Numeration systems on a regular language : Arithmetic operations, recognizability and formal power series

Michel Rigo

Institut de Mathématiques, Université de Liège,
Grande Traverse 12 (B 37), B-4000 Liège, Belgium.
M.Rigo@ulg.ac.be

Abstract

Generalizations of numeration systems in which $\mathbb{N}$ is recognizable by a finite automaton are obtained by describing a lexicographically ordered infinite regular language $L \subset \Sigma^*$. For these systems, we obtain a characterization of recognizable sets of integers in terms of $\mathbb{N}$-rational formal series. After a study of the polynomial regular languages, we show that, if the complexity of L is $\Theta(n^l)$ (resp. if L is the complement of a polynomial language), then multiplication by $\lambda \in \mathbb{N}$ preserves recognizability only if $\lambda = \beta^{l+1}$ (resp. if $\lambda \neq (\#\Sigma)^\beta$) for some $\beta \in \mathbb{N}$. Finally, we obtain sufficient conditions for the notions of recognizability for abstract systems and some positional number systems to be equivalent.

1 Introduction

A usual way of representing non-negative integers is to consider a strictly increasing sequence $(U_n)_{n\in\mathbb{N}}$ of positive integers and use the greedy algorithm to represent each natural number x by a word $d = d_n \cdots d_0$ such that $x = d_n U_n + \cdots + d_0 U_0$; the word d is the *normalized representation* of x. These conventions lead to the so-called *positional numeration systems*, see for instance [9].

The recognizability of subsets of $\mathbb{N}$ is extensively studied in a lot of recent papers [4,5,11,15,19]. A set of integers is said to be *recognizable* if the normalized representations of its elements are accepted by a finite automaton. The case when $\mathbb{N}$ is recognizable is of special interest because then it is very easy to decide wether or not a given word represents an integer. A necessary condition for $\mathbb{N}$ to be recognizable is that $(U_n)_{n\in\mathbb{N}}$ is the solution of a linear recurrence

Preprint submitted to Elsevier Preprint

relation [19]. Sufficient conditions for the recognizability of $\mathbb{N}$ are discussed in [12,14].

The use of the greedy algorithm has an important consequence. If x and y are two non-negative integers such that $x < y$ then the normalized representation of x is lexicographically less than the one of y.

Having in mind this property, a generalization of positional numeration systems for which $\mathbb{N}$ is recognizable was recently introduced in [13]. An *abstract numeration system* is a triple $S = (L, \Sigma, <)$ where L is an infinite regular language over a totally ordered alphabet $(\Sigma, <)$. The lexicographic ordering of L gives a one-to-one correspondence r_S between the set $\mathbb{N}$ of natural numbers and the language L. The word $\mathrm{r}_S(x)$ of L is the *S-representation* of the integer x. Having generalized numeration at our disposal systems, it is natural to be interested in the corresponding S-recognizable subsets of $\mathbb{N}$. A subset $X \subset \mathbb{N}$ is called *S-recognizable* if $\mathrm{r}_S(X)$ is a regular subset of L. The first properties of these sets are given in [13,16,17].

In the present paper S-recognizable sets are characterized in terms of rational series in the noncommuting variables $\sigma \in \Sigma$ and with coefficients in $\mathbb{N}$. In particular, we show that $\sum_{n\in\mathbb{N}} n\, \mathrm{r}_S(n)$ is $\mathbb{N}$-rational (this kind of result is also discussed in [2,6]). Using classical results about rational series, we obtain a generalization of the fact given in [13] that ultimately periodic sets are S-recognizable for any numeration system S. This result is quite remarkable in the frame of a possible generalization of the famous Cobham's theorem [7].

In this paper, our main purpose is to study the stability of the S-recognizability under arithmetic operations like addition and multiplication by a constant. If addition preserves the S-recognizability then multiplication by 2 also preserves the S-recognizability. So, a natural question about the stability of the recognizability arises. When does the multiplication by an integer λ preserve the recognizability ?

It is well known that for positional numeration systems with an integer base p the problem of addition and multiplication by a constant is completely settled. The p-recognizable sets are exactly those defined in the first order structure $\langle \mathbb{N}, +, V_p \rangle$ (see for instance [5]). It is obvious that addition and multiplication by a constant are definable in the Presburger arithmetic $\langle \mathbb{N}, + \rangle$. They thus preserve p-recognizability. In fact, the problem of the recognizability of addition for positional number systems is only settled for the class of positional numeration systems built on a linear reccurent sequence $(U_n)_{n\in\mathbb{N}}$ such that the characteristic polynomial of $(U_n)_{n\in\mathbb{N}}$ is the minimal polynomial of a Pisot number [4,11].

On the other hand, it is shown in [13] that for the abstract numeration system $S = (a^*b^*, \{a, b\}, a < b)$, the multiplication by a non-negative integer λ

transforms the S-recognizable sets into S-recognizable sets if and only if λ is an odd perfect square. In particular, the multiplication by 2 does not preserve S-recognizability.

Notice that the language a^*b^* has a polynomial complexity (the complexity function $\rho_L(n)$ of a language L counts the number of words of length n in L). We generalize here this result for polynomial languages as follows: if a numeration system is built on a regular language with complexity in $\Theta(n^l)$ then the multiplication by λ preserves the recognizability only if λ is of the form β^{l+1} for some integer β. (As a consequence, the addition cannot be a regular map for numeration systems on polynomial regular languages.)

Our proof relies on the following property of polynomial regular languages. Denote by $v_L(n)$, or simply v_n if the context is clear, the number of words of length not exceeding n in L. Assume that the complexity of L is $\Theta(n^l)$. Then the sequence $(v_n/n^{l+1})_{n\in\mathbb{N}}$ converges to a strictly positive limit. (Notice that, in contrast, the sequence $(\rho_L(n)/n^l)_{n\in\mathbb{N}}$ generally does not converge.)

It is well known that the complexity function of a regular language is $\Theta(n^l)$ or of order $2^{\Theta(n)}$ (in the latter case, the language is said to be exponential) [20].

For abstract numeration systems built on the complement of a polynomial language, we show that multiplication by a power of the cardinality of the alphabet does not preserve recognizability.

Finally, assume that S is an abstract numeration system built on an exponential language L with exponential complement. Under some assumptions on the complexity functions of the languages accepted from the different states of a deterministic finite automaton accepting L, we show that S-recognizability is equivalent to U-recognizablity where U is some positional numeration system related to a Pisot number. Consequently, addition and multiplication by a constant preserve S-recognizability. An example of an abstract system satisfying these assumptions is a generalization of the Fibonacci system [4].

This paper is organized as follows. It begins with a section recalling basic facts in formal language theory and introducing abstract numeration systems. Next, in Section 3, for the sake of simplicity, we assume that the complexity of the language of the numeration system S is a polynomial of degree l with rational coefficients. Even for a language of exact polynomial complexity, we construct a subset $X \subset \mathbb{N}$ which is recognizable and we prove that λX is not recognizable for any $\lambda \in \mathbb{N} \setminus \{n^{l+1} : n \in \mathbb{N}\}$. Section 4 is devoted to properties of regular languages with complexity in $\Theta(n^l)$. It extend in some way the results of [20]. In particular, we prove the convergence of $(v_n/n^{l+1})_{n\in\mathbb{N}}$. In Section 5, having all the necessary material of Section 4 and the scheme of Section 3, we develop the general case of multiplication by a constant for a regular language with $\Theta(n^l)$ complexity. The canvas of the proof is the same as for exact polynomials but we

need sharper discussions. Section 6 is devoted to the problem of multiplication by a constant for abstract numeration systems built on the complement of polynomial languages and Section 7 is concerned with the relation between positional numeration systems and abstract systems. Finally, the last section deals with the other topic of this paper: the characterization of S-recognizable sets in terms of rational series. This last section is independent of the previous ones and could be read separately.

2 Basic facts

We denote by Σ^* the free monoid (with identity ε) generated by the linearly ordered finite set $\Sigma = \{\sigma_1 < \cdots < \sigma_p\}$. For a set S, $\#S$ is the cardinality of S and for a string $w \in \Sigma^*$, $|w|$ is the length of w.

If the reader is not familiar with regular languages and finite automata, see for instance [8,21] where these notions are well detailed.

Let $L \subseteq \Sigma^*$ be a regular language; the minimal automaton of L is a 5-tuple $M_L = (K, s, F, \Sigma, \delta)$ where K is the set of states, s is the initial state, F is the set of final states and $\delta : K \times \Sigma \to K$ is the transition function. We often write $k.\sigma$ instead of $\delta(k, \sigma)$. Recall that the elements of K are the *derivatives* [8, III.5]

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

The state k is equal to $w^{-1}.L$ if and only if $k = s.w$; $w^{-1}.L$ being then the set L_k of words accepted by M_L from k. In particular, $L = L_s$.

We denote $u_l(k)$ the number $\#(L_k \cap \Sigma^l)$ of words of length l belonging to L_k and $v_l(k)$ the number of words of length at most l belonging to L_k,

$$v_l(k) = \sum_{i=0}^{l} u_i(k).$$

Notice that the notations L_k, $u_l(k)$ and $v_l(k)$ are relevant to any DFA (deterministic finite automaton) accepting L.

Let us recall a useful property of regular languages. It is often used to show that a particular language is not regular.

Proposition 1 *Let $L \subset \Sigma^*$ be a regular language. The set $|L| = \{|w| \, : \, w \in L\}$ is a finite union of arithmetic progressions.*

PROOF. First, if $\Gamma = \{\gamma\}$ then the regular languages over Γ are the languages of the form $\{\gamma^n : n \in A\}$ where A is a finite union of arithmetic progressions. Next, consider the length-preserving homomorphism $k : \Sigma = \{\sigma_1, \ldots, \sigma_p\} \to \Gamma : \sigma_i \mapsto \gamma$. It is clear that k maps regular languages over Σ onto regular languages over Γ.

The linear order on Σ induces a *lexicographic ordering* on Σ^*. Let x and y be in Σ^*. We say that $x < y$ if $|x| < |y|$ or if $|x| = |y|$ and there exist letters $\alpha < \beta$ such that $x = w\alpha x'$ and $y = w\beta y'$.

An extension of numeration systems in which the set of representations is regular is the following.

Definition 2 [13] An *(abstract) numeration system* or *numeration system on a regular language* is a triple $(L, \Sigma, <)$ where L is an infinite regular language over a linearly ordered finite alphabet $(\Sigma, <)$. The lexicographic ordering of L gives a one-to-one correspondence r_S between the set $\mathbb{N}$ of natural numbers and the language L.

For each $n \in \mathbb{N}$, $\mathrm{r}_S(n)$ is the $(n+1)^{th}$ word of L with respect to the lexicographic ordering and is called the *S-representation* of n. For $w \in L$, we set $\mathrm{val}_S(w) = \mathrm{r}_S^{-1}(w)$ and we call it the *numerical value* of w.

The mappings val_S and r_S are sometimes called *ranking* and *unranking* in the literature [6].

This way of representing integers generalizes linear numeration systems in which $\mathbb{N}$ is recognizable by finite automata [4,5,11,12,14,19]. Examples of such systems are the numeration systems defined by a recurrence relation whose characteristic polynomial is the minimum polynomial of a Pisot number (i.e., an algebraic integer $\alpha > 1$ such that its Galois conjugates have modulus less than one) [4]. (Indeed, with this hypothesis, the set of representations of the integers is a regular language.) The standard numeration systems with integer base and also the Fibonacci system belong to this class.

Definition 3 Let S be a numeration system. A subset X of $\mathbb{N}$ is *S-recognizable* if $\mathrm{r}_S(X)$ is recognizable by a finite automaton.

Let $S = (L, \Sigma, <)$ be a numeration system. Each $k \in K$ for which L_k is infinite leads to the numeration system $S_k = (L_k, \Sigma, <)$. The applications r_{S_k} and val_{S_k} are simply denoted r_k and val_k if the context is clear. If L_k is finite, then the applications r_k and val_k are defined as in the infinite case but the domain of the former restricts to $\{0, \ldots, \#L_k - 1\}$.

With these notations, we can recall a very useful proposition.

Lemma 4 [13] *Let $S = (L, \Sigma, <)$ and $M = (K, s, F, \Sigma, \delta)$ be a DFA accepting L. If σw belongs to L_k, $k \in K$, $\sigma \in \Sigma$, $w \in \Sigma^+ = \Sigma^* \setminus \{\varepsilon\}$, then*

$$\mathrm{val}_k(\sigma w) = \mathrm{val}_{k.\sigma}(w) + v_{|w|}(k) - v_{|w|-1}(k.\sigma) + \sum_{\sigma' < \sigma} u_{|w|}(k.\sigma').$$

3 Multiplication for exact polynomial languages

In [13], we proved that for the numeration system $S = (a^*b^*, \{a, b\}, a < b)$, the multiplication by a non-negative integer λ transforms the S-recognizable sets into S-recognizable sets if and only if λ is an odd perfect square.

In this section, we study the family of abstract number systems built on regular languages with polynomial complexity function. This step contains the main ideas for the discussion of arbitrary polynomial languages (i.e., languages with complexity function bounded by a polynomial).

The proof of the following lemma is left to the reader.

Lemma 5 *Let $f : \mathbb{N} \to \mathbb{N}$ be a strictly increasing function such that $f(\mathbb{N})$ is a finite union of arithmetic progressions. There exist $y_0 \in \mathbb{N}$ and $\Gamma \in \mathbb{N} \setminus \{0\}$ such that $\forall y \geq y_0$, $y \in f(\mathbb{N}) \Leftrightarrow y + \Gamma \in f(\mathbb{N})$. If $k = f^{-1}(y_0 + \Gamma) - f^{-1}(y_0)$. Then, for all $x \geq f^{-1}(y_0)$, $n \in \mathbb{N}$,*

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

Definition 6 The *complexity function* of a language $L \subseteq \Sigma^*$ is

$$\rho_L : \mathbb{N} \to \mathbb{N} : n \mapsto \#(\Sigma^n \cap L).$$

In the following of this section, we assume that we deal with "true" complexity functions, i.e., if ρ_L is a polynomial belonging to $\mathbb{Q}[x]$ and $n \in \mathbb{N}$ then $\rho_L(n)$ is a non-negative integer. We equally use the notation $\rho_L(n)$, $u_n(s)$ or even u_n provided the context is clear.

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

Lemma 7 *If H is a polynomial such that $\forall x \in \mathbb{N} \setminus \{0\}$, $H(x) \in \mathbb{Z}$ then $H(\mathbb{Z}) \subseteq \mathbb{Z}$.*

PROOF. One can proceed by induction on the degree of H.

Proposition 8 *Let $L \subset \Sigma^*$ be a regular language such that*

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

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

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

PROOF. One can build a polynomial $P \in \mathbb{Q}[x]$ of degree $l+1$ such that $P(0) = 0$ and for all $n \geq 1$, $P(n+1) = P(n) + \rho_L(n)$.

Indeed, let $P(x) = b_{l+1}\, x^{l+1} + \cdots + b_1\, x + b_0$. The conditions on P gives the following triangular system

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

This polynomial P has some useful properties. We have the polynomial identity $P(x+1) = P(x) + \rho_L(x)$ for $x \in \mathbb{N} \setminus \{0\}$. Then it holds for $x \in \mathbb{R}$ if we extend the definition of ρ_L to $\rho_L : \mathbb{R} \to \mathbb{R} : x \mapsto a_l\, x^l + \cdots + a_0$. By Lemma 7, $P(1) = \rho_L(0) = a_0 \in \mathbb{Z}$. One shows by induction on $n \in \mathbb{N}$ that $P(n)$ (resp. $P(-n)$) is an integer since $\rho_L(\mathbb{N}) \subset \mathbb{N}$ (resp. since $\rho_L(\mathbb{Z}) \subset \mathbb{Z}$ by Lemma 7).

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

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

Indeed, an integer x has a representation of length n if $v_{n-1} \leq x < v_n$ and

$$v_n = \sum_{i=1}^{n} \rho_L(i) + 1 = \sum_{i=1}^{n} [P(i+1) - P(i)] + 1 = P(n+1) - P(1) + 1.$$

Notice that $\mathrm{r}_S(P(\mathbb{N}))$ is a translation of the set $\mathcal{I}(L, <)$ of the first words of each length (let $Y, Z \subset \mathbb{N}$, Y is a *translation* of Z if there exists a constant $t \in \mathbb{N}$ such that $z \in Z \Leftrightarrow z + t \in Y$). Therefore $X = P(\mathbb{N})$ is S-recognizable, see [13,19].

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

For n large enough, we first show that

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

The first inequality is obvious. In view of (1), to satisfy the second inequality, it suffices to check whether

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

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

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

The coefficient of n^{l+1} in the last expression is $b_{l+1}\,(\lambda^{(l+1)/l} - \lambda) > 0$. So, there exists n_0 such that for all $n \geq n_0$, this polynomial expression of degree $l+1$ is strictly positive and $|\mathrm{r}_S(\lambda\, P(n))| < \lambda^{1/l}\, n$.

If n is sufficiently large, we show that

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

Let $i = |\mathrm{r}_S(\lambda\, P(n))|$. In view of (1), one has to verify that

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

By definition of P and by (1), one has

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

Therefore it is sufficient to check whether $P(i) - a_0 + \lambda\, \rho_L(n) > P(i+1) - a_0$, which occurs if and only if

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

To verify that this inequality holds, remember that $a_l > 0$ and for $n \geq n_0$, $1 \leq \frac{i}{n} < \lambda^{1/l}$. Thus one studies the quotient $\frac{\lambda\, \rho_L(n) - \rho_L(i)}{n^l}$ when $n \to +\infty$,

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

So there exists $n_0' \geq n_0$ such that for all $n \geq n_0'$, $|\mathrm{r}_S(\lambda\, P(n+1))| > |\mathrm{r}_S(\lambda\, P(n))|$.

Assume that $\mathrm{r}_S(\lambda\, P(\mathbb{N}))$ is regular then by Proposition 1, the set $|\mathrm{r}_S(\lambda\, P(\mathbb{N}))|$ is a finite union of arithmetic progressions. We may apply Lemma 5; indeed, the function $|\mathrm{r}_S(\lambda\, P(.))|$ is strictly increasing for $n \geq n_0'$ and there exist l_0 and $\Gamma_\lambda \in \mathbb{N} \setminus \{0\}$ (simply written Γ) such that $\forall l \geq l_0,\ l \in |\mathrm{r}_S(\lambda\, P(\mathbb{N}))| \Leftrightarrow l + \Gamma \in |\mathrm{r}_S(\lambda\, P(\mathbb{N}))|$. Let $n_1 \geq n_0'$ be such that $|\mathrm{r}_S(\lambda\, P(n_1))| > l_0$. By Lemma 5, there exists $k_\lambda \in \mathbb{N} \setminus \{0\}$ (simply written k) such that for all $n \geq n_1$ and for all $\alpha \in \mathbb{N}$,

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

Let $i = |\mathrm{r}_S(\lambda\, P(n))|$. In view of (1), one has

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

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

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

Notice that this latter coefficient vanishes only if $\lambda = \left(\frac{\Gamma}{k}\right)^{l+1}$. By hypothesis, this case is excluded (notice that $\frac{\Gamma}{k} \in \mathbb{Q}\backslash\mathbb{N} \Rightarrow (\frac{\Gamma}{k})^{l+1} \notin \mathbb{N}$). So, $k \neq \frac{\Gamma}{\lambda^{1/(l+1)}}$ and we have the condition

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

But $\lambda\, P(n + \alpha k) - P(i + \alpha\Gamma + 1) + a_0 \leq 0$ for all $\alpha \in \mathbb{N}$. Then we have simultaneously the condition

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

which leads to a contradiction.

In Proposition 8, we exhibit a recognizable set $X = P(\mathbb{N})$ such that $|\mathrm{r}_S(\lambda\, P(\mathbb{N}))|$ is not a finite union of arithmetic progressions. When we consider the case $\lambda = \beta^{l+1}$, $\beta \in \mathbb{N} \setminus \{0, 1\}$, we cannot find easily a subset X which is recognizable and such that $\lambda\, X$ is not. The next proposition shows that $|\mathrm{r}_S(\beta^{l+1} P(\mathbb{N}))|$ is ultimately an arithmetic progression if ρ_L is a polynomial of degree l.

Proposition 9 *With the assumptions and notations of Proposition 8, there exists $C \in \mathbb{Z}$ such that for n large enough,*

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

The proof of this proposition does not highlight new constructions or results and is thus omitted.

4 Properties of polynomial regular languages

It is shown in [20] that the complexity function $\rho_L(n)$ of polynomial regular language L is $\Theta(n^l)$ for some $l \in \mathbb{N}$. In this section, we have a closer look at these complexity functions. In particular, we show that the sequence $(v_n/n^{l+1})_{n\in\mathbb{N}}$ converges if the complexity of L is $\Theta(n^l)$.

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

Definition 10 [20] Let $M = (K, s, F, \Sigma, \delta)$ be a DFA. A word $w = w_1 \cdots w_n \in \Sigma^*$ is said to be *t-tiered*, $t \geq 0$, with respect to M if the state transition sequence of w is given by

$$(s.w_1)(s.w_1w_2)\cdots(s.w_1\cdots w_n) = \alpha\beta_1^{d_1}\gamma_1\cdots\beta_t^{d_t}\gamma_t$$

where

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

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

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

The next lemma is just a refinement of [20, Lemma 1]. We simply remark that one can consider an ultimately periodic sequence n_i such that $\rho_L(n_i) \geq b_0\, n_i^l$ for some positive constant b_0.

Lemma 11 *If L is a regular language such that $\rho_L(n)$ is $\Theta(n^l)$ for some integer l then there exist constants b_0 and C and a strictly increasing infinite sequence $n_0, n_1, \ldots, n_i, \ldots$ of integers such that for all $i \in \mathbb{N}$, $\rho_L(n_i) \geq b_0\, n_i^l$ and $n_{i+1} - n_i = C$.*

PROOF. In view of Lemmas 2–4 of [20] about the structure of a polynomial language, it is obvious that there exists a word $w \in L$ which is $(l+1)$-tiered, $w = x\, y_1^{d_1}\, z_1 \ldots y_{l+1}^{d_{l+1}}\, z_{l+1}$. Let $C = |y_1| \ldots |y_{l+1}|$. As shown in [20], there exists a constant b_0 such that the number of words of length $n_t = |x z_1 \ldots z_{l+1}| + t\, C$ is greater than $b_0\, n_t^l$ for any integer t.

Recall (see [3]) that the finite sum of integer powers is given by

$$\sum_{i=0}^{n} i^p = \frac{(n+B+1)^{p+1} - B^{p+1}}{p+1} \tag{2}$$

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

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

This formula will be useful in the next lemma and in the proof of Theorem 14.

Lemma 12 *If L is a regular language such that $\rho_L(n)$ is $\Theta(n^l)$ then $v_n = \sum_{i=0}^{n} \rho_L(i)$ is $\Theta(n^{l+1})$. Moreover, there exists a constant J such that $v_{n_i} \geq J\, n_i^{l+1}$ for the sequence $n_0, n_1, \ldots, n_i, \ldots$ of Lemma 11.*

PROOF. i) There exist N_0 and a constant b_1 such that for all $n \geq N_0$, $\rho_L(n) \leq b_1\, n^l$. If one replaces b_1 by a bigger constant then the latter inequality holds for all n. For n sufficiently large, there exists a constant K such that

$$v_n = \sum_{i=0}^{n} \rho_L(i) \leq b_1 \sum_{i=0}^{n} i^l \leq K\, n^{l+1}.$$

ii) With the sequence n_i of Lemma 11, one has

$$v_{n_i} = \sum_{j=0}^{n_i} \rho_L(j) \geq \sum_{j=0}^{i} \rho_L(n_j) \geq b_0 \sum_{j=0}^{i} (n_0 + j\, C)^l \geq b_0\, C^l \sum_{j=0}^{i} j^l.$$

Since $n_i = n_0 + i\, C$, there exists a constant $J > 0$ such that

$$v_{n_i} \geq J\, n_i^{l+1}$$

holds for i sufficiently large.

So far, we have a sequence n_i such that $n_i = n_0 + i\,C$ and constants b_0, b_1, K and J such that for n and i sufficiently large,

$$\begin{cases} \rho_L(n) \;\leq b_1\, n^l \\ \rho_L(n_i) \geq b_0\, n_i^l \end{cases} \text{ and } \begin{cases} v_n \;\leq K\, n^{l+1} \\ v_{n_i} \geq J\, n_i^{l+1} \end{cases} . \tag{3}$$

Now, we give an interesting result about the convergence of the sequence $(\frac{v_n}{n^{l+1}})_{n\in\mathbb{N}}$ when L is a polynomial language. A remarkable fact is that the limit always exists. Although this is generally not the case for the sequence $(\frac{\rho_L(n)}{n^l})_{n\in\mathbb{N}}$. Consider for instance the language $W = a^*b^* \setminus (\{a,b\}^2)^*$. It is obvious that $\rho_W(2n+1) = 2n+2$, $\rho_W(2n) = 0$ and $v_{2n} = v_{2n+1} = (n+1)^2$.

Lemma 13 *Let $\rho_1, \ldots, \rho_k, \theta_1, \ldots, \theta_k, \Phi_1, \ldots, \Phi_k$ be real numbers such that for all $i \neq j$, $\theta_i \not\equiv \theta_j \pmod{2\pi}$ and for all j, $\rho_j \neq 0$. There exists $\varepsilon > 0$ such that*

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

for an infinite sequence of integers n.

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

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

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

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

Therefore one has

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

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

We are now able to prove the convergence of $(v_n/n^{l+1})_{n\in\mathbb{N}}$. This result and its proof were suggested by P. Lecomte.

Theorem 14 *If L is a regular language such that $\rho_L(n)$ is $\Theta(n^l)$ then the sequence $(\frac{v_n}{n^{l+1}})_{n\in\mathbb{N}}$ converges to a strictly positive limit. Moreover, 1 is a root of the characteristic polynomial of the sequence $(\rho_L(n))_{n\in\mathbb{N}}$ with a multiplicity at least equal to $l+1$.*

PROOF. Since L is regular, the sequence $(\rho_L(n))_{n\in\mathbb{N}}$ satisfies a recurrence relation with coefficients in $\mathbb{Z}$. A proof of this result can be found in [1]; it is due to the fact that the series $f_L(X) = \sum_{n\geq 0} \rho_L(n)\, X^n$ is $\mathbb{N}$-rational. Therefore, we can write $\rho_L(n)$ as a finite sum,

$$\rho_L(n) = \sum_i P_i(n)\, z_i^n \tag{4}$$

where the P_i's are polynomials of degree less than α_i and the z_i's are distinct complex numbers.

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

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

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

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

So $R_n \to 0$ if $n \to +\infty$. Therefore, by Lemma 13, there exists $\varepsilon > 0$ and an infinite sequence of integers such that

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

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

Consequently, in (4) the degree of the polynomials P_k's corresponding to the numbers of modulus one cannot exceed l and there is no z_i of modulus greater than one. So there exist polynomials Q_j's of degree not exceeding l such that

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

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

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

Let $q_j = \frac{\partial^l}{\partial^l z} \frac{Q_j(z)}{l!}$ (i.e., q_j is the coefficient of n^l in $Q_j(n)$; notice that q_j could be zero). We have

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

and by definition of v_n, $\frac{v_n}{n^{l+1}}$ can be written

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

We have, by (2)

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

and

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

because the degree of $Q_j(p) - q_j\, p^l$ is less than l and

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

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

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

On the other hand,

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

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

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

Indeed, let us consider one of the terms appearing in $T(n)$, say $P(n)\, z^n$ with $|z| = \delta < 1$ and $P(n) = \sum_{i=0}^{q} a_i\, n^i$, $q \in \mathbb{N}$. So

$$\left|\frac{1}{n^{l+1}} \sum_{p=0}^{n} P(p)\, z^p\right| \leq \sum_{i=0}^{q} \frac{|a_i|}{n^{l+1}} \sum_{p=0}^{n} \delta^p\, p^i$$

Hence the limit by applying the same kind of computation as for the previous limit and using the fact that $\delta < 1$.

To conclude, q_0 cannot vanish. Otherwise, $\lim_{n\to\infty} \frac{v_n}{n^{l+1}} = 0$, which is a contradiction with Lemma 12. Thus, 1 has necessary, a multiplicity at least $l+1$ as root of the characteristic polynomial of the recurrence.

5 Multiplication for arbitrary polynomial languages

Thanks to the results of the previous section, we generalize Proposition 8 to arbitrary regular languages of polynomial complexity.

Theorem 15 *Let $L \subset \Sigma^*$ be a regular language such that $\rho_L(n)$ is $\Theta(n^l)$ for some integer l. If $\lambda \in \mathbb{N} \setminus \{n^{l+1} : n \in \mathbb{N}\}$, then there exists a subset X of $\mathbb{N}$ such that $\mathrm{r}_S(X)$ is regular and that $\mathrm{r}_S(\lambda\, X)$ is not.*

PROOF. In this proof, we use the sequence n_i and the constants C, J, K, b_0 and b_1 of (3). By definition of an abstract numeration system, it is clear that for n sufficiently large, $n+1 \leq |\mathrm{r}_S(v_n)| \leq n+C+1$ since for C consecutive values of $\rho_L(n)$ at least one of them does not vanish. (Notice that if $\rho_L(n) > 0$ for all n, then $|\mathrm{r}_S(v_n)| = n+1$.) Recall also that $|\mathrm{r}_S(x)| = n$ iff $v_{n-1} \leq x < v_n$.

i) Assume that the integer constant λ is strictly greater than $\left(\frac{K}{J}\right)^l$. We show that for n large enough,

$$n+1 \leq |\mathrm{r}_S(\lambda\, v_n)| \leq \lceil \lambda^{1/l}\, n \rceil + C - 1 < \lambda^{1/l}\, n + C. \tag{5}$$

It is sufficient to show that $\lambda\, v_n < v_{\lceil \lambda^{1/l}\, n\rceil + C-1}$. By Lemma 12, there exists $k \in \{\lceil \lambda^{1/l}\, n\rceil, \ldots, \lceil \lambda^{1/l}\, n\rceil + C - 1\}$ such that $v_k \geq J\, k^{l+1}$. Moreover the function $n \mapsto v_n$ is increasing. So,

$$v_{\lceil \lambda^{1/l}\, n\rceil + C-1} \geq J\, \lceil \lambda^{1/l}\, n\rceil^{l+1} \geq J\, \lambda^{\frac{l+1}{l}}\, n^{l+1}.$$

Moreover, by Lemma 12, $\lambda\, v_n \leq \lambda\, K\, n^{l+1}$. By the choice of λ, it is clear that $\lambda\, K\, n^{l+1} < J\, \lambda^{\frac{l+1}{l}}\, n^{l+1}$.

ii) In Lemma 11 and Lemma 12, we have introduced the constants b_0 and b_1 such that $b_0 \leq b_1$. Let $s \in \mathbb{N}\backslash\{0\}$ be such that $s\, b_0 > b_1$. We show that the function

$$i \mapsto |\mathrm{r}_S(\lambda\, v_{n_{si}-1})|$$

is strictly increasing for i sufficiently large. So, we have to show that

$$|\mathrm{r}_S(\lambda\, v_{n_{s(i+1)}-1})| = |\mathrm{r}_S(\lambda\, v_{n_{si}+sC-1})| > |\mathrm{r}_S(\lambda\, v_{n_{si}-1})|.$$

Let $k = |\mathrm{r}_S(\lambda\, v_{n_{si}-1})|$ then $v_{k-1} \leq \lambda\, v_{n_{si}-1} < v_k$ and we must show that

$$\lambda\, v_{n_{si}+sC-1} = \lambda\, v_{n_{si}-1} + \lambda \sum_{j=0}^{sC-1} \rho_L(n_{si}+j) \geq v_k = v_{k-1} + \rho_L(k).$$

So, it is sufficient to show that $\lambda \sum_{j=0}^{sC-1} \rho_L(n_{si}+j) \geq \rho_L(k)$. In view of (5), we have $k < \lambda^{1/l}(n_{si}-1) + C$. Therefore $\rho_L(k) < b_1\, [\lambda^{1/l}(n_{si}-1) + C]^l$. On the other hand,

$$\lambda \sum_{j=0}^{sC-1} \rho_L(n_{si}+j) \geq \lambda \sum_{j=0}^{s-1} \underbrace{\rho_L(n_{si}+j\,C)}_{\geq b_0\, (n_{si}+j\,C)^l} \geq \lambda\, b_0\, s\, n_{si}^l.$$

To conclude this part, notice that the coefficient of n_{si}^l in $b_1\, [\lambda^{1/l}(n_{si}-1)+C]^l$ is $b_1\, \lambda$ and by choice of s, we have $b_1\, \lambda < \lambda\, b_0\, s$. So the inequality holds for i sufficiently large.

iii) Consider the subset

$$X = \{v_{n_{si}-1} : i \in \mathbb{N}\} = \{v_{n_0+siC-1} : i \in \mathbb{N}\}.$$

Since $\rho_L(n_0 + siC) > 0$, then $\mathrm{r}_S(v_{n_0+siC-1})$ is the first word of length $n_0 + siC$ and

$$\mathrm{r}_S(X) = \mathrm{r}_S\left(\{v_n : n \in \mathbb{N}\}\right) \cap \Sigma^{n_0} \left(\Sigma^{sC}\right)^*.$$

So X is a S-recognizable subset of $\mathbb{N}$ [19].

Assume that λX is recognizable. Therefore, by Proposition 1, $|\mathrm{r}_S(\lambda X)|$ is a finite union of arithmetic progressions. In view of ii), we can apply Lemma 5 and obtain two integral constants Γ and k such that for all $\alpha \in \mathbb{N}$,

$$|\mathrm{r}_S(\lambda\, v_{n_0+sC(i+\alpha\, k)-1})| = |\mathrm{r}_S(\lambda\, v_{n_0+sCi-1})| + \alpha\, \Gamma.$$

Or equivalently, if we set $z = |\mathrm{r}_S(\lambda\, v_{n_0+sCi-1})|$ then

$$v_{z+\alpha\,\Gamma-1} \leq \lambda\, v_{n_0+sC(i+\alpha\, k)-1} < v_{z+\alpha\,\Gamma}. \tag{6}$$

First consider the left inequality in (6), with the same argument as in i), we obtain

$$v_{z+\alpha\,\Gamma-1} \geq J\,(z+\alpha\,\Gamma-C)^{l+1}.$$

On the other hand,

$$\lambda\, v_{n_0+sC(i+\alpha\, k)-1} \leq \lambda\, K\,(n_0+sCi+sCk\,\alpha-1)^{l+1}.$$

Since α can be arbitrary large, we focus on the terms of the form α^{l+1}. Then we obtain the following condition,

$$J\,\Gamma^{l+1} \leq \lambda\, K\,(sCk)^{l+1} \text{ or } \lambda \geq \frac{J}{K}\left(\frac{\Gamma}{sCk}\right)^{l+1}. \tag{7}$$

If we consider the right inequality in (6), we have from the inequalities in (3), $v_{z+\alpha\,\Gamma} \leq K\,(z+\alpha\,\Gamma)^{l+1}$ and also

$$\lambda\, v_{n_0+sC(i+\alpha\, k)-1} \geq \lambda\, J\,(n_0+sCi+sCk\,\alpha-C)^{l+1}.$$

If we focus on terms of α^{l+1}, we obtain

$$\lambda \leq \frac{K}{J}\left(\frac{\Gamma}{sCk}\right)^{l+1}. \tag{8}$$

iv) By Theorem 14, $(\frac{v_n}{n^{l+1}})_{n\in\mathbb{N}}$ converges to a limit $a > 0$. Consider the sequences

$$K_m = a + \frac{1}{m} \text{ and } J_m = a - \frac{1}{m}.$$

As a consequence of Theorem 14, if m is given then for n large enough, $v_n \leq (a+\frac{1}{m})n^{l+1}$ and $v_n \geq (a-\frac{1}{m})n^{l+1}$. In other words, for a given m there exist i_m and n_m such that for all $i \geq i_m$, $v_{n_i} \geq J_m\, n_i^{l+1}$ and for all $n \geq n_m$, $v_n \leq K_m\, n^{l+1}$. These two inequalities can replace the ones in (3). So if we replace K by K_m and J by J_m, the previous points i), ii) and iii) remain true for n sufficiently large.

For m large enough, the condition $\lambda > \left(\frac{K_m}{J_m}\right)^l$ given in i) is equivalent to $\lambda \geq 2$ and the conditions (7) and (8) may be replaced by a unique condition

$$\lambda = \left(\frac{\Gamma}{sCk}\right)^{l+1}$$

which contradicts the hypothesis (remember that Γ, s, C and k are integers).

This theorem has a direct corollary.

Corollary 16 *Under the assumptions of Theorem 15, the addition is not a regular map (i.e., the graph of the application $(x, y) \mapsto x + y$ is not regular).*

6 Multiplication and complement of polynomial languages

In the previous sections, we have considered multiplication for numeration systems based upon a polynomial language. If the complexity function of a regular language is not bounded by a polynomial then it is of order $2^{\Theta(n)}$ and the language is said to be *exponential* (see [20]). The class of exponential languages splits into two subclasses according whether the complement of a language is polynomial or not.

In this section, we have a closer look at numeration systems constructed over an exponential regular language such that its complement has a complexity function bounded by a polynomial. We show that for such systems, multiplication by a constant generally does not preserve recognizability.

We begin with the example of $\Sigma^* \setminus L$ where L is the polynomial language a^*b^* and $\Sigma = \{a, b\}$. Thus, with $S = (\Sigma^* \setminus L, \{a, b\}, a < b)$, we compute the representations of $2\,v_n$ and obtain Table 1 (for an algorithm of representation, see [13]).

In view of this table, it appears that the number of leading b's in the representation is increasing. Furthermore, it seems that the length of the tail also increases. Let us show that this observation is true and can be generalized.

Definition 17 Let $L \subset \Sigma^*$ and $x \in \Sigma^*$, we set $L_x = \{w \in L : w = xy\}$. Any confusion with the notation L_k where k is a state of a DFA is cleared by the context.

It is clear that $L_x \subseteq L$. So $\rho_{L_x}(n) \leq \rho_L(n)$ and ρ_{L_x} is $O(n^l)$ whenever ρ_L is $O(n^l)$.

Table 1
First terms of $2\,v_n$ for $S = (\{a,b\}^* \setminus a^*b^*, \{a,b\}, a < b)$.

n	$2\,v_n$	$\mathrm{r}_S(2\,v_n) = b^k\,aw$	k	$\lvert w\rvert$
1	0	$b\,a$	1	0
2	2	$b\,aa$	1	1
3	10	$b\,aab$	1	2
4	32	$b\,abab$	1	3
5	84	$bb\,aaaa$	2	3
6	198	$bb\,ababa$	2	4
7	438	$bbb\,aaabb$	3	4
8	932	$bbb\,abbabb$	3	5
9	1936	$bbbb\,abaaba$	4	5
10	3962	$bbbbb\,aabaaa$	5	5
11	8034	$bbbbb\,abbbbab$	5	6
12	16200	$bbbbbb\,abbaaab$	6	6

In our example, for $0 \le k < n$, we have

$$\rho_{(\Sigma^*\setminus L)_{b^{n-k}}}(n) = \rho_{\Sigma^*_{b^{n-k}}}(n) - \rho_{L_{b^{n-k}}}(n) = 2^k - 1.$$

The complexity function $\rho_{(\Sigma^*\setminus L)}(n)$ of the language associated to the system S is $2^n - n - 1$. So the sequence v_n associated to $\Sigma^* \setminus L$ is

$$v_n = \sum_{i=0}^{n} \rho_{(\Sigma^*\setminus L)}(i) = 2^{n+1} - \frac{n(n+3)}{2} - 2.$$

The words of $\mathrm{r}_S(\{v_n : n \in \mathbb{N}\})$ are the first words of each length in $\Sigma^* \setminus L$. So $\{v_n : n \in \mathbb{N}\}$ is S-recognizable. Recall that $|\mathrm{r}_S(x)| = n \Leftrightarrow v_{n-1} \le x < v_n$. For n large enough, it is obvious that $v_n \le 2\,v_n < v_{n+1}$. Then $|\mathrm{r}_S(2\,v_n)| = n + 1$.

Let us show that $\{2\,v_n : n \in \mathbb{N}\}$ is not S-recognizable. For each n there exists a unique i such that

$$\rho_{(\Sigma^*\setminus L)_{b^{n-i+2}}}(n+1) = 2^{i-1} - 1 < \underbrace{v_{n+1} - 2\,v_n}_{=n(n+1)/2} \le 2^i - 1 = \rho_{(\Sigma^*\setminus L)_{b^{n-i+1}}}(n+1).$$

Then $\mathrm{r}_S(2\,v_n) = b^{n-i+1}az$ with $|z| = i - 1$. Notice that, as a function of n, i is increasing but grows slower than n (in fact, it has a logarithmic growth). So $n - i \to +\infty$.

Assume that $\mathcal{L} = \mathrm{r}_S(\{2\,v_n : n \in \mathbb{N}\})$ is accepted by an automaton with q states. There exist n_0, i_0 and $t \ge 0$ such that $\mathrm{r}_S(2\,v_{n_0}) = b^{q+t}az_0$ with $|z_0| = i_0$.

By the pumping lemma (see for instance [21, Lemma 4.1]), there exists $\alpha > 0$ such that

$$\forall m \in \mathbb{N},\ b^{q+t+m\alpha} a z_0 \in \mathcal{L}.$$

In this last expression, z_0 has a constant length i_0 independent of m. A contradiction.

In view of this example, we state the following theorem. Recall that the complexity of any polynomial language is $\Theta(n^l)$ for some l.

Theorem 18 *Let $\Sigma = \{\sigma_1 < \cdots < \sigma_{s-1} < \beta\}$, $s \geq 2$ and $L \subset \Sigma^*$ be a regular language such that $\rho_L(n)$ is $\Theta(n^l)$. If $S = (\Sigma^* \setminus L, \Sigma, <)$ then there exists an S-recognizable set $X \subset \mathbb{N}$ such that for all $j \geq 1$, $s^j X$ is not S-recognizable.*

PROOF. For $0 \leq k < n$, we have

$$\rho_{(\Sigma^* \setminus L)_{\beta^n - k}}(n) = \rho_{\Sigma^*_{\beta^n - k}}(n) - \rho_{L_{\beta^n - k}}(n) = s^k - \underbrace{\rho_{L_{\beta^n - k}}(n)}_{\in O(n^l)}.$$

To avoid any misunderstanding, v_n is the sequence associated to the language $\Sigma^* \setminus L$ of the numeration S and $v_L(n)$ is related to L. So, $v_L(n) = \sum_{i=0}^{n} \rho_L(i)$ and

$$v_n = \sum_{i=0}^{n} \rho_{(\Sigma^* \setminus L)}(i) = \frac{s^{n+1} - 1}{s - 1} - v_n(L).$$

We take $X = \mathrm{r}_S(\{v_n : n \in \mathbb{N}\})$, a recognizable set. We have, for n sufficiently large and for all $j \geq 1$,

$$v_{n+j-1} \leq s^j v_n < v_{n+j}.$$

Indeed, $v_{n+j} - s^j\, v_n = s^j\, v_L(n) - v_L(n+j) + \frac{s^j - 1}{s-1}$. By Theorem 14, there exists $a > 0$ such that $v_L(n) \sim a\, n^{l+1}$ (let f and g be two functions, one writes $f \sim g$ if $\lim_{x \to \infty} \frac{f(x)}{g(x)} = 1$). So $v_{n+j} - s^j\, v_n \sim (s^j - 1) a\, n^{l+1}$. On the other hand, $s^j\, v_n - v_{n+j-1} = s^{n+j} + v_L(n+j-1) - s^j\, v_L(n) - \frac{s^j - 1}{s-1}$ has an exponential dominant term. Then $|\mathrm{r}_S(s^j\, v_n)| = n + j$.

For all n sufficiently large, there exists a unique i such that

$$\underbrace{\rho_{(\Sigma^* \setminus L)_{\beta^{n+j-i+1}}}(n+j)}_{= s^{i-1} - \rho_{L_{\beta^{n+j-i+1}}}(n+j)} < v_{n+j} - s^j\, v_n \leq \underbrace{\rho_{(\Sigma^* \setminus L)_{\beta^{n+j-i}}}(n+j)}_{= s^i - \rho_{L_{\beta^{n+j-i}}}(n+j)} \tag{9}$$

Then $\mathrm{r}_S(s^j\, v_n) = \beta^{n+j-i} \sigma z$ with $|z| = i - 1$ and $\sigma \neq \beta$. Notice that as a function of n, i is increasing and not bounded. To show that $n - i \to +\infty$ if

$n \to +\infty$: assume that $n - i$ is bounded, divide all members of (9) by s^n, let $n \to +\infty$ and obtain a contradiction.

Suppose that $\mathrm{r}_S(\{s^j X\})$ is accepted by an automaton with q states. There exist n_0, i_0 and $t \geq 0$ such that $\mathrm{r}_S(s^j v_{n_0}) = \beta^{q+t}\sigma z_0$ with $|z_0| = i_0$ and $\sigma \neq \beta$. Then using the pumping lemma [21, Lemma 4.1], we obtain a contradiction.

7 Relation with positional numeration systems

In this section, we give sufficient conditions to achieve the computation of an U-representation of an integer from its S-representation, where U is some positional numeration system related to a sequence of integers. In particular, we obtain sufficient conditions to guarantee the stability of the S-recognizability after addition and multiplication by a constant.

Recall that a strictly increasing sequence $U = (U_n)_{n\in\mathbb{N}}$ of integers such that $U_0 = 1$ and $\frac{U_{n+1}}{U_n}$ is bounded, defines a *positional numeration system* [9].

Definition 19 If $U = (U_n)_{n\in\mathbb{N}}$ is a positional numeration system and $w = w_n \cdots w_0$, a word over an alphabet $B \subset \mathbb{Z}$. We define the *numerical value* of w as

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

The word w is said to be an *U-representation* of $\pi_U(w)$. Notice that an integer can have more than one U-representation.

If x is an integer, the U-representation of x obtained by the greedy algorithm is denoted by $\mathrm{rep}_U(x)$ and belongs to A_U^* where $A_U = \{0, \ldots, Q\}$ is the *canonical alphabet* of the system U, $Q < \max \frac{U_{n+1}}{U_n}$, $Q \in \mathbb{N}$. A set $X \subset \mathbb{N}$ is said *U-recognizable* if $\mathrm{rep}_U(X)$ is regular. For any alphabet C of integers, one can define a partial function called *normalization* [10]

$$\nu_{U,C} : C^* \to A_U^* : z \mapsto \mathrm{rep}_U(\pi_U(z)).$$

Let us recall some definitions. A *2-tape automaton* over $A^* \times B^*$ (also called *transducer*) is a directed graph with edges labelled by elements of $A^* \times B^*$. The automaton is finite if the set of edges is finite. A 2-tape automaton is said *letter-to-letter* if the edges are labelled by elements of $A \times B$. A relation $\mathcal{R} \subset A^* \times B^*$ is said to be *computable by a finite 2-tape automaton* if there exists a finite 2-tape automaton over $A^* \times B^*$ such that the set of labels of paths starting in an initial state and ending in a final state is equal to $\mathcal{R}$.

Finally, a function is computable by a finite 2-tape automaton if its graph is computable by a finite 2-tape automaton.

Proposition 20 *Let $L \subset \Sigma^*$ be a regular language, $M = (K, s, F, \Sigma, \delta)$ be a DFA accepting L and $S = (L, \Sigma, <)$. Let $U = (U_n)_{n\in\mathbb{N}}$ be a sequence of integers such that $U_0 = 1$. If there exist $k, \alpha \in \mathbb{N} \setminus \{0\}$, $e_{p,i} \in \mathbb{Z}$ ($p \in K$, $i = 0, \ldots, k-1$) such that for all states $p \in K$ and all $n \in \mathbb{N}$*

$$\alpha\, u_{n+k-1}(p) = \sum_{i=0}^{k-1} e_{p,i}\, U_{n+i}. \tag{10}$$

Then there exist a finite alphabet $B \subset \mathbb{Z}$ and a finite letter-to-letter 2-tape automaton which compute a function $g : L \to B^*$ such that $|w| = |g(w)|$ and

$$\alpha\, \mathrm{val}_S(w) = \pi_U(g(w)).$$

Remark 21 The function g of the previous proposition is injective. If v and w are two words of L such that $g(v) = g(w)$ then $\mathrm{val}_S(v) = \mathrm{val}_S(w)$. So the conclusion, since val_S is a one-to-one correspondence.

PROOF. We consider words of length at least k. Indeed, there is only a finite number of words of length less than k and they can be treated separately. Let $w = w_{k+l} \cdots w_{k-1} w_{k-2} \cdots w_0$ be a word of L of length $k + l + 1$ with $l \geq -1$. To compute $\mathrm{val}_s(w)$, we apply Lemma 4 on the first $l + 2$ letters of w. A first application of the lemma gives

$$\mathrm{val}_s(w) = \mathrm{val}_{s.w_{k+l}}(w_{k+l-1} \cdots w_0) + v_{k+l}(s) - v_{k+l-1}(s.w_{k+l}) + \sum_{\sigma<w_{k+l}} u_{k+l}(s.\sigma).$$

Continuing this way, we obtain for $\mathrm{val}_s(w)$

$$\sum_{\sigma<w_{k+l}} u_{k+l}(s.\sigma) + \sum_{i=-1}^{l} u_{k+i}(s) + \sum_{i=-1}^{l-1} \sum_{\sigma<w_{k+i}} u_{k+i}(s.w_{k+l} \ldots w_{k+i+1}\sigma)$$
$$+\mathrm{val}_{s.w_{k+l}\ldots w_{k-1}}(w_{k-2} \ldots w_0) + v_{k-2}(s) - v_{k-2}(s.w_{k+l} \ldots w_{k-1}).$$

Recall that the notation $p.\sigma$ is written in place of $\delta(p, \sigma)$. We will denote by C_w the sum of the last three terms. For all $q \in K$, $p \in K \setminus \{s\}$ and $\sigma \in \Sigma$, let us define

$$\beta_{q,p,\sigma} = \#\{\sigma' < \sigma : q.\sigma' = p\}$$

and

$$\beta_{q,s,\sigma} = 1 + \#\{\sigma' < \sigma : q.\sigma' = s\}.$$

With these notations, we can rewrite $\mathrm{val}_s(w)$ as

$$C_w + \sum_{p\in K} \beta_{s,p,w_{k+l}}\, u_{k+l}(p) + \sum_{i=-1}^{l-1} \sum_{p\in K} \beta_{s.w_{k+l}\ldots w_{k+i-1},p,w_{k+i}}\, u_{k+i}(p).$$

Therefore, using (10), we have

$$\begin{aligned}\alpha\,\mathrm{val}_s(w) = \quad & \alpha\, C_w + \sum_{j=0}^{k-1} \underbrace{\sum_{p\in K} \beta_{s,p,w_{k+l}}\, e_{p,j}}_{=\lambda_{l,j}}\, U_{l+j+1} \\ & + \sum_{i=-1}^{l-1} \sum_{j=0}^{k-1} \underbrace{\sum_{p\in K} \beta_{s.w_{k+l}\ldots w_{k+i-1},p,w_{k+i}}\, e_{p,j}}_{=\lambda_{i,j}}\, U_{i+j+1}.\end{aligned}$$

It is obvious that the $\lambda_{i,j}$'s take their values in a finite set R. Therefore sums of $k-1$ elements of R also take their values in a finite set, say T. Notice that the $\lambda_{i,j}$'s (resp. the $\lambda_{l,j}$'s) are completely determined by the letter w_{k+i} (resp. w_{k+l}) and the state $s.w_{k+l}\ldots w_{k+i-1}$ reached after the reading of the first letters of w (resp. the state s). Therefore, we extend the notation $\lambda_{i,j}$ to a meaningful one:

$$\lambda_{q,\sigma,j} = \sum_{p\in K} \beta_{q,p,\sigma}\, e_{p,j} \tag{11}$$

with $q \in K$, $\sigma \in \Sigma$ and $j = 0, \ldots, k-1$.

We are now able to build a finite letter-to-letter 2-tape automaton $\mathcal{M}$ over $\Sigma^* \times B^*$ with $B \subset \mathbb{Z}$ some finite alphabet. The formula expressing $\alpha\mathrm{val}_s(w)$ can be interpreted in the following way. The reading of w_{k+i}, $l \le i \le -1$, provides the decomposition of $\alpha\,\mathrm{val}_s(w)$ with $\lambda_{i,k-1}\, U_{k+i}$; $\lambda_{i,k-2}\, U_{k+i-1}$; ...; $\lambda_{i,0}\, U_{i+1}$. The reading of w_{k+i} gives a coefficient $\lambda_{i,k-1}$ for U_{k+i}. The other $k-1$ coefficients can be viewed as "remainders". Roughly speaking, if we have already read the word $t = w_{k+l}\ldots w_{k+i+1}$ and if we are reading $\sigma = w_{k+i}$, then we have to consider the state $s.t.$ (Therefore it seems natural to mimic M in $\mathcal{M}$.) The coefficients $\lambda_{i,k-1}; \ldots; \lambda_{i,0}$ are nothing else but $\lambda_{s.t,\sigma,k-1}; \ldots; \lambda_{s.t,\sigma,0}$.

Thereby we can give a precise definition of $\mathcal{M}$. The set of states is $\mathcal{K} = K \cup \{f\} \times \underbrace{T \times \cdots \times T}_{k-1}$ where f does not belong to K and is the unique final state of $\mathcal{M}$. The copies of T will be used to store the "remainders". The start state is $(s, 0, \ldots, 0)$. The transition relation $\Delta : \mathcal{K} \times (\Sigma \times B) \to \mathcal{K}$ is defined as follows. If $p \in K$, $\sigma \in \Sigma$,

$$\begin{aligned}&\Delta((p,\gamma_{k-2},\dots,\gamma_0),(\sigma,\lambda_{p,\sigma,k-1}+\gamma_{k-2}))\\ =&(p.\sigma;\lambda_{p,\sigma,k-2}+\gamma_{k-3};\dots;\lambda_{p,\sigma,1}+\gamma_0;\lambda_{p,\sigma,0})\end{aligned}$$

These transitions compute an output $x_{k+l}\dots x_{k-1}$ from $w_{k+l}\dots w_{k-1}$. The alphabet B is finite since T is finite.

But we have still to read the last $k-1$ letters of w. For each state $p\in K$, $D_p = L_p\cap\Sigma^{k-1}$ is finite (recall that L_p are the words accepted from p). So, for each state $p\in K$ and each word $w_{k-2}\dots w_0\in D_p$, we construct an edge from $(p,\gamma_{k-2},\dots,\gamma_0)$ to f labelled by $(w_{k-2}\dots w_0,\gamma_{k-2}\cdots\gamma_1(\gamma_0+C_w))$. (This kind of edge can naturally be split in $k-1$ elementary edges using $k-2$ new states.) Indeed, notice that C_w is a constant which only depends on the state $s.w_{k+l}\dots w_{k-1}$ reached (the first component in $\mathcal{K}$) and the remaining word $w_{k-2}\dots w_0$.

Corollary 22 *Let $S=(L,\Sigma,<)$ and let the hypothesis and notations of Proposition 20 be satisfied. If the sequence U defines a positional numeration system such that the normalization function $\nu_{U,B}$ is computable by finite letter-to-letter 2-tape automaton then $X\subset\mathbb{N}$ is S-recognizable if and only if αX is U-recognizable.*

PROOF. Let the regular language $\mathcal{G}\subset(\Sigma\times B)^*\cap(L\times B^*)$ be the graph of the function g defined in Proposition 20. We denote by $p_1:\Sigma\times B\to\Sigma$ and $p_2:\Sigma\times B\to B$ the canonical projections. Let

$$Y = p_2[p_1^{-1}(\mathrm{r}_S(X))\cap\mathcal{G}].$$

If X is S-recognizable then $Y\subset B^*$ is regular and $\pi_U(Y)=\alpha X$. So αX is U-recognizable since $\nu_{U,B}(Y)$ is regular.

Conversely, if $\mathrm{rep}_U(\alpha X)$ is regular then $\nu_{U,B}^{-1}\circ\mathrm{rep}_U(\alpha X)$ is also regular. For each $y\in\alpha X$, $\nu_{U,B}^{-1}\circ\mathrm{rep}_U(y)$ may contain more than one element but only one is in $p_2(\mathcal{G})$. So the set

$$p_1\left(p_2^{-1}[\nu_{U,B}^{-1}\circ\mathrm{rep}_U(\alpha X)]\cap\mathcal{G}\right)$$

is regular and equal to $\mathrm{r}_S(X)$.

Corollary 23 *Let $S=(L,\Sigma,<)$ and let the hypothesis and notations of Proposition 20 be satisfied. If the sequence U satisfies a linear recurrence relation*

$$U_n = d_1U_{n-1}+\cdots+d_mU_{n-m}, d_i\in\mathbb{Z}, d_m\neq 0, n\geq m$$

such that its characteristic polynomial is the minimal polynomial of a Pisot number then $X\subset\mathbb{N}$ is S-recognizable if and only if X is U-recognizable.

PROOF. It is well known that for the system U the normalization $\nu_{U,C}$ is computable by a finite letter-to-letter 2-tape automaton for any alphabet C (see [11]). So by the previous corollary, X is S-recognizable if and only if αX is U-recognizable. Another well-known fact related to Pisot numeration systems is that a subset X is U-recognizable if and only if it is definable in the structure $\langle \mathbb{N}, +, V_U \rangle$ (see [4]). In particular, multiplication by a constant α is definable in $\langle \mathbb{N}, + \rangle$. So αX is definable in the structure if and only if X is definable.

Example 24 Consider the language $L \subset \{a, b, c\}^*$ of the words that do not contain aa. Its minimal automaton M_L is given in Figure 1. As usual, the start state is indicated by an unlabeled arrow and the final states by double circles. The sequences associated to the different states satisfy the relation

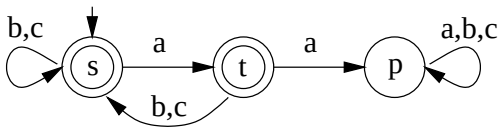

Fig. 1. The minimal automaton of L.

$$u_{n+2} = 2u_{n+1} + 2u_n, \forall n \in \mathbb{N}$$

with the initial conditions $u_0(s) = 1$, $u_1(s) = 3$, $u_0(t) = 1$, $u_1(t) = 2$, $u_0(p) = u_1(p) = 0$. The sequence U of Proposition 20 is given by $(u_n(s))_{n\in\mathbb{N}}$. For all $n \in \mathbb{N}$, we have the relations

$$\begin{cases} u_{n+1}(s) = 1\,u_{n+1}(s) + 0\,u_n(s) \Rightarrow e_{s,0} = 0,\ e_{s,1} = 1 \\ u_{n+1}(t) = 0\,u_{n+1}(s) + 2\,u_n(s) \Rightarrow e_{t,0} = 2,\ e_{t,1} = 0 \\ u_{n+1}(p) = 0\,u_{n+1}(s) + 0\,u_n(s) \Rightarrow e_{p,0} = 0,\ e_{p,1} = 0 \end{cases}$$

Notice that the characteristic polynomial of the recurrence relation of $u_n(s)$ is $x^2 - 2x - 2 = (x - 1 + \sqrt{3})(x - 1 - \sqrt{3})$. So $U = (u_n(s))_{n\in\mathbb{N}}$ is a positional numeration system associated to the Pisot number $1 + \sqrt{3}$. From M_L, we compute the 3×3 matrices $B_\sigma = (\beta_{q,r,\sigma})_{q,r=s,t,p}$, $\sigma \in \Sigma$:

$$B_a = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, B_b = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}, B_c = \begin{pmatrix} 2 & 1 & 0 \\ 2 & 0 & 0 \\ 1 & 1 & 2 \end{pmatrix}$$

If $E = (e_{q,i})_{q=s,t,p;i=0,1}$ then it follows from (11) that $(B_\sigma E)_{q,i} = \lambda_{q,\sigma,i}$. We have

$$B_a E = \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, B_b E = \begin{pmatrix} 2 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, B_c E = \begin{pmatrix} 2 & 2 \\ 0 & 2 \\ 2 & 1 \end{pmatrix}$$

To obtain the complete transducer, with the notations of the proof of Proposition 20, we have to compute the C_w namely

$$C_{q,\sigma} = \mathrm{val}_q(\sigma) + v_0(s) - v_0(q)$$

for q and σ such that $q.\sigma \in F$. Finally we have in Figure 2 the finite letter-to-letter automaton built from M_L and the $\lambda_{q,\sigma,i}$'s.

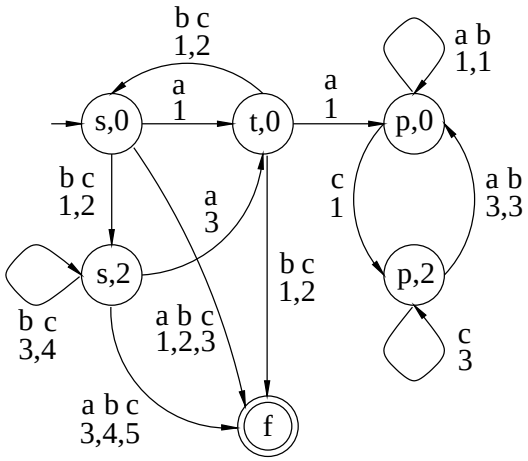

Fig. 2. The transducer computing g.

We can do the same construction for the language $L' = a^+\{a,b\}^*$. Its minimal automaton $M_{L'}$ is given in Figure 3. The sequence U of Proposition 20 is

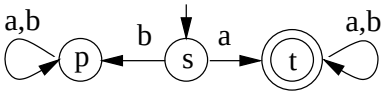

Fig. 3. The minimal automaton of $L' = a^+\{a,b\}^*$.

given by $u_n(t) = 2^n$. So here, the Pisot number involved is 2 and it is multiplicatively independent from $1+\sqrt{3}$. So from [15], the only subsets which are simultaneously recognizable in $(L, \{a,b,c\}, a < b < c)$ and $(L', \{a,b\}, a < b)$ are the arithmetic progressions.

Remark 25 Let $J = a\{a,b\}^* \cup \{a,b\}^*bb\{a,b\}^*$. Notice that J is an exponential language with exponential complement. Its minimal automaton M_J is given in Figure 4. We consider the numeration system $S = (J, \{a,b\}, a < b)$ and we show that

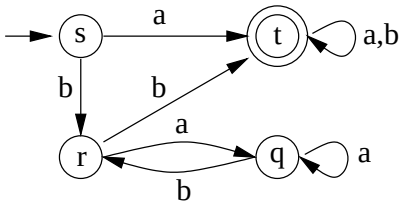

Fig. 4. The minimal automaton of $J = a\{a,b\}^* \cup \{a,b\}^*bb\{a,b\}^*$.

i) there exists no linear recurrent sequence associated to a Pisot number such that the condition (10) of proposition 20 is satisfied for all states of M_J
ii) the set $X = \{v_n(s) : n \in \mathbb{N}\}$ is S-recognizable but $2X$ is not.

One can check that for all $n \geq 1$, $u_n(t) = 2^n$ and

$$u_n(s) = 2^n - \frac{\sqrt{5}}{5}\left(\frac{1+\sqrt{5}}{2}\right)^n + \frac{\sqrt{5}}{5}\left(\frac{1-\sqrt{5}}{2}\right)^n.$$

So i) holds. To check ii), we use the same technique as in Theorem 18. One can verify that

$$v_{n+1}(s) - 2v_n(s) = 1 - \frac{\sqrt{5}}{5}\left(\frac{1-\sqrt{5}}{2}\right)^n + \frac{\sqrt{5}}{5}\left(\frac{1+\sqrt{5}}{2}\right)^n$$

has an exponential dominant term. Furthermore, for all n large enough there exists i such that

$$\rho_{J_{b^{i+1}}}(n) = 2^{n-i-1} < v_{n+1}(s) - 2v_n(s) \leq 2^{n-i} = \rho_{J_{b^i}}(n)$$

and $n-i \to +\infty$ if $n \to +\infty$. One can conclude as in Theorem 18; $\mathrm{r}_S(2\,v_n(s)) = b^iaz$ with $|z| = n-i-1$.

8 Recognizable formal power series

This last section is independent from the results obtained in the other sections. Here, we characterize the S-recognizable subsets of $\mathbb{N}$ in terms of $\mathbb{N}$-rational series. First, let us recall some basic definitions about rational and recognizable power series.

Let R be a semiring, a formal power series is an application $T : \Sigma^* \to R$. It can be written as a formal sum

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

the (T,w)'s are the *coefficients* of T, $(T,w) = T(w) \in R$. We mainly adopt the terminology of [1] concerning semirings, rational and recognizable series. The

reader can also see [18]. The set of applications $T : \Sigma^* \to R$ is denoted $R\langle\langle\Sigma\rangle\rangle$. One can endow $R\langle\langle\Sigma\rangle\rangle$ with the operations of sum and Cauchy product. For $T_1, T_2 \in R\langle\langle\Sigma\rangle\rangle$, the *sum* $T_1 + T_2$ is defined by $T_1 + T_2(w) = (T_1, w) + (T_2, w)$ and the *(Cauchy) product* $(T_1T_2, w) = \sum_{xy=w}(T_1, x)(T_2, y)$. These operations induces a semiring structure to power series.

Definition 26 A sequence $(T_n)_{n\in\mathbb{N}}$ of elements of $R\langle\langle\Sigma\rangle\rangle$ *converges* to the limit T if for all k there exists m such that

$$|w| \leq k, j > m \Rightarrow (T_j, w) = (T, w).$$

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

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

exists. This limit is called the *quasi inverse* of T.

Definition 27 A subsemiring of $R\langle\langle\Sigma\rangle\rangle$ is *rationally closed* iff it contains the quasi inverse of every quasi regular element. The family of *R-rational series* over Σ is the smallest rationally closed subset of $R\langle\langle\Sigma\rangle\rangle$ which contains all polynomials.

As a consequence of this latter definition, any R-rational series can be obtained from polynomials by a finite number of applications of sum, Cauchy product and quasi inversion.

Definition 28 A series $T \in R\langle\langle\Sigma\rangle\rangle$ is *R-recognizable* if there exist $n \in \mathbb{N}\backslash\{0\}$, a morphism $\mu : \Sigma^* \to R^{n\times n}$ and two matrices $\lambda \in R^{1\times n}$, $\gamma \in R^{n\times 1}$ such that for all $w \in \Sigma^*$

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

In that case, (λ, μ, γ) is said to be a *linear representation* of T.

According to the celebrated *Schützenberger's Representation Theorem* the class of R-rational and R-recognizable formal power series coincide.

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

$$u^{-1}T = \sum_{w\in\Sigma^*} (T, uw)\, w.$$

In other words, $(u^{-1}T, w) = (T, uw)$.

It is shown in [1] that the series $\sum_{w\in X^*} \pi_2(w)\, w \in \mathbb{N}\langle\langle x\rangle\rangle$ is $\mathbb{N}$-rational. In the last expression, X is the alphabet $\{x_0, x_1\}$ and if $w = x_{i_k} \cdots x_{i_0}$ then $\pi_2(w) = 2^k\, i_k + \cdots + 2\, i_1 + i_0$ is the numerical value in base two of w.

Here, we obtain the same result for any numeration system on a regular language. Another proof of this result can be found in [6] where complexity problems are discussed.

Proposition 29 *Let $S = (L, \Sigma, <)$ be a numeration system. The formal series*

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

is $\mathbb{N}$-recognizable.

PROOF. Let $M_L = (K, s, F, \Sigma, \delta)$ be the minimal automaton of L. For $k, l \in K$, $\sigma \in \Sigma$, we introduce the following series in $\mathbb{N}\langle\langle \Sigma \rangle\rangle$

$$\begin{aligned}
T_k &= \sum_{w \in L_k, w \neq \varepsilon} [\mathrm{val}_k(w) - v_{|w|-1}(k)]\, w \\
U_{l,k} &= \sum_{w \in L_l, w \neq \varepsilon} u_{|w|}(k)\, w \\
U'_{l,k} &= \sum_{w \in L_l} u_{|w|}(k)\, w \\
V_{l,k} &= \sum_{w \in L_l, w \neq \varepsilon} v_{|w|-1}(k)\, w
\end{aligned}$$

$$W_{k,\sigma} = \begin{cases} [\mathrm{val}_k(\sigma) - v_0(k)]\, \varepsilon \text{ if } \sigma \in L_k \\ 0 \qquad\qquad\qquad \text{otherwise.} \end{cases}$$

If $k, l \in K$, $\alpha, \sigma \in \Sigma$, then we have the following relations

$$i)\ \sigma^{-1} T_k = T_{k.\sigma} + \sum_{\sigma' < \sigma} U_{k.\sigma, k.\sigma'} + W_{k,\sigma}$$

$$ii)\ \ \sigma^{-1} U_{l,k} = \sum_{\alpha \in \Sigma} U'_{l.\sigma, k.\alpha} \qquad iii)\ \sigma^{-1} U'_{l,k} = \sum_{\alpha \in \Sigma} U'_{l.\sigma, k.\alpha}$$

$$iv)\ \ \sigma^{-1} V_{l,k} = V_{l.\sigma,k} + U'_{l.\sigma,k} \qquad v)\ \sigma^{-1} W_{k,\alpha} = 0.$$

To check relation $i)$, one has to compute $(T_k, \sigma w)$. Notice that $\sigma w \in L_k$ iff $w \in L_{k.\sigma}$. Use Lemma 4 and treat the case $w = \varepsilon$ separately.

For relations ii) and iii), if σw belongs to L_l then $w \in L_{l.\sigma}$ and

$$(U_{l,k}, \sigma w) = u_{|w|+1}(k) = \sum_{\alpha \in \Sigma} u_{|w|}(k.\alpha).$$

In iv), one observes that $v_{|w|}(k) = v_{|w|-1}(k) + u_{|w|}(k)$. Relation v) is immediate.

Therefore the submodule $\mathcal{R}$ of $\mathbb{N}\langle\langle\Sigma\rangle\rangle$ finitely generated by the series T_k's, $U_{l,k}$'s, $U'_{l,k}$'s, $V_{l,k}$'s, $W_{k,\sigma}$'s is stable for the operation $T \mapsto \sigma^{-1}T$, $\sigma \in \Sigma$. By associativity of the operation $T \mapsto w^{-1}T$, this module is stable. By [1, Prop. 1, p. 18], the series of $\mathcal{R}$ are $\mathbb{N}$-recognizable.

To conclude the proof, notice that

$$T_k + V_{k,k} = \sum_{w \in L_k, w \neq \varepsilon} \mathrm{val}_k(w)\, w = \sum_{w \in L_k} \mathrm{val}_k(w)\, w.$$

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

Example 30 We consider the numeration system $S = (a^*b^*, \{a, b\}, a < b)$. We obtain a linear representation (λ, μ, γ) of $\mathcal{F}_S$:

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

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

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

Inspired by the definition of U-automata given in [4], we have the following characterization of the regular subsets of a regular language. The proof of this result is omitted. It is no more difficult than an exercise in automata theory. (See Section III.5 of [8] for the properties of the minimal automaton).

Lemma 31 *Let $L \subset \Sigma^*$ be a regular language and $M_L = (Q_L, s_L, F_L, \Sigma, \delta_L)$ be its minimal automaton. If $M_K = (Q_K, s_K, F_K, \Sigma, \delta_K)$ is the minimal automaton of a regular language $K \subset L$ then there exists a morphism h of automata between M_K and M_L defined as follows*

$$h : Q_K \to Q_L,$$

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

With this lemma, we can generalize Proposition 29 and obtain a characterization of the S-recognizable sets.

Theorem 32 *Let $S = (L, \Sigma, <)$ be a numeration system, a set $X \subseteq \mathbb{N}$ is S-recognizable if and only if the formal series*

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

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

PROOF. The condition is sufficient. The support of a recognizable series belonging to $\mathbb{N}\langle\langle \Sigma \rangle\rangle$ is a regular language [1, Lemme 2, p. 49].

The condition is necessary. By Lemma 31, one has a morphism $h : M_X \to M_L$ where M_X (resp. M_L) is the minimal automaton of $\mathrm{r}_S(X)$ (resp. L). We proceed as in the proof of Proposition 29. Let Q_X be the set of states of M_X; for $k, l \in K$, $\sigma \in \Sigma$, we introduce the following series

$$\begin{aligned}
T_k &= \sum_{w \in L_k, w \neq \varepsilon} [\mathrm{val}_{h(k)}(w) - v_{|w|-1}(h(k))]\, w \\
U_{l,k} &= \sum_{w \in L_l, w \neq \varepsilon} u_{|w|}(h(k))\, w \\
U'_{l,k} &= \sum_{w \in L_l} u_{|w|}(h(k))\, w \\
V_{l,k} &= \sum_{w \in L_l, w \neq \varepsilon} v_{|w|-1}(h(k))\, w
\end{aligned}$$

$$W_{k,\sigma} = \begin{cases} [\mathrm{val}_{h(k)}(\sigma) - v_0(h(k))]\, \varepsilon & \text{if } \sigma \in L_k \\ 0 & \text{otherwise.} \end{cases}$$

We conclude as in Proposition 29.

In [13], it is shown that for any numeration system S, arithmetic progressions are always S-recognizable. Using formal series, we can obtain a generalization of this result. Here, the language L is not necessary lexicographically ordered.

Proposition 33 *Let $L \subset \Sigma^*$ be an infinite regular language and $\alpha : L \to \mathbb{N}$ be a one-to-one correspondence. If*

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

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

PROOF. After division p can be written as $a\,q + r$ with $0 \leq r < q$. Let us first assume that $r = 0$. Consider the congruence of the semiring $\langle \mathbb{N}, +, \cdot, 0, 1 \rangle$ defined by $n \sim n + q$. We denote by $\mathcal{N}$ the finite semiring $\mathbb{N}/\sim$ and by φ the canonical morphism $\varphi : \mathbb{N} \to \mathcal{N}$. The characteristic series of L, $\underline{L} = \sum_{w \in L} w$, is recognizable (see [1, Prop. 1, p. 51]). So

$$U = \varphi(T + \underline{L}) = \sum_{w \in L} \varphi(\alpha(w) + 1)\, w \in \mathcal{N}\langle\langle \Sigma \rangle\rangle$$

is rational (see [1, Lemme 1, p. 49]). Since $\mathcal{N}$ is finite and U is rational, the set

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

is a regular language (see [1, Prop. 2, p. 52]). To conclude this first part, observe that if $a \geq 1$ then

$$\alpha^{-1}(p + \mathbb{N}\, q) = \alpha^{-1}(\mathbb{N}\, q) \setminus \alpha^{-1}(\{n\, q : 0 \leq n < a\}).$$

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

If $r \neq 0$ then consider the series $U = \varphi(T)$ and

$$U^{-1}(\{\varphi(p)\}) = \alpha^{-1}(p + \mathbb{N}\, q) \cup \alpha^{-1}(\{n\, q + r : 0 \leq n < a\}).$$

We conclude as in the previous case.

Corollary 34 *Arithmetic progressions are S-recognizable for any numeration system S.*

PROOF. This is a direct consequence of Propositions 29 and 33.

9 Acknowledgments

The author warmly thanks P. Mathonet for fruitful discussions in the polynomial case, J.-P. Allouche for advice and also P. Lecomte for his support and improvements in many proofs. The author is grateful to the referees for advice in improving the presentation of this paper and for a simplification in the proof of Theorem 14.

References

[1] J. Berstel, C. Reutenauer, *Les séries rationnelles et leurs languages*, études et recherches en informatique, Masson, 1984.

[2] A. Bertoni, D. Bruschi, M. Goldwurm, Ranking and formal power series, Algebraic and computing treatment of noncommutative power series (Lille, 1988), *Theoret. Comput. Sci.* **79** (1991) 25–35.

[3] C. B. Boyer, Pascal's Formula for the Sums of the Powers of the Integers, *Scripta Math.* **9** (1943) 237–244.

[4] V. Bruyère, G. Hansel, Bertrand numeration systems and recognizability, *Theoret. Comput. Sci.* **181** (1997) 17–43.

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

[6] C. Choffrut, W. Goldwurm, Rational transductions and complexity of counting problems, *Math. Systems Theory* **28** (1995) 437–450.

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

[8] S. Eilenberg, *Automata, Languages and Machines*, Vol. A, Academic Press, New York, 1974.

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

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

[11] C. Frougny, B. Solomyak, On representation of integers in linear numeration systems, in *Ergodic theory of Z^d actions* (Warwick, 1993–1994), 345–368, London Math. Soc. Lecture Note Ser. **228**, Cambridge Univ. Press, Cambridge, 1996.

[12] M. Hollander, Greedy numeration systems and regularity, *Theory Comput. Syst.* **31** (1998) 111–133.

[13] P. B. A. Lecomte, M. Rigo, Numeration systems on a regular language, *Theory of Comput. Systems.* **34** (2001) 27–44.

[14] N. Loraud, β-shift, systèmes de numération et automates, *J. Théor. Nombres Bordeaux* **7** (1995) 473–498.

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

[16] M. Rigo, Generalization of automatic sequences for numeration systems on a regular language, *Theor. Comput. Sci.* **244** (2000) 271–281.

[17] M. Rigo, Construction of regular languages and recognizability of polynomials, *preprint* (1999), `http://xxx.lanl.gov/abs/cs.CC/9908018`.

[18] A. Salomaa, Formal Languages and Power Series, *Handbook of theoretical computer science*, Vol. B, Elsevier, Amsterdam, (1990), 103–132.

[19] J. Shallit, Numeration systems, linear recurrences, and regular sets, *Inform. and Comput.*, **113** (1994) 331–347.

[20] A. Szilard, S. Yu, K. Zhang, J. Shallit, Characterizing regular languages with polynomial densities, *Proceedings of the 17th International Symposium on Mathematical Foundations of Computer Science*, *Lect. Notes in Comp. Sci.* **629** (1992) 494–503.

[21] S. Yu, Regular languages, *Handbook of formal languages*, Vol. 1, 41–110, Springer, Berlin, 1997.