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
January 17, 2000

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 rational
formal series. We also 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 and \( U \)-recognizability to be equivalent,
where \( U \) is some positional numeration system related to a sequence
of integers.

1 Introduction

According to [9], a \textit{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 \( L \). A subset \( X \subset \mathbb{N} \) is called
\( S \)-\textit{recognizable} if \( r_S(X) \) is a regular subset of \( L \).

We first characterize the \( S \)-recognizable subsets of \( \mathbb{N} \) in terms of rational
series in the noncommuting variables \( \sigma \in \Sigma \) and with coefficients in \( \mathbb{N} \). In
particular, we show that \( \sum_{n \in \mathbb{N}} n r_S(n) \) is rational (this kind of result is also
discussed in [2,3]). Using classical results about rational series, we obtain a generalization of the fact given in [9] that ultimately periodic sets are $S$-recognizable for any numeration system $S$.

Our main purpose is related to 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 $\lambda$ preserve the recognizability?

It is well known that for positional numeration systems in 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 [4,5]). It is obvious that addition and multiplication by a constant are definable in the Presburger arithmetic. Therefore, $p$-recognizability is preserved.

On the other hand, using the specific structure of the language $a^*b^*$, it is shown in [9] that for the numeration system $S = (a^*b^*, \{a, b\}, a < b)$, the multiplication by a non-negative integer $\lambda$ transforms the $S$-recognizable sets into $S$-recognizable sets if and only if $\lambda$ is a perfect square. Then 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 $L$). So, it is natural to check whether a numeration system on a polynomial language preserves the recognizability of a set after multiplication by a constant. For $a^*b^*$, perfect squares play a special role. Does there exist a similar set for an arbitrary language in $\Theta(n^l)$? We get the following result: if $S$ is a numeration system built on a regular language with complexity in $\Theta(n^l)$ then the multiplication by $\lambda$ preserves the recognizability only if $\lambda = \beta^{l+1}$ for some integer $\beta$. As a consequence, the addition cannot be a regular map for numeration systems on polynomial regular languages.

In order to prove this, we proceed in two steps. In section 4, we assume that the complexity of the language is a polynomial of degree $l$ with rational coefficients. With such a language, we exhibit a subset $X$ which is recognizable and we prove that $\lambda X$ is not recognizable for any $\lambda \in \mathbb{N} \setminus \{n^{l+1} : n \in \mathbb{N}\}$. In section 5, we consider the general case.

In this study of polynomial regular languages, we have obtained an interesting result about a special sequence associated to a language. We denote by $v_L(n)$, or simply $v_n$ if the context is clear, the number of words of length not exceeding $n$ belonging to $L$. In section 6, we show that if 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. It is surprising to notice that, in contrast, the sequence $(\rho_L(n)/n^l)_{n \in \mathbb{N}}$ generally does not converge.

The end of this paper is mainly related to exponential languages. In section $3$, we consider numeration systems on the complement of a polynomial language. As in the polynomial case, we find a recognizable set $X$ and constants $\lambda$ such that $\lambda X$ is not recognizable. Here, the $\lambda$'s are powers of the cardinality of the alphabet.

In the last section, we study relations between some positional numeration system $U$ and a system $S$ on a regular language $L$. We give sufficient conditions for the equivalence of $S$-recognizability and $U$-recognizability. These conditions are strongly dependent on the language $L$ and the recognizability of the normalization in $U$. Using these conditions, we give two examples of numeration systems on an exponential language such that addition and multiplication by a constant preserve $S$-recognizability.

2 Basic definitions and notations

We denote by $\Sigma^*$ the free monoid (with identity $\varepsilon$) generated by $\Sigma$. For a set $S$, $\#S$ is the cardinality of $S$ and for a string $w \in \Sigma^*$, $|w|$ is the length of $w$.

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 $\delta : K \times \Sigma \rightarrow K$ is the transition function. We often write $k.\sigma$ instead of $\delta(k, \sigma)$. Recall that the elements of $K$ are the derivatives $[7, III.5]

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

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 $L$. 

3
The lexicographic ordering can be used to compare words of different length. Let $x$ and $y$ be two words. 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 1** A numeration system is a triple $(L, \Sigma, <)$ where $L$ is an infinite regular language over a totally ordered finite alphabet $(\Sigma, <)$ (see [9]).

The lexicographic ordering of $L$ gives a one-to-one correspondence between the set $\mathbb{N}$ of natural numbers and the language $L$.

For each $n \in \mathbb{N}$, $r_S(n)$ 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 $\text{val}_S(w) = r^{-1}_S(w)$ and we call it the numerical value of $w$.

The mappings $\text{val}_S$ and $r_S$ are sometimes called ranking and unranking in the literature.

This way of representing integers generalizes linear numeration systems in which $\mathbb{N}$ is recognizable by finite automata. 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 2** Let $S$ be a numeration system. A subset $X$ of $\mathbb{N}$ is $S$-recognizable if $r_S(X)$ 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 $\text{val}_{S_k}$ are simply denoted $r_k$ and $\text{val}_k$ if the context is clear. If $L_k$ is finite, the applications $r_k$ and $\text{val}_k$ 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 3** Let $S = (L, \Sigma, <)$ and $M = (K, s, F, \Sigma)$ be a DFA accepting $L$. If $\sigma w$ belongs to $L_k$, $k \in K$, $\sigma \in \Sigma$, $w \in \Sigma^+$, then

$$\text{val}_k(\sigma w) = \text{val}_{k, \sigma}(w) + v_{|w|}(k) - v_{|w|-1}(k, \sigma) + \sum_{\sigma' < \sigma} u_{|w|}(k, \sigma').$$
3 Recognizable formal power series

Let \( R \) be a semiring, a formal power series \( T : \Sigma^* \to R \) can be written as a formal sum

\[
T = \sum_{w \in \Sigma^*} (T, w) w.
\]

We mainly adopt the terminology of [1] concerning semirings, rational and recognizable series. 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 \( \sum_{w \in X^*} \pi_2(w) w \in \mathbb{N}\langle\langle \Sigma \rangle\rangle \) is rational. In the last expression, \( X \) is the alphabet \( \{x_0, x_1\} \) and if \( w = x_{i_k} \cdots x_{i_0} \) then \( \pi_2(w) = 2^{i_k} + \cdots + 2 i_1 + i_0 \) 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 [8] where complexity problems are discussed.

**Proposition 4** Let \( S = (L, \Sigma, <) \) be a numeration system. The formal series

\[
\mathcal{F}_S = \sum_{w \in L} \text{val}_S(w) w \in \mathbb{N}\langle\langle \Sigma \rangle\rangle
\]

is 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 of \( \mathbb{N}\langle\langle \Sigma \rangle\rangle \)

\[
T_k = \sum_{w \in L_k, w \neq \varepsilon} [\text{val}_k(w) - v_{|w|-1}(k)] w
\]

\[
U_{l,k} = \sum_{w \in L_l, w \neq \varepsilon} u_{|w|}(k) w
\]

\[
U'_{l,k} = \sum_{w \in L_l} u_{|w|}(k) w
\]

\[
V_{l,k} = \sum_{w \in L_l, w \neq \varepsilon} v_{|w|-1}(k) w
\]
\[ W_{k,\sigma} = \begin{cases} \lfloor \text{val}_k(\sigma) - v_0(k) \rfloor, & \text{if } \sigma \in L_k \\ 0, & \text{otherwise.} \end{cases} \]

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

\[
\begin{align*}
\text{i)} & \quad \sigma^{-1}T_k = T_{k,\sigma} + \sum_{\sigma' < \sigma} U_{k,\sigma,k,\sigma'} + W_{k,\sigma} \\
\text{ii)} & \quad \sigma^{-1}U_{l,k} = \sum_{\alpha \in \Sigma} U'_{l,\sigma,k,\alpha} \\
\text{iii)} & \quad \sigma^{-1}U'_{l,k} = \sum_{\alpha \in \Sigma} U'_{l,\sigma,k,\alpha} \\
\text{iv)} & \quad \sigma^{-1}V_{l,k} = V_{l,\sigma,k} + U'_{l,\sigma,k} \\
\text{v)} & \quad \sigma^{-1}W_{k,\alpha} = 0.
\end{align*}
\]

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 3 and treat the case \( w = \varepsilon \) separately.

For relations ii) and iii), if \( \sigma w \) 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 \( \mathbb{N}[\langle \Sigma \rangle] \) 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 \([1, \text{Prop. 1, p. 18}]\), the series of \( \mathcal{R} \) are recognizable.

To conclude the proof, notice that

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

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

**Example 1** We consider the numeration system \( S = (a^*b^*, \{a, b\}, a < b) \).

We obtain a linear representation \((\lambda, \mu, \gamma)\) for \( \mathcal{F}_S \):

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

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

\[ \text{val}_S(w) = \lambda \mu(w) \gamma. \]
Inspired by the definition of $U$-automata given in [4], we have the following characterization of the regular subsets of a regular language.

**Lemma 5** 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

$$h : Q_K \to Q_L,$$

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

**Proof.** A state of $M_K$ is a derivative of $K$ of the form

$$u^{-1}.K = \{ v \in \Sigma^* : uv \in K \}.$$

Since $K \subset L$, then $u^{-1}.K \subset u^{-1}.L$. We consider the morphism $h : Q_K \to Q_L$ defined by $h(q) = u^{-1}.L$ if $q = u^{-1}.K$ for some $u$. We can verify the properties of $h$ using the definition of the minimal automaton [3, III.5],

1. $\delta_K(q, \sigma) = \sigma^{-1}.q = \sigma^{-1}.(u^{-1}.K)$ for some $u \in \Sigma^*$, $q \in Q_K$, $\sigma \in \Sigma$. So $\delta_K(q, \sigma) = (u\sigma)^{-1}.K$ and $(u\sigma)^{-1}.L = \sigma^{-1}.h(q) = \delta_L(h(q), \sigma)$.
2. $s_K = \varepsilon^{-1}.K$ and $s_L = \varepsilon^{-1}.L$.
3. A state $q = u^{-1}.K$ belongs to $F_K$ if $\varepsilon \in u^{-1}.K$ therefore $\varepsilon \in u^{-1}.L$ and $h(q) = u^{-1}.L \in F_L$.

$\square$

With this lemma, we can generalize Proposition 4 and obtain a characterization of the $S$-recognizable sets.

**Theorem 6** Let $S = (L, \Sigma, <)$ be a numeration system, a set $X \subseteq \mathbb{N}$ is $S$-recognizable if and only if the formal series

$$\sum_{w \in \text{rs}(X)} \text{val}_S(w) w \in \mathbb{N} \langle \langle \Sigma \rangle \rangle$$

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

The condition is necessary. By Lemma 3, one has a morphism $h : M_X \rightarrow M_L$ where $M_X$ (resp. $M_L$) is the minimal automaton of $r_S(X)$ (resp. $L$). We proceed as in the proof of Proposition 4. Let $K$ be the set of states of $M_X$; for $k, l \in K$, $\sigma \in \Sigma$, we introduce the following series

$$T_k = \sum_{w \in L_k, w \neq \varepsilon} [\text{val}_{h(k)}(w) - v_{|w|-1}(h(k))] w$$

$$U_{l,k} = \sum_{w \in L_l, w \neq \varepsilon} u_{|w|}(h(k)) w$$

$$U'_{l,k} = \sum_{w \in L_l} u_{|w|}(h(k)) w$$

$$V_{l,k} = \sum_{w \in L_l, w \neq \varepsilon} v_{|w|-1}(h(k)) w$$

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

We conclude as in Proposition 4. □

In [9], 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 7** Let $L \subset \Sigma^*$ be an infinite regular language and $\alpha : L \rightarrow \mathbb{N}$ be a one-to-one correspondence. If

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

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

Proof. Assume $p = 0$. 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 $\varphi$ the canonical morphism $\varphi : \mathbb{N} \rightarrow \mathcal{N}$. The characteristic series of $L$, $L = \sum_{w \in L} w$, is recognizable (see [1, Prop. 1, p. 51]). So

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

8
is rational (see [1, Lemme 1, p. 49]). Since $\mathcal{N}$ is finite and $U$ is rational, the set

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

is a regular language (see [1, Prop. 2, p. 52]).

If $p \neq 0$ and $p < q$, then consider the series $U = \varphi(T)$ and the set $U^{-1}(\{\varphi(p)\})$. □

**Corollary 8** Arithmetic progressions are $S$-recognizable for any numeration system $S$.

**Proof.** This is a direct consequence of Propositions 4 and 7. □

**Remark 1** One can easily characterize the congruences $\sim$ of the semiring $(\mathbb{N}, +, 0)$ with finite index $q > 1$. The canonical morphism is denoted by $\varphi$.

First notice that $\varphi(0) \neq \varphi(1)$. Since $\mathbb{N}/\sim$ is finite, there exist $x, y \in \mathbb{N}$ such that $x + y \sim x$. Let

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

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

By definition of $x_0$ and $y_0$, if $z < x_0$ then $\varphi^{-1}\varphi(z) = \{z\}$

Therefore the congruences of $\mathbb{N}$ with finite index are generated by the relation $n \sim n + y_0$ for $n$ sufficiently large. So we cannot refine Proposition 7 with the same kind of proof because it uses explicitly the finiteness of $\mathbb{N}/\sim$.

**4 Multiplication for exact polynomial languages**

In [9], we proved that for the numeration system $S = (a^*b^*, \{a, b\}, a < b)$, the multiplication by a non-negative integer $\lambda$ transforms the $S$-recognizable sets into $S$-recognizable sets if and only if $\lambda$ is a perfect square.

In this section, we study the family of regular languages with polynomial complexity function. This step contains the main ideas leading to the case of an arbitrary polynomial language (i.e. a language with complexity function bounded by a polynomial). But it is simpler to handle since we only deal with polynomials.
Lemma 9 Let $f : \mathbb{N} \to \mathbb{N}$ be a strictly increasing function such that $f(\mathbb{N})$ is a finite union of arithmetic progressions (i.e. there exist $y_0$ and $\Gamma$ such that $\forall y \geq y_0, y \in f(\mathbb{N}) \Leftrightarrow y + \Gamma \in f(\mathbb{N})$).

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

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

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

$$f(x_0 + k) = f(f^{-1}(y_0) + k) = f(f^{-1}(y_0 + \Gamma)) = f(x_0) + \Gamma.$$ 

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

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

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

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

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

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

 Definition 10 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, we assume that we deal with “true” complexity functions, i.e. if $\rho_L$ 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 11 If $H$ is a polynomial such that $\forall n \in \mathbb{N} \setminus \{0\}$, $H(n) \in \mathbb{Z}$ then $H(\mathbb{Z}) \subseteq \mathbb{Z}$. 

10
Proof. We proceed by induction on the degree of $H$. If $H$ is a polynomial of degree one then one has $H(n) = an + b$ with $a, b \in \mathbb{Z}$ and $H(\mathbb{Z}) \subseteq \mathbb{Z}$.

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

**Theorem 12** 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 $\Sigma$ 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 $r_S(X)$ is regular and that $r_S(\lambda X)$ 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

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

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 $\rho_L$ to $\rho_L : \mathbb{R} \to \mathbb{R} : x \mapsto a_l x^l + \cdots + a_0$. By Lemma [1], $P(1) = \rho_L(0) = a_0 \in \mathbb{Z}$. 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 [1]).

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

$$|r_S(x)| = n \Leftrightarrow x \in [P(n) - a_0 + 1, P(n + 1) - a_0].$$

(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 \( r_S(P(\mathbb{N})) \) is a translation of the set \( \mathcal{I}(L, <) \) of the first words of each length. Therefore \( X = P(\mathbb{N}) \) is \( S \)-recognizable, see [4, 13].

Let \( \lambda \in \mathbb{N} \setminus \{0, 1\} \). Our aim is to show that \( \lambda P(\mathbb{N}) \) is not \( S \)-recognizable.

For \( n \) large enough, we first show that

\[
  n \leq |r_S(\lambda P(n))| < \lambda^{1/l} n.
\]

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

\[
  \lambda P(n) < P(\lambda^{1/l} n) - a_0 + 1.
\]

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,

\[
  P(\lambda^{1/l} n) - \lambda P(n) - a_0 + 1 = b_{l+1}(\lambda^{1/l} n)^{l+1} - \lambda b_{l+1} n^{l+1} + Q(\lambda^{1/l} n + 1) - \lambda Q(n) - a_0 + 1.
\]

The coefficient of \( n^{l+1} \) 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 \( |r_S(\lambda P(n))| < \lambda^{1/l} n \).

If \( n \) is sufficiently large, we show that

\[
  |r_S(\lambda P(n+1))| > |r_S(\lambda P(n))|.
\]

Let \( i = |r_S(\lambda P(n))| \). 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 \( n \to +\infty \),

\[
  a_l \left[ \lambda - \left( \frac{i}{n} \right)^l \right] + \cdots \left[ \frac{a_k}{n^{l-k}} \right] \left( \lambda - \left( \frac{i}{n} \right)^k \right) + \cdots + \frac{a_0}{n^l} (\lambda - 1) \quad \text{is bounded.}
\]
So there exists $n_0' \geq n_0$ such that for all $n \geq n_0'$, $|r_S(\lambda P(n + 1))| > |r_S(\lambda P(n))|$.

Assume that $r_S(\lambda P(\mathbb{N}))$ is regular then the set $|r_S(\lambda P(\mathbb{N}))|$ is a finite union of arithmetic progressions. We may apply Lemma 3, indeed, the function $|r_S(\lambda P(.))|$ is strictly increasing in $\{n : n \geq n_0'\}$ and there exist $l_0$ and $\Gamma_\lambda$ (simply written $\Gamma$) such that $\forall l \geq l_0$, $l \in |r_S(\lambda P(\mathbb{N}))| \Leftrightarrow l + \Gamma \in |r_S(\lambda P(\mathbb{N}))|$. Let $n_1 \geq n_0'$ be such that $|r_S(\lambda P(n_1))| > l_0$. By Lemma 3, there exists $\lambda_k$ (simply written $\lambda$) such that for all $n \geq n_1$ and for all $\alpha \in \mathbb{N}$,

$$|r_S(\lambda P(n + \alpha k))| = |r_S(\lambda P(n))| + \alpha \Gamma.$$

Let $i = |r_S(\lambda P(n))|$. In view of (3), one has

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

Since $\lambda P(n + \alpha k) - P(i + \alpha \Gamma) + a_0 - 1$ must be positive for all $\alpha \in \mathbb{N}$, the coefficient of the greatest power of $\alpha$, $\alpha^{l+1}$, must be strictly positive. This coefficient is

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

and we have the condition

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

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

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

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

which leads to a contradiction. \(\square\)

In Theorem 12, we exhibit a recognizable set $X = P(\mathbb{N})$ such that $|r_S(\lambda P(\mathbb{N}))|$ 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 $|r_S(\beta^{l+1} P(\mathbb{N}))|$ is a finite union of arithmetic progressions whether $\rho_L$ is a polynomial of degree $l$. 13
Proposition 13  With the assumptions and notations of Theorem 12, there exists \( C \in \mathbb{Z} \) such that for \( n \) large enough,

\[
|\rho_S(\beta^{l+1}P(n))| = \beta n + C.
\]

Proof. In the proof of Theorem 12, we introduced a polynomial \( P(x) = b_{l+1} x^{l+1} + \cdots + b_1 x \) such that \( P(n+1) - P(n) = \rho_L(n) \). In view of (1), we have to find an integer \( C \) such that for \( n \) large enough

\[
P(\beta n + C + 1) - a_0 - \beta^{l+1} P(n) \geq 0 \quad (2)
\]

\[
\beta^{l+1} P(n) - P(\beta n + C) + a_0 - 1 \geq 0 \quad (3)
\]

The coefficient of \( n^{l+1} \) vanishes in (2) and (3). The coefficient of \( n^l \) in (2) is

\[
\beta^l [a_l (C+1) + b_l (1-\beta)]
\]

with \( a_l = b_{l+1} (l+1) \). It is strictly increasing with \( C \) and equals zero for

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

The same coefficient in (3) is

\[
-\beta^l [a_l C + b_l (1-\beta)].
\]

It is strictly decreasing with \( C \) and equals zero for \( C = C_2 := C_1 + 1 \).

If \( C_1 \) and \( C_2 \) are not integers then there exists \( C \in \mathbb{Z} \) such that for \( n \) large enough, the coefficients of terms of maximal degree are both strictly positive.

Otherwise, one has to consider the integer case \( C = C_1 \) or \( C = C_2 \) (it is obvious that any other \( C \) leads to a strictly negative expression for (2) or (3)). Moreover, if \( C = C_1 \) (resp. \( C = C_2 \)) then (3) (resp. (2)) is satisfied for \( n \) large enough.

Notice that for \( i = 1, \ldots, l - 1 \) the coefficient of \( n^i \) in (2) with \( C = C_1 \) is the opposite of the coefficient of \( n^i \) in (3) with \( C = C_2 \) since \( C_2 = C_1 + 1 \). Notice also that the independent term in (3) for \( C = C_1 \) is \( P(C_2) - a_0 \). In (3) for \( C = C_2 \) this term is \( -P(C_2) + a_0 - 1 \). Thus we can write (2) with \( C = C_1 \) as

\[
A_{l-1} n^{l-1} + \cdots + A_1 n + P(C_2) - a_0
\]

and (3) with \( C = C_2 \) as

\[
-A_{l-1} n^{l-1} - \cdots - A_1 n - P(C_2) + a_0 - 1.
\]

If there exists \( i \) such that \( A_i \neq 0 \) then let \( j = \max_{A_i \neq 0} i \). If \( A_j > 0 \) (resp. \( A_j < 0 \)) then one takes \( C = C_1 \) (resp. \( C = C_2 \)).

Now, assume that \( A_i = 0 \) for \( i = 1, \ldots, l - 1 \). If \( P(C_2) - a_0 \geq 0 \) then one takes \( C = C_1 \). Otherwise, \( -P(C_2) + a_0 \) is a strictly positive integer (remember the properties of \( P \) obtained in the proof of Theorem 12). Therefore \( -P(C_2) + a_0 - 1 \geq 0 \) and one takes \( C = C_2 \). \( \square \)
5 Multiplication and polynomial languages

Here we obtain the generalization of Theorem \[12\] for an arbitrary regular language of polynomial complexity. In the same time, 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 an infinite sequence \(n_0, n_1, \ldots, n_i, \ldots\) 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 \(\Omega(g(n))\). Let \(x \in A^*\) and \(y \in B^*\), with \(A\) and \(B\) two finite alphabets. If \(|x| = |y| + i\), \(i \in \mathbb{N}\) then \((x, y)^\# = (x, ^iy)\) where \(^i\) is a new symbol which does not belong to \(A \cup B\). If \(|y| = |x| + i\) then \((x, y)^\# = (^ix, y)\).

This operation can be extended to \(n\)-uples of words. Let \(R\) be a relation over \(A^* \times B^*\). We say that \(R\) is regular if \(R^\#\) is a regular language. This definition can be extended to \(n\)-ary relations. A map is regular if its graph is regular.

**Theorem 14** Let \(L \subset \Sigma^*\) be a regular language such that \(p_L(n)\) 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 \(r_S(X)\) is regular and that \(r_S(\lambda X)\) is not.

This theorem has a direct corollary.

**Corollary 15** Under the assumptions of Theorem \[14\], the addition is not a regular map (i.e. the graph of the application \((x, y) \mapsto x + y\) is not regular).

**Proof.** By Theorem \[12\], there exists a subset \(X\) of \(\mathbb{N}\) such that \(X\) is \(S\)-recognizable and \(2X\) is not. Assume that the graph of the addition

\[
\mathcal{G} = \{(r_S(x), r_S(y), r_S(x + y))^\# : x, y \in \mathbb{N}\}
\]

is regular. Let \(p_3\) be the canonical homomorphism defined by \(p_3(x, y, z) = z\). It is clear that the set \(A = \{(r_S(x), r_S(x), w)^\# : x \in X, w \in \Sigma^*\}\) is regular. Therefore

\[
A \cap \mathcal{G} = \{(r_S(x), r_S(x), r_S(2x))^\# : x \in X\}
\]

is regular. Thus \(p_3(A \cap \mathcal{G}) = r_S(2X)\) is also regular, a contradiction. \(\Box\)

In the following, we will use the term of \(k\)-tiered word and the results obtained in \[14\] about the complexity of regular polynomial languages.
The first lemma is just a refinement of [14, Lemma 1]. We simply remark that one can consider an ultimately periodic sequence \( n_i \) such that \( \rho_L(n_i) \geq b_0 n_i^l \).

**Lemma 16** 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 an infinite sequence \( n_0, n_1, \ldots, n_i, \ldots \) 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.** It is obvious that there exists a word \( w \in L \) which is \( (l + 1) \)-tiered (see [14, Lemmas 2-4]), \( w = x y_1^{d_1} z_1 \cdots y_{l+1}^{d_{l+1}} z_{l+1} \). Let \( C = |y_1| \cdots |y_{l+1}| \). As shown in [14], there exists a constant \( b_0 \) such that the number of words of length \( n_t = |xz_1 \cdots z_{l+1}| + tC \) is greater than \( b_0 n_t^l \) for any integer \( t \). \( \Box \)

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

\[
\sum_{i=0}^{n} i^p = \frac{(n + B + 1)^{p+1} - B^{p+1}}{p+1}
\]

where all terms of the form \( B^m \) are replaced with the corresponding Bernoulli numbers \( B_m \). This formula will be useful in the next lemma.

**Lemma 17** If \( \rho_L(n) \) 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 16.

**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 16 one has

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

Since \( n_i = n_0 + i C \), then \( n_i \) is a linear function of \( i \) and for \( i \) large enough, there exists a constant \( J \) such that

\[
v_{n_i} \geq J n_i^{l+1}.
\]
So, at this stage, we have a sequence \( n_i \) such that \( n_i = n_0 + iC \) and constants \( b_0, b_1, K \) and \( J \) such that for \( n \) and \( i \) sufficiently large,

\[
\begin{align*}
\rho_L(n) & \leq b_1 n^i \\
\rho_L(n_i) & \geq b_0 n_i^i \\
v_n & \leq K n^{l+1} \\
v_{n_i} & \geq J n_i^{l+1}
\end{align*}
\]

Before going further in the proof of Theorem 14, 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 \( W = a^*b^* \cap (\{a, b\}^*)^* \). It is obvious that \( \rho_W(2n + 1) = 2n + 2 \), \( \rho_W(2n) = 0 \) and \( v_{2n} = v_{2n+1} = (n + 1)^2 \).

**Lemma 18** Let \( \rho_1, \ldots, \rho_k, \theta_1, \ldots, \theta_k, \Phi_1, \ldots, \Phi_k \) be real numbers such that for all \( i \neq j \), \( \theta_i \neq \theta_j \) 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 \( k(n) \) such that

\[
\rho_j e^{i (k(n) \theta_j + \Phi_j)} \to z_j \quad \text{and} \quad |z_j| = \rho_j \neq 0.
\]

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

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

Therefore one has

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

This equality leads to a contradiction since the Vandermonde determinant does not vanish. \( \square \)

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 19 If $L$ is a regular language such that $\rho_L(n)$ is $\Theta(n^l)$ then the sequence $(\frac{n^l}{n})_{n \in \mathbb{N}}$ 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 equal to $l + 1$.

Proof. The sequence $(\rho_L(n))_{n \in \mathbb{N}}$ satisfies a recurrence relation. Therefore, if $z_i$ is a root of multiplicity $\alpha_i$ of the characteristic polynomial of $(\rho_L(n))_{n \in \mathbb{N}}$ then one can write

$$
\rho_L(n) = \sum_i P_i(n) z_i^n
$$

where $P_i(n)$ is a polynomial of degree less than $\alpha_i$. Moreover $\rho_L(n)$ is $\Theta(n^l)$; in other words, we have a constant $K$ such that

$$
\frac{\rho_L(n)}{n^l} \leq K.
$$

This latter inequality has important consequences.

i) We first show that $|z_i| > 1$ implies $P_i = 0$. Otherwise, let $\tau = \sup_i |z_i|$ and $d$ the maximal degree of polynomials $P_i$ corresponding to the different roots of modulus $\tau$. So we can write

$$
\frac{\rho_L(n)}{n^l} = \tau^n \frac{n^d}{n^l} |c_1 e^{in\theta_1} + \cdots + c_t e^{in\theta_t} + R_n|.
$$

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

$$
R_n = \frac{1}{\tau^n n^d} \left( \sum_{j:|z_j|<\tau} P_j(n) z_j^n + \sum_{j:|z_j|=\tau} h_d(P_j(n)) z_j^n \right)
$$

where

$$
h_d : \mathbb{C}[z] \to \mathbb{C}[z] : P(z) \mapsto P(z) - \frac{D^d P(z)}{d!} z^d.
$$

So $R_n \to 0$ if $n \to +\infty$. Therefore, by Lemma 18, there exists an infinite sequence of integers such that

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

For $n$ large enough, $|R_n| \leq \varepsilon/2$ and $\left| \frac{\rho_L(n)}{n^l} \right| \geq \tau^n n^{d-1} \frac{\varepsilon}{2}$ occurs infinitely often which contradicts (3).
ii) In the same way, one can verify that if \( |z_i| = 1 \) then the degree of the corresponding polynomial \( P_i \) cannot exceed \( l \).

iii) If we are interested in the behaviour of \( v_n/n^{l+1} \) when \( n \to +\infty \), then in the expression (4), we simply focus on the terms of the form \( \frac{D_i^l P_i(n)}{l!} n^l z_i^n \) for \( i \) such that \( |z_i| = 1 \). Indeed, any other term in \( \rho_L(n) \) provides \( v_n/n^{l+1} \) with a term which converges to zero (all these terms are included in \( R_n' \)). So, if we assume that \( 1 = z_0 \) has a multiplicity \( l + 1 \) and if \( z_1 = e^{i\theta_1}, \ldots, z_t = e^{i\theta_t} \) are the other roots of modulus one with \( q_j = \frac{D_i^l P_j(n)}{l!} \), \( j = 0, \ldots, t \) and \( \theta_0 = 0 \); then one can write

\[
\rho_L(n) = \sum_{j=1}^{t} q_j n^l e^{i\theta_j} + q_0 n^l + R_n'
\]

with \( \theta_j \neq 0 \) and

\[
R_n' = \sum_{j=0}^{t} h_l(P_j(n))e^{i\theta_j} + \sum_{j:|z_j|<1} P_j(n)z_j^n.
\]

Therefore, it is easy to see that

\[
\frac{v_n}{n^{l+1}} = \sum_{j=1}^{t} q_j \left( \frac{1}{n^{l+1}} \sum_{k=0}^{n} \frac{k^l e^{i\theta_j}}{n^{l+1}} + q_0 \frac{1}{n^{l+1}} \sum_{k=0}^{n} k^l + \frac{1}{n^{l+1}} \sum_{k=0}^{n} R_k \right).
\]

Moreover, we see that 1 has, necessary, a multiplicity \( l + 1 \); otherwise, \( \frac{v_n}{n^{l+1}} \to 0 \), which is a contradiction with Lemma [17]. □

**Proof of Theorem 14.** By definition of a numeration system, it is clear that for \( n \) sufficiently large, \( n + 1 \leq |r_S(v_n)| \leq n + C + 1 \) 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 \( |r_S(v_n)| = n + 1 \).) Recall also that \( |r_S(x)| = n \) iff \( v_{n-1} \leq x < v_n \). In this proof, we use the sequence \( n_i \) and the constants \( J, K, b_0 \) and \( b_1 \) introduced in the previous propositions.

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

\[
n + 1 \leq |r_S(\lambda v_n)| \leq \lceil \lambda^{1/l} n \rceil + C - 1 < \lambda^{1/l} n + C.
\]
It is sufficient to show that $\lambda v_n < v_{[\lambda^{1/l} n] + C - 1}$. By Lemma 17, there exists $k \in \{[\lambda^{1/l} n], \ldots, [\lambda^{1/l} n] + C - 1\}$ such that $v_k \geq J k^{l+1}$. Moreover the function $n \mapsto v_n$ is increasing. So,

$$v_{[\lambda^{1/l} n] + C - 1} \geq J [\lambda^{1/l} n]^{l+1} \geq J \frac{\lambda^{l+1}}{l} n^{l+1}.$$

Moreover, by Lemma 17, $\lambda v_n \leq \lambda K n^{l+1}$. By the choice of $\lambda$, it is clear that $\lambda K n^{l+1} < J \frac{\lambda^{l+1}}{l} n^{l+1}$.

ii) In Lemma 16 and Lemma 17, we have introduced two constants $b_0$ and $b_1$ such that $b_0 \leq b_1$. Let $s \in \mathbb{N} \setminus \{0\}$ such that $s b_0 > b_1$. Here, we show that the function

$$i \mapsto |r_{S}(\lambda v_{n si - 1})|$$

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

$$|r_{S}(\lambda v_{n si + 1, 1} - 1)| = |r_{S}(\lambda v_{n si + s C - 1})| > |r_{S}(\lambda v_{n si - 1})|.$$

Let $k = |r_{S}(\lambda v_{n si - 1})|$ then $v_{k-1} \leq v_{n si - 1} < v_k$ and we must show that

$$\lambda v_{n si + s C - 1} = \lambda v_{n si - 1} + \lambda \sum_{j=0}^{s C - 1} \rho_{L}(n si + j) \geq v_k = v_{k-1} + \rho_{L}(k).$$

So, it is sufficient to show that $\lambda \sum_{j=0}^{s C - 1} \rho_{L}(n si + j) \geq \rho_{L}(k)$. In view of (8), $k < \lambda^{1/l}(n si - 1) + C$. Therefore $\rho_{L}(k) < b_1 [\lambda^{1/l}(n si - 1) + C]^l$. On the other hand,

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

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

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

Since $\rho_{L}(n_0 + si C) > 0$, then $r_{S}(v_{n_0 + si C - 1})$ is the first word of length $n_0 + si C$ and

$$r_{S}(X) = r_{S} (\{v_n : n \in \mathbb{N}\}) \cap \sum_{n_0}^{\infty} (\sum^n_{C})^s.$$
So $X$ is a $S$-recognizable subset of $\mathbb{N}$.

Assume that $\lambda X$ is recognizable. Therefore, $|r_S(\lambda X)|$ is a finite union of arithmetic progressions. In view of ii), we can apply Lemma 9 and obtain two integral constants $\Gamma$ and $k$ such that for all $\alpha \in \mathbb{N}$,

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

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

$$v_{z+\alpha \Gamma-1} \leq \lambda v_{n_0+sC(i+\alpha k)-1} < v_{z+\alpha \Gamma}. \quad (7)$$

First consider the left inequality in (7), 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,

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

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

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

If we consider the right inequality in (7), we have $v_{z+\alpha \Gamma} \leq K (z + \alpha \Gamma)^{l+1}$ and also

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

If we focus on terms in $\alpha^{l+1}$, we obtain

$$\lambda \leq \frac{K}{J} \left( \frac{\Gamma}{sCk} \right)^{l+1}. \quad (9)$$

iv) By Theorem 19, $(\frac{v_i}{n^{l+1}})_{n \in \mathbb{N}}$ converges to a limit $a > 0$. Consider the sequences

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

For a given $m$ there exist $i_m$ and $n_m$ such that for $i \geq i_m$, $v_i \geq J_m n_i^{l+1}$ and for $n \geq n_m$, $v_n \leq K_m n^{l+1}$. 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 $\lambda > \left( \frac{Km}{J_m} \right)^l$ given in i) is equivalent to $\lambda \geq 2$ and the conditions (8) and (9) may be replaced by a unique condition

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

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

\[\square\]

6 Multiplication and complement of polynomial languages

In the previous sections, we have considered multiplication for numeration systems based on 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. 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 on 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 $2v_n$ and obtain Table 6 (for an algorithm of representation, see [9]).

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 20 Let $L \subset \Sigma^*$ and $x \in \Sigma^*$, we set $L_x = \{w \in L : w = xy\}$. It is clear that $L_x \subseteq L$. So $\rho_{L_x}(n) \leq \rho_L(n)$ and $\rho_{L_x}$ is $O(n^l)$ whenever $\rho_L$ is $O(n^l)$.

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

$$\rho(\Sigma^* \setminus L)_{n-k}(n) = \rho_{\Sigma^* \setminus L}(n) - \rho_{L_{b^n-k}}(n) = 2^k - 1.$$
Table 1: first terms of $2v_n$ for $S = ([a, b] \setminus a^*b^*, \{a, b\}, a < b)$. 

| $n$ | $2v_n$ | $r_{S}(2v_n) = b^kaw$ | $k$ | $|w|$ |
|-----|--------|----------------------|-----|------|
| 1   | 0      | $b$                  | 1   | 0    |
| 2   | 2      | $baa$                | 1   | 1    |
| 3   | 10     | $baab$               | 1   | 2    |
| 4   | 32     | $babab$              | 1   | 3    |
| 5   | 84     | $bbaaaa$             | 2   | 3    |
| 6   | 198    | $bbababa$            | 2   | 4    |
| 7   | 438    | $bbbaaabb$           | 3   | 4    |
| 8   | 932    | $bbabbabb$           | 3   | 5    |
| 9   | 1936   | $bbababaabaabaaba$   | 4   | 5    |
| 10  | 3962   | $bbbbaaabaabaabaaba$ | 5   | 5    |
| 11  | 8034   | $bbbbabbbbbab$       | 5   | 6    |
| 12  | 16200  | $bbbbbabbaaba$       | 6   | 6    |
| 13  | 32556  | $bbbbbbabaababaaba$  | 7   | 6    |
| 14  | 65294  | $bbbbbbbbabaabbaaba$ | 8   | 6    |
| 15  | 130798 | $bbbbbbbbaaabbbbbab$ | 9   | 7    |
| 16  | 261836 | $bbbbbbbbabbbabbb$   | 9   | 7    |
| 17  | 523944 | $bbbbbbbbbbbbabbaabaaba$ | 10 | 7    |
| 18  | 1048194| $bbbbbbbbbbbbababababa$ | 11 | 7    |
| 19  | 2096730| $bbbbbbbbbbbbabbaaaab$ | 12 | 7    |
| 20  | 4193840| $bbbbbbbbbbbbbabbaab$ | 13 | 7    |
| 21  | 8388100| $bbbbbbbbbbbbbbbaaaabbaaa$ | 14 | 7    |

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

23
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 \( r_S(\{v_n : n \in \mathbb{N}\}) \) are the first words of each length in \( \Sigma^* \setminus L \). So \( \{v_n : n \in \mathbb{N}\} \) is \( S \)-recognizable. Recall that \( |r_S(x)| = n \iff v_{n-1} \leq x < v_n. \) For \( n \) large enough, it is obvious that \( v_n \leq 2v_n < v_{n+1}. \) Then \( |r_S(2v_n)| = n + 1 \).

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

\[
\rho_{(\Sigma^* \setminus L)_{\beta n - i+1}}(n) = 2^{i-1} - 1 < v_{n+1} - 2v_n \leq 2^i - 1 = \rho_{(\Sigma^* \setminus L)_{\beta n - i}}(n).
\]

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

Assume that \( L = r_S(\{2v_n : n \in \mathbb{N}\}) \) is accepted by an automaton with \( q \) states. There exist \( n_0, i_0 \) and \( t \geq 0 \) such that \( r_S(2v_{n_0}) = b^{q+t}az_0 \) with \( |z_0| = i_0 \). By the pumping lemma, there exists \( \alpha > 0 \) such that

\[
\forall m \in \mathbb{N}, \ b^{q+t+ma}az_0 \in L.
\]

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 21** 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 \( j \geq 1, \ s^jX \) is not \( S \)-recognizable.

**Proof.** For \( 0 \leq k < n \), we have

\[
\rho_{(\Sigma^* \setminus L)_{\beta n - k}}(n) = \rho_{\Sigma^*_{\beta n - k}}(n) - \rho_{r_{\beta n - k}}(n) = s^k - \rho_{L_{\beta n - k}}(n) \in O(n^l).
\]
To avoid any misunderstanding, $v_n$ is the sequence associated to the language $\Sigma^* \setminus L$ of the numeration $S$ and $v_n(L)$ is related to $L$. So, $v_L(n) = \sum_{i=0}^{n} \rho_L(i)$ and

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

We take $X = r_S(\{v_n : n \in \mathbb{N}\})$, an $S$-recognizable set. We have, for $n$ sufficiently large,

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

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 \ref{thm:sufficient-cond-stab}, there exists $a > 0$ such that $v_L(n) \sim a n^{l+1}$. 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 $|r_S(s^j v_n)| = n + j$.

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

$$\rho(\Sigma^* \setminus L)_{\beta n-i+1}(n) < v_{n+j} - s^j v_n \leq \rho(\Sigma^* \setminus L)_{\beta n-i}(n)$$

Then $r_S(s^j v_n) = \beta^{n-i} \sigma z$ with $|z| = i + j - 1$ 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 (10) by $s^n$. Let $n \to +\infty$ and obtain a contradiction.

Suppose that $r_S(\{s^j X\})$ is accepted by an automaton with $q$ states. There exist $n_0$, $i_0$ and $t \geq 0$ such that $r_S(s^j v_{n_0}) = \beta^{q+t} \sigma z_0$ with $|z_0| = i_0$ and $\sigma \neq \beta$. Then using the pumping lemma, 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.

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 $R \subset A^* \times B^*$ 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 $R$. Finally, a function is computable by a finite 2-tape automaton if its graph is computable by a finite 2-tape automaton.

**Definition 22** If $U = (U_n)_{n \in \mathbb{N}}$ is a sequence of integers and $x = x_n \ldots x_0$, a word over an alphabet $B \subset \mathbb{Z}$. We define the numerical value of $x$ as

$$\pi_U(x) = \sum_{i=0}^{n} x_i U_i.$$ 

Notice that different words can have the same numerical value.

**Proposition 23** 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 state $p \in K$ and all $n \in \mathbb{N}$

$$\alpha u_{n+k-1}(p) = \sum_{i=0}^{k-1} e_{p,i} U_{n+i}.$$ (11)

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

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

**Remark 2** The function $g$ of the previous theorem is injective. If $v$ and $w$ are two words of $L$ such that $g(v) = g(w)$ then $\text{val}_S(v) = \text{val}_S(w)$. So the conclusion, since $\text{val}_S$ 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 $w = w_{k+l} \ldots w_{k-1} w_{k-2} \ldots w_0$ be a word of $L$ of length $k + l + 1$ with $l \geq -1$. We compute $l+2$ applications of Lemma $\Box$ on $\text{val}_S(w)$ and we obtain

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

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

Therefore, using (11), we have

$$\alpha \text{val}_s(w) = \alpha C_w + \sum_{j=0}^{k-1} \sum_{p \in K} \beta_{s,p,w_{k+i}} e_{p,j} U_{i+j+1} = \lambda_{i,j}$$

$$+ \sum_{i=-1}^{l-1} \sum_{j=0}^{k-1} \sum_{p \in K} \beta_{s,w_{k+i-1}p,w_{k+i}} e_{p,j} U_{i+j+1} = \lambda_{i,j}$$

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 $s.w_{k+i} \ldots w_{k+i-1}$ reached after the lecture of the first letters of $w$ (resp. the state $s$). Therefore, we extend the notation $\lambda_{i,j}$ to a meaningful one:

$$\lambda_{q,\sigma,j} = \sum_{p \in K} \beta_{q,p,\sigma} e_{p,j} \quad (12)$$

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 $\alpha \text{val}_s(w)$ can be interpreted in the following way. The reading of $w_{k+i}$, $l \leq i \leq -1$, provides the decomposition of $\alpha \text{val}_s(w)$ with $\lambda_{i,k-1}U_{k+i}$.
The reading of \( w_{k+i} \) gives a coefficient \( \lambda_{i,k-2} \) 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 \( \lambda_{i,k-3} \ldots \lambda_{i,0} \) 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 \( K = K \cup \{ f \} \times T \times \cdots \times T \) 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 \( \Delta : K \times (\Sigma \times B) \to K \) is defined as follows. If \( p \in K, \sigma \in \Sigma \),

\[
\Delta((p, \gamma_{k-2}, \ldots, \gamma_0), (\sigma, \lambda_{p,\sigma,k-1} + \gamma_{k-2})) = (p;\sigma; \lambda_{p,\sigma,k-2} + \gamma_{k-3} \ldots \lambda_{p,\sigma,1} + \gamma_0; \lambda_{p,\sigma,0})
\]

These transitions compute an output \( x_{k+l} \ldots x_{k-1} \) from \( w_{k+l} \ldots w_{k-1} \). The alphabet \( B \) is finite since \( T \) is finite. But we have still to read the last \( k-1 \) letters of \( w \). For each state \( p \in K \), \( D_p = L_p \cap \Sigma^{k-1} \) is finite (recall that \( L_p \) are the words accepted from \( p \)). So, for each state \( p \in K \) and each word \( w_{k-2} \ldots w_0 \in D_p \), we construct an edge from \( (p, \gamma_{k-2}, \ldots, \gamma_0) \) to \( f \) labelled by \( (w_{k-2} \ldots w_0, \gamma_{k-2} \cdot \cdot \cdot \gamma_1 (\gamma_0 + C_w)) \). (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 \( s.w_{k+l} \ldots w_{k-1} \) reached (the first component in \( K \)) and the remaining word \( w_{k-2} \ldots w_0 \).

**Remark 3** The states of \( M \) satisfy the same recurrence relation of degree \( l \). A practical way to check (11) is to seek a final state \( f \in F \) such that

\[
\det \begin{pmatrix} u_0(f) & \cdots & u_{l-1}(f) \\ \vdots & \ddots & \vdots \\ u_{l-1}(f) & \cdots & u_{2l-2}(f) \end{pmatrix} \neq 0.
\]

If such an \( f \) exists then for all \( p \in K \), there exist \( c_{p,i} \in \mathbb{Q} \) such that

\[
u_{n+l-1}(p) = \sum_{i=0}^{l-1} c_{p,i} u_{n+i}(f)
\]

and (11) can be easily obtained.
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. If \( x \) is an integer, the \( U \)-representation of \( x \) obtained by the greedy algorithm is denoted by \( \rho_U(x) \) and belongs to \( A'_U \), where \( A_U = \{0, \ldots, Q\} \) is the canonical alphabet of the system \( U \), \( Q < \max \frac{U_{n+1}}{U_n} \). A set \( X \subset \mathbb{N} \) is said \( U \)-recognizable if \( \rho_U(X) \) is regular. For any alphabet \( C \) of integers, one can define a partial function called normalization

\[
\nu_{U,C} : C^* \to A_U^* : z \mapsto \rho_U(\pi_U(z)).
\]

**Corollary 24** Let \( S = (L, \Sigma, <) \). With the hypothesis and notations of Proposition 23, 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 \( \alpha X \) is \( U \)-recognizable.

*Proof.* Let the regular language \( G \subset (\Sigma \times B)^* \) be the graph of the function \( q \) defined in Proposition 23. We denote by \( p_1 : \Sigma \times B \to \Sigma \) and \( p_2 : \Sigma \times B \to B \) the canonical homomorphisms of projection. Let

\[
Y = p_2[p_1^{-1}(r_S(X)) \cap G].
\]

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

Conversely, if \( \rho_U(\alpha X) \) is regular then \( \nu_{U,B}^{-1} \circ \rho_U(\alpha X) \) is also regular. For each \( y \in \alpha X \), \( \nu_{U,B}^{-1} \circ \rho_U(y) \) can take more than one value but only one is in \( p_2(G) \). So the set

\[
p_1 \left( p_2^{-1}[\nu_{U,B}^{-1} \circ \rho_U(\alpha X)] \cap G \right)
\]

is regular and equal to \( r_S(X) \). \( \Box \)

**Corollary 25** Let \( S = (L, \Sigma, <) \). With the hypothesis and notations of Proposition 24, if the sequence \( U \) satisfies a linear recurrence relation

\[
U_n = d_1 U_{n-1} + \cdots + d_m U_{n-m}, d_i \in \mathbb{Z}, d_m \neq 0, n \geq m
\]

such that its characteristic polynomial is the minimal polynomial of a Pisot number then \( X \subset \mathbb{N} \) is \( S \)-recognizable if and only if \( X \) is \( U \)-recognizable.
Proof. It is well known that for such a system $U$ the normalization $\nu_{U,C}$ is computable by finite letter-to-letter 2-tape automaton for any alphabet $C$ (see [8]). So by the previous corollary, $X$ is $S$-recognizable if and only if $\alpha X$ 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 $\alpha$ is definable in $\langle \mathbb{N}, + \rangle$. So $\alpha X$ is definable in the structure if and only if $X$ is definable. □

Remark 4 Let $S = (L, \Sigma, \prec)$ and $S' = (L, \Sigma, \prec')$ be two systems which only differ by the ordering of the alphabet. If the hypothesis of Proposition 23 and Corollary 24 are satisfied then a set $X$ is $S$-recognizable if and only if it is $S'$-recognizable. In other words, recognizable sets are independent of the ordering of the alphabet.

Example 2 Consider the language $L \subset \{a, b, c\}^*$ of the words that do not contain $aa$. Its minimal automaton $M_L$ is given on 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

Figure 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 23 can be played 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 satisfied by $u_n(s)$ 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 $3 \times 3$ matrices $B_\sigma = (\beta_{q,r,\sigma})_{q,r=\sigma,s,\sigma}$, $\sigma \in \Sigma$:

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

If $E = (e_{q,i})_{q=s,t,p;i=0,1}$ then it follows from (12) 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 23, we have to compute the $C_w$ namely

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

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

![Figure 2: The transducer computing $g$.](image)

We can do the same construction for the language $L' = a^+\{a, b\}^*$. Its minimal automaton $M_{L'}$ is given on Figure 3. The sequence $U$ of Proposition 23 can be played by $u_n(t) = 2^n$. So here, the Pisot number involved is 2 and it is multiplicatively independent with $1 + \sqrt{3}$. So from [10], 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 5 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 on Figure 4. We consider the numeration system $S = (J, \{a,b\}, a < b)$

![Figure 3: The minimal automaton of $L' = a^+\{a,b\}^*$.

![Figure 4: The minimal automaton of $J = a\{a,b\}^* \cup \{a,b\}^* bb\{a,b\}^*$.

and we show that

i) we cannot find a linear recurrent sequence associated to a Pisot number such that the condition (11) of proposition 23 is satisfied for all state 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 21. 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

$$\rho_{J^*_{\psi+1}}(n) = 2^{n-i-1} < v_{n+1}(s) - 2v_n(s) \leq 2^{n-i} = \rho_{J^*_{\psi}}(n)$$

32
and \( n - i \to +\infty \) if \( n \to +\infty \). One can conclude as in Theorem 21; \( r_S(2v_n(s)) = b^iaz \) with \( |z| = n - i - 1 \).

8 Acknowledgments

The author would like to warmly thank P. Mathonet for fruitful discussions in the polynomial case and also P. Lecomte for his support and improvements in many proofs.

References

[1] J. Berstel, C. Reutenauer, *Les séries rationnelles et leurs languages*, 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** No 1 (1991) 25–35.

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

[4] V. Bruyère, G. Hansel, Bertrand numeration systems and recognizability, *Theor. 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** No 5 (1995) 437–450.

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

[8] 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.

[9] P. B. A. Lecomte, M. Rigo, Numeration systems on a regular language, *preprint* (1999), see also [http://xxx.lanl.gov/abs/cs.0H/9903005](http://xxx.lanl.gov/abs/cs.0H/9903005).

[10] F. Point, V. Bruyère, On the Cobham-Semenov theorem. *Theory Comput. Syst.* **30** No. 2 (1997), 197–220.
[11] M. Rigo, Generalization of automatic sequences for numeration systems on a regular language, preprint (1999), see also http://xxx.lanl.gov/abs/cs.CC/9906017.

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

[13] J. Shallit, Numeration systems, linear recurrences, and regular sets, Information and Computation, 113 No 2 (1994) 331–347.

[14] 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.

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