HAL open science

On Cobham's theorem

Fabien Durand, Michel Rigo

To cite this version:

Fabien Durand, Michel Rigo. On Cobham's theorem. Automata: from Mathematics to Applications, European Math. Soc., 2011. hal-00605375

HAL Id: hal-00605375

https://hal.science/hal-00605375v1

Submitted on 1 Jul 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On Cobham's theorem

Fabien Durand[1] *and Michel Rigo*[2]

[1]LAMFA - Université de Picardie Jules Verne - CNRS UMR 6140
33 rue Saint Leu, F-80039 Amiens cedex 1, France
email: `fabien.durand@u-picardie.fr`

[2]Université de Liège, Institut de Mathématiques
12 Grande Traverse (B37), B-4000 Liège, Belgique
email: `M.Rigo@ulg.ac.be`

2010 Mathematics Subject Classification: 68Q45, 68R15, 11B85, 11A67, 11U05, 37B20

Key words: Numeration systems, recognizable sets of integers, Cobham's theorem, ultimate periodicity, substitutions, dynamical systems.

Contents

1 Introduction 2
2 Numeration basis 4
3 Automatic sequences 7
4 Multidimensional extension and first order logic 10
4.1 Subsets of $\mathbb{N}^d$ . . . 10
4.2 Logic and k-definable sets . . . 11
5 Numeration systems and substitutions 13
5.1 Substitutive sets and abstract numeration systems . . . 13
5.2 Cobham's theorem for substitutive sets . . . 15
5.3 Density, syndeticity and bounded gaps . . . 18
5.3.1 The length of the iterates . . . 19
5.3.2 Letters and words appear with bounded gaps . . . 20
5.4 Ultimate periodicity . . . 21
5.5 The case of fixed points . . . 23
5.6 Back to numeration systems . . . 23
5.6.1 Polynomially growing abstract numeration systems . . . 24
5.6.2 Bertrand basis and ω_α-substitutive words . . . 24
6 Cobham's theorem in various contexts 24
6.1 Regular sequences . . . 24
6.2 Algebraic setting and quasi-automatic functions . . . 25

6.3 Real numbers and verification of infinite-state systems 25
6.4 Dynamical systems and subshifts 26
6.5 Tilings 28
6.5.1 From definable sets 28
6.5.2 Self-similar tilings 30
6.6 Toward Cobham's theorem for the Gaussian integers 30
6.7 Recognizability over $\mathbb{F}_q[X]$ 31
7 Decidability issues 31
8 Acknowledgments 32
References 32

1 Introduction

In this chapter we essentially focus on the representation of non-negative integers in a given numeration system. The main role of such a system — like the usual integer base k numeration system — is to replace numbers or more generally sets of numbers by their corresponding representations, *i.e.*, by words or by languages. First we consider integer base numeration systems to present the main concepts but rapidly we will introduce non-standard systems and their relationships with substitutions.

Let $k \in \mathbb{N}_{\geqslant 2}$ be an integer where $\mathbb{N}_{\geqslant 2}$ denotes the set of non-negative integers larger or equal to 2. The set $\{0, \ldots, k\}$ is denoted by $[\![0, k]\!]$. If we do not allow leading zeroes when representing numbers, the function mapping a non-negative integer n onto its k-ary representation $\mathrm{rep}_k(n) \in [\![0, k-1]\!]^*$ is a one-to-one correspondence. In the literature, one also finds notation like $\langle n \rangle_k$, $(n)_k$ or $\rho_k(n)$ instead of $\mathrm{rep}_k(n)$. In particular, 0 is assumed to be represented by the empty word ε. Hence any set $X \subseteq \mathbb{N}$ is associated with the language $\mathrm{rep}_k(X)$ consisting of the k-ary representations of the elements of X.

It is natural to study the relation existing between the arithmetic or number-theoretic properties of integers and the syntactical properties of the corresponding representations in a given numeration system. We focus on those sets $X \subseteq \mathbb{N}$ for which a finite automaton can be used to decide for any given word w over $[\![0, k-1]\!]$ whether or not w belongs to $\mathrm{rep}_k(X)$. Sets having the property that $\mathrm{rep}_k(X)$ is regular[1] are called *k-recognizable* sets. Such a set can be considered as a particularly simple set because using the k-ary numeration system it has a somehow elementary algorithmic description. In the framework of infinite-state systems verification, one also finds the terminology of *Number Decision Diagram* or *NDD* [122].

The essence of Cobham's theorem is to express that the property for a set to be recognizable by a finite automaton *strongly depends* on the choice of the base and more generally on the considered numeration system. Naturally this fact leads to and motivates the introduction and the study of recognizable sets in non-standard numeration systems. Considering alternative numeration systems may provide new recognizable sets and these non-standard systems also have applications in computer arithmetic [59]. Last but not least, the proof of Cobham's theorem is non-trivial and relies on quite elaborate argu-

[1] We use the terminology of regular language, instead of rational language.

ments.

Now let us state this celebrated result from 1969 and give all the needed details and definitions. Several surveys have been written on the same subject, see [25, 26, 28, 98].

Theorem 1.1 (Cobham's theorem [35]). *Let $k, \ell \geqslant 2$ be two multiplicatively independent integers. A set $X \subseteq \mathbb{N}$ is both k-recognizable and ℓ-recognizable if and only if it is ultimately periodic.*

In the various contexts that we will describe, showing that an ultimately periodic set is recognizable is always the easy direction to prove. See Remark 1.3. So we focus on the other direction.

Definition 1.1. A subset of $\mathbb{N}$ is *ultimately periodic* if it is the union of a finite set and a finite number of infinite arithmetic progressions. In particular, X is ultimately periodic if and only if there exist $N \geqslant 0$ and $p \geqslant 1$ such that for all $n \geqslant N$, $n \in X \Leftrightarrow n + p \in X$. Recall that an *arithmetic progression* is a set of the kind $a\mathbb{N} + b := \{an + b \mid n \geqslant 0\}$.

Definition 1.2. Let $\alpha, \beta > 1$ be two real numbers. If the equation $\alpha^m = \beta^n$ with $m, n \in \mathbb{N}$ has only the trivial integer solution $m = n = 0$, then α and β are said to be *multiplicatively independent*. Otherwise, α and β are said to be *multiplicatively dependent*.

Let $k, \ell \geqslant 2$ be two integers. Notice that k and ℓ are multiplicatively independent if and only if $\log k / \log \ell$ is irrational. Note that for k and ℓ to be multiplicatively dependent it is not enough that k and ℓ share exactly the same prime factors occurring in their decomposition. For instance, 6 and 18 are multiplicatively independent. But coprime integers are multiplicatively independent.

The irrationality of $\log k / \log \ell$ is a crucial point in the proof of Cobham's theorem (see Subsection 5.3). Recall that if $\theta > 0$ is irrational, then the set $\{\{n\theta\} \mid n > 0\}$ of fractional parts of the multiples of θ is dense in $[0, 1]$. For a proof of the so-called *Kronecker's theorem*, see [66].

Remark 1.2. The fact for two integers to be multiplicatively dependent is an equivalence relation $\mathfrak{M}$ over $\mathbb{N}_{\geqslant 2}$. If k and ℓ are multiplicatively dependent, then there exist a minimal $q \geqslant 2$ and two positive integers m, n such that $k = q^m$ and $\ell = q^n$. Let us give the first (with respect to their minimal element) few equivalence classes for $\mathfrak{M}$ partitioning $\mathbb{N}_{\geqslant 2}$: $[2]_{\mathfrak{M}}$, $[3]_{\mathfrak{M}}$, $[5]_{\mathfrak{M}}$, $[6]_{\mathfrak{M}}$, $[7]_{\mathfrak{M}}$, $[10]_{\mathfrak{M}}$, $[11]_{\mathfrak{M}}$, $[12]_{\mathfrak{M}}$, ...

Remark 1.3. We show that if a set $X \subseteq \mathbb{N}$ is ultimately periodic then, for all $k \geqslant 2$, X is k-recognizable. In the literature, one also finds the terminology of a *recognizable set* X (without any mention to a base), meaning that X is k-recognizable for all $k \geqslant 2$. Note that a finite union of regular languages is again a regular language. Hence it is enough to check that $\mathrm{rep}_k(a\mathbb{N} + b)$ is regular with $0 \leqslant b < a$. We can indeed assume that $b < a$ because if we add or remove a finite number of words to a regular language, we still have a regular language. Consider a DFA having $Q = [\![0, a-1]\!]$ as its set of states. For all $i \in Q$, $d \in [\![0, k-1]\!]$, the transitions are given by

$$i \xrightarrow{d} ki + d \mod a.$$

The initial state is 0 and the unique final state is b. As an example, a DFA accepting exactly binary representations of the integers congruent to 3 mod 4 is given in Figure 1. A study of the minimal automaton recognizing such divisibility criteria expressed in an

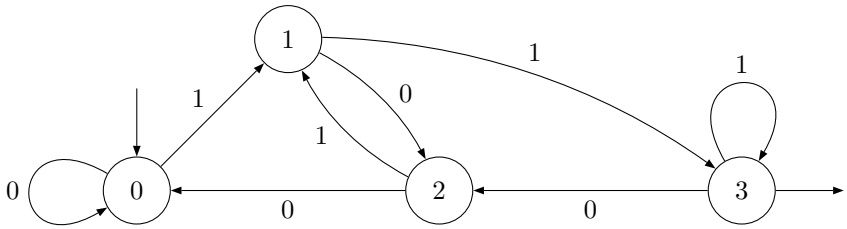

Figure 1. A finite automaton accepting $\mathrm{rep}_2(4\mathbb{N}+3)$.

integer base is given in [3]. See also the discussion in [109, Prologue]. The fact that a divisibility criterion exists in every base for any fixed divisor was already observed by Pascal in [97, pp. 84–89].

2 Numeration basis

It is remarkable that the recognizability of ultimately periodic sets extends to wider contexts (see Proposition 2.6 and Theorem 5.1). Let us introduce a first generalization of the integer base numeration system.

Definition 2.1. A *numeration basis* is a sequence $U = (U_n)_{n\geqslant 0}$ of integers such that U is increasing, $U_0 = 1$ and that the set $\{U_{i+1}/U_i \mid i \geqslant 0\}$ is bounded. This latter condition ensures the finiteness of the alphabet of digits used to represent integers. If $w = w_\ell \cdots w_0$ is a word over a finite alphabet $A \subset \mathbb{Z}$ then the numerical value of w is

$$\pi_{A,U}(w) = \sum_{i=0}^{\ell} w_i\, U_i.$$

Using the greedy algorithm [57], any integer n has a unique *(normal) U-representation* $\mathrm{rep}_U(n) = w_\ell \cdots w_0$ which is a finite word over a minimal finite alphabet called the *canonical alphabet* of U and denoted by A_U. The normal U-representation satisfies

$$\pi_{A_U,U}(\mathrm{rep}_U(n)) = n \text{ and for all } i \in [\![0, \ell-1]\!],\ \pi_{A_U,U}(w_i \cdots w_0) < U_{i+1}.$$

Again, $\mathrm{rep}_U(0) = \varepsilon$. See [85, Chapter 7] or Ch. Frougny and J. Sakarovitch's chapter in [12, Chapter 2]. A subset $X \subseteq \mathbb{N}$ is *U-recognizable* if $\mathrm{rep}_U(X)$ is accepted by a finite automaton. Let $B \subset \mathbb{Z}$ be a finite alphabet. If $w \in B^*$ is such that $\pi_{B,U}(w) \geqslant 0$, then the function mapping w onto $\mathrm{rep}_U(\pi_{B,U}(w))$ is called *normalization*.

Definition 2.2. A numeration basis U is said to be *linear* if there exist $k \in \mathbb{N} \setminus \{0\}$, $d_1, \ldots, d_k \in \mathbb{Z}$, $d_k \neq 0$, such that, for all $n \geqslant k$, $U_n = d_1 U_{n-1} + \cdots + d_k U_{n-k}$. The

polynomial $P_U(X) = X^k - d_1X^{k-1} - \cdots - d_{k-1}X - d_k$ is called the *characteristic polynomial* of U.

Definition 2.3. Recall that a *Pisot-Vijayaraghavan number* is an algebraic integer $\beta > 1$ whose Galois conjugates have modulus strictly less than one. We say that $U = (U_n)_{n\geqslant 0}$ is a *Pisot numeration system* if the numeration basis U is linear and $P_U(X)$ is the minimal polynomial of a Pisot number β. Integer base numeration systems are particular cases of Pisot systems. For instance, see [27] where it is shown that most properties related to k-recognizable sets, $k \in \mathbb{N}_{\geqslant 2}$, can be extended to Pisot systems. In such a case, there exists some $c > 0$ such that $|U_n - c\,\beta^n| \to 0$, as n tends to infinity.

Example 2.1. Consider the Fibonacci sequence defined by $U_0 = 1$, $U_1 = 2$ and $U_{n+2} = U_{n+1}+U_n$ for all $n \geqslant 0$. A word over $\{0, 1\}$ is a U-representation if and only if it belongs to the language $L = 1\{0, 01\}^* \cup \{\varepsilon\}$. For instance 10110 is not a U-representation. Since $\pi_{A_U,U}(10110) = 13$, the normalization maps 10110 to $\mathrm{rep}_U(13) = 100000$. The characteristic polynomial of this linear numeration basis is the minimal polynomial of the Pisot number $(1 + \sqrt{5})/2$. This Pisot numeration system is presented in [123].

The following result is an easy exercise but also can be carried on in a wider context.

Theorem 2.1. [115] *Let U be a numeration basis. If $\mathbb{N}$ is U-recognizable, then U is linear.*

Definition 2.4. [13] A *Bertrand numeration basis* U is a numeration basis satisfying the following property: $w \in \mathrm{rep}_U(\mathbb{N})$ if and only if, for all $n \in \mathbb{N}$, $w0^n \in \mathrm{rep}_U(\mathbb{N})$. It is a natural condition satisfied by all integer base $k \geqslant 2$ systems. For instance, the sequence defined by $U_0 = 1$, $U_1 = 3$ and, for all $n \geqslant 0$, $U_{n+2} = U_{n+1} + U_n$ is not a Bertrand numeration basis because $\mathrm{rep}_U(2) = 2$, but $\pi_{A_U,U}(20) = 6$ and $\mathrm{rep}_U(6) = 102$.

Let $\alpha > 1$ be a real number. The notion of α-expansion was introduced by Parry in [96], (also see Rényi's paper [104]). See again [85, Chapter 7]. All $x \in [0, 1]$ can be uniquely written in the following way:

$$x = \sum_{n\geqslant 1} a_n\alpha^{-n}, \tag{2.1}$$

with $x_1 = x$ and for all $n \geqslant 1$, $a_n = \lfloor \alpha\, x_n \rfloor$ and $x_{n+1} = \{\alpha x_n\}$, where $\lfloor\cdot\rfloor$ stands for the integer part. The sequence $d_\alpha(x) = (a_n)_{n\geqslant 1}$ is the *α-expansion* of x and $L(\alpha)$ denotes the set of finite words having an occurrence in some sequences $d_\alpha(x)$, $x \in [0, 1]$. Let $d_\alpha(1) = (t_n)_{n\geqslant 1}$. If there exist $N \geqslant 0$, $p > 0$ such that, for all $n \geqslant N$, $t_{n+p} = t_n$ then α is said to be a *Parry number*, sometimes called a *β-number* (for more details or information about these numbers, see [96] or [58]). Observe that integers greater or equal to 2 are Parry numbers.

The following result relates Bertrand numeration systems to languages defined by some real number.

Theorem 2.2 (A. Bertrand-Mathis [14]). *Let U be a numeration basis. It is a Bertrand numeration basis if and only if there exists a real number $\alpha > 1$ such that $\mathrm{rep}_U(\mathbb{N}) = L(\alpha)$. In this case, if U is linear then α is a root of the characteristic polynomial of U.*

Theorem 2.3 (A. Bertrand-Mathis [13]). *Let $\alpha > 1$ be a real number. The language $L(\alpha)$ is regular if and only if α is a Parry number.*

Associated with a Parry number β, one can define the notion of beta-polynomial. For details see [68] or [12, Chapter 2]. First we define the *canonical beta-polynomial*. If $d_\beta(1)$ is eventually constant and equal to 0: $d_\beta(1) = t_1 \cdots t_m 0^\omega$, with $t_m \neq 0$, then we set $G_\beta(X) = X^m - \sum_{i=1}^m t_i X^{m-i}$ and $r = m$. Otherwise, $d_\beta(1)$ is eventually periodic: $d_\beta(1) = t_1 \cdots t_m (t_{m+1} \cdots t_{m+p})^\omega$, with m and p being minimal. Then we set $G_\beta(X) = X^{m+p} - \sum_{i=1}^{m+p} t_i X^{m+p-i} - X^m + \sum_{i=1}^m t_i X^{m-i}$ and $r = p$. Let β be a Parry number. An *extended beta-polynomial* is a polynomial of the form $H_\beta(X) = G_\beta(X)(1 + X^r + \cdots + X^{rk})X^n$ for $k, n \in \mathbb{N}$.

Proposition 2.4. [68] *Let U be a linear numeration basis with dominant root β, i.e., $\lim_{n\to\infty} U_{n+1}/U_n = \beta$ for some $\beta > 1$. If $\mathrm{rep}_U(\mathbb{N})$ is regular, then β is a Parry number.*

Theorem 2.5 (M. Hollander [68]). *Let U be a linear numeration basis whose dominant root β is a Parry number.*

- *If $d_\beta(1)$ is infinite and eventually periodic, then $\mathrm{rep}_U(\mathbb{N})$ is regular if and only if U satisfies an extended beta-polynomial for β.*
- *If $d_\beta(1)$ is finite of length m, then: if U satisfies an extended beta-polynomial for β then $\mathrm{rep}_U(\mathbb{N})$ is regular; and conversely if $\mathrm{rep}_U(\mathbb{N})$ is regular, then U satisfies either an extended beta-polynomial for β, $H_\beta(X)$, or a polynomial of the form $(X^m - 1)H_\beta(X)$.*

Ultimately periodic sets are recognizable for any linear numeration basis.

Proposition 2.6 (Folklore [12, 85]). *Let $a, b \geqslant 0$. If $U = (U_n)_{n\geqslant 0}$ is a linear numeration basis, then*

$$\pi^{-1}_{A_U,U}(a\,\mathbb{N} + b) = \left\{ c_\ell \cdots c_0 \in A_U^* \mid \sum_{k=0}^{\ell} c_k\, U_k \in a\,\mathbb{N} + b \right\}$$

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

To conclude this section, consider again the integer base numeration systems.

Example 2.2. The set $P_2 = \{2^n \mid n \geqslant 0\}$ of powers of two is trivially 2-recognizable because $\mathrm{rep}_2(P_2) = 10^*$. Since the difference between any two consecutive elements in P_2 is of the kind $2^{n+1} - 2^n = 2^n$, P_2 is not ultimately periodic. As a consequence of Cobham's theorem, P_2 is for instance neither 3-recognizable nor 5-recognizable.

One could also consider the case when the two bases k and ℓ are multiplicatively dependent. This case is much easier and can be considered as an exercise.

Proposition 2.7. *Let $k, \ell \geqslant 2$ be two multiplicatively dependent integers. A set $X \subseteq \mathbb{N}$ is k-recognizable if and only if it is ℓ-recognizable.*

The theorem of Cobham implies that ultimately periodic sets are the only infinite sets that are k-recognizable for every $k \geqslant 2$. We have seen so far that there exist sets (like the set P_2 of powers of two) that are only recognizable for some specific bases: exactly all bases belonging to a unique equivalence class for the equivalence relation $\mathfrak{M}$ over $\mathbb{N}_{\geqslant 2}$. To see that a given infinite ordered set $X = \{x_0 < x_1 < x_2 < \cdots\}$ is k-recognizable for *no* base $k \geqslant 2$ at all, we can use results like the following one, where the behavior of the ratio (*resp.* difference) of any two consecutive elements in X is studied through the quantities

$$\mathbf{R}_X = \limsup_{i\to\infty} \frac{x_{i+1}}{x_i} \text{ and } \mathbf{D}_X = \limsup_{i\to\infty} (x_{i+1} - x_i) .$$

Theorem 2.8 (Gap theorem [36]). *Let $k \geqslant 2$. If $X \subseteq \mathbb{N}$ is a k-recognizable infinite subset of $\mathbb{N}$, then either $\mathbf{R}_X > 1$ or $\mathbf{D}_X < +\infty$.*

Corollary 2.9. *Let $a \in \mathbb{N}_{\geqslant 2}$. The set of primes and the set $\{n^a \mid n \geqslant 0\}$ are never k-recognizable for any integer base $k \geqslant 2$.*

Proofs of the Gap theorem and its corollary can also be found in [51]. For more results on primes, see also the chapter "Automata in number theory" of this handbook.

Definition 2.5. An infinite ordered set $X = \{x_0 < x_1 < x_2 < \cdots\}$ such that $\mathbf{D}_X < +\infty$ is said to be *syndetic* or with *bounded gaps*: there exists $C > 0$ such that for all $n \geqslant 0$, $x_{n+1} - x_n < C$. In particular, any ultimately periodic set is syndetic. The converse does not hold, see for instance Example 3.1.

Remark 2.10. Note that syndeticity occurs in various contexts like in ergodic theory. As an example, a subset of an Abelian group G is said to be syndetic if finitely many translates of it cover G. The term "syndetic" was first quoted in [62]. Note that in [64] the following result is proved. Let $\alpha, \beta > 1$ be multiplicatively independent real numbers. If a set $X \subseteq \mathbb{N}$ is α-recognizable and β-recognizable, for the Bertrand numeration systems based respectively on the real numbers α and β in the sense of [14] and Theorem 2.2, then X is syndetic.

Cobham's original proof of Theorem 1.1 appeared in [35] and we quote [51] "*The proof is correct, long and hard. It is a challenge to find a more reasonable proof of this fine theorem*". Then G. Hansel proposed a simpler presentation in [63], also one can see [98] or the dedicated chapter in [9] for an expository presentation. Prior to these last two references, one should read [108]. Usually a first step to prove Cobham's theorem is to show the syndeticity of the considered set. See Section 5.3.

3 Automatic sequences

As explained in Corollary 3.3 presented in this section, the formalism of k-recognizable sets is equivalent to the one of k-automatic sequences[2]. Let us recall briefly what they are.

[2] We indifferently use the terms sequence and infinite word.

An infinite word $x = (x_n)_{n \geqslant 0} \in B^{\mathbb{N}}$ over an alphabet B is said to be k-*automatic* if there exists a DFAO (deterministic finite automaton with output) over the alphabet $[\![0, k-1]\!]$, $(Q, [\![0, k-1]\!], \cdot, q_0, B, \tau)$ such that, for all $n \geqslant 0$,

$$x_n = \tau(q_0 \cdot \mathrm{rep}_k(n)) \ .$$

The transition function is $\cdot : Q \times [\![0, k-1]\!] \to Q$ and can easily be extended to $Q \times [\![0, k-1]\!]^*$ by $q \cdot \varepsilon = q$ and $q \cdot wa = (q \cdot w) \cdot a$. The output function is $\tau : Q \to B$. Roughly speaking, the nth term of the sequence is obtained by feeding a DFAO with the k-ary representation of n. For a complete and comprehensive exposition on k-automatic sequences and their applications see the book [9]. We equally use the terms of sequences or (right-) infinite words. For more information about combinatorics on words, see [84, 85] or also J. Cassaigne and F. Nicolas' chapter in [12, Chapter 4].

Definition 3.1. Let $\sigma : A^* \to A^*$ be a morphism, *i.e.*, $\sigma(uv) = \sigma(u)\sigma(v)$ for all $u, v \in A^*$. Naturally such a map can be defined on A^ω. A finite or infinite word x such that $\sigma(x) = x$ is said to be a *fixed point* of σ. A morphism $\sigma : A^* \to A^*$ is completely determined by the images of the letters in A. In particular, if there exists $k \geqslant 0$ such that for all $a \in A$, $|\sigma(a)| = k$, then σ is said to be of k-*uniform* or simply *uniform*. A 1-uniform morphism is called a *coding*. If there exist a letter $a \in A$ and a word $u \in A^+$ such that $\sigma(a) = au$ and moreover, if $\lim_{n \to +\infty} |\sigma^n(a)| = +\infty$, then σ is said to be *prolongable* on a or to be a *substitution*. Let $\sigma : A^* \to A^*$ be a morphism prolongable on a. We have

$$\sigma(a) = a\,u,\ \sigma^2(a) = a\,u\,\sigma(u),\ \sigma^3(a) = a\,u\,\sigma(u)\,\sigma^2(u), \ldots \ .$$

Since for all $n \in \mathbb{N}$, $\sigma^n(a)$ is a prefix of $\sigma^{n+1}(a)$ and because $|\sigma^n(a)|$ tends to infinity when $n \to +\infty$, the sequence $(\sigma^n(a))_{n \geqslant 0}$ converges (for the usual product topology on words, see for instance (6.2)) to an infinite word denoted by $\sigma^\infty(a)$ and given by

$$\sigma^\infty(a) := \lim_{n \to +\infty} \sigma^n(a) = a\,u\,\sigma(u)\,\sigma^2(u)\,\sigma^3(u) \cdots \ .$$

This infinite word is a fixed point of σ. An infinite word obtained in this way by iterating a prolongable morphism is said to be *purely substitutive* (or *pure morphic*). If $\sigma : A^* \to B^*$ is a non-erasing morphism, it can be extended to a map from $A^{\mathbb{N}}$ to $B^{\mathbb{N}}$ as follows. If $x = x_0 x_1 \cdots$ is an infinite word over A, then the sequence of words $(\sigma(x_0 \cdots x_{n-1}))_{n \geqslant 0}$ is easily seen to be convergent towards an infinite word over B. Its limit is denoted by $\sigma(x) = \sigma(x_0)\sigma(x_1)\sigma(x_2) \cdots$. If $x \in A^{\mathbb{N}}$ is purely substitutive and if $\tau : A \to B$ is a coding, then the word $y = \tau(x)$ is said to be *substitutive*.

Another result due to A. Cobham is the following one, see [36]. The idea is to canonically associated with any k-uniform morphism a DFA over $[\![0, k-1]\!]$.

Theorem 3.1. *Let $k \geqslant 2$. A sequence $x = (x_n)_{n \geqslant 0} \in B^{\mathbb{N}}$ is k-automatic if and only if there exists a k-uniform morphism $\sigma : A^* \to A^*$ prolongable on a letter $a \in A$ and a coding $\tau : A \to B$ such that $x = \tau(\sigma^\infty(a))$.*

Theorem 3.2 (Eilenberg [51]). *A sequence $x = (x_n)_{n \geqslant 0}$ is k-automatic if and only if its* k-kernel $N_k(x) = \{(x_{k^e n + d})_{n \geqslant 0} \mid e \geqslant 0,\ 0 \leqslant d < k^e\}$ *is finite.*

Definition 3.2. The *characteristic sequence* $\mathbb{1}_X \in \{0,1\}^{\mathbb{N}}$ of a set $X \subseteq \mathbb{N}$ is defined by $\mathbb{1}_X(n) = 1$ if and only if $n \in X$.

An infinite word $x \in A^\omega$ is *ultimately periodic* if there exist two finite words $u \in A^*$ and $v \in A^+$ such that $x = uv^\omega$. If $u = \varepsilon$, x is *periodic*. Obviously, a set $X \subseteq \mathbb{N}$ is ultimately periodic if and only if $\mathbb{1}_X$ is an ultimately periodic word over $\{0,1\}$. In that case, there exist two finite words $u \in \{0,1\}^*$ and $v \in \{0,1\}^+$ such that $\mathbb{1}_X = uv^\omega$. In particular, $|v|$ is a period of X. If u and v are chosen of minimal length, then $|u|$ (resp. $|v|$) is said to be the *preperiod* or *index* of X (resp. the *period* of X). If $u = \varepsilon$, X is *(purely) periodic*. Periodic sets are in particular ultimately periodic.

Corollary 3.3. *Let $k \geqslant 2$. If $x = (x_n)_{n\geqslant 0} \in B^{\mathbb{N}}$ is a k-automatic sequence then, for all $b \in B$, the set $\{n \geqslant 0 \mid x_n = b\}$ is k-recognizable. Conversely, if a set $X \subseteq \mathbb{N}$ is k-recognizable, then its characteristic sequence is k-automatic.*

Theorem 3.4 (Cobham's theorem, version 2). *Let $k, \ell \geqslant 2$ be two multiplicatively independent integers. An infinite word $x = (x_n)_{n\geqslant 0} \in B^{\mathbb{N}}$ is both k-automatic and ℓ-automatic if and only if it is ultimately periodic.*

Remark 3.5. Using the framework of k-automatic sequences instead of the formalism of k-recognizable sets turns out to be useful. For instance, consider the *complexity function* of an infinite word x which maps $n \in \mathbb{N}$ onto the number $p_x(n)$ of distinct factors of length n occurring in x. Morse–Hedlund's theorem states that x is ultimately periodic if and only if p_x is bounded by some constant. This result appeared first in [90]. Proofs can be found in classical textbooks like [9, 84].

It is also well known that for a k-automatic sequence x, $p_x \in \mathcal{O}(n)$, again see the seminal paper [36]. This latter result can be used to show that particular sets are not k-recognizable for any $k \geqslant 2$: for instance, those sets whose characteristic sequence $\mathbb{1}_X$ has a complexity function such that $\lim_{n\to+\infty} p_{\mathbb{1}_X}(n)/n = +\infty$. For the behavior of p_x in the substitutive case, see the survey [4] or [12, Chapter 4].

Example 3.1. Iterating the morphism $\sigma : 0 \mapsto 01, 1 \mapsto 10$, we get the *Thue–Morse word* $(t_n)_{n\geqslant 0} = \sigma^\infty(0) = 0110100110010110100101100110\cdots$. For an account on this celebrated word, see [8] and [56, Chapter 2]. It is a 2-automatic word, the nth letter in the word is 0 if and only if $\mathrm{rep}_2(n)$ contains an even number of 1's. This word is generated by the DFAO represented in Figure 2. In particular, the set

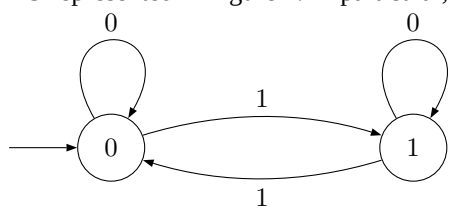

Figure 2. A DFAO generating the Thue–Morse word.

$$X_2 = \left\{ n \in \mathbb{N} \mid \mathrm{rep}_2(n) = c_t \cdots c_0 \text{ and } \sum_{i=0}^{t} c_i \equiv 0 \pmod 2 \right\}$$

is 2-recognizable. The Thue–Morse word is not ultimately periodic (see for instance [23] or [39] where the complexity function of this word is studied carefully) and therefore X_2 is k-recognizable only for those k of the form 2^m, $m \in \mathbb{N}_{\geqslant 1}$. Nevertheless, one can notice that X_2 is syndetic.

4 Multidimensional extension and first order logic

4.1 Subsets of $\mathbb{N}^d$

To extend the concept of k-recognizability to subsets of $\mathbb{N}^d$, $d \geqslant 2$, it is natural to consider d-tuples of k-ary representations. To get d words of the same length that have to be read simultaneously by an automaton, the shortest ones are padded with leading zeroes. We extend the definition of rep_k to a map of domain $\mathbb{N}^d$ as follows. If $n_1, \ldots, n_d$ are non-negative integers, we consider the word

$$\mathrm{rep}_k(n_1, \ldots, n_d) := \begin{pmatrix} 0^{m-|\mathrm{rep}_k(n_1)|}\,\mathrm{rep}_k(n_1) \\ \vdots \\ 0^{m-|\mathrm{rep}_k(n_d)|}\,\mathrm{rep}_k(n_d) \end{pmatrix} \in \left([\![0, k-1]\!]^d\right)^*$$

where $m = \max\{|\mathrm{rep}_k(n_1)|, \ldots, |\mathrm{rep}_k(n_d)|\}$. A subset X of $\mathbb{N}^d$ is *k-recognizable* if the corresponding language $\mathrm{rep}_k(X)$ is accepted by a finite automaton over the alphabet $[\![0, k-1]\!]^d$ which is the Cartesian product of d copies of $[\![0, k-1]\!]$. This automaton is reading d digits at a time (one for each component): this is why we need d words of the same length.

Example 4.1. Consider the automaton depicted in Figure 3 (the sink is not represented). It accepts $(\varepsilon, \varepsilon)$ and all pairs of words of the kind $(u0, 0u)$ where $u \in 1\{0,1\}^*$. This means that the set $\{(2n, n) \mid n \geqslant 0\}$ is 2-recognizable.

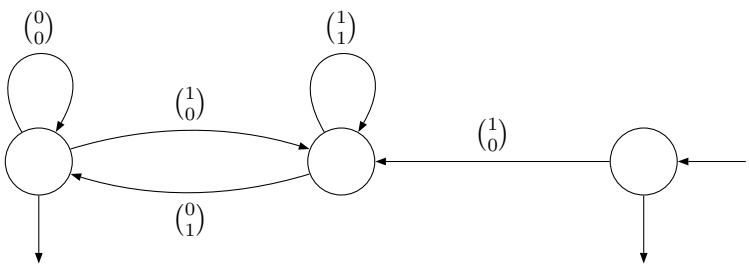

Figure 3. A DFA recognizing $\{(2n, n) \mid n \geqslant 0\}$.

Note that the notion of k-automatic sequence and Theorem 3.1 have been extended accordingly in [111, 112] where the images by a morphism of letters are d-dimensional cubes of size k.

Extending the concept of ultimately periodic sets to subsets of $\mathbb{N}^d$, with $d \geqslant 2$, is at first glance not so easy. We use bold face letters to represent elements in $\mathbb{N}^d$. For

instance, one could take the following definition of a (purely) periodic subset $X \subseteq \mathbb{N}^d$. There exists a non-zero element $\mathbf{p} \in \mathbb{N}^d$ such that $\mathbf{x} \in X$ if and only if $\mathbf{x} + \mathbf{p} \in X$. As we will see (Remark 4.2, Proposition 6.9 and Theorem 6.11), it turns out that this definition does not fit to the extension of Cobham's theorem in d dimensions. Therefore we will consider sets definable in $\langle \mathbb{N}, + \rangle$. Let us mention *Nivat's conjecture* connecting such a notion of periodicity in higher dimensions with the notion of block complexity as introduced in Remark 3.5: let $X \subset \mathbb{Z}^2$, if there exist positive integers n_1, n_2 such that $p_X(n_1, n_2) \leqslant n_1 n_2$ then X is periodic, where $p_X(n_1, n_2)$ counts the number of distinct blocks of size $n_1 \times n_2$ occurring in X. See [92] and in particular [102] for details and pointers to the existing bibliography.

4.2 Logic and k-definable sets

The formalism of first order logic is probably the best suited to present a natural extension (in the sense of Cobham's theorem) of the definition of ultimately periodic sets in d dimensions. See [100, 101] or the survey [16]. In the *Presburger arithmetic* $\langle \mathbb{N}, + \rangle$, the variables range over $\mathbb{N}$ and we have at our disposal the connectors $\wedge, \vee, \neg, \rightarrow, \leftrightarrow$, the equality symbol $=$ and the quantifiers $\forall$ and $\exists$ that can only be applied to variables. This is the reason we speak of first order logic; in second order logic, quantifiers can be applied to relations, and in monadic second order logic, only variables and unary relations, *i.e.*, sets, may be quantified. If a variable is not within the scope of any quantifier, this variable is said to be *free*. Formulas are build inductively from terms and atomic formulas. Here details have been omitted, see for instance [28, Section 3.1]. For instance, order relations $<, \leqslant, \geqslant$ and $>$ can be added to the language by noticing that $x \leqslant y$ is equivalent to

$$(\exists z)(y = x + z). \tag{4.1}$$

In the same way, constants can also be added. For instance, $x = 0$ is equivalent to $(\forall y)(x \leqslant y)$ and $x = 1$ is equivalent to $\neg(x = 0) \wedge (\forall y)(\neg(y = 0) \rightarrow (x \leqslant y))$. In general, the *successor* function $\mathbf{S}(x) = y$ of x is defined by

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

For a complete account on the interactions between first order logic and k-recognizable sets, see the excellent survey [28].

Remark 4.1. We mainly discuss the case $\langle \mathbb{N}, + \rangle$ but all developments can be made for $\langle \mathbb{Z}, +, \leqslant \rangle$. Note that if the variables belong to $\mathbb{Z}$ then it is no longer possible to define $\leqslant$ as in (4.1). So this order relation has to be added to the structure. The constant 0 can be defined by $x + x = x$.

Let $\varphi(x_1, \ldots, x_d)$ be a formula with d free variables $x_1, \ldots, x_d$. Interpreting φ in $\langle \mathbb{N}, + \rangle$ permits one to define the set of d-tuples of non-negative integers for which the formula holds true:

$$\{(r_1, \ldots, r_d) \mid \langle \mathbb{N}, + \rangle \models \varphi[r_1, \ldots, r_d]\}.$$

We write $\langle \mathbb{N}, + \rangle \models \varphi[r_1, \ldots, r_d]$ if $\varphi(x_1, \ldots, x_d)$ is satisfied in $\langle \mathbb{N}, + \rangle$ when interpreting

x_i by r_i for all $i \in \{1, \ldots, d\}$. For the reader having no background in logic and model theory, the first chapters of [50] are worth reading.

Remark 4.2. The ultimately periodic sets of $\mathbb{N}$ are exactly the sets that are definable in the Presburger arithmetic. It is obvious that ultimately periodic sets of $\mathbb{N}$ are definable. For instance, the set of even integers can be defined by $\varphi(x) \equiv (\exists y)(x = y + y)$. Since constants can easily be defined, it is easy to write a formula for any arithmetic progression. As an example, the formula $\varphi(x) \equiv (\exists y)(x = \mathbf{S}(\mathbf{S}(y + y + y)))$ defines the progression $3\mathbb{N} + 2$. In particular, multiplication by a fixed constant is definable in $\langle \mathbb{N}, + \rangle$. Note that it is a classical result that the theory of $\langle \mathbb{N}, +, \times \rangle$ is undecidable, see for instance [15].

Adding congruences modulo any integer m permits quantifier elimination, which means that any formula expressed in the Presburger arithmetic is equivalent to a formula using only $\wedge$, $\vee$, $=$, $<$ and congruences, see [100, 101]. Presentations can also be found in [52, 80].

Theorem 4.3 (Presburger). *The structure $\langle \mathbb{N}, +, <, (\equiv_m)_{m>0} \rangle$ admits elimination of quantifiers.*

This result can be used to prove that the theory of $\langle \mathbb{N}, + \rangle$ is decidable. This can be done using the formalism of automata, see for instance [28].

Corollary 4.4. *Any formula $\varphi(x)$ in the Presburger arithmetic $\langle \mathbb{N}, + \rangle$ defines an ultimately periodic set of $\mathbb{N}$.*

Let $k \geqslant 2$. We add to the structure $\langle \mathbb{N}, + \rangle$ a function V_k defined by $V_k(0) = 1$ and for all $x > 0$, $V_k(x)$ is the greatest power of k dividing x. As an example, we have $V_2(6) = 2$, $V_2(20) = 4$ and $V_2(2^n) = 2^n$ for all $n \geqslant 0$. Again the theory of $\langle \mathbb{N}, +, V_k \rangle$ can be shown to be decidable [28]. The next result shows that, as for the k-automatic sequences, the logical framework within the richer structure $\langle \mathbb{N}, +, V_k \rangle$ gives an equivalent presentation of the k-recognizable sets in any dimension. Proofs of the next three theorems can again be found in [28] where a full account of the different approaches used to prove Theorem 4.5 is presented. For Büchi's original paper, see [29].

Theorem 4.5 (Büchi's theorem). *Let $k \geqslant 2$ and $d \geqslant 1$. A set $X \subseteq \mathbb{N}^d$ is k-recognizable if and only if it can be defined by a first order formula $\varphi(x_1, \ldots, x_d)$ of $\langle \mathbb{N}, +, V_k \rangle$.*

For instance the set P_2 introduced in Example 2.2 can be defined by the formula $\varphi(x) \equiv V_2(x) = x$. Note that Theorem 4.5 holds for Pisot numeration systems given in Definition 2.3, see [27] where the function V_k is modified accordingly. This is partially based on the fact that in a Pisot numeration system the normalization function is realized by a finite automaton, see [58], which allows one to consider addition of integers: first perform addition digit-wise without any carry, then normalize the result.

Theorem 4.6 (Cobham's theorem, version 3). *Let $k, \ell \geqslant 2$ be two multiplicatively independent integers. A set $X \subseteq \mathbb{N}$ can be defined by a first order formula in $\langle \mathbb{N}, +, V_k \rangle$ and by a first order formula in $\langle \mathbb{N}, +, V_\ell \rangle$ if and only if it can be defined by a first order formula in $\langle \mathbb{N}, + \rangle$.*

This theorem still holds in higher dimensions and is called the Cobham–Semenov theorem. In this respect, the notion of subset of $\mathbb{N}^d$ definable in the Presburger arithmetic $\langle \mathbb{N}, + \rangle$ is the right extension of periodicity in a multidimensional setting. For Semenov's original paper, see [113].

Theorem 4.7 (Cobham–Semenov theorem). *Let $k, \ell \geqslant 2$ be two multiplicatively independent integers. A set $X \subseteq \mathbb{N}^d$ can be defined by a first order formula in $\langle \mathbb{N}, +, V_k \rangle$ and by a first order formula in $\langle \mathbb{N}, +, V_\ell \rangle$ if and only if it can be defined by a first order formula in $\langle \mathbb{N}, + \rangle$.*

Subsets of $\mathbb{N}^d$ defined by a first order formula in $\langle \mathbb{N}, + \rangle$ are characterized in [61]. The nice criterion of Muchnik appeared first in 1991 and is given in [91]. See Proposition 6.9 for its precise statement. Using this latter characterization, a proof of Theorem 4.7 is presented in [28]. The logical framework has given rise to several works. Let us mention chronologically [118, 119] and [88, 89]. In [89, Section 5] the authors interestingly show how to reduce Semenov's theorem to Cobham's theorem: "Nothing new in higher dimensions". Also extensions to non-standard numeration systems are considered in [99] and [15]. In this latter paper, the Cobham–Semenov theorem is proved for two Pisot numeration systems.

5 Numeration systems and substitutions

5.1 Substitutive sets and abstract numeration systems

In Sections 4.1 and 4.2, we have mainly extended the notion of recognizability to subsets of $\mathbb{N}^d$. Now we consider another extension of recognizability. In Corollary 3.3, we have seen that a k-recognizable set has a characteristic sequence generated by a uniform substitution and the application of an extra coding. It is pretty easy to define sets of integers encoded by a characteristic sequence generated by an arbitrary substitution and an extra coding, that is whose characteristic sequence is morphic. This generalization permits one to reach a larger class of infinite words, hence a larger class of sets of integers.

Example 5.1. Consider the morphism $\sigma : \{a, b, c\}^* \to \{a, b, c\}^*$ given by $\sigma(a) = abcc$, $\sigma(b) = bcc$, $\sigma(c) = c$ and the coding $\tau : a, b \mapsto 1, c \mapsto 0$. We get

$$\sigma^\infty(a) = abccbccccbcccccccbcccccccccbcccccccccccbcc \cdots$$

and $\tau(\sigma^\infty(a)) = 0100100001000000100000000100000000000100 \cdots$. Using the special form of the images by σ of b and c, it is not difficult to see that the difference between the position of the nth b and the $(n+1)$st b in $\sigma^\infty(a)$ is $2n+1$. Hence $\tau(\sigma^\infty(a))$ is the characteristic sequence of the set of squares and it is substitutive. From Corollary 2.9 the set of squares is never k-recognizable for any integer base k.

Definition 5.1. As a natural extension of the concept of recognizability, we may consider sets $X \subseteq \mathbb{N}$ having a characteristic sequence $\mathbb{1}_X$ which is (purely) substitutive. Such a set is said to be a *(purely) substitutive set*. In particular k-recognizable sets are substitutive.

With Theorem 5.2 it will turn out that the formalism of substitutive sets is equivalent to the one of abstract numeration systems.

Definition 5.2. [81] An *abstract numeration system* or *ANS* is a triple $\mathcal{S} = (L, A, <)$ where L is an infinite regular language over a totally ordered alphabet $(A, <)$. The map $\mathrm{rep}_{\mathcal{S}} : \mathbb{N} \to L$ is the one-to-one correspondence mapping $n \in \mathbb{N}$ onto the $(n+1)$th word in the genealogically ordered language L, which is called the *$\mathcal{S}$-representation* of n. In particular, a set $X \subseteq \mathbb{N}$ is *$\mathcal{S}$-recognizable*, if $\mathrm{rep}_{\mathcal{S}}(X)$ is regular, and, $\mathbb{N}$ is trivially S-recognizable because $\mathrm{rep}_S(\mathbb{N}) = L$. Recall that in the *genealogical order* (also called *radix* or *military* order), words are first ordered by increasing length and for words of the same length, one uses the lexicographic ordering induced by the order $<$ on A.

Example 5.2. Consider the language $L = a^*b^* \cup a^*c^*$ with $a < b < c$. The first words in L are $\varepsilon, a, b, c, aa, ab, ac, bb, cc, aaa, aab, aac, abb, \ldots$. This means that for the ANS $\mathcal{S}$ built on L, 0 is represented by ε, 1 by a, 2 by b, 3 by c, 4 by aa, *etc*. Since L contains exactly $2n+1$ words of length n for all $n \geqslant 0$, we have that n^2 is represented by a^n for all $n \geqslant 0$. In particular, the set $\{n^2 \mid n \geqslant 0\}$ is $\mathcal{S}$-recognizable because a^* is regular. It is well known that in a regular language L, the set of the first words of each length in the genealogically ordered language L is regular, see [115].

Pisot numeration systems are special cases of ANS. Indeed, if the numeration basis $U = (U_n)_{n \geqslant 0}$ defines a Pisot numeration system, then $\mathrm{rep}_U(\mathbb{N})$ is regular.

Example 5.3. Consider the Fibonacci sequence and the language $L = 1\{0, 01\}^* \cup \{\varepsilon\}$ defined in Example 2.1. To get the representation of an integer n, one can either decompose n using the greedy algorithm or, order genealogically the words in L and take the $(n+1)$th element.

Theorem 5.1. [81] *Let $\mathcal{S} = (L, A, <)$ be an abstract numeration system. Any ultimately periodic set is $\mathcal{S}$-recognizable.*

Note that in [78], it is in particular proved that this latter result cannot be extended to context-free languages. Specific cases of $\mathcal{S}$-recognizable sets are discussed in P. Lecomte and M. Rigo's chapter in [12, Chapter 3]. We have an extension of Theorem 3.1.

Theorem 5.2. *Let $x = (x_n)_{n \geqslant 0}$ be an infinite word over an alphabet B. This word is substitutive if and only if there exists an abstract numeration system $\mathcal{S} = (L, A, <)$ such that x is $\mathcal{S}$-automatic,* i.e.*, there exists a DFAO $(Q, A, \cdot, \{q_0\}, B, \tau)$ such that for all $n \geqslant 0$, $x_n = \tau(q_0 \cdot \mathrm{rep}_{\mathcal{S}}(n))$.*

A proof of this result is given in [105, 107] and a comprehensive treatment is given in [12, Chapter 3]. In that context, we also obtain an extension of Corollary 3.3.

Corollary 5.3. *Let $x = (x_n)_{n \geqslant 0}$ be an infinite substitutive word over an alphabet B. There exists an ANS $\mathcal{S}$ such that for all $b \in B$, $\{n \geqslant 0 \mid x_n = b\}$ is $\mathcal{S}$-recognizable. Conversely, if a set $X \subseteq \mathbb{N}$ is $\mathcal{S}$-recognizable, then its characteristic sequence is $\mathcal{S}$-automatic.*

Corollary 5.4. *A set $X \subseteq \mathbb{N}$ is substitutive if and only if there exists an ANS $\mathcal{S}$ such that X is $\mathcal{S}$-recognizable.*

5.2 Cobham's theorem for substitutive sets

In the context of substitutive sets of integers, *how could a Cobham-like theorem be expressed, i.e., what is playing the role of a base*? Assume that there exist two purely substitutive infinite words $x \in A^\omega$ and $y \in B^\omega$ respectively generated by the morphisms $\sigma : A^* \to A^*$ prolongable on $a \in A$ and $\tau : B^* \to B^*$ prolongable on $b \in B$, *i.e.*, $\sigma^\infty(a) = x$ and $\tau^\infty(b) = y$. Consider two codings $\lambda : A \to \{0,1\}$ and $\mu : B \to \{0,1\}$ such that $\lambda(x) = \mu(y)$. This situation corresponds to the case where a set (here, given by its characteristic word) is recognizable in two *a priori* different numeration systems.

If $A = B$ and $\tau = \sigma^m$ for some $m \geqslant 1$, nothing particular can be said about the infinite word $\lambda(x)$: iterating σ or σ^m from the same prolongable letter leads to the same fixed point. So we must introduce a notion analogous to the one of multiplicatively independent bases related to the substitutions σ and λ.

Definition 5.3. Let $\sigma : A^* \to A^*$ be a substitution over an alphabet A. The matrix $\mathbf{M}_\sigma \in \mathbb{N}^{A\times A}$ associated with σ is called the *incidence matrix* of σ and is defined by

$$\text{for all } a, b \in A,\ (\mathbf{M}_\sigma)_{a,b} = |\sigma(b)|_a\ .$$

A square matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$ with entries in $\mathbb{R}_{\geqslant 0}$ is *irreducible* if, for all i, j, there exists k such that $(\mathbf{M}^k)_{i,j} > 0$. A square matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$ with entries in $\mathbb{R}_{\geqslant 0}$ is *primitive* if there exists k such that, for all i, j, we have $(\mathbf{M}^k)_{i,j} > 0$. Similarly, a substitution over the alphabet A is *irreducible* (resp. *primitive*) if its incidence matrix is irreducible (resp. primitive). Otherwise stated, a substitution $\sigma : A^* \to A^*$ is primitive, if there exists an integer $n \geqslant 1$ such that, for all $a \in A$, all the letters in A appear in the image of $\sigma^n(a)$.

Let us denote by $\mathbf{P}$ the abelianisation map (or Parikh map) which maps a word w over $A = \{a_1, \ldots, a_r\}$ on the r-tuple ${}^t(|w|_{a_1}, \ldots, |w|_{a_r})$. The matrix $\mathbf{M}_\sigma$ can be defined by its columns:

$$\mathbf{M}_\sigma = \big(\mathbf{P}(\sigma(a_1)) \quad \cdots \quad \mathbf{P}(\sigma(a_r))\big)\ ,$$

and it satisfies:

$$\text{for all } w \in A^*,\ \mathbf{P}(\sigma(w)) = \mathbf{M}_\sigma \mathbf{P}(w)\ .$$

Remark 5.5. If a matrix $\mathbf{M}$ is primitive, the celebrated theorem of Perron can be used, see standard textbooks like [72] or [60, 114]. A presentation is also given in [83]. To recap some of the key points, $\mathbf{M}$ has a unique dominating real eigenvalue $\beta > 0$ and there exists an eigenvector with positive entries associated with β. Also, for all i, j, there exists $c_{i,j}$ such that $(\mathbf{M}^n)_{i,j} = c_{i,j}\beta^n + o(\beta^n)$. For instance, primitiveness of $\mathbf{M}_\sigma$ implies the existence of the frequency of any factor occurring in any fixed point of σ. Note that

$$\text{if } \mathbf{P}(w) = {}^t(p_1, \ldots, p_r),\ \text{then } |w| = \sum_{i=1}^{r} p_i. \tag{5.1}$$

Hence, for all $n \geqslant 0$, $|\sigma^n(a_j)|$ is obtained by summing up the entries in the jth column of $\mathbf{M}_\sigma^n$. If σ is primitive then there exists some C_j such that $|\sigma^n(a_j)| = C_j\beta^n + o(\beta^n)$. In particular, if σ is prolongable on a, then $|\sigma^n(a)| \sim C\beta^n$, for some $C > 0$.

In the general case of a matrix $\mathbf{M}$ with non-negative entries, one can use the Perron–Frobenius theorem for each of the irreducible components of $\mathbf{M}$ (they correspond to the strongly connected components of the associated graph, also called communicating classes). Thus any non-negative matrix $\mathbf{M}$ has a real eigenvalue α which is greater or equal to the modulus of any other eigenvalue. We call α the *dominating eigenvalue* of $\mathbf{M}$. Moreover, if we exclude the case where $\alpha = 1$, then there exists a positive integer p such that $\mathbf{M}^p$ has a dominating eigenvalue α^p which is a Perron number, see [83, p. 369]. A *Perron number* is an algebraic integer $\alpha > 1$ such that all its algebraic conjugates have modulus less than α. In particular, if we replace a prolongable substitution σ such that $\mathbf{M}_\sigma$ has a dominating eigenvalue $\alpha > 1$, with a convenient power σ^p of σ, we can assume that the dominating eigenvalue of σ is a Perron number.

Definition 5.4. Let $\sigma : A^* \to A^*$ be a substitution prolongable on $a \in A$ such that *all* letters of A have an occurrence in $\sigma^\infty(a)$. Let $\alpha > 1$ be the dominating eigenvalue of the incidence matrix of σ. Let $\phi : A \to B^*$ be a coding. We say $\phi(\sigma^\infty(a))$ is an *α-substitutive infinite word (with respect to σ)*. In view of Definition 5.1, this notion can be applied to subsets of $\mathbb{N}$. If moreover σ is primitive, then $\phi(\sigma^\infty(a))$ is said to be a *primitive α-substitutive infinite word (w.r.t. σ)*.

Observe that k-automatic infinite words are k-substitutive infinite words.

Example 5.4. Consider the substitution σ defined by $\sigma(a) = aa0a$, $\sigma(0) = 01$ and $\sigma(1) = 10$. Its dominating eigenvalue is 3. It is prolongable both on a, 0 and 1. The fixed point x of σ starting with 0 is the Thue-Morse sequence (see Example 3.1). Definition 5.4 does not implies that x is 3-substitutive because a does not appear in x. But the fixed point y of σ starting with a is 3-substitutive.

Example 5.5. Consider the so-called *Tribonacci word*, which is the unique fixed point of $\sigma : a \mapsto ab, b \mapsto ac, c \mapsto a$. See [117, 56]. The incidence matrix of σ is

$$\mathbf{M}_\sigma = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} .$$

One can check that $\mathbf{M}_\sigma^3$ contains only positive entries. So the matrix is primitive. Let $\alpha_T \simeq 1.839$ be the unique real root of the characteristic polynomial $-X^3 + X^2 + X + 1$ of $\mathbf{M}_\sigma$. The Tribonacci word $T = abacabaab\cdots$ is primitive α_T-substitutive. Let $\tau : a \mapsto 1, b, c \mapsto 0$ be a coding. The word $\tau(T)$ is the characteristic sequence of a primitive α_T-substitutive set of integers $\{0, 2, 4, 6, 7, \ldots\}$.

To explain the substitutive extension of Cobham's theorem we need the following definition.

Definition 5.5. Let $\mathcal{S}$ be a set of prolongable substitutions and x be an infinite word. If x is an α-substitutive infinite word w.r.t. a substitution σ belonging to $\mathcal{S}$, then x is said to be *α-substitutive with respect to $\mathcal{S}$*.

Let us consider the following Cobham-like statement depending on two sets $\mathcal{S}$ and $\mathcal{S}'$ of prolongable substitutions. It is useful to describe chronologically known results generalizing Cobham's theorem in terms of substitutions leading to the most general statement for all substitutions.

Statement $(\mathcal{S}, \mathcal{S}')$. Let $\mathcal{S}$ and $\mathcal{S}'$ be two sets of prolongable substitutions. Let α and β be two multiplicatively independent Perron numbers. Let $x \in A^\omega$ where A is a finite alphabet. Then the following are equivalent:

(1) the infinite word x is both α-substitutive w.r.t. $\mathcal{S}$ and β-substitutive w.r.t. $\mathcal{S}'$;
(2) the infinite word x is ultimately periodic.

Note that this statement excludes 1-substitutions, *i.e.*, substitutions with a dominating eigenvalue equal to 1, because Perron numbers are larger than 1. The case of 1-substitutive infinite words will be mentioned in Subsection 5.6. Also notice that the substitutions we are dealing with can be *erasing*, *i.e.*, at least one letter is sent onto the empty word. But from a result in [34, 9, 71], we can assume that the substitutions are non-erasing. Note that α and α^k are multiplicatively dependent.

Proposition 5.6. [49] *Let x be an α-substitutive infinite word. Then, there exists an integer $k \geqslant 1$ such that x is α^k-substitutive with respect to a non-erasing substitution.*

The implication (2) $\Rightarrow$ (1) in the above general statement is not difficult to obtain as mentioned in Remark 1.3 for the uniform situation.

Proposition 5.7. [47] *Let x be an infinite word over a finite alphabet and α be a Perron number. If x is periodic (resp. ultimately periodic) then x is primitive α-substitutive (resp. α-substitutive).*

Definition 5.6. Let $\sigma : A^* \to A^*$ and $\tau : B^* \to B^*$ be two substitutions. We say that σ *projects* on τ if there exists a coding $\phi : A \to B$ such that

$$\phi \circ \sigma = \tau \circ \phi \,. \tag{5.2}$$

The implication (1) $\Rightarrow$ (2) in Statement $(\mathcal{S}, \mathcal{S}')$ is known in many cases described below:

(i) When $\mathcal{S} = \mathcal{S}'$ is the set of *uniform* substitutions, this is the classical theorem of Cobham.
(ii) In [53] S. Fabre proves the statement when $\mathcal{S}$ is the set of uniform substitutions and $\mathcal{S}'$ is a set of non-uniform substitutions related to some non-standard numeration systems.
(iii) When $\mathcal{S} = \mathcal{S}'$ is the set of *primitive substitutions*, the statement is proved in [44]. The proof is based on a characterization of primitive substitutive sequences using the notion of return word [43].
(iv) When $\mathcal{S} = \mathcal{S}'$ is the set of *substitutions projecting on primitive substitutions*, the statement is proved in [45]. This result is applied to generalize (ii). Using a characterization of U-recognizable sets of integers for a Bertrand numeration basis U

[54], the main result of [45] extends Cobham's theorem for some large family of non-standard numeration systems. This latter result includes a result obtained previously in [15] for Pisot numeration systems.

(v) Definition 5.8 and Theorem 5.17 describe the situation where $\mathcal{S} = \mathcal{S}' = \mathcal{S}_{\text{good}}$ (defined later). It includes all known and previously described situations for substitutions.

(vi) In [42], Statement $(\mathcal{S}, \mathcal{S}')$ is proven for the most general case that is $\mathcal{S}$ and $\mathcal{S}'$ are both the set of all substitutions. The final argument is based on a fine study of return words for non-primitive substitutive sequences.

Example 5.6. The Tribonacci word T is purely substitutive but is k-automatic for no integer $k \geqslant 2$. Proceed by contradiction. Assume that there exists an integer $k \geqslant 2$ such that T is k-automatic. Then T is both k-substitutive and primitive α_T-substitutive. By Theorem 5.17, T must be ultimately periodic but it is not the case. The factor complexity of T is $p_T(n) = 2n + 1$. By the Morse–Hedlund theorem, see Remark 3.5, T is not ultimately periodic.

Let $L(x)$ be the set of all factors of the infinite word x. In [55], the following generalization of Cobham's theorem is proved.

Theorem 5.8. *Let $k, \ell \geqslant 2$ be two multiplicatively independent integers. Let x be a k-automatic infinite word and y be a ℓ-automatic infinite word. If $L(x) \subset L(y)$, then x is ultimately periodic.*

The same result is valid in the primitive case.

Theorem 5.9. [44] *Let x and y be respectively a primitive α-substitutive infinite word and a primitive β-substitutive infinite word such that $L(x) = L(y)$. If α and β are multiplicatively independent, then x and y are periodic.*

Note that under the hypothesis of Theorem 5.9 x and y are primitive substitutive infinite words. Thus $L(x) = L(y)$ whenever $L(x) \subset L(y)$. Observe that if y is the fixed point starting with a and x the fixed point starting with 0 of the substitution σ defined in Example 5.4, then $L(x) \subset L(y)$ but x is not ultimately periodic.

In Sections 5.3 and 5.4 we give the main arguments to prove Statement $(\mathcal{S}_{\text{good}}, \mathcal{S}_{\text{good}})$.

5.3 Density, syndeticity and bounded gaps

The proofs of most of the generalizations of Cobham's theorem are divided into two parts.

(i) Dealing with a subset X of integers, we have to prove that X is syndetic. Equivalently, dealing with an infinite word x, we have to prove that the letters occurring infinitely many times in x appear with bounded gaps.

(ii) In the second part of the proof, the ultimate periodicity of X or x has to be carried out.

This section is devoted to the description of the main arguments that lead to the complete treatment of (i).

In the original proof of Cobham one of the main arguments is that as k and ℓ are multiplicatively independent (we refer to Theorem 1.1) the set $\{k^n/\ell^m \mid n, m \in \mathbb{N}\}$ is dense in $[0, +\infty)$. In the uniform case these powers refer to the length of the iterates of the substitutions. Indeed, suppose $\sigma : A^* \to A^*$ is a k-uniform substitution. Then for any $a \in A$ we have $|\sigma^n(a)| = k^n$. Unsurprisingly, to be able to treat the non-uniform case, it is important to know that the set

$$\left\{ \frac{\mid \sigma^n(a) \mid}{\mid \tau^m(b) \mid} \mid n, m \in \mathbb{N} \right\}$$

is dense in $[0, +\infty)$, for some $a, b \in A$. We explain below that $|\sigma^n(a)|$ and $|\tau^m(b)|$ are governed by the dominating eigenvalue of their incidence matrices. First we focus on part (i) and consider infinite words.

5.3.1 The length of the iterates The length of the iterates are described in the following lemma. Note that it includes erasing substitutions and substitutions with a dominating eigenvalue equal to 1. Observe that for the substitution σ defined by $0 \mapsto 001$ and $1 \mapsto 11$ we have $|\sigma^n(0)| = (n+2)2^{n-1}$ and $|\sigma^n(1)| = 2^n$ showing that the situation is different from the uniform case. It can easily be described using the Jordan normal form of the incidence matrix $\mathbf{M}_\sigma$. Discussion of the following result can be found in [12, Section 4.7.3].

Lemma 5.10 (Chapter III.7 in [110]). *Let $\sigma : A \to A^*$ be a substitution. For all $a \in A$ one of the two following situations occur*

(1) there exists $N \in \mathbb{N}$ such that for all $n > N$, $|\sigma^n(a)| = 0$, or,
(2) there exist $d(a) \in \mathbb{N}$ and real numbers $c(a), \theta(a)$ such that

$$\lim_{n \to +\infty} \frac{|\sigma^n(a)|}{c(a)\, n^{d(a)}\, \theta(a)^n} = 1.$$

Moreover, in the situation (2)*, for all $i \in \{0, \ldots, d(a)\}$ there exists a letter $b \in A$ appearing in $\sigma^j(a)$ for some $j \in \mathbb{N}$ and such that*

$$\lim_{n \to +\infty} \frac{|\sigma^n(b)|}{c(b)\, n^i\, \theta(a)^n} = 1.$$

Definition 5.7. Let σ be a non-erasing substitution. For all $a \in A$, the pair $(d(a), \theta(a))$ defined in Lemma 5.10 is called the *growth type* of a. If (d, θ) and (e, β) are two growth types, we say that (d, θ) is *less than* (e, β) (or $(d, \theta) < (e, \beta)$) whenever $\theta < \beta$ or, $\theta = \beta$ and $d < e$.

Consequently if the growth type of $a \in A$ is less than the growth type of $b \in A$ then $\lim_{n \to +\infty} |\sigma^n(a)|/|\sigma^n(b)| = 0$. We say that $a \in A$ is a *growing letter* if $(d(a), \theta(a)) > (0, 1)$ or equivalently, if $\lim_{n \to +\infty} |\sigma^n(a)| = +\infty$.

We set $\Theta := \max\{\theta(a) \mid a \in A\}$, $D := \max\{d(a) \mid \forall a \in A : \theta(a) = \Theta\}$ and $A_{max} := \{a \in A \mid \theta(a) = \Theta, d(a) = D\}$. The dominating eigenvalue of $\mathbf{M}_\sigma$ is Θ. We

say that the letters of A_{max} are *of maximal growth* and that (D, Θ) is the *growth type* of σ. Consequently, we say that a substitutive infinite word y is (D, Θ)-*substitutive* if the underlying substitution is of growth type (D, Θ). Observe that, due to Lemma 5.10, any substitutive sequence is (D, Θ)-substitutive for some pair (D, Θ).

Observe that if $\Theta = 1$, then in view of the last part of Lemma 5.10, there exists at least one non-growing letter of growth type $(0, 1)$. Otherwise stated, if a letter has polynomial growth, then there exists at least one non-growing letter. Consequently σ is *growing* (*i.e.*, all its letters are growing) if and only if $\theta(a) > 1$ for all $a \in A$. We define

$$\lambda_\sigma : A^* \to \mathbb{R},\ u_0 \cdots u_{n-1} \mapsto \sum_{i=0}^{n-1} c(u_i)\, \mathbb{1}_{A_{max}}(u_i)\ ,$$

where $c : A \to \mathbb{R}_+$ is defined in Lemma 5.10. From Lemma 5.10 we deduce the following lemma.

Lemma 5.11. *For all $u \in A^*$, we have $\lim_{n\to+\infty} |\sigma^n(u)|/n^D\Theta^n = \lambda_\sigma(u)$.*

We say that the word $u \in A^*$ is of *maximal growth* if $\lambda_\sigma(u) \neq 0$.

Corollary 5.12. *Let σ be a substitution of growth type (D, Θ). For all $k \geqslant 1$, the growth type of σ^k is (D, Θ^k).*

5.3.2 Letters and words appear with bounded gaps Recall that the first step for Cobham's theorem is to prove that the letters occurring infinitely many times appear with bounded gaps. In our context, this implies the same property for words. Moreover, we can relax the multiplicative independence hypothesis in order to include 1-substitutions. Note that 1 and $\alpha > 1$ are multiplicatively dependent.

Theorem 5.13. [49] *Let $d, e \in \mathbb{N} \setminus \{0\}$ and $\alpha, \beta \in [1, +\infty)$ such that $(d, \alpha) \neq (e, \beta)$ and satisfying one of the following three conditions:*

(i) *α and β are multiplicatively independent;*
(ii) *$\alpha, \beta > 1$ and $d \neq e$;*
(iii) *$(\alpha, \beta) \neq (1, 1)$ and, $\beta = 1$ and $e \neq 0$, or, $\alpha = 1$ and $d \neq 0$.*

Let C be a finite alphabet. If $x \in C^\omega$ is both (d, α)-substitutive and (e, β)-substitutive then the words occurring infinitely many times in x appear with bounded gaps.

The main argument used to prove this in [49] is the following.

Theorem 5.14. *Let $d, e \in \mathbb{N}$ and $\alpha, \beta \in [1, +\infty)$. The set*

$$\Omega = \left\{ \frac{\alpha^n n^d}{\beta^m m^e} \mid n, m \in \mathbb{N} \right\}$$

is dense in $[0, +\infty)$ if and only if one of the following three conditions holds:

(i) *α and β are multiplicatively independent;*
(ii) *$\alpha, \beta > 1$ and $d \neq e$;*
(iii) *$\beta = 1$ and $e \neq 0$, or, $\alpha = 1$ and $d \neq 0$.*

Sketch of the proof of Theorem 5.13. We only consider the case where α and β are multiplicatively independent.

Let $\sigma : A^* \to A^*$ be a substitution prolongable on a letter a' having growth type (d, α). Let $\tau : B^* \to B^*$ be a substitution prolongable on a letter b' having growth type (e, β). Let $\phi : A \to C$ and $\psi : B \to C$ be two codings such that $\phi(\sigma^\infty(a')) = \psi(\tau^\infty(b')) = x$. Using Proposition 5.6 we may assume that σ and τ are non-erasing. Suppose there is a letter a having infinitely many occurrences in x but that appears with unbounded gaps. Then the letters in $\phi^{-1}(\{a\})$ appear with unbounded gaps. To avoid extra technicalities (a complete treatment is considered in [49]), we assume that there is a letter in $\phi^{-1}(\{a\})$ having maximal growth. Then, it is quite easy to construct, for all $n \in \mathbb{N}$, a word w_n of length $c_1 n^d \alpha^n$, appearing in y at the index $c_2 n^d \alpha^n$, that does not contain any letter of $\phi^{-1}(\{a\})$. On the other hand, using a kind of pumping lemma for substitutions, one can show that there is a letter of $\psi^{-1}(\{a\})$ in z at the index $c_3 n^e \beta^n$. Therefore, using Theorem 5.14, the letter a appears in a word $\phi(w_n)$ for some n. This is not possible.

Now let us explain how to extend this result for a single letter to words. It uses what is called in [103] the *substitutions of the words of length* n. Let u be a word of length n occurring infinitely often in x. To prove that u appears with bounded gaps in x, it suffices to prove that the letter 1 appears with bounded gaps in the infinite word $t \in \{0,1\}^{\mathbb{N}}$ defined by

$$t_i = \begin{cases} 1, & \text{if } x_i \cdots x_{i+n-1} = u; \\ 0, & \text{otherwise.} \end{cases}$$

Let A^n be the set of words of length n over A. The infinite word $y^{(n)} = (y_i \cdots y_{i+n-1})_{i \geqslant 0}$ over the alphabet A^n is a fixed point of the substitution $\sigma_n : (A^n)^* \to (A^n)^*$ defined, for all $(a_1 \cdots a_n)$ in A^n, by

$$\sigma_n((a_1 \cdots a_n)) = (b_1 \cdots b_n)(b_2 \cdots b_{n+1}) \cdots (b_{|\sigma(a_1)|} \cdots b_{|\sigma(a_1)|+n-1})$$

where $\sigma(a_1 \cdots a_n) = b_1 \cdots b_k$. For details, see Section V.4 in [103].

Let $\rho : A^n \to A^*$ be the coding defined by $\rho((b_1 \cdots b_n)) = b_1$ for all $(b_1 \cdots b_n) \in A^n$. We have $\rho \circ \sigma_n = \sigma \circ \rho$, and then $\rho \circ \sigma_n^k = \sigma^k \circ \rho$. Hence, if σ is of growth type (d, α) then $y^{(n)}$ is (d, α)-substitutive. Let $f : A^n \to \{0, 1\}$ be the coding defined by

$$f((b_1 \cdots b_n)) = \begin{cases} 1, & \text{if } b_1 \cdots b_n = u; \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to see that $f(y^{(n)}) = t$, hence t is (d, α)-substitutive. Then one proceeds in the same way with τ and uses the result for letters to conclude the proof. □

5.4 Ultimate periodicity

Definition 5.8. Let $\sigma : A^* \to A^*$ be a substitution. If there exists a sub-alphabet $B \subseteq A$ such that for all $b \in B$, $\sigma(b) \in B^*$, then the substitution $\tau : B^* \to B^*$ defined by the restriction $\tau(b) = \sigma(b)$, for all $b \in B$, is a *sub-substitution* of σ. Note that σ is in particular a sub-substitution of itself.

The substitution σ having α as dominating eigenvalue is a *"good" substitution*, if it

has a *primitive* sub-substitution whose dominating eigenvalue is α. So let us stress the fact that to be a "good" substitution, the sub-substitution has to be primitive and have the same dominating eigenvalue as the original substitution. We let $\mathcal{S}_{\text{good}}$ denote the set of good substitutions.

Remark 5.15. For all growing substitutions σ, there exists an integer k such that σ^k has a primitive sub-substitution. Hence by taking a convenient power of σ, the substitution can always be assumed to have a primitive sub-substitution.

Note that primitive substitutions and uniform substitutions are good substitutions. Now consider the substitution $\sigma : \{a, 0, 1\}^* \to \{a, 0, 1\}^*$ given by $\sigma : a \mapsto aa0, 0 \mapsto 01, 1 \mapsto 0$. Its dominating eigenvalue is 2 and it has only one primitive sub-substitution ($0 \mapsto 01$, $1 \mapsto 0$) whose dominating eigenvalue is $(1 + \sqrt{5})/2$, hence it is not a good substitution.

Remark 5.16. Let $\sigma : A^* \to A^*$ and $\tau : B^* \to B^*$ be two substitutions such that σ projects on τ, recall (5.2) for the definition of projection. There exists a coding $\phi : A \to B$ such that $\phi \circ \sigma = \tau \circ \phi$. Note that $\phi \circ \sigma^n = \tau^n \circ \phi$. If τ is primitive, then it follows that σ belongs to $\mathcal{S}_{\text{good}}$.

Theorem 5.17. *Let α and β be two multiplicatively independent Perron numbers. Let $x \in A^\omega$ where A is a finite alphabet. Then the following are equivalent:*

(i) *the infinite word x is both α-substitutive w.r.t. $\mathcal{S}_{\text{good}}$ and β-substitutive w.r.t. $\mathcal{S}_{\text{good}}$;*
(ii) *the infinite word x is ultimately periodic.*

Proof. Let $\sigma : B^* \to B^*$ (resp. $\tau : C^* \to C^*$) be a substitution in $\mathcal{S}_{\text{good}}$ having α (resp. β) as its dominating eigenvalue and ϕ (resp. ψ) be a coding such that $x = \phi(\sigma^\infty(b))$ for some $b \in B$ (resp. $x = \psi(\tau^\infty(c))$ for some $c \in C$).

Let us first suppose that both substitutions are growing. In this way, taking a power if needed, we can suppose that they have primitive sub-substitutions.

By Theorem 5.13, the factors occurring infinitely many times in x appear with bounded gaps. Hence for any primitive and growing sub-substitutions $\overline{\sigma}$ and $\overline{\tau}$ of σ and of τ respectively, we have $\phi(L(\overline{\sigma})) = \psi(L(\overline{\tau})) = L$. Using Theorem 5.9 it follows that L is periodic, *i.e.*, there exists a shortest word u, appearing infinitely many times in x, such that $L = L(u^\omega)$. Thus u appears with bounded gaps. Let $\mathcal{R}_u$ be the set of return words to u. A word w is a *return word* to u if $wu \in L(x)$, u is a prefix of wu and u has exactly two occurrences in wu. Since u appears with bounded gaps, the set $\mathcal{R}_u$ is finite. There exists an integer N such that all words $wu \in L(x_N x_{N+1} \cdots)$ appear infinitely many times in x for all $w \in \mathcal{R}_u$. Hence these words appear with bounded gaps in x. We set $t = x_N x_{N+1} \cdots$ and we will prove that t is periodic. Consequently x would be ultimately periodic. We can suppose that u is a prefix of t. Then t is a concatenation of return words to u. Let w be a return word to u. It appears with bounded gaps hence it appears in some $\phi(\overline{\sigma}^n(a))$, where $\overline{\sigma}$ is a primitive and growing sub-substitution, and there exist two words, p and q, and an integer i such that $wu = pu^i q$. As $|u|$ is the least period of L it must be that $wu = u^i$. It follows that $t = u^\omega$.

If, for example, σ is non-growing, then a result of J.-J. Pansiot [94] asserts that either by modifying in a suitable way σ and ϕ (in that case α could be replaced by a power

of α) we can suppose σ is growing or $L(\sigma^\infty(b))$ contains the language of a periodic infinite word. We have treated the first case before. For the second case it suffices to use Theorem 5.13. $\square$

Suppose α and β are multiplicatively independent real numbers and that x is a α-substitutive infinite word w.r.t. $\mathcal{S}_{\mathrm{good}}$ and y is a β-substitutive infinite word w.r.t. $\mathcal{S}_{\mathrm{good}}$ satisfying $L(x) \subset L(y)$. Then the conclusion of Theorem 5.8 is far from true. It suffices to look at Example 5.4 and the observation made after Theorem 5.9.

Remark 5.18. The Statement $(\mathcal{S}, \mathcal{S}')$ remains open when $\mathcal{S}$ is the set of substitutions which are not good. Nevertheless there are cases where we can say more. For example, if x is both α-substitutive and β-substitutive (with α and β being multiplicatively independent), and, $L(x)$ contains the language of a periodic sequence then, from Theorem 5.13, we deduce that x is ultimately periodic.

Moreover, as we will see in the next section, this statement holds true in the purely substitutive context.

5.5 The case of fixed points

Now let restrict ourselves to the purely substitutive case. In this setting Cobham's theorem holds true. Note that in the statement of the following result, α and β are necessarily Perron numbers. Moreover, since the substitutions are growing, then α and β must be larger than one.

Theorem 5.19. *Let $\sigma : A^* \to A^*$ and $\tau : A^* \to A^*$ be two non-erasing growing substitutions prolongable on $a \in A$ with respective dominating eigenvalues α and β. Suppose that all letters of A appear in $\sigma^\infty(a)$ and in $\tau^\infty(a)$ and that α and β are multiplicatively independent. If $x = \sigma^\infty(a) = \tau^\infty(a)$, then x is ultimately periodic.*

Proof. Thanks to Remark 5.15, we may assume that σ has a primitive sub-substitution. Using Theorem 5.13, the letters appearing infinitely often in x appear with bounded gaps. Let $\overline{\sigma} : \overline{A} \to \overline{A}$ be a primitive sub-substitution of σ. Let $c \in \overline{A}$. Suppose that there exists a letter b, appearing infinitely many times in x, which does not belong to $\overline{A}$. Then the word $\sigma^n(c) = \overline{\sigma}^n(c)$ does not contain b and b could not appear with bounded gaps. Consequently all letters (and in particular a letter of maximal growth) appearing infinitely often in x belong to $\overline{A}$. Hence $\overline{\sigma}$ also has α as dominating eigenvalue and σ is a "good" substitution. In the same way τ is a "good" substitution. Theorem 5.17 concludes the proof. $\square$

5.6 Back to numeration systems

Let $\mathcal{S}$ be an abstract numeration system. There is no reason for the substitutions describing characteristic words of $\mathcal{S}$-recognizable sets (see Corollary 5.4) to be primitive. To obtain a Cobham type theorem for families of abstract numeration systems, one has to interpret Theorem 5.17 in this formalism.

5.6.1 Polynomially growing abstract numeration systems Here we only mention the following result. The paper [40] is also of interest. It is well-known that the growth function counting the number of words of length n in a regular language is either *polynomial, i.e.,* in $\mathcal{O}(n^k)$ for some integer k or *exponential, i.e.,* in $\Omega(\theta^n)$ for some $\theta > 1$.

Proposition 5.20. [49] *Let $\mathcal{S} = (L, A, <)$ (resp. $\mathcal{T} = (M, B, \prec)$) be an abstract numeration system where L is a polynomial regular language (resp. M is an exponential regular language). A set X of integers is both $\mathcal{S}$-recognizable and $\mathcal{T}$-recognizable if and only if X is ultimately periodic.*

5.6.2 Bertrand basis and ω_α-substitutive words Let U be a Bertrand numeration basis such that $\mathrm{rep}_U(\mathbb{N}) = L(\alpha)$ where α is a Parry number which is not an integer. In [54] a substitution denoted by ω_α is defined. The importance of this substitution is justified by Theorem 5.21. If $d_\alpha(1) = t_1 \cdots t_n 0^\omega$, $t_n \neq 0$, then ω_α is defined on the alphabet $\{1, \ldots, n\}$ by

$$1 \mapsto 1^{t_1}2,\ \ldots,\ n-1 \mapsto 1^{t_{n-1}}n,\ n \mapsto 1^{t_n}\ .$$

If $d_\alpha(1) = t_1 \cdots t_n (t_{n+1} t_{n+2} \cdots t_{n+m})^\omega$, where n and m are minimal and where $t_{n+1} + t_{n+2} + \cdots + t_{n+m} \neq 0$, then ω_α is defined on the alphabet $\{1, \cdots, n+m\}$ by

$$1 \mapsto 1^{t_1}2,\ \ldots,\ n+m-1 \mapsto 1^{t_{n+m-1}}(n+m),\ n+m \mapsto 1^{t_{n+m}}(n+1)\ .$$

In both cases the substitution ω_α is primitive and has α as dominating eigenvalue. A substitution that projects (see Definition 5.6) on ω_α is called a ω_α-*substitution* and we call each infinite word which is the image under a coding of a fixed point of a ω_α-substitution a ω_α-*substitutive* infinite word (α-automatic infinite word in [54]).

Theorem 5.21. [54, Corollary 1] *Let U be a Bertrand numeration basis such that $\mathrm{rep}_U(\mathbb{N}) = L(\alpha)$ where α is a Parry number. A set $X \subset \mathbb{N}$ is U-recognizable if and only if its characteristic sequence $\mathbb{1}_X$ is ω_α-substitutive.*

Remark 5.16 and Theorem 5.17 imply the following result.

Theorem 5.22. [45] *Let U and V be two Bertrand numeration systems. Let α and β be two multiplicatively independent Parry numbers such that $\mathrm{rep}_U(\mathbb{N}) = L(\alpha)$ and $\mathrm{rep}_V(\mathbb{N}) = L(\beta)$. A set $X \subseteq \mathbb{N}$ is U-recognizable and V-recognizable if and only if X is ultimately periodic.*

6 Cobham's theorem in various contexts

6.1 Regular sequences

Regular sequences as presented in [6, 7, 9] are a generalization of automatic sequences for sequences taking infinitely many values. Many examples of such sequences are given in the first two references. Also see [41] for a generalization of the notion of automaticity

in the framework of group actions. Let R be a commutative ring. Let $k \geqslant 2$. Consider a sequence $x = (x_n)_{n \geqslant 0}$ taking values in some R-module. If the R-module generated by all sequences in the k-kernel $N_k(x)$ is finitely generated (recall Theorem 3.2) then the sequence x is said to be (R, k)-*regular*.

Theorem 6.1 (Cobham–Bell theorem [10]). *Let R be a commutative ring*[3]*. Let k, ℓ be two multiplicatively independent integers. If a sequence $x \in R^{\mathbb{N}}$ is both (R, k)-regular and (R, ℓ)-regular, then it satisfies a linear recurrence over R.*

6.2 Algebraic setting and quasi-automatic functions

In [32] G. Christol characterized p-recognizable sets in terms of formal power series.

Theorem 6.2. *Let p be a prime number and $\mathbb{F}_p$ be the field with p elements. A subset $A \subset \mathbb{N}$ is p-recognizable if and only if $f(X) = \sum_{n \in A} X^n \in \mathbb{F}_p[[X]]$ is algebraic over $\mathbb{F}_p(X)$.*

This was applied to Cobham's theorem in [33] to obtain an algebraic version.

Theorem 6.3. *Let A be a finite alphabet, $x \in A^{\mathbb{N}}$, and, K_1 and K_2 be two finite fields with different characteristics. Let $\alpha_1 : A \to K_1$ and $\alpha_2 : A \to K_2$ be two one-to-one maps. If $f(X) = \sum_{n \in \mathbb{N}} \alpha_1(x_n) X^n \in K_1[[X]]$ is algebraic over $K_1(X)$ and $f(X) = \sum_{n \in \mathbb{N}} \alpha_2(x_n) X^n \in K_2[[X]]$ is algebraic over $K_2(X)$ then $f(X)$ is rational.*

Quasi-automatic functions are introduced by Kedlaya in [74]. Also see [75] where Christol's theorem is generalized to Hahn's generalized power series. In this algebraic setting, an extension of Cobham's theorem is proved by Adamczewski and Bell in [1]. Details are given in the chapter "Automata in number theory" of this handbook.

6.3 Real numbers and verification of infinite-state systems

Sets of numbers recognized by finite automata arise when analyzing systems with unbounded mixed variables taking integer or real values. Therefore systems such as timed or hybrid automata are considered [17]. One needs to develop data structures representing sets manipulated during the exploration of infinite state systems. For instance, it is often needed to compute the set of reachable configurations of such a system. Let $k \geqslant 2$ be an integer. Considering separately integer and fractional parts, a real number $x > 0$ can be decomposed as

$$x = \sum_{i=0}^{d} c_i \, k^i + \sum_{i=1}^{+\infty} c_{-i} \, k^{-i}, \; c_i \in [\![0, k-1]\!], \; i \leqslant d, \tag{6.1}$$

and gives rise to the infinite word $c_d \cdots c_0 \star c_{-1} c_{-2} \cdots$ over $[\![0, k-1]\!] \cup \{\star\}$ which is a k-*ary representation* of x. Note that rational numbers of the kind p/k^n have two k-ary

[3] Note that in [6] the ground ring R is assumed to be Noetherian (every ideal in R is finitely generated), but this extra assumption is not needed in the above statement.

representations, one ending with 0^ω and one with $(k-1)^\omega$. For the representation of negative elements, one can consider base k-complements or signed number representations [77], the sign being determined by the most significant digit which is thus 0 or $k-1$ (and this digit may be repeated an arbitrary number of times). For definition of Büchi and Muller automata, see the first part of this handbook.

Definition 6.1. A set $X \subseteq \mathbb{R}$ is *k-recognizable* if there exists a Büchi automaton accepting all the k-ary representations of the elements in X. Such an automaton is called a *Real Number Automaton* or *RNA*.

These notions extend naturally to subsets of $\mathbb{R}^d$ and to *Real Vector Automata* or *RVA*. Also the Büchi theorem 4.5 holds for a suitable structure $\langle \mathbb{R}, \mathbb{Z}, +, <, V_k \rangle$, see [22].

Theorem 6.4. [21] *If $X \subseteq \mathbb{R}^d$ is definable by a first-order formula in $\langle \mathbb{R}, \mathbb{Z}, +, < \rangle$, then X written in base $k \geqslant 2$ is accepted by a weak deterministic RVA $\mathcal{A}$.*

Weakness means that each strongly connected component of $\mathcal{A}$ contains only accepting states or only non-accepting states.

Theorem 6.5. [18] *Let $k, \ell \geqslant 2$ be two multiplicatively independent integers. If $X \subseteq \mathbb{R}$ is both k- and ℓ-recognizable by two weak deterministic RVA, then it is definable in $\langle \mathbb{R}, \mathbb{Z}, +, < \rangle$.*

The extension of the Cobham–Semenov theorem for subsets of $\mathbb{R}^d$ in this setting is discussed in [20], see also [24] for a comprehensive presentation. The case of two coprime bases was first considered in [18]. Surprisingly, if the multiplicatively independent bases $k, \ell \geqslant 2$ share the same prime factors, then there exists a subset of $\mathbb{R}$ that is both k- and ℓ-recognizable but not definable in $\langle \mathbb{R}, \mathbb{Z}, +, < \rangle$, see [19]. This shows a main difference between recognizability of subsets of real numbers written in base k for (general) Büchi automata and weak deterministic RVA. Though written in a completely different language, a similar result was independently obtained in [2]. This latter paper is motivated by the study of some fractal sets.

6.4 Dynamical systems and subshifts

In this section we would like to express a Cobham-type theorem in terms of dynamical systems called substitutive subshifts. Theorem 5.9 will appear as a direct corollary of these developments.

We first need some definitions.

A *dynamical system* is a pair (X, S) where X is a compact metric space and S a continuous map from X onto itself. The dynamical system (X, S) is *minimal* whenever X and the empty set are the only S-invariant closed subsets of X, that is, $S(X) = X$. We say that a minimal system (X, S) is *periodic* whenever X is finite.

Let (X, S) and (Y, T) be two dynamical systems. We say that (Y, T) is a *factor* of (X, S) if there is a continuous and onto map $\phi : X \to Y$ such that $\phi \circ S = T \circ \phi$ (ϕ is

called a *factor map*). If ϕ is one-to-one we say that ϕ is an *isomorphism* and that (X, S) and (Y, T) are *isomorphic*.

Let A be an alphabet. We endow A^ω with the infinite product of the discrete topologies. It is a metric space where the metric is given by

$$d(x, y) = \frac{1}{2^n} \text{ with } n = \inf\{k \mid x_k \neq y_k\}, \tag{6.2}$$

where $x = (x_n)_{n\geqslant 0}$ and $y = (y_n)_{n\geqslant 0}$ are two elements of A^ω. A *subshift* on A is a pair $(X, T_{|X})$ where X is a closed T-invariant subset of A^ω and T is the *shift transformation* $T : A^\omega \to A^\omega,\ (x_n)_{n\geqslant 0} \mapsto (x_{n+1})_{n\geqslant 0}$.

Let u be a word over A. The set $[u]_X = \{x \in X \mid x_0 \cdots x_{|u|-1} = u\}$ is a *cylinder*. The family of these sets is a base of the induced topology on X. When there is no misunderstanding, we write $[u]$ and T instead of $[u]_X$ and $T_{|X}$.

Let $x \in A^\omega$. The set $\{y \in A^\omega \mid L(y) \subseteq L(x)\}$ is denoted $\Omega(x)$. It is clear that $(\Omega(x), T)$ is a subshift. We say that $(\Omega(x), T)$ is the *subshift generated* by x. When x is a sequence, we have $\Omega(x) = \overline{\{T^n x \mid n \in \mathbb{N}\}}$. Observe that $(\Omega(x), T)$ is minimal if and only if x is uniformly recurrent, i.e., all its factors occur infinitely often in x and for each factor u of x, there exists a constant K such that the distance between two consecutive occurrences of u in x is bounded by K.

Let ϕ be a factor map from the subshift (X, T) on the alphabet A onto the subshift (Y, T) on the alphabet B. Here $x_{[i,j]}$ denotes the word $x_i \cdots x_j$, $i \leqslant j$. The Curtis–Hedlund–Lyndon theorem [83, Thm. 6.2.9] asserts that ϕ is a *sliding block code*: there exists an *r-block map* $f : A^r \to B$ such that $(\phi(x))_i = f(x_{[i,i+r-1]})$ for all $i \in \mathbb{N}$ and $x \in X$. We shall say that f is a *block map associated to* ϕ and that f *defines* ϕ. If $u = u_0 u_1 \cdots u_{n-1}$ is a word of length $n \geqslant r$ we define $f(u)$ by $(f(u))_i = f(u_{[i,i+r-1]})$, $i \in \{0, 1, \cdots, n - r + 1\}$. Let C denote the alphabet A^r and $Z = \{(x_{[i,r+i-1]})_{i\geqslant 0} \mid (x_n)_{n\geqslant 0} \in X\}$. It is easy to check that the subshift (Z, T) is isomorphic to (X, T) and that f induces a 1-block map (a coding) from C onto B which defines a factor map from (Z, T) onto (Y, T).

We can now state a Cobham-type theorem for subshifts generated by substitutive sequences. Observe that it implies Theorem 5.9 and Statement $(\mathcal{S}, \mathcal{S}')$ when $\mathcal{S} = \mathcal{S}'$ is the set of primitive substitutions.

Theorem 6.6. *Let (X, T) and (Y, T) be two subshifts generated respectively by a primitive α-substitutive sequence x and by a primitive β-substitutive sequence y. Suppose (X, T) and (Y, T) both factorize to the subshift (Z, T). If α and β are multiplicatively independent then (Z, T) is periodic.*

Below we give a sketch of the proof, which involves the concept of an ergodic measure. An *invariant measure* for the dynamical system (X, S) is a probability measure μ, on the σ-algebra $\mathcal{B}(X)$ of Borel sets, with $\mu(S^{-1}B) = \mu(B)$ for all $B \in \mathcal{B}(X)$; the measure is *ergodic* if every S-invariant Borel set has measure 0 or 1. The set of invariant measures for (X, S) is denoted by $\mathcal{M}(X, S)$. The system (X, S) is *uniquely ergodic* if $\#(\mathcal{M}(X, S)) = 1$. For expository books on subshifts and/or ergodic theory, see [37, 76, 83, 103, 79].

It is well known that the subshifts generated by primitive substitutive sequences are uniquely ergodic [103].

Let $\phi : X \to Z$ and $\psi : Y \to Z$ be two factor maps. Suppose that (Z, T) is not periodic. We will prove that α and β are multiplicatively independent.

Let μ and λ be the unique ergodic measures of (X, T) and (Y, T) respectively. It is not difficult to see that (Z, T) is also generated by a primitive substitutive sequence and consequently is uniquely ergodic. Let δ be its unique ergodic measure. We notice that $\phi\mu$ defined by $\phi\mu(A) = \mu(\phi^{-1}(A))$, for all Borel sets A of Z, and $\psi\lambda$ defined by $\psi\lambda(A) = \mu(\psi^{-1}(A))$, for all Borel sets A of Z, are invariant measures for (Z, T). Hence $\phi\mu = \delta = \psi\lambda$. Let us give more details about these measures in order to conclude the proof.

Theorem 6.7. [69] *Let (Ω, T) be a subshift generated by a primitive purely γ-substitutive sequence and m be its unique ergodic measure. Then, the measures of cylinders in Ω lie in a finite union of geometric progressions. There exists a finite set $\mathcal{F}$ of positive real numbers such that*

$$\{m(C) \mid C \text{ cylinder of } X\} \subset \bigcup_{n\in\mathbb{N}} \gamma^{-n}\mathcal{F} .$$

In conjunction with the next result and using the pigeon hole principle we will conclude the proof.

Proposition 6.8. [46] *Let (Ω, T) be a subshift generated by a primitive substitutive sequence on the alphabet A. There exists a constant K such that for any block map $f : A^{2r+1} \to B$, we have $\#(f^{-1}(\{u\})) \leqslant K$ for all u appearing in some sequences of $f(\Omega)$.*

From these last two results we deduce that there exist two sets of numbers $\mathcal{F}_X$ and $\mathcal{F}_Y$ such that

$$\begin{aligned}\{\delta(C) \mid C \text{ cylinder of } Z\} =& \{\mu(\phi^{-1}(C)) \mid C \text{ cylinder of } Z\}\\ =& \{\lambda(\psi^{-1}(C)) \mid C \text{ cylinder of } X\}\\ \subset& \left(\bigcup_{n\in\mathbb{N}} \alpha^{-n}\mathcal{F}_X\right) \bigcap \left(\bigcup_{n\in\mathbb{N}} \beta^{-n}\mathcal{F}_Y\right).\end{aligned}$$

The sets $\mathcal{F}_X$ and $\mathcal{F}_Y$ being finite, there exist two cylinder sets U and V of Z, $a \in \mathcal{F}_X$, $b \in \mathcal{F}_Y$ and n, m, r, s four distinct positive integers, such that

$$a\alpha^{-n} = \delta(U) = b\beta^{-m} \text{ and } a\alpha^{-r} = \delta(V) = b\beta^{-s} .$$

Consequently α and β are multiplicatively dependent.

6.5 Tilings

6.5.1 From definable sets Let A be a finite alphabet. An *array* in $\mathbb{N}^d$ is a map $\mathcal{T} : \mathbb{N}^d \to A$. It can be viewed as a tiling of $\mathbb{R}^d_+$. The collection of all these arrays is $A^{\mathbb{N}^d}$. For all $\mathbf{x} \in \mathbb{N}^d$, let $|\mathbf{x}|$ denote the sum of the coordinates of $\mathbf{x}$ and $B(\mathbf{x}, r)$ be the set $\{(y_1, \ldots, y_d) \in \mathbb{N}^d \mid 0 \leqslant y_i - x_i < r, 1 \leqslant i \leqslant d\}$.

We say $\mathcal{T}$ is *periodic* (resp. *ultimately periodic*) if there exists $\mathbf{p} \in \mathbb{N}^d$ such that $\mathcal{T}(\mathbf{x}+\mathbf{p}) = \mathcal{T}(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{N}^d$ (resp. for all large enough $\mathbf{x}$). We also need another notion of periodicity. We say that $Z \subset \mathbb{N}^d$ is $\mathbf{p}$-*periodic inside* $X \subset \mathbb{N}^d$ if for any $\mathbf{x} \in X$ with $\mathbf{x}+\mathbf{p} \in X$ we have

$$\mathbf{x} \in Z \text{ if and only if } \mathbf{x}+\mathbf{p} \in Z\ .$$

We say that Z is *locally periodic* if there exists a non-empty finite set $V \subset \mathbb{N}^d$ of non-zero vectors such that for some $K > \max\{|\mathbf{v}| \mid \mathbf{v} \in V\}$ and $L \geqslant 0$ one has:

$$(\forall \mathbf{x} \in \mathbb{N}^d, |\mathbf{x}| \geqslant L)(\exists \mathbf{v} \in V)(Z \text{ is } \mathbf{v}\text{-periodic inside } B(\mathbf{x}, K))\ .$$

Observe that for $d = 1$, local periodicity is equivalent to ultimate periodicity. We say $\mathcal{T}$ is *pseudo-periodic* if for all $a \in A$, $\mathcal{T}^{-1}(a)$ is locally periodic and every $(d-1)$-section of $\mathcal{T}^{-1}(a)$, say $S(i,n) = \{\mathbf{x} \in \mathcal{T}^{-1}(a) \mid x_i = n\}$, $1 \leqslant i \leqslant d$ and $n \in \mathbb{N}$, is pseudo-periodic (ultimately periodic when $d-1=1$). The following criterion is due to Muchnik, see [91] for the proof.

Proposition 6.9. *Let $E \subset \mathbb{N}^d$ and $\mathcal{T} : \mathbb{N}^d \to \{0,1\}$ be its characteristic function. The following are equivalent:*

(i) E is definable in the Presburger arithmetic;
(ii) $\mathcal{T}$ is pseudo-periodic;
(iii) for all $a \in \{0,1\}$, there exist $n \in \mathbb{N}$, $\mathbf{v}_i \in \mathbb{N}^d$ and finite sets $V_i \subset \mathbb{N}^d$, $0 \leqslant i \leqslant n$ such that

$$\mathcal{T}^{-1}(a) = V_0 \cup \left(\bigcup_{1 \leqslant i \leqslant n} \left(\mathbf{v}_i + \sum_{\mathbf{v} \in V_i} \mathbb{N}\mathbf{v} \right) \right)\ .$$

Let p be a positive integer and A be a finite alphabet. A p-*substitution* (or substitution if we do not need to specify p) is a map $S : A \to A^{B_p}$ where $B_p = B(\mathbf{0}, p) = \Pi_{i=1}^d \{0, \cdots, p-1\}$. The substitution S can be considered as a function from $A^{\mathbb{N}^d}$ into itself by setting

$$S((\mathcal{T}(\mathbf{x})) = [S(\mathcal{T}(\mathbf{y}))](\mathbf{z}), \text{ for all } \mathcal{T} \in A^{\mathbb{N}^d}$$

where $\mathbf{y} \in \mathbb{N}^d$ and $\mathbf{z} \in B_p$ are the unique vectors satisfying $\mathbf{x} = p\mathbf{y} + \mathbf{z}$.

In the same way, we can define $S : A^{B_{p^n}} \to A^{B_{p^{n+1}}}$. We remark that $S^n(a) = S(S^{n-1}(a))$ for all $a \in A$ and $n > 0$. We say $\mathcal{T}$ is *generated by a p-substitution* if there exist a coding ϕ and a fixed point $\mathcal{T}_0$ of a p-substitution such that $\mathcal{T} = \phi \circ \mathcal{T}_0$.

In [30] the authors proved the following theorem, which is analogous to Theorem 3.1.

Theorem 6.10. *Let $p \geqslant 2$ and $d \geqslant 1$. A set $E \subset \mathbb{N}^d$ is p-recognizable if and only if the characteristic function of E is generated by a p-substitution.*

Hence we can reformulate the Cobham–Semenov theorem as follows [113].

Theorem 6.11 (Cobham–Semenov theorem, Version 2). *Let p and q be two multiplicatively independent integers greater or equal to 2. Then, the array $\mathcal{T}$ is generated by both a p-substitution and a q-substitution if and only if $\mathcal{T}$ is pseudo-periodic.*

A dynamical proof of this can be given as for the unidimensional case, see [48] for the primitive case.

6.5.2 Self-similar tilings In [38], a Cobham-like theorem is expressed in terms of self-similar tilings of $\mathbb{R}^d$ with a proof using ergodic measures, see [116] for more about self-similar tilings. From the point of view of dynamical systems, the main result in [93] is also a Cobham-like theorem for self-similar tilings.

6.6 Toward Cobham's theorem for the Gaussian integers

I. Kátai and J. Szabó proved in [73] that the sequences $((-p+i)^n)_{n\geqslant 0}$ and $((-p-i)^n)_{n\geqslant 0}$ give rise to numeration systems whose set of digits is $\{0, 1, \ldots, p^2\}$, $p \in \mathbb{N} \setminus \{0\}$. It is an exercise to check that when $p \in \mathbb{N} \setminus \{0\}$ and $q \in \mathbb{N} \setminus \{0\}$ are different then $-p + i$ and $-q + i$ are multiplicatively independent. Therefore one could expect a Cobham-type theorem for the set of Gaussian integers $\mathbb{G} = \{a + ib \mid a, b \in \mathbb{Z}\}$. A subset $S \subset \mathbb{G}$ is *periodic* if there exists $h \in \mathbb{G}$ such that, for all $g \in \mathbb{G}$, $s \in S$ if and only if $s + gh \in S$. G. Hansel and T. Safer conjectured in [65] the following:

Conjecture 6.12. *Let p and q be two different positive integers and $S \in \mathbb{G}$. Then the following are equivalent.*

(i) The set S is $(-p + i)$-recognizable and $(-q + i)$-recognizable;
(ii) There exists a periodic set P such that the symmetric difference set $S\Delta P$ is finite.

The proof that (ii) implies (i) is easy. They tried to prove the other implication using the following (classical) steps:

(1) $D_{p,q} = \left\{ \frac{(-p+i)^n}{(-q+i)^m} \mid n, m \in \mathbb{Z} \right\}$ is dense in $\mathbb{C}$.
(2) S is syndetic
(3) S is periodic up to some finite set.

They succeeded in proving (ii) as given by the next result.

Theorem 6.13. *Let p and q be two positive integers such that the set $D_{p,q}$ is dense in $\mathbb{C}$. Let $S \subset \mathbb{G}$ be $(-p + i)$-recognizable and $(-q + i)$-recognizable. Then, S is syndetic.*

Let us make some observations about the density of the set $D_{p,q}$. Let $-p + i = ae^{i\theta}$ and $-q + i = be^{i\phi}$.

Proposition 6.14. *The following are equivalent.*

(i) The set $D_{p,q}$ is dense in $\mathbb{C}$;
(ii) The set $D_{p,q}$ is dense on the circle: $\{e^{i\theta} \mid \theta \in \mathbb{R}\} \subset \overline{D_{p,q}}$;
(iii) The following numbers are rationally independent (or linearly dependent over $\mathbb{Q}$):

$$\frac{\ln b}{\ln a},\ \frac{\theta}{2\pi}\frac{\ln b}{\ln a} - \frac{\phi}{2\pi},\ 1\ .$$

The equivalence between (i) and (iii) is proven in [65] from an easy computation. The equivalence between (i) and (ii) comes from the fact that $p^2 + 1$ and $q^2 + 1$ are multiplicatively independent, see [65, Prop. 2]. As an example, take $p = 1$ and $p = 2$. Then, $a = \sqrt{2}$, $b = \sqrt{5}$, $\theta = \frac{3\pi}{4}$ and $\phi = \arctan(-\frac{1}{2})$. Proving the density of $D_{1,2}$ is equivalent to proving that $\ln 5/\ln 2$, $\arctan(1/2)/\pi$ and 1 are rationally independent. In [65] the authors observe that the Four Exponential Conjecture, see [120], would imply that $D_{p,q}$ is dense in $\mathbb{C}$.

Conjecture 6.15 (Four Exponential Conjecture). *Let $\{\lambda_1, \lambda_2\}$ and $\{x_1, x_2\}$ be two pairs of rationally independent complex numbers. Then, one of the numbers $e^{\lambda_1 x_1}$, $e^{\lambda_1 x_2}$, $e^{\lambda_2 x_1}$, $e^{\lambda_2 x_2}$ is transcendental.*

6.7 Recognizability over $\mathbb{F}_q[X]$

Using the analogy existing between $\mathbb{Z}$ and the ring of polynomials over a finite field $\mathbb{F}_q$ of positive characteristic, one can easily define B-recognizable sets of polynomials [106]. In [121] characterization of these sets in a convenient logical structure analogous to Theorem 4.5 is given. A family of sets of polynomials recognizable in all polynomial bases is described in [106, 121]. We can again conjecture a Cobham-like theorem.

7 Decidability issues

So far we have seen that ultimately periodic sets have a very special status in the context of numeration systems (recall Proposition 2.6, Theorem 5.1 or Theorems 5.17 and 5.19). They can be described using a finite amount of data (two finite words for the preperiodic and the periodic parts). Let us settle down once more to the usual integer base numeration system. Let $X \subseteq \mathbb{N}$ be a k-recognizable set of integers given by a DFA accepting $\mathrm{rep}_k(X)$. Is there an algorithmic decision procedure which permits one to decide for any such set X, whether or not X is ultimately periodic? For an integer base, the problem was solved positively in [70]. The main ideas are the following ones. Given a DFA $\mathcal{A}$ accepting a k-recognizable set $X \subseteq \mathbb{N}$, the number of states of $\mathcal{A}$ gives an upper bound on the possible index and period for X. Consequently, there are finitely many candidates to check. For each such pair (i, p) of candidates, produce a DFA for all possible corresponding ultimately periodic sets and compare it with $\mathcal{A}$. Using non-deterministic finite automata, the same problem was solved in [5]. With the formalism of first order logic the problem becomes trivial. If a set $X \subseteq \mathbb{N}$ is k-recognizable, then using Theorem 4.5 it is definable by a formula $\varphi(x)$ in $\langle \mathbb{N}, +, V_k \rangle$ and X is ultimately periodic if and only if $(\exists p)(\exists N)(\forall x)(x \geqslant N \wedge (\varphi(x) \leftrightarrow \varphi(x + p)))$. Since we have a decidable theory, it is decidable whether this latter sentence is true [28, Prop. 8.2]. The problem can be extended to $\mathbb{Z}^d$ and was discussed in [91]. It is solved in polynomial time in [82]. In view of Theorem 5.1 the question is extended to any abstract numeration system. Let $\mathcal{S}$ be an abstract numeration system. Given a DFA accepting an $\mathcal{S}$-recognizable set $X \subseteq \mathbb{N}$. Decide whether or not X is ultimately periodic. Some special cases have been solved

positively in [31, 11]. Using Corollary 5.3, the same question can be asked in terms of morphisms. Given a morphism $\sigma : A^* \to A^*$ prolongable on a letter a and a coding $\tau : A \to B$, decide whether or not $\tau(\sigma^\infty(a))$ is ultimately periodic. It is the *HD0L (ultimate) periodicity problem*. The purely substitutive case was solved independently in [95] and [67]. Note that the general substitutive case is still open (one has to give a decision procedure for any abstract numeration system). Also see [86, 87] where decidability questions about almost-periodicity are considered. A word is *almost periodic* if factors occurring infinitely often have a bounded distance between occurrences (but some factors may occur only finitely often).

8 Acknowledgments

We would like to warmly thank Valérie Berthé, Alexis Bès, Véronique Bruyère, Christiane Frougny, Julien Leroy and Narad Rampersad for their useful comments after reading a preliminary version of this work.

References

[1] B. Adamczewski and J. Bell. Function fields in positive characteristic: expansions and Cobham's theorem. *J. Algebra*, 319:2337–2350, 2008.

[2] B. Adamczewski and J. Bell. An analogue of Cobham's theorem for fractals. *Trans. Amer. Math. Soc.*, 363:4421–4442, 2011.

[3] B. Alexeev. Minimal DFAs for testing divisibility. *J. Comput. Syst. Sci.*, 69:235–243, 2004.

[4] J.-P. Allouche. Sur la complexité des suites infinies. *Bull. Belg. Math. Soc.*, 1:133–143, 1994.

[5] J.-P. Allouche, N. Rampersad, and J. Shallit. Periodicity, repetitions, and orbits of an automatic sequence. *Theoret. Comput. Sci.*, 410(30-32):2795–2803, 2009.

[6] J.-P. Allouche and J. Shallit. The ring of k-regular sequences. *Theoret. Comput. Sci.*, 98(2):163–197, 1992.

[7] J.-P. Allouche and J. Shallit. The ring of k-regular sequences. II. *Theoret. Comput. Sci.*, 307(1):3–29, 2003. Words.

[8] J.-P. Allouche and J. O. Shallit. The ubiquitous Prouhet-Thue-Morse sequence. In C. Ding, T. Helleseth, and H. Niederreiter, editors, *Sequences and Their Applications, Proceedings of SETA '98*, pages 1–16. Springer-Verlag, 1999.

[9] J.-P. Allouche and J. O. Shallit. *Automatic Sequences, Theory, Applications, Generalizations*. Cambridge University Press, 2003.

[10] J. P. Bell. A generalization of Cobham's theorem for regular sequences. *Sém. Lothar. Combin.*, 54A:Art. B54Ap, 15 pp. (electronic), 2005/07.

[11] J. P. Bell, E. Charlier, A. S. Fraenkel, and M. Rigo. A decision problem for ultimately periodic sets in non-standard numeration systems. *Internat. J. Algebra Comput.*, 19:809–839, 2009.

[12] V. Berthé and M. Rigo, editors. *Combinatorics, Automata and Number Theory*, volume 135 of *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, 2009.

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

[14] A. Bertrand-Mathis. Comment écrire les nombres entiers dans une base que n'est pas entière. *Acta Math. Hung.*, 54:237–241, 1989.

[15] A. Bès. An extension of the Cobham-Semënov theorem. *J. Symbolic Logic*, 65(1):201–211, 2000.

[16] A. Bès. A survey of arithmetical definability. *Bull. Belg. Math. Soc. Simon Stevin*, (suppl.):1–54, 2001. A tribute to Maurice Boffa.

[17] B. Boigelot, L. Bronne, and S. Rassart. An improved reachability analysis method for strongly linear hybrid systems. In *Proc. 9th CAV*, volume 1254 of *Lecture Notes in Computer Science*, pages 167–177, 1997.

[18] B. Boigelot and J. Brusten. A generalization of Cobham's theorem to automata over real numbers. *Theoret. Comput. Sci.*, 410(18):1694–1703, 2009.

[19] B. Boigelot, J. Brusten, and V. Bruyère. On the sets of real numbers recognized by finite automata in multiple bases. In *Proc. 35th ICALP (Reykjavik)*, volume 5126 of *Lecture Notes in Computer Science*, pages 112–123. Springer-Verlag, 2008.

[20] B. Boigelot, J. Brusten, and J. Leroux. A generalization of Semenov's theorem to automata over real numbers. In R. A. Schmidt, editor, *Automated Deduction, 22nd International Conference, CADE 2009, McGill University, Montreal*, volume 5663 of *Lecture Notes in Computer Science*, pages 469–484, 2009.

[21] B. Boigelot, S. Jodogne, and P. Wolper. An effective decision procedure for linear arithmetic over the integers and reals. *ACM Trans. Comput. Log.*, 6(3):614–633, 2005.

[22] B. Boigelot, S. Rassart, and P. Wolper. On the expressiveness of real and integer arithmetic automata. In *Proc. 25th ICALP (Aalborg)*, volume 1443 of *Lecture Notes in Computer Science*, pages 152–163. Springer, 1998.

[23] S. Brlek. Enumeration of factors in the Thue-Morse word. *Discrete Appl. Math.*, 24:83–96, 1989.

[24] J. Brusten. *On the sets of real vectors recognized by finite automata in multiple bases*. PhD thesis, University of Liège, Fac. Sciences Appliquées, 2011.

[25] V. Bruyère. Automata and numeration systems. *Sém. Lothar. Combin.*, 35:Art. B35b, approx. 19 pp. (electronic), 1995.

[26] V. Bruyère. On Cobham's theorem. Thematic term on semigroups, algorithms, automata and languages, School on automata and languages, Coimbra, Portugal, 2001.

[27] V. Bruyère and G. Hansel. Bertrand numeration systems and recognizability. *Theoret. Comput. Sci.*, 181(1):17–43, 1997. Latin American Theoretical INformatics (Valparaíso, 1995).

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

[29] J. R. Büchi. Weak secord-order arithmetic and finite automata. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 6:66–92, 1960. Reprinted in S. Mac Lane and D. Siefkes, eds., *The Collected Works of J. Richard Büchi*, Springer-Verlag, 1990, pp. 398–424.

[30] A. Černý and J. Gruska. Modular trellises. In G. Rozenberg and A. Salomaa, editors, *The Book of L*, pages 45–61. Springer-Verlag, 1986.

[31] E. Charlier and M. Rigo. A decision problem for ultimately periodic sets in non-standard numeration systems. In *Proc. of the 33th International Symposium: MFCS (Torun)*, volume 5162 of *Lecture Notes in Computer Science*, pages 241–252, 2008.

[32] G. Christol. Ensembles presque périodiques k-reconnaissables. *Theoret. Comput. Sci.*, 9:141–145, 1979.

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

[34] A. Cobham. On the Hartmanis-Stearns problem for a class of tag machines. In *IEEE Conference Record of 1968 Ninth Annual Symposium on Switching and Automata Theory*, pages 51–60, 1968. Also appeared as IBM Research Technical Report RC-2178, August 23 1968.

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

[36] A. Cobham. Uniform tag sequences. *Math. Systems Theory*, 6:164–192, 1972.

[37] I. P. Cornfeld, S. V. Fomin, and Y. G. Sinaĭ. *Ergodic theory*, volume 245 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, New York, 1982.

[38] M. I. Cortez and F. Durand. Self-similar tiling systems, topological factors and stretching factors. *Discrete Comput. Geom.*, 40:622–640, 2008.

[39] A. de Luca and S. Varricchio. Some combinatorial properties of the Thue-Morse sequence and a problem in semigroups. *Theoret. Comput. Sci.*, 63(3):333–348, 1989.

[40] V. Diekert and D. Krieger. Some remarks about stabilizers. *Theoret. Comput. Sci.*, 410(30-32):2935–2946, 2009.

[41] A. W. M. Dress and F. von Haeseler. A semigroup approach to automaticity. *Ann. Comb.*, 7(2):171–190, 2003.

[42] F. Durand. Cobham's theorem for substitutions. *J. Eur. Math. Soc. (JEMS)*. to appear.

[43] F. Durand. A characterization of substitutive sequences using return words. *Discrete Math.*, 179(1-3):89–101, 1998.

[44] F. Durand. A generalization of Cobham's theorem. *Theory Comput. Syst.*, 31(2):169–185, 1998.

[45] F. Durand. Sur les ensembles d'entiers reconnaissables. *J. Théor. Nombres Bordeaux*, 10(1):65–84, 1998.

[46] F. Durand. Linearly recurrent subshifts have a finite number of non-periodic subshift factors. *Ergodic Theory Dynam. Systems*, 20(4):1061–1078, 2000.

[47] F. Durand. A theorem of Cobham for non primitive substitutions. *Acta Arith.*, 104:225–241, 2002.

[48] F. Durand. Cobham-Semenov theorem and $\mathbb{N}^d$-subshifts. *Theoret. Comput. Sci.*, 391(1-2):20–38, 2008.

[49] F. Durand and M. Rigo. Syndeticity and independent substitutions. *Adv. in Appl. Math.*, 42:1–22, 2009.

[50] H.-D. Ebbinghaus, J. Flum, and W. Thomas. *Mathematical logic*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, second edition, 1994.

[51] S. Eilenberg. *Automata, Languages, and Machines*, volume A. Academic Press, 1974.

[52] H. B. Enderton. *A mathematical introduction to logic*. Academic Press, New York, 1972.

[53] S. Fabre. Une généralisation du théorème de Cobham. *Acta Arith.*, 67:197–208, 1994.

[54] S. Fabre. Substitutions et β-systèmes de numération. *Theoret. Comput. Sci.*, 137(2):219–236, 1995.

[55] I. Fagnot. Sur les facteurs des mots automatiques. *Theoret. Comput. Sci.*, 172:67–89, 1997.

[56] N. P. Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.

[57] A. S. Fraenkel. Systems of numeration. *Amer. Math. Monthly*, 92:105–114, 1985.

[58] C. Frougny. Representations of numbers and finite automata. *Math. Systems Theory*, 25(1):37–60, 1992.

[59] C. Frougny. Non-standard number representation: computer arithmetic, beta-numeration and quasicrystals. In *Physics and theoretical computer science*, volume 7 of *NATO Secur. Sci. Ser. D Inf. Commun. Secur.*, pages 155–169. IOS, Amsterdam, 2007.

[60] F. R. Gantmacher. *The Theory of Matrices*. Chelsea, 1960.

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

[62] W. H. Gottschalk and G. A. Hedlund. *Topological Dynamics*, volume 36 of *AMS Colloquium Publications*. Amer. Math. Soc., 1955.

[63] G. Hansel. A propos d'un théorème de Cobham. In D. Perrin, editor, *Actes de la Fête des Mots*, pages 55–59. Greco de Programmation, CNRS, Rouen, 1982.

[64] G. Hansel. Systèmes de numération indépendants et syndéticité. *Theoret. Comput. Sci.*, 204:119–130, 1998.

[65] G. Hansel and T. Safer. Vers un théorème de Cobham pour les entiers de Gauss. *Bull. Belg. Math. Soc. Simon Stevin*, 10:723–735, 2003.

[66] G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers*. Oxford University Press, 5th edition, 1985.

[67] T. Harju and M. Linna. On the periodicity of morphisms on free monoids. *RAIRO Inform. Théor. App.*, 20:47–54, 1986.

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

[69] C. Holton and L. Q. Zamboni. Directed graphs and substitutions. *Theory Comput. Syst.*, 34(6):545–564, 2001.

[70] J. Honkala. A decision method for the recognizability of sets defined by number systems. *RAIRO Inform. Théor. App.*, 20:395–403, 1986.

[71] J. Honkala. On the simplification of infinite morphic words. *Theoret. Comput. Sci.*, 410(8-10):997–1000, 2009.

[72] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[73] I. Kátai and J. Szabó. Canonical number systems for complex integers. *Acta Sci. Math. (Szeged)*, 37(3-4):255–260, 1975.

[74] K. S. Kedlaya. The algebraic closure of the power series field in positive characteristic. *Proc. Amer. Math. Soc.*, 129(12):3461–3470, 2001.

[75] K. S. Kedlaya. Finite automata and algebraic extensions of function fields. *J. Théor. Nombres Bordeaux*, 18(2):379–420, 2006.

[76] B. P. Kitchens. *Symbolic dynamics*. Universitext. Springer-Verlag, Berlin, 1998. One-sided, two-sided and countable state Markov shifts.

[77] D. E. Knuth. *The art of computer programming. Vol. 2*. Addison-Wesley Publishing Co., Reading, Mass., second edition, 1981. Seminumerical algorithms, Addison-Wesley Series in Computer Science and Information Processing.

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

[79] P. Kůrka. *Topological and symbolic dynamics*, volume 11 of *Cours Spécialisés*. Société Mathématique de France, Paris, 2003.

[80] L. Latour. *Presburger Arithmetic: from Automata to Formulas*. PhD thesis, University of Liège, Fac. Sciences Appliquées, 2006.

[81] P. B. A. Lecomte and M. Rigo. Numeration systems on a regular language. *Theory Comput. Systems*, 34:27–44, 2001.

[82] J. Leroux. A polynomial time Presburger criterion and synthesis for number decision diagrams. In *IEEE Symposium on Logic in Computer Science (LICS 2005)*, IEEE Computer Society, pages 147–156, 2005.

[83] D. Lind and B. Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, 1995.

[84] M. Lothaire. *Combinatorics on Words*, volume 17 of *Encyclopedia of Mathematics and Its Applications*. Addison-Wesley, 1983.

[85] M. Lothaire. *Algebraic Combinatorics on Words*, volume 90 of *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, 2002.

[86] A. Maes. Morphisms and almost-periodicity. *Discrete Appl. Math.*, 86(2-3):233–248, 1998.

[87] A. Maes. More on morphisms and almost-periodicity. *Theoret. Comput. Sci.*, 231(2):205–215, 2000. Universal machines and computations (Metz, 1998).

[88] C. Michaux and R. Villemaire. Cobham's theorem seen through Büchi's theorem. In *Automata, languages and programming (Lund, 1993)*, volume 700 of *Lecture Notes in Computer Science*, pages 325–334. Springer, Berlin, 1993.

[89] C. Michaux and R. Villemaire. Presburger arithmetic and recognizability of sets of natural numbers by automata: New proofs of Cobham's and Semenov's theorems. *Ann. Pure Appl. Logic*, 77:251–277, 1996.

[90] M. Morse and G. A. Hedlund. Symbolic dynamics. *Amer. J. Math.*, 60:815–866, 1938.

[91] A. A. Muchnik. The definable criterion for definability in Presburger arithmetic and its applications. *Theoret. Comput. Sci.*, 290(3):1433–1444, 2003.

[92] M. Nivat. invited talk at ICALP, bologna, 1997.

[93] N. Ormes, C. Radin, and L. Sadun. A homeomorphism invariant for substitution tiling spaces. *Geom. Dedicata*, 90:153–182, 2002.

[94] J.-J. Pansiot. Complexité des facteurs des mots infinis engendrés par morphismes itérés. In *Automata, languages and programming (Antwerp, 1984)*, volume 172 of *Lecture Notes in Computer Science*, pages 380–389. Springer, Berlin, 1984.

[95] J.-J. Pansiot. Decidability of periodicity for infinite words. *RAIRO Inform. Théor. App.*, 20:43–46, 1986.

[96] W. Parry. On the β-expansions of real numbers. *Acta Math. Acad. Sci. Hung.*, 11:401–416, 1960.

[97] B. Pascal. *Œuvres complètes*. Seuil, 1963. The treatise *De numeris multiplicibus*, written with the other arithmetical treatises before 1654, was published by Guillaume Desprez in 1665.

[98] D. Perrin. Finite automata. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science, Volume B:Formal Models and Semantics*, pages 1–57. Elsevier — MIT Press, 1990.

[99] F. Point and V. Bruyère. On the Cobham-Semenov theorem. *Theory Comput. Syst.*, 30(2):197–220, 1997.

[100] M. Presburger. Über die volständigkeit eines gewissen systems der arithmetik ganzer zahlen, in welchem die addition als einzige operation hervortritt. *C. R. Premier congrès des Mathématiciens des pays slaves, Varsovie*, pages 92–101, 1929.

[101] M. Presburger. On the completeness of a certain system of arithmetic of whole numbers in which addition occurs as the only operation. *Hist. Philos. Logic*, 12:225–233, 1991.

[102] A. Quas and L. Zamboni. Periodicity and local complexity. *Theoret. Comput. Sci.*, 319:229–240, 2004.

[103] M. Queffélec. *Substitution dynamical systems—spectral analysis*, volume 1294 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1987.

[104] A. Rényi. Representations for real numbers and their ergodic properties. *Acta Math. Acad. Sci. Hung.*, 8:477–493, 1957.

[105] M. Rigo. Generalization of automatic sequences for numeration systems on a regular language. *Theoret. Comput. Sci.*, 244:271–281, 2000.

[106] M. Rigo. Syntactical and automatic properties of sets of polynomials over finite fields. *Finite Fields Appl.*, 14(1):258–276, 2008.

[107] M. Rigo and A. Maes. More on generalized automatic sequences. *J. Automata, Languages, and Combinatorics*, 7:351–376, 2002.

[108] M. Rigo and L. Waxweiler. A note on syndeticity, recognizable sets and Cobham's theorem. *Bull. European Assoc. Theor. Comput. Sci.*, 88:169–173, February 2006.

[109] J. Sakarovitch. *Éléments de théorie des automates*. Vuibert, 2003. English corrected edition: *Elements of Automata Theory*, Cambridge University Press, 2009.

[110] A. Salomaa and M. Soittola. *Automata-theoretic aspects of formal power series*. Springer-Verlag, New York, 1978. Texts and Monographs in Computer Science.

[111] O. Salon. Suites automatiques à multi-indices. In *Séminaire de Théorie des Nombres de Bordeaux*, pages 4.01–4.27, 1986-1987.

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

[113] A. L. Semenov. The Presburger nature of predicates that are regular in two number systems. *Sibirsk. Mat. Ž.*, 18(2):403–418, 479, 1977. In Russian. English translation in *Siberian J. Math.* **18** (1977), 289–300.

[114] E. Seneta. *Non-negative matrices and Markov chains*. Springer Series in Statistics. Springer, New York, 2006.

[115] J. O. Shallit. Numeration systems, linear recurrences, and regular sets. *Inform. Comput.*, 113:331–347, 1994.

[116] B. Solomyak. Dynamics of self-similar tilings. *Ergodic Theory Dynam. Systems*, 17(3):695–738, 1997.

[117] B. Tan and Z.-Y. Wen. Some properties of the Tribonacci sequence. *European J. Combin.*, 28(6):1703–1719, 2007.

[118] R. Villemaire. Joining k- and l-recognizable sets of natural numbers. In *STACS 92 (Cachan, 1992)*, volume 577 of *Lecture Notes in Computer Science*, pages 83–94. Springer, Berlin, 1992.

[119] R. Villemaire. The theory of $\langle \mathbf{N}, +, V_k, V_l \rangle$ is undecidable. *Theoret. Comput. Sci.*, 106(2):337–349, 1992.

[120] M. Waldschmidt. *Diophantine approximation on linear algebraic groups*, volume 326 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 2000. Transcendence properties of the exponential function in several variables.

[121] L. Waxweiler. *Caractère reconnaissable d'ensembles de polynômes à coefficients dans un corps fini*. PhD thesis, University of Liège, 2009. http://hdl.handle.net/2268/11381.

[122] P. Wolper and B. Boigelot. Verifying systems with infinite but regular state spaces. In *Computer aided verification (Vancouver, BC, 1998)*, volume 1427 of *Lecture Notes in Computer Science*, pages 88–97. Springer, Berlin, 1998.

[123] E. Zeckendorf. Représentation des nombres naturels par une somme de nombres de Fibonacci ou de nombres Lucas. *Bull. Soc. Roy. Liège*, 41:179–182, 1972.

Abstract. Let $k \geqslant 2$ be an integer. A set X of integers is k-recognizable if the language of k-ary representations of the elements in X is accepted by a finite automaton. The celebrated theorem of Cobham from 1969 states that if a set of integers is both k-recognizable and ℓ-recognizable, then it is a finite union of arithmetic progressions. We present several extensions of this result to non-standard numeration systems, we describe the relationships with substitutive and automatic words and list Cobham-type results in various contexts.

Index

abelianisation map, 15
abstract numeration system, 13
alphabet
    canonical, 4
arithmetic
    Presburger, 11
    progression, 3
array
    periodic, 27
    pseudo-periodic, 27
    ultimately periodic, 27
automaton
    real number, 24
    real vector, 24

basis
    numeration, 4
Bertrand numeration basis, 5
beta-polynomial
    canonical, 5
    extended, 5
block map, 25
bounded gap, 7

canonical
    beta-polynomial, 5
canonical alphabet, 4
characteristic
    polynomial, 5
    sequence, 8
Cobham theorem, 2
code
    sliding block, 25
coding, 8
complexity function, 9
constant length
    morphism, 8
cylinder, 25

decision problem
    HD0L (ultimate) periodicity problem, 29
dominating eigenvalue, 15
dynamical system, 25
    factor, 25
    isomorphism, 25
    minimal, 25
    periodic, 25

erasing substitution, 16
ergodic measure, 26
expansion
    α-expansion, 5
extended
    beta-polynomial, 5

factor map, 25
fixed point, 8
free
    variable, 11

genealogical order, 13
good substitution, 21
growing
    letter, 19
    substitution, 19
growth
    maximal, 19
growth type, 19
    less, 19

HD0L (ultimate) periodicity problem, 29

incidence matrix, 15
integer
    multiplicatively dependent, 3
    multiplicatively independent, 3
invariant
    measure, 25
irreducible
    matrix, 15
    substitution, 15

k-recognizable set, 2

k-kernel, 8
Kronecker's theorem, 3

letter-to-letter
  morphism, 8
linear numeration basis, 4

map
  abelianisation, 15
  Parikh, 15
matrix
  incidence, 15
  irreducible, 15
  primitive, 15
maximal growth, 19
  word, 19
measure
  ergodic, 26
  invariant, 25
  uniquely ergodic, 26
military order, 13
minimal
  dynamical system, 25
morphism
  constant length, 8
  fixed point, 8
  letter-to-letter, 8
  prolongable, 8
Morse-Hedlund theorem, 9
multiplicatively
  dependent, 3
  independent, 3

NDD, 2
Nivat's conjecture, 10
normalization, 4
number
  beta-number
    β-number, 5
  Parry, 5
  Perron, 15
  Pisot, 4
Number Decision Diagram, 2
numeration basis
  Bertrand, 5
numeration basis, 4
  linear, 4
numeration system
  abstract, 13

order
  genealogical, 13
  military, 13
  radix, 13

p-substitution, 27
Parikh map, 15
Parry number, 5
period, 8
periodic
  array, 27
  dynamical system, 25
  inside X, 27
  locally periodic, 27
  word, 8
Perron
  number, 15
  theorem, 15
Perron–Frobenius' theorem, 15
Pisot
  number, 4
  numeration system, 4
preperiod, 8
Presburger arithmetic, 11
primitive
  matrix, 15
  substitution, 15
  substitutive w.r.t. σ, 15
progression
  arithmetic, 3
prolongable morphism, 8
purely substitutive word, 8

radix order, 13
Real Number Automaton, 24
Real Vector Automaton, 24
recognizable set, 3
regular
  sequence, 23
representation
  $\mathcal{S}$-, 13
  normal, 4

U-, 4
return word, 21
RNA, 24
RVA, 24

sequence
automatic, 7
set
k-recognizable, 2
k-recognizable, 10, 24
locally periodic, 27
periodic inside X, 27
purely substitutive, 13
recognizable, 3
S-recognizable, 13
substitutive, 13
syndetic, 7
U-recognizable, 4
ultimately periodic, 3
shift transformation, 25
sliding block code, 25
sub-substitution, 21
subshift, 25
substitution, 8
erasing, 16
good, 21
growing, 19
irreducible, 15
ω_α-substitution, 23
primitive, 15
projection, 17
sub-substitution, 21
substitutive w.r.t. $\mathcal{S}$, 16
substitutive w.r.t. σ, 15
substitutive word, 8
successor, 11
syndetic set, 7

theorem
Kronecker, 3
Perron, 15
Perron–Frobenius, 15
Thue–Morse word, 9
Tribonacci word, 16

ultimately periodic, 3
array, 27
word, 8
uniquely ergodic
measure, 26

variable
free, 11

word
almost periodic, 29
automatic, 7
characteristic, 8
maximal growth, 19
ω_α-substitutive, 23
periodic, 8
primitive substitutive w.r.t. σ, 15
return, 21
substitutive, 8
purely, 8
substitutive w.r.t. $\mathcal{S}$, 16
substitutive w.r.t. σ, 15
Thue–Morse, 9
Tribonacci, 16
ultimately periodic, 8