complex but we can decide if it is true or false.

■

8.2 Proof of the Theorem of Cobham-Semenov

In this section, we give Muchnik's proof for the theorem of Cobham-Semenov (Theorem 7.7). The proof is not exactly the same as in [54]; some parts are modified or shortened and other ones are given in more detail. It uses the definability criterion stated in Theorem 8.1.

Theorem 8.3 *Let $m \geqslant 1$, let $p, q \geqslant 2$ be multiplicatively independent integers. If $M \subseteq \mathbb{N}^m$ is p- and q-recognizable, then its characteristic sequence is definable in $\langle \mathbb{N}, + \rangle$.*

We first recall properties of the relation $\sim_{p,M}$ associated with any set $M \subseteq \mathbb{N}^m$ (see Section 4.3 or [24]). Let $\overline{n}, \overline{m} \in \mathbb{N}^m$, then

$$\overline{n} \sim_{p,M} \overline{m} \quad \Leftrightarrow \quad [\ \overline{n}p^k + \overline{r} \in M \Leftrightarrow \overline{m}p^k + \overline{r} \in M \quad \forall k \geqslant 0, \forall \overline{r} \in \mathbb{N}^m, |\overline{r}| < p^k\ ] \ .$$

Set M is p-recognizable if and only if $\sim_{p,M}$ has finite index, M is a union of some equivalence classes of $\sim_{p,M}$. The relation $\sim_{p,M}$ is *p-stable*, i.e.,

$$\overline{n} \sim_{p,M} \overline{m} \quad \Rightarrow \quad \overline{n}p^k + \overline{r} \sim_{p,M} \overline{m}p^k + \overline{r}$$

for $|\overline{r}| < p^k$.

The following lemmas will be useful.

Lemma 8.4 *Let $m \geqslant 1$ and $p, q \geqslant 2$. If $M \subseteq \mathbb{N}^m$ is p- and q-recognizable, then any equivalence class C of $\sim_{p,M}$ is also p- and q-recognizable.*

Proof. To simplify the notations, we limit the proof to the case $m = 1$.

From the properties above, C is clearly p-recognizable. Consider the minimal automaton $\mathcal{A}(L) = (Q, \{q_0\}, F, T)$ of the set

$$L = \{\ w \in \{0, \ldots, p-1\}^* \ \mid \ [w]_p \in M\ \} \ .$$

The relation $\sim_{p,M}$ is the translation to $\mathbb{N}$ of the right-congruence $\sim_L$ defined on $\{0, \ldots, p-1\}^*$. The class C of $\sim_{p,M}$ corresponds to a class C_L of $\sim_L$. By construction of $\mathcal{A}(L)$, C_L is some of its states, that we denote by r. Moreover, for any state $r' \neq r$, there exists a word u, *depending on* r', such that $T(r, u)$ is final and $T(r', u)$ is not, or the opposite (see Section 2). We then define the subset of $\{0, \ldots, p-1\}^*$

$$\begin{array}{rcll} L(r') & = & \{w \mid wu \in L\} & \text{if } T(r,u) \text{ is final} \ , \\ & = & \{w \mid wu \notin L\} & \text{otherwise} \ . \end{array}$$

One verifies that

$$C_L = \bigcap_{r' \neq r} L(r') \ .$$

Coming back to $\mathbb{N}$, this gives

$$C = \bigcap_{r' \neq r} M(r') \ .$$

where $M(r') = \{[w]_p \mid w \in L(r')\}$. Each set $M(r')$ is q-recognizable using Corollary 6.4. Indeed M is q-definable and $wu \in L$ if and only if $[w]_p \cdot p^{|u|} + [u]_p \in M$ ($p^{|u|}$ is a constant). So C is itself q-recognizable.

■

Lemma 8.5 *Let $p, q \geqslant 2$ be multiplicatively independent integers. Then for any real numbers r_1, r_2 with $0 \leqslant r_1 < r_2$, there exist arbitrarily large integers k, l such $r_1 < q^l/p^k < r_2$.*

The proof of this lemma can be found in [58] for example. It is based on Kronecker's theorem which states that if θ is an irrational number, then the fractional parts of its multiples $n\theta, n \geqslant 0$, are dense in the interval $[0, 1]$ (see [38]).

We are now going to prove Theorem 8.3.

Proof.

The idea of the proof is the following.
We have to prove that M is locally periodic and all the sections of M are definable in $\langle \mathbb{N}, + \rangle$. The proof will use induction on m; this is necessary to prove the definability of the sections. However, the property of local periodicity will be proved directly, independently of the induction.
We first show that M is locally periodic but for the p^k-neighbourhoods $\mathcal{N}(\overline{n}p^k, p^k)$ of $\overline{n}p^k$ only, (for some well-chosen power p^k). This will be possible using an elaborate equivalence relation $\sim_{p,q,M}$ defined from the relations $\sim_{p,M}$ and $\sim_{q,M}$.
In a way to reach all the neighbourhoods, we repeat the previous argument to the following subset M' of $\mathbb{N}^{2m}$, instead of M :

$$M' = \{ \ (\overline{n}_1, \overline{n}_2) \ \mid \ \overline{n}_1 + \overline{n}_2 \in M \ \} \ .$$

A well-chosen projection on $\mathbb{N}^m$ of the neighbourhoods $\mathcal{N}((\overline{n}_1, \overline{n}_2)p^k, p^k)$ in $\mathbb{N}^{2m}$ will yield all the neighbourhoods of $\mathbb{N}^m$.
We now go into the details of the proof.

(A) By Lemma 8.4, any equivalence class C of $\sim_{p,M}$ is p- and q-recognizable. In particular, any $\sim_{q,C}$ has finite index. We define a new equivalence relation $\sim_{p,q,M}$ over $\mathbb{N}^m$, from the relations $\sim_{p,M}$ and $\sim_{q,C}$:

$$\overline{n} \sim_{p,q,M} \overline{m} \quad \Leftrightarrow \quad [\ \overline{n} \sim_{q,C} \overline{m}, \text{ for all classes } C\ ] \ .$$

In other words, $\sim_{p,q,M}$ is the finite intersection (over the classes C of $\sim_{p,M}$) of the relations $\sim_{q,C}$. The new relation has interesting properties : $\sim_{p,q,M}$ has finite index and

1. $\overline{n} \sim_{p,q,M} \overline{m} \quad \Rightarrow \quad \overline{n} \sim_{p,M} \overline{m}$.
2. $\overline{n} \sim_{p,q,M} \overline{m} \quad \Rightarrow \quad \overline{n} \sim_{q,M} \overline{m}$.

The first property is proved as follows. Let C be the equivalent class of $\sim_{p,M}$ containing $\overline{n}$. Then

$$\overline{n} \sim_{p,q,M} \overline{m} \quad \Rightarrow \quad \overline{n} \sim_{q,C} \overline{m} \quad \Rightarrow \quad \overline{m} \in C \ .$$

And if $\overline{n}, \overline{m} \in C$, then $\overline{n} \sim_{p,M} \overline{m}$. For the second property, the proof is the following one. If $\overline{n} \sim_{p,q,M} \overline{m}$, then $\overline{n} \sim_{q,C} \overline{m}$ for all classes C of $\sim_{p,M}$, and also $\overline{n}q^l + \overline{r} \sim_{q,C} \overline{m}q^l + \overline{r}$. Thus

$$\overline{n}q^l + \overline{r} \in M \quad \Rightarrow \quad \overline{n}q^l + \overline{r} \in C \text{ for some } C \subseteq M \quad \Rightarrow \quad \overline{m}q^l + \overline{r} \in C \subseteq M \ .$$

The conclusion follows.

(B) We here prove that M is locally periodic but for some neighbourhoods only. These neighbourhoods have the particular form $\mathcal{N}(\overline{n}p^k, p^k)$.

(a) For any infinite class D of $\sim_{p,q,M}$ we choose a pair $(\overline{x}_D, \overline{y}_D)$ of elements of D, simply denoted $(\overline{x}, \overline{y})$, such that

1. for all $\overline{\alpha} \in \mathbb{Z}^m$ with $|\overline{\alpha}| \leqslant 1$, if $\overline{x} + \overline{\alpha}, \overline{y} + \overline{\alpha} \in \mathbb{N}^m$, then $\overline{x} + \overline{\alpha} \sim_{p,q,M} \overline{y} + \overline{\alpha}$,
2. there exists $\overline{n} \in \mathbb{N}^m \backslash \{\overline{0}\}$ such that $\overline{x} + \overline{n} = \overline{y}$, i.e., $\overline{x} < \overline{y}$.

This pair $(\overline{x}, \overline{y})$ always exists because $\sim_{p,q,M}$ has finite index and any set of $\mathbb{N}^m$ of mutually incomparable elements is finite [25]. More precisely, given $\overline{z} \in D$, for any $\overline{\alpha} \in \mathbb{Z}^m$ such that $|\overline{\alpha}| \leqslant 1$ and $\overline{z} + \overline{\alpha} \in \mathbb{N}^m$, let $D_{\overline{\alpha}}$ be the class of $\sim_{p,q,M}$ containing $\overline{z} + \overline{\alpha}$. Now list these $\overline{\alpha}$'s and the related classes $D_{\overline{\alpha}}$. As D is infinite, there are infinitely many $\overline{z} \in D$ with the same list of classes $D_{\overline{\alpha}}$. We choose $\overline{x}$ and $\overline{y}$ among them.
Then using Lemma 8.5, we choose k and l such that $1 < q^l/p^k < 1+\epsilon$, or equivalently

$$0 < q^l - p^k < \epsilon p^k \tag{1}$$

where ϵ satisfies the following condition : for any chosen $\overline{y} = (y_1, \ldots, y_m)$, for all $1 \leqslant i \leqslant m$,

$$(1 + y_i)\epsilon < 1 \ . \tag{2}$$

In particular, as $\overline{x} < \overline{y}$, we have

$$x_i \epsilon < 1 \ . \tag{3}$$

(b) We first prove that $(\overline{y} - \overline{x})(q^l - p^k)$ is a period for M inside the particular neighbourhood $\mathcal{N}(\overline{x}p^k, p^k)$: precisely for all $\overline{m} \in \mathcal{N}(\overline{x}p^k, p^k)$,

$$\overline{m} \in M \quad \Leftrightarrow \quad \overline{m} + (\overline{y} - \overline{x})(q^l - p^k) \in M \ .$$

Let $\overline{m} = \overline{x}p^k + \overline{r} \in \mathcal{N}(\overline{x}p^k, p^k)$. We define $\overline{x}' = (x'_1, \ldots, x'_m)$ and $\overline{r}' = (r'_1, \ldots, r'_m)$ in $\mathbb{N}^m$ such that for any $1 \leqslant i \leqslant m$, x'_i and r'_i are respectively the quotient and the remainder of the division of $x_i p^k + r_i$ by q^l. Then

$$\overline{x}'q^l + \overline{r}' = \overline{x}p^k + \overline{r} \quad |\overline{r}'| < q^l \ . \tag{4}$$

We have

$$\overline{x}' = \overline{x} + \overline{\alpha} \quad \text{with } |\overline{\alpha}| \leqslant 1 \text{ and } \overline{\alpha} \leqslant \overline{0} \ .$$

Indeed as $q^l > p^k$, then $x'_i \leqslant x_i$ for any $1 \leqslant i \leqslant m$, by (1) and (4). On the other hand, $x_i - 1 \leqslant x'_i$ otherwise $x'_i \leqslant x_i - 2$ and by (1), (3),

$$x'_i q^l + r'_i < (x_i - 2)q^l + q^l \leqslant x_i p^k + x_i(q^l - p^k) - q^l < x_i p^k \ ,$$

which is impossible by (4). Therefore by (4)

$$(\overline{x} + \overline{\alpha})q^l + \overline{r}' = \overline{x}p^k + \overline{r} \ . \tag{5}$$

We have $\overline{x} + \overline{\alpha}, \overline{y} + \overline{\alpha} \in \mathbb{N}^m$ and as $\overline{x} + \overline{\alpha} \sim_{p,q,M} \overline{y} + \overline{\alpha}$,

$$\overline{x}p^k + \overline{r} \in M \quad \Leftrightarrow \quad (\overline{y} + \overline{\alpha})q^l + \overline{r}' \in M \tag{6}$$

using (5) and property 2 of $\sim_{p,q,M}$.
Now let $\overline{y}' = (y'_1, \ldots, y'_m)$ and $\overline{s} = (s_1, \ldots, s_m)$ in $\mathbb{N}^m$ be such that for any $1 \leqslant i \leqslant m$, y'_i and s_i are respectively the quotient and the remainder of the division of $(y_i + \alpha_i)q^l + r'_i$ by p^k, that is

$$\overline{y}' p^k + \overline{s} = (\overline{y} + \overline{\alpha})q^l + \overline{r}' \quad |\overline{s}| < p^k \ . \tag{7}$$

We have

$$\overline{y}' = \overline{y} + \overline{\beta} \quad \text{with } |\overline{\beta}| \leqslant 1 \text{ and } \overline{\alpha} \leqslant \overline{\beta} \ .$$

Indeed as $p^k < q^l$, then $y_i + \alpha_i \leqslant y'_i$ for all $1 \leqslant i \leqslant m$, by (1) and (7). We have also $y'_i \leqslant y_i + 1$ otherwise $y_i + 2 \leqslant y'_i$ and by (1), (2) (remember $\overline{\alpha} \leqslant \overline{0}$),

$$y'_i p^k + s_i \geqslant (y_i + 2)p^k \geqslant (y_i + \alpha_i)q^l + y_i(p^k - q^l) + 2p^k > (y_i + \alpha_i)q^l + q^l \ ,$$

which is in contradiction with (7). Hence by (7)

$$(\overline{y} + \overline{\beta})p^k + \overline{s} = (\overline{y} + \overline{\alpha})q^l + \overline{r}' \ . \tag{8}$$

As $\overline{y} + \overline{\beta}, \overline{x} + \overline{\beta} \geqslant \overline{x} + \overline{\alpha} \in \mathbb{N}^m$ and as $\overline{y} + \overline{\beta} \sim_{p,q,M} \overline{x} + \overline{\beta}$, it follows from property 1 of $\sim_{p,q,M}$ that

$$(\overline{y} + \overline{\alpha})q^l + \overline{r}' \in M \quad \Leftrightarrow \quad (\overline{x} + \overline{\beta})p^k + \overline{s} \in M \ .$$

Together with (6), this yields

$$\overline{x}p^k + \overline{r} \in M \quad \Leftrightarrow \quad (\overline{x} + \overline{\beta})p^k + \overline{s} \in M \quad \Leftrightarrow \quad \overline{x}p^k + \overline{r} + (\overline{y} - \overline{x})(q^l - p^k) \in M \ .$$

Indeed using (8) and (5)

$$\begin{aligned} \overline{\beta}p^k + \overline{s} - \overline{r} &= \overline{\beta}p^k + (\overline{y} + \overline{\alpha})q^l - (\overline{y} + \overline{\beta})p^k + \overline{r}' - \overline{r} \\ &= (\overline{y} + \overline{\alpha})q^l - \overline{y}p^k + \overline{x}p^k - (\overline{x} + \overline{\alpha})q^l \\ &= (\overline{y} - \overline{x})(q^l - p^k) \ . \end{aligned}$$

(c) For any infinite class D of $\sim_{p,q,M}$, we denote by $\overline{v}_D$ the period $(\overline{y} - \overline{x})(q^l - p^k)$. The relation $\sim_{p,q,M}$ has finite index, then there exists $U \geqslant 0$ such that, as soon as $|\overline{n}| \geqslant U$, the class D of $\sim_{p,q,M}$ containing $\overline{n}$ is infinite. Let us now show that M is $\overline{v}_D$-periodic inside the p^k-neighbourhood $\mathcal{N}(\overline{n}p^k, p^k)$ of $\overline{n}$. Let $\overline{m}, \overline{m} + \overline{v}_D \in \mathcal{N}(\overline{n}p^k, p^k)$.

More precisely $\overline{m} = \overline{n}p^k + \overline{r}$ with $|\overline{r}| < p^k$ and $|\overline{r} + \overline{v}_D| < p^k$. Using (b), we have the following equivalences

$$\begin{aligned} \overline{m} \in M &\Leftrightarrow \overline{n}p^k + \overline{r} \in M \\ &\Leftrightarrow \overline{x}_D p^k + \overline{r} \in M \\ &\Leftrightarrow \overline{x}_D p^k + \overline{r} + \overline{v}_D \in M \\ &\Leftrightarrow \overline{n}p^k + \overline{r} + \overline{v}_D \in M \\ &\Leftrightarrow \overline{m} + \overline{v}_D \in M \ . \end{aligned}$$

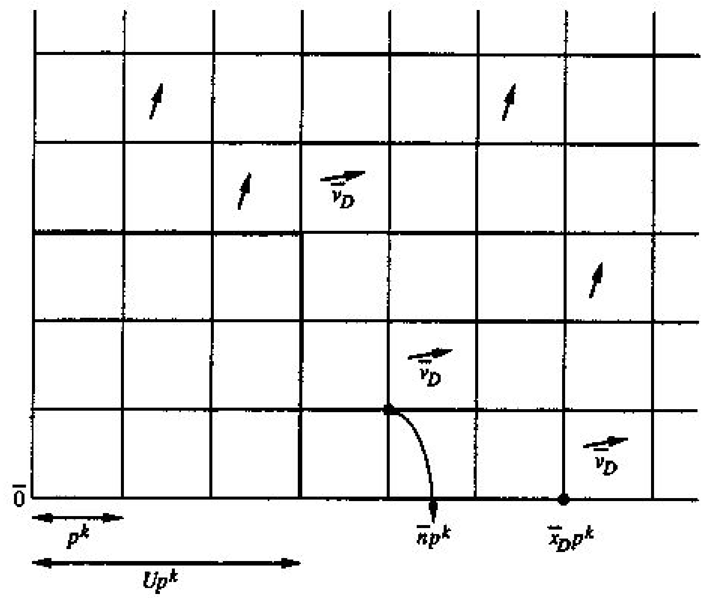

Figure 12. The local periodicity of M inside the neighbourhoods $\mathcal{N}(\overline{n}p^k, p^k)$

(C) To prove the local periodicity of M inside all the neighbourhoods, we transfer the results of (B) to the subset of $\mathbb{N}^{2m}$

$$M' = \{(\overline{n}_1, \overline{n}_2) \mid \overline{n}_1 + \overline{n}_2 \in M\}$$

which is p- and q-recognizable by Corollary 6.4. We add the condition on ϵ that

$$\epsilon < \frac{1}{6 \sum |\overline{y}_D|} \ . \tag{9}$$

We denote by V' the finite set of periods $\overline{v}_D = (\overline{y} - \overline{x})(q^l - p^k)$. We are going to show that M is locally periodic with the set of periods

$$V = \{\overline{v} = \overline{v}_1 + \overline{v}_2 \mid (\overline{v}_1, \overline{v}_2) \in V'\}$$

and parameters $K = \lceil p^k/3 \rceil$ and $L = Up^k$ (the constant U was defined in (c)). We first verify that $|V| < K$ with (1) and (9).

$$|V| = \sum_{(\overline{v}_1, \overline{v}_2) \in V'} |\overline{v}_1 + \overline{v}_2| \leqslant 2 \sum_{\overline{v}_D \in V'} |\overline{v}_D| \leqslant 2(q^l - p^k) \sum |\overline{y}_D - \overline{x}_D| < p^k/3 \ .$$

Let $\overline{u} \in \mathbb{N}^m, |\overline{u}| \geqslant L$. We write $\overline{u}$ as $\overline{u} = \overline{n}p^k + \overline{r}, \quad |\overline{r}| < p^k$. We decompose $\overline{r}$ in $\mathbb{N}^m$ such that $\overline{r} = \overline{r}_1 + \overline{r}_2$ with $|\overline{r}_1| < 2p^k/3$ and $|\overline{r}_2| \leqslant p^k/3$ (see Figure 13).

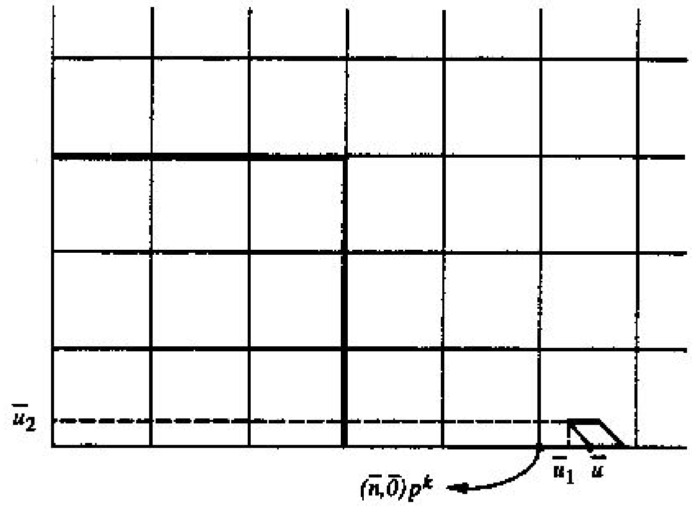

Figure 13. Projection from $\mathbb{N}^{2m}$ to $\mathbb{N}^m$

Then

$$\overline{u} = \overline{u}_1 + \overline{u}_2$$

with $\overline{u}_1 = \overline{n}p^k + \overline{r}_1, \overline{u}_2 = \overline{r}_2$. This corresponds in $\mathbb{N}^{2m}$ to the point

$$(\overline{u}_1, \overline{u}_2) = (\overline{n}, \overline{0})p^k + (\overline{r}_1, \overline{r}_2) \ .$$

As $|(\overline{n}, \overline{0})| \geqslant U$, there exists a period $\overline{v}_D = (\overline{v}_1, \overline{v}_2)$ in V' for which M' is $\overline{v}_D$-periodic inside $\mathcal{N}((\overline{n}, \overline{0})p^k, p^k)$. We will prove that this implies, in $\mathbb{N}^m$, that M is $(\overline{v} = \overline{v}_1 + \overline{v}_2)$-periodic inside $\mathcal{N}(\overline{u}, K)$. Let $\overline{u} + \overline{t}$ and $\overline{u} + \overline{t} + \overline{v}$ in $\mathcal{N}(\overline{u}, K)$. As $|\overline{v}_D| < p^k/3$ by (1) and (9), the points

$$\begin{aligned}
(\overline{u}_1, \overline{u}_2 + \overline{t}) &= (\overline{n}, \overline{0})p^k + (\overline{r}_1, \overline{r}_2 + \overline{t}) \\
(\overline{u}_1, \overline{u}_2 + \overline{t}) + (\overline{v}_1, \overline{v}_2) &= (\overline{n}, \overline{0})p^k + (\overline{r}_1 + \overline{v}_1, \overline{r}_2 + \overline{t} + \overline{v}_2)
\end{aligned}$$

both belong to $\mathcal{N}((\overline{n}, \overline{0})p^k, p^k)$. Consequently

$$\overline{u} + \overline{t} \in M \Leftrightarrow (\overline{u}_1, \overline{u}_2 + \overline{t}) \in M' \Leftrightarrow (\overline{u}_1, \overline{u}_2 + \overline{t}) + (\overline{v}_1, \overline{v}_2) \in M' \Leftrightarrow \overline{u} + \overline{t} + \overline{v} \in M \ .$$

Hence (B) and (C) show that M is locally periodic.

(D) We now finish the proof. If $m = 1$, then M is locally periodic and any of its sections is a constant which is trivially definable in $\langle \mathbb{N}, + \rangle$. Let $m > 1$. Again M is locally periodic. Its sections are definable in $\langle \mathbb{N}, + \rangle$ by the induction hypothesis. Indeed, any section has dimension $m - 1$ and it is p- and q-recognizable by Corollary 6.4.

■

9 Related Work

There exist different possible generalizations of the notion of p-recognizable sequence (see the survey [1]). One generalization is replacing p-substitutions $f : B \to B^p$ by *nonuniform* substitutions (the images of the letters of B have different lengths). One well-known example is the Fibonacci substitution defined on $B = \{0, 1\}$ by $f(0) = 01, f(1) = 0$. This is related to *nonstandard representations* of numbers generalizing p-ary expansions. For instance, any positive integer can be written as $\sum a_n f_n$ where f_n is the n-th Fibonacci number and a_n is 0 or 1. Nonstandard representations of numbers are studied in [56, 30, 6]. For connections between nonuniform substitutions and nonstandard representations of integers, see [61, 70, 29]. See also [17, 47, 41] for decision problems related to generalized number systems. Another generalization of p-recognizable sequence is the notion of *p-regular* sequence introduced in [2]. A sequence $\mathbf{s}$ has its values s_n in a noetherian (possibly infinite) ring R instead of a finite alphabet A. It is p-regular if and only if the R-module generated by its p-kernel $\{(s_{np^k+r})_{n\geqslant 0} \mid k \geqslant 0, r < p^k\}$, is finitely generated. The reference [2] contains the following conjecture : if a sequence $\mathbf{s}$ is p- and q-regular, and p and q are multiplicatively independent, then the formal power series $\sum_{n\geqslant 0} s_n x^n$ is rational.

In Section 7, we classified the sequences into three groups, the last one containing sequences recognizable in no base. The first proofs of unrecognizability of the set of squares (in any base) are logical ones [9, 26]. Later, Ritchie gives a more direct combinatorial proof [63]. The references [51, 16, 24] introduce methods based on the asymptotic behavior of functions dealing with the gaps between successive elements of $M \subseteq \mathbb{N}$. These *gap theorems* easily show that the set of squares or of prime numbers are never p-recognizable, for any $p \geqslant 2$. For properties of star-height 0 of p-recognizable sequences for suitable bases p, see also the work [22].

In the present paper, we considered the free monoid $(\{0, 1, \ldots, p-1\}^m)^*$ for defining p-recognizable sets of $\mathbb{N}^m$. The monoid $(\{0, 1, \ldots, p-1\}^*)^m$, which is not free, could be used for defining another concept of p-recognizable subset of $\mathbb{N}^m$. For $m = 2$, $M \subseteq \mathbb{N}^2$ is called p-recognizable in this context if and only if the set $L = \{((n)_p, (m)_p) \mid (n, m) \in M\}$ is a *recognizable* subset of $\{0, 1, \ldots, p-1\}^* \times \{0, 1, \ldots, p-1\}^*$. Mezei's theorem then states that L is a finite union of sets $L_1 \times L_2$ where L_1, L_2 are recognizable sets of $\{0, 1, \ldots, p-1\}^*$ (see [24, p. 68]). This leads to another version of the theorem of Cobham-Semenov [37] : *Let $p, q \geqslant 2$ be two multiplicatively independent integers, if $M \subseteq \mathbb{N}^2$ is p- and q-recognizable, then M is a recognizable subset of* $\mathbb{N}^2$ (and not only a rational subset of $\mathbb{N}^2$ as in Theorem 7.7).

Rational (rather than recognizable) subsets of the monoid $\{0, \ldots, p-1\}^* \times \{0, \ldots, p-1\}^*$ are also much studied in relation with nonstandard representations of numbers [31, 32]. The related automata, called *transducers*, have their transitions labelled by pairs of words $(u, v) \in \{0, \ldots, p-1\}^* \times \{0, \ldots, p-1\}^*$, rather than by pairs of letters as done in this paper (see [24]). For the Fibonacci base $(f_n)_{n\geqslant 0}$ for example, a number can have several representations; the function of normalization which transforms any representation into the *normal* one (obtained by the usual algorithm) can be realized by a finite transducer [31].

The study of p-recognizable sequences leads to interesting transcendence results in number theory. If p is a prime number and $\mathbf{s}$ a p-algebraic sequence which is not ultimately periodic, then the real number $\sum s_n p^{-n}$ is *transcendental* [45]. This gives a method to easily generate transcendental numbers, as $\sum 2^{-2^n}$ for example. See [18] for connections between number theory and finite automata.

In the context of algebraic formal power series, other interesting results concern *diagonals* of series [33, 19, 23]. Among them, one states that for a finite field K, the series $S(x) \in K[[x]]$ is algebraic over $K[x]$ if and only if it is the diagonal of some *rational* formal power series $T(x,y) = \sum t_{n,m} x^n y^m \in K[[x,y]]$, i.e., $S(x) = \sum t_{n,n} x^n$. Another result is : let K be a field with characteristic $p \neq 0$, if $S(x,y) \in K[[x,y]]$ is algebraic over $K[x,y]$, then its diagonal is also algebraic (this is also true for more than 2 variables x, y) [19]. When K is finite, this second theorem follows from the works of [64], it is also a corollary of Theorem 5.1. Indeed the sequence $\mathbf{s}$ associated with $S(x,y)$ is p-definable by a formula $\varphi(x,y)$ of $\langle \mathbb{N}, +, V_p \rangle$. The formula $\varphi(x,x)$ of $\langle \mathbb{N}, +, V_p \rangle$ shows that the diagonal of $S(x,y)$ is algebraic.

The theories of the structures $\langle \mathbb{N}, + \rangle$ and $\langle \mathbb{N}, +, V_p \rangle$ are decidable. It is no longer true for the structure $\langle \mathbb{N}, +, V_p, V_q \rangle$ with p and q multiplicatively independent integers. The structure is indeed equivalent to $\langle \mathbb{N}, +, \cdot \rangle$ [73, 74]. This undecidability result is in a certain way a counterpart to the theorem of Cobham-Semenov, as indicated by Figure 14. See also [21] for related work.

$$\begin{tikzcd}
 & \langle \mathbb{N}, +, V_p \rangle \arrow[dr] & \\
\langle \mathbb{N}, + \rangle \arrow[ur] \arrow[dr] & & \langle \mathbb{N}, +, V_p, V_q \rangle \\
 & \langle \mathbb{N}, +, V_q \rangle \arrow[ur] &
\end{tikzcd}$$

Figure 14. The theory $Th(\langle \mathbb{N}, +, V_p, V_q \rangle)$ is undecidable

In the same spirit, the decidability of the weak-monadic second-order theory $\langle \mathbb{N}, S \rangle$ extended by one relation or by one function, is studied in [27, 69]. Semenov proves that this decision problem comes back to rather special decision problems for rational sets [69] (see also [4] for a complete proof). For instance, the weak-monadic second-order structure $\langle \mathbb{N}, S \rangle$ enriched by the function $y = 2x$, already allows to interpret all the first-order formulae of $\langle \mathbb{N}, +, \cdot \rangle$ and thus is an undecidable theory [27].

Several papers study in details rational subsets of $\mathbb{N}^m$ (those which are semi-linear or equivalently definable in $\langle \mathbb{N}, + \rangle$ according to Theorem 7.4). These sets are *unambigously rational*. In other words they are finite unions of points and of disjoint cones, with cones generated by linearly independent vectors [25, 44]. This property generally holds in any commutative monoid [25]. Rational subsets of $\mathbb{N}^m$ are classified and characterized in [57] with respect to logical formulae. Another hierarchy based on rational transductions is proposed for $\mathbb{N}^2$ in [5]. See also the reference [34] for a theorem of Fine and Wilf in $\mathbb{N}^2$.

Appendix

We prove in this appendix that *any $M \subseteq \mathbb{N}^m$ definable in $\langle \mathbb{N}, + \rangle$ is locally periodic and has all its sections definable in $\langle \mathbb{N}, + \rangle$.*

The condition on the sections is trivial (see Corollary 6.4).

By a classical result [59], we can suppose that the formula $\varphi(x_1,\ldots,x_m)$ defining M is a Boolean combination of the formulae

$$\begin{array}{rclll} t_i(x_1,\ldots,x_m) & \geqslant & c_i & & 1\leqslant i\leqslant r\ , \\ t_j(x_1,\ldots,x_m) & = & e_j & \bmod d & r<j\leqslant s\ , \end{array}$$

where c_i, e_j, d, m are constants (see also Section 7.2).

Any equation $t_i(x_1,\ldots,x_m)=\sum u_{i,k}x_k = c_i$, $1\leqslant i\leqslant r$, defines an hyperplane orthogonal to the vector $\overline{u}_i=(u_{i,1},\ldots,u_{i,m})$. Formulae $t_i(x_1,\ldots,x_m)\geqslant c_i$ divide $\mathbb{N}^m$ into domains. We show how to construct a finite set V of periods for M, from these inequalities. Let I be a subset of the set of indices $\{1,\ldots,r\}$. We say that I is a *first-type* set if the family of vectors $(\overline{u}_i)_{i\in I}$ generates $\mathbb{N}^m$, otherwise I is called a *second-type* set. For any set $I\subseteq\{1,\ldots,r\}$ of second type, let $\overline{v}\in\mathbb{N}^m\backslash\{\overline{0}\}$ be such that $\overline{v}=(v_1,\ldots,v_m)$ is orthogonal to any $\overline{u}_i, i\in I$. We can suppose that the components v_j of $\overline{v}$ are all divisible by d. These vectors $\overline{v}$ form the set V (see also Figure 11).

Let $K>|V|$, K is the size of the neighbourhoods. Consider a K-neighbourhood $\mathcal{N}(\overline{n},K)$, suppose that the hyperplanes $t_i(x_1,\ldots,x_m)=c_i$ intersecting it are indexed by a set I of the first type. The intersection of these hyperplanes contains at most one point. As the size of the neighbourhood is the constant K, the intersection point is close to $\mathcal{N}(\overline{n},K)$. Moreover, in $\mathbb{N}^m$ there is a finite number of such intersection points. So it is possible to choose $L\geqslant 0$ such that any neighbourhood $\mathcal{N}(\overline{n},K)$, with $|\overline{n}|\geqslant L$, intersects a set of hyperplanes indexed by a second-type set only.

Now let $\overline{n}\in\mathbb{N}^m$, $|\overline{n}|\geqslant L$. Then, there is $\overline{v}\in V$ such that $\overline{v}$ is parallel to each of the hyperplanes intersecting $\mathcal{N}(\overline{n},K)$. Vector $\overline{v}$ is a period for $\mathcal{N}(\overline{n},K)$. Indeed let $\overline{m},\overline{m}+\overline{v}\in\mathcal{N}(\overline{n},K)$. Consider the inequality $t_i(\overline{x})=t_i(x_1,\ldots,x_m)\geqslant c_i$ associated with an hyperplane intersecting $\mathcal{N}(\overline{n},K)$. Then

$$t_i(\overline{m})\geqslant c_i \quad\Leftrightarrow\quad t_i(\overline{m}+\overline{v})\geqslant c_i$$

as $\overline{v}$ is parallel to this hyperplane. By construction, d divides all the components of $\overline{v}$. it follows that for any formula $t_j(\overline{x})=e_j \bmod d, r<j\leqslant s$, we have

$$t_j(\overline{m})=e_j \bmod d \quad\Leftrightarrow\quad t_j(\overline{m}+\overline{v})=e_j \bmod d\ .$$

This proves that $\overline{m}\in M$ if and only if $\overline{m}+\overline{v}\in M$.

Acknowledgements

We are thankful to A. Semenov who informed us about Muchnik's preprint [54]. K. Archangelsky kindly translated this paper into English. We thank J.-P. Allouche, J. Berstel, M. Boffa, C. Frougny and F. Point for fruitful discussions. We also thank D. Perrin and J. Sakarovitch for pointing out some of the references. We are grateful to the referees for valuable advices in improving the presentation of this paper.

References

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

[2] J.-P. Allouche, J. Shallit, The ring of k-regular sequences, *Theoret. Comput. Sci* **98** (1991) 163–197.

[3] J. Barwise, An introduction to first-order logic, in : *Handbook of Mathematical Logic*, J. Barwise, Ed., North Holland, Amsterdam (1977) 5–46.

[4] P.T. Bateman, C.G. Jockusch, A.R. Woods, Decidability and undecidability with a predicate for the primes, *J. Symbolic Logic* **58** (1993) 672–687.

[5] J. Berstel, Une hiérarchie des parties rationnelles de $\mathbb{N}^2$, *Math. Systems Theory* **7** (1973) 114–137.

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

[7] M. Boffa, personal communication to V. Bruyère (1985).

[8] V. Bruyère, Entiers et automates finis, *Mémoire de fin d'études*, Université de Mons (1985).

[9] J.R. Büchi, Weak second-order arithmetic and finite automata, *Z. Math. Logik Grundlag. Math.* **6** (1960) 66–92.

[10] A. Černý, J. Gruska, Modular Trellises, in : *The book of L*, G. Rozenberg and A. Salomaa, Eds., Springer Verlag (1985) 45–61.

[11] G. Cherlin, F. Point, On extensions of Presburger arithmetic, Proc. 4th Easter Model Theory conference, Gross Köris 1986 Seminarberichte 86, Humboldt Universität zu Berlin (1986) 17–34.

[12] G. Christol, Ensembles presque periodiques k-reconnaissables, *Theoret. Comput. Sci* **9** (1979) 141-145.

[13] G. Christol, T. Kamae, M. Mendès France, G. Rauzy, Suites algébriques, automates et substitutions, *Bull. Soc. Math. France* **108** (1980) 401–419.

[14] A. Church, A note on the Entscheidungsproblem, *J. Symbolic Logic* **1** (1936) 40–41, 101–102.

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

[16] A. Cobham, Uniform tag sequences, *Math. Systems Theory* **6** (1972) 164–192.

[17] K. Culik II, A. Salomaa, Ambiguity and decision problems concerning number systems, *Inform. and Control* **56** (1983) 139–153.

[18] M. Dekking, M. Mendès-France, A.J. Van der Poorten, Folds I, II, III, *Math. Intelligencer* **4** (1982) 130–138, 173–181, 190–195.

[19] P. Deligne, Intégration sur un cycle évanescent, *Invent. Math.* **76** (1984) 129–143.

[20] F. Delon, Personal communication to C. Michaux and F. Point (1991).

[21] F. Delon, Pour l'arithmétique faible de Penzin, $\mathbb{Z}$ est indécidable et $\mathbb{Q}$ est décidable, *preprint* (1994).

[22] A. de Luca, A. Restivo, Star-free sets of integers, *Theoret. Comput. Sci.* **43** (1986) 265–275.

[23] J. Denef, L. Lipshitz, Algebraic power series and diagonals, *J. Number Theory* **26** (1987) 46–67.

[24] S. Eilenberg, *Automata, Languages and Machines*, Vol. A, Academic Press, New-York (1974).

[25] S. Eilenberg, M.-P. Schützenberger, Rational sets in commutative monoids, *J. Algebra* **13** (1969) 173–191.

[26] C.C. Elgot, Decision problems of finite automata design and related arithmetics, *Trans. Amer. Math. Soc.* **98** (1961) 21–51.

[27] C.C. Elgot, M.O. Rabin, Decidability and undecidability of second (first) order theory of (generalized) successor, *J. Symbolic Logic* **31** (1966) 169–184.

[28] H.E. Enderton, *An introduction to mathematical logic*, Academic Press (1972).

[29] S. Fabre, Substitution et indépendance des systèmes de numération, *Thesis*, Université Aix-Marseille II (1992).

[30] A.S. Fraenkel, Systems of numeration, *Amer. Math. Monthly* **92** (1985) 105–114.

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

[32] C. Frougny, J. Sakarovitch, From the Fibonacci numeration system to the golden mean base and some generalizations, *Proc. FPSAC'93*, A. Barlotti, M. Delest and R. Pinzani, Eds. (1993) 231–244.

[33] H. Furstenberg, Algebraic functions over finite fields, J. Algebra **7** (1967) 271–277.

[34] R. Giancarlo, F. Mignosi, Generalizations of the periodicity theorem of Fine and Wilf, to appear in Proc. CAAP'94, *Lecture Notes in Comput. Sci.* (1994).

[35] S. Ginsburg, E.H. Spanier, Semigroups, Presburger formulas and languages, *Pacific J. Math.* **16** (1966) 285–296.

[36] G. Hansel, A propos d'un théorème de Cobham, in : *Actes de la fête des mots*, D. Perrin, Ed., Greco de Programmation, CNRS, Rouen (1982).

[37] G. Hansel, Personal communication to V. Bruyère (1993).

[38] G.H. Hardy, E.M. Wright, *An introduction to the theory of numbers*, Oxford University Press, Oxford, 5th ed. (1979).

[39] T. Harju, M. Linna, On the periodicity of morphisms on free monoids, *RAIRO Inform. Théor. Appl.* **20** (1986) 47–54.

[40] B.R. Hodgson, Décidabilité par automate fini, *Ann. Sci. Math. Québec* **7** (1983) 39–57.

[41] J. Honkala, Bases and ambiguity of number systems, *Theoret. Comput. Sci.* **31** (1984) 61–71.

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

[43] J.E. Hopcroft, J.D. Ullman, *Introduction to automata theory, languages and computation*, Addison-Wesley (1979).

[44] R. Ito, Every semilinear set is a finite union of disjoint linear sets, *J. Comput. System Sci.* **3** (1969) 221–231.

[45] J.H. Loxton, A.J. van der Poorten, Arithmetic properties of the solutions of a class of functional equations, *J. Reine Angew. Math.* **330** (1982) 159–195.

[46] R. MacNaughton, review of [9], *J. Symbolic Logic* **28** (1963) 100–102.

[47] H. Maurer, A. Salomaa, D. Wood, L codes and number systems, *Theoret. Comput. Sci.* **22** (1983) 331–346.

[48] C. Michaux, F. Point, Les ensembles k-reconnaissables sont définissables dans $\langle \mathbb{N}, +, V_k \rangle$, *C. R. Acad. Sci. Paris* **303** (1986) 939–942.

[49] C. Michaux, R. Villemaire, Cobham's theorem seen through Büchi theorem, Proc. Icalp'93, *Lecture Notes in Comput. Sci.* **700** (1993) 325–334.

[50] C. Michaux, R. Villemaire, Presburger arithmetic and recognizability of natural numbers by automata : new proofs of Cobham's and Semenov's theorems, in preparation (1993).

[51] M. Minsky, S. Papert, Unrecognizable sets of numbers, *J. Assoc. Comput. Mach.* **13** (1966) 281–286.

[52] M. Morse, G.A. Hedlund, Symbolic dynamics, *Amer. J. Math.* **60** (1938) 815–866.

[53] M. Morse, G.A. Hedlund, Symbolic dynamics II. Sturmian trajectories *Amer. J. Math.* **62** (1940) 1–42.

[54] A. Muchnik, Definable criterion for definability in Presburger Arithmetic and its application, *preprint in Russian*, Institute of New Technologies (1991).

[55] J.-J. Pansiot, Decidability of periodicity for infinite words, *RAIRO Inform. Théor. Appl.* **20** (1986) 43–46.

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

[57] P. Péladeau, Logically defined subsets of $\mathbb{N}^k$, *Theoret. Comput. Sci.* **93** (1992) 169–183.

[58] D. Perrin, Finite automata, in : *Handbook of Theoretical Computer Science*, vol. B, J. Van Leeuwen, Ed., Elsevier (1990) 2–57.

[59] M. Presburger, Über die Volständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt, *C. R. 1er congrès des Mathématiciens des pays slaves*, Varsovie (1929) 92–101.

[60] M.O. Rabin, Decidable theories, in : *Handbook of Mathematical Logic*, J. Barwise, Ed., North Holland, Amsterdam (1977) 595–629.

[61] G. Rauzy, Nombres algébriques et substitutions, *Bull. Soc. Math. France* **110** (1984) 147–178.

[62] C. Reutenauer, Démonstration du théorème de Cobham sur les ensembles de nombres reconnaissables par automate fini, d'après Hansel, in : *Séminaire d'Informatique Théorique*, Année 1983-84, Université Paris 7, Paris (1984) 217–224.

[63] R.W. Ritchie, Finite automata and the set of squares, *J. Assoc. Comput. Mach.* **10** (1963) 528–531.

[64] O. Salon, Suites automatiques à multi-indices et algébricité, *C. R. Acad. Sci. Paris* **305** (1987) 501–504.

[65] O. Salon, Suites automatiques à multi-indices, *Séminaire de Théorie des Nombres de Bordeaux*, Exposé **4** (1986-1987) (suivi d'un appendice par J. Shallit).

[66] A.L. Semenov, Presburgerness of predicates regular in two number systems (in Russian), *Sibirsk. Mat. Zh.* 18 (1977) 403–418, English translation, *Siberian Math. J.* **18** (1977) 289–299.

[67] A.L. Semenov, On certain extensions of the arithmetic of addition of natural numbers, *Izv. Akad. Nauk. SSSR ser. Mat.* **43** (1979) 1175–1195, English translation, *Math. USSR-Izv.* **15** (1980) 401–418.

[68] A.L. Semenov, Logical theories of one-place functions on the set of natural numbers (in Russian), *Izv. Akad. Nauk. SSSR ser. Mat.* **47** (1983) 623–658, English translation, *Math. USSR-Izv.* **22** (1984) 587–618.

[69] A.L. Semenov, Decidability of monadic theories, in : Proc. MFCS'84, M.P. Chytil and V. Koubek, Eds., *Lecture Notes in Comput. Sci.* **176** (1984) 162–175.

[70] J. Shallit, A generalization of automatic sequences, *Theoret. Comput. Sci.* **61** (1988) 1–16.

[71] W. Thomas, Automata on infinite objects, in : *Handbook of Theoretical Computer Science*, vol. B, J. Van Leeuwen, Ed., Elsevier (1990) 135–191.

[72] L. van den Dries, The field of reals with a predicate for the powers of two, *Manuscripta Math.* **54** (1985) 187–195.

[73] R. Villemaire, Joining k- and l-recognizable sets of natural numbers, Proc. Stacs'92, *Lecture Notes in Comput. Sci.* **577** (1992) 83–94.

[74] R. Villemaire, The theory of $\langle \mathbb{N}, +, V_k, V_l \rangle$ is undecidable, *Theoret. Comput. Sci.* **106** (1992) 337–349.

Véronique Bruyère, Christian Michaux
Université de Mons-Hainaut
15, Avenue Maistriau
B-7000 Mons
Belgique

Georges Hansel
Université de Rouen
Laboratoire d'Informatique de Rouen
Place Blondel
F-76134 Mont Saint-Aignan
France

Roger Villemaire
Université du Québec à Montréal
Département de Mathématiques et d'Informatique
C.P. 8888, succ. centre-ville
Montréal (Québec)
Canada H3C 3P8

Index