

Université de Liège
Faculté des Sciences
Département de Mathématiques



Positional Numeration Systems: Ultimate Periodicity, Complexity and Automatic Sequences

Adeline MASSUIR

Dissertation présentée en vue
de l'obtention du grade de
Docteur en Sciences
Avril 2021

*À Papy et Mamy,
mes belles étoiles*

Abstract

This dissertation thesis is made up of three distinct parts, connected especially by complexity notion, factorial complexity as well as state complexity. We study positional numeration systems and recognizable sets through decision problems and automatic sequences.

The first part is devoted to the following problem: given a numeration system U and a finite automaton accepting U -representations of a set $X \subseteq \mathbb{N}$, can we decide whether the set X is ultimately periodic (i.e. a finite union of arithmetic progressions)? We prove that this problem is decidable for a large class of numeration systems based on linear recurrent sequences. Thanks to the given automaton, we bound the possible periods of X via an arithmetical study of the linear recurrent sequence, as well as p -adic methods.

The second part is dealing with the set of non-negative integers whose base-2 representation contains an even number of 1, called the Thue-Morse set and denoted by \mathcal{T} . We study of the minimal automaton of the base- 2^p expansions of sets of the form $m\mathcal{T} + r$, where m and p are positive integers and r a remainder between 0 and $m-1$. In particular, we give the state complexity of such sets. The proposed method is constructive and general for any b -recognizable set of integers. As an application, we get a procedure to decide whether a 2^p -recognizable set given via an automaton is a set of the form $m\mathcal{T} + r$.

Finally, in the third part, we study properties of automatic sequences based on Parry and Bertrand numeration systems. We show that Parry-automatic sequences, like Pisot-automatic sequences (and thus in particular like b -automatic sequences) have a sublinear factor complexity. Furthermore, we exhibit a Bertrand-automatic sequence whose factor complexity is quadratic. We also prove that, contrarily to Pisot-automatic sequences, the image of a Parry-automatic sequence under a uniform morphism is not always a Parry-automatic sequence. The same happens for periodic deletion of letters. Last, we give the generalization to multidimensional sequences of a well-known result: a sequence is U -automatic if and only if its U -kernel is

finite, U being such that the numeration language is regular.

Résumé

Cette dissertation est composée de trois parties distinctes, liées par des notions de complexité, que ce soit la complexité factorielle ou la complexité en nombre d'états. Nous étudions les systèmes de numération de position et les ensembles reconnaissables à travers des problèmes de décision ainsi que les suites automatiques.

La première partie est dédiée au problème suivant: étant donné un système de numération U et un automate fini acceptant les écritures en base U d'un ensemble $X \subseteq \mathbb{N}$, peut-on décider si l'ensemble X est ultimement périodique (i.e. une union finie de progressions arithmétiques)? Nous démontrons que ce problème est décidable pour une grande classe de systèmes de numération définis à partir de suites linéaires récurrentes. Grâce à l'automate donné, nous bornons les périodes possibles pour X via une étude arithmétique de la suite linéaire récurrente, ainsi qu'à l'aide de méthodes p -adiques.

La deuxième partie est consacrée à l'ensemble des entiers positifs ou nuls dont l'écriture en base 2 contient un nombre pair de 1, appelé ensemble de Thue-Morse et noté \mathcal{T} . Nous étudions l'automate minimal des écritures en base 2^p des ensembles de la forme $m\mathcal{T} + r$, où m et p sont des entiers positifs et r un reste compris entre 0 et $m-1$. En particulier, nous donnons la complexité en états de tels ensembles. La méthode proposée est constructive et générale pour n'importe quel ensemble b -reconnaisable d'entiers. Comme application, nous obtenons une procédure pour décider si un ensemble 2^p -reconnaisable donné via un automate est un ensemble de la forme $m\mathcal{T} + r$.

Finalement, dans la troisième partie, nous étudions des propriétés des suites automatiques basées sur des systèmes de numération de Parry et de Bertrand. Nous montrons que les suites Parry-automatiques ont, comme les suites Pisot-automatiques (et donc en particulier comme les suites b -automatiques) une complexité factorielle sous-linéaire. D'autre part, nous exhibons une suite Bertrand-automatique dont la complexité factorielle est quadratique. Nous démontrons également que, contrairement aux suites Pisot-automatiques, l'image d'une suite Parry-automatique par un morphisme

uniforme n'est plus nécessairement Parry-automatique. Il en va de même pour la suppression périodique de lettres. Enfin, nous donnons la généralisation aux suites multidimensionnelles d'un résultat bien connu: une suite est U -automatique si et seulement si son U -noyau est fini, U étant tel que le langage de la numération est régulier.

Remerciements

Mes premiers remerciements vont évidemment à Michel RIGO pour son encadrement durant ces cinq dernières années. Malgré un emploi du temps chargé, il a toujours su se montrer disponible et m'a permis de mener à bien ce projet de doctorat.

Émilie CHARLIER m'a également beaucoup apporté durant ces dernières années. Je tiens à la remercier non seulement de faire partie de mon jury, mais surtout pour sa bienveillance et sa sympathie.

Ensuite, je remercie sincèrement Juha HONKALA, Julien LEROY, Victor MARSAULT et Eric ROWLAND de me faire l'honneur de faire partie de mon jury.

Durant cette thèse, j'ai eu l'opportunité de discuter avec de nombreux chercheurs. Je souhaite en particulier remercier Jarkko PELTOMÄKI avec qui j'ai eu le plaisir de collaborer lors de ma première publication, ainsi que Jérôme LEROUX pour son accueil à Bordeaux.

Merci à ma grande sœur de thèse Manon STIPULANTI pour son aide et son soutien précieux. Je tiens également à adresser des remerciements particuliers à Célia CISTERNINO pour nos agréables séances de travail ainsi que nos conversations, mathématiques et autres.

Beaucoup de personnes ont contribué à mon parcours par leur présence et leur soutien. Je ne peux les citer tous ici, mais je tiens à remercier Marine, Luc, Julie, Anne-Lise et Céline d'être toujours là, dans les bons comme dans les mauvais moments.

Ma famille est sans conteste un pilier central dans ma vie. Merci à mes parents pour leur soutien et leurs encouragements perpétuels. Merci à mes frères, sœur et belles-sœurs pour tout. Merci également à Clara, Julien, Elena et Milo, qui par leur seule présence me font oublier les soucis. Je tiens à remercier mes beaux-parents pour leur accueil toujours bienveillant. Et bien entendu, merci à Christophe d'avoir relu ce manuscrit, mais surtout de me supporter au quotidien et d'être un incroyable roc sur lequel me reposer.

Enfin, je tiens à remercier mon grand-père, qui nous a quittés il y a

quelques mois à peine. Merci Papy pour nos rendez-vous du mercredi soir, pour tes blagues et ta bonne humeur constante. Mais surtout merci pour ce modèle de courage et de persévérance que tu m'as donné. J'espère que tu es fier de ce que j'ai accompli aujourd'hui.

Table of contents

Abstract	i
Résumé	iii
Remerciements	v
Table of contents	vii
Introduction	1
1 Basics	7
1.1 Words and languages	7
1.2 Automata	11
1.3 Positional numeration systems	18
1.4 Abstract numeration systems	31
1.5 Rational series	32
1.6 Automatic sequences	35
1.7 Some material about p -adic numbers	38
2 Ultimate periodicity problem for linear numeration systems	45
2.1 Introduction	45
2.2 Our setting	47
2.3 Some useful lemmas	49
2.4 Number of states	56
2.4.1 Factors of the period that are coprime with a_0	57
2.4.2 Prime factors of the period that divide a_0 but do not divide all the coefficients of the recurrence relation	58
2.4.3 Prime factors of the period that divide all the coefficients of the recurrence relation	61
2.5 Cases we can deal with	65
2.5.1 The gcd of the coefficients of the recurrence relation is 1	65

2.5.2	The gcd of the coefficients of the recurrence relation is larger than 1	67
2.6	An incursion into p -adic analysis	71
2.6.1	A third-order sequence	71
2.6.2	A fourth-order sequence	78
2.7	Concluding remarks	79
3	Minimal automaton for multiplying and translating the Thue–Morse set	83
3.1	Introduction	83
3.2	Method	86
3.3	Construction of the intermediate automata	88
3.3.1	The automaton $\mathcal{A}_{\mathcal{T}, 2^p}$	88
3.3.2	The automaton $\mathcal{A}_{m,r,b}$	90
3.3.3	The projected automaton $\Pi(\mathcal{A}_{m,r,b})$	92
3.3.4	The product automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T}, 2^p}$	93
3.3.5	The projection $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T}, 2^p})$ of the product automaton	93
3.4	Properties of the intermediate automata	96
3.4.1	Properties of $\mathcal{A}_{\mathcal{T}, 2^p}$	96
3.4.2	Properties of $\mathcal{A}_{m,r,b}$	96
3.4.3	Properties of $\Pi(\mathcal{A}_{m,r,b})$	97
3.4.4	Properties of $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T}, 2^p}$	100
3.4.5	Properties of $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T}, 2^p})$	101
3.5	Minimization of $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T}, 2^p})$	102
3.5.1	Definition of the classes	102
3.5.2	Looking for the empty classes	106
3.5.3	States of the same class are indistinguishable	108
3.5.4	States of different classes are distinguishable	112
3.5.5	The minimal automaton of $\text{val}_{2^p}^{-1}(m\mathcal{T} + r)$	115
3.6	A decision procedure	116
3.7	A direct description of the classes when $r = 0$	117
3.8	Replacing \mathcal{T} by its complement $\overline{\mathcal{T}}$	120
3.9	Conclusion and perspectives	122
4	Automatic sequences based on Parry or Bertrand numeration systems	125
4.1	Introduction	125
4.2	Factor complexity	128
4.3	Closure properties	132

4.4	Multidimensional sequences	139
4.5	Open problem	143
5	Perspectives	145
	Appendices	149
A	Examples of Parry numeration systems	151
B	Computations for Section 4.3	159
	Bibliography	169
	List of Figures	177

Introduction

Formal language theory and numeration systems are two important sides of discrete mathematics. The present dissertation is about both and the connection between them. On the one hand, formal language theory deals with finite sets, called *alphabets*, whose elements are *letters*. Concatenating letters forms (*finite*) *words*. Then, a *formal language* is simply a set of words over an alphabet. Among languages, one can point out *regular* languages: these are the ones accepted by a *finite automaton*. Roughly speaking, an automaton is an elementary computer, it is the simplest object in the Chomsky hierarchy (see [30]). An example is given in Figure 0.1. We say that this automaton

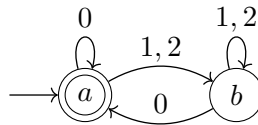


Figure 0.1: A finite automaton.

has two *states*: a and b . Since there is an incoming edge, state a is *initial*: when one feeds the automaton with some word, we start the reading in the state a . In this example, one feeds the automaton with words defined over the alphabet $\{0, 1, 2\}$. Because state a is made of two concentric circles, it is also *final*: if the reading of a word ends in state a , the word is *accepted*, otherwise it is not. Thus, the word 120 is accepted while the word 102 is not:

$$a \xrightarrow{1} b \xrightarrow{2} b \xrightarrow{0} a \quad \text{and} \quad a \xrightarrow{1} b \xrightarrow{0} a \xrightarrow{2} b$$

On the other hand, a *numeration system* is a way to *represent* numbers by words. For instance, in our everyday life, we use the base 10: letters are the digits between 0 and 9 and we form words by placing these letters according powers of 10. For example, the number four hundred fifty six is *represented* as the word 456 (i.e. the concatenation of the three letters 4, 5 and 6): 6 is

the units digit, 5 the tens digit and 4 the hundreds digit. But instead, we could use powers of 3: we represent numbers by concatenating letters among 0, 1 and 2. In the base-3 numeration system, the number four hundred fifty six is represented by the word 121220, because

$$456 = 1 \times 3^5 + 2 \times 3^4 + 1 \times 3^3 + 2 \times 3^2 + 2 \times 3 + 0 \times 1.$$

From representations of numbers and their combinatoric properties, one can deduce properties of the original number. As an example, a number is divisible by 5 if and only if its base-10 representation ends with 0 or 5. A number is divisible by 3 if and only if the sum of the letters of the representation is also a multiple of 3. This criterion is not much complicated. However, if one considers the base-3 representation of the same number, one can directly see if it is a multiple of 3 or not: it suffices that the representation in base 3 ends with 0. In fact, we know that for any $n \in \mathbb{N}$, there is a simple criterion to determine if a number given by its base- b representation is a multiple of n or not. More precisely, for any finite union of sets of the form $m\mathbb{N} + r$, called *ultimately periodic set*, there is a simple criterion for each base b . By “simple criterion”, we mean that there is a finite automaton accepting the base- b representations of the integers of the set. For example, the automaton depicted in Figure 0.1 accepts all the words that are base-3 representations of a non-negative integer which is divisible by 3. Moreover, Cobham’s theorem [31] of 1969 states that ultimately periodic sets are exactly the ones for whom there is a simple criterion for any base b .

Cobham’s theorem gave rise to numerous works about ultimately periodic sets by mathematicians, see the introduction of Chapters 2 and 3. It is also the starting point of research about *positional numeration systems*. See for example [53]. A positional numeration system is defined by an increasing sequence $U = (U_i)_{i \in \mathbb{N}}$ starting with 1 and such that the quotient of two consecutive terms is bounded: $C_U = \sup_{i \in \mathbb{N}} [U_{i+1}/U_i] < \infty$. The *alphabet of the numeration*, i.e. the authorized range for the letters, is given by $\{0, \dots, C_U - 1\}$. A positive integer n is represented by the word $w_{\ell-1} \dots w_0$, which is called the *greedy U -representation* of n and denoted $\text{rep}_U(n)$, if

$$\sum_{i=0}^{\ell-1} w_i U_i = n, \quad w_{\ell-1} \neq 0 \quad \text{and} \quad \forall j \in \{1, \dots, \ell\}, \quad \sum_{i=0}^{j-1} w_i U_i < U_j.$$

A set X of integers is *U -recognizable* if the language $\text{rep}_U(X)$ is regular. In that case, we say that the automaton accepting $\text{rep}_U(X)$ *recognizes* the set X . As an example, the Fibonacci sequence $F = (F_i)_{i \in \mathbb{N}}$ defined by $F_0 = 1, F_1 = 2$ and $F_{i+2} = F_{i+1} + F_i$ for all $i \geq 0$ forms the *Fibonacci numeration system*. In

this system, the greedy representation of a positive integer cannot contain the factor 11. More generally, given a positional numeration system U , the set of all greedy representations, denoted by $\text{rep}_U(\mathbb{N})$, is the *numeration language*. It is often convenient to work with numeration systems whose numeration language is regular. Indeed, this hypothesis ensures that ultimately periodic sets are U -recognizable, see [42, 73]. Moreover, it provides us with a simple criterion to verify whether a representation is greedy or not.

Other numeration systems deserve special interest, such as Pisot numeration systems, Parry numeration systems and Bertrand numeration systems. Precise definitions are given in Section 1.3. All these kinds of numeration systems are extensions from one another: each integer base numeration system is Pisot, each Pisot numeration system is Parry and each Parry numeration system is a Bertrand numeration system. The distinctive characteristic of Bertrand numeration systems is that greediness is preserved when adding or removing arbitrary many ending zeros [11]. For these systems, the numeration language is not necessarily regular, but it is the case for Parry numeration systems [12]. For Pisot numeration systems, we have in addition a characterization of recognizable sets in terms of first order logic [15]. It allows us to prove that multiplication by a constant and addition preserve recognizability for Pisot systems. Note that as the name suggests, a Pisot numeration system is canonically built from a Pisot number, as we will see in Section 1.3.

In this thesis, we present original results obtained in [27] in Chapter 2, results of [23, 24] and of the conference paper [22] in Chapter 3 and then the results of [56] are developed in Chapter 4.

The aim of the first chapter is to recall classical definitions and results. We start with words, languages and automata. Then, we restate well-known material about positional numeration systems. Next, we briefly introduce abstract numeration systems, rational series and automatic sequences. This chapter ends with some background on p -adic numbers.

The second chapter is devoted to the following problem: given a recognizable set of non-negative integers, can we decide whether or not this set is a finite union of arithmetic progressions? First solved by Honkala for integer base systems [43], many authors gave decision procedures following different strategies for positional numeration systems under various hypotheses, as we explain in Section 2.1. In this chapter, we show that the above problem is decidable for a wide class of positional numeration systems. In particular, we give a procedure for systems for which no procedure was known up to now. Of course, we also make some assumptions on the considered numeration systems $U = (U_i)_{i \in \mathbb{N}}$: the numeration language must be regular, there

are arbitrary large gaps between consecutive terms of the sequence U and the gap sequence $(U_{i+1} - U_i)_{i \in \mathbb{N}}$ must be ultimately non-decreasing. Our strategy is similar to Honkala's: given a U -recognizable set of positive integers X , we bound the admissible preperiod and period of X thanks to the automaton accepting $\text{rep}_U(X)$. Then, we have only a finite number of equality tests to do. In order to bound the possible period π_X of X , we proceed step by step. Since the numeration language is regular, the sequence U satisfies a linear recurrence relation. Decomposing π_X in prime factors, we separate primes in three classes: the factors that do not divide the last coefficient of the recurrence, the ones that divide the last coefficient of the recurrence but not all coefficients simultaneously, and the ones that divide all the coefficients of the recurrence relation. We bound these three classes distinctly. With the help of these bounds, we give a decision procedure in the case where the gcd of the coefficients of the recurrence is 1. When there is a prime dividing all the coefficients, our answer relies on a celebrated hard problem of p -adic analysis. We illustrate our method on examples thanks to p -adic techniques.

In the third chapter, we consider integer base numeration systems. Still in the idea of the previous decision problem, our objective is to answer the following question: given a recognizable set of positive integers via a finite automaton, can we decide whether this set is of the form $mX + r$ for some set of integers X , also given via a finite automaton? To solve this problem, we study the minimal automaton accepting $\text{rep}_b(mX + r)$, and in particular the associated state complexity, for any b -recognizable subset X of \mathbb{N} . The case $X = \mathbb{N}$ and $r = 0$ was examined by B. Alexeev in 2004 in [1]. Our aim is to generalize his results. Our work starts with the *Thue-Morse set* \mathcal{T} , the set of integers whose base-2 representation contains an even number of occurrences of the digit 1. The Thue-Morse set is intrinsically linked to the famous Thue-Morse sequence: this infinite word is the characteristic sequence of the Thue-Morse set [2, 5, 64]. Thanks to Cobham's theorem [31], one can prove that this set is only recognizable in bases which are a power of 2. The purpose of this chapter is to give a complete description of the minimal automaton recognizing $m\mathcal{T} + r$ in a base which is a power of 2, for any positive multiple m and remainder r . Our method is constructive. The key idea is to make use of pairs of integers: we first construct an automaton recognizing $\mathcal{T} \times \mathbb{N}$. Then, we build an automaton accepting the language $\text{val}_b^{-1}(\{(n, mn + r) \mid n \in \mathbb{N}\})$ where b is a power of 2. Next, we make the product of the two previous automata, and end our construction by projecting the label of each transition of the last automaton on its second component. In this work, we study every intermediate automaton, before defining the states of the desired minimal automaton: classes of the Myhill-Nerode equivalence

relation. The description of these classes has a nice form in the particular case where $r = 0$. In the end of the chapter, we also study the state complexity of multiplication and dilation of the complementary of the Thue-Morse set. We conclude with a conjecture about the state complexity for a large class of sets.

The fourth chapter deals with automatic sequences and can thus be linked to Chapter 2 and positional numeration systems. A U -automatic sequence is an infinite word for which there exists a finite automaton with output (DFAO for short) such that the n^{th} letter of the word is the output of the DFAO when feeding it with the U -representation of n for a numeration system U [6]. Recall that there is a hierarchy of numeration systems:

Integer base systems \subsetneq Pisot systems (with convenient initial conditions)
 \subsetneq Parry systems
 \subsetneq Bertrand systems with a regular numeration language.

We study properties of automatic sequences and their limitations according to the type of numeration system in consideration. First, we look at factor complexity: we show that any Parry-automatic sequence has a sublinear factor complexity, and that this property cannot be extended to Bertrand-automatic sequences: we give such a sequence with superlinear complexity. Note that this property was already known for the integer base numeration system since 1972 thanks to Cobham [32]. Then, we look at closure properties. We present a Parry numeration system U and U -automatic sequences that are not closed under taking image by a uniform substitution, or by periodic deletion. In particular, it gives another proof of the non-existence of a first-order logical characterization for Parry-automatic sequences (see [38]). We conclude this chapter with the generalisation to multidimensional sequences of a well-known result: a sequence is U -automatic if and only if its U -kernel is finite [67].

The fifth chapter gives some open problems.

Our research work was often driven by developing many examples. We think that this experimental material could be useful for other researcher, so we added a first appendix at the end of this dissertation presenting several Parry numeration systems and properties presented in this thesis, such as the dominant root β associated with the system, the β -expansion of 1, or the associated automaton.

Finally, in the second appendix we present the `Mathematica` code used to compute approximations necessary for proofs of several results in Chapter 4.

Chapter 1

Basics

This first chapter is devoted to outline the necessary background for a clear understanding of this text. Most of our notation and results are standard, but we collect them here for the ease of reference. At the beginning of every section, we refer the interested reader to relevant books or chapters of books.

We start with basic definitions of formal language theory and usual properties and definitions of automata theory. Next, we introduce positional numeration systems, and focus particularly on integer bases, as well as Parry, Bertrand and Pisot numeration systems. We also give the definition of an abstract numeration system. Section 1.5 is devoted to introduce formal power series. Then, we briefly discuss automatic sequences and conclude with a small introduction to p -adic numbers.

Throughout this text, we denote by \mathbb{N} the set of non-negative integers $\{0, 1, 2, \dots\}$. We also refer to the set of integers (resp. rational numbers, real numbers) as \mathbb{Z} (resp. \mathbb{Q} , \mathbb{R}). Moreover, we set $\mathbb{N}_0 = \mathbb{N} \setminus \{0\}$ and $\mathbb{Z}_0 = \mathbb{Z} \setminus \{0\}$. We also set $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$. Furthermore, if i and j are two integers such that $i \leq j$, we denote by $\llbracket i, j \rrbracket$ the set $\{i, i+1, \dots, j-1, j\}$. We will also make use of the notation O and Θ : $f(n)$ is in $O(g(n))$ if there is some $k > 0$ such that for n large enough, $|f(n)| \leq k|g(n)|$. Moreover, $f(n)$ is in $\Theta(g(n))$ if there are $k, \ell > 0$ such that $k|g(n)| \leq |f(n)| \leq \ell|g(n)|$ for n large enough.

1.1 Words and languages

We briefly introduce the basic terminology on words. The interested reader can find more details in [52].

Definition 1.1.1. An *alphabet* is a non-empty finite set, whose elements are called *letters*. A *word* over an alphabet Σ is a finite or infinite sequence of letters in Σ . We denote by ε the empty sequence and call it the *empty word*. Given a finite word w , its *length*, denoted by $|w|$, is the number of letters composing w . If w is a finite word over the alphabet Σ and $a \in \Sigma$, we let $|w|_a$ be the number of occurrences of a in w . For a positive integer n and an alphabet Σ , we write Σ^n for the set of all words over Σ of length n . If $w = w_0 \cdots w_{|w|-1}$ where w_i are letters, then we let $w^R = w_{|w|-1} \cdots w_0$ denote the *reversal* or *mirror* of w . The set of finite (resp. infinite) words over an alphabet Σ is denoted by Σ^* (resp. Σ^ω). Note that for a unary alphabet $\{a\}$, we simply write a^* instead of $\{a\}^*$. We set $\Sigma^+ = \Sigma^* \setminus \{\varepsilon\}$. A *language* over an alphabet Σ is a subset of Σ^* .

Remark 1.1.2. In the following, *word* with no specification stands for *finite word*. We also refer to infinite words as *sequences*. Moreover, according to the context, we will index letters of words from left to right, like in the previous definition, or from right to left, as it will be the case in Definition 1.3.1 for example.

Example 1.1.3. Let $\Sigma = \{a, m, o, t, u\}$ be the alphabet composed of the five letters a, m, o, t and u . The finite word $w = \text{automata}$ is of length $|w| = 8$, and we have $|w|_a = 3$ and $|w|_o = 1$, since we can find three times the letter a in w and only one o . The mirror of w is $w^R = \text{atamotua}$.

Definition 1.1.4. If $u = u_0 \cdots u_{m-1}$ and $v = v_0 \cdots v_{n-1}$ are two finite words over an alphabet Σ , the *concatenation of u and v* is the finite word w of length $|u| + |v|$ defined by $w = w_0 \cdots w_{m+n-1}$ where

$$w_i = \begin{cases} u_i & \text{if } i \in \llbracket 0, m-1 \rrbracket \\ v_{i-m} & \text{if } i \in \llbracket m, m+n-1 \rrbracket. \end{cases}$$

We denote the concatenation of u and v by $u \cdot v$, or simply uv when the context is clear. In a similar way, we can define the concatenation of a finite word with an infinite word. For a non-negative integer n and a finite word w over an alphabet Σ , the concatenation of n copies of w is denoted by w^n and defined by induction: $w^0 = \varepsilon$ and $w^{n+1} = w^n w$ for all $n \in \mathbb{N}$. Similarly, let w^ω be the concatenation of infinitely many copies of w . An infinite word $\mathbf{w} \in \Sigma^\omega$ is *ultimately periodic* if there are two finite words $u, v \in \Sigma^*$ such that $\mathbf{w} = uv^\omega$. If u and v are taken minimal, the *preperiod* of \mathbf{w} is the integer $|u|$ and the *period* of \mathbf{w} is the integer $|v|$. In the special case where $u = \varepsilon$, \mathbf{w} is said *purely periodic* or simply *periodic*. When $\mathbf{w} \in \Sigma^\omega$ is not ultimately periodic, it is said to be *aperiodic*.

Example 1.1.5. Over the binary alphabet $\Sigma = \{a, b\}$, the concatenation of the words aba and bab gives the word $ababab = (ab)^3$.

Definition 1.1.6. Let w be a word (finite or infinite) over an alphabet Σ . A *factor* or *subword* of w is a finite word u such that there exist two words $x \in \Sigma^*$ and $y \in \Sigma^* \cup \Sigma^\omega$ verifying $w = xuy$. A *prefix* (resp. *suffix*) of w is a word u (resp. v) such that there exists $v \in \Sigma^* \cup \Sigma^\omega$ (resp. $u \in \Sigma^*$) satisfying $w = uv$.

Example 1.1.7. In the finite word *automata*, *auto* is a prefix, *tom* is a factor and *omata* a suffix.

An important concept in Chapter 4 is the one of *factor complexity*. The *factor complexity function* of an infinite word \mathbf{x} , denoted $p_{\mathbf{x}}$, is a function from \mathbb{N} to \mathbb{N} that associates to each $n \in \mathbb{N}$ the number of factors of length n occurring in \mathbf{x} . More background will be given in Chapter 4, Section 4.2.

When one has a total order on an alphabet Σ , one can extend this order to Σ^* or $\Sigma^* \cup \Sigma^\omega$. In the following, we need two special orders: lexicographical and genealogical orders.

Definition 1.1.8. Let $(\Sigma, <)$ be a totally ordered alphabet. The order $<$ on Σ extends to an order on Σ^ω , called the *lexicographical order*, as follows. If \mathbf{u}, \mathbf{v} are two infinite words over Σ , then \mathbf{u} is said to be *lexicographically less* than \mathbf{v} , which is denoted $\mathbf{u} <_{\text{lex}} \mathbf{v}$, if there are $x \in \Sigma^*$, $\mathbf{y}, \mathbf{z} \in \Sigma^\omega$ and $a, b \in \Sigma$ such that we have $\mathbf{u} = xay$, $\mathbf{v} = xbz$ and $a < b$. This order extends to $\Sigma^* \cup \Sigma^\omega$ by replacing finite words t over Σ by $t\Diamond \in (\Sigma \cup \{\Diamond\})^\omega$, where \Diamond is a letter not belonging to Σ and is assumed to verify $\Diamond < a$ for all $a \in \Sigma$. We write $u \leq_{\text{lex}} v$ for two words u and v satisfying either $u <_{\text{lex}} v$ or $u = v$.

Considering alphabetical order, the lexicographical order is the one used in dictionaries.

Definition 1.1.9. Let $(\Sigma, <)$ be a totally ordered alphabet. The order $<$ on Σ extends to an order on Σ^* , called the *genealogical order*, as follows. If u and v are two finite words over Σ , then u is said to be *genealogically less* than v , and we write $u <_{\text{gen}} v$, if they satisfy either $|u| < |v|$, or $|u| = |v|$ and $u <_{\text{lex}} v$. We write $u \leq_{\text{gen}} v$ for two words u and v satisfying either $u <_{\text{gen}} v$ or $u = v$.

Example 1.1.10. In the lexicographical order, *past* comes after *future*, but in the genealogical order, *future* is after *past*. Considering a binary alphabet $\Sigma = \{a, b\}$ totally ordered by $a < b$, we have $aaba <_{\text{lex}} abb <_{\text{lex}} abba$, but $abb <_{\text{gen}} aaba <_{\text{gen}} abba$.

In Chapter 2, we will make take advantage of a strategy useful for definite languages. We restate here the definition.

Definition 1.1.11. Let $n \in \mathbb{N}$. A language L over an alphabet Σ is *weakly n -definite* if for any $x, y \in \Sigma^*$ satisfying $|x| \geq n$, $|y| \geq n$ and having the same suffix of length n , $x \in L$ if and only if $y \in L$.

Let $n \geq 1$. A language L over an alphabet Σ is *n -definite* if it is weakly n -definite and not weakly $(n-1)$ -definite.

Let us conclude this section by the notion of morphism.

Definition 1.1.12. Let Σ and Δ be two alphabets. A *morphism* (or *substitution*) is a map $\mu: \Sigma^* \rightarrow \Delta^*$ satisfying $\mu(uv) = \mu(u)\mu(v)$ for all $u, v \in \Sigma^*$. In particular, $\mu(\varepsilon) = \varepsilon$, and μ is completely determined by the image of the letters of Σ . Let $k \in \mathbb{N}_0$. When $|\mu(a)| = k$ for all $a \in \Sigma$, we say that μ is *k -uniform*. In particular, a 1-uniform morphism is called a *coding*. If for some $a \in \Sigma$, $\mu(a) = \varepsilon$, then μ is said to be *erasing*, otherwise it is said *non-erasing*.

In order to extend these definitions to infinite words, we need to describe a distance on words, see [12, Chapter 1].

Definition 1.1.13. Let Σ be an alphabet and let \mathbf{x}, \mathbf{y} be two infinite words over Σ . Let $\mathbf{x} \wedge \mathbf{y}$ denote the longest common prefix of \mathbf{x} and \mathbf{y} . Then the *distance* d between \mathbf{x} and \mathbf{y} is defined by

$$d(\mathbf{x}, \mathbf{y}) = \begin{cases} 0, & \text{if } \mathbf{x} = \mathbf{y}, \\ 2^{-|\mathbf{x} \wedge \mathbf{y}|} & \text{otherwise.} \end{cases}$$

Let Σ be an alphabet. Since Σ^ω is a (complete) metric space, it makes sense to speak of convergent sequences of infinite words. The sequence $(\mathbf{x}_i)_{i \in \mathbb{N}}$ of infinite words over Σ converges to $\mathbf{x} \in \Sigma^\omega$ if for all $\epsilon > 0$, there exists $I \in \mathbb{N}$ such that for all $i \geq I$, $d(\mathbf{x}_i, \mathbf{x}) < \epsilon$. Let $\diamond \notin \Sigma$. We say that the sequence $(y_i)_{i \in \mathbb{N}}$ of finite words over Σ converges to an infinite word $\mathbf{y} \in \Sigma^\omega$ if the sequence of infinite words $(y_i \diamond^\omega)_{i \in \mathbb{N}}$ converges to \mathbf{y} . Intuitively, the sequence $(y_i)_{i \in \mathbb{N}}$ of finite words over Σ converges to an infinite word $\mathbf{y} \in \Sigma^\omega$ if every prefix of \mathbf{y} is a prefix of all but a finite number of the words y_i .

Now, we are able to extend morphisms to infinite words. If $\mu: \Sigma^* \rightarrow \Delta^*$ is a non-erasing morphism, it can be extended to a map from Σ^ω to Δ^ω as follows: if $\mathbf{x} = x_0x_1\cdots$ is an infinite word over Σ , then the sequence of words $(\mu(x_0 \cdots x_i))_{i \in \mathbb{N}}$ is convergent towards an infinite word over Δ . Its limit is denoted $\mu(\mathbf{x}) = \mu(x_0)\mu(x_1)\cdots$. In the following, morphisms will always be defined on letters, and we consider implicitly their extension to infinite words.

Definition 1.1.14. Let $\mu: \Sigma^* \rightarrow \Sigma^*$ be a morphism. A finite or infinite word x is a *fixed point* of μ if $\mu(x) = x$.

Example 1.1.15. Consider the 2-uniform morphism

$$\theta: \{0, 1\}^* \rightarrow \{0, 1\}^*: 0 \mapsto 01, 1 \mapsto 10.$$

The fixed point of θ starting with 0 is the infinite word

$$\mathbf{t} = 0110100110010110 \dots$$

and is called the *Thue-Morse word*. This word is well-known to be aperiodic (see for example [53, p.113]).

Definition 1.1.16. Let $\mu: \Sigma^* \rightarrow \Sigma^*$ be a morphism and let $a \in \Sigma$. Then μ is *prolongeable* on a if $\lim_{n \rightarrow +\infty} |\mu^n(a)| = +\infty$ and if there is a non-empty word $u \in \Sigma^*$ such that $\mu(a) = au$. In this case, for all $n \in \mathbb{N}$, $\mu^n(a)$ is a prefix of $\mu^{n+1}(a)$ and since $|\mu^n(a)|$ tends to infinity when n tends to infinity, the sequence $(\mu^n(a))_{n \in \mathbb{N}}$ converges to an infinite word, denoted $\mu^\omega(a)$, given by

$$\mu^\omega(a) = \lim_{n \rightarrow +\infty} \mu^n(a) = au\mu(u)\mu^2(u) \dots$$

This infinite word is a fixed point of μ . A *purely morphic* word is an infinite word obtained by iterating a prolongeable morphism. If $\mathbf{x} \in \Sigma^\omega$ is purely morphic and if $\tau: \Sigma \rightarrow \Delta$ is a coding, then the word $\mathbf{y} = \tau(\mathbf{x})$ is said to be *morphic*.

Two famous decision problems are related to (purely) morphic words.

Problem 1.1.17 (HD0L periodicity problem). Given a morphism μ and a coding τ such that μ is prolongeable on a letter a , decide whether or not the the infinite word $\tau(\mu^\omega(a))$ is ultimately periodic.

The *D0L periodicity problem* is a restricted case of the HD0L periodicity problem, since we consider only the morphism μ , without the coding τ . The D0L periodicity problem was shown to be decidable in [41] and [62]. We will get back to the HD0L periodicity problem in Chapter 2, Section 2.7 and in Chapter 3, Section 3.1.

1.2 Automata

An automaton is in some way a very simple machine processing information. Here, we recall fundamental definitions and properties needed in this work. For more on automata theory, see [70].

Definition 1.2.1. A *deterministic automaton* is a 5-tuple

$$\mathcal{A} = (Q, q_0, F, \Sigma, \delta)$$

where

- Q is a non-empty set, called the set of *states*,
- q_0 is a special element of Q , called the *initial state*,
- $F \subset Q$ is the set of *final states*,
- Σ is an alphabet,
- $\delta: Q \times \Sigma \rightarrow Q$ is the *transition function*.

Note that δ can be partial. When the transition function is total, we say that the automaton is *complete*. The transition function δ naturally extends to a possibly partial function on $Q \times \Sigma^*$ by setting $\delta(q, \varepsilon) = q$ for all states q and $\delta(q, aw) = \delta(\delta(q, a), w)$, for all $q \in Q, a \in \Sigma, w \in \Sigma^*$. A deterministic automaton is *finite* (resp. *infinite*) if its set of states is finite (resp. infinite). In the present dissertation, we will use the abbreviation DFA for “deterministic finite automaton”. If $p, q \in Q$ and $w = w_0 \cdots w_n \in \Sigma^*$, $w_0, \dots, w_n \in \Sigma$, are such that $\delta(p, w) = q$, then we say that the execution

$$p \xrightarrow{w_0} p_1 \xrightarrow{w_1} \cdots \xrightarrow{w_{n-1}} p_n \xrightarrow{w_n} q$$

is a *path* from p to q of label w .

DFA's can be represented by oriented graphs. The states appear as nodes, and if $p, q \in Q$ and $a \in \Sigma$ are such that $\delta(p, a) = q$, then there is an edge from p to q labelled by a . The initial state is symbolised by an incoming arrow, and final states are designated by two concentric circles around the node.

Example 1.2.2. Consider the DFA $\mathcal{A} = (\{1, 2, 3, 4\}, 1, \{2\}, \{a, b\}, \delta)$ where the transition function δ is given by the following table:

	1	2	3	4
a	4	2	3	4
b	2	3	2	4

Since δ is a total function, \mathcal{A} is a complete DFA. In Figure 4.2, the graph representation of \mathcal{A} is depicted.

Definition 1.2.3. An automaton is said *irreducible* if the associated graph is strongly connected. It is *primitive* if there exists an integer N such that for any two states q, q' , there is a word w of length N (depending on q, q') such that $\delta(q, w) = q'$ (where δ is the transition function of the automaton).

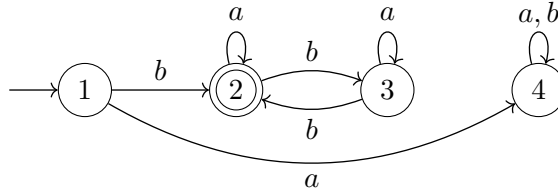


Figure 1.1: A deterministic finite automaton.

Proposition 1.2.4. *An automaton is primitive if and only if it is irreducible and aperiodic (the gcd of lengths of the cycles going through any state is 1).*

Since a DFA \mathcal{A} can be depicted by an oriented graph, one can associate with it an adjacency matrix, denoted by $Adj(\mathcal{A})$. The entry $(Adj(\mathcal{A}))_{i,j}^n$ counts the number of distinct paths of length n from state i to state j (see [50, Chapter 2]). The matrix $Adj(\mathcal{A})$ being obviously a square matrix, one can compute its characteristic polynomial and its zeros: the *eigenvalues* of $Adj(\mathcal{A})$ (and by extension, of \mathcal{A}). The following statement is a piece of Perron's theorem (see for example [6, Section 8.3]).

Proposition 1.2.5. *If \mathcal{A} is a primitive automaton, then it admits a strictly positive eigenvalue λ whose module is strictly greater than the module of any other eigenvalue of \mathcal{A} . The real number λ is called the Perron eigenvalue of \mathcal{A} . Moreover, one has, for all i, j , $(Adj(\mathcal{A}))_{i,j}^n = \Theta(\lambda^n)$.*

Let us now get back to basic properties of automata.

Definition 1.2.6. Let $\mathcal{A} = (Q, q_0, F, \Sigma, \delta)$ be a deterministic finite automaton. A finite word w over Σ is *accepted by \mathcal{A}* if $\delta(q_0, w) \in F$. The set of words accepted by \mathcal{A} is the *language accepted by \mathcal{A}* . It is denoted by $L(\mathcal{A})$. The *language accepted from the state $q \in Q$* , denoted by $L_q(\mathcal{A})$ or simply L_q when the context is clear, is the set of words accepted by the automaton $(Q, q, F, \Sigma, \delta)$. In particular, $L(\mathcal{A}) = L_{q_0}$.

Example 1.2.7. The word $babb$ is accepted by the automaton \mathcal{A} of Example 1.2.2, but the word bb is not. Indeed, we can easily check that we have $\delta(1, babb) = \delta(2, abb) = \delta(2, bb) = \delta(3, b) = 2 \in F$. Furthermore, $\delta(1, bb) = \delta(2, b) = 3 \notin F$. The language accepted by \mathcal{A} is the set of words over $\{a, b\}$ starting with b and containing an odd number of b . The language accepted from state 2 is the set of words over $\{a, b\}$ containing an even number of b . The language accepted from state 4 is the empty set.

A lot of properties are interesting about automata. Let's point out some of them that will be necessary in Chapter 3.

Definition 1.2.8. Let $\mathcal{A} = (Q, q_0, F, \Sigma, \delta)$ be a deterministic finite automaton. A state $q \in Q$ is *accessible* if there is $w \in \Sigma^*$ such that $\delta(q_0, w) = q$. Otherwise stated, q is accessible if q can be reached from the initial state. A state $q \in Q$ is *coaccessible* if one can reach a final state from it, i.e. if there is a word $w \in \Sigma^*$ such that $\delta(q, w) \in F$. The automaton \mathcal{A} is said *accessible* (resp. *coaccessible*) if all its states are accessible (resp. coaccessible). An automaton is *trim* if it is both accessible and coaccessible. Two states p and q are *distinguishable* (resp. *indistinguishable*) if $L_p \neq L_q$ (resp. if $L_p = L_q$). The automaton \mathcal{A} is *reduced* if any two distinct states are distinguishable. We say that \mathcal{A} has *disjoint states* if the languages accepted from different states are disjoint: for distinct states p and q , we have $L_p \cap L_q = \emptyset$.

Remark 1.2.9. Note that in particular, any coaccessible DFA having disjoint states is reduced. Moreover, in a reduced DFA, there can be at most one non-coaccessible state.

Example 1.2.10. The automaton depicted in Figure 1.2 is a trim deterministic automaton, but it is not complete, since the transition function is not defined for the pair $(1, a)$. The automaton of Figure 4.2 is not coaccessible, since one can't reach a final state from state 4, but it is accessible. Both automata are reduced. Note that these two automata accept the same language.

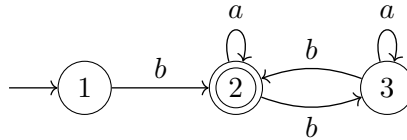


Figure 1.2: A trim deterministic finite automaton.

Another useful concept is the notion of non-deterministic automaton.

Definition 1.2.11. A *non-deterministic automaton* is a 5-tuple

$$\mathcal{A} = (Q, I, F, \Sigma, \Delta)$$

where

- Q, F and Σ are defined as for a deterministic automaton,

- $I \subseteq Q$ is a non-empty set, called the *set of initial states*,
- $\Delta \subseteq Q \times \Sigma^* \times Q$ is a non-empty set, called the *transition relation*.

A non-deterministic automaton is *finite* (resp. *infinite*) if its set of states is finite (resp. infinite). A word w is *accepted* by \mathcal{A} if there is a *path* from an initial state to a final state labelled by w . Otherwise stated, w is accepted if there are a positive integer n , finite words w_1, w_2, \dots, w_n and states $q_0, q_1, \dots, q_n \in Q$ with $q_0 \in I$ and $q_n \in F$ such that $w = w_1 \cdot w_n$ and

$$(q_0, w_1, q_1), (q_1, w_2, q_2), \dots, (q_{n-1}, w_n, q_n)$$

belong to Δ . The *language accepted by \mathcal{A}* , denoted $L(\mathcal{A})$, is the set of words accepted by \mathcal{A} . We write NFA for “non-deterministic finite automaton”.

A transition graph can be depicted for NFAs in the same way than for DFAs: states are represented by nodes, and for any $p, q \in Q$ and $w \in \Sigma^*$ such that $(p, w, q) \in \Delta$, there is an edge from p to q labelled by w . Note that there can be several initial states, thus several incoming arrows.

Since a DFA is in particular a NFA, one could think that there are more languages accepted by NFAs than by DFAs, but the following proposition prove it is wrong.

Proposition 1.2.12. *A language is accepted by a NFA if and only if it is accepted by a DFA.*

Definition 1.2.13. A language is *regular* if it is accepted by a finite automaton.

Regular languages benefit from closure properties.

Proposition 1.2.14. *The class of regular languages is stable for concatenation, union, intersection, complementation, reversal and image by morphism.*

The next result is useful to reject the regularity of a language.

Proposition 1.2.15 (Pumping lemma). *If L is a regular language over an alphabet Σ , then there exists a positive integer k such that any word w in L of length $|w| \geq k$ can be decomposed as $w = xyz$, where x, y, z are finite words over Σ satisfying $y \neq \varepsilon$, $|xy| \leq k$ and $xy^n z \in L$ for all $n \geq 0$.*

If fact, if \mathcal{A} is a DFA accepting L , then in the previous proposition, one can take k to be the number of states of \mathcal{A} .

As seen in Example 1.2.10, different deterministic automata may accept the same language. Among them, one can point out the minimal automaton of this language.

Definition 1.2.16. Let Σ be an alphabet and $L \subseteq \Sigma^*$. The *Myhill-Nerode equivalence relation*, denoted by \sim_L , is the relation on Σ^* defined by

$$u \sim_L v \Leftrightarrow (\forall w \in \Sigma^*, uw \in L \Leftrightarrow vw \in L).$$

If u is a finite word over Σ , then we set $u^{-1}L = \{w \in \Sigma^* : uw \in L\}$. In other words, $u^{-1}L$ is the set of finite words over Σ which, when concatenated with u , form a word belonging to L .

Remark 1.2.17. Note that if L is a language over Σ and $u, v \in \Sigma^*$, then we have $u \sim_L v \Leftrightarrow u^{-1}L = v^{-1}L$.

Lemma 1.2.18. Let Σ be an alphabet, $L \subseteq \Sigma^*$ and $u, v \in \Sigma^*$. We have

$$(uv)^{-1}L = v^{-1}(u^{-1}L).$$

Definition 1.2.19. Let Σ be an alphabet. The *minimal automaton* of a language $L \subseteq \Sigma^*$ is the deterministic automaton

$$\mathcal{A}_L = (Q_L, q_{0,L}, F_L, \Sigma, \delta_L)$$

with

- $Q_L = \{u^{-1}L : u \in \Sigma^*\}$,
- $q_{0,L} = \varepsilon^{-1}L = L$,
- $F_L = \{u^{-1}L : u \in L\}$,
- $\forall q \in Q_L, \forall a \in \Sigma, \delta_L(q, a) = a^{-1}q$.

Note that thanks to Lemma 1.2.18, the transition function can be extended to $Q_L \times \Sigma^*$ by $\delta_L(q, w) = w^{-1}q, \forall q \in Q_L, w \in \Sigma^*$. The *trim minimal automaton* of a language L is the minimal automaton of L from which the only possible non-coaccessible state (called *sink state*) is removed. Notice that \mathcal{A}_L is not necessarily finite.

Proposition 1.2.20. The minimal automaton of a language accepts this language.

In Chapter 3, the following characterization of minimal automata will be central.

Proposition 1.2.21. Let \mathcal{A} be a complete deterministic automaton. Then \mathcal{A} is minimal if and only if it is accessible and reduced.

The next proposition justifies the term *minimal*: the minimal automaton of a language is the one containing the least number of states among those accepting the language.

Proposition 1.2.22. *Let L be a language over an alphabet Σ and let \mathcal{A} be a deterministic automaton accepting L and whose set of states is Q . Then $\text{Card } Q_L \leq \text{Card } Q$.*

The following theorem is a characterization of regular languages.

Theorem 1.2.23 (Myhill and Nerode). *A language L is regular if and only if the Myhill-Nerode equivalence relation \sim_L is of finite index. Otherwise stated, a language L is regular if and only if its minimal automaton \mathcal{A}_L is finite.*

Thanks to the previous proposition and theorem, one can define the state complexity of a regular language.

Definition 1.2.24. The *state complexity* of a regular language L is the number of states of its minimal automaton \mathcal{A}_L .

Let us now introduce the notion of deterministic finite automaton with output, useful to define automatic sequences in Section 1.6.

Definition 1.2.25. A *deterministic finite automaton with output* (or DFAO for short) is a 6-tuple

$$\mathcal{A} = (Q, q_0, \Sigma, \delta, \Gamma, \tau)$$

where

- Q, q_0, Σ and δ are defined as in a DFA,
- Γ is the *output alphabet*,
- $\tau: Q \rightarrow \Gamma$ is the *output function*.

The *output* corresponding to the *input* $w \in \Sigma^*$ is $\tau(\delta(q_0, w))$.

One can also represent a transition graph for a DFAO: it is the same as for a DFA, except that there is no final state, thus no concentric circles. Instead, for each state there is an additional outgoing arrow, labelled by the output corresponding to the state it comes from.

We conclude this section by the definition of a (finite) transducer. It is in some way similar to a DFAO. Note that we define here a particular model of finite transducer, but there is a wide literature about it. For example, “our” transducers are called *sequential transducers* in [12].

Definition 1.2.26. A finite *transducer* \mathcal{T} is a 6-tuple

$$\mathcal{T} = (Q, q_0, \Sigma, \delta, \Gamma, \tau)$$

where

- Q, q_0, Σ and δ are defined as in a DFA,
- Γ is the *output alphabet*,
- $\tau: Q \times \Sigma \rightarrow \Gamma^*$ is the *output function*.

A transducer can be viewed as a way to define functions: to every *input word* $w = w_1 \cdots w_n \in \Sigma^*$, $w_i \in \Sigma$ for all $i \in \llbracket 1, n \rrbracket$, the transducer \mathcal{T} associates an *output word* $\mathcal{T}(w) \in \Gamma^*$ defined by

$$\tau(q_0, w_1) \tau(\delta(q_0, w_1), w_2) \tau(\delta(q_0, w_1 w_2), w_3) \cdots \tau(\delta(q_0, w_1 \cdots w_{n-1}), w_n).$$

Proposition 1.2.27. *The image of a regular language by a finite transducer is a regular language.*

1.3 Positional numeration systems

In this section, we present one of the main concepts used in the present dissertation: numeration systems. Briefly, a numeration system is a way to represent numbers with digits, or letters in \mathbb{N} . We introduce linear, Parry, Bertrand, Pisot numeration systems, the integer base and related properties. We refer the interested reader to [53, Chapter 7] and [12, Chapter 2].

Let us first consider the representation of integers.

Definition 1.3.1. A *positional numeration system* (or simply a *numeration system*) is an increasing sequence $U = (U_i)_{i \in \mathbb{N}}$ of integers such that $U_0 = 1$ and $C_U = \sup_{i \geq 0} \lceil U_{i+1}/U_i \rceil$ is finite. We let Σ_U be the integer alphabet $\llbracket 0, C_U - 1 \rrbracket$ and call it the *alphabet of the numeration*. In this particular alphabet, the letters are often called *digits*. The *greedy U -representation* (or *greedy U -expansion*) of the positive integer n is the finite word $w_{\ell-1} \cdots w_0$ over Σ_U , denoted $\text{rep}_U(n)$, satisfying

$$\sum_{i=0}^{\ell-1} w_i U_i = n, \quad w_{\ell-1} \neq 0 \quad \text{and} \quad \forall j \in \llbracket 1, \ell \rrbracket, \sum_{i=0}^{j-1} w_i U_i < U_j.$$

We set $\text{rep}_U(0)$ to be the empty word ε . The language $\text{rep}_U(\mathbb{N})$ is called the *numeration language*. The *U -numerical valuation* $\text{val}_U: \mathbb{Z}^* \rightarrow \mathbb{N}$ maps

a word $w_{\ell-1} \cdots w_0$ over any alphabet of integers to the number $\sum_{i=0}^{\ell-1} w_i U_i$. If $\text{val}_U(w) = n$, we say that w is a U -representation of n . Note that this representation is not necessarily greedy. Clearly, the function $\text{val}_U \circ \text{rep}_U$ is the identity from \mathbb{N} to \mathbb{N} . When the context is clear, the letter U will be omitted: we will talk of *greedy representation*, *greedy expansion*, *numerical valuation* and *representation*.

Remark 1.3.2. Note that we write greedy U -representations with most significant digit first (MSDF convention): the leftmost digit is associated with the largest U_i occurring in the decomposition.

The following example is a classical one and will be central in Chapter 3. It contains our daily numeration system: the base 10.

Example 1.3.3 (Integer base). Let $b \geq 2$ be an integer. The *integer base- b numeration system* is the positional numeration system built on the sequence

$$U_b = (b^i)_{i \in \mathbb{N}}.$$

The alphabet is in this case $\Sigma_b = \Sigma_{U_b} = \llbracket 0, b-1 \rrbracket$ and the numeration language is

$$\text{rep}_{U_b}(\mathbb{N}) = (\Sigma_b \setminus \{0\})\Sigma_b^* \cup \{\varepsilon\}.$$

For this numeration system, we set $\text{rep}_b = \text{rep}_{U_b}$. We also set $\text{val}_b = \text{val}_{U_b}$ and we call greedy U_b -expansions as *base- b expansions*.

Another classical example is the following. It is based on the Fibonacci sequence.

Example 1.3.4 (Fibonacci). Let $F = (F_i)_{i \in \mathbb{N}} = (1, 2, 3, 5, 8, \dots)$ be the sequence of Fibonacci numbers defined by

$$F_0 = 1, F_1 = 2 \text{ and } F_{i+2} = F_{i+1} + F_i \forall i \in \mathbb{N}.$$

We have $\Sigma_F = \{0, 1\}$ and the Fibonacci numeration language $\text{rep}_F(\mathbb{N})$ is $1\{0, 01\}^* \cup \{\varepsilon\}$, thus greedy representations are words over $\{0, 1\}$ that do not contain the factor 11. For instance, we have $\text{rep}_F(11) = (10100)$ and $\text{val}_F(1001) = 5 + 1 = 6 = \text{val}_F(111)$. Remark that 111 is not greedy, thus it does not belong to the numeration language.

Before looking at real numbers, let us just notice that the genealogical order (Definition 1.1.9) coincides with the classical order in \mathbb{N} .

Proposition 1.3.5. *Let U be a numeration system. For all $m, n \in \mathbb{N}$, we have*

$$m < n \Leftrightarrow \text{rep}_U(m) <_{\text{gen}} \text{rep}_U(n).$$

Note that in the previous statement, the genealogical order applies on greedy representations and the classical order on their values in \mathbb{N} . Given a numeration system U , the result is not true for all words: we do not have that

$$u <_{\text{gen}} v \Leftrightarrow \text{val}_U(u) < \text{val}_U(v)$$

for all words $u, v \in \Sigma_U^*$. Consider the Fibonacci numeration system of Example 1.3.4, $u = 11$ and $v = 100$. Then $u <_{\text{gen}} v$ since $|u| = 2 < 3 = |100|$, but $\text{val}_F(11) = 3 = \text{val}_F(100)$.

There is a link between the representation of integers and the representation of real numbers.

Definition 1.3.6. Let $\beta > 1$ be a real number. The β -*expansion* (or β -*representation*) of a real number $x \in [0, 1]$ is the sequence of non-negative integers $d_\beta(x) = (x_i)_{i \geq 1}$ that satisfies

$$x = \sum_{i=1}^{+\infty} x_i \beta^{-i}$$

and which is the maximal element in \mathbb{N}^ω having this property with respect to the lexicographic order over \mathbb{N} . Notice that β -expansions can be obtained by a greedy algorithm and they only contain letters (or digits) over the alphabet $\Sigma_\beta = \llbracket 0, \lceil \beta \rceil - 1 \rrbracket$. Also note that if a representation ends with infinitely many zeros, then it is sometimes convenient to omit those zeros and the representation is said to be *finite*. By $\text{Fact}(D_\beta)$, we denote the set of finite factors occurring in the β -expansion of the real numbers in $[0, 1]$.

Example 1.3.7. Consider the golden ratio $\varphi = \frac{1+\sqrt{5}}{2}$ and let $x = \frac{1}{2}$. Then $d_\varphi(\frac{1}{2})$ starts with 0100. Indeed, one has

$$\begin{aligned} x_1 &= \lfloor \varphi x \rfloor = \left\lfloor \frac{1 + \sqrt{5}}{4} \right\rfloor = 0; \\ x_2 &= \lfloor \varphi(\varphi x - x_1) \rfloor = \lfloor \varphi^2 x \rfloor = \left\lfloor \frac{6 + 2\sqrt{5}}{8} \right\rfloor = 1; \\ x_3 &= \lfloor \varphi(\varphi(\varphi x - x_1) - x_2) \rfloor = \lfloor \varphi(\varphi^2 x - 1) \rfloor = \left\lfloor \frac{1}{2} \right\rfloor = 0; \\ x_4 &= \lfloor \varphi(\varphi(\varphi(\varphi x - x_1) - x_2) - x_3) \rfloor = \lfloor \varphi^2(\varphi^2 x - 1) \rfloor = \left\lfloor \frac{1 + \sqrt{5}}{4} \right\rfloor = 0. \end{aligned}$$

Let $n \geq 1$ be an integer. The idea to get the digit x_n is the following: take x and multiply it by φ . This operation shifts the φ -expansion of x “to the left”.

Then remove x_1 . The φ -expansion of the obtained number is now $(x_i)_{i \geq 2}$. Then you apply the same procedure until n : multiply by φ the last obtained number and remove x_i (up to $i = n-1$) and finally, take the integer part.

As for non-negative integers in Proposition 1.3.5, the order between real numbers is given by the lexicographic order (Definition 1.1.8) between their β -expansions.

Proposition 1.3.8. *Let $\beta > 1$ be a real number and $x, y \in [0, 1)$. Then*

$$x < y \Leftrightarrow d_\beta(x) <_{\text{lex}} d_\beta(y).$$

Another representation of 1 is interesting and useful for what comes next: the quasi-greedy expansion.

Definition 1.3.9. Let $\beta > 1$ be a real number. If $d_\beta(1) = t_1 \cdots t_m 0^\omega$, with $t_1, \dots, t_m \in \Sigma_\beta$ and $t_m \neq 0$ (in other words if $d_\beta(1)$ is finite), we set $d_\beta^*(1) = (t_1 \cdots t_{m-1} (t_m - 1))^\omega$. Otherwise, we set $d_\beta^*(1) = d_\beta(1)$. An equivalent definition is to set $d_\beta^*(1) = \lim_{x \rightarrow 1^-} d_\beta(x)$. We say that $d_\beta^*(1)$ is the *quasi-greedy β -expansion* of 1.

Example 1.3.10. Consider the golden ratio $\varphi = \frac{1+\sqrt{5}}{2}$. Thanks to the equality $1 = \frac{1}{\varphi} + \frac{1}{\varphi^2}$, one can easily see that $d_\varphi(1) = 110^\omega$ (or simply 11) and thus $d_\varphi^*(1) = (10)^\omega$.

With every real number $\beta > 1$, we associate canonically a numeration system as follows.

Definition 1.3.11. Let $\beta > 1$ be a real number such that $d_\beta^*(1) = (t_j)_{j \geq 1}$. The *numeration system* $U_\beta = (U_i)_{i \in \mathbb{N}}$ *canonically associated with β* is defined by

$$U_i = t_1 U_{i-1} + \cdots + t_i U_0 + 1, \forall i \geq 0.$$

Note that if $\beta = b \in \mathbb{N}_{\geq 2}$, one has $d_b(1) = (b-1)^\omega$, hence $d_b^*(1) = (b-1)^\omega$. We can show by induction that the system U_β is the integer base numeration system from Example 1.3.3.

Notice that for integers, the quasi-greedy expansion of 1 is ultimately periodic. Numbers with such property are known as Parry numbers.

Definition 1.3.12. A *Parry number* is a real number $\beta > 1$ such that $d_\beta^*(1)$ is ultimately periodic.

Definition 1.3.13. A numeration system U is a *Parry numeration system* if there is a Parry number β such that $U = U_\beta$.

A common characteristic of Parry numeration systems is that they are linear.

Definition 1.3.14. A numeration system $U = (U_i)_{i \in \mathbb{N}}$ is said to be *linear* if it ultimately satisfies a homogeneous linear recurrence relation with integer coefficients: there are $k \geq 1, a_{k-1}, \dots, a_0 \in \mathbb{Z}$ such that $a_0 \neq 0$ and $N \geq 0$ such that for all $i \geq N$,

$$U_{i+k} = a_{k-1}U_{i+k-1} + \dots + a_0U_i.$$

The polynomial $X^N(X^k - a_{k-1}X^{k-1} - \dots - a_0)$ is the *characteristic polynomial* of the system, the integer k is the *order* of the recurrence.

Lemma 1.3.15. *If β is a Parry number, then the canonical numeration system U_β is linear.*

Proof. Let β be a Parry number. Then either $d_\beta(1) = t_1 \dots t_k, t_k \neq 0$, either $d_\beta(1) = t_1 \dots t_m(t_{m+1} \dots t_{m+k})^\omega$ where k and m are taken minimal.

First, suppose that $d_\beta(1) = t_1 \dots t_k, t_k \neq 0$. Then

$$d_\beta^*(1) = (t_1 \dots t_{k-1}(t_k - 1))^\omega = (s_j)_{j \in \mathbb{N}_0},$$

where

$$s_j = \begin{cases} t_j \bmod k & \text{if } j \not\equiv 0 \pmod{k}, \\ t_k - 1 & \text{if } j \equiv 0 \pmod{k}. \end{cases}$$

Since $U_\beta = (U_i)_{i \in \mathbb{N}}$ is the numeration system canonically associated with β , one has

$$U_i = s_1U_{i-1} + \dots + s_iU_0 + 1, \forall i \geq 0.$$

In particular,

$$\begin{aligned} U_0 &= 1, \\ U_i &= s_1U_{i-1} + \dots + s_iU_0 + 1 \\ &= t_1U_{i-1} + \dots + t_iU_0 + 1, \forall i \in \llbracket 1, k-1 \rrbracket. \end{aligned}$$

Moreover, if $\ell \geq 0$, then

$$\begin{aligned} U_{k+\ell} &= s_1U_{k+\ell-1} + \dots + s_{k-1}U_{\ell+1} + s_kU_\ell + \dots + s_{k+\ell-1}U_1 + s_{k+\ell}U_0 + 1 \\ &= t_1U_{k+\ell-1} + \dots + t_{k-1}U_{\ell+1} + t_kU_\ell - U_\ell + \sum_{j=1}^{\ell} s_{k+j}U_{\ell-j} + 1 \\ &= t_1U_{k+\ell-1} + \dots + t_{k-1}U_{\ell+1} + t_kU_\ell - U_\ell + \sum_{j=1}^{\ell} s_jU_{\ell-j} + 1 \\ &= t_1U_{k+\ell-1} + \dots + t_{k-1}U_{\ell+1} + t_kU_\ell - U_\ell + U_\ell \\ &= t_1U_{k+\ell-1} + \dots + t_{k-1}U_{\ell+1} + t_kU_\ell. \end{aligned}$$

Now, suppose that $d_\beta(1) = t_1 \cdots t_m (t_{m+1} \cdots t_{m+k})^\omega$ with k and m being chosen minimal. In this case, $d_\beta^*(1) = t_1 \cdots t_m (t_{m+1} \cdots t_{m+k})^\omega = (s_j)_{j \in \mathbb{N}_0}$, where

$$s_j = \begin{cases} t_j & \text{if } j \in \llbracket 1, m+k \rrbracket, \\ t_{m+\ell} & \text{if } j = m + qk + \ell, \text{ with } \ell \in \llbracket 1, k \rrbracket, q > 0. \end{cases}$$

Since $U_\beta = (U_i)_{i \in \mathbb{N}}$ is the numeration system canonically associated with β , one has

$$U_i = s_1 U_{i-1} + \cdots + s_i U_0 + 1, \quad \forall i \geq 0.$$

From this we derive as previously an expression for U_i , $i \in \llbracket 0, m+k-1 \rrbracket$, the initial conditions. Let $\ell \geq 0$. One has

$$\begin{aligned} U_{m+k+\ell} &= s_1 U_{m+k+\ell-1} + \cdots + s_{m+k} U_\ell + s_{m+k+1} U_{\ell-1} + \cdots + s_{m+k+\ell} U_0 + 1 \\ &= t_1 U_{m+k+\ell-1} + \cdots + t_{m+k} U_\ell + \sum_{j=1}^{\ell} s_{m+k+j} U_{\ell-j} + 1. \end{aligned}$$

Moreover,

$$\begin{aligned} \sum_{j=1}^{\ell} s_{m+k+j} U_{\ell-j} + 1 &= \sum_{j=1}^{\ell} s_{m+j} U_{\ell-j} + 1 \\ &= \sum_{i=1}^{m+\ell} s_i U_{\ell+m-i} + 1 - s_1 U_{\ell+m-1} - \cdots - s_m U_\ell \\ &= U_{\ell+m} - t_1 U_{\ell+m-1} - \cdots - t_m U_\ell. \end{aligned}$$

Hence

$$U_{m+k+\ell} = t_1 U_{m+k+\ell-1} + \cdots + t_{m+k} U_\ell + U_{\ell+m} - t_1 U_{\ell+m-1} - \cdots - t_m U_\ell.$$

□

Example 1.3.16. Every integer is a Parry number. The golden ratio φ is a Parry number. Indeed, Example 1.3.10 shows that $d_\varphi^*(1) = (10)^\omega$. It is straightforward to deduce from Definition 1.3.11 that the associated Parry numeration system is the Fibonacci system of Example 1.3.4 defined by the recurrence $F_{i+2} = F_{i+1} + F_i$ and the initial conditions $F_0 = 1, F_1 = 2$. One could also use the proof of Lemma 1.3.15, since $d_\varphi(1) = 11$.

Parry numeration systems are quite manageable, thanks to the following theorem.

Theorem 1.3.17 (Parry [63]). *Let $\beta > 1$ be a real number. A sequence $x = (x_i)_{i \geq 1}$ over \mathbb{N} is the β -expansion of a real number in $[0, 1)$ if and only if $(x_{n+i})_{i \geq 1}$ is lexicographically less than $d_\beta^*(1)$ for all $n \geq 0$.*

As a consequence of this result, to any Parry number β one can canonically associate a deterministic finite automaton $\mathcal{A}_\beta = (Q_\beta, q_0, F_\beta, \Sigma_\beta, \delta_\beta)$ accepting the language $\text{Fact}(D_\beta)$. This automaton \mathcal{A}_β has a special form. Let us set $d_\beta^*(1) = t_1 \cdots t_i(t_{i+1} \cdots t_{i+p})^\omega$ where $i \geq 0$ and $p \geq 1$ are the minimal period and preperiod respectively. The set of states Q_β of \mathcal{A}_β is $\{q_0, \dots, q_{i+p-1}\}$. All states are final. For every $j \in \llbracket 1, i+p \rrbracket$, we have t_j edges $q_{j-1} \rightarrow q_0$ labelled by $0, \dots, t_j-1$ and, for $j < i+p$, one edge $q_{j-1} \rightarrow q_j$ labelled by t_j . There is also an edge $q_{i+p-1} \rightarrow q_i$ labelled by t_{i+p} . See for instance [39, 66]. Note that in [53, Theorem 7.2.13], \mathcal{A}_β is shown to be the trim minimal automaton of $\text{Fact}(D_\beta)$.

Example 1.3.18. The automaton canonically associated to an integer $b \geq 2$ is made of a single state with loops labelled by $0, 1, \dots, b-1$. Considering the golden ratio, the automaton \mathcal{A}_φ is depicted in Figure 1.3. For more examples,

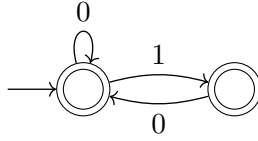


Figure 1.3: The canonical automaton accepting $\text{Fact}(D_\varphi)$.

we refer the reader to the Appendix.

The following statement about the automaton associated with a Parry number is well-known, see for example [50]. Recall Definition 1.2.3 and Proposition 1.2.4.

Lemma 1.3.19. *Let β be a Parry number. The automaton \mathcal{A}_β is primitive.*

Proof. The periodic part of $d_\beta^*(1)$ contains at least a non-zero digit. Consequently, there is a path from every state of \mathcal{A}_β to the initial state q_0 . Moreover, there is a loop on q_0 with label 0. Hence \mathcal{A}_β is irreducible and aperiodic. The conclusion follows. \square

The next definition points out another special type of numeration systems, called Bertrand numeration systems.

Definition 1.3.20. A numeration system U is a *Bertrand numeration system* if for all $w \in \Sigma_U^+$, $w \in \text{rep}_U(\mathbb{N}) \Leftrightarrow w0 \in \text{rep}_U(\mathbb{N})$.

Example 1.3.21. The integer base b of Example 1.3.3 is a Bertrand numeration system. The Fibonacci numeration system of Example 1.3.4 is also a Bertrand numeration system. It is enough to notice that adding or removing zeros at the end of a representation does not change its greediness, since the condition of being greedy is to not contain the factor 11. If we slightly modify the Fibonacci system by taking the initial conditions $U_0 = 1, U_1 = 3$, we get a numeration system $(U_i)_{i \in \mathbb{N}} = (1, 3, 4, 7, 11, 18, 29, 47, \dots)$, which is no longer a Bertrand system. Indeed, 2 is the greedy representation of an integer, but 20 is not, because $\text{rep}_U(\text{val}_U(20)) = 102$.

Example 1.3.22. Consider the numeration system B given by the recurrence $B_{i+1} = 3B_i + 1$ for all $i \in \mathbb{N}$ and $B_0 = 1$. For this numeration system, we have $0^* \text{rep}_B(\mathbb{N}) = \{0, 1, 2\}^* (\{\varepsilon\} \cup 30^*)$ (see [42, p. 131]). The automaton accepting the language $0^* \text{rep}_B(\mathbb{N})$ is depicted in Figure 1.4. By its simple form, it is obvious that it is a Bertrand numeration system. Notice that the sequence $(B_i)_{i \in \mathbb{N}}$ also satisfies the homogeneous linear recurrence $B_{i+2} = 4B_{i+1} - 3B_i$.

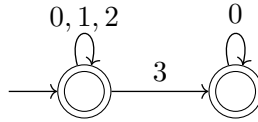


Figure 1.4: The canonical automaton accepting $\{0, 1, 2\}^* (\{\varepsilon\} \cup 30^*)$.

The following statement is a consequence of Bertrand's theorem (see [11] or [12, Chapter 2]).

Proposition 1.3.23. Let $\beta > 1$ be a real number. The numeration system U_β canonically associated with β satisfies

$$0^* \text{rep}_{U_\beta}(\mathbb{N}) = \text{Fact}(D_\beta).$$

Corollary 1.3.24. For all $\beta > 1$, the canonical system U_β associated with β is a Bertrand numeration system.

Proof. It is enough to notice that $w \in \text{Fact}(D_\beta)$ if and only if $w0 \in \text{Fact}(D_\beta)$ for all $w \in \Sigma_\beta^*$. \square

One can prove that every Parry numeration system is a Bertrand numeration system.

Lemma 1.3.25. *The set of Parry numeration systems is a strict subset of the set of Bertrand numeration systems.*

Proof. Thanks to Corollary 1.3.24, we already know that every Parry numeration system is a Bertrand numeration system. Now consider the Bertrand numeration system $B = (B_i)_{i \in \mathbb{N}}$ of Example 1.3.22. We will show that there is no $\beta > 1$ such that $B = U_\beta$. Proceed by contradiction. Assume that there exists β such that B is the numeration system canonically associated with β . The greatest word of length n for the lexicographical order in $0^* \text{rep}_B(\mathbb{N})$ is 30^{n-1} . Consequently, we have $1 = 3/\beta$. The Parry numeration system U_3 is the classical base-3 system and $0^* \text{rep}_{U_3}(\mathbb{N}) = \{0, 1, 2\}^*$, which differs from $0^* \text{rep}_B(\mathbb{N})$. This is a contradiction. \square

Notice that, thanks to Proposition 1.3.23, the automaton canonically associated with a Parry number β accepts the language of the numeration system U_β . Otherwise stated, the numeration language associated with a Parry numeration system is regular.

In Chapter 4, we will make use of a third kind of numeration systems, associated with Pisot numbers. Recall that the conjugates of an algebraic number are the other roots of its minimal polynomial.

Definition 1.3.26. A *Pisot number* is an algebraic integer $\beta > 1$ whose conjugates have modulus strictly less than 1.

Definition 1.3.27. A numeration system U is a *Pisot numeration system* if there is a Pisot number β such that $U = U_\beta$.

Example 1.3.28. Every integer $b \geq 2$ is a Pisot number (since it has no conjugate). The golden ratio is a Pisot number, hence the Fibonacci numeration system is a Pisot numeration system.

In fact, every Pisot number is a Parry number, as shown in [10, 72]. However, there are Parry numbers that are not Pisot numbers, as stated in the following.

Lemma 1.3.29. *The set of Pisot numeration systems is a strict subset of the set of Parry numeration systems.*

Proof. Since Pisot numbers are Parry numbers, every Pisot numeration system is a Parry numeration system. Moreover, consider the numeration system U defined by $U_0 = 1, U_1 = 4, U_2 = 15, U_3 = 54$ and

$$U_{i+4} = 3U_{i+3} + 2U_{i+2} + 3U_i \quad \forall i \geq 0,$$

see [38, Example 3]. The characteristic polynomial has two real roots β and γ and two complex roots with modulus less than 1. We have $\beta \approx 3.61645$ and $\gamma \approx -1.09685$. Hence U is not a Pisot numeration system. However, β is a Parry number, since $d_\beta(1) = 3203$. We have $U = U_\beta$ and U is a Parry numeration system. \square

The numeration system of Example 2.1.2 is also a Parry but not Pisot numeration system.

It is often convenient to work with numeration systems U such that the numeration language is regular: we want to be able to check with a DFA whether or not a word is a valid greedy U -representation.

Definition 1.3.30. Let U be a numeration system. A set X of non-negative integers is *U -recognizable* if the language $\text{rep}_U(X)$ over Σ_U is regular. In the case where U is the integer base- b numeration system, we say that X is *b -recognizable*.

Let \mathcal{B} be the set of Bertrand numeration systems and let \mathcal{R} be the set of numeration systems U whose numeration language $\text{rep}_U(\mathbb{N})$ is regular. The three sets $\mathcal{B} \cap \mathcal{R}$, $\mathcal{B} \setminus \mathcal{R}$ and $\mathcal{R} \setminus \mathcal{B}$ are non-empty. For instance, the modified Fibonacci system of Example 1.3.21 belongs to $\mathcal{R} \setminus \mathcal{B}$. All Parry numeration systems and the Bertrand numeration system of Example 1.3.22 belong to $\mathcal{B} \cap \mathcal{R}$. If β is not a Parry number, for instance when β is transcendental, then the numeration language $\text{rep}_{U_\beta}(\mathbb{N})$ is not regular, even though U_β is a Bertrand system. Hence $\mathcal{B} \setminus \mathcal{R}$ is non-empty. The situation is represented in Figure 1.5.

We will often make the assumption that we are dealing with positional numeration systems such that the numeration language is regular. This is one of our minimal assumptions in Chapter 2, and it is particularly important when we will deal with finite U -kernels in Chapter 4 Section 4.4. In fact, this assumption is somewhat restrictive. Indeed, the next proposition is a particular case of a theorem of Shallit [73].

Proposition 1.3.31. *Let U be a numeration system. If \mathbb{N} is U -recognizable, then the sequence U satisfies a linear recurrence relation over \mathbb{Z} . Otherwise stated, U is a linear numeration system.*

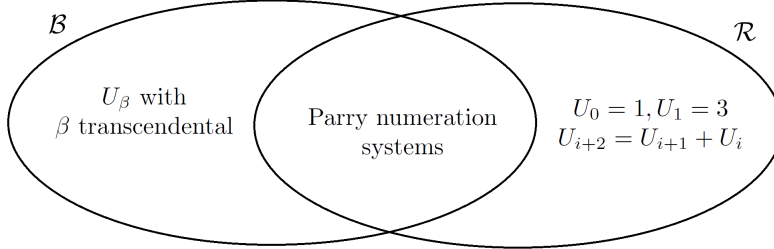


Figure 1.5: Comparison of the sets \mathcal{B} and \mathcal{R} .

Remark that the converse does not hold in general ([73]). Indeed, the numeration system $U = (U_i)_{i \in \mathbb{N}}$ defined by $U_i = (i+1)^2$ for all $i \in \mathbb{N}$ is linear since it satisfies $U_{i+3} = 3U_{i+2} - 3U_{i+1} + U_i \forall i \in \mathbb{N}$, but the associated numeration language is not regular, as shown in [21, Example 1.6.9]. However, note that in [42] and [51], the authors give sufficient conditions.

We will also make use of a folklore property of ultimately periodic sets.

Definition 1.3.32. A set of integers X is *ultimately periodic* if there are $\alpha, \pi \in \mathbb{N}$ with $\pi > 0$ such that for all $n \geq \alpha$, we have $n \in X \Leftrightarrow n + \pi \in X$. If the integers α and π are minimal for this property, then we say that α is the *preperiod* of X and π is the *period* of X .

Notice that this definition is consistent with Definition 1.1.4. Indeed, with each set of non-negative integers X , one can associate its characteristic sequence $\mathbf{1}_X \in \{0, 1\}^{\mathbb{N}}$. This sequence is an infinite word over $\{0, 1\}$ and X is an ultimately periodic set if and only if the sequence $\mathbf{1}_X$ is an ultimately periodic word in the sense of Definition 1.1.4.

Proposition 1.3.33. Let m, r be non-negative integers and let $U = (U_i)_{i \in \mathbb{N}}$ be a linear numeration system. The language

$$\text{val}_U^{-1}(m\mathbb{N} + r) = \{w \in \Sigma_U^* : \text{val}_U(w) \in m\mathbb{N} + r\}$$

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

Proof. Regularity is stable if we add or remove a finite number of words in the language. Thus, we can assume that $0 \leq r < m$. Since U is linear,

the sequence $(U_i \bmod m)_{i \in \mathbb{N}}$ is ultimately periodic. Let α be the preperiod and π the period. The following DFA recognizes the reversal of the language $\text{val}_U^{-1}(m\mathbb{N} + r)$. States are pairs (p, q) such that $p, q \in \mathbb{N}, 0 \leq p < m$, and $0 \leq q < \alpha + \pi$. The initial state is $(0, 0)$. Final states are the ones whose first component is r . The alphabet is the set Σ_U . Transitions are defined as follows. For all $j \in \Sigma_U, \forall p \in \llbracket 0, m-1 \rrbracket, \forall q \in \llbracket 0, \alpha + \pi - 2 \rrbracket$, we have a transition from the state (p, q) of label j to the state $(jU_q + p \bmod m, q+1)$. Moreover, we have a transition from state $(p, \alpha + \pi - 1)$ of label j to the state $(jU_{\alpha + \pi - 1} + p \bmod m, \alpha)$.

Note that the greediness of the accepted words is not checked, since the construction only relies on the numerical value of the words.

For the particular case, it is enough to consider the intersection

$$\text{rep}_U(\mathbb{N}) \cap \text{val}_U^{-1}(m\mathbb{N} + r)$$

of two regular languages. □

Example 1.3.34. Consider the numeration system $U = (U_i)_{i \in \mathbb{N}}$ defined by $U_0 = 1, U_1 = 7$ and the recurrence $U_{i+2} = 7U_{i+1} - 2U_i$. One has $\Sigma_U = \llbracket 0, 6 \rrbracket$. Let us describe a DFA accepting $\text{val}_U^{-1}(4\mathbb{N} + 1)$. The sequence $(U_i \bmod 4)_{i \in \mathbb{N}}$ is given by 13^ω . Hence its preperiod α is 1 and so is its period π . States of the DFA given in the proof of Proposition 1.3.33 are the pairs

$$(0, 0), (1, 0), (2, 0), (3, 0), (0, 1), (1, 1), (2, 1), (3, 1).$$

The initial state is $(0, 0)$ and there are two final states: $(1, 0)$ and $(1, 1)$. Note that, following the construction of Proposition 1.3.33, the states $(1, 0), (2, 0)$ and $(3, 0)$ are not accessible, hence we will not consider them anymore. Let us have a look at transition relations. Let $j \in \llbracket 0, 6 \rrbracket$. From the state $(0, 0)$, there is an edge labelled by j to the state $(jU_0 + 0 \bmod 4, 1) = (j \bmod 4, 1)$. From state $(0, 1)$, one goes to the state $(jU_1 + 0 \bmod 4, 1) = (7j \bmod 4, 1)$ when reading j . There is a transition of label j from the state $(1, 1)$ to the state $(jU_1 + 1 \bmod 4, 1) = (7j + 1 \bmod 4, 1)$, and so on. The automaton is depicted in Figure 1.6.

Let us conclude this section by some properties of the integer base, with which we will deal in Chapter 3. In base- b numeration systems, the numeration language $\text{rep}_b(\mathbb{N})$ is always regular. Otherwise stated, \mathbb{N} is b -recognizable for every integer $b \geq 2$. Moreover, thanks to Cobham's theorem, it is known that the sets with this property (being recognizable in any integer base $b \geq 2$) are exactly the ultimately periodic sets.

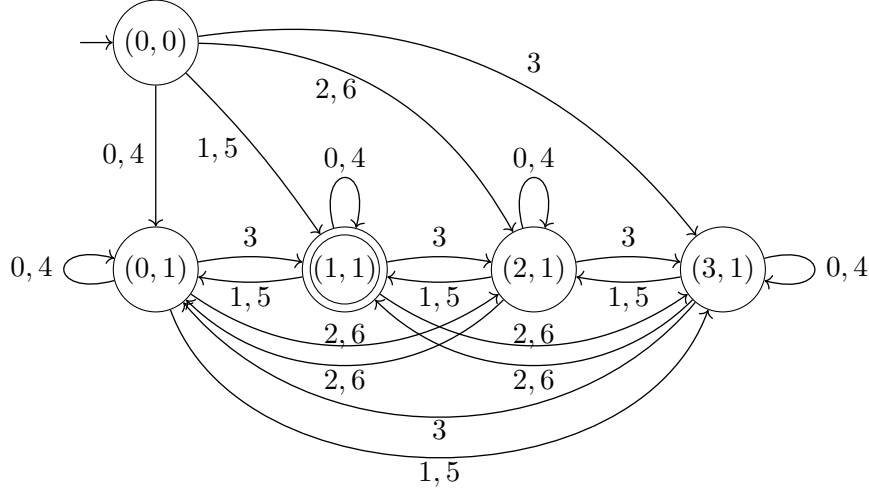


Figure 1.6: A DFA accepting $\text{val}_U^{-1}(4\mathbb{N}+1)$ ($U_{i+2} = 7U_{i+1} - 2U_i, U_0 = 1, U_1 = 7$).

Definition 1.3.35. Two positive integers p, q are said *multiplicatively independent* if $p^a = q^b \Rightarrow a = b = 0$. They are said *multiplicatively dependent* otherwise.

Theorem 1.3.36 (Cobham [31]).

- Let b, b' be two multiplicatively independent integers. Then a subset of \mathbb{N} is both b -recognizable and b' -recognizable if and only if it is ultimately periodic.
- Let b, b' be two multiplicatively dependent integers. Then a subset of \mathbb{N} is b -recognizable if and only if it is b' -recognizable.

This famous theorem is the starting point of many research problems, but we will discuss it later.

Note that given an integer base $b \geq 2$ and a positive integer n , one can easily derive from the base- b expansion of a positive integer its base- b^n expansion and vice versa. Indeed, from a base- b^n expansion, it is enough to send every digit (belonging to $\llbracket 0, b^n - 1 \rrbracket$) onto its base- b expansion, concatenate these words of length n and remove potential leading zeros. The other way around, given a base- b expansion of a natural number, one can cut the word into blocks of letters, each block having a length n (adding leading zeros if necessary). Then, every block of size n is sent onto its image by val_b . Since each block is of length n , one gets *digits* between 0 and $b^n - 1$.

Let us make a few comments about the numerical valuation in base b , val_b . Recall that this function is defined over \mathbb{Z}^* . It can be interesting to restrict the definition domain to Σ_b^* . Indeed, we have $\text{val}_b^{-1}(X) \cap \Sigma_b^* = 0^* \text{rep}_b(X)$ for any subset X of \mathbb{N} . Working with representations of integers, it is sometimes convenient to add extra leading zeros. In particular, we can say that a subset X of \mathbb{N} is b -recognizable if and only if the language $\text{val}_b^{-1}(X)$ is regular, since no word in $\text{rep}_b(X)$ begins with 0. This will be useful in Chapter 3: in this chapter, we will always consider automata accepting $\text{val}_b^{-1}(X)$ instead of $\text{rep}_b(X)$. We thus introduce the following definition.

Definition 1.3.37. The *state complexity* of a b -recognizable subset X of \mathbb{N} with respect to the base b is the state complexity of the language $\text{val}_b^{-1}(X)$.

Note that our choice of $\text{val}_b^{-1}(X)$ rather than $\text{rep}_b(X)$ makes no significant difference to state complexity, since the state complexity of the languages $\text{rep}_b(X)$ and $\text{val}_b^{-1}(X)$ differ at most by 1.

The case of integer base has already been widely studied, as we will explain later. Lots of properties are thus known. The next one will be helpful for our considerations in Chapter 3, see for example [16].

Proposition 1.3.38. Let $b \geq 2$ be an integer. For any $m, t \in \mathbb{N}$, if the set X is b -recognizable, then so is $mX + t$.

Finally, in the present dissertation, we will need to represent not only natural numbers, but also pairs of natural numbers. If $u = u_1 \cdots u_n \in \Sigma_1^*$ and $v = v_1 \cdots v_n \in \Sigma_2^*$ are words of the same length n , then we use the notation (u, v) to designate the word $(u_1, v_1) \cdots (u_n, v_n)$ of length n over the alphabet $\Sigma_1 \times \Sigma_2$. For $(m, n) \in \mathbb{N}^2$, we write

$$\text{rep}_U(m, n) = (0^{\ell - |\text{rep}_U(m)|} \text{rep}_U(m), 0^{\ell - |\text{rep}_U(n)|} \text{rep}_U(n))$$

where $\ell = \max\{|\text{rep}_U(m)|, |\text{rep}_U(n)|\}$. Otherwise stated, we add leading zeros to the shortest expansion (if any) in order to obtain two words of the same length. Finally, for a subset X of \mathbb{N}^2 , we write

$$\text{val}_U^{-1}(X) = (0, 0)^* \text{rep}_U(X).$$

1.4 Abstract numeration systems

In this thesis, we only work with positional numeration systems. As explained previously, among those, we are particularly interested in positional numeration systems such that the numeration language is regular. Indeed, this

property allows us to verify with a finite automaton whether a given word is a valid greedy representation. Moreover, for a positional numeration system U , the numeration language is regular if and only if all ultimately periodic sets are U -recognizable, cf. Proposition 1.3.33. As P. Lecomte and M. Rigo in [48], one could go the other way around: one takes an infinite regular language L over an alphabet Σ to build a numeration system, considering this language L as the valid representations of integers, instead of starting with a sequence U of integers and searching for conditions so that the numeration language is regular.

Definition 1.4.1. An *abstract numeration system* is a triplet

$$S = (L, \Sigma, <)$$

where L is an infinite regular language, called the *numeration language*, written over a totally ordered alphabet $(\Sigma, <)$. One can enumerate the words in L using the genealogical order $<_{\text{gen}}$ induced by the order $<$ on Σ . This gives a one-to-one correspondence $\text{rep}_S: \mathbb{N} \rightarrow L$ mapping any non-negative integer to the $(n+1)^{\text{th}}$ word in L . A set X of integers is said *S -recognizable* if $\text{rep}_S(X)$ is regular.

Even though we mainly interest ourselves to positional numeration systems in this text, some results are well-known in the general framework of abstract numeration systems. Since any positional numeration system with a regular numeration language is an abstract numeration system (which is one of our hypotheses in the present dissertation), we can make use of these results. Every such statement will be restated in our context in the following.

1.5 Rational series

In Chapter 2, we will make use of formal power series. For the sake of completeness, we restate here some necessary basic definitions and properties. Details can be found for example in [8] and [71]. Note that we only define formal power series with coefficients in $\mathbb{R}, \mathbb{R}_+, \mathbb{Z}$ or \mathbb{N} , which is sufficient for the present dissertation, but rational series are defined in general over a semiring (and a finite alphabet).

Definition 1.5.1. Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{R}_+, \mathbb{Z}, \mathbb{N}\}$. The *set of formal power series in x with coefficients in \mathbb{K}* , denoted by $\mathbb{K}[[x]]$, is defined as follows. The elements of $\mathbb{K}[[x]]$, called *formal power series*, are infinite expressions of the form

$$S(x) = \sum_{i=0}^{+\infty} s_i x^i,$$

where $s_i \in \mathbb{K}$ for all $i \in \mathbb{N}$.

Given two formal power series $S(x) = \sum_{i=0}^{+\infty} s_i x^i$ and $T(x) = \sum_{i=0}^{+\infty} t_i x^i$, their *sum* is given by

$$(S + T)(x) = \sum_{i=0}^{+\infty} (s_i + t_i) x^i$$

and their *product* by

$$(ST)(x) = \sum_{i=0}^{+\infty} \sum_{j+k=i} (s_j \cdot t_k) x^i.$$

One can also define the multiplication by a scalar $r \in \mathbb{K}$:

$$(rS)(x) = \sum_{i=0}^{+\infty} (r \cdot s_i) x^i.$$

A formal series S is *proper* if the coefficient of x^0 is equal to 0. In this case, if $n \in \mathbb{N}$ and

$$S^n(x) = \sum_{i=0}^{+\infty} t_i x^i,$$

then $t_i = 0$ for all $i < n$. This implies that the sum

$$S^*(x) = \sum_{n=0}^{+\infty} S^n(x)$$

exists. It is called the *star* of S .

The *rational operations* in $\mathbb{K}[[x]]$ are the sum, the product, the multiplication by a scalar and the star operation. A subset of $\mathbb{K}[[x]]$ is *rationally closed* if it is closed for the rational operations. The smallest subset containing a subset A of $\mathbb{K}[[x]]$ and which is rationally closed is called the *rational closure* of A . The set of polynomials is a strict subset of $\mathbb{K}[[x]]$. We denote it by $\mathbb{K}[x]$. A formal series is \mathbb{K} -*rational* if it is in the rational closure of $\mathbb{K}[x]$.

The following statement is classical, see [71, Corollary 9.2].

Proposition 1.5.2. *Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{Z}\}$. A formal power series $S(x) = \sum_{i=0}^{+\infty} s_i x^i$ in x with coefficients in \mathbb{K} is \mathbb{K} -rational if and only if the sequence $(s_i)_{i \in \mathbb{N}}$ satisfies a linear recurrent equation with coefficients in \mathbb{K} .*

Let $S(x) = \sum_{i=0}^{+\infty} s_i x^i$ be a \mathbb{R} -rational series and consider the polynomial $P(x) = a_0 + a_1 x + \cdots + a_k x^k$, the characteristic polynomial of the linear

recurrence satisfied by the sequence $(s_i)_{i \in \mathbb{N}}$. We can deduce from the previous proposition that any \mathbb{R} -rational series can be expressed as a quotient of two polynomials. Moreover, we can suppose that the denominator is the reciprocal polynomial of P : $Q(x) = a_k + a_{k-1}x + \cdots + a_1x^{k-1} + a_0x^k$. Let us show it on an example. Consider the rational series $S(x) = \sum_{i=0}^{+\infty} s_i x^i$ where $s_0 = 1, s_1 = 1$ and $s_{i+2} = 2s_{i+1} + 3s_i$ for all $i \in \mathbb{N}$. In this case, $P(x) = x^2 - 2x - 3$ and we have

$$\begin{aligned} S(x) &= \sum_{i=0}^{+\infty} s_i x^i = s_0 + s_1 x + \sum_{i=0}^{+\infty} s_{i+2} x^{i+2} \\ &= 1 + x + \sum_{i=0}^{+\infty} (2s_{i+1} + 3s_i) x^{i+2} \\ &= 1 + x + 2x \sum_{i=0}^{+\infty} s_{i+1} x^{i+1} + 3x^2 \sum_{i=0}^{+\infty} s_i x^i \\ &= 1 + x + 2x(S(x) - s_0) + 3x^2 S(x). \end{aligned}$$

Hence

$$S(x) = \frac{1-x}{1-2x-3x^2}.$$

Note that the numerator only depends on the initial conditions. The other way around, from a rational fraction (i.e. the quotient of two polynomials) over \mathbb{R} , one can derive a formal power series. Indeed, if $Q, R \in \mathbb{R}[x]$, then one obtain a formal power series by carrying out the (long) Euclidean division of Q by R . If R is the reciprocal polynomial of a linear recurrent sequence, such a division is always possible. Lets us state this result formally (cf. [8, Proposition 1.1]).

Proposition 1.5.3. *A series S is \mathbb{R} -rational in and only if there exist polynomials P and Q in $\mathbb{R}[x]$ with $Q(0) = 1$ such that S is the power series expansion of the rational function P/Q .*

Let us now give a characterization of \mathbb{R}_+ -rational series with Soittola's theorem (see [71, Theorem 10.2] or [9]).

Consider a \mathbb{R} -rational series $S(x) = \sum_{i=0}^{+\infty} s_i x^i$. Then the sequence $(s_i)_{i \in \mathbb{N}}$ satisfies a linear recurrent equation. With this equation, one can associate its characteristic polynomial, which admits zeros, called *eigenvalues*. (Note that these eigenvalues are the inverse of the poles of the fraction given above.) A \mathbb{R} -rational series admits a *dominating eigenvalue* if there is among its eigenvalues a unique eigenvalue having maximal modulus.

Definition 1.5.4. Let S_0, \dots, S_{p-1} be formal power series in $\mathbb{R}[[x]]$. The *merge* of these series is the formal power series defined by

$$S(x) = \sum_{i=0}^{p-1} x^i S_i(x^p).$$

Otherwise stated, if $n = mp + i$ where $i \in \llbracket 0, p-1 \rrbracket$, then the coefficient of x^n in $S(x)$ is given by the coefficient of x^m in $S_i(x)$.

Theorem 1.5.5 (Soittola). *A power series over \mathbb{R}_+ is \mathbb{R}_+ -rational if and only if it is the merge of polynomials and of \mathbb{R} -rational series having a dominating eigenvalue.*

We conclude this section with Schützenberger’s theorem of 1961, see [8, Theorem 7.1].

Definition 1.5.6. Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{R}_+\}$. A formal power series $S(x) = \sum_{i=0}^{+\infty} s_i x^i$ in $\mathbb{K}[[x]]$ is \mathbb{K} -recognizable if there exist an integer $n \geq 1$, a matrix $A \in \mathbb{K}^{n \times n}$ and two vectors $X \in \mathbb{K}^{1 \times n}$ and $Y \in \mathbb{K}^{n \times 1}$ such that for all $i \geq 0$, one has $s_i = X A^i Y$.

Theorem 1.5.7 (Schützenberger). *Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{R}_+\}$. A formal power series is \mathbb{K} -recognizable if and only if it is \mathbb{K} -rational.*

1.6 Automatic sequences

The aim of this short section is to introduce basic definitions and properties about automatic sequences needed for a good understanding of Chapter 4, Section 4.4. For a survey on automatic sequences, we refer the reader to [6].

Definition 1.6.1. Let U be a numeration system. We say that an infinite word $\mathbf{x} = (x_i)_{i \in \mathbb{N}}$ over an alphabet Σ is U -automatic (i.e. is a U -automatic sequence) if there is a complete DFAO $(Q, q_0, \Sigma_U, \delta, \Sigma, \tau)$ with transition function $\delta: Q \times \Sigma_U \rightarrow Q$ and output function $\tau: Q \rightarrow \Sigma$ such that there is a loop of label 0 over the initial state q_0 and

$$x_i = \tau(\delta(q_0, \text{rep}_U(i))), \forall i \in \mathbb{N}.$$

The infinite word \mathbf{x} is b -automatic if $U = (b^i)_{i \in \mathbb{N}}$ for an integer $b \geq 2$. We say that an infinite word \mathbf{x} is *Pisot-automatic* (resp. *Parry-automatic*, resp. *Bertrand-automatic*) if U is a Pisot numeration system (resp. a Parry numeration system, resp. a Bertrand numeration system).

Thanks to Cobham, we know exactly the b -automatic sequences, see [6, Theorem 6.3.2].

Theorem 1.6.2. *Let $b \geq 2$ be an integer. A sequence is b -automatic if and only if it is the image under a coding of a fixed point of a b -uniform morphism.*

Properties of Parry-automatic sequences are discussed in [36]. The next result is classical for abstract numeration systems, see for instance [65]. We restate it in our context.

Theorem 1.6.3. *Let U be a numeration system such that $\text{rep}_U(\mathbb{N})$ is regular. An infinite word $\mathbf{x} = (x_i)_{i \in \mathbb{N}}$ over Σ is U -automatic if and only if, for all $\sigma \in \Sigma$, the set $\{j \geq 0 : x_j = \sigma\}$ is U -recognizable. Otherwise stated, \mathbf{x} is U -automatic if and only if, for all $\sigma \in \Sigma$, the set $\{\text{rep}_U(j) : j \in \mathbb{N}, x_j = \sigma\}$ is regular.*

Proof. Let us set $\mathcal{F}_U(\mathbf{x}, \sigma) = \{\text{rep}_U(j) : j \in \mathbb{N}, x_j = \sigma\}$.

First, suppose that \mathbf{x} is U -automatic. Then \mathbf{x} is generated by a DFAO $\mathcal{A} = (Q, q_0, \Sigma_U, \delta, \Sigma, \tau)$. Let $\sigma \in \Sigma$. Let $L(\mathcal{B})$ be the language recognized by the DFA $\mathcal{B} = (Q, q_0, F, \Sigma_U, \delta)$, where the set of final states F only contains the states q such that $\tau(q) = \sigma$. Then $\mathcal{F}_U(\mathbf{x}, \sigma)$ is regular, since it is the intersection of the two regular languages $L(\mathcal{B})$ and $\text{rep}_U(\mathbb{N})$.

Suppose now that for all $\sigma \in \Sigma$, the set $\mathcal{F}_U(\mathbf{x}, \sigma)$ is regular. Let us denote $\Sigma = \{\sigma_1, \dots, \sigma_n\}$. If $m \neq \ell$,

$$\mathcal{F}_U(\mathbf{x}, \sigma_m) \cap \mathcal{F}_U(\mathbf{x}, \sigma_\ell) = \emptyset \quad \text{and} \quad \text{rep}_U(\mathbb{N}) = \bigcup_{\ell=1}^n \mathcal{F}_U(\mathbf{x}, \sigma_\ell).$$

By hypothesis, for all $\ell \in \llbracket 1, n \rrbracket$, $\mathcal{F}_U(\mathbf{x}, \sigma_\ell)$ is recognized by a DFA, say $\mathcal{B}_\ell = (Q_\ell, q_{0,\ell}, F_\ell, \Sigma_U, \delta_\ell)$. From all these automata, we can build a DFAO $\mathcal{A} = (Q, q_0, \Sigma_U, \delta, \Sigma, \tau)$ to generate \mathbf{x} using the numeration system U . Let us set $Q = Q_1 \times \dots \times Q_n$, $q_0 = (q_{0,1}, \dots, q_{0,n})$. For all state $(q_1, \dots, q_n) \in Q$ and for all $a \in \Sigma_U$, we set $\delta((q_1, \dots, q_n), a) = (\delta_1(q_1, a), \dots, \delta_n(q_n, a))$. If there is a unique ℓ such that $q_\ell \in F_\ell$, then $\tau((q_1, \dots, q_n)) = \sigma_\ell$, otherwise the state can not be reached by a word of $\text{rep}_U(\mathbb{N})$ and the corresponding output is meaningless. We can conclude, since the word \mathbf{x} is obtained from U and the DFAO \mathcal{A} . \square

Let $b \geq 2$ be an integer. The b -kernel of an infinite word $\mathbf{x} = (x_i)_{i \in \mathbb{N}}$ over Σ is the set of its subsequences of the form

$$\{(x_{b^e i + d})_{i \in \mathbb{N}} \in \Sigma^\omega : e \geq 0, 0 \leq d < b^e\}.$$

Observe that an element of the b -kernel is obtained by considering those indices whose base- b expansions end with $\text{rep}_b(d)$ (possibly preceded by some zeros to get a suffix of length e). With this in mind, we introduce the more general U -kernel of an infinite word. Note that this definition is from [67]. In [25], the authors give another definition of the kernel.

Definition 1.6.4. Let U be a numeration system and $s \in \Sigma_U^*$ be a finite word. Define the ordered set of integers

$$\mathcal{K}_s = \text{val}_U(0^* \text{rep}_U(\mathbb{N}) \cap \Sigma_U^* s) = \{k(s, 0) < k(s, 1) < \dots\}.$$

Depending on s , it is possible for this set to be finite or empty. The U -kernel of an infinite word $\mathbf{x} = (x_i)_{i \in \mathbb{N}}$ over Σ is the set

$$\ker_U(\mathbf{x}) = \{(x_{k(s,i)})_{i \in \mathbb{N}} : s \in \Sigma_U^*\}.$$

With the above remark, this set can contain finite or even empty subsequences.

Example 1.6.5. Consider the Fibonacci numeration system F of Example 1.3.4 defined by $F_{i+2} = F_{i+1} + F_i$ and $F_0 = 1, F_1 = 2$. Recall that $\Sigma_F = \{0, 1\}$ and $\text{rep}_F(\mathbb{N}) = 1\{0, 01\}^* \cup \{\varepsilon\}$. Let $\mathbf{x} = (x_i)_{i \in \mathbb{N}}$ be the F -automatic sequence built through the DFAO depicted in Figure 1.7, where the output function applied to a state is the name of the state. The first

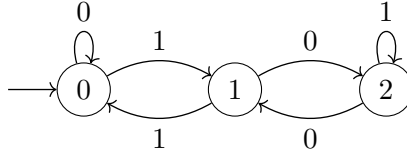


Figure 1.7: A DFAO generating a Fibonacci-automatic sequence.

terms of the sequence \mathbf{x} are

$$\mathbf{x} = 012122011202020101120120200122010120101120012121202 \dots$$

One has

$$\mathcal{K}_\varepsilon = \text{val}_F((0^* 1\{0, 01\}^* \cup \{\varepsilon\}) \cap \{0, 1\}^* \varepsilon) = \text{val}_F(0^* 1\{0, 01\}^* \cup \{\varepsilon\}) = \mathbb{N}.$$

Hence $k(\varepsilon, n) = n \forall n \in \mathbb{N}$ and the sequence \mathbf{x} belongs to the U -kernel of \mathbf{x} . Let us also take a look at the suffix $s = 1$. One has

$$\begin{aligned} \mathcal{K}_1 &= \text{val}_F(0^* \text{rep}_F(\mathbb{N}) \cap \{0, 1\}^* 1) = \text{val}_F(0^* \text{rep}_F(\mathbb{N}) 01) \\ &= \{1, 4, 6, 9, 12, 14, 17, 19, 22, 25, \dots\}. \end{aligned}$$

Hence the sequence $1202001220101012122 \dots$ is an element of $\ker_F(\mathbf{x})$.

To compute completely the F -kernel of the sequence \mathbf{x} , we use the technique introduced in the proof of [67, Proposition 7], which is similar to the proof of Proposition 4.4.3 in one dimension. The F -kernel of the sequence \mathbf{x} is

$$\ker_F(\mathbf{x}) = \{\emptyset, \mathbf{x}, \mathbf{x}_{(0\,1)}, \mathbf{x}_{(0\,2)}, \mathbf{x}_{(1\,2)}, \mathbf{x}_{(0\,1\,2)}, \mathbf{x}_{(0\,2\,1)}\},$$

where the sequence $\mathbf{x}_{(m\,n)}$ ($m, n \in \{0, 1, 2\}$) is the sequence obtained by replacing m by n and n by m in the sequence \mathbf{x} and in the same way, the sequence $\mathbf{x}_{(m\,n\,\ell)}$ ($m, n, \ell \in \{0, 1, 2\}$) is the sequence obtained from the sequence \mathbf{x} by replacing every occurrence of m by n , of n by ℓ and of ℓ by m .

The next two results have been obtained in the general framework of abstract numeration systems (see [67, Propositions 7 and 9]).

Proposition 1.6.6. *Let U be a numeration system such that $\text{rep}_U(\mathbb{N})$ is regular. An infinite word \mathbf{x} is U -automatic if and only if its U -kernel is finite.*

Proposition 1.6.7. *Let U be a numeration system. If an infinite word is U -automatic, then it is reversal- U -automatic, i.e. its n^{th} term is obtained by reading the reversal of $\text{rep}_U(n)$ in a DFAO.*

Remark 1.6.8. Notice that the proof of the latter result only relies on classical constructions on automata defined from the DFAO generating the U -automatic sequence. The same construction applies in multidimensional setting, which will be defined in Chapter 4, Section 4.4.

1.7 Some material about p -adic numbers

This section is devoted to introduce basic definitions and some properties about p -adic numbers that will be useful in Chapter 2 to characterize linear recurrent sequences for which all coefficients are multiple of the same prime, and to tackle some examples. We refer the interested reader to [40] and [68].

In the sequel of this section, p is a prime number. Let us define an absolute value $|\cdot|_p$ on \mathbb{Z} .

Definition 1.7.1. The p -adic valuation of a non-zero integer n , denoted $\nu_p(n)$, is the exponent of the highest power of p dividing n . Otherwise stated, for all $n \in \mathbb{Z}_0$, $\nu_p(n)$ is the unique positive integer satisfying

$$n = p^{\nu_p(n)} m \quad \text{with} \quad \gcd(p, m) = 1.$$

We also set $\nu_p(0) = +\infty$. Thus, ν_p can be seen as a function from \mathbb{Z} to $\mathbb{N} \cup \{+\infty\}$.

Definition 1.7.2. The p -adic absolute value on \mathbb{Z} is defined by

$$|n|_p = \begin{cases} p^{-\nu_p(n)} & \text{if } n \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Example 1.7.3. We have

$$\begin{aligned} \nu_2(3) = \nu_2(5) = 0, \nu_2(2) = 1, \nu_2(8) = 3 = \nu_2(24), \\ \nu_3(2) = 0, \nu_3(9) = 2, \nu_3(6) = 1, \end{aligned}$$

hence

$$\begin{aligned} |3|_2 = |5|_2 = 2^{-0} = 1, |2|_2 = \frac{1}{2}, |8|_2 = |24|_2 = \frac{1}{8}, \\ |2|_3 = 1, |9|_3 = \frac{1}{9}, |6|_3 = \frac{1}{3}. \end{aligned}$$

Note that for all $n \in \mathbb{Z}$, $|n|_p \leq 1$.

One can naturally extend this absolute value to \mathbb{Q} as in the following definition.

Definition 1.7.4. The p -adic absolute value on \mathbb{Z} extends to \mathbb{Q} by setting

$$\left| \frac{m}{n} \right|_p = \frac{|m|_p}{|n|_p} \quad \forall m, n \in \mathbb{Z}, n \neq 0.$$

Proposition 1.7.5. For all $m, n \in \mathbb{Q}$, we have $|m \cdot n|_p = |m|_p \cdot |n|_p$. Moreover, the p -adic absolute value is non-archimedean, i.e.

$$|m + n|_p \leq \max \{ |m|_p, |n|_p \}.$$

Furthermore, if $|m|_p \neq |n|_p$, the previous inequality is an equality.

Example 1.7.6. We have

$$\begin{aligned} \left| \frac{3}{2} \right|_2 = \frac{|3|_2}{|2|_2} = \frac{1}{\frac{1}{2}} = 2, \left| \frac{3}{2} \right|_3 = \frac{\frac{1}{3}}{1} = \frac{1}{3}, \\ \left| \frac{8}{24} \right|_2 = 1, \left| \frac{9}{6} \right|_3 = \frac{1}{3}, \\ \left| \frac{8}{24} + \frac{3}{2} \right|_2 = \left| \frac{11}{6} \right|_2 = \frac{1}{\frac{1}{2}} = 2 = \max \left\{ \left| \frac{8}{24} \right|_2, \left| \frac{3}{2} \right|_2 \right\}. \end{aligned}$$

In fact, the field \mathbb{Q} is not complete with respect to the p -adic absolute value: there are Cauchy sequences that do not have a limit. Let us consider a completion of \mathbb{Q} .

Definition 1.7.7. The field of p -adic numbers, denoted by \mathbb{Q}_p , is the completion of \mathbb{Q} with respect to the p -adic absolute value.

In particular, \mathbb{Q}_p is an extension field of \mathbb{Q} , and we can extend the p -adic absolute value on \mathbb{Q} to \mathbb{Q}_p , which we will still denote $|\cdot|_p$, and which is non-archimedean. Note that \mathbb{Q} is dense in \mathbb{Q}_p with respect to $|\cdot|_p$. Moreover, the p -adic valuation ν_p also extends to \mathbb{Q}_p : for all $x \in \mathbb{Q}_p \setminus \{0\}$, there is an integer $\nu_p(x)$ such that $|x|_p = p^{-\nu_p(x)}$. As before, one can extend it to \mathbb{Q}_p by setting $\nu_p(0) = +\infty$.

Remark 1.7.8. Contrarily to \mathbb{R} , a series $\sum_{i \geq 0} \gamma_i$ converges in \mathbb{Q}_p if and only if $\lim_{i \rightarrow +\infty} |\gamma_i|_p = 0$, since \mathbb{Q}_p is a complete non-archimedean field with respect to the p -adic absolute value.

Among p -adic numbers, one can point out special numbers: the p -adic integers.

Definition 1.7.9. The set of p -adic integers, denoted \mathbb{Z}_p , is the closed unit ball

$$\mathbb{Z}_p = \{x \in \mathbb{Q}_p : |x|_p \leq 1\}.$$

Note that \mathbb{Z}_p is a subring of \mathbb{Q}_p . Moreover, as expected, the inclusion $\mathbb{Z} \hookrightarrow \mathbb{Z}_p$ is dense.

Proposition 1.7.10. *The set of ordinary integers \mathbb{Z} is dense in \mathbb{Z}_p with respect to the p -adic absolute value. Conversely, every Cauchy sequence in $\mathbb{Z}^{\mathbb{N}}$ has a limit in \mathbb{Z}_p with respect to the p -adic absolute value.*

Remark 1.7.11. The topology of \mathbb{Q}_p is closely related to its algebraic structure. For example, if $x, y \in \mathbb{Q}_p$, then

$$|x - y|_p \leq p^{-n} \quad \text{if and only if} \quad x - y \in p^n \mathbb{Z}_p.$$

With this in mind, we can see that the sets $a + p^n \mathbb{Z}_p$ with $a \in \mathbb{Q}$ and $n \in \mathbb{Z}$ are closed balls in \mathbb{Q}_p with center a and radius p^{-n} .

With Definition 1.7.7, it is quite hard to imagine a non-trivial p -adic number. We can make a description of p -adic numbers in terms of p -adic expansions (or p -adic developments) as follows.

Proposition 1.7.12. *Every $\zeta \in \mathbb{Q}_p$ can be written in the form*

$$\begin{aligned}\zeta &= d_{-N}p^{-N} + \cdots + d_{-1}p^{-1} + d_0 + d_1p + d_2p^2 + \cdots \\ &= \sum_{i \geq -N} d_i p^i,\end{aligned}$$

with $N \in \mathbb{Z}$ and $d_i \in \llbracket 0, p-1 \rrbracket$ for all $i \geq -N$. This representation is unique.

Note that the p -adic integers are exactly those whose p -adic expansion involve non-negative powers of p . A nice property is that for ordinary non-negative integers, the p -adic development corresponds to the usual base- p expansion.

Just as for \mathbb{Q} , we have the following statement.

Proposition 1.7.13. *The field \mathbb{Q}_p is not algebraically closed.*

In view of the previous proposition, we will have to make use of the splitting field of a particular polynomial in Chapter 2, Section 2.6. To know whether a polynomial is irreducible or not, there is the Eisenstein criterion.

Proposition 1.7.14 (Eisenstein irreducibility criterion). *Let*

$$P(X) = a_0 + a_1X + \cdots + a_nX^n$$

be a polynomial satisfying the conditions

- $|a_n|_p = 1$,
- $|a_i|_p < 1$ for $0 \leq i < n$,
- $|a_0|_p = 1/p$.

Then $P(X)$ is irreducible over \mathbb{Q}_p .

The next statement, known as Hensel's Lemma, is helpful to decide whether a polynomial has roots in \mathbb{Z}_p .

Theorem 1.7.15 (Hensel's Lemma). *Let $P(X) = a_0 + a_1X + \cdots + a_nX^n$ be a polynomial whose coefficients are in \mathbb{Z}_p . Suppose that there is a p -adic integer $\zeta \in \mathbb{Z}_p$ such that*

$$|P(\zeta)|_p < |P'(\zeta)|_p^2,$$

where $P'(X)$ is the formal derivative of $P(X)$. Then there is a unique p -adic integer η such that $P(\eta) = 0$ and $|\eta - \zeta|_p < |P'(\zeta)|_p$.

To be precise, recall that for a formal derivative of a polynomial, there is not limit process involved: the formal derivative of the polynomial

$$P(X) = a_0 + a_1X + \cdots + a_nX^n$$

is the polynomial $P'(X) = a_1 + 2a_2X + \cdots + na_nX^{n-1}$.

As seen previously, \mathbb{Q}_p is not algebraically closed. Hence we can consider its algebraic closure $\overline{\mathbb{Q}_p}$. One can extend the p -adic absolute value to $\overline{\mathbb{Q}_p}$ (we refer the interested reader to [40, Section 5.3] for the construction). However, $\overline{\mathbb{Q}_p}$ is not complete with respect to this absolute value. Hence one constructs a completion \mathbb{C}_p .

Proposition 1.7.16. *There are a field \mathbb{C}_p and an absolute value $|\cdot|$ on \mathbb{C}_p such that*

- \mathbb{C}_p contains $\overline{\mathbb{Q}_p}$ and the restriction of $|\cdot|$ to $\overline{\mathbb{Q}_p}$ coincides with the p -adic absolute value,
- \mathbb{C}_p is complete with respect to $|\cdot|$,
- $\overline{\mathbb{Q}_p}$ is dense in \mathbb{C}_p .

The basic ideas of functions remain unchanged when we go to the p -adic numbers. There are no "intervals", hence usually functions are defined on open or closed balls. This is the case for the p -adic logarithm and the p -adic exponential.

Definition 1.7.17. The p -adic logarithm is defined on $\{x \in \mathbb{Z}_p : |x-1|_p < 1\}$ by

$$\log_p(x) = \sum_{i=1}^{+\infty} (-1)^{i+1} \frac{(x-1)^i}{i}.$$

The p -adic exponential is defined on $\{x \in \mathbb{Z}_p : |x|_p < p^{-1/(p-1)}\}$ by

$$\exp_p(x) = \sum_{i=0}^{+\infty} \frac{x^i}{i!}.$$

In this dissertation, \log_p will denote the p -adic logarithm, while \log is the classical logarithm in base p .

Of course, usual properties of logarithms and exponentials also hold in this context. In particular, we have the following.

Proposition 1.7.18. *The p -adic logarithm \log_p is an isomorphism from the multiplicative group $\{x \in \mathbb{Z}_p : |x-1|_p < p^{-1/(p-1)}\}$ to the additive group $\{x \in \mathbb{Z}_p : |x|_p < p^{-1/(p-1)}\}$, and its inverse map is the p -adic exponential \exp_p .*

Let us conclude with a theorem giving a link between functions defined by power series and functions defined by polynomials. In what comes next, if $f(X) = \sum_{i=0}^I a_i X^i$ (I may be $+\infty$), we use the following notation

$$\|f(X)\| = \max_i |a_i|.$$

Theorem 1.7.19 (p -adic Weierstrass Preparation Theorem). *Consider a power series $f(X) = \sum_{i=0}^{+\infty} a_i X^i$ with coefficients in \mathbb{Q}_p such that $a_i \rightarrow 0$ as $i \rightarrow +\infty$, so that $f(x)$ converges for $x \in \mathbb{Z}_p$. Let I be the number defined by the conditions*

$$|a_I|_p = \max_{i \geq 0} |a_i|_p \quad \text{and} \quad |a_i|_p < |a_I|_p \text{ for all } i > I.$$

Then there are a polynomial

$$g(X) = b_0 + b_1 X + \cdots + b_I X^I$$

of degree I with coefficients in \mathbb{Q}_p and a power series

$$h(X) = 1 + c_1 X + c_2 X^2 + \cdots$$

with coefficients in \mathbb{Q}_p satisfying

- $f(X) = g(X)h(X)$,
- $|b_I|_p = \max_{0 \leq i \leq I} |b_i|_p$,
- $\lim_{i \rightarrow +\infty} c_i = 0$, so that $h(x)$ converges for $x \in \mathbb{Z}_p$,
- $|c_i|_p < 1$ for all $i \geq 1$,
- $\|f(X) - g(X)\| < 1$.

In particular, $h(X)$ has no zero in \mathbb{Z}_p .

Note that, since $h(X)$ has no zero in \mathbb{Z}_p , the zeros of $f(X)$ in \mathbb{Z}_p are exactly the same as the zeros of $g(X)$.

Chapter 2

Ultimate periodicity problem for linear numeration systems

2.1 Introduction

The material of this chapter appears in [27]. We address the following decision problem. Our aim is to prove that this problem is decidable for a large class of numeration systems.

Problem 2.1.1. Given a linear numeration system U and a deterministic finite automaton \mathcal{A} whose accepted language is contained in the numeration language $\text{rep}_U(\mathbb{N})$, decide whether the subset X of \mathbb{N} that is recognized by \mathcal{A} is ultimately periodic, i.e. whether or not X is a finite union of arithmetic progressions (along a finite set).

This question about ultimately periodic sets is motivated by the theorem of Cobham (Theorem 1.3.36), stating that these sets are precisely the only ones that are recognizable in all integer base numeration systems. In fact, these sets are also exactly the sets definable by a first-order formula in the Presburger arithmetic $\langle \mathbb{N}, + \rangle$ [16]. Cobham's result has been extended to various settings. Indeed, inspired by this seminal result, many descriptions of b -recognizable sets were given, e.g. morphic, algebraic and logical characterizations [14, 16, 32], extensions of these to systems based on a Pisot number [15], the normalization map [37] or the possible growth functions [28, 35]. See also [34] for a survey.

Recall that we use MSDF convention (see Remark 1.3.2). Considering least significant digit first would not affect decidability, since a language is regular if and only if its reversal is.

Let us quickly review cases where the decision problem is known to be decidable. The problem was first solved by Honkala for integer base systems [43], the proof relying on number theoretic results. Another way to tackle the problem is the one of [47], where the authors bound the syntactic complexity of ultimately periodic sets written in base b . Thanks to a deep analysis of the structure of the automata accepting ultimately periodic sets in [55, 13, 54], an efficient procedure is now known for integer base systems. For Pisot numeration systems (and in particular for integer base systems), one can make use of first-order logic and the decidable extension $\langle \mathbb{N}, +, V_U \rangle$ of Presburger arithmetic [15]. (If $U = (U_i)_{i \in \mathbb{N}}$, then $V_U(n)$ is the smallest U_i appearing in the greedy representation of n with a non-zero coefficient.) Given a U -recognizable set X , there is a first order formula φ in $\langle \mathbb{N}, +, V_U \rangle$ describing X . The formula

$$(\exists N)(\exists p)(\forall n \geq N)(\varphi(n) \Leftrightarrow \varphi(n + p))$$

expresses when X is ultimately periodic, N being a preperiod and p a period of X . The logic formalism can be applied to systems such that the addition is recognizable by an automaton, i.e. the set $\{(x, y, z) \in \mathbb{N}^3 : x + y = z\}$ is U -recognizable. This is the case for Pisot numeration systems [37].

When addition is not known to be U -recognizable, other techniques must be found. The problem was shown in [7] to be decidable for some non-Pisot linear numeration systems satisfying a gap condition, $\lim_{i \rightarrow +\infty} U_{i+1} - U_i$ is infinite, and a more technical condition, $\lim_{m \rightarrow +\infty} N_U(m) = +\infty$, where $N_U(m)$ is the number of residue classes that appear infinitely often in the sequence $(U_i \bmod m)_{i \in \mathbb{N}}$. An example of such system is given by the relation $U_{i+3} = 3U_{i+2} + 2U_{i+1} + 3U_i$.

In view of the above summary, we are looking for a decision procedure that may be applied to non-Pisot linear numeration systems such that $N_U(m)$ does not tend to infinity when $m \rightarrow +\infty$. Hence we want to take into account systems where we are not able to apply a decision procedure based on first-order logic nor on the technique from [7]. We have some minimal assumptions that we will discuss in Section 2.2. We follow Honkala's original scheme: if a DFA \mathcal{A} is given as input (the question being whether the corresponding recognized subset of \mathbb{N} is ultimately periodic), the number of states of \mathcal{A} should provide an upper bound on the admissible preperiods and periods. If there is a finite number of such pairs to test, then we build a DFA $\mathcal{A}_{N,p}$ for each pair (N, p) and one can test whether or not the two automata \mathcal{A} and $\mathcal{A}_{N,p}$ accept the same language. This provides us with a decision procedure. In other words, the idea is that if the given DFA has few states, then it cannot accept an ultimately periodic set with a large minimal period.

Example 2.1.2. Consider the numeration system defined by

$$G_{i+4} = 2G_{i+3} + 2G_{i+2} + 2G_i.$$

The largest root β of the characteristic polynomial is roughly 2.804, and -1.134 is another root of modulus larger than one. Note that $d_\beta(1) = 2202$. Hence, with the initial conditions 1, 3, 9, 25 given by Definition 1.3.11, this numeration system is the numeration system G_β canonically associated with β and is thus a Parry non-Pisot numeration system. Moreover, $\text{rep}_G(\mathbb{N})$ is a regular language over $\{0, 1, 2\}$. When m is a power of 2, there is a unique congruence class visited infinitely many times by the sequence $(G_i \bmod m)_{i \in \mathbb{N}}$ because $G_i \equiv 0 \pmod{2^r}$ for large enough i . For such an example, $N_G(m)$ does not tend to infinity and thus the previously known decision procedures may not be applied. This is a perfect candidate for which no decision procedures are known.

This chapter is organized as follows. In Section 2.2, we give our assumptions on the considered numeration systems. In Section 2.3, we collect several useful known results on periodic sets and U -representations. In particular, we relate the length of the U -representation of an integer to its value. Then in Section 2.4, we discuss cases to bound the admissible periods. There are three kinds of prime factors of the admissible periods: those that divide all the coefficients of the recurrence, those that do not but divide the last coefficient of the recurrence, and those that do not divide the last coefficient of the recurrence. This section is the main point of the chapter. In Section 2.5, we apply the results of the previous sections. First, we obtain a decision procedure when the gcd of the coefficients of the recurrence relation is 1, see Theorem 2.5.1. This extends the scope of results from [7]. On the other hand, if there are primes dividing all the coefficients, our approach heavily relies on quite general arithmetic properties of linear recurrence relations. It has therefore inherent limitations because of notoriously difficult results in p -adic analysis such as finding bounds on the growth rate of blocks of zeros in p -adic numbers of a special logarithmic form. We discuss the question and give illustrations of these p -adic techniques in Section 2.6. We end this chapter with some concluding remarks.

2.2 Our setting

Throughout this chapter, we let $U = (U_i)_{i \in \mathbb{N}}$ be a numeration system. We have minimal assumptions on this considered linear numeration system U .

(H1) \mathbb{N} is U -recognizable,

(H2) there are arbitrary large gaps between consecutive terms:

$$\limsup_{i \rightarrow +\infty} (U_{i+1} - U_i) = +\infty,$$

(H3) the gap sequence $(U_{i+1} - U_i)_{i \in \mathbb{N}}$ is ultimately non-decreasing:

$$\exists N \geq 0 \forall i \geq N, U_{i+1} - U_i \leq U_{i+2} - U_{i+1}.$$

Let us make a few remarks. Hypothesis (H1) gives sense to our decision problem. Indeed, under that assumption, ultimately periodic sets are U -recognizable (Proposition 1.3.33), and in particular, there is a DFA \mathcal{A} whose accepted language is contained in the numeration language. Moreover, thanks to Proposition 1.3.31, the sequence $(U_i)_{i \in \mathbb{N}}$ must satisfy a linear recurrent equation. Hence we can set

$$U_{i+k} = a_{k-1}U_{i+k-1} + \cdots + a_0U_i, \quad (2.1)$$

for all $i \in \mathbb{N}$ with k minimal.

The assumptions (H2) and (H3) imply that $\lim_{i \rightarrow +\infty} (U_{i+1} - U_i) = +\infty$. Note that however, in many cases, even if $\lim_{i \rightarrow +\infty} (U_{i+1} - U_i) = +\infty$, the gap sequence may decrease from time to time.

The reason why we introduce (H3) is that, as showed in Lemma 2.3.6, if $10^n w$ is a greedy representation and if n is large enough, then $10^m w$ is also a valid greedy representation for all $m \geq n$. Otherwise stated, as soon as the greediness property is fulfilled, one can shift the leading 1 at every larger index. This is not always the case, as we will see in Example 2.3.7. This property will be crucial in the proofs of Propositions 2.4.1 and 2.4.6 as well as Theorem 2.4.15, where we construct U -representations with leading 1's in convenient positions.

Note that numeration systems $U = (U_i)_{i \in \mathbb{N}}$ satisfying (H1) and such that all the coefficients of the linear recurrent equation are greater than or equal to 0 also satisfy (H2) and (H3). Indeed, the sequence $(U_{i+1} - U_i)_{i \in \mathbb{N}}$ satisfies the same linear recurrent equation than U and is thus ultimately increasing.

Example 2.1.2 satisfies the assumptions (H1), (H2) and (H3).

Example 2.2.1. Our toy example that we will have in mind in Sections 2.4.3, 2.5.2 and 2.6.1 is given by the recurrence $H_{i+3} = 12H_{i+2} + 6H_{i+1} + 12H_i$. Even though the system is associated with a Pisot number, it is still interesting because $N_H(m)$ does not tend to infinity (thus we cannot follow the decision procedure from [7]) and the gcd of the coefficients of the recurrence is larger than 1 (see Section 2.5.2). Let $r \geq 1$, then $H_i \equiv 0 \pmod{2^r}$ (resp. $H_i \equiv 0$

(mod 3^r) for large enough i , thus if m is a power of 2 or 3, $N_H(m)$ is finite. By taking the initial conditions 1, 13, 163, the language of greedy H -representations is regular. Indeed, this choice of initial conditions corresponds to the canonical system associated to the dominant root of the polynomial $X^3 - 12X^2 - 6X - 12$, hence it is a Bertrand numeration system (see Section 1.3 and particularly Definition 1.3.11).

2.3 Some useful lemmas

The following lemma is a simple consequence of the minimality of the period chosen to represent an ultimately periodic set, see also [43].

Lemma 2.3.1. *Let X be an ultimately periodic set of period π and let i, j be integers greater than or equal to the preperiod of X . If $i \not\equiv j \pmod{\pi}$, then there is $r < \pi$ such that either $i + r \in X$ and $j + r \notin X$, or $i + r \notin X$ and $j + r \in X$.*

Proof. Suppose that $i < j$ (we proceed similarly for the other case) and putting aside the trivial case $\pi = 1$, suppose $\pi > 1$. Proceed by contradiction: suppose that for all $t \in \llbracket 0, \pi-1 \rrbracket$, $i + t \in X \Leftrightarrow j + t \in X$. Let $p \in \llbracket 0, \pi-1 \rrbracket$ be such that $p \equiv j - i \pmod{\pi}$. Let $n \in \mathbb{N}$ be such that $n \geq i$ (so that n is greater than the preperiod of X too). Then $\exists r \in \llbracket 0, \pi-1 \rrbracket$ such that $n \equiv i + r \pmod{\pi}$. Thus $n + p \equiv i + r + p \equiv j + r \pmod{\pi}$. Then, $n + p \in X \Leftrightarrow j + r \in X \Leftrightarrow i + r \in X \Leftrightarrow n \in X$, but π is minimal with this property, hence a contradiction. \square

Our assumption (H2) allows us to extend greedy U -representations with some extra leading digits.

Lemma 2.3.2. *Let U be a numeration system satisfying (H2). Then for all $i \in \mathbb{N}$ and all $L \geq i$, there exists $\ell \geq L$ such that*

$$10^{\ell - |\text{rep}_U(t)|} \text{rep}_U(t), \quad t = 0, \dots, U_i - 1$$

are greedy U -representations. Otherwise stated, if the word w is a greedy U -representation, then there are arbitrary large r such that the word $10^r w$ is also a greedy U -representation.

Proof. Let $i \in \mathbb{N}, L \geq i$. We have $\text{rep}_U(U_i) = 10^i$. Hence $\text{rep}_U(U_i - 1)$ is the greatest word of length i with respect to the genealogical order. Since we suppose (H2), there exists $\ell \geq L$ such that

$$U_{\ell+1} - U_\ell > U_i - 1.$$

Thus the word $10^{\ell-i} \text{rep}_U(U_i-1)$ is the greedy U -representation of the integer $U_\ell + U_i - 1 < U_{\ell+1}$. Hence the conclusion. \square

When \mathbb{N} is U -recognizable, using a pumping-like argument, we can give an upper bound on the number of zeros to be inserted.

Lemma 2.3.3. *Let U be a numeration system satisfying (H1) and (H2). Then there is an integer constant $C > 0$ such that if w is a greedy U -representation, then for some $\ell < C$, $10^\ell w$ is also a greedy U -representation.*

Proof. By assumption (H1), there is a DFA, say with C states, accepting the numeration language $\text{rep}_U(\mathbb{N})$. Let w be a greedy U -representation. Then from Lemma 2.3.2, there is $r \geq C$ such that $10^r w \in \text{rep}_U(\mathbb{N})$. The path of label $10^r w$ starting from the initial state is accepting. Since $r \geq C$, a state is visited a least twice when reading the block 0^r . Thus there is an accepting path of label $10^\ell w$ with $\ell < C$. \square

Let us introduce a constant Z .

Lemma 2.3.4. *Let U be a numeration system satisfying (H2) and (H3). Then there exists $R \geq N$ (where N is the constant given in (H3)) such that*

$$U_{R+1} - U_R \geq U_{i+1} - U_i$$

for all $i \leq R$.

Definition 2.3.5. We set

$$Z = \max\{R, C\}$$

where C is the constant given in Lemma 2.3.3 and R in Lemma 2.3.4.

Note that all these constants can be effectively computed. Indeed, C can be deduced from the automaton accepting the language of the numeration. Then, assuming that N is given in input with the numeration system, R can be computed by an exhaustive search and finally, one has to choose $Z = \max\{R, C\}$.

Thanks to (H3), we can add many zeros to greedy representations and obtain new greedy representations.

Lemma 2.3.6. *Let U be a numeration system satisfying (H1), (H2) and (H3). If w is a greedy U -representation, then for all $z \geq Z$, $10^z w$ is also a greedy U -representation.*

Proof. Let w be a greedy U -representation. By Lemma 2.3.3, there is $\ell < C$ such that $10^\ell w$ is a greedy U -representation. Let $i = \ell + |w|$. Let $n = \text{val}_U(w)$. We have $U_i + n < U_{i+1}$.

- If $i \geq Z$, then

$$\begin{aligned} U_{i+1} + n &= U_{i+1} - U_i + U_i + n \\ &\leq U_{i+2} - U_{i+1} + U_i + n \\ &< U_{i+2}. \end{aligned}$$

Hence $U_j + n < U_{j+1}$ for all $j \geq i$. Otherwise stated, $10^{\ell'} w$ is a greedy U -representation for all $\ell' \geq \ell$. In particular, since $\ell < Z$, for all $z \geq Z$, $10^z w$ is a greedy U -representation.

- If $i < Z$, then

$$\begin{aligned} U_Z + n &= U_Z - U_i + U_i + n \\ &< U_Z - U_i + U_{i+1} \\ &\leq U_{Z+1} - U_{i+1} + U_{i+1} \\ &\leq U_{Z+1}. \end{aligned}$$

Hence $10^{Z-|w|} w$ is a greedy U -representation. We conclude by applying the first part of the proof. □

Example 2.3.7. The sequence $1, 2, 4, 5, 16, 17, 64, 65, \dots$ is a solution of the linear recurrence $U_{i+4} = 5U_{i+2} - 4U_i$, but it does not satisfies (H3). The property stated in Lemma 2.3.6 does not hold: only some shifts to the left of leading coefficient 1 lead to valid greedy U -representations. The word 1001 is the greedy expansion of 6, but for all $t \geq 1$, the word $1(00)^t 1001$ is not a greedy representation.

Example 2.3.8. The sequence $1, 2, 3, 4, 8, 12, 16, 32, 48, 64, 128, \dots$ is a solution of the linear recurrence $U_{i+3} = 4U_i$. The associated numeration language $0^* \text{rep}_U(\mathbb{N})$ is the set of suffixes of $\{000, 001, 010, 100\}^*$, hence (H1) holds. For all $i \in \mathbb{N}$, $U_{i+1} - U_i = 4^{\lfloor i/3 \rfloor}$. Therefore, (H2) and (H3) are also verified.

Under assumption (H1), the formal series $\sum_{i \in \mathbb{N}} U_i X^i$ is \mathbb{R}_+ -rational. Indeed, U_i is the number of words of length less than or equal to i in the regular language $\text{rep}_U(\mathbb{N})$. To the automaton accepting $\text{rep}_U(\mathbb{N})$, say with m states numbered from 1 to m , one can associate an adjacency matrix $A \in \mathbb{N}^{m \times m}$, a

row vector $X \in \mathbb{N}^{1 \times m}$ where the k^{th} element of X is 1 if the state k is initial and 0 otherwise, and one can associate a column vector $Y \in \mathbb{N}^{m \times 1}$ where the k^{th} element of Y is 1 if the state k is final and 0 otherwise. Thus, if V_i is the number of words of length i in $\text{rep}_U(\mathbb{N})$, since $\mathbb{N} \subseteq \mathbb{R}_+$, the series $\sum_{i \in \mathbb{N}} V_i X^i$ is \mathbb{R}_+ -recognizable, hence \mathbb{R}_+ -rational, by Theorem 1.5.7. To conclude, it suffices to compute the product of this latter series with the series $\sum_{i=0}^{+\infty} X^i$.

One can therefore make use of Soittola's theorem 1.5.5. We thus define the following quantities.

Definition 2.3.9. We introduce an integer u and a real number β depending only on the numeration system. From Soittola's theorem, there are an integer $u \geq 1$, real numbers $\beta_0, \dots, \beta_{u-1} \geq 1$, a positive integer I_1 and non-zero polynomials P_0, \dots, P_{u-1} such that for $r \in \llbracket 0, u-1 \rrbracket$ and $i \geq I_1$,

$$U_{ui+r} = P_r(i)\beta_r^i + Q_r(i)$$

where $\frac{Q_r(i)}{\beta_r^i} \rightarrow 0$ when $i \rightarrow +\infty$. Since $(U_i)_{i \in \mathbb{N}}$ is increasing, for $r < s < u$ and for all $i \geq I_1$, we have

$$U_{ui+r} < U_{ui+s} < U_{u(i+1)+r}.$$

Thus, dividing each side of the first inequality by β_r^i , we obtain

$$P_r(i) + \frac{Q_r(i)}{\beta_r^i} < P_s(i) \frac{\beta_s^i}{\beta_r^i} + \frac{Q_s(i)}{\beta_r^i}.$$

If $\beta_r > \beta_s$, then letting i go to infinity gives

$$\frac{Q_r(i)}{\beta_r^i} \rightarrow 0, \quad \frac{\beta_s^i}{\beta_r^i} \rightarrow 0 \quad \text{and} \quad \frac{Q_s(i)}{\beta_r^i} \rightarrow 0,$$

which is impossible since P_r is a non-zero polynomial with positive dominant term. Hence $\beta_r \leq \beta_s$. In the same way, the second inequality shows that $\beta_s \leq \beta_r$. We conclude that we must have $\beta_0 = \dots = \beta_{u-1}$, that we denote by β . In a similar way, we can show that $\deg(P_0) = \dots = \deg(P_{u-1})$, that we denote d . Otherwise stated, for $r \in \llbracket 0, u-1 \rrbracket$, $U_{ui+r} \sim c_r i^d \beta^i$ for some constant c_r . Let T be such that $c_T = \max_{0 \leq r < u} c_r$. In other words, we highlight with T a subsequence $(U_{ui+T})_{i \in \mathbb{N}}$ with the maximal dominant coefficient.

Since $(U_i)_{i \in \mathbb{N}}$ is increasing and $\frac{Q_T(i)}{\beta^i} \rightarrow 0$ when $i \rightarrow +\infty$, there is $I_2 > 0$ such that $P_T(i) > 0$, for all $i \geq I_2$. Moreover, there is $I_3 > 0$ such that P_T is non-decreasing "after I_3 ". Finally, let I be the positive integer $\max\{I_1, I_2, I_3\}$.

Note that if a numeration system has a dominant root, i.e. the minimal recurrence relation satisfied by $(U_i)_{i \in \mathbb{N}}$ has a unique root $\beta > 1$, possibly with multiplicity greater than 1, of maximum modulus, then $u = 1$.

Lemma 2.3.10. *With the notation of Definition 2.3.9, if $\beta > 1$, then there are non-negative constants K and L such that for all $n \in \mathbb{N}$,*

$$|\text{rep}_U(n)| < u \log_\beta(n) + K$$

and

$$|\text{rep}_U(n)| > u \log_\beta(n) - u \log_\beta(P_T(\log_\beta(n) + K/u)) - L.$$

This lemma shows that the length of the greedy U -representation of n grows at most like $u \log_\beta(n)$. If P_T is a constant polynomial, the lower bound is of the form $u \log_\beta(n) + L'$ for some constant L' . From the result, we may express the weaker information (on ratios instead of differences) that $|\text{rep}_U(n)| \sim u \log_\beta(n)$.

Proof. We have $|\text{rep}_U(n)| = \ell$ if and only if $U_{\ell-1} \leq n < U_\ell$. We make use of Definition 2.3.9 for u, β, T and I . Let $j = \lfloor \frac{\ell-1-T}{u} \rfloor$. Suppose that n is large enough so that $j \geq I$. Since U is increasing and $j \geq I$,

$$U_{\ell-1} \geq U_{ju+T} = P_T(j)\beta^j + Q_T(j).$$

We get

$$\log_\beta(n) \geq \log_\beta(U_{\ell-1}) \geq j + \log_\beta(P_T(j)) + \log_\beta \left(1 + \frac{Q_T(j)}{P_T(j)\beta^j} \right).$$

Note that, up to an increasing of n , we can suppose that $1 + \frac{Q_T(j)}{P_T(j)\beta^j} > 0$ (since $\frac{Q_T(i)}{\beta^i} \rightarrow 0$ when $i \rightarrow +\infty$ and P_T is non-decreasing after I), so that the last logarithm in the above inequality is well-defined.

Hence

$$j \leq \log_\beta(n) - \log_\beta(P_T(j)) - \log_\beta \left(1 + \frac{Q_T(j)}{P_T(j)\beta^j} \right).$$

Moreover, $j > \frac{\ell-1-T}{u} - 1 \geq \frac{\ell-u}{u} - 1 \geq \frac{\ell}{u} - 2$. We obtain

$$\ell < u(j+2) \leq u \log_\beta(n) + 2u - u \log_\beta(P_T(j)) - u \log_\beta \left(1 + \frac{Q_T(j)}{P_T(j)\beta^j} \right).$$

Since $j \geq I$ and P_T is non-decreasing after I , we get

$$\ell < u(j+2) \leq u \log_\beta(n) + 2u - u \log_\beta(P_T(I)) - u \log_\beta \left(1 + \frac{Q_T(j)}{P_T(j)\beta^j} \right).$$

Finally, since $\frac{Q_T(i)}{\beta^i} \rightarrow 0$ when $i \rightarrow +\infty$, there is a constant $K \geq 0$ such that

$$\ell < u(j+2) \leq u \log_\beta(n) + K.$$

We supposed that n is large enough so that $j \geq I$ and $1 + \frac{Q_T(j)}{P_T(j)\beta^j} > 0$. Note that there is only a finite number of integers not fulfilling these conditions. Hence, possibly increasing the value of the constant K , we can assume that the above inequality is satisfied for any integer n .

We proceed similarly to get a lower bound for ℓ . Let $k = \lfloor \frac{\ell-T}{u} \rfloor$. Observe that $j \leq k$, hence $k \geq I$. Since U is increasing, we have

$$U_\ell < U_{u(k+1)+T} = P_T(k+1)\beta^{k+1} + Q_T(k+1).$$

We obtain

$$\log_\beta(n) < \log_\beta(U_\ell) < k+1 + \log_\beta(P_T(k+1)) + \log_\beta\left(1 + \frac{Q_T(k+1)}{P_T(k+1)\beta^{k+1}}\right).$$

As in the first part of the proof, we can suppose that n is large enough to get $1 + \frac{Q_T(k+1)}{P_T(k+1)\beta^{k+1}} > 0$.

Observe that $k \leq j+1$. Hence, from the first part, we get

$$k+1 \leq j+2 \leq \log_\beta(n) + \frac{K}{u}.$$

Since $k \leq \frac{\ell-T}{u} \leq \frac{\ell}{u}$, we have

$$\ell \geq uk > u \log_\beta(n) - u - u \log_\beta(P_T(k+1)) - u \log_\beta\left(1 + \frac{Q_T(k+1)}{P_T(k+1)\beta^{k+1}}\right).$$

We have $k+1 > k \geq I$ and recall that P_T is non-decreasing after I , hence

$$P_T(k+1) \leq P_T\left(\log_\beta(n) + \frac{K}{u}\right).$$

Hence

$$\begin{aligned} \ell &> u \log_\beta(n) - u - u \log_\beta\left(P_T\left(\log_\beta(n) + \frac{K}{u}\right)\right) \\ &\quad - u \log_\beta\left(1 + \frac{Q_T(k+1)}{P_T(k+1)\beta^{k+1}}\right). \end{aligned}$$

Furthermore, since $\frac{Q_T(i)}{\beta^i} \rightarrow 0$ when $i \rightarrow +\infty$ and P_T is non-decreasing after I , there is a constant $L \geq 0$ such that

$$\ell > u \log_\beta(n) - u \log_\beta\left(P_T\left(\log_\beta(n) + \frac{K}{u}\right)\right) - L.$$

As in the first part of the proof, we only considered the n such that $j \geq I$ and $1 + \frac{Q_T(k+1)}{P_T(k+1)\beta^{k+1}} > 0$. Possibly increasing the value of L , we can assume that the above inequality is satisfied for all integers n . \square

Example 2.3.11. Consider the sequence of integers $1, 2, 6, 12, 36, 72, \dots$ defined by $U_0 = 1$, $U_{2i+1} = 2U_{2i}$ and $U_{2i+2} = 3U_{2i+1}$. Then for all $i \in \mathbb{N}$, $U_{i+2} = 6U_i$. It is easily seen that $U_{2i} = 6^i$ and $U_{2i+1} = 2 \cdot 6^i$. With the notation of Definition 2.3.9, $u = 2, \beta = 6, d = 0$ and $P_T = c_T = 2$. The language $0^* \text{rep}_U(\mathbb{N})$ is made of words where in even (resp. odd) positions when reading from right to left (i.e. with LSDF convention), we can write $0, 1$ (resp. $0, 1, 2$). If $|\text{rep}_U(n)| = 2\ell + 1$, then $U_{2\ell} = 6^\ell \leq n < U_{2\ell+1} = 2 \cdot 6^\ell$, thus on the one hand $|\text{rep}_U(n)| \leq 2\log_6(n) + 1$ and on the other hand $|\text{rep}_U(n)| > 2\log_6(\frac{n}{2}) + 1 = 2\log_6(n) - 2\log_6(2) + 1$. If $|\text{rep}_U(n)| = 2\ell$, then $U_{2\ell-1} = 2 \cdot 6^{\ell-1} \leq n < U_{2\ell} = 6^\ell$, hence $6^\ell \leq 6\frac{n}{2}$, which implies $|\text{rep}_U(n)| \leq 2\log_6(3n) = 2\log_6(n) + 2\log_6(3)$ and $|\text{rep}_U(n)| > 2\log_6(n)$.

Example 2.3.12. Consider the sequence $1, 3, 8, 20, 48, 112, \dots$ defined by $U_0 = 1$, $U_1 = 3$ and $U_{i+2} = 4U_{i+1} - 4U_i$. Then $U_i = (\frac{i}{2} + 1)2^i$. With the notation of Definition 2.3.9, $u = 1, \beta = 2, d = 1$ and $P_T(X) = \frac{X}{2} + 1$. If $|\text{rep}_U(n)| = \ell$ then $U_{\ell-1} = (\frac{\ell-1}{2} + 1)2^{\ell-1} \leq n < U_\ell = (\frac{\ell}{2} + 1)2^\ell$, thus $|\text{rep}_U(n)| < \log_2(n) + 1$. Indeed, $\ell \leq \log_2(n) + 1 - \log_2(\frac{\ell-1}{2} + 1)$, and $\log_2(\frac{\ell-1}{2} + 1) \geq 0 \Leftrightarrow \ell \geq 1$. Moreover,

$$|\text{rep}_U(n)| > \log_2(n) - \log_2(\frac{\ell}{2} + 1) > \log_2(n) - \log_2(\frac{1}{2}\log_2(n) + \frac{3}{2}).$$

We get $K = 1$ and $P_T(\log_2(n) + K/u) = \frac{1}{2}\log_2(n) + \frac{3}{2}$, with the notation of Lemma 2.3.10.

As shown by the next result, it is enough to obtain a bound on the admissible period of X . In [7, Proposition 41], the result is given for abstract numeration systems. We restate it in our context.

Proposition 2.3.13. *Let U be a numeration system satisfying (H1), let X be an ultimately periodic set of non-negative integers and let \mathcal{A}_X be a DFA with $\#Q_X$ states accepting $\text{rep}_U(X)$. Then the preperiod α_X of X is bounded by a computable constant J depending only on the number of states of \mathcal{A}_X and the period π_X of X .*

Proof. Let \mathcal{B} be the minimal automaton of the numeration language and denote by Q its set of states.

Let $U = (U_i)_{i \in \mathbb{N}}$. By Proposition 1.3.31, the sequence $(U_i)_{i \in \mathbb{N}}$ satisfies a linear recurrent equation. Hence the sequence $(U_i \bmod \pi_X)_{i \in \mathbb{N}}$ is ultimately periodic. Let a (resp. p) denote its preperiod (resp. its period).

For α_X large enough, one has $|\text{rep}_U(\alpha_X - 1)| > \#Q_X \cdot \#Q$. Suppose that $\alpha_X - 1 \in X$. Then by the pumping lemma (Proposition 1.2.15) applied to the product automaton $\mathcal{A}_X \times \mathcal{B}$, there are x, y, z such that $\text{rep}_U(\alpha_X - 1) = xyz$, $y \neq \varepsilon$, $|xy| \leq \#Q_X \cdot \#Q$ and $xy^n z \in \text{rep}_U(X)$ for all $n \geq 0$. Now if $\alpha_X - 1 \notin X$, $\text{rep}_U(\alpha_X - 1)$ is accepted by the automaton $\overline{\mathcal{A}_X \times \mathcal{B}}$, which is the automaton $\mathcal{A}_X \times \mathcal{B}$ where the status final/non-final of every state has been exchanged. By the pumping lemma, there exist words x, y, z such that $\text{rep}_U(\alpha_X - 1) = xyz$, $y \neq \varepsilon$, $|xy| \leq \#Q_X \cdot \#Q$ and $xy^n z \notin \text{rep}_U(X)$ for all $n \geq 0$. In both cases, $|xy|$ is bounded by a constant, thus one has $|z| > a$ if α_X is large enough.

In both cases, since $|z| > a$ and since the sequence $(U_i \bmod \pi_X)_{i \in \mathbb{N}}$ is ultimately periodic of preperiod a and period p , one has for all $\ell \geq 0$

$$\text{val}_U(xy^{\ell \pi_X p} yz) \equiv \text{val}_U(xyz) \pmod{\pi_X}.$$

Now, we use the minimality of α_X to obtain a contradiction. First, suppose that $\alpha_X - 1 \in X$. Then by minimality of α_X , one has $\alpha_X - 1 + n\pi_X \notin X$ for all $n \geq 1$. Then for $\ell > 0$

$$xy^{\ell \pi_X p} yz \in \text{rep}_U(X)$$

by the pumping lemma. But this word represents an integer of the form $\alpha_X - 1 + n\pi_X$ with $n > 0$, which cannot belong to X .

Secondly, suppose that $\alpha_X - 1 \notin X$. Then by minimality of α_X , one has $\alpha_X - 1 + n\pi_X \in X$ for all $n \geq 1$. Then for $\ell > 0$

$$xy^{\ell \pi_X p} yz \notin \text{rep}_U(X)$$

by the pumping lemma. But this word represents an integer of the form $\alpha_X - 1 + n\pi_X$ with $n > 0$, which belongs to X .

Remark that J can be effectively computed as follows. The constant J must be chosen such that $\alpha_X > J$ implies $|\text{rep}_U(\alpha_X - 1)| - \#Q_X \cdot \#Q > a$. Since the numeration system U , the period π_X and the number of states of $\#Q_X$ are given, a and $\text{rep}_U(n)$ for all $n \geq 0$ can be effectively computed. \square

2.4 Number of states

We follow Honkala's strategy introduced in [43]. A DFA \mathcal{A}_X accepting $\text{rep}_U(X)$ is given as input. Assuming that X is ultimately periodic, the number of states of \mathcal{A}_X should provide a computable upper bound on the possible period and preperiod of X . This should leave us with a finite number

of candidates to test. Thanks to Proposition 1.3.33, one therefore builds a DFA for each pair of admissible preperiod and period. Equality of regular languages being decidable, we compare the language accepted by this DFA and the one accepted by \mathcal{A}_X . If an agreement is found, then X is ultimately periodic, otherwise it is not. Thanks to Proposition 2.3.13, we only focus on the admissible periods.

Recall Equation 2.1:

$$U_{i+k} = a_{k-1}U_{i+k-1} + \cdots + a_0U_i.$$

Assume that the minimal automaton \mathcal{A}_X of $\text{rep}_U(X)$ is given. Let π_X be a potential period for X . We consider the prime decomposition of π_X . There are three types of prime factors:

- (P1) those that do not divide a_0 ,
- (P2) those that divide a_0 but that do not simultaneously divide all the coefficients of the recurrence relation,
- (P3) the remaining ones are the primes dividing all the coefficients of the recurrence relation.

Our strategy is to bound these three types of factors separately.

2.4.1 Factors of the period that are coprime with a_0

The next result shows that, given \mathcal{A}_X , the possible period cannot have a large factor coprime with a_0 : it provides a bound on this kind of factor that may occur in a candidate period.

Proposition 2.4.1. *Assume (H1), (H2) and (H3). Let $X \subseteq \mathbb{N}$ be an ultimately periodic U -recognizable set and let q be a divisor of the period π_X such that $\gcd(q, a_0) = 1$. Then the minimal automaton of $\text{rep}_U(X)$ has at least q states.*

Proof. Since $\gcd(q, a_0) = 1$, the sequence $(U_i \bmod q)_{i \in \mathbb{N}}$ is purely periodic. In particular, 1 occurs infinitely often in this sequence.

Let us define q integers $k_1, \dots, k_q \geq 0$ and q words $w_1, \dots, w_q \in \{0, 1\}^*$ of the following form

$$w_j = 10^{k_j} 10^{k_{j-1}} \dots 10^{k_1} 0^{|\text{rep}_U(\pi_X)|}.$$

Thanks to Lemmas 2.3.2 and 2.3.6, we may impose the following conditions.

- First, k_1 is taken large enough to ensure that $\text{val}_U(w_1)$ is larger than the preperiod of X and $10^{k_1} \text{rep}_U(\pi_X)$ is a valid greedy U -representation.
- Secondly, k_2, \dots, k_q are taken large enough to ensure that $w_j \in \text{rep}_U(\mathbb{N})$ for all j .
- Thirdly, we can choose k_1, \dots, k_q so that the 1's occur at indices m such that $U_m \equiv 1 \pmod{q}$.

Observe that $\text{val}_U(w_j) \equiv j \pmod{q}$. Since q divides π_X , the words w_1, \dots, w_q have pairwise distinct values modulo π_X .

Let $i, j \in \{1, \dots, q\}$ such that $i \neq j$. By Lemma 2.3.1, we can assume that there exists $r_{i,j} < \pi_X$ such that $\text{val}_U(w_i) + r_{i,j} \in X$ and $\text{val}_U(w_j) + r_{i,j} \notin X$ (the symmetric situation can be handled similarly). In particular, one has $|\text{rep}_U(r_{i,j})| \leq |\text{rep}_U(\pi_X)|$. Consider the two words

$$w_i 0^{-|\text{rep}_U(r_{i,j})|} \text{rep}_U(r_{i,j}) \quad \text{and} \quad w_j 0^{-|\text{rep}_U(r_{i,j})|} \text{rep}_U(r_{i,j})$$

where, in the above notation, it should be understood that we replace the rightmost zeros in w_i and w_j by $\text{rep}_U(r_{i,j})$. The first word belongs to $\text{rep}_U(X)$ and the second does not. Consequently, the number of states of the minimal automaton of $\text{rep}_U(X)$ is at least q : w_1, \dots, w_q belong to pairwise distinct Nerode equivalence classes. \square

2.4.2 Prime factors of the period that divide a_0 but do not divide all the coefficients of the recurrence relation

We depart from the strategy developed in [7] and now turn to a particular situation where a prime factor p of the candidate period for X is such that, for some integer $\mu \geq 1$, the sequence $(U_i \bmod p^\mu)_{i \in \mathbb{N}}$ has a period containing a non-zero element. Again, this will provide us with an upper bound on p and its exponent in the prime decomposition of the period.

Definition 2.4.2. We say that an ultimately periodic sequence has a *zero period* if it has period 1 and the repeated element is 0. Otherwise stated, the sequence has a tail of zeros.

Remark 2.4.3. Let $\mu \geq 1$. Note that if the periodic part of $(U_i \bmod p^\mu)_{i \in \mathbb{N}}$ contains a non-zero element, then the same property holds for the sequence $(U_i \bmod p^{\mu'})_{i \in \mathbb{N}}$ with $\mu' \geq \mu$.

Furthermore, assume that for infinitely many μ , $(U_i \bmod p^\mu)_{i \in \mathbb{N}}$ has a zero period. Then from the previous paragraph, we conclude that $(U_i \bmod p^\mu)_{i \in \mathbb{N}}$ has a zero period for all $\mu \geq 1$.

Example 2.4.4. We give a sequence where only finitely many sequences modulo p^μ have a zero period. Take the sequence $U_0 = 1, U_1 = 4, U_2 = 8$ and $U_{i+2} = U_{i+1} + U_i$ for $i \in \mathbb{N}_0$, then the sequence $(U_i \bmod 2^\mu)_{i \geq 0}$ has a zero period for $\mu = 1, 2$ because of the particular initial conditions. But it is easily checked that it has a non-zero period for all $\mu \geq 3$. Indeed, the sequence $(U_i \bmod 8)_{i \in \mathbb{N}}$ is given by $1(404)^\omega$.

The next result is a special instance of [7, Theorem 30] and its proof turns out to be much simpler. It precisely describes the case where a zero period occurs infinitely often.

Theorem 2.4.5. *Let p be a prime. The sequence $(U_i \bmod p^\mu)_{i \in \mathbb{N}}$ has a zero period for all $\mu \geq 1$ if and only if all the coefficients a_0, \dots, a_{k-1} of the linear relation (2.1) are divisible by p .*

Proof. It is clear that if a_0, \dots, a_{k-1} are divisible by p , then for any choice of initial conditions, U_k, \dots, U_{2k-1} are divisible by p , hence U_{2k}, \dots, U_{3k-1} are divisible by p^2 , and so on. Otherwise stated, for all $\mu \geq 1$ and all $i \geq \mu k$, U_i is divisible by p^μ .

We turn to the converse. Since the sequence $(U_i)_{i \geq 0}$ is linearly recurrent, the power series

$$U(x) = \sum_{i \geq 0} U_i x^i$$

is rational. By assumption, $(U_i \bmod p^\mu)_{i \geq 0}$ has a zero period for all $\mu \geq 1$. Otherwise stated, with the p -adic absolute value notation, $|U_i|_p \leq p^{-\mu}$ for large enough i , i.e. $|U_i|_p \rightarrow 0$ as $i \rightarrow +\infty$. Applying Remark 1.7.8, the series $U(x)$ converges in \mathbb{Q}_p in the closed unit disc. Therefore, the poles $\rho_1, \dots, \rho_r \in \mathbb{C}_p$ of $U(x)$ must satisfy $|\rho_j|_p > 1$ for $1 \leq j \leq r$.

Let $P(x) = 1 - a_{k-1}x - \dots - a_0x^k$ be the reciprocal polynomial of the linear recurrence relation (2.1). By minimality of the order k of the recurrence, the roots of P are precisely the poles of $U(x)$ with the same multiplicities. If we factor

$$P(x) = (1 - \delta_1 x) \cdots (1 - \delta_k x),$$

each of the δ_j is one of the $\frac{1}{\rho_1}, \dots, \frac{1}{\rho_r}$. For $n > 0$, the coefficient of x^n is an integer equal to a sum of product of elements of p -adic absolute value less than 1. Since $|m + n|_p \leq \max\{|m|_p, |n|_p\}$ (see Proposition 1.7.5), this coefficient is an integer with a p -adic absolute value less than 1, i.e. a multiple of p . \square

Thanks to Remark 2.4.3 and Theorem 2.4.5, if p is a prime not dividing all the coefficients of the recurrence relation (2.1) then there is a least integer

λ (depending only on p) such that $(U_i \bmod p^\lambda)_{i \in \mathbb{N}}$ has a period containing a non-zero element.

Proposition 2.4.6. *Assume (H1), (H2) and (H3). Let p be a prime not dividing all the coefficients of the recurrence relation (2.1) and let $\lambda \geq 1$ be the least integer such that $(U_i \bmod p^\lambda)_{i \in \mathbb{N}}$ has a period containing a non-zero element. If $X \subseteq \mathbb{N}$ is an ultimately periodic U -recognizable set with period $\pi_X = p^\mu \cdot r$ where $\mu \geq \lambda$ and r is not divisible by p , then the minimal automaton of $\text{rep}_U(X)$ has at least $p^{\mu-\lambda+1}$ states.*

Proof. We will make use of the following observation. Let $n \geq 1$. In the additive group $(\mathbb{Z}/p^n\mathbb{Z}, +)$, an integer a has order p^s with $0 \leq s \leq n$ if and only if $a = p^{n-s} \cdot m$ where m is not divisible by p .

By assumption, $(U_i \bmod p^\lambda)_{i \in \mathbb{N}}$ has a period containing a non-zero element R of order $\text{ord}_{p^\lambda}(R) = p^\theta$ for some θ such that $0 < \theta \leq \lambda$. Consider a large enough index K such that U_K is in the periodic part of $(U_i \bmod p^\mu)_{i \in \mathbb{N}}$ and $U_K \equiv R \pmod{p^\lambda}$. Using the above observation twice, one has first $U_K \equiv m \cdot p^{\lambda-\theta} \pmod{p^\lambda}$ for some m coprime with p and therefore, U_K has order $\text{ord}_{p^\mu}(U_K) = p^{\mu-\lambda+\theta}$ modulo p^μ .

We can again apply the same construction as in the proof of Proposition 2.4.1. We define words of the form

$$w_j = 10^{k_j} 10^{k_{j-1}} \dots 10^{k_1} 0^{|\text{rep}_U(\pi_X)|}$$

with the same properties, except for the second one: the 1's occur at indices t such that $U_t \equiv U_K \pmod{p^\mu}$. Note that

$$\text{val}_U(w_j) \equiv j \cdot U_K \pmod{p^\mu}.$$

Hence the number of distinct numerical values modulo p^μ that are taken by those words is given by the order of U_K in $\mathbb{Z}/p^\mu\mathbb{Z}$, i.e. $p^{\mu-\lambda+\theta}$. Since $\theta \geq 1$, the words $w_1, \dots, w_{p^{\mu-\lambda+1}}$ have pairwise distinct values modulo p^μ . Then they also have pairwise distinct values modulo π_X .

Let $i, j \in \{1, \dots, p^{\mu-\lambda+1}\}$ such that $i \neq j$. As in the proof of Proposition 2.4.1, by Lemma 2.3.1, we can suppose that there is $r_{i,j} < \pi_X$ such that $\text{val}_U(w_i) + r_{i,j} \in X$ and $\text{val}_U(w_j) + r_{i,j} \notin X$ (the other case is handled in like manner). In particular, $|\text{rep}_U(r_{i,j})| \leq |\text{rep}_U(\pi_X)|$. Consider the two words

$$w_i 0^{-|\text{rep}_U(r_{i,j})|} \text{rep}_U(r_{i,j}) \quad \text{and} \quad w_j 0^{-|\text{rep}_U(r_{i,j})|} \text{rep}_U(r_{i,j})$$

where the above notation means that we replace the rightmost zeros in w_i and w_j by $\text{rep}_U(r_{i,j})$. The first word belongs to $\text{rep}_U(X)$ and the second does not. Consequently, the number of states of the minimal automaton of $\text{rep}_U(X)$ is at least $p^{\mu-\lambda+1}$: $w_1, \dots, w_{p^{\mu-\lambda+1}}$ belong to pairwise distinct Nerode equivalence classes. \square

2.4.3 Prime factors of the period that divide all the coefficients of the recurrence relation

We can factor the period π_X as

$$\pi_X = m_X \cdot p_1^{\mu_1} \cdots p_t^{\mu_t} \quad (2.2)$$

where every p_j divides all the coefficients of the recurrence relation (2.1) (factors of type (P3)) and for every prime factor q of m_X , at least one of the coefficients of the recurrence relation (2.1) is not divisible by q . Otherwise stated, the factor m_X collects the prime factor of types (P1) and (P2).

Remark 2.4.7. There is a finite number of primes dividing all the coefficients of the recurrence relation. Thus, we only have to obtain an upper bound on the corresponding exponents μ_1, \dots, μ_t that may appear in (2.2).

Lemma 2.4.8. [43, Lemma 6] *Let X be an ultimately periodic set with period (2.2). Then there exists $r \in \llbracket 0, m_X - 1 \rrbracket$ such that $X \cap (m_X \mathbb{N} + r)$ is ultimately periodic of period $\pi_\nu = m_X \cdot p_1^{\nu_1} \cdots p_t^{\nu_t}$ with $\max_{i \in \llbracket 1, t \rrbracket} \mu_i = \max_{i \in \llbracket 1, t \rrbracket} \nu_i$.*

Remark 2.4.9. The quantity r in the previous lemma is not necessarily unique. To avoid ambiguity, we always consider the smallest possible r and the associated exponents ν_1, \dots, ν_t .

Definition 2.4.10. Let $j \in \llbracket 1, t \rrbracket$ and $\mu \geq 1$. From Theorem 2.4.5, the sequence $(U_i \bmod p_j^\mu)_{i \in \mathbb{N}}$ has a zero period. We denote by $\mathbf{f}_{p_j}(\mu)$ the length of the preperiod, i.e. $U_{\mathbf{f}_{p_j}(\mu)-1} \not\equiv 0 \pmod{p_j^\mu}$ and $U_i \equiv 0 \pmod{p_j^\mu}$ for all $i \geq \mathbf{f}_{p_j}(\mu)$.

Example 2.4.11. Let us consider the numeration system from Example 2.1.2. The sequence $(G_i \bmod 2)_{i \in \mathbb{N}}$ is $1, 1, 1, 1, 0^\omega$. Hence $\mathbf{f}_2(1) = 4$. The sequence $(G_i \bmod 4)_{i \in \mathbb{N}}$ is $1, 3, 1, 3, 2, 0, 2, 2, 0^\omega$. Hence $\mathbf{f}_2(2) = 8$. Continuing this way, we have $\mathbf{f}_2(3) = 12$ and $\mathbf{f}_2(4) = 16$.

Note that \mathbf{f}_{p_j} is non-decreasing: $\mathbf{f}_{p_j}(\mu + 1) \geq \mathbf{f}_{p_j}(\mu)$.

Definition 2.4.12. We denote by F_X the maximum of the values $\mathbf{f}_{p_j}(\nu_j)$ for $j \in \llbracket 1, t \rrbracket$:

$$F_X = \max_{1 \leq j \leq t} \mathbf{f}_{p_j}(\nu_j).$$

Otherwise stated, F_X is the least index such that for all $i \geq F_X$ and all $j \in \llbracket 1, t \rrbracket$, $U_i \equiv 0 \pmod{p_j^{\nu_j}}$. By the Chinese remainder theorem, F_X is also the least index such that for all $i \geq F_X$, $U_i \equiv 0 \pmod{\frac{\pi_\nu}{m_X}}$.

Example 2.4.13. Consider the numeration system from Example 2.2.1. Here we have two prime factors 2 and 3 to take into account. Computations show that $f_2(1) = 3$, $f_2(2) = 5$, $f_2(3) = 7$ and $f_3(1) = 3$, $f_3(2) = 6$, $f_3(3) = 9$. Assume that we are interested in a period $\pi_\nu = 72 = 2^3 \cdot 3^2$. With the above definition, $F_X = \max(f_2(3), f_3(2)) = 7$. One can easily check that $(H_i \bmod 72)_{i \in \mathbb{N}}$ is 1, 13, 19, 30, 54, 48, 36, 0^ω .

We introduce a quantity γ_{m_X} which only depends on the numeration system U and the number m_X defined in (2.2). Since we are only interested in decidable issues, there is no need to find a sharp estimate on this quantity.

Definition 2.4.14. Under (H1), for each $s \in \llbracket 0, m-1 \rrbracket$, a DFA accepting the language $0^* \text{rep}_U(m\mathbb{N} + s)$ can be effectively built (see Proposition 1.3.33). We let γ_m denote the maximum of the number of states of these DFAs for $s \in \llbracket 0, m-1 \rrbracket$.

A crucial point in the proof of next statement is that a digit 1 occurs for U_{F_X-1} in a specific word we construct. The proof makes use of the same kind of arguments built for definite languages (Definition 1.1.11) as in [47, Lemma 2.1].

Theorem 2.4.15. *Assume (H1), (H2) and (H3). Let $X \subseteq \mathbb{N}$ be an ultimately periodic U -recognizable set with period π_X factored as in (2.2). Assume that $F_X - 1 - |\text{rep}_U(\pi_\nu - 1)| \geq Z$, where Z is the constant given in Definition 2.3.5. Also assume that F_X is greater than the preperiod of $(U_i \bmod m_X)_{i \in \mathbb{N}}$. Then there is a positive constant C such that the minimal automaton of $0^* \text{rep}_U(X)$ has at least $\frac{|\text{rep}_U(\pi_\nu - 1)| + 1}{\gamma_{m_X}}$ states.*

This result will provide us with an upper bound on μ_1, \dots, μ_t (details are given in Section 2.5.2). Since m_X has been bounded in the first part of this chapter, if $\max(\mu_1, \dots, \mu_t) \rightarrow \infty$, then $\pi_\nu \rightarrow \infty$ but therefore the number of states of the minimal automaton of $0^* \text{rep}_U(X)$ should increase.

Proof. We may apply Lemma 2.3.6 and consider the given positive constant Z : we will assume that if w is a greedy U -representation, then, for all $z \geq Z$, $10^z w$ also belongs to $\text{rep}_U(\mathbb{N})$.

Let r be the quantity of Lemma 2.4.8. Then the set $Y = X \cap (m_X \mathbb{N} + r)$ has period π_ν . Let \mathcal{A}_Y be the minimal automaton of $0^* \text{rep}_U(Y)$. We will provide a lower bound on the number of states of this automaton. By definition of F_X , we have $U_{F_X-1} \not\equiv 0 \pmod{\frac{\pi_\nu}{m_X}}$. Let g large enough so that

- $g \geq Z$

- U_{F_X+g} is larger than the preperiod of Y
- $g+1$ is a multiple of the period of $(U_i \pmod{m_X})_{i \in \mathbb{N}}$.

Consider

$$n_1 = \text{val}_U((10^g)^{m_X} 10^{F_X-1}) = \sum_{i=0}^{m_X} U_{F_X+(g+1)i-1}$$

$$n_2 = \text{val}_U(10^{F_X+g}) = U_{F_X+g}.$$

Observe that n_1 and n_2 are both congruent to $U_{F_X-1} \pmod{m_X}$. However, modulo $\frac{\pi_\nu}{m_X}$, n_1 is congruent to $U_{F_X-1} \not\equiv 0$ but n_2 is congruent to 0. Consequently, n_1 and n_2 are not congruent modulo π_ν . By Lemma 2.3.1 applied to the set Y , we may suppose that there exists $s < \pi_\nu$ such that

$$n_1 + s \in Y \quad \text{and} \quad n_2 + s \notin Y$$

(the symmetrical situation can be treated in the same way). Let us set $\ell_X = |\text{rep}_U(\pi_\nu-1)|$. Note that $|\text{rep}_U(s)| \leq \ell_X$ and then by assumption, $F_X - 1 - |\text{rep}_U(s)| \geq F_X - 1 - \ell_X \geq Z$. Thanks to Lemma 2.3.6, both words

$$u = (10^g)^{m_X} 10^{F_X-1-|\text{rep}_U(s)|} \text{rep}_U(s)$$

and

$$v = 10^g 00^{F_X-1-|\text{rep}_U(s)|} \text{rep}_U(s)$$

are greedy U -representations. For all $\ell \geq 0$, define an equivalence relation E_ℓ on the set of states of \mathcal{A}_Y :

$$E_\ell(q, q') \Leftrightarrow (\forall x \in \Sigma_U^*) [|x| \geq \ell \Rightarrow (\delta(q, x) \in \mathcal{F} \Leftrightarrow \delta(q', x) \in \mathcal{F})]$$

where δ (resp. \mathcal{F}) is the transition function (resp. the set of final states) of \mathcal{A}_Y . Let us denote the number of equivalence classes of E_ℓ by P_ℓ . Clearly, $E_\ell(q, q')$ implies $E_{\ell+1}(q, q')$, and thus $P_\ell \geq P_{\ell+1}$.

Let $i \in \llbracket 0, \ell_X \rrbracket$. By assumption, $\ell_X < F_X$. Since u and v have the same suffix of length F_X-1 , we can factorize these words as

$$u = u_i w_i \quad \text{and} \quad v = v_i w_i$$

where $|w_i| = i$. Let q_0 be the initial state of \mathcal{A}_Y . By construction, one has $\delta(q_0, u_i w_i) \in \mathcal{F}$ whereas $\delta(q_0, v_i w_i) \notin \mathcal{F}$, hence the states $\delta(q_0, u_i)$ and $\delta(q_0, v_i)$ are not in relation with respect to E_i . But for all $j > i$, they satisfy E_j . It is enough to show that

$$E_{i+1}(\delta(q_0, u_i), \delta(q_0, v_i)). \tag{2.3}$$

Figures 2.1 and 2.2 can help the reader. Let x be such that $|x| = i + t$, with $t \geq 1$. Let p be the prefix of $\text{rep}_U(s)$ of length $|\text{rep}_U(s)| - i$, this prefix p being empty whenever this difference is negative. If we replace w_i by x in u and v , we get

$$u_ix = (10^g)^{m_X} 10^{F_X-1-|px|+t} px \quad \text{and} \quad v_ix = 10^g 00^{F_X-1-|px|+t} px.$$

Then

$$\text{val}_U(u_ix) - \text{val}_U(v_ix) = U_{F_X+t-1} + \sum_{i=2}^{m_X} U_{F_X+(g+1)i+t-1}.$$

This quantity is congruent to 0 (mod m_X) and by definition of F_X , it is also congruent to 0 (mod $\frac{\pi_\nu}{m_X}$). Hence, $\text{val}_U(u_ix)$ and $\text{val}_U(v_ix)$ belong to the periodic part of Y and they differ by a multiple of the period π_ν . Therefore, $\text{val}_U(u_ix)$ belongs to Y if and only if $\text{val}_U(v_ix)$ also does.

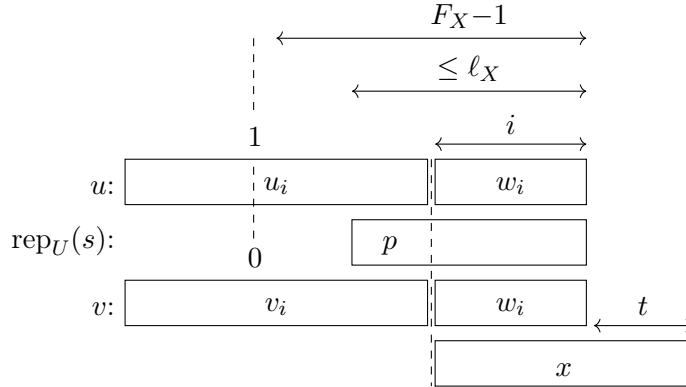


Figure 2.1: The different words (case where $i \leq |\text{rep}_U(s)|$).

In order to obtain (2.3), it remains to show that either both u_ix and v_ix are valid greedy U -representations or both are not. If the word px is not a greedy U -representation then neither u_ix nor v_ix can be valid. Assume now that px is a greedy U -representation. Note that in both situations described in Figures 2.1 and 2.2, $|px| \leq \ell_X + t$. Thanks to the assumption, we obtain $F_X - 1 - |px| + t \geq F_X - 1 - \ell_X \geq Z$. The greediness of px and Lemma 2.3.6 imply that $10^{F_X-1-|px|+t} px$ is a greedy U -representation. Since $g \geq Z$, u_ix is also a greedy U -representation and the same observation trivially holds for v_ix .

We conclude that

$$P_0 > P_1 > \cdots > P_{\ell_X} \geq 1.$$

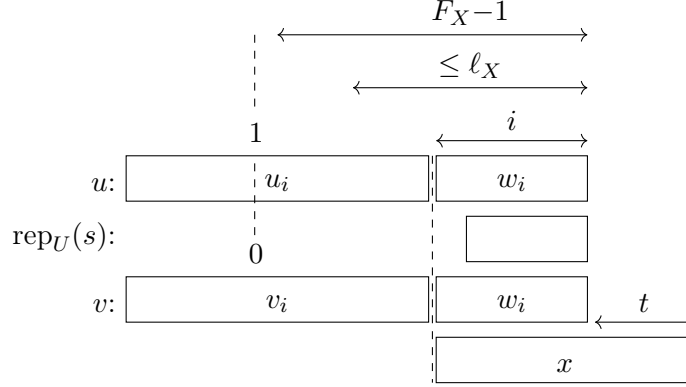


Figure 2.2: The different words (case where $i > |\text{rep}_U(s)|$).

Since P_0 is the number of states of \mathcal{A}_Y , the automaton \mathcal{A}_Y has at least $\ell_X + 1$ states.

Finally, denote by \mathcal{A}_X and \mathcal{A}_r the minimal automata of $0^* \text{rep}_U(X)$ and $0^* \text{rep}_U(m_X \mathbb{N} + r)$ respectively. The number of states of \mathcal{A}_r is bounded by γ_{m_X} . The DFA $\mathcal{A}_{X \cap (m_X \mathbb{N} + r)}$ is a quotient of the product automaton $\mathcal{A}_X \times \mathcal{A}_r$, hence the number of states of $\mathcal{A}_{X \cap (m_X \mathbb{N} + r)}$ is at most the number of states of \mathcal{A}_X times γ_{m_X} . We thus obtain that the number of states of \mathcal{A}_X is at least $\frac{\ell_X + 1}{\gamma_{m_X}}$. \square

2.5 Cases we can deal with

This section is divided in two parts. First, we solve the decision problem for numeration systems such that all the coefficients of the recurrence relation are coprime. This first case allows us to handle new systems for which no procedure was known. Secondly, we give a decision procedure for numeration systems such that the gcd of the coefficients of the recurrence relation is greater than 1 and with an extra hypothesis. Finally, we discuss this additional assumption.

2.5.1 The gcd of the coefficients of the recurrence relation is 1

In this case, for any ultimately periodic set X , the factorization of the period π_X given in (2.2) has the special form $\pi_X = m_X$ and the addressed decision problem turns out to be decidable.

Theorem 2.5.1. *Let U be a linear numeration system satisfying (H1), (H2) and (H3) and such that the gcd of the coefficients of the recurrence relation (2.1) is 1. Given a DFA \mathcal{A} accepting a language contained in the numeration language $\text{rep}_U(\mathbb{N})$, it is decidable whether this DFA recognizes an ultimately periodic set.*

Proof. Assume that X is an ultimately periodic set with period π_X . Let p be a prime that divides π_X . Either p divides the last coefficient of the recurrence relation a_0 , or it does not.

In the latter case, thanks to Proposition 2.4.1, for any $n \geq 1$, if p^n divides π_X then p^n is bounded by the number of states of \mathcal{A} .

In the former case, p divides a_0 . Note that there is only a finite number of such primes. By assumption, p does not divide all the coefficients of the recurrence relation. Then thanks to Theorem 2.4.5, there is $\mu \geq 1$ such that the periodic part of the sequence $(U_i \bmod p^\mu)_{i \in \mathbb{N}}$ contains a non-zero element. Let λ be the least such μ . By an exhaustive search, one can determine the value of λ : one finds the period of a sequence $(U_i \bmod p^\mu)_{i \in \mathbb{N}}$ as soon as two k -tuples $(U_i \bmod p^\mu, \dots, U_{i+k-1} \bmod p^\mu)$ are identical (remind that k is the order of the recurrence). We then apply Proposition 2.4.6. For any $n \geq 1$, if p^n divides π_X then either $n < \lambda$ or $p^{n-\lambda+1}$ is bounded by the number of states of \mathcal{A} .

The previous discussion provides us with an upper bound on π_X , i.e. on the admissible periods for X . Then from Proposition 2.3.13, associated with each admissible period, there is a computable bound for the corresponding admissible preperiods for X . We conclude that there is a finite number of pairs of candidates for the preperiod and period of X . Similar to Honkala's scheme, we therefore have a decision procedure by enumerating a finite number of candidates. For each pair (α, π) of possible preperiods and periods, there are $2^{\alpha 2^\pi}$ corresponding ultimately periodic sets X . For each such candidate X , we build a DFA accepting $\text{rep}_U(X)$ and compare it with \mathcal{A} . We can conclude since equality of regular languages is decidable. \square

There are recurrence relations with that property but that were not handled in [7]. Take [7, Example 33]

$$U_{i+5} = 6U_{i+4} + 3U_{i+3} - U_{i+2} + 6U_{i+1} + 3U_i, \quad \forall i \geq 0.$$

For this recurrence relation, $N_U(3^i) \not\rightarrow \infty$ (the proof relies on [7, Theorem 30]). The characteristic polynomial has the dominant root $\beta = 3 + 2\sqrt{3}$ and it also has three roots of modulus 1. Therefore, no decision procedure was known. But thanks to Theorem 2.5.1, we can handle such new cases under our

mild assumptions (H1), (H2) and (H3). Indeed, by applying Bertrand's theorem with the initial conditions 1, 7, 45, 291, 1881 (we have $d_\beta^*(1) = (62)^\omega$), the numeration language $0^* \text{rep}_U(\mathbb{N})$ is the set of words over $\{0, 1, \dots, 6\}$ avoiding the factors 63, 64, 65, 66, hence (H1) holds. Moreover, it is easily checked that for all $i \geq 0$, $U_{i+1} - U_i \geq 5U_i$. Therefore, the system U also satisfies (H2) and (H3).

2.5.2 The gcd of the coefficients of the recurrence relation is larger than 1

If X is an ultimately periodic set with period $\pi_X = m_X \cdot p_1^{\mu_1} \cdots p_t^{\mu_t}$ with $t \geq 1$ as in (2.2), then the quantity F_X is well-defined. Theorem 2.4.15 has a major assumption. The quantity

$$n_X = F_X - 1 - |\text{rep}_U(\pi_X - 1)|$$

should be larger than some positive constant Z , which only depends on the numeration system U .

Theorem 2.5.2. *Let U be a linear numeration system satisfying (H1), (H2) and (H3), and such that the gcd of the coefficients of the recurrence relation (2.1) is larger than 1. Let Z be the constant given in Definition 2.3.5. Assume there is a computable positive integer D such that for all ultimately periodic sets X of period $\pi_X = m_X \cdot p_1^{\mu_1} \cdots p_t^{\mu_t}$ as in (2.2) with $t \geq 1$, if $\max(\mu_1, \dots, \mu_t) \geq D$ then $n_X \geq Z$. Then, given a DFA \mathcal{A} accepting a language contained in the numeration language $\text{rep}_U(\mathbb{N})$, it is decidable whether this DFA recognizes an ultimately periodic set.*

Proof. Assume that X is an ultimately periodic set with period, as in (2.2), $\pi_X = m_X \cdot p_1^{\mu_1} \cdots p_t^{\mu_t}$. Note that there are only finitely many primes dividing all the coefficients of the recurrence relation (2.1), hence the possible p_1, \dots, p_t belongs to a finite set depending only on the numeration system U .

Applying the same reasoning as in the proof of Theorem 2.5.1, m_X is bounded by a constant B deduced from \mathcal{A} . Thus the quantity γ_{m_X} introduced in Definition 2.4.14 is also bounded.

Consider the greatest preperiod P of the sequences $(U_i \pmod{b})_{i \in \mathbb{N}}$, $b \in \llbracket 1, B \rrbracket$. Then by definition of F_X , there exists a computable constant D' such that if $\max(\mu_1, \dots, \mu_t) \geq D'$, then F_X is greater than P .

By hypothesis, there is a computable positive integer constant D such that if $\max(\mu_1, \dots, \mu_t) \geq D$ then $n_X \geq Z$. Let $E = \max(D, D')$. The number of t -uples (μ_1, \dots, μ_t) in $\llbracket 0, E-1 \rrbracket^t$ is finite. Hence there is a finite number of periods π_X of the form $m_X \cdot p_1^{\mu_1} \cdots p_t^{\mu_t}$ with m_X bounded and

(μ_1, \dots, μ_t) in this set. We can enumerate them and proceed as in the last paragraph of the proof of Theorem 2.5.1.

We may now assume that $\max(\mu_1, \dots, \mu_t) \geq E$. Thanks to the assumption, $n_X \geq Z$. Moreover, F_X is greater than P . We are thus able to apply Theorem 2.4.15¹: it provides a bound on π_ν and thus on the possible exponents μ_1, \dots, μ_t depending only on \mathcal{A} . We conclude in the same way as in the proof of Theorem 2.5.1. \square

In the last part of this section, we present a possible way to tackle new examples of numeration systems by applying Theorem 2.5.2. We stress the fact that when π_X is increasing then both terms F_X and $|\text{rep}_U(\pi_\nu - 1)|$ are increasing. If $\beta > 1$ (see Definition 2.3.9), then the growth of the second one has a logarithmic bound thanks to Lemma 2.3.10, hence we need insight on $\mathbf{f}_{p_j}(\mu)$ to be able to guarantee $n_X \geq Z$.

There is a clear link between the p_j -adic valuation ν_{p_j} (cf. Definition 1.7.1) and \mathbf{f}_{p_j} : for all non-negative integers μ and M ,

$$\mathbf{f}_{p_j}(\mu) = M \iff (\nu_{p_j}(U_{M-1}) < \mu \wedge \forall i \geq M, \nu_{p_j}(U_i) \geq \mu).$$

Remark 2.5.3. With our Example 2.2.1 and initial conditions 1, 2, 3, computing the first few values of $\nu_2(H_i)$ might suggest that it is bounded by a function of the form $\frac{i}{2} + c$, for some constant c . Nevertheless, computing more terms we get the following pairs $(i, \nu_2(H_i))$: (67, 44), (2115, 1070), (10307, 5172), (534595, 267318), (2631747, 1315896). The constant c suggested by each of these points is respectively $\frac{21}{2}$, $\frac{25}{2}$, $\frac{37}{2}$, $\frac{41}{2}$, $\frac{45}{2}$, which is increasing. This example explains the second term $g(i)$ in the function bounding $\nu_{p_j}(U_i)$ in the next statement.

In the next lemma, the reader can think about logarithm function instead of a general function g . Indeed, for any $\epsilon > 0$, $\log(i) < \epsilon i$ for large enough i . We also keep context and notation from (2.2).

Lemma 2.5.4. *Let $j \in \llbracket 1, t \rrbracket$. Assume that there are $\alpha, \epsilon \in \mathbb{R}_{>0}$ and a non-decreasing function g such that*

$$\nu_{p_j}(U_i) < \lfloor \alpha i \rfloor + g(i)$$

for all $i \in \mathbb{N}$ and there exists M such that $g(i) < \epsilon i$ for all $i > M$. Then, for large enough μ ,

$$\mathbf{f}_{p_j}(\mu) > \frac{\mu}{\alpha + \epsilon}.$$

¹Considering leading zeros or not do not change the reasoning.

Proof. By definition of the p -adic valuation, $p_j^{\nu_{p_j}(U_i)} \mid U_i$ and $p_j^{\nu_{p_j}(U_i)+1} \nmid U_i$. Thus, by definition of \mathbf{f}_{p_j} , for all i ,

$$\mathbf{f}_{p_j}(\nu_{p_j}(U_i) + 1) \geq i + 1.$$

For all μ , since the functions $x \mapsto \lfloor \alpha x \rfloor$ and g are non-decreasing and since the first one tends to infinity, there exists i such that

$$\lfloor \alpha i \rfloor + g(i) \leq \mu < \lfloor \alpha(i+1) \rfloor + g(i+1).$$

Take μ large enough so that $i \geq M$. Using the right-hand side inequality, $\mu < \alpha(i+1) + \epsilon(i+1)$ and we get

$$i > \frac{\mu}{\alpha + \epsilon} - 1.$$

Using the left-hand side inequality, $\mu \geq \lfloor \alpha i \rfloor + g(i) > \nu_{p_j}(U_i)$. Since we have integers on both sides, $\mu \geq \nu_{p_j}(U_i) + 1$. Since \mathbf{f}_{p_j} is non-decreasing, for all large enough μ ,

$$\mathbf{f}_{p_j}(\mu) \geq \mathbf{f}_{p_j}(\nu_{p_j}(U_i) + 1) \geq i + 1 > \frac{\mu}{\alpha + \epsilon}.$$

□

To apply Theorem 2.5.2, we are looking for some constant D such that $\max\{\mu_1, \dots, \mu_t\} \geq D \Rightarrow n_X \geq Z$. To find this D , let us first look for a lower bound for n_X . On the one hand, suppose that for each $j \in \llbracket 1, t \rrbracket$, there exist $\alpha_j, \epsilon_j, g_j$ and M_j as in the above lemma. Then, if ν_1, \dots, ν_t are large enough,

$$F_X = \max_{j \in \llbracket 1, t \rrbracket} \mathbf{f}_{p_j}(\nu_j) > \max_{j \in \llbracket 1, t \rrbracket} \left(\frac{\nu_j}{\alpha_j + \epsilon_j} \right) \geq \frac{\max_{j \in \llbracket 1, t \rrbracket} \nu_j}{\max_{j \in \llbracket 1, t \rrbracket} (\alpha_j + \epsilon_j)}.$$

Secondly, let u and β as in Definition 2.3.9. Assume that $\beta > 1$. Applying Lemma 2.3.10, there is a constant K such that

$$|\text{rep}_U(\pi_\nu - 1)| \leq u \log_\beta \left(m_X \prod_{j=1}^t p_j^{\nu_j} \right) + K.$$

The right hand side is

$$\begin{aligned} u \sum_{j=1}^t \nu_j \log_\beta(p_j) + u \log_\beta(m_X) + K \\ \leq u \left(\max_{j \in \llbracket 1, t \rrbracket} \nu_j \right) \sum_{j=1}^t \log_\beta p_j + u \log_\beta(m_X) + K. \end{aligned}$$

Consequently,

$$n_X \geq \max_{j \in \llbracket 1, t \rrbracket} \nu_j \left(\frac{1}{\max_{j \in \llbracket 1, t \rrbracket} (\alpha_j + \epsilon_j)} - u \sum_{j=1}^t \log_\beta p_j \right) - u \log_\beta(m_X) - K - 1.$$

We thus obtain a lower bound for n_X . If π_X tends to infinity (and assuming that the corresponding factor m_X remains bounded as explained in the proof of Theorem 2.5.2), then the quantity $\max_{j \in \llbracket 1, t \rrbracket} \mu_j = \max_{j \in \llbracket 1, t \rrbracket} \nu_j$ must also tend to infinity. Hence we are able to conclude, i.e. n_X tends to infinity and in particular, n_X will become larger than Z (the constant from Definition 2.3.5) whenever

$$\frac{1}{\max_{j \in \llbracket 1, t \rrbracket} (\alpha_j + \epsilon_j)} > u \sum_{j=1}^t \log_\beta p_j. \quad (2.4)$$

Actually, we don't need n_X tending to infinity, since we have the weaker requirement $n_X \geq Z$. The constant D from Theorem 2.5.2 can be obtained as follows. To ensure that $n_X \geq Z$, it is enough to have

$$\max_{j \in \llbracket 1, t \rrbracket} \mu_j \geq \frac{Z + K + 1}{\frac{1}{\max_{j \in \llbracket 1, t \rrbracket} (\alpha_j + \epsilon_j)} - u \sum_{j=1}^t \log_\beta p_j} \quad (2.5)$$

and the right hand side only depends on the numeration system U .

As a conclusion, we simply define the constant D as the right hand side in (2.5) and, under the assumption of Lemma 2.5.4 about the behaviour of the p_j -adic valuations of $(U_i)_{i \in \mathbb{N}}$, the decision procedure of Theorem 2.5.2 may thus be applied. From a practical point of view, even though n_X tending to infinity is not required, testing (2.4) is relatively easy with estimations as seen in the following remark. This is not a formal proof, simply rough computations suggesting what could be the value of α in Lemma 2.5.4.

Remark 2.5.5. One can make some computational experiments. Take the numeration system of Example 2.1.2. If we compute $\nu_2(G_i)$, the values for $41 \leq i \leq 60$ are given by

10, 10, 10, 11, 13, 11, 11, 12, 12, 12, 12, 13, 14, 13, 13, 14, 14, 14, 14, 15.

Hence, one can conjecture that $\alpha_1 = \frac{1}{4}$ and the above condition (2.4) becomes ($u = 1$), assuming ϵ_1 to be negligible,

$$4 > \log_{2.804}(2) \simeq 0.672.$$

Take the numeration system of Example 2.2.1. If we compute $\nu_2(H_i)$, the values for $41 \leq i \leq 60$ are given by

$$24, 20, 21, 21, 24, 22, 23, 23, 27, 24, 25, 25, 28, 26, 27, 27, 33, 28, 29, 29$$

and, similarly, for $\nu_3(H_i)$

$$13, 14, 14, 14, 15, 15, 15, 16, 17, 16, 17, 17, 17, 18, 18, 18, 19, 20, 19, 20.$$

Hence, one can conjecture that $\alpha_1 = \frac{1}{2}$ and $\alpha_2 = \frac{1}{3}$. The recurrence has a real dominant root $\beta \simeq 12.554$ (hence $u = 1$). Assuming ϵ_1 and ϵ_2 to be negligible, the condition (2.4) is therefore

$$2 > \log_{12.554}(2) + \log_{12.554}(3) \simeq 0.708.$$

2.6 An incursion into p -adic analysis

In this section, we discuss the requirement on the p -adic valuation given in Lemma 2.5.4. To that end, we reconsider our toy example.

2.6.1 A third-order sequence

Throughout this section, let $H_{i+3} = 12H_{i+2} + 6H_{i+1} + 12H_i$ with initial conditions $H_0 = 1, H_1 = 13, H_2 = 163$ be the sequence of Example 2.2.1. There are two primes dividing all the coefficients of the recurrence relation: 2 and 3. The 3-adic valuation of H_i has a simple structure.

Theorem 2.6.1. *For all $i \in \mathbb{N}$, we have*

$$\nu_3(H_i) = \left\lfloor \frac{i}{3} \right\rfloor + \begin{cases} 0 & \text{if } i \not\equiv 4 \pmod{9} \\ 1 & \text{if } i \equiv 4 \pmod{9}. \end{cases}$$

Proof. Let $T_i = H_i/3^{\frac{i-2}{3}}$ for all $i \in \mathbb{N}$. Since $H_{i+3} = 12H_{i+2} + 6H_{i+1} + 12H_i$, the sequence $(T_i)_{i \in \mathbb{N}}$ satisfies for all $i \in \mathbb{N}$ the recurrence relation given by $T_{i+3} = 4 \cdot 3^{2/3}T_{i+2} + 2 \cdot 3^{1/3}T_{i+1} + 4T_i$. Moreover, the initial terms are $T_0 = 3^{2/3}, T_1 = 13 \cdot 3^{1/3}, T_2 = 163$, thus it follows that $T_i \in \mathbb{Z}[3^{1/3}]$ for all $i \in \mathbb{N}$. Modulo $9\mathbb{Z}[3^{1/3}]$, one computes that the sequence $(T_i)_{i \in \mathbb{N}}$ is periodic with period 27, since this sequence can be written uv^ω where v is the finite word

$$\begin{array}{cccccccc} 3^{2/3} & 4 \cdot 3^{1/3} & 1 & 7 \cdot 3^{2/3} & 3 \cdot 3^{1/3} & 1 & 2 \cdot 3^{2/3} & 2 \cdot 3^{1/3} & 4 \\ 3^{2/3} & 3^{1/3} & 7 & 7 \cdot 3^{2/3} & 3 \cdot 3^{1/3} & 7 & 8 \cdot 3^{2/3} & 5 \cdot 3^{1/3} & 1 \\ 3^{2/3} & 7 \cdot 3^{1/3} & 4 & 7 \cdot 3^{2/3} & 3 \cdot 3^{1/3} & 4 & 5 \cdot 3^{2/3} & 8 \cdot 3^{1/3} & 7. \end{array}$$

Therefore the sequence $(\nu_3(T_i))_{i \in \mathbb{N}}$ of 3-adic valuations is

$$\frac{2}{3}, \frac{1}{3}, 0, \frac{2}{3}, \frac{4}{3}, 0, \frac{2}{3}, \frac{1}{3}, 0, \dots$$

with period 9. (Note that we use the natural extension of ν_3 to a function $\nu_3: \mathbb{Z}[3^{1/3}] \rightarrow \frac{1}{3}\mathbb{Z}$.) Equivalently,

$$\nu_3(T_i) = \left\lfloor \frac{i}{3} \right\rfloor - \frac{i-2}{3} + \begin{cases} 0 & \text{if } i \not\equiv 4 \pmod{9} \\ 1 & \text{if } i \equiv 4 \pmod{9}. \end{cases}$$

It follows that

$$\nu_3(H_i) = \frac{i-2}{3} + \nu_3(T_i) = \left\lfloor \frac{i}{3} \right\rfloor + \begin{cases} 0 & \text{if } i \not\equiv 4 \pmod{9} \\ 1 & \text{if } i \equiv 4 \pmod{9} \end{cases}$$

for all $i \in \mathbb{N}$. □

Theorem 2.6.1 implies $\frac{i-2}{3} \leq \nu_3(H_i) \leq \frac{i+2}{3}$ for all $i \in \mathbb{N}$. In particular, $\nu_3(H_i) < \left\lfloor \frac{i}{3} \right\rfloor + 2$, hence the condition of Lemma 2.5.4 is satisfied, and therefore for every $\epsilon > 0$ we have

$$f_3(\mu) > \frac{\mu}{\frac{1}{3} + \epsilon}$$

for large enough μ . This takes care of one of the two primes dividing all the coefficients of the recurrence relation. We still have to discuss $\nu_2(H_i)$.

Unfortunately, Theorem 2.6.1 is not representative of the behaviour of $\nu_p(s_i)$ for a general sequence $(s_i)_{i \in \mathbb{N}}$ satisfying a linear recurrence with constant coefficients. For instance, the 2-adic valuation of the sequence $(H_i)_{i \in \mathbb{N}}$ is (much) more complicated. To study the more general setting, we will make use of the field of p -adic numbers (see Chapter 1 Section 1.7).

Recall that $|\text{rep}_p(n)|$ is the number of digits in the standard base- p representation of n (cf. Example 1.3.3). For all $n \geq 1$, we can bound $\nu_p(n)$ as

$$\nu_p(n) \leq |\text{rep}_p(n)| - 1 = \lfloor \log_p(n) \rfloor \leq \log_p(n).$$

Proposition 2.6.3 below gives the analogous upper bound on $\nu_p(n - \zeta)$ when ζ is a p -adic integer whose sequence of base- p digits does not have blocks of consecutive zeros that grow too quickly.

Definition 2.6.2. Let p be a prime and let $\zeta \in \mathbb{Z}_p \setminus \mathbb{N}$. Write $\zeta = \sum_{i \geq 0} d_i p^i$, where each $d_i \in \llbracket 0, p-1 \rrbracket$. For all $a \in \mathbb{N}$, let $\ell_\zeta(a) \geq 0$ be maximal j such that $0 = d_a = d_{a+1} = \dots = d_{a+j-1}$.

Proposition 2.6.3. *Let p be a prime and let $\zeta \in \mathbb{Z}_p \setminus \mathbb{N}$. If there are real numbers C, D such that $C > 0$, $D \geq -(C + 1)$, and $\ell_\zeta(a) \leq Ca + D$ for all $a \geq 2$, then $\nu_p(n - \zeta) \leq (2C + D + 2) \log_p(n)$ for all $n \geq p$.*

Proof. Write $\zeta = \sum_{i \geq 0} d_i p^i$, where each $d_i \in \llbracket 0, p-1 \rrbracket$. For all $a \in \mathbb{N}$, define the integer $\zeta_a = \zeta \pmod{p^a} = \sum_{i=0}^{a-1} d_i p^i$. Then $\nu_p(\zeta_a - \zeta) = a + \ell_\zeta(a)$.

Let $n \geq p$ and $a = |\text{rep}_p(n)| \geq 2$. Since $\zeta \notin \mathbb{N}$, $n - \zeta \neq 0$, the p -adic valuation $b = \nu_p(n - \zeta)$ is well-defined. There are two cases.

If $n \leq \zeta_b$, then in fact $n = \zeta_b$. Indeed, $n \leq \zeta_b < p^b$, thus $n \neq \zeta_b$ implies $n - \zeta_b \not\equiv 0 \pmod{p^b}$, which contradicts $b = \nu_p(n - \zeta)$. Thus $n = \zeta_b$. Since $|\text{rep}_p(n)| = a$ and $n = \zeta_b$, we have $0 = d_a = \dots = d_{b-1}$. Therefore we have $\zeta_a = \zeta_b = n \geq p^{a-1}$, and

$$\frac{\nu_p(n - \zeta)}{\log_p(n)} = \frac{\nu_p(\zeta_a - \zeta)}{\log_p(\zeta_a)} \leq \frac{a + \ell_\zeta(a)}{\log_p(p^{a-1})} \leq \frac{a + Ca + D}{(a-1)} \leq 2 + 2C + D,$$

where the final inequality follows from $1 + C + D \geq 0$.

If $n > \zeta_b$, then $n = \zeta_b + p^b m$ for some positive integer m . Therefore $n \geq p^b$, hence

$$\frac{\nu_p(n - \zeta)}{\log_p(n)} \leq \frac{b}{\log_p(p^b)} = 1 < 1 + C \leq 2 + 2C + D$$

if $b \geq 1$ and $\frac{\nu_p(n - \zeta)}{\log_p(n)} = 0 < 2 + 2C + D$ if $b = 0$. □

We now focus on the sequence of 2-adic valuations $(\nu_2(H_i))_{i \in \mathbb{N}}$. Note that computations in what follows can be done with a computing system like **Mathematica**. To analyse the 2-adic behaviour of $(H_i)_{i \in \mathbb{N}}$, we construct a piecewise interpolation of H_i to \mathbb{Z}_2 using the method described by Rowland and Yassawi in [69]. Let $P(x) = x^3 - 12x^2 - 6x - 12$ be the characteristic polynomial of $(H_i)_{i \in \mathbb{N}}$. The polynomial $P(x)$ has a unique root $\beta_1 \in \mathbb{Z}_2$ satisfying $\beta_1 \equiv 2 \pmod{4}$. Indeed, since $|P(2)|_2 = \frac{1}{64} < \frac{1}{4} = |P'(2)|_2^2$, we can apply Hensel's Lemma (Theorem 1.7.15): there is a unique $\beta_1 \in \mathbb{Z}_2$ such that $P(\beta_1) = 0$ and $|\beta_1 - 2|_2 < |P'(2)|_2 = 1/2$. Therefore there is a unique β_1 such that $\nu_2(\beta_1 - 2) > 1$, i.e. so that $4 \mid \beta_1 - 2$.

Polynomial division shows that $P(x)$ factors in $\mathbb{Z}_2[x]$ as

$$P(x) = (x - \beta_1) (x^2 + (\beta_1 - 12)x + (\beta_1^2 - 12\beta_1 - 6)).$$

One checks that $P(x)$ has no roots in \mathbb{Z}_2 congruent to 0, 1, 3, 4, 5, or 7

modulo 8: $P(x) \equiv x^3 + 4x^2 + 2x + 4 \pmod{8}$, hence

x	$P(x) \pmod{8}$
0	4
1	$1 + 4 + 2 + 4 \equiv 3$
3	$27 + 36 + 6 + 4 \equiv 1$
4	$64 + 64 + 8 + 4 \equiv 4$
5	$125 + 100 + 10 + 4 \equiv 7$
7	$343 + 196 + 14 + 4 \equiv 5$

Since β_1 has multiplicity 1, this implies that the splitting field K of $P(x)$ is a quadratic extension of \mathbb{Q}_2 . Let β_2 and β_3 be the other two roots of $P(x)$ in $K = \mathbb{Q}_2(\beta_2)$. Since $\beta_1 \equiv 2 \pmod{4}$, the 2-adic absolute value of β_1 is $|\beta_1|_2 = \frac{1}{2}$. Using the quadratic factor of $P(x)$ and an approximation to β_1 , one computes $|\beta_2|_2 = |\beta_3|_2 = \frac{1}{\sqrt{2}}$.

Let $c_1, c_2, c_3 \in K$ be such that

$$H_i = c_1\beta_1^i + c_2\beta_2^i + c_3\beta_3^i$$

for all $i \in \mathbb{N}$. Using the initial conditions, we get

$$\begin{aligned} c_1 &= \frac{-H_0\beta_2\beta_3 + H_1(\beta_2 + \beta_3) - H_2}{(\beta_2 - \beta_1)(\beta_1 - \beta_3)} \\ c_2 &= \frac{-H_0\beta_3\beta_1 + H_1(\beta_3 + \beta_1) - H_2}{(\beta_3 - \beta_2)(\beta_2 - \beta_1)} \\ c_3 &= \frac{-H_0\beta_1\beta_2 + H_1(\beta_1 + \beta_2) - H_2}{(\beta_1 - \beta_3)(\beta_3 - \beta_2)}. \end{aligned}$$

Recall that $H_0 = 1, H_1 = 13, H_2 = 163$. Using this information, one computes $|c_1|_2 = 2$ and $|c_2|_2 = 2\sqrt{2} = |c_3|_2$. Factoring out β_2^i gives

$$H_i = \beta_2^i \left(c_1 \left(\frac{\beta_1}{\beta_2} \right)^i + c_2 + c_3 \left(\frac{\beta_3}{\beta_2} \right)^i \right). \quad (2.6)$$

For further considerations (see the proof of Theorem 2.6.7), we need to study the proximity of $c_2 + c_3 \left(\frac{\beta_3}{\beta_2} \right)^i$ to 0. To analyse the size of $c_2 + c_3 \left(\frac{\beta_3}{\beta_2} \right)^i$, we interpret $\left(\frac{\beta_3}{\beta_2} \right)^i$ as a function of a 2-adic variable. For this we need the 2-adic exponential \mathbf{exp}_2 and logarithm \mathbf{log}_2 (see Definition 1.7.17). One can check with a computation that one has $|\left(\frac{\beta_3}{\beta_2} \right)^4 - 1|_2 = \frac{1}{8} < \frac{1}{2} = 2^{-1/(2-1)}$. Therefore, for all $m \geq 0$ and $r \in \{0, 1, 2, 3\}$,

$$\begin{aligned} \left(\frac{\beta_3}{\beta_2} \right)^{r+4m} &= \left(\frac{\beta_3}{\beta_2} \right)^r \left(\frac{\beta_3}{\beta_2} \right)^{4m} \\ &= \left(\frac{\beta_3}{\beta_2} \right)^r \mathbf{exp}_2 \mathbf{log}_2 \left(\left(\frac{\beta_3}{\beta_2} \right)^{4m} \right) \\ &= \left(\frac{\beta_3}{\beta_2} \right)^r \mathbf{exp}_2 \left(m \mathbf{log}_2 \left(\left(\frac{\beta_3}{\beta_2} \right)^4 \right) \right). \end{aligned}$$

Denote $L = \log_2((\frac{\beta_3}{\beta_2})^4)$. Using the power series for \log_2 , one computes $|L|_2 = \frac{1}{8}$. For all $x \in \mathbb{Z}_2[\beta_2]$ and $r \in \{0, 1, 2, 3\}$, define

$$f_r(r + 4x) = c_2 + c_3 \left(\frac{\beta_3}{\beta_2}\right)^r \exp_2(Lx).$$

For all $x \in \mathbb{Z}_2$, we have $|Lx|_2 = \frac{1}{8}|x|_2 \leq \frac{1}{8} < \frac{1}{2} = 2^{-1/(2-1)}$, thus f_r is well-defined on $r + 4\mathbb{Z}_2$. The four functions f_0, f_1, f_2, f_3 constitute a piecewise interpolation of $c_2 + c_3 \left(\frac{\beta_3}{\beta_2}\right)^i$. Namely, $c_2 + c_3 \left(\frac{\beta_3}{\beta_2}\right)^i = f_{i \bmod 4}(i)$ for all $i \in \mathbb{N}$.

The equation $f_r(r + 4x) = 0$ is equivalent to

$$\exp_2(Lx) = -\frac{c_2}{c_3} \left(\frac{\beta_2}{\beta_3}\right)^r.$$

For $r \in \{0, 2, 3\}$, one computes $\left| -\frac{c_2}{c_3} \left(\frac{\beta_2}{\beta_3}\right)^r - 1 \right|_2 \geq \frac{1}{2} = 2^{-1/(2-1)}$, hence there is no solution x for these values of r . For $r = 1$, $\left| -\frac{c_2}{c_3} \left(\frac{\beta_2}{\beta_3}\right)^r - 1 \right|_2 = \frac{1}{16} < \frac{1}{2}$, thus there is a unique solution, namely $x = \frac{1}{L} \log_2\left(-\frac{c_2\beta_2}{c_3\beta_3}\right)$, which has size $|x|_2 = \frac{1}{2}$.

Definition 2.6.4. With the above notation, let

$$\zeta = 1 + 4\frac{1}{L} \log_2\left(-\frac{c_2\beta_2}{c_3\beta_3}\right).$$

In this case, one has $f_1(\zeta) = 0$ and $|\zeta|_2 = 1$ (using Proposition 1.7.5). Remark that ζ is a computable number, and one computes $\zeta \equiv 660098850944665 \pmod{2^{50}}$.

Proposition 2.6.5. *The quantity ζ given in Definition 2.6.4 belongs to \mathbb{Z}_2 .*

Proof. Let $\sigma: K \rightarrow K$ be the Galois automorphism that non-trivially permutes β_2 and β_3 . The formulas for c_2 and c_3 imply $\frac{c_2}{c_3} \cdot \frac{\sigma(c_2)}{\sigma(c_3)} = 1$; this implies

$$\begin{aligned} \log_2\left(-\frac{c_2\beta_2}{c_3\beta_3}\right) + \sigma\left(\log_2\left(-\frac{c_2\beta_2}{c_3\beta_3}\right)\right) &= \log_2\left(\frac{c_2\beta_2}{c_3\beta_3} \cdot \frac{\sigma(c_2)\beta_3}{\sigma(c_3)\beta_2}\right) \\ &= \log_2(1) = 0. \end{aligned}$$

Similarly,

$$\log_2\left(\left(\frac{\beta_3}{\beta_2}\right)^4\right) + \sigma\left(\log_2\left(\left(\frac{\beta_3}{\beta_2}\right)^4\right)\right) = \log_2(1) = 0.$$

Therefore

$$\frac{\log_2\left(-\frac{c_2\beta_2}{c_3\beta_3}\right)}{\log_2\left(\left(\frac{\beta_3}{\beta_2}\right)^4\right)} = \frac{-\sigma\left(\log_2\left(-\frac{c_2\beta_2}{c_3\beta_3}\right)\right)}{-\sigma\left(\log_2\left(\left(\frac{\beta_3}{\beta_2}\right)^4\right)\right)} = \sigma\left(\frac{\log_2\left(-\frac{c_2\beta_2}{c_3\beta_3}\right)}{\log_2\left(\left(\frac{\beta_3}{\beta_2}\right)^4\right)}\right)$$

is invariant under σ and thus is an element of \mathbb{Q}_2 . It follows from $|\zeta|_2 = 1$ that $\zeta \in \mathbb{Z}_2$. \square

By Proposition 2.6.3, the growth rate of $\nu_2(H_i)$ is determined by the approximability of

$$\zeta = \cdots 10010110000101101100111101100001101111011010011001_2$$

by non-negative integers.

Conjecture 2.6.6. *Let $\zeta \in \mathbb{Z}_2$ be defined as in Definition 2.6.4. The lengths of the 0 blocks of the 2-adic digits of ζ satisfy $\ell_\zeta(a) \leq \frac{2}{95}a + \frac{18}{5}$ for all $a \geq 0$.*

Conjecture 2.6.6 is weak in the sense that it is almost certainly far from sharp. One expects the digits of ζ to be randomly distributed, in which case $\ell_\zeta(a) = \log_2(a) + O(1)$. Indeed, among the first 1000 base-2 digits of ζ , the longest block of 0s has length 10. However, results concerning digits of irrational numbers are notoriously difficult to prove. Bugeaud and Kekeç [19, Theorem 1.7] give a lower bound on the number of non-zero digits among the first a digits of an irrational algebraic number in \mathbb{Q}_p . However, there are no known results of this form for transcendental numbers.

The above conjectural bound was obtained by computing the line through $\ell_\zeta(19) = 4$ and $\ell_\zeta(304) = 10$. If Conjecture 2.6.6 is true, then an explicit formula for $\nu_2(H_i)$ is given by the following theorem. In particular, the approximation $\zeta \equiv 660098850944665 \pmod{2^{50}}$ is sufficient to compute $\nu_2(H_i)$ for all $i \leq 2^{49}$.

Theorem 2.6.7. *Let $\zeta \in \mathbb{Z}_2$ be defined as in Definition 2.6.4. Conjecture 2.6.6 implies that for all $i \geq 10$,*

$$\nu_2(H_i) = \left\lfloor \frac{i-1}{2} \right\rfloor + \begin{cases} \nu_2(i-\zeta) & \text{if } i \equiv 1 \pmod{4} \\ 0 & \text{if } i \not\equiv 1 \pmod{4}. \end{cases}$$

Proof. We start as in the proof of Theorem 2.6.1. Let $T_i = H_i/2^{\frac{i}{2}-1}$ for all $i \in \mathbb{N}$. Since $H_{i+3} = 12H_{i+2} + 6H_{i+1} + 12H_i$, the sequence $(T_i)_{i \in \mathbb{N}}$ satisfies the recurrence $T_{i+3} = 6\sqrt{2}T_{i+2} + 3T_{i+1} + 3\sqrt{2}T_i$. The initial terms are $T_0 = 2, T_1 = 13\sqrt{2}, T_2 = 163$, thus it follows that $T_i \in \mathbb{Z}[\sqrt{2}]$ for all $i \in \mathbb{N}$. Modulo $2\mathbb{Z}[\sqrt{2}]$, the sequence $(T_i)_{i \geq 2}$ is periodic with period 4: $1, \sqrt{2}, 1, 0, 1, \sqrt{2}, 1, 0, \dots$. It follows that if $i \geq 2$ and $i \not\equiv 1 \pmod{4}$ then

$$\begin{aligned} \nu_2(H_i) &= \frac{i}{2} - 1 + \nu_2(T_i) = \frac{i}{2} - 1 + \begin{cases} 0 & \text{if } i \equiv 0 \pmod{4} \\ 0 & \text{if } i \equiv 2 \pmod{4} \\ \frac{1}{2} & \text{if } i \equiv 3 \pmod{4} \end{cases} \\ &= \left\lfloor \frac{i-1}{2} \right\rfloor. \end{aligned}$$

It remains to determine $\nu_2(H_i)$ when $i \equiv 1 \pmod{4}$. We continue to use the splitting field K of $P(x)$, the 2-adic numbers $\beta_1, \beta_2, \beta_3, c_1, c_2, c_3$ and the function f_1 defined before Definition 2.6.4. When $i \equiv 1 \pmod{4}$, Equation (2.6) gives

$$|H_i|_2 = 2^{-\frac{i}{2}} \left| c_1 \left(\frac{\beta_1}{\beta_2} \right)^i + f_1(i) \right|_2.$$

To obtain a simpler formula for $|H_i|_2$, we compare the sizes of the two terms being added and use the particular case of Proposition 1.7.5. For the first, we have $\left| c_1 \left(\frac{\beta_1}{\beta_2} \right)^i \right|_2 = 2^{1-\frac{i}{2}}$. For the second,

$$|f_1(i)|_2 = \left| c_2 + \frac{c_3 \beta_3}{\beta_2} \mathbf{exp}_2(L \cdot \frac{i-1}{4}) \right|_2.$$

Since the function $f_1(1+4x) = c_2 + \frac{c_3 \beta_3}{\beta_2} \mathbf{exp}_2(Lx)$ has a unique zero $\frac{\zeta-1}{4}$, the p -adic Weierstrass preparation theorem (Theorem 1.7.19) implies the existence of a power series h with coefficients in K such that $h(0) = 1$, $|h(x)|_2 = 1$ for all $x \in \mathbb{Z}_2[\beta_2]$, and

$$f_1(1+4x) = \frac{c_2 + \frac{c_3 \beta_3}{\beta_2}}{-\frac{\zeta-1}{4}} \left(x - \frac{\zeta-1}{4} \right) h(x).$$

(Recall that the p -adic Weierstrass preparation theorem implies the existence of a polynomial g and a power series h such that $f_1(x) = g(x)h(x)$, where g has exactly the same zeros as f , and where the coefficient of X^0 in h is 1.) Therefore

$$\begin{aligned} |f_1(i)|_2 &= \left| \frac{c_2 + \frac{c_3 \beta_3}{\beta_2}}{-\frac{\zeta-1}{4}} \right|_2 \left| \frac{i-1}{4} - \frac{\zeta-1}{4} \right|_2 \\ &= \sqrt{2} |i-\zeta|_2. \end{aligned}$$

Conjecture 2.6.6 and Proposition 2.6.3 imply $|i-\zeta|_2 \geq i^{-536/95}$ for all $i \geq 2$. The functions $2^{1-\frac{i}{2}}$ and $\sqrt{2}i^{-536/95}$ intersect at $i \approx 70.21$. For all $i \geq 73$ such that $i \equiv 1 \pmod{4}$,

$$\left| c_1 \left(\frac{\beta_1}{\beta_2} \right)^i \right|_2 = 2^{1-\frac{i}{2}} < \sqrt{2}i^{-536/95} \leq |f_1(i)|_2$$

and therefore

$$|H_i|_2 = 2^{-\frac{i}{2}} \left| c_1 \left(\frac{\beta_1}{\beta_2} \right)^i + f_1(i) \right|_2 = 2^{-\frac{i}{2}} |f_1(i)|_2 = 2^{\frac{1-i}{2}} |i-\zeta|_2. \quad (2.7)$$

Moreover, an explicit computation shows that we have $2^{1-\frac{i}{2}} < \sqrt{2}|i-\zeta|_2$ for all $i \equiv 1 \pmod{4}$ satisfying $13 \leq i \leq 69$, hence $|H_i|_2 = 2^{\frac{1-i}{2}} |i-\zeta|_2$ for these values as well. Therefore $\nu_2(H_i) = \frac{i-1}{2} + \nu_2(i-\zeta)$ for all $i \equiv 1 \pmod{4}$ verifying $i \geq 13$. \square

Corollary 2.6.8. *Conjecture 2.6.6 implies that $\nu_2(H_i) \leq \frac{i}{2} + \frac{536}{95} \log_2(i)$ for all $i \geq 10$.*

Proof. Since $H_i \neq 0$ for all $i \in \mathbb{N}$, we have $|H_i|_2 \neq 0$ for all $i \in \mathbb{N}$. Moreover, $|f_1(\zeta)|_2 = 0$. Thus Equation (2.7) implies $\zeta \notin \mathbb{N}$. Conjecture 2.6.6 and Proposition 2.6.3 imply $\nu_2(i - \zeta) \leq \frac{536}{95} \log_2(i)$ for all $i \geq 2$. By Theorem 2.6.7, $\nu_2(H_i) \leq \frac{i}{2} + \frac{536}{95} \log_2(i)$ for all $i \geq 10$. \square

This is sufficient to apply Lemma 2.5.4. Under Conjecture 2.6.6, we have the right behaviour for both $\nu_2(H_i)$ and $\nu_3(H_i)$.

2.6.2 A fourth-order sequence

The general case is even more complicated than getting Definition 2.6.4 and Theorem 2.6.7. For example, consider the sequence $(G_i)_{i \in \mathbb{N}}$ from Example 2.1.2 given by the initial conditions $G_0 = 1, G_1 = 3, G_2 = 9, G_3 = 25$ and the recurrence $G_{i+4} = 2G_{i+3} + 2G_{i+2} + 2G_i$. There is only one prime number dividing all the coefficients of the recurrence relation: 2. By the Eisenstein criterion 1.7.14, the characteristic polynomial $P(x) = x^4 - 2x^3 - 2x^2 - 2$ is irreducible over \mathbb{Q}_2 . Let K be the splitting field of $P(x)$ over \mathbb{Q}_2 . Let $\beta_1, \beta_2, \beta_3, \beta_4$ be the four roots of $P(x)$ in K , and let c_1, c_2, c_3, c_4 be the elements of K such that $G_i = \sum_{j=1}^4 c_j \beta_j^i$ for all $i \in \mathbb{N}$.

To compute with the roots β_i , we would want to write K as a simple extension $\mathbb{Q}_2(\alpha)$. For this, we need to determine the degree d of the extension and a polynomial $Q(x) \in \mathbb{Q}_2[x]$ of degree d such that $Q(x)$ is irreducible over \mathbb{Q}_2 and $Q(\alpha) = 0$. Then we could compare the sizes $|\beta_j|_2$ of the roots to each other. Experiments suggest that $|\beta_1|_2 = |\beta_2|_2 = |\beta_3|_2 = |\beta_4|_2 = 2^{-1/4}$ and $|(\frac{\beta_j}{\beta_1})^8 - 1|_2 = \frac{1}{4} < \frac{1}{2} = 2^{-1/(2-1)}$ for each $j \in \{2, 3, 4\}$. Assuming this is the case, $G_i/\beta_1^i = \sum_{j=1}^4 c_j (\frac{\beta_j}{\beta_1})^i$ can be interpolated piecewise to \mathbb{Z}_2 using 8 analytic functions. However, we can unfortunately not solve the equation $c_1 + b_2 \exp_2(L_2x) + b_3 \exp_2(L_3x) + b_4 \exp_2(L_4x) = 0$ explicitly, as we solved $c_2 + c_3 (\frac{\beta_3}{\beta_2})^r \exp_2(Lx) = 0$ in Section 2.6.1. Instead, we could use the p -adic Weierstrass preparation theorem to determine the number of solutions and compute approximations to them. However, we would also need to determine which of these solutions belong to \mathbb{Z}_2 . We do not carry out this step here, but this would give an analogue of Definition 2.6.4, with some finite set Y of 2-adic integers. If the blocks of zeros in the digit sequences of each $\zeta \in Y$ satisfy $\ell_\zeta(a) \leq Ca + D$ for some C, D as in Conjecture 2.6.6, then Proposition 2.6.3 gives an upper bound on $\nu_2(G_i)$. This same approach applies to a general constant-recursive sequence and a general prime p .

2.7 Concluding remarks

The case of integer base- b numeration systems (see Example 1.3.3) is not treated in this thesis. Let $b \geq 2$. Assume first for the sake of simplicity that b is a prime. Consider the sequence $U_b = (b^i)_{i \in \mathbb{N}}$. If X is an ultimately periodic set with period $\pi_X = b^\lambda$ for some λ , then with our notation $m_X = 1$ and $|\text{rep}_U(\pi_X - 1)| = \lambda$. The sequence $(b^i \bmod b^\lambda)_{i \geq 0}$ has a zero period and $f_b(\lambda) = \lambda$. Hence we don't have the required assumption to apply Theorem 2.4.15: for every such set X , $n_X = F_X - 1 - |\text{rep}_U(\frac{\pi_X}{m_X} - 1)| = -1$. Let us also point out that the technique of Propositions 2.4.1 or 2.4.6 cannot be applied: adding 1 as a most significant digit will not change the value of a representation modulo π_X when words are too long: $U_i \equiv 0 \pmod{b^\lambda}$ for large enough i . Of course, integer base systems can be handled with other decision procedures [13, 15, 43, 47, 54, 55]. If the base b is now a composite number of the form $p_1^{s_1} \cdots p_t^{s_t}$, the same observation holds. The length of the non-zero preperiod of $(b^i \bmod p_j^\mu)_{i \in \mathbb{N}}$ is $\lfloor \frac{\mu}{s_j} \rfloor$. Taking again an ultimately periodic set with period $\pi_X = b^\lambda$, we get $m_X = 1$ and $f_{p_j}(\lambda s_j) = \lambda$, hence $F_X = \lambda$ and we still have $|\text{rep}_U(\pi_X - 1)| = \lambda$, thus $n_X = -1$.

A similar situation occurs in a slightly more general setting: the merge of r sequences that ultimately behave like b^i . Let $b \geq 2$, $k \geq 1$, $N \geq 0$. If the recurrence relation is of the form $U_{i+k} = bU_i$ for $i \geq N$ (as for instance in Example 2.3.8), then again $n_X \not\rightarrow \infty$ as $\pi_X \rightarrow \infty$. Indeed, if X is an ultimately periodic set with period $\pi_X = b^\lambda$, then $m_X = 1$ and applying Lemma 2.3.10 (here the polynomial P_T with the notation of Definition 2.3.9 is just a constant and $u = k$), $|\text{rep}_U(\pi_X - 1)| \geq u\lambda - L$, for some constant L , and with the same reasoning as for a composite integer base, $F_X \leq N + u\lambda$. Thus n_X remains bounded for all λ . Hence there is no way to ensure that n_X can be larger than Z .

Trying to figure out the limitations of our decision procedure and assuming that we are under the assumption of Lemma 2.5.4, this type of linear numeration systems is the only one that we were able to find where our procedure cannot be applied. Moreover, as shown by the following proposition, these systems are sufficiently close to the classical base- b system hence usual decision procedures can still be applied. It is an open problem to determine if there are linear numeration systems satisfying (H1), (H2) and (H3) where the decision procedure may not be applied and not of the above type.

Example 2.7.1. Take $b = 4$, $k = 2$ and $N = 0$. Start with the first two values 1 and 3. We get the sequence 1, 3, 4, 12, 16, 48, 64, ... We have $f_2(\mu) = \mu$ if μ is even and $f_2(\mu) = \mu + 1$ if μ is odd. Hence, for a set of period $\pi_X = 4^\lambda$, $F_X = f_2(2\lambda) = 2\lambda$. Moreover, $|\text{rep}_U(4^\lambda - 1)| = 2\lambda$. Thus $n_X = -1$ for all λ .

Proposition 2.7.2. *Let $b \geq 2$, $u \geq 1$, $N \geq 0$. Let U be a linear numeration system $U = (U_i)_{i \in \mathbb{N}}$ such that $U_{i+u} = bU_i$ for all $i \geq N$. If a set is U -recognizable then it is b -recognizable. Moreover, given a DFA accepting $\text{rep}_U(X)$ for some set X , we can compute a DFA accepting $\text{rep}_b(X)$.*

Proof. We build in two steps a sequence of transducers reading least significant digit first that maps any U -representation $c_{\ell-1} \cdots c_1 c_0 \in \Sigma_U^*$ (here written with the usual convention that the most significant digit is on the left) to the corresponding b -ary representation. Adding leading zeros, we may assume that the length ℓ of the U -representation is of the form $N + mu$. The idea is to read the first $N + u$ (least significant) digits and to output a single digit (over a finite alphabet in \mathbb{N}) equal to

$$d_0 = \text{val}_U(c_{N+u-1} \cdots c_0).$$

Then we process blocks of size u . Each such block of the form

$$c_{N+(j+1)u-1} \cdots c_{N+ju}$$

gives as output a single digit equal to

$$d_j = c_{N+(j+1)u-1}U_{N+u-1} + \cdots + c_{N+ju}U_N.$$

Hence the digits d_0, d_1, \dots, d_{m-1} all belong to the finite set

$$\{\text{val}_U(w) : w \in \Sigma_U^* \text{ and } |w| \leq N + u\}.$$

From the form of the recurrence, we have

$$\text{val}_U(c_{N+mu-1} \cdots c_0) = \sum_{j=0}^{m-1} d_j b^j = \text{val}_b(d_{m-1} \cdots d_0).$$

Thus this transducer \mathcal{T} maps any U -representation to a non-classical b -ary representation of the same integer. Precisely, when a DFA accepting $\text{rep}_U(X)$ is given, we build a DFA accepting the language

$$L = 0^* \text{rep}_U(X) \cap \{w \in \Sigma_U^* : |w| \equiv N \pmod{u}, |w| \geq N\}.$$

Recall that if L is a regular language then its image $\mathcal{T}(L)$ by a transducer is again regular, cf. Proposition 1.2.27. Moreover, $\text{val}_b(\mathcal{T}(L)) = X$.

Then, it is a classical result that normalization in base b , i.e. mapping a representation over a non-canonical finite set of digits to the canonical expansion over $\Sigma_b = \{0, \dots, b-1\}$ can be achieved by a transducer \mathcal{N} ([37] or [66, p. 104]). To conclude with the proof, we compose these two transducers and consider the image $\mathcal{N}(0^* \mathcal{T}(L)) = 0^* \text{rep}_b(X)$. \square

With the above proposition, the decision problem for the merge of sequences ultimately behaving like b^i (such as the numeration systems of Examples 2.3.8 and 2.3.11) can be reduced to the usual decision problem for integer bases.

Consider the numeration system of Example 2.3.7 defined by initial conditions 1, 2, 4, 5 and $U_{i+4} = 5U_{i+2} - 4U_i$ for all $i \in \mathbb{N}$. We showed that it does not satisfy (H3). Therefore our decision procedure cannot be applied. Could we weaken our hypotheses? More generally, can we give a decision procedure for Problem 2.1.1 in general?

Even more generally, can we give a decision procedure for a similar problem stated for abstract numeration systems?

Problem 2.7.3. Given an abstract numeration system S and a set X of non-negative integers which is S -recognizable, is it decidable whether or not X is ultimately periodic?

This problem was shown to be equivalent to the HD0L periodicity problem [67, 44] (see Problem 1.1.17). As we will see in the introduction of the next chapter, the HD0L periodicity problem was shown to be decidable in its full generality [33, 58], but the proofs do not provide convenient algorithm, in the sense that it cannot easily be implemented. Can we give a decision procedure which provides an “practical” algorithm?

Chapter 3

Minimal automaton for multiplying and translating the Thue–Morse set

3.1 Introduction

The material of this chapter is taken from [22, 23, 24]. We continue to investigate decision problems related to numeration systems, but we focus here on the integer base. In this particular case, the decision problem 2.1.1 is rephrased : given an automaton accepting the language of the base- b expansions of a set $X \subseteq \mathbb{N}$, is it decidable whether X is a finite union of arithmetic progressions? As explained in the introduction of Chapter 2, several authors gave decision procedures for this problem ([4, 16, 43, 55]). Moreover, a multidimensional version of this problem was shown to be decidable in a way based on logical methods [16, 60, 49] (the multidimensional problem is to decide whether a b -recognizable subset of \mathbb{N}^d is definable within the Presburger arithmetic $\langle \mathbb{N}, + \rangle$).

With any set of integers X is naturally associated an infinite word, which is its characteristic sequence $\chi_X : n \mapsto 1$ if $n \in X$, $n \mapsto 0$ otherwise. Thus, to a finite union of arithmetic progressions corresponds an ultimately periodic infinite word. Therefore, the HD0L ultimate periodicity problem consisting in deciding whether a given morphic word (i.e. the image under a coding of the fixed point of a morphism) is ultimately periodic is a generalization of the periodicity problem for b -recognizable sets mentioned in the previous paragraph. The HD0L ultimate periodicity problem was shown to be decidable in its full generality [33, 58]. The proofs rely on return words, primitive

substitutions or evolution of Rauzy graphs. However, these methods do not provide algorithms that could be easily implemented. In addition, they do not allow us to obtain an algorithm for the multidimensional generalization of the periodicity problem. Therefore, a better understanding of the inner structure of automata arising from numeration systems remains a powerful tool to obtain efficient decision procedures.

Recall that the general idea is as follows. Suppose that $\mathcal{L} = \{L_i : i \in \mathbb{N}\}$ is a collection of regular languages and that we want to decide whether some particular regular language L belongs to \mathcal{L} . Now, suppose that we are able to explicitly give a lower bound on the state complexities (see Definition 1.2.24) of the languages in \mathcal{L} , i.e. for each given N , we can effectively produce a bound $B(N)$ such that for all $i > B(N)$, the state complexity of L_i is greater than N . Then the announced problem is decidable: if N is the state complexity of the given language L , then only the finitely many languages $L_0, \dots, L_{B(N)}$ have to be compared with L .

The state complexity of a b -recognizable set (i.e. the number of states of the minimal automaton accepting the b -expansions of its elements, cf. Definition 1.3.37) is closely related to the length of the logical formula describing this set. Short formulas are crucial in order to produce efficient mechanical proofs by using for example the Walnut software [59, 74]. There are several ways to improve the previous decision procedure. One of them is to use precise knowledge of the structure of the involved automata. This idea was successfully used in the papers [13, 55]. In [29], the structure of automata accepting the greedy expansions of $m\mathbb{N}$ for a wide class of non-standard numeration systems, and in particular, estimations of the state complexity of $\text{rep}(m\mathbb{N})$ are given. Another way of improving this procedure is to have at our disposal the exact state complexities of the languages in \mathcal{L} . Finding an exact formula is a much more difficult problem than finding good estimates. However, some results in this direction are known. For instance, it is proved in [29] that for the Fibonacci numeration system (see Example 1.3.4), the state complexity of $m\mathbb{N}$ is exactly $2m^2$. A complete description of the minimal automaton recognizing $m\mathbb{N}$ in any integer base b was given in [1] and the state complexity of $m\mathbb{N}$ with respect to the base b is shown to be exactly

$$\frac{m}{\gcd(m, b^N)} + \sum_{t=0}^{N-1} \frac{b^t}{\gcd(m, b^t)}$$

where N is the smallest integer α such that $\frac{m-b^\alpha}{\gcd(m, b^\alpha)} < \frac{m}{\gcd(m, b^{\alpha+1})}$.

For all the above mentioned reasons, the study of the state complexity of b -recognizable sets deserves special interest. In the present chapter, we

propose ourselves to initiate a study of the state complexity of sets of the form $mX + r$, for any recognizable subset X of \mathbb{N} (with respect to a given base b), any multiple m and any remainder r . In doing so, we aim at generalizing the previous framework concerning the case $X = \mathbb{N}$ only. Our study starts with the Thue-Morse set \mathcal{T} of the so-called *evil numbers* [2], i.e. the natural numbers whose base-2 expansions contain an even number of occurrences of the digit 1. The characteristic sequence of this set corresponds to the ubiquitous Thue-Morse word $\mathbf{t} = 0110100110010110 \dots$, which is the fixed point starting with 0 of the morphism $\theta : 0 \mapsto 01, 1 \mapsto 10$, cf. Example 1.1.15. This infinite word is one of the archetypical aperiodic automatic words, see the surveys [5, 64]. Many number-theoretic works devoted to sets of integers defined thanks to the Thue-Morse word exist, such as the study of additive and multiplicative properties, or iterations and sums of such sets [3, 17, 57]. In this vein, the set \mathcal{T} seems to be a natural candidate to start with. The goal of this chapter is to provide a complete characterization of the minimal automata recognizing the sets $m\mathcal{T} + r$ for any multiple m and remainder r , and any base b which is a power of 2 (other bases are not relevant with the choice of the Thue-Morse set in view of Cobham's theorem). In fact, this work was first done in the special case where $r = 0$ (see [22]). Surprisingly, the description of the left quotients (i.e. the states of the minimal automaton, see Definition 1.2.19) are quite different when $r = 0$ than in the general case.

This chapter has the following organization. In Section 3.2, we state our main result and expose the method that will be carried out for its proof. More precisely, we present the steps of our construction of the minimal automaton accepting the base- 2^p expansions of the elements of $m\mathcal{T} + r$ for any positive integers p and m , and any remainder $r \in \llbracket 0, m-1 \rrbracket$. In Section 3.3, we give the details of the construction of the intermediate automata. In particular, we study the transitions of each automaton. Thus, at the end of Section 3.3, we are provided with an automaton recognizing the desired language. Then in Section 3.4, we study the properties of the built automata that will be needed for proving the announced state complexity result. The minimization procedure of the last automaton is handled in Section 3.5. This part is the most technical one and it deeply relies on the properties of the intermediate automata proved in the previous sections. In Section 3.6, we show that as an application of our results, we obtain a procedure to decide whether a 2^p -recognizable set given via an automaton is a set of the form $m\mathcal{T} + r$. In Section 3.7, we explicitly give the correspondence between the description of the minimal automaton recognizing $m\mathcal{T}$ obtained in [22] and that given in the present work in the particular case where $r = 0$. In Section 3.8, we show that the minimal automaton recognizing $m\overline{\mathcal{T}} + r$, where $\overline{\mathcal{T}}$ is the complement of the

Thue-Morse set \mathcal{T} , is obtained directly from the one recognizing $m\mathcal{T}+r$ by moving the initial state. As a consequence, the state complexities of $m\overline{\mathcal{T}}+r$ and $m\mathcal{T}+r$ coincide. Finally, in Section 3.9, we discuss future work and give two related open problems.

3.2 Method

The *Thue-Morse set*, denoted by \mathcal{T} , is the set of all natural numbers whose base-2 expansion contains an even number of occurrences of the digit 1:

$$\mathcal{T} = \{n \in \mathbb{N} : |\text{rep}_2(n)|_1 \in 2\mathbb{N}\}.$$

The Thue-Morse set \mathcal{T} is 2-recognizable since the language $\text{val}_2^{-1}(\mathcal{T})$ is accepted by the automaton depicted in Figure 3.1.

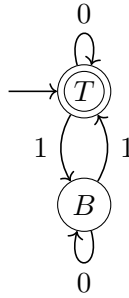


Figure 3.1: The Thue-Morse set is 2-recognizable.

More precisely, the Thue-Morse set \mathcal{T} is 2^p -recognizable for all $p \in \mathbb{N}_0$ and is not b -recognizable for any other base b . This is a consequence of the famous Cobham’s theorem (see Theorem 1.3.36). Indeed, it is easily seen that, for each $p \in \mathbb{N}_0$, the language $\text{val}_{2^p}^{-1}(\mathcal{T})$ is accepted by the DFA $(\{T, B\}, T, \{T\}, \Sigma_{2^p}, \delta)$ where for all $X \in \{T, B\}$ and all $a \in \Sigma_{2^p}$,

$$\delta(X, a) = \begin{cases} X & \text{if } a \in \mathcal{T} \\ \overline{X} & \text{else,} \end{cases}$$

where $\overline{T} = B$ and $\overline{B} = T$. For example this automaton is depicted in Figure 3.2 for $p = 2$.

Remark 3.2.1. The notation T and B stand for “Top” and “Bottom”. Representing these automata vertically will be helpful in what happens next.

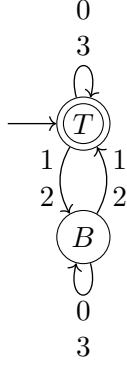


Figure 3.2: The Thue-Morse set is 4-recognizable.

Notation 3.2.2. In order to avoid a systematic case separation, we introduce the following notation: for $X \in \{T, B\}$ and $n \in \mathbb{N}$, we define

$$X_n = \begin{cases} X & \text{if } n \in \mathcal{T} \\ \overline{X} & \text{else.} \end{cases}$$

With this notation, we can simply rewrite the definition of the transition function δ as $\delta(X, a) = X_a$.

Thanks to Proposition 1.3.38, for any $m, t \in \mathbb{N}$ and $p \in \mathbb{N}_{\geq 1}$, the set $m\mathcal{T} + t$ is 2^p -recognizable. The aim of this chapter is to show the following result.

Theorem 3.2.3. *Let m, p be positive integers and $r \in \llbracket 0, m-1 \rrbracket$. Then the state complexity of $m\mathcal{T} + r$ with respect to the base 2^p is equal to*

$$2k + \left\lceil \frac{z}{p} \right\rceil$$

if $m = k2^z$ with k odd.

As an example, if m is odd, then the state complexity of $m\mathcal{T} + r$ (where $r \in \llbracket 0, m-1 \rrbracket$) is always $2m$. The minimal automaton recognizing $3\mathcal{T}$ in base 2 is depicted in Figure 3.3, it has $2 \cdot 3 = 6$ states as expected. In Figure 3.13, one can find the minimal automaton of $\text{val}_4^{-1}(6\mathcal{T} + 2)$, with $2 \cdot 3 + \lceil \frac{1}{2} \rceil = 7$ states.

Our proof of Theorem 3.2.3 is constructive. In order to describe the minimal DFA of $\text{val}_{2^p}^{-1}(m\mathcal{T} + r)$, we will successively construct several automata.

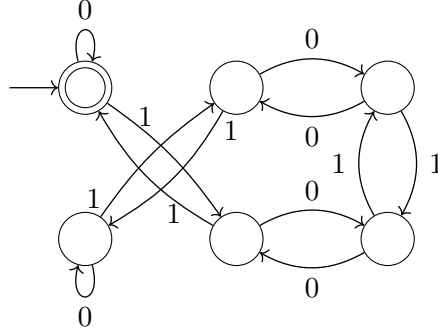


Figure 3.3: The minimal automaton of $\text{val}_2^{-1}(3\mathcal{T})$.

Except for the last one, the constructed automata accept pairs of representations of integers. First, we build a DFA $\mathcal{A}_{\mathcal{T},2^p}$ accepting the language

$$\text{val}_{2^p}^{-1}(\mathcal{T} \times \mathbb{N}).$$

Then we build a DFA $\mathcal{A}_{m,r,b}$ accepting the language

$$\text{val}_b^{-1}(\{(n, mn + r) : n \in \mathbb{N}\}).$$

Note that we do the latter step for any integer base b and not only for powers of 2. Next, we consider the product automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$. This DFA accepts the language

$$\text{val}_{2^p}^{-1}(\{(t, mt + r) : t \in \mathcal{T}\}).$$

Finally, a finite automaton $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$ accepting $\text{val}_{2^p}^{-1}(m\mathcal{T} + r)$ is obtained by projecting the label of each transition in $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ onto its second component. At each step of our construction, we check that the automaton under consideration is minimal (and hence deterministic) and the ultimate step precisely consists in a minimization procedure.

Notation 3.2.4. From now on, we fix some positive integers m, p and some remainder $r \in \llbracket 0, m-1 \rrbracket$. We also let z and k be the unique integers such that $m = k2^z$ with k odd. Finally we let $R = |\text{rep}_{2^p}(r)|$ and $N = \max\{\lceil \frac{z}{p} \rceil, R\}$.

3.3 Construction of the intermediate automata

3.3.1 The automaton $\mathcal{A}_{\mathcal{T},2^p}$

First, we build a DFA $\mathcal{A}_{\mathcal{T},2^p}$ accepting the language $\text{val}_{2^p}^{-1}(\mathcal{T} \times \mathbb{N})$. This DFA is a modified version of the automaton accepting $\text{val}_{2^p}^{-1}(\mathcal{T})$ defined in

the previous section. Namely, we replace each transition labelled by $a \in \Sigma_{2^p}$ by 2^p copies of itself labelled by (a, b) , for each $b \in \Sigma_{2^p}$. Formally,

$$\mathcal{A}_{\mathcal{T}, 2^p} = (\{T, B\}, T, \{T\}, \Sigma_{2^p} \times \Sigma_{2^p}, \delta_{\mathcal{T}, 2^p})$$

where, for all $X \in \{T, B\}$ and all $a, b \in \Sigma_{2^p}$, we have $\delta_{\mathcal{T}, 2^p}(X, (a, b)) = X_a$. For example, the automata $\mathcal{A}_{\mathcal{T}, 2}$ and $\mathcal{A}_{\mathcal{T}, 4}$ are depicted in Figure 3.4.

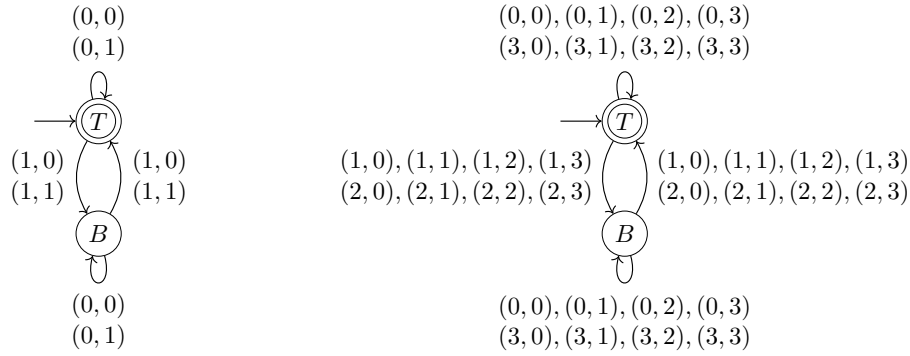


Figure 3.4: The automata $\mathcal{A}_{\mathcal{T}, 2}$ (left) and $\mathcal{A}_{\mathcal{T}, 4}$ (right).

Lemma 3.3.1. *Let $u, v \in \Sigma_{2^p}^*$. Then $\text{val}_{2^p}(uv) \in \mathcal{T}$ if and only if, either $\text{val}_{2^p}(u) \in \mathcal{T}$ and $\text{val}_{2^p}(v) \in \mathcal{T}$, or $\text{val}_{2^p}(u) \notin \mathcal{T}$ and $\text{val}_{2^p}(v) \notin \mathcal{T}$.*

Proof. Let $\tau: \Sigma_{2^p}^* \rightarrow \Sigma_{2^p}^*$ be the p -uniform morphism defined on every letter $a \in \Sigma_{2^p}$ by $\tau(a) = 0^{p-|\text{rep}_2(a)|} \text{rep}_2(a)$. Then, for all $w \in \Sigma_{2^p}^*$, we have $\text{val}_{2^p}(w) = \text{val}_2(\tau(w))$. Therefore, $\text{val}_{2^p}(w) \in \mathcal{T}$ if and only if $|\tau(w)|_1 \in 2\mathbb{N}$. Since τ is a morphism, we have $|\tau(uv)|_1 = |\tau(u)|_1 + |\tau(v)|_1$. Hence $|\tau(uv)|_1$ is even if and only if $|\tau(u)|_1$ and $|\tau(v)|_1$ are both even or both odd. \square

Lemma 3.3.2. *For all $X \in \{T, B\}$ and $(u, v) \in (\Sigma_{2^p} \times \Sigma_{2^p})^*$, we have*

$$\delta_{\mathcal{T}, 2^p}(X, (u, v)) = X_{\text{val}_{2^p}(u)}.$$

Proof. We do the proof by induction on $|(u, v)|$. The case $|(u, v)| = 0$ is trivial. Now let $X \in \{T, B\}$ and let $(ua, vb) \in (\Sigma_{2^p} \times \Sigma_{2^p})^*$ with $a, b \in \Sigma_{2^p}$. We suppose that the result is satisfied for (u, v) and we show that it is also true for (ua, vb) . Let $Y = \delta_{\mathcal{T}, 2^p}(X, (u, v))$. By induction hypothesis, we have $Y = X_{\text{val}_{2^p}(u)}$. Thus we obtain

$$\delta_{\mathcal{T}, 2^p}(X, (ua, vb)) = \delta_{\mathcal{T}, 2^p}(Y, (a, b)) = Y_a = (X_{\text{val}_{2^p}(u)})_a = X_{\text{val}_{2^p}(ua)}$$

where we have used Lemma 3.3.1 for the last step. \square

3.3.2 The automaton $\mathcal{A}_{m,r,b}$

In this section, we consider an arbitrary integer base $b \geq 2$. Let

$$\mathcal{A}_{m,r,b} = (\llbracket 0, m-1 \rrbracket, 0, \{r\}, \Sigma_b \times \Sigma_b, \delta_{m,r,b})$$

where the (partial) transition function $\delta_{m,r,b}$ is defined as follows: for all $i, j \in \llbracket 0, m-1 \rrbracket$ and $d, e \in \Sigma_b$, we set

$$\delta_{m,r,b}(i, (d, e)) = j \iff bi + e = md + j.$$

The DFA $\mathcal{A}_{m,r,b}$ accepts the language $\text{val}_b^{-1}(\{(n, mn + r) : n \in \mathbb{N}\})$. We refer the interested reader to [75]. For example, the automaton $\mathcal{A}_{6,2,4}$ is depicted in Figure 3.5.

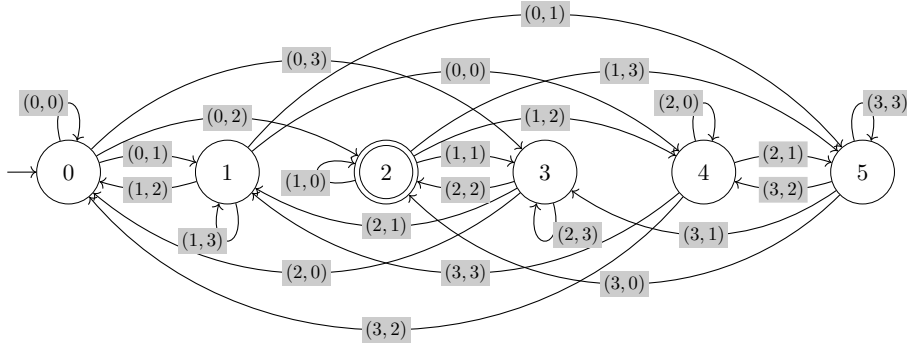


Figure 3.5: The automaton $\mathcal{A}_{6,2,4}$ accepts the language $\text{val}_4^{-1}(\{(n, 6n + 2) : n \in \mathbb{N}\})$.

Note that the automaton $\mathcal{A}_{m,r,b}$ is not complete (see Lemma 3.3.3). Also note that there is always a loop labelled by $(0, 0)$ on the initial state 0.

Lemma 3.3.3. *For each $i \in \llbracket 0, m-1 \rrbracket$ and $e \in \Sigma_b$, there exist unique $d \in \Sigma_b$ and $j \in \llbracket 0, m-1 \rrbracket$ such that $\delta_{m,r,b}(i, (d, e)) = j$.*

Proof. Indeed, d and j are unique since they are the quotient and remainder of the Euclidean division of $bi + e$ by m . We still have to check that $d < b$. We have

$$bi + e = md + j \iff d = \frac{bi + e - j}{m}.$$

Since $i \leq m-1$, $j \geq 0$ and $e < b$, we have

$$\frac{bi + e - j}{m} < \frac{b(m-1) + b}{m} = b.$$

□

Lemma 3.3.4. *For $i, j \in \llbracket 0, m-1 \rrbracket$ and $(u, v) \in (\Sigma_b \times \Sigma_b)^*$, we have*

$$\delta_{m,r,b}(i, (u, v)) = j \iff b^{|(u,v)|} i + \text{val}_b(v) = m \text{val}_b(u) + j.$$

Proof. We do the proof by induction on $n = |(u, v)|$. If n is equal to 0, the result is clear. Now let $i, j \in \llbracket 0, m-1 \rrbracket$ and let $(du, ev) \in (\Sigma_b \times \Sigma_b)^*$ with $d, e \in \Sigma_b$ and $|(u, v)| = n$. We suppose that the result is satisfied for (u, v) and we show that it is also true for (du, ev) . We use the notation $\text{DIV}(x, y)$ and $\text{MOD}(x, y)$ to designate the quotient and the remainder of the Euclidean division of x by y (thus, we have $\text{DIV}(x, y) = \lfloor \frac{x}{y} \rfloor$). By definition of the transition function, we have $\delta_{m,r,b}(i, (du, ev)) = j$ if and only if

$$d = \text{DIV}(bi + e, m) \text{ and } \delta_{m,r,b}(\text{MOD}(bi + e, m), (u, v)) = j.$$

By using the induction hypothesis, we have

$$\begin{aligned} \delta_{m,r,b}(bi + e - md, (u, v)) &= j \\ \iff b^n (bi + e - md) + \text{val}_b(v) &= m \text{val}_b(u) + j \\ \iff b^{n+1} i + \text{val}_b(ev) &= m \text{val}_b(du) + j. \end{aligned}$$

To be able to conclude the proof, we still have to show that

$$b^{n+1} i + \text{val}_b(ev) = m \text{val}_b(du) + j \tag{3.1}$$

implies

$$d = \text{DIV}(bi + e, m).$$

Thus, suppose that (3.1) is true. Then

$$b^{n+1} i + b^n e + \text{val}_b(v) = m(b^n d + \text{val}_b(u)) + j.$$

Since $\text{val}_b(u)$ and $\text{val}_b(v)$ are less than b^n and since $d \geq 0$, $j < m$ and $b^n d + \text{val}_b(u) \geq 0$, we obtain

$$\begin{aligned} d &= \text{DIV}(b^n d + \text{val}_b(u), b^n) \\ &= \text{DIV}(\text{DIV}(b^{n+1} i + b^n e + \text{val}_b(v), m), b^n) \\ &= \text{DIV}(\text{DIV}(b^{n+1} i + b^n e + \text{val}_b(v), b^n), m) \\ &= \text{DIV}(bi + e, m) \end{aligned}$$

as desired. \square

It is easily checked that Lemma 3.3.3 extends from letters to words.

Lemma 3.3.5. *For each $i \in \llbracket 0, m-1 \rrbracket$ and $v \in \Sigma_b^*$, there exist unique $u \in \Sigma_b^*$ and $j \in \llbracket 0, m-1 \rrbracket$ such that $\delta_{m,r,b}(i, (u, v)) = j$. In particular, the word u must have the same length as the word v , and hence $\text{val}_b(u) < b^{|v|}$.*

3.3.3 The projected automaton $\Pi(\mathcal{A}_{m,r,b})$

In this section again, $b \geq 2$ is an arbitrary integer base. We consider the automaton obtained by projecting the label of each transition of $\mathcal{A}_{m,r,b}$ onto its second component. We denote by $\Pi(\mathcal{A}_{m,r,b})$ the automaton obtained thanks to this projection. Thanks to Lemma 3.3.3, the automaton $\Pi(\mathcal{A}_{m,r,b})$ is deterministic and complete. We denote by $\delta_{m,r,b}^\Pi$ the corresponding transition function. For example, the automaton $\Pi(\mathcal{A}_{6,2,4})$ is depicted in Figure 3.6.

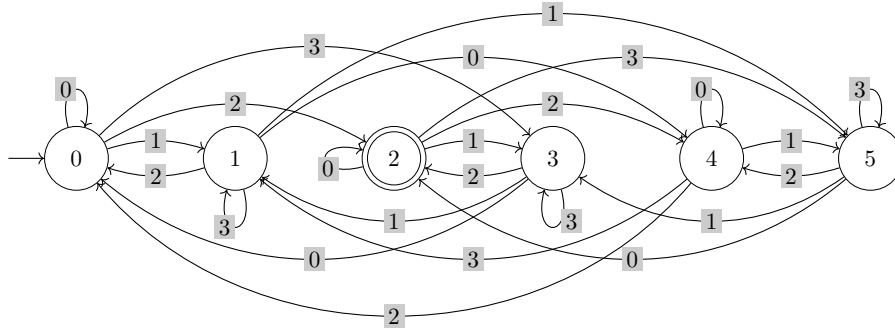


Figure 3.6: The projected automaton $\Pi(\mathcal{A}_{6,2,4})$.

Note that this automaton corresponds actually the “classical” construction of an automaton accepting $m\mathbb{N} + r$ in base b , see for example [1]. Indeed, in this natural construction, states are any possible remainder $j \bmod m$. The initial state is 0 and the only final state is r . Finally, transitions are defined as follows: from a state $i \in \llbracket 0, m-1 \rrbracket$, there is a transition of label $e \in \Sigma_b$ to the state $bi + e \bmod m$.

Lemma 3.3.6. *For $i, j \in \llbracket 0, m-1 \rrbracket$ and $v \in \Sigma_b^*$, we have*

$$\delta_{m,r,b}^\Pi(i, v) = j \iff b^{|v|}i + \text{val}_b(v) \equiv j \pmod{m}.$$

Proof. Let $i, j \in \llbracket 0, m-1 \rrbracket$ and $v \in \Sigma_b^*$. If $\delta_{m,r,b}^\Pi(i, v) = j$, then there is a word u of the same length as v such that $\delta_{m,r,b}(i, (u, v)) = j$. By Lemma 3.3.4, we get $b^{|v|}i + \text{val}_b(v) \equiv j \pmod{m}$. Conversely, suppose that there exists some $\ell \in \mathbb{N}$ such that $b^{|v|}i + \text{val}_b(v) = m\ell + j$. Since $i \leq m-1$, $\text{val}_b(v) < b^{|v|}$ and $j \geq 0$, we necessarily have

$$\ell = \frac{b^{|v|}i + \text{val}_b(v) - j}{m} < \frac{b^{|v|}(m-1) + b^{|v|}}{m} = b^{|v|}.$$

Hence $|\text{rep}_b(\ell)| \leq |v|$ and the word $u = 0^{|v|-|\text{rep}_b(\ell)|} \text{rep}_b(\ell)$ has length $|v|$ and is such that $\text{val}_b(u) = \ell$. The conclusion follows from Lemma 3.3.4. \square

3.3.4 The product automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$

In this section, we study the product automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$. Since the states of $\mathcal{A}_{m,r,2^p}$ are numbered from 0 to $m-1$ and those of $\mathcal{A}_{\mathcal{T},2^p}$ are T and B , we denote the states of the product automaton by

$$(0, T), \dots, (m-1, T) \text{ and } (0, B), \dots, (m-1, B).$$

The transitions of $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ are defined as follows. For $i, j \in \llbracket 0, m-1 \rrbracket$, $X, Y \in \{T, B\}$ and $d, e \in \Sigma_{2^p}$, there is a transition labelled by (d, e) from the state (i, X) to the state (j, Y) if and only if

$$2^p i + e = md + j \quad \text{and} \quad Y = X_d.$$

We denote by δ_{\times} the (partial) transition function of this product automaton. The initial state is $(0, T)$ and the only final state is (r, T) .

Lemma 3.3.7. *For all $i, j \in \llbracket 0, m-1 \rrbracket$, $X, Y \in \{T, B\}$ and all pair of words $(u, v) \in (\Sigma_{2^p} \times \Sigma_{2^p})^*$, we have $\delta_{\times}((i, X), (u, v)) = (j, Y)$ if and only if*

$$2^p |(u, v)| i + \text{val}_{2^p}(v) = m \text{val}_{2^p}(u) + j \quad \text{and} \quad Y = X_{\text{val}_{2^p}(u)}.$$

Proof. It is enough to combine Lemmas 3.3.2 and 3.3.4. \square

In Figure 3.7, we have depicted the automaton $\mathcal{A}_{6,2,4} \times \mathcal{A}_{\mathcal{T},4}$, as well as the automata $\mathcal{A}_{6,2,4}$ and $\mathcal{A}_{\mathcal{T},4}$, which have been placed in such a way that the labels of the product automaton can be easily deduced. Here and in the next figures, states are named iX instead of (i, X) for clarity.

3.3.5 The projection $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$ of the product automaton

Now, we provide a DFA accepting the language $\text{val}_{2^p}^{-1}(m\mathcal{T} + r)$. This automaton is denoted by $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$ and is defined from the automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ by only keeping the second component of the label of each transition. Formally, the states of $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$ are

$$(0, T), \dots, (m-1, T) \text{ and } (0, B), \dots, (m-1, B),$$

the initial state is $(0, T)$, the only final state is (r, T) , and the transitions are defined as follows. For $i, j \in \llbracket 0, m-1 \rrbracket$, $X, Y \in \{T, B\}$ and $e \in \Sigma_{2^p}$, there is a transition labelled by e from the state (i, X) to the state (j, Y) if and only if there exists $d \in \Sigma_{2^p}$ such that

$$2^p i + e = md + j \quad \text{and} \quad Y = X_d.$$

We denote by δ_{\times}^{Π} the (partial) transition function of this product automaton.

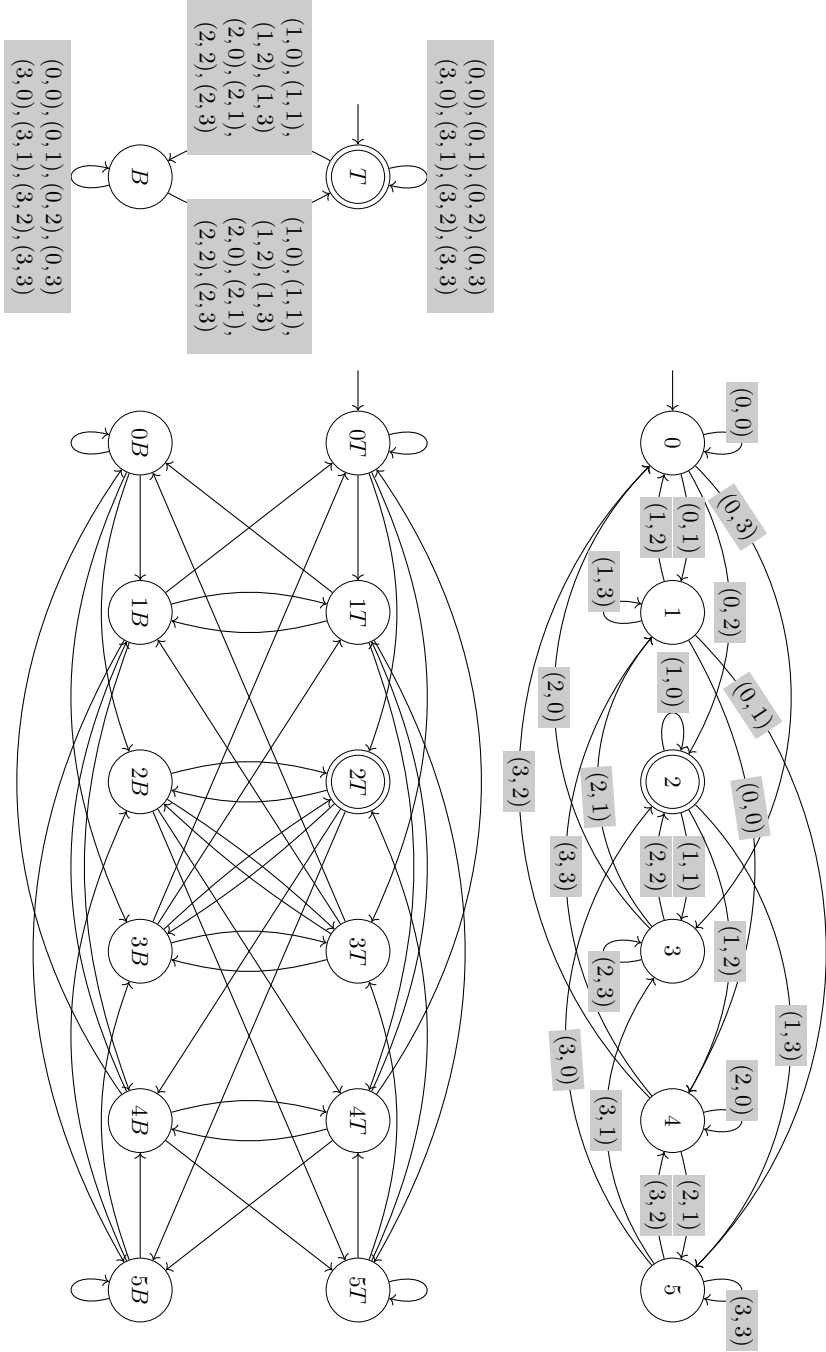


Figure 3.7: The product automaton $\mathcal{A}_{6,2,4} \times \mathcal{A}_{T,4}$

Example 3.3.8. The automata $\mathcal{A}_{6,2,4} \times \mathcal{A}_{\mathcal{T},4}$ and $\Pi(\mathcal{A}_{6,2,4} \times \mathcal{A}_{\mathcal{T},4})$ are depicted in Figures 3.7 and 3.8 respectively. In Figure 3.8, all edges labelled by 0 (1, 2 and 3 respectively) are represented in black (blue, red and green respectively).

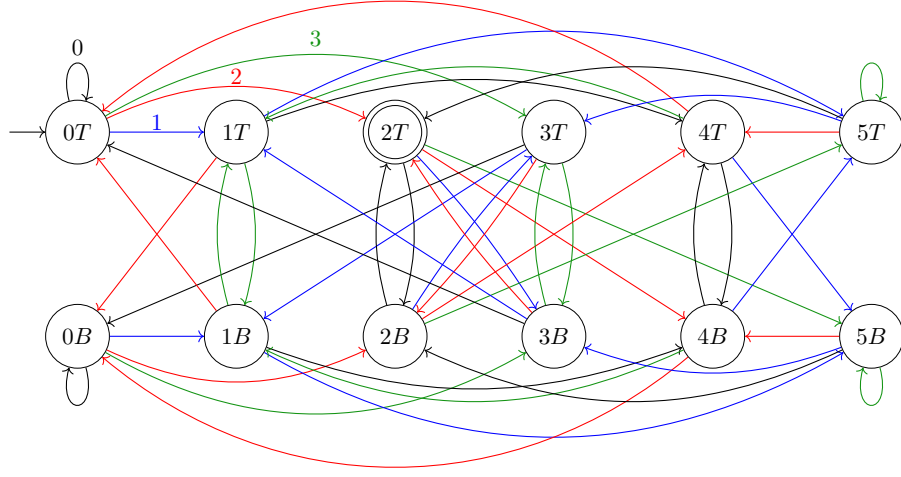


Figure 3.8: The projected automaton $\Pi(\mathcal{A}_{6,2,4} \times \mathcal{A}_{\mathcal{T},4})$.

Lemma 3.3.9. For all $i, j \in \llbracket 0, m-1 \rrbracket$, $X, Y \in \{T, B\}$ and $v \in \Sigma_{2^p}^*$, we have $\delta_{\times}^{\Pi}((i, X), v) = (j, Y)$ if and only if there exists $d \in \mathbb{N}$ such that

$$2^{p|v|}i + \text{val}_{2^p}(v) = md + j \quad \text{and} \quad Y = X_d.$$

Remark 3.3.10. Note that in the statement of Lemma 3.3.9, the integer d is necessarily less than $2^{p|v|}$. This is due to the fact that if v is the label of some path in the projected automaton $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$, there must be a word u of the same length d as v such that the pair (u, v) is the label of a path in the automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$. This can be deduced directly from the computation: if $2^{p|v|}i + \text{val}_{2^p}(v) = md + j$ (with $i, j \in \llbracket 0, m-1 \rrbracket$) then

$$d = \frac{2^{p|v|}i + \text{val}_{2^p}(v) - j}{m} < \frac{2^{p|v|}(i+1)}{m} \leq 2^{p|v|}.$$

Proof of Lemma 3.3.9. We have $\delta_{\times}^{\Pi}((i, X), v) = (j, Y)$ if and only if there exists some word $u \in \Sigma_{2^p}^*$ with $|u| = |v|$ such that $\delta_{\times}((i, X), (u, v)) = (j, Y)$. Take $d = \text{val}_{2^p}(u)$. The conclusion follows from Lemma 3.3.7, a similar argument as in the proof of Lemma 3.3.6 and Remark 3.3.10. \square

3.4 Properties of the intermediate automata

Now we prove some properties of the automata $\mathcal{A}_{\mathcal{T},2^p}$, $\mathcal{A}_{m,r,b}$, $\Pi(\mathcal{A}_{m,r,b})$, $\mathcal{A}_{\mathcal{T},2^p} \times \mathcal{A}_{m,r,2^p}$ and $\Pi(\mathcal{A}_{\mathcal{T},2^p} \times \mathcal{A}_{m,r,2^p})$ that will be useful for our concerns.

3.4.1 Properties of $\mathcal{A}_{\mathcal{T},2^p}$

Lemma 3.4.1. *For all $X, Y \in \{T, B\}$ and $(u, v) \in (\Sigma_{2^p} \times \Sigma_{2^p})^*$, we have*

$$\delta_{\mathcal{T},2^p}(X, (u, v)) = Y \iff \delta_{\mathcal{T},2^p}(\overline{X}, (u, v)) = \overline{Y}.$$

Proof. This directly follows from Lemma 3.3.2. \square

Proposition 3.4.2. *The automaton $\mathcal{A}_{\mathcal{T},2^p}$ is deterministic, complete, accessible, coaccessible and has disjoint states. In particular, it is the minimal automaton of $\text{val}_{2^p}^{-1}(\mathcal{T} \times \mathbb{N})$.*

Proof. These properties are all straightforward verifications. \square

3.4.2 Properties of $\mathcal{A}_{m,r,b}$

In this section, we study the properties of the automaton $\mathcal{A}_{m,r,b}$ for any integer base b , not especially for a base which is a power of 2.

Lemma 3.4.3. *For each $i \in \llbracket 0, m-1 \rrbracket$, there is a path from the state i to the state 0 in the automaton $\mathcal{A}_{m,r,b}$.*

Proof. Let $i \in \llbracket 0, m-1 \rrbracket$. We are looking for a word (u, v) that leads from i to 0 in $\mathcal{A}_{m,r,b}$, that is such that $b^ni + \text{val}_b(v) = m \text{val}_b(u)$ where $n = |(u, v)|$. Equivalently, we have to find $n \in \mathbb{N}$ and $d, e \in \llbracket 0, b^n-1 \rrbracket$ such that we have $b^ni + e = md$.

We claim that for all non-negative integer n and $i \in \llbracket 0, m-1 \rrbracket$, there are $d, e \in \llbracket 0, b^n-1 \rrbracket$ such that $b^ni + e = md$ if and only if the following two inequalities hold

$$\left\lceil \frac{b^ni}{m} \right\rceil - \frac{b^n}{m} < \frac{b^ni}{m} \leq b^n - 1. \quad (3.2)$$

First, suppose that $d, e \in \llbracket 0, b^n-1 \rrbracket$ are such that $b^ni + e = md$. Then we obtain that $\frac{b^ni}{m} = d - \frac{e}{m} \leq d \leq b^n - 1$. Moreover $\frac{b^ni}{m} \leq d = \frac{b^ni+e}{m} < \frac{b^n(i+1)}{m}$. Since d is an integer, we get that $\left\lceil \frac{b^ni}{m} \right\rceil < \frac{b^n(i+1)}{m}$. Conversely, suppose that the two inequalities (3.2) hold. Let $d = \left\lceil \frac{b^ni}{m} \right\rceil$ and $e = md - b^ni$. It is enough to show that $d, e \in \llbracket 0, b^n-1 \rrbracket$. Clearly $d, e \in \mathbb{N}$. From the inequality on the right, we get $d \leq b^n - 1$ and from that on the left, we get

$$e = md - b^ni < b^n(i+1) - b^ni = b^n.$$

This proves the claim.

For a given $i \in \llbracket 0, m-1 \rrbracket$, the inequalities (3.2) may not be satisfied for small n but they are both satisfied for all n large enough. Indeed, for the first inequality of (3.2), we have $\lceil \frac{b^n i}{m} \rceil - \frac{b^n}{m} < \frac{b^n i}{m} + 1 - \frac{b^n}{m} = \frac{b^n(i-1)+m}{m}$. But $\frac{b^n(i-1)+m}{m} < \frac{b^n i}{m}$ if and only if $m < b^n$. For the second inequality, we have $\frac{b^n i}{m} \leq \frac{b^n(m-1)}{m} = b^n - \frac{b^n}{m}$. But $b^n - \frac{b^n}{m} < b^n - 1$ if and only if $m < b^n$. Hence, $m < b^n$ implies (3.2). Since m and b are fixed and $b \geq 1$, for large enough n , (3.2) holds. \square

Proposition 3.4.4. *The automaton $\mathcal{A}_{m,r,b}$ is deterministic, accessible, coaccessible and has disjoint states.*

Proof. The automaton $\mathcal{A}_{m,r,b}$ is clearly deterministic (cf. Lemma 3.3.3). Then, by Lemma 3.3.4, for all $i \in \llbracket 0, m-1 \rrbracket$, we have $\delta_{m,r,b}(0, \text{rep}_b(0, i)) = i$. Therefore, $\mathcal{A}_{m,r,b}$ is accessible.

Let us now prove coaccessibility. Let $i \in \llbracket 0, m-1 \rrbracket$. By Lemma 3.4.3, there is a path from the state i to the state 0. Moreover, thanks to accessibility, there is a path from state 0 to state r . Thus it is enough to concatenate the two paths to obtain a path from state i to state r .

Finally, let $i, j \in \llbracket 0, m-1 \rrbracket$ and let $(u, v) \in L_i \cap L_j$. By Lemma 3.3.4, we have

$$b^{|(u,v)|}i + \text{val}_b(v) = m \text{val}_b(u) + r \quad \text{and} \quad b^{|(u,v)|}j + \text{val}_b(v) = m \text{val}_b(u) + r,$$

which implies that $i = j$. We have thus obtained that $i \neq j \Rightarrow L_i \cap L_j = \emptyset$, i.e. that $\mathcal{A}_{m,r,b}$ has disjoint states. \square

Recall that in a reduced DFA, there can be at most one non-coaccessible state, cf. Remark 1.2.9. Thus, we deduce from Proposition 3.4.4 that the *trim minimal* automaton of the language $\text{val}_b^{-1}(\{(n, mn + r) : n \in \mathbb{N}\})$ is indeed $\mathcal{A}_{m,r,b}$, that is the automaton obtained by removing the only non-coaccessible state from its minimal automaton.

3.4.3 Properties of $\Pi(\mathcal{A}_{m,r,b})$

Proposition 3.4.5. *The automaton $\Pi(\mathcal{A}_{m,r,b})$ is deterministic, complete, accessible and coaccessible.*

Proof. The accessibility and coaccessibility of the automaton $\Pi(\mathcal{A}_{m,r,b})$ are straightforward consequences of Proposition 3.4.4. In order to see that it is deterministic and complete, observe that for every state $i \in \llbracket 0, m-1 \rrbracket$ and every digit $e \in \Sigma_b$, there is a transition from i to the state $bi + e \pmod{m}$ labelled by e (see Lemma 3.3.3). \square

The automaton $\Pi(\mathcal{A}_{m,r,b})$ is not minimal in general: it is minimal if and only if m and b are coprime or if $m = 2$. This is a corollary of the main theorem in [1]. We give here a direct proof. In view of Proposition 3.4.5, it is enough to show that the automaton $\Pi(\mathcal{A}_{m,r,b})$ is reduced if and only if m and b are coprime or $m = 2$. The case $m = 2$ is trivial, since the automaton is composed of two states, one is final and the other one is not. If m and b are coprime, we have a stronger property than minimality, as shown in Proposition 3.4.6 below (recall Definition 1.2.8). Now turn to the converse. Suppose that $\gcd(m, b) \neq 1$ and that $m \neq 2$. Then one can find a prime p and non-negative integers q, n, α, β such that $b = qp^\alpha, m = np^\beta$. Then if $e \in \llbracket 0, b-1 \rrbracket$, the transitions from states 0 and $\frac{m}{p}$ of label e lead to the same state. Indeed, one has

$$b \cdot 0 + e \equiv e \pmod{m} \quad \text{and} \quad b \frac{m}{p} + e \equiv qp^{\alpha-1}m + e \equiv e \pmod{m},$$

and we then apply Lemma 3.3.6. A similar computation prove that the states 1 and $\frac{m}{p} + 1$ reach the same states when reading a letter $e \in \llbracket 0, b-1 \rrbracket$. Thus the only possible way for the states 0 and $\frac{m}{p}$ (resp. 1 and $\frac{m}{p} + 1$) to be distinguishable is that one is final and the other is not. Since $m \neq 2$, one has $\#\{0, 1, \frac{m}{p}, \frac{m}{p} + 1\} \geq 3$. If 0 is final, then the states 1 and $\frac{m}{p} + 1$ are indistinguishable and the automaton is not reduced (cf. Definition 1.2.8). If 1 is final and $\frac{m}{p} \neq 1$, the states 0 and $\frac{m}{p}$ are indistinguishable. If finally 1 is final and $\frac{m}{p} = 1$, the states 0 and $\frac{m}{p} + 1$ are indistinguishable.

Proposition 3.4.6. *If m and b are coprime, then the automaton $\Pi(\mathcal{A}_{m,r,b})$ has disjoint states and hence it is the minimal automaton of $\text{val}_b^{-1}(m\mathbb{N} + r)$.*

Proof. Let $i, j \in \llbracket 0, m-1 \rrbracket$ and let $v \in \Sigma_b^*$ be a word accepted from both i and j in $\Pi(\mathcal{A}_{m,r,b})$. By Lemma 3.3.5, there are unique words u and u' of the same length as v such that, in the automaton $\mathcal{A}_{m,r,b}$, (u, v) and (u', v) are accepted from i and j respectively. By Lemma 3.3.4, it is equivalent to say that

$$b^{|v|}i + \text{val}_b(v) = m \text{val}_b(u) + r \quad \text{and} \quad b^{|v|}j + \text{val}_b(v) = m \text{val}_b(u') + r.$$

Thus, we have

$$b^{|v|}i - m \text{val}_b(u) = b^{|v|}j - m \text{val}_b(u'). \quad (3.3)$$

Therefore $m \text{val}_b(u) \equiv m \text{val}_b(u') \pmod{b^{|v|}}$. By using the hypothesis of coprimality of m and b , we obtain that $\text{val}_b(u) \equiv \text{val}_b(u') \pmod{b^{|v|}}$. Since $\text{val}_b(u)$ and $\text{val}_b(u')$ are both less than $b^{|v|}$, we obtain $\text{val}_b(u) = \text{val}_b(u')$. Finally, we get from (3.3) that $i = j$. We have thus obtained that if $i \neq j$, then $L_i \cap L_j = \emptyset$, i.e. that $\Pi(\mathcal{A}_{m,r,b})$ has disjoint states. \square

To end this section, we prove some useful properties of the automaton $\Pi(\mathcal{A}_{m,r,b})$ under the more restrictive hypotheses of this work: $b = 2^p$ and $m = k2^z$ with k odd. In this particular case, assuming that $k > 1$, we are able to explicitly provide a word w_i that leads from the state i to the state 0 in the projected automaton $\Pi(\mathcal{A}_{m,r,b})$, see Lemma 3.4.8.

In the following, we set $K = |\text{rep}_{2^p}((k-1)2^z)|$, provided that $k > 1$. Afterwards we define a permutation σ of the integers in $\llbracket 0, k-1 \rrbracket$ by setting $\sigma(i) = -2^{pK-z}i \pmod{k}$. Note that σ permutes the integers $0, 1, \dots, k-1$ because k is odd. Further, we define w_i to be the unique word of length K representing $\sigma(i)2^z$ in base 2^p :

$$w_i = 0^{K-|\text{rep}_{2^p}(\sigma(i)2^z)|} \text{rep}_{2^p}(\sigma(i)2^z)$$

for each $i \in \llbracket 0, k-1 \rrbracket$. Note that the words w_i are well-defined since, by the choice of K , we have $\sigma(i)2^z \leq (k-1)2^z < 2^{pK}$ for every $i \in \llbracket 0, k-1 \rrbracket$.

Lemma 3.4.7. *If $k > 1$ then $pK \geq z$.*

Proof. We have

$$K = \lfloor \log_{2^p}((k-1)2^z) \rfloor + 1 = \left\lfloor \log_{2^p}(k-1) + \frac{z}{p} \right\rfloor + 1 \geq \left\lfloor \frac{z}{p} \right\rfloor + 1 \geq \left\lceil \frac{z}{p} \right\rceil.$$

Thus $pK \geq p \left\lceil \frac{z}{p} \right\rceil \geq z$. □

Lemma 3.4.8. *Suppose that $k > 1$ and let $i \in \llbracket 0, k-1 \rrbracket$. Then the word w_i leads from the state i to the state 0 in the automaton $\Pi(\mathcal{A}_{m,r,2^p})$. Otherwise stated, $\delta_{m,r,2^p}^\Pi(i, w_i) = 0$.*

Proof. The word w_i has length K and from Lemma 3.4.7, we know that $pK \geq z$. By Lemma 3.3.6, we have

$$\begin{aligned} \delta_{m,r,2^p}^\Pi(i, w_i) = 0 &\iff 2^{pK}i + \text{val}_{2^p}(w_i) \equiv 0 \pmod{m} \\ &\iff 2^{pK}i + \sigma(i)2^z \equiv 0 \pmod{k2^z} \\ &\iff 2^{pK-z}i + \sigma(i) \equiv 0 \pmod{k}. \end{aligned}$$

The result follows from the definition of σ . □

Lemma 3.4.9. *Suppose that $k > 1$ and let $i, j \in \llbracket 0, k-1 \rrbracket$. For any $\ell \in \mathbb{N}$, the word*

$$w_i(\text{rep}_{2^p}(m))^\ell \text{rep}_{2^p}(r)$$

is accepted from j in the automaton $\Pi(\mathcal{A}_{m,r,2^p})$ if and only if $i = j$.

Proof. Let $\ell \in \mathbb{N}$ and, for each $i \in \llbracket 0, k-1 \rrbracket$, let $y_i = w_i(\text{rep}_{2^p}(m))^\ell \text{rep}_{2^p}(r)$. Recall that we have set $R = |\text{rep}_{2^p}(r)|$. Further, set $M = |\text{rep}_{2^p}(m)|$. Then $|y_i| = K + \ell M + R$ and from Lemma 3.4.7, we know that $pK \geq z$. Therefore, we have

$$\begin{aligned}
& j2^{p|y_i|} + \text{val}_{2^p}(y_i) \equiv r \pmod{m} \\
\iff & j2^{p(K+\ell M+R)} + \text{val}_{2^p}(w_i)2^{p(\ell M+R)} + \sum_{s=0}^{\ell-1} m2^{p(sM+R)} + r \equiv r \pmod{m} \\
\iff & j2^{p(K+\ell M+R)} + \sigma(i)2^{z+p(\ell M+R)} \equiv 0 \pmod{k2^z} \\
\iff & j2^{p(K+\ell M+R)-z} + \sigma(i)2^{p(\ell M+R)} \equiv 0 \pmod{k} \\
\iff & j2^{p(K+\ell M+R)-z} - i2^{p(K+\ell M+R)-z} \equiv 0 \pmod{k} \\
\iff & j \equiv i \pmod{k} \\
\iff & j = i.
\end{aligned}$$

The conclusion follows from Lemma 3.3.6. \square

3.4.4 Properties of $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$

Lemma 3.4.10. *For all $i, j \in \llbracket 0, m-1 \rrbracket$, $X, Y \in \{T, B\}$ and all pair of words $(u, v) \in (\Sigma_{2^p} \times \Sigma_{2^p})^*$, we have*

$$\delta_{\mathcal{T},2^p}((i, X), (u, v)) = (j, Y) \iff \delta_{\mathcal{T},2^p}((i, \bar{X}), (u, v)) = (j, \bar{Y}).$$

Proof. This directly follows from Lemma 3.4.1. \square

Lemma 3.4.11. *Let $i \in \llbracket 0, m-1 \rrbracket$ and $X \in \{T, B\}$. The word $\text{rep}_{2^p}(1, m)$ is the label of a path from the state $(0, X)$ to the state $(0, \bar{X})$ in $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$.*

Proof. This directly follows from Lemma 3.3.7. \square

Proposition 3.4.12. *The automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ accepts the language $\text{val}_{2^p}^{-1}(\{(t, mt+r) : t \in \mathcal{T}\})$, is deterministic, accessible, coaccessible and has disjoint states.*

Proof. By construction of the product automaton and since

$$\{(n, mn+r) : n \in \mathbb{N}\} \cap (\mathcal{T} \times \mathbb{N}) = \{(t, mt+r) : t \in \mathcal{T}\},$$

we get that the product automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ accepts the language

$$\text{val}_{2^p}^{-1}(\{(t, mt+r) : t \in \mathcal{T}\}).$$

Since the automata $\mathcal{A}_{m,r,2^p}$ and $\mathcal{A}_{\mathcal{T},2^p}$ are deterministic, so is the product automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$. By Lemma 3.3.7, we can check that for every $i \in \llbracket 0, m-1 \rrbracket$, the states (i, T) and (i, B) are accessible thanks to the word $\text{rep}_{2^p}(0, i)$ and $\text{rep}_{2^p}(1, m+i)$ respectively. Hence, $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ is accessible.

To prove coaccessibility, we now fix some $i \in \llbracket 0, m-1 \rrbracket$ and $X \in \{T, B\}$. By Lemma 3.4.8, we know that there is a word w_i that leads from the state i to the state 0 in the automaton $\Pi(\mathcal{A}_{m,r,2^p})$. Thus, there is a word u of the same length as w_i such that the word (u, w_i) leads from i to 0 in $\mathcal{A}_{m,r,2^p}$. Now, by reading (u, w_i) from (i, X) in $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$, we reach either the state $(0, T)$ or the state $(0, B)$. If we reach $(0, T)$, then the concatenation $(u, w_i) \text{rep}_b(0, r)$ leads from the state (i, X) to (r, T) in $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$. If we reach $(0, B)$ instead, then we may apply Lemma 3.4.11 in order to obtain that the concatenation $(u, w_i) \text{rep}_b(1, m) \text{rep}_b(0, r)$ leads from (i, X) to (r, T) in $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$. This proves that $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ is coaccessible.

The fact that $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ has disjoint states follows from Propositions 3.4.2 and 3.4.4. \square

3.4.5 Properties of $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$

Lemma 3.4.13. *For all $i, j \in \llbracket 0, m-1 \rrbracket$, $X, Y \in \{T, B\}$ and $v \in \Sigma_{2^p}^*$, we have*

$$\delta_{\times}^{\Pi}((i, X), v) = (j, Y) \implies \delta_{m,r,2^p}^{\Pi}(i, v) = j.$$

Proof. This is a direct verification. \square

Lemma 3.4.14. *For every $i \in \llbracket 0, m-1 \rrbracket$, the states (i, T) and (i, B) are disjoint in the projected automaton $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$.*

Proof. Proceed by contradiction and suppose that a word v over Σ_{2^p} is accepted from both (i, T) and (i, B) in $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$ for some integer $i \in \llbracket 0, m-1 \rrbracket$. Then there are words u and u' over Σ_{2^p} of length $|v|$ such that, in $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$, the words (u, v) and (u', v) are accepted from (i, T) and (i, B) respectively. But from Lemma 3.3.5, we must have $u = u'$. Hence the word (u, v) is accepted from both (i, T) and (i, B) in $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$, contradicting that this automaton has disjoint states (see Proposition 3.4.12). \square

Proposition 3.4.15. *The automaton $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$ accepts the language $\text{val}_{2^p}^{-1}(m\mathcal{T}+r)$, is deterministic, complete, accessible and coaccessible.*

Proof. By construction, $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$ accepts $\text{val}_{2^p}^{-1}(m\mathcal{T}+r)$ (see Section 3.2). The fact that this automaton is deterministic and complete follows

from Lemma 3.3.3. Since $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ is accessible and coaccessible, so is $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$. \square

As we will see in the following section, this automaton is in general not minimal.

3.5 Minimization of $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$

We start by defining some classes of states of $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$. Our aim is twofold. (Recall Definition 1.2.8.) First, we will prove that those classes consist in *indistinguishable* states, i.e. accepting the same language. Secondly, we will show that states belonging to different classes are *distinguishable*, i.e. accept different languages. Otherwise stated, these classes correspond to the left quotients $w^{-1}L$ (cf. Definitions 1.2.16 and 1.2.19) where w is any finite word over the alphabet Σ_{2^p} and $L = \text{val}_{2^p}^{-1}(m\mathcal{T}+r)$.

3.5.1 Definition of the classes

Recall that $R = |\text{rep}_{2^p}(r)|$ and $N = \max\{\lceil \frac{z}{p} \rceil, R\}$ (see Notation 3.2.4). The classes we are going to define are closely related to the base 2^p -expansion of the remainder r with some additional leading zeros. More precisely, we have to consider the word $0^{N-R}\text{rep}_{2^p}(r)$, which is the unique word over the alphabet Σ_{2^p} with length N and 2^p -value r . This word is equal to the 2^p -expansion $\text{rep}_{2^p}(r)$ if and only if $N = R$, i.e. $\lceil \frac{z}{p} \rceil \leq R$. Recall Notation 3.2.2.

Definition 3.5.1. For $\alpha \in \llbracket 0, N \rrbracket$, we define

$$C'_\alpha = \begin{cases} \{(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell \frac{m}{2^{p\alpha}}, T_\ell) : 0 \leq \ell \leq 2^{p\alpha}-1\} & \text{if } \alpha \leq \frac{z}{p} \\ \{(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell k, T_\ell) : 0 \leq \ell \leq 2^z-1\} & \text{if } \alpha \geq \frac{z}{p}. \end{cases}$$

Note that when $\alpha = \frac{z}{p}$ (it can only happen when z is divisible by p), the two cases of the definition coincide.

Let us comment the previous definition, which may seem quite technical at first. The first elements of the sets C'_α are the integer part of the remainder r divided by increasing powers of the base 2^p , i.e. r divided by $2^{p\alpha}$ for the set indexed by α . The further elements of the set C'_α are obtained by adding to the first element $\lfloor \frac{r}{2^{p\alpha}} \rfloor$ integer multiples of $\frac{m}{2^{p\alpha}}$ so that the greatest element so-obtained is still less than m , provided that m is divisible by $2^{p\alpha}$. When m is no longer divisible by $2^{p\alpha}$, i.e. when $\alpha > \frac{z}{p}$, then we add integer multiples of k , which is the odd part of m . In particular, if m is odd, i.e. if $z = 0$, then all the sets C'_α are reduced to a single state: $C'_\alpha = \{(\lfloor \frac{r}{2^{p\alpha}} \rfloor, T)\}$. Finally,

remark that since $R = |\text{rep}_{2^p}(r)|$ and $N \geq R$, we have $\lfloor \frac{r}{2^{pN}} \rfloor = 0$ and $C'_N = \{(\ell k, T_\ell) : 0 \leq \ell \leq 2^z - 1\}$.

We will see in Lemma 3.5.16 that the states in C'_α are exactly those from which there is a path labelled by the suffix of length α of $0^{N-R} \text{rep}_{2^p}(r)$ to the state (r, T) .

Example 3.5.2. Let $m = 24$ and $p = 2$. We have $k = 3$ and $z = 3$. Let us consider the extremal possible values of the remainder r . For $r = 23$, we have $\text{rep}_4(23) = 113$, $R = 3$ and $N = \max\{\lceil \frac{3}{2} \rceil, 3\} = 3$. Thus, the sets defined above are

$$\begin{aligned} C'_0 &= \{(23, T)\} \\ C'_1 &= \{(5, T), (11, B), (17, B), (23, T)\} \\ C'_2 &= \{(1, T), (4, B), (7, B), (10, T), (13, B), (16, T), (19, T), (22, B)\} \\ C'_3 &= \{(0, T), (3, B), (6, B), (9, T), (12, B), (15, T), (18, T), (21, B)\}. \end{aligned}$$

For instance, we have the following two distinct paths, which are labelled by $0^{N-R} \text{rep}_4(23) = 113$:

$$(0, T) \xrightarrow{1} (1, T) \xrightarrow{1} (5, T) \xrightarrow{3} (23, T)$$

and

$$(3, B) \xrightarrow{1} (13, B) \xrightarrow{1} (5, T) \xrightarrow{3} (23, T).$$

For $r = 0$, we have $R = 0$ and $N = \max\{\lceil \frac{3}{2} \rceil, 0\} = 2$. In this case, the sets C'_α are

$$\begin{aligned} C'_0 &= \{(0, T)\} \\ C'_1 &= \{(0, T), (6, B), (12, B), (18, T)\} \\ C'_2 &= \{(0, T), (3, B), (6, B), (9, T), (12, B), (15, T), (18, T), (21, B)\}. \end{aligned}$$

For instance, we have the following paths, labelled by $0^{N-R} \text{rep}_4(0) = 00$:

$$(0, T) \xrightarrow{0} (0, T) \xrightarrow{0} (0, T)$$

and

$$(3, B) \xrightarrow{0} (12, B) \xrightarrow{0} (0, T).$$

The sets C'_α are not necessarily disjoint as Example 3.5.2 shows. In order to obtain the desired classes of states of the automaton $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$, we consider the following definition.

Definition 3.5.3. For $\alpha \in \llbracket 0, N \rrbracket$, we define

$$C_\alpha = C'_\alpha \setminus \bigcup_{\beta=0}^{\alpha-1} C'_\beta.$$

As we will see in Example 3.5.5, if we take a look at Example 3.5.2 with $r = 23$, then one has $C_\alpha = C'_\alpha$ for all α , except that $C_1 = C'_1 \setminus \{(23, T)\}$.

Let us define a second type of classes. The idea behind this definition is that these classes are "too far" from the remainder r with respect to the division by consecutive powers of the base 2^p , in the sense that these states do not accept any suffix of $0^{N-R} \text{rep}_{2^p}(r)$.

Definition 3.5.4. For $(j, X) \in (\llbracket 0, k-1 \rrbracket \times \{T, B\}) \setminus \{(0, T)\}$, we define

$$D'_{(j,X)} = \{(j + \ell k, X_\ell) : 0 \leq \ell \leq 2^z - 1\}$$

and

$$D_{(j,X)} = D'_{(j,X)} \setminus \bigcup_{\alpha=0}^N C_\alpha.$$

As we already observed, all states of the form $(\ell k, T_\ell)$ appear in the set C'_N and thus, also in the union of the sets C_α . This is the reason why the sets $D'_{(0,T)}$ and $D_{(0,T)}$ are not defined, i.e. $(j, X) \neq (0, T)$ in the previous definition.

We will refer to the sets of states C_α and $D_{(j,X)}$ as *classes* of states. Let us make some preliminary observations concerning the previous definitions.

The classes C_α and $D_{(j,X)}$ are pairwise disjoint: the intersection of any two such classes is empty. Moreover, the non-empty classes C_α and $D_{(j,X)}$ form a partition of the whole set of states $\llbracket 0, m-1 \rrbracket \times \{T, B\}$ of $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$. Note that if m is odd, i.e. if $z = 0$, then the sets $D'_{(j,X)}$ are reduced to a single state. If m is a power of 2, i.e. if $k = 1$, then there is no set of the form $D'_{(j,X)}$ and $D_{(j,X)}$ with $j \geq 1$.

Example 3.5.5. Let us resume Example 3.5.2. For $r = 23$, the classes

defined above are

$$\begin{aligned}
C_0 &= \{(23, T)\} \\
C_1 &= \{(5, T), (11, B), (17, B)\} \\
C_2 &= \{(1, T), (4, B), (7, B), (10, T), (13, B), (16, T), (19, T), (22, B)\} \\
C_3 &= \{(0, T), (3, B), (6, B), (9, T), (12, B), (15, T), (18, T), (21, B)\} \\
D_{(1,T)} &= \emptyset \\
D_{(2,T)} &= \{(2, T), (5, B), (8, B), (11, T), (14, B), (17, T), (20, T), (23, B)\} \\
D_{(0,B)} &= \{(0, B), (3, T), (6, T), (9, B), (12, T), (15, B), (18, B), (21, T)\} \\
D_{(1,B)} &= \{(1, B), (4, T), (7, T), (10, B), (13, T), (16, B), (19, B), (22, T)\} \\
D_{(2,B)} &= \{(2, B), (8, T), (14, T), (20, B)\}.
\end{aligned}$$

Note that $2k + \lceil \frac{z}{p} \rceil = 2 \cdot 3 + \lceil \frac{3}{2} \rceil = 8$ of them are non-empty. In Figure 3.9, the automaton $\Pi(\mathcal{A}_{24,23,4} \times \mathcal{A}_{\mathcal{T},4})$ is represented without the transitions, and the states are coloured with respect to these classes.

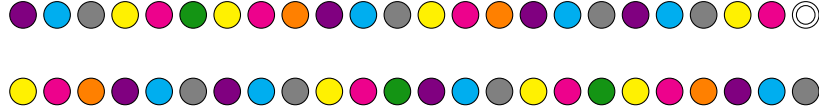


Figure 3.9: The classes of the projected automaton $\Pi(\mathcal{A}_{24,23,4} \times \mathcal{A}_{\mathcal{T},4})$.

Now if we consider $r = 0$, these classes are

$$\begin{aligned}
C_0 &= \{(0, T)\} \\
C_1 &= \{(6, B), (12, B), (18, T)\} \\
C_2 &= \{(3, B), (9, T), (15, T), (21, B)\} \\
D_{(1,T)} &= \{(1, T), (4, B), (7, B), (10, T), (13, B), (16, T), (19, T), (22, B)\} \\
D_{(2,T)} &= \{(2, T), (5, B), (8, B), (11, T), (14, B), (17, T), (20, T), (23, B)\} \\
D_{(0,B)} &= \{(0, B), (3, T), (6, T), (9, B), (12, T), (15, B), (18, B), (21, T)\} \\
D_{(1,B)} &= \{(1, B), (4, T), (7, T), (10, B), (13, T), (16, B), (19, B), (22, T)\} \\
D_{(2,B)} &= \{(2, B), (5, T), (8, T), (11, B), (14, T), (17, B), (20, B), (23, T)\}.
\end{aligned}$$

In this case, they are all non-empty and there are 8 of them. In Figure 3.10, the states of the automaton $\Pi(\mathcal{A}_{24,0,4} \times \mathcal{A}_{\mathcal{T},4})$ are coloured with respect to these classes.

Our aim is to prove that the non-empty classes defined above correspond exactly to the left quotients of the language $\text{val}_{2^p}^{-1}(m\mathcal{T} + r)$.

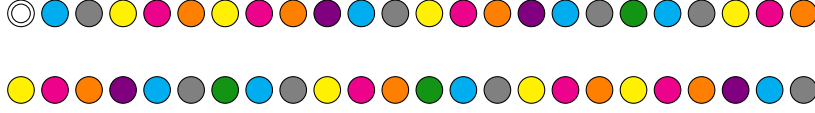


Figure 3.10: The classes of the projected automaton $\Pi(\mathcal{A}_{24,0,4} \times \mathcal{A}_{7,4})$.

3.5.2 Looking for the empty classes

Proposition 3.5.6. *For all $\alpha \in \llbracket 0, N \rrbracket$, the classes C_α are non-empty.*

Proof. The sequence $(\lfloor \frac{r}{2^{p\alpha}} \rfloor)_{\alpha \in \llbracket 0, N \rrbracket}$ is (strictly) decreasing for $\alpha \in \llbracket 0, R \rrbracket$ and is equal to 0 for $\alpha \in \llbracket R, N \rrbracket$. In particular, note that $\alpha = R$ is the first value for which $\lfloor \frac{r}{2^{p\alpha}} \rfloor = 0$. Therefore $(\lfloor \frac{r}{2^{p\alpha}} \rfloor, T_0) = (\lfloor \frac{r}{2^{p\alpha}} \rfloor, T) \in C_\alpha$ for every $\alpha \in \llbracket 0, R \rrbracket$ (see Figure 3.11).

C'_0	(r, T)	
C'_1	$(\lfloor \frac{r}{2^p} \rfloor, T)$	
\vdots	\vdots	
C'_R	$(0, T)$	
C'_{R+1}	$(0, T)$	$(\frac{m}{2^{p(R+1)}}, B)$
C'_{R+2}	$(0, T)$	$(\frac{m}{2^{p(R+2)}}, B)$
\vdots	\vdots	
C'_N	$(0, T)$	$(\frac{m}{2^{pN}}, B)$

Figure 3.11: The elements of the sets C'_α up to the first belonging to the classes C_α .

If $N = R$ then we are done (in this case, the part of Figure 3.11 below the line is empty).

Now suppose that $N = \lceil \frac{z}{p} \rceil > R$. Then $(\lfloor \frac{r}{2^{p\alpha}} \rfloor, T_0) = (0, T) \in C'_\alpha \setminus C_\alpha$ for every $\alpha \in \llbracket R+1, N \rrbracket$. We claim that, for each such α , the "next" element of C'_α (i.e. the element corresponding to $\ell = 1$) indeed belongs to C_α . Let us fix some $\alpha \in \llbracket R+1, N \rrbracket$.

First, suppose that $\alpha \leq \frac{z}{p}$. We must prove that $(\frac{m}{2^{p\alpha}}, T_1) = (\frac{m}{2^{p\alpha}}, B) \notin C'_\beta$ for $\beta < \alpha$. If $\beta < \alpha$ then $\frac{m}{2^{p\alpha}} < \frac{m}{2^{p\beta}} \leq \lfloor \frac{r}{2^{p\beta}} \rfloor + \frac{m}{2^{p\beta}}$. This shows that if the state $(\frac{m}{2^{p\alpha}}, B)$ belongs to some C'_β with $\beta < \alpha$, then its first component has to be $\lfloor \frac{r}{2^{p\beta}} \rfloor$. In other words, $(\lfloor \frac{r}{2^{p\beta}} \rfloor, T_0)$ is the only pair in C'_β such that the first component might be equal to the first component of $(\frac{m}{2^{p\alpha}}, T_1)$. However the second components of these pairs clearly differ: $T_0 = T$ and $T_1 = B$.

We have just shown that $(\frac{m}{2^{p\alpha}}, T_1) \notin C'_\beta$ for any $\beta < \alpha$, hence that $(\frac{m}{2^{p\alpha}}, T_1)$ indeed belongs to C_α in the case where $\alpha \leq \frac{z}{p}$.

The remaining case is when $\alpha > \frac{z}{p}$. In this case, we must show that one has $(k, B) \notin C'_\beta$ for $\beta < \alpha$. Since $\alpha \leq N = \lceil \frac{z}{p} \rceil$, we have $\lceil \frac{z}{p} \rceil \geq \alpha > \frac{z}{p}$. Thus $\alpha = \lceil \frac{z}{p} \rceil$. Therefore, if $\beta < \alpha$, then $\beta < \frac{z}{p}$. Thereby we may apply the same reasoning as in the previous paragraph. Since $\beta < \frac{z}{p}$, one has $k = \frac{m}{2^z} < \frac{m}{2^{p\beta}} \leq \lfloor \frac{r}{2^{p\beta}} \rfloor + \frac{m}{2^{p\beta}}$. Hence $(\lfloor \frac{r}{2^{p\beta}} \rfloor, T_0)$ is the only pair in C'_β such that the first component may be equal to the first component of (k, B) . Clearly, the second components of these pairs are different. We can thus conclude that (k, B) belongs to C_α in the case where $\alpha > \frac{z}{p}$. \square

Lemma 3.5.7. *We have $|\text{rep}_{2^p}(m-1)| - \lceil \frac{z}{p} \rceil \in \llbracket 0, k-1 \rrbracket$.*

Proof. Observe that

$$|\text{rep}_{2^p}(m)| = \lfloor \log_{2^p}(m) \rfloor + 1 = \left\lfloor \log_{2^p}(k) + \frac{z}{p} \right\rfloor + 1. \quad (3.4)$$

If m is a power of 2^p , otherwise stated if $k = 1$ and p divides z , then $|\text{rep}_{2^p}(m-1)| = |\text{rep}_{2^p}(m)| - 1 = \frac{z}{p}$ and the result is clear.

Now, suppose that m is not a power of the base 2^p . In this case, we have $|\text{rep}_{2^p}(m-1)| = |\text{rep}_{2^p}(m)|$. From (3.4) we get that $|\text{rep}_{2^p}(m)| \geq \lceil \frac{z}{p} \rceil$. Let us show that $|\text{rep}_{2^p}(m)| - \lceil \frac{z}{p} \rceil \leq k-1$. If $k = 1$ then p does not divide z (otherwise m would be a power of 2^p) and we get from (3.4) that $|\text{rep}_{2^p}(m)| = \lfloor \frac{z}{p} \rfloor + 1 = \lceil \frac{z}{p} \rceil$. In the case where $k = 3$ and $p = 1$, we obtain $|\text{rep}_{2^p}(m)| = \lfloor \log_2(3) + z \rfloor + 1 = 2 + z = k-1 + \lceil \frac{z}{p} \rceil$. In all other cases, that is if $k \geq 5$ or $(k = 3 \text{ and } p \geq 2)$, we can check that $\log_{2^p}(k) < k-2$. Therefore we have $|\text{rep}_{2^p}(m)| \leq \lfloor k-2 + \frac{z}{p} \rfloor + 1 = k-1 + \lfloor \frac{z}{p} \rfloor$. \square

Proposition 3.5.8. *The empty classes $D_{(j,X)}$ are exactly those of the form $D_{(\lfloor \frac{r}{2^{p\alpha}} \rfloor, T)}$ with $p\alpha \geq z$.*

Proof. The k classes $D_{(0,B)}, \dots, D_{(k-1,B)}$ are all non-empty. Indeed, for any $j \in \llbracket 0, k-1 \rrbracket$, the state (j, B) does not belong to any C_α .

Now, let $j \in \llbracket 1, k-1 \rrbracket$. We have to show that all classes of the form $D_{(\lfloor \frac{r}{2^{p\alpha}} \rfloor, T)}$ with $p\alpha \geq z$ are empty (observe that if $p\alpha \geq z$ then $\lfloor \frac{r}{2^{p\alpha}} \rfloor \leq k-1$), and that the other classes $D_{(j,T)}$ are non-empty.

If $(j, T) \notin \cup_{\alpha=0}^N C'_\alpha$, then the class $D_{(j,T)}$ is non-empty since it contains (j, T) . In this case, $j \neq \lfloor \frac{r}{2^{p\alpha}} \rfloor$ for any $\alpha \in \llbracket 0, N \rrbracket$. Now, suppose that there is some $\alpha \in \llbracket 0, N \rrbracket$ such that $(j, T) \in C'_\alpha$. Since $j < k$, this α is unique and $j = \lfloor \frac{r}{2^{p\alpha}} \rfloor$. We have to show that $D_{(j,T)}$ is empty if and only if $p\alpha \geq z$.

Clearly, $p\alpha \geq z$ implies that $D'_{(j,T)} = \{(j + \ell k, T_\ell) : \ell \in \llbracket 0, 2^z - 1 \rrbracket\} = C'_\alpha$, and hence that $D_{(j,T)}$ is empty.

Now suppose that $p\alpha < z$. We show that the second element $(j + k, B)$ of the set $D'_{(j,T)}$ does not belong to any set C'_β , and hence indeed belongs to the class $D_{(j,T)}$. Let $\beta \in \llbracket 0, N \rrbracket$ and suppose to the contrary that $(j + k, B) \in C'_\beta$. Since $j + k \in \llbracket 0, 2k - 1 \rrbracket$, the state $(j + k, B)$ must be either the first or the second element of the set C'_β . But since $B \neq T_0 = T$, it has to be the second. If $p\beta < z$, then we obtain $j + k = \lfloor \frac{r}{2^{p\beta}} \rfloor + k2^{z-p\beta} \geq 2k$, a contradiction. Thus $p\beta \geq z$ and $j + k = \lfloor \frac{r}{2^{p\beta}} \rfloor + k$. But this implies that $\lfloor \frac{r}{2^{p\alpha}} \rfloor = j = \lfloor \frac{r}{2^{p\beta}} \rfloor$. Since $\beta > \alpha$, this means that $j = 0$, a contradiction. \square

Corollary 3.5.9. *There are exactly $N - \lceil \frac{z}{p} \rceil$ empty classes among the $2k - 1$ classes $D_{(j,X)}$.*

Proof. By Proposition 3.5.8, the k classes $D_{(0,B)}, \dots, D_{(k-1,B)}$ are non-empty and we have to count the number of classes of the form $D_{(\lfloor \frac{r}{2^{p\alpha}} \rfloor, T)}$ with $p\alpha \geq z$ among the $k - 1$ classes $D_{(1,T)}, \dots, D_{(k-1,T)}$. Note that by Lemma 3.5.7 and by definition of N , we have $N - \lceil \frac{z}{p} \rceil \in \llbracket 0, k - 1 \rrbracket$.

Equivalently, we have to count the elements $\alpha \in \llbracket \lceil \frac{z}{p} \rceil, N \rrbracket$ such that $\lfloor \frac{r}{2^{p\alpha}} \rfloor \neq 0$. Similarly to the proof of Proposition 3.5.6, we consider two cases (also see Figure 3.11). If $N = \lceil \frac{z}{p} \rceil$ then there is no such α at all since $\lfloor \frac{r}{2^{pN}} \rfloor = 0$. If $N = R > \lceil \frac{z}{p} \rceil$, then the suitable α are exactly those in $\llbracket \lceil \frac{z}{p} \rceil, R - 1 \rrbracket$, and there are exactly $R - 1 - \lceil \frac{z}{p} \rceil + 1 = N - \lceil \frac{z}{p} \rceil$ of them. Hence the conclusion. \square

3.5.3 States of the same class are indistinguishable

Given two states (j, X) and (j', X') of the automaton $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{T,2^p})$, in order to prove that (j, X) and (j', X') are indistinguishable, we have to prove that $L_{(j,X)} = L_{(j',X')}$. The general procedure that we use goes as follows. Pick some word $v \in \Sigma_{2^p}^*$ and let $n = |v|$ and $e = \text{val}_{2^p}(v)$. By Lemma 3.3.9, the word v is accepted from the state (j, X) if and only if there exists some $d \in \mathbb{N}$ such that

$$2^{pn}j + e = md + r \quad \text{and} \quad X_d = T.$$

Similarly, the word v is accepted from the state (j', X') if and only if there exists some $d' \in \mathbb{N}$ such that

$$2^{pn}j + e = md' + r \quad \text{and} \quad X_{d'} = T.$$

But then, observe that there is only one possible pair of candidates for d and d' : we necessarily have

$$d = \frac{2^{pn}j + e - r}{m} \quad \text{and} \quad d' = \frac{2^{pn}j' + e - r}{m}. \quad (3.5)$$

Therefore, proving that

$$L_{(j,X)} = L_{(j',X')}$$

is equivalent to proving that for all $n \in \mathbb{N}$ and $e \in \llbracket 0, 2^{pn}-1 \rrbracket$, we have

$$(d \in \mathbb{N} \text{ and } X_d = T) \iff (d' \in \mathbb{N} \text{ and } (X')_{d'} = T)$$

where d and d' are given by (3.5). Moreover, note that such d and d' are always greater than or equal to $-\frac{r}{m}$, hence they are greater than -1 . Thus, provided that d and d' are integers, we know that they are necessary non-negative. Similarly, thanks to Remark 3.3.10, d and d' must be less than 2^{pn} . For these reasons, in the forthcoming proofs (namely, in Lemmas 3.5.10 and 3.5.12), we need to verify that $d, d' \in \mathbb{Z}$ but we don't need to check that $0 \leq d, d' < 2^{pn}$.

Our first aim is to show that all states in the same class $D_{(j,X)}$ accept the same language. We start with a lemma that will be used several times. Note that this lemma does not only concern the classes $D_{(j,X)}$ since we can have $(j, X) = (0, T)$ in the statement.

Lemma 3.5.10. *Let $j \in \llbracket 0, k-1 \rrbracket$, $\ell \in \llbracket 0, 2^z-1 \rrbracket$ and $X \in \{T, B\}$. For all $n \in \mathbb{N}$ such that $pn \geq z$, we have*

$$L_{(j,X)} \cap (\Sigma_{2^p})^n = L_{(j+\ell k, X_\ell)} \cap (\Sigma_{2^p})^n.$$

Proof. Let $n \in \mathbb{N}$ such that $pn \geq z$ and let $e \in \llbracket 0, 2^{pn}-1 \rrbracket$. Set

$$d = \frac{2^{pn}j + e - r}{m} \quad \text{and} \quad d' = \frac{2^{pn}(j + \ell k) + e - r}{m}.$$

Following the procedure described above, we have to prove that one has $(d \in \mathbb{N} \text{ and } X_d = T) \iff (d' \in \mathbb{N} \text{ and } (X_\ell)_{d'} = T)$. Moreover, since $d' = d + \frac{2^{pn}\ell k}{m} = d + \ell 2^{pn-z}$ and since $pn \geq z$, d is an integer if and only if so is d' . Furthermore,

$$d \leq \frac{2^{pn}j + e}{m} < \frac{2^{pn}(j+1)}{m} \leq \frac{2^{pn}k}{m} = 2^{pn-z}. \quad (3.6)$$

If $d, d' \in \mathbb{N}$ then $\text{rep}_2(d') = \text{rep}_2(\ell)0^{pn-z-|\text{rep}_2(d)|}\text{rep}_2(d)$, and $X_d = (X_\ell)_{d'}$. \square

Proposition 3.5.11. *Let $(j, X) \in (\llbracket 0, k-1 \rrbracket \times \{T, B\}) \setminus \{(0, T)\}$. Then any two states in $D_{(j, X)}$ accept the same language.*

Proof. Let $\ell, \ell' \in \llbracket 0, 2^z - 1 \rrbracket$. It suffices to show that if $(j + \ell k, X_\ell) \in D_{(j, X)}$ then $L_{(j + \ell k, X_\ell)} \subseteq L_{(j + \ell' k, X_{\ell'})}$. Thus, suppose that $(j + \ell k, X_\ell) \notin \cup_{\alpha=0}^N C_\alpha$. Let $n \in \mathbb{N}$ and $e \in \llbracket 0, 2^{pn} - 1 \rrbracket$. Set $d = \frac{2^{pn}(j + \ell k) + e - r}{m}$ and assume that $d \in \mathbb{N}$ and $(X_\ell)_d = T$. Then $X_\ell = T_d$. If $pn < z$ then $\frac{r-e}{2^{pn}} = \lfloor \frac{r}{2^{pn}} \rfloor$ because $\frac{r-e+dm}{2^{pn}} = j + k\ell$ is an integer, m is divisible by 2^{pn} and $e \in \llbracket 0, 2^{pn} - 1 \rrbracket$. Therefore, if $pn < z$ then we get that

$$(j + \ell k, X_\ell) = \left(\frac{r - e + dm}{2^{pn}}, T_d \right) = \left(\left\lfloor \frac{r}{2^{pn}} \right\rfloor + d \frac{m}{2^{pn}}, T_d \right) \in C'_n$$

which contradicts our assumption. Hence $pn \geq z$ and the conclusion follows from Lemma 3.5.10. \square

Note that the proof of Proposition 3.5.11 shows that no word shorter than $\lfloor \frac{z}{p} \rfloor$ is accepted from a state of a class $D_{(j, X)}$. However, such words may be accepted from a state of one of the classes C_α (see Lemma 3.5.16 below).

Now we turn to the classes C_α . The proof is divided into several technical lemmas.

Lemma 3.5.12. *For every $\alpha \in \llbracket 0, N \rrbracket$, any two states in C'_α accept the same words of length at least α .*

Proof. Let $\alpha \in \llbracket 0, N \rrbracket$. First, we do the case $\alpha \leq \frac{z}{p}$. By definition of the sets C'_α , it suffices to show that for all $\ell \in \llbracket 0, 2^{p\alpha} - 1 \rrbracket$ and $n \geq \alpha$, we have $L_{(\lfloor \frac{r}{2^{p\alpha}} \rfloor, T)} \cap (\Sigma_{2^p})^n = L_{(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell \frac{m}{2^{p\alpha}}, T_\ell)} \cap (\Sigma_{2^p})^n$. Thus, let $\ell \in \llbracket 0, 2^{p\alpha} - 1 \rrbracket$, $n \geq \alpha$ and $e \in \llbracket 0, 2^{pn} - 1 \rrbracket$. Then set

$$d = \frac{2^{pn} \lfloor \frac{r}{2^{p\alpha}} \rfloor + e - r}{m} \quad \text{and} \quad d' = \frac{2^{pn} (\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell \frac{m}{2^{p\alpha}}) + e - r}{m}.$$

We have to prove that $(d \in \mathbb{N} \text{ and } T_d = T) \iff (d' \in \mathbb{N} \text{ and } (T_\ell)_{d'} = T)$. Since $\lfloor \frac{r}{2^{p\alpha}} \rfloor < \frac{m}{2^{p\alpha}} = k2^{z-p\alpha}$ and $z - p\alpha \geq 0$, we obtain that $\lfloor \frac{r}{2^{p\alpha}} \rfloor + 1 \leq \frac{m}{2^{p\alpha}}$. Then

$$d < \frac{2^{pn} (\lfloor \frac{r}{2^{p\alpha}} \rfloor + 1)}{m} \leq 2^{p(n-\alpha)}. \quad (3.7)$$

Since $d' = d + \ell 2^{p(n-\alpha)}$ and $n \geq \alpha$, it follows that d is an integer if and only if so is d' . Moreover, in the case where both d and d' are in \mathbb{N} then we have $\text{rep}_2(d') = \text{rep}_2(\ell) 0^{p(n-\alpha) - |\text{rep}_2(d)|} \text{rep}_2(d)$, hence $T_d = (T_\ell)_{d'}$.

Next, suppose that $\alpha > \frac{z}{p}$. Then, we must show that for all $\ell \in \llbracket 0, 2^z - 1 \rrbracket$ and $n \geq \alpha$, we have $L_{(\lfloor \frac{r}{2^{p\alpha}} \rfloor, T)} \cap (\Sigma_{2^p})^n = L_{(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell k, T_\ell)} \cap (\Sigma_{2^p})^n$. Since $\lfloor \frac{r}{2^{p\alpha}} \rfloor < \frac{m}{2^{p\alpha}} = k2^{z-p\alpha} < k$, the conclusion follows from Lemma 3.5.10. \square

Lemma 3.5.13. *Let $\alpha \in \llbracket 0, N \rrbracket$.*

1. *If $p\alpha \leq z$ then no state in C_α accepts any words of length $< \alpha$.*
2. *If $p\alpha > z$ then no state in C_α accepts any words of length $\leq \lfloor \frac{z}{p} \rfloor$.*

Proof. Let us prove the first item. Suppose that $p\alpha \leq z$ and that there is a word over Σ_{2^p} of length $\beta < \alpha$ that is accepted from a state of the form $(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell \frac{m}{2^{p\alpha}}, T_\ell)$ with $\ell \in \llbracket 0, 2^{p\alpha}-1 \rrbracket$, i.e. from a state in C'_α . This means that there is $e \in \llbracket 0, 2^{p\beta}-1 \rrbracket$ such that if we set

$$d = \frac{2^{p\beta}(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell \frac{m}{2^{p\alpha}}) + e - r}{m},$$

then $d \in \mathbb{N}$ and $(T_\ell)_d = T$. But then

$$\left(\left\lfloor \frac{r}{2^{p\alpha}} \right\rfloor + \ell \frac{m}{2^{p\alpha}}, T_\ell \right) = \left(\frac{r - e + dm}{2^{p\beta}}, T_d \right) = \left(\left\lfloor \frac{r}{2^{p\beta}} \right\rfloor + d \frac{m}{2^{p\beta}}, T_d \right) \in C'_\beta.$$

Since $\beta < \alpha$, the state $(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell \frac{m}{2^{p\alpha}}, T_\ell)$ does not belong to C_α . (To show that $\frac{r-e}{2^{p\beta}} = \lfloor \frac{r}{2^{p\beta}} \rfloor$, we proceed as in the proof of Proposition 3.5.11.)

Now we prove the second part. Suppose that $p\alpha > z$ and that there is a word over Σ_{2^p} of length $\beta \leq \lfloor \frac{z}{p} \rfloor$ that is accepted from a state of the form $(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell k, T_\ell)$ with $\ell \in \llbracket 0, 2^z-1 \rrbracket$, i.e. from a state in C'_α . This means that there is $e \in \llbracket 0, 2^{p\beta}-1 \rrbracket$ such that if we set

$$d = \frac{2^{p\beta}(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell k) + e - r}{m},$$

then $d \in \mathbb{N}$ and $(T_\ell)_d = T$. But then

$$\left(\left\lfloor \frac{r}{2^{p\alpha}} \right\rfloor + \ell k, T_\ell \right) = \left(\frac{r - e + dm}{2^{p\beta}}, T_d \right) = \left(\left\lfloor \frac{r}{2^{p\beta}} \right\rfloor + d \frac{m}{2^{p\beta}}, T_d \right) \in C'_\beta.$$

Therefore the state $(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell k, T_\ell)$ does not belong to C_α . \square

Lemma 3.5.14. *If $N = \lceil \frac{z}{p} \rceil$, then no state in C_N accepts any words of length $< N$.*

Proof. This is a reformulation of Lemma 3.5.13 with $\alpha = N$. \square

We are now ready to prove that two states belonging to any given class C_α are indistinguishable.

Proposition 3.5.15. *For every $\alpha \in \llbracket 0, N \rrbracket$, any two states in C_α accept the same language.*

Proof. Let $\alpha \in \llbracket 0, N \rrbracket$. From Lemma 3.5.12, it is enough to consider words of length smaller than α and from the first item of Lemma 3.5.13, we may suppose that $p\alpha > z$. If $N = \lceil \frac{z}{p} \rceil$, then we must have $\alpha = N$ and we are done thanks to Lemma 3.5.14. Thus, we may also assume that $N = R > \lceil \frac{z}{p} \rceil$. Under these assumptions, $\lfloor \frac{r}{2^{p\alpha}} \rfloor < k$ and the first state of C_α is $(\lfloor \frac{r}{2^{p\alpha}} \rfloor, T)$ (see Figure 3.11). Thus, we have to show that for all $\ell \in \llbracket 0, 2^z - 1 \rrbracket$ such that the state $(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell k, T_\ell)$ indeed belongs to C_α and all $n < \alpha$, we have

$$L_{(\lfloor \frac{r}{2^{p\alpha}} \rfloor, T)} \cap (\Sigma_{2^p})^n = L_{(\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell k, T_\ell)} \cap (\Sigma_{2^p})^n.$$

If $pn < z$ then both languages are empty by the second item of Lemma 3.5.13. If $pn \geq z$ then the equality follows from Lemma 3.5.10. \square

3.5.4 States of different classes are distinguishable

In this section, we show that, in the projected automaton $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$, states belonging to different classes C_α or $D_{(j,X)}$ are pairwise distinguishable, that is, for any two such states, there is a word which is accepted from exactly one of them.

The following lemma shows that the states in a set C'_α are exactly those that lead to states of the set $C'_{\alpha-1}$ by reading the letter $r_{\alpha-1} \in \Sigma_{2^p}$, where $0^{N-R} \text{rep}_{2^p}(r) = r_{N-1} \cdots r_1 r_0$.

Lemma 3.5.16. *Let $0^{N-R} \text{rep}_{2^p}(r) = r_{N-1} \cdots r_1 r_0$ and let $\alpha \in \llbracket 0, N \rrbracket$. Then*

$$C'_\alpha = \{(i, X) \in \llbracket 0, m-1 \rrbracket \times \{T, B\} : \delta_\times^\Pi((i, X), r_{\alpha-1} \cdots r_1 r_0) = (r, T)\}.$$

Proof. First, we consider the case where $p\alpha \leq z$. Pick some $(i, X) \in C'_\alpha$. By definition, there exists $\ell \in \llbracket 0, 2^{p\alpha} - 1 \rrbracket$ such that $(i, X) = (\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell \frac{m}{2^{p\alpha}}, T_\ell)$. Observe that $\lfloor \frac{r}{2^{p\alpha}} \rfloor = \text{val}_{2^p}(r_{N-1} \cdots r_{\alpha+1} r_\alpha)$. Then

$$2^{p\alpha} \left(\left\lfloor \frac{r}{2^{p\alpha}} \right\rfloor + \ell \frac{m}{2^{p\alpha}} \right) + \text{val}_{2^p}(r_{\alpha-1} \cdots r_1 r_0) = \ell m + r.$$

Since $(T_\ell)_\ell = T$, from Lemma 3.3.9 we get $\delta_\times^\Pi((i, X), r_{\alpha-1} \cdots r_1 r_0) = (r, T)$. The other way around, let $(i, X) \in \llbracket 0, m-1 \rrbracket \times \{T, B\}$ be a state such that $\delta_\times^\Pi((i, X), r_{\alpha-1} \cdots r_1 r_0) = (r, T)$. Then there exists some $d \in \llbracket 0, 2^{p\alpha} - 1 \rrbracket$ such that

$$2^{p\alpha} i + \text{val}_{2^p}(r_{\alpha-1} \cdots r_1 r_0) = md + r \quad \text{and} \quad X_d = T.$$

We obtain

$$i = \frac{1}{2^{p\alpha}} \left(md + r - \text{val}_{2^p}(r_{\alpha-1} \cdots r_1 r_0) \right) = d \frac{m}{2^{p\alpha}} + \left\lfloor \frac{r}{2^{p\alpha}} \right\rfloor.$$

Observe that $X_d = T$ is equivalent to $X = T_d$. This proves that $(i, X) \in C'_\alpha$.

Secondly, we consider the case where $p\alpha > z$. Pick some $(i, X) \in C'_\alpha$. There exists $\ell \in \llbracket 0, 2^z - 1 \rrbracket$ such that $(i, X) = (\lfloor \frac{r}{2^{p\alpha}} \rfloor + \ell k, T_\ell)$. Then

$$2^{p\alpha} \left(\left\lfloor \frac{r}{2^{p\alpha}} \right\rfloor + \ell k \right) + \text{val}_{2^p}(r_{\alpha-1} \cdots r_1 r_0) = \ell k 2^{p\alpha} + r = \ell 2^{p\alpha-z} m + r.$$

Since $(T_\ell)_{\ell 2^{p\alpha-z}} = (T_\ell)_\ell = T$, we obtain $\delta_{\times}^\Pi((i, X), r_{\alpha-1} \cdots r_1 r_0) = (r, T)$. Conversely, let (i, X) be some state of the set $\llbracket 0, m-1 \rrbracket \times \{T, B\}$ such that $\delta_{\times}^\Pi((i, X), r_{\alpha-1} \cdots r_1 r_0) = (r, T)$. Then there exists some $d \in \llbracket 0, 2^{p\alpha}-1 \rrbracket$ such that

$$2^{p\alpha} i + \text{val}_{2^p}(r_{\alpha-1} \cdots r_1 r_0) = md + r \quad \text{and} \quad X_d = T.$$

From the first part, we get

$$md = 2^{p\alpha} i - (r - \text{val}_{2^p}(r_{\alpha-1} \cdots r_1 r_0)) = 2^{p\alpha} i - 2^{p\alpha} \left\lfloor \frac{r}{2^{p\alpha}} \right\rfloor,$$

hence $kd = 2^{p\alpha-z} (i - \lfloor \frac{r}{2^{p\alpha}} \rfloor)$. Since k is odd, d must be a multiple of $2^{p\alpha-z}$. We obtain

$$i = \frac{d}{2^{p\alpha-z}} k + \left\lfloor \frac{r}{2^{p\alpha}} \right\rfloor$$

and $X_{\frac{d}{2^{p\alpha-z}}} = X_d = T$. Since $\frac{d}{2^{p\alpha-z}} \in \llbracket 0, 2^z - 1 \rrbracket$, we get that $(i, X) \in C'_\alpha$. \square

Proposition 3.5.17. *For every $\alpha \in \llbracket 0, N \rrbracket$, the class C_α is distinguishable from all the other classes.*

Proof. First, we show that the classes C_α are distinguishable among them. From Proposition 3.5.6, we know that these classes are all non-empty. Let $\alpha, \beta \in \llbracket 0, N \rrbracket$ such that $\alpha < \beta$ and let $(i, X) \in C_\alpha$ and $(j, Y) \in C_\beta$. We show that $L_{(i,X)} \neq L_{(j,Y)}$. By definition of the classes, the state (i, X) belongs to C'_α and since $\alpha < \beta$, the state (j, Y) does not belong to C'_α . We get from Lemma 3.5.16 that the suffix s of length α of the word $0^{N-R} \text{rep}_{2^p}(r)$ is accepted from (i, X) but not from (j, Y) . Hence $s \in L_{(i,X)} \setminus L_{(j,Y)}$.

Secondly, we show that the classes C_α are distinguishable from all the non-empty classes of the form $D_{(i,X)}$. Let $\alpha \in \llbracket 0, N \rrbracket$. By definition, any state in a class $D_{(i,X)}$ cannot belong to C'_α . Similarly to what precedes, the conclusion follows from Lemma 3.5.16. \square

It remains to show that the non-empty classes $D_{(j,X)}$ are distinguishable from each other. Recall that when m is a power of 2, i.e. when $k = 1$, there is no class of the form $D_{(j,X)}$.

Proposition 3.5.18. *Suppose that $k > 1$ and let $(i, X), (j, Y)$ be two distinct states of $(\llbracket 0, k-1 \rrbracket \times \{T, B\}) \setminus \{(0, T)\}$ such that the classes $D_{(i,X)}$ and $D_{(j,Y)}$ are both non-empty. Then $D_{(i,X)}$ and $D_{(j,Y)}$ are distinguishable.*

Proof. We already know from the previous section that the states of $D_{(i,X)}$ (resp. $D_{(j,Y)}$) are indistinguishable. Therefore, it suffices to show that $L_{(i,X)} \neq L_{(j,Y)}$.

First, suppose that $i = j$. Then $X \neq Y$ by hypothesis and the states (i, X) and (j, Y) are disjoint by Lemma 3.4.14. Since $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$ is coaccessible by Proposition 3.4.15, we obtain that the states (i, X) and (j, Y) are distinguishable.

Now suppose that $i \neq j$. By Lemma 3.4.9, the word $w_i \text{rep}_{2^p}(r)$ is accepted from i in the automaton $\Pi(\mathcal{A}_{m,r,2^p})$ but is not accepted from j . Then, there are a word u_1 of length $|w_i|$ and a word u_2 of length R such that the word $(u_1, w_i)(u_2, \text{rep}_{2^p}(r))$ is accepted from i in the automaton $\mathcal{A}_{m,r,2^p}$ but is not accepted from j . By Lemma 3.4.10, this word is accepted either from (i, T) or from (i, B) in the automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$, but is not accepted neither from (j, T) nor from (j, B) . Now, two cases are possible.

First, suppose that $(u_1, w_i)(u_2, \text{rep}_{2^p}(r))$ is accepted from (i, X) in the automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$. Then, in the projection $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$, the word $w_i \text{rep}_{2^p}(r)$ is accepted from (i, X) but not from (j, Y) . Thus, the word $w_i \text{rep}_{2^p}(r)$ distinguishes the states (i, X) and (j, Y) .

Secondly, suppose that $(u_1, w_i)(u_2, \text{rep}_{2^p}(r))$ is accepted from (i, \overline{X}) in $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$. Let $(i', X') = \delta_{\times}((i, \overline{X}), (u_1, w_i))$. In this case we have $\delta_{\times}((i', X'), (u_2, \text{rep}_{2^p}(r))) = (r, T)$. In particular $\delta_{m,r,2^p}^{\Pi}(i, w_i) = i'$, hence $i' = 0$ by Lemma 3.4.8. Now, by using Lemma 3.4.10 and Lemma 3.4.11 successively, we obtain

$$\begin{aligned} & \delta_{\times}((i, X), (u_1, w_i) \text{rep}_{2^p}(1, m)(u_2, \text{rep}_{2^p}(r))) \\ &= \delta_{\times}((0, \overline{X'}), \text{rep}_{2^p}(1, m)(u_2, \text{rep}_{2^p}(r))) \\ &= \delta_{\times}((0, X'), (u_2, \text{rep}_{2^p}(r))) \\ &= (r, T). \end{aligned}$$

This shows that the word $w_i \text{rep}_{2^p}(m) \text{rep}_{2^p}(r)$ is accepted from (i, X) in $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$. From Lemmas 3.4.9 and 3.4.13, this word cannot be accepted from (j, Y) , hence it distinguishes the states (i, X) and (j, Y) . \square

3.5.5 The minimal automaton of $\text{val}_{2^p}^{-1}(m\mathcal{T} + r)$

We are ready to construct the minimal automaton of $\text{val}_{2^p}^{-1}(m\mathcal{T} + r)$. Since the states of $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$ that belong to the same class C_α or $D_{(j,X)}$ are indistinguishable, they can be glued together in order to define a new automaton $\mathcal{M}_{m,r,\mathcal{T},2^p}$ that still accepts the same language.

The formal definition of $\mathcal{M}_{m,r,\mathcal{T},2^p}$ is as follows. The alphabet is Σ_{2^p} . The states are the classes C_α for $\alpha \in \llbracket 0, N \rrbracket$ and the non-empty classes $D_{(j,X)}$ for $(j, X) \in (\llbracket 0, k-1 \rrbracket \times \{T, B\}) \setminus \{(0, T)\}$. The class C_R is the initial state and the only final state is the class C_0 . Note that $(0, T) \in C_R$ and that $(r, T) \in C_0$. The transitions of $\mathcal{M}_{m,r,\mathcal{T},2^p}$ are defined as follows: there is a transition labelled by a letter a in Σ_{2^p} from a class J_1 to a class J_2 if and only if in the automaton $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$, there is a transition labelled by a from a state of J_1 to a state of J_2 .

Example 3.5.19. In Figure 3.12, the classes of $\Pi(\mathcal{A}_{6,2,4} \times \mathcal{A}_{\mathcal{T},4})$ are coloured in white, blue, grey, yellow, fuchsia, orange and purple.

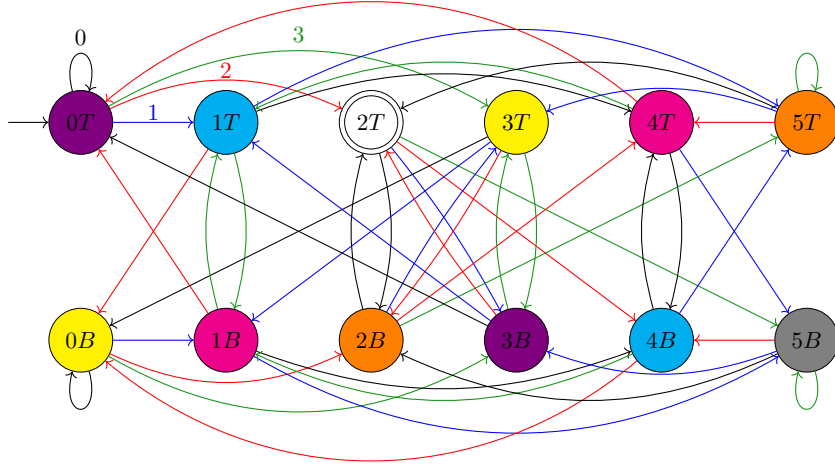
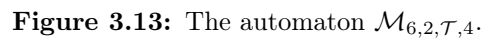


Figure 3.12: The classes of the automaton of $\Pi(\mathcal{A}_{6,2,4} \times \mathcal{A}_{\mathcal{T},4})$.

Figure 3.13 depicts the minimal automaton $\mathcal{M}_{6,2,\mathcal{T},4}$ of $\text{val}_4^{-1}(6\mathcal{T} + 2)$, where states corresponding to the same color are glued together to form a single state.

Theorem 3.5.20. *Let p and m be positive integers. Then the automaton $\mathcal{M}_{m,r,\mathcal{T},2^p}$ is the minimal automaton of the language $\text{val}_{2^p}^{-1}(m\mathcal{T} + r)$.*



We are now ready to prove Theorem 3.2.3.

Example 3.5.21. The minimal automaton of the language $\text{val}_4^{-1}(6\mathcal{T}+2)$ has 7 states (see Figure 3.13). We can indeed compute that $2 \cdot 3 + \lceil \frac{1}{2} \rceil = 7$.

3.6 A decision procedure

As an application of Theorem 3.2.3, we obtain the following decision procedure.

Corollary 3.6.1. *Given any 2^p -recognizable set Y (via a finite automaton \mathcal{A} recognizing it), it is decidable whether $Y = m\mathcal{T} + r$ for some $m \in \mathbb{N}$ and $r \in \llbracket 0, m-1 \rrbracket$.*

Proof. Let Y be a 2^p -recognizable set given thanks to a complete DFA that accepts the language of the 2^p -expansions of its elements. We can minimize the DFA and hence compute the state complexity M of Y (with respect to the base 2^p). Let us decompose the possible multiples m as $k2^z$ with k odd. By Theorem 3.2.3, it is sufficient to test the equality between Y and $m\mathcal{T} + r$ for the finitely many values of pairs (k, z) such that $2k + \lceil \frac{z}{p} \rceil = M$ and the finitely many $r \in \llbracket 0, m-1 \rrbracket$. For each couple (m, r) that has to be tested, we can directly use our description of the minimal automaton of $\text{val}_{2^p}^{-1}(m\mathcal{T} + r)$ (this is Theorem 3.5.20). This concludes the proof since the equality of two regular languages is decidable. \square

Remark 3.6.2. We think, even though we do not have any proof yet, that the previous decision procedure could be run in quadratic time with respect to the number of states of the automaton given in entry.

3.7 A direct description of the classes when $r = 0$

In the conference paper [22], we described the automaton $\mathcal{M}_{m,r,\mathcal{T},2^p}$ in the particular case where $r = 0$, i.e. for the exact multiples of \mathcal{T} . The construction was similar, but the way we build the classes of states was different. Therefore, we can give another description of the classes C_α and $D_{(j,X)}$ for $r = 0$ which is easier than the descriptions from Definitions 3.5.3 and 3.5.4 in the sense that the classes are built in a direct way, without having to remove some states a posteriori. Since the proofs in [22] are similar to the ones in the present dissertation, we give here another proof of the equivalence of the descriptions.

Note that if $r = 0$ then $R = 0$ and $N = \lceil \frac{z}{p} \rceil$.

Proposition 3.7.1. *Suppose that $r = 0$.*

- We have $C_0 = \{(0, T)\}$.
- For each $\alpha \in \llbracket 1, N-1 \rrbracket$, we have

$$C_\alpha = \bigcup_{\beta=(\alpha-1)p}^{\alpha p-1} \{(k2^{z-\beta-1} + \ell k2^{z-\beta}, B_\ell) : \ell \in \llbracket 0, 2^\beta-1 \rrbracket\}.$$

- We have

$$C_N = \bigcup_{\beta=\left(\left\lceil \frac{z}{p} \right\rceil - 1\right)p}^{z-1} \{(k2^{z-\beta-1} + \ell k2^{z-\beta}, B_\ell) : \ell \in \llbracket 0, 2^\beta - 1 \rrbracket\}.$$

- For $(j, X) \in (\llbracket 0, k-1 \rrbracket \times \{T, B\}) \setminus \{(0, T)\}$, we have

$$D_{(j,X)} = \{(j + \ell k, X_\ell) : \ell \in \llbracket 0, 2^z - 1 \rrbracket\}.$$

Proof. A direct verification shows that $C_0 = C'_0 = \{(0, T)\}$. Now, let us show that for all $\alpha \in \llbracket 1, N \rrbracket$, we have $C'_{\alpha-1} \subset C'_\alpha$. If $\alpha \leq z/p$, then for all $\ell \in \llbracket 0, 2^{p(\alpha-1)} - 1 \rrbracket$, we have

$$(\ell k 2^{z-p(\alpha-1)}, T_\ell) = (\ell 2^p k 2^{z-p\alpha}, T_\ell)$$

and $\ell 2^p \geq 0$, $\ell 2^p < 2^{p\alpha-p} 2^p = 2^{p\alpha}$ and $T_{\ell 2^p} = T_\ell$. If $\alpha \geq z/p$, then $\alpha = \left\lceil \frac{z}{p} \right\rceil$, $\alpha - 1 \leq z/p$ and for all $\ell \in \llbracket 0, 2^{p(\alpha-1)} - 1 \rrbracket$

$$(\ell k 2^{z-p(\alpha-1)}, T_\ell) = (\ell 2^{z-p(\alpha-1)} k, T_\ell)$$

and $\ell 2^{z-p(\alpha-1)} \geq 0$, $\ell 2^{z-p(\alpha-1)} < 2^{p\alpha-p} 2^{z-p\alpha+p} = 2^z$ and $T_{\ell 2^{z-p(\alpha-1)}} = T_\ell$. Hence $C'_{\alpha-1} \subset C'_\alpha$. Note that thanks to Proposition 3.5.6, the inclusion is strict. We can deduce that for all $\alpha \in \llbracket 1, N \rrbracket$, we have

$$C_\alpha = C'_\alpha \setminus \bigcup_{\beta=0}^{\alpha-1} C'_\beta = C'_\alpha \setminus C'_{\alpha-1}.$$

Moreover, $C'_N = \{(\ell k, T_\ell) : \ell \in \llbracket 0, 2^z - 1 \rrbracket\}$. This suffices to show that

$$D_{(j,X)} = \{(j + \ell k, X_\ell) : \ell \in \llbracket 0, 2^z - 1 \rrbracket\}$$

for any $(j, X) \in (\llbracket 0, k-1 \rrbracket \times \{T, B\}) \setminus \{(0, T)\}$.

It remains to show the equality for the classes C_α , $\alpha \in \llbracket 1, N \rrbracket$. First, suppose that $\alpha \in \llbracket 1, N-1 \rrbracket$. We have

$$C'_{\alpha-1} = \{(\ell' k 2^{z-p\alpha+p}, T_{\ell'}) : \ell' \in \llbracket 0, 2^{p\alpha-p} - 1 \rrbracket\}.$$

Moreover,

$$\begin{aligned} C'_\alpha &= \{(\ell k 2^{z-p\alpha}, T_\ell) : \ell \in \llbracket 0, 2^{p\alpha} - 1 \rrbracket\} \\ &= \left\{ \left(\frac{\ell}{2^p} k 2^{z-p\alpha+p}, T_\ell \right) : \ell \in \llbracket 0, 2^{p\alpha} - 1 \rrbracket \right\}. \end{aligned}$$

Thus a simple verification shows that the states of the class C_α are the states $(\ell k 2^{z-p\alpha}, T_\ell)$ of C'_α where ℓ is not divisible by 2^p . We also have

$$\begin{aligned}
& \bigcup_{\beta=(\alpha-1)p}^{\alpha p-1} \left\{ (k 2^{z-\beta-1} + \ell k 2^{z-\beta}, B_\ell) : \ell \in \llbracket 0, 2^\beta - 1 \rrbracket \right\} \\
&= \bigcup_{\beta=(\alpha-1)p}^{\alpha p-1} \left\{ (k(2\ell + 1) 2^{z-\beta-1}, B_\ell) : \ell \in \llbracket 0, 2^\beta - 1 \rrbracket \right\} \\
&= \bigcup_{\beta=(\alpha-1)p}^{\alpha p-1} \left\{ (k\ell' 2^{z-\beta-1}, T_{\ell'}) : \ell' \in \llbracket 1, 2^{\beta+1} - 1 \rrbracket, \ell' \text{ odd} \right\} \\
&= \bigcup_{\gamma=(\alpha-1)p+1}^{\alpha p} \left\{ (k\ell' 2^{z-\gamma}, T_{\ell'}) : \ell' \in \llbracket 1, 2^\gamma - 1 \rrbracket, \ell' \text{ odd} \right\}.
\end{aligned}$$

Now, let $\ell \in \llbracket 0, 2^{\alpha p} - 1 \rrbracket$ be such that ℓ is not divisible by 2^p (and thus the state $(\ell k 2^{z-p\alpha}, T_\ell)$ belongs to C_α). We are looking for $\gamma \in \llbracket (\alpha-1)p + 1, \alpha p \rrbracket$ and $\ell' \in \llbracket 1, 2^\gamma - 1 \rrbracket$, ℓ' odd, such that

$$(\ell k 2^{z-p\alpha}, T_\ell) = (\ell' k 2^{z-\gamma}, T_{\ell'}).$$

Since ℓ is not divisible by 2^p , there is an odd number $x \geq 1$ and a number $y \in \llbracket 0, p-1 \rrbracket$ such that $\ell = x 2^y$. It suffices to choose

$$\gamma = p\alpha - y \quad \text{and} \quad \ell' = x.$$

In this case, we have

$$\ell' k 2^{z-\gamma} = x k 2^{z-p\alpha+y} = \ell k 2^{z-p\alpha}.$$

Moreover, since $\ell = \ell' 2^y$, we have $T_\ell = T_{\ell'}$. It remains to show that γ and ℓ' are in the good intervals and ℓ' is odd. On the one hand, we have $\gamma = p\alpha - y > p\alpha - p = p(\alpha - 1)$ and $\gamma = p\alpha - p \leq p\alpha$. On the other hand, ℓ' is odd because x is, $\ell' = x \geq 1$ and

$$\ell' = x < 2^\gamma \iff \frac{\ell}{2^y} < 2^{p\alpha-y} \iff \ell < 2^{p\alpha}.$$

Let us now turn to the other inclusion. Let $\gamma \in \llbracket (\alpha-1)p + 1, \alpha p \rrbracket$ and $\ell' \in \llbracket 1, 2^\gamma - 1 \rrbracket$ with ℓ' odd. Since $\gamma \leq \alpha p$, we have

$$(k\ell' 2^{z-\gamma}, T_{\ell'}) = (\ell' 2^{p\alpha-\gamma} k 2^{z-p\alpha}, T_{\ell' 2^{p\alpha-\gamma}}).$$

Furthermore, $\ell' 2^{p\alpha-\gamma} \geq 2^{p\alpha-\gamma} \geq 0$. We also have $\ell' 2^{p\alpha-\gamma} < 2^\gamma 2^{p\alpha-\gamma} = 2^{p\alpha}$. Finally, since ℓ' is odd, $\ell' 2^{p\alpha-\gamma}$ is not divisible by 2^p if and only if we have $p\alpha - \gamma < p$, which is equivalent to $\gamma > p\alpha - p$, hence the conclusion.

The proof of the case $\alpha = N$ is similar. \square

3.8 Replacing \mathcal{T} by its complement $\overline{\mathcal{T}}$

If we are interested in the set $\overline{\mathcal{T}} = \mathbb{N} \setminus \mathcal{T}$ instead of \mathcal{T} , we can use the same construction that we described and studied for \mathcal{T} . We only have to exchange the final/non-final status of the states in the automaton $\mathcal{A}_{\mathcal{T}}$. In this section, we show that we may instead directly obtain the minimal automaton of the language $\text{val}_{2^p}^{-1}(m\overline{\mathcal{T}} + r)$ from that of $\text{val}_{2^p}^{-1}(m\mathcal{T} + r)$.

Example 3.8.1. Let us push further our running example by considering now $\overline{\mathcal{T}}$ instead of \mathcal{T} . The classes of states are defined similarly by exchanging T and B everywhere. In Figure 3.14, we have depicted the classes of the corresponding projected product automaton, which we denote by $\Pi(\mathcal{A}_{6,2,4} \times \mathcal{A}_{\overline{\mathcal{T}},4})$. Figure 3.15 depicts the minimal automaton $\mathcal{M}_{6,2,\overline{\mathcal{T}},4}$ of

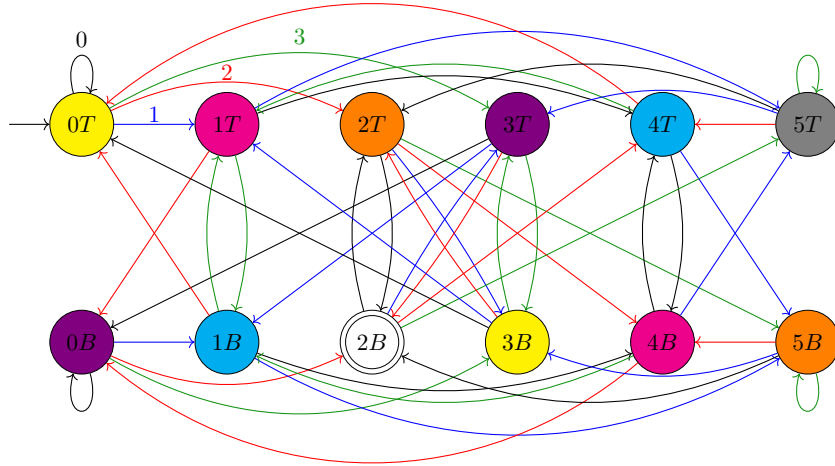


Figure 3.14: The classes of the automaton of $\Pi(\mathcal{A}_{6,2,4} \times \mathcal{A}_{\overline{\mathcal{T}},4})$.

$\text{val}_4^{-1}(6\overline{\mathcal{T}} + 2)$, where states corresponding to the same color are glued together to form a single state. Since the classes of states have been modified but the edges are unchanged, the minimal automaton obtained by gluing the sets of the same classes together is not a symmetric version of the automaton $\mathcal{M}_{m,r,\mathcal{T},2^p}$ we obtained starting from the set \mathcal{T} ; compare Figures 3.13 and 3.15. Nevertheless, observe that the automaton of Figure 3.15 can be obtained from the one of Figure 3.13 by replacing the initial state (in purple) by the yellow state. Also observe that, in the automaton of Figure 3.13, the yellow state is reached from the initial state by reading the word $\text{rep}_4(6) = 12$.

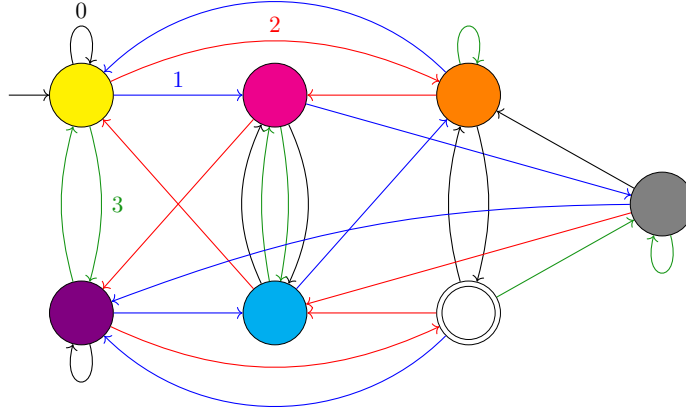


Figure 3.15: The automaton $\mathcal{M}_{6,2,\overline{\mathcal{T}},4}$.

This fact is always true and is proved in Proposition 3.8.2.

In the next proposition, we show that the minimal automaton of the language $\text{val}^{-1}(m\overline{\mathcal{T}}+r)$ can be obtained directly from the minimal automaton of $\text{val}_{2^p}^{-1}(m\mathcal{T}+r)$ by only moving the initial state.

Proposition 3.8.2. *The minimal automaton of $\text{val}_{2^p}^{-1}(m\overline{\mathcal{T}}+r)$ is obtained by replacing the initial state of the automaton $\mathcal{M}_{m,r,\mathcal{T},2^p}$ by the state that is reached by reading $\text{rep}_{2^p}(m)$ from the initial state.*

Proof. Consider the automaton $\mathcal{M}_{m,r,\mathcal{T},2^p}$. By construction, its states are sets of states (called classes) of the automaton $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$. Using Lemma 3.3.9, for each $X \in \{T, B\}$, there is a path labelled by $\text{rep}_{2^p}(m)$ going from $(0, X)$ to $(0, \overline{X})$ in $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$, and hence the same holds for the corresponding classes of states in $\mathcal{M}_{m,r,\mathcal{T},2^p}$.

First, let us show that the obtained automaton is again minimal. By only changing the initial state of any minimal DFA, we keep a DFA that is complete and reduced. Furthermore, the obtained DFA is still accessible since we have seen in the previous paragraph that there is a path from the class of $(0, B)$ to the class of $(0, T)$, which is precisely the initial state in $\mathcal{M}_{m,r,\mathcal{T},2^p}$.

It remains to show that the language L accepted from the class of $(0, B)$ in the automaton $\mathcal{M}_{m,r,\mathcal{T},2^p}$ is equal to $\text{val}_{2^p}^{-1}(m\overline{\mathcal{T}}+r)$. By construction, L is equal to the language $L_{(0,B)}$ accepted from the state $(0, B)$ in the automaton $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$ and we already know that $L_{(0,T)} = \text{val}_{2^p}^{-1}(m\mathcal{T}+r)$.

Let $w \in \Sigma_{2^p}^*$. We know that $w \in L_{(0,B)} \iff \text{rep}_{2^p}(m)w \in L_{(0,T)}$. Thus, we only have to show $\text{val}_{2^p}(w) \in m\overline{\mathcal{T}} + r \iff m2^{p|w|} + \text{val}_{2^p}(w) \in m\mathcal{T} + r$. In both cases, we must have that $\text{val}_{2^p}(w) = mq + r$ with $q \in \mathbb{N}$. Since $q \leq \text{val}_{2^p}(w) < 2^{p|w|}$, we have $\text{rep}_2(2^{p|w|} + q) = 10^{p|w| - |\text{rep}_2(q)|} \text{rep}_2(q)$. This shows that $q \in \overline{\mathcal{T}} \iff 2^{p|w|} + q \in \mathcal{T}$, hence the conclusion. \square

Note that the minimal automaton of $\text{val}_{2^p}^{-1}(m\overline{\mathcal{T}} + r)$ can also be obtained by replacing the initial state T of the automaton $\mathcal{A}_{\mathcal{T}, 2^p}$ (cf. Section 3.3.1) by the state B and then apply the same strategy than we did for $\text{val}_{2^p}^{-1}(m\mathcal{T} + r)$.

Corollary 3.8.3. *Let m, p be positive integers and $r \in \llbracket 0, m-1 \rrbracket$. Then the state complexity of $m\overline{\mathcal{T}} + r$ with respect to the base 2^p is equal to $2k + \left\lceil \frac{z}{p} \right\rceil$ if $m = k2^z$ with k odd.*

3.9 Conclusion and perspectives

Our method is constructive and in principle, it may be applied to any b -recognizable set $X \subseteq \mathbb{N}$. However, the product automaton $\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{X,2^p}$ recognizing the bidimensional set $\{(n, mn + r) : n \in X\}$ is not minimal in general. As an example, consider the 2-recognizable set X of powers of 2: $X = \{2^n : n \in \mathbb{N}\}$. The product automaton $\mathcal{A}_{3,0,2} \times \mathcal{A}_{X,2}$ of our construction (for $m = 3$, $r = 0$ and $b = 2$) has 6 states but is clearly not minimal since it is easily checked that the automaton of Figure 3.16 is the trim minimal automaton recognizing the set $\{(2^n, 3 \cdot 2^n) : n \in \mathbb{N}\}$. This illustrates that, in

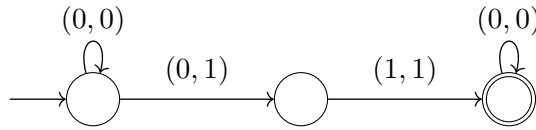


Figure 3.16: Minimal automaton recognizing the set $\{(2^n, 3 \cdot 2^n) : n \in \mathbb{N}\}$.

general, the minimization procedure is not only needed in the final projection $\Pi(\mathcal{A}_{m,r,2^p} \times \mathcal{A}_{X,2^p})$ as it is the case in the present work.

Nevertheless, we conjecture that the phenomenon described in this work for the Thue–Morse set also appears for all b -recognizable sets of the form

$$X_{b,c,M,R} = \{n \in \mathbb{N} : |\text{rep}_b(n)|_c \equiv R \pmod{M}\}$$

where b is an integer base, c is any digit in Σ_b , M is an integer greater than or equal to 2 and R is any possible remainder in $\llbracket 0, M-1 \rrbracket$. More precisely,

we conjecture that whenever the base b is a prime power, i.e. $b = q^p$ for some prime q , then the state complexity of $mX_{b,c,M,R} + r$ is given by the formula $Mk + \lceil \frac{z}{p} \rceil$ where k is the part of the multiple m that is prime to the base b , i.e. $m = kq^z$ with $\gcd(k, q) = 1$. Note that the set $\overline{\mathcal{T}}$ is of this form: $\overline{\mathcal{T}} = \{n \in \mathbb{N} : |\text{rep}_2(n)|_1 \equiv 0 \pmod{2}\}$. Another example is the following one. Consider $b = 3, c = 2, M = 3, R = 0$ and $m = 2, r = 0$. Our conjecture announce that the state complexity of the set $2X_{3,2,3,0}$ with respect to the base 3 is $3 \cdot 2 + \lceil \frac{0}{1} \rceil = 6$. Applying the same method as in the present chapter, we get the automata of Figure 3.17. For the sake of clarity, one omits the labels in the product automaton, since they can easily be deduced from the two other automata. Finally, by projecting each label of

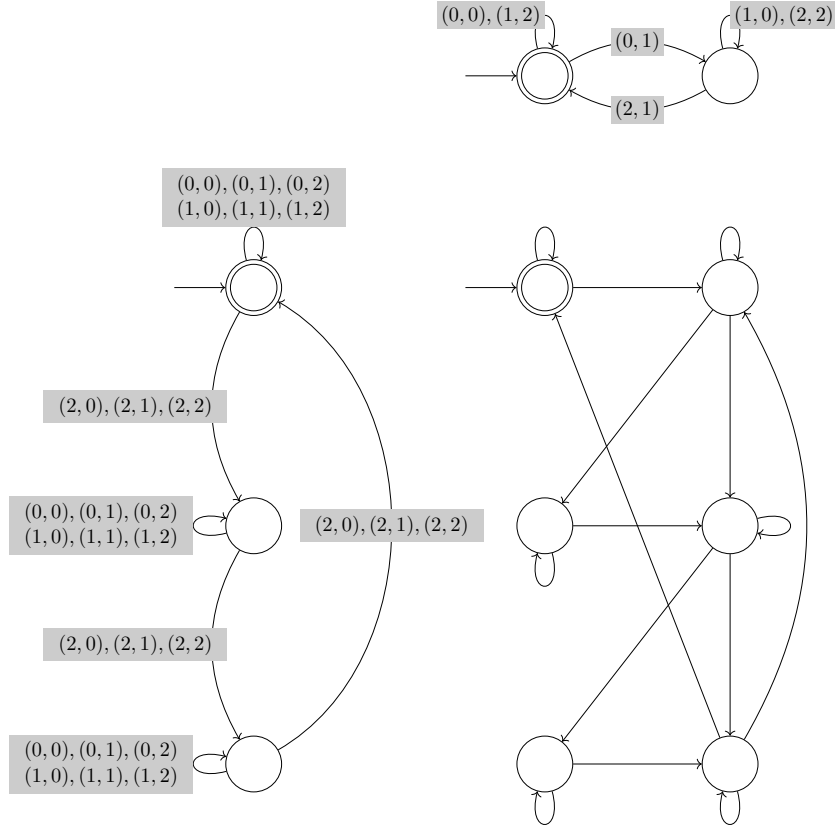


Figure 3.17: Construction of the automaton recognizing $2X_{3,2,3,0}$ in base 3.

the product automaton onto its second component, one gets the automaton depicted in Figure 3.18. One can verify that it is minimal, and it has 6 states, as announced by the conjecture.

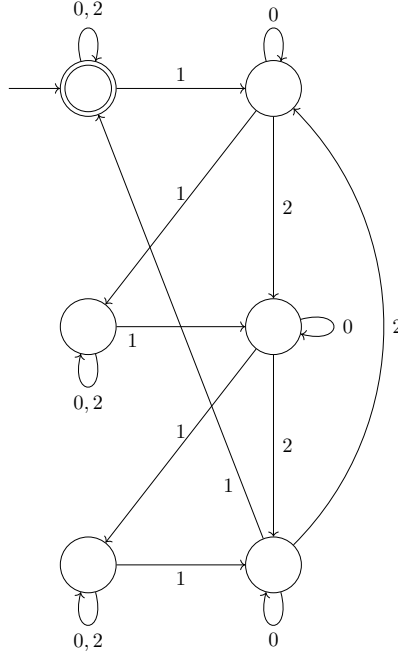


Figure 3.18: Minimal automaton of the base-3 expansions of the set $2X_{3,2,3,0}$.

Another potential future research direction in the continuation of the present work is to consider automata reading the expansions of numbers with least significant digit first. Both reading directions are relevant to different problems. For example, it is easier to compute addition thanks to an automaton reading expansions from “right to left” rather than from “left to right”. On the opposite, if we have in mind to generalize our problems to b -recognizable sets of real numbers (see for instance [14, 20, 26]), then the relevant reading direction is the one with most significant digit first. Further, there is no intrinsic reason why the state complexity from “left to right” should be the same as (or even close to) the one obtained from “right to left” since in general, it is well-known that the state complexity of an arbitrary language can greatly differ from the one of its reversed language. One can cite the papers [13] and [54]. For example, consider the language $L_n = 1(0+1)^n 1(0+1+\varepsilon)^n 0^*$ and its mirror K_n . The state complexity of L_n grows linearly with n , while the state complexity of K_n grows exponentially with n [54]. However, evaluating L_n as LSDF encodings or K_n as MSDF encodings gives the same finite (ultimately periodic) set.

Chapter 4

Automatic sequences based on Parry or Bertrand numeration systems

4.1 Introduction

The content of this chapter can be found in [56]. As shown by Theorem 1.6.3, there is a strong link between automatic sequences and recognizable sets of integers. For example, the Thue-Morse word \mathbf{t} defined in the previous chapter is 2-automatic, since it is generated by the DFAO of Figure 4.1. Moreover,

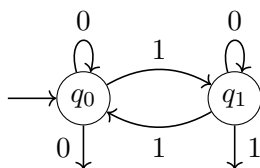


Figure 4.1: DFAO generating the Thue-Morse word.

we previously saw that the sets \mathcal{T} and $\overline{\mathcal{T}}$ are 2-recognizable. Furthermore, considering an integer base $b \geq 2$, a set $X \subseteq \mathbb{N}$ is b -recognisable if and only if its characteristic sequence is b -automatic. A characteristic sequence is defined over $\{0, 1\}$, but one can study automatic sequences defined over bigger alphabets. In this chapter, we are going to look at automatic sequences based on particular numeration systems and their properties. Roughly speaking, an automatic sequence is an infinite word over a finite alphabet such that its n^{th} symbol is obtained as the output given by a deterministic finite automaton

fed with the representation of n in a suitable numeration system. Precise definitions are given in Chapter 1, Section 1.6.

If we consider the usual base- b numeration system, then we get the family of b -automatic sequences. Thanks to a theorem of Cobham (Theorem 1.6.2), we know that these words are images under a coding of a fixed point of a b -uniform morphism. On a larger scale, if one considers abstract numeration systems (Definition 1.4.1) based on a regular language (see for instance [12, Chapter 3] or [66]), then we get exactly the family of morphic words (i.e. images under a coding of a fixed point of an arbitrary substitution). Between these two extremes, we have the automatic sequences based on Pisot, Parry and Bertrand numeration systems, and we have the following hierarchy (cf. Chapter 1, Section 1.3):

$$\begin{aligned} \text{Integer base systems} &\subsetneq \text{Pisot systems} \subsetneq \text{Parry systems} \\ &\subsetneq \text{Bertrand systems with a regular numeration language} \\ &\subsetneq \text{Abstract numeration systems.} \end{aligned}$$

Abstract numeration systems are uniquely based on the genealogical ordering of the words belonging to a regular language. This is contrasting with the more restricted case, treated in this chapter, of positional numeration systems based on an increasing sequence of integers: a digit occurring in n^{th} position is multiplied by the n^{th} element of the underlying sequence. As an example, consider the DFAO depicted in Figure 4.2 defined over the alphabet $\llbracket 0, 3 \rrbracket$, where the output function applied to a state gives the name of the state. If

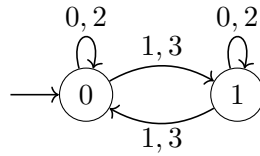


Figure 4.2: DFAO generating several automatic sequences.

one feeds this DFAO with a word $w \in \llbracket 0, 3 \rrbracket^*$, then the output is 0 if the sum of the digits appearing in w is even, the output is 1 otherwise. This DFAO generates automatic sequences based on different positional numeration systems. Table 4.3 takes some examples. The first one is an integer base system. The second one is the Fibonacci numeration system (Example 1.3.4), then one has the Parry non-Pisot numeration system of Lemma 1.3.29. Finally, we give a Bertrand non-Parry numeration system. Its is from Example 1.3.22. The initial conditions can be deduced from Definition 1.3.11. The last column

Type	Numeration system	Automatic sequence
Integer	$U_{i+1} = 2U_i$	01101001100101...
Pisot	$U_{i+2} = U_{i+1} + U_i$	01110100100011...
Parry	$U_{i+4} = 3U_{i+3} + 2U_{i+2} + 3U_i$	01011010010110...
Bertrand	$U_{i+1} = 3U_i + 1$	01011010010111...

Table 4.3: Examples of automatic sequences

of Table 4.3 gives the sequences obtained by feeding the DFAO of Figure 4.2 with the U -representations of integers (one forgets the transitions of label 2 and 3 when the alphabet of the numeration is $\llbracket 0, 1 \rrbracket$).

The Pisot-automatic sequences behave in many respects like b -automatic sequences [15, 37, 66]: if β is a Pisot number, then $\text{rep}_{U_\beta}(\mathbb{N})$ is regular, normalization $w \mapsto \text{rep}_{U_\beta}(\text{val}_{U_\beta}(w))$ (and thus addition) is computable by a finite automaton. Moreover, U_β -recognizable sets (β being either an integer base, either a Pisot number) are characterized in terms of first-order logic. This characterization in the b -automatic case is due to Büchi [18] and it was generalized to the Pisot case by Bruyère and Hansel [15]; also see [66, Chap. 3] and the references therein. Now by using the logical characterization, it is particularly straightforward to show that both the class of b -automatic sequences and the class of Pisot-automatic sequences enjoy many closure properties. For instance, both classes are closed under taking images by a uniform substitution and under periodic deletion of letters. For b -automatic sequences, these are classical results of Cobham [32]. The proofs of these results presented in [6, Chap. 6.8] are straightforward to generalize to Pisot-automatic sequences given the logical characterization of [15]. Indeed, as explained in Chapter 2, Section 2.1, one can make use of first-order logic and $\langle \mathbb{N}, +, V_U \rangle$ for Pisot-numeration systems, where $V_U(0) = U_0 = 1$ and for $n \neq 0$, $V_U(n)$ is the smallest U_i appearing in the U -representation of n with a non zero coefficient. Having Theorem 1.6.3 in mind, it is quite easy to transfer proofs of closure properties of b -automatic sequences (like in the beginning of Section 4.3) to Pisot-automatic sequences. For more closure properties, see [6, Chap. 6.8].

In this chapter, we study if some properties common to b -automatic sequences and Pisot-automatic sequences also hold for Parry-automatic sequences or more general automatic sequences. In a sense, we show that the generalization to Pisot numeration systems is the broadest possible generalization (regarding the numeration systems considered in this thesis) if the goal is to preserve the many good properties of b -automatic sequences.

It has been known before that a logical characterization no longer exists for Parry-automatic sequences. This follows from [38, Example 3]; we shall return to this matter in Section 4.3. We show that the closure properties mentioned above break when generalizing from Pisot to Parry and obtain as a corollary yet another proof showing that no logical characterization indeed exists for these sequences.

In combinatorics on words and in symbolic dynamics, the factor complexity of infinite words is often of interest. It was famously shown by Pansiot [61] that the factor complexity of an infinite word generated by a substitution is in one of the following five classes: $\Theta(1)$, $\Theta(n)$, $\Theta(n \log \log n)$, $\Theta(n \log n)$, or $\Theta(n^2)$. Previously, it has been known that the factor complexity of a b -automatic sequence is sublinear (that is, it is in $\mathcal{O}(1)$ or $\Theta(n)$) [32], [6, Theorem 10.3.1]. We extend this result and show that the factor complexity of any Parry-automatic sequence is sublinear. In contrast, we show by an explicit example that there exists a Bertrand-automatic sequence of superlinear complexity.

A well-known result concerning b -automatic sequences is their characterization in terms of the b -kernel originally due to Eilenberg [35]. This was generalized in [67] for all sequences associated with abstract numeration systems. The multidimensional version of this generalization [67, Proposition 32] however needs an additional assumption that is not required in the b -automatic case. We show in this chapter that this additional assumption is unnecessary also for positional numeration systems with a regular numeration language. Note that in [25], the authors give another definition of the kernel of a sequence associated with an abstract numeration system.

This chapter is organized as follows. In Section 4.2, we study the factor complexity of Parry-automatic sequences. Then in Section 4.3 we show that the closure properties of Pisot-automatic sequences do not hold for Parry-automatic sequences. In Section 4.4, the relationship of U -automaticity and the finiteness of the U -kernel is studied in the multidimensional setting. We conclude this chapter with an open problem.

4.2 Factor complexity

The *factor complexity function* $p_{\mathbf{x}}(n)$ of an infinite word \mathbf{x} counts the number of factors of length n occurring in \mathbf{x} . For more on factor complexity, see [12, Chapter 3].

Lemma 4.2.1. *For any infinite word \mathbf{x} over an alphabet Σ and any coding*

$\tau: \Sigma \rightarrow \Delta$, where Δ is an alphabet, we have

$$p_{\tau(\mathbf{x})}(n) \leq p_{\mathbf{x}}(n)$$

for all $n \in \mathbb{N}$.

Proof. Let $\mathbf{x} = x_0x_1x_2\cdots$ and $n \in \mathbb{N}$. Let $\mathbf{y} = \tau(\mathbf{x}) = y_0y_1\cdots$ and let $y_iy_{i+1}\cdots y_{i+n-1}$ be a factor of length n of \mathbf{y} . Then, since τ is a coding, $y_iy_{i+1}\cdots y_{i+n-1} = \tau(x_ix_{i+1}\cdots x_{i+n-1})$. Thus every factor of length n of $\tau(\mathbf{x})$ is completely determined by a factor of length n of \mathbf{x} . Hence the conclusion. \square

Let us recall the following classical result of Cobham.

Proposition 4.2.2 ([32],[6]). *Let $\mu: \Sigma^* \rightarrow \Sigma^*$ be a b -uniform morphism prolongeable on $a \in \Sigma$ and let $\tau: \Sigma \rightarrow \Delta$ be a coding. Let $\mathbf{x} = \tau(\mu^\omega(a))$. Then $p_{\mathbf{x}}(n) \leq b\ell^2n$ for all $n \geq 1$, where $\ell = \#\Sigma$. In particular, the factor complexity function of a b -automatic sequence is sublinear.*

Proof. By Lemma 4.2.1, it suffices to prove the upper bound for

$$\mathbf{y} = \mu^\omega(a).$$

Let $\mathbf{y} = y_0y_1y_2\cdots$. Let $n \geq 1$, let r be such that $b^{r-1} \leq n < b^r$ and let $y_iy_{i+1}\cdots y_{i+n-1}$ be a factor of \mathbf{y} of length n . Let $j = \lfloor \frac{i}{b^r} \rfloor$. Then $y_i\cdots y_{i+n-1}$ is a factor of $y_{jb^r}\cdots y_{(j+1)b^r}\cdots y_{(j+2)b^r-1}$. Moreover,

$$y_{jb^r}\cdots y_{(j+2)b^r-1} = \mu^r(y_jy_{j+1}),$$

thus $y_i\cdots y_{i+n-1}$ is completely determined by $i \pmod{b^r}$, y_j and y_{j+1} . There are b^r possibilities for $i \pmod{b^r}$, ℓ possibilities for y_j and ℓ possibilities for y_{j+1} . Hence

$$p_{\mathbf{y}}(n) \leq b^r\ell^2 \leq b\ell^2n.$$

\square

Next we generalize this result. For the proof, we need the following definition and proposition, as well as Pansiot's theorem; see [61], [12, Theorem 4.7.47].

Definition 4.2.3. Let $\mu: \Sigma^* \rightarrow \Sigma^*$ be a substitution. If there is a number $\alpha \geq 1$ such that $|\mu^n(a)| = \Theta(\alpha^n)$ for all $a \in \Sigma$, then we say that μ is *quasi-uniform*.

Proposition 4.2.4. *The factor complexity of a fixed point of a quasi-uniform substitution is sublinear.*

Theorem 4.2.5 (Pansiot). *Let \mathbf{x} be a purely morphic word. Then one of the following holds:*

- $p_{\mathbf{x}}(n) = \Theta(1)$,
- $p_{\mathbf{x}}(n) = \Theta(n)$,
- $p_{\mathbf{x}}(n) = \Theta(n \log \log n)$,
- $p_{\mathbf{x}}(n) = \Theta(n \log n)$, or
- $p_{\mathbf{x}}(n) = \Theta(n^2)$.

In order to generalize Proposition 4.2.2, recall Definition 1.2.3 and Proposition 1.2.5.

Theorem 4.2.6. *The factor complexity function of a Parry-automatic sequence is sublinear.*

Proof. Let U be a Parry numeration system having canonical automaton \mathcal{A} , and let \mathbf{x} be a U -automatic sequence generated by a DFAO \mathcal{B} . Recall that \mathcal{B} is complete and has a loop labelled by 0 on its initial state (cf. Definition 1.6.1). The product automaton $\mathcal{A} \times \mathcal{B}$ has $Q_{\mathcal{A} \times \mathcal{B}} = Q_{\mathcal{A}} \times Q_{\mathcal{B}}$ as set of states, the initial state q_0 is the pair made of the initial states of \mathcal{A} and \mathcal{B} , and the transition function is given by

$$\delta_{\mathcal{A} \times \mathcal{B}}((q, q'), i) = (\delta_{\mathcal{A}}(q, i), \delta_{\mathcal{B}}(q', i)).$$

We consider the automaton $\mathcal{A} \times \mathcal{B}$ as a DFAO by setting that the output function τ maps a state $(q_{\mathcal{A}}, q_{\mathcal{B}})$ of $\mathcal{A} \times \mathcal{B}$ to the output of the state $q_{\mathcal{B}}$ of \mathcal{B} . It is clear that \mathbf{x} is generated by $\mathcal{A} \times \mathcal{B}$.

Based on the automaton $\mathcal{A} \times \mathcal{B}$, we can build a substitution μ and consider the output function τ as a coding such that $\mathbf{x} = \tau(\mu^\omega(q))$ for some state $q \in Q_{\mathcal{A} \times \mathcal{B}}$. The construction is classical, see for instance [66, Lemma 2.28]. The substitution μ is defined as follows

$$\begin{aligned} \mu((q_{\mathcal{A}}, q_{\mathcal{B}})) &= (\delta_{\mathcal{A}}(q_{\mathcal{A}}, 0), \delta_{\mathcal{B}}(q_{\mathcal{B}}, 0))(\delta_{\mathcal{A}}(q_{\mathcal{A}}, 1), \delta_{\mathcal{B}}(q_{\mathcal{B}}, 1)) \\ &\quad \cdots (\delta_{\mathcal{A}}(q_{\mathcal{A}}, C_U - 1), \delta_{\mathcal{B}}(q_{\mathcal{B}}, C_U - 1)). \end{aligned}$$

In the latter expression, since \mathcal{A} is in general not complete, if $\delta_{\mathcal{A}}(q_{\mathcal{A}}, j)$ is undefined, then $(\delta_{\mathcal{A}}(q_{\mathcal{A}}, j), \delta_{\mathcal{B}}(q_{\mathcal{B}}, j))$ is replaced by ε . Notice that the substitution μ is defined over the alphabet $Q_{\mathcal{A} \times \mathcal{B}}$. Since $\mathcal{A} \times \mathcal{B}$ has a loop with label $(0, 0)$ on its initial state q_0 , iterating μ on this state generates the sequence of states $\mu^\omega(q_0)$ in $\mathcal{A} \times \mathcal{B}$ reached from the initial state by the words of $\text{rep}_U(\mathbb{N})$ in genealogical order.

Since every state of a canonical automaton of a Parry numeration system is final, the coding τ is non-erasing. Then by Lemma 4.2.1, the factor complexity of \mathbf{x} is at most the factor complexity of $\mu^\omega(q_0)$, hence it is sufficient to show that a fixed point of μ has sublinear complexity. This is accomplished as follows. First we establish that there is a number α such that $|\mu^n(q)| = \Theta(\alpha^n)$ for every state $q \in Q_{\mathcal{A} \times \mathcal{B}}$. In other words, we show that the substitution μ is quasi-uniform. It then follows from Proposition 4.2.4 that the factor complexity of a fixed point of μ is sublinear.

Let us define a projection mapping $\varphi: Q_{\mathcal{A} \times \mathcal{B}} \rightarrow Q_{\mathcal{A}}$ by setting for any state $(q_{\mathcal{A}}, q_{\mathcal{B}})$ of $\mathcal{A} \times \mathcal{B}$, $\varphi((q_{\mathcal{A}}, q_{\mathcal{B}})) = q_{\mathcal{A}}$. By the definition of the product automaton $\mathcal{A} \times \mathcal{B}$, we have $\varphi(\delta_{\mathcal{A} \times \mathcal{B}}((q_{\mathcal{A}}, q_{\mathcal{B}}), a)) = \delta_{\mathcal{A}}(\varphi((q_{\mathcal{A}}, q_{\mathcal{B}})), a)$ for all letter a and all states $q_{\mathcal{A}}$ and $q_{\mathcal{B}}$.

Recall that given an automaton \mathcal{C} with adjacency matrix $Adj(\mathcal{C})$, the entry $(Adj(\mathcal{C}))_{i,j}^n$ counts the number of distinct paths of length n from state i to state j (cf. Chapter 1, Section 1.2). Let $(q_{\mathcal{A}}, q_{\mathcal{B}})$ be a state of $\mathcal{A} \times \mathcal{B}$ and consider all paths of length n starting from this state. These paths can be identified with their edge labels. Given such a path with edge label w , we find by applying the projection mapping φ a path in \mathcal{A} with edge label w starting at the state $q_{\mathcal{A}}$. Conversely, given a path of length n in \mathcal{A} with edge label w starting at state $q_{\mathcal{A}}$, there is a path with edge label w in $\mathcal{A} \times \mathcal{B}$ starting at the state $(q_{\mathcal{A}}, q_{\mathcal{B}})$ because the automaton \mathcal{B} is complete (cf. Definition 1.6.1). Denoting the total number of paths of length n starting at a state q of $\mathcal{A} \times \mathcal{B}$ by $K_q(n)$, we have thus argued that

$$K_q(n) = \sum_{r \in Q_{\mathcal{A} \times \mathcal{B}}} (Adj(\mathcal{A} \times \mathcal{B}))_{q,r}^n = \sum_{s \in Q_{\mathcal{A}}} (Adj(\mathcal{A}))_{\varphi(q),s}^n.$$

The canonical automaton of a Parry numeration system is primitive (cf. Lemma 1.3.19), we have for each i and j that $(Adj(\mathcal{A}))_{i,j}^n = \Theta(\alpha^n)$, where α is the Perron-Frobenius eigenvalue of \mathcal{A} (see Proposition 1.2.5). Thus $K_q(n) = \Theta(\alpha^n)$. By rephrasing the number $K_q(n)$ in terms of substitutions, we have $|\mu^n(q)| = K_q(n)$. Hence $|\mu^n(q)| = \Theta(\alpha^n)$, and we get the conclusion. \square

Notice that in fact we showed in the proof of Theorem 4.2.6 that for each Parry-automatic sequence \mathbf{x} there exist a coding τ and a quasi-uniform substitution μ such that $\mathbf{x} = \tau(\mu^\omega(a))$ for a letter a . This should be contrasted with the fact that b -automatic sequences are codings of fixed points of *uniform* substitutions.

As showned in Lemma 1.3.25, there are Bertrand numeration systems that are not Parry numeration systems. We show that Theorem 4.2.6 does

not generalize to Bertrand-automatic sequences. In this aim, we need the following definition and theorem (see [12, Theorem 4.7.66]).

Definition 4.2.7. Let Σ be an alphabet, let $\mu: \Sigma^* \rightarrow \Sigma^*$ be a morphism and let $x \in \Sigma^*$. We say that x is *bounded* under μ if the sequence $(|\mu^n(x)|)_{n \in \mathbb{N}}$ is bounded.

Theorem 4.2.8. Let Σ be an alphabet and let $\mathbf{x} \in \Sigma^\omega$ be a purely morphic word. Let $\mu: \Sigma^* \rightarrow \Sigma^*$ be a morphism that generates \mathbf{x} . If \mathbf{x} is not ultimately periodic and if infinitely many distinct factors of \mathbf{x} are bounded under μ , then $p_{\mathbf{x}}(n) = \Theta(n^2)$.

Theorem 4.2.9. There exists a Bertrand-automatic sequence with superlinear factor complexity.

Proof. Consider the numeration system given by initial condition $B_0 = 1$ and the recurrence $B_{i+1} = 3B_i + 1$ for all $i \in \mathbb{N}$. In Example 1.3.22, it was shown that this numeration system is a Bertrand numeration system.

The substitution associated with the canonical automaton, depicted in Figure 1.4, is $\mu: a \mapsto aaab, b \mapsto b$; see in the proof of Theorem 4.2.6 how this substitution is defined. Let \mathbf{x} be the infinite fixed point of μ . Observe that \mathbf{x} is a Bertrand-automatic sequence. It is easy to see that $ab^n a$ occurs in \mathbf{x} for all $n \in \mathbb{N}$. Thus \mathbf{x} is aperiodic and there are infinitely many bounded factors occurring in \mathbf{x} . It follows by Theorem 4.2.8 that the factor complexity of \mathbf{x} is quadratic. \square

4.3 Closure properties

It is easy to see that the image of a b -automatic sequence $\mathbf{x} \in \Sigma^\omega$ under a substitution $\mu: \Sigma^* \rightarrow \Delta^*$ of constant length ℓ is again a b -automatic sequence. Indeed, Theorem 1.6.3 implies that for all $a \in \Sigma$ there is a first-order formula $\varphi_a(i)$ in $\langle \mathbb{N}, +, V_b \rangle$ which holds if and only if $\mathbf{x}_i = a$. Let us then define for each $c \in \Delta$ a formula $\psi_c(i)$ that holds if and only if $\mu(\mathbf{x})_i = c$. For every i there are unique q and r such that $0 \leq r < \ell$ and $i = \ell q + r$. For each $a \in \Sigma$, we can construct a formula $\sigma_a(r)$ that holds if and only if $\mu(a)$ contains the letter c at position r (indexing from 0). Setting

$$\psi_c(i) = (\exists q)(\exists r < \ell)(i = \ell q + r \wedge \bigvee_{a \in \Sigma} (\varphi_a(q) \wedge \sigma_a(r)))$$

certainly has the desired effect. Notice that this is indeed a formula in $\langle \mathbb{N}, +, V_b \rangle$ since ℓ is constant. Therefore it follows from Theorem 1.6.3 that $\mu(\mathbf{x})$ is b -automatic. For a proof not based on logic, see [6, Corollary 6.8.3].

Example 4.3.1. Assume $A = \{a, c\}$, $B = \{d, e\}$, $\ell = 3$ and set $\mu(a) = dde$, $\mu(c) = ede$. In this case, the formula $\psi_d(i)$ is given by

$$(\exists q)(\exists r < 3)(i = 3q + r \wedge [(\varphi_a(q) \wedge (r = 0 \vee r = 1)) \vee (\varphi_c(q) \wedge r = 1)]).$$

The same construction can be applied to numeration systems canonically associated with a Pisot number [15]. Here, we show that this closure property does not hold for Parry-automatic sequences.

Theorem 4.3.2. *There is a Parry numeration system U such that the class of U -automatic sequences is not closed under taking image by a uniform morphism.*

Throughout this section, we shall consider a specific numeration system U given by the recurrence

$$U_{i+4} = 3U_{i+3} + 2U_{i+2} + 3U_i \quad \forall i \in \mathbb{N}, \quad (4.1)$$

with initial values $U_0 = 1$, $U_1 = 4$, $U_2 = 15$, and $U_3 = 54$ (cf. [38, Example 3]). The characteristic polynomial has two real roots ϱ and γ and two complex roots with modulus less than 1. We have $\varrho \approx 3.61645$ and $\gamma \approx -1.09685$. Thus from the basic theory of linear recurrent sequences, we have $U_i \sim c\varrho^i$ for some constant c . A simple verification shows that the characteristic polynomial of the recurrence is the minimal polynomial of ϱ hence, in particular, γ is an algebraic conjugate of ϱ . Since $|\gamma| > 1$, the number ϱ is not a Pisot number. It is however a Parry number, as it is readily checked that $d_\varrho(1) = 3203$. Thus U is a Parry numeration system. Moreover, we have $U = U_\varrho$. Recall that $\text{rep}_U(\mathbb{N})$ is regular as this holds for all Parry numeration systems (cf. Chapter 1, Section 1.3). Note that the `Mathematica` code of the computations presented in this section can be found in Appendix B.

Consider the characteristic sequence \mathbf{x} of the set $\{U_i : i \in \mathbb{N}\}$:

$$\mathbf{x} = 010010000000000100000000000 \dots$$

From Theorem 1.6.3, this sequence is U -automatic. We consider the constant length substitution $\mu: 0 \mapsto 0^t, 1 \mapsto 10^{t-1}$ with $t \geq 4$. Observe that $\mu(\mathbf{x})$ is the characteristic sequence of $\{tU_i : i \in \mathbb{N}\}$. The multiplier 4 is the first interesting value to consider because $\text{rep}_U(\{jU_i : i \in \mathbb{N}\}) = j0^*$ for $j = 2, 3$, and we trivially get U -recognizable sets. Our aim is to show that $\mu(\mathbf{x})$ is not U -automatic (see Corollary 4.3.5). This will prove Theorem 4.3.2.

We begin with a lemma and an auxiliary result that is of independent interest. The following lemma is technical and is obtained by adapting [72, Lemma 2.2] to our situation. Since ϱ is an algebraic number of degree 4, it is well-known that every element in $\mathbb{Q}(\varrho)$ can be expressed as a polynomial in ϱ of degree at most 3 with coefficients in \mathbb{Q} .

Lemma 4.3.3. *Let $x \in [0, 1) \cap \mathbb{Q}(\varrho)$, and write*

$$x = q^{-1} \sum_{i=0}^3 p_i \varrho^i$$

for integers q and p_i , $q > 0$. If $d_\varrho(x) = x_1 x_2 x_3 \cdots$ is ultimately periodic, then

$$q^{-1} \sum_{i=0}^3 p_i \gamma^i = \sum_{i=1}^{+\infty} x_i \gamma^{-i}.$$

Proposition 4.3.4. *Let $r \geq 2$ be an integer. If t is an integer such that $4 \leq t \leq \lfloor \varrho^r \rfloor$, then the ϱ -expansion of the number t/ϱ^r is aperiodic.*

Proof. Let us first make the additional assumption that $t \geq \lceil \varrho^{r-1} \rceil$ and prove the result in this case. Set $x = t/\varrho^r = q^{-1} \sum_{i=0}^3 p_i \varrho^i$, and assume for a contradiction that $d_\varrho(x)$ is ultimately periodic. Write $d_\varrho(x) = x_1 x_2 \cdots$. Since ϱ and γ are conjugates,

$$\frac{t}{\gamma^r} = q^{-1} \sum_{i=0}^3 p_i \gamma^i$$

and it follows from Lemma 4.3.3 that

$$\frac{t}{\gamma^r} = \sum_{i=1}^{+\infty} x_i \gamma^{-i}.$$

In other words, for any positive integer k , we have

$$t = \sum_{i=1}^{+\infty} x_i \gamma^{-i+r} = S_{1,k} + S_{k+1,+\infty}, \quad (4.2)$$

where $S_{m,n} = \sum_{i=m}^n x_i \gamma^{-i+r}$. Since $x_i \leq 3$ for all $i \geq 1$ and since γ is negative, we can remove odd powers of γ and obtain

$$S_{r+1,+\infty} = \sum_{i=1}^{+\infty} x_{i+r} \gamma^{-i} \leq 3 \sum_{i=1}^{+\infty} \gamma^{-2i} = \frac{3\gamma^{-2}}{1-\gamma^{-2}} < 15. \quad (4.3)$$

Similarly by discarding the odd terms and estimating $x_i \leq 3$, we obtain

$$S_{1,r} = \sum_{i=0}^{r-1} x_{r-i} \gamma^i \leq \frac{3(1-\gamma^{2(k+1)})}{1-\gamma^2}, \quad (4.4)$$

where k is the largest integer such that $2k \leq r - 1$. Combining (4.2), (4.3), and (4.4) with our assumption $t \geq \lceil \varrho^{r-1} \rceil$, we obtain that

$$\varrho^{r-1} < \frac{3(1-\gamma^{2(k+1)})}{1-\gamma^2} + 15. \quad (4.5)$$

The left side of (4.5) clearly increases faster than the right side when $r \rightarrow \infty$ since $\varrho \approx 3.62$ and $\gamma^2 \approx 1.20$. Using these approximations, it is straightforward to compute that for $r = 4$ the left side of (4.5) is approximately 47 while the right side is only approximately 22. Hence it must be that $r \leq 3$.

We are thus left with a few cases we have to deal with separately. The idea is the same, but we need to actually compute some digits x_i . Suppose first that $r = 3$. Like previously, we see that

$$S_{4,+\infty} = \gamma^{-2} \sum_{i=4}^{+\infty} x_{i+3} \gamma^{-i+2} \leq 3\gamma^{-4}/(1-\gamma^{-2}) < 12.28.$$

Since $14 = \lceil \varrho^2 \rceil \leq t \leq \lfloor \varrho^3 \rfloor = 47$, by enumerating all possibilities for the word $x_1 x_2 x_3$ (within the given range for t), we see that $f(t) = t - S_{1,3}$ is minimized when $t = \lceil \varrho^2 \rceil = 14$.

t	$d_\varrho(t/\varrho^3)$	$t - S_{1,3}$
14	100...	12.797
15	101...	12.797
16	102...	12.797
17	110...	16.894
18	111...	16.894
\vdots	\vdots	\vdots
44	311...	40.488
45	312...	40.488
46	313...	40.488
47	320...	45.584

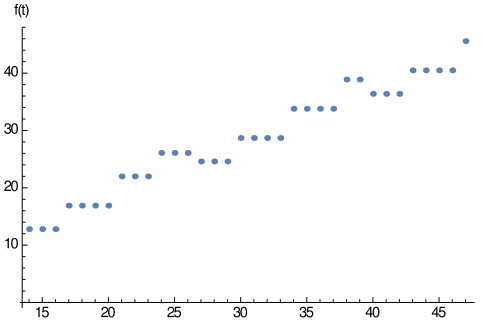


Figure 4.4: Values of $t - S_{1,3}$

In this case, $x_1 x_2 x_3 = 100$ and $t - S_{1,3} > 12.79$. This contradicts (4.2). Suppose then that $r = 2$. We proceed as above, but now we are interested in the number $t - S_{1,12}$ instead. By enumerating all possibilities, we see that $t - S_{1,12}$ is minimized for $t = \lceil \varrho \rceil = 4$. Then $x_1 \cdots x_{12} = 101111202300$ and $t - S_{1,12} > 6.12$. Since $S_{13,+\infty} < 6$, we get a contradiction.

Suppose finally that $4 \leq t < \lceil \varrho^{r-1} \rceil$. If $r = 2$, then $\lceil \varrho^{2-1} \rceil = 4$, thus we may suppose that $r > 2$. As $t < \varrho^{r-1}$, we see that $\varrho \cdot t / \varrho^r < 1$ meaning that $x_1 = 0$. Thus $x_2 x_3 \dots$ is the ϱ -expansion of t / ϱ^{r-1} . Inductively it follows that this expansion is aperiodic, hence the conclusion. \square

We now have tools to prove that the set $\{tU_i : i \in \mathbb{N}\}$ is not U -recognizable for $t \geq 4$.

Corollary 4.3.5. *The set $\{tU_i : i \in \mathbb{N}\}$ is not U -recognizable for $t \geq 4$. In other words, its characteristic sequence $\mu(\mathbf{x})$ is not U -automatic.*

Proof. Let $t \geq 4$, and suppose that $\lceil \varrho^{r-1} \rceil \leq t \leq \lfloor \varrho^r \rfloor$ for some $r \geq 2$. Recall that $U_i \sim c\varrho^i$ for some positive constant c . By some simple asymptotic analysis, it is easy to see that $U_{i+r-1} < tU_i < U_{i+r}$ for i large enough. Hence, for i large enough, $\text{rep}_U(tU_i)$ is a word of length $i + r$ (starting with a non-zero digit). Let $k > 0$. We show that, for large enough i , $\text{rep}_U(tU_i)$ and $d_\varrho(t/\varrho^r)$ have the same prefix of length k . See Table 4.5 for an example. Assume that

$$\text{rep}_U(tU_i) = x_1 \cdots x_k x_{k+1} \cdots x_{i+r}.$$

The extremal values for $x_{k+1} \cdots x_{i+r}$ are 0^{i+r-k} and $\text{rep}_U(U_{i+r-k} - 1)$ due to the greediness of the representations. Hence

$$0 \leq tU_i - x_1 U_{i+r-1} - \cdots - x_k U_{i+r-k} < U_{i+r-k}.$$

Dividing by U_{i+r} and letting i tend to infinity, we get

$$0 \leq \frac{t}{\varrho^r} - \frac{x_1}{\varrho} - \cdots - \frac{x_k}{\varrho^k} < \frac{1}{\varrho^k}.$$

Otherwise stated, the first k digits of $d_\varrho(t/\varrho^r)$ are $x_1 \cdots x_k$.

Now proceed by contradiction and assume that $\{\text{rep}_U(tU_i) : i \in \mathbb{N}\}$ is accepted by a finite deterministic automaton. By a classical pumping argument, there are $u, v, w \in \Sigma_U^*$, with v non-empty, such that uv^jw is accepted by this automaton for all $j \geq 0$. Hence, $d_\varrho(t/\varrho^r)$ should be of the form uv^ω contradicting Proposition 4.3.4. \square

Corollary 4.3.5 is interesting because it shows that addition in U is not computable by a finite automaton. Indeed, if this was the case, then surely multiplication by any constant would be computable by a finite automaton contrary to Corollary 4.3.5. This result is not new: it already appears in [38, Example 3]. The conclusion is that addition in a Parry numeration system is not necessarily computable by a finite automaton. This shows in particular

i	$\text{rep}_U(4U_i)$
0	1 0
1	1 0 1
2	1 0 1 1
3	1 0 1 1 1
4	1 0 1 1 1 1
5	1 0 1 1 1 1 2
6	1 0 1 1 1 1 2 0
7	1 0 1 1 1 1 2 0 3
8	1 0 1 1 1 1 2 0 2 3
9	1 0 1 1 1 1 2 0 2 3 0

Table 4.5: Representations of the first $4U_i$.

that Parry-recognizable sets do not have a characterization based on first-order logic like Pisot-recognizable sets have. This is a considerable defect of Parry numeration systems that are not Pisot.

Let us then describe why a word obtained from a b -automatic sequence by periodically deleting letters is still b -automatic. Suppose that \mathbf{x} is a b -automatic sequence over Σ , and let \mathbf{y} be the word obtained from \mathbf{x} by keeping only the letters at positions $0, t, 2t, 3t, \dots$ for a fixed integer $t \geq 2$. In other words, we have $\mathbf{y}_i = \mathbf{x}_{ti}$. As mentioned at the beginning of this section, for each $a \in \Sigma$, there is a first-order formula $\varphi_a(i)$ in $\langle \mathbb{N}, +, V_b \rangle$ such that it holds if and only if $\mathbf{x}_i = a$. By substituting i by ti in $\varphi_a(i)$, we obtain a new first-order formula in $\langle \mathbb{N}, +, V_b \rangle$ such that it holds if and only if $\mathbf{y}_i = a$. It follows from Theorem 1.6.3 that \mathbf{y} is b -automatic. Again, a similar construction works in the Pisot case. See also [6, Theorem 6.8.1].

Let us next show that the class of U -automatic sequences is not closed under periodic deletion. Consider the characteristic sequence \mathbf{y} of the set $\{U_i/2 : i \in \mathbb{N} \text{ and } U_i \in 2\mathbb{N}\}$:

$$\mathbf{y} = 0010000000000000000000001 \dots$$

This sequence \mathbf{y} is obtained from the characteristic sequence \mathbf{x} of the set $\{U_i : i \in \mathbb{N}\}$ by removing its every second letter. Indeed, $\mathbf{y}_i = \mathbf{x}_{2i}$ hence $\mathbf{y}_i = 1$ if and only if $2i$ belongs to $\{U_j : j \in \mathbb{N}\}$. We will show that \mathbf{y} is not U -automatic, which will prove the following theorem.

Theorem 4.3.6. *There is a Parry numeration system U such that the class of U -automatic sequences is not closed under periodic deletion.*

Let us begin with the following result.

Proposition 4.3.7. *The ϱ -expansion of $1/2$ is aperiodic.*

Proof. Assume for a contradiction that $d_\varrho(1/2) = x_1x_2\cdots$ is ultimately periodic. As in the proof of Proposition 4.3.4, we obtain that

$$\frac{1}{2} = \sum_{i=1}^{+\infty} x_i \gamma^{-i} = S_{1,k} + S_{k+1,+\infty},$$

where $S_{m,n} = \sum_{i=m}^n x_i \gamma^{-i}$. One can compute that the first 21 digits of $d_\varrho(1/2)$ are $x_1 \cdots x_{21} = 123102303001010220123$. This computation actually needs some extra accuracy. It is sufficient to know that 3.61645454325 are correct initial digits for ϱ . Using this information on $x_1 \cdots x_{21}$, it is computed that

$$S_{1,21} < -2.20.$$

Since γ is negative and $x_i \leq 3$ for all $i \geq 1$, we obtain that

$$S_{22,+\infty} \leq \frac{3\gamma^{-22}}{1-\gamma^{-2}} < 2.33.$$

The two preceding inequalities show that $1/2 < -2.20 + 2.33 = 0.13$, which is obviously absurd. \square

Interestingly the ϱ -expansion of $1/3$ is ultimately periodic. Indeed, it can be shown that $d_\varrho(1/3) = 10(2212)^\omega$.

Corollary 4.3.8. *The set $\{U_i/2 : i \in \mathbb{N} \text{ and } U_i \in 2\mathbb{N}\}$ is not U -recognizable. In other words, its characteristic sequence \mathbf{y} is not U -automatic.*

Proof. We follow steps similar to those of the proof of Corollary 4.3.5. From (4.1), it is clear that $U_{i-1} < \lfloor U_i/2 \rfloor < U_i$ for $i \geq 1$, so that $\text{rep}_U(\lfloor U_i/2 \rfloor)$ is a word of length i . Let $k > 0$. We show that, for large enough i , $\text{rep}_U(\lfloor U_i/2 \rfloor)$ and $d_\varrho(1/2)$ have the same prefix of length k . Assume that

$$\text{rep}_U(\lfloor U_i/2 \rfloor) = x_1 \cdots x_k x_{k+1} \cdots x_i.$$

Again, the extremal possible values for $x_{k+1} \cdots x_i$ are 0^{i-k} and $\text{rep}_U(U_{i-k}-1)$ due to the greediness of the representations. Therefore

$$0 \leq \lfloor U_i/2 \rfloor - x_1 U_{i-1} - \cdots - x_k U_{i-k} < U_{i-k}.$$

Clearly $\lfloor U_i/2 \rfloor / U_i \xrightarrow{i \rightarrow +\infty} 1/2$ thus, dividing by U_i and letting i tend to infinity, we obtain

$$0 \leq \frac{1}{2} - \frac{x_1}{\varrho} - \dots - \frac{x_k}{\varrho^k} < \frac{1}{\varrho^k}.$$

Hence the first k digits of $d_\varrho(1/2)$ are $x_1 \cdots x_k$. This means that the words of the language $\{\text{rep}_U(U_i/2) : i \in \mathbb{N} \text{ and } U_i \in 2\mathbb{N}\}$ share longer and longer prefixes with $d_\varrho(1/2)$.

The results follows by an argument similar to the final paragraph of the proof of Corollary 4.3.5: if $\{\text{rep}_U(U_i/2) : i \in \mathbb{N} \text{ and } U_i \in 2\mathbb{N}\}$ is accepted by a finite deterministic automaton, then $d_\varrho(1/2)$ is ultimately periodic, and this is impossible by Proposition 4.3.7. \square

Notice that the proof in fact shows that the set $\{\lfloor U_i/2 \rfloor : i \in \mathbb{N}\}$ is not U -recognizable. Even though \mathbf{y} is not U -automatic, we suspect that the word obtained from \mathbf{x} , the characteristic sequence of $\{U_i : i \in \mathbb{N}\}$, by keeping only the letters at indices that are divisible by 3 is U -automatic. This would follow from our conjecture that $\{\text{rep}_U(U_n/3) : i \in \mathbb{N} \text{ and } U_i \equiv 0 \pmod{3}\}$ equals $11 + 10(2212)^*(3 + 23 + 222 + 2213)$, but we have not attempted to prove this rigorously. Notice that U_i is divisible by 3 when $i \geq 2$.

4.4 Multidimensional sequences

By Proposition 1.6.6, an infinite word is U -automatic with respect to a numeration system U with $\text{rep}_U(\mathbb{N})$ regular if and only its U -kernel is finite. Moreover, this is true more generally for abstract numeration systems. The generalization of this result to multidimensional sequences $\mathbf{x} = (x_{i,j})_{i,j \in \mathbb{N}}$ (see [67, Proposition 32]) is however slightly problematic as an extra assumption on the projections $(x_{\ell,j})_{j \in \mathbb{N}}$ and $(x_{i,\ell})_{i \in \mathbb{N}}$ is required. This extra assumption is however unnecessary for positional numeration systems considered in this work.

For the sake of clarity, we limit our presentation to two-dimensional sequences. We will consider finite automata reading pairs of digits. In particular, a pair of words can be read only if the two components have the same length. As explained in Chapter 1, with positional numeration systems, when considering two representations of different length, the shorter is padded with leading zeros. For general abstract numeration systems an additional padding letter needs to be added, and this causes some complications (cf. [67, Definition 30]).

Definition 4.4.1. Let U be a numeration system. A 2-dimensional word $\mathbf{x} = (x_{i,j})_{i,j \in \mathbb{N}}$ over an alphabet Σ is U -automatic if there exists a complete

DFAO $(Q, q_0, \Sigma_U \times \Sigma_U, \delta, \Sigma, \tau)$ with transition function $\delta: Q \times (\Sigma_U \times \Sigma_U) \rightarrow Q$ and output function $\tau: Q \rightarrow \Sigma$ such that $\delta(q_0, (0, 0)) = q_0$ and

$$x_{i,j} = \tau(\delta(q_0, (0^{\ell-|\text{rep}_U(i)|} \text{rep}_U(i), 0^{\ell-|\text{rep}_U(j)|} \text{rep}_U(j)))), \quad \forall i, j \in \mathbb{N},$$

where $\ell = \max\{|\text{rep}_U(i)|, |\text{rep}_U(j)|\}$. The 2-dimensional word \mathbf{x} is said *b-automatic* (resp. *Parry-automatic*, *Bertrand-automatic*) if $U = (b^i)_{i \in \mathbb{N}}$ for an integer $b \geq 2$ (resp. U is a Parry numeration system, U is a Bertrand numeration system).

Definition 1.6.4 is extended as follows (we make use of the notation $k(s, i)$ introduced therein).

Definition 4.4.2. The *U-kernel* of a 2-dimensional word $\mathbf{x} = (x_{i,j})_{i,j \in \mathbb{N}}$ over Σ is the set

$$\ker_U(\mathbf{x}) = \{(x_{k(s,i), k(t,j)})_{i,j \in \mathbb{N}} \in \Sigma^{\mathbb{N}^2} : s, t \in \Sigma_U^*, |s| = |t|\}.$$

Let us then state and prove the result mentioned above.

Proposition 4.4.3. *Let U be a numeration system such that the numeration language $\text{rep}_U(\mathbb{N})$ is regular. A 2-dimensional word $\mathbf{x} = (x_{i,j})_{i,j \in \mathbb{N}}$ is U -automatic if and only if its U -kernel is finite.*

Proof. Let $\mathbf{x} = (x_{i,j})_{i,j \in \mathbb{N}}$ be a 2-dimensional word. For $s \in \Sigma_U$, define

$$\mathcal{L}(s) = 0^* \text{rep}_U(\mathbb{N}) \cdot s^{-1} = \{w \in \Sigma_U^* : ws \in 0^* \text{rep}_U(\mathbb{N})\}.$$

One can genealogically order the set $\mathcal{L}(s)$ for all $s \in \Sigma_U^*$. Note that $\mathcal{L}(s)$ is not necessarily infinite: it can be finite or empty. Also notice that the set

$$J = \{\mathcal{L}(s) : s \in \Sigma_U^*\} \subseteq 2^{\Sigma_U^*}$$

is finite, because by hypothesis $\text{rep}_U(\mathbb{N})$ is regular, hence accepted by a finite automaton. Set $m = \#J$. We thus have $\#\{(\mathcal{L}(s), \mathcal{L}(t)) : s, t \in \Sigma_U^*\} = m^2$.

First, suppose that the 2-dimensional word \mathbf{x} is U -automatic: there is a complete DFAO $(Q, q_0, \Sigma_U \times \Sigma_U, \delta, \Sigma, \tau)$ that generates \mathbf{x} . Let $s, t \in \Sigma_U^*$ such that $|s| = |t|$ and $\mathcal{L}(s), \mathcal{L}(t) \neq \emptyset$. Since $\mathcal{L}(s)$ and $\mathcal{L}(t)$ are genealogically ordered, one can associate to the couple $[s, t]$ a sequence of states

$$(q[s, t]_{i,j})_{i,j \in \mathbb{N}}$$

defined by

$$q[s, t]_{i,j} = \delta(q_0, (0^{\ell-|v_{s,i}|} v_{s,i}, 0^{\ell-|v_{t,j}|} v_{t,j})) \quad \forall i, j \in \mathbb{N}$$

where $v_{s,i}$ is the $(i+1)^{\text{th}}$ word of $\mathcal{L}(s)$, $v_{t,j}$ is the $(j+1)^{\text{th}}$ word of $\mathcal{L}(t)$ and $\ell = \max\{|v_{s,i}|, |v_{t,j}|\}$. Note that this latter sequence could possibly be finite. There are at most m^2 distinct sequences $(q[s, t]_{i,j})_{i,j \in \mathbb{N}}$. Since we are interested in the U -kernel of \mathbf{x} , we have to consider the sequence of states

$$(\delta(q[s, t]_{i,j}, (s, t)))_{i,j \in \mathbb{N}}$$

which is, up to the application of τ , an element of the U -kernel of \mathbf{x} . For any pair $(u, w) \in \Sigma_U^* \times \Sigma_U^*$ of words such that $\mathcal{L}(u) = \mathcal{L}(s)$, $\mathcal{L}(w) = \mathcal{L}(t)$ and $|u| = |w|$, we also have to consider the sequence

$$(\delta(q[s, t]_{i,j}, (u, w)))_{i,j \in \mathbb{N}},$$

which is, also up to an application of τ , an element of the kernel. For a given pair $[s, t]$, there are at most $(\#Q)^{\#Q}$ distinct sequences of this type. Indeed, for every state $q \in Q$, $\delta(q, (u, w))$ can take at most $\#Q$ values. Therefore the cardinality of the U -kernel of \mathbf{x} is at most $m^2 \cdot (\#Q)^{\#Q}$.

Conversely, let K denote the U -kernel of \mathbf{x} and suppose that it is finite. Let us define a DFAO \mathcal{M} with states set

$$Q = J \times J \times K,$$

transition function δ , output function τ and initial state

$$q_0 = (0^* \text{rep}_U(\mathbb{N}), 0^* \text{rep}_U(\mathbb{N}), (x_{i,j})_{i,j \in \mathbb{N}}) = (\mathcal{L}(\varepsilon), \mathcal{L}(\varepsilon), (x_{k(\varepsilon,i), k(\varepsilon,j)})_{i,j \in \mathbb{N}}).$$

For a state $q = (\mathcal{L}(s), \mathcal{L}(t), (x_{k(s,i), k(t,j)})_{i,j \in \mathbb{N}})$ in Q , with $|s| = |t|$, and each pair (a, b) of digits in $\Sigma_U \times \Sigma_U$, we set

$$\delta(q, (a, b)) = (\mathcal{L}(as), \mathcal{L}(bt), (x_{k(as,i), k(bt,j)})_{i,j \in \mathbb{N}}).$$

For other types of states, i.e. $(\mathcal{L}(s), \mathcal{L}(t), (x_{k(s',i), k(t',j)})_{i,j \in \mathbb{N}})$ with $s \neq s'$ or $t \neq t'$, we leave the transition function undefined as it is clear that such states are not reachable from the initial state q_0 .

We have to check that the transition function δ is well-defined. Assume that

$$(\mathcal{L}(s), \mathcal{L}(t), (x_{k(s,i), k(t,j)})_{i,j \in \mathbb{N}}) = (\mathcal{L}(s'), \mathcal{L}(t'), (x_{k(s',i), k(t',j)})_{i,j \in \mathbb{N}})$$

with $|s| = |t|$ and $|s'| = |t'|$. For all $(a, b) \in \Sigma_U \times \Sigma_U$, we need to show that

$$(\mathcal{L}(as), \mathcal{L}(bt), (x_{k(as,i), k(bt,j)})_{i,j \in \mathbb{N}}) = (\mathcal{L}(as'), \mathcal{L}(bt'), (x_{k(as',i), k(bt',j)})_{i,j \in \mathbb{N}}).$$

For the first two components, the result follows from the definition: indeed, $\mathcal{L}(as) = \mathcal{L}(s) \cdot a^{-1}$ for any letter a . For the third component, we want

to prove that $x_{k(as,i),k(bt,j)} = x_{k(as',i),k(bt',j)}$ for all $i, j \in \mathbb{N}$. We know that $\mathcal{L}(s) = \mathcal{L}(s')$, $\mathcal{L}(t) = \mathcal{L}(t')$ and $x_{k(s,i),k(t,j)} = x_{k(s',i),k(t',j)}$ for all $i, j \in \mathbb{N}$. Let us enumerate the words of $\mathcal{L}(s) \setminus 0\Sigma_U^*$ in genealogical order \prec :

$$\mathcal{L}(s) \setminus 0\Sigma_U^* = \{r_{s,0} \prec r_{s,1} \prec r_{s,2} \prec \dots\}.$$

Similarly, we write

$$\mathcal{L}(t) \setminus 0\Sigma_U^* = \{r_{t,0} \prec r_{t,1} \prec r_{t,2} \prec \dots\}.$$

Note that if s is a valid U -representation, then $r_{s,0} = \varepsilon$ and similarly for $r_{t,0}$. Let $m, n \in \mathbb{N}$. Since $r_{s,m}$ and $r_{s,n}$ do not start with a zero digit, we have

$$r_{s,m} \prec r_{s,n} \Leftrightarrow \text{val}_U(r_{s,m}0^{|s|}) < \text{val}_U(r_{s,n}0^{|s|}),$$

and an analogous equivalence holds for $r_{t,m}$ and $r_{t,n}$. The subsequence $(x_{k(s,i),k(t,j)})_{i,j \in \mathbb{N}}$ is the same as the sequence

$$(x_{\text{val}_U(r_{s,i}), \text{val}_U(r_{t,j})})_{i,j \in \mathbb{N}}$$

because by Definition 1.6.4, $k(s, i)$ (resp. $k(t, j)$) is the $(i+1)^{\text{th}}$ (resp. the $(j+1)^{\text{th}}$) integer belonging to $\mathcal{K}_s = \text{val}_U(0^* \text{rep}_U(\mathbb{N}) \cap \Sigma_U^* s)$ (resp. \mathcal{K}_t). Notice that words in $\mathcal{L}(s)$ (resp. $\mathcal{L}(t)$) starting with 0 do not provide any new indices. Thus when building the subsequence, we can limit ourselves to words not starting with 0. If we select in $\mathcal{L}(s) \setminus 0\Sigma_U^*$ all words ending with a , we get exactly $(\mathcal{L}(as) \setminus 0\Sigma_U^*)a$, which is equal to $(\mathcal{L}(as') \setminus 0\Sigma_U^*)a$ because $\mathcal{L}(as) = \mathcal{L}(as')$. Let $i \in \mathbb{N}$ and $r_{as,i}$ be the $(i+1)^{\text{th}}$ word in $\mathcal{L}(as) \setminus 0\Sigma_U^*$. Suppose that the $(i+1)^{\text{th}}$ word in $(\mathcal{L}(as) \setminus 0\Sigma_U^*)a$, which is $r_{as,i}a$, occurs as the $(m+1)^{\text{th}}$ word $r_{s,m}$ in $\mathcal{L}(s) \setminus 0\Sigma_U^*$. Then $r_{s,m}$ also occurs as the $(m+1)^{\text{th}}$ word $r_{s',m}$ in $\mathcal{L}(s') \setminus 0\Sigma_U^*$. With our notation, we have

$$r_{as,i}a = r_{s,m}, \quad \text{val}_U(r_{as,i}as) = \text{val}_U(r_{s,m}s), \quad \text{and} \quad k(as, i) = k(s, m) = k(s', m).$$

We can make similar observations for the other component. Supposing that $r_{bt,j} = r_{t,n}$ for some n , we thus have

$$x_{k(as,i),k(bt,j)} = x_{k(s,m),k(t,n)} = x_{k(s',m),k(t',n)} = x_{k(as',i),k(bt',j)},$$

where the central equality comes from our initial assumption. Therefore we have shown that δ is well-defined.

From our definition of the transition function δ , the accessible part of \mathcal{M} is limited to states q of the form

$$(\mathcal{L}(s), \mathcal{L}(t), (x_{k(s,i),k(t,j)})_{i,j \in \mathbb{N}})$$

with $|s| = |t|$. For such a state q , we set

$$\tau(q) = x_{k(s,0),k(t,0)}.$$

Notice that the preceding arguments show that τ is also well-defined. To conclude the proof, let us prove that if s, t are two words of the same length in $0^* \text{rep}_U(\mathbb{N})$, then

$$\tau(\delta(q_0, (s^R, t^R))) = x_{\text{val}_U(s), \text{val}_U(t)}.$$

(Recall that s^R and t^R respectively denote the reversals of the words s and t , cf. Chapter 1, Section 1.1.) Reading (s^R, t^R) from q_0 leads to the state $(\mathcal{L}(s), \mathcal{L}(t), (x_{k(s,i),k(t,j)})_{i,j \in \mathbb{N}})$. Since $s, t \in 0^* \text{rep}_U(\mathbb{N})$, we have that ε belongs to $\mathcal{L}(s)$ and $\mathcal{L}(t)$. It is clear that $k(s, 0) = \text{val}_U(s)$ and $k(t, 0) = \text{val}_U(t)$. We have thus proved that \mathbf{x} is reversal- U -automatic. It follows from Proposition 1.6.7 and Remark 1.6.8 that \mathbf{x} is U -automatic. \square

4.5 Open problem

In the first part of this chapter, we studied the factor complexity of automatic sequences. We showed first that Pisot-automatic sequences have a sublinear factor complexity, like b -automatic sequences do. We also proved that this property do not extend to Bertrand-automatic sequences, since we provided a Bertrand-automatic sequence with superlinear complexity. However, in view of Pansiot's theorem (Theorem 4.2.5), can we give a Bertrand-automatic sequences with factor complexities $\Theta(n \log \log n)$ or $\Theta(n \log n)$?

Chapter 5

Perspectives

Let us end this dissertation by a brief summary of what has been achieved and some potential future research questions.

In Chapter 2, we tackle the following problem:

Problem 5.0.1. Given a linear numeration system U and a deterministic finite automaton \mathcal{A} whose accepted language is contained in the numeration language $\text{rep}_U(\mathbb{N})$, decide whether the subset X of \mathbb{N} that is recognized by \mathcal{A} is ultimately periodic.

First solved by Honkala in 1986 for integer base systems [43], many authors gave different decision procedures for positional numeration systems under various hypotheses. In this vein, we required the following hypotheses on the positional numeration system $U = (U_i)_{i \in \mathbb{N}}$:

- (H1) the numeration language must be regular,
- (H2) there are arbitrary large gaps between consecutive terms of the sequence U ,
- (H3) the gap sequence is ultimately non-decreasing.

The first natural question arising is: how could we extend our decision procedure to other positional numeration systems? Otherwise stated, could we weaken our hypotheses?

Secondly, our strategy requires the quantity n_X defined in Section 2.5.2 to be larger than some positive constant Z . We saw in Section 2.7 that this condition is not satisfied for integer base systems and “disguised integer base”, i.e. sequences of the form $U_{i+k} = bU_i$ for $i \geq N$, with $b \geq 2, k \geq 1, N \geq 0$. Are they the only sequences not verifying $n_X \geq Z$? Related to this question,

could we imagine a strategy that would work for integer base systems and non-standard numeration systems as well?

Finally, what can be said about the time complexity of our procedure? The strategy of looking at all the possible admissible pairs of preperiod/period is probably far from being optimal. Except for integer base systems [54, 13], not much work has been done in that direction.

The third chapter is the beginning of the study of the following decision problem:

Problem 5.0.2. Given a recognizable set of positive integers via an automaton recognizing it, can we decide whether this set is of the form $mX + r$ for some set of integers X ?

Our strategy has similarity with the one of Chapter 2. Indeed, in Chapter 2, we give bounds on the state complexity of ultimately periodic sets. In Chapter 3, we want to provide the minimal automaton of $\text{val}_b^{-1}(mX + r)$ and thus the exact state complexity.

The case $X = \mathbb{N}$ and $r = 0$ was first studied by Alexeev in [1]. In this dissertation, we provide the minimal automaton recognizing in a base which is a power of 2 sets of the form $m\mathcal{T} + r$, where \mathcal{T} is the set of positive integers whose base-2 expansions contain an even number of occurrences of the digit 1. Can we extend our results? Can we replace \mathcal{T} by any b -recognizable set and provide the associated state complexity? We conjecture that for sets of the form

$$X_{b,c,M,r} = \{n \in \mathbb{N} : |\text{rep}_b(n)|_c \equiv R \pmod{M}\}$$

where b is an integer base, c is any digit in $\llbracket 0, b-1 \rrbracket$, M is an integer greater than or equal to 2 and R is any possible remainder in $\llbracket 0, M-1 \rrbracket$, when $b = q^p$ for some prime q , then the state complexity of $mX_{b,c,M,R} + r$ with respect to the base b is $Mk + \lceil \frac{z}{p} \rceil$, where $m = kq^z$ with $\gcd(k, q) = 1$.

Keeping Chapter 2 in mind, could we give the state complexity (or a bound) for U -expansions of sets of the form $mX + r$, where U is a positional numeration system? Some work in this direction has already been done in [29] for $r = 0$ and $X = \mathbb{N}$.

Finally, in Chapter 4, we studied U -automatic sequences and their properties. We looked at factor complexity, closure properties and interested ourselves to multidimensional sequences and their U -kernel.

We showed that Parry-automatic sequences have, like Pisot-automatic sequences, a sublinear factor complexity. We also exhibit a Bertrand-automatic sequence with superlinear complexity. Having Pansiot's theorem in the back

of your mind, the following question is natural: are there Bertrand-automatic sequences with factor complexities $\Theta(n \log \log n)$ or $\Theta(n \log n)$?

Appendices

Appendix A

Examples of Parry numeration systems

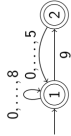
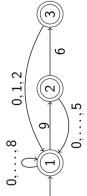
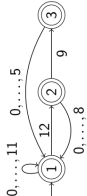
This appendix is made of examples of Parry numeration systems. We give several arrays organised as follows. The first kind of table consists in recurrences, initial conditions, the associated characteristic polynomial, the dominant root β , the β -expansion of 1, the quasi-greedy β -expansion of 1 and finally the automaton \mathcal{A}_β associated to the numeration system. Among all these systems, only one is not a Pisot numeration system. The associated β is coloured in green, it is the system from Example 2.1.2. Note that other Parry non-Pisot numeration systems are given in the main text of the present dissertation, as in the proof of Lemma 1.3.29.

The second kind of table gets a look back at the same numeration systems than in the first table, but this time we interest ourselves to the sequences $(U_i \bmod m)_{i \in \mathbb{N}}$ for well-chosen multiples m . With Chapter 2 in mind, we give, when available, the state complexity and the syntactic complexity of $m\mathbb{N}$ in base U .

Recurrence	Initial conditions	Characteristic polynomial	Dominant root β	$d_\beta(1)$	$d_\beta^*(1)$	\mathcal{A}_β
$U_{i+3} = 2U_{i+2} + U_i$	1, 3, 7	$x^3 - 2x^2 - 1$	≈ 2.20557	2010^ω	$(200)^\omega$	
$U_{i+2} = U_{i+1} + U_i$	1, 2	$x^2 - x - 1$	$\frac{1 + \sqrt{5}}{2}$	110^ω	$(10)^\omega$	
$U_{i+4} = U_{i+3} + U_{i+2} + U_{i+1} + U_i$	1, 2, 4, 8	$x^4 - x^3 - x^2 - x - 1$	≈ 1.92756	11110^ω	$(1110)^\omega$	
$U_{i+2} = 2U_{i+1} + 2U_i$	1, 3	$x^2 - 2x - 2$	$1 + \sqrt{3}$	220^ω	$(21)^\omega$	
$U_{i+2} = 6U_{i+1} + 3U_i$	1, 7	$x^2 - 6x - 3$	$3 + 2\sqrt{3}$	630^ω	$(62)^\omega$	
$U_{i+2} = 6U_{i+1} - 4U_i$	1, 6	$x^2 - 6x + 4$	$3 + \sqrt{5}$	511111110^ω	$(51111110)^\omega$	

Recurrence	Initial conditions	Characteristic polynomial	Dominant root β	$d_\beta(1)$	$d_\beta^*(1)$	\mathcal{A}_β
$U_{i+2} = 10U_{i+1} - 5U_i$	1, 10	$x^2 - 10x + 5$	$5 + 2\sqrt{5}$	9444450^ω	$(944444)^\omega$	
$U_{i+2} = 7U_{i+1} + 2U_i$	1, 8	$x^2 - 7x - 2$	$\frac{7 + \sqrt{57}}{2}$	720^ω	$(71)^\omega$	
$U_{i+2} = 7U_{i+1} - 2U_i$	1, 7	$x^2 - 7x + 2$	$\frac{7 + \sqrt{41}}{2}$	6444450^ω	$(644444)^\omega$	
$U_{i+4} = 2U_{i+3} + 2U_{i+2} + 2U_i$	1, 3, 9, 25	$x^4 - 2x^3 - 2x^2 - 2$	≈ 2.80399	22020^ω	$(2201)^\omega$	
$U_{i+2} = 10U_{i+1} + 5U_i$	1, 11	$x^2 - 10x - 5$	$5 + \sqrt{30}$	$10\ 5\ 0^\omega$	$(10\ 4)^\omega$	

Recurrence	Initial conditions	Characteristic polynomial	Dominant root β	$d_\beta(1)$	$d_\beta^*(1)$	\mathcal{A}_β
$U_{i+2} = 10U_{i+1} + 4U_i$	1, 11	$x^2 - 10x - 4$	$5 + \sqrt{29}$	$10\ 4\ 0^\omega$	$(10\ 3)^\omega$	
$U_{i+2} = 5U_{i+1} + 2U_i$	1, 6	$x^2 - 5x - 2$	$\frac{5 + \sqrt{33}}{2}$	520^ω	$(51)^\omega$	
$U_{i+2} = 6U_{i+1} - 3U_i$	1, 6	$x^2 - 6x + 3$	$3 + \sqrt{6}$	52^ω	52^ω	
$U_{i+2} = 5U_{i+1} + 3U_i$	1, 6	$x^2 - 5x - 3$	$\frac{5 + \sqrt{37}}{2}$	530^ω	$(52)^\omega$	
$U_{i+2} = 12U_{i+1} + 3U_i$	1, 13	$x^2 - 12x - 3$	$6 + \sqrt{39}$	$12\ 3\ 0^\omega$	$(12\ 2)^\omega$	
$U_{i+2} = 9U_{i+1} + 3U_i$	1, 10	$x^2 - 9x - 3$	$\frac{9 + \sqrt{93}}{2}$	930^ω	$(92)^\omega$	

Recurrence	Initial conditions	Characteristic polynomial	Dominant root β	$d_\beta(1)$	$d_\beta^*(1)$	\mathcal{A}_β
$U_{i+2} = 9U_{i+1} + 6U_i$	1, 10	$x^2 - 9x - 6$	$\frac{9 + \sqrt{105}}{2}$	960^ω	$(95)^\omega$	
$U_{i+3} = 9U_{i+2} + 6U_{i+1} + 3U_i$	1, 10, 97	$x^3 - 9x^2 - 6x - 3$	≈ 9.65371	9630^ω	$(962)^\omega$	
$U_{i+3} = 12U_{i+2} + 9U_{i+1} + 6U_i$	1, 13, 166	$x^3 - 12x^2 - 9x - 6$	≈ 12.7432	$12\ 9\ 6\ 0^\omega$	$(12\ 9\ 5)^\omega$	

Recurrence	$(U_i)_{i \in \mathbb{N}}$	m	$(U_i \pmod{m})_{i \in \mathbb{N}}$	$\#\mathcal{A}$	$\#\mathcal{M}$
$U_{i+3} = 2U_{i+2} + U_i$	1, 3, 7, 15, 33, 73, 161, 355, ...	2	1^ω	2	21
		3	$(1010012102211)^\omega$	351	3162
		4	$(133311)^\omega$	96	867
		5	$(1320331032414204004311013134411)^\omega$	3875	34878
$U_{i+2} = U_{i+1} + U_i$	1, 2, 3, 5, 8, 13, 21, 34, ...	2	$(101)^\omega$	12	50
		3	$(12022101)^\omega$	72	290
		4	$(123101)^\omega$	96	386
		5	$(12303314044320224101)^\omega$	500	2002
$U_{i+4} = U_{i+3} + U_{i+2} + U_{i+1} + U_i$	1, 2, 4, 8, 15, 29, 56, 108, ...	2	$(10001)^\omega$	80	1284
$U_{i+2} = 2U_{i+1} + 2U_i$	1, 3, 8, 22, 60, 164, 448, 1224, ...	6	$13(240)^\omega$	112	329
		12	$1\ 3\ 8\ 10\ (0\ 8\ 4)^\omega$	350	996
		24	$1\ 3\ 8\ 22\ 12\ 20\ (16\ 0\ 8)^\omega$	1093	2821
$U_{i+2} = 6U_{i+1} + 3U_i$	1, 7, 45, 291, 1881, 12159, ...	12	$17(93)^\omega$	131	131
		15	$17(06691290121239309963603312612)^\omega$	3605	3605
$U_{i+2} = 6U_{i+1} - 4U_i$	1, 6, 32, 168, 880, 4608, ...	6	$1(0204)^\omega$	74	74
		2	10^ω	4	4
		8	160^ω	9	9
		12	$16(8040)^\omega$	112	112

Recurrence	$(U_i)_{i \in \mathbb{N}}$	m	$(U_i \pmod{m})_{i \in \mathbb{N}}$
$U_{i+2} = 10U_{i+1} - 5U_i$	1, 10, 95, 900, 8525, ...	10	$1(05)^\omega$
		15	1 (10 5 0 5 5 10 10 0 10) $^\omega$
		20	1 (10 15 0 5) $^\omega$
$U_{i+2} = 7U_{i+1} + 2U_i$	1, 8, 58, 422, 3070, ...	4	102^ω
		6	$1(24)^\omega$
		8	1026^ω
		16	1 8 10 6 (14) $^\omega$
$U_{i+2} = 7U_{i+1} - 2U_i$	1, 7, 47, 315, 2111, ...	4	13^ω
		8	$17(73)^\omega$
		16	1 7 15 (11 15 3 7) $^\omega$
		32	1 7 15 27 (31 3 23 27 15 19 7 11) $^\omega$
$U_{i+4} = 2U_{i+3} + 2U_{i+2} + 2U_i$	1, 3, 9, 25, 70, 196, ...	6	$1331(4440440200042044200222202204000240224004)^\omega$
		10	$1395(06024464822620886840)^\omega$
		12	$1391104106(4408000480448008888088040008408840044440)^\omega$
$U_{i+2} = 10U_{i+1} + 5U_i$	1, 11, 115, 1205, ...	10	115^ω
		15	1 11 (10 5) $^\omega$
		20	1 11 (15 5 15) $^\omega$
$U_{i+2} = 10U_{i+1} + 4U_i$	1, 11, 114, 1184, ...	6	$15(02240442)^\omega$
		10	$11(4466)^\omega$
		12	1 11 6 (88404480) $^\omega$
		18	$111(614241210486281612416146814101216102)^\omega$
$U_{i+2} = 5U_{i+1} + 2U_i$	1, 6, 32, 172, 924, ...	6	$1(024)^\omega$
		8	1604^ω
		10	$1(62244886)^\omega$
		16	1 6 0 12 (12 4) $^\omega$

Recurrence	$(U_i)_{i \in \mathbb{N}}$	m	$(U_i \pmod{m})_{i \in \mathbb{N}}$
$U_{i+2} = 6U_{i+1} - 3U_i$	1, 6, 33, 180, 981, ...	6 12 15	$1(03)^\omega$ $1(690)^\omega$ $1(6306)^\omega$
$U_{i+2} = 5U_{i+1} + 3U_i$	1, 6, 33, 183, 1014, ...	6 9 12	$1(033)^\omega$ $16(63)^\omega$ $1(693639)^\omega$
$U_{i+2} = 12U_{i+1} + 3U_i$	1, 13, 159, 1947, ...	6 12 15 24	113^ω $11(3399)^\omega$ $1\ 13\ (9\ 12\ 6\ 3)^\omega$ $1\ 13\ (15\ 3\ 9\ 21)^\omega$
$U_{i+2} = 9U_{i+1} + 3U_i$	1, 10, 93, 867, 8082, ...	6 12 15	$14(330)^\omega$ $1\ 10\ (9\ 3\ 6\ 3\ 9\ 6)^\omega$ $1\ 10\ (3\ 12\ 12\ 9\ 12\ 0\ 6\ 9\ 3\ 9\ 0\ 12\ 3\ 3\ 6\ 3\ 0\ 9\ 6\ 6\ 12\ 6\ 0)^\omega$
$U_{i+2} = 9U_{i+1} + 6U_i$	1, 10, 96, 924, 8892, ...	15 21 24	$1\ 10\ (6\ 9\ 12\ 12\ 0\ 12\ 3\ 9\ 0\ 9\ 6\ 3\ 3\ 0\ 3\ 12\ 6\ 6\ 0)^\omega$ $1\ 10\ (12\ 0\ 9\ 18\ 6\ 15\ 3)^\omega$ $1\ 10\ 0\ (12)^\omega$
$U_{i+3} = 9U_{i+2} + 6U_{i+1} + 3U_i$	1, 10, 97, 936, 9036, ...	6 12 15	$141(0033303)^\omega$ $1\ 10\ 1\ (0\ 0\ 3\ 3\ 9\ 0\ 3\ 6\ 0\ 9\ 3\ 9\ 6\ 9)^\omega$ $1\ 10\ 7\ (6\ 6\ 6\ 3\ 6\ 0\ 0\ 3\ 12\ 6\ 0\ 12)^\omega$
$U_{i+3} = 12U_{i+2} + 9U_{i+1} + 6U_i$	1, 13, 166, 2115, ...	12 15 18	$1\ 1\ 10\ (3\ 0\ 3\ 6)^\omega$ $1\ 13\ 1\ (0\ 12\ 0\ 3\ 3\ 6\ 12\ 6\ 6\ 3\ 6)^\omega$ $1\ 13\ 4\ 9\ 6\ 15\ (0\ 9)^\omega$

Appendix B

Computations for Section 4.3

In Chapter 4, several proofs use numerical approximations. Indeed, in Section 4.3 we study the numeration system based on the linear recurrent sequence $(U_i)_{i \in \mathbb{N}}$ defined by

$$U_{i+4} = 3U_{i+3} + 2U_{i+2} + 3U_i \quad \forall i \in \mathbb{N}$$

with initial values $U_0 = 1, U_1 = 4, U_2 = 15$ and $U_3 = 54$. This system U is interesting because the class of U -automatic sequences is not closed under taking image by a uniform morphism (see Theorem 4.3.2). The proof of this latter result needs some previous work, done with processing some computations. This appendix is devoted to provide the `Mathematica` code used for our approximations, in particular in Propositions 4.3.4 and 4.3.7.

About Section 4.3: Closure properties

We consider the numeration system
based on the following recurrence:

```
In[1]:= U[i_] := U[i] = 3 U[i - 1] + 2 U[i - 2] + 3 U[i - 4];
        U[0] = 1;
        U[1] = 4;
        U[2] = 15;
        U[3] = 54;
```

```
In[2]:= Table[U[i], {i, 0, 20}]
        table
```

```
Out[2]= {1, 4, 15, 54, 195, 705, 2550, 9222, 33 351, 120 612, 436 188,
        1 577 454, 5 704 791, 20 631 117, 74 611 497, 269 829 087, 975 824 628,
        3 529 025 409, 12 762 559 974, 46 155 218 001, 166 918 247 835}
```

It's characteristic polynomial has four roots:

```
In[3]:= Solve[X^4 - 3 X^3 - 2 X^2 - 3 == 0, X]
        résous
```

```
Out[3]= {{X -> -1.10...}, {X -> 3.62...},
        {X -> 0.240... - 0.836... i}, {X -> 0.240... + 0.836... i}}
```

The two complex roots have a modulus less than 1:

```
In[4]:= Abs[X /. N[{Solve[X^4 - 3 X^3 - 2 X^2 - 3 == 0, X] [[3]],
        valeur a... [v... résous
        Solve[X^4 - 3 X^3 - 2 X^2 - 3 == 0, X] [[4]]}]]]
        résous
```

```
Out[4]= {0.869653, 0.869653}
```

There are also two real roots:

$$\begin{aligned} \text{In}[5]:= \rho = & \frac{3}{4} + \frac{1}{4 \sqrt[4]{\frac{3}{43-128 \left(\frac{2}{-1177+9 \sqrt{18721}} \right)^{1/3} + 2 \cdot 2^{2/3} \left(-1177+9 \sqrt{18721} \right)^{1/3}}}}} + \\ & \frac{1}{2} \sqrt[3]{\left(\frac{43}{6} + \frac{32}{3} \left(\frac{2}{-1177+9 \sqrt{18721}} \right)^{1/3} - \right.} \\ & \quad \left. \frac{1}{3} \left(\frac{1}{2} \left(-1177+9 \sqrt{18721} \right) \right)^{1/3} + \right.} \\ & \quad \left. \frac{51}{2} \sqrt[3]{\left(3 / \left(43-128 \left(\frac{2}{-1177+9 \sqrt{18721}} \right)^{1/3} + \right. \right.} \right. \\ & \quad \left. \left. \left. 2 \times 2^{2/3} \left(-1177+9 \sqrt{18721} \right)^{1/3} \right) \right) \right) \right) \Bigg); \end{aligned}$$

In[6]:= **N**[ρ]
[valeur numérique]

Out[6]= 3.61645

$$\begin{aligned} \text{In}[7]:= \gamma = & \frac{3}{4} + \frac{1}{4 \sqrt[4]{\frac{3}{43-128 \left(\frac{2}{-1177+9 \sqrt{18721}} \right)^{1/3} + 2 \cdot 2^{2/3} \left(-1177+9 \sqrt{18721} \right)^{1/3}}}}} - \\ & \frac{1}{2} \sqrt[3]{\left(\frac{43}{6} + \frac{32}{3} \left(\frac{2}{-1177+9 \sqrt{18721}} \right)^{1/3} - \right.} \\ & \quad \left. \frac{1}{3} \left(\frac{1}{2} \left(-1177+9 \sqrt{18721} \right) \right)^{1/3} + \right.} \\ & \quad \left. \frac{51}{2} \sqrt[3]{\left(3 / \left(43-128 \left(\frac{2}{-1177+9 \sqrt{18721}} \right)^{1/3} + \right. \right.} \right. \\ & \quad \left. \left. \left. 2 \times 2^{2/3} \left(-1177+9 \sqrt{18721} \right)^{1/3} \right) \right) \right) \right) \Bigg); \end{aligned}$$

In[8]:= **N**[γ]
[valeur numérique]

Out[8]= -1.09685

The polynomial of the recurrence is indeed
the minimal polynomial of ρ :

```
In[9]:= MinimalPolynomial[ρ]
      polynôme minimal
Out[9]= -3 - 2 ρ12 - 3 ρ13 + ρ14 &
```

■ Computations for Proposition 4.3.4

About equation (4.3):

```
In[10]:= N[ (3 ρ-2) / (1 - ρ-2) ]
      valeur numérique
Out[10]= 14.773
```

Let us compute the square of γ :

```
In[11]:= N[ρ2]
      valeur numérique
Out[11]= 1.20307
```

Now, we look at equation (4.5) for $r=4$. In this case, $k=1$.

```
In[12]:= r = 4; k = 1;
In[13]:= N[ρr (r - 1)]
      valeur numérique
Out[13]= 47.2987
In[14]:= N[3 (1 - ρ(2k+2)) / (1 - ρ2) + 15]
      valeur numérique
Out[14]= 21.6092
```

Consier now the case $r=3$ (hence $k=1$).

```
In[15]:= Clear[r, k];
      efface
In[16]:= r = 3; k = 1;
```

One the one hand, one has

```
In[17]:= N[3 ρ(-4) / (1 - ρ(-2))]
      valeur numérique
Out[17]= 12.2794
```

Moreover,


```
In[18]:= {Ceiling[ $\rho^2$ ], Floor[ $\rho^3$ ]}
```

entier supérieur entier inférieur

```
Out[18]= {14, 47}
```

thus t can take any integer value between 14 and 47.

For these values of t , we want to minimize $t-S_{\{1,3\}}$.

In this aim, we must compute the first values of $d_{\rho}(t/\rho^3)$

(which can be found via a classical algorithm, cf. Example 1.3.7).

```
In[19]:= Table[x = t/ $\rho^3$ ; list = {}; i = 1;
table
{t, Flatten[{While[i < 4,
aplatis    tant que
AppendTo[list, Floor[ $\rho x$ ]];
appose à    entier inférieur
x =  $\rho x$  - Floor[ $\rho x$ ];
entier inférieur
i++]; list]},
N[t - {list}.Reverse[Table[ $\gamma^i$ , {i, 0, 2}]], 10]
valeur numéri...    renverses    table
},
{t, 14, 47}]
```

```
Out[19]= {{14, {1, 0, 0}, {12.79692683}}, {15, {1, 0, 1}, {12.79692683}},
{16, {1, 0, 2}, {12.79692683}}, {17, {1, 1, 0}, {16.89377375}},
{18, {1, 1, 1}, {16.89377375}}, {19, {1, 1, 2}, {16.89377375}},
{20, {1, 1, 3}, {16.89377375}}, {21, {1, 2, 0}, {21.99062067}},
{22, {1, 2, 1}, {21.99062067}}, {23, {1, 2, 2}, {21.99062067}},
{24, {1, 3, 0}, {26.08746759}}, {25, {1, 3, 1}, {26.08746759}},
{26, {1, 3, 2}, {26.08746759}}, {27, {2, 0, 0}, {24.59385367}},
{28, {2, 0, 1}, {24.59385367}}, {29, {2, 0, 2}, {24.59385367}},
{30, {2, 1, 0}, {28.69070059}}, {31, {2, 1, 1}, {28.69070059}},
{32, {2, 1, 2}, {28.69070059}}, {33, {2, 1, 3}, {28.69070059}},
{34, {2, 2, 0}, {33.78754751}}, {35, {2, 2, 1}, {33.78754751}},
{36, {2, 2, 2}, {33.78754751}}, {37, {2, 2, 3}, {33.78754751}},
{38, {2, 3, 0}, {38.88439443}}, {39, {2, 3, 1}, {38.88439443}},
{40, {3, 0, 0}, {36.39078050}}, {41, {3, 0, 1}, {36.39078050}},
{42, {3, 0, 2}, {36.39078050}}, {43, {3, 1, 0}, {40.48762742}},
{44, {3, 1, 1}, {40.48762742}}, {45, {3, 1, 2}, {40.48762742}},
{46, {3, 1, 3}, {40.48762742}}, {47, {3, 2, 0}, {45.58447434}}}
```

```
In[20]:= m = Flatten[Map[# [3] &, %]]
          [aplatis] [applique]

Out[20]:= {12.79692683, 12.79692683, 12.79692683, 16.89377375, 16.89377375,
          16.89377375, 16.89377375, 21.99062067, 21.99062067, 21.99062067,
          26.08746759, 26.08746759, 26.08746759, 24.59385367, 24.59385367,
          24.59385367, 28.69070059, 28.69070059, 28.69070059, 28.69070059,
          33.78754751, 33.78754751, 33.78754751, 33.78754751, 38.88439443,
          38.88439443, 36.39078050, 36.39078050, 36.39078050, 40.48762742,
          40.48762742, 40.48762742, 45.58447434}
```

For the sake of clarity, we represent the graph of the function $f(t)=t-S_{\{1,3\}}$ (for t between 14 and 47).

```
In[21]:= Show[ListPlot[Table[{i + 13, m[[i]]}, {i, 1, Length[m]}]],
              [mon· [tracé de li· [table] [longueur]
              AxesLabel → {"t", "f(t)"}, PlotLabel → None,
              [titre d'axe] [titre de tracé] [aucun]
              LabelStyle → {GrayLevel[0]}]
              [style d'étiquette] [niveau de gris]

Out[21]:=
```

This conclude the case $r=3$. If $r=2$, then $k=0$.

```
In[22]:= Clear[r, k, m, list, x];
          [efface]
```

```
In[23]:= r = 2; k = 0;
```

Then one has

```

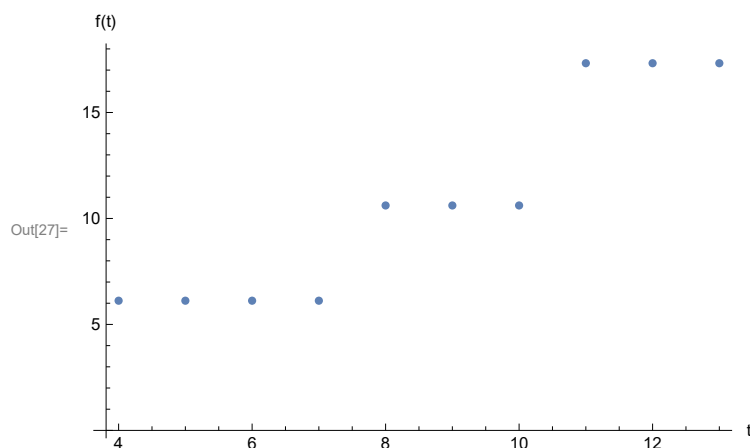
In[24]:= {Ceiling[ρ], Floor[ρ^2]}
          |entier supérieur |entier inférieur
Out[24]= {4, 13}

In[25]:= Table[x = t/ρ^2; list = {}; i = 1;
          |table
          {t, Flatten[{While[i < 13,
          |aplatis |tant que
          AppendTo[list, Floor[ρ x]];
          |appose à |entier inférieur
          x = ρ x - Floor[ρ x];
          |entier inférieur
          i++; list}],
          N[t - {list}.Reverse[Table[γ^i, {i, -10, 1}]], 10]
          |valeur numéri... |renverses |table
          },
          {t, 4, 13}]
Out[25]= {{4, {1, 0, 1, 1, 1, 1, 2, 0, 2, 3, 0, 0}}, {6.119167485}},
          {5, {1, 1, 1, 1, 1, 1, 2, 0, 2, 3, 0, 0}}, {6.119167485}},
          {6, {1, 2, 1, 1, 1, 1, 2, 0, 2, 3, 0, 0}}, {6.119167485}},
          {7, {1, 3, 1, 1, 1, 1, 2, 0, 2, 3, 0, 0}}, {6.119167485}},
          {8, {2, 0, 2, 2, 2, 3, 0, 2, 3, 1, 2, 0}}, {10.61282538}},
          {9, {2, 1, 2, 2, 2, 3, 0, 2, 3, 1, 2, 0}}, {10.61282538}},
          {10, {2, 2, 2, 2, 2, 3, 0, 2, 3, 1, 2, 0}}, {10.61282538}},
          {11, {3, 0, 0, 1, 3, 1, 3, 0, 0, 1, 2, 0}}, {17.32461539}},
          {12, {3, 1, 0, 1, 3, 1, 3, 0, 0, 1, 2, 0}}, {17.32461539}},
          {13, {3, 2, 0, 1, 3, 1, 3, 0, 0, 1, 2, 0}}, {17.32461539}}}

In[26]:= m = Flatten[Map[#[[3]] &, %]]
          |aplatis |applique
Out[26]= {6.119167485, 6.119167485, 6.119167485, 6.119167485, 10.61282538,
          10.61282538, 10.61282538, 17.32461539, 17.32461539, 17.32461539}

```

```
In[27]:= Show[ListPlot[Table[{i + 3, m[[i]]}, {i, 1, Length[m]}]],
  mon... tracé de li... table longueur
  AxesLabel -> {"t", "f(t)"}, PlotLabel -> None,
  titre d'axe titre de tracé aucun
  LabelStyle -> {GrayLevel[0]}]
  style d'étiquette niveau de gris
```



On the other hand,

```
In[28]:= N[3 * 2^(-12) / (1 - 2^(-2))]
  valeur numérique
```

```
Out[28]:= 5.8615
```

■ Computations for Proposition 4.3.7

```
In[29]:= Clear[x, list, r, k, m];
  efface
```

The first 21 digits of $d_{\rho}(1/2)$ can be computed like previously:

```
In[30]:= list = {}; x = 1/2; i = 1;
```

```
Flatten[{While[i < 22,
  aplatis tant que
  AppendTo[list, Floor[ρ x]];
  appose à entier inférieur
  x = ρ x - Floor[ρ x];
  entier inférieur
  i++];
list}]
```

```
Out[31]:= {1, 2, 3, 1, 0, 2, 3, 0, 3, 0, 0, 1, 0, 1, 0, 2, 2, 0, 1, 2, 3}
```

We are now able to compute $S_{\{1,21\}}$:

```
In[32]:= N[{list}.Table[ $\gamma^i$ , {i, 1, 21}], 10]
          [valeur nu... [table
Out[32]:= {-2.203818911}
```

And finally,

```
In[33]:= N[ $3 \gamma^{-22} / (1 - \gamma^{-2})$ ],
          [valeur numérique
Out[33]:= 2.32567
```


Bibliography

- [1] B. Alexeev, *Minimal DFA for testing divisibility*, J. Comput. System Sci. **69** (2004), 2, 235–243.
- [2] J.-P. Allouche, *Thue, combinatorics on words and conjectures inspired by the Thue-Morse sequence*, J. Théor. Nombres Bordeaux **27** (2015), 375–388.
- [3] J.-P. Allouche, B. Cloitre, V. Shevelev, *Beyond odious and evil*, Aequationes Math. **90** (2016), 2, 341–353.
- [4] J.-P. Allouche, N. Rampersad, J. Shallit, *Periodicity, repetitions and orbits of an automatic sequence*, Theoret. Comput. Sci. **410** (2009), 30–32, 2795–2803.
- [5] J.-P. Allouche, J. Shallit, *The ubiquitous Prouhet-Thue-Morse sequence*, Sequences and their Applications (Singapore, 1998), Springer, London, 1999, 1–16.
- [6] J.-P. Allouche, J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge University Press, Cambridge, 2003.
- [7] J. Bell, É. Charlier, A. Fraenkel, M. Rigo, *A decision problem for ultimately periodic sets in nonstandard numeration system*, Internat. J. Algebra Comput. **19** (2009), 6, 809–839.
- [8] J. Berstel, Ch. Reutenauer, *Rational Series and Their Languages*, Electronic edition, 2006.
- [9] J. Berstel, Ch. Reutenauer, *Another proof of Soittola’s theorem*, Theoret. Comput. Sci. **393** (2007), 196–203.
- [10] A. Bertrand, *Développements en base Pisot et répartition modulo 1*, C.R. R. Acad. Sci. Paris **285** (1977), Sér. A–B, A419–A421.

- [11] A. Bertrand-Mathis, *Comment écrire les nombres entiers dans une base qui n'est pas entière*, Acta Math. Hungar. **54** (1989), 237–241.
- [12] V. Berthé, M. Rigo, eds, *Combinatorics, Automata and Number Theory*, Encyclopedia of Mathematics and its Applications, vol. 135, Cambridge University Press, Cambridge, 2010.
- [13] B. Boigelot, I. Mainz, V. Marsault, M. Rigo, *An efficient algorithm to decide periodicity of b -recognizable sets using MSDF convention*, 44th International Colloquium on Automata, Languages and Programming (Warsaw, July 10–14, 2017)(Ioannis Chatzigiannakis and Piotr Indyk and Fabian Kuhn and Anca Muscholl), Leibniz International Proceedings in Informatics, 80, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, 2017, Paper No. 118, 14.
- [14] B. Boigelot, S. Rassart, P. Wolper, *On the expressiveness of real and integer arithmetic automata (extended abstract)*, ICALP (Aalborg, Denmark, July 13–17, 1998), Lecture Notes in Comput. Sci., 1443, Springer, Berlin, 1998, 152–163.
- [15] V. Bruyère, G. Hansel, *Bertrand numeration systems and recognizability*, Theoret. Comput. Sci. **181** (1997), 17–43.
- [16] V. Bruyère, G. Hansel, C. Michaux and R. Villemaire, *Logic and p -recognizable sets of integers*, Bull. Belg. Math. Soc. Simon Stevin **1** (1994), 191–238.
- [17] M. Bucci, N. Hindman, S. Puzynina, L. Q. Zamboni, *On additive properties of sets defined by the Thue-Morse word*, J. Combin. Theory Ser. A **120** (2013), 6, 1235–1245.
- [18] J. R. Büchi, *Weak second-order arithmetic and finite automata*, Z. Math. Logik Grundlag. Math. **6** (1960), 66–92.
- [19] Y. Bugeaud, G. Kekeç, *On Mahler's classification of p -adic numbers*, Bulletin of the Australian Mathematical Society **98** (2018), 203–211.
- [20] É. Charlier, *First-order logic and numeration systems*, Sequences, groups and number theory, Trends Maths., Birkhäuser/Springer, Cham, 2018.
- [21] É. Charlier, *Abstract Numeration Systems: Recognizability, Decidability, Multidimensional S -Automatic Words, and Real Numbers*, PhD thesis, University of Liège, Belgium, 2009.

- [22] É. Charlier, C. Cisternino, A. Massuir, *State complexity of the multiples of the Thue-Morse set*, Tenth International Symposium on Games, Automata, Logics and Formal Verification (Bordeaux, September 2–3, 2019), Electronic Proceedings in Theoretical Computer Science, 305, Open Publishing Association, 2019, 34–49.
- [23] É. Charlier, C. Cisternino, A. Massuir, *State complexity of the multiples of the Thue-Morse set*, 2019, long version, arXiv:1903.06114.
- [24] É. Charlier, C. Cisternino, A. Massuir, *Minimal automaton for multiplying and translating the Thue-Morse set*, 2019, to appear in Electronic Journal of Combinatorics, arXiv:1910.08543.
- [25] É. Charlier, C. Cisternino, M. Stipulanti, *Regular sequences and synchronized sequences in abstract numeration systems*, 2020, arXiv:2012.04969.
- [26] É. Charlier, J. Leroy, M. Rigo, *An analogue of Cobham’s theorem for graph directed iterated function systems*, Adv. Math. **280** (2015), 86–120.
- [27] É. Charlier, A. Massuir, E. Rowland, M. Rigo, *Ultimate periodicity problem for linear numeration systems*, 2020, submitted, arXiv:2007.08147.
- [28] É. Charlier, N. Rampersad, *The growth function of S -recognizable sets*, Theoret. Comput. Sci. **412** (2011), 39, 5400–5408.
- [29] É. Charlier, N. Rampersad, M. Rigo, L. Waxweiler, *The minimal automaton recognizing $m\mathbb{N}$ in a linear numeration system*, Integers **11B** (2011), A4, 1–24.
- [30] N. Chomsky, *On Certain Formal Properties of Grammars*, Information and Control **2** (1959), 137–167.
- [31] A. Cobham, *On the base-dependence of sets of numbers recognizable by finite automata*, Math. Systems Theory **3** (1969), 186–192.
- [32] A. Cobham, *Uniform tag sequences*, Math. Systems Theory **6** (1972).
- [33] F. Durand, *Decidability of the HD0L ultimate periodicity problem*, RAIRO Theor. Inform. Appl. **47** (2013), 2, 201–214.
- [34] F. Durand, M. Rigo, *On Cobham’s theorem*, Handbook of Automata, to appear.

- [35] S. Eilenberg, *Automata, languages and machines*, Vol. A, Pure and Applied Mathematics, 58, Academic Press, New York, 1974.
- [36] S. Fabre, *Substitutions et β -systèmes de numération*, Theoret. Comput. Sci. **137** (1995), 219–236.
- [37] Ch. Frougny, *Representations of numbers and finite automata*, Math. Systems Theory **25** (1992), 37–60.
- [38] Ch. Frougny, *On the sequentiality of the successor function*, Inform. and Comput. **139.1** (1997), 17–38.
- [39] Ch. Frougny, B. Solomyak, *On the representation of integers in linear numeration systems*, Ergodic Theory of \mathbb{Z}^d -Actions (Warwick, 1993–1994), London Math. Soc. Lecture Note Ser., 228, Cambridge University Press, Cambridge, 1996, 345–368.
- [40] F. Q. Gouvêa, *p -adic Numbers : An Introduction*, second edition, Springer, Berlin, 1997.
- [41] T. Harju, M. Linna, *On the periodicity of morphisms on free monoids*, RAIRO Theor. Inform. Appl. **20** (1986), 1, 47–54.
- [42] M. Hollander, *Greedy Numeration Systems and Regularity*, Theory Comput. Syst. **31.2** (1998), 111–133.
- [43] J. Honkala, *A decision method for the recognizability of sets defined by number systems*, Theoretic. Inform. Appl. **20** (1986), 395–403.
- [44] J. Honkala, M. Rigo, *Decidability questions related to abstract numeration systems*, Discrete Math. **285** (2004), 1–3, 329–333.
- [45] J. Hopcroft, *An $n \log n$ algorithm for minimizing states in a finite automaton*, International Symposium on the Theory of Machines and Computations (Haifa, Israel, August 16–19, 1971), Theory of Machines and Computations, Academic Press, New York, 1971, 189–196.
- [46] J. Hopcroft, R. Karp, *A linear algorithm for testing equivalence of finite automata*, Technical report, University of California, 1971, 71–114.
- [47] A. Lacroix, N. Rampersad, M. Rigo, É. Vandomme, *Syntactic complexity of ultimately periodic sets of integers and application to a decision procedure*, Fund. Infor. **116** (2012), 175–187.

- [48] P. Lecomte, M. Rigo, *Numeration systems on a regular language*, Theory Comput. Syst. **34** (2001), 1, 27–44.
- [49] J. Leroux, *A polynomial time Presburger criterion and synthesis for number decision diagrams*, 20th IEEE Symposium on Logic in Computer Science (Chicago, IL, USA, 26–29 June 2005), IEEE Computer Society, Chicago, IL, USA, 2005, 147–156.
- [50] D. Lind, B. Marcus, *An Introduction to Symbolic Dynamics and Coding*, Cambridge University Press, Cambridge, 1995.
- [51] N. Loraud, *β -shift, systèmes de numération et automates*, J. Théor. Nombres Bordeaux **7** (1995), 2, 473–498.
- [52] M. Lothaire, *Combinatorics on Words*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 1997.
- [53] M. Lothaire, *Algebraic Combinatorics on Words*, Encyclopedia of Math. and Its Applications, vol. 90, Cambridge University Press, Cambridge, 2002.
- [54] V. Marsault, *An efficient algorithm to decide periodicity of b -recognizable sets using LSDF convention*, Log. Methods Comput. Sci. **15** (2019), Paper No. 8, 30.
- [55] V. Marsault, J. Sakarovitch, *Ultimate periodicity of b -recognizable sets: a quasilinear procedure*, Developments in Language Theory (Marne-la-Vallée, June 18–21, 2013), Lect. Notes in Comput. Sci., 7907, Springer, Heidelberg, 2013, 362–373.
- [56] A. Massuir, J. Peltomäki, M. Rigo, *Automatic sequences based on Parry or Bertrand numeration systems*, Advances in Applied Mathematics **108** (2019), 11–30.
- [57] Ch. Mauduit, *Multiplicative properties of the Thue-Morse sequence*, Period. Math. Hungar. **43** (2001), 1–2, 137–153.
- [58] I. V. Mitrofanov, *Periodicity of morphic words*, Fundam. Prikl. Mat. **18** (2013), 4, 107–119.
- [59] H. Mousavi, *Walnut*, available at <https://cs.uwaterloo.ca/~shallit/walnut.html>, 2015, arXiv:1603.06017.

- [60] A. A. Muchnik, *The definable criterion for definability in Presburger arithmetic and its applications*, Theoret. Comput. Sci. **290** (2003), 3, 1433–1444.
- [61] J.-J. Pansiot, *Complexité des facteurs des mots infinis engendrés par morphismes itérés*, 11th International Conference on Automata, Languages and Programming (Antwerp, Belgium, July 16–20, 1984), Lecture Notes in Computer Science, 172, Springer-Verlag, 1984, 380–389.
- [62] J.-J. Pansiot, *Decidability of periodicity for infinite words*, RAIRO Theor. Inform. Appl. **20** (1986), 1, 43–46.
- [63] W. Parry, *On the β -expansions of real numbers*, Acta Math. Acad. Sci. Hungar. **11** (1960), 401–416.
- [64] M. Queffélec, *Questions around the Thue-Morse sequence*, Unif. Distrib. Theory **13** (2018), 1, 1–25.
- [65] M. Rigo, *Generalization of automatic sequences for numeration systems on a regular language*, Theoret. Comput. Sci. **244** (2000), 271–281.
- [66] M. Rigo, *Formal Languages, Automata and numeration Systems vol. 2: Applications to recognizability and decidability*, ISTE, London; John Wiley & Sons, Inc., Hoboken, NJ, 2014.
- [67] M. Rigo, A. Maes, *More on generalized automatic sequences*, J. of Automata, Languages and Combinatorics **7** (2002), 351–376.
- [68] A. M. Robert, *A course in p -adic analysis*, Springer, New York, 2000.
- [69] E. Rowland, R. Yassawi, *p -adic asymptotic properties of constant-recursive sequences*, Indag. Math. **28** (2017), 205–220.
- [70] J. Sakarovitch, *Elements of Automata Theory*, Cambridge University Press, Cambridge, 2009. Translated from the 2003 French original by Reuben Thomas.
- [71] A. Salomaa, M. Soittola, *Automata-theoretic aspects of formal power series* Texts and Monographs in Computer Science, Springer-Verlag, New York, Heidelberg, Berlin, 1978.
- [72] K. Schmidt, *On periodic expansions of Pisot numbers and Salem numbers*, Bull. London Math. Soc. **12** (1980), 269–278.

- [73] J. Shallit, *Numeration systems, linear recurrences and regular sets*, Inform. and Comput. **113.2** (1994), 331–347.
- [74] J. Shallit, *Enumeration and automatic sequences*, Pure Math. Appl. (PU.M.A.) **25** (2015), 1, 96–106.
- [75] L. Waxweiler, *Caractère reconnaissable d'ensembles des polynômes à coefficients dans un corps fini*, PhD thesis, University of Liège, Belgium, 2009.

List of Figures

0.1	A finite automaton.	1
1.1	A deterministic finite automaton.	13
1.2	A trim deterministic finite automaton.	14
1.3	The canonical automaton accepting $\text{Fact}(D_\varphi)$	24
1.4	The canonical automaton accepting $\{0, 1, 2\}^*(\{\varepsilon\} \cup 30^*)$	25
1.5	Comparison of the sets \mathcal{B} and \mathcal{R}	28
1.6	A DFA accepting $\text{val}_U^{-1}(4\mathbb{N}+1)$ ($U_{i+2} = 7U_{i+1} - 2U_i, U_0 = 1, U_1 = 7$).	30
1.7	A DFAO generating a Fibonacci-automatic sequence.	37
2.1	The different words (case where $i \leq \text{rep}_U(s) $).	64
2.2	The different words (case where $i > \text{rep}_U(s) $).	65
3.1	The Thue-Morse set is 2-recognizable.	86
3.2	The Thue-Morse set is 4-recognizable.	87
3.3	The minimal automaton of $\text{val}_2^{-1}(3\mathcal{T})$	88
3.4	The automata $\mathcal{A}_{\mathcal{T},2}$ (left) and $\mathcal{A}_{\mathcal{T},4}$ (right).	89
3.5	The automaton $\mathcal{A}_{6,2,4}$	90
3.6	The projected automaton $\Pi(\mathcal{A}_{6,2,4})$	92
3.7	The product automaton $\mathcal{A}_{6,2,4} \times \mathcal{A}_{\mathcal{T},4}$	94
3.8	The projected automaton $\Pi(\mathcal{A}_{6,2,4} \times \mathcal{A}_{\mathcal{T},4})$	95
3.9	The classes of the projected automaton $\Pi(\mathcal{A}_{24,23,4} \times \mathcal{A}_{\mathcal{T},4})$	105
3.10	The classes of the projected automaton $\Pi(\mathcal{A}_{24,0,4} \times \mathcal{A}_{\mathcal{T},4})$	106
3.11	The elements of the sets C'_α up to the first belonging to the classes C_α	106
3.12	The classes of the automaton of $\Pi(\mathcal{A}_{6,2,4} \times \mathcal{A}_{\mathcal{T},4})$	115
3.13	The automaton $\mathcal{M}_{6,2,\mathcal{T},4}$	116
3.14	The classes of the automaton of $\Pi(\mathcal{A}_{6,2,4} \times \mathcal{A}_{\overline{\mathcal{T}},4})$	120
3.15	The automaton $\mathcal{M}_{6,2,\overline{\mathcal{T}},4}$	121
3.16	Minimal automaton recognizing the set $\{(2^n, 3 \cdot 2^n) : n \in \mathbb{N}\}$	122

3.17	Construction of the automaton recognizing $2X_{3,2,3,0}$ in base 3.	123
3.18	Minimal automaton of the base-3 expansions of the set $2X_{3,2,3,0}$.	124
4.1	DFAO generating the Thue-Morse word.	125
4.2	DFAO generating several automatic sequences.	126
4.4	Values of $t - S_{1,3}$	135