Université de Liège
Faculté des sciences

Abstract numeration systems on a regular language and recognizability

Michel RIGO

Dissertation originale présentée en vue
de l'obtention du grade académique de
docteur en sciences
Mars 2001

Remerciements

Je voudrais en premier lieu exprimer toute ma gratitude à Pierre Lecomte qui m'a non seulement initié à la théorie des langages formels mais qui a également voulu me faire découvrir les joies de la recherche en mathématique. Ce travail n'aurait jamais pu aboutir sans son aide constante et les nombreuses et intéressantes discussions que nous avons partagées.

Ensuite, je voudrais remercier Véronique Bruyère et Jean-Paul Allouche. Leurs questions pertinentes, l'intérêt manifesté à l'égard de mes recherches et les petits courriers électroniques provenant souvent de l'autre bout du monde ont été autant de catalyseurs à ce travail. Je les remercie aussi d'avoir accepté, dans un premier temps, de faire partie de mon comité de thèse et ensuite, d'être membre de mon jury.

Je remercie messieurs les professeurs Pascal Laubin, Marc Roubens et Pierre Wolper de me faire l'honneur de prendre part à mon jury.

Enfin, je remercie pour leur soutien qui d'une façon ou d'une autre a contribué à l'élaboration de ce travail, Christelle Xhurdebise, Eliane et Georges Henderyckx, Marie Wuidar et Suzanne Wiame.

Contents

Introduction

On peut affirmer que c'est vers la fin des années soixante, avec les travaux de Cobham [**19**], qu'a débuté l'étude des relations liant les propriétés arithmétiques des ensembles d'entiers aux propriétés syntaxiques des langages formels constitués par leurs représentations dans un système de numération.

Depuis plus d'une dizaine d'années, de nombreuses avancées ont été réalisées dans ce domaine [**10**], [**14**], [**26**], [**30**], [**37**], [**63**], ... On y étudie non seulement les numérations classiques comme le système décimal ou encore le système binaire mais, d'une manière générale, on s'intéresse aux *systèmes de numération de position*. Dans un tel système, on décompose un entier positif n comme une combinaison linéaire

$$n = \sum_{i=0}^{p} d_i\, U_i,\ \ d_0, \ldots, d_p \in \mathbb{N}, d_p \neq 0$$

d'éléments d'une suite strictement croissante $(U_n)_{n\in\mathbb{N}}$ d'entiers commençant par $U_0 = 1$. On représente alors n par le *mot* $d_p \cdots d_0$. Afin que l'alphabet des chiffres d_i soit fini, on suppose le rapport $\frac{U_{n+1}}{U_n}$ borné. Le système de numération est alors complètement spécifié lorsqu'on dispose d'un algorithme calculant la décomposition ci-dessus pour chaque n. Le plus souvent, il s'agit de l'algorithme d'Euclide, encore appelé algorithme glouton. Les systèmes classiques, en base k, s'obtiennent en appliquant cet algorithme à la suite $(U_n)_{n\in\mathbb{N}} = (k^n)_{n\in\mathbb{N}}$.

Une des préoccupations centrales dans les travaux concernant les systèmes de numération est l'étude de la *reconnaissabilité* des ensembles d'entiers, c'est-à-dire de la régularité du langage formé des représentations de leurs éléments dans le système considéré. Pour la plupart, les résultats obtenus relèvent de thèmes généraux brièvement décrits ci-dessous.

Il y a tout d'abord, la caractérisation des parties reconnaissables dans un système fixé et la détermination de systèmes pour lesquels un ensemble d'entiers donné est reconnaissable.

En particulier, un effort important a été consenti pour décrire les systèmes pour lesquels $\mathbb{N}$ tout entier est reconnaissable (en partie sans doute en raison du fait que si l'ensemble des représentations des entiers est régulier, il existe des algorithmes très simples permettant de décider si un mot représente ou non un nombre). Le fait que $\mathbb{N}$ soit reconnaissable impose à la suite $(U_n)_{n\in\mathbb{N}}$ de vérifier une relation de récurrence linéaire à coefficients constants (*systèmes linéaires*) [**63**]. Des conditions suffisantes, s'exprimant en termes de propriétés des polynômes satisfaits par cette relation, ont été obtenues dans [**37**] et [**43**].

Les systèmes de numération linéaires pour lesquels les problèmes ci-dessus sont les mieux cernés, sont à ce jour ceux dont le polynôme caractéristique est le polynôme minimum d'un nombre de Pisot [**14**], [**33**]. Pour ces systèmes, on dispose par exemple de plusieurs descriptions pratiques des parties reconnaissables.

La reconnaissabilité d'un ensemble d'entiers est fortement liée au système de numération utilisé pour le représenter. En effet, le théorème de Cobham stipule que sous des hypothèses très larges, les seules parties simultanément reconnaissables dans deux systèmes donnés sont les unions finies de progressions arithmétiques [**19**]. Ce théorème a été largement étendu à des numérations non standards, citons [**15**], [**25**], [**26**], [**35**], [**48**], [**54**], [**66**] et [**67**].

Une autre problématique envisagée dans la littérature est l'étude de la stabilité du caractère reconnaissable vis-à-vis des opérations arithmétiques. Il est bien connu que la multiplication ne peut garantir la conservation de ce caractère. Par contre, pour les numérations classiques en base entière, l'addition préserve trivialement la reconnaissabilité et on peut espérer que cela soit encore vrai pour de larges classes de systèmes. En tout cas, les systèmes basés sur un nombre de Pisot forment la classe la plus riche actuellement connue de systèmes ayant cette propriété [**14**], [**30**].

Les systèmes étudiés jusqu'à présent ont sans exception la propriété d'être monotones : l'application associant à un nombre sa représentation est strictement croissante pour l'ordre naturel sur $\mathbb{N}$ et l'ordre lexicographique induit par l'ordre des chiffres. Dès lors, indépendamment de l'algorithme spécifique utilisé pour obtenir la représentation d'un nombre, la numération est en fait entièrement caractérisée par le langage formé de toutes les représentations et cette propriété de monotonie. En effet, celle-ci impose que chaque entier soit représenté par le mot dont il est le numéro dans l'énumération du langage fournie par l'ordre lexicographique.

Fort de cette constatation, nous avons introduit dans [**39**] la notion de *système de numération abstrait* $S = (L, \Sigma, <)$ où L est un langage dénombrable sur l'alphabet Σ et où $<$ est un ordre total sur ce dernier. Dans un tel système, la représentation d'un entier est l'élément de L dont il est le rang dans l'énumération de L par ordre lexicographique croissant. Nous faisons également l'hypothèse que L est régulier. En effet, il nous semble d'une part que la reconnaissabilité de $\mathbb{N}$ occupe une place importante dans les résultats mentionnés sommairement ci-dessus et, d'autre part, vu le rôle particulier des progressions arithmétiques mis en lumière par le théorème de Cobham, nous souhaitons que celles-ci soient reconnaissables et c'est le cas si et seulement si $\mathbb{N}$ l'est [**39**].

Le présent travail est consacré à l'étude de la reconnaissabilité dans le cadre des systèmes de numération abstraits ainsi qu'à l'extension de ceux-ci à la représentation des nombres réels.

Nous obtenons plusieurs caractérisations originales de la reconnaissabilité (notamment en termes de séries rationnelles en variables non commutatives et en termes de mots morphiques). Nous montrons également que des ensembles qui ne sont reconnaissables dans aucun système de position à base entière le sont toujours dans des systèmes abstraits appropriés (par exemple les ensembles de puissances d'entiers) et que par contre, les nombres premiers ne sont jamais reconnaissables. Cependant, vu la très grande généralité des systèmes introduits, on ne peut espérer obtenir des propriétés liées à la reconnaissabilité aussi riches que dans les systèmes de position comme, par exemple, ceux associés à un nombre de Pisot. En effet, l'addition ne conserve généralement pas la reconnaissabilité. A cet égard, il est frappant de constater combien la complexité du langage sur lequel est construit un système de numération abstrait, joue un rôle fondamental. Pour les langages polynomiaux et les exponentiels à complémentaire polynomial, la multiplication par des constantes — et *a fortiori* l'addition — ne préserve pas la reconnaissabilité. Seuls les langages exponentiels à complémentaire exponentiel sont donc susceptibles de donner lieu à une addition régulière [**58**].

Il nous semble qu'un système de numération doit non seulement permettre de représenter les entiers mais également les réels. Nous inspirant de la manière dont ceux-ci sont décrits dans les systèmes en base entière, nous proposons pour une large classe de systèmes abstraits, une extension aux nombres réels dont nous donnons les premières propriétés.

Passons à présent en revue l'organisation des chapitres du présent travail.

Dans le premier chapitre, nous installons les notations et propriétés de base relatives aux systèmes de numération de position. Nous définissons les systèmes de numération abstraits et fournissons pour ceux-ci des algorithmes permettant l'un de calculer le nombre représenté par un mot et l'autre de déterminer le mot représentant un nombre. Ce dernier est une généralisation de l'algorithme glouton dans laquelle les fonctions de complexité[1] des dérivés du langage utilisé se substituent à la suite $(U_n)_{n\in\mathbb{N}}$ des systèmes de position.

[1]La fonction de complexité d'un langage $L \subset \Sigma^*$ compte le nombre de mots de longueur n dans L, $\mathbf{u}_n(L) : n \mapsto \#(L \cap \Sigma^n)$.

Le deuxième chapitre est principalement consacré aux ensembles reconnaissables dans un système abstrait. On y montre que les progressions arithmétiques sont toujours reconnaissables et que les translations par des constantes préservent la reconnaissabilité. Outre leur intérêt propre, ces résultats sont utilisés par la suite. Par après, nous construisons pour chaque exponentielle polynôme

$$f(n) = \sum_{i=1}^{k} P_i(n)\, \alpha_i^n,\ \ \alpha_i \in \mathbb{N},\ \ P_i \in \mathbb{Q}[x],\ \ P_i(\mathbb{N}) \subset \mathbb{N}$$

un système abstrait $(L, \Sigma, <)$ dans lequel $f(\mathbb{N})$ est reconnaissable. La méthode consiste essentiellement à faire en sorte que $f(\mathbb{N})$ soit représenté par l'ensemble des plus petits mots de chaque longueur de L, ce à quoi on parvient en choisissant L pour avoir $\mathbf{u}_n(L) = f(n+1) - f(n)$. Pour terminer ce chapitre, on établit que les ensembles reconnaissables peuvent être caractérisés par des séries formelles rationnelles en variables non commutatives.

Il est clair qu'une condition nécessaire à la stabilité de la régularité vis-à-vis de l'addition est la conservation du caractère reconnaissable des ensembles d'entiers après multiplication par une constante. Dans le troisième chapitre, après une étude détaillée de la fonction de complexité, nous démontrons que si un système de numération est construit sur un langage polynomial de complexité $\Theta(n^l)$, alors un multiplicateur assure la stabilité du caractère reconnaissable seulement s'il est puissance $(l+1)$-ième d'un naturel, prouvant dès lors que l'addition n'est pas régulière pour les langages polynomiaux. Cette condition ne peut être suffisante car pour le système construit sur le langage a^*b^*, la stabilité est conservée si et seulement si le multiplicateur est un carré parfait *impair*. Pour un langage exponentiel à complémentaire polynomial, on prouve que les multiplicateurs pouvant assurer la stabilité du caractère reconnaissable ne sont jamais puissance du cardinal de l'alphabet. Enfin, nous obtenons des conditions suffisantes pour que l'addition soit reconnaissable dans un système basé sur un langage exponentiel à complémentaire exponentiel. Les systèmes ainsi obtenus se ramènent par transduction à des systèmes de position basés sur un nombre de Pisot.

Cobham a montré que les ensembles reconnaissables pour une base entière k sont exactement ceux dont la suite caractéristique est *k-automatique*, i.e. l'image par un morphisme lettre-à-lettre du point fixe d'un morphisme uniforme de longueur k [**20**]. Dans le quatrième chapitre de ce travail, nous généralisons les suites automatiques aux systèmes abstraits (ceci étend notamment le cas traité dans [**62**]). On parle alors de *suites S-automatiques* où S est une numération abstraite basée sur un langage régulier. Nous étudions non seulement les propriétés intrinsèques de ces suites mais aussi leurs relations avec les parties reconnaissables. Il s'avère ainsi qu'un ensemble est reconnaissable dans un système abstrait S si et seulement si sa suite caractéristique est S-automatique. On montre que la complexité d'une suite S-automatique (au sens nombre de facteurs de longueur n apparaissant dans une suite infinie) est $O(n^2)$ et on en déduit que l'ensemble des nombres premiers n'est reconnaissable dans aucun système généralisé. Pour terminer ce chapitre, nous montrons que toute suite automatique sur l'alphabet $\{0, 1\}$ est morphique et réciproquement, que tout prédicat morphique est la suite caractéristique d'un ensemble d'entiers reconnaissable pour un système abstrait que l'on peut construire effectivement.

Dans le cinquième chapitre, encore en relation avec la fonction de complexité d'un langage L, on montre que si $\mathbf{u}_n(L)$ est borné par une constante, alors les parties reconnaissables pour les numérations construites sur L sont exactement les unions finies de progressions arithmétiques. Nous avons dès lors conservation du caractère reconnaissable pour l'addition mais aussi pour le changement d'ordre sur l'alphabet. Concernant cette dernière opération, nous montrons qu'en général, le caractère reconnaissable d'une partie de $\mathbb{N}$ est une propriété qui dépend de l'ordre total placé sur l'alphabet induisant l'ordre lexicographique sur le langage de la numération.

Pour terminer, dans le sixième chapitre, nous nous intéressons à la représentation des nombres réels. De manière classique, la partie entière du nombre réel à représenter est donnée par un mot fini alors que la partie fractionnaire est quant à elle représentée par un mot infini (pouvant éventuellement se terminer par des zéros). On peut voir ce mot infini comme limite d'une suite de mots finis où chaque mot est préfixe du suivant, la suite des approximations numériques fournies par ces mots convergeant vers le réel donné. Nous montrons comment étendre les systèmes de numération abstraits pour permettre la représentation des nombres réels à l'aide de mots infinis, d'une manière qui généralise de façon naturelle la description des réels en base entière. Nous mettons alors en place des hypothèses assurant la convergence numérique des approximations données par les éléments w_n d'une suite de mots de L convergeant vers un mot infini w et pour qu'il y ait assez de telles suites pour représenter un intervalle réel. Il est intéressant de noter que nous sommes en présence de deux types de convergence : numérique d'une part et en mots d'autre part. Signalons encore que la propriété de monotonie de l'application associant à un entier sa représentation s'étend à l'application donnant la représentation des réels. De plus, on montre que cette dernière est uniformément continue. En particulier, les représentations abstraites obtenues en utilisant le langage des représentations normalisées de l'unique système de Bertrand basé sur un nombre de Pisot $\theta > 1$ coïncident avec les mots infinis obtenus comme θ-développements de nombres réels. Dans un développement en base entière, un entier possède une voire deux représentations. Ici, pour nos systèmes abstraits, de multiples cas de figure sont déterminés : nombre de représentations fini, dénombrable et même non dénombrable. Une fois encore, la clé du raisonnement réside dans l'étude de la fonction de complexité.

Pour conclure, signalons que de nombreux résultats donnés dans ce travail ont été induits par l'expérimentation par des moyens informatiques. Dès lors, on trouve en annexe des procédures implémentant les algorithmes de représentation et diverses constructions en relation avec la théorie des automates.

Introduction

It can be said that, in the late sixties, with Cobham's works [**19**], started the study of the relationships linking the arithmetical properties of sets of integers to the syntactic properties of formal languages constituted by their representations in a numeration system.

For more than ten years a lot of progress has been made in this field [**10**], [**14**], [**26**], [**30**], [**37**], [**63**], ... Are studied not only classical numeration systems, such as the decimal or the binary system, but generally speaking, *positional numeration systems.* In such a system, a positive integer n is decomposed as a linear combination

$$n = \sum_{i=0}^{p} d_i\, U_i,\ \ d_0, \ldots, d_p \in \mathbb{N}, d_p \neq 0$$

of elements of a strictly increasing sequence $(U_n)_{n\in\mathbb{N}}$ of integers beginning with $U_0 = 1$. The integer n is then represented by the *word* $d_p \cdots d_0$. In order that the alphabet of digits d_i's is finite, the ratio $\frac{U_{n+1}}{U_n}$ is bounded. The numeration system is then completely specified when we have at hand an algorithm computing the above decomposition for every n. Euclid's algorithm, also known as greedy algorithm, is mostly used. We get the classic system with an integer base k, using this algorithm with the sequence $(U_n)_{n\in\mathbb{N}} = (k^n)_{n\in\mathbb{N}}$.

One of the main concerns in the works about the numeration systems is studying the *recognizability* of sets of integers, i.e., the regularity of the language made up of the representations of their elements in a given system. For most of them, the obtained results are related to the general themes shortly described below.

There is first the characterization of the recognizable parts in a fixed system and the determination of the systems for which a given set of integers is recognizable.

A great effort has been made especially to describe the systems in which the whole set $\mathbb{N}$ is recognizable (partly due to the fact that, if the set of representations of all integers is regular, there are very simple algorithms making it possible to decide if a word stands or does not stand for a number). The fact that $\mathbb{N}$ is recognizable demands that the sequence $(U_n)_{n\in\mathbb{N}}$ verifies a linear recurrent relation with constant coefficients (*linear systems*) [**63**]. Sufficient conditions in terms of the properties of polynomials satisfied by this relation have been obtained in [**37**] and [**43**].

The linear systems, for which the above problems have been figured out best, are nowadays those the characteristic polynomial of which is the minimal polynomial of a Pisot number [**14**], [**33**]. For these systems, we have several practical descriptions of the recognizable parts at our disposal.

The recognizability of a set of integers is strongly linked to the numeration system that is used to represent it. Actually Cobham's theorem stipulates that, under very wide hypotheses, the only simultaneous recognizable parts in two given systems are the finite unions of arithmetic progressions [**19**]. This theorem has been extended widely to non standard numeration systems, see [**15**], [**25**], [**26**], [**35**], [**48**], [**54**], [**66**] and [**67**].

Another problem which has been taken into account in the literature is the study of the stability of the recognizability under arithmetic operations. It is well known that the multiplication does not preserve recognizability. On the other hand, for classical numeration systems with an integer base, the addition preserves the recognizability trivially and, hopefully, it is also true for wider classes of systems. Anyway, the systems based on a Pisot number make up the currently known largest class of systems having this property.

All the up to now studied systems have got the property of being monotonous: the application mapping a number onto its representation is strictly increasing for the natural order on $\mathbb{N}$ and the lexicographical order induced by the ordering of the digits. Thus, independently from the specific algorithm used to compute the representation of a number, the numeration system is actually wholly characterized by the language made up of all the representations and this property of monotony. Indeed, this latter property demands that each integer is represented by the word whose number it is in the enumeration of the language described by the lexicographic ordering.

In the knowledge of this statement we have introduced in [**39**], the notion of *abstract numeration system* $S = (L, \Sigma, <)$ in which L is a countable language over the alphabet Σ and where $<$ is a total order on the latter. In such a system, the representation of an integer is the element of L of which it is the rank in the enumeration of L by increasing lexicographic ordering. We also assume that L is regular. Indeed, we think that the recognizability of $\mathbb{N}$ is very important in the above briefly given results and that, in view of the special role of arithmetic progressions rendered by Cobham's theorem, we wish they are recognizable and it is, if and only if $\mathbb{N}$ is recognizable [**39**].

This work is dedicated to the study of recognizability in the frame of abstract numeration systems as well as to their expansion for the representation of real numbers.

We achieve several original characterizations of the recognizability (namely in terms of rational series in non-commuting variables and of morphic words). We show too that some sets that are not recognizable in any positional system with an integer base, are always recognizable in suitable abstract systems (for example sets of powers of integers) and that, on the other hand, the set of prime numbers is never recognizable. However, as there is a very large generality in the introduced systems, we cannot hope to obtain properties related to recognizability that are as rich as in positional systems such as the ones related to a Pisot number. Indeed, addition does not generally preserve recognizability. In that respect, it is worth noticing the essential role of the complexity function of the language upon which an abstract system is built. In case of polynomial languages and exponential languages with polynomial complement, multiplication by constants — and *a fortiori* addition — does not preserve recognizability. Only exponential languages with exponential complement are thus likely to give a regular addition [**58**]. We think that a numeration system has to make it possible to represent not only integers but also real numbers. Using the way those are described in systems with an integer base, we suggest for a large class of abstract systems, an extension to the representation of real numbers for which we give the first properties.

Let us now go over the organization of the chapters in this work.

In the first chapter we settle the notations and basic properties connected with positional numeration systems. We define the abstract numeration systems and give, for those, algorithms on the one hand to compute the number represented by a word and on the other hand to determine the word representing a number. The latter is a generalization of the greedy algorithm where the complexity functions[2] of the derivatives of the used language substitute for the sequence $(U_n)_{n\in\mathbb{N}}$ of positional systems.

The second chapter is mainly about the recognizable sets in an abstract system. In this chapter, we show that arithmetic progressions are always recognizable and that translations by constants preserve recognizability. Beside their own importance, these results will be used later. Moreover, we build for each exponential polynomial function

$$f(n)=\sum_{i=1}^{k} P_i(n)\,\alpha_i^n,\ \ \alpha_i\in\mathbb{N},\ \ P_i\in\mathbb{Q}[x],\ \ P_i(\mathbb{N})\subset\mathbb{N}$$

[2]The complexity function of a language $L\subset\Sigma^*$ counts the number of words of length n in L, $\mathbf{u}_n(L):n\mapsto\#(L\cap\Sigma^n)$.

an abstract system $(L, \Sigma, <)$ in which $f(\mathbb{N})$ is recognizable. The method consists mainly in having $f(\mathbb{N})$ represented by the set of the smallest words of each length in L, which we achieve when we choose L so that we have $\mathbf{u}_n(L) = f(n+1) - f(n)$. We end up this chapter in proving that recognizable sets can be characterized by rational formal series in non-commuting variables.

It is obvious that a necessary condition for the stability of regularity after addition is the preservation of the recognizability of sets of integers under multiplication by a constant. In the third chapter we give a detailed study of the complexity function and show that if a numeration system is built on a polynomial language of $\Theta(n^l)$ complexity then a multiplicator ensures the stability of the recognizability only if it is an $(l+1)^{th}$ power of a natural number, proving thus that addition is not regular for polynomial languages. This condition cannot be sufficient as, for the system built upon the language a^*b^*, the stability is preserved if and only if the multiplicator is an *odd* perfect square. For an exponential language with a polynomial complement, we prove that the multiplicators that can ensure the stability of the recognizability are never a power of the cardinality of the alphabet. Eventually we get sufficient conditions so that the addition is recognizable in a system built on an exponential language with an exponential complement. The thus obtained systems lead back by transduction to positional systems related to a Pisot number.

Cobham has shown that recognizable sets in an integer base k are exactly the ones for which the characteristic sequence is k-*automatic*, i.e., the image by a letter-to-letter morphism of the fixed point of a uniform morphism of length k [**20**]. In the fourth chapter of this work, we generalize automatic sequences to abstract numeration systems (this widens the problem studied in [**62**]). So we speak of S-*automatic sequences* where S is an abstract system built on a regular language. We study not only the intrinsic properties of these sequences but also their relationships with recognizable parts. It turns out that a set is recognizable in an abstract system S if and only if its characteristic sequence is S-automatic. We show that the complexity of an S-automatic sequence (in the meaning the number of factors of length n in an infinite sequence) is $O(n^2)$ and from that it may be deduced that the set of prime numbers is not recognizable in any generalized numeration system. We end this chapter with proving that any automatic sequence over the alphabet $\{0,1\}$ is morphic and conversely, that any morphic predicate is the characteristic sequence of a set of integers which is recognizable in an abstract system that can be effectively built.

In the fifth chapter, still in connection with the complexity function of a language L, we show that if $\mathbf{u}_n(L)$ is bounded by a constant, then the recognizable parts for the numeration system built upon L are exactly the finite unions of arithmetic progressions. We have thus conservation of the recognizability under addition but also for the change of order on the alphabet. As far as the latter is concerned, we show that the recognizability of a subset of $\mathbb{N}$ is a property that generally depends on the total ordering of the alphabet inducing the lexicographic ordering of the language of the numeration.

In the sixth and last chapter, we go into the representation of real numbers. In a usual way, the integer part of the real number to be represented is given by a finite word whereas the fractional part is represented by an infinite word (that may be ending with zeroes). We can consider this infinite word as the limit of a sequence of finite words where each word is the prefix of the next one and the sequence of numerical approximations given by these words converges to the given real number. We show how the abstract numeration systems can be extended to enable the representation of real numbers by infinite words, in a way that generalizes naturally the description of real numbers in an integer base. Then we set up hypotheses ensuring the numerical convergence of the approximations given by the elements w_n of a sequence of words in L converging to an infinite word w and so that there are enough such sequences to represent an interval of real numbers. It is interesting to notice that we face up two types of convergence: on the one hand a numerical convergence and on the other hand a word convergence. Let us underline the fact that the property of monotony of the application mapping an integer onto its representation is extended to the application giving the representation of real numbers. Moreover, we show that the latter is uniformly continuous. In particular, the abstract representations obtained through the language of the normalized representations of the unique Bertrand system based on a Pisot number $\theta > 1$ coincide with the infinite words obtained as θ-developments of real numbers. In case of an expansion in an integer base, a real number has got one or even two representation(s). Here different cases are determined for our abstract systems: a finite, countable or uncountable number of representations. Once again, the key argument is found in the study of the complexity function.

Eventually let us underline the fact that many results given in this work have been induced by computer experiments. Thus, procedures implementing representation algorithms and various constructions related to automata theory can be found in an appendix.

CHAPTER I

Basics

In this introductory chapter, the first two sections are devoted to recall well-known definitions and results in formal languages theory. We also recall some important results on positional numeration systems. Next, we introduce abstract numeration systems on a regular language and set the notations used throughout this work. In particular, we study the structure of a lexicographically ordered regular language L and the link with the functions counting the number of words accepted from the different states of a deterministic finite automaton $\mathcal{A}$ recognizing L (i.e., the complexity function of the languages accepted from the different states of $\mathcal{A}$). The material of the sections related to abstract numeration systems was introduced by P. Lecomte and the author in [**39**].

1. Words, languages and automata

We recall some definitions. More details can be found in [**28**], [**53**] or [**70**]. An *alphabet* is a finite set of *symbols* or *letters*. We will denote alphabets by capital case Greek letters. A *word* over Σ is the concatenation of a finite number of letters of Σ. The set of words over Σ is denoted Σ^*. It is the free monoid generated by Σ with respect to the concatenation of symbols as monoid operation. The identity, or *empty word*, is denoted ε. The length of a word w is denoted $|w|$. A *language* over Σ is a subset of Σ^*. The concatenation of two languages M and L is the set

$$ML = \{vw \mid v \in M,\ w \in L\}.$$

The *Kleene's star* L^* of a language L is the language of all concatenations of an arbitrary number of elements in L,

$$L^* = \bigcup_{n \geq 0} L^n,\ L^0 = \{\varepsilon\}.$$

We now introduce the class of regular languages.

Definition I.1.1. The set R_Σ of *regular expressions* over the alphabet Σ is the language over $\Sigma \cup \{0, e, +, (,), *\}$ defined inductively by

(1) $\forall \sigma \in \Sigma : \sigma \in R_\Sigma$
(2) $0, e \in R_\Sigma$
(3) if $x, y \in R_\Sigma$ then (xy), $(x + y)$ and x^* belong to R_Σ.

We define an application $\alpha : R_\Sigma \to 2^{\Sigma^*}$ in the following way,

(1) $\forall \sigma \in \Sigma : \alpha(\sigma) = \{\sigma\}$
(2) $\alpha(0) = \emptyset$ and $\alpha(e) = \{\varepsilon\}$
(3) if $x, y \in R_\Sigma$ then
- $\alpha((xy)) = \alpha(x)\,\alpha(y)$,
- $\alpha((x+y)) = \alpha(x) \cup \alpha(y)$,
- $\alpha(x^*) = (\alpha(x))^*$.

Definition I.1.2. If x is a regular expression over Σ then $\alpha(x) \in 2^{\Sigma^*}$ is said to be a *regular language*. In the literature, one also finds the terms *rational expression* and *rational language*. In this work, we will not distinguish a regular expression from the corresponding regular language. Two regular expressions x and y are said to be *equivalent* if $\alpha(x) = \alpha(y)$.

We set forth the following result characterizing the class of regular languages.

Theorem I.1.3. *The family of regular languages over an alphabet Σ is the smallest subset of 2^{Σ^*} containing $\emptyset$, $\{\sigma\}$ for all $\sigma \in \Sigma$ and closed under union, concatenation and Kleene's star.*

Regular languages can be described in two ways. Regular expressions can be viewed as generators of regular languages and finite automata as acceptors.

Definition I.1.4. A *deterministic finite automaton* or DFA is a quintuple $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ where Q is the finite set of states, Σ is the input alphabet, $\delta : Q \times \Sigma \to Q$ is the state transition function, $s \in Q$ is the starting state and $F \subseteq Q$ is the set of final states. The function δ can be extended to $Q \times \Sigma^*$ by $\delta(p, \varepsilon) = p$ and $\delta(p, \sigma w) = \delta(\delta(p, \sigma), w)$, for any $p \in Q$, $\sigma \in \Sigma$ and $w \in \Sigma^*$. We will often write $p.w = q$ instead of $\delta(p, w) = q$.

Definition I.1.5. A *non-deterministic finite automaton* (NDFA) is a quintuple $\mathcal{A} = (Q, \Sigma, E, I, F)$ where Q, Σ and F are defined exactly the same way as for a DFA, $I \subseteq Q$ is the set of starting states and the finite set $E \subseteq Q \times \Sigma^* \times Q$ is the transition relation.

Automata will be as usual depicted by directed graphs. The states will be indicated by circles. The starting states will be marked by an arrow pointing toward the state. The final states will be denoted by double circles. In a DFA, if $\delta(p, \sigma) = q$, $p, q \in Q$, $\sigma \in \Sigma$, then there is an edge labeled by σ from p to q. In an NDFA, we have the same kind of construction when $(p, w, q) \in E$, $w \in \Sigma^*$.

A word $w \in \Sigma^*$ is *accepted* by a finite automaton (deterministic or not) if there exists a path labeled by w starting in an initial state and ending in a final state. A language L is *accepted* by a finite automaton $\mathcal{A}$ if L is the set of words accepted by $\mathcal{A}$.

THEOREM I.1.6 (Kleene's theorem). *A language is regular if and only if it is accepted by a finite automaton.*

THEOREM I.1.7. *The class of regular languages is closed under the operations of union, intersection, complementation, concatenation, morphism, inverse morphism and mirror image.*

DEFINITION I.1.8. Among the DFA accepting a given regular language $L \subset \Sigma^*$, one distinguishes the *minimal automaton*

$$\mathcal{A}_L = (Q_L, \Sigma, \delta_L, s_L, F_L)$$

of L. The states of $\mathcal{A}_L$ are the *derivatives*

$$w^{-1}.L = \{v \in \Sigma^* \mid wv \in L\}.$$

One sometimes finds the notation $D_w L$ in the literature [**70**]. The starting state s_L is $L = \varepsilon^{-1}.L$, the set of final states is

$$F_L = \{p \in Q_L \mid \varepsilon \in p\} = \{w^{-1}.L \mid w \in L\}$$

and the transition function is defined by $\delta_L(p, \sigma) = \sigma^{-1}.p$. More about minimal automaton can be found in Section III.5 of [**28**].

Let $w \in \Sigma^*$ and $p \in Q_L$. If $s_L.w = \delta_L(s_L, w) = p$ then $w^{-1}.L$ is the language L_p of words accepted by $\mathcal{A}_L$ from the state p (i.e., the set of words which are label of a path starting in p and ending in F_L),

$$L_p = \{w \in \Sigma^* \mid p.w \in F_L\}.$$

REMARK I.1.9. Notice that the set L_p of words accepted from the state p can be defined for any finite automaton (deterministic or not).

In this work, we will also encounter other kinds of automata. Let us define them.

DEFINITION I.1.10. A *deterministic finite automaton with output* or simply a DFAO is a sextuple $\mathcal{A} = (Q, \Sigma, \delta, s, \Delta, \tau)$ where Q, Σ, δ and s are defined exactly the same way as for a DFA, Δ is the output alphabet and $\tau : Q \to \Delta$ is the output function. The output produced by $\mathcal{A}$ on a word $w \in \Sigma^*$ is

$$\tau(\delta(s, w)).$$

DEFINITION I.1.11. A 2-*tape automaton* or *transducer* is an NDFA $\mathcal{A} = (Q, \Sigma^* \times \Gamma^*, E, I, F)$ with edges labeled by elements of the monoid $\Sigma^* \times \Gamma^*$. If the set of edges E is finite (and thus Q is finite), $\mathcal{A}$ is said to be *finite*. A 2-tape automaton is said to be *letter-to-letter* if the edges are labeled by elements of $\Sigma \times \Gamma$, that is, by couple of letters. A relation $R \subset \Sigma^* \times \Gamma^*$ is said to be *computable by a finite* 2-*tape automaton* if there exists a finite 2-tape automaton such that the set of labels starting in I and ending in F is R. A function $f : \Sigma^* \to \Gamma^*$ is *computable by a finite* 2-*tape automaton* if its graph $\hat{f} \subset \Sigma^* \times \Gamma^*$ is computable by a finite 2-tape automaton. These definitions can be extended to infinite sequences of letters.

Example I.1.12. The automaton depicted in Figure I.1 is a letter-to-letter transducer computing the relation $(a^2, b^2)^* \cup (a, c)(a^2, c^2)^*$. In this example, $\Sigma = \{a\}$ and $\Gamma = \{b, c\}$.

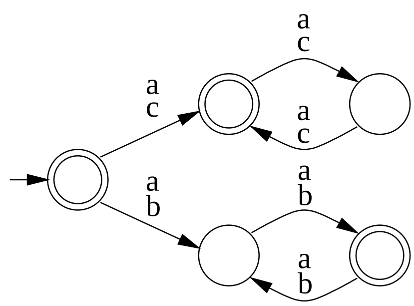

Figure I.1. A transducer computing the relation $(a^2, b^2)^* \cup (a, c)(a^2, c^2)^*$.

2. Positional numeration systems

With a sequence of integers $1 = U_0 < U_1 < U_2 < \cdots$, any natural number x can be written as a linear combination of the U_n's with natural coefficients,

$$x = d_n\, U_n + \cdots + d_0\, U_0,\ \ d_0, \ldots, d_n \in \mathbb{N}, d_n \neq 0.$$

The fact that $U_0 = 1$ makes sure that we have at least the trivial decomposition $x = x\, U_0$. Among the possible decompositions of an integer x, a special decomposition is the one obtained by the greedy algorithm in the following way. Let n be such that $U_n \leq x < U_{n+1}$ and $x_0 = x$. For $i = 0, \ldots, n$, the quotient and the remainder of the Euclidian division of x_i by U_{n-i} are respectively d_{n-i} and x_{i+1}. Moreover, if the ratio $\frac{U_{n+1}}{U_n}$ is bounded then the coefficients d_i's computed through the greedy algorithm belong to the finite set $\{0, \ldots, a\}$ where a is the greatest integer less than $\sup \frac{U_{n+1}}{U_n}$.

For a given sequence $(U_n)_{n\in\mathbb{N}}$, we can therefore associate any natural number x with a canonical word $d_n \cdots d_0$ representing x. So, we have the following definition of a numeration system.

Definition I.2.1. A *positional numeration system* is a strictly increasing sequence $U = (U_n)_{n\in\mathbb{N}}$ of integers such that $U_0 = 1$ and that the ratio $\frac{U_{n+1}}{U_n}$ is bounded. A numeration system U is said to be *linear* if the sequence $(U_n)_{n\in\mathbb{N}}$ satisfies a linear recurrence relation with integer coefficients,

$$U_n = c_{k-1}\, U_{n-1} + \cdots + c_0\, U_{n-k},\ \ c_0, \ldots, c_{k-1} \in \mathbb{Z}, c_0 \neq 0.$$

Let $\Delta \subset \mathbb{Z}$ be an alphabet and $w = w_n \cdots w_0 \in \Delta^*$. We denote by

$$\pi_U(w) = \sum_{i=0}^{n} w_i\, U_i$$

the *numerical value* of w. If $x \in \mathbb{N}$ is such that $\pi_U(w) = x$ then we say that w is a *U-representation* of x. Observe that an integer x can have more than one U-representation. Among these representations, we distinguish the *normalized U-representation* of x computed through the use of the greedy algorithm (see [**29**]) and denoted by $\rho_U(x)$. By convention, the normalized representation of 0 is the empty word and we assume that a normalized representation has no leading zeroes. We introduce a partial function [**30**],

$$\nu_{\Delta,U} : \Delta^* \to \rho_U(\mathbb{N})$$

called *normalization*, in the following way. If $w \in \Delta^*$ is such that $\pi_U(w) \in \mathbb{N}$ then $\nu_{\Delta,U}(w) = \rho_U(\pi_U(w))$.

EXAMPLE I.2.2. With $U_n = 2^n$ for all $n \in \mathbb{N}$, we obtain the binary system. The normalized representation in base 2 of 83 is

$$\rho_2(83) = 1010011$$

because

$$83 = \mathbf{1}.64 + \mathbf{0}.32 + \mathbf{1}.16 + \mathbf{0}.8 + \mathbf{0}.4 + \mathbf{1}.2 + \mathbf{1}.1.$$

Another representation of 83 is "210003" because

$$83 = \mathbf{2}.32 + \mathbf{1}.16 + \mathbf{0}.8 + \mathbf{0}.4 + \mathbf{0}.2 + \mathbf{3}.1.$$

In particular, $\nu_{\{0,1,2,3\},U}(210003) = 1010011$.

More generally, if $U_n = k^n$ for all $n \in \mathbb{N}$, $k \geq 2$, then we have the base k system or k-ary system. In this case, we use the notation ρ_k and π_k instead of ρ_U and π_U to specify the base k.

Observe that (normalized) representations of integers in base k are words over $\{0, \ldots, k-1\}$. So a set X of numbers gives the language $\rho_k(X)$ of normalized representations of the integers belonging to X and we say that this set is *k-recognizable* if $\rho_k(X)$ is regular. The notion of recognizability related to a numeration system will take a central position throughout this work.

We now give an example of a recognizable set of integers. The set of even integers is 2-recognizable since

$$\rho_2(2\mathbb{N}) = 1\{0,1\}^*0 \cup \{\varepsilon\}.$$

The representations in base 2 of even integers are accepted by the automaton in Figure I.2 (the sink has not been represented).

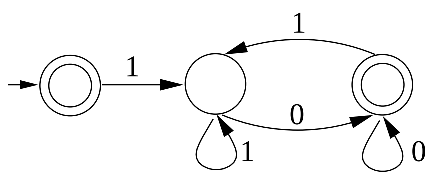

FIGURE I.2. Automaton accepting even integers in base 2.

For a survey on the k-recognizable sets and their properties, see [**13**] or [**15**]. In particular, it is worth noting that arithmetic progressions are k-recognizable for any k. In other words, any divisibility criterion can be viewed as a syntaxical property of the normalized representations in a specified base.

In this introductory chapter, we have to recall one of the most famous theorems for numeration systems with integer bases — Cobham's theorem — which says that the only subsets of $\mathbb{N}$ which are simultaneously k-recognizable and l-recognizable, k and l being multiplicatively independent, are ultimately periodic (see [**19**]). A series of recent papers is devoted to the generalization of this result for non-standard positional systems [**15**], [**25**], [**26**], [**35**], [**54**], [**48**], [**66**], [**67**].

Example I.2.3. Another example of a positional number system is the *Fibonacci system* defined by the linear recurrent sequence

$$\begin{cases} U_0 &= 1, \\ U_1 &= 2, \\ U_{n+2} &= U_{n+1} + U_n, \ n \geq 0. \end{cases}$$

The first terms of the sequence are

$$1, 2, 3, 5, 8, 13, 21, 34, 55, \ldots$$

The normalized U-representation of 17 is $\rho_U(17) = 100101$ since

$$17 = \mathbf{1}.13 + \mathbf{0}.8 + \mathbf{0}.5 + \mathbf{1}.3 + \mathbf{0}.2 + \mathbf{1}.1.$$

As a consequence of the greedy algorithm, the set of all normalized representations is the set of words over $\{0, 1\}$ which do not contain two consecutive ones.

Definition I.2.4. As for systems with an integer base k, if U is an arbitrary positional number system, we say that a set $X \subset \mathbb{N}$ is *U-recognizable* if $\rho_U(X)$ is accepted by a finite automaton.

A series of recent papers is devoted to U-recognizable subsets of $\mathbb{N}$ [**14**], [**33**], [**35**], [**63**]. In particular, the U-recognizability of $\mathbb{N}$ has been extensively studied. Indeed, the case when $\mathbb{N}$ is recognizable is of special interest because then it is very easy to decide whether or not a given word represents an integer. Under quite general assumptions, it is shown in [**63**] that for $\mathbb{N}$ to be U-recognizable, it is necessary that the numeration system $U = (U_n)_{n \in \mathbb{N}}$ satisfies a linear recurrence equation. In [**37**], a sufficient condition is given in terms of the polynomials of the recurrence that $(U_n)_{n \in \mathbb{N}}$ satisfies (it generalizes the case studied in [**43**]).

A wide class of positional numeration systems which has very interesting properties is the class of systems defined by a linear recurrent sequence such that its characteristic polynomial is the minimal polynomial of a Pisot number. (A *Pisot number*, or *Pisot-Vijayaraghavan* *number*, is an algebraic integer greater than 1 with all its Galois conjugates having modulus less than one. Recall that an *algebraic integer* is a root of a monic polynomial with integral coefficients.) It is clear that systems with an integer base and the Fibonacci system belong to this class. For instance, the characteristic polynomial of the recurrent sequence defining the Fibonacci system is

$$P(X) = X^2 - X - 1.$$

It is the minimal polynomial of $\frac{1+\sqrt{5}}{2}$ and the other root of this polynomial is $\frac{1-\sqrt{5}}{2}$ of modulus less than one.

If U is a linear numeration system the characteristic polynomial of which is the minimal polynomial of a Pisot number then $\mathbb{N}$ is U-recognizable (a proof of this result can be found in [**14**] and [**33**]). For these systems, we have a nice characterization of the U-recognizable subsets of $\mathbb{N}$. Let us introduce the logical structure $\langle \mathbb{N}, +, V_U \rangle$ where $V_U(x) = y$ means that y is the smallest U_n appearing in the normalized U-representation of x with a non-null digit. We say that $X \subseteq \mathbb{N}$ is *U-definable* if there exists a first order formula of $\langle \mathbb{N}, +, V_U \rangle$ defining X. If U is the k-ary system, we use the term *k-definable* set and the notation V_k.

THEOREM I.2.5. [**14**] *Let X be a subset of $\mathbb{N}$ and $U = (U_n)_{n\in\mathbb{N}}$ be a linear numeration system such that the characteristic polynomial of U is the minimal polynomial of a Pisot number. The set X is U-recognizable if and only if X is U-definable.*

Another property related to the normalization is the following.

PROPOSITION I.2.6. [**33**] *Let $U = (U_n)_{n\in\mathbb{N}}$ be a linear numeration system such that its characteristic polynomial is the minimal polynomial of a Pisot number and $\Delta \subset \mathbb{Z}$ be a finite alphabet. The set*

$$\{(v, w) \in \Delta^* \times \rho_U(\mathbb{N}) \mid \nu_{\Delta,U}(v) = w\}^1$$

is recognizable by a finite letter-to-letter automaton, i.e., the normalization function is computable by a finite 2-tape automaton.

For more about U-recognizability and linear numeration systems whose characteristic polynomial is the minimal polynomial of a Pisot number, see [**14**].

3. k-automatic sequences

A class of infinite sequences is related to the representation of integers and to the recognizable sets of integers. The construction of the sequences belonging to this class is based on the representation of nonnegative integers in an integer base k. A given integer n is represented

[1] To obtain couples of words of the same length, instead of (v, w), we have to consider the couples $(0^{n-|v|}v, 0^{n-|w|}w)$ where $n = \max\{|v|, |w|\}$. Indeed, if $|v|$ differs from $|w|$ then (v, w) cannot be read by a 2-tape letter-to-letter automaton.

in the base k using the greedy algorithm and we obtain a word $\rho_k(n)$ over the alphabet $\{0,\ldots,k-1\}$. Next, $\rho_k(n)$ is given to a deterministic finite automaton with output to obtain the n^{th} term of a sequence, which is said to be a *k-automatic sequence*. In [**20**], the sequences are said to be *uniform tag sequences* (they are fixed points of a uniform morphism of length k).

DEFINITION I.3.1. Let $k>1$. The *k-kernel* of a sequence $(x_n)_{n\in\mathbb{N}}$ is the set of sub-sequences

$$\{n\mapsto x_{k^d\,n+r}\,|\,d\geq 0,\ 0\leq r<k^d\}.$$

THEOREM I.3.2. [**62**] *A sequence is k-automatic if and only if its k-kernel is finite.*

Recall that the *characteristic sequence* of a subset X of $\mathbb{N}$ is the sequence $(\chi_n^X)_{n\in\mathbb{N}}$ defined by

$$\begin{cases}\chi_n^X=1, & \text{if } n\in X;\\ \chi_n^X=0, & \text{otherwise.}\end{cases}$$

We have the following characterization of the k-recognizable subsets of integers.

PROPOSITION I.3.3. [**20**, Theorem 3] *A subset $X\subset\mathbb{N}$ is k-recognizable if and only if its characteristic sequence $(\chi_n^X)_{n\in\mathbb{N}}$ is k-automatic.*

The next proposition will be useful.

PROPOSITION I.3.4. [**20**, Theorem 4] *The set of k-automatic sequences is closed under finitary sequential transduction, i.e., closed under function computed by a finite 2-tape letter-to-letter automaton.*

4. Numeration systems on a regular language

In this section, we generalize positional numeration systems U for which $\rho_U(\mathbb{N})$ is regular. In particular, we generalize linear numeration systems the characteristic polynomial of which is the minimal polynomial of a Pisot number.

A total order on Σ induces a lexicographic ordering of Σ^*.

DEFINITION I.4.1. Let $(\Sigma,<)$ be a totally ordered alphabet and v,w be in Σ^n, $n\geq 1$. We say that v is *lexicographically less* than w, and we write $v<w$, if there exists $u,v',w'\in\Sigma^*$ and $\sigma,\gamma\in\Sigma$ such that $v=u\sigma v'$, $w=u\gamma w'$ and $\sigma<\gamma$. If v and w are words of different length then v is *lexicographically less* than w if $|v|<|w|$. In the literature, this ordering is sometimes called “radix order”, “genealogical order” or “military order”.

The use of the greedy algorithm in positional numeration systems has a trivial consequence.

PROPOSITION I.4.2. *Let U be a positional numeration system. Let v and w be normalized U-representations of two integers $x = \pi_U(v)$ and $y = \pi_U(w)$. Then*

$$x < y \Leftrightarrow v < w$$

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

Instead of using a sequence of integers and an algorithm to compute representations, we can find another way to represent integers. Observe that describing an arbitrary infinite language over a totally ordered alphabet according to the lexicographic ordering gives a one-to-one correspondence between $\mathbb{N}$ and this language. Doing this, the application rep that maps a natural number onto its representation is strictly increasing if one endows the set of words, $\mathrm{rep}(\mathbb{N})$, with the lexicographic order (this is an assumption in [**63**] and it is also relevant with Proposition I.4.2). Among the possibly recognizable sets of integers, $\mathbb{N}$ is of special interest. For instance, if $\mathbb{N}$ is recognizable, then one can easily check whether a word over the alphabet of the digits represents an integer or not. Taking this into account, we choose to describe an infinite regular language according to the lexicographic ordering. So we have the following definition.

DEFINITION I.4.3. An *abstract numeration system* or *numeration system on a regular language* is a triple $S = (L, \Sigma, <)$ where L is an infinite regular language over the totally ordered alphabet $(\Sigma, <)$. Enumerating the elements of L lexicographically with respect to $<$ leads to a one-to-one map rep_S from $\mathbb{N}$ onto L. To any non-negative integer n, it assigns the $(n+1)^{th}$ word of L, its *S-representation*, while the inverse map val_S sends any word belonging to L onto its *numerical value.*

Having generalized numeration systems at our disposal, it is natural to be interested in the corresponding recognizable subsets of $\mathbb{N}$.

DEFINITION I.4.4. Let $S = (L, \Sigma, <)$ be a numeration system. A subset $X \subset \mathbb{N}$ is said to be *S-recognizable* if $\mathrm{rep}_S(X)$ is a regular subset of L.

EXAMPLE I.4.5. Let $\Sigma = \{a, b\}$, $a < b$, and $L = a^*b^*$. We consider the numeration system $S = (L, \Sigma, <)$. Table I.1 gives the first words of the ordered regular language a^*b^*. For instance,

$$\mathrm{rep}_S(4) = ab \text{ and } \mathrm{val}_S(aaa) = 6.$$

In a positional system, each digit has its own weight. Observe that this is generally not the case for an abstract numeration system.

Numeration systems on a regular language generalize linear numeration systems whose characteristic polynomial is the minimal polynomial of a Pisot number. If U is a positional system having this "Pisot property" then $\rho_U(\mathbb{N})$ is regular and we have Proposition I.4.2. So, we can describe the regular language $\rho_U(\mathbb{N})$ according to the lexicographic ordering induced by the natural ordering of the digits and we obtain an

$\mathbb{N}$	a^*b^*
0	ε
1	a
2	b
3	aa
4	ab
5	bb
6	aaa
7	aab
8	abb
$\vdots$	$\vdots$

TABLE I.1. S-representations with $S = (a^*b^*, \{a, b\}, a < b)$.

equivalent abstract numeration system. Moreover, we can completely forget the sequence $(U_n)_{n\in\mathbb{N}}$ and the algorithm of representation. They are just extra data devised to compute the function rep : $n \mapsto \rho_U(n)$ in some "practical" manner.

EXAMPLE I.4.6. Consider the Fibonacci system introduced in Example I.2.3. The set of all normalized representations is

$$L = \{\varepsilon\} \cup 1\{0, 01\}^*.$$

So the positional Fibonacci system and the abstract numeration system $S = (L, \{0, 1\}, 0 < 1)$ are equivalent. In these systems, 17 is represented by "100101" because $\pi_U(100101) = 17$ but also because 100101 is the 18^{th} word of the lexicographically ordered language L.

EXAMPLE I.4.7. Let $k \geq 2$ and Σ be the totally ordered alphabet of digits $\{1 < \cdots < k\}$. The abstract numeration system

$$S = (\Sigma^*, \Sigma, <)$$

is said to be the *k-adic numeration system* and the $(n+1)^{th}$ word of Σ^* is said to be the *k-adic representation* of n (see page 303 of [**38**]). It is worth noting that if $w = w_m \cdots w_0 \in \Sigma^*$ then

$$\mathrm{val}_S(w) = \sum_{i=0}^{m} w_i\, k^i.$$

Observe that this system can be viewed as a positional numeration system defined by the sequence $(U_n)_{n\in\mathbb{N}} = (k^n)_{n\in\mathbb{N}}$. In this case, the k-adic representation of an integer is not computed by the greedy algorithm (otherwise, we would obtain the classical representation in base k over the alphabet $\{0, \ldots, k-1\}$) but it can be easily shown that each integer x has a unique decomposition

$$x = d_n\, k^n + \cdots + d_0\, k^0$$

with the coefficients d_i's belonging to $\{1,\dots,k\}$. With our notations, it is clear that the functions val_S and π_k coincide. Therefore, the restriction of π_k to the language $\{1,\dots,k\}^*$ is an increasing one-to-one and onto mapping.

EXAMPLE I.4.8. In a positional system U, allowing leading zeroes in the representation w of an integer has no consequence on its numerical value, $\pi_U(w) = \pi_U(0^n w)$, $\forall n \in \mathbb{N}$. This is not the case for abstract numeration systems because they are dependent on the lexicographic ordering. This is shown in Table I.2 where we have chosen the binary system. One can see that if we allow leading zeroes, the words 0, 00

$\mathbb{N}$	$1\{0,1\}^* \cup \{\varepsilon\}$	$\{0,1\}^*$
0	ε	ε
1	1	0
2	10	1
3	11	00
4	100	01
5	101	10
6	110	11
7	111	000
$\vdots$	$\vdots$	$\vdots$

TABLE I.2. Allowing leading zeroes changes the abstract system of numeration.

and 000 have different numerical values with respect to an abstract system. Observe that the numeration system on the language $\{0,1\}^*$ corresponds, up to a homomorphism, to the *dyadic numeration system.* Consider the homomorphism $h : \{0,1\} \to \{1,2\}$ defined by $h(i) = i+1$, $i = 0, 1$. The application of h is represented in Table I.3.

$w \in \{0,1\}^*$	$h(w)$	$\pi_2(h(w))$
ε	ε	0
0	1	$\mathbf{1}.2^0 = 1$
1	2	$\mathbf{2}.2^0 = 2$
00	11	$\mathbf{1}.2^1 + \mathbf{1}.2^0 = 3$
01	12	$\mathbf{1}.2^1 + \mathbf{2}.2^0 = 4$
10	21	$\mathbf{2}.2^1 + \mathbf{1}.2^0 = 5$
11	22	$\mathbf{2}.2^1 + \mathbf{2}.2^0 = 6$
000	111	$\mathbf{1}.2^2 + \mathbf{1}.2^1 + \mathbf{1}.2^0 = 7$
$\vdots$	$\vdots$	$\vdots$

TABLE I.3. Dyadic numeration system.

5. First properties of abstract numeration systems

This section is mainly devoted to algorithms for the computation of val_S and rep_S. To obtain these algorithms, we have to study the structure of an ordered regular language.

Let $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ be a DFA accepting a language $L \subset \Sigma^*$. We define some sequences related to the states of $\mathcal{A}$. If p is a state of $\mathcal{A}$ then $\mathbf{u}_n(p)$ is the number of words of length n accepted from the state p, i.e.,

$$\mathbf{u}_n(p) = \#(L_p \cap \Sigma^n)$$

and $\mathbf{v}_n(p)$ is the number of words of length at most n accepted from p, that is,

$$\mathbf{v}_n(p) = \#(L_p \cap \Sigma^{\leq n}) = \sum_{i=0}^{n} \mathbf{u}_i(p).$$

Observe that the function which maps n onto $\mathbf{u}_n(p)$ is the *complexity function* of the language L_p (i.e., the function which maps $n \in \mathbb{N}$ onto the number of words of length n belonging to L_p). In the following of this work, we mix up, whenever possible, the function $n \mapsto \mathbf{u}_n(p)$ and the sequence $(\mathbf{u}_n(p))_{n\in\mathbb{N}}$ provided that it does not lead to any confusion. A well-known fact in automata theory is that the sequences $(\mathbf{u}_n(p))_{n\in\mathbb{N}}$, and thus the sequences $(\mathbf{v}_n(p))_{n\in\mathbb{N}}$, satisfy a linear recurrent equation. (A proof of this result is based on the fact that the series $f_{L_p}(X) = \sum_{n\geq 0} \mathbf{u}_n(p)\, X^n$ is $\mathbb{N}$-rational, see [**7**]. Another proof can be be given in terms of graph theory by counting the number of paths of length n starting in a state p and ending in a final state: the characteristic polynomial of the incidence matrix of the automaton is satisfied by the sequence $(\mathbf{u}_n(p))_{n\in\mathbb{N}}$.)

If we are only interested in the complexity function of a language L, we simply write $\mathbf{u}_n(L)$ or $\mathbf{u}_n$ provided that it does not lead to any confusion. Similarly, we allow the notation $\mathbf{v}_n$ or $\mathbf{v}_n(L)$ to denote the number of words of length not exceeding n in L.

When we consider a numeration system $S = (L, \Sigma, <)$ or more specifically a DFA $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ accepting L, each state $p \in Q$ for which $L_p = \{w \in \Sigma^* \,|\, p.w \in F\}$ is infinite leads to a numeration system $S_p = (L_p, \Sigma, <)$. The functions rep_{S_p} and val_{S_p} are simply denoted by rep_p and val_p if the context is clear. If L_p is finite, the functions rep_p and val_p are defined as in the infinite case but the domain of the former is restricted to $\{0, \ldots, \#L_p - 1\}$.

Example I.5.1. Consider the DFA depicted in Figure I.3. We have $L_s = a^*bab^*$, $L_t = ab^*$, $L_p = b^*$ and $L_q = \emptyset$ and Table I.4 gives the first representations in these numeration systems.

Observe that the numerical value of a word w is equal to the number of words in the language which are lexicographically less than w.

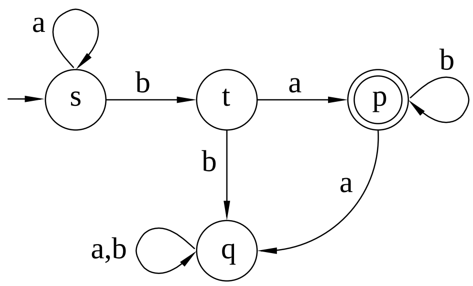

FIGURE I.3. A DFA and the languages L_p's.

$\mathbb{N}$	L_s	L_t	L_p
0	ba	a	ε
1	aba	ab	b
2	bab	abb	bb
3	$aaba$	$abbb$	bbb
$\vdots$	$\vdots$	$\vdots$	$\vdots$

TABLE I.4. The languages accepted from the different states.

Taking this into account, the next Lemma binds the numerical value of a word and the different sequences $(\mathbf{u}_n(p))_{n\in\mathbb{N}}$ and $(\mathbf{v}_n(p))_{n\in\mathbb{N}}$.

LEMMA I.5.2. *Let $S = (L, \Sigma, <)$ be an abstract numeration system and $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ be a DFA accepting L. If zw belongs to L_p, $z, w \in \Sigma^+ = \Sigma^* \setminus \{\varepsilon\}$, then* $\mathrm{val}_p(zw)$ *is equal to*

$$\boxed{\mathrm{val}_{p.z}(w) + \mathbf{v}_{|zw|-1}(p) - \mathbf{v}_{|w|-1}(p.z) + \sum_{\substack{z'<z \\ |z'|=|z|}} \mathbf{u}_{|w|}(p.z').} \tag{1}$$

Proof. We have to compute the number of words belonging to L_p and lexicographically less than zw. There are three kinds of such words. The first consists of words of length less than zw and counts $\mathbf{v}_{|zw|-1}(p)$ elements. The second consists of words of length $|zw|$ admitting the prefix z. Since a word $z'w'$ belongs to L_p if and only if w' belongs to $L_{p.z'}$, we see that there are $\mathrm{val}_{p.z}(w) - \mathbf{v}_{|w|-1}(p.z)$ such words. It is clear that there are

$$\begin{aligned} &\#\{x \in L_p : x = z'w', |z'| = |z|, |w'| = |w| \text{ and } z' < z\} \\ &= \sum_{\substack{z'<z \\ |z'|=|z|}} \mathbf{u}_{|w|}(p.z') \end{aligned}$$

words of the third kind.

□

REMARK I.5.3. Taking z to be a letter in Lemma I.5.2 leads easily to an effective algorithm to compute val_S. If σw belongs to L_p, $\sigma \in \Sigma$, $w \in \Sigma^+$, then

$$\mathrm{val}_p(\sigma w) = \mathrm{val}_{p.\sigma}(w) + \mathbf{v}_{|w|}(p) - \mathbf{v}_{|w|-1}(p.\sigma) + \sum_{\sigma'<\sigma} \mathbf{u}_{|w|}(p.\sigma'). \tag{2}$$

If σ is a letter belonging to L_p, it is obvious that

$$\mathrm{val}_p(\sigma) = \mathbf{u}_0(p) + \sum_{\sigma'<\sigma} \mathbf{u}_0(p.\sigma'). \tag{3}$$

So, applying several times (2), we obtain the following decomposition of the numerical value of a word $w = w_l \cdots w_1$ of length l belonging to L_p

(4)

$$\begin{aligned}\mathrm{val}_p(w) \quad = \quad & \mathbf{v}_{l-1}(p) + \sum_{\sigma<w_l} \mathbf{u}_{l-1}(p.\sigma) + \cdots + \sum_{\sigma<w_2} \mathbf{u}_1(p.w_l \cdots w_3\sigma) \\ & -\mathbf{v}_0(p.w_l \cdots w_2) + \mathrm{val}_{p.w_l\cdots w_2}(w_1).\end{aligned}$$

Thus, using (3) and the definition of $\mathbf{v}_{l-1}(p)$, we have

$$\boxed{\mathrm{val}_p(w) = \sum_{q\in Q} \sum_{i=0}^{|w|-1} \beta_{q,i}(p,w)\, \mathbf{u}_i(q)} \tag{5}$$

for some $\beta_{q,i}(p,w) \in \mathbb{N}$ and the following proposition is obvious.

PROPOSITION I.5.4. *Let $S = (L, \Sigma, <)$ be a numeration system and $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ be a DFA accepting L. If $p, q \in Q$, $w \in L_p$ and $i < |w|$, then the coefficients of (5) are such that*

$$\beta_{q,i}(p,w) < \#\Sigma + \delta_{p,q}$$

where $\delta_{p,q}$ is the Kronecker's symbol.

EXAMPLE I.5.5. Consider the numeration system of Example I.4.5. The minimal automaton of a^*b^* is represented in Figure I.4. It is easy

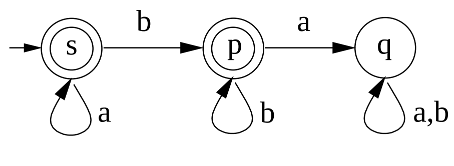

FIGURE I.4. The minimal automaton of a^*b^*.

to see that, $L_s = a^*b^*$, $L_p = b^*$, $L_q = \emptyset$ and $\forall n \in \mathbb{N}$,

$$\left\{\begin{array}{lcl}\mathbf{u}_n(s) & = & n+1\\ \mathbf{u}_n(p) & = & 1\\ \mathbf{u}_n(q) & = & 0\end{array}\right. \quad \text{and} \quad \left\{\begin{array}{lcl}\mathbf{v}_n(s) & = & \frac{1}{2}(n+1)(n+2)\\ \mathbf{v}_n(p) & = & n+1\\ \mathbf{v}_n(q) & = & 0.\end{array}\right.$$

Using (4), one obtains, for instance,

$$\begin{aligned}\mathrm{val}_S(ab^2) &= \mathbf{v}_2(s)+\mathbf{u}_1(s.aa)-\mathbf{v}_0(s.ab)+\mathrm{val}_{s.ab}(b)\\ &= 6+2-1+1=8.\\ \mathrm{val}_S(a^2b^3) &= \mathbf{v}_4(s)+\mathbf{u}_2(s.a^3)+\mathbf{u}_1(s.a^2ba)-\mathbf{v}_0(s.a^2b^2)+\mathrm{val}_{s.a^2b^2}(b)\\ &= 15+3+0-1+1=18.\end{aligned}$$

To conclude this first chapter, we give an algorithm to compute the S-representation of an integer x. It is some generalization of the greedy algorithm involving the complexity functions of the languages accepted from the different states of the minimal automaton of L.

Let $S=(L,\Sigma,<)$ be a numeration system with $\Sigma=\{\sigma_1<\cdots<\sigma_k\}$ and $\mathcal{A}_L=(Q_L,\Sigma,\delta_L,s_L,F_L)$ be the minimal automaton of L. Recall that for the minimal automaton of L, the set $\sigma^{-1}.L$ of words that concatenated with σ belong to L is $\delta_L(s_L,\sigma)=s_L.\sigma$.

It is clear that

$$|\mathrm{rep}_S(x)|=\inf_n\{n\mid x<\mathbf{v}_n(s_L)\}. \tag{6}$$

This observation is enlightened by the following example.

Example I.5.6. Consider the languages $L=a^*b^*$ and

$$M=\{w\in a^*b^*:|w|\equiv 1 \mod 2\}.$$

Notice that $\mathbf{u}_{2n}(M)=0$, $\forall n\in\mathbb{N}$. Table I.5 gives the positions of the first $\mathbf{v}_n$ for L and M.

		L		M
0		ε	$\mathbf{v}_0(M)$	a
1	$\mathbf{v}_0(L)$	a		b
2		b	$\mathbf{v}_1(M)=\mathbf{v}_2(M)$	aaa
3	$\mathbf{v}_1(L)$	aa		aab
4		ab		abb
5		bb		bbb
6	$\mathbf{v}_2(L)$	aaa	$\mathbf{v}_3(M)=\mathbf{v}_4(M)$	$aaaaa$
$\vdots$		$\vdots$		$\vdots$

Table I.5. Relation between $(\mathbf{v}_n)_{n\in\mathbb{N}}$ and a language.

Set $|\mathrm{rep}_S(x)|=n$; then $x-\mathbf{v}_{n-1}$ is the number of words of length n belonging to L and less than $\mathrm{rep}_S(x)$.

Table I.6 sketches the structure of an ordered language L for words of length n. To determine the first letter of the representation, we have to compute the number N_j^n of words of length n belonging to L and

$\mathbf{v}_{n-1}(s_L)$	σ_1	$\sigma_1 \quad \cdots$
	$\vdots$	$\vdots$
$\mathbf{v}_{n-1}(s_L) + \mathbf{u}_{n-2}(s_L.\sigma_1\sigma_1)$		$\sigma_2 \quad \cdots$
		$\vdots$
	$\vdots$	
	σ_1	$\sigma_k \quad \cdots$
$\mathbf{v}_{n-1}(s_L) + \mathbf{u}_{n-1}(s_L.\sigma_1)$	σ_2	$\sigma_1 \quad \cdots$
	$\vdots$	$\vdots$
	σ_2	$\sigma_k \quad \cdots$
	$\vdots$	
$\mathbf{v}_{n-1}(s_L) + \sum_{i=1}^{k-1} \mathbf{u}_{n-1}(s_L.\sigma_i)$	σ_k	$\sigma_1 \quad \cdots$
	$\vdots$	$\vdots$
$\mathbf{v}_{n-1}(s_L) + \sum_{i=1}^{k-1} \mathbf{u}_{n-1}(s_L.\sigma_i) + \mathbf{u}_{n-2}(s_L.\sigma_k\sigma_1)$		$\sigma_2 \quad \cdots$
		$\vdots$
	$\vdots$	
$\mathbf{v}_n(s_L) - 1$	σ_k	$\sigma_k \quad \cdots$

TABLE I.6. Structure of an ordered language for words of length n.

beginning with σ_1 or σ_2 or $\cdots$ or σ_j $(j \leq k)$

$$N_j^n = \sum_{i=1}^{j} \mathbf{u}_{n-1}(\sigma_i^{-1}.L).$$

If $N_{j-1}^n \leq x - \mathbf{v}_{n-1} < N_j^n$ then the first letter of $\mathrm{rep}_S(x)$ is σ_j. We proceed in the same way to determine the other letters of the representation. Hence, the following algorithm computes the S-representation w of a given integer x. See Appendix C for a practical implementation.

ALGORITHM I.5.7. $\mathrm{rep}_S : \mathbb{N} \to L : x \mapsto \mathrm{rep}_S(x) = w$.

Let n be such that $\mathbf{v}_{n-1} \leq x < \mathbf{v}_n$,
$p \leftarrow s_L$
$r \leftarrow x - \mathbf{v}_{n-1}$
$w \leftarrow \varepsilon$
for i **ranging from** 1 **to** n **do**
    $j \leftarrow 1$
    while $r \geq \mathbf{u}_{n-i}[\delta_L(p, \sigma_j)]$ **do**
        $r \leftarrow r - \mathbf{u}_{n-i}[\delta_L(p, \sigma_j)]$
        $j \leftarrow j + 1$
    $p \leftarrow \delta_L(p, \sigma_j)$
    $w \leftarrow \mathrm{concatenate}(w, \sigma_j)$.

CHAPTER II

Recognizability

In this chapter, we study the first properties concerning the recognizability of subsets of $\mathbb{N}$ for numeration systems on a regular language.

In the first section, we show that for any numeration system S, the S-recognizability of a set is conserved under translation by a constant. Next, we show that ultimately periodic sets are always recognizable, for any abstract numeration system on a regular language. The material of the first two sections can be found in [**39**].

Consider a specific subset X of $\mathbb{N}$. One can ask the following: is X recognizable in any abstract numeration system ? As a consequence of Cobham's theorem, the answer is negative except for the ultimately periodic sets. Therefore, the question becomes: can we build a particular system S for X to be S-recognizable. This kind of question is treated in the third section of this chapter where we are interested in the recognizability of $P(\mathbb{N})$ when P is a polynomial. We prove that for any $P \in \mathbb{Q}[x]$ such that $P(\mathbb{N}) \subset \mathbb{N}$, there exists a numeration system S such that $P(\mathbb{N})$ is S-recognizable. The content of this section can be found in [**56**].

In the fourth section, we generalize the technique of the previous section and obtain abstract numeration systems that recognize exponential polynomial functions. In other words, for any function of the form

$$f(n) = \sum_{i=1}^{k} P_i(n)\, \alpha_i^n$$

where $P_i \in \mathbb{Q}[x]$ is such that $P(\mathbb{N}) \subset \mathbb{N}$ and $\alpha \in \mathbb{N}$, there exists a numeration system S such that $f(\mathbb{N})$ is S-recognizable.

The last section of this chapter is devoted to the characterization of the recognizable subsets of $\mathbb{N}$ in terms of $\mathbb{N}$-rational formal power series: a subset $X \subseteq \mathbb{N}$ is S-recognizable if and only if the series

$$\sum_{w \in \mathrm{rep}_S(X)} \mathrm{val}_S(w)\, w$$

is $\mathbb{N}$-rational.

1. Translation by a constant

Here we show that the S-recognizability of a set of non-negative integers is conserved under translation by a constant. The proof of this result reveals some interesting link between abstract and positional numeration systems.

Proposition II.1.1. *Let $S = (L, \Sigma, <)$ be a numeration system. For each natural number t, $X + t$ is S-recognizable if and only if $X \subset \mathbb{N}$ is S-recognizable.*

Proof. Let $\Sigma = \{\sigma_1 < \cdots < \sigma_k\}$ and $h : \Sigma^* \to \{1, \ldots, k\}^*$ be the homomorphism defined by $h(\sigma_i) = i$. It is cleat that

$$\Phi_k = \pi_k \circ h : \Sigma^* \to \mathbb{N}$$

is a strictly increasing one-to-one and onto mapping (see Example I.4.7, if w is the $(n+1)^{th}$ word of Σ^* then $h(w)$ is the k-adic representation of n).

Let us show that $A \subset \Sigma^*$ is regular if and only if $\Phi_k(A) \subset \mathbb{N}$ is k-recognizable. Let ν be the normalization function which maps representations of an integer n onto its normalized k-ary representation $\rho_k(n)$. If A is regular over Σ, by Theorem I.1.7 and Proposition I.2.6, $\nu(h(A))$ is regular over $\{0, \ldots, k-1\}$, in other words, the set $\pi_k(\nu(h(A))) = \Phi_k(A)$ is k-recognizable. Conversely, if $B \subset \mathbb{N}$ is k-recognizable, then $\Phi_k^{-1}(B) = h^{-1}(\nu^{-1}(\rho_k(B)) \cap \{1, \ldots, k\}^*)$ is regular since for $x \in \mathbb{N}$, $\nu^{-1}(\rho_k(x))$ can contain more than one element but each natural number x has a unique representation over $\{1, \ldots, k\}$, its k-adic representation.

By Theorem I.2.5, $A \subset \Sigma^*$ is regular if and only if $\Phi_k(A) \subset \mathbb{N}$ is k-definable. Let $\mathcal{N}_k$ be the k-definable set $\Phi_k(L)$ (by definition of S, $L \subset \Sigma^*$ is regular). We define the successor function by

$$\mathrm{Succ}_L : L \to L : w \mapsto \mathrm{rep}_S(\mathrm{val}_S(w) + 1).$$

The function

$$\mathrm{Succ}_{\mathcal{N}_k} = \Phi_k \circ \mathrm{Succ}_L \circ \Phi_k^{-1}$$

is the restriction to $\mathcal{N}_k$ of the function $f : \mathbb{N} \to \mathbb{N} : x \mapsto y = f(x)$ defined in $\langle \mathbb{N}, +, V_k \rangle$ by the formula

$$(y \in \mathcal{N}_k) \wedge (x < y) \wedge (\forall z)(z \in \mathcal{N}_k \wedge x < z) \to (y \leq z).$$

So $\mathrm{Succ}_{\mathcal{N}_k}$ maps k-definable sets onto k-definable sets and Succ_L transforms regular subsets of L into regular subsets of L. The commutative diagram on Figure II.1 summarizes the situation.

$$\begin{tikzcd}
\mathrm{rep}_S(\mathbb{N}) = L \arrow[r, "\Phi_k"] \arrow[d, "\mathrm{Succ}_L"'] & \mathcal{N}_k \subset \mathbb{N} \arrow[d, "\mathrm{Succ}_{\mathcal{N}_k}"] \\
\mathrm{rep}_S(\mathbb{N} \setminus \{0\}) \subset L & \mathrm{Succ}_{\mathcal{N}_k}(\mathcal{N}_k) \subset \mathcal{N}_k \arrow[l, "\Phi_k^{-1}"]
\end{tikzcd}$$

Figure II.1. The application Succ_L.

Assume now that X is S-recognizable, i.e., that $\text{rep}_S(X)$ is a regular set. Then $\text{rep}_S(X+t) = \text{Succ}_L^t(\text{rep}_S(X))$ is regular. The converse can be obtained in a similar way by observing that if $\Phi_k(\text{rep}_S(X+1)) \subset \mathcal{N}_k$ is definable by a formula φ of $\langle \mathbb{N}, +, V_k \rangle$ then the set

$$\{y \in \mathcal{N}_k \mid (\exists x)(\varphi(x)) \wedge (x = \text{Succ}_{\mathcal{N}_k}(y))\} = \Phi_k(\text{rep}_S(X))$$

is also definable.

□

EXAMPLE II.1.2. In this example, we emphasize the different functions encountered in Proposition II.1.1. Consider again the numeration system related to a^*b^*. We obtain Table II.1.

$\mathbb{N}$	$\text{rep}_S(\mathbb{N}) = L$	$h(L)$	$\nu(h(L))$	$\Phi_2(L) = \mathcal{N}_2$
0	ε	ε	ε	0
1	a	1	1	1
2	b	2	10	2
3	aa	11	11	3
4	ab	12	100	4
5	bb	22	110	6
6	aaa	111	111	7
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

TABLE II.1. The application $\Phi_2 = \pi_2 \circ h = \pi_2 \circ \nu \circ h$.

One can observe that $\mathcal{N}_2 = \mathbb{N}$ if and only if $L = \Sigma^*$. Here $ba \notin L$ and thus $5 \notin \mathcal{N}_2$.

The normalization which maps representations over $\{0,1,2\}$ onto the normalized binary representations is computed by the automaton depicted in Figure II.2 (the sink has not been represented). This automaton recognizes the set

$$\{(v,w) \in \{0,1,2\}^* \times \{0,1\}^* : |v| = |w|, \rho_2(\pi_2(v)) = w\}$$

where words are read from right to left. For instance, $\binom{0202}{1010}$ and $\binom{0122}{1010}$

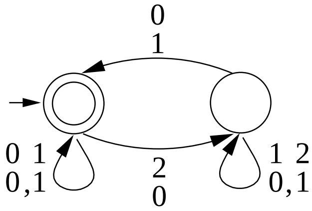

FIGURE II.2. Transducer computing normalization over $\{0,1,2\}$ in base 2.

are recognized by this automaton. In fact,"202" is a representation of ten, "1010" is its (normalized) binary representation and "122" is its dyadic representation. We use leading zeroes to obtain couples of words of the same length.

Remark II.1.3. The reader could be seduced by the logical arguments introduced in the proof of Proposition II.1.1. But it is difficult to use a logical definition of the recognizable subsets of $\mathbb{N}$ because these sets are viewed through the application $\Phi_k = \pi_k \circ h$, where k is the size of the alphabet.

For instance, the use of a bigger alphabet changes the set $\mathcal{N}_k$ itself ($\mathbb{N}$ is trivially recognizable since $\text{rep}_S(\mathbb{N}) = L$). Roughly speaking, $\mathcal{N}_k$ contains the positions of the words of L in the lexicographically ordered language $\{\sigma_1, \ldots, \sigma_k\}^*$. Consider again the language a^*b^* but now viewed as a subset of $\{a,b,c\}^*$ instead of $\{a,b\}^*$. In this case, one considers 3-adic representations and 3-recognizable sets. Therefore, the first terms of $\mathcal{N}_3$ are this time

$$0, 1, 2, 4, 5, 8, 13, \ldots$$

because c (3), ac (6), ba (7), bc (9), ca (10), cb (11), cc (12) do not belong to a^*b^*. For instance, $h(ab) = 12$ and $\pi_3(12) = \mathbf{1}.3^1 + \mathbf{2}.3^0 = 5$. Compare the results obtained here with the ones in Table II.1. For example, one can see that $3 \in \mathcal{N}_2$ but $3 \notin \mathcal{N}_3$ for the same language a^*b^*.

Another example is to consider the multiplication by 2. The application $x \mapsto 2x$ is quite difficult to be defined logically! Indeed, we have to work in $\mathcal{N}_k$ instead of $\mathbb{N}$: if $x' = \pi_k(h(\text{rep}_S(x)))$ is the i^{th} element of $\mathcal{N}_k$ for some i, then $(2x)' = \pi_k(h(\text{rep}_S(2x)))$ has to be defined as the $(2i)^{th}$ element of $\mathcal{N}_k$. We shall see in Chapter III that such a logical definition cannot exist for an arbitrary numeration system.

2. Arithmetic progressions

Let $S = (L, \Sigma, <)$ be a numeration system. Having in mind a possible generalization of Cobham's theorem, it is a quite remarkable fact that every arithmetic progression is S-recognizable (a special case of this result has been obtained separately in [**45**]).

Theorem II.2.1. *Let $S = (L, \Sigma, <)$ be a numeration system and r, d be two non-negative integers. The arithmetic progression $r + \mathbb{N}d$ is S-recognizable.*

The proof below exploits directly the regularity of L. Another proof at the end of this section uses the constructions introduced in the proof of Proposition II.1.1

Proof. We can assume that $r < d$. We show that the minimal automaton $\mathcal{A}_M = (Q_M, \Sigma, \delta_M, s_M, F_M)$ of $M = \text{rep}_S(r + \mathbb{N}d)$ is finite. Its states are the sets

$$w^{-1}.M = \{x \in \Sigma^* : \text{val}_S(wx) \equiv r \bmod d\},\ w \in \Sigma^*.$$

Let $\mathcal{A}_L = (Q_L, \Sigma, \delta_L, s_L, F_L)$ be the minimal automaton of L. Since L is a regular language, $\mathcal{A}_L$ is finite and it is well known (see justification on page 12) that the sequence $(\mathbf{u}_n(s_L))_{n\in\mathbb{N}} = (\#(L \cap \Sigma^n))_{n\in\mathbb{N}}$ satisfies a linear recurrent relation with coefficients in $\mathbb{Z}$. Moreover, $\mathbf{v}_{n+1}(s_L) - \mathbf{v}_n(s_L) = \mathbf{u}_{n+1}(s_L)$. So the sequence $(\mathbf{v}_n(s_L))_{n\in\mathbb{N}}$ is also the solution of a linear recurrence equation and is therefore ultimately periodic in the finite ring $\mathbb{Z}/(d)$, say of period t. By Lemma I.5.2, one has

$$\mathrm{val}_S(wx) = \mathrm{val}_{s_L.w}(x) + \mathbf{v}_{|wx|-1}(s_L) - \mathbf{v}_{|x|-1}(s_L.w) + \sum_{\substack{w'<w\\|w|=|w'|}} \mathbf{u}_{|x|}(s_L.w')$$

where the point in expression like $s_L.w$ represents the transition function δ_L of $\mathcal{A}_L$. This latter automaton being finite, $s_L.w$ can only take a finite number of values in Q_L. In $\mathbb{Z}/(d)$ and for $|w|$ large enough, the term $\mathbf{v}_{|wx|-1}(s_L)$ can be written as $\mathbf{v}_{C+|x|+i}(s_L)$ for some $i \in \{0,\ldots,t-1\}$ and for some constant C since $(\mathbf{v}_n(s_L))_{n\in\mathbb{N}}$ is ultimately periodic. Still working in $\mathbb{Z}/(d)$, the term $\sum_{\substack{w'<w\\|w|=|w'|}} \mathbf{u}_{|x|}(s_L.w')$ can be written as $\sum_{p\in Q_L} j_p\, \mathbf{u}_{|x|}(p)$ for some j_p's belonging to $\mathbb{Z}/(d)$.

So, for $|w|$ large enough, any set $w^{-1}.M$ is of the form

$$\{x \,:\, \mathrm{val}_q(x) + \mathbf{v}_{C+|x|+i}(s_L) - \mathbf{v}_{|x|-1}(q) + \sum_{p\in Q_L} j_p\, \mathbf{u}_{|x|}(p) \equiv r \bmod d\}$$

for some $q \in Q_L$, $j_p \in \{0,\ldots,d-1\}$ and $i \in \{0,\ldots,t-1\}$. So, there is finitely many sets of this kind and the set of states of the minimal automaton of $\mathcal{A}_M$, $\{w^{-1}.M \mid w \in \Sigma^*\}$, is finite.

$\square$

Remark II.2.2. We can give an explicit method to construct an NDFA accepting $\mathrm{rep}_S(r + \mathbb{N}d)$. The key of this method rests again on the ultimate periodicity in $\mathbb{Z}/(d)$ of the sequences $(\mathbf{u}_n)_{n\in\mathbb{N}}$ and $(\mathbf{v}_n)_{n\in\mathbb{N}}$. Let $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ be a DFA accepting L. For each $p \in Q$, there exist minimal constants a_p, b_p, e_p and f_p belonging to $\mathbb{N}$ such that $b_p, f_p \geq 1$,

$$\forall n \geq a_p,\ \mathbf{u}_n(p) \equiv \mathbf{u}_{n+b_p}(p) \bmod d$$

and

$$\forall n \geq e_p,\ \mathbf{v}_n(p) \equiv \mathbf{v}_{n+f_p}(p) \bmod d.$$

Set $\mathcal{M}$ to be the least common multiple of the b_p's and f_p's and

$$\mathcal{L} = \max\{\sup_{p\in Q} a_p, \sup_{p\in Q} e_p + 1\}.^1$$

[1]The consideration of $e_p + 1$ instead of e_p is due to the term $\mathbf{v}_{|w|-1}(p.\sigma)$ in the expression of $\mathrm{val}_p(\sigma w)$ given by (2) (see page 14).

With these notations, the formulation (2) of Lemma I.5.2 shows that if $|w| \geq \mathcal{L}$, then $\mathrm{val}_p(\sigma w)$ is equal modulo d to $\mathrm{val}_{p.\sigma}(w)$ plus some remainder depending only on p, σ and $|w| \bmod \mathcal{M}$. Thus, we consider the NDFA $\mathcal{B} = (Q' \cup \{f\}, \Sigma, E, I, F)$ where

$$\left\{ \begin{array}{rcl} Q' & = & Q \times \{0, \ldots, d-1\} \times \{0, \ldots, \mathcal{M}-1\} \\ I & = & \{(s, 0, i) \,|\, i = 0, \ldots, \mathcal{M}-1\} \\ F & = & \{f\} \text{ where } f \notin Q'. \end{array} \right.$$

The first component of a state of $\mathcal{B}$ is used to mimic the behavior of $\mathcal{A}$. The numerical value modulo d given by the already read letters is stored by the second component. The length modulo $\mathcal{M}$ of the remaining part of the word to be read is stored in the last component of the state. The transition relation of $\mathcal{B}$ is such that

$$\left\{ \begin{array}{ll} ((p, i, j), \sigma, (\delta(p, \sigma), i', j-1)) \in E, & \text{if } j \in \{1, \ldots \mathcal{M}-1\}; \\ ((p, i, j), \sigma, (\delta(p, \sigma), i', \mathcal{M}-1)) \in E, & \text{if } j = 0. \end{array} \right.$$

where the unique i', depending on p, σ and j, is easily computed through the use of Lemma I.5.2. If $x \in L_p \cap \Sigma^{\mathcal{L}}$ and $i \in \{0, \ldots, d-1\}$ are such that $\mathrm{val}_p(x) + i \equiv r \bmod d$ then

$$((p, i, \mathcal{L} \bmod \mathcal{M}), x, f) \in E.$$

The finite language $\mathrm{rep}_S(r + \mathbb{N}\, d) \cap \Sigma^{<\mathcal{L}}$ is treated separately.

The reading of a word w of length greater than $\mathcal{L}$ could be started in any of the initial states of $\mathcal{B}$. But notice that only one of these states has to be chosen (with respect to $|w|$) to reach the final state f at the end of the reading of w.

Example II.2.3. We apply the previously described method to obtain an NDFA recognizing $\mathrm{rep}_S(3\mathbb{N} + 1)$ where S is the numeration system constructed on the language L of the words over $\{a, b\}$ having an even number of b. The minimal automaton of L is depicted in Figure II.3.

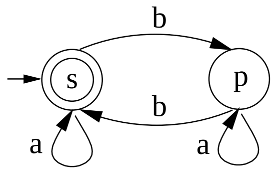

Figure II.3. DFA accepting words with an even number of b.

We have

$$\left\{ \begin{array}{l} \mathbf{u}_n(s) = 2^{n-1},\ n \geq 1 \\ \mathbf{u}_0(s) = 1 \end{array} \right. \quad \text{and} \quad \left\{ \begin{array}{l} \mathbf{u}_n(p) = 2^{n-1},\ n \geq 1 \\ \mathbf{u}_0(p) = 0. \end{array} \right.$$

So, in $\mathbb{Z}/(3)$, $\forall n \geq 1$, $\mathbf{u}_n(s) = \mathbf{u}_n(p) = (-1)^{n-1}$ and $\forall n \in \mathbb{N}$, $\mathbf{v}_n(s) = (-1)^n$ and $\mathbf{v}_n(p) = (-1)^n - 1$. With the notation of the previous remark, $\mathcal{L} = 1$ and $\mathcal{M} = 2$. From (2) page 14, it follows the next relations modulo 3. If $|w| \geq 1$,

$$\begin{array}{rcll} \mathrm{val}_s(aw) & = & \mathrm{val}_s(w)+ & (-1)^{|w|+1} \\ \mathrm{val}_s(bw) & = & \mathrm{val}_p(w)+ & (-1)^{|w|}+1 \\ \mathrm{val}_p(aw) & = & \mathrm{val}_p(w)+ & (-1)^{|w|+1} \\ \mathrm{val}_p(bw) & = & \mathrm{val}_s(w)+ & \underbrace{(-1)^{|w|}-1}_{\substack{\text{depends on} \\ |w| \bmod 2}}. \end{array}$$

With these relations, one obtains easily the main part of the transition relation:

	(s,0,0)	(s,1,0)	(s,2,0)	(s,0,1)	(s,1,1)	(s,2,1)
a	(s,2,1)	(s,0,1)	(s,1,1)	(s,1,0)	(s,2,0)	(s,0,0)
b	(p,2,1)	(p,0,1)	(p,1,1)	(p,0,0)	(p,1,0)	(p,2,0)

	(p,0,0)	(p,1,0)	(p,2,0)	(p,0,1)	(p,1,1)	(p,2,1)
a	(p,2,1)	(p,0,1)	(p,1,1)	(p,1,0)	(p,2,0)	(p,0,0)
b	(s,0,1)	(s,1,1)	(s,2,1)	(s,1,0)	(s,2,0)	(s,0,0)

For instance, $((s,1,0),b,(p,0,1)) \in E$ because in the minimal automaton of L, $s.b = p$ and when $|w| \equiv 0 \bmod 2$ then $1 + (-1)^{|w|} + 1 \equiv 0 \bmod 3$. To conclude, observe that $\mathrm{val}_s(a) = 1$, $b \notin L_s$, $a \notin L_p$ and $\mathrm{val}_p(b) = 0$. So, $((s,0,1),a,f)$ and $((p,1,1),b,f)$ also belong to the relation. The automaton recognizing $\mathrm{rep}_S(3\mathbb{N}+1)$ is depicted in Figure II.4.

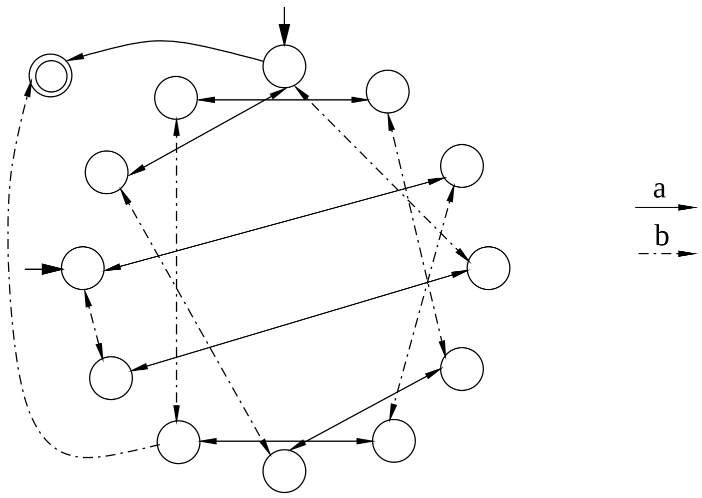

FIGURE II.4. An NDFA accepting $\mathrm{rep}_S(3\mathbb{N}+1)$.

To reach the final state, the words of even (resp. odd) length have to be read from the initial state $(s,0,0)$ on the left (resp. $(s,0,1)$ on the top) in Figure II.4. If the reading of a word begins in the wrong initial state, then no path reaches the final state.

To conclude this section, here is another proof of the fact that $r+\mathbb{N}d$ is S-recognizable where S is the numeration system $(L,\Sigma,<)$.

Proof. We use the notations of Proposition II.1.1. With $k=|\Sigma|$, the set $\Phi_k(L)$ is $|\Sigma|$-recognizable. By Proposition I.3.3, the characteristic sequence χ of $\Phi_k(L)$ is $|\Sigma|$-automatic. To conclude, use Proposition I.3.4 and observe that the characteristic sequence of $\Phi_k(\mathrm{rep}_S(r+\mathbb{N}d))$ is the image of χ under the *frying pan* finite transducer depicted in Figure II.5 (the tail has r nodes and the head counts d of them). In

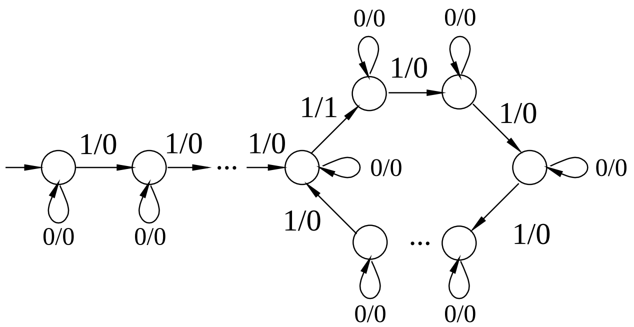

FIGURE II.5. A frying pan transducer.

the representation of a finitary sequential transformation, a label of the form α/β means the reading of α and the writing of β.

□

3. Polynomials

In a positional numeration system with an integer base, the set of perfect squares is not recognizable, see page 311 of [**28**]. On the other hand, the characteristic sequence of this set is a *morphic predicate*, i.e., it can be generated by iterating a non-uniform morphism, see page 141 of [**45**] or page 81 of this work. Here, in the frame of abstract numeration systems, we give an example of a system recognizing the set of perfect squares. This example is based on the following classical result in formal languages theory.

LEMMA II.3.1. [**63**] *Let L be a regular language over the totally ordered alphabet* $(\Sigma,<)$. *The set* $\mathrm{Min}(L,<)$ *(resp.* $\mathrm{Max}(L,<)$*) obtained by taking from all the words of L of the same length only the first (resp. last) one in the lexicographic order is regular.*

EXAMPLE II.3.2. Consider the language $L=a^*b^*\cup a^*c^*$. Using the previous lemma, one shows easily that the set of squares is S-recognizable in the system $S=(L,\{a,b,c\},a<b<c)$. Table II.2 gives the first words of L.

Observe that $(n+1)^2-n^2=2n+1$ is exactly the complexity function $\mathbf{u}_n(L)$ of the language L. So $\mathrm{rep}_S(\{n^2 : n\in\mathbb{N}\})$ is $\mathrm{Min}(L,<)=a^*$.

$\mathbb{N}$	$a^*b^* \cup a^*c^*$
0	$\boldsymbol{\varepsilon}$
1	**a**
2	b
3	c
4	**aa**
5	ab
6	ac
7	bb
8	cc
9	**aaa**
10	aab
$\vdots$	$\vdots$

TABLE II.2. The language $a^*b^* \cup a^*c^*$.

After seeing this example, J.-P. Allouche asked the following question. Is it possible to generalize the result on recognizability of the perfect squares to the set $\{n^k : n \in \mathbb{N}\}$, $k > 2$? Furthermore, if P is a polynomial belonging to $\mathbb{N}[x]$ (resp. $\mathbb{Z}[x]$ or $\mathbb{Q}[x]$) such that $P(\mathbb{N}) \subset \mathbb{N}$ then can one find a numeration system S such that $P(\mathbb{N})$ is S-recognizable ?

We answer affirmatively to all these questions. The constructions encountered in this section use the same technique as the one of the previous example: build a regular language L such that

$$\mathbf{u}_n(L) = P(n+1) - P(n).$$

The present section is organized as follows. First, we give an explicit iterative method to obtain regular languages $L^{(k)}$ such that the number of words of length n is exactly n^k (in [**70**] it is said that such languages can be easily obtained but we need our construction for later purposes). Next, we increase gradually the difficulty. We begin with the case $P \in \mathbb{N}[x]$ which is quite simple since we only deal with the operation of addition. Next we consider $P \in \mathbb{Z}[x]$ and the problem of subtraction must be resolved. The proof in the case of negative coefficients rests on our construction of the languages $L^{(k)}$. Finally, we face up the most general case, $P \in \mathbb{Q}[x]$ and the problem of division is solved through the use of Theorem II.2.1. In each of these last three steps, we give an instructive short example of construction.

3.1. Languages with complexity n^k. Let us first recall two definitions.

DEFINITION II.3.3. If v and w are two words in Σ^* then the *shuffle* of v and w is the language $v \sqcup\!\sqcup\, w$ of words $v_1 w_1 \dots v_n w_n$ such that

$$v = v_1 \cdots v_n,\ w = w_1 \cdots w_n,\ v_i, w_i \in \Sigma^*,\ 1 \leq i \leq n,\ n \geq 1.$$

If $L, M \subseteq \Sigma^*$ then the *shuffle* of the two languages is the language

$$L \sqcup\!\sqcup\, M = \{w \in \Sigma^* \mid w \in x \sqcup\!\sqcup\, y, \text{for some } x \in L, y \in M\}.$$

Recall that if L, M are regular then $L \sqcup\!\sqcup\, M$ is also regular (see for instance Proposition 3.5 of [**28**]).

DEFINITION II.3.4. Let $L \subseteq \Sigma^*$. Then Σ is the *minimal alphabet of* L if for each $\sigma \in \Sigma$, there exists a word w in L containing σ, $|w|_\sigma \neq 0$.

We want to build regular languages $L^{(k)}$ such that $\mathbf{u}_n(L^{(k)}) = n^k$, $k \in \mathbb{N}$. To that end, we define regular languages $M^{(k)}$ such that $\mathbf{u}_n(M^{(k)}) = (n+1)^{k-1}$, $k \geq 2$. The first two languages $L^{(0)}$ and $L^{(1)}$ are, for example, $L^{(0)} = a^*$ and $L^{(1)} = a^+ b^*$.

Let $k \geq 2$. Let us assume that we have $L^{(0)}, \dots, L^{(k-1)}$. One has

$$\mathbf{u}_n(M^{(k)}) = \sum_{j=0}^{k-1} \binom{k-1}{j} n^j.$$

Therefore, $M^{(k)}$ can be obtained as a finite union of regular languages $L^{(j)}$'s over distinct alphabets, $j < k$. That is

$$M^{(k)} = \bigcup_{j=0}^{k-1} \bigcup_{i=1}^{\binom{k-1}{j}} L_i^{(j)} \tag{7}$$

where $\mathbf{u}_n(L_i^{(j)}) = n^j$. If σ_k does not belong to the minimal alphabet of $M^{(k)}$, then we can define $L^{(k)}$ as

$$L^{(k)} = M^{(k)} \sqcup\!\sqcup\, \{\sigma_k\}. \tag{8}$$

Indeed, for each of the $(n+1)^{k-1}$ words w of length n in $M^{(k)}$, $w \sqcup\!\sqcup\, \sigma_k$ contains $n+1$ words of length $n+1$. So there are exactly $(n+1)^k$ words of length $n+1$ in $L^{(k)}$.

As an example, we give the nine words of length 3 in $L^{(2)}$. First, we have $M^{(2)} = a^* \cup b^+ c^*$ and Table II.3 shows the situation.

$M^{(2)} \cap \{a,b,c\}^2$	$L^{(2)} = M^{(2)} \sqcup\!\sqcup\, \{d\}$
aa	aad, ada, daa
bb	bbd, bdb, dbb
bc	bcd, bdc, dbc

TABLE II.3. The operation of shuffle.

In what follows, $M^{(k)}$ and $L^{(k)}$ will refer to the languages defined in (7) and (8) respectively.

REMARK II.3.5. Let l_k be the size of the minimal alphabet of $L^{(k)}$. The construction of $L^{(k)}$ gives

$$\begin{cases} l_0 = 1,\ l_1 = 2, \\ l_k = 1 + \sum_{j=0}^{k-1} \binom{k-1}{j} l_j,\ \forall k \geq 2. \end{cases}$$

By direct inspection, one can check that $l_2 = 4$, $l_3 = 10$, $l_4 = 30$, $l_5 = 104 < 5!$ and for $k = 6, \ldots, 9$, $l_k < k!$. Let $k \geq 9$. One has easily, by induction on k, the following upper bound

$$l_k < \sum_{j=0}^{k-1} j! \binom{k-1}{j} = e\,\Gamma(k,1) < e\,(k-1)!$$

where $\Gamma(k,1)$ is the incomplete gamma function defined by

$$\Gamma(a,b) = \int_b^{+\infty} t^{a-1} e^{-t}\, dt.$$

There are certainly several ways to improve the size of the minimal alphabet of a language K such that $\mathbf{u}_n(K) = n^k$. For instance, $a^*b^* \sqcup\!\sqcup \{c\}$ has the same complexity function as $L^{(2)} = (a^* \cup b^+c^*) \sqcup\!\sqcup \{d\}$ but is over a smaller alphabet. This simple modification would change l_2 and thus would change l_n for all $n \geq 3$. The question of finding the minimal size of the alphabet of K is beyond the concern of the present section. But we can give a lower bound for l_k. Indeed, for all n, it is clear that $l_k^n \geq n^k$ (with an alphabet of size l_k, there are at most l_k^n words of length n and we need at least n^k of them). Therefore, we have a lower bound $2^{k/2}$ on the size of the alphabet of a language containing n^k words of length n. Moreover, there is a systematic construction to get a regular language K over an alphabet with 2^k letters such that $\mathbf{u}_n(K) = n^k$. Consider the matrix

$$A_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

One has $(A_1^n)_{1,2} = n$. For $k \geq 2$, A_k is the direct product of the matrices A_1 and A_{k-1}, i.e. $A_k = A_1 \otimes A_{k-1}$. For instance,

$$A_2 = \begin{pmatrix} 1\,A_1 & 1\,A_1 \\ 0\,A_1 & 1\,A_1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Then $(A_k^n)_{1,2^k} = n^k$. This matrix A_k can be viewed as the transition matrix of a DFA over an alphabet of 2^k letters. For instance, a DFA for A_2 is depicted in Figure II.6. Consequently, there exists a language K over a 2^k letters alphabet such that $\mathbf{u}_n(K) = n^k$.

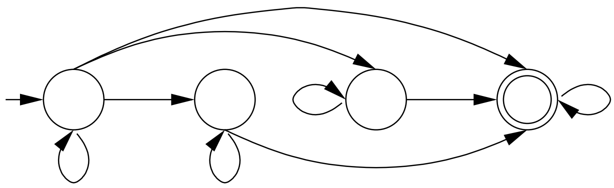

FIGURE II.6. DFA with a transition matrix A_2.

Nevertheless, in what follows, the main thing is that $L^{(k)}$ is built in (8) with one last operation of shuffle with a new letter σ_k.

REMARK II.3.6. After reading an earlier version of [**56**], J. Shallit suggested another construction of a language K such that $\mathbf{u}_n(K) = n^k$. It uses the following result (see Section 6.5 of [**12**])

$$n^k = \sum_{t=0}^{k} t!\, S(k,t) \binom{n}{t}$$

where $S(k,t)$ are the Stirling numbers of the second kind[2]. The language over $\{a,b\}$ with all strings of length n containing exactly t occurrences of the letter b is regular and has a complexity $\mathbf{u}_n = \binom{n}{t}$. Therefore a union of such languages over distinct alphabets gives the language K.

This construction is perhaps simpler than the construction of $L^{(k)}$ but uses a larger alphabet. The size of the minimal alphabet is

$$\max_{t=0,\dots,k} t!\, S(k,t)$$

and a lower bound is given by $k!$. We shall not use it in the following because the operation of shuffle given in (8) is needed in our proof of Lemma II.3.10.

3.2. Recognizability of polynomials belonging to $\mathbb{N}[x]$. The main idea is that we have to find a regular language L such that the positions of the first words of each length in the lexicographically ordered language L are the values taken by the polynomial.

PROPOSITION II.3.7. *Let $P \in \mathbb{N}[x]$. If $P(\mathbb{N}) \subset \mathbb{N}$ then $P(\mathbb{N})$ is S-recognizable for some abstract numeration system S.*

Proof. If P is constant then the result is obvious. So we may assume that P is a non-constant polynomial.

This proof exploits the following observation. If there exists $n_0 \in \mathbb{N}$ and a regular language L such that the position of the first word of length n_0 in the ordered language L is $P(n_0)$ and that for all $n \geq n_0$, $\mathbf{u}_n(L) = P(n+1) - P(n) \geq 0$, then the position of the first word of length $n \geq n_0$ in L is $P(n)$.

[2]The Stirling number of the second kind $S(k,t)$ is the number of ways of partitioning a set of k elements in t non-empty sets.

Since translation by a constant does not alter the recognizability of a set (see Proposition II.1.1), we can assume that $P(0) = 0$.

For each $n \in \mathbb{N}$, $P(n) < P(n+1)$ and the polynomial $P(n+1)-P(n)$ only contains powers of n with non-negative integral coefficients. Thus the construction of a regular language L such that for all $n \geq 1$,

$$\mathbf{u}_n(L) = P(n+1) - P(n) \tag{9}$$

can be achieved by union of languages $L^{(k)}$ over distinct alphabets Γ_k. It is clear that if $P(1) > 1$ then (9) is not satisfied for $n = 0$ because $\mathbf{u}_0(L) \leq 1$. We fix a total order $<$ on $\Sigma = \cup_k \Gamma_k$ and let $S = (L, \Sigma, <)$.

Finite modifications of a regular language do not alter its regularity. So we can assume that the first word w of length 2 in the lexicographically ordered language L, i.e., $\mathrm{Min}(L \cap \Sigma^2, <) = \{w\}$, is such that $\mathrm{val}_S(w) = P(2)$. To that end, the alphabet Σ should maybe be extended. Notice that we have to consider words of length 2 instead of words of length 1 because $P(1)$ could be greater than one and therefore cannot possibly be represented by the first word of length 1.

Let $n \geq 2$. Since $\mathbf{u}_n(L) = P(n+1) - P(n)$, it is clear that the numerical value of the first word of length n is $P(n)$ and

$$\mathrm{rep}_S(P(\mathbb{N}) \setminus \{P(0), P(1)\}) = \mathrm{Min}(L, <) \cap \Sigma^{\geq 2}.$$

By Lemma II.3.1, $P(\mathbb{N})$ is S-recognizable (two words should maybe be added to a regular language for the S-representations of $P(0)$ and $P(1)$).

□

Example II.3.8. Let $P(x) = 2\,x^2 + 3\,x$. Then

$$P(x+1) - P(x) = 4\,x + 5.$$

We consider the language L which is made up of four copies of $L^{(1)}$ and five copies of $L^{(0)}$. Observe that with five copies of $L^{(0)}$, *we obtain five words of any positive length but the only one empty word* ε. To ensure that $\mathrm{rep}_S(P(2)) = \mathrm{rep}_S(14)$ is the first word of length 2 in L, we add to our language four new words of length 1 (we possibly have to add four letters to the alphabet). This remark applies for all the following constructions: if one uses n copies of $L^{(0)}$ then add $n-1$ words of length 1 and treat the case $n = 1$ separately. So here we can take

$$L = \bigcup_{i=1}^{4} a_i^+ b_i^* \,\cup\, \bigcup_{i=1}^{5} c_i^* \,\cup\, \{b_1, b_2, b_3, b_4\}.$$

Table II.4 shows the first words of L. The lexicographic ordering is induced by the ordering $a_1 < \cdots < a_4 < b_1 < \cdots < b_4 < c_1 < \cdots < c_5$.

Corollary II.3.9. *Let $k \in \mathbb{N} \setminus \{0, 1\}$. There exists a numeration system S such that the set $\{x^k : x \in \mathbb{N}\}$ is S-recognizable.*

$\mathbb{N}$	L
$P(0) = 0$	$\boldsymbol{\varepsilon}$
1	a_1
$\vdots$	$\vdots$
4	a_4
$P(1) = 5$	$\mathbf{b_1}$
$\vdots$	$\vdots$
8	b_4
9	c_1
$\vdots$	$\vdots$
13	c_5
$P(2) = 14$	$\mathbf{a_1 a_1}$
15	$a_1 b_1$
16	$a_2 a_2$
$\vdots$	$\vdots$
21	$a_4 b_4$
22	$c_1 c_1$
$\vdots$	$\vdots$
26	$c_5 c_5$
$P(3) = 27$	$\mathbf{a_1 a_1 a_1}$
$\vdots$	$\vdots$
$P(4) = 44$	$\mathbf{a_1 a_1 a_1 a_1}$
$\vdots$	$\vdots$

TABLE II.4. Numeration system recognizing $P(x) = 2\,x^2 + 3\,x$.

3.3. Recognizability of polynomials belonging to $\mathbb{Z}[x]$. First, we set forth a lemma to get rid of the problem of coefficients belonging to $\mathbb{Z}$ instead of $\mathbb{N}$.

LEMMA II.3.10. *Let k and α be two positive integers. There exists a regular language $\mathcal{L}$ such that $\mathbf{u}_n(\mathcal{L}) = n^k - \alpha\, n^{k-1}$ for all $n \geq \alpha$ and $\mathbf{u}_n(\mathcal{L}) = 0$ for $n < \alpha$.*

Proof. Let us assume that $k \geq 2$. Let Σ be the minimal alphabet of $M^{(k)}$. From the construction given in (8), one has $L^{(k)} = M^{(k)} \sqcup\!\sqcup \{\sigma_k\}$ where $\sigma_k \notin \Sigma$. For $i = 1, \ldots, n$, $L^{(k)}$ has exactly n^{k-1} words of length n with σ_k in position i. From this observation, the language

$$\mathcal{L} = L^{(k)} \setminus \bigcup_{j=0}^{\alpha-1} \Sigma^*\, \sigma_k\, \Sigma^j$$

has exactly $n^k - \alpha\, n^{k-1}$ words of length n for $n \geq \alpha$. Notice that $\mathbf{u}_n(\mathcal{L}) = 0$ if $n < \alpha$.

If $k = 1$, it suffices to consider the language $\mathcal{L} = a^\alpha\, a^+ b^*$.

□

Proposition II.3.11. *Let $P \in \mathbb{Z}[x]$. If $P(\mathbb{N}) \subset \mathbb{N}$ then $P(\mathbb{N})$ is S-recognizable for some numeration system $S = (L, \Sigma, <)$.*

Proof. We proceed as in Proposition II.3.7 and consider the polynomial $Q(n) = P(n+1) - P(n)$. Observe that since $P(\mathbb{N}) \subset \mathbb{N}$, the coefficient of the dominant power in P is positive and thus the same remark holds for Q. By adding extra terms of the form $x^j - x^j$, if $\deg(Q) = k$ then $Q(x)$ can be written as

$$x^{i_1+1} - a_{i_1}\, x^{i_1} + \cdots + x^{i_r+1} - a_{i_r}\, x^{i_r} + \sum_{l=0}^{k} b_l\, x^l$$

where $i_1, \ldots, i_r \in \{0, \ldots, k-1\}$, $a_{i_1}, \ldots, a_{i_r} \in \mathbb{N} \setminus \{0\}$ and $b_0, \ldots, b_k \in \mathbb{N}$. Let $\alpha = \sup_{j=1,\ldots,r} a_{i_j}$. Using Lemma II.3.10, for $j = 1, \ldots, r$ we construct languages $\mathcal{L}_j$'s such that for all $n \geq \alpha$, $\mathbf{u}_n(\mathcal{L}_j) = n^{i_j+1} - a_{i_j}\, n^{i_j}$. By union of languages $\mathcal{L}_j$'s and $L^{(l)}$'s, we can construct a regular language L such $\forall n \geq \alpha$, $\mathbf{u}_n(L) = Q(n)$.

We can assume that L contains exactly $P(\alpha)$ words of length not greater than $\alpha - 1$. This can be achieved by adding or removing a finite number of words from the language L (this operation does not alter the regularity of L). Let S be a numeration system built upon the ordered regular language L. The first word of length α has numerical value equal to $P(\alpha)$ and $\forall n \geq \alpha$, $\mathbf{u}_n(L) = P(n+1) - P(n)$. Then one has

$$\mathrm{rep}_S(\{P(n) : n \geq \alpha\}) = \mathrm{Min}(L, <) \cap \Sigma^{\geq \alpha}.$$

To conclude we have to add a finite number of words for the representation of $P(0), \ldots, P(\alpha - 1)$ and

$$\mathrm{rep}_S(P(\mathbb{N})) = (\mathrm{Min}(L, <) \cap \Sigma^{\geq \alpha}) \cup \{\mathrm{rep}_S(P(0)), \ldots, \mathrm{rep}_S(P(\alpha - 1))\}.$$

By Lemma II.3.1, $\mathrm{rep}_S(P(\mathbb{N}))$ is regular.

□

Example II.3.12. Let $P(x) = x^4 - 3\,x^2 - 2\,x + 5$. Then

$$\begin{aligned} Q(n) = P(n+1) - P(n) &= 4\,x^3 + 6\,x^2 - 2\,x - 4 \\ &= 4\,x^3 + 5\,x^2 + x^2 - 3\,x + x - 4. \end{aligned}$$

With four copies of $L^{(3)}$, five copies of $L^{(2)}$ and using Lemma II.3.10, one can construct a regular language L such that[3]

$$\mathbf{u}_n(L) = \begin{cases} 4\,n^3 + 6\,n^2 - 2\,n - 4 & ,\text{if } n \geq 4; \\ 4\,n^3 + 5\,n^2 & ,\text{otherwise.} \end{cases}$$

We have $P(4) = 205$ and the number of words of length not greater than 3 belonging to L is 214 thus we remove 9 words of length not exceeding 3 in L. Therefore, the first word of length 4 in L is the representation of $P(4)$ and

$$\text{rep}_S(\{P(n) : n \geq 4\}) = \text{Min}(L, <) \cap \Sigma^{\geq 4} \tag{10}$$

is a regular subset of L. Since $\{P(0), \dots, P(3)\}$ is equal to $\{1, 5, 53\}$, we add the second, the 6^{th} and the 54^{th} word of L to (10) to obtain $\text{rep}_S(P(\mathbb{N}))$.

Example II.3.13. We begin another example which shows how to obtain a correct expression for $\mathbf{u}_n(L)$ in a trickier situation. Let $P(x) = x^5 - 4\,x^3 - 2\,x^2 + 8$, then

$$Q(x) = 5\,x^4 + 9\,x^3 + x^3 - 3\,x^2 + x^2 - 12\,x + x - 5.$$

To construct a language L, we use five copies of $L^{(4)}$, nine copies of $L^{(3)}$ and apply Lemma II.3.10 three times. Thus

$$\mathbf{u}_n(L) = \begin{cases} Q(n) & ,\ \text{if } n \geq 12; \\ 5\,n^4 + 10\,n^3 - 3\,n^2 + n - 5 & ,\ \text{if } 12 > n \geq 5; \\ 5\,n^4 + 10\,n^3 - 3\,n^2 & ,\ \text{if } 5 > n \geq 3; \\ 5\,n^4 + 9\,n^3 & ,\ \text{otherwise.} \end{cases}$$

3.4. Recognizability of polynomials belonging to $\mathbb{Q}[x]$. Finally, we obtain the theorem of recognizability in the general case.

Theorem II.3.14. *Let $P \in \mathbb{Q}[x]$. If $P(\mathbb{N}) \subset \mathbb{N}$ then $P(\mathbb{N})$ is S-recognizable for some numeration system $S = (L, \Sigma, <)$.*

Proof. Let

$$P(x) = \frac{a_k}{b_k}\,x^k + \frac{a_{k-1}}{b_{k-1}}\,x^{k-1} + \cdots + \frac{a_0}{b_0}$$

with $b_0, \dots, b_k, a_k \in \mathbb{N} \setminus \{0\}$ and $a_0, \dots, a_{k-1} \in \mathbb{Z}$. Let m be the least common multiple of $b_0, \dots, b_k$. One has

$$P = \frac{P'}{m}$$

with $P' \in \mathbb{Z}[x]$. By hypothesis $P(\mathbb{N}) \subset \mathbb{N}$; thus $P'(\mathbb{N}) \subset m\,\mathbb{N}$. As in Proposition II.3.11, there exist a constant α and a language $L' \subset \Sigma^*$ such that $\forall n \geq \alpha$,

$$\mathbf{u}_n(L') = P'(n+1) - P'(n) = m[P(n+1) - P(n)].$$

[3]By Lemma II.3.10, there exist languages $\mathcal{L}_1$ and $\mathcal{L}_2$ such that $\mathbf{u}_n(\mathcal{L}_1) = n^2 - 3n$ if $n \geq 3$ and $\mathbf{u}_n(\mathcal{L}_2) = n - 4$ if $n \geq 4$. Observe that if $n = 3$, $\mathbf{u}_n(\mathcal{L}_1) > 0$, $\mathbf{u}_n(\mathcal{L}_2) = 0$ and $4\,n^3 + 5\,n^2 + n^2 - 3\,n = 4\,n^3 + 5\,n^2$.

We modify L' (by adding or removing a finite number of words) to have

$$\sum_{i=0}^{\alpha-1} \mathbf{u}_i(L') = m\, P(\alpha).$$

In other words, if $\{w\} = \mathrm{Min}(L' \cap \Sigma^\alpha, <)$ then $\mathrm{val}_{S'}(w) = P'(\alpha)$ for the numeration system $S' = (L', \Sigma, <)$ where $<$ is a total ordering of Σ. By Theorem II.2.1, the arithmetic progression $m\,\mathbb{N}$ is S'-recognizable. Consequently, $L = \mathrm{rep}_{S'}(m\,\mathbb{N})$ is a regular language such that

$$\sum_{i=0}^{\alpha-1} \mathbf{u}_i(L) = P(\alpha) \text{ and } \forall n \geq \alpha,\ \mathbf{u}_n(L) = P(n+1) - P(n).$$

Indeed, to obtain L one takes in the lexicographically ordered language L' the words at position $im+1$, $i \in \mathbb{N}$. Since the first word of length α in L' is the first word of length α in L and its position in the lexicographically ordered language L is $P(\alpha)$, then we conclude as in Proposition II.3.7 by using Lemma II.3.1.

□

Example II.3.15. Let

$$\begin{aligned} P(x) &= \frac{x^4}{3} - 2\,x^3 + \frac{37}{6}\,x^2 - \frac{17}{2}\,x + 4 \\ &= \frac{1}{3}\,(x-7)\,x^2\,(x+1) + \frac{17}{2} x\,(x-1) + 4. \end{aligned}$$

It is clear that $P(\mathbb{N}) \subset \mathbb{N}$ since one of the numbers x, $x-7$ or $x+1$ must be divisible by 3 and one of the numbers x or $x-1$ must be divisible by 2. We have $m = 6$ and

$$\begin{aligned} P'(n+1) - P'(n) &= 8\,n^3 - 24\,n^2 + 46\,n - 24 \\ &= 7\,n^3 + 45\,n + n^3 - 24\,n^2 + n - 24. \end{aligned}$$

Using seven copies of $L^{(3)}$, 45 copies of $L^{(1)}$ and applying Lemma II.3.10 twice, we construct a language L' such that

$$\mathbf{u}_n(L') = \begin{cases} 6\,(P(n+1) - P(n)) & \text{, if } n \geq 24; \\ 7\,n^3 + 45\,n & \text{, otherwise.} \end{cases}$$

The number of words of length not greater than 23 in L' is 545652 and $6\,P(24) = 517776$. Thus we remove 27876 words from $L' \cap \Sigma^{\leq 23}$. In this new lexicographically ordered language, we only take the words at position $6i+1$, $i \in \mathbb{N}$, to obtain the regular language L. Thus the $[P(24)+1]^{th}$ word of L is the first word of length 24 belonging to L and

$$\mathbf{u}_n(L) = P(n+1) - P(n) \text{ if } n \geq 24.$$

Hence,

$$\mathrm{rep}_S(\{P(n) : n \geq 24\}) = \mathrm{Min}(L, <) \cap \Sigma^{\geq 24}.$$

Eventually we have as usual to add a finite number of words for the representation of $P(0), \ldots, P(23)$.

Remark II.3.16. We want to emphasize the importance of the constructions encountered in this section. Indeed, when one considers an arbitrary regular language L, the only infinite subsets of $L \subset \Sigma^*$ that are regular are related to ordering considerations: $\mathrm{Min}(L,<)$, $\mathrm{Max}(L,<)$ and eventually $\mathrm{rep}_S(r + d\,\mathbb{N})$ with $S = (L, \Sigma, <)$. Checking the regularity of $\mathrm{Min}(L,<)$ will be used as a powerful tool to make sure of the non-regularity of L itself (see Chapter III).

4. Exponential polynomial functions

Proceeding in the same way as in the previous section, we show that for any function of the form

$$f(n) = \sum_{i=1}^{k} P_i(n)\, \alpha_i^n,$$

where the P_i's are polynomials with rational coefficients such that $P_i(\mathbb{N})$ is included in $\mathbb{N}$ and the α_i's are non-negative integers, there exists an abstract numeration system S, such that $f(\mathbb{N})$ is S-recognizable. It is interesting to note that each predicate $\mathcal{P} = \{Q(n)\,\alpha^n \mid n \in \mathbb{N}\}$ for $\alpha \geq 0$ and Q a polynomial with non-negative integer values is morphic [**16**].

Proposition II.4.1. *Let $\alpha \in \mathbb{N} \setminus \{0,1\}$. There exists a numeration system S such that the set $\{\alpha^n : n \in \mathbb{N}\}$ is S-recognizable.*

Proof. We have to build a regular language L such that

$$\mathbf{u}_n(L) = \alpha^{n+1} - \alpha^n = (\alpha - 1)\,\alpha^n.$$

This can be achieved by using $\alpha - 1$ distinct copies of Σ^*, where Σ is an alphabet of cardinality α.

□

Proposition II.4.2. *Let $\alpha \in \mathbb{N} \setminus \{0,1\}$ and $P \in \mathbb{N}[x]$ such that $P(\mathbb{N}) \subset \mathbb{N}$. There exists a numeration system S such that the set*

$$\{P(n)\,\alpha^n : n \in \mathbb{N}\}$$

is S-recognizable.

Proof. We have to build a regular language L such that

$$\mathbf{u}_n(L) = P(n+1)\,\alpha^{n+1} - P(n)\,\alpha^n = [\alpha\, P(n+1) - P(n)]\,\alpha^n.$$

It is obvious that $\alpha\, P(n+1) - P(n) \in \mathbb{N}[x]$. It is enough to show how to build a regular language $L^{(k,\alpha)}$ containing exactly $n^k\,\alpha^n$ words of length $n > k \geq 1$. First, we build $L^{(1,\alpha)}$. Let Σ be such that $|\Sigma| = \alpha$. With α distinct copies of Σ^*, we obtain a language $M_{1,1}$ such that $\mathbf{u}_{n-1}(M_{1,1}) = \alpha^n$ if $n \geq 2$ (for each copy of Σ^*, one has $\mathbf{u}_{n-1}(\Sigma^*) = \alpha^{n-1}$ and there are α copies). If a does not belong to the minimal alphabet of $M_{1,1}$ then $L^{(1,\alpha)}$ can be defined as

$$L^{(1,\alpha)} = M_{1,1} \sqcup\!\sqcup \{a\}.$$

	description	*complexity*
$M_{3,1}$	α^3 copies of Σ^*	$\mathbf{u}_{n-3} = \alpha^n$
$M_{3,2}$	$M_{3,1} \sqcup\!\sqcup \{a_1\} \bigcup 2$ copies of $M_{2,1}$	$\mathbf{u}_{n-2} = n\,\alpha^n$
$M_{3,3}$	$M_{3,2} \sqcup\!\sqcup \{a_2\} \bigcup 1$ copy of $M_{2,2}$	$\mathbf{u}_{n-1} = n^2\,\alpha^n$
$L^{(3,\alpha)}$	$M_{3,3} \sqcup\!\sqcup \{a_3\}$	$\mathbf{u}_n = n^3\,\alpha^n$

TABLE II.5. The construction of $L^{(3,\alpha)}$, $(n \geq 4)$.

Indeed, if $n \geq 2$, for each of the α^n words w of length $n-1$ in $M_{1,1}$, $w \sqcup\!\sqcup a$ contains n words of length n. Next, we build $L^{(2,\alpha)}$. With α^2 distinct copies of Σ^*, we obtain a language $M_{2,1}$ such that $\mathbf{u}_{n-2}(M_{2,1}) = \alpha^n$ if $n \geq 3$. If a_1 does not belong to the minimal alphabet of $M_{2,1}$ then $\mathbf{u}_{n-1}(M_{2,1} \sqcup\!\sqcup \{a_1\}) = (n-1)\,\alpha^n$ for $n \geq 3$. Therefore,

$$M_{2,2} = (M_{2,1} \sqcup\!\sqcup \{a_1\}) \cup M_{1,1} \text{ is such that } \mathbf{u}_{n-1}(M_{2,2}) = n\,\alpha^n,\ n \geq 3$$

where the union is made from languages over distinct alphabets. If a_2 is a new symbol,

$$L^{(2,\alpha)} = M_{2,2} \sqcup\!\sqcup \{a_2\}.$$

Continuing this way, we can build $L^{(k,\alpha)}$ using the previously defined languages $M_{i,j}$'s and k operations of shuffle with new letters. For instance, the construction of $L^{(3,\alpha)}$ is summarized in Table II.5.

□

REMARK II.4.3. Observe that the last step in the building of $L^{(k,\alpha)}$ is the shuffle of $M_{k,k}$ and a new symbol that does not belong to the minimal alphabet of $M_{k,k}$. Moreover,

$$\mathbf{u}_{n-1}(M_{k,k}) = n^{k-1}\,\alpha^n.$$

So, with the same construction as in Lemma II.3.10 and Proposition II.3.11, we can consider polynomials belonging to $\mathbb{Z}[x]$. Proceeding as in Theorem II.3.14, we can assume that the polynomials belong to $\mathbb{Q}[x]$. Thus the following result is obvious.

THEOREM II.4.4. *Let P_i be polynomials belonging to $\mathbb{Q}[x]$ such that $P_i(\mathbb{N}) \subset \mathbb{N}$ and α_i be non-negative integers, $i = 1, \ldots, k$, $k \geq 1$. Set*

$$f(n) = \sum_{i=1}^{k} P_i(n)\,\alpha_i^n.$$

There exists a numeration system S such that $f(\mathbb{N})$ is S-recognizable.

5. Recognizable formal power series

We now characterize the S-recognizable subsets of $\mathbb{N}$ in terms of rational series in the noncommuting variables $\sigma \in \Sigma$ and with coefficients in $\mathbb{N}$. In particular, we show that $\sum_{n\in\mathbb{N}} n\,\mathrm{rep}_S(n)$ is rational (this kind of result is also discussed in [**8**] and [**17**]). Using classical results on rational series, we obtain a generalization of the fact that ultimately periodic sets are S-recognizable for any numeration system S. Let us start with a small introduction to rational and recognizable formal power series.

A formal power series with coefficients in the semiring $\langle \mathcal{R}, +, \cdot \rangle$ is a mapping $\mathfrak{T} : \Sigma^* \to \mathcal{R} : w \mapsto (\mathfrak{T}, w)$. It can be written as a formal sum

$$\mathfrak{T} = \sum_{w \in \Sigma^*} (\mathfrak{T}, w)\, w.$$

The set of formal power series of Σ with coefficients in $\mathcal{R}$ is denoted $\mathcal{R}\langle\langle \Sigma \rangle\rangle$. If one equips this set with the operations of *sum*,

$$(\mathfrak{T}_1 + \mathfrak{T}_2, w) = (\mathfrak{T}_1, w) + (\mathfrak{T}_2, w)$$

and *(Cauchy) product*,

$$(\mathfrak{T}_1\mathfrak{T}_2, w) = \sum_{v_1 v_2 = w} (\mathfrak{T}_1, v_1) \cdot (\mathfrak{T}_2, v_2),$$

then $\mathcal{R}\langle\langle \Sigma \rangle\rangle$ is a semiring.

Definition II.5.1. A *polynomial* of $\mathcal{R}\langle\langle \Sigma \rangle\rangle$ is a formal power series $\mathfrak{T}$ such that the set

$$\{w \in \Sigma^* : (\mathfrak{T}, w) \neq 0\}$$

is finite.

We mainly adopt the terminology of [**7**] concerning semirings, rational and recognizable series. The reader can also see [**60**].

Let us recall some definitions.

Definition II.5.2. A sequence $(\mathfrak{T}_n)_{n \in \mathbb{N}}$ of elements in $\mathcal{R}\langle\langle \Sigma \rangle\rangle$ converges to the limit $\mathfrak{T}$ if for all l there exists N_0 such that

$$|w| \leq l, n > N_0 \Rightarrow (\mathfrak{T}_n, w) = (\mathfrak{T}, w).$$

If $\mathfrak{T} \in \mathcal{R}\langle\langle \Sigma \rangle\rangle$ is *quasi-regular* (i.e., $(\mathfrak{T}, \varepsilon) = 0$) then the sequence $\mathfrak{T}, \mathfrak{T}^2, \mathfrak{T}^3, \ldots$ converges to 0 and

$$\lim_{n \to \infty} \sum_{k=1}^{n} \mathfrak{T}^k$$

exists. This limit is called the *quasi-inverse* of $\mathfrak{T}$.

Definition II.5.3. A subsemiring of $\mathcal{R}\langle\langle \Sigma \rangle\rangle$ is *rationally closed* iff it contains the quasi-inverse of every quasi-regular element. The family of $\mathcal{R}$-*rational* series over Σ is the smallest rationally closed subset of $\mathcal{R}\langle\langle \Sigma \rangle\rangle$ which contains all polynomials.

As a consequence of this latter definition, any $\mathcal{R}$-rational series can be obtained from polynomials by a finite number of sum, product and quasi-inversion.

Definition II.5.4. A series $\mathfrak{T} \in \mathcal{R}\langle\langle\Sigma\rangle\rangle$ is *$\mathcal{R}$-recognizable* if there exist $n \in \mathbb{N} \setminus \{0\}$, a morphism $\mu : \Sigma^* \to \mathcal{R}^{n\times n}$ and two matrices $\lambda \in \mathcal{R}^{1\times n}$ and $\gamma \in \mathcal{R}^{n\times 1}$ such that for all $w \in \Sigma^*$

$$(\mathfrak{T}, w) = \lambda\, \mu(w)\, \gamma.$$

In that case, (λ, μ, γ) is a *linear representation* of $\mathfrak{T}$.

According to the celebrated *Schützenberger's Theorem* the class of $\mathcal{R}$-rational and $\mathcal{R}$-recognizable formal power series coincides (a proof of this result can be found in [**7**]).

Finally, recall that for each word $v \in \Sigma^*$ and for each formal series $\mathfrak{T}$, one associates the series $v^{-1}\mathfrak{T}$ defined by

$$v^{-1}\mathfrak{T} = \sum_{w\in\Sigma^*} (\mathfrak{T}, vw)\, w.$$

In other words, $(v^{-1}\mathfrak{T}, w) = (\mathfrak{T}, vw)$.

It is shown in [**7**] that the series

$$\sum_{w\in\{0,1\}^*} \pi_2(w)\, w \in \mathbb{N}\langle\langle\{0,1\}\rangle\rangle$$

is rational. Here, we generalize this result for any numeration system on a regular language. Another proof of the following proposition can be found in [**17**] where complexity issues are discussed.

Proposition II.5.5. *Let $S = (L, \Sigma, <)$ be a numeration system. The formal series*

$$\mathfrak{F}_S = \sum_{w\in L} \mathrm{val}_S(w)\, w \in \mathbb{N}\langle\langle\Sigma\rangle\rangle$$

is $\mathbb{N}$-recognizable.

Proof. Let $\mathcal{A}_L = (Q_L, \Sigma, \delta_L, s_L, F_L)$ be the minimal automaton of L. For $p, q \in Q_L$, $\sigma \in \Sigma$, we introduce the following series in $\mathbb{N}\langle\langle\Sigma\rangle\rangle$

$$\begin{aligned}
\mathfrak{T}_p &= \sum_{w\in L_p, w\neq\varepsilon} [\mathrm{val}_p(w) - \mathbf{v}_{|w|-1}(p)]\, w\\
\mathfrak{U}_{q,p} &= \sum_{w\in L_q, w\neq\varepsilon} \mathbf{u}_{|w|}(p)\, w\\
\mathfrak{U}'_{q,p} &= \sum_{w\in L_q} \mathbf{u}_{|w|}(p)\, w\\
\mathfrak{V}_{q,p} &= \sum_{w\in L_q, w\neq\varepsilon} \mathbf{v}_{|w|-1}(p)\, w
\end{aligned}$$

$$\mathfrak{W}_{p,\sigma} = \begin{cases} [\mathrm{val}_p(\sigma) - \mathbf{v}_0(p)]\, \varepsilon & \text{, if } \sigma \in L_p;\\ 0 & \text{, otherwise.}\end{cases}$$

If $p, q \in Q_L$, $\sigma, \alpha \in \Sigma$, then we have the following relations

$$\begin{array}{rrcl}
\text{(i)} & \sigma^{-1}\mathfrak{T}_p & = & \mathfrak{T}_{p.\sigma} + \sum_{\alpha<\sigma} \mathfrak{U}_{p.\sigma,p.\alpha} + \mathfrak{W}_{p,\sigma} \\
\text{(ii)} & \sigma^{-1}\mathfrak{U}_{q,p} & = & \sum_{\alpha\in\Sigma} \mathfrak{U}'_{q.\sigma,p.\alpha} \\
\text{(iii)} & \sigma^{-1}\mathfrak{U}'_{q,p} & = & \sum_{\alpha\in\Sigma} \mathfrak{U}'_{q.\sigma,p.\alpha} \\
\text{(iv)} & \sigma^{-1}\mathfrak{V}_{q,p} & = & \mathfrak{V}_{q.\sigma,p} + \mathfrak{U}'_{q.\sigma,p} \\
\text{(v)} & \sigma^{-1}\mathfrak{W}_{p,\alpha} & = & 0.
\end{array}$$

To check relation (i), one has to compute $(\mathfrak{T}_p, \sigma w)$. Notice that $\sigma w \in L_p$ if and only if $w \in L_{p.\sigma}$. Use Lemma I.5.2 and treat the case $w = \varepsilon$ separately. For the relations (ii) and (iii), if σw belongs to L_q then $w \in L_{q.\sigma}$ and

$$(\mathfrak{U}_{q,p}, \sigma w) = \mathbf{u}_{|w|+1}(p) = \sum_{\alpha\in\Sigma} \mathbf{u}_{|w|}(p.\alpha).$$

In (iv), one has to observe that $\mathbf{v}_{|w|}(p) = \mathbf{v}_{|w|-1}(p) + \mathbf{u}_{|w|}(p)$. Checking relation (v) is immediate.

Therefore the submodule $\mathcal{O}$ of $\mathbb{N}\langle\langle\Sigma\rangle\rangle$ finitely generated by the series $\mathfrak{T}_p$'s, $\mathfrak{U}_{q,p}$'s, $\mathfrak{U}'_{q,p}$'s, $\mathfrak{V}_{q,p}$'s, $\mathfrak{W}_{p,\sigma}$'s is stable for the operation $\mathfrak{C} \mapsto \sigma^{-1}\mathfrak{C}$, $\sigma \in \Sigma$. By associativity of the operation $\mathfrak{C} \mapsto w^{-1}\mathfrak{C}$, this module is stable. By Proposition 1, page 18 of [**7**], the series in $\mathcal{O}$ are $\mathbb{N}$-recognizable. To conclude the proof, notice that

$$\mathfrak{T}_p + \mathfrak{V}_{p,p} = \sum_{w\in L_p, w\neq\varepsilon} \mathrm{val}_p(w)\, w = \sum_{w\in L_p} \mathrm{val}_p(w)\, w.$$

Indeed, if $\varepsilon \in L_p$ then $\mathrm{val}_p(\varepsilon) = 0$.

□

Example II.5.6. We consider again the abstract numeration system $S = (a^*b^*, \{a,b\}, a < b)$. We obtain a linear representation (λ, μ, γ) for $\mathfrak{F}_S$:

$$\lambda = (1\ 0\ 0)\,,\ \mu(a) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix},\ \mu(b) = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix},\ \gamma = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

where $\mu : \{a,b\}^* \to \mathbb{N}^{3\times 3}$ is a morphism of monoids. Thus, one has

$$\mathrm{val}_S(w) = \lambda\, \mu(w)\, \gamma.$$

Considering the definition of U-automata given in [**14**], we have the following characterization of the regular subsets of a regular language.

Lemma II.5.7. *Let $L \subset \Sigma^*$ be a regular language. Its minimal automaton is $\mathcal{A}_L = (Q_L, \Sigma, \delta_L, s_L, F_L)$. If $\mathcal{A}_K = (Q_K, \Sigma, \delta_K, s_K, F_K)$ is the minimal automaton of a regular language $K \subset L$ then*

$$h : Q_K \to Q_L : u^{-1}.K \mapsto u^{-1}.L$$

is a morphism h of automata between $\mathcal{A}_K$ and $\mathcal{A}_L$, i.e.,

$$\begin{cases} h(\delta_K(q,\sigma)) = \delta_L(h(q),\sigma), \ \sigma \in \Sigma, \ q \in Q_K, \\ h(s_K) = s_L, \\ h(F_K) \subseteq F_L. \end{cases}$$

Proof. The proof is an immediate consequence of the definition of the minimal automaton given in page 3.

□

With this lemma, we can generalize Proposition II.5.5 and obtain a characterization of the S-recognizable sets.

THEOREM II.5.8. *Let $S = (L, \Sigma, <)$ be a numeration system, a set $X \subseteq \mathbb{N}$ is S-recognizable if and only if the formal series*

$$\sum_{w \in \mathrm{rep}_S(X)} \mathrm{val}_S(w)\, w \in \mathbb{N}\langle\langle \Sigma \rangle\rangle$$

is $\mathbb{N}$-recognizable or $\mathbb{N}$-rational.

Proof. The condition is sufficient. The support of a recognizable series belonging to $\mathbb{N}\langle\langle \Sigma \rangle\rangle$ is a regular language (Lemme 2, page 49 of [**7**]).

The condition is necessary. By Lemma II.5.7, one has a morphism $h : \mathcal{A}_{\mathrm{rep}(X)} \to \mathcal{A}_L$ where $\mathcal{A}_{\mathrm{rep}(X)}$ (resp. $\mathcal{A}_L$) is the minimal automaton of $\mathrm{rep}_S(X)$ (resp. L). We proceed as in the proof of Proposition II.5.5. Let $Q_{\mathrm{rep}(X)}$ be the set of states of $\mathcal{A}_{\mathrm{rep}(X)}$; for $p, q \in Q_{\mathrm{rep}(X)}$, $\sigma \in \Sigma$, we introduce the following series

$$\begin{aligned}
\mathfrak{T}_p &= \sum_{w \in L_p, w \neq \varepsilon} [\mathrm{val}_{h(p)}(w) - \mathbf{v}_{|w|-1}(h(k))]\, w \\
\mathfrak{U}_{q,p} &= \sum_{w \in L_q, w \neq \varepsilon} \mathbf{u}_{|w|}(h(p))\, w \\
\mathfrak{U}'_{q,p} &= \sum_{w \in L_q} \mathbf{u}_{|w|}(h(p))\, w \\
\mathfrak{V}_{q,p} &= \sum_{w \in L_q, w \neq \varepsilon} \mathbf{v}_{|w|-1}(h(p))\, w
\end{aligned}$$

$$\mathfrak{W}_{p,\sigma} = \begin{cases} [\mathrm{val}_{h(p)}(\sigma) - \mathbf{v}_0(h(p))]\, \varepsilon & , \text{ if } \sigma \in L_p; \\ 0 & , \text{ otherwise.} \end{cases}$$

We conclude as in Proposition II.5.5.

□

In the second section of this chapter, it was shown that for any numeration system S, arithmetic progressions are always S-recognizable. Using formal series, we can obtain a generalization of this result. Here, the language L is not necessarily lexicographically ordered.

PROPOSITION II.5.9. *Let $L \subset \Sigma^*$ be an infinite regular language and $\alpha : L \to \mathbb{N}$ be a one-to-one correspondence. If*

$$\mathfrak{F}_\alpha = \sum_{w \in L} \alpha(w)\, w \in \mathbb{N}\langle\langle\Sigma\rangle\rangle$$

is $\mathbb{N}$-recognizable then $\alpha^{-1}(m + \mathbb{N}\, d)$ is a regular language.

PROOF. After division, m can be written as $q\,d + r$ with $0 \le r < d$. Let us assume first that $r = 0$. Consider the congruence of the semiring $\langle \mathbb{N}, +, \cdot, 0, 1\rangle$ defined by $n \sim n + d$. Observe that the semiring $\mathbb{N}/\!\sim$ is finite. We denote by φ the canonical morphism $\varphi : \mathbb{N} \to \mathbb{N}/\!\sim$. The characteristic series of L,

$$\underline{L} = \sum_{w \in L} w,$$

is recognizable (see Proposition 1, page 51 of [**7**]). So,

$$\mathfrak{U} = \varphi(\mathfrak{F}_\alpha + \underline{L}) = \sum_{w \in L} \varphi(\alpha(w) + 1)\, w$$

is rational (see Lemme 1, page 49 of [**7**]). Since $\mathbb{N}/\!\sim$ is finite and $\mathfrak{U}$ is rational, the set

$$\mathfrak{U}^{-1}(\{\varphi(1)\}) = \{w \in \Sigma^* : (\mathfrak{U}, w) = \varphi(1)\} = \alpha^{-1}(\mathbb{N}\, d)$$

is a regular language (see Proposition 2, page 52 of [**7**]). To conclude this first part, observe that if $q \ge 1$ then

$$\alpha^{-1}(m + \mathbb{N}\, d) = \alpha^{-1}(\mathbb{N}\, d) \setminus \alpha^{-1}(\{n\, d : 0 \le n < q\}).$$

(Removing a finite number of words from a regular language preserves its regularity.)

If $r \neq 0$ then

$$\mathfrak{U}^{-1}(\{\varphi(r)\}) = \alpha^{-1}(m + \mathbb{N}\, d) \cup \alpha^{-1}(\{n\, d + r : 0 \le n < q\})$$

where $\mathfrak{U} = \varphi(\mathfrak{F}_\alpha)$. We conclude as in the previous case. □

REMARK II.5.10. Observe in the previous proposition that as a consequence of the $\mathbb{N}$-recognizability of $\mathfrak{F}_\alpha$, the language L is necessarily regular (Lemme 2, page 49 of [**7**]).

COROLLARY II.5.11. *Arithmetic progressions are S-recognizable for any numeration system S.*

Proof. This is a direct consequence of Propositions II.5.5 and II.5.9. □

REMARK II.5.12. In the proof of Proposition II.5.9, we use the finiteness of $\mathbb{N}/\!\sim$, where $\sim$ is the congruence defined by $n \sim n + d$.

It would be interesting to characterize the congruences $\sim$ of the semiring $\langle \mathbb{N}, +, \cdot, 0, 1\rangle$ with finite index $p > 1$. Indeed, if there is a congruence $\sim$ different from $n \sim n + d$ such that $\mathbb{N}/\!\sim$ is finite, then we should obtain, using the same proof and thanks to Proposition II.5.5, new S-recognizable sets for any abstract numeration system S.

Let us assume that $\sim$ is a congruence with finite index $p > 1$. The canonical morphism is denoted by $\varphi : \mathbb{N} \to \mathbb{N}/\sim$. First notice that $\varphi(0) \neq \varphi(1)$, because $p > 1$. Since $\mathbb{N}/\sim$ is finite, there exist $x, y \in \mathbb{N}$ such that $x + y \sim x$. Let

$$y_0 = \min\{y > 0 | \exists x : x \sim x + y\} \text{ and } x_0 = \min\{x | x \sim x + y_0\}.$$

For all $n \in \mathbb{N}$ and $i = 0, \ldots, y_0 - 1$, one has $x_0 + i \sim x_0 + i + ny_0$. It is obvious that if $y_0 > 1$ then for $i, j \in \{0, \ldots, y_0 - 1\}$, $i \neq j$, one has $x_0 + i \not\sim x_0 + j$. By definition of x_0 and y_0, if $z < x_0$ then $\varphi^{-1}\varphi(z) = \{z\}$

Therefore the congruences of $\mathbb{N}$ with finite index are generated by the relation $n \sim n + y_0$ for n sufficiently large. So with the canvas of the proof of Proposition II.5.9, we can only reach the ultimately periodic sets. Once again, it can be considered as an interpretation of Cobham's theorem. The only subsets, possibly recognizable in any number system, are ultimately periodic.

CHAPTER III

Multiplication by a constant

The main purpose of this chapter is to study the stability of S-recognizability under addition and multiplication by a constant.

It is well known that for linear numeration systems $U = (U_n)_{n\in\mathbb{N}}$ such that the characteristic polynomial of $(U_n)_{n\in\mathbb{N}}$ is the minimal polynomial of a Pisot number, the problem of addition and multiplication by a constant is completely settled. The U-recognizable sets are exactly those defined in the first order structure $\langle \mathbb{N}, +, V_U\rangle$ (see Theorem I.2.5). It is obvious that addition and multiplication by a constant are definable in the Presburger arithmetic $\langle \mathbb{N}, +\rangle$. Therefore, U-recognizability is preserved under addition or multiplication by a constant.

Having generalized numeration systems at our disposal, we can consider the effect of addition on S-recognizable sets. If addition preserves S-recognizability then multiplication by 2 also preserves the S-recognizability. So, a natural question about the stability of S-recognizability arises. When does the multiplication by an integer λ preserve the S-recognizability ?

In the first section of this chapter, we show that for the numeration system $S = (a^*b^*, \{a, b\}, a < b)$, the multiplication by a non-negative integer λ transforms the S-recognizable sets into S-recognizable sets if and only if λ is an odd perfect square. As a consequence, addition cannot be a regular map for an arbitrary numeration system on a regular language.

The second section is devoted to recall results on the complexity function $\mathbf{u}_n(L)$ of regular languages L (see [**64**]). Indeed, the properties related to the stability of recognizability under multiplication by a constant are linked to the complexity function of the regular languages on which numeration systems are built. Recall that the complexity function of a regular language is either $\Theta(n^l)$ or of order $2^{\Omega(n)}$ [**64**]. In the first case, the language is said to be *polynomial*. Otherwise, it is said to be *exponential*. We show that if L is a polynomial language with $\Theta(n^l)$ complexity then the sequence $(\frac{\mathbf{v}_n}{n^{l+1}})_{n\in\mathbb{N}}$ converges to a strictly positive limit. In contrast, the sequence $(\frac{\mathbf{u}_n}{n^l})_{n\in\mathbb{N}}$ generally does not converge.

Notice that for the particular language a^*b^*, the only multiplicators preserving the recognizability are squares and this language has a polynomial complexity of degree one

$$\mathbf{u}_n(a^*b^*) = n + 1.$$

This observation is generalized by the following result: if S is a numeration system built on a regular language with $\Theta(n^l)$ complexity then the multiplication by λ preserves the S-recognizability only if $\lambda = \beta^{l+1}$ for some integer β.

For the sake of simplicity, we prove this result in two steps. In the third section, we assume that the complexity of the language of the numeration system S is a polynomial of degree l with rational coefficients. With such a language, we underline a subset $X \subset \mathbb{N}$ which is S-recognizable and we prove that λX is not S-recognizable for any $\lambda \in \mathbb{N} \setminus \{n^{l+1} : n \in \mathbb{N}\}$. The primary idea of the proof is the same as the one on which rests the construction of polynomial regular languages recognizing polynomial images of $\mathbb{N}$ found in the second chapter of this work. Next, thanks to the results of Section 2 and the scheme given in Section 3, we consider in Section 4 the general case of multiplication by a constant for an arbitrary polynomial language with $\Theta(n^l)$ complexity.

The end of this chapter is mainly related to exponential languages. In the fifth section, we consider numeration systems built on the complement of a polynomial language. As in the polynomial case, we find a recognizable set X and constants λ such that λX is not recognizable. Here, the λ's are the powers of the cardinality of the alphabet.

In the last section of this chapter, we study relations between some positional numeration systems U and an abstract system S on a regular exponential language L with exponential complement. We give sufficient conditions for the equivalence of S-recognizability and U-recognizability. These conditions are strongly dependent on the language L (on the complexity functions of the languages accepted from the different states of $\mathcal{A}_L$) and the recognizability of the normalization in U. Using these conditions, we give two examples of abstract numeration systems on an exponential language such that addition and multiplication by a constant preserve S-recognizability. One of these systems is a generalization of the well-known Fibonacci numeration system.

1. Multiplication in a^*b^*

We show that multiplication by a constant does not generally preserve recognizability. To that end, we use the extensively studied system $S = (a^*b^*, \{a, b\}, a < b)$, for which it is easy to see that

$$\mathrm{val}_S(a^p b^q) = \frac{1}{2}(p+q)(p+q+1) + q. \tag{11}$$

Remark III.1.1. Observe that the r.h.s. is the well-known function of Peano [**69**].

It would suffice to show that, say, multiplication by two does not preserve recognizability but here we are lucky enough to get more.

THEOREM III.1.2. [**39**] *Let S be the system $(a^*b^*, \{a,b\}, a<b)$ and let $\alpha \in \mathbb{N}$. Multiplication by α transforms the S-recognizable sets into S-recognizable sets if and only if α is an odd perfect square.*

Proof. *(i) Sketch.* If α is not a perfect square, we show that for a suitably chosen r,

$$\mathcal{L}_\alpha^r = a^r b^* \cap \mathrm{rep}_S(\alpha\,\mathrm{val}_S(a^*))$$

is infinite whereas the set of lengths $|\mathcal{L}_\alpha^r|$ only contains finite arithmetic progressions so that $\mathrm{rep}_S(\alpha \mathrm{val}_S(a^*))$ is not even context-free, thanks to Parikh's theorem [**42**].

If $\alpha = \beta^2$, $\mathbb{N}^2$ is divided into $\beta+1$ regions R_i in each of which an explicit formula for the function $M : (p,q) \mapsto (r,s)$ such that $\alpha\,\mathrm{val}_S(a^pb^q) = \mathrm{val}_S(a^rb^s)$ can be supplied. These regions come from length considerations: given a word of length l and of numerical value x, there are $\beta+1$ possible lengths for the word of value αx. When α is an odd perfect square, the fact that multiplication by α preserves the regularity of the subsets of a^*b^* then comes from an easy lemma. If $\alpha = (2\gamma)^2$ then we show that there exist constants j and k such that for n large enough, $\mathrm{rep}_S(\alpha\,\mathrm{val}_S(a^n)) = a^{j+n\gamma}b^{k+n\gamma}$. So that $\mathrm{rep}_S(\alpha\,\mathrm{val}_S(a^*))$ is not regular.

We give a visual example of these regions: consider the multiplication by 25 of the set of integers represented by $a^{3i}b^{4j}$, $0 \le i,j \le 30$. In this example, a point of coordinates (p,q) represents the word a^pb^q of numerical value $\frac{1}{2}(p+q)(p+q+1)+q$. The effect of the multiplication by 25 is depicted in Figure III.1. The result is an intricate set of points without any clear regularity. This regularity seems only to appear through the different regions as shown in Figures III.2 and III.3. The left (resp. right) row contains the different regions before (resp. after) multiplication.

(ii) Case of a non-perfect-square. Let α be a non-perfect-square integer. We have

$$l \in |\mathcal{L}_\alpha^r| \Leftrightarrow \exists p : \ \mathrm{val}_S(a^rb^{l-r}) = \alpha\,\mathrm{val}_S(a^p).$$

In other words, $l \in |\mathcal{L}_\alpha^r|$ if and only if

$$[2\,(r+t)+3]^2 - \alpha\,(2\,p+1)^2 = 8\,r + 9 - \alpha \tag{12}$$

for some p, where $t = l - r$.

To guarantee that $|\mathcal{L}_\alpha^r|$ is infinite, we choose r in such a way that

$$X^2 - \alpha\,Y^2 = 8\,r + 9 - \alpha \tag{13}$$

has infinitely many solutions (X,Y) with X, Y odd. To that end, it suffices to choose r such that $8r+9-\alpha > 0$ and that the equation (13) admits a solution $(x,1)$ with x odd (cf. Appendix concerning Pell's

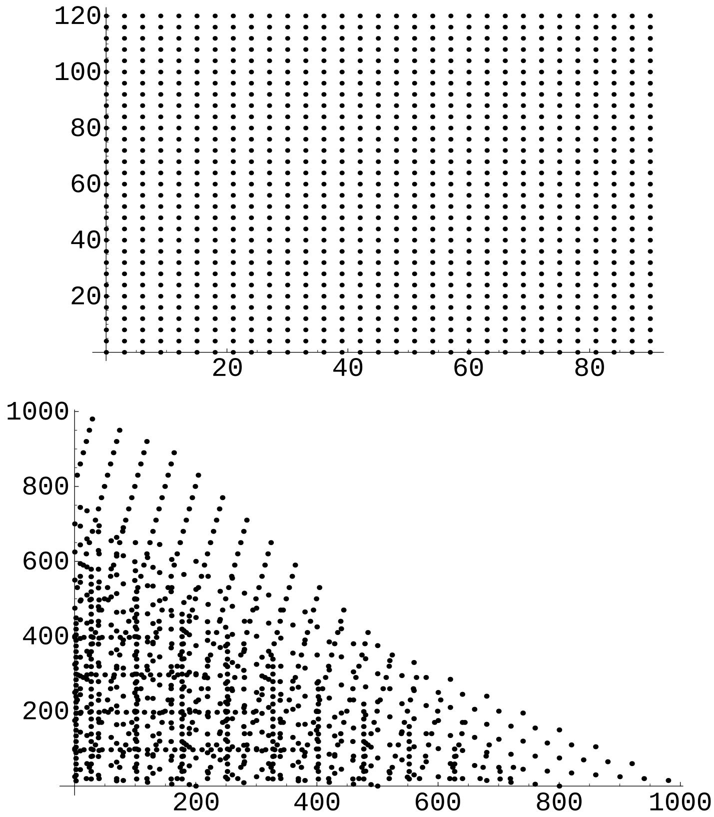

FIGURE III.1. The multiplication by 25 in a^*b^*.

equation). This can be achieved with r of the form z^2. Indeed, the equation $x^2 - 8z^2 = 9$ has infinitely many solutions given by

$$\begin{pmatrix} x_0 \\ z_0 \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}, \begin{pmatrix} x_{i+1} \\ z_{i+1} \end{pmatrix} = \begin{pmatrix} 3 & 8 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_i \\ z_i \end{pmatrix}, \forall i \in \mathbb{N}.$$

The x_i's are odd. We choose i such that $8z_i^2 + 9 - \alpha > 0$ and take $x = x_i$.

The set of the solutions of (13) with odd components is a finite union of sequences $(X_n^{(j)}, Y_n^{(j)})_{n\in\mathbb{N}}$, $j = 1, \ldots, m$, such that $X_n^{(j)} > C^n$ for some $C > 1$ (cf. Appendix).

We are now in a position to show that $|\mathcal{L}_\alpha^r|$ only contains finite arithmetic progressions. Suppose on the contrary that it contains an infinite progression, then there exist $\lambda, \mu \in \mathbb{N}, \mu > 0$, and, for each $t \in \mathbb{N}$, indices $n_t \in \mathbb{N}, j_t \in \{1, \ldots, m\}$ such that

$$\lambda + \mu\, t = X_{n_t}^{(j_t)} > C^{n_t}.$$

Given t, the sequence $n_0, \ldots, n_{mt}$ contains at least t distinct numbers. Therefore

$$\forall t \in \mathbb{N},\ \lambda + \mu m t > C^t,$$

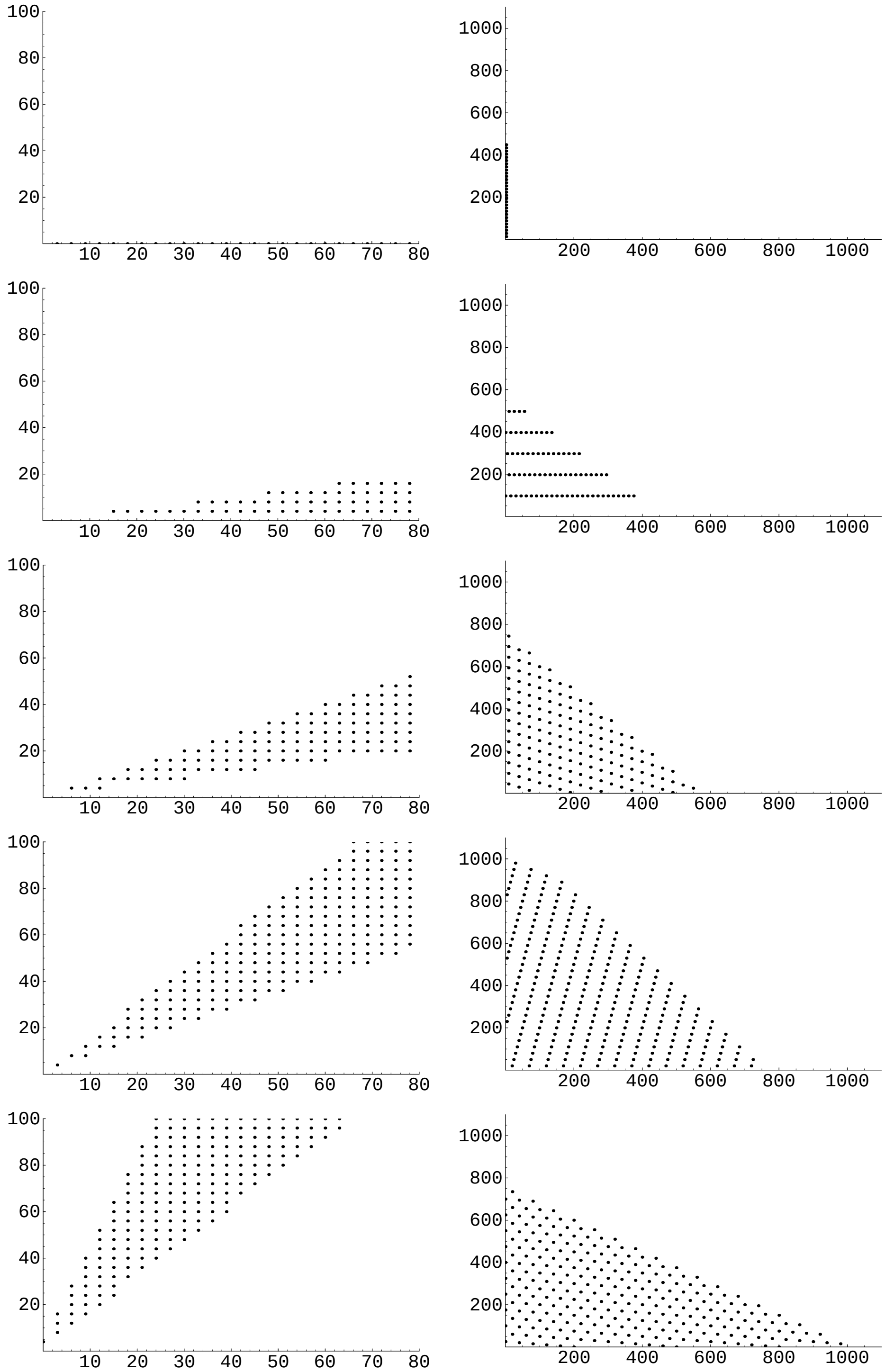

FIGURE III.2. The regions R_i before and after multiplication by 25.

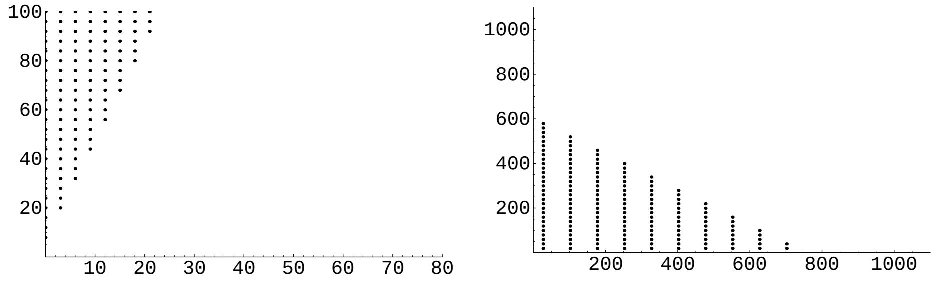

FIGURE III.3. The last region R_i.

a contradiction.

(iii) The case of an odd perfect square. Let $\alpha = \beta^2$ and β be an odd integer.

We want to compute r, t such that $\alpha\,\mathrm{val}_S(a^p b^q) = \mathrm{val}_S(a^r b^t)$, i.e.,

$$[2(r+t)+3]^2 - \beta^2[2(p+q)+3]^2 = 8r - 8p\beta^2 - 9(\beta^2-1).$$

Let $l = p+q$, $l' = r+t$. Observe that $l' \geq \beta\, l$. Indeed, let x be an integer, $|\mathrm{rep}_S(x)| = l$ if and only if,

$$1 + \cdots + l \leq x < 1 + \cdots + (l+1)$$

and it is clear that $1 + \cdots + \beta l \leq \beta^2(1 + \cdots + l)$. So, $|\mathrm{rep}_S(\beta^2 x)| \geq \beta l$ and we can write $l' = \beta l + u$ for some $u \geq 0$. Obviously,

$$\mathrm{val}_S(a^l) \leq \mathrm{val}_S(a^p b^q) \leq \mathrm{val}_S(b^l).$$

Then

$$\alpha l(l+1) \leq 2\alpha \mathrm{val}_S(a^p b^q) \leq \alpha l(l+3)$$

and

$$l'(l'+1) \leq 2\mathrm{val}_S(a^r b^t) \leq l'(l'+3).$$

Therefore, $l'(l'+1) \leq \beta^2 l(l+3)$ and $\beta^2 l(l+1) \leq l'(l'+3)$. Replacing l' with $\beta l + u$ in these latter inequalities and since l can be arbitrarily large, we obtain

$$\frac{\beta-3}{2} \leq u < \frac{3\beta-1}{2}.$$

From this, it follows easily that

$$r + s = \beta(p+q) + \left\lfloor \frac{\beta}{2} \right\rfloor + i$$

and thus

$$\left\{ \begin{array}{rcl} r = r_i(p,q) & := & \beta(i+1)p - \beta(\beta-i-1)q + \frac{1}{8}[(\beta+2i+2)^2 - 9] \\ t = t_i(p,q) & := & -\beta i p + \beta(\beta-i)q - \frac{1}{8}[(\beta+2i)^2 - 9] - 1 \end{array} \right.$$

for some $i \in \{-1, \ldots, \beta-1\}$. These equations together with the conditions $r, t \geq 0$ define $\beta+1$ regions R_i which divide $\mathbb{N}^2$.

The regular subsets of a^*b^* are the finite unions of sets of the form

$$D = \{a^{y+fz} b^{w+gx} : f, g \geq 0\},$$

$w, x, y, z \geq 0$. Substituting $y + fz$ and $w + gx$ in place of p and q, respectively, in $r_i(p,q)$ and $t_i(p,q)$, one sees that

$$D' = \mathrm{rep}_S[\alpha\,\mathrm{val}_S(D \cap R_i)]$$

is of the form (14) of Lemma III.1.3 below, the matrix A being

$$A = \begin{pmatrix} z\beta(i+1) & -x\beta(\beta - i - 1) \\ -z\beta i & x\beta(\beta - i) \end{pmatrix}.$$

One can apply the lemma to see that D' is regular except if $i = -1$ or $xz = 0$. In these cases, D' is easily shown to be regular by direct inspection. Indeed, if $i = -1$ and $x \neq 0$, one has

$$r \geq 0 \Leftrightarrow \beta^2(w + gx) \leq \frac{1}{8}(\beta^2 - 9).$$

Therefore, g has an upper bound and can only take a finite number of values, say $g_1, \ldots, g_n$. For each of these values, the condition $s \geq 0$ is given by

$$\beta(y + fz) \geq -\beta(\beta+1)(w + g_j x) + \frac{1}{8}[(\beta - 2)^2 - 9] + 1$$

and this gives a lower bound for f. If $x = 0$, there is no condition on g and we may have a lower bound for f. If $i \neq -1$, $x = 0$ or $z = 0$ may give lower bound for f and g. In all these cases, the corresponding languages of the form $a^n b^m$ are regular.

(iv) The case of an even perfect square. Let $\alpha = \beta^2$ with $\beta = 2\gamma$, $\gamma \in \mathbb{N} \setminus \{0\}$. Using the same kind of computation as in *(iii)*, we have, for p large enough, $\mathrm{rep}_S(\alpha\,\mathrm{val}_S(a^p)) = a^r b^t$ with

$$\begin{cases} r &= \gamma p + \frac{1}{2}\gamma(\gamma+1) - 1 \\ t &= \gamma p - \frac{1}{2}\gamma(\gamma - 1). \end{cases}$$

Therefore, $\mathrm{rep}_S(\alpha\,\mathrm{val}_S(a^*))$ is context-free but not regular.

□

Lemma III.1.3. *Let A be a non-singular $p \times p$ integral matrix. For $i = 1, \ldots, p$, set*

$$h_i(\vec{n}) = A_{i1}n_1 + \cdots + A_{ip}n_p - b_i,$$

where $\vec{n} = (n_1, \ldots, n_p) \in \mathbb{N}^p$ and $b_1, \ldots, b_p \in \mathbb{Z}$. If the entries of $\det(A)A^{-1}$ *are non-negative, then the language*

$$\mathcal{L} = \{a_1^{h_1} \cdots a_p^{h_p} : h_1(\vec{n}) \geq 0, \ldots, h_p(\vec{n}) \geq 0, \vec{n} \in \mathbb{N}^p\} \tag{14}$$

is a regular subset of $a_1^* \cdots a_p^*$.

Proof. If $\vec{n} \in \mathbb{N}^p$ satisfies $h_i(\vec{n}) \geq 0$ then $(A\vec{n})_i = b_i + u_i$, i.e.,

$$n_i = \sum_{j=1}^{p} (A^{-1})_{ij}(b_j + u_j), \tag{15}$$

for some $u_i \in \mathbb{N}$.

We need to describe those $\vec{u} = (u_1, \ldots, u_p) \in \mathbb{N}^p$ for which (15) defines non-negative integers n_i.

If $\det(A) < 0$, the entries of A^{-1} are negative, there are finitely many such $\vec{u}$ and $\mathcal{L}$ is finite. If $\det(A) > 0$, $(A^{-1})_{ij} \geq 0$, for large enough u_j's, (15) thus defines positive numbers n_i but it remains to make sure that they are integers. To that purpose, since $A^{-1} = \mathfrak{A}/\det(A)$, where the entries of $\mathfrak{A}$ are natural numbers, it is necessary and sufficient that the remainders $r_j \in \{0, \ldots, \det(A) - 1\}$ of the division of u_j by $\det(A)$ satisfy

$$\sum_{j=1}^{p} \mathfrak{A}_{ij}(b_j + r_j) \equiv 0 \pmod{\det(A)}.$$

There is a finite number of such $(r_1, \ldots, r_p)$ so that $\mathcal{L}$ is a finite union of regular languages of the form

$$\left(a_1^{\det(A)}\right)^* a_1^{s_1 \det(A) + r_1} \cdots \left(a_p^{\det(A)}\right)^* a_p^{s_p \det(A) + r_p}.$$

(The s_j's are chosen to guarantee that the u_j's are large enough for the corresponding n_i's to be non-negative.)

□

Theorem III.1.2 has a direct corollary. Let $x \in \Sigma^*$ and $y \in \Delta^*$, with Σ and Δ two finite alphabets. If $|x| = |y| + i$, $i \in \mathbb{N}$ then $(x, y)^\$ = (x, \$^i y)$ where \$ is a new symbol which does not belong to $\Sigma \cup \Delta$. If $|y| = |x| + i$ then $(x, y)^\$ = (\$^i x, y)$. This operation can be extended to n-uples of words. Let R be a relation over $\Sigma^* \times \Delta^*$. We say that R is *regular* if $R^\$$ is a regular language. This definition can be extended to n-ary relations. A mapping $f : \Sigma^* \to \Delta^*$ is *regular* if its graph is regular.

Corollary III.1.4. *For the numeration system S built on a^*b^*, the addition is not a regular map (i.e., the graph of the application $(x, y) \mapsto x + y$ is not regular).*

Proof. By Theorem III.1.2, there exists a subset X of $\mathbb{N}$ such that X is S-recognizable and $2X$ is not. Assume that the graph of the addition

$$\hat{\mathcal{G}} = \{(\mathrm{rep}_S(x), \mathrm{rep}_S(y), \mathrm{rep}_S(x + y))^\$: x, y \in \mathbb{N}\}$$

is regular. Let p_3 be the canonical homomorphism defined by

$$p_3(x, y, z) = z.$$

It is clear that the set $A = \{(\mathrm{rep}_S(x), \mathrm{rep}_S(x), w)^\$: x \in X, w \in \Sigma^*\}$ is regular. Therefore

$$A \cap \hat{\mathcal{G}} = \{(\mathrm{rep}_S(x), \mathrm{rep}_S(x), \mathrm{rep}_S(2x))^\$: x \in X\}$$

is regular. Thus $p_3(A \cap \hat{\mathcal{G}}) = \mathrm{rep}_S(2X)$ is also regular. This is a contradiction. □

2. About the complexity of regular languages

In this section, we extend in some way the results of [**64**] concerning the complexity function of polynomial regular languages. We also show that the sequence $(\mathbf{v}_n(L)/n^{l+1})_{n\in\mathbb{N}}$ converges to a strictly positive limit if the complexity of L is $\Theta(n^l)$. This result could be surprising because generally, the sequence $(\mathbf{u}_n(L)/n^l)_{n\in\mathbb{N}}$ does not converge.

Let us recall some notations. Let $f(n)$ and $g(n)$ be two functions, it is said that $f(n)$ is $O(g(n))$ if there exist positive constants c and n_0 such that for all $n \geq n_0$, $f(n) \leq c\,g(n)$; $f(n)$ is $\Omega(g(n))$ if there exist a strictly positive constant c and a strictly increasing infinite sequence $n_0, n_1, \ldots, n_i, \ldots$ of integers such that for all $i \in \mathbb{N}$, $f(n_i) \geq c\,g(n_i)$. The function $f(n)$ is $\Theta(g(n))$ if $f(n)$ is $O(g(n))$ and $\Omega(g(n))$.

The class of regular languages splits into two subclasses according whether the complexity function is bounded by a polynomial or is an exponential function of order $2^{\Omega(n)}$. The first subclass is the class of *polynomial regular languages* and the second is the class of *exponential languages*. The gap between polynomial and exponential languages is rendered by the following theorem.

Theorem III.2.1. [**64**, Theorem 6] *There does not exist a regular language such that its complexity function is neither $O(n^k)$, for some integer k, nor $2^{\Omega(n)}$.*

We now have a closer look at the polynomial languages. The lemmas recalled here are taken from [**64**].

Definition III.2.2. Let $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ be a DFA. For any word $w = w_1 \cdots w_n$ of length n over Σ, the *state transition sequence* of $\mathcal{A}$ on w, is the sequence of states

$$STS_{\mathcal{A}}(w) = q_{i_0} \cdots q_{i_n}$$

where $q_{i_0} = s$ and $\delta(q_{i_k}, w_{k+1}) = q_{i_{k+1}}$ for $k = 0, \ldots, n-1$.

Definition III.2.3. Let $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ be a DFA. A word $w \in \Sigma^*$ is said to be *t-tiered*, $t \geq 0$, with respect to $\mathcal{A}$ if the state transition sequence of w is given by

$$STS_{\mathcal{A}}(w) = \alpha\beta_1^{d_1}\gamma_1 \cdots \beta_t^{d_t}\gamma_t$$

where

1) $0 \leq |\alpha| \leq \#Q$

and for each i, $1 \leq i \leq t$,

2) $\beta_i = q_{i,0} \cdots q_{i,k_i}$ and $\gamma_i = q_{i,0} r_{i,1} \cdots r_{i,l_i}$, $0 \leq k_i, l_i \leq \#Q$, where the q's and r's are states of $\mathcal{A}$.
3) $q_{i,0}$ appears only as the first state in β_i and γ_i,
4) $d_i > 0$.

Lemma III.2.4. *Let L be the regular language accepted by the DFA $\mathcal{A} = (Q, \Sigma, \delta, s, F)$. If there exists a word $w \in L$ which is k-tiered with respect to $\mathcal{A}$ then the complexity function of L is $\Omega(n^{k-1})$.*

Lemma III.2.5. *Let L be a regular language accepted by some DFA $\mathcal{A}$. If the complexity function of L is $O(n^k)$ for some integer $k \geq 0$, then each word of L is t-tiered with respect to $\mathcal{A}$ for some non-negative $t \leq k+1$.*

Lemma III.2.6. *Let L be a regular language accepted by a DFA $\mathcal{A} = (Q, \Sigma, \delta, s, F)$. If there exists $k \geq 0$ such that each word of L is t-tiered with respect to $\mathcal{A}$, for some $t \leq k$, then L can be represented as a finite union of regular expressions of the form*

$$xy_1^* z_1 \cdots y_t^* z_t$$

with $0 \leq t \leq k$, $|x|, |z_1|, \ldots, |z_t| < \#Q$ and $|y_1|, \ldots, |y_t| \leq \#Q$.

Lemma III.2.7. *If L is the language defined by the regular expression $xy_1^* z_1 \cdots y_t^* z_t$, where $t > 0$ and $x, y_1, z_1, \ldots, y_t, z_t$ are all words over Σ, then the complexity function of L is $O(n^{t-1})$.*

In view of these lemmas, it is clear that the complexity function of a polynomial regular language is $\Theta(n^k)$ for some k and a language L is such that its complexity function is $O(n^k)$ if and only if L can be represented as a finite union of expressions of the form

$$xy_1^* z_1 \cdots y_t^* z_t$$

with $0 \leq t \leq k+1$.

The next lemma is just an improvement of Lemma III.2.4. We simply notice that one can consider an ultimately periodic sequence n_i such that $\mathbf{u}_{n_i}(L) \geq b_0\, n_i^l$ for some positive constant b_0.

Lemma III.2.8. *If L is a regular language such that its complexity function is $\Theta(n^l)$ for some integer l then there exist constants b_0 and C and a strictly increasing infinite sequence $n_0, n_1, \ldots, n_i, \ldots$ of integers such that for all $i \in \mathbb{N}$, $\mathbf{u}_{n_i}(L) \geq b_0\, n_i^l$ and $n_{i+1} - n_i = C$.*

Proof. Let $\mathcal{A}$ be a DFA accepting L. In view of the previous lemmas, it is obvious that there exists a word $w \in L$ which is $(l+1)$-tiered with respect to $\mathcal{A}$,

$$w = x\, y_1^{d_1}\, z_1 \ldots y_{l+1}^{d_{l+1}}\, z_{l+1}.$$

Let $C = |y_1| \ldots |y_{l+1}|$ and $C_i = \frac{C}{|y_i|}$, $1 \leq i \leq l+1$. For an arbitrary integer $t > 0$, let $n_t = |xz_1 \ldots z_{l+1}| + t\,C$. For any $l+1$ arbitrary non-negative integers $t_1, \ldots, t_{l+1}$ such that $t_1 + \cdots + t_{l+1} = t$, the word

$$x\, y_1^{t_1 C_1}\, z_1 \cdots y_{l+1}^{t_{l+1} C_{l+1}}\, z_{l+1}$$

belongs to L and its length is n_t. Indeed,

$$\begin{aligned} & |x\, y_1^{t_1 C_1}\, z_1 \cdots y_{l+1}^{t_{l+1} C_{l+1}}\, z_{l+1}| \\ = \quad & |x z_1 \ldots z_{l+1}| + t_1 \underbrace{C_1 |y_1|}_{=C} + \cdots + t_{l+1} \underbrace{C_{l+1} |y_{l+1}|}_{=C} \\ = \quad & |x z_1 \ldots z_{l+1}| + (t_1 + \cdots + t_{l+1})\, C \end{aligned}$$

The number of words of length n_t belonging to L is at least equal to the number $N^{l+1}(t)$ of $(l+1)$-uples of integers such that

$$N^{l+1}(t) = \#\left\{ (t_1, \ldots, t_{l+1}) \in \mathbb{N}^{l+1} \ : \ \sum_{i=1}^{l+1} t_i = t \right\}.$$

It is shown in Theorem 29 page 80 and Exercise 87 page 102 of [**21**] that

$$\mathbf{u}_{n_t}(L) \geq N^{l+1}(t) = \left(\begin{array}{c} t+l \\ t \end{array} \right) = \left(\begin{array}{c} t+l \\ l \end{array} \right) > \frac{t^l}{l!}.$$

Hence the conclusion, since n_t is a linear function of t.

□

REMARK III.2.9. Let us recall Skolem's theorem (see Chapter 4 of [**7**] or [**34**]). *Let*

$$\sum_{n \geq 0} a_n\, X^n$$

be a rational series with coefficients in $\mathbb{Q}$*. The set* $\mathcal{Z}_a = \{n \in \mathbb{N} \,|\, a_n = 0\}$ *is ultimately periodic.*

Therefore, the ultimate periodicity of the sequence n_i in Lemma III.2.8 is quite natural when one recall that $\mathbf{u}_n(L)$ is the solution of a linear recurrent equation. For decidable problems related to the set $\mathcal{Z}_a$ see [**6**] or [**49**].

Recall (see [**11**]) that the finite sum of integer powers is given by

$$\sum_{i=0}^{n} i^p = \frac{(n+B+1)^{p+1} - B^{p+1}}{p+1} \tag{16}$$

where all terms of the form B^m are replaced with the corresponding *Bernoulli numbers* B_m which are usually defined by the identity

$$\frac{x}{e^x - 1} = \sum_{m=0}^{\infty} \frac{B_m\, x^m}{m!}.$$

The formula (16) will be useful in the proof of Theorem III.2.12.

REMARK III.2.10. We give below a useful result on the convergence of the sequence $(\frac{\mathbf{v}_n(L)}{n^{l+1}})_{n\in\mathbb{N}}$ when L is a language of complexity $\Theta(n^l)$. We prove that the limit always exists. Although this is generally not the case for the sequence $(\frac{\mathbf{u}_n(L)}{n^l})_{n\in\mathbb{N}}$. Consider for instance the language

$$W = a^* b^* \cap (\{a,b\}^2)^*.$$

It is obvious that $\mathbf{u}_{2n+1}(W) = 0$, $\mathbf{u}_{2n}(W) = 2n+1$ and $\mathbf{v}_{2n}(L) = \mathbf{v}_{2n+1}(L) = (n+1)^2$. It is clear that $(\frac{\mathbf{u}_n(W)}{n})_{n\in\mathbb{N}}$ does not converge since it contains two subsequences converging to different limits,

$$\lim_{n\to\infty} \frac{\mathbf{u}_{2n}(W)}{n} = 2,\ \lim_{n\to\infty} \frac{\mathbf{u}_{2n+1}(W)}{n} = 0$$

and

$$\lim_{n\to\infty} \frac{\mathbf{v}_n(W)}{n^2} = 1.$$

LEMMA III.2.11. *Let $\rho_1,\dots,\rho_k,\theta_1,\dots,\theta_k,\Phi_1,\dots,\Phi_k$ be real numbers such that for all $i \neq j$, $\theta_i \not\equiv \theta_j \pmod{2\pi}$ and for all j, $\rho_j \neq 0$. There exists $\varepsilon > 0$ such that*

$$M_n = |\rho_1\, e^{i(n\theta_1+\Phi_1)} + \cdots + \rho_k\, e^{i(n\theta_k+\Phi_k)}| > \varepsilon$$

for an infinite sequence of integers n.

Proof. Assume that for all $\varepsilon > 0$, $M_n \geq \varepsilon$ only for a finite number of integers n. In other words, $M_n \to 0$. By successive applications of Bolzano-Weierstrass'theorem, there exist complex numbers $z_1,\dots,z_k$ and a subsequence $t(n)$ such that

$$\rho_j\, e^{i\,(t(n)\,\theta_j+\Phi_j)} \to z_j \text{ and } |z_j| = |\rho_j| \neq 0.$$

Since $M_n \to 0$, then $\sum_{j=1}^k z_j = 0$. For $l = 0,\dots,k-1$, one gets in the same manner

$$\sum_{j=1}^k \rho_j\, e^{i\,[(t(n)+l)\,\theta_j+\Phi_j]} \to \sum_{j=1}^k z_j\, e^{i\,l\,\theta_j} = 0.$$

Therefore one has

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ e^{i\theta_1} & e^{i\theta_2} & \dots & e^{i\theta_k} \\ \vdots & \vdots & & \vdots \\ e^{i\,(k-1)\theta_1} & e^{i\,(k-1)\theta_2} & \dots & e^{i\,(k-1)\theta_k} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

This equality leads to a contradiction since the Vandermonde determinant does not vanish.

□

We are now able to prove the convergence of $(\mathbf{v}_n(L)/n^{l+1})_{n\in\mathbb{N}}$. This result and its proof were suggested by P. Lecomte. Recall that if L is regular, then the sequence $(\mathbf{u}_n(L))_{n\in\mathbb{N}}$ satisfies a linear recurrence relation (a justification of this fact was given in page 12).

THEOREM III.2.12. *If L is a polynomial regular language such that $\mathbf{u}_n(L)$ is $\Theta(n^l)$ then the sequence $(\frac{\mathbf{v}_n}{n^{l+1}})_{n\in\mathbb{N}}$ converges to a strictly positive limit. Moreover, 1 is a root of the characteristic polynomial of the sequence $(\mathbf{u}_n(L))_{n\in\mathbb{N}}$ with a multiplicity at least equal to $l+1$.*

Proof. The sequence $(\mathbf{u}_n(L))_{n\in\mathbb{N}}$ satisfying a linear recurrence relation, it can be written as a finite sum,

$$\mathbf{u}_n(L)=\sum_i P_i(n)\, z_i^n \tag{17}$$

where the P_i's are polynomials and the z_i's are distinct complex numbers.

Let $\tau=\sup_i|z_i|$ and d be the maximal degree of polynomials P_k's corresponding to the different numbers of modulus τ. Let us take the following notations. Let $z_1=\tau\, e^{i\theta_1},\ldots,z_r=\tau\, e^{i\theta_r}$ be the numbers of modulus τ having a corresponding polynomial P_k of degree d (the coefficient of n^d in $P_k(n)$ is denoted by c_k). We may assume that $\theta_j\not\equiv\theta_k \pmod{2\pi}$ for $j\neq k$, $j,k\in\{1,\ldots,r\}$. Let $z_{r+1},\ldots,z_s$ be the other numbers of modulus τ having a corresponding polynomial P_k of degree less than d. Finally, $z_{s+1},\ldots,z_t$ are the numbers of modulus less than τ. So we can write

$$\left|\frac{\mathbf{u}_n(L)}{n^l}\right|=\frac{\tau^n\, n^d}{n^l}\,|c_1\, e^{in\theta_1}+\cdots+c_r\, e^{in\theta_r}+R_n|.$$

In the last expression, R_n is made up of two sorts of terms, namely

$$R_n=\frac{1}{\tau^n\, n^d}\left(\sum_{j=1}^{r}(P_j(n)-c_j\, n^d)\, z_j^n+\sum_{j=r+1}^{t}P_j(n)\, z_j^n\right)$$

So, $R_n\to 0$ if $n\to+\infty$. Therefore, by Lemma III.2.11, there exist $\varepsilon>0$ and an infinite sequence of integers such that

$$\left|\frac{\mathbf{u}_n(L)}{n^l}\right|\geq\frac{\tau^n\, n^d}{n^l}(\varepsilon-|R_n|).$$

For n large enough, $|R_n|\leq\varepsilon/2$ and $|\frac{\mathbf{u}_n(L)}{n^l}|\geq\tau^n\, n^{d-l}\,\frac{\varepsilon}{2}$ occurs infinitely often. If $\tau>1$ or if $\tau=1$ and $d>l$, we obtain a contradiction with the hypothesis that $\mathbf{u}_n(L)$ is $\Theta(n^l)$.

Consequently, in (17) the degree of the polynomials P_j's corresponding to the numbers z_j's of modulus one cannot exceed l and there is no z_i of modulus greater than one. So there exist polynomials Q_j's of degree not exceeding l such that

$$\mathbf{u}_n(L)=\sum_{j=0}^{k}Q_j(n)\, e^{in\theta_j}+T(n)$$

with $\theta_0=0$ and for $i\neq j$, $\theta_i\not\equiv\theta_j \pmod{2\pi}$ and

$$T(n)=\sum_{i:|z_i|<1}P_i(n)\, z_i^n.$$

Let q_j be the coefficient of n^l in $Q_j(n)$; notice that q_j could be zero, $q = 0, \ldots, k$. We have

$$\mathbf{u}_n(L) = q_0\, n^l + \sum_{j=1}^{k} q_j\, e^{in\theta_j}\, n^l + \sum_{j=0}^{k} (Q_j(n) - q_j\, n^l)\, e^{in\theta_j} + T(n)$$

and by definition of $\mathbf{v}_n$, $\frac{\mathbf{v}_n(L)}{n^{l+1}}$ can be written

$$\begin{aligned} & q_0 \frac{\sum_{p=0}^{n} p^l}{n^{l+1}} + \sum_{j=1}^{k} q_j \frac{\sum_{p=0}^{n} e^{ip\theta_j}\, p^l}{n^{l+1}} \\ + & \sum_{j=0}^{k} \frac{\sum_{p=0}^{n} (Q_j(p) - q_j\, p^l)\, e^{ip\theta_j}}{n^{l+1}} + \frac{\sum_{p=0}^{n} T(p)}{n^{l+1}}. \end{aligned}$$

We have, by (16)

$$\lim_{n\to\infty} \frac{\sum_{p=0}^{n} p^l}{n^{l+1}} = \frac{1}{l+1}$$

and

$$\lim_{n\to\infty} \frac{\sum_{p=0}^{n} (Q_j(p) - q_j\, p^l)\, e^{ip\theta_j}}{n^{l+1}} = 0$$

because the degree of $Q_j(p) - q_j\, p^l$ is less than l. For the second term,

$$\lim_{n\to\infty} \frac{\sum_{p=0}^{n} e^{ip\theta_j}\, p^l}{n^{l+1}} = 0.$$

The computation of this latter limit can be achieved by applying $(z\frac{\partial}{\partial z})^l$ to $\sum_{p=0}^{n} z^p$. Indeed, in the one hand, one has

$$\left(z\frac{\partial}{\partial z}\right)^l \sum_{p=0}^{n} z^p = \sum_{p=0}^{n} z^p\, p^l.$$

On the other hand,

$$\left(z\frac{\partial}{\partial z}\right)^l \frac{z^{n+1}-1}{z-1} = \sum_{k=0}^{l} n^k\, z^n \frac{R_k(z)}{(z-1)^{l+1-k}} + \frac{R(z)}{(z-1)^{l+1}}$$

where the R_k's and R are polynomials of degree less than $l+2$. Observe that if $z = e^{i\theta_j}$ then the moduli of the fractions in the r.h.s. equation are bounded. Finally,

$$\lim_{n\to\infty} \frac{\sum_{p=0}^{n} T(p)}{n^{l+1}} = 0.$$

To obtain this limit, let us consider one of the term appearing in $T(n)$, say $P(n)\, z^n$, where $|z| = \zeta < 1$ and $P(n) = \sum_{i=0}^{d} a_i\, n^i$, $d \in \mathbb{N}$. We have

$$\begin{aligned}\left|\frac{1}{n^{l+1}}\sum_{p=0}^{n} P(p)\, z^p\right| &\leq \frac{1}{n^{l+1}}\sum_{p=0}^{n}\zeta^p \sum_{i=0}^{d}|a_i|\, p^i\\ &\leq \sum_{i=0}^{d}\frac{|a_i|}{n^{l+1}}\sum_{p=0}^{n}\zeta^p\, p^i.\end{aligned}$$

We conclude with the same kind of computation as in the previous limit and using the fact that $\zeta < 1$.

Eventually q_0 cannot vanish. Otherwise, $\lim_{n\to\infty}\frac{\mathbf{v}_n(L)}{n^{l+1}} = 0$, which is a contradiction with Lemma III.2.13 below. Thus, 1 has at least a multiplicity $l+1$ as a root of any polynomial satisfying the recurrence.

□

LEMMA III.2.13. *Let L be a regular language such that its complexity function is $\Theta(n^l)$. The sequence $\mathbf{v}_n(L) = \sum_{i=0}^{n}\mathbf{u}_i(L)$ is $\Omega(n^{l+1})$.*

Proof. With the sequence n_i of Lemma III.2.8, we have

$$\mathbf{v}_{n_i}(L) = \sum_{j=0}^{n_i}\mathbf{u}_j(L) \geq \sum_{j=0}^{i}\mathbf{u}_{n_j}(L) \geq b_0 \sum_{j=0}^{i}(n_0 + j\, C)^l \geq b_0\, C^l \sum_{j=0}^{i} j^l.$$

Since $n_i = n_0 + i\, C$, there exists a constant $M > 0$ such that for i large enough,

$$\mathbf{v}_{n_i}(L) \geq M\, n_i^{l+1}.$$

□

REMARK III.2.14. Theorem III.2.12 can be applied not only to the complexity function of polynomial regular languages but to any linear recurrent sequence $(\mathbf{u}_n)_{n\in\mathbb{N}}$ which is $\Theta(n^l)$ with a possible $q_0 = 0$.

REMARK III.2.15. So far, we have taken stock of the notations for a regular language with $\Theta(n^l)$ complexity. From Lemma III.2.8, we have an ultimately periodic sequence n_i of period C such that

$$\boxed{n_i = n_0 + i\, C}$$

and a constant b_0 such that

$$\boxed{\mathbf{u}_{n_i}(L) \geq b_0\, n_i^l.}$$

Since $\mathbf{u}_n(L)$ is $O(n^l)$ there exists a constant $b_1 \geq b_0$ such that for n large enough,

$$\boxed{\mathbf{u}_n(L) \leq b_1\, n^l.}$$

By Theorem III.2.12, $\lim_{n\to\infty}\frac{\mathbf{v}_n(L)}{n^{l+1}} = a > 0$. Consequently, for any constant $K > a$, there exists n_K such that

$$\boxed{\forall n \geq n_K,\ \mathbf{v}_n(L) \leq K\, n^{l+1}}$$

and for any constant $J < a$, there exists n_J such that

$$\boxed{\forall n \geq n_J,\ \mathbf{v}_n(L) \geq J\, n^{l+1}.}$$

3. Multiplication for exact polynomial languages

Here, we study the multiplication by a constant for abstract numeration systems built on regular languages L with a polynomial complexity function. This step contains the main ideas for the discussion of arbitrary polynomial languages (i.e., languages with complexity function bounded by a polynomial).

LEMMA III.3.1. *Let $f : \mathbb{N} \to \mathbb{N}$ be a strictly increasing function such that $f(\mathbb{N})$ is a finite union of arithmetic progressions, i.e., there exist $y_0 \in f(\mathbb{N})$ and $\Gamma \in \mathbb{N} \setminus \{0\}$ such that for all $y \geq y_0$,*

$$y \in f(\mathbb{N}) \Leftrightarrow y + \Gamma \in f(\mathbb{N}).$$

Let $k = f^{-1}(y_0 + \Gamma) - f^{-1}(y_0)$. Then for all $x \geq f^{-1}(y_0)$, $n \in \mathbb{N}$,

$$f(x + nk) = f(x) + n\Gamma.$$

Proof. Let $x_0 = f^{-1}(y_0)$. We have by definition of k,

$$f(x_0 + k) = f(x_0) + \Gamma.$$

It is sufficient to show that if $x \geq x_0$ then

$$f(x + k) = f(x) + \Gamma \Rightarrow f(x + k + 1) = f(x + 1) + \Gamma.$$

Since f is strictly increasing,

$$f(x + k + 1) > f(x + k) = f(x) + \Gamma.$$

Since $f(\mathbb{N})$ is ultimately periodic of period Γ, there exists $v \geq x_0$ such that $f(v) = f(x + k + 1) - \Gamma > f(x)$. Then $v \geq x + 1$. There exists $u \in \mathbb{N}$ such that $f(u) = f(x + 1) + \Gamma > f(x) + \Gamma = f(x + k)$.

Now, let us assume that $v > x + 1$. Therefore $f(v) > f(x + 1)$ and

$$f(x + k + 1) = f(v) + \Gamma > f(x + 1) + \Gamma = f(u) > f(x + k).$$

So we have $x+k+1 > u > x+k$ which is a contradiction and $v = x+1$. □

REMARK III.3.2. Lemma III.3.1 will be used in the proofs of Theorem III.3.4 and Theorem III.4.1 with the same scheme. We take a well-chosen S-recognizable set X and let $Y = \lambda X = \{y_1, y_2, \ldots\}$. The choice of the set X makes sure that $|\mathrm{rep}_S(y_i)| < |\mathrm{rep}_S(y_{i+1})|$. To apply Lemma III.3.1, we assume that Y is S-recognizable. Therefore, $|\mathrm{rep}_S(Y)|$ has to be ultimately periodic and we obtain a contradiction.

The next lemma will be useful when applied to a complexity function.

LEMMA III.3.3. *If H is a polynomial such that*

$$\forall x \in \mathbb{N} \setminus \{0\},\ H(x) \in \mathbb{Z}$$

then $H(\mathbb{Z}) \subseteq \mathbb{Z}$.

Proof. We proceed by induction on the degree of H. If H is a polynomial of degree one then we have $H(x) = a\,x + b$ with $a, b \in \mathbb{Z}$ and $H(\mathbb{Z}) \subseteq \mathbb{Z}$.

Assume that the result holds for polynomials of degree $k \geq 1$. If H is a polynomial of degree $k+1$, then $R(x) = H(x+1) - H(x)$ is a polynomial of degree k and $R(\mathbb{N}) \subset \mathbb{Z}$. Therefore $R(\mathbb{Z}) \subseteq \mathbb{Z}$ and $H(0) = H(1) - R(0) \in \mathbb{Z}$. We can conclude by induction on $x < 0$ because $H(x) = H(x+1) - R(x)$.

□

THEOREM III.3.4. *Let $L \subset \Sigma^*$ be a regular language such that*

$$\mathbf{u}_n(L) = \begin{cases} a_l\, n^l + \cdots + a_1\, n + a_0 & \text{, if } n > 0; \\ 1 & \text{, otherwise} \end{cases}$$

where the a_i's belong to $\mathbb{Q}$ and $a_l > 0$. Let $\prec$ be an ordering of the alphabet Σ and $S = (L, \Sigma, \prec)$ be the corresponding numeration system.

If $\lambda \in \mathbb{N} \setminus \{n^{l+1} : n \in \mathbb{N}\}$, then there exists a subset X of $\mathbb{N}$ such that $\mathrm{rep}_S(X)$ is regular and that $\mathrm{rep}_S(\lambda\, X)$ is not.

REMARK III.3.5. To avoid any misunderstanding in the proof of Theorem III.3.4, we use the notation

$$\mathbf{u}_L(n) : \mathbb{N} \to \mathbb{N} : n \mapsto \#(L \cap \Sigma^n)$$

to denote the complexity function of L. This notation turns out to be easier than $\mathbf{u}_n(L)$ where n is only an index when we need a function with n as an argument.

Proof. One can build a polynomial $P \in \mathbb{Q}[x]$ of degree $l+1$ such that $P(0) = 0$ and for all $n \geq 1$, $P(n+1) = P(n) + \mathbf{u}_L(n)$. This polynomial is some kind of "discrete primitive" of $\mathbf{u}_L(n)$. Indeed, let $P(x) = b_{l+1}\, x^{l+1} + \cdots + b_1\, x + b_0$. The conditions on P give the following triangular system

$$\begin{cases} a_l & = & b_{l+1}\,(l+1) \\ a_{l-1} & = & b_{l+1}\,(l+1)\,\frac{l}{2} + b_l\, l \\ & \vdots & \\ a_0 & = & b_{l+1} + \cdots + b_1 \\ b_0 & = & 0. \end{cases}$$

This polynomial P has some useful properties. We have the polynomial identity $P(x+1) = P(x) + \mathbf{u}_L(x)$ for $x \in \mathbb{N} \setminus \{0\}$. Then it holds for $x \in \mathbb{R}$ if we extend the definition of the function $\mathbf{u}_L$ to

$$\mathbf{u}_L : \mathbb{R} \to \mathbb{R} : x \mapsto a_l\, x^l + \cdots + a_0.$$

By Lemma III.3.3, $P(1) = \mathbf{u}_L(0) = a_0 \in \mathbb{Z}$. One shows by induction on $n \in \mathbb{N}$ that $P(n)$ (resp. $P(-n)$) is an integer since $\mathbf{u}_L(\mathbb{N}) \subset \mathbb{N}$ (resp. since $\mathbf{u}_L(\mathbb{Z}) \subset \mathbb{Z}$ by Lemma III.3.3).

Let $x \in \mathbb{N} \setminus \{0\}$, notice that

$$|\mathrm{rep}_S(x)| = n \Leftrightarrow x \in [P(n) - a_0 + 1, P(n+1) - a_0]. \tag{18}$$

Indeed, an integer x has a representation of length n if $\mathbf{v}_{n-1} \le x < \mathbf{v}_n$ (see (6) page 15) and

$$\mathbf{v}_n = 1 + \sum_{i=1}^{n} \mathbf{u}_L(i) = 1 + \sum_{i=1}^{n} [P(i+1) - P(i)] = P(n+1) - P(1) + 1.$$

Notice that $\mathrm{rep}_S(P(\mathbb{N}))$ is a translation of the set $\mathrm{Min}(L,<)$ of the first words of each length. Therefore, by Lemma II.3.1 and Proposition II.1.1, $X = P(\mathbb{N})$ is S-recognizable.

Let $\lambda \in \mathbb{N} \setminus \{n^{l+1} : n \in \mathbb{N}\}$. Our aim is to show that $\lambda\, P(\mathbb{N})$ is not S-recognizable.

• For n large enough, we first show that

$$n \le |\mathrm{rep}_S(\lambda\, P(n))| < \lfloor \lambda^{1/l} n \rfloor \le \lambda^{1/l} n.$$

The first inequality is obvious. In view of (18), to satisfy the second inequality, it suffices to check whether

$$\lambda\, P(n) < P(\lfloor \lambda^{1/l}\, n \rfloor) - a_0 + 1.$$

We can write $P(n)$ as $b_{l+1}\, n^{l+1} + Q(n)$ with $b_{l+1} > 0$ and Q being a polynomial of degree not exceeding l. Then,

$$\begin{aligned} & P(\lfloor \lambda^{1/l}\, n \rfloor) - \lambda\, P(n) - a_0 + 1 \\ = \; & b_{l+1}(\lfloor \lambda^{1/l}\, n \rfloor)^{l+1} - \lambda\, b_{l+1}\, n^{l+1} + Q(\lfloor \lambda^{1/l}\, n \rfloor) - \lambda\, Q(n) - a_0 + 1 \\ > \; & b_{l+1}((\lambda^{1/l}\, n - 1)^{l+1} - \lambda\, n^{l+1}) + O(n^l) \end{aligned} \tag{*}$$

because $\lambda^{1/l}\, n - \lfloor \lambda^{1/l}\, n \rfloor < 1$. The coefficient of n^{l+1} in (*) is

$$b_{l+1}\,(\lambda^{(l+1)/l} - \lambda) > 0.$$

So, there exists n_0 such that for all $n \ge n_0$, the expression (*) is strictly positive and $|\mathrm{rep}_S(\lambda\, P(n))| < \lambda^{1/l}\, n$.

• If n is sufficiently large, we show that

$$|\mathrm{rep}_S(\lambda\, P(n+1))| > |\mathrm{rep}_S(\lambda\, P(n))|.$$

Let $i = |\mathrm{rep}_S(\lambda\, P(n))|$. In view of (18), we have to check that

$$\lambda\, P(n+1) > P(i+1) - a_0.$$

By definition of P and by (18), one has

$$\lambda\, P(n+1) = \lambda\, P(n) + \lambda\, \mathbf{u}_L(n) > P(i) - a_0 + \lambda\, \mathbf{u}_L(n).$$

Therefore it is sufficient to check whether

$$P(i) - a_0 + \lambda\, \mathbf{u}_L(n) > P(i+1) - a_0$$

that occurs if and only if

$$\lambda\, \mathbf{u}_L(n) - \mathbf{u}_L(i) > 0$$

i.e., if and only if

$$a_l\,(\lambda n^l - i^l) + \cdots + a_k\,(\lambda n^k - i^k) + \cdots + a_0\,(\lambda - 1) > 0.$$

To check that this inequality holds, remember that $a_l > 0$ and by the previous point of this proof and the definition of i, there exists n_0 such that for $n \geq n_0$, $1 \leq \frac{i}{n} < \lambda^{1/l}$. Since $\lambda\, \mathbf{u}_L(n) - \mathbf{u}_L(i)$ can be written

$$n^l \left(\underbrace{a_l \left[\lambda - \left(\frac{i}{n}\right)^l \right]}_{>0} + \cdots + \underbrace{\frac{a_k}{n^{l-k}}}_{\to 0} \underbrace{\left[\lambda - \left(\frac{i}{n}\right)^k \right]}_{is\ bounded} + \cdots + \underbrace{\frac{a_0}{n^l}}_{\to 0} (\lambda - 1) \right),$$

it is clear that $\lim_{n\to\infty} \frac{\lambda\, \mathbf{u}_L(n) - \mathbf{u}_L(i)}{n^l} > 0$. Thus there exists $n_0' \geq n_0$ such that for all $n \geq n_0'$, $|\mathrm{rep}_S(\lambda\, P(n+1))| > |\mathrm{rep}_S(\lambda\, P(n))|$.

• Assume that $\mathrm{rep}_S(\lambda\, P(\mathbb{N}))$ is regular. The set $|\mathrm{rep}_S(\lambda\, P(\mathbb{N}))|$ is a finite union of arithmetic progressions. We may apply Lemma III.3.1 because the function $|\mathrm{rep}_S(\lambda\, P(.))|$ is strictly increasing for $n \geq n_0'$. Thus there exist l_0 and $\Gamma \in \mathbb{N}\setminus\{0\}$ (depending on λ) such that $\forall l \geq l_0$,

$$l \in |\mathrm{rep}_S(\lambda\, P(\mathbb{N}))| \Leftrightarrow l + \Gamma \in |\mathrm{rep}_S(\lambda\, P(\mathbb{N}))|.$$

Let $n_1 \geq n_0'$ be such that $|\mathrm{rep}_S(\lambda\, P(n_1))| > l_0$. By Lemma III.3.1, there exists $k \in \mathbb{N}\setminus\{0\}$ (depending on λ) such that for all $n \geq n_1$ and for all $\alpha \in \mathbb{N}$,

$$|\mathrm{rep}_S(\lambda\, P(n + \alpha k))| = |\mathrm{rep}_S(\lambda\, P(n))| + \alpha\Gamma.$$

Let $i = |\mathrm{rep}_S(\lambda\, P(n))|$. In view of (18), one has

$$P(i + \alpha\Gamma) - a_0 + 1 \leq \lambda\, P(n + \alpha k) \leq P(i + \alpha\Gamma + 1) - a_0.$$

If one considers the l.h.s. inequality, $\lambda\, P(n+\alpha k) - P(i + \alpha\Gamma) + a_0 - 1$ must be non-negative for all $\alpha \in \mathbb{N}$. Consequently, the coefficient of the greatest power of α, α^{l+1}, appearing in this polynomial expression in α must be non-negative. This coefficient is

$$b_{l+1}\,(\lambda\, k^{l+1} - \Gamma^{l+1})$$

Notice that this latter coefficient vanishes only if $\lambda = \left(\frac{\Gamma}{k}\right)^{l+1}$. By hypothesis, this case is excluded. Indeed, suppose on the contrary that $\lambda = \left(\frac{\Gamma}{k}\right)^{l+1}$. If $\frac{\Gamma}{k} \in \mathbb{N}$ then $\lambda \in \{n^{l+1} : n \in \mathbb{N}\}$ and this case has to be excluded. Otherwise, $\frac{\Gamma}{k} \in \mathbb{Q}\setminus\mathbb{N}$ and thus $\lambda = (\frac{\Gamma}{k})^{l+1} \notin \mathbb{N}$, which is also impossible, since we consider multiplication by a non-negative integer. So we have the condition

$$k > \frac{\Gamma}{\lambda^{1/(l+1)}}.$$

But $\lambda\, P(n+\alpha k) - P(i+\alpha\Gamma+1)+a_0 \le 0$ for all $\alpha\in\mathbb{N}$. The coefficient of the greatest power of α is also $b_{l+1}\,(\lambda\, k^{l+1}-\Gamma^{l+1})$ and must be strictly negative. Then we have simultaneously the condition

$$k<\frac{\Gamma}{\lambda^{1/(l+1)}},$$

which leads to a contradiction and $\mathrm{rep}_S(\lambda P(\mathbb{N}))$ is not regular.

□

In Theorem III.3.4, we have underlined a recognizable set $X = P(\mathbb{N})$ such that $|\mathrm{rep}_S(\lambda\, P(\mathbb{N}))|$ is not a finite union of arithmetic progressions. When we consider the case $\lambda=\beta^{l+1}$, $\beta\in\mathbb{N}\setminus\{0,1\}$, we cannot find easily a subset X which is recognizable and such that $\lambda\, X$ is not. The next proposition shows that $|\mathrm{rep}_S(\beta^{l+1}P(\mathbb{N}))|$ is a finite union of arithmetic progressions if the complexity function of L is a polynomial of degree l. The author would like to thanks P. Mathonet for his help in the development of the following proof.

Proposition III.3.6. *With the assumptions and notations of Theorem III.3.4, there exists $C\in\mathbb{Z}$ such that for n large enough,*

$$|\mathrm{rep}_S(\beta^{l+1}P(n))| = \beta\, n + C. \tag{19}$$

Proof. In the proof of Theorem III.3.4, we have introduced a polynomial $P(x)=b_{l+1}\,x^{l+1}+\cdots+b_1\,x$ such that $P(n+1)-P(n)=\mathbf{u}_L(n)$. In view of (18), to satisfy (19), we have to find an integer C such that for n large enough

$$P(\beta\, n + C+1) - a_0 - \beta^{l+1}\,P(n) \;\ge\; 0 \tag{20}$$

$$\beta^{l+1}\,P(n) - P(\beta\, n+C) + a_0 - 1 \;\ge\; 0. \tag{21}$$

The coefficient of n^{l+1} vanishes in (20) and (21). The coefficient of n^l in (20) is $\beta^l\,[a_l\,(C+1)+b_l\,(1-\beta)]$ with $a_l=b_{l+1}\,(l+1)$. It is strictly increasing with C and equals zero for

$$C=C_1:=\frac{b_l\,(\beta-1)-a_l}{a_l}.$$

The coefficient of n^l in (21) is $-\beta^l\,[a_l\,C+b_l\,(1-\beta)]$. It is strictly decreasing with C and equals zero for $C=C_2:=C_1+1$. This situation is explained in Figure III.4.

If C_1 and C_2 are not integers then there exists $C\in]C_1,C_2[\cap\mathbb{Z}$ such that the coefficients of terms of maximal degree are both strictly positive (and the two inequalities are satisfied for n large enough).

Otherwise, one has to consider the integer case $C=C_1$ or $C=C_2$ (it is obvious that any other C leads to a strictly negative expression for (20) or (21)). Moreover, if $C=C_1$ (resp. $C=C_2$) then (21) (resp. (20)) is satisfied for n large enough.

Notice that for $i=1,\ldots,l-1$ the coefficient of n^i in (20) with $C=C_1$ is the opposite of the coefficient of n^i in (21) with $C=C_2$ since $C_2=C_1+1$. Notice also that the independent term in (20) for

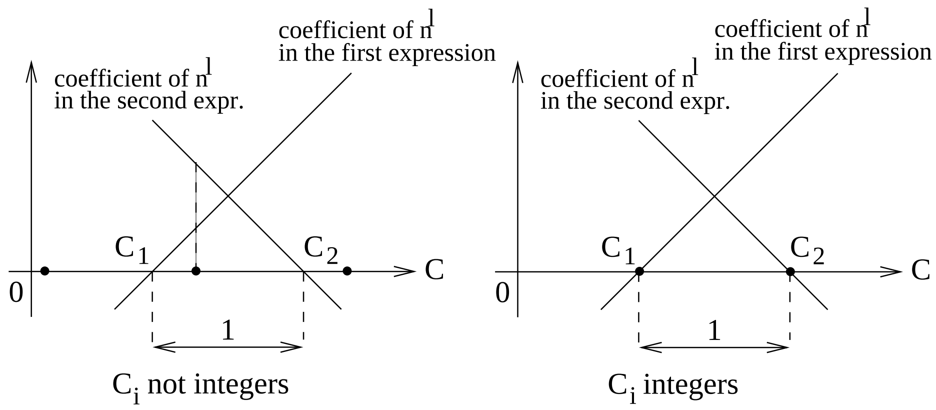

FIGURE III.4. The coefficients of n^l in (20) and (21)

$C = C_1$ is $P(C_2) - a_0$. In (21) for $C = C_2$ this term is $-P(C_2) + a_0 - 1$. Thus we can write (20) with $C = C_1$ as

$$A_{l-1}\, n^{l-1} + \cdots + A_1\, n + P(C_2) - a_0$$

and (21) with $C = C_2$ as

$$-A_{l-1}\, n^{l-1} - \cdots - A_1\, n - P(C_2) + a_0 - 1.$$

If there exists i such that $A_i \neq 0$ then let $j = \max_{A_i \neq 0} i$. If $A_j > 0$ (resp. $A_j < 0$) then one takes $C = C_1$ (resp. $C = C_2$).

Now, let us assume that $A_i = 0$ for $i = 1, \ldots, l-1$. If $P(C_2) - a_0 \geq 0$ then one takes $C = C_1$. Otherwise, $-P(C_2) + a_0$ is a strictly positive integer (remember the properties of P obtained in the proof of Theorem III.3.4). Therefore $-P(C_2) + a_0 - 1 \geq 0$ and one takes $C = C_2$.

□

4. Multiplication for arbitrary polynomial languages

Thanks to the material developed in the two previous sections, we obtain the generalization of Theorem III.3.4 for an arbitrary polynomial regular language.

THEOREM III.4.1. *Let $L \subset \Sigma^*$ be a regular language such that $\mathbf{u}_n(L)$ is $\Theta(n^l)$ for some integer l. If $\lambda \in \mathbb{N} \setminus \{n^{l+1} : n \in \mathbb{N}\}$, then there exists a subset X of $\mathbb{N}$ such that $\mathrm{rep}_S(X)$ is regular and that $\mathrm{rep}_S(\lambda X)$ is not. In other words, multiplication by a constant λ conserves recognizability only if λ is of the form n^{l+1}, for some $n \in \mathbb{N}$.*

Proof. To simplify the notations, we write $\mathbf{u}_n$ and $\mathbf{v}_n$ instead of $\mathbf{u}_n(L)$ and $\mathbf{v}_n(L)$ since we are not interested in the complexity functions of the different states but only in the complexity of L itself.

• *(i) Sketch of the proof.* We have the sequence n_i and the constant l depending only on the regular language L. We can also fix two constants J and K such that the last two inequalities of Remark III.2.15

are satisfied for n large enough. In Theorem III.3.4, the first step was to show that

$$n \leq |\mathrm{rep}_S(\lambda\, P(n))| < \lambda^{1/l} n.$$

Here we will show that, if $\lambda > (\frac{K}{J})^l$ then

$$n+1 \leq |\mathrm{rep}_S(\lambda\, \mathbf{v}_n)| < \lambda^{1/l}\, n + C.$$

The second step was to show that

$$|\mathrm{rep}_S(\lambda\, P(n+1))| > |\mathrm{rep}_S(\lambda\, P(n))|.$$

Here, since $\mathbf{u}_n$ periodically vanishes, we consider the subsequence $\mathbf{v}_{n_i}$ and show that the function

$$i \mapsto |\mathrm{rep}_S(\lambda\, \mathbf{v}_{n_{si}-1})|$$

is strictly increasing for i large enough and for some suitably chosen s. In the third step, we assume that $\mathrm{rep}_S(\lambda\{\mathbf{v}_{n_{si}-1} : i \in \mathbb{N}\})$ is regular and, as in Theorem III.3.4, use Lemma III.3.1 to obtain a contradiction thanks to the set of lengths. To conclude the proof, a last step is necessary to get rid of the assumption $\lambda > (\frac{K}{J})^l$. This can be achieved through the use of Theorem III.2.12.

• *(ii) Preliminaries.* As a consequence of Lemma III.2.8, it is clear that for n sufficiently large, $n+1 \leq |\mathrm{rep}_S(\mathbf{v}_n)| \leq n+C+1$ since for C consecutive values of $\mathbf{u}_n$ at least one of them does not vanish. (Recall that $|\mathrm{rep}_S(x)| = n$ iff $\mathbf{v}_{n-1} \leq x < \mathbf{v}_n$ (see (6) page 15) and therefore if $\mathbf{u}_n > 0$ for all n, then $|\mathrm{rep}_S(\mathbf{v}_n)| = n+1$ and the situation is simpler to handle.)

• *(iii)* Assume that the integer constant λ is strictly greater than $\left(\frac{K}{J}\right)^l$. We show that for n large enough,

$$n+1 \leq |\mathrm{rep}_S(\lambda\, \mathbf{v}_n)| \leq \lceil \lambda^{1/l}\, n \rceil + C - 1 < \lambda^{1/l}\, n + C. \tag{22}$$

It is sufficient to show that $\lambda\, \mathbf{v}_n < \mathbf{v}_{\lceil \lambda^{1/l} n \rceil + C - 1}$. For n large enough, $\mathbf{v}_{\lceil \lambda^{1/l} n \rceil} \geq J\, (\lceil \lambda^{1/l}\, n \rceil)^{l+1} \geq J\, (\lambda^{1/l}\, n)^{l+1}$ (see Remark III.2.15). Moreover the function $n \mapsto \mathbf{v}_n$ is increasing. So,

$$\mathbf{v}_{\lceil \lambda^{1/l} n \rceil + C - 1} \geq J\, \lambda^{\frac{l+1}{l}}\, n^{l+1}.$$

Moreover, for n large enough, $\lambda\, \mathbf{v}_n \leq \lambda\, K\, n^{l+1}$ (see Remark III.2.15). By the choice of λ, it is clear that $\lambda\, K\, n^{l+1} < J\, \lambda^{\frac{l+1}{l}}\, n^{l+1}$.

• *(iv)* Let $s \in \mathbb{N}\setminus\{0\}$ such that $s\, b_0 > b_1$ where b_0 and b_1 are the constants related to $\mathbf{u}_n$ given in Remark III.2.15. Let $a = \lim_{n\to\infty} \frac{\mathbf{v}_n(L)}{n^{l+1}}$. We show that for any J, K such that $J < a < K$ and for any $\lambda > \left(\frac{K}{J}\right)^l$, the function

$$i \mapsto |\mathrm{rep}_S(\lambda\, \mathbf{v}_{n_{si}-1})|$$

is strictly increasing for i sufficiently large. So we have to show that[1]

$$|\mathrm{rep}_S(\lambda\,\mathbf{v}_{n_{s(i+1)}-1})| = |\mathrm{rep}_S(\lambda\,\mathbf{v}_{n_{si}+sC-1})| > |\mathrm{rep}_S(\lambda\,\mathbf{v}_{n_{si}-1})|.$$

Let $g = |\mathrm{rep}_S(\lambda\,\mathbf{v}_{n_{si}-1})|$ then $\mathbf{v}_{g-1} \le \lambda\,\mathbf{v}_{n_{si}-1} < \mathbf{v}_g$ and we must show that

$$\lambda\,\mathbf{v}_{n_{si}+sC-1} = \lambda\,\mathbf{v}_{n_{si}-1} + \lambda \sum_{j=0}^{sC-1} \mathbf{u}_{n_{si}+j} \ge \mathbf{v}_g = \mathbf{v}_{g-1} + \mathbf{u}_g.$$

So, it is sufficient to show that $\lambda \sum_{j=0}^{sC-1} \mathbf{u}_{n_{si}+j} \ge \mathbf{u}_g$. In view of (22), we have $g < \lambda^{1/l}(n_{si}-1)+C$. Therefore $\mathbf{u}_g < b_1\,[\lambda^{1/l}(n_{si}-1)+C]^l$. On the other hand,

$$\lambda \sum_{j=0}^{sC-1} \mathbf{u}_{n_{si}+j} \ge \lambda \sum_{j=0}^{s-1} \underbrace{\mathbf{u}_{n_{si}+j\,C}}_{\ge b_0\,(n_{si}+j\,C)^l} \ge \lambda\, b_0\, s\, n_{si}^l.$$

To conclude this part, notice that the coefficient of n_{si}^l in

$$b_1\,[\lambda^{1/l}(n_{si}-1)+C]^l$$

is $b_1\,\lambda$ and by the choice of s, we have $b_1\,\lambda < \lambda\, b_0\, s$. So the inequality holds for i sufficiently large.

- *(v)* Consider the subset

$$X = \{\mathbf{v}_{n_{si}-1} : i \in \mathbb{N}\} = \{\mathbf{v}_{n_0+siC-1} : i \in \mathbb{N}\}.$$

By Lemma III.2.8, $\mathbf{u}_{n_0+siC} > 0$. Thus $\mathrm{rep}_S(\mathbf{v}_{n_0+siC-1})$ is the first word of length n_0+siC and

$$\begin{aligned}\mathrm{rep}_S(X) &= \mathrm{rep}_S\left(\{\mathbf{v}_n : n \in \mathbb{N}\}\right) \cap \Sigma^{n_0}\left(\Sigma^{sC}\right)^* \\ &= \mathrm{Min}(L,<) \cap \Sigma^{n_0}\left(\Sigma^{sC}\right)^*\end{aligned}$$

is regular and X is an S-recognizable subset of $\mathbb{N}$ (by Lemma II.3.1). Assume that $\lambda\,X$ is S-recognizable. Therefore, $|\mathrm{rep}_S(\lambda\,X)|$ is a finite union of arithmetic progressions. In view of *(iv)*, we can apply Lemma III.3.1 and obtain two integer constants Γ and k (depending on λ) such that for all $\alpha \in \mathbb{N}$ and for i large enough

$$|\mathrm{rep}_S(\lambda\,\mathbf{v}_{n_0+sC(i+\alpha\,k)-1})| = |\mathrm{rep}_S(\lambda\,\mathbf{v}_{n_0+sCi-1})| + \alpha\,\Gamma.$$

Or in an equivalent manner, if we set $z = |\mathrm{rep}_S(\lambda\,\mathbf{v}_{n_0+sCi-1})|$ then

$$\mathbf{v}_{z+\alpha\Gamma-1} \le \lambda\,\mathbf{v}_{n_0+sC(i+\alpha\,k)-1} < \mathbf{v}_{z+\alpha\Gamma}. \tag{23}$$

First consider the left inequality in (23). For i large enough,

$$\mathbf{v}_{z+\alpha\Gamma-1} \ge J\,(z+\alpha\,\Gamma-1)^{l+1}.$$

On the other hand,

$$\lambda\,\mathbf{v}_{n_0+sC(i+\alpha\,k)-1} \le \lambda\,K\,(n_0+sCi+sCk\,\alpha-1)^{l+1}.$$

[1]Recall that n_i is ultimately periodic, hence $n_{s(i+1)}-1 = n_{si}+sC-1$.

Since α can be arbitrary large, we focus on the terms of the form α^{l+1} and we obtain the condition $J\,\Gamma^{l+1} \leq \lambda\, K\,(sCk)^{l+1}$, i.e.,

$$\lambda \geq \frac{J}{K}\left(\frac{\Gamma}{sCk}\right)^{l+1}. \tag{24}$$

If we consider the right inequality in (23), then we have, using the inequalities in Remark III.2.15,

$$\mathbf{v}_{z+\alpha\,\Gamma} \leq K\,(z+\alpha\,\Gamma)^{l+1}$$

and also

$$\lambda\,\mathbf{v}_{n_0+sC(i+\alpha\,k)-1} \geq \lambda\,J\,(n_0+sCi+sCk\,\alpha-1)^{l+1}.$$

If we focus on the terms of the form α^{l+1}, we obtain

$$\lambda \leq \frac{K}{J}\left(\frac{\Gamma}{sCk}\right)^{l+1}. \tag{25}$$

• *(vi)* By Theorem III.2.12, the sequence $(\frac{\mathbf{v}_n}{n^{l+1}})_{n\in\mathbb{N}}$ converges to a strictly positive limit denoted by a. We consider the sequences

$$K_m = a + \frac{1}{m} \text{ and } J_m = a - \frac{1}{m}.$$

Consequently, if m is given, then for n large enough $\mathbf{v}_n \leq K_m\,n^{l+1}$ and $\mathbf{v}_n \geq J_m\,n^{l+1}$ because $J_m < a < K_m$. So if we replace K by K_m and J by J_m, the previous points *(iii)*, *(iv)* and *(v)* remain true for i sufficiently large.

When m tends to infinity, the condition $\lambda > \left(\frac{K_m}{J_m}\right)^l$ given in *(iii)* is equivalent to $\lambda \geq 2$ because λ is an integer; from the conditions (24) and (25), we deduce that

$$\lambda = \left(\frac{\Gamma}{sCk}\right)^{l+1}$$

which contradicts the hypothesis (remember that Γ, s, C and k are integers) and therefore λX is not S-recognizable.

□

This latter theorem has a direct corollary. Its proof is the same as the one of Corollary III.1.4.

Corollary III.4.2. *Under the assumptions of Theorem III.4.1, the addition is not a regular map.*

5. Complement of polynomial languages

The class of exponential languages splits into two subclasses according to the fact that the complement of a language is polynomial or not. In this section, we have a closer look at the numeration systems built over an exponential regular language such that its complement has a complexity function bounded by a polynomial. We show that for such systems, multiplication by a constant does not generally preserve recognizability. It is interesting to notice that the method used to check the non-recognizability of a set differs from the one encountered in the previous sections. We use here a so-called pumping lemma stated below (a proof of this result can be found in [**70**]).

LEMMA III.5.1. *Let $L \subset \Sigma^*$ be a regular language. There exists a constant k depending only on L such that for each $w \in L$, $|w| \geq k$, there exist $x, y, z \in \Sigma^*$ such that*

(1) $w = xyz$,
(2) $|xy| \leq k$, $|y| > 0$,
(3) *for all* $n \in \mathbb{N}$, $xy^n z \in L$.

We begin with the example of $L = \Sigma^* \setminus M$ where M is the polynomial language a^*b^* and $\Sigma = \{a, b\}$. Thus, with $S = (L, \{a, b\}, a < b)$, we compute the representations of $2\,\mathbf{v}_n(L)$ and obtain Table III.1.

In view of this table, it appears that the number of leading b's in the representation is increasing. Furthermore, it seems that the length of the tail also increases. Let us show that this observation is true and can be generalized. But first we set a useful notation.

DEFINITION III.5.2. For $L \subset \Sigma^*$ and $x \in \Sigma^*$, we set

$$L_x = \{w \in L : w = xy\}.$$

Any confusion with the notation L_p where p is a state of a DFA is cleared from the context. Notice that if $\sigma \in \Sigma$ then $L_{\sigma^{i+1}} \subset L_{\sigma^i}$, $i \in \mathbb{N}$.

It is clear that $L_x \subseteq L$. So $\mathbf{u}_n(L_x) \leq \mathbf{u}_n(L)$ and $\mathbf{u}_n(L_x)$ is $O(n^l)$ whenever $\mathbf{u}_n(L)$ is $O(n^l)$.

In our example, for $0 \leq k < n$, we have

$$\mathbf{u}_n(L_{b^{n-k}}) = \mathbf{u}_n(\Sigma^*_{b^{n-k}}) - \mathbf{u}_n(M_{b^{n-k}}) = 2^k - 1.$$

The complexity function $\mathbf{u}_n(L) = \mathbf{u}_n(\Sigma^* \setminus M)$ of the language L associated to the system S is $2^n - n - 1$. So

$$\mathbf{v}_n(L) = \sum_{i=0}^{n} \mathbf{u}_i(L) = 2^{n+1} - \frac{n\,(n+3)}{2} - 2.$$

The words of $\mathrm{rep}_S(\{\mathbf{v}_n(L) : n \in \mathbb{N}\})$ are the first words of each length in L. So $\{\mathbf{v}_n(L) : n \in \mathbb{N}\}$ is S-recognizable. Let us show that $\{2\,\mathbf{v}_n(L) : n \in \mathbb{N}\}$ is not S-recognizable.

Recall that

$$|\mathrm{rep}_S(x)| = n \Leftrightarrow \mathbf{v}_{n-1}(L) \leq x < \mathbf{v}_n(L).$$

n	$2\,\mathbf{v}_n(L)$	$\mathrm{rep}_S(2\,\mathbf{v}_n(L)) = b^k aw$	k	$\lvert w\rvert$
1	0	ba	1	0
2	2	baa	1	1
3	10	$baab$	1	2
4	32	$babab$	1	3
5	84	$bbaaaa$	2	3
6	198	$bbababa$	2	4
7	438	$bbbaaabb$	3	4
8	932	$bbbabbabb$	3	5
9	1936	$bbbbabaaba$	4	5
10	3962	$bbbbbaabaaa$	5	5
11	8034	$bbbbbabbbbab$	5	6
12	16200	$bbbbbbabbaaab$	6	6
13	32556	$bbbbbbbabaabaa$	7	6
14	65294	$bbbbbbbbaababba$	8	6
15	130798	$bbbbbbbbbaaaabbb$	9	6
16	261836	$bbbbbbbbbabbbabbb$	9	7
17	523944	$bbbbbbbbbbabbaabba$	10	7
18	1048194	$bbbbbbbbbbbabababaa$	11	7
19	2096730	$bbbbbbbbbbbbabaaaaab$	12	7
20	4193840	$bbbbbbbbbbbbbaababbab$	13	7
21	8388100	$bbbbbbbbbbbbbbaaabbaaa$	14	7
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

TABLE III.1. The first $2\,\mathbf{v}_n$ in $\{a,b\}^* \setminus a^*b^*$.

Since $\mathbf{v}_{n+1}(L) - 2\,\mathbf{v}_n(L) = n(n+1)/2$, it is obvious that for n large enough, $\mathbf{v}_n(L) \le 2\,\mathbf{v}_n(L) < \mathbf{v}_{n+1}(L)$. Thus $|\mathrm{rep}_S(2\,\mathbf{v}_n(L))| = n+1$.

For each n there exists a unique i (depending on n) such that[2]

$$\mathbf{u}_{n+1}(L_{b^{n+2-i}}) = 2^{i-1}-1 < \underbrace{\mathbf{v}_{n+1}(L) - 2\,\mathbf{v}_n(L)}_{=n(n+1)/2} \le 2^i - 1 = \mathbf{u}_{n+1}(L_{b^{n+1-i}}).$$

Then $\mathrm{rep}_S(2\,\mathbf{v}_n(L)) = b^{n+1-i}az$ with $|z| = i-1$. This phenomenon can be enlightened by the following example given in Table III.2 where $n = i = 3$ and the representation of $2\,\mathbf{v}_3(L)$ begins with exactly one b. As a function of n, i is increasing but

$$\log(\frac{n(n+1)}{2}+1) \le i < \log(\frac{n(n+1)}{2}+1)+1.$$

So

$$\lim_{n\to\infty} n - i = +\infty.$$

[2]Since $2\,\mathbf{v}_n(L)$ is represented by a word of length $n+1$, we compare this representation with the first word of length $n+2$ and of numerical value $\mathbf{v}_{n+1}(L)$ to obtain the number of leading b's in $\mathrm{rep}_S(2\,\mathbf{v}_n(L))$ as shown in Table III.2.

5	$aaba$			
	$abaa$			
	$abab$			
	$abba$			
	$baaa$			$\mathbf{u}_4(L_b) = 7$
$2\,\mathbf{v}_3(L) = 10$	$baab$			$\downarrow$
	$baba$			
	$babb$			
	$bbaa$		$\mathbf{u}_4(L_{b^2}) = 3$	
	$bbab$		$\downarrow$	
15	$bbba$	$\mathbf{u}_4(L_{b^3}) = 1$	$\uparrow$	$\uparrow$
$\mathbf{v}_4(L) = 16$	$aaaba$			
$\vdots$	$\vdots$			

TABLE III.2. Structure of $\Sigma^* \setminus a^*b^*$.

Assume that $\mathcal{L} = \text{rep}_S(\{2\,\mathbf{v}_n : n \in \mathbb{N}\})$ is accepted by an automaton with q states. There exist n_0, i_0 and $t \geq 0$ such that $\text{rep}_S(2\,\mathbf{v}_{n_0}(L)) = b^{q+t}az_0$ with $|z_0| = i_0$. By the pumping lemma (Lemma III.5.1), there exists $\alpha > 0$ such that

$$\forall m \in \mathbb{N},\ b^{q+t+m\alpha}az_0 \in \mathcal{L}.$$

In this last expression, z_0 has a constant length i_0 independent of m, which is a contradiction.

In view of this example, we state the following theorem.

THEOREM III.5.3. *Let $\Sigma = \{\sigma_1 < \cdots < \sigma_{t-1} < \tau\}$, $t \geq 2$ and $L \subset \Sigma^*$ be a regular language such that $\mathbf{u}_n(L)$ is $\Theta(n^l)$. Let $S = (\Sigma^* \setminus L, \Sigma, <)$. There exists an S-recognizable set $X \subset \mathbb{N}$ such that for all $j \geq 1$, $t^j X$ is not S-recognizable.*

Proof. For $0 \leq k < n$, we have

$$\mathbf{u}_n((\Sigma^* \setminus L)_{\tau^{n-k}}) = \mathbf{u}_n(\Sigma^*_{\tau^{n-k}}) - \mathbf{u}_n(L_{\tau^{n-k}}) = t^k - \underbrace{\mathbf{u}_n(L_{\tau^{n-k}})}_{\in O(n^l)}.$$

To avoid any misunderstanding, $\mathbf{v}_n$ is the sequence associated to the language $\Sigma^* \setminus L$ of the numeration S and $\mathbf{v}_n(L)$ is related to L. So, $\mathbf{v}_n(L) = \sum_{i=0}^{n} \mathbf{u}_i(L)$ and

$$\mathbf{v}_n = \sum_{i=0}^{n} \mathbf{u}_i(\Sigma^* \setminus L) = \frac{t^{n+1} - 1}{t - 1} - \mathbf{v}_n(L).$$

We take $X = \text{rep}_S(\{\mathbf{v}_n : n \in \mathbb{N}\})$, as a recognizable set. Let $j \geq 1$. We have for n sufficiently large,

$$\mathbf{v}_{n+j-1} \leq t^j \mathbf{v}_n < \mathbf{v}_{n+j}.$$

Indeed, $\mathbf{v}_{n+j} - t^j\,\mathbf{v}_n = t^j\,\mathbf{v}_n(L) - \mathbf{v}_{n+j}(L) + \frac{t^j-1}{t-1}$. By Theorem III.2.12, there exists $a > 0$ such that[3] $\mathbf{v}_n(L) \sim a\,n^{l+1}$. So, we have

$$\mathbf{v}_{n+j} - t^j\,\mathbf{v}_n \sim (t^j - 1)\,a\,n^{l+1}.$$

On the other hand,

$$t^j\,\mathbf{v}_n - \mathbf{v}_{n+j-1} = t^{n+j} + \mathbf{v}_{n+j-1}(L) - t^j\,\mathbf{v}_n(L) - \frac{t^j-1}{t-1}$$

has an exponential dominant term. Then $|\mathrm{rep}_S(t^j\,\mathbf{v}_n)| = n + j$.

For all n sufficiently large, there exists a unique i (depending on n) such that

$$(26)\qquad \underbrace{\mathbf{u}_{n+j}((\Sigma^*\setminus L)_{\tau^{n+j-i+1}})}_{=t^{i-1}-\mathbf{u}_{n+j}(L_{\tau^{n+j-i+1}})} < \mathbf{v}_{n+j} - t^j\,\mathbf{v}_n \le \underbrace{\mathbf{u}_{n+j}((\Sigma^*\setminus L)_{\tau^{n+j-i}})}_{=t^{i}-\mathbf{u}_{n+j}(L_{\tau^{n+j-i}})}$$

Then $\mathrm{rep}_S(t^j\,\mathbf{v}_n) = \tau^{n+j-i}\sigma z$ with $|z| = i-1$ and $\sigma \neq \tau$. Notice that as a function of n, i is increasing and not bounded. To show that $n-i \to +\infty$ if $n \to +\infty$, let us assume that $n-i$ is bounded and divide all members of (26) by t^n. Let $n \to +\infty$ and obtain a contradiction.

Suppose that $\mathrm{rep}_S(\{t^j X\})$ is accepted by an automaton with q states. There exist n_0, i_0 and $t \geq 0$ such that $\mathrm{rep}_S(t^j\mathbf{v}_{n_0}) = \tau^{q+t}\sigma z_0$ with $|z_0| = i_0$ and $\sigma \neq \tau$. Then using the pumping lemma (Lemma III.5.1), we obtain a contradiction.

□

6. Exponential languages with exponential complement

In this section, we give sufficient conditions to achieve the computation of a U-representation of an integer from its S-representation, where U is some positional numeration system related to a sequence of integers. In particular, we obtain sufficient conditions to guarantee the stability of the S-recognizability under addition and multiplication by a constant. These conditions are related to the normalization function.

PROPOSITION III.6.1. *Let $L \subset \Sigma^*$ be a regular language accepted by the DFA $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ and $S = (L, \Sigma, <)$ be an abstract numeration system. Let $U = (U_n)_{n\in\mathbb{N}}$ be a sequence of integers such that $U_0 = 1$. If there exist $k, \alpha \in \mathbb{N}\setminus\{0\}$, $e_{p,i} \in \mathbb{Z}$ ($p \in Q$, $i = 0, \ldots, k-1$) such that for all states $p \in Q$ and all $n \in \mathbb{N}$*

$$(27)\qquad \alpha\,\mathbf{u}_{n+k-1}(p) = \sum_{i=0}^{k-1} e_{p,i}\,U_{n+i}.$$

Then there exist a finite alphabet $B \subset \mathbb{Z}$ and a finite letter-to-letter 2-tape automaton which computes a function $g : L \to B^*$ such that $|w| = |g(w)|$ and

$$\alpha\,\mathrm{val}_S(w) = \pi_U(g(w)).$$

[3]Let f, g two functions, $f \sim g$ iff $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$.

In other words, $g(w)$ is a U-representation over B of $\mathrm{val}_S(w)$.

REMARK III.6.2. The function g of the previous proposition is injective. If v and w are two words of L such that $g(v) = g(w)$ then $\mathrm{val}_S(v) = \mathrm{val}_S(w)$. Thus the conclusion, since val_S is a one-to-one correspondence.

Proof. We consider words in L of length at least k. Indeed, there is only a finite number of words of length less than k and to take them into account, we simply have to do finite modifications to the transducer that we will obtain at the end of this proof. Let

$$w = w_{k+l} \dots w_{k-1} w_{k-2} \dots w_0$$

be a word in L of length $k + l + 1$ with $l \geq -1$. We apply on the first $l+2$ letters of w the formula (2) of Lemma I.5.2 with respect to $\mathcal{A}$ (see page 13) and we obtain the following expression for $\mathrm{val}_s(w)$,

$$\sum_{\sigma < w_{k+l}} \mathbf{u}_{k+l}(s.\sigma) + \sum_{i=-1}^{l} \mathbf{u}_{k+i}(s) + \sum_{i=-1}^{l-1} \sum_{\sigma < w_{k+i}} \mathbf{u}_{k+i}(s.w_{k+l} \dots w_{k+i+1}\sigma)$$
$$+\mathrm{val}_{s.w_{k+l} \dots w_{k-1}}(w_{k-2} \dots w_0) + \mathbf{v}_{k-2}(s) - \mathbf{v}_{k-2}(s.w_{k+l} \dots w_{k-1}).$$

In this latter expression, we have written $\sum_{i=-1}^{l} \mathbf{u}_{k+i}(s) + \mathbf{v}_{k-2}(s)$ instead of $\mathbf{v}_{k+l}(s)$. Let us also recall that the notation $p.\sigma$ is written instead of $\delta(p, \sigma)$. We will denote by C_w the sum of the last three terms. For all $q \in Q$, $p \in Q \setminus \{s\}$ and $\sigma \in \Sigma$, let us define

$$\beta_{q,p,\sigma} = \#\{\sigma' < \sigma : q.\sigma' = p\}$$

and

$$\beta_{q,s,\sigma} = 1 + \#\{\sigma' < \sigma : q.\sigma' = s\}.$$

With these notations, we can rewrite $\mathrm{val}_s(w)$ as

$$C_w + \sum_{p \in Q} \beta_{s,p,w_{k+l}} \, \mathbf{u}_{k+l}(p) + \sum_{i=-1}^{l-1} \sum_{p \in Q} \beta_{s.w_{k+l} \dots w_{k+i-1},p,w_{k+i}} \, \mathbf{u}_{k+i}(p).$$

Therefore, using (27) we have

$$\begin{aligned} \alpha \, \mathrm{val}_s(w) = \quad & \alpha \, C_w + \sum_{j=0}^{k-1} \sum_{p \in Q} \beta_{s,p,w_{k+l}} \, e_{p,j} \, U_{l+j+1} \\ & + \sum_{i=-1}^{l-1} \sum_{j=0}^{k-1} \sum_{p \in Q} \beta_{s.w_{k+l} \dots w_{k+i-1},p,w_{k+i}} \, e_{p,j} \, U_{i+j+1}. \end{aligned}$$

Set now for $q \in Q$, $\sigma \in \Sigma$ and $j = 0, \dots, k-1$,

$$\lambda_{q,\sigma,j} = \sum_{p \in Q} \beta_{q,p,\sigma} \, e_{p,j}. \tag{28}$$

It is obvious that these numbers $\lambda_{q,\sigma,j}$'s take their values in a finite set R. Therefore sums of k elements of R also take their values in a finite set, say T.

We are now able to build a finite letter-to-letter 2-tape automaton $\mathcal{M}$ over $\Sigma^* \times B^*$ with $B \subset \mathbb{Z}$ a finite alphabet. The formula expressing $\alpha\,\mathrm{val}_s(w)$ is given by

$$\boxed{\alpha\, C_w + \sum_{j=0}^{k-1} \lambda_{s,w_{k+l},j}\, U_{l+j+1} + \sum_{i=-1}^{l-1} \sum_{j=0}^{k-1} \lambda_{s.w_{k+l}\ldots w_{k+i-1},w_{k+i},j}\, U_{i+j+1}.}$$

and can be interpreted in the following way. The reading of w_{k+i}, $l \leq i \leq -1$, provides the decomposition of $\alpha\,\mathrm{val}_s(w)$ with

$$\lambda_{q,w_{k+i},k-1}\, U_{k+i};\ \lambda_{q,w_{k+i},k-2}\, U_{k+i-1};\ \ldots;\ \lambda_{q,w_{k+i},0}\, U_{i+1}$$

where $q = s$ if $i = l$ and $q = s.w_{k+l} \ldots w_{k+i-1}$, otherwise. The reading of w_{k+i} gives a coefficient $\lambda_{q,w_{k+i},k-1}$ for U_{k+i}. The other $k-1$ coefficients can be viewed as “carries”. Roughly speaking, if we have already read the word $t = w_{k+l} \ldots w_{k+i+1}$ and if we are reading $\sigma = w_{k+i}$, then the computation of the coefficients λ's of $U_{k+i}, \ldots, U_{i+1}$ depends only on the knowledge of σ and the state $s.t$. Therefore it seems natural to mimic $\mathcal{A}$ in $\mathcal{M}$.

Thereby we give a precise definition of $\mathcal{M}$. The set of states is $\mathcal{K} = Q \cup \{f\} \times \underbrace{T \times \cdots \times T}_{k-1}$ where f does not belong to Q and is the unique final state of $\mathcal{M}$. The copies of T will be used to store the $k-1$ “carries”. The start state is $(s, 0, \ldots, 0)$. The transition relation $\Delta : \mathcal{K} \times (\Sigma \times B) \to \mathcal{K}$ of $\mathcal{M}$ is defined as follows. If $p \in Q$, $\sigma \in \Sigma$,

$$\begin{aligned} &\Delta((p, \gamma_{k-2}, \ldots, \gamma_0), (\sigma, \lambda_{p,\sigma,k-1} + \gamma_{k-2})) \\ = \ &(p.\sigma; \lambda_{p,\sigma,k-2} + \gamma_{k-3}; \ldots; \lambda_{p,\sigma,1} + \gamma_0; \lambda_{p,\sigma,0}) \end{aligned}$$

These transitions compute an output $x_{k+l} \ldots x_{k-1}$ from $w_{k+l} \ldots w_{k-1}$. The alphabet B is finite since T is finite.

But we have still to read the last $k-1$ letters of w. For each state $p \in Q$, $D_p = L_p \cap \Sigma^{k-1}$ is finite (recall that L_p are the words accepted from p). So, for each state $p \in Q$ and each word $w_{k-2} \ldots w_0 \in D_p$, we construct an edge from $(p, \gamma_{k-2}, \ldots, \gamma_0)$ to f labeled by

$$(w_{k-2} \ldots w_0, \gamma_{k-2} \cdots \gamma_1(\gamma_0 + C_w)).$$

(This kind of edge can naturally be split in $k-1$ elementary edges using $k-2$ new states.) Indeed, notice that C_w is a constant which only depends on the reached state $s.w_{k+l} \ldots w_{k-1}$ (the first component in $\mathcal{K}$) and the remaining word $w_{k-2} \ldots w_0$.

□

Remark III.6.3. The complexity functions of languages accepted from the different states of the DFA $\mathcal{A}$ accepting L satisfy the same recurrence relation of degree l (they only differ by the initial conditions). A practical way of checking (27) is seeking a final state $f \in F$ such that

$$\det\begin{pmatrix} \mathbf{u}_0(f) & \cdots & \mathbf{u}_{l-1}(f) \\ \vdots & & \vdots \\ \mathbf{u}_{l-1}(f) & \cdots & \mathbf{u}_{2l-2}(f) \end{pmatrix} \neq 0.$$

If such an f exists then for all $p \in Q$, there exist $c_{p,i} \in \mathbb{Q}$ such that

$$\mathbf{u}_{n+l-1}(p) = \sum_{i=0}^{l-1} c_{p,i}\, \mathbf{u}_{n+i}(f)$$

and (27) can be easily obtained. In fact, the integer α of (27) can play the role of the least common multiple of the denominators appearing in the $c_{p,i}$'s.

Corollary III.6.4. *Let $S = (L, \Sigma, <)$ be an abstract numeration system and let the hypothesis and notations of Proposition III.6.1 be satisfied. If the sequence $U = (U_n)_{n\in\mathbb{N}}$ defines a positional numeration system such that the normalization function $\nu_{U,B}$ is computable by finite letter-to-letter 2-tape automaton then $X \subset \mathbb{N}$ is S-recognizable if and only if αX is U-recognizable.*

Proof. Let the regular language $\hat{g} \subset (\Sigma \times B)^* \cap (L \times B^*)$ be the graph of the function g defined in Proposition III.6.1. We denote by $p_1 : \Sigma \times B \to \Sigma$ and $p_2 : \Sigma \times B \to B$ the canonical homomorphisms of projection. Let

$$Y = p_2[p_1^{-1}(\mathrm{rep}_S(X)) \cap \hat{g}].$$

If X is S-recognizable then by Theorem I.1.7, $Y \subset B^*$ is regular and by Proposition III.6.1, $\pi_U(Y) = \alpha X$. So αX is U-recognizable since $\nu_{U,B}(Y)$ is computable by a finite letter-to-letter transducer.

Conversely, if $\rho_U(\alpha X)$ is regular then $\nu_{U,B}^{-1} \circ \rho_U(\alpha X)$ is also regular. For each $y \in \alpha X$, $\nu_{U,B}^{-1} \circ \rho_U(y)$ may contain more than one element, but only one is in $p_2(\hat{g})$. So the set

$$p_1\left(p_2^{-1}[\nu_{U,B}^{-1} \circ \rho_U(\alpha X)] \cap \hat{g}\right)$$

is regular and equal to $\mathrm{rep}_S(X)$.

□

Corollary III.6.5. *Let $S = (L, \Sigma, <)$ be an abstract numeration system and let the hypothesis and notations of Proposition III.6.1 be satisfied. If the sequence U satisfies a linear recurrence relation*

$$U_n = d_1 U_{n-1} + \cdots + d_m U_{n-m}, d_i \in \mathbb{Z}, d_m \neq 0, n \geq m$$

such that its characteristic polynomial is the minimal polynomial of a Pisot number then $X \subset \mathbb{N}$ is S-recognizable if and only if X is U-recognizable.

Proof. It is well known that for the system U the normalization $\nu_{U,C}$ is computable by a finite letter-to-letter 2-tape automaton for any finite alphabet $C \subset \mathbb{Z}$ (see Proposition I.2.6). So by the previous corollary,

X is S-recognizable if and only if αX is U-recognizable. Another well-known fact related to Pisot numeration systems is that a subset X is U-recognizable if and only if it is definable in the structure $\langle \mathbb{N}, +, V_U\rangle$ (see Theorem I.2.5). In particular, multiplication by a constant α is definable in $\langle \mathbb{N}, +\rangle$. So αX is definable in the structure if and only if X is definable. Indeed, if αX is definable by a formula φ then $X = \{x \in \mathbb{N} \,|\, (\exists y)(\varphi(y)) \wedge (\alpha\, x = y)\}$.

□

Example III.6.6. Consider the language L over $\{a, b, c\}$ of the words that do not contain aa. Its minimal automaton $\mathcal{A}_L$ is given in Figure III.5. This example can be viewed as some kind of *generalized Fibonacci system.* (In the Fibonacci system, one considers only a two letters alphabet.) The sequences associated to the different states

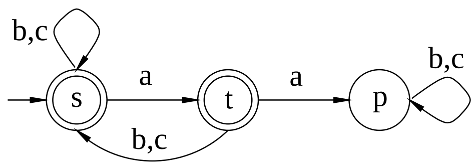

Figure III.5. The minimal automaton for the generalized 3-letters Fibonnaci system.

satisfy the relation

$$\mathbf{u}_{n+2} = 2\mathbf{u}_{n+1} + 2\mathbf{u}_n, \forall n \in \mathbb{N}$$

with the initial conditions $\mathbf{u}_0(s) = 1$, $\mathbf{u}_1(s) = 3$, $\mathbf{u}_0(t) = 1$, $\mathbf{u}_1(t) = 2$, $\mathbf{u}_0(p) = \mathbf{u}_1(p) = 0$. The sequence $U = (U_n)_{n\in\mathbb{N}}$ of Proposition III.6.1 is given by $(\mathbf{u}_n(s))_{n\in\mathbb{N}}$. For all $n \in \mathbb{N}$, we have the relations

$$\left\{\begin{array}{llllllll} \mathbf{u}_{n+1}(s) & = 1 & \mathbf{u}_{n+1}(s) & + & 0 & \mathbf{u}_n(s) & , & e_{s,0} = 0, \quad e_{s,1} = 1 \\ \mathbf{u}_{n+1}(t) & = 0 & \mathbf{u}_{n+1}(s) & + & 2 & \mathbf{u}_n(s) & , & e_{t,0} = 2, \quad e_{t,1} = 0 \\ \mathbf{u}_{n+1}(p) & = 0 & \mathbf{u}_{n+1}(s) & + & 0 & \mathbf{u}_n(s) & , & e_{p,0} = 0, \quad e_{p,1} = 0 \end{array}\right.$$

Notice that the characteristic polynomial of the recurrence relation of $\mathbf{u}_n(s)$ is

$$x^2 - 2x - 2 = (x - 1 + \sqrt{3})(x - 1 - \sqrt{3}).$$

So $U = (\mathbf{u}_n(s))_{n\in\mathbb{N}}$ is a positional numeration system associated to the Pisot number $1 + \sqrt{3}$. From $\mathcal{A}_L$, we compute the 3×3 matrices $B_\sigma = (\beta_{q,r,\sigma})_{q,r\in\{s,t,p\}}$, $\sigma \in \Sigma$:

$$B_a = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, B_b = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}, B_c = \begin{pmatrix} 2 & 1 & 0 \\ 2 & 0 & 0 \\ 1 & 1 & 2 \end{pmatrix}$$

See the proof of Proposition III.6.1 for the definition of $\beta_{q,r,\sigma}$. If $E = (e_{q,i})_{q\in\{s,t,p\};i\in\{0,1\}}$ then it follows from (28) that $(B_\sigma E)_{q,i} = \lambda_{q,\sigma,i}$. We have

$$E = \begin{pmatrix} 0 & 1 \\ 2 & 0 \\ 0 & 0 \end{pmatrix}$$

and

$$B_a E = \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, B_b E = \begin{pmatrix} 2 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, B_c E = \begin{pmatrix} 2 & 2 \\ 0 & 2 \\ 2 & 1 \end{pmatrix}.$$

To obtain the complete transducer, with the notations of the proof of Proposition III.6.1, we have to compute the C_w namely

$$C_{q,\sigma} = \text{val}_q(\sigma) + \mathbf{v}_0(s) - \mathbf{v}_0(q)$$

for q and σ such that $q.\sigma$ is a final state. Finally we have in Figure III.6 the finite letter-to-letter automaton built from $\mathcal{A}_L$ and the $\lambda_{q,\sigma,i}$'s. For

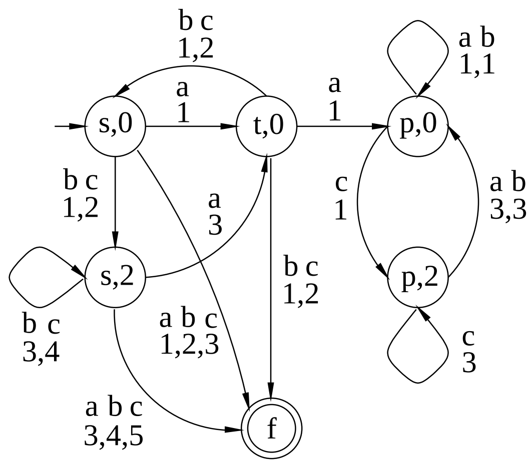

FIGURE III.6. The transducer computing g.

instance, since the first line of $B_a E$ is $(0,1)$, we have, starting in the state (s,α), an edge labeled by $(a, \mathbf{1}+\alpha)$ pointing towards the state $(s.a, \mathbf{0})$, $\alpha = 0, 2$. This transducer can be used in the following way. Notice that the first terms of $U = (\mathbf{u}_n)_{n\in\mathbb{N}}$ are

$$1, 3, 8, 22, 60, \ldots$$

By enumeration of the words of L, we get $\text{val}_S(ba) = 6$, $\text{val}_S(aba) = 12$ and $\text{val}_S(abca) = 39$. Starting in the initial state of the transducer and ending in its final state, we obtain the couples

$$\begin{pmatrix} b & a \\ 1 & 3 \end{pmatrix}, \begin{pmatrix} a & b & a \\ 1 & 1 & 1 \end{pmatrix}, \begin{pmatrix} a & b & c & a \\ 1 & 1 & 2 & 3 \end{pmatrix}.$$

To conclude this example, observe

$$\begin{cases} \pi_U(13) &= \mathbf{1}.3+\mathbf{3}.1=6 \\ \pi_U(111) &= \mathbf{1}.8+\mathbf{1}.3+\mathbf{1}.1=12 \\ \pi_U(1123) &= \mathbf{1}.22+\mathbf{1}.8+\mathbf{2}.3+\mathbf{3}.1=39. \end{cases}$$

We can do the same construction for the language $L'=a^+\{a,b\}^*$. Its minimal automaton $\mathcal{A}_{L'}$ is given in Figure III.7. The sequence U of

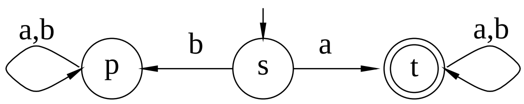

FIGURE III.7. The minimal automaton of $a^+\{a,b\}^*$.

Proposition III.6.1 is given by $\mathbf{u}_n(t)=2^n$. So here, the involved Pisot number is 2 and it is multiplicatively independent from $1+\sqrt{3}$. So, from [**54**], the only subsets which are simultaneously recognizable in $(L,\{a,b,c\},a<b<c)$ and $(L',\{a,b\},a<b)$ are the finite union of arithmetic progressions.

Although the conditions of Proposition III.6.1 are only sufficient conditions, the next remark shows us an example of a system which does not satisfy all these conditions and such that S-recognizability is not closed under multiplication by 2.

REMARK III.6.7. Let $J=a\{a,b\}^*\cup\{a,b\}^*bb\{a,b\}^*$. Notice that J is an exponential language with exponential complement. Its minimal automaton $\mathcal{A}_J$ is given in Figure III.8. We consider the numeration

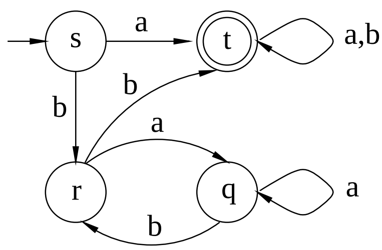

FIGURE III.8. The minimal automaton of $a\{a,b\}^*\cup\{a,b\}^*bb\{a,b\}^*$.

system $S=(J,\{a,b\},a<b)$ and we show that

(i) there exists no linear recurrent sequence associated to a Pisot number such that the condition (27) of Proposition III.6.1 is satisfied for all states of $\mathcal{A}_J$

(ii) the set $X=\{\mathbf{v}_n(s):n\in\mathbb{N}\}$ is S-recognizable but $2X$ is not.

One can check that for all $n\geq 1$, $\mathbf{u}_n(t)=2^n$ and

$$\mathbf{u}_n(s)=2^n-\frac{\sqrt{5}}{5}\left(\frac{1+\sqrt{5}}{2}\right)^n+\frac{\sqrt{5}}{5}\left(\frac{1-\sqrt{5}}{2}\right)^n.$$

So (i) holds. To check (ii), we use the same technique as in Theorem III.5.3. One can verify that

$$\mathbf{v}_{n+1}(s) - 2\,\mathbf{v}_n(s) = 1 - \frac{\sqrt{5}}{5}\left(\frac{1-\sqrt{5}}{2}\right)^n + \frac{\sqrt{5}}{5}\left(\frac{1+\sqrt{5}}{2}\right)^n$$

has an exponential dominant term. Furthermore, for all n large enough there exists i such that

$$\mathbf{u}_{n+1}(J_{b^{i+2}}) = 2^{n-i-1} < \mathbf{v}_{n+1}(s) - 2\,\mathbf{v}_n(s) \leq 2^{n-i} = \mathbf{u}_{n+1}(J_{b^{i+1}})$$

and $n - i \to +\infty$ if $n \to +\infty$. One can conclude as in Theorem III.5.3; $\mathrm{rep}_S(2\,\mathbf{v}_n(s)) = b^{i+1}az$ with $|z| = n - i - 1$.

We end this chapter with a concluding remark.

Remark III.6.8. In this chapter, we have studied the stability of S-recognizability under multiplication by a constant in relation with the complexity of the regular languages on which the systems are built. With a small exception, the only regular languages for which multiplication preserves recognizability, are exponential languages with exponential complement. This result itself is interesting if one would like to build an abstract numeration system having useful arithmetic properties. This result can also be connected to the case of positional systems. The problem of addition in the frame of positional system is not settled yet, even for linear positional systems having a characteristic polynomial that is not the minimal polynomial of a Pisot number.

The last section of this chapter shows that the only class of "good" numeration systems with respect to arithmetic operations, seems once again to be the class of linear numeration systems such that their characteristic polynomial is the minimal polynomial of a Pisot number. (If it is not exactly this class, the abstract systems encountered are strongly related to positional systems having the "Pisot property". See the assumptions of Proposition III.6.1.) All our attempts in the building of an abstract numeration system in which addition and multiplication by a constant do preserve recognizability, have lead to the assumptions of Proposition III.6.1 and a Pisot number.

Although the question remains open, this chapter is a testimony that the only "good" class of positional numeration systems is the class of linear systems such that their characteristic polynomial is the minimal polynomial of a Pisot number. It is therefore challenging to find an abstract numeration system which cannot be related, using Proposition III.6.1, to a Pisot number and for which addition and multiplication by a constant do preserve recognizability.

One can also notice, in view of Theorem III.4.1, that if the complexity function of a regular language is bounded by a constant, i.e. by a polynomial of degree zero, (a language with such a complexity function is said to be *slender*) then the possible multiplicators for which recognizability holds are of the form n^1, $n \in \mathbb{N}$. Therefore, the slender languages have to be discussed for their own properties. This situation is studied in Chapter V (see Theorem V.2.3). It is shown that for abstract numeration systems built on a slender regular language, addition and multiplication by a constant preserve recognizability.

CHAPTER IV

S-Automatic sequences

For a given subset X of $\mathbb{N}$, a question arises naturally. Is this set S-recognizable for some system S ? This kind of question is answered in Chapter 2 when X is an exponential polynomial image of $\mathbb{N}$. An interesting question could be the following: is there a system S such that the set of primes is S-recognizable ?

To answer this question we generalize Proposition I.3.3 and show that a subset of $\mathbb{N}$ is S-recognizable if and only if its characteristic sequence can be generated by an "automatic" method. The term automatic refers, as we shall further see, to a generalization of the k-automatic sequences for numeration systems on a regular language.

The k-automatic sequences are well-known and have been studied extensively since the 70's [**3**], [**20**], [**23**] and have already been generalized in many different ways [**1**], [**62**]. In particular, to generalize the k-automatic sequences, J. Shallit considers some kind of linear numeration system instead of the standard numeration system with an integer base k [**62**]. Two properties of Shallit's systems are precisely that the set of all the representations is regular and that the lexicographic ordering is respected.

Here, instead of giving $\rho_k(n)$ to a deterministic finite automaton with output, we feed it with $\mathrm{rep}_S(n)$ to obtain an output which is the n^{th} term of an S-*automatic sequence* for an abstract numeration system S.

Having thus introduced the concept of S-automatic sequences, we can learn their intrinsic properties but also use them as a tool to check if a subset of $\mathbb{N}$ is S-recognizable. The material of this chapter can be found mainly in [**57**]. In the first section, we recall some definitions and we introduce a running example which could be instructive for the reader not familiar with automatic sequences. In the second section, we adapt the classical results concerning the fiber and the kernel of an automatic sequence.

In the third section, we show that an S-automatic sequence is always generated by a substitution (i.e., an iterated non-uniform morphism followed by the application of another morphism). From this, we deduce that the number of distinct factors of length l in an S-automatic sequence is $O(l^2)$. We also show how to build S-automatic sequences with at least the same complexity that infinite words obtained by iterated morphisms.

A. Cobham showed the equivalence between the k-automatic sequences and the sequences obtained by iterating a uniform morphism (also called uniform tag system) [**20**]. Thanks to the results developed in the third section, the fourth section shows the equivalence between morphic predicates and S-automatic characteristic sequences.

In the fifth section, we shall be able to show that for any numeration system S, the set of primes is never S-recognizable. To be S-recognizable, the characteristic sequence of the set must be generated by a substitution. We use some results of C. Mauduit about the complexity of the infinite words obtained by substitution [**46**], [**47**].

1. Some definitions

Let Σ be a finite alphabet. We denote by Σ^ω the set of right-infinite words over Σ. A right-infinite word x over Σ is a sequence $(x_n)_{n\in\mathbb{N}}$ of elements belonging to Σ.

Definition IV.1.1. Let $S=(L,\Sigma,<)$ be an abstract numeration system and Δ be a finite alphabet. A sequence $x\in\Delta^\omega$ is *S-automatic* if there exists a DFAO $\mathcal{A}=(Q,\Sigma,\delta,s,\Delta,\tau)$ such that for all $n\in\mathbb{N}$,

$$x_n=\tau(\delta(s,\mathrm{rep}_S(n))).$$

If the context is clear, we write $\tau(w)$ in place of $\tau(\delta(s,w))$.

Remark IV.1.2. A subset $X\subset\mathbb{N}$ is S-recognizable if and only if its characteristic sequence $(\chi_n^X)_{n\in\mathbb{N}}\in\{0,1\}^\omega$ is S-automatic.

Two additional methods for generating infinite sequences will be used in the sequel.

Definition IV.1.3. Let $\varphi:\Sigma\to\Sigma^*$ be a morphism of monoids such that for some $\sigma\in\Sigma$, $\varphi(\sigma)\in\sigma\Sigma^*$. The word

$$x_\varphi=\varphi^\omega(\sigma)=\lim_{n\to\infty}\varphi^n(\sigma)$$

is a fixed point of φ and we say that x_φ is *generated by an iterated morphism.* A morphism is *uniform* if $|\varphi(\sigma_1)|=\ldots=|\varphi(\sigma_n)|$, $\Sigma=\{\sigma_1,\ldots,\sigma_n\}$.

Example IV.1.4. We give two classical examples of words generated by an iterated morphism. The first is the *Thue-Morse word* (also known as *Prouhet-Thue-Morse sequence*) obtained by iterating the uniform morphism ψ defined on $\{a,b\}$ by $\psi(a)=ab$ and $\psi(b)=ba$. The first applications of ψ are shown in Table IV.1.

The second is the *Fibonacci word* generated by the non-uniform morphism φ defined by $\varphi(a)=ab$ and $\varphi(b)=a$. The first applications of φ are shown in Table IV.2.

Definition IV.1.5. A *substitution* T is a triple (φ,h,c) such that $\varphi:\Sigma\to\Sigma^*$ and $h:\Sigma\to\Delta^*$ are morphisms of monoids. Moreover

$$\begin{aligned} \psi(a) &= ab \\ \psi^2(a) &= abba \\ \psi^3(a) &= abbabaab \\ \psi^4(a) &= abbabaabbaababba \\ &\vdots \end{aligned}$$

TABLE IV.1. The Thue-Morse word.

$$\begin{aligned} \varphi(a) &= ab \\ \varphi^2(a) &= aba \\ \varphi^3(a) &= abaab \\ \varphi^4(a) &= abaababa \\ &\vdots \end{aligned}$$

TABLE IV.2. The Fibonacci word.

$c \in \Sigma$, $\varphi(c) \in c\Sigma^*$ and for any $\sigma \in \Sigma$, $h(\sigma) = \varepsilon$ or $h(\sigma) \in \Delta$ (h is said to be a *weak coding*). We said that the word

$$x_T = h(\varphi^\omega(c))$$

over Δ is *generated by the substitution T*.

If $h(\sigma) = \varepsilon$ for some σ then h is said to be *erasing* otherwise h is said to be *non-erasing*.

EXAMPLE IV.1.6. We consider the following substitution (φ, h, γ) to build the characteristic sequence of the perfect squares (see [**45**]),

$$\varphi : \left\{ \begin{array}{lll} \gamma & \mapsto & \gamma abc \\ a & \mapsto & a \\ b & \mapsto & bcc \\ c & \mapsto & c \end{array} \right. , \quad h : \left\{ \begin{array}{lll} \gamma & \mapsto & 1 \\ a & \mapsto & 1 \\ b & \mapsto & 0 \\ c & \mapsto & 0 \end{array} \right.$$

$$\begin{aligned} \varphi(\gamma) &= \gamma abc \\ \varphi^2(\gamma) &= \gamma abcabccc \\ \varphi^3(\gamma) &= \gamma abcabcccabccccc \\ \varphi^4(\gamma) &= \gamma abcabcccabcccccabccccccc \\ &\vdots \end{aligned}$$

$$h(\varphi^\omega(\gamma)) = 1100100001000000100000000\ldots$$

TABLE IV.3. A substitution generating the characteristic sequence of the perfect squares.

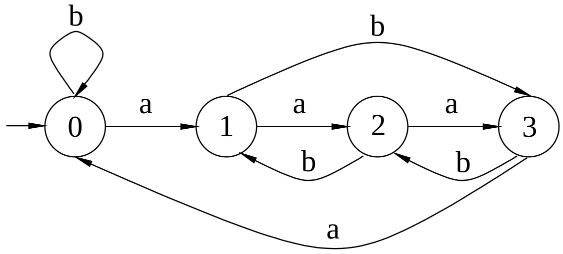

FIGURE IV.1. A deterministic finite automaton with output.

1.1. A running example. We consider the numeration system $S = (a^*b^*, \{a, b\}, a < b)$, the alphabets $\Sigma = \{a, b\}$, $\Delta = \{0, 1, 2, 3\}$ and the following DFAO

The first words of a^*b^* are

$$\varepsilon, a, b, aa, ab, bb, aaa, aab, abb, bbb, \ldots$$

Therefore, by feeding the automaton in Figure IV.1 with these words, we obtain the first terms of the sequence $x \in \Delta^\omega$,

$$x = 0102303120023101012302303120312023100231012301 0123\ldots.$$

REMARK IV.1.7. The sequence x is not ultimately periodic. We can observe that the distance between two occurrences of the block '00' is not bounded. Indeed,

$$\tau(w) = 0 \Leftrightarrow \exists r, t \in \mathbb{N} : w = a^{4r}b^t \tag{29}$$

thus a block '00' comes from two consecutive words b^{4r-1} and a^{4r}, $r \geq 1$ and the number of words of length n in a^*b^* is $\mathbf{u}_n = n + 1$, $n \in \mathbb{N}$.

REMARK IV.1.8. The sequence x is not generated by an iterated morphism φ. First, observe that

$$\tau(w) = \left\{ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \right\} \Leftrightarrow \exists r, t \in \mathbb{N} : w = \left\{ \begin{array}{l} a^{4r+1}b^{3t},\ a^{4r+2}b^{3t+1},\ a^{4r+3}b^{3t+2} \\ a^{4r+1}b^{3t+2},\ a^{4r+2}b^{3t},\ a^{4r+3}b^{3t+1} \\ a^{4r+1}b^{3t+1},\ a^{4r+2}b^{3t+2},\ a^{4r+3}b^{3t}. \end{array} \right.$$

Assume that there exists a morphism φ such that $x = \lim_{n\to+\infty} \varphi^n(0)$.

1) If $\varphi(0) \in 0102\Delta^*$ then the block '0102' must appear at least twice in x since '0' appears twice in x. If the first '0' of the block is obtained from a word $a^{4r}b^t$ with $r \geq 1$ then the second '0' is obtained from $a^{4r-2}b^{t+2}$. In view of (29), this leads to a contradiction. If the first '0' is obtained from b^t with $t \geq 1$ then the second '0' comes from a^tb and we have $t = 4y$. The '2' is obtained from $a^{4y-1}b^2$, which also leads to a contradiction.
2) If $\varphi(0) = 01$ then in view of the first terms of x,

$$\varphi(1) \in 023031200231\Delta^*.$$

We show that '023031200' appears only once in x. Suppose that we can find another block of this kind. Thus the last two '0' come from words b^{4r-1} and a^{4r} with $r \geq 2$. Since we consider all the words of a^*b^* in ascending lexicographic order, the first '0' of the block comes from a^7b^{4r-8}, which is in contradiction with (29).

3) If $\varphi(0) = 010$ then
$$\varphi(1) \in 23031200231\Delta^*$$
and $\varphi(010) \in 01023031200\Delta^*$. The block '010' appears at least twice in x but we know that '023031200' appears only once.

We shall further see that x is generated by a substitution.

2. First results about S-automatic sequences

Some classical results on k-automatic sequences can be easily restated [**20**], [**23**].

Definition IV.2.1. Let $a \in \Delta$ and $S = (L, \Sigma, <)$, the *S-fiber* $\mathcal{F}_S(x, a)$ of a sequence $x \in \Delta^\omega$ is defined as follows
$$\mathcal{F}_S(x, a) = \{\mathrm{rep}_S(n) \,:\, x_n = a\}.$$

Proposition IV.2.2. *Let x be an infinite sequence over Δ and $S = (L, \Sigma, <)$ be an abstract numeration system. The sequence x is S-automatic if and only if for all $a \in \Delta$, $\mathcal{F}_S(x, a)$ is a regular subset of L.*

Proof. If x is S-automatic then we have a DFAO $\mathcal{A} = (Q, \Sigma, \delta, s, \Delta, \tau)$ which is used to generate x. Let $L(\mathcal{A}')$ be the language recognized by the DFA $\mathcal{A}' = (Q, \Sigma, \delta, s, F)$ where the set of final states F only contains the states q such that $\tau(q) = a$. Then $\mathcal{F}_S(x, a)$ is regular since it is the intersection of the two regular sets $L(\mathcal{A}')$ and L.

The condition is sufficient. Let $\Delta = \{a_1, \ldots, a_n\}$. Notice that if $i \neq j$, $\mathcal{F}_S(x, a_i) \cap \mathcal{F}_S(x, a_j) = \emptyset$ and $L = \cup_{i=1}^n \mathcal{F}_S(x, a_i)$. For all $i = 1, \ldots, n$, $\mathcal{F}_S(x, a_i)$ is accepted by a DFA $\mathcal{A}_i = (Q_i, \Sigma, s_i, \delta_i, F_i)$. From these automata we build a DFAO $\mathcal{A} = (Q, \Sigma, s, \delta, \Delta, \tau)$ to generate x using the numeration system S. The set Q is $Q_1 \times \cdots \times Q_n$, the initial state is $(s_1, \ldots, s_n)$. For all states $(q_1, \ldots, q_n) \in Q$ and for all $\sigma \in \Sigma$, $\delta((q_1, \ldots, q_n), \sigma) = (\delta_1(q_1, \sigma), \ldots, \delta_n(q_n, \sigma))$. If there is a unique i such that $q_i \in F_i$ then $\tau((q_1, \ldots, q_n)) = a_i$ otherwise the state cannot be reached by a word of L and the associated output is meaningless. The sequence x is obtained from S and the DFAO $\mathcal{A}$. Hence the result.

□

The notion of k-kernel of a k-automatic sequence can be transposed as follows.

Definition IV.2.3. Let $S = (L, \Sigma, <)$ and x be an infinite sequence. As in Definition III.5.2, for each $w \in \Sigma^*$, we set
$$L_w = \{v \in L \,|\, \exists z \in \Sigma^* : v = wz\}.$$

One can enumerate L_w lexicographically with respect to $<$,

$$L_w = \{wz_0 < wz_1 < \ldots\}.$$

Thus for each $w \in \Sigma^*$, one can construct the subsequence $n \mapsto x_{\mathrm{val}_S(wz_n)}$ (notice that the subsequence can be finite or even empty). The set

$$\{n \mapsto x_{\mathrm{val}_S(wz_n)} \,:\, w \in \Sigma^*\}$$

is the S-*kernel* of x.

Proposition IV.2.4. *Let $S = (L, \Sigma, <)$ be a numeration system. A sequence $x \in \Delta^\omega$ is S-automatic if and only if its S-kernel is finite.*

Proof. If x is S-automatic, then we have a DFAO $\mathcal{A} = (Q, \Sigma, \delta, s, \Delta, \tau)$ used to generate x. We define the equivalence relation $\sim_1$ over Σ^* by $v \sim_1 w$ if and only if $\delta(s, v) = \delta(s, w)$. In the same way, the minimal automaton of L provides us with an equivalence relation $\sim_2$. The two relations have a finite index. So the relation $\sim_{1,2}$, given by $v \sim_{1,2} w$ if and only if $v \sim_1 w$ and $v \sim_2 w$, has also a finite index. Notice that the classes of $\sim_{1,2}$ give exactly the sequences $n \mapsto x_{\mathrm{val}_S(wz_n)}$. Indeed, $v \sim_2 w$ implies that $\{z \in \Sigma^* : vz \in L\} = \{z \in \Sigma^* : wz \in L\}$ thus $L_v = \{vz_0 < vz_1 < \ldots\}$ and $L_w = \{wz_0 < wz_1 < \ldots\}$ with the same $z_0, z_1, \ldots$.

The condition is sufficient. We show how to build a DFAO. The states are the subsequences $q_w = (n \mapsto x_{\mathrm{val}_S(wz_n)})$. The initial state is q_ε (i.e., the subsequence obtained from the empty word). The transition function δ is given by $\delta(q_w, \sigma) = q_{w\sigma}$ and the output function τ is given by $\tau(q_w) = x_{\mathrm{val}_S(w)}$.

□

We now study some operations on S-automatic sequences.

Definition IV.2.5. The sequence $(y_n)_{n\in\mathbb{N}}$ is a *finite modification* of the sequence $(x_n)_{n\in\mathbb{N}}$ if

$$\exists p, q \in \mathbb{N} : \forall n \in \mathbb{N},\ y_{p+n} = x_{q+n}.$$

Proposition IV.2.6. *Let $S = (L, \Sigma, <)$ be an abstract numeration system. The set of S-automatic sequences is closed under finite modifications.*

Proof. A sequence $x \in \Delta^w$ is S-automatic if and only if for all $a \in \Delta$, $\mathcal{F}_S(x, a)$ is a regular language (see Proposition IV.2.2); in other words, if $\mathrm{val}_S(\mathcal{F}_S(x, a))$ is S-recognizable. By Proposition II.1.1, the S-recognizability of this latter set is conserved by translation. Therefore, for all $a \in \Delta$, $\mathcal{F}_S(y, a)$ is also a regular language. Hence the conclusion.

□

Definition IV.2.7. A sequence $(y_n)_{n\in\mathbb{N}}$ is obtained by *periodic deletation* of the sequence $(x_n)_{n\in\mathbb{N}}$ if there exist $a, b, j_0, \ldots, j_{a-1}$, $1 \leq a < b$, $0 \leq j_0 < \cdots < j_{a-1} < b$ such that

$$y_{an+i} = x_{bn+j_i},\ 0 \leq i < a,\ n \in \mathbb{N}.$$

Proposition IV.2.8. *Let $S = (L, \Sigma, <)$ be an abstract numeration system and $x = (x_n)_{n\in\mathbb{N}}$ be an S-automatic sequence. If the sequence $y = (y_n)_{n\in\mathbb{N}}$ is obtained by periodic deletation of x then there exists a numeration S' such that y is S'-automatic. Moreover, the set of S-automatic sequences is not closed under periodic deletation.*

Proof. The sequence x is generated through the use of a DFAO $\mathcal{A}$ and the language L. To obtain y, we can consider the same DFAO $\mathcal{A}$ but the language

$$L' = \bigcup_{i=0}^{a-1} \mathrm{rep}_S(b\mathbb{N} + j_i).$$

By Theorem II.2.1, this latter language is regular. It is obvious that $\mathcal{A}$ and $S' = (L', \Sigma, <)$ produce y.

To prove the second part, we consider the following DFAO depicted in Figure IV.2 and the language a^*b^*. If we assume that $a < b$, this

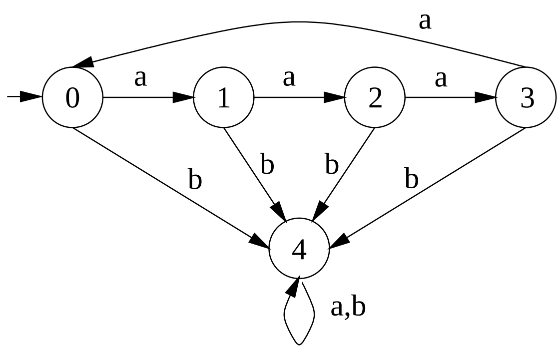

Figure IV.2. A DFAO.

automaton produces the sequence

$$x = 0142443444044441444442444444344444440\ldots$$

which is S-automatic for the system S built on a^*b^* and

$$\mathcal{F}_S(x, 0) = (a^4)^*.$$

In view of (11),

$$\mathrm{val}_S((a^4)^*) = \{2n\,(4n+1) \mid n \in \mathbb{N}\}.$$

We delete from x the letters at odd positions and obtain the sequence

$$y = 04434044444444434440\ldots$$

Therefore,

$$\mathcal{F}_S(y, 0) = \{\mathrm{rep}_S(n\,(4n+1)) \mid n \in \mathbb{N}\}.$$

To conclude the proof, we show that this latter language is not regular. Assume that $\mathcal{F}_S(y, 0)$ is regular. Therefore, $\mathcal{F}_S(y, 0) \cap a^*b^4$ must be regular. A word $a^p b^4$ belongs to $\mathcal{F}_S(y,0)$ if and only if there exists n such that

$$n\,(4n+1) = \frac{1}{2}(p+4)(p+5)+4$$

in other words, such that

$$(8n+1)^2 - 2(2p+9)^2 = 63.$$

We use the same technique as in the proof of Proposition III.1.2. The minimal solution of the Pell's equation $U^2 - 2V^2 = 1$ is $(u,v) = (3,2)$ and $X^2 - 2Y^2 = 63$ has the solution $(X_0, Y_0) = (9,3)$ and

$$\begin{pmatrix} X_{i+1} \\ Y_{i+1} \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \end{pmatrix}.$$

Thus for each $n \in \mathbb{N}$, $X_{2n} \in 8\mathbb{N}+1$. We conclude as in the proof of Proposition III.1.2 and obtain a contradiction.

□

3. Complexity of S-automatic sequences

The *complexity function p_x of an infinite sequence x* maps $n \in \mathbb{N}$ to the number $p_x(n)$ of distinct factors of length n occurring at least once in x. In this section, we shall show that since every S-automatic sequence is generated by a substitution, its complexity is in $O(n^2)$.

Recall that an infinite word w generated by an iterated morphism has a complexity such that

$$c_1 f(n) \le p_w(n) \le c_2 f(n)$$

where $f(n)$ is one of the following functions: 1, n, $n \log\log n$, $n \log n$ or n^2 [**50**]. For a survey on the complexity function, see for instance [**2**].

The next remark shows that an S-automatic sequence can reach at least the same complexity as a word generated by a morphism.

Remark IV.3.1. For every infinite word w generated by an iterated morphism φ over an alphabet Δ, there exists an S-automatic sequence u such that for all $n \in \mathbb{N}$, $p_w(n) \le p_u(n)$.

We show how to proceed on the following example,

$$\Delta = \{0,1\},\ \varphi : \begin{cases} 0 \mapsto 0101 \\ 1 \mapsto 11. \end{cases}$$

It is well known that the complexity function p_w of $w = \varphi^\omega(0)$ is $O(n \log\log n)$ [**50**]. To the morphism φ, we associate a finite automaton $\mathcal{A}$ ($\mathcal{A}$ is not deterministic because the morphism is not uniform). The set of states is Δ, all the states are final and the transition function δ is obtained by reading the productions of φ from left to right. For this purpose, we introduce a new ordered alphabet Σ such that $\#\Sigma = \sup_{x\in\Delta} |\varphi(x)|$. Here, 0 gives the initial state (because we consider the word $\varphi^\omega(0)$) and 1 the other state. Thus with $\Sigma = \{a < b < c < d\}$, we have

$$\left\{\begin{array}{lcl}\delta(0,a) & = & [\varphi(0)]_1 = 0\\ \delta(0,b) & = & [\varphi(0)]_2 = 1\\ \delta(0,c) & = & [\varphi(0)]_3 = 0\\ \delta(0,d) & = & [\varphi(0)]_4 = 1\end{array}\right. \text{ and } \left\{\begin{array}{lcl}\delta(1,a) & = & [\varphi(1)]_1 = 1\\ \delta(1,b) & = & [\varphi(1)]_2 = 1.\end{array}\right.$$

Then $\mathcal{A}$ is the automaton given in Figure IV.3.

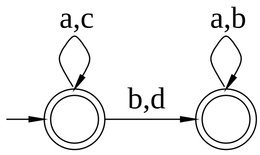

FIGURE IV.3. The automaton associated to the morphism φ.

The language accepted by $\mathcal{A}$ is $L = \{a,c\}^*\{b,d\}\{a,b\}^* \cup \{a,c\}^*$. So, the numeration system S is $(L, \Sigma, a < b < c < d)$. This kind of construction can also be found in [**45**]. Now, from $\mathcal{A}$, we simply build the DFAO $\mathcal{A}'$ given in Figure IV.4.

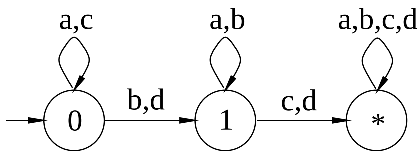

FIGURE IV.4. The DFAO $\mathcal{A}'$.

Its output is easily computed. The third state can have any output because it is never reached through a word belonging to L. One remarks that the S-automatic sequence obtained with $\mathcal{A}'$ and S is

$$u = \varphi(0)\varphi^2(0)\varphi^3(0)\ldots$$

and thus every factor of $w = \varphi^\omega(0)$ belongs to u.

We now show that every S-automatic sequence is generated by a substitution. First, we set forth an obvious Lemma.

LEMMA IV.3.2. *Let* $\Sigma = \{\sigma_1 < \ldots < \sigma_n\}$, $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ *be a DFA and* $\alpha \notin Q$. *The morphism* $\varphi_{\mathcal{A}} : Q \cup \{\alpha\} \to (Q \cup \{\alpha\})^*$ *defined by*

$$\left\{\begin{array}{l}\alpha \mapsto \alpha s\\ q \mapsto \delta(q,\sigma_1)\ldots\delta(q,\sigma_n),\ q \in Q\end{array}\right.$$

produces the sequence $((x_\varphi)_n)_{n\in\mathbb{N}}$ *of the states reached by the words of* Σ^* *i.e.,* $\forall i \in \mathbb{N} \setminus \{0\}$, $(x_\varphi)_i = \delta(s, w_i)$ *where* w_i *is the* i^{th} *element of* $(\Sigma^*, <)$.

EXAMPLE IV.3.3. Here is an application of Lemma IV.3.2. Consider the automaton $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ depicted in Figure IV.5. By

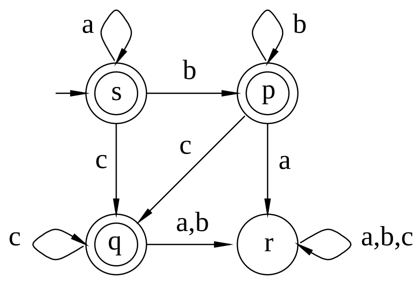

FIGURE IV.5. The minimal automaton of $a^*b^*c^*$.

the previous lemma, we obtain

$$\varphi : \begin{cases} \alpha & \mapsto & \alpha s \\ s & \mapsto & spq \\ p & \mapsto & rpq \\ q & \mapsto & rrq \\ r & \mapsto & rrr \end{cases}$$

and the sequence of the states reached by the words of $\{a, b, c\}^*$ is

$$x_\varphi = \alpha sspqspqrpqrrqspqrpqrrqrrrrpqrrq\ldots.$$

PROPOSITION IV.3.4. *Every S-automatic sequence is generated by a substitution.*

Proof. Let $S = (L, \Sigma, <)$, $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ be a DFA accepting L and x be an S-automatic sequence obtained with the DFAO $\mathcal{M} = (Q', \Sigma, \delta', s', \Delta, \tau)$. From these two automata, we build the product automaton $\mathcal{P} = (Q \times Q', \Sigma, (s, s'), \eta)$ where $\eta((q, q'), \sigma) = (\delta(q, \sigma), \delta'(q', \sigma))$. We shall not explicitly write the final states of $\mathcal{P}$. By Lemma IV.3.2, we associate to this automaton a morphism $\varphi_{\mathcal{P}} : (Q \times Q') \cup \{\alpha\} \to ((Q \times Q') \cup \{\alpha\})^*$. To conclude the proof, we make up the erasing morphism $h : (Q \times Q') \cup \{\alpha\} \to \Delta^*$ defined by

$$\begin{cases} h(\alpha) & = \varepsilon & \\ h((q, q')) & = \varepsilon & \text{, if } q \notin F; \\ & = \tau(q') & \text{, otherwise.} \end{cases}$$

Indeed, $\varphi_{\mathcal{P}}^{\omega}(\alpha)$ is the sequence of the states reached by the words of Σ^* in $\mathcal{P}$. But we are only interested in the words belonging to L and in the corresponding output of $\mathcal{M}$. Thus x is generated by $(\varphi_{\mathcal{P}}, h, \alpha)$.

$\square$

Dealing with erasing morphisms whenever one wants to determine the complexity function of a sequence can turn out to be painful. So the next lemma allows us to get rid of erasing morphisms. (The author thanks J.-P. Allouche who has pointed out this result.)

LEMMA IV.3.5. [**18, 4**] *If f and g are arbitrary morphisms such that $f(g^\omega(a))$ is an infinite word, then there exist a non-erasing morphism k and a coding h (i.e., a letter-to-letter morphism h) such that*

$$f(g^\omega(a)) = h(k^\omega(a)).$$

THEOREM IV.3.6. *The complexity of an S-automatic sequence is in $O(n^2)$. Moreover, there exists an S-automatic sequence $y = (y_n)_{n\in\mathbb{N}}$ and a positive constant d' such that $\forall n > 0$, $p_y(n) \geq d'n^2$.*

Proof. Let $x = (x_n)_{n\in\mathbb{N}}$ be an S-automatic sequence. By Proposition IV.3.4, x is generated by a substitution (φ, h, α) and by Lemma IV.3.5 we can assume that h is non-erasing. The word $x_\varphi = \varphi^\omega(\alpha)$ is generated by an iterated morphism. Then $p_{x_\varphi}(n) \leq d\,n^2$. Remember that if v, w are two infinite words and if h is a non-erasing morphism such that $h(v) = w$ then there exist positive constants a, b such that $p_w(n) \leq a\,p_v(n+b)$ [**50**]. Hence the conclusion, since $x = h(x_\varphi)$.

We show that there exist a language L over an ordered alphabet and a DFAO such that the corresponding automatic sequence $y = (y_n)_{n\in\mathbb{N}}$ has a complexity function $p_y(n) \geq d'n^2$.

The morphism

$$\varphi : \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 12 \\ 2 \mapsto 2 \end{cases}$$

generates the word

$$w = \varphi^\omega(0) = 0112122122212222122222122222\ldots.$$

Since 2 is a bounded letter (i.e. $|\varphi^n(2)|$ is bounded) and 2^n is a factor of w for an arbitrary n, there exists a positive constant d' such that $p_w(n) \geq d'\,n^2$ (see [**50**]). Using the same technique as in Remark IV.3.1, we construct an S-automatic sequence y such that $p_y(n) \geq p_w(n)$. One finds easily that the regular language used in the numeration system S is $L = a^* \cup a^*ba^* \cup a^*ba^*ba^*$.

$\square$

To conclude this section, we refine in a very simple way Proposition IV.3.4 to give a characterization of S-automatic sequences.

DEFINITION IV.3.7. Let $T = (\varphi, h, c)$ and $T' = (\varphi', h', c')$ be two substitutions such that $\varphi : \Sigma \to \Sigma^*$, $h : \Sigma \to \Delta^*$, $\varphi' : \Sigma' \to \Sigma'^*$ and $h' : \Sigma' \to \Delta'^*$. A *morphism of substitutions* $m : T \to T'$ is a surjective morphism $m : \Sigma \cup \Delta \to \Sigma' \cup \Delta'$ such that

(1) $m(c) = c'$, $m(\Sigma) = \Sigma'$, $m(\Delta) = \Delta'$
(2) $m(\varphi(\sigma)) = \varphi'(m(\sigma))$, $\forall \sigma \in \Sigma$
(3) $m(h(\sigma)) = h'(m(\sigma))$, $\forall \sigma \in \Sigma$.

EXAMPLE IV.3.8. There exists a morphism m between the following substitutions (φ, h, a) and (φ', h', a'),

$$\varphi : \begin{cases} a & \mapsto & abc \\ b & \mapsto & bd \\ c & \mapsto & ac \\ d & \mapsto & dbc \end{cases} \quad h : \begin{cases} a & \mapsto & 0 \\ b & \mapsto & 1 \\ c & \mapsto & \varepsilon \\ d & \mapsto & 2 \end{cases}$$

$$\varphi' : \begin{cases} a' & \mapsto & a'b'c' \\ b' & \mapsto & b'a' \\ c' & \mapsto & a'c' \end{cases} \quad h' : \begin{cases} a' & \mapsto & 0' \\ b' & \mapsto & 1' \\ c' & \mapsto & \varepsilon \end{cases}$$

It is clear that, one can find

$$m : \begin{cases} a & \mapsto & a' \\ b & \mapsto & b' \\ c & \mapsto & c' \\ d & \mapsto & a' \end{cases} \quad m : \begin{cases} 0 & \mapsto & 0' \\ 1 & \mapsto & 1' \\ 2 & \mapsto & 0' \end{cases}$$

For a regular language L on the totally ordered alphabet $(\Sigma, <)$ and for a DFAO $\mathcal{D} = (Q, \Sigma, \delta, s, \Delta, \tau)$, one can make up the *canonical substitution* $T_{(L,<,\mathcal{D})}$ by proceeding in the same way as in Proposition IV.3.4 with $\mathcal{A}$ equal to the minimal automaton $\mathcal{A}_L$ of L and the DFAO $\mathcal{M}$ equal to a reduced and accessible copy of $\mathcal{D}$.

To reduce $\mathcal{D}$, one has to merge the states p, q such that for all $w \in \Sigma^*$, $\tau(\delta(p, w)) = \tau(\delta(q, w))$.

DEFINITION IV.3.9. A substitution T is an $(L, <, \mathcal{D})$-*substitution* if there exists a morphism $m : T \to T_{(L,<,\mathcal{D})}$. This kind of construction has already been introduced in [**14**] for linear numeration systems based on a Pisot number.

The next theorem is obvious and we state it without proof.

THEOREM IV.3.10. *Let $S = (L, \Sigma, <)$. The sequence $x \in \Delta^\omega$ is S-automatic if and only if x is generated by a $(L, <, \mathcal{D})$-substitution for some DFAO $\mathcal{D}$.*

4. Equivalence with morphic predicates

In this section, we obtain the converse of Proposition IV.3.4. Consequently, the S-automatic sequences and the infinite words obtained by substitution are the same. In particular, the recognizable subsets of $\mathbb{N}$ are exactly the morphic predicates. Let us fix the terminology of [**45**].

DEFINITION IV.4.1. The *characteristic word* of a predicate P over $\mathbb{N}$ is the infinite sequence $\chi^P = (x_n)_{n\in\mathbb{N}}$ such that $x_n = 1$ iff $n \in P$ and $x_n = 0$ otherwise. A predicate is said to be *morphic* if its characteristic word is generated by a substitution.

EXAMPLE IV.4.2. In view of Example IV.1.6, the predicate $P = \{n^2 \mid n \in \mathbb{N}\}$ is morphic.

The main issue of the present section is the following proposition. It results from a joint work with A. Maes.

PROPOSITION IV.4.3. *Every infinite word generated by a substitution is an S-automatic sequence for some abstract numeration system S.*

Proof. Let (φ, h, a) be a substitution. Thanks to Lemma IV.3.5, we can assume that $\varphi : \Sigma \to \Sigma^+$ is non-erasing and that $h : \Sigma \to \Delta$ is a coding. As in Remark IV.3.1, to the morphism φ we associate an automaton $\mathcal{A}$ over an alphabet $\Gamma = \{\gamma_1, \ldots, \gamma_r\}$ with $r = \max_{\sigma\in\Sigma} |\varphi(\sigma)|$. The set of states of $\mathcal{A}$ is Σ. Let $L \subset \Gamma^*$ be the regular language accepted by $\mathcal{A}$. Since $\varphi(a) \in a\Sigma^+$, it is clear that if $w \in L$ then $\gamma_1 w \in L$. Indeed, the transitions of $\mathcal{A}$ are given by the images of φ and therefore, the initial state of $\mathcal{A}$ has a loop labeled by the first letter of Γ.

To enlighten the proof, we consider the following short example. With the previous notation, $\Sigma = \{a, b, c\}$, $\Delta = \{0, 1, 2\}$,

$$\varphi : \Sigma \to \Sigma^+ : \left\{ \begin{array}{lcl} a & \mapsto & abc \\ b & \mapsto & bc \\ c & \mapsto & aac \end{array} \right. \quad \text{and } h : \Sigma \to \Delta : \left\{ \begin{array}{lcl} a & \mapsto & 0 \\ b & \mapsto & 1 \\ c & \mapsto & 2. \end{array} \right.$$

The automaton $\mathcal{A}$ over $\Gamma = \{\alpha, \beta, \gamma\}$ is given in Figure IV.6.

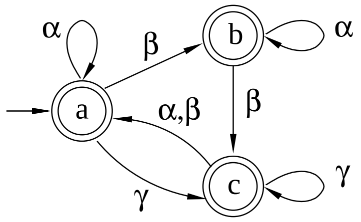

FIGURE IV.6. The automaton $\mathcal{A}$ associated to φ.

If φ is not uniform then the transition function of $\mathcal{A}$ is partial. In our example, $b.\gamma$ is not defined because $|\varphi(b)| = 2$. If we apply without any adaptation Lemma IV.3.2, instead of the sequence of states reached by the words of Γ^*, we obtain the sequence of states reached by the words of L. Indeed, by construction, $\mathcal{A}$ has no sink and all its states are final. Continuing our example, we have a morphism ψ obtained from $\mathcal{A}$ (or equivalently from φ since $\mathcal{A}$ is derived from φ). Let $\mu \notin \Sigma$,

$$\psi : \left\{ \begin{array}{lcl} \mu & \mapsto & \mu a \\ a & \mapsto & \varphi(a) = abc \\ b & \mapsto & \varphi(b) = bc \\ c & \mapsto & \varphi(c) = aac. \end{array} \right.$$

The main point is to compare the sequence $\psi^\omega(\mu)$ of the states reached by the words of L and the infinite word $\varphi^\omega(a)$,

$$\begin{array}{lllllllll}
\varphi^\omega(a) = & & a & & bc & & bcaac & & bcaacabcabca\cdots \\
\psi^\omega(\mu) = & \mu & a & \underbrace{a}_{w_1} & bc & \underbrace{abc}_{w_2} & bcaac & \underbrace{abcbcaac}_{w_3} & bcaacabcabca\cdots
\end{array}$$

Notice that $w_2 = \varphi(a)$ is obtained by applying ψ to w_1 and $w_3 = \varphi(abc)$ is obtained by applying ψ to w_2. For each $n \in \mathbb{N} \setminus \{0\}$, there exists a sub-word w_n in $\psi^\omega(\mu)$ that is derived from $a = w_1$ by application of ψ^{n-1}. The first elements of $\psi^\omega(\mu)$, except μ, are the states reached by

$$\varepsilon, \alpha, \beta, \gamma, \alpha\alpha, \alpha\beta, \alpha\gamma, \beta\alpha, \beta\beta, \gamma\alpha, \gamma\beta, \gamma\gamma, \alpha\alpha\alpha, \ldots$$

Roughly speaking, the two sequences $\varphi^\omega(a)$ and $\psi^\omega(\mu)$ would coincide if we removed from $\psi^\omega(\mu)$ the states of $\mathcal{A}$ reached by the words in L beginning with α. This observation is general. Indeed, in the definition of φ, $\varphi(a)$ always begins with a. It simply ensures the convergence of $\varphi^n(a)$ when n tends to infinity. On the other hand, to any substitution (φ, h, a), the associated infinite word $\psi^\omega(\mu)$ always begins with μaa (the states reached by ε and γ_1 are a). This second letter a is a "seed" producing sub-words w_i in $\psi^\omega(\mu)$. By construction of ψ, these sub-words correspond exactly to the states reached by the words of L beginning with γ_1.

So it is clear that if we consider the language $L \setminus \gamma_1\Gamma^*$ then the corresponding numeration system S and the DFAO built with $\mathcal{A}$ where the output of a state $\sigma \in \Sigma$ is $h(\sigma)$ produce the S-automatic sequence $h(\varphi^\omega(a))$.

□

This latter result has an important corollary.

Corollary IV.4.4. *The recognizable subsets of* $\mathbb{N}$ *for abstract numeration systems and the morphic predicates coincide.*

Proof. Let S be a numeration system. A subset $X \subset \mathbb{N}$ is S-recognizable if and only if its characteristic sequence χ^X is S-automatic. By Proposition IV.3.4 and Proposition IV.4.3, any S-automatic sequence over $\{0, 1\}$ is a morphic predicate.

□

Remark IV.4.5. It is well known that the multiplication by a constant c does not preserve generally the S-recognizability of a set of integers. But thanks to the characterization of the recognizable subsets of $\mathbb{N}$ given in the previous corollary, we can build from an abstract numeration system S and an S-recognizable subset $X \subset \mathbb{N}$, a new system S' such that cX is S'-recognizable. Indeed, χ^X is S-recognizable and is therefore generated by a substitution (φ, h, c) where $h : \Sigma \to \{0, 1\}$

is a coding (thanks to Lemma IV.3.5). We can replace h by h' where

$$h'(\sigma) = \begin{cases} 1(0^{c-1}) & \text{, if } h(\sigma) = 1; \\ 0^c & \text{, if } h(\sigma) = 0. \end{cases}$$

By Lemma IV.3.5, there exists a substitution (ψ, g, d) such that

$$g(\psi^\omega(d)) = h'(\varphi^\omega(c)) = \chi^{cX}.$$

In other words, cX is a morphic predicate and is S'-recognizable for some system S'. Notice that the computation of S' can be effectively achieved (the proof of Proposition IV.4.3 is constructive).

5. Application to S-recognizable sets of integers

Proposition IV.3.4 gives a necessary condition for a set X of integers to be S-recognizable. The characteristic sequence $\chi^X \in \{0,1\}^\omega$ has to be generated by a substitution. Thus this proposition can be used as an interesting tool to show that a subset of $\mathbb{N}$ is not S-recognizable for any numeration system S.

In the sequel of this chapter, $\mathcal{P}$ is the set of primes and $\chi^\mathcal{P}$ is its characteristic sequence. We show that $\mathcal{P}$ is never S-recognizable. But first we build by hand a subset of $\mathbb{N}$ which cannot be S-recognizable because its characteristic sequence is too complex.

Example IV.5.1. For $n \geq 3$, consider the $\binom{n}{3}$ words belonging to $\{0,1\}^n$ that contain exactly three '1' and concatenate these lexicographically ordered words to obtain the word w_{n-3}. To conclude, let us consider the infinite word

$$w = w_0 w_1 w_2 \dots = \underbrace{111}_{w_0}\underbrace{0111\,1011\,1101\,1110}_{w_1}\underbrace{00111\,01011\,\dots}_{w_2}\dots$$

By construction, it is obvious that for all positive constants C, there exists n_0 such that $\forall n \geq n_0 : p_w(n) > Cn^2$. Thus w cannot be generated by a substitution and the corresponding subset

$$W = \{0, 1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 14, 15, 16, 17, 21, 22, 23, 25, 27, \dots\},$$

such that $\chi^W = w$, is never S-recognizable.

Proposition IV.5.2. *For any numeration system S, $\mathcal{P}$ is not S-recognizable.*

Proof. In [**46**], [**47**], C. Mauduit shows using some density arguments that $\chi^\mathcal{P} \in \{0,1\}^\omega$ is not generated by a substitution (φ, h, α) where h sends all the letters on 0 except one. A slight adaptation of the proof leads to the conclusion for any letter-to-letter morphism h. For the sake of completeness, a proof of Proposition IV.5.2 can be found in the appendix.

□

CHAPTER V

Some topics on S-recognizability

In the first section of this chapter, we study the stability of recognizability under the change of ordering of the alphabet. Thanks to the language $\{a,b\}^* \setminus a^*b^*$, we show that this operation does not generally preserve recognizability.

In the second section, we study the class of slender languages. For a numeration built on such a language, it is shown that the recognizable sets of integers are exactly the ultimately periodic sets. Consequently addition, multiplication by a constant and the change of ordering do preserve recognizability.

In the third section, we give some examples related to Cobham's theorem. We consider two systems built on exponential language having the same dominant root (that is, two supposedly "dependent" systems) and we find subsets of integers recognizable in a system and not recognizable in the other system.

1. Changing the ordering

In this section, we give some results concerning the changing of ordering. This operation induces a transformation from $\mathbb{N}$ onto $\mathbb{N}$ and a transformation in the language on which numeration systems are built. We study the conservation of S-recognizability under such operations. It is convenient to introduce notation for the changing of numeration systems. Given systems $S = (L, \Sigma, <)$ and $T = (L', \Sigma', \prec)$, we set

$$\xi_{S,T} = \mathrm{rep}_T \circ \mathrm{val}_S : L \to L' \text{ and } \xi'_{S,T} = \mathrm{val}_T \circ \mathrm{rep}_S : \mathbb{N} \to \mathbb{N}.$$

If the underlying S and T are known from the context, we simply write ξ and ξ'.

EXAMPLE V.1.1. Consider the abstract numeration systems $S = (a^*b^*, \{a,b\}, a < b)$ and $T = (a^*b^*, \{a,b\}, b < a)$. The first words are given in Table V.1. In this example, $\xi_{S,T}(aab) = abb$ and $\xi'_{S,T}(5) = 3$. It is clear that

$$\xi_{S,T}(a^i b^j) = a^j b^i,\ \ i,j \in \mathbb{N}.$$

Thus, $\xi_{S,T}(w) = h(w^R)$, where w^R is the *reversal* of w and $h : \{a,b\} \to \{a,b\}$ is the homomorphism defined by $h(a) = b$ and $h(b) = a$. Therefore, if $X \subset \mathbb{N}$ is S-recognizable then X is also T-recognizable.

$\mathbb{N}$	S	T
0	ε	ε
1	a	b
2	b	a
3	aa	bb
4	ab	ab
5	bb	aa
6	aaa	bbb
7	aab	abb
8	abb	aab
9	bbb	aaa
$\vdots$	$\vdots$	$\vdots$

TABLE V.1. Changing the ordering in a^*b^*.

In spite of the previous example, the change of ordering of the alphabet does not generally preserve the recognizability as we shall see in the case where $\Sigma = \{a, b\}$ and $L = \Sigma^* \setminus a^*b^*$.

LEMMA V.1.2. *Let $n \in \mathbb{N}$. For $U = (\Sigma^*, \Sigma, a < b)$ and $V = (\Sigma^*, \Sigma, b \prec a)$ one has*

$$\xi'_{U,V}(n) = 3.2^l - n - 3,$$

where $l = |\mathrm{rep}_U(n)|$.

Proof. Observe that if $w_1 < \cdots < w_{2^l}$ then $w_{2^l} \prec \cdots \prec w_1$. Moreover $2^l - 1 \leq n \leq 2^{l+1} - 2$. It is enlightened by Table V.2 which shows the words of length 3. Hence

$\mathbb{N}$	U	V
7	aaa	bbb
8	aab	bba
9	aba	bab
10	abb	baa
11	baa	abb
12	bab	aba
13	bba	aab
14	bbb	aaa

TABLE V.2. Changing the ordering in Σ^*.

$$\xi'_{U,V}(n) = 2^{l+1} - 2 - [n - (2^l - 1)].$$

□

PROPOSITION V.1.3. *Let* $\Sigma = \{a,b\}$ *and* $L = \Sigma^* \setminus a^*b^*$. *For all* $n \geq 2$, *if* $l = |\mathrm{rep}_U(n-1)|$ *then*

$$\xi_{S,T}(ba\,b^n) = ab\,a^{n-l-1}\,b\,\mathrm{rep}_U(n-1),$$

where $S = (L,\Sigma,a<b)$, $T = (L,\Sigma,b \prec a)$ *and* U, V *are defined as in Lemma V.1.2. In particular,* $\mathrm{val}_S(ba\,b^2b^*)$ *is not* T*-recognizable.*

Proof. The minimal automaton $\mathcal{A}_L$ of L is depicted in Figure V.1. Therefore $L_p = \Sigma^*$,

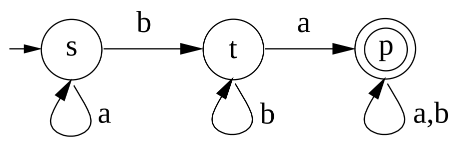

FIGURE V.1. The minimal automaton of $\Sigma^* \setminus a^*b^*$.

$$\begin{cases} \mathbf{u}_0(s) = \mathbf{u}_1(s) = 0, \\ \mathbf{u}_n(s) = 2^n - n - 1, \quad \forall n \geq 2, \end{cases}$$

while $\mathbf{u}_n(t) = 2^n - 1$ for all $n \in \mathbb{N}$.

In L, there are $\mathbf{v}_{n+1}(s)$ words of length not greater than $n+1$, $\mathbf{u}_{n+1}(s)$ words of length $n+2$ beginning with a and $\mathbf{u}_n(p) - 1$ words of length $n+2$ beginning with ba. Hence, the number of words belonging to L and lexicographically less than $ba\,b^n$ is

$$\mathrm{val}_S(ba\,b^n) = \sum_{i=2}^{n+1}(2^i - i - 1) + 2^{n+1} + 2^n - n - 3.$$

We sketch the computation of $\mathrm{val}_T[ab\,a^{n-l-1}\,b\,\mathrm{rep}_U(n-1)]$ obtained by using Lemma I.5.2

$$\begin{aligned}
& \mathrm{val}_t[a^{n-l-1}\,b\,\mathrm{rep}_U(n-1)] + \sum_{i=2}^{n+1}(2^i - i - 1) + 2^n + n \\
= \; & \mathrm{val}_p[a^{n-l-2}\,b\,\mathrm{rep}_U(n-1)] + \sum_{i=2}^{n+1}(2^i - i - 1) + 2^{n+1} - 1 \\
\vdots \; & \\
= \; & \mathrm{val}_p[b\,\mathrm{rep}_U(n-1)] + \sum_{i=2}^{n+1}(2^i - i - 1) + 2^{n+1} - 1 + \sum_{i=l+2}^{n-1} 2^i \\
= \; & \mathrm{val}_p[\mathrm{rep}_U(n-1)] + \sum_{i=2}^{n+1}(2^i - i - 1) + 2^{n+1} - 1 + \sum_{i=l+2}^{n-1} 2^i + 2^l \\
= \; & \xi'_{U,V}(n-1) + \sum_{i=2}^{n+1}(2^i - i - 1) + 2^{n+1} - 1 + 2^n - 3.2^l.
\end{aligned}$$

Hence the value of $\xi_{S,T}(ba\,b^n)$, in view of Lemma V.1.2. Applying the pumping lemma (Lemma III.5.1), it is now straightforward to check that $\mathrm{val}_S(ba\,b^2b^*)$ is not T-recognizable.

□

Although, the class of recognizable sets of integers is not closed under the change of ordering of the alphabet, we can exhibit some classes of languages for which this operation keeps recognizable sets unchanged. We also review some situations encountered throughout this work where the change of ordering has no effect.

PROPOSITION V.1.4. *Let $S = (L, \Sigma, <)$ and $T = (L, \Sigma, \prec)$ be two numeration systems. Let n_0 be a non-negative integer. If for all states p and q of $\mathcal{A}_L = (Q, \Sigma, \delta, s, F)$ and all $n \geq n_0$,*

$$\mathbf{u}_n(p) = \mathbf{u}_n(q),$$

then $X \subset \mathbb{N}$ is S-recognizable if and only if X is T-recognizable.

Proof. Assume that $\Sigma = \{\sigma_1 < \cdots < \sigma_p\} = \{\sigma_{\nu_1} \prec \cdots \prec \sigma_{\nu_p}\}$ where ν is a permutation of $\{1, \ldots, p\}$. We prove that the graph

$$\hat{\xi} = \{(x, y) \in L \times L : \mathrm{val}_S(x) = \mathrm{val}_T(y)\}$$

of $\xi_{S,T}$ is a regular language over the alphabet $\Sigma \times \Sigma$, showing that

$$\mathrm{rep}_T(X) = p_2(\hat{\xi} \cap p_1^{-1}(\mathrm{rep}_S(X))$$

is regular if and only if $\mathrm{rep}_S(X)$ is regular, where $p_1, p_2 : (\Sigma \times \Sigma)^* \to \Sigma^*$ are the canonical homomorphisms of projection.

Let (x, y) belonging to $\hat{\xi}$. The two systems S and T have the same sequence $(\mathbf{v}_n)_{n \in \mathbb{N}}$, thus $|x| = |y|$.

By Algorithm I.5.7, if $|x| \geq n_0$ then

$$x = \underbrace{\sigma_{i_1} \ldots \sigma_{i_l}}_{\alpha} \beta \text{ and } y = \underbrace{\sigma_{i_{\nu_1}} \ldots \sigma_{i_{\nu_l}}}_{\alpha'} \beta'$$

where $|\beta| = |\beta'| = n_0$, $\beta \in L_{s.\alpha}$, $\beta' \in L_{s.\alpha'}$ and[1]

$$\mathrm{val}_{S_{s.\alpha}}(\beta) = \mathrm{val}_{T_{s.\alpha'}}(\beta').$$

To conclude, it is then sufficient to observe that the words of $\hat{\xi}$ of length at least n_0 are exactly the words accepted by the following non-deterministic finite automaton. The set of states is $(Q \times Q) \cup \{f\}$. The initial state is (s, s). The new symbol f denotes the unique final state. According to what precedes, there are two kinds of transitions. First, those of label $(\sigma_i, \sigma_{\nu_i})$ mapping the state (q, q') onto $(q.\sigma_i, q'.\sigma_{\nu_i})$. Second, those of label (β, β') mapping (q, q') onto f, provided that $|\beta| = |\beta'| = n_0$, $\beta \in L_q$, $\beta' \in L_{q'}$ and $\mathrm{val}_{S_q}(\beta) = \mathrm{val}_{T_{q'}}(\beta')$.

□

EXAMPLE V.1.5. The language L over the alphabet $\Sigma = \{a, b\}$ consisting of the words containing an even number of b's satisfies the

[1] If S is a numeration system built on a regular language L accepted by a DFA $\mathcal{A}$ then S_q denotes the system built on the language L_q, q being a state of $\mathcal{A}$.

hypothesis of Proposition V.1.4. Its minimal automaton was given in Figure II.3 page 22 and

$$\mathbf{u}_0(s) = 1, \ \mathbf{u}_0(p) = 0,$$
$$\mathbf{u}_n(s) = \mathbf{u}_n(p) = 2^{n-1}, \ \forall n \geq 1.$$

Hence the transducer depicted in Figure V.2 computing the function $\xi_{S,T}$ where $S = (L, \Sigma, a < b)$ and $T = (L, \Sigma, b < a)$.

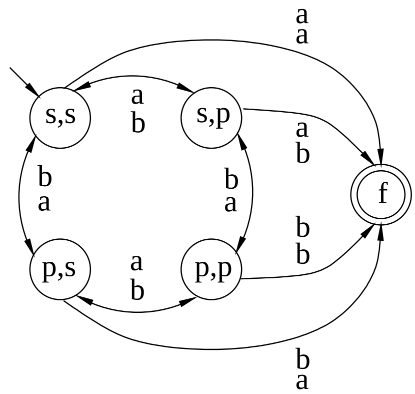

Figure V.2. A transducer computing $\xi_{S,T}$.

In the next proposition we give equivalent formulations of the assumption of Proposition V.1.4. They are expressed in terms of the incidence matrix M of the minimal automaton $\mathcal{A}_L$ of L. Recall that it is the matrix defined by

$$M_{i,j} = \sum_{t=1}^{n} \delta_{q_i.\sigma_t, q_j}, \ 1 \leq i, j \leq \kappa,$$

where the σ_t's and the q_i's denote the n letters and the κ states of $\mathcal{A}_L$ respectively.

We denote by f_L the characteristic vector of the set of final states of $\mathcal{A}_L$:

$$(f_L)_i = \begin{cases} 1 & , \text{ if } q_i \in F; \\ 0 & , \text{ otherwise.} \end{cases}$$

Observe that

$$(M^m f)_i = \sum_{j=1}^{\kappa} (M^m)_{i,j} f_j = \mathbf{u}_m(i). \tag{30}$$

Proposition V.1.6. *Let L be a regular language over an alphabet Σ and let $\mathcal{A}_L = (Q, \Sigma, \delta, s, F)$ be its minimal automaton. Let m be the multiplicity of zero as root of the minimal polynomial of M. Let $r > m$. The following assertions are equivalent:*

(1) $\forall n \geq r, \forall q, q' \in Q, \ \mathbf{u}_n(q) = \mathbf{u}_n(q')$,
(2) $\forall n \geq m, \forall q, q' \in Q, \ \mathbf{u}_n(q) = \mathbf{u}_n(q')$,

(3) $\exists \lambda \in \mathbb{N} \setminus \{0\}$*:* $M^m f = \lambda\, \vec{v}$*, with* $\vec{v} = (1, \ldots, 1)^T$.
In particular, $\forall q \in Q$, $\forall i \geq 0$, $\mathbf{u}_{m+i}(q) = (\#\Sigma)^i \mathbf{u}_m(q)$.

Proof. This follows immediately from (30) and the well-known fact that any polynomial that is annihilated by M is the characteristic polynomial of a linear recurrence equation satisfied by each of the sequences $\mathbf{u}_n(k)$.

□

Here is another easy characterization of the languages for which the assumption of Proposition V.1.4 holds true.

PROPOSITION V.1.7. *Let L be a regular language over an alphabet Σ. Then L satisfies the hypothesis of Proposition V.1.4 if and only if there exist $n_0, \alpha_0 \in \mathbb{N}$ such that for all $w \in \Sigma^*$,* $\#((w^{-1}.L) \cap \Sigma^{n_0}) = \alpha_0$.

Another class of languages for which the recognizable sets of integers are independent of the ordering of the alphabet is the class of slender languages. This class has in fact a wider underlying property and is discussed in the next section.

REMARK V.1.8. *Recognizability of exponential polynomial functions.* We have shown in Sections 3 and 4 of Chapter II, how to build an abstract numeration system $S = (L, \Sigma, <)$ that recognizes the set $f(\mathbb{N})$ where

$$f(n) = \sum_{i=1}^{k} P_i(n)\, \alpha_i^n.$$

The construction was made up to make sure that

$$\mathrm{rep}_S(f(\mathbb{N})) = \mathrm{Min}(L, <).$$

Changing the ordering of Σ does not change the complexity function of the language L and $\mathrm{rep}_S(f(\mathbb{N})) = \mathrm{Min}(L, \prec)$ is regular whatever the ordering $\prec$ is. Indeed, Lemma II.3.1 holds true for any ordering of the alphabet.

REMARK V.1.9. *Relation between abstract and positional systems.* The definition of the $\lambda_{q,\sigma,j}$'s given in (28) depends on the ordering of the alphabet since the $\beta_{q,p,\sigma}$ given in the proof of Proposition III.6.1 are defined by

$$\beta_{q,p,\sigma} = \#\{\sigma' < \sigma : q.\sigma' = p\}.$$

So the transducer obtained in this proof depends on the ordering of the alphabet but whatever the ordering is, a transducer can be obtained. For instance, with the assumptions and notations of Corollary III.6.5, a subset X is U-recognizable if and only if it is S-recognizable for any ordering of the alphabet.

2. Slender languages

The set of slender languages is a second class of languages for which the recognizable sets of integers are independent of the ordering of the alphabet. This property has been introduced in [**39**] and for the sake of completeness, we recall its original proof in this section. But in fact, the slender languages make up a particular class of languages with respect to abstract numeration systems: the recognizable subsets are exactly the ultimately periodic subsets of $\mathbb{N}$.

Definition V.2.1. [**5**] Let d be a positive integer. The language L is said to be *d-slender* if

$$\forall n \geq 0,\ \mathbf{u}_n(L) \leq d,$$

L is said to be *slender* if it is d-slender for some d. A regular language is slender if and only if for some $k \geq 1$ and words x_i, y_i, z_i, $1 \leq i \leq k$,

$$L = \bigcup_{i=1}^{k} x_i\, y_i^* z_i.$$

In this case, L is said to be a *union of single loops* [**52**], [**63**]. Moreover, we can assume that the sets $x_i\, y_i^* z_i$ are pairwise disjoint.

The next proposition was first stated in [**39**]. We leave it in this work, only for the construction given in its proof. In fact, it is a consequence of a wider result given below.

Proposition V.2.2. *Let d be a positive integer. Let L be a regular d-slender language. Let $S = (L, \Sigma, <)$ and $T = (L, \Sigma, \prec)$ be two numeration systems. If $X \subset \mathbb{N}$ is S-recognizable, then X is T-recognizable.*

Proof. As in the proof of Proposition V.1.4, we show that the graph $\widehat{\xi_{S,T}}$ of the change of systems is regular. Using Lemma II.3.1, we define iteratively the regular languages $I_{i,<}$ and $I_{i,\prec}$ by

$$\left\{\begin{array}{lcl} I_{1,<} & = & \mathrm{Min}(L,<) \\ I_{1,\prec} & = & \mathrm{Min}(L,\prec), \end{array}\right.$$

and, for $i = 2, \ldots, d$,

$$\left\{\begin{array}{lcl} I_{i,<} & = & \mathrm{Min}[L \setminus (\bigcup_{j=1}^{i-1} I_{j,<}), <] \\ I_{i,\prec} & = & \mathrm{Min}[L \setminus (\bigcup_{j=1}^{i-1} I_{j,\prec}), \prec]. \end{array}\right.$$

Since for all $x \in L$, $|x| = |\xi(x)|$, the graph of ξ is thus given by

$$\hat{\xi} = \bigcup_{j=1}^{d} \left[(I_{j,<} \times I_{j,\prec}) \cap (\Sigma \times \Sigma)^* \right].$$

$\square$

THEOREM V.2.3. *Let $L \subset \Sigma^*$ be a slender language and $S = (L, \Sigma, <)$. A set $X \subset \mathbb{N}$ is S-recognizable if and only if X is a finite union of arithmetic progressions.*

Proof. By the characterization of the slender languages, we have

$$L = \bigcup_{i=1}^{k} x_i\, y_i^* z_i \cup F_0,\ \ x_i, z_i \in \Sigma^*, y_i \in \Sigma^+$$

where the sets $x_i\, y_i^* z_i$ are pairwise disjoint and F_0 is a finite set.

(i) The sequence $(\mathbf{u}_n(L))_{n\in\mathbb{N}}$ is ultimately periodic of period

$$C = \mathrm{lcm}_j |y_j|.$$

Moreover, for n large enough, if $x_i\, y_i^n\, z_i$ is the l^{th} word of length $|x_i\, z_i| + n\,|y_i|$ then $x_i\,(y_i^{n+C/|y_i|})\, z_i$ is the l^{th} word of length $|x_i\, z_i| + n\,|y_i| + C$. In other words, for n sufficiently large, the structures of the ordered sets of words of length n and $n + C$ are the same.

(ii) The regular subsets of L are of the form

$$\bigcup_{j\in J} x_j\, (y_j^{\alpha_j})^* z_j \cup F_0'$$

where $J \subset \{1, \ldots, k\}$, $\alpha_j \in \mathbb{N}$ for $j \in J$ and F_0' is a finite subset of L.

(iii) If X is S-recognizable, then $\mathrm{rep}_S(X)$ is a regular subset of L and we can apply *(ii)*. In view of *(i)*, it is clear that X is ultimately periodic. The converse is immediate by Theorem II.2.1.

□

This theorem has direct corollaries.

COROLLARY V.2.4. *Let S be a numeration system built on a slender language. If $X \subset \mathbb{N}$ is S-recognizable then αX is S-recognizable for all $\alpha \in \mathbb{N}$.*

REMARK V.2.5. This corollary is relevant with Theorem III.4.1 since the complexity function of a slender language is bounded by a constant.

COROLLARY V.2.6. *Let S be a numeration system built on a slender language. If X and Y are S-recognizable subsets of $\mathbb{N}$ then $X + Y$ is S-recognizable.*

3. Dependent systems and recognizability

For the representation of integers in an abstract numeration system, a lot of open questions remain. The most challenging is certainly the possible generalization of Cobham's theorem. We know from Theorem II.2.1 that the ultimately periodic sets are recognizable in any system but what could be the "multiplicatively independent condition" ? If the regular languages on which numeration systems are built, are exponential with a unique dominant root, a guess would be to consider two systems to be independent if their respective dominant roots are multiplicatively independent. Having in mind Corollary IV.4.4, a possible path would be to connect our results with Durand's results [**25**] and [**26**].

It is well know that if m and n are multiplicatively dependent integers then m-recognizable sets and n-recognizable sets coincide (any formula definable in $\langle \mathbb{N}, +, V_n \rangle$ is definable in $\langle \mathbb{N}, +, V_m \rangle$ and the converse also holds true). In this small section, we consider exponential languages having the same unique dominant root. We exhibit subsets which are recognizable in a numeration system and which are not recognizable in another system having the same dominant root. Roughly speaking, we have possibly "dependent" systems with different recognizable sets of integers.

The technique used here is the same as the one found in Theorem III.5.3. In the different examples, we consider complement of polynomial languages of different degree, next complement of polynomial languages of the same degree and finally exponential language with exponential complement.

EXAMPLE V.3.1. Let $\Sigma = \{a, b\}$,

$$L_1 = \Sigma^* \setminus a^*ba^*ba^*ba^*$$

and

$$L_2 = \Sigma^* \setminus a^*b^*.$$

We consider the systems $S_i = (L_i, \Sigma, a < b)$, $i = 1, 2$. An easy computation shows that

$$\mathbf{u}_n(L_1) = 2^n - \frac{n^3}{6} + \frac{n^2}{2} - \frac{n}{3}$$

and

$$\mathbf{u}_n(L_2) = 2^n - n - 1.$$

Thus, the two systems have the same dominant root. We consider the subset X of $\mathbb{N}$ such that

$$\mathrm{rep}_{S_1}(X) = \mathrm{Min}(L_1, <) = a^*.$$

One has

$$X = \mathrm{val}_{S_1}(a^*) = \{\frac{1}{24}(-24 + 3.2^{n+4} - 2\,n + n^2 + 2\,n^3 - n^4) : n \in \mathbb{N}\}.$$

We show that $\mathrm{rep}_{S_2}(X)$ is not a regular language using the same technique as in Theorem III.5.3. The first words of $\mathrm{rep}_{S_2}(X)$ are given in Table V.3. We see that the number of leading b's seems to increase as well as the length of the tail. We have

$$\mathbf{v}_n(L_2) = 2^{n+1} - 2 - \frac{n(n+1)}{2} - n.$$

For n large enough, the n^{th} element of X,

$$X_n = 2^{n+1} - \frac{n}{12} + \frac{n^2}{24} + \frac{n^3}{12} - \frac{n^4}{24} - 1,$$

X	L_2
385	b $ababbba$
813	b $abbaabaa$
1717	b $abbbbabbb$
3600	bb $aababbbba$
7476	bb $abbaaabbbb$
15382	bbb $aaabbbbbbb$
31402	bbb $abbaabaaaba$
63715	$bbbb$ $aababbababb$
128691	$bbbb$ $abbbabaabbaa$
259083	$bbbbb$ $abaababbabba$
520411	$bbbbbb$ $aaabbaabbaab$
1043730	$bbbbbb$ $abbabbbbaabaa$
2091166	$bbbbbbb$ $abaabbaaaabab$
4186988	$bbbbbbbb$ $aabaaabbabaab$
8379752	$bbbbbbbb$ $abbbbaabbbbbaa$
16766589	$bbbbbbbbb$ $ababbbbababaab$
33541781	$bbbbbbbbbb$ $aabbbbbbabbaba$
67093913	$bbbbbbbbbbb$ $aaabbabbbbbaaa$
134200177	$bbbbbbbbbbb$ $abbbbaabbbababb$

TABLE V.3. The first words of $\text{rep}_{S_2}(X)$

has a representation in S_2 of length n since
$$\mathbf{v}_{n-1}(L_2) \le X_n \le \mathbf{v}_n(L_2).$$
With the notation given in Definition III.5.2, it is clear that for each n, there exists a unique k such that
$$\underbrace{\mathbf{u}_n((L_2)_{b^{k+1}})}_{=2^{n-k-2}} < \underbrace{\mathbf{v}_n(L_2) - X_n}_{\sim n^4/24} \le \underbrace{\mathbf{u}_n((L_2)_{b^k})}_{=2^{n-k-1}}.$$
As a function of n, k is increasing and it is clear that $n-k \to \infty$ if $n \to \infty$ since the growth of k is logarithmic. We conclude using the so-called pumping lemma (Lemma III.5.1).

EXAMPLE V.3.2. We consider the languages $L_1 = \Sigma^* \setminus a^*ba^*$ and $L_2 = \Sigma^* \setminus a^*b^*$ where $\Sigma = \{a, b\}$. So,
$$\mathbf{u}_n(L_1) = 2^n - n \text{ and } \mathbf{u}_n(L_2) = 2^n - n - 1.$$
Let us consider the set $X \subset \mathbb{N}$ such that
$$\text{rep}_{S_1}(X) = \text{Min}(L_1, <) = a^*$$
and $X_n = 2^{n+1} - 1 - \frac{1}{2}n(n+1)$. One can verify that for n large enough, $|\text{rep}_{S_2}(X_n)| = n+1$. In L_2, the first word of length $n+1$ begins with $n-1$ letters a followed by ba (notice that $\mathbf{u}_2(L_2) = 1$). Next, there are $\mathbf{u}_3(L_2)$ words beginning with at least $n-2$ letters a and so on.

Therefore $\mathrm{rep}_{S_2}(X_n)$ begins with exactly $n-i$ letters a followed by a tail of length $i+1$ if

$$\mathbf{v}_n(L_2)+\mathbf{u}_{i-1}(L_2)\leq X_n<\mathbf{v}_n(L_2)+\mathbf{u}_i(L_2)$$

i.e., if

$$2^{i-1}-i-1\leq n<2^i-i-2.$$

Hence the conclusion.

EXAMPLE V.3.3. In this last example, we consider exponential languages with exponential complement. Let $\Sigma=\{a,b\}$, $L_1=a\Sigma^*$ and $L_2=a\Sigma^*\cup\Sigma^*bb\Sigma^*$. One has,

$$\mathbf{u}_n(L_2)=2^n-\frac{\sqrt{5}}{5}\left(\frac{1+\sqrt{5}}{2}\right)^n+\frac{\sqrt{5}}{5}\left(\frac{1-\sqrt{5}}{2}\right)^n.$$

We set X such that

$$\mathrm{rep}_{S_2}(X)=\mathrm{Min}(\mathrm{L}_2,<)$$

and

$$X_n=2^{n+1}+\frac{3\sqrt{5}-5}{10}\left(\frac{1-\sqrt{5}}{2}\right)^n-\frac{3\sqrt{5}+5}{10}\left(\frac{1+\sqrt{5}}{2}\right)^n-1.$$

For n large enough, one has

$$\mathbf{v}_n(L_1)\leq X_n<\mathbf{v}_{n+1}(L_1)$$

thus $|\mathrm{rep}_{S_1}(X_n)|=n+1$ and

$$\underbrace{\mathbf{u}_{n+1}((L_1)_{ab^{i+1}})}_{2^{n-i-1}}<\underbrace{\mathbf{v}_{n+1}(L_1)-X_n}_{\sim((1+\sqrt{5})/2)^n}\leq\underbrace{\mathbf{u}_{n+1}((L_1)_{ab^i})}_{=2^{n-i}}.$$

As a function of n, i is increasing and $n-i\to\infty$ if $n\to\infty$.

CHAPTER VI

Representing real numbers

The other chapters of this work were wholly related to the representation of natural numbers. In a classical numeration system, the representation of a non-negative real number x can be achieved in two steps. First the representation of its integer part $\lfloor x \rfloor$ is computed by the greedy algorithm. Next, an infinite word over a finite alphabet of digits is used to represent its fractional part $\{x\}$. In this chapter, we give some results concerning the representation of non-negative real numbers using an abstract numeration system. In particular, we have to set up hypotheses for the convergence of the numerical approximations given by an infinite word. Once again, these hypotheses rely on the complexity function of regular languages. The here obtained results generalize classical results in an integer base and in particular, give a deeper understanding of the fact that a rational number could have two different representations in a k-ary numeration system. Here, a real number can have a finite number, a countable number or even an uncountable number of abstract representations. The number of representations is determined by the complexity functions of the languages accepted from the different states of the minimal automaton. We also show that if $\theta > 1$ is a Pisot number then classical θ-developments and our representations of real numbers for the language of the numeration associated to the Bertrand numeration system related to θ coincide. It is interesting to note that we shall be confronted with two different kinds of convergence: numerical convergence of real numbers but also convergence of words to an infinite word. The material of this chapter was introduced in [**40**].

1. Some definitions

If w is a finite (resp. a right infinite) word, we enumerate the letters of w starting with 0 and if $k \le l < |w|$ (resp. $k \le l$), $w[k,l]$ denotes the factor of w starting at the k^{th} position and ending at the l^{th} one. We simply write w_j instead of $w_{[j,j]}$, $0 \le j < |w|$.

A sequence $(w_n)_{n\in\mathbb{N}}$ of words belonging to Σ^* converges to an infinite word $w \in \Sigma^\omega$ and we write $\lim_{n\to\infty} w_n = w$, if

$$\forall l \in \mathbb{N}, \exists N \in \mathbb{N} : \forall n > N,\ w_n[0,l] = w[0,l].$$

In the same way, a sequence $(w_n)_{n\in\mathbb{N}}$ of infinite words converges to $w \in \Sigma^\omega$ if the same condition is satisfied. One can endow Σ^ω with a distance d defined as follows. Let $x, y \in \Sigma^\omega$, $d(x,y) = 2^{-n}$ where $n = \inf\{j : x_j \neq y_j\}$ and $n = \infty$ if $x = y$. With this distance, Σ^ω is a topological space and the previous definition of a convergent sequence of elements of Σ^ω in terms of common prefixes is relevant. Observe that any finite word $x \in \Sigma^*$ can be viewed as an infinite word $x\zeta^\omega \in (\Sigma \cup \{\zeta\})^\omega$, $\zeta \notin \Sigma$. So a sequence $(w_n)_{n\in\mathbb{N}}$ of finite words converges to an infinite word w if and only if $(w_n\zeta^\omega)_{n\in\mathbb{N}}$ converges to w in the topological space $(\Sigma \cup \{\zeta\})^\omega$.

We shall use the notation $\exists^\omega n$ as an abbreviation for "there exist infinitely many n".

2. Positioning the problem

In the classical numeration system in base 10, a real number $r \in \mathbb{R} \setminus \mathbb{N}$ is represented by an infinite word v which can be decomposed in three parts,

$$v = v_1, v_2\, v_3.$$

The integer part of r is represented by v_1. Its fractional part is represented by v_2 which contains a finite number of zeroes and v_3 which is an infinite word in $\{1,\ldots,9\}\{0,\ldots,9\}^\omega$. As an example,

$$\frac{16501}{1100} = \underbrace{15}_{v_1}, \underbrace{000}_{v_2}\underbrace{909090\cdots}_{v_3}.$$

Observe that an infinite word $v_3 = a_1a_2\cdots$ such that $a_1 \neq 0$ can be viewed as the representation of a real number x in $[\frac{1}{10}, 1]$. The prefixes $a_1 \cdots a_j$, $j \geq 1$, of this word make up a sequence of finite words convergent to the infinite word v_3. Moreover, each of these prefixes gives a numerical approximation of r,

$$\frac{a_1}{10} + \cdots + \frac{a_j}{10^j} = \frac{\pi_{10}(a_1\cdots a_j)}{10^j}. \tag{31}$$

The sequence of these approximations is convergent to x. To represent the interval $]0,1[$, we have to represent not only $[\frac{1}{10}, 1]$ but $[\frac{1}{10^{p+1}}, \frac{1}{10^p}]$ for each $p \in \mathbb{N}$. It is the reason of the presence of v_2. If $10^{-p-1} \leq x \leq 10^{-p}$ then $10^p\, x \in [\frac{1}{10}, 1]$, this latter number being represented by a word of the kind of v_3. To obtain the representation of x instead of $10^p\, x$, it is enough to know the exponent p. Therefore, $v_2 = 0^p$ can be viewed as a "coding" of p.

In this chapter, we want to extend in a natural way this representation of real numbers to abstract numeration systems. Let us consider the system $S = (L, \Sigma, <)$. The word v_3 representing an element of $[\frac{1}{10}, 1]$ can be replaced in a first guess by an infinite word in Σ^ω having an infinite number of prefixes w_n in L. In (31), the numerical approximation given by $a_1 \cdots a_j$ has its numerator equal to the numerical value of $a_1 \cdots a_j$ and the denominator, 10^j, is exactly the number of words of length not greater than j in the language of the numeration

$\{1,\ldots,9\}\{0,\ldots,9\}^* \cup \{0\}$. Consequently, for the numerical approximation given by w_n, we suggest to replace (31) with

$$\frac{\mathrm{val}_S(w_n)}{\mathbf{v}_{|w_n|}(L)}.$$

Instead of an infinite word w having an infinite number of prefixes in L, we can consider the weaker condition of a sequence $(w_n)_{n\in\mathbb{N}}$ of words in L convergent to w and such that $(|w_n|)_{n\in\mathbb{N}}$ is strictly increasing. (In this case w_n is not necessarily a prefix of w.) A part of this chapter is dedicated to the convergence of the numerical sequence $\left(\frac{\mathrm{val}_S(w_n)}{\mathbf{v}_{|w_n|}}\right)_{n\in\mathbb{N}}$. It is obvious that

$$\frac{\mathbf{v}_{|w_n|-1}}{\mathbf{v}_{|w_n|}} \le \frac{\mathrm{val}_S(w_n)}{\mathbf{v}_{|w_n|}} \le \frac{\mathbf{v}_{|w_n|}-1}{\mathbf{v}_{|w_n|}}.$$

As a consequence of Theorem III.2.12, if L is a polynomial regular language then for any sequence $(w_n)_{n\in\mathbb{N}}$ increasing in length,

$$\lim_{n\to\infty} \frac{\mathrm{val}_S(w_n)}{\mathbf{v}_{|w_n|}(L)} = 1.$$

Consequently, for the representation of real numbers, we have to consider abstract systems built on an exponential language. Roughly speaking, if the complexity function of L has a dominant term of the form θ^n, $\theta > 1$, the interval $[\frac{1}{10}, 1]$ obtained in base 10 will be replaced by $[\frac{1}{\theta}, 1]$ (details will follow in the other sections). As for the decimal system, the representation of the interval $[\theta^{-p-1}, \theta^{-p}]$, $p \in \mathbb{N}$, relies simply on some conventions.

To conclude, it is interesting to note that classical θ-representations and representations of real numbers in an abstract numeration system built on a Pisot number θ coincide (see Section 8). Recall that for any real numbers $\theta > 1$ and $x \ge 0$, any sequence $(\epsilon_n)_{n\in\mathbb{N}}$ of non-negative integers such that $x = \sum_{i=0}^{\infty} \epsilon_i\, \theta^{-i}$ is a *θ-representation* of x. A distinguished θ-representation is the *θ-development* of x. It is computed through the use of the greedy algorithm: $x_0 = \lfloor x \rfloor$, $r_0 = \{x\}$ and for $n > 0$, $x_n = \lfloor \theta\, r_{n-1} \rfloor$ and $r_n = \{\theta\, r_{n-1}\}$ with

$$\lfloor x \rfloor = \sup\{y \in \mathbb{N} \mid y \le x\}$$

and

$$\{x\} = x - \lfloor x \rfloor.$$

So, $x = \sum_{i=0} r_i\, \theta^{-i}$. (For more, see [**55**].)

Example VI.2.1. Consider again the numeration system S introduced in Example III.6.6. We shall see that

$$\lim_{n\to\infty} \frac{\mathrm{val}_S((ba)^n c)}{\mathbf{v}_{2n+1}(s)} = \frac{1}{1+\sqrt{3}} + \frac{3}{9+5\sqrt{3}} \simeq 0.53589838486224541.$$

Table VI.1 gives some numerical approximations of that number expressed in S.

w	$\mathrm{val}_S(w)$	$\mathbf{v}_{\|w\|}$	$\mathrm{val}_S(w)/\mathbf{v}_{\|w\|}$
bc	8	12	0.66666666666666667
bac	19	34	0.55882352941176471
$babc$	52	94	0.55319148936170213
$babac$	139	258	0.53875968992248062
$bababc$	380	706	0.53824362606232295
$bababac$	1035	1930	0.53626943005181347
$babababc$	2828	5274	0.53621539628365567
$babababac$	7723	14410	0.53594725884802221
$bababababc$	21100	39370	0.53594107188214376
$bababababac$	57643	107562	0.53590487346832524
$babababababc$	157484	293866	0.53590411956469956
$babababababac$	430251	802858	0.53589924992962641
$\vdots$	$\vdots$	$\vdots$	$\vdots$

TABLE VI.1. Some numerical approximations.

DEFINITION VI.2.2. Let L be a language. We denote by $L_\infty \subset \Sigma^\omega$ the set of infinite words having an infinite number of prefixes belonging to L,

$$L_\infty = \{w \in \Sigma^\omega | \exists^\omega n : w[0,n] \in L\}.$$

Let us introduce the set

$$\mathcal{L}_\infty = \{w \in \Sigma^\omega | \exists (w_n)_{n\in\mathbb{N}} \in L^{\mathbb{N}} : \lim_{n\to\infty} w_n = w\}.$$

Notice that $L_\infty \subset \mathcal{L}_\infty$.

EXAMPLE VI.2.3. Consider the language L accepted by its minimal automaton depicted in Figure VI.1. For instance, it is clear that $(ab)^\omega$

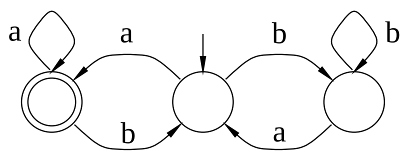

FIGURE VI.1. The minimal automaton of a language L such that $L_\infty \neq \mathcal{L}_\infty$.

belongs to L_∞. One also has $(ba)^\omega \in \mathcal{L}_\infty$ because

$$\lim_{n\to\infty} (ba)^n a = (ba)^\omega$$

but this infinite word does not belong to L_∞.

Some natural problems related to the representation of real numbers addressed in this chapter are:

- Characterizing regular languages L for which L_∞ or $\mathcal{L}_\infty$ is uncountably infinite.

- Giving assumptions insuring the existence of
$$\lim_{n\to\infty} \frac{\mathrm{val}_S(w_n)}{\mathbf{v}_{|w_n|}}.$$
- Showing that any sequence $(w_n)_{n\in\mathbb{N}}$ of words belonging to L such that $\lim_{n\to\infty} w_n = w \in \mathcal{L}_\infty$ gives a sequence of approximations converging to a unique real r_w.
- Showing that for some interval $I \subset \mathbb{R}$, any $x \in I$ can be represented by an infinite limit word of $\mathcal{L}_\infty$.

3. Regular languages with uncountably infinite L_∞ or $\mathcal{L}_\infty$

Since we want to represent an interval of $\mathbb{R}$, it is natural to seek under which conditions L_∞ or $\mathcal{L}_\infty$ has uncountably many elements. In the other sections of this chapter, we shall mainly be interested in the set $\mathcal{L}_\infty$.

3.1. Languages with uncountable $\mathcal{L}_\infty$. Here it is a characterization of the regular languages with uncountable $\mathcal{L}_\infty$ in terms of automata.

PROPOSITION VI.3.1. *The set $\mathcal{L}_\infty$ is uncountably infinite if and only if there exist two distinct cycles $(p_1, \ldots, p_r, p_1)$ and $(q_1, \ldots, q_t, q_1)$ in a DFA accepting L such that*

(1) $p_1 = q_1$
(2) *there is an accessible*[1] *state in* $\{p_1, \ldots, p_r, q_1, \ldots, q_t\}$
(3) *there is a coaccessible*[2] *state in* $\{p_1, \ldots, p_r, q_1, \ldots, q_t\}$.

Proof. The condition is sufficient. Let c (resp. d) be the accessible (resp. coaccessible) state given in the assumptions. Thus, there exist $w, w' \in \Sigma^*$ such that $s.w = c$ and that $d.w'$ is a final state. The cycle $(p_1, \ldots, p_r, p_1)$ is labeled by a word $\alpha_0 = m_1 \cdots m_r$, that is
$$p_1 \xrightarrow{m_1} p_2 \xrightarrow{m_2} p_3 \cdots p_{r-1} \xrightarrow{m_{r-1}} p_r \xrightarrow{m_r} p_1.$$
In the same way, the cycle $(q_1, \ldots, q_t, q_1)$ is labeled by a word α_1. It is clear that there exist two words, v, v' such that $c.v = p_1$ and $p_1.v' = d$. Let $f : \mathbb{N} \to \{0, 1\}$ be a function. With this function, we build the following sequence $(w_n)_{n\in\mathbb{N}}$ of words belonging to L, $\forall i \in \mathbb{N}$,
$$w_i = wv\alpha_{f(0)} \cdots \alpha_{f(i)} v'w'.$$
This sequence is convergent to a limit word w_f of $\mathcal{L}_\infty$. If $f \neq g$ then it is obvious that $w_f \neq w_g$. Moreover, the set of functions of domain $\mathbb{N}$ and codomain $\{0, 1\}$ is uncountable.

We now show the converse. Let us assume that any state lying on a path starting in s and ending in a final state belongs to at most one

[1] A state q is *accessible*, if there exists a path starting in the initial state and ending in q.

[2] A state q is *coaccessible*, if there exists a path from q to a final state.

cycle. In other words, if $xyz \in L$, $y \neq \varepsilon$ and if $s.x$ belongs to a cycle $(s.x, p_2, \dots, p_r, s.x)$ and $s.xy$ belongs to a cycle $(s.xy, q_2, \dots, q_t, s.xy)$ then $\{s.x, p_2, \dots, p_r\} \cap \{s.xy, q_2, \dots, q_t\} = \emptyset$. Therefore L is a finite union of languages of the form

$$\lambda_1 \mu_1^* \lambda_2 \mu_2^* \cdots \lambda_k \mu_k^* \lambda_{k+1},\ \lambda_i, \mu_i \in \Sigma^*. \tag{32}$$

If $m \in \mathcal{L}_\infty$ then by definition, there exist words of L having an arbitrary long common prefix with m. These words belong to a set of the form (32), so m has one of the following forms

$$\lambda_1\mu_1^\omega,\ \lambda_1\mu_1^{n_1}\lambda_2\mu_2^\omega, \dots, \lambda_1\mu_1^{n_1}\lambda_2\mu_2^{n_2}\cdots\mu_{k-1}^{n_{k-1}}\lambda_k\mu_k^\omega,\ n_1, \dots, n_{k-1} \in \mathbb{N}.$$

There is a countable number of such words. So $\mathcal{L}_\infty$ is a finite union of countable sets, a contradiction.

□

Remark VI.3.2. A finite union of languages of form (32) is a polynomial language (see Chapter III Section 2). So a regular language L such that $\mathcal{L}_\infty$ is uncountable, is exponential.

3.2. Languages with uncountable L_∞. An obvious adaptation of the previous proof leads to the following characterization of the rational languages with uncountable L_∞.

Proposition VI.3.3. *The set L_∞ is uncountably infinite if and only if there exist two distinct cycles $(p_1, \dots, p_r, p_1)$ and $(q_1, \dots, q_t, q_1)$ in a DFA accepting L such that*

(1) $p_1 = q_1$
(2) *there is an accessible state in* $\{p_1, \dots, p_r, q_1, \dots, q_t\}$
(3) *there exist $i \leq r$, $j \leq t$ such that p_i and q_j are final states.*

For instance, the language of Example III.6.6 is such that $L_\infty = \mathcal{L}_\infty$ is uncountably infinite.

Remark VI.3.4. The subset L_∞ of $\mathcal{L}_\infty$ is interesting in terms of Büchi automaton. Recall that a *Büchi automaton* is a nondeterministic finite automaton which accepts infinite words if reading the word from left to right, a final state is reached infinitely often (see for instance [**65**]). So we can use the minimal automaton of L as a Büchi automaton to recognize L_∞.

4. The basic assumptions

From now on, we shall assume the following.

Hypotheses. The set $\mathcal{L}_\infty$ is uncountable and for all state[3] q of $\mathcal{A}_L$, either

[3]In the following of this chapter, we use by convenience the minimal automaton of L but this automaton could be replaced by any accessible DFA accepting the same language.

(i) $\exists N_q \in \mathbb{N} : \forall n > N_q,\ \mathbf{u}_n(q) = 0$

or

(ii) $\exists \theta_q \geq 1,\ P_q(x) \in \mathbb{R}[x]\ ,\ b_q > 0 :$

$$\lim_{n\to\infty} \frac{\mathbf{u}_n(q)}{P_q(n)\,\theta_q^n} = b_q.$$

Notation. By Remark VI.3.2, situation (i) never occurs for the initial state s. Let us denote $\theta_s > 1$, P_s and b_s respectively by θ, P and a_s.

Consequence 1. For any state q such that (ii) occurs, either $\theta_q < \theta$ or either $\theta_q = \theta$ and $d(P_q) \leq d(P)$, where $d(P)$ is the degree of the polynomial P. Indeed, let us assume that there exist $\theta_q > 1$ and $P_q(x) \in \mathbb{R}[x]$ such that

$$\frac{P_q(n)\,\theta_q^n}{P(n)\,\theta^n}$$

is unbounded and that

$$\lim_{n\to\infty} \frac{\mathbf{u}_n(q)}{P_q(n)\,\theta_q^n} = b_q > 0.$$

Since $\mathcal{A}_L$ is accessible (this is a property of the minimal automaton), there exists a constant i such that $\mathbf{u}_n(s) \geq \mathbf{u}_{n-i}(q)$. So,

$$\frac{\mathbf{u}_n(s)}{P_q(n)\,\theta_q^n} \geq \frac{\mathbf{u}_{n-i}(q)}{P_q(n-i)\,\theta_q^{n-i}}\ \frac{1}{\theta_q^i}\ \frac{P_q(n-i)}{P_q(n)} \to \frac{b_q}{\theta_q^i} > 0.$$

This is a contradiction, because a subsequence of

$$\frac{\mathbf{u}_n(s)}{P_q(n)\,\theta_q^n} = \frac{\mathbf{u}_n(s)}{P(n)\,\theta^n}\ \frac{P(n)\,\theta^n}{P_q(n)\,\theta_q^n}$$

converges to 0.

Consequence 2. The limit $\lim_{n\to\infty} \frac{\mathbf{u}_n(q)}{P(n)\,\theta^n}$ exists for any state q and we denote by a_q its non-negative value. Indeed,

$$\frac{\mathbf{u}_n(q)}{P(n)\,\theta^n} = \underbrace{\frac{\mathbf{u}_n(q)}{P_q(n)\,\theta_q^n}}_{\to b_q}\ \frac{P_q(n)\,\theta_q^n}{P(n)\,\theta^n}$$

where the second factor in the r.h.s. is bounded.

Remark VI.4.1. By Remark VI.3.2, we know that the language is exponential thus it is quite natural to hope that $\mathbf{u}_n(s) \sim P(n)\,\theta^n$. In the same way, any sequence $\mathbf{u}_n(q)$ satisfies some recurrence relation with coefficients in $\mathbb{Z}$ and therefore, we can also expect that $\mathbf{u}_n(q) \sim P_q(n)\,\theta_q^n$. But notice that, with these hypotheses, we loose at least languages of the form $M \cap (\Sigma^j)^*$ where M is rational and $n > 1$ (in this case $\mathbf{u}_n(s) = 0$ infinitely often and the sequence $(\frac{\mathbf{u}_n(q)}{P_q(n)\theta_q^n})_{n\in\mathbb{N}}$ is not convergent).

EXAMPLE VI.4.2. A class of languages satisfying the hypotheses introduced in this section is the set of languages for which the assumptions of Proposition III.6.1 and Corollary III.6.5 are verified. If L is such a language and if θ is the Pisot number introduced in Corollary III.6.5 then for every state q of $\mathcal{A}_L$, $\frac{\mathbf{u}_n(q)}{\theta^n}$ converges when n tends to infinity.

5. Convergence of numerical approximations

In this section, we show that if $(w_n)_{n\in\mathbb{N}}$ is a convergent sequence of words belonging to L then the numerical sequence $(\frac{\mathrm{val}_S(w_n)}{\mathbf{v}_{|w_n|}})_{n\in\mathbb{N}}$ is also convergent. So, it will be meaningful to say that a word of $\mathcal{L}_\infty$ is a representation of a real number (or an S-representation if we want to refer to the numeration system S).

REMARK VI.5.1. Recall (5). If $\mathcal{A}=(Q,\Sigma,\delta,s,F)$ is a DFA accepting L, $p\in Q$ and $w\in L_p$ then

$$\mathrm{val}_p(w)=\sum_{q\in Q}\sum_{i=0}^{|w|-1}\beta_{q,|w|-i-1}(p,w)\,\mathbf{u}_i(q).$$

Observe that the most significant coefficients, i.e., the coefficients of the $\mathbf{u}_{|w|-1}(q)$'s are here indexed by zero. It is worth noting that if x and xy belong to L_p, it is obvious that for any state q and for $i=0,\ldots,|x|-1$,

$$\beta_{q,i}(p,x)=\beta_{q,i}(p,xy).$$

In other words, the first n letters of a word w determine completely the coefficients $\beta_{q,i}(p,w)$ for $i=0,\ldots,n-1$.

EXAMPLE VI.5.2. Consider the language of Example III.6.6. Using (5), we have

$$\begin{aligned}
\mathrm{val}_s(abc) &= \mathbf{1}\,\mathbf{u}_2(s)+\mathbf{1}\,\mathbf{u}_1(s)+\mathbf{2}\,\mathbf{u}_0(s)+\mathbf{1}\,\mathbf{u}_0(t)+\mathbf{1}\,\mathbf{u}_1(p)\\
\mathrm{val}_s(abcba) &= \mathbf{1}\,\mathbf{u}_4(s)+\mathbf{1}\,\mathbf{u}_3(s)+\mathbf{2}\,\mathbf{u}_2(s)+1\,\mathbf{u}_1(s)+1\,\mathbf{u}_0(s)+\mathbf{1}\,\mathbf{u}_2(t)\\
&\quad +1\,\mathbf{u}_1(t)+\mathbf{1}\,\mathbf{u}_3(p).
\end{aligned}$$

In the following, we shall be interested only in the numerical value of words y (i.e., $\mathrm{val}_S(y)=\mathrm{val}_s(y)$). So the notation $\beta_{q,i}(p,y)$ can be replaced by a simpler one, $\beta_{q,i}(y)$, since p will be fixed to the initial state s. As a consequence of the latter remark, notice that to any $w\in\mathcal{L}_\infty$, it corresponds, for each state q, a unique sequence $(\beta_{q,n}(w))_{n\in\mathbb{N}}$ of integers. Therefore, if $(w_n)_{n\in\mathbb{N}}$ is a sequence of words of L convergent to w,

$$\lim_{n\to\infty}\frac{\mathrm{val}_S(w_n)}{\mathbf{v}_{|w_n|}}=\lim_{n\to\infty}\sum_{q\in Q}\frac{\sum_{i=0}^{n}\beta_{q,n-i}(w)\,\mathbf{u}_i(q)}{\mathbf{v}_n(s)}.$$

Indeed, the sequence in the l.h.s. is a subsequence of the one in the r.h.s.; if the latter is convergent, then the former converges to the same limit.

In the last part of this section, we fix the limit word w. So, we simplify the notation and replace $\beta_{q,j}(w)$ by $\beta_{q,j}$ since it does not lead to confusion.

PROPOSITION VI.5.3. *If q is a state of $\mathcal{A}_L$ such that $a_q > 0$ then*

i.
$$\lim_{n\to\infty} \frac{\sum_{i=0}^{n} \mathbf{u}_i(q)}{\sum_{i=0}^{n} \mathbf{u}_i(s)} = \frac{a_q}{a_s}$$

ii.
$$\lim_{n\to\infty} \frac{\mathbf{u}_n(q)}{\sum_{i=0}^{n} \mathbf{u}_i(q)} = \frac{\theta - 1}{\theta}$$

iii.
$$\lim_{n\to\infty} \frac{\sum_{i=0}^{n} \beta_{q,n-i}\, \mathbf{u}_i(q)}{\mathbf{u}_n(q)} = \sum_{j=0}^{\infty} \beta_{q,j}\, \theta^{-j}.$$

Otherwise, $a_q = 0$ and
$$\lim_{n\to\infty} \frac{\sum_{i=0}^{n} \beta_{q,n-i}\, \mathbf{u}_i(q)}{\mathbf{v}_n(s)} = 0.$$

Proof. Let us assume $a_q > 0$. Let r be the degree of P. Then $P(n) = \alpha\, n^r + Q(n)$ with $d(Q) < r$ and some $\alpha > 0$. We have,
$$\frac{\mathbf{u}_n(q)}{\alpha\, n^r \theta^n} - \frac{\mathbf{u}_n(q)}{P(n)\theta^n} = \frac{\mathbf{u}_n(q)}{P(n)\theta^n}\, \frac{Q(n)}{\alpha\, n^r} \to 0$$
because $\frac{\mathbf{u}_n(q)}{P(n)\theta^n} \to a_q$. We may thus assume that
$$\lim_{n\to\infty} \frac{\mathbf{u}_n(q)}{n^r \theta^n} = a_q$$
replacing a_q by $\alpha\, a_q$ if necessary. So for $p \in \{q, s\}$, there exists a sequence $(\alpha_n^p)_{n\in\mathbb{N}}$ convergent to 1 such that

$$\mathbf{u}_n(p) = \alpha_n^p\, a_p\, n^r \theta^n. \tag{33}$$

For $k > 1$, there exists some $K > 1$ such that

$$n > K \Rightarrow \alpha_n^s,\ \alpha_n^q \in \left[1 - \frac{1}{k}, 1 + \frac{1}{k}\right]. \tag{34}$$

The three limits can be computed using the same technique, searching convenient upper and lower bounds. This can be achieved by the use of (34). Here, we only give a proof for the third limit because it contains all major arguments. We set
$$z_n = \frac{\sum_{i=0}^{n} \beta_{q,n-i}\, \mathbf{u}_i(q)}{\mathbf{u}_n(q)}.$$

Let $\epsilon > 0$. We first give an upper bound to z_n. Using (33), one has for any pair (k, K) satisfying (34),

$$\begin{aligned} z_n &\le \frac{a_q \sum_{i=K+1}^{n} \beta_{q,n-i}\, \alpha_i^q\, i^r\, \theta^i}{a_q\, \alpha_n^q\, n^r\, \theta^n} + \frac{\sum_{i=0}^{K} \beta_{q,n-i}\, \mathbf{u}_i(q)}{\mathbf{u}_n(q)} \\ &\le \frac{k+1}{k-1} \sum_{i=K+1}^{n} \beta_{q,n-i}\, (\frac{i}{n})^r\, \theta^{i-n} + \frac{\sum_{i=0}^{K} \beta_{q,n-i}\, \mathbf{u}_i(q)}{\mathbf{u}_n(q)} \end{aligned}$$

Moreover,

$$\begin{aligned} \sum_{i=K+1}^{n} \beta_{q,n-i}\, (\frac{i}{n})^r\, \theta^{i-n} &= \sum_{i=0}^{n-K-1} \beta_{q,i}\, (1 - \frac{i}{n})^r\, \theta^{-i} \\ &= \sum_{i=0}^{n-K-1} \beta_{q,i}\, \theta^{-i} \\ &\quad \underbrace{+ \sum_{j=1}^{r} \binom{r}{j}\, n^{-j} \sum_{i=0}^{n-K-1} \beta_{q,i}\, (-i)^j\, \theta^{-i}}_{:=\xi_n}. \end{aligned}$$

The series $\sum_{i=0}^{\infty} \beta_{q,i}\, \theta^{-i}$ is uniformly convergent and may be differentiated term by term. Thus

$$(\theta \frac{d}{d\theta})^j \sum_{i=0}^{\infty} \beta_{q,i}\, \theta^{-i} = \sum_{i=0}^{\infty} \beta_{q,i}\, (-i)^j \theta^{-i}$$

is bounded. Notice that, in the expression of ξ_n, r is a constant and $\xi_n \in O(n^{-1})$. So, if K is fixed, $\xi_n \to 0$ as n tends to infinity. We can write,

$$\begin{aligned} z_n &\le (1 + \frac{2}{k-1})(\sum_{i=0}^{n-K-1} \beta_{q,i}\, \theta^{-i} + \xi_n) + \frac{\sum_{i=0}^{K} \beta_{q,n-i}\, \mathbf{u}_i(q)}{\mathbf{u}_n(q)} \\ &\le \sum_{i=0}^{n} \beta_{q,i}\, \theta^{-i} + \frac{2}{k-1} \sum_{i=0}^{\infty} \beta_{q,i}\, \theta^{-i} + (1 + \frac{2}{k-1})\, \xi_n \\ &\quad + \frac{\sum_{i=0}^{K} \beta_{q,n-i}\, \mathbf{u}_i(q)}{\mathbf{u}_n(q)}. \end{aligned}$$

The series $\sum_{i=0}^{\infty} \beta_{q,i}\, \theta^{-i}$ is bounded. Let k large enough to insure the second term in the latter expression is less than ϵ. Taking such a k, we choose K to satisfy (34). Thus, for n large enough, the last two terms in the latter expression are also less than ϵ. So,

$$z_n - \sum_{i=0}^{n} \beta_{q,i}\, \theta^{-i} \le 3\, \epsilon.$$

For the upper bound, we obtain

$$\begin{aligned} z_n &\geq (1-\frac{2}{k+1})\sum_{i=0}^{n-K-1}\beta_{q,i}\,(1-\frac{i}{n})^r\,\theta^{-i} \\ &\geq \sum_{i=0}^{n}\beta_{q,i}\,\theta^{-i} - \sum_{i=n-K}^{n}\beta_{q,i}\,\theta^{-i} + \xi_n - \frac{2}{k+1}\sum_{i=0}^{\infty}\beta_{q,i}\,\theta^{-i} - \frac{2\,\xi_n}{k+1} \end{aligned}$$

The fourth term in the latter expression is greater than $-\epsilon$ as are the others but the first, for n large enough. Hence

$$z_n - \sum_{i=0}^{n}\beta_{q,i}\,\theta^{-i} \geq -4\,\epsilon$$

and the conclusion.

Assume now that $a_q = 0$. We have to show that $d_n := \frac{\sum_{i=0}^{n}\beta_{q,n-i}\,\mathbf{u}_i(q)}{\mathbf{v}_n(s)}$ tends to 0 as n tends to infinity. If $\mathbf{u}_n(q) = 0$ for all n sufficiently large, the result is obvious. Otherwise, we have P_q and θ_q such that

$$\lim_{n\to\infty}\frac{\mathbf{u}_n(q)}{P_q(n)\,\theta_q^n} = b_q > 0.$$

If $\theta_q = 1$, d_n is the ratio of a function bounded by polynomial and of an exponential function. Otherwise, either $\theta_q = \theta$ and $t = d(Q) < d(P) = r$ or $1 < \theta_q < \theta$. In these two cases, one proceeds as above. For $p \in \{q, s\}$, there exists a sequence (α_n^p) convergent to 1 such that

$$\mathbf{u}_n(s) = \alpha_n^s\,a_s\,n^r\theta^n \text{ and } \mathbf{u}_n(q) = \alpha_n^q\,b_q\,n^t\theta_q^n.$$

It is clear that $d_n \geq 0$ and one has to find a convenient upper bound to conclude as in the previous cases.

□

The latter proposition has some direct corollaries.

Corollary VI.5.4. *One has*

$$\lim_{n\to\infty}\frac{\mathrm{val}_S(w_n)}{\mathbf{v}_{|w_n|}} = \frac{\theta-1}{\theta}\sum_{q\in K}\frac{a_q}{a_s}\sum_{j=0}^{\infty}\beta_{q,j}\,\theta^{-j}.$$

Corollary VI.5.5. *One has*

$$\lim_{n\to\infty}\frac{\mathbf{v}_{n-1}(s)}{\mathbf{v}_n(s)} = \frac{1}{\theta}.$$

Proof. Indeed,

$$\frac{\mathbf{v}_{n-1}(s)}{\mathbf{v}_n(s)} = 1 - \frac{\mathbf{u}_n(s)}{\mathbf{v}_n(s)} \to 1 - \frac{\theta-1}{\theta}.$$

□

Corollary VI.5.6. *If $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are two sequences of words belonging to L which converge to the same limit word w then*

$$\lim_{n\to\infty}\frac{\mathrm{val}_S(x_n)}{\mathbf{v}_{|x_n|}}=\lim_{n\to\infty}\frac{\mathrm{val}_S(y_n)}{\mathbf{v}_{|y_n|}}.$$

Proof. The two sequences give the same sequence $(\beta_{q,n}(w))_{n\in\mathbb{N}}$. One concludes using Corollary VI.5.4.

□

6. Representation of real numbers

Recall that if $w\in L\cap\Sigma^n$ then $\mathrm{val}_S(w)\in[\mathbf{v}_{n-1}(s),\mathbf{v}_n(s)-1]$. Therefore, if we admit to represent a real number by an infinite word and if we consider the words w of length n as approximations $\frac{\mathrm{val}_S(w)}{|w|}$ of real numbers, then these numbers necessarily belong to $[\frac{1}{\theta},1]$, since $\frac{\mathbf{v}_{n-1}}{\mathbf{v}_n}\to\frac{1}{\theta}$ and $\frac{\mathbf{v}_n-1}{\mathbf{v}_n}\to 1$ as $n\to\infty$.

These considerations appear in the classical system with an integer base p. It is obtained by describing the language

$$\{1,\dots,p-1\}\{0,\dots,p-1\}^*\cup\{0\}$$

according to the natural ordering of the digits. In this system, an infinite word $w=a_1a_2\cdots$, $a_1\neq 0$, is used to represent the real $\sum_{i=1}^{\infty}a_i\,p^{-i}$. Consequently, we can only represent the interval $[\frac{1}{p},1]$. Let x be a real number in $[0,\frac{1}{p}]$. There exists t such that $\frac{1}{p}\le p^t\,x\le 1$. So the representation of x can be achieved by representing $p^t\,x$ and storing t in some way. For the base p, the infinite word representing x is thus preceded by t zeroes and the interval $[0,1]$ is decomposed in intervals of the form $[p^{-t-1},p^{-t}]$, $t\in\mathbb{N}$. In an abstract numeration system, we have the same phenomenom with p replaced by θ.

Let $x\in]\frac{1}{\theta},1[$. We can build a sequence $(w_n)_{n\in\mathbb{N}}$ of words satisfying the following conditions (35),

$$\boxed{\left\{\begin{array}{ll}\diamond & \forall n\in\mathbb{N},\ w_n\in L\\ \diamond & (|w_n|)_{n\in\mathbb{N}}\text{ is strictly increasing}\\ \diamond & \lim_{n\to\infty}\frac{\mathrm{val}_s(w_n)}{\mathbf{v}_{|w_n|}}=x.\end{array}\right.}\tag{35}$$

Indeed, for n large enough, there exists $w\in L\cap\Sigma^n$ such that

$$\frac{\mathrm{val}_S(w)}{\mathbf{v}_n(s)}\le x<\frac{\mathrm{val}_S(w)+1}{\mathbf{v}_n(s)}.$$

The length of this interval containing x is trivially $1/\mathbf{v}_n(s)$ which tends to zero. Such a sequence of words does not converge necessarily to a limit word but the following lemma shows that we can always extract a convergent subsequence.

Lemma VI.6.1. *Every infinite sequence $(w_n)_{n\in\mathbb{N}}$ of words belonging to L has a subsequence $(w_{k(n)})_{n\in\mathbb{N}}$ convergent to an element of $\mathcal{L}_\infty$.*

Proof. Let $A_0 = \{w_n : n \in \mathbb{N}\}$ and $(l_n)_{n\ge 1}$ be a strictly increasing sequence of positive integers. There exists $p_1 \in \Sigma^{l_1}$ which is a prefix of an infinite number of elements of A_0. We define the infinite set

$$A_1 = p_1(p_1^{-1}.L) \cap A_0 = \{w_n \in A_0 : \exists z \in \Sigma^*,\ w_n = p_1 z\}$$

and $w_{k(1)}$ as the smallest word of A_1. Similarly, there exists $p_2 \in \Sigma^{l_2}$ which is a prefix of an infinite number of elements of A_1. Observe that p_1 is prefix of p_2. Continuing this way, we obtain a convergent subsequence $(w_{k(n)})_{n\in\mathbb{N}}$.

□

Let $x \in]\frac{1}{\theta}, 1[$, our aim is to characterize the set Q_x of words of $\mathcal{L}_\infty$ that are the limits of sequences satisfying (35). Roughly speaking, Q_x is the set of all infinite words representing x. If w belongs to Q_x then w is said to be an S-*representation* of x.

We can already notice that for a given x, the set Q_x depends on the ordering of the alphabet. We shall not emphasize this observation and in the following, we assume that the ordering of the alphabet is fixed.

6.1. Defining the intervals I_W. The characterization of Q_x is simplified by the introduction of some special real intervals.

Let $l \in \mathbb{N}$ and $\mathcal{W}_l$ be the set of words of length l which are prefixes of an infinite number of words belonging to L (i.e., words $W \in \Sigma^l$ such that $W^{-1}.L$ is infinite). We can enumerate this set lexicographically,

$$\mathcal{W}_l = \{W_1 < \ldots < W_k\},\ k \le \#\Sigma^l.$$

Notice that, since L is infinite, $\mathcal{W}_l$ is never empty.

For l fixed and n large enough, any word of $L \cap \Sigma^{\ge n}$ has a prefix in $\mathcal{W}_l$. Figure VI.2 sketches the situation. In particular, the number of words of length n belonging to L and having a prefix of length l lexicographically less than W is $\mathbf{v}_{n-1}(s) + \sum_{V<W,\ V\in\mathcal{W}_l} \mathbf{u}_{n-l}(s.V)$. For $W \in \mathcal{W}_l$, we set

$$\alpha_W^n = \frac{\mathbf{v}_{n-1}(s)}{\mathbf{v}_n(s)} + \sum_{V<W,\ V\in\mathcal{W}_l} \frac{\mathbf{u}_{n-l}(s.V)}{\mathbf{v}_n(s)}$$

and

$$I_W^n = \left[\alpha_W^n, \alpha_W^n + \frac{\mathbf{u}_{n-l}(s.W)}{\mathbf{v}_n(s)}\right].$$

With our notations, using Corollary VI.5.4 and Corollary VI.5.5, we have

$$\lim_{n\to\infty} \frac{\mathbf{u}_{n-l}(q)}{\mathbf{v}_n(s)} = \frac{a_q}{a_s}\ \frac{\theta-1}{\theta^{l+1}}.$$

We denote by I_W the limit of the I_W^n's when n tends to infinity, i.e.,

$$I_W = \left[\frac{1}{\theta} + \frac{\theta-1}{\theta^{l+1}} \sum_{V<W,\ V\in\mathcal{W}_l} \frac{a_{s.V}}{a_s}, \frac{1}{\theta} + \frac{\theta-1}{\theta^{l+1}} \sum_{V\le W,\ V\in\mathcal{W}_l} \frac{a_{s.V}}{a_s}\right].$$

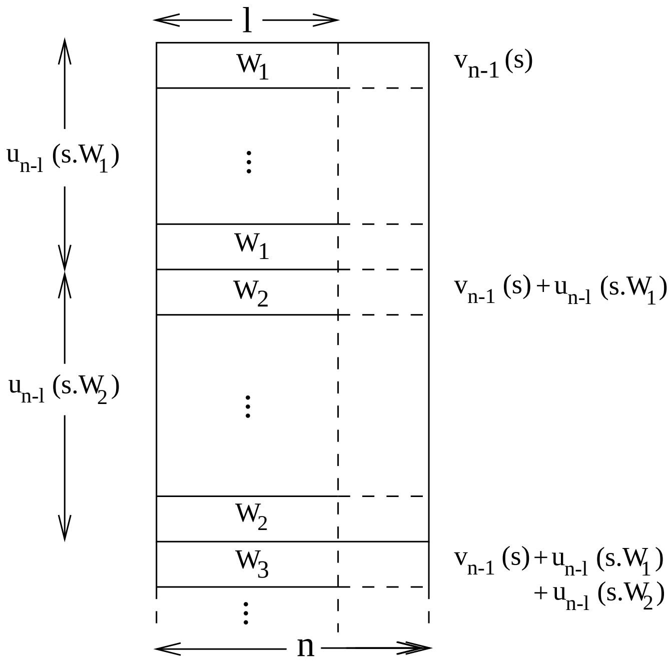

FIGURE VI.2. The ordered set $L \cap \Sigma^n$ and the corresponding numerical values.

By convention, if $W \in \Sigma^l \setminus \mathcal{W}_l$ then $I_W = \emptyset$. Notice that if $W \in \mathcal{W}_l$ is such that $a_{s.W} = 0$ then I_W is reduced to a unique element.

REMARK VI.6.2. Observe that a real number y belongs to I_W if and only if $y = \lim_{n\to\infty} \frac{\mathrm{val}(w_n)}{\mathbf{v}_{|w_n|}}$ for a sequence $(w_n)_{n\in\mathbb{N}}$ of words of L convergent to an element of $W\Sigma^\omega$.

We also have, for all $l \in \mathbb{N}$,

$$\left[\frac{1}{\theta}, 1\right] = \bigcup_{W\in\mathcal{W}_l} I_W. \tag{36}$$

Indeed, the upper bound of an interval is equal to the lower bound of the next interval. Moreover, the lower bound of the first interval is $\lim_{n\to\infty} \frac{\mathbf{v}_{n-1}}{\mathbf{v}_n} = \frac{1}{\theta}$ and the upper bound of the last interval is $\lim_{n\to\infty} (\frac{\mathbf{v}_{n-1}}{\mathbf{v}_n} + \sum_\sigma \frac{\mathbf{u}_{n-1}(s.\sigma)}{\mathbf{v}_n}) = 1$.

REMARK VI.6.3. Let $l \in \mathbb{N}$ and $W \in \mathcal{W}_l$. Observe that the length of the interval I_W is $O(\theta^{-l-1})$.

LEMMA VI.6.4. *If x, $y \in \Sigma^*$, then*

$$I_{xy} \subseteq I_x.$$

Proof. Let $\Sigma = \{\sigma_1 < \ldots < \sigma_p\}$. Notice that $I_x = \emptyset$ implies $I_{xy} = \emptyset$ (if $x^{-1}.L$ is finite or empty so is $(xy)^{-1}.L = y^{-1}.(x^{-1}.L)$). Let us assume that $I_x \neq \emptyset$. With our notations,

$$I_x^n = \left[\alpha_x^n, \alpha_x^n + \frac{\mathbf{u}_{n-|x|}(s.x)}{\mathbf{v}_n(s)}\right].$$

In this proof, we take n large enough to guarantee that any word of $L \cap \Sigma^{\geq n}$ has a prefix in $\mathcal{W}_{|x|+1}$. It is enough to observe that, if $(x\sigma_1)^{-1}.L$ is infinite then $\alpha_{x\sigma_1}^n = \alpha_x^n$ and

$$I_{x\sigma_1}^n = \left[\alpha_x^n, \alpha_x^n + \frac{\mathbf{u}_{n-|x|-1}(s.x\sigma_1)}{\mathbf{v}_n(s)}\right]$$

indeed, among the words of length n beginning with x, the first ones begin with $x\sigma_1$ and there are $\mathbf{u}_{n-|x|-1}(s.x\sigma_1)$ such words. Similarly, for $i = 2, \ldots, p$, if $(x\sigma_i)^{-1}.L$ is infinite then

$$I_{x\sigma_i}^n = \left[\alpha_x^n + \frac{\sum_{j=1}^{i-1} \mathbf{u}_{n-|x|-1}(s.x\sigma_j)}{\mathbf{v}_n(s)}, \alpha_x^n + \frac{\sum_{j=1}^{i} \mathbf{u}_{n-|x|-1}(s.x\sigma_j)}{\mathbf{v}_n(s)}\right].$$

Notice also that $\sum_{j=1}^{p} \mathbf{u}_{n-|x|-1}(s.x\sigma_j) = \mathbf{u}_{n-|x|}(s.x)$. So, for n sufficiently large, $I_{x\sigma_i}^n \subseteq I_x^n$. Hence the conclusion.

□

Corollary VI.6.5. *We have for $w \in \Sigma^*$,*

$$\bigcup_{\sigma \in \Sigma} I_{w\sigma} = I_w.$$

6.2. Some trees $\mathcal{T}_x$. Now, we shall introduce infinite trees related to Q_x. These trees are constructed on the intervals I_W and we show that the infinite paths in the tree are the S-representations of the real x.

Definition VI.6.6. To any real $x \in]\frac{1}{\theta}, 1[$, we associate a unique tree $\mathcal{T}_x$ defined as follows.

- The root (i.e., the node of level 0) of $\mathcal{T}_x$ is I_ε, where ε is the empty word.
- The nodes of level $l \geq 1$ are the I_W such that $W \in \mathcal{W}_l$ and $x \in I_W$.
- An edge connects a node I_V of level $l \geq 0$ to a node I_W of level $l+1$ if and only if there exists $\sigma \in \Sigma$ such that $W = V\sigma$, i.e., $I_W \subset I_V$. This edge is labeled by σ.

A *path* in $\mathcal{T}_x$ is a sequence of consecutive connected nodes starting in the root. To a (finite or infinite) path P, one associates a unique (finite or infinite) word formed by the consecutive labels of the edges in P.

As a consequence of Corollary VI.6.5, each node of $\mathcal{T}_x$ has at least one son and at most $\#\Sigma$ sons. The tree has some other properties.

Proposition VI.6.7. *Let $x \in]\frac{1}{\theta}, 1[$, the tree $\mathcal{T}_x$ contains an infinite number of nodes and at least an infinite path.*

Proof. The number of nodes is infinite because, by (36), for each $l \in \mathbb{N}$, there exists at least one $W \in \mathcal{W}_l$ such that $x \in I_W$. An infinite tree, the nodes of which have a bounded number of sons, has always an infinite path.

$\square$

PROPOSITION VI.6.8. *Let $x \in]\frac{1}{\theta}, 1[$. The set of words labeling infinite paths in $\mathcal{T}_x$ is Q_x.*

Proof. Let P be a label of an infinite path in $\mathcal{T}_x$. By definition of $\mathcal{T}_x$, one has $x \in I_{P_0}$ (P_0 is the first letter of P). By definition of $\mathcal{W}_1$, an infinite number of words belonging to L have P_0 as a prefix. Let w_1 be the smallest of these words. For the same reasons, $x \in I_{P[0,1]}$ and an infinite number of words of L begin with $P[0,1]$. Let $w_2 \in L$ be the smallest word of length greater than $|w_1|$ beginning with $P[0,1]$. Continuing this way, we obtain a sequence $(w_n)_{n\in\mathbb{N}}$ of words strictly increasing in length which converges to P. To show that $P \in Q_x$, we still have to verify that the third condition of (35) is satisfied. Let $\epsilon > 0$. By Remark VI.6.3, for l large enough, the length of $I_{P[0,l]}$ is less than $\epsilon/4$. So for m large enough, say $m > M_1$, the length of $I^m_{P[0,l]}$ is less than $\epsilon/2$. Since $x \in I_{P[0,l]}$, for m sufficiently large, say $m > M_2$, $I^m_{P[0,l]} \cap]x - \epsilon/2, x + \epsilon/2[\neq \emptyset$. There exists N_1 such that, for all $n > N_1$, $|w_n| > \sup\{M_1, M_2\}$. Since $w_n \to P$, there exists N_2 such that, for all $n > N_2$, w_n starts with $P[0,l]$. So, for all $n > \sup\{N_1, N_2\}$,

$$\frac{\mathrm{val}_S(w_n)}{\mathbf{v}_{|w_n|}} \in I^{|w_n|}_{P[0,l]} \text{ and } \left|\frac{\mathrm{val}_S(w_n)}{\mathbf{v}_{|w_n|}} - x\right| < \epsilon.$$

Let $w \in Q_x$. By definition, there exists a sequence $(w_n)_{n\in\mathbb{N}}$ satisfying (35) and such that $(w_n)_{n\in\mathbb{N}}$ is convergent to w. By Remark VI.6.2, for any $l \in \mathbb{N}$, $x \in I_{w[0,l]}$ and $I_{w[0,l]}$ is a node of $\mathcal{T}_x$. Thus w is the label of an infinite path in $\mathcal{T}_x$.

$\square$

COROLLARY VI.6.9. *Let $x \in]\frac{1}{\theta}, 1[$. Assume that for each l, there exists $W_{(l)} \in \mathcal{W}_l$ such that $x \in (I_{W_{(l)}})^\circ$. All the sequences $(w_n)_{n\in\mathbb{N}}$ satisfying (35) converge. Their limits are equal each to the other. In other words, Q_x contains a unique element.*

Proof. Let $l \geq 1$. If $x \in (I_{W_{(l)}})^\circ$ then for any word $W \in \mathcal{W}_l \setminus \{W_{(l)}\}$, $x \notin I_W$. So each node in $\mathcal{T}_x$ has exactly one son.

$\square$

6.3. The endpoints of I_W. If $x \in]\frac{1}{\theta}, 1[$ does not satisfy the assumptions of Corollary VI.6.9 then it is the endpoint of some interval I_W. In this section we study such a case. To that purpose, we shall distinguish special elements in Q_x.

Since Σ is an ordered alphabet, the set Σ^ω is also totally ordered by the lexicographic order. Recall that for that order $u < v$ if $u = t\sigma u'$ and $v = t\beta v'$ with $t \in \Sigma^*$, $\sigma, \beta \in \Sigma$, $u', v' \in \Sigma^\omega$ and $\sigma < \beta$.

Proposition VI.6.10. *If $x \in]\frac{1}{\theta}, 1[$ is the endpoint of I_W then* $\inf Q_x$ *(resp.* $\sup Q_x$*) is an ultimately periodic word (i.e., the smallest (resp. greatest) word of Q_x is of the form $x(y)^\omega$ for some $x \in \Sigma^*$ and $y \in \Sigma^+$).*

Proof. Let us assume that x is the upper bound of I_W, $W \in \mathcal{W}_l$. We may assume that I_W has non empty interior. This follows from (36) and from the definition of the intervals. (If $I_W \neq \emptyset$ has an empty interior then it is reduced to one point; observe that in this case $a_{s.W} = 0$.) Note that no $U \in \mathcal{W}_l$ is such that $x \in I_U$ and $U < W$.

By Corollary VI.6.5, there exists a unique letter σ_0 such that x is the upper bound of $I_{W\sigma_0}$ with $(I_{W\sigma_0})^\circ$ non empty. (Notice that if $\beta < \sigma_0$ then $x \notin I_{W\beta}$.) In terms of paths in the minimal automaton of L, we were in a state $q_0 = s.W$ and we chose the greatest letter σ_0 of the alphabet such that $a_{q_0.\sigma_0} > 0$. So, we define $q_1 = q_0.\sigma_0$. Continuing this way we obtain a sequence $(q_n)_{n\in\mathbb{N}}$ of states (each element being completely determined by the previous one) and from this sequence, we compute a sequence $(\sigma_n)_{n\in\mathbb{N}}$ of letters (a letter σ_n being determined by the state q_n). The automaton $\mathcal{A}_L$ is finite hence $(q_n)_{n\in\mathbb{N}}$ and then $(\sigma_n)_{n\in\mathbb{N}}$ are ultimately periodic. It is clear that $\sigma_0\sigma_1\cdots = \sup Q_x$. One can do the same proof with x being the lower bound of some interval.

$\square$

As a consequence of this result, an endpoint of $]\frac{1}{\theta}, 1[$ has at least two different representations.

Proposition VI.6.11. *The endpoints of I_W are algebraic numbers.*

Proof. For all state q of $\mathcal{M}_L$, it is well known that the sequence $\mathbf{u}_n(q)$ verifies some recurrence relation with coefficients in $\mathbb{Z}$. The real θ_q is a root of the characteristic polynomial of the recurrence and is therefore an algebraic number. We set $\theta_q = \theta_{q,1}$, $P_q = P_{q,1}$. We have

$$\mathbf{u}_n(q) = P_{q,1}(n)\theta_{q,1}^n + \cdots + P_{q,k}(n)\theta_{q,k}^n$$

where $\theta_{q,1}, \ldots, \theta_{q,k}$ are the roots of the characteristic polynomial of the recurrence with multiplicity $m_1, \ldots, m_k$ and $d(P_{q,i}) < m_i$ for $i = 1, \ldots, k$. The coefficients of the polynomials $P_{q,i}$'s (and specially b_q) are completely determined by the initial integral conditions $\mathbf{u}_0(q), \ldots, \mathbf{u}_m(q)$ where m is the degree of the recurrence. So, b_q and therefore a_q are also algebraic reals. An endpoint of I_W is obtained from the a_q's and θ_q's by quotients and sums and is then algebraic.

$\square$

7. Some examples of representation

EXAMPLE VI.7.1. Consider the language and the automaton of Example III.6.6. One can show that

$$\begin{cases} \mathbf{u}_n(s) &= (1+\sqrt{3})^n \frac{2+\sqrt{3}}{2\sqrt{3}} + (1-\sqrt{3})^n \frac{\sqrt{3}-2}{2\sqrt{3}} \\ \mathbf{u}_n(t) &= (1+\sqrt{3})^{n+1} \frac{1}{2\sqrt{3}} + (1-\sqrt{3})^n \frac{3+\sqrt{3}}{6\,(2+\sqrt{3})} \\ \mathbf{u}_n(p) &= 0 \\ \mathbf{v}_n(s) &= (1+\sqrt{3})^n \frac{3\sqrt{3}+5}{6} + (1-\sqrt{3})^n \frac{5-3\sqrt{3}}{6} - \frac{2}{3}. \end{cases}$$

So $\theta = 1+\sqrt{3}$ and for $l \geq 1$,

$$\begin{aligned} \lim_{n\to\infty} \frac{\mathbf{u}_{n-l}(s)}{\mathbf{v}_n(s)} &= \frac{3+2\sqrt{3}}{(5+3\sqrt{3})\,(1+\sqrt{3})^l} \\ \lim_{n\to\infty} \frac{\mathbf{u}_{n-l}(t)}{\mathbf{v}_n(s)} &= \frac{3}{(9+5\sqrt{3})\,(1+\sqrt{3})^{l-1}}. \end{aligned}$$

The upper bound of I_a is

$$x = \frac{1}{\theta} + \lim_{n\to\infty} \frac{\mathbf{u}_{n-1}(s.a)}{\mathbf{v}_n(s)} = \frac{1}{1+\sqrt{3}} + \frac{3}{9+5\sqrt{3}}.$$

It is equal to the lower bound of I_b. Figure VI.3 shows $\mathcal{T}_x$. In this case,

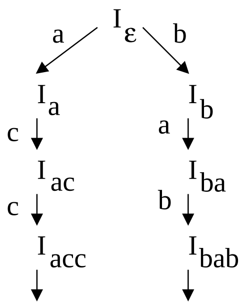

FIGURE VI.3. The tree $\mathcal{T}_x$, with $x = \frac{1}{1+\sqrt{3}} + \frac{3}{9+5\sqrt{3}}$.

$Q_x = \{a(c)^\omega, (ba)^\omega\}$. Another example of this kind is given by

$$x = \frac{1}{\theta} + \sum_{w\in H} \lim_{n\to\infty} \frac{\mathbf{u}_{n-3}(s.w)}{\mathbf{v}_n(s)} = \frac{1}{\theta} + \lim_{n\to\infty} \frac{\mathbf{u}_{n-1}(s.a) + \mathbf{u}_{n-3}(s.bab)}{\mathbf{v}_n(s)}$$

where $H = \{aba, abb, abc, aca, acb, acc, bab\}$. So,

$$Q_x = \{bab(c)^\omega, bac(ab)^\omega\}.$$

Notice that the two infinite words have the same prefix ba.

EXAMPLE VI.7.2. If we want to find back the classical numeration system in base $k \geq 2$, we use the regular language

$$L = \{1,\dots,k-1\}\{0,\dots,k-1\}^* \cup \{0\}$$

accepted by the DFA in Figure VI.4. The first intervals are

$$I_1 = \left[\frac{1}{k}, \frac{2}{k}\right], \; \dots, \; I_{k-1} = \left[\frac{k-1}{k}, 1\right]$$

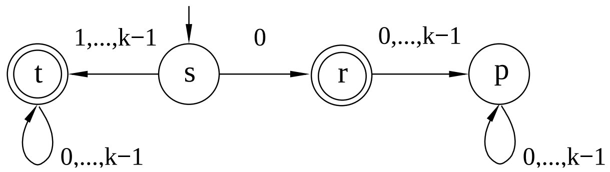

FIGURE VI.4. The minimal automaton for the k-ary system.

and if $i = 1, \ldots, k-1$,

$$I_{i1} = \left[\frac{i}{k}, \frac{i}{k} + \frac{1}{k^2}\right], \ \ldots, \ I_{i(k-1)} = \left[\frac{i}{k} + \frac{k-1}{k^2}, \frac{i+1}{k}\right].$$

On Figure VI.5, we have represented some trees when $k = 10$. The

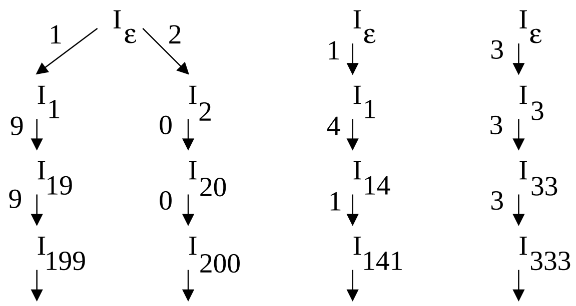

FIGURE VI.5. The trees $\mathcal{T}_{2/10}$, $\mathcal{T}_{\sqrt{2}-1}$ and $\mathcal{T}_{1/3}$.

sets $Q_{\sqrt{2}-1}$ and $Q_{1/3}$ contain only one element (this is a consequence of Corollary VI.6.9) but $Q_{2/10} = \{1(9)^\omega, 2(0)^\omega\}$ (Proposition VI.6.10). It is obvious that for systems in base k, Q_x contains two elements if and only if $x = r/k^n$ for some positive integers n and $r < k^n$. Otherwise, Q_x contains exactly one element.

EXAMPLE VI.7.3. This third example was constructed to obtain an infinite Q_x. The language L used here is based on the one found in Example III.6.6. Its minimal automaton is depicted on Figure VI.6. In this language, the number of words beginning with a or c has an exponential growth as in Example III.6.6 but $b\Sigma^* \cap L = b^+a^*$ has a polynomial growth. So, $\mathbf{u}_n(s.b)/\mathbf{v}_n(s)$ tends to zero when n goes to infinity but for all n, $\mathbf{u}_n(s.b) = n + 1 > 0$. We have

$$I_a = [\frac{1}{\theta}, x], \ I_b = [x, x] \text{ and } I_c = [x, 1]$$

where $x = \frac{1}{1+\sqrt{3}} + \frac{3}{9+5\sqrt{3}}$. The tree $\mathcal{T}_x$ is depicted on Figure VI.7. Notice that $Q_x = \{a(b)^\omega, c(ab)^\omega\} \cup \{b^n(a)^\omega : n \geq 1\}$ is infinite.

EXAMPLE VI.7.4. In the previous example Q_x was countable. Here we show that we can also obtain a set Q_x being uncountably infinite.

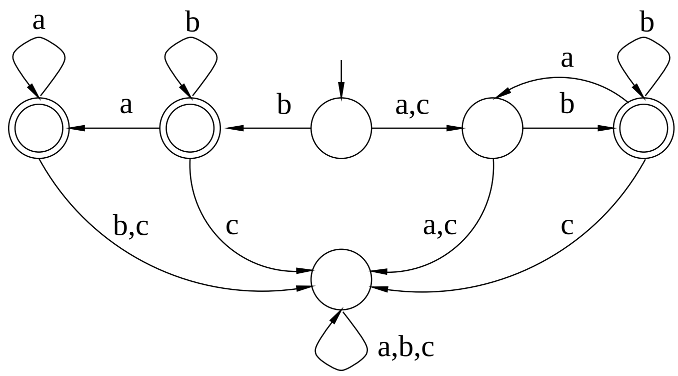

FIGURE VI.6. The minimal automaton of L.

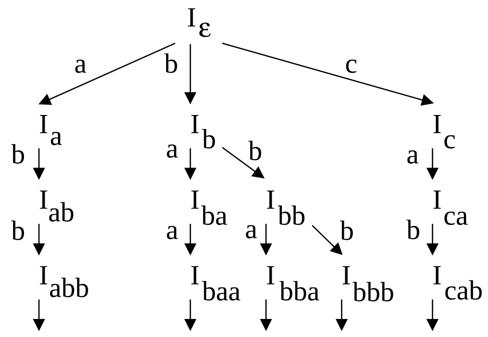

FIGURE VI.7. The tree $\mathcal{T}_x$, with $x = \frac{1}{1+\sqrt{3}} + \frac{3}{9+5\sqrt{3}}$.

Let $\Sigma = \{a, b, c\}$, $\Gamma = \{d, e\}$. Consider the language

$$L = \{a, c\}\,\Sigma^* \cup b\,\Gamma^* \subset \{a < b < c < d < e\}^*.$$

It is obvious that for all n, $\mathbf{u}_n(s.b) = 2^n$, $\mathbf{u}_n(s.a) = 3^n$ and for $n \geq 1$, $\mathbf{u}_n(s) = 3^{n-1} + 2^{n-1}$ and $\mathbf{v}_n(s) = 2^n - 1 + \frac{1}{2}(3^n - 1)$. So for $l \geq 1$,

$$\lim_{n\to\infty} \frac{\mathbf{u}_{n-l}(s.b)}{\mathbf{v}_n(s)} = 0 \text{ and } \lim_{n\to\infty} \frac{\mathbf{u}_{n-l}(s.a)}{\mathbf{v}_n(s)} = \frac{2}{3^{l+1}}.$$

Let $x = \frac{1}{3} + \frac{2}{3^2}$,

$$Q_x = \{a(c)^\omega, c(a)^\omega\} \cup b\,\Gamma^\omega$$

is uncountably infinite.

8. Equivalence with θ-development

We consider a Pisot number $\theta > 1$. To this number corresponds a unique positional Bertrand number system $U = (U_n)_{n\in\mathbb{N}}$ having its characteristic polynomial equal to the minimal polynomial of θ. We denote by L the language $\rho_U(\mathbb{N})$ of all normalized representations. We show that this latter language satisfies the hypotheses given in Section 4 and that the representations of real numbers in the abstract numeration system built upon L and the θ-developments of numbers in $[\frac{1}{\theta}, 1]$

coincide. This section has the following articulation. First, we begin with an introductory example and consider the golden mean $\theta = \frac{1+\sqrt{5}}{2}$. Next, we recall basic facts related to θ-developments and positional systems. Thanks to these results, we show the announced equivalence.

DEFINITION VI.8.1. A positional numeration system $U = (U_n)_{n\in\mathbb{N}}$ is said to be a *Bertrand numeration system* if

$$\forall n \in \mathbb{N}, w\, 0^n \in \rho_U(\mathbb{N}) \Leftrightarrow w \in \rho_U(\mathbb{N}).$$

As examples, k-ary number systems and the Fibonacci system are Bertrand systems.

Let $\theta > 1$ be a Pisot number. To this number corresponds a class of positional linear number systems such that their characteristic polynomial is the minimal polynomial of θ. The systems in this class differ only by the initial conditions $U_0, \ldots, U_k$. Among these systems exactly one is a Bertrand number system and it shall be denoted by $U_\theta = (U_n)_{n\in\mathbb{N}}$ or simply U if the context is clear. We shall further see that the initial conditions defining U_θ are derived from the θ-development of one.

EXAMPLE VI.8.2. Consider the Fibonacci system U introduced in Example I.2.3. It is well known that

$$\rho_U(\mathbb{N}) = 1\{0, 01\}^* \cup \{\varepsilon\}.$$

Hence, it is obvious that the Fibonacci system is the Bertrand numeration system related to the golden mean $\theta = \frac{1+\sqrt{5}}{2}$. The minimal automaton of $\rho_U(\mathbb{N})$ is depicted in Figure VI.8. The complexity func-

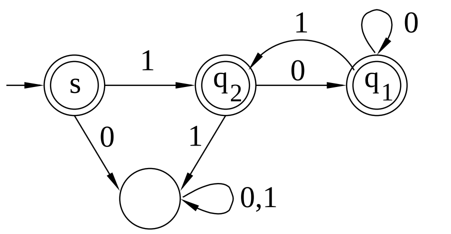

FIGURE VI.8. The minimal automaton accepting normalized representations in the Fibonacci system.

tion of the different states are given by

$$\mathbf{u}_n(s) = \begin{cases} \frac{\sqrt{5}}{5}\,\theta^n - \frac{\sqrt{5}}{5}\left(\frac{1-\sqrt{5}}{2}\right)^n & ;\ \text{if } n \geq 1, \\ 1 & ;\ \text{if } n = 0, \end{cases}$$

$$\mathbf{u}_n(q_2) = \mathbf{u}_{n+1}(s),\ \forall n \in \mathbb{N}$$

and

$$\mathbf{u}_n(q_1) = \frac{\sqrt{5}}{5}\,\theta^{n+2} + \frac{5 - 3\sqrt{5}}{10}\left(\frac{1-\sqrt{5}}{2}\right)^n,\ \forall n \in \mathbb{N}.$$

Therefore,

$$\mathbf{v}_n(s) = \frac{5+3\sqrt{5}}{10}\,\theta^n + \frac{5+3\sqrt{5}}{10}\left(\frac{1-\sqrt{5}}{2}\right)^n, \ \forall n \in \mathbb{N}.$$

With these sequences, we can compute the endpoints of the intervals I_w defined in Section 6 of the present chapter. The first intervals are

$I_\varepsilon, I_1, I_{10}, I_{100}, I_{101}, I_{1000}, I_{1001}, I_{1010}, I_{10000}, I_{10001}, I_{10010}, I_{10100}, I_{10101}, \ldots$

Observe that

$$\lim_{n\to\infty} \frac{\mathbf{u}_{n-i}(q_2)}{\mathbf{v}_n(s)} = \frac{\sqrt{5}}{5}\,\frac{10}{5+3\sqrt{5}}\,\theta^{1-i} = \theta^{-i-1}$$

and in the same way,

$$\lim_{n\to\infty} \frac{\mathbf{u}_{n-i}(q_1)}{\mathbf{v}_n(s)} = \theta^{-i}.$$

For instance, we can compute the endpoints of the three intervals corresponding to words of length four: I_{1000}, I_{1001} and I_{1010}. The upper bound of I_{1000} is the lower bound of I_{1001} and is equal to

$$\frac{1}{\theta} + \lim_{n\to\infty} \frac{\mathbf{u}_{n-4}(s.1000)}{\mathbf{v}_n(s)} = \frac{1}{\theta} + \frac{1}{\theta^4}.$$

The upper bound of I_{1001} is equal to the lower bound of I_{1010} and is equal to

$$\frac{1}{\theta} + \frac{1}{\theta^4} + \lim_{n\to\infty} \frac{\mathbf{u}_{n-4}(s.1001)}{\mathbf{v}_n(s)} = \frac{1}{\theta} + \frac{1}{\theta^4} + \frac{1}{\theta^5}$$

Since θ is a root of $X^2 - X - 1$, this latter endpoint is equal to $\frac{1}{\theta} + \frac{1}{\theta^3}$. Finally the upper bound of I_{1010} is

$$\frac{1}{\theta} + \frac{1}{\theta^3} + \lim_{n\to\infty} \frac{\mathbf{u}_{n-4}(s.1010)}{\mathbf{v}_n(s)} = \frac{1}{\theta} + \frac{1}{\theta^3} + \frac{1}{\theta^4} = \frac{1}{\theta} + \frac{1}{\theta^2} = 1.$$

With our conventions, a real number belonging to $[\frac{1}{\theta}, \frac{1}{\theta} + \frac{1}{\theta^4}]$ (resp. $[\frac{1}{\theta} + \frac{1}{\theta^4}, \frac{1}{\theta} + \frac{1}{\theta^3}]$ or $[\frac{1}{\theta} + \frac{1}{\theta^3}, 1]$) has a representation having a prefix 1000 (resp. 1001 or 1010). On the other hand, if we use the greedy algorithm to compute the θ-development of real numbers in $[\frac{1}{\theta}, 1]$, we obtain exactly the same intervals. For instance, if x belongs to $[\frac{1}{\theta} + \frac{1}{\theta^4}, \frac{1}{\theta} + \frac{1}{\theta^3}[$ then it is obvious that the θ-development of x, $e_\theta(x)$, begins with 1001.

It is clear that at each step of the procedure, the intervals I_w and the intervals derived from the greedy algorithm computing θ-developments are the same. So if we use $\rho_U(\mathbb{N})$ or the greedy algorithm then we obtain the same representation. (For the abstract numeration system built on $\rho_U(\mathbb{N})$, a real number can have two representations if it is the endpoint of some interval; hence, the representations coincide if we choose for each endpoint its greatest representation with respect to the lexicographic ordering.)

Consider an arbitrary Pisot number θ. It is well known that the θ-development of one is finite or ultimately periodic [**61**]. In the first case,

$$e_\theta(1) = t_1 \cdots t_m$$

and we define, as usual,

$$e_\theta^*(1) = (t_1 \cdots t_{m-1}(t_m - 1))^\omega.$$

It is clear that we still have

$$1 = \frac{t_1}{\theta} + \frac{t_2}{\theta^2} + \cdots + \frac{t_m - 1}{\theta^m} + \frac{t_1}{\theta^{m+1}} + \cdots .$$

In the second case, there exist minimal integers $N \geq 0$, $p \geq 1$ such that

$$e_\theta(1) = t_1 \cdots t_N \, (t_{N+1} \cdots t_{N+p})^\omega.$$

Remark VI.8.3. When the θ-development of one is finite or ultimately periodic, then a particular polynomial naturally arises. In the first case, we introduce the polynomial

$$b(X) = X^m - \sum_{i=1}^{m} t_i \, X^{m-i}.$$

It is clear that $b(\theta) = 0$. If $e_\theta(1)$ is ultimately periodic then we define

$$b(X) = X^{N+p} - \sum_{i=1}^{N+p} t_i \, X^{N+p-i} - X^N + \sum_{i=1}^{N} t_i \, X^{N-i}.$$

It is clear that $b(\theta) = 0$ since

$$1 = \frac{t_1}{\theta} + \cdots + \frac{t_N}{\theta^N} + \left(\frac{t_{N+1}}{\theta^{N+1}} + \cdots + \frac{t_{N+p}}{\theta^{N+p}}\right)\left(1 + \frac{1}{\theta^p} + \frac{1}{\theta^{2p}} + \cdots\right).$$

Since N and p have been chosen to be minimal, then we say that $b(X)$ is the *canonical beta polynomial* for θ.

Let $\beta > 1$ be a real number. The set D_β of all β-developments of numbers in $[0, 1]$ is characterized as follows.

Theorem VI.8.4. [**51**] *Let $\beta > 1$ be a real number. A sequence $(x_n)_{n\geq 1}$ belongs to D_β if and only if for all $i \in \mathbb{N}$, the shifted sequence $(x_{n+i})_{n\geq 1}$ is lexicographically less than the sequence $e_\beta(1)$ or $e_\beta^*(1)$ whenever $e_\beta(1)$ is finite.*

For any real number $\beta > 1$, we denote by $F(D_\beta)$, the set of finite factors of the sequences in D_β. Bertrand numeration systems are characterized by Bertrand's Theorem given below. Notice that U is not necessarily linear.

Theorem VI.8.5. [**10**] *Let $U = (U_n)_{n\in\mathbb{N}}$ be a positional numeration system. Then U is a Bertrand numeration system if and only if there exists a real number $\beta > 1$ such that $0^*\rho_U(\mathbb{N}) = F(D_\beta)$. In this case,* *if* $e_\beta(1) = (d_n)_{n\ge 1}$ *(or* $e^*_\beta(1) = (d_n)_{n\ge 1}$ *whenever* $e_\beta(1)$ *is finite) then* $U_0 = 1$ *and*

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

Let $\theta > 1$ be a Pisot number. In what follows, we assume that $e_\theta(1)$ is ultimately periodic

$$e_\theta(1) = t_1 \cdots t_N\, (t_{N+1} \cdots t_{N+p})^\omega.$$

The Bertrand numeration system $U_\theta = (U_n)_{n\in\mathbb{N}}$ belonging to the class of positional systems related to θ is a linear numeration system satisfying the recurrence relation

$$\begin{aligned} U_n \ &= \ t_1\, U_{n-1} + \cdots + t_{p-1}\, U_{n-p+1} + (t_p + 1)\, U_{n-p} \\ &\quad + (t_{p+1} - t_1)\, U_{n-p-1} + \cdots + (t_{N+p} - t_N)\, U_{n-N-p},\ \ n \ge N + p. \end{aligned}$$

In other words, $(U_n)_{n\in\mathbb{N}}$ satisfies the canonical beta polynomial of θ. In what follows, θ is fixed and we denote U_θ simply by U.

The main point is the following. Since θ is a Pisot number, the set $F(D_\theta) = 0^*\rho_U(\mathbb{N})$ is recognizable by a finite automaton $\mathcal{A}$ [**33**] (the θ-shift is sofic). This automaton has $N + p$ states $q_1, \ldots, q_{N+p}$. For each $i \in \{1, \ldots, N + p\}$, there are edges labeled by $0, 1, \ldots, t_i - 1$ from q_i to q_1, and an edge labeled t_i from q_i to q_{i+1} if $i < N + p$. Finally, there is an edge labeled t_{N+p} from q_{N+p} to q_{N+1}. All states are final and q_1 is the initial state. The set $F(D_\theta)$ is recognized by the automaton depicted in Figure VI.9 (the sink is not represented).

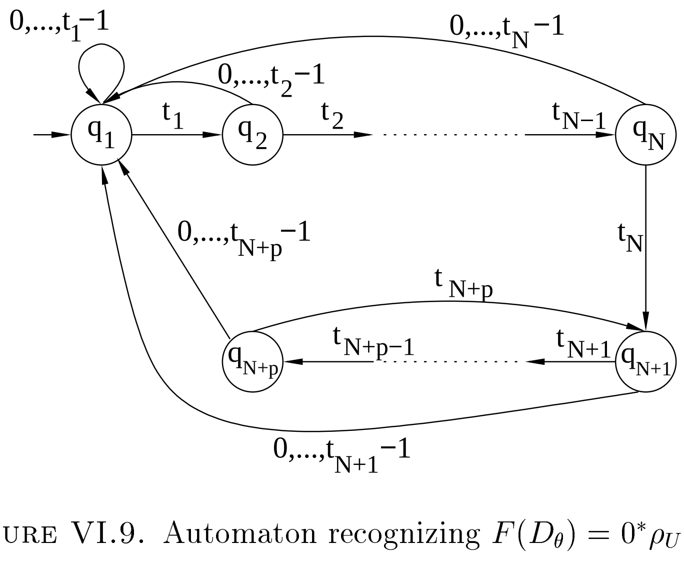

FIGURE VI.9. Automaton recognizing $F(D_\theta) = 0^*\rho_U(\mathbb{N})$.

REMARK VI.8.6. We can check that the characteristic polynomial of the incidence matrix A of $\mathcal{A}$,

$$A = \left(\begin{array}{cccc|ccccc}
t_1 & 1 & & & & & & & \\
t_2 & 0 & \ddots & & & & & & \\
\vdots & \vdots & \ddots & 1 & & & & & \\
t_N & 0 & \cdots & 0 & 1 & & & & \\
\hline
t_{N+1} & 0 & \cdots & 0 & 0 & 1 & & & \\
t_{N+2} & 0 & \cdots & 0 & 0 & 0 & \ddots & & \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & 1 & \\
t_{N+p-1} & 0 & & 0 & 0 & 0 & \cdots & 0 & 1 \\
t_{N+p} & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0
\end{array}\right)$$

is the canonical beta polynomial of θ. Hence for each $i \in \{1, \ldots, N+p\}$, the sequence $\mathbf{u}_n(q_i)$ satisfies the same recurrence relation as $(U_n)_{n\in\mathbb{N}}$. We could use this information but we are lucky to get more thanks to the characterization of $\mathcal{A}$.

In an abstract numeration system, allowing leading zeroes changes the representations. Therefore, we modify slightly the automaton $\mathcal{A}$ to obtain an automaton $\mathcal{A}'$ recognizing $\rho_U(\mathbb{N})$. We add a new state s. There are edges labeled by $1, \ldots, t_1 - 1$ from s to q_1 and an edge labeled t_1 from s to q_2. This state s is the initial state of $\mathcal{A}'$ and is also final. The automaton $\mathcal{A}'$ is sketched in Figure VI.10.

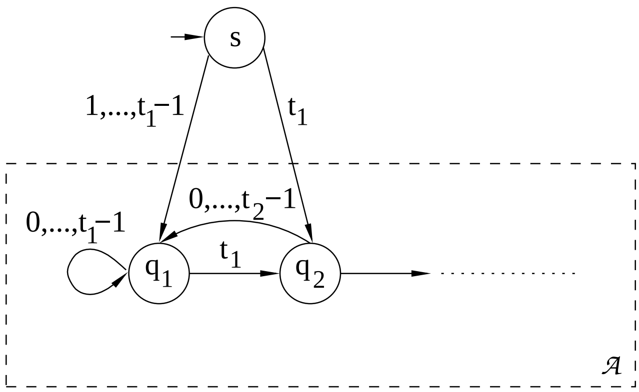

FIGURE VI.10. Automaton recognizing $\rho_U(\mathbb{N})$.

Our task is now to determine the different sequences $\mathbf{u}_n(q_i)$. To that end, we use the specific form of $\mathcal{A}'$. The first word of length $n+1$ in $\rho_U(\mathbb{N})$ is $1(0)^n$ and its numerical value is U_n. In the same way, $1(0)^{n-1}$ is the first word of length n. Therefore,

$$\mathbf{u}_n(s) = U_n - U_{n-1}.$$

Since θ is a Pisot number and the characteristic polynomial of U is the minimal polynomial of θ, there exists a real number γ such that

$$U_n \sim \gamma\,\theta^n.$$

For $n \geq 1$, it is clear that

$$\mathbf{u}_n(s) = (t_1 - 1)\,\mathbf{u}_{n-1}(q_1) + \mathbf{u}_{n-1}(q_2)$$

and

$$\mathbf{u}_n(q_1) = t_1\,\mathbf{u}_{n-1}(q_1) + \mathbf{u}_{n-1}(q_2) = \mathbf{u}_n(s) + \mathbf{u}_{n-1}(q_1).$$

Consequently,

$$\mathbf{u}_n(q_1) = \mathbf{v}_n(s).$$

We also have $\mathbf{u}_{n-1}(q_2) = \mathbf{u}_n(q_1) - t_1\,\mathbf{u}_{n-1}(q_1)$ and thus

$$\mathbf{u}_n(q_2) = \mathbf{v}_{n+1}(s) - t_1\,\mathbf{v}_n(s).$$

But $\mathbf{u}_n(q_2) = t_2\,\mathbf{u}_{n-1}(q_1) + \mathbf{u}_{n-1}(q_3)$. So we find

$$\mathbf{u}_n(q_3) = \mathbf{v}_{n+2}(s) - t_1\,\mathbf{v}_{n+1}(s) - t_2\,\mathbf{v}_n(s).$$

Continuing this way, for $i \leq N + p$

$$\mathbf{u}_n(q_i) = \mathbf{v}_{n+i-1}(s) - t_1\,\mathbf{v}_{n+i-2}(s) - t_2\,\mathbf{v}_{n+i-3}(s) - \cdots - t_{i-1}\mathbf{v}_n(s).$$

We are now able to determine the endpoints of the intervals I_w. It is clear that

$$\mathbf{v}_n(s) = U_n \sim \gamma\,\theta^n.$$

Therefore, for $i \in \mathbb{N}$

$$\lim_{n\to\infty} \frac{\mathbf{u}_{n-i}(q_1)}{\mathbf{v}_n(s)} = \theta^{-i}$$

and in the same manner,

$$\lim_{n\to\infty} \frac{\mathbf{u}_{n-i}(q_2)}{\mathbf{v}_n(s)} = \theta^{1-i} - t_1\,\theta^{-i}.$$

Continuing this way, for $j \leq N + p$ and $i \in \mathbb{N}$

$$\lim_{n\to\infty} \frac{\mathbf{u}_{n-i}(q_j)}{\mathbf{v}_n(s)} = \theta^{j-i-1} - t_1\,\theta^{j-i-2} - \cdots - t_{j-1}\theta^{-i}.$$

We can now compute the different intervals. The first words in $\rho_U(\mathbb{N})$ are

$$1, \ldots, (t_1 - 1), t_1, 10, \ldots, 1t_1, 20, \ldots, (t_1 - 1)t_1, t_1 0, \ldots, t_1 t_2, \ldots$$

We have the intervals corresponding to words of length one

$$I_j = [\frac{j}{\theta}, \frac{j}{\theta} + \lim_{n\to\infty} \frac{\mathbf{u}_{n-1}(q_1)}{\mathbf{v}_n(s)}] = [\frac{j}{\theta}, \frac{j+1}{\theta}],\ j < t_1$$

and

$$I_{t_1} = [\frac{t_1}{\theta}, \frac{t_1}{\theta} + \lim_{n\to\infty} \frac{\mathbf{u}_{n-1}(q_2)}{\mathbf{v}_n(s)}] = [\frac{t_1}{\theta}, 1].$$

For the words of length two, if $j,k<t_1$ then

$$I_{jk}=[\frac{j}{\theta}+\frac{k}{\theta^2},\frac{j}{\theta}+\frac{k}{\theta^2}+\lim_{n\to\infty}\frac{\mathbf{u}_{n-2}(q_1)}{\mathbf{v}_n(s)}]=[\frac{j}{\theta}+\frac{k}{\theta^2},\frac{j}{\theta}+\frac{k+1}{\theta^2}]$$

and

$$I_{jt_1}=[\frac{j}{\theta}+\frac{t_1}{\theta^2},\frac{j}{\theta}+\frac{t_1}{\theta^2}+\lim_{n\to\infty}\frac{\mathbf{u}_{n-2}(q_2)}{\mathbf{v}_n(s)}]=[\frac{j}{\theta}+\frac{t_1}{\theta^2},\frac{j+1}{\theta}].$$

For the words of length two beginning with t_1, we have

$$I_{t_1j}=[\frac{t_1}{\theta}+\frac{j}{\theta^2},\frac{t_1}{\theta}+\frac{j+1}{\theta^2}],\ j<t_2$$

and

$$I_{t_1t_2}=[\frac{t_1}{\theta}+\frac{t_2}{\theta^2},\frac{t_1}{\theta}+\frac{t_2}{\theta^2}+\lim_{n\to\infty}\frac{\mathbf{u}_{n-2}(q_3)}{\mathbf{v}_n(s)}]=[\frac{t_1}{\theta}+\frac{t_2}{\theta^2},1].$$

On the other hand, we can associate intervals to the θ-developments of real numbers. These developments are computed through the use of the greedy algorithm. For instance, if $x\in[\frac{1}{\theta},\frac{2}{\theta}[$ then it is obvious that $e_\theta(x)$ begins with 1. By Theorem VI.8.4, $e_\theta(1)$ induces the definition of intervals related to θ-developments and it is clear that these intervals coincide exactly with the intervals I_w.

Remark VI.8.7. If $e_\theta(1)$ is finite, that is if $e_\theta(1)=t_1\cdots t_m$ with $t_m\neq 0$ and $t_n=0$ for $n>m$, then the construction of $\mathcal{A}$ still holds with $N=m$ and $p=0$. All the edges from q_m lead to q_1 and are labeled by $0,\ldots,t_m-1$. Indeed, by Theorem VI.8.4, the shifted sequences in D_θ are less than $e_\theta^*(1)$. The automaton $\mathcal{A}$ is represented in Figure VI.11. As an example, the golden mean is such that $e_\theta(1)=11$ and

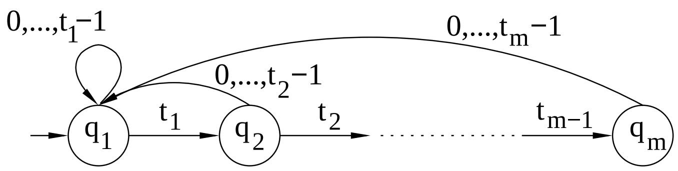

Figure VI.11. Automaton recognizing $0^*\rho_U(\mathbb{N})$ when $e_\theta(1)$ is finite.

the automaton recognizing $\rho_U(\mathbb{N})$ is depicted in Figure VI.8.

To conclude this section, we give a small example showing that generally θ-developments and abstract representations do not coincide.

Example VI.8.8. We consider the language $L=0^*1\{01,0\}^*\cup\{\varepsilon\}$. It is simply the language of the normalized representations in the Fibonacci system when we allow leading zeroes. The minimal automaton of L in Figure VI.12 differs slightly from the one depicted in Figure VI.8. In particular, the sequences $(\mathbf{u}_n(q_1))_{n\in\mathbb{N}}$ and $(\mathbf{u}_n(q_2))_{n\in\mathbb{N}}$ are the same as in Example VI.8.2. The real number θ related to this system

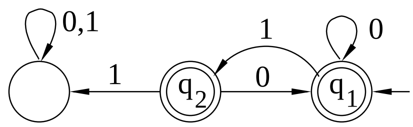

FIGURE VI.12. The minimal automaton of $0^*1\{01,0\}^* \cup \{\varepsilon\}$.

is once again the golden mean. Consider the real number
$$x = \frac{1}{\theta} + \lim_{n\to\infty} \frac{\mathbf{u}_{n-1}(q_1.0)}{\mathbf{v}_n(q_1)} = \frac{1}{\theta} + \frac{1}{\theta^3}.$$
It is the upper bound of I_0 and the lower bound of I_1. Using the language L, we obtain two representations of x,
$$(01)^\omega \text{ and } 1(0)^\omega.$$
On the other hand,
$$e_\theta(x) = 101.$$

9. Uniform continuity

In this section, we consider the application $g : \mathcal{L}_\infty \to]\frac{1}{\theta}, 1[$ which maps an infinite word w, limit of a convergent sequence $(w_n)_{n\in\mathbb{N}}$ of words of L, to the real number $\lim_{n\to\infty} \frac{\mathrm{val}_S(w_n)}{\mathbf{v}_{|w_n|}}$. In other words, g maps a representation onto its numerical value.

PROPOSITION VI.9.1. *The function g is uniformly continuous.*

Proof. Let $w, w' \in \mathcal{L}_\infty$. By Corollary VI.5.4, we have $|g(w) - g(w')|$ equal to
$$\frac{\theta-1}{\theta}\left|\sum_{q\in K}\frac{a_q}{a_s}\sum_{j=0}^{\infty}\beta_{q,j}(w)\,\theta^{-j} - \sum_{q\in K}\frac{a_q}{a_s}\sum_{j=0}^{\infty}\beta_{q,j}(w')\,\theta^{-j}\right|.$$
If w and w' have the same prefix of length M, then $\beta_{q,j}(w) = \beta_{q,j}(w')$ for $j = 0, \ldots, M-1$ and
$$\begin{aligned}
|g(w) - g(w')| &\le \frac{\theta-1}{\theta}\sum_{q\in K}\frac{a_q}{a_s}\left|\sum_{j=M}^{\infty}\beta_{q,j}(w)\,\theta^{-j} - \sum_{j=M}^{\infty}\beta_{q,j}(w')\,\theta^{-j}\right| \\
&\le \frac{\theta-1}{\theta^{M+1}}\sum_{q\in K}\frac{a_q}{a_s}\left|\sum_{j=0}^{\infty}\beta_{q,j+M}(w)\,\theta^{-j} - \sum_{j=0}^{\infty}\beta_{q,j+M}(w')\,\theta^{-j}\right| \\
&\le \frac{\theta-1}{\theta^{M+1}}\sum_{q\in K}\frac{a_q}{a_s}\,2\,(\#\Sigma)\sum_{j=0}^{\infty}\theta^{-j} \\
&\le \frac{2\,(\#\Sigma)}{\theta^M}\sum_{q\in K}\frac{a_q}{a_s}.
\end{aligned}$$

Since $\theta > 1$, for any $\varepsilon > 0$, the latter bound can be taken less than ε if M is large enough. Hence the conclusion because $d(w, w') = 2^{-M}$.

□

PROPOSITION VI.9.2. *The function g is monotone: if $w, x \in \mathcal{L}_\infty$ and $w < x$ then*

$$g(w) \leq g(x).$$

Proof. There exists $l \in \mathbb{N}$ such that $w_{[0,l-1]} = x_{[0,l-1]}$ and $w_{[l,l]} < x_{[l,l]}$. By definition of the intervals,

$$y \in I_{w[0,l]}, z \in I_{x[0,l]} \Rightarrow y \leq z.$$

To conclude, observe that $g(w) \in I_{w[0,l]}$ and $g(x) \in I_{x[0,l]}$.

□

APPENDIX A

About Pell's equation

An equation in integers of the form

$$X^2 - \alpha Y^2 = N$$

where α is not a perfect square and $N > 0$, is said to be a *Pell's equation*. In this section, we explain how to solve such equations.

The next proposition summarizes some well-known facts that are used in the proof of Theorem III.1.2. The reader will find in [**22**] or in [**59**] the material necessary to achieve its proof.

Proposition A.1.1. *Assume that $\alpha \in \mathbb{N}$ is not a perfect square and that $N > 0$ is a natural number.*

(i) *The set of solutions $(X, Y) \in \mathbb{N}^2$ of the equation $X^2 - \alpha Y^2 = N$ is the (finite) union of the sequences $(X_n, Y_n)_{n \in \mathbb{N}}$ defined by*

$$\begin{pmatrix} X_{i+1} \\ Y_{i+1} \end{pmatrix} = \begin{pmatrix} u & \alpha\, v \\ v & u \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \end{pmatrix}, \tag{37}$$

$$\forall i \in \mathbb{N}, \text{ and } 0 < X_0 \leq u\sqrt{N},$$

where the couple $(u, v) \in \mathbb{N}^2$ is the minimal non-trivial solution of $U^2 - \alpha V^2 = 1$, i.e., that for which $u > 1$ is the smallest.

(ii) *Each component of any solution $(X_n, Y_n)_{n \in \mathbb{N}}$ of (37) is a solution of*

$$Z_{i+2} = 2uZ_{i+1} - Z_i, \ \forall i \in \mathbb{N}.$$

In particular, X_{2n}, X_{2n+1}, Y_{2n} and Y_{2n+1} are of the same parity as X_0, X_1, Y_0 and Y_1 respectively.

(iii) *For any solution $(X_n, Y_n)_{n \in \mathbb{N}}$ of (37), one has $X_n > u^n$.*

In view of the previous proposition, one has to study the equation $U^2 - \alpha V^2 = 1$. (Notice that α cannot be a perfect square since the difference of two perfect squares is never equal to 1 except for $1^2 - 0^2$.) The minimal non-trivial solution of $U^2 - \alpha V^2 = 1$ is given by the development of $\sqrt{\alpha}$ in continued fractions (see for instance, Chapter IV of [**22**]). It is well known that any quadratic irrational has an ultimately periodic continued fraction of the form

$$\sqrt{\alpha} = (q_0, q_1, \ldots, q_n, 2q_0, q_1, \ldots, q_n, 2q_0, q_1 \ldots).$$

This latter writing is a shortcut for

$$\sqrt{\alpha} = q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \cfrac{1}{q_3 + \cfrac{1}{\ddots}}}}$$

The convergents of a continued fraction are

$$\frac{A_0}{B_0} = \frac{q_0}{1},\ \frac{A_1}{B_1} = q_0 + \frac{1}{q_1},\ \frac{A_2}{B_2} = q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2}}, \ldots$$

If the length l of the period $(q_1, \ldots, q_n, 2q_0)$ is even, the minimal solution (u, v) of $U^2 - \alpha V^2 = 1$ is given by (A_{l-1}, B_{l-1}) (for a proof of this result see pages 108–109 of [**22**]).

Example A.1.2. Search the minimal solution of $U^2 - 21\,V^2 = 1$. One has

$$\sqrt{21} = (4, \underbrace{1, 1, 2, 1, 1, 8,}_{\text{length } l=6} 1, 1, 2, 1, 1, 8, \ldots).$$

The first convergents are

$$\frac{A_0}{B_0} = \frac{4}{1},\ \frac{A_1}{B_1} = 4 + \frac{1}{1} = \frac{5}{1},\ \frac{A_2}{B_2} = 4 + \cfrac{1}{1 + \cfrac{1}{1}} = \frac{9}{2}, \ldots$$

Thus one gets $\frac{A_5}{B_5} = \frac{55}{12}$ and $(u, v) = (55, 12)$.

Otherwise, the length l of the period $(q_1, \ldots, q_n, 2q_0)$ is odd and the minimal solution (u, v) of $U^2 - \alpha V^2 = 1$ is given by (A_{2l-1}, B_{2l-1}).

Example A.1.3. Search the minimal solution of $U^2 - 29\,V^2 = 1$. One has

$$\sqrt{29} = (5, \underbrace{2, 1, 1, 2, 10,}_{\text{length } l=5} 2, 1, 1, 2, 10, \ldots).$$

The first convergents are

$$\frac{A_0}{B_0} = \frac{5}{1},\ \frac{A_1}{B_1} = 5 + \frac{1}{2} = \frac{11}{2},\ \frac{A_2}{B_2} = 5 + \cfrac{1}{2 + \cfrac{1}{1}} = \frac{16}{3}, \ldots$$

Continuing this way, one obtains $\frac{A_9}{B_9} = \frac{9801}{1820}$ and $(u, v) = (9801, 1820)$.

APPENDIX B

Proof of proposition IV.5.2

In [**47**], C. Mauduit shows that the characteristic sequence $\chi^{\mathcal{P}}$ of the set of primes is not generated by a substitution (φ, h, a_1) where $\varphi : \Sigma = \{a_1, \ldots, a_g\} \to \Sigma^*$ and $h : \Sigma \to \{0, 1\}$ sends all the letters of Σ onto 0 except one. Here, we consider the case where h sends more than one letter of Σ onto 1. In the first part of this appendix, we give, for the sake of completeness, a variation of Mauduit's proof. Next, we explain the modifications.

1. Preliminaries

First, recall the *prime number theorem* (see for instance [**36**]). The *prime counting function* $\pi_{\mathcal{P}}(n)$ counts the number of primes less than n and the theorem stipulates that

$$\pi_{\mathcal{P}}(n) \sim \frac{n}{\log n}.$$

Let $\varphi : \Sigma = \{a_1, \ldots, a_g\} \to \Sigma^*$ be an homomorphism such that $\varphi(a_1) \in a_1\Sigma^+$. We denote by

$$l_i(n) = |\varphi^n(a_i)|,\ i = 1, \ldots, g,$$

the length of the image of φ^n on the i^{th} letter of the alphabet. We denote by $D_i(N)$, the number of a_i's appearing in the first N letters of the infinite word $\varphi^\omega(a_1)$.

With φ, one can build a matrix

$$(M_{ij})_{(i,j)\in\{1,\ldots g\}^2}$$

such that M_{ij} is the number of a_j's in the word $\varphi(a_i)$. We fix $a_t \in \Sigma$ and denote by $I(t)$ the subset of $\{1, \ldots, g\}$ such that $i \in I(t)$ if and only if there exists an integer n such that $M^n_{it} > 0$. In other words, $i \in I(t)$ if a_t appears in $\varphi^n(a_i)$ for some n. We denote by r the spectral radius[1] of M and by $d + 1$ the maximal order of the Jordan's blocks associated to r. We denote by $M^{I(t)}$ the matrix $(M_{ij})_{(i,j)\in I(t)^2}$. For this latter matrix, we define r_t and d_t as for M.

[1]Notice that the elements of M are non-negative. So M has a real non-negative eigenvalue r and any other eigenvalue λ of M is such that $|\lambda| \le r$.

Proposition B.1.1. [**47**, Prop. 1, p. 184] *There exist non-zero real numbers C and C_t which are algebraic over $\mathbb{Q}$ such that*

$$\begin{array}{rcl} M_{1t}^n & \sim & C_t\, n^{d_t}\, r_t^n \\ l_1(n) & \sim & C\, n^d\, r^n. \end{array}$$

2. First part of the proof

For the sake of simplicity, since we consider the only infinite word $\varphi^\omega(a_1)$, we will denote $l_1(n)$ by l (l depends on n). Since all the letters are sent onto 0 except one, say a_t, we will simply denote $D_t(l_1(n))$ by D (depending on n).

Proof. Assume that $\chi^{\mathcal{P}}$ is generated by such a substitution (φ, h, a_1). With these notations, one has

$$\lim_{n\to\infty} \frac{D \log l}{l} = 1, \tag{38}$$

$$\lim_{n\to\infty} \frac{D}{C_t\, n^{d_t}\, r_t^n} = 1, \tag{39}$$

$$\lim_{n\to\infty} \frac{l}{C\, n^d\, r^n} = 1. \tag{40}$$

The first equation is the translation of the prime number theorem and the last two equations are given by Proposition B.1.1.

By [**46**, Cor. 1, p. 242], there exist $(\phi, \psi) \in \mathbb{R} \times [0,1]$ and α, $\alpha' \in]0, +\infty[$ such that, for n large enough,

$$\alpha\, (\log l)^\phi\, l^\psi \le D \le \alpha'\, (\log l)^\phi\, l^\psi$$

with $\phi \in \mathbb{Z}(\psi)$. Therefore, using (38), there exists constants β and β' such that

$$0 < \beta' \le (\log l)^{\phi+1}\, l^{\psi-1} \le \beta < \infty. \tag{41}$$

- If $\psi < 1$, then

$$\lim_{n\to\infty} (\log l)^{\phi+1}\, l^{\psi-1} = 0$$

which contradicts the l.h.s. of (41).

- If $\psi = 1$ and $\phi+1 > 0$ (resp. $\phi+1 < 0$) then $(\log l)^{\phi+1} \to +\infty$ (resp. $(\log l)^{\phi+1} \to 0$) if n tends to infinity which also contradicts (41).

- The last case, $\psi = 1$ and $\phi = -1$ splits into two sub-cases. First, assume that $r > 1$. A refinement of Proposition B.1.1 gives

$$\left\{ \begin{array}{rcl} \phi & = & d_t - d \log_r r_t \\ \psi & = & \log_r r_t \end{array} \right.$$

and thus, $r_t = r$ and $d_t = d - 1$, since $\psi = 1$ and $\phi = -1$. In view of (38), (39) and (40),

$$\lim_{n\to\infty} \frac{D \log l}{l}\, \frac{l}{C\, n^d\, r^n}\, \frac{C_t\, n^{d_t}\, r_t^n}{D} = 1.$$

Therefore, in our case,

$$\lim_{n\to\infty} \frac{C_t \log l}{C\, n} = 1.$$

Taking the logarithm of (40) leads us to

$$\lim_{n\to\infty} \frac{C_t}{C\, n}(\log C + d \log n + n \log r) = 1$$

and therefore,

$$\lim_{n\to\infty} \frac{C_t}{C} \log r = 1.$$

In this latter expression, C_t, C and r are constants, so

$$\log r = \frac{C}{C_t}.$$

Since C and C_t are algebraic over $\mathbb{Q}$, $\frac{C}{C_t}$ is also algebraic over $\mathbb{Q}$. This leads to a contradiction, $r > 1$ is algebraic over $\mathbb{Q}$ and by Hermite-Lindemann theorem [**24**, pp. 128–137], $\log r$ is transcendental.

If $r = 1$ (the second sub-case) then $\phi = 0$ and this case was already discussed.

□

3. Generalization of the proof

Let (φ, h, a_1) be a substitution such that

$$\varphi : \Sigma = \{a_1, \ldots, a_k, a_{k+1}, \ldots, a_g\} \to \Sigma^*$$

and to simplify the notations, we assume that

$$h(a_i) = \begin{cases} 1, & \text{if } i = 1, \ldots, k; \\ 0, & \text{otherwise.} \end{cases}$$

We need to extend the notations of the previous section. For $i = 1, \ldots, k$, we have $D_i = D_i(l_1(n))$ and $l = l_1(n)$. Notice that D_i and l depend on n.

Proof. We assume that $\chi^{\mathcal{P}}$ is generated by (φ, h, a_1). Equation (38) given by the prime number theorem, becomes

$$\lim_{n\to\infty} (D_1 + \cdots + D_k) \frac{\log l}{l} = 1. \tag{42}$$

For $i = 1, \ldots, k$, by Corollary 1 page 242 of [**46**], there exist $(\phi_i, \psi_i) \in \mathbb{R} \times [0, 1]$ and α_i, $\beta_i \in]0, +\infty[$ such that, for n large enough,

$$\alpha_i\, (\log l)^{\phi_i}\, l^{\psi_i} \le D_i \le \beta_i\, (\log l)^{\phi_i}\, l^{\psi_i} \tag{43}$$

with $\phi_i \in \mathbb{Z}(\psi_i)$. If we set

$$\alpha = \inf_i \alpha_i \text{ and } \beta = \sup_i \beta_i.$$

Then, for $i \in \{1, \dots, k\}$,

$$\frac{D_i}{\beta} \leq (\log l)^{\phi_i}\, l^{\psi_i} \leq \frac{D_i}{\alpha}.$$

Therefore,

$$\frac{1}{\beta}\,(D_1 + \cdots + D_k)\,\frac{\log l}{l} \leq \sum_{i=1}^{k} (\log l)^{\phi_i+1}\, l^{\psi_i-1} \leq \frac{1}{\alpha}\,(D_1 + \cdots + D_k)\,\frac{\log l}{l}.$$

In view of (42), we have for n large enough

$$0 < \frac{1}{2\beta} \leq \sum_{i=1}^{k} (\log l)^{\phi_i+1}\, l^{\psi_i-1} \leq \frac{2}{\alpha} < \infty. \tag{44}$$

If for all $i \in \{1, \dots, k\}$, $(\psi_i, \phi_i) \neq (1, -1)$, one can conclude as in the previous section by obtaining a contradiction. Otherwise, we assume that there exists $s \in \{1, \dots, k\}$ such that

$$\begin{cases} (\psi_i, \phi_i) & = & (1, -1), \; 1 \leq i \leq s; \\ (\psi_i, \phi_i) & \neq & (1, -1), \; \text{otherwise}. \end{cases}$$

With the same arguments as in the previous section, for $i > s$, it is clear that $(\log l)^{\phi_i+1}\, l^{\psi_i-1}$ cannot tend to infinity, otherwise

$$\lim_{n\to\infty} \sum_{i=1}^{k} (\log l)^{\phi_i+1}\, l^{\psi_i-1} = \infty,$$

which is not possible in view of (44). So for $i > s$,

$$\lim_{n\to\infty} (\log l)^{\phi_i+1}\, l^{\psi_i-1} = \lim_{n\to\infty} (\log l)^{\phi_i}\, l^{\psi_i}\, \frac{\log l}{l} = 0$$

and in view of (43), for $i > s$,

$$\lim_{n\to\infty} D_i\, \frac{\log l}{l} = 0.$$

Therefore, (42) becomes

$$\lim_{n\to\infty} (D_1 + \cdots + D_s)\, \frac{\log l}{l} = 1.$$

And with (40),

$$\lim_{n\to\infty} (D_1 + \cdots + D_s)\, \frac{\log l}{C\, n^d\, r^n} = 1.$$

As in the previous section, one has

$$r_1 = \cdots = r_s = r \text{ and } d_1 = \cdots = d_s = d - 1.$$

So,

$$\lim_{n\to\infty} \left(\sum_{i=1}^{s} \frac{D_i}{C\, n^{d_i+1}\, r_i^n}\right) \log l = 1.$$

Taking the logarithm of (40) gives

$$\lim_{n\to\infty} \frac{1}{C} \, (\sum_{i=1}^{s} \frac{D_i}{n^{d_i} \, r_i^n}) \, \frac{1}{n} \, (\log C + d \, \log n + n \, \log r) = 1.$$

Equation (39) can be applied for $t = 1, \ldots, s$, so

$$\lim_{n\to\infty} \frac{C_1 + \cdots + C_s}{C} \, \log r = 1.$$

And therefore,

$$\log r = \frac{C_1 + \cdots + C_s}{C},$$

the r.h.s. is algebraic over $\mathbb{Q}$ and by Hermite-Lindemann theorem, the l.h.s. is transcendental, a contradiction.

□

APPENDIX C

Implementation with Mathematica

A lot of results given in this work were discovered through the extensive use of computer simulations. In this appendix, we give an implementation of different algorithms related to regular languages and abstract numeration systems. We have chosen Mathematica[1] for the implementation because we can easily use arbitrary large numbers and lists of arbitrary objects. In the following sections, we give source code written with Mathematica. For the reader not familiar with the syntax and basic structures like arrays or lists, see for instance [**68**].

1. Data structure for automata

A NDFA $\mathcal{A}$ is a 5-uple (Q, Σ, E, I, F). We have to code this structure. Observe that a DFA is a special case of a NDFA. In this chapter, we only use elementary NDFA. A NDFA is *elementary* if its edges are labeled by single letters or by the empty word ε. If a NDFA has edges labeled by words, then it can be easily made elementary by adding new states. For instance, if $p \stackrel{ab}{\rightarrow} q$ then this transition can be replaced by $p \stackrel{a}{\rightarrow} p' \stackrel{b}{\rightarrow} q$.

We define an ordered list `alphabet` containing the different letters of Σ (with respect to the ordering) and an extra symbol for the empty word (in a NDFA, we can have ε-transitions). This extra symbol has no special role in the data structure, except as a reminder. The states q_i of $\mathcal{A}$ are enumerated by $1, \ldots, \#Q$ and the i^{th} element of the list `transMat` contains the ordered list of the sets of states reached from the i^{th} state of $\mathcal{A}$ by reading the different letters of Σ with respect to its ordering (`transMat` stands for *transition matrix*). In other words, `k` belongs to the set `transMat[[i,j]]` if and only if $q_i.\sigma_j = q_k$. The lists `finState` and `iniState` are equal respectively to F and I.

EXAMPLE C.1.1. For instance, consider the NDFA $\mathcal{E}$ depicted in Figure C.1, with a set of states $\{1, 2, 3\}$. This automaton has the following transition table:

	a	b	ε
1	2	1	3
2	3	–	–
3	1, 2	3	–

[1] Mathematica® is a trademark of Wolfram Research.

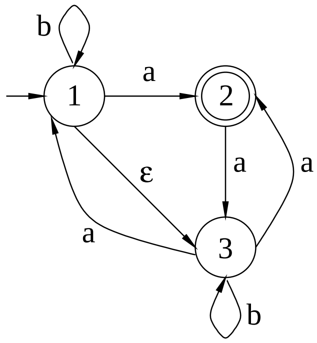

Figure C.1. A NDFA with three states and ε-transitions.

With our definitions, we get the following implementation:

```
alphabet={"a","b","ε"};
transMat={{{2},{1},{3}},{{3},{},{}},{{1,2},{3},{}}};
finState={2};
iniState={1};
```

Remark C.1.2. For the coding of a DFA, `iniState` has to contain a single element and `transMat[[i]]` is a list of arrays, each containing a single element; the last array being empty (no ε-transition). So the next function `IsDFA` returns the value `True` if the actual automaton is deterministic and `False`, otherwise.

```
IsDFA:=Module[{i,j,x}
   x=True;
   If[Length[iniState]!=1,x=False,
    For[i=1,(i<=Length[transMat])&&(x==True),
     If[Length[transMat[[i,Length[alphabet]]]]>0,x=False];
     For[j=1,(j<Length[alphabet])&&(x==True),
      If[Length[transMat[[i,j]]]!=1,x=False];
      j++];
     i++];
   ];
   x
  ]
```

2. Mirror of an automaton

The mirror of an automaton $\mathcal{A} = (Q, \Sigma, E, I, F)$ is the automaton $\mathcal{A}^R = (Q, \Sigma, E^R, F, I)$ where

$$(p, \sigma, q) \in E \Leftrightarrow (q, \sigma, p) \in E^R.$$

In the following procedure, if `k` belongs to `transMat[[i,j]]` then the new transition matrix of the mirror automaton built in `temp` must be such that `i` belongs to `temp[[k,j]]`. So we have the following program.

```
Mirror:=Module[{i,j,k,temp},
   temp=Table[{},{i,1,Length[transMat]},
```

```
    {j,1,Length[alphabet]}];
   For[k=1,k<=Length[transMat],
    For[i=1,i<=Length[transMat],
     For[j=1,j<=Length[alphabet],
      If[MemberQ[transMat[[i,j]], k],
       temp[[k,j]]=Union[temp[[k,j]],{i}]];
      j++];
     i++];
    k++];
   transMat=temp;
   temp=finState;
   finState=iniState;
   iniState=temp;
  ];
```

Example C.2.1. Continuing Example C.1.1, the mirror automaton of $\mathcal{E}$ is depicted in Figure C.2. After application of `Mirror`, we get

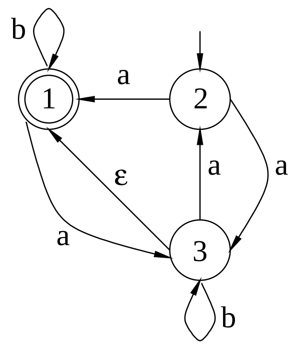

Figure C.2. The mirror automaton $\mathcal{E}^R$.

```
transMat={{{3},{1},{}},{{1, 3},{},{}},{{2},{3},{1}}}
finState={1}
iniState={2}
```

3. Determinization of an NDFA

In this section, we implement the *subset construction* to build, from a NDFA $\mathcal{A}$ accepting a language L, a DFA accepting the same language. (See for instance [**70**], for the description of this classical method.) In the first procedure of this section, we consider the case of ε-transitions and define a new list `statesWithoutRead` of sets of states. A state q_j, coded by the integer `j`, belongs to the set `statesWithoutRead[[i]]` if and only if $q_i.\varepsilon = q_j$, that is if q_j can be reached from q_i without the reading of a letter.

```
EpsilonTrans:=Module[{i,j,k,x,stateStatus},
  statesWithoutRead={};
  For[k=1,k<=Length[transMat],
```

```
   stateStatus = Table[0,{i,1,Length[transMat]}];
   stateStatus[[k]] = 1;    (* The state k is visited. *)
   (* Code : 0 = None, 1 = Visited, 2 = To be visited. *)

   (* Is there some epsilon-transitions from k ?
      If it is the case, the states must be visited. *)

   If[Length[x=transMat[[k,Length[alphabet]]]]>0,
    For[i=1,i<=Length[x],
     If[stateStatus[[x[[i]]]]==0,
      stateStatus[[x[[i]]]]=2;];
     i++];

    (* While there is some state to be visited, a state is
       visited and checked for epsilon-transitions. *)

    While[Length[Position[stateStatus,2]]>0,
     j=Position[stateStatus,2][[1,1]];
     stateStatus[[j]]=1;
     If[Length[x=transMat[[j,Length[alphabet]]]]>0,
      For[i=1,i<=Length[x],
       If[stateStatus[[x[[i]]]]==0,
        stateStatus[[x[[i]]]]=2];
       i++]];
    ];
   ];

   (* The states visited are the states reached
      from k by epsilon-transitions. *)

   statesWithoutRead=Append[statesWithoutRead,
     Flatten[Position[stateStatus,1]]];
   k++];
]
```

The following procedure implements the subset construction. Notice that there is a call to the procedure `EpsilonTrans`.

```
Determinization:=
 Module[{states={Sort[iniState]},c=0,numberStates=1,
   i,j,k,x,y,newTrans,newFin},
  EpsilonTrans;
  (* newTrans is the transition matrix of the
     determinized automaton.                   *)
  newTrans={};

  While[c!=numberStates,
   c++;
   newTrans=Append[newTrans,
```

```
     Table[{},{i,1,Length[alphabet]}]];

   (* y contains the states reached from states[[c]]
      by epsilon-transitions. *)
   y={};
   For[k=1,k<=Length[states[[c]]],
    y=Union[y,statesWithoutRead[[states[[c,k]]]]];
    k++];

   (* The subset construction.*)
   For[j=1,j<Length[alphabet],
    x={};
    For[i=1,i<=Length[y],
     x=Union[x,transMat[[y[[i]],j]]];
     i++];
    x=Sort[x];
    If[MemberQ[states,x]==False,states=Append[states,x];
     numberStates=numberStates+1;];
    newTrans[[c,j]]=Flatten[Position[states,x]];
    j++];
  ];
  transMat=newTrans;
  iniState=Flatten[Position[states,Sort[iniState]]];

  (* Determining the final states. *)
  newFin={};
  For[i=1,i<=Length[states],
   If[Length[Intersection[states[[i]],finState]]>0,
    newFin=Union[newFin,{i}]];
   i++];
  finState=newFin;
 ]
```

Example C.3.1. Continuing Example C.1.1. After application of the procedure `Determinization` to $\mathcal{E}^R$, we find

```
transMat={{{2},{3},{}},{{4},{2},{}},{{3},{3},{}},
          {{5},{2},{}},{{5},{2},{}}}
iniState={1}
finState={2,5}
```

Indeed, the subset construction gives by hand the following table:

States	a	b
$\{2\}$	$\{1,3\}$	$\emptyset$
$\{1,3\}$	$\{2,3\}$	$\{1,3\}$
$\emptyset$	$\emptyset$	$\emptyset$
$\{2,3\}$	$\{1,2,3\}$	$\{1,3\}$
$\{1,2,3\}$	$\{1,2,3\}$	$\{1,3\}$

This automaton is depicted in Figure C.3

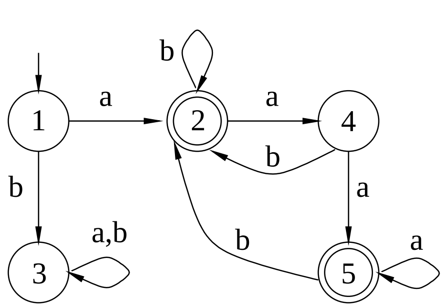

FIGURE C.3. The determinization of $\mathcal{E}^R$.

REMARK C.3.2. It is well known that if $\mathcal{A}$ is a DFA accepting the regular language L then applying twice the procedures `Mirror` and `Determinization` gives the minimal automaton of L (see pp. 43–49 of [**28**]). Hence we continue Example C.1.1 and want to obtain the minimal automaton of the language $L(\mathcal{E})$ accepted by $\mathcal{E}$. First, we have to apply `Determinization`. Next, we need two successive applications of `Mirror` and `Determinization` to obtain the automaton depicted in Figure C.4. The languages accepted from the different states are

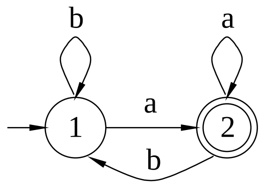

FIGURE C.4. The minimal automaton of $L(\mathcal{E})$.

$L(\mathcal{E}) = L_1 = \{a, b\}^* a$ and $L_2 = \{\varepsilon\} \cup L_1$.

4. Implementing the function val_q

Having now the minimal automaton[2] $\mathcal{A}_L$ of a regular language L over an ordered alphabet $(\Sigma, <)$ at our disposal, we can compute the first terms of the sequences $(\mathbf{u}_n(q))_{n\in\mathbb{N}}$ and $(\mathbf{v}_n(q))_{n\in\mathbb{N}}$ for the different states q of $\mathcal{A}_L$. With these sequences, we can use the formula (4) of Lemma I.5.2 to obtain a function computing the values of the function val_q. We consider two arrays `dataU` and `dataV` such that `dataU[[i,j]]`$=\mathbf{u}_{j-1}(q_i)$ and `dataV[[i,j]]`$=\mathbf{v}_{j-1}(q_i)$. We also define two functions `U[q,n]` and `V[q,n]` to access the data and such that `U[q,n]`$=\mathbf{u}_n(q)$ and `V[q,n]`$=\mathbf{v}_n(q)$.

[2]To compute val_q, we simply need a DFA, but the computation of rep_q requires the use of the minimal automaton. So, from now on, we will assume that we have the minimal automaton of the considered language.

First, we build an array `Adjac` for the *adjacency matrix* of the automaton, $(\texttt{Adjac})_{i,j} = n$ if and only if there exist n edges labeled by elements of Σ from q_i to q_j.

```
MatAdj:=Module[{i,j},
  Adjac=Table[0,{i,1,Length[transMat]},
    {j,1,Length[transMat]}];
  For[i=1,i<=Length[transMat],
   For[j=1,j<Length[alphabet],
    Adjac[[i,transMat[[i,j,1]]]]++;
    j++];
   i++];
 ]
```

The recurrence relation satisfied by $\mathbf{u}_n(q)$ can be computed through the use the characteristic polynomial of the matrix `Adjac`. To compute the initial conditions of the recurrence, we have to remark that $\mathbf{u}_n(q)$ is the number of paths of length n starting in q and ending in a final state. So this number is equal, in $\texttt{Adjac}^n$, to the sum of the elements at the line corresponding to q and in the rows corresponding to a final state.

```
Init:=Module[{i,j,k,x,y,temp,mat,coeff,caractPol,l},
  dataU = {}; dataV = {}; MatAdj;

  (* We put the coefficients of the recurrence satisfied
     by u in the list RecCoeff. *)

  caracteristicPol=
   Det[Adjac-z IdentityMatrix[Length[Adjac]]];
  l=Exponent[caracteristicPol,z];
  coeff=-Coefficient[caracteristicPol,z^l];
  RecCoeff=
   Table[Coefficient[caracteristicPol,z^(l-i)]/coeff,
    {i,1,l-1}];
  RecCoeff=
   Append[RecCoeff,(caracteristicPol/.z->0)/coeff];

  (* Computation of the first terms in dataU and dataV,
     the initial conditions of the recurrence. *)

  For[k=1,k<=Length[Adjac],
   x=Table[0,{i,1,Length[RecCoeff]}];
   y=Table[0,{i,1,Length[RecCoeff]}];
   If[MemberQ[finState,k],x[[1]]=1;y[[1]]=1];
   For[j=1,j<Length[RecCoeff],
     temp=0;
     mat=MatrixPower[Adjac,j];
     For[i=1,i<=Length[finState],
      temp=temp+mat[[k,finState[[i]]]];
      i++];
```

```
     x[[j+1]]=temp;
     y[[j+1]]=y[[j]]+temp;
     j++];
    dataU=Append[dataU,x];
    dataV=Append[dataV,y];
    k++];
 ]
```

After application of Init to the minimal automaton depicted in Figure C.4, we have

```
dataU={{0,1},{1,1}}
dataV={{0,1},{1,2}}
RecCoeff={2,0}
```

For instance,

$$\left\{\begin{array}{rcl}\mathbf{u}_{n+2}(q_2) & = & 2\,\mathbf{u}_{n+1}(q_2)+0\,\mathbf{u}_n(q_2),\ \forall n\geq 0,\\ \mathbf{u}_0(q_2) & = & 1,\\ \mathbf{u}_1(q_2) & = & 1,\end{array}\right.$$

Access to dataU and dataV is made through the use of the following functions. If terms do not exist in dataU or dataV, they are added to the corresponding list.

```
U[q_,n_]:=Module[{j,k},
  If[n<Length[dataU[[q]]],dataU[[q,n+1]],
   For[j=Length[dataU[[q]]]+1,j<=n+1,
    dataU[[q]]=Append[dataU[[q]],
      Sum[RecCoeff[[k]]dataU[[q,j-k]],
       {k,1,Length[RecCoeff]}]];
    j++];
   dataU[[q,n+1]]
  ]
 ]

V[q_,n_]:=Module[{j,k},
  If[n<Length[dataV[[q]]],dataV[[q,n+1]],
   U[q,n];
   For[j=Length[dataV[[q]]]+1,j<=n+1,
    dataV[[q]]=Append[dataV[[q]],
      dataV[[q,j-1]]+dataU[[q,j]]];
    j++];
   dataV[[q,n+1]]
  ]
 ]
```

We are now able to implement the function $\mathrm{val}_q(w)$.

```
val[q_,w_]:=Module[{temp,l,i,j,k,state},
  state=q;
  l=StringLength[w];
  If[l>0,
```

```
   temp=V[q,l-1];
   For[i=1,i<=l,
    j=Position[alphabet,StringTake[w,{i}]][[1,1]];
    If[j>1,temp=temp+
       Sum[U[transMat[[state,k]][[1]],l-i],{k,1,j-1}]];
    state=transMat[[state,j]][[1]];
    i++];,
   temp=0;];
  (* If w is not in Lq then val gives a negative value. *)
  If[MemberQ[finState,state],temp,-1]
 ]
```

5. Implementing the function rep_q

Here, we implement Algorithm I.5.7 for the computation of $\mathrm{rep}_q(x)$ very easily. Indeed, we have at our disposal, the minimal automaton of the language and the complexity sequences of the different states. Recall that q is a state of the minimal automaton and is thus coded by an integer `q`.

```
rep[q_,x_]:=Module[{n=0,p,r,w,i,j},
  While[x>=V[q,n],n++];
  p=q;
  r=x-V[q,n-1];
  w="";
  For[i=1,i<=n,
   j=1;
   While[r>=U[transMat[[p,j]][[1]],n-i],
    r=r-U[transMat[[p,j]][[1]],n-i];
    j++;
   ];
  p=transMat[[p, j]][[1]];
  w=w<>alphabet[[j]];
  i++];
  w
 ]
```

Notice that this procedure will not end if L_q is finite and $x \geq \#L_q$. For the sake of simplicity, we leave this procedure as it is but it could be improved to handle this kind of error (one can seek in the automaton the cycles reaching or containing a final state, L_q is finite if q cannot reach a cycle of this kind).

6. Real numbers

Sequences of words in L (and more generally words of L_q) give approximations of real numbers, see Table VI.1 page 110. The approximation given by a word $w \in L_q$ is computed by

```
RVal[q_,w_]:=val[q,w]/V[q,StringLength[w]]
```

Let us introduce two small macros.

```
Succ[q_,w_]:=rep[q,val[q,w]+1]
Pred[q_,w_]:=rep[q,val[q,w]-1]
```

In the following procedure, we want to obtain successive approximations in $L_q \cap \Sigma^{\leq leng}$ of a real number x (in the appropriate range, see Chapter VI for details) such that

$$\lim_{n\to\infty} \frac{\mathrm{val}_q(w_n)}{\mathbf{v}_{|w_n|}(q)} = x.$$

In the first part of the procedure, we find the greatest word of minimal length with a numerical approximation less or equal to x. Next from an approximation w, we build the next one. For the sake of simplicity, we assume that each word in L_t is a prefix of another word in L_t, for any state t. To speed up the computations of the next best approximation, we take the concatenation of w and $\mathrm{rep}_{s.w}(1)$ as a plausible approximation (this is justified by Corollary VI.6.5) and then try to improve it. This proceduce gives an array (stored in `Approx`) containing the different approximations.

```
Approximate[q_,x_,leng_]:=Module[{n,w,state,i,Approx},
    (* initialisation *)
    n=0;
    While[U[q,n]==0,n++];
    (* w is the greatest word of length n *)
    w=rep[q,V[q,n]-1];
    While[RVal[q,w]>x,w=Pred[q,w]];
    Approx={w};

    While[StringLength[w]<leng,
     (* simulation of the reading of w *)
     state=iniState[[1]];
     For[i=1,i<=StringLength[w],
      state=transMat[[state,
          Position[alphabet,StringTake[w,{i}]][[1,1]]]][[1]];
      i++];
     (* state is the state reached by w *)

     w=w<>rep[state,1];
     If[RVal[q, w]>x,
      While[RVal[q,w]>x,w=Pred[q,w]],
      While[RVal[q,w]<=x,w=Succ[q,w]];w=Pred[q,w]];
     Approx = Append[Approx, w];
    ];
    Approx
   ]
```

An application of `Approximate[1,0.7,15]` with the minimal automaton of $L(\mathcal{E})$ depicted in Figure C.4 gives Table C.1.

$w \in L_1$	$\mathrm{val}_1(w)/\mathbf{v}_{\lvert w\rvert}(1)$
a	0
aba	0.57142857142857142857
$abba$	0.66666666666666666667
$abbaa$	0.67741935483870967742
$abbaba$	0.69841269841269841270
$abbaaba$	0.69291338582677165354
$abbaabba$	0.69803921568627450980
$abbaabbaa$	0.69863013698630136986
$abbaabbaba$	0.69990224828934506354
$abbaabbaaba$	0.69956033219345383488
$abbaabbaabba$	0.69987789987789987790
$abbaabbaabbaa$	0.69991454034916371627
$abbaabbaabbaba$	0.69999389611182323140

TABLE C.1. Application of `Approximate[1,0.7,15]`.

7. Other operations on automata

In this section, we define some procedures which are useful to build new automata accepting new languages from previously defined automata.

First, let us assume that we have an automaton (DFA or NDFA) accepting a language L, the next procedure builds a NDFA accepting L^*.

```
Star:=Module[{i},
  If[Length[finState]>0,
   For[i=1,i<=Length[finState],
    transMat[[finState[[i]],Length[alphabet]]]=
     Union[transMat[[finState[[i]],Length[alphabet]]],
      iniState];
    i++];
  ];
  transMat=Append[transMat,
    Append[Table[{},{i,1,Length[alphabet]-1}],iniState]];
  iniState={Length[transMat]};
  finState=Union[finState,{Length[transMat]}];
 ]
```

Let us assume that we have an automaton (DFA or NDFA) accepting a language L over Σ, the next procedure builds a DFA accepting $\Sigma^* \setminus L$.

```
Complem:=Module[{i},
  Determinization;
  finState=Complement[Table[i,{i,1,Length[transMat]}],
    finState];
 ]
```

The following procedures need two automata over the same alphabet. If $\mathcal{A}$ is an automaton accepting L_1 and $\mathcal{B}$ is an automaton accepting L_2, the following procedure gives a unique NDFA over Σ accepting $L_1 L_2$. The elements defining an automaton are denoted by `transMat`i, `finState`i and `iniState`i, $i = 1, 2$. For instance, we can consider two copies of the same automaton $\mathcal{E}$ introduced in Example C.1.1.

```
alphabet={"a","b","ε"};
transMat1={{{2},{1},{3}},{{3},{},{}},{{1,2},{3},{}}};
finState1={2};
iniState1={1};
transMat2=transMat1;
finState2=finState1;
iniState2=iniState1;
```

We have the following procedure for the concatenation.

```
Concat:=Module[{i},
  transMat2=transMat2+Length[transMat1];
  iniState2=iniState2+Length[transMat1];
  iniState=iniState1;
  finState=finState2+Length[transMat1];
  transMat=Join[transMat1,transMat2];
  If[Length[finState1]>0,
   For[i=1,i<=Length[finState1],
     transMat[[finState1[[i]],Length[alphabet]]]=
      Union[transMat[[finState1[[i]],Length[alphabet]]],
       iniState2];
     i++];
  ];
 ]
```

Applying this procedure gives an automaton accepting $L(\mathcal{E})L(\mathcal{E})$,

```
transMat={{{2},{1},{3}},{{3},{},{4}},{{1,2},{3},{}},
           {{5},{4},{6}},{{6},{},{}},{{4,5},{6},{}}}
iniState={1}
finState={5}
```

In the next procedure, from $\mathcal{A}$ accepting L_1 and $\mathcal{B}$ accepting L_2, we build a NDFA accepting $L_1 \cup L_2$.

```
Uni:=Module[{},
  transMat2=transMat2+Length[transMat1];
  iniState2=iniState2+Length[transMat1];
  finState2=finState2+Length[transMat1];
  finState=Union[finState1,finState2];
  transMat=Join[transMat1,transMat2];
  transMat=Append[transMat,
    Append[Table[{},{i,1,Length[alphabet]-1}],
     Union[iniState1,iniState2]]];
  iniState={Length[transMat]}; ]
```

The following procedure computes the product $\mathcal{A} \times \mathcal{B}$ of two automata $\mathcal{A}$ and $\mathcal{B}$. The language accepted by this new automaton is $L(\mathcal{A}) \cap L(\mathcal{B})$. If q_i is the i^{th} state of $\mathcal{A}$ and q_j the j^{th} state of $\mathcal{B}$, then the position of the state (q_i, q_j) in the set of states of $\mathcal{A} \times \mathcal{B}$ is defined by

$$(i-1).\#Q_{\mathcal{B}} + j$$

where $\#Q_{\mathcal{B}}$ is the number of states of $\mathcal{B}$. If the set of initial states of $\mathcal{A}$ (resp. $\mathcal{B}$) is $I_{\mathcal{A}}$ (resp. $I_{\mathcal{B}}$) and the set of final states is $F_{\mathcal{A}}$ (resp. $F_{\mathcal{B}}$) then the set of initial states of $\mathcal{A} \times \mathcal{B}$ is $I_{\mathcal{A}} \times I_{\mathcal{B}}$ and the set of final states is $F_{\mathcal{A}} \times F_{\mathcal{B}}$.

```
Inter:=Module[{i,j,k,l,m},
  transMat=Table[Table[{},{i,1,Length[alphabet]}],
    {j,1,Length[transMat1]*Length[transMat2]}];
  For[i=1,i<=Length[transMat1],
   For[j=1,j<=Length[transMat2],
    For[k=1,k<=Length[alphabet],
     If[Length[transMat1[[i,k]]]*
        Length[transMat2[[j,k]]]>0,
      For[l=1,l<=Length[transMat1[[i,k]]],
        For[m=1,m<=Length[transMat2[[j,k]]],
         AppendTo[transMat[[(i-1)*Length[transMat2]+j,k]],
          (transMat1[[i,k,l]]-1)*Length[transMat2]+
           transMat2[[j,k,m]]];
         m++];
        l++];
      ];
      k++];
     j++];
    i++];
   iniState={};
   For[l=1,l<=Length[iniState1],
    For[m=1,m<=Length[iniState2],
     AppendTo[iniState,(iniState1[[l]]-1)*
        Length[transMat2]+iniState2[[m]]];
     m++];
    l++];
   finState={};
   If[Length[finState1]*Length[finState2]>0,
    For[l=1,l<=Length[finState1],
      For[m=1,m<=Length[finState2],
       finState = Union[finState,{(finState1[[l]]-1)*
            Length[transMat2]+finState2[[m]]}];
       m++];
      l++];
   ];
  ]
```

List of Figures

List of Tables

Bibliography

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

[2] J.-P. Allouche, Sur la complexité des suites infinies, *Journées Montoises* (Mons, 1992), *Bull. Belg. Math. Soc.* **1** (1994), 133–143.

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

[4] J.-P. Allouche, J. Shallit, *Automatic Sequences : Theory and Applications*, to appear.

[5] M. Andraşiu, J. Dassow, G. Păun, A. Salomaa, Language-theoretic problems arising from Richelieu cryptosystems, *Theoret. Comput. Sci.* **116** (1993), 339–357.

[6] J. Berstel, M. Mignotte, Deux propriétés décidables des suites récurrentes linéaires, *Bull. Soc. Math. France* **104** (1976), 175–184.

[7] J. Berstel, C. Reutenauer, *Les séries rationnelles et leurs languages*, études et recherches en informatique, Masson, Paris, (1984).

[8] A. Bertoni, D. Bruschi, M. Goldwurm, Ranking and formal power series, *Algebraic and computing treatment of noncommutative power series* (Lille, 1988), *Theoret. Comput. Sci.* **79** (1991), 25–35.

[9] A. Bertrand, Développements en base de Pisot et répartition modulo 1, *C. R. Acad. Sci. Paris Sér. A-B* **289** (1979), A1–A4.

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

[11] C. B. Boyer, Pascal's Formula for the Sums of the Powers of the Integers, *Scripta Math.* **9** (1943), 237–244.

[12] R. A. Brualdi, *Introductory Combinatorics*, North-Holland, New-York-Oxford-Amsterdam, (1977).

[13] V. Bruyère, Automata and numeration systems, *Sém. Lothar. Combin.* **35** (1995), `http://www.mat.univie.ac.at/~slc/`.

[14] V. Bruyère, G. Hansel, Bertrand numeration systems and recognizability, *Theoret. Comput. Sci.* **181** (1997), 17–43.

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

[16] O. Carton, W. Thomas, The monadic theory of morphic infinite words and generalizations, preprint (2000).

[17] C. Choffrut, W. Goldwurm, Rational transductions and complexity of counting problems. *Math. Systems Theory* **28** (1995), 437–450.

[18] A. Cobham, On the Hartmanis-Stearns problem for a class of tag machines, *IEEE Conference Record of 1968 Ninth Annual Symposium on Switching and Automata Theory*, schenectady (1968), 51–60.

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

[20] A. Cobham, Uniform tag sequences, *Math. Systems Theory* **6** (1972), 186–192.

[21] D. I. A. Cohen, *Basic Techniques of Combinatorial Theory*, John Wiley & Sons, New-York-Chichester-Brisbane, (1978).

[22] H. Davenport, *The Higher Arithmetic*, sixth edn., Cambridge University Press, Cambridge, (1992).

[23] F. M. Dekking, M. Mendès France, A. van der Poorten, Folds!, *Math. Intelligencer* **4** (1982), 130–138, 173–181, 190–195.

[24] H. Dörrie, *100 Great Problems of Elementary Mathematics*, second edn., Dover Ed., New-York, (1965).

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

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

[27] A. Ehrenfeucht, K. P. Lee, G. Rozenberg, Subword complexity of various classes of deterministic developmental languages without interactions, *Theoret. Comp. Sci.* **1** (1975), 59–75.

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

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

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

[31] C. Frougny, On the sequentiality of the successor function, *Information and Computation* **139** (1997), 17–38.

[32] C. Frougny, Numeration systems, in *Algebraic Combinatorics on words*, Cambridge University Press, Cambridge, to appear.

[33] C. Frougny, B. Solomyak, On representation of integers in linear numeration systems, in *Ergodic theory of Z^d actions* (Warwick, 1993–1994), 345–368, London Math. Soc. Lecture Note Ser. **228**, Cambridge University Press, Cambridge, (1996).

[34] G. Hansel, Une démonstration simple du théorème de Skolem-Mahler-Lech, *Theoret. Comput. Sci.* **43** (1986), 91–98.

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

[36] G. H. Hardy, E. M. Wright, *An introduction to the theory of numbers*, fourth edn., Oxford University Press, London, (1965).

[37] M. Hollander, Greedy numeration systems and regularity, *Theory Comput. Syst.* **31** (1998), 111–133.

[38] L. Kari, G. Rozenberg, A. Salomaa, L Systems, in *Handbook of formal languages*, Vol. 1, G. Rozenberg and A. Salomaa Eds., Springer, Berlin, (1997), 254–328.

[39] P. Lecomte, M. Rigo, Numeration systems on a regular language, *Theory Comput. Syst.* **34** (2001), 27–44.

[40] P. Lecomte, M. Rigo, On the representation of real number using rational languages, preprint (2000).

[41] W. J. LeVeque, *Topics in Number Theory*, Addison-Wesley, Massachusetts, (1956).

[42] H. R. Lewis, C. H. Papadimitriou, *Elements of the Theory of Computation*, Prentice-Hall, Englewood Cliffs, NJ, (1981).

[43] N. Loraud, β-shift, systèmes de numération et automates, *J. Théor. Nombres Bordeaux* **7** (1995), 473–498.

[44] A. Maes, An automata-theoretic decidability proof for first-order theory of $\langle \mathbb{N}, <, P \rangle$ with morphic predicate P, *Journées Montoises d'Informatique Théorique* (Mons, 1998), *J. Autom. Lang. Comb.* **4** (1999), 229–245.

[45] A. Maes, *Morphic predicates and applications to the decidability of arithmetic theories*, Doctoral dissertation, Université de Mons-Hainaut (1999).

[46] C. Mauduit, Sur l'ensemble normal des substitutions de longueur quelconque, *J. Number Theory* **29** (1988), 235–250.

[47] C. Mauduit, Propriétés arithmétiques des substitutions, *Séminaire de Théorie des Nombres, Paris,* (1989–90), 177–190, *Progr. Math.* **102**, Birkhäuser Boston, Boston, MA, 1992.

[48] C. Michaux, 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** (1996), 251-277.

[49] M. Mignotte, Some effective results about linear recursive sequences, *Automata, languages and programming Fifth Internat. Colloq.* (Udine, 1978), *Lect. Notes in Comput. Sci.* **62** (1978), 322–329.

[50] J.-J. Pansiot, Complexité des facteurs des mots infinis engendrés par morphismes itérés, *Lect. Notes in Comput. Sci.* **172** (1984), 380–389.

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

[52] G. Păun, A. Salomaa, Thin and slender languages, *Discrete Appl. Math.* **61** (1995), 257–270.

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

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

[55] A. Rényi, Representations for real numbers and their ergodic properties, *Acta Math. Acad. Sci. Hungar.* **8** (1957), 477–493.

[56] M. Rigo, Construction of regular languages and recognizability of polynomials, preprint (1999).

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

[58] M. Rigo, Numeration systems on a regular language : Arithmetic operations, recognizability and formal power series, to appear in *Theoret. Comput. Sci.* (2001).

[59] H. E. Rose, *A Course in Number Theory*, second edn., Oxford Science, Oxford, (1996).

[60] A. Salomaa, Formal Languages and Power Series, in *Handbook of theoretical computer science*, Vol. B, J. Van Leeuwen Ed., Elsevier, Amsterdam, (1990), 103–132.

[61] K. Schmidt, On periodic expansions of Pisot numbers and Salem numbers, *Bull. London Math. Soc.* **12** (1980), 269–278.

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

[63] J. Shallit, Numeration systems, Linear recurrences, and Regular sets, *Information and Computation* **113** (1994), 331–347.

[64] A. Szilard, S. Yu, K. Zhang, J. Shallit, Characterizing regular languages with polynomial densities, *Proceedings of the 17th International Symposium on Mathematical Foundations of Computer Science*, *Lect. Notes in Comput. Sci.* **629** (1992), 494–503.

[65] W. Thomas, Automata on infinite objects, in *Handbook of theoretical computer science*, Vol. B, J. Van Leeuwen Ed., Elsevier, Amsterdam, (1990), 133–191.
[66] R. Villemaire, The theory of $\langle \mathbb{N}, +, V_k, V_l \rangle$ is undecidable. *Theoret. Comput. Sci.* **106** (1992), 337–349.
[67] R. Villemaire, Joining k- and l-recognizable sets of natural numbers, *STACS 92* (Cachan, 1992), *Lect. Notes in Comput. Sci.* **577** (1992), 83–94.
[68] S. Wolfram, *The Mathematica book*, third edn., Wolfram Media, Inc., Cambridge University Press, Cambridge, (1996).
[69] S. Yablonski, *Introduction aux mathématiques discrètes*, Ed. Mir, Moscou, (1983).
[70] S. Yu, Regular languages, in *Handbook of formal languages*, Vol. 1, G. Rozenberg and A. Salomaa Eds., Springer, Berlin, (1997), 41–110.

Index