The Nondeterministic Complexity of a Finite Automaton

Imre Simon*
Instituto de Matemática e Estatística
Universidade de São Paulo
01498 São Paulo, SP, Brasil

February 6, 1987

Abstract

We define the nondeterministic complexity of a finite automaton and show that there exist, for any integer $p \geq 1$, automata which need $\Theta(k^{1/p})$ nondeterministic transitions to spell words of length k. This leads to a subdivision of the family of recognizable $\mathcal{M}$-subsets of a free monoid into a hierarchy whose members are indexed by polynomials, where $\mathcal{M}$ denotes the Min–Plus semiring.

1 Introduction

Let us consider a (nondeterministic) finite automaton $\mathcal{A} = (Q, I, T)$ over an alphabet A. Here, Q is the set of states of $\mathcal{A}$ and subsets I and T of Q are the initial and terminal states. In $\mathcal{A}$ we distinguish between deterministic and nondeterministic edges as follows: an edge (p, a, q) is deterministic if there are no other edges (p, a, r), with $r \neq q$. For every word s accepted by $\mathcal{A}$ we associate a natural, denoted $s\|\mathcal{A}\|$, which is the least number of nondeterministic edges needed to spell s from I to T. The function $\|\mathcal{A}\|$ thus defined will be called the nondeterministic complexity of $\mathcal{A}$. A similar concept was introduced by Kintala and Fischer in [6]; they use it mainly for more complex devices, such as Turing machines.

*Part of this work was done while the author was visiting the Institut de Programmation, Université Pierre et Marie Curie, Paris VI, in 1986. That visit was partially supported by FAPESP.

An important property of function $\|\mathcal{A}\|$ is given by the finiteness of the set of coefficients $s\|\mathcal{A}\|$, when s ranges over the set of words accepted by $\mathcal{A}$. In this case we say that $\|\mathcal{A}\|$ is limited. Notice that a decision procedure for this property has been given by Hashiguchi in [4, 5].

The following question has been posed by J. Goldstine around 1977 [9]: Is it true that for any automaton $\mathcal{A}$ either $\|\mathcal{A}\|$ is limited or there is a constant $c > 0$ such that for every large enough k some word s accepted by $\mathcal{A}$ satisfies:

$$|s| \leq k \qquad \text{and} \qquad c|s| \leq s\|\mathcal{A}\|.$$

Remembering that $s\|\mathcal{A}\| \leq |s|$ for every s accepted by $\mathcal{A}$, this property would imply that the maximum coefficient of words of length at most k is a function which grows linearly with k.

We shall give a negative answer to this question. Indeed, we will exhibit automata $\mathcal{A}_p$, for every $p \geq 1$, such that $\|\mathcal{A}_p\|$ is unlimited but for any s accepted by $\mathcal{A}_p$, $s\|\mathcal{A}_p\| \leq c_p|s|^{1/p}$, for some constant c_p depending only on p.

Throughout the paper we shall look at the complexity function $\|\mathcal{A}\|$ of automaton $\mathcal{A}$ as a recognizable $\mathcal{M}$-subset of a free monoid (see Section 2 for definitions). This leads, in Section 5, to the subdivision of the family of recognizable $\mathcal{M}$-subsets of a free monoid into a hierarchy whose members are indexed by polynomials. These polynomials reflect the relationship between coefficients and lengths of words accepted by an $\mathcal{M}$-A-automaton $\mathcal{A}$. The complexity functions of automata $\mathcal{A}_p$ will be used to show that the hierarchy is a proper one.

2 Recognizable $\mathcal{M}$-subsets of a free monoid

At this point we need to formalize the discussion. Initially we will look at the nondeterministic complexity functions from a more general viewpoint: that of recognizable $\mathcal{M}$-subsets of a free monoid.

In Chapter VI of [3], Eilenberg formalized the concept of sets with multiplicities in a semiring K and developped the theory of recognizable K-subsets of a free monoid. This theory will be the general framework for our studies in the sequel.

We recall from [3] that a K-subset S of a set X is a function $S\colon\ X \to K$. The K-subsets of a free monoid A^* can also be viewed as formal series with coefficients in K and variables in A [3, 2].

We will consider the Min-Plus semiring $\mathcal{M}$ which has support $\mathcal{M} = \mathbb{N} \cup \infty$ and operations $a \oplus b = \min\{a, b\}$ and $a \otimes b = a + b$. The operations of $\mathbb{N}$ are extended to $\mathcal{M}$ in the usual way and the identities of $\oplus$ and $\otimes$ are, respectively, ∞ and 0. Notice that $\mathcal{M}$ is a complete positive commutative semiring [3].

The semiring $\mathcal{M}$ is used for cost minimization in Operations Research. Our use of $\mathcal{M}$ is strongly related to a decision procedure for the finiteness of finitely generated matrix semigroups over $\mathcal{M}$, given in [8] and recently extended by Mascle in [7] to other semirings.

Consider a finite automaton $\mathcal{A} = (Q, I, T)$, over the alphabet A, whose edges have multiplicities in $\mathcal{M}$, i.e. $\mathcal{A}$ is an $\mathcal{M}$-A-automaton with state set Q, whose initial and terminal states are I and T respectively. The edges of $\mathcal{A}$ will be seen as an $\mathcal{M}$-subset of $Q \times A \times Q$,

$$E\colon\; Q \times A \times Q \to \mathcal{M},$$

where $(p, a, q)E$ is the multiplicity of the edge from $p \in Q$ to $q \in Q$ labeled by $a \in A$. The set of useful edges of $\mathcal{A}$ is given by the support of E:

$$\{\, (p, a, q) \in Q \times A \times Q \mid (p, a, q)E \neq \infty \,\}.$$

A path c of $\mathcal{A}$ is a continuous sequence of edges,

$$c = (p_0, a_1, p_1)(p_1, a_2, p_2) \ldots (p_{n-1}, a_n, p_n),$$

for $p_i \in Q$ and $a_i \in A$. The label of c, here denoted by $|c|$, is the word $a_1 a_2 \ldots a_n$ in A^*, while its multiplicity, denoted by $\|c\|$, is the sum of multiplicities of its edges,

$$\|c\| = \sum (p_i, a_{i+1}, p_{i+1})E.$$

Again, a path is useful iff its multiplicity is not ∞. A useful path from I to T is called successful.

The set recognized by $\mathcal{A}$, denoted by $|\mathcal{A}|$, is the set of words which are labels of successful paths. On the other hand, the behavior of $\mathcal{A}$, denoted by $\|\mathcal{A}\|$, is an $\mathcal{M}$-subset of A^*,

$$\|\mathcal{A}\|\colon\; A^* \to \mathcal{M},$$

associating to each word $s \in A^*$ its multiplicity (or coefficient), $s\|\mathcal{A}\|$, given by

$$s\|\mathcal{A}\| = \min\{\, \|c\| \mid c \text{ is a successful path with label } |c| = s \,\}.$$

In other words, the multiplicity of s is the minimum of the multilicities of paths spelling s from I to T. A successful path with label s and multiplicity $s\|\mathcal{A}\|$ is called victorious.

The family of behaviors of $\mathcal{M}$-A-automata is denoted by $\mathrm{Rec}_{\mathcal{M}}\, A^*$. Its elements are called recognizable $\mathcal{M}$-subsets of A^*.

For $n \geq 1$, let $M_n K$ denote the multiplicative monoid of $n \times n$ matrices over the semiring K. For a matrix x in $M_n K$ we denote the coefficient of x in row i and column j by (i, x, j). Let $\mathcal{A} = (Q, I, T)$ be an $\mathcal{M}$-A-automaton with $n = \operatorname{card} Q$ states. For simplicity let us assume that $Q = \{1, 2, \ldots, n\}$.

A useful construction associates a submonoid of $M_n\mathcal{M}$ to $\mathcal{A}$ as follows: Let

$$f:\ A \to M_n\mathcal{M}$$

be the function defined by

$$(p, af, q) = (p, a, q)E.$$

We also denote by f the unique extension of f to a morphism

$$f:\ A^* \to M_n\mathcal{M}.$$

It is easy to see that for every $s \in A^*$ and $p, q \in Q$,

$$(p, sf, q) = \min\{\, \|c\| \mid c \text{ is a path from } p \text{ to } q \text{ with label } |c| = s\,\}.$$

Thus, in particular,

$$s\|\mathcal{A}\| = \min\{\, (i, sf, t) \mid i \in I,\ t \in T\,\}.$$

This value can also be obtained by regarding I and T as a row vector $1 \times n$ and a column vector $n \times 1$, respectively, with coefficients

$$(1, I, i) = \begin{cases} 0 & \text{if } i \in I \\ \infty & \text{if } i \notin I \end{cases} \qquad \text{and} \qquad (t, T, 1) = \begin{cases} 0 & \text{if } t \in T \\ \infty & \text{if } t \notin T \end{cases}.$$

Then,

$$s\|\mathcal{A}\| = I(sf)T.$$

3 Relations between coefficients and lengths

Our purpose is to study the relationship between coefficients, $s\|\mathcal{A}\|$, and lengths, $|s|$, of words s recognized by an $\mathcal{M}$-A-automaton $\mathcal{A}$. To do this, we define two measures of this relationship.

Let $F\colon A^* \to \mathcal{M}$ be an $\mathcal{M}$-subset of A^*. For F and for $m \geq 0$, we define

$$\mathrm{sh}(F, m) = \min\{\, |s| \mid s \in A^*,\, m \leq sF < \infty \,\}.$$

Thus, $\mathrm{sh}(F, m)$ is the minimum length needed to achieve a finite coefficient which exceeds m. For F and for $k \geq 0$, we also define

$$\mathrm{cm}(F, k) = \max\{\, sF \mid s \in A^*,\, |s| \leq k,\, sF < \infty \,\}.$$

Thus, $\mathrm{cm}(F, k)$ is the maximum finite coefficient of words of length at most k. In both cases we consider undefined the min and max of an empty set. Hence, in general, sh and cm are both partial functions.

Functions sh and cm can be thought as being inverse functions. Indeed, even though they are generally step-functions, hence not bijective, their restrictions to the appropriate corners are inverse bijections. More precisely, let us consider $\mathbb{N}^2$ ordered by the relation

$$(k, m) \leq (k', m') \qquad \text{iff} \qquad k \leq k' \text{ and } m \geq m'.$$

Then, for any minimal element (k, m) in the set

$$\{\, (|s|, sF) \mid s \in A^*,\, sF < \infty \,\}$$

we have the relations

$$\mathrm{cm}(F, \mathrm{sh}(F, m)) = m \qquad \text{and} \qquad \mathrm{sh}(F, \mathrm{cm}(F, k)) = k.$$

The $\mathcal{M}$-subset F is said to be *limited* if A^*F is finite. Or, equivalently, the finite coefficients of F are bounded. F is said to be *unlimited* if it is not limited. Note that if F is limited then $\mathrm{sh}(F, m)$ is undefined for sufficiently large m and $\mathrm{cm}(F, k)$ is either always undefined or it is defined and bounded for sufficiently large k. By contrast, if F is unlimited then $\mathrm{sh}(F, m)$ is always defined and unbounded while $\mathrm{cm}(F, k)$ is defined and unbounded for sufficiently large k.

Next we consider the assymptotic behavior of these complexity measures of F. We use the O, Ω and Θ notations, extended for partial functions. More precisely, we say that $f(m) \in O(g(m))$ (or, equivalently, $g(m) \in \Omega(f(m))$) iff there exist $c > 0$ and m_0 such that for every $m \geq m_0$ if $f(m)$ and $g(m)$ are both defined then $f(m) \leq cg(m)$. It follows that if F is limited then $\mathrm{sh}(F, m)$ belongs to every $O(g(m))$ and to every $\Omega(g(m))$. By contrast, $\mathrm{cm}(F, k)$ belongs to $O(g(k))$ (resp. $\Omega(g(k))$) iff $O(g(k))$ (resp. $\Omega(g(k))$) contains an identically constant function with nonnull value.

These complexity measures for F are related by the following properties, where $0 \leq m, k < \infty$, and we assume that both $\mathrm{sh}(F, m)$ and $\mathrm{cm}(F, k)$ are defined:

$$\mathrm{sh}(F, m) \leq k \qquad \text{iff} \qquad \mathrm{cm}(F, k) \geq m.$$

It follows that, for any $p > 0$ and F,

$$\mathrm{sh}(F, m) \in \Omega(m^p) \qquad \text{iff} \qquad \mathrm{cm}(F, k) \in O(k^{1/p})$$

and

$$\mathrm{sh}(F, m) \in O(m^p) \qquad \text{iff} \qquad F \text{ is limited or } \mathrm{cm}(F, k) \in \Omega(k^{1/p}).$$

4 The nondeterministic complexity of a finite automaton

Let $\mathcal{A} = (Q, I, T)$ be a (not necessarily deterministic) finite automaton over the alphabet A. We say that edge (p, a, q) of $\mathcal{A}$ is *deterministic* if there are no other edges (p, a, r) in $\mathcal{A}$, with $r \neq q$.

We will convert $\mathcal{A}$ into an $\mathcal{M}$-A-automaton by defining the multiplicity of (p, a, q) in $Q \times A \times Q$ by

$$(p, a, q)E = \begin{cases} 0 & \text{if } (p, a, q) \text{ is a deterministic edge} \\ 1 & \text{if } (p, a, q) \text{ is not a deterministic edge} \\ \infty & \text{if } (p, a, q) \text{ is not an edge of } \mathcal{A}. \end{cases}$$

The behavior $\|\mathcal{A}\|$ of $\mathcal{A}$ will be called the *nondeterministic complexity* of the finite automaton $\mathcal{A}$. Thus, $s\|\mathcal{A}\|$ is the minimum number of nondeterministic edges needed to spell s from I to T.

Now we show that there exists a finite automaton $\mathcal{A}_p$, for each $p \geq 1$, with nondeterministic complexity $\|\mathcal{A}_p\|$, for which $\mathrm{sh}(\|\mathcal{A}_p\|, m)$ can be expressed as a polynomial of degree p in m. It follows that $\mathrm{sh}(\|\mathcal{A}_p\|, m)$ and $\mathrm{cm}(\|\mathcal{A}_p\|, k)$ are in $\Theta(m^p)$ and $\Theta(k^{1/p})$, respectively.

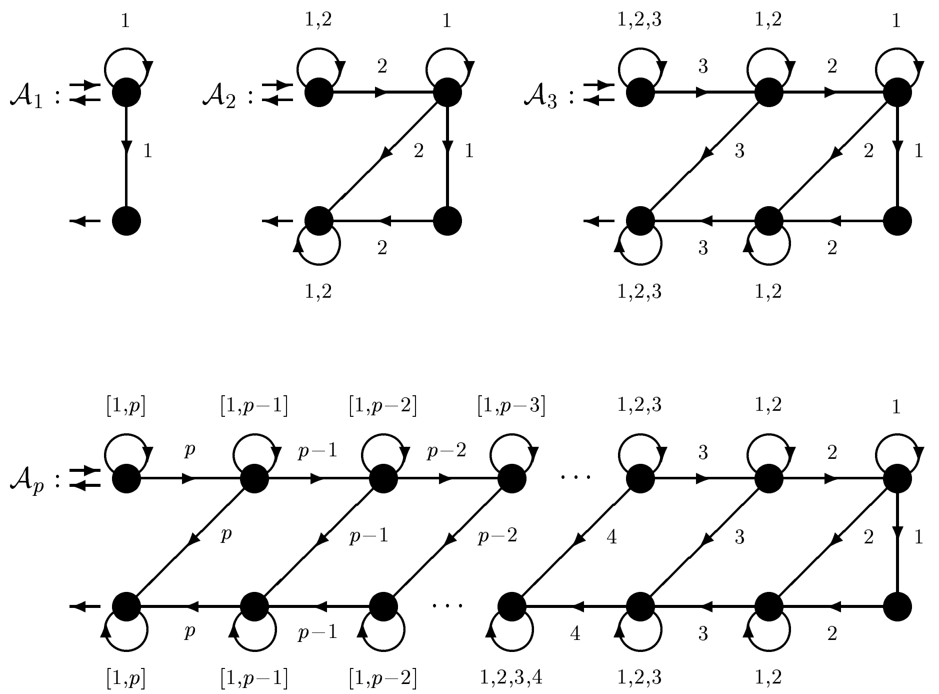

Figure 1: Automata $\mathcal{A}_p$

Let $p \geq 1$ be an integer. Automaton $\mathcal{A}_p = (Q_p, I_p, T_p)$ has $2p$ states, it is over the p-letter alphabet $A_p = \{ 1, 2, \ldots, p \}$, and has edge set E_p:

$$Q_1 = \{ u_1, v_1 \}, \qquad Q_p = Q_{p-1} \cup \{ u_p, v_p \} \quad (p > 1),$$

$$I_p = \{ u_p \}, \qquad T_p = \{ u_p, v_p \},$$

$$E_1 = \{ (u_1, 1, u_1), (u_1, 1, v_1) \},$$

$$\begin{aligned} E_p \quad = \quad & E_{p-1} \cup \{ (q, a, q) \mid q \in \{ u_p, v_p \},\, a \in A_p \} \cup \\ & \cup \{ (u_p, p, u_{p-1}), (u_{p-1}, p, v_p), (v_{p-1}, p, v_p) \} \quad (p > 1). \end{aligned}$$

Figure 1 shows the sequence $\mathcal{A}_p$ of automata.

Notice that the set N_p of nondeterministic edges of $\mathcal{A}_p$ has exactly $2p$ elements and is given by:

$$N_1 = E_1, \qquad N_p = N_{p-1} \cup \{ (u_p, p, u_p), (u_p, p, u_{p-1}) \} \quad (p > 1).$$

Proposition 1 *For any $p \geq 1$, the nondeterministic complexity of $\mathcal{A}_p$ satisfies:*

$$\mathrm{sh}(\|\mathcal{A}_p\|, m) = \sum_{k=1}^{p} \binom{m}{k}.$$

Proof. Note initially that, $|\mathcal{A}_p| = A_p^*$, for every $p \geq 1$. In order to study the function $\|\mathcal{A}_p\|$ let us define $X(p,m)$ by

$$X(p,m) = \{\, s \in A_p^* \mid s\|\mathcal{A}_p\| \geq m \,\} \quad (p \geq 1,\ m \geq 0).$$

Then, $\mathrm{sh}(\|\mathcal{A}_p\|, m) = \min\{\, |s| \mid s \in X(p,m) \,\}$.

Next, we show that the sets $X(p,m)$ satisfy the recurrence relations

$$\begin{aligned}
X(1,m) &= 1^m 1^*, \quad (m \geq 0)\\
X(p,0) &= A_p^*, \qquad (p > 1)\\
X(p,m) &= A_{p-1}^* pX(p-1,m-1)pX(p-1,m-2)\dots pX(p-1,1)pA_p^*,\\
&\quad (p > 1,\ m \geq 1) \qquad\qquad (1)
\end{aligned}$$

The first two relations clearly hold. To show (1) let initially s be a word in $X(p,m)$. Notice that the path c spelling s from u_p to u_p has multiplicity $\|c\| = |s|_p$. Thus, the hypothesis $s\|\mathcal{A}_p\| \geq m$ implies that s has a factorization

$$s = t_0 p t_1 p \dots t_{m-1} p t_m,$$

with $t_i \in A_{p-1}^*$, for $i \in [0, m-1]$, and $t_m \in A_p^*$.

Now, for $i \in [0, m-1]$, let $m_i = t_i\|\mathcal{A}_{p-1}\|$; we claim that $s\|\mathcal{A}_p\| \leq i + m_i$. Indeed, let c_1 be the path that spells $s_1 = t_0 p t_1 p \dots t_{i-1}$ from u_p to u_p. Then, $\|c_1\| = |s_1|_p = i - 1$. Let $c_2 = (u_p, p, u_{p-1})$. Then, $\|c_2\| = 1$. Let c_3 be a path of multiplicity $\|c_3\| = m_i$, from u_{p-1} to T_{p-1}, spelling t_i. The existence of c_3 is guaranteed by the hypothesis. Note that whatever the terminus w of c_3, there exists a path $c_4 = (w, p, v_p)$ of multiplicity $\|c_4\| = 0$. Finally, let c_5 be the path that spells $s_5 = t_{i+1} p t_{i+2} \dots p t_m$ from v_p to v_p. The multiplicity of c_5 is $\|c_5\| = 0$. Thus, $c = c_1 c_2 c_3 c_4 c_5$ is a path, of multiplicity $\|c\| = i + m_i$, spelling s from u_p to v_p. This proves the claim.

By assumption, $m \leq s\|\mathcal{A}_p\|$. From the claim, $s\|\mathcal{A}_p\| \leq i + m_i$. Thus, $t_i\|\mathcal{A}_{p-1}\| = m_i \geq m - i$, hence $t_i \in X(p-1, m-i)$. It follows that s belongs to the expression on the right hand side of (1).

To see the reverse inclusion, let

$$s = t_0 p t_1 p \dots t_{m-1} p t_m,$$

with $t_0 \in A_{p-1}^*$, $t_i \in X(p-1, m-i)$, for $i \in [1, m-1]$, and $t_m \in A_p^*$. We wish to prove that $s\|\mathcal{A}_p\| \geq m$. Let c be a victorious path spelling s in $\mathcal{A}_p$. Then, $\|c\| = s\|\mathcal{A}_p\|$. Let w be the terminus of c. If $w = u_p$ then $\|c\| = |s|_p \geq m$ and we are done. If, on the other hand, $w = v_p$ then c has a factorization

$$c : u_p \xrightarrow{s_1} u_p \xrightarrow{p} u_{p-1} \xrightarrow{s_2} v_p,$$

for some s_1 and s_2, with $s = s_1 p s_2$. If $|s_1|_p \geq m-1$ then $\|c\| \geq |s_1|_p + 1 \geq m$, hence we are done; otherwise, for some $i \in [1, m-1]$, $s_1 = t_0 p t_1 \dots p t_{i-1}$ and $s_2 = t_i p t_{i+1} \dots p t_m$. Then, for some states w_1 and w_2, and for s_3 such that $s_2 = t_i p s_3$, we have the following factorization of c:

$$c : u_p \xrightarrow{s_1} u_p \xrightarrow{p} u_{p-1} \xrightarrow{t_i} w_1 \xrightarrow{p} w_2 \xrightarrow{s_2'} v_p.$$

From the construction of $\mathcal{A}_p$ we conclude that $w_2 = v_p$ and that w_1 is either u_{p-1} or v_{p-1}. In both cases, the hypothesis $t_i \in X(p-1, m-i)$ implies that the multiplicity of the segment of c spelling t_i is at least $m - i$. Computing the multiplicity of c we have:

$$s\|\mathcal{A}_p\| = \|c\| \geq |s_1|_p + 1 + m - i = m.$$

This completes the proof of the recurrence relations.

To finish the proof of the proposition we observe that $\mathrm{sh}(\|\mathcal{A}_p\|, m)$ is just the length of a shortest word in $X(p, m)$; thus,

$$\begin{aligned}
\mathrm{sh}(\|\mathcal{A}_1\|, m) &= m, \quad (m \geq 0) \\
\mathrm{sh}(\|\mathcal{A}_p\|, 0) &= 0, \quad (p > 1) \\
\mathrm{sh}(\|\mathcal{A}_p\|, m) &= m + \sum_{i=1}^{m-1} \mathrm{sh}(\|\mathcal{A}_{p-1}\|, m-i), \quad (p > 1,\ m \geq 1) \qquad (2)
\end{aligned}$$

Now, we prove by induction on $m + p$ that

$$\mathrm{sh}(\|\mathcal{A}_p\|, m) = \sum_{k=1}^{p} \binom{m}{k}.$$

Indeed, for $p = 1$ or $m = 0$ we have nothing to prove. Let then $p > 1$ and $m \geq 1$. Subtracting $\mathrm{sh}(\|\mathcal{A}_p\|, m-1)$ from (2) we have:

$$\mathrm{sh}(\|\mathcal{A}_p\|, m) - \mathrm{sh}(\|\mathcal{A}_p\|, m-1) = 1 + \mathrm{sh}(\|\mathcal{A}_{p-1}\|, m-1).$$

Then, using the induction hypothesis,

$$\begin{aligned}
\mathrm{sh}(\|\mathcal{A}_p\|, m) &= 1 + \sum_{k=1}^{p} \binom{m-1}{k} + \sum_{k=1}^{p-1} \binom{m-1}{k} = \\
&= \binom{m-1}{0} + \sum_{k=1}^{p} \binom{m-1}{k} + \sum_{k=2}^{p} \binom{m-1}{k-1} = \\
&= \sum_{k=1}^{p} \binom{m-1}{k} + \sum_{k=1}^{p} \binom{m-1}{k-1} = \\
&= \sum_{k=1}^{p} \binom{m}{k}.
\end{aligned}$$

This completes the proof of our proposition. □

Corollary 2 *For any $p \geq 1$, the nondeterministic complexity of $\mathcal{A}_p$ satisfies:*

$$\mathrm{sh}(\|\mathcal{A}_p\|, m) \in \Theta(m^p) \qquad \textit{and} \qquad \mathrm{cm}(\|\mathcal{A}_p\|, k) \in \Theta(k^{1/p}).$$

Proof. The theorem implies that $\mathrm{sh}(\|\mathcal{A}_p\|, m)$ is a poynomial of degree p in m. Thus, $\mathrm{sh}(\|\mathcal{A}_p\|, m) \in \Theta(m^p)$. The other property follows from the relationships of Section 3 between $\mathrm{sh}(\|\mathcal{A}_p\|, m)$ and $\mathrm{cm}(\|\mathcal{A}_p\|, k)$. □

Automaton $\mathcal{A}_p$ is over a p-letter alphabet, but the same effect can also be obtained for a two letter alphabet. This is done by codifying the $\mathcal{M}$-subsets $\|\mathcal{A}_p\|$ of A_p^* over a two letter alphabet $B = \{a, b\}$. We shall do this maintaining recognizability. Even more is true: the codified $\mathcal{M}$-subsets will be the nondeterministic complexity functions of appropriate automata. At the same time the sh functions of the original and codified $\mathcal{M}$-subsets will turn out to be Θ-equivalent.

More precisely, let us fix $p \geq 1$ and let $\kappa_p \colon A_p^* \to B^*$ be the morphism defined by

$$q\kappa_p = ab^q a \qquad (1 \leq q \leq p).$$

Then $A_p\kappa_p$ is a finite biprefix code and κ_p is injective [1]. Hence, for $t \in B^*$, $t\kappa_p^{-1}$ is either empty or it is a singleton.

The image $F\kappa$ of a K-subset F of A^* by a nonerasing morphism κ can be defined in general and it is known that if F is recognizable then so is $F\kappa$ (see [3, page 140] or [2, page 15]). Some care is needed with this notation since $F\kappa$ is not the composition of F with κ.

In our case, let $F\colon A_p^* \to \mathcal{M}$ be an $\mathcal{M}$-subset of A_p^*. Then, since κ_p is injective, $F\kappa_p\colon B^* \to \mathcal{M}$ is given by

$$t(F\kappa_p) = \begin{cases} (t\kappa_p^{-1})F & \text{if } t \in A_p^*\kappa_p \\ \infty & \text{otherwise} \end{cases} \qquad (t \in B^*).$$

Proposition 3 *Let $p \geq 1$ be an integer and let F be an $\mathcal{M}$-subset of A_p^*. Then*

$$\mathrm{sh}(F\kappa_p, m) \in \Theta(\mathrm{sh}(F, m)).$$

Proof. Note that κ_p is a code which satisfies, for any $a \in A_p$,

$$3 \leq |a\kappa_p| \leq p + 2.$$

Then, clearly, for any $m \geq 0$, $\mathrm{sh}(F, m)$ is defined iff $\mathrm{sh}(F\kappa_p, m)$ is defined; further, if they are both defined then

$$3\,\mathrm{sh}(F, m) \leq \mathrm{sh}(F\kappa_p, m) \leq (p + 2)\,\mathrm{sh}(F, m).$$

This proves the proposition. □

We mention an interesting problem. The statement of Proposition 3 is not true if we substitute the sh function by its "inverse" cm. Is it true for cm if we require F to be recognizable?

Proposition 4 *Let $p \geq 1$ be an integer and let $\mathcal{A}$ be an A_p-automaton. Then there exists a B-automaton $\mathcal{B}$ such that $\|\mathcal{B}\| = \|\mathcal{A}\|\kappa_p$, where $\|\mathcal{A}\|$ and $\|\mathcal{B}\|$ are the nondeterministic complexity functions of $\mathcal{A}$ and $\mathcal{B}$ respectively.*

Proof. Assume that $\mathcal{A} = (Q, I, T)$, with $Q = \{1, 2, \ldots, n\}$ for some n. We put $\mathcal{B} = (Q_{\mathcal{B}}, I, T)$, with

$$Q_{\mathcal{B}} = Q \cup Q \times \{0, 1, \ldots, p\}.$$

For each $q \in Q$, $\mathcal{B}$ has the edges

$$E_q = \{(q, a, (q, 0))\} \cup \{((q, i), b, (q, i + 1)) \mid 0 \leq i < p\}.$$

Further, for each edge (q, i, r) of $\mathcal{A}$, $\mathcal{B}$ has an edge $((q, i), a, r)$. This construction amounts to substituting each state of $\mathcal{A}$ by the configuration represented in Figure 2, where we also show an edge in $\mathcal{A}$ and its corresponding edge in $\mathcal{B}$.

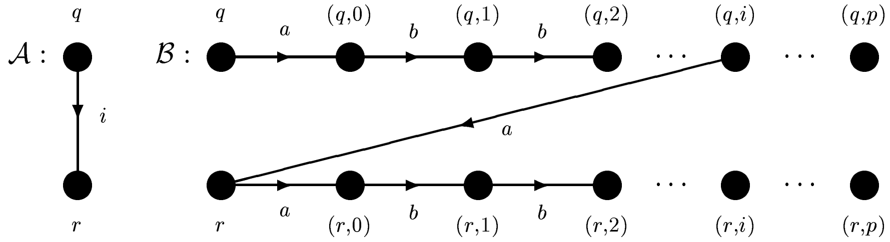

Figure 2: States and edges of $\mathcal{A}$ and $\mathcal{B}$

Note that all edges in E_q are deterministic and that the edge $((q,i),a,r)$ of $\mathcal{B}$ is deterministic iff the edge (q,i,r) of $\mathcal{A}$ is deterministic. Further, for $q,r \in Q$, $\mathcal{A}$ has an edge (q,i,r) iff $\mathcal{B}$ has a path c from q to r with label $|c| = ab^i a$.

These observations imply that for every $s \in A_p^*$ and for every $q,r \in Q$ there exists a bijection between the (q,r)-paths in $\mathcal{A}$ with label s and the (q,r)-paths in $\mathcal{B}$ with label $s\kappa_p$. Further, this bijection preserves the multiplicities. This implies the claim of the proposition. $\square$

Finally, the results of this section can be summarized in the following theorem.

Theorem 5 *Let B be a two letter alphabet. For any $p \geq 1$ there exists a B-automaton $\mathcal{B}$ whose nondeterministic complexity $\|\mathcal{B}\|$ satisfies*

$$\mathrm{sh}(\|\mathcal{B}\|, m) \in \Theta(m^p) \quad \textit{and} \quad \mathrm{cm}(\|\mathcal{B}\|, k) \in \Theta(k^{1/p}).$$

5 A hierarchy of recognizable $\mathcal{M}$-subsets

We fix an alphabet A and define a family $\mathcal{H}_p$ $(p \geq 0)$ of subsets of $\mathrm{Rec}_{\mathcal{M}} A^*$ as follows:

$$\mathcal{H}_p = \{ F \in \mathrm{Rec}_{\mathcal{M}} A^* \mid \mathrm{sh}(F,m) \in O(m^p) \}.$$

Thus, $\mathcal{H}_p$ is the family of F's for which the function $\mathrm{sh}(F,m)$ grows not faster than a p-degree polynomial in m. In particular, $\mathcal{H}_0$ is the family of limited recognizable $\mathcal{M}$-subsets of A^*.

In the sequel we will show that for an alphabet with at least two letters the sequence $\mathcal{H}_p$ forms a hierarchy which exhausts $\mathrm{Rec}_{\mathcal{M}} A^*$. However, for a one letter alphabet the hierarchy collapses and $\mathrm{Rec}_{\mathcal{M}} A^* = \mathcal{H}_0 \cup \mathcal{H}_1$. In particular, our results imply that Goldstine's problem, stated in the Introduction, has an affirmative answer only for one letter alphabets.

Initially we show that every F in $\mathrm{Rec}_{\mathcal{M}}\, A^*$ belongs to some $\mathcal{H}_p$. To do this we use a characterization of limited $\mathcal{M}$-subsets obtained recently by Hashiguchi [5]. This characterization makes use of certain sequences of words which can be described in terms of rational expressions which do not use the union operation.

A *multiplicative rational expression over an alphabet* A is defined inductively:

(1) the empty word 1 is a multiplicative rational expression,

(2) any letter a in A is a multiplicative rational expression,

(3) if e_1 and e_2 are multiplicative rational expressions then $(e_1 e_2)$ is a multiplicative rational expression,

(4) if e_1 is a multiplicative rational expression then $(e_1)^*$ is a multiplicative rational expression.

As usual we dispense with unnecessary parenthesis and associate a subset of A^* to each multiplicative rational expression. Further, to each multiplicative rational expression e we associate its height eh, a natural number that measures the largest number of nested applications of the star operation. We shall alternatively look at e as a function

$$e\colon\ \mathbb{N} \to A^*,$$

where me is the m-th term of the sequence of words of our interest.

We define eh and me inductively:

(1) if $e = 1$ then $eh = 0$ and $me = 1$ for every $m \geq 0$;

(2) if $e = a$, for $a \in A$, then $eh = 0$ and $me = a$ for every $m \geq 0$;

(3) if $e = e_1 e_2$ then $eh = \max\{\, e_1 h, e_2 h \,\}$ and $me = (me_1)(me_2)$ for every $m \geq 0$;

(4) if $e = e_1^*$ then $eh = 1 + e_1 h$ and $me = (me_1)^m$ for every $m \geq 0$.

It can be seen that me is a word in the subset of A^* denoted by e, more precisely, me is the word obtained by replacing every occurrence of $*$ in e by the exponent m. These concepts are related by the following lemma.

Lemma 6 *For every multiplicative rational expression e there exists a polynomial $P_e(m)$, of degree eh, such that, for every $m \geq 0$, $|me| = P_e(m)$.*

Proof. It suffices to define $P_e(m)$ as follows:

(1) if $e = 1$ then $P_e(m) = 0$;

(2) if $e = a$, for $a \in A$, then $P_e(m) = 1$;

(3) if $e = e_1e_2$ then $P_e(m) = P_{e_1}(m) + P_{e_2}(m)$;

(4) if $e = e_1^*$ then $P_e(m) = mP_{e_1}(m)$.

The lemma is now established by a simple inductive argument. □

To state Hashiguchi's theorem we need to restrict the family of $\mathcal{M}$-A-automata. We say that an $\mathcal{M}$-A-automaton is *simple* if every edge of $\mathcal{A}$ has its multiplicity in the set $\{0, 1, \infty\}$.

Theorem 7 (Hashiguchi) *The behavior of a simple $\mathcal{M}$-A-automaton $\mathcal{A}$ is not limited iff there exists a multiplicative rational expression e over A such that, for every $m \geq 1$, $m \leq me\|\mathcal{A}\| < \infty$.*

This deep theorem is proved in [5]. Shortly we will publish a new, possibly simpler, proof of it. Our proof will also contain an algorithm to decide whether $\|\mathcal{A}\|$ is limited or not and if it is not then we will exhibit explicitly a multiplicative rational expression which proves that $\|\mathcal{A}\|$ is not limited.

To illustrate Hashiguchi's theorem the reader can verify that the multiplicative rational expressions e_p satisfy the statement of that theorem for the nondeterministic complexity functions $\|\mathcal{A}_p\|$ of Proposition 1, where

$$e_1 = 1^* \quad \text{and} \quad e_p = (pe_{p-1})^* \quad (p > 1).$$

Now we remove the hypothesis of simplicity in Hashiguchi's theorem. To do this we define the *projection* of an $\mathcal{M}$-A-automaton $\mathcal{A} = (Q, I, T)$ as being the $\mathcal{M}$-A-automaton $\mathcal{B} = (Q, I, T)$ in which, for every $(p, a, q) \in Q \times A \times Q$,

$$(p, a, q)E_{\mathcal{B}} = \begin{cases} 0 & \text{if } (p, a, q)E_{\mathcal{A}} = 0, \\ 1 & \text{if } 0 < (p, a, q)E_{\mathcal{A}} < \infty, \\ \infty & \text{if } (p, a, q)E_{\mathcal{A}} = \infty. \end{cases}$$

Clearly the projection of $\mathcal{A}$ is a simple $\mathcal{M}$-A-automaton. Further, for $\mathcal{A}$ we denote by $\Delta_{\mathcal{A}}$ the maximum finite multiplicity of edges of $\mathcal{A}$. Formally,

$$\Delta_{\mathcal{A}} = \max(\{0\} \cup (Q \times A \times Q)E_{\mathcal{A}} - \infty).$$

Proposition 8 *Let $\mathcal{A}$ be an $\mathcal{M}$-A-automaton and let $\mathcal{B}$ be its projection. Then, for every $s \in A^*$,*

$$s\|\mathcal{B}\| \leq s\|\mathcal{A}\| \leq \Delta_{\mathcal{A}} s\|\mathcal{B}\|.$$

Proof. To see that $s\|\mathcal{B}\| \leq s\|\mathcal{A}\|$ consider a victorious path c in $\mathcal{A}$ with label s. Then,

$$s\|\mathcal{B}\| \leq \|c\|_{\mathcal{B}} \leq \|c\|_{\mathcal{A}} = s\|\mathcal{A}\|.$$

Conversely, let c be a victorious path in $\mathcal{B}$ with label s. Then,

$$s\|\mathcal{A}\| \leq \|c\|_{\mathcal{A}} \leq \Delta_{\mathcal{A}}\|c\|_{\mathcal{B}} = \Delta_{\mathcal{A}} s\|\mathcal{B}\|.$$

This completes the proof of the proposition. □

Corollary 9 *A recognizable $\mathcal{M}$-subset F of A^* is not limited iff there exists a multiplicative rational expression e over A such that, for every $m \geq 1$, $m \leq meF < \infty$.*

Proof. Clearly the existence of an expression e, as asserted, implies that F is not limited. Conversely, let $\mathcal{A}$ be an $\mathcal{M}$-A-automaton with behavior F and let $\mathcal{B}$ be the projection of $\mathcal{A}$. If $\|\mathcal{A}\|$ is not limited then, by Proposition 8, $\|\mathcal{B}\|$ is not limited either. Since $\mathcal{B}$ is simple there exists, by Theorem 7, a multiplicative rational expression e such that, for every $m \geq 1$, $m \leq me\|\mathcal{B}\| < \infty$. By Proposition 8, for every $m \geq 1$,

$$m \leq me\|\mathcal{B}\| \leq me\|\mathcal{A}\| = meF \leq \Delta_{\mathcal{A}} me\|\mathcal{B}\| < \infty.$$

This proves the corollary. □

Proposition 10 *For every recognizable $\mathcal{M}$-subset F of A there exists a $p \geq 0$ such that $F \in \mathcal{H}_p$.*

Proof. If F is limited then $F \in \mathcal{H}_0$, else there exists, by Corollary 9, a multiplicative rational expression e over A such that, for every $m \geq 1$, $m \leq meF < \infty$. Then, for every $m \geq 1$, $\mathrm{sh}(F, m) \leq |me|$. By Lemma 6 there exists a polynomial $P_e(m)$, of degree $p = eh$, such that, for every $m \geq 0$, $|me| = P_e(m)$. Hence, $\mathrm{sh}(F, m) \in O(m^p)$ and $F \in \mathcal{H}_p$. □

Finally, we show that for a one letter alphabet A, $\mathrm{Rec}_{\mathcal{M}}\, A^* = \mathcal{H}_0 \cup \mathcal{H}_1$. Initially we observe that in this case the behavior of an $\mathcal{M}$-A-automaton is completely determined by the subsemigroup of $M_n\mathcal{M}$ generated by one matrix. We study this in the next proposition where all undefined notations can be found in [8].

Lemma 11 *For every matrix x in $M_n\mathcal{M}$, there exist $c_1, c_2, c_3 > 0$ such that, for every $k \geq 1$ and for every $1 \leq i, j \leq n$, if $(i, x^k, j) < \infty$ then either $(i, x^k, j) \leq c_1$ or $c_2 k \leq (i, x^k, j) \leq c_3 k$. Further, there exist integers $k_0, d \geq 1$ such that, for every $k \geq k_0$ and for every $1 \leq i, j \leq n$, if $c_1 < (i, x^k, j) < \infty$ then $c_1 < (i, x^{k+d}, j) < \infty$.*

Proof. We observe initially that $x\Psi$ is a matrix in the finite monoid $M_n\mathcal{N}$. Hence, for some $d \geq 1$, $f = x^d\Psi$ is idempotent. Let

$$e = x^d, \qquad M = \{\, x^q \mid 0 \leq q \leq d \,\}\Delta,$$

$$c_1 = 3M, \quad c_2 = 1/(2d) \quad \text{and} \quad c_3 = 2M/d.$$

Let us fix i and j, such that $1 \leq i, j \leq n$ and let us consider $k \geq 1$. Let p and q be integers such that $k = pd + q$ and $0 \leq q < d$. Then, $x^k = e^p x^q$. Let $1 \leq l \leq n$ be such that

$$(i, x^k, j) = (i, e^p, l) + (l, x^q, j).$$

If $(i, x^k, j) = \infty$ then we have nothing to prove; thus we assume $(i, x^k, j) < \infty$. Consequently, $(i, e^p, l) < \infty$ and $(l, x^q, j) < \infty$. By items (i), (ii) and (v) of Proposition 1 of [8],

$$(e\Psi\imath)^p \leq e^p \leq M(e\Psi\imath)^p,$$

since $e\delta \leq M$ by construction.

On the other hand, by Proposition 6 of [8], applied for $f = e\Psi$, we have

$$\text{either} \quad (i, (e\Psi\imath)^p, l) \leq 2 \quad \text{or} \quad (i, (e\Psi\imath)^p, l) = p.$$

Thus,

$$\text{either} \quad (i, e^p, l) \leq 2M \quad \text{or} \quad p \leq (i, e^p, l) \leq Mp.$$

Since $(l, x^q, j) \leq M$ it follows that

$$\text{either} \quad (i, x^k, j) \leq 2M + M = c_1 \quad \text{or} \quad p \leq (i, x^k, j) \leq Mp + M.$$

For $k < d$, $(i, x^k, j) \leq c_1$ by the choice of c_1. Let us assume that $k \geq d$. Since $p = \lfloor k/d \rfloor$ it follows that $k/(2d) \leq p$ and $Mp + M \leq 2Mk/d = c_3 k$. Thus,

$$\text{either} \quad (i, x^k, j) \leq c_1 \quad \text{or} \quad c_2 k \leq (i, x^k, j) \leq c_3 k.$$

Finally, let us assume that $k \geq c_1 d + 3d + 1 = k_0$ and that $c_1 < (i, x^k, j) < \infty$. Then, $p \geq 3 + c_1$, $(i, (e\Psi\imath)^p, l) = p$ and, from Proposition 6 of [8],

$(i, (e\Psi\imath)^{p+1}, l) = p + 1$. It follows that $(i, x^{k+d}, j) < \infty$. Assume, for a contradiction, that $(i, x^{k+d}, j) \leq c_1$. Then, there exists $1 \leq l' \leq n$, such that

$$(i, x^{k+d}, j) = (i, e^{p+1}, l') + (l', x^q, j).$$

If $(i, (e\Psi\imath)^p, l') \leq 2$ then, arguing as before, $(i, x^k, j) \leq 2M + M = c_1$, a contradiction. Hence, $(i, (e\Psi\imath)^p, l') = p \geq 3$. By Proposition 6 of [8], $(i, (e\Psi\imath)^{p+1}, l') = p + 1$, and it follows that

$$c_1 < p + 1 \leq (i, x^{k+d}, j) \leq c_1,$$

a contradiction. This completes the proof of the lemma. □

Proposition 12 *For a one letter alphabet A,*

$$\mathrm{Rec}_{\mathcal{M}}\, A^* = \mathcal{H}_0 \cup \mathcal{H}_1.$$

Proof. Let $\mathcal{A} = (Q, I, T)$ be an $\mathcal{M}$-A-automaton, where $A = \{a\}$. Without loss of generality we may assume that $\mathcal{A}$ is normalized in the sense of Eilenberg [3, page 138], hence both I and T are singletons. Then, performing the construction of Section 2, there results a matrix $x \in M_n\mathcal{M}$ and indices $1 \leq i, j \leq n$, such that, for every $k \geq 1$,

$$a^k \|\mathcal{A}\| = (i, x^k, j).$$

If $\|\mathcal{A}\|$ is limited then $\|\mathcal{A}\| \in \mathcal{H}_0$. Assume now that $\|\mathcal{A}\|$ is not limited and let us apply the previous lemma. Let k_1 be the least integer $k_1 \geq k_0$, such that $c_1 < (i, x^{k_1}, j) < \infty$. Then, for every $r \geq 1$, $c_1 < (i, x^{k_1+rd}, j) < \infty$; hence,

$$c_2(k_1 + rd) \leq (i, x^{k_1+rd}, j) \leq c_3(k_1 + rd).$$

Thus, for every $m \geq c_2k_1$ there exists an $r \geq 0$ such that

$$m/c_2 \leq k_1 + rd < m/c_2 + d.$$

Then,

$$m \leq c_2(k_1 + rd) \leq (i, x^{k_1+rd}, j) = a^{k_1+rd}\|\mathcal{A}\|.$$

Hence,

$$\mathrm{sh}(\|\mathcal{A}\|, m) \leq k_1 + rd < m/c_2 + d.$$

This implies that $\mathrm{sh}(\|\mathcal{A}\|, m) \in O(m)$; i.e. $\|\mathcal{A}\| \in \mathcal{H}_1$, as required. □

We finish the paper by summarizing its main results in a theorem.

Theorem 13 *For an alphabet A with at least two letters*

$$\mathrm{Rec}_{\mathcal{M}}\, A^* = \bigcup_{p \geq 0} \mathcal{H}_p,$$

and for every $p \geq 1$ there exists a nondeterministic complexity function in $\mathcal{H}_p - \mathcal{H}_{p-1}$.

For a one letter alphabet A,

$$\mathrm{Rec}_{\mathcal{M}}\, A^* = \mathcal{H}_0 \cup \mathcal{H}_1.$$

ACKNOWLEDGEMENTS. This work first took form in Paris during a visit to the Laboratoire d'Informatique Thèorique et Programmation, which friendly atmosphere stimulated the author. I thank especially George Hansel for permission and help to use his programming system to compute monoids of matrices, which had a definite influence on the results reported here. I also thank Christian Choffrut and Nami Kobayashi for useful discussions. Finally, I thank Yoshiharu Kohayakawa, Mario Setsuo Sakumoto and Tomasz Kowaltowski for advice in typesetting the manuscript and the Centro de Computação da Unicamp for permission to use their laser printer.

References

[1] J. Berstel and D. Perrin, *Theory of Codes,* Academic Press, New York, 1985.

[2] J. Berstel and C. Reutenauer, *Les Séries Rationnelles et leurs Langages,* Masson, Paris, 1984.

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

[4] K. Hashiguchi, Limitedness Theorem on Finite Automata with Distance Functions, *J. Comput. System Sci.* 24:2 (1982), 233–244.

[5] K. Hashiguchi, Improved Limitedness Theorems on Finite Automata with Distance Functions, manuscript, 1986.

[6] C. M. R. Kintala and P. Fischer, Computations with a Restricted Number of Nondeterministic Steps, in *Proc. of the Ninth Annual ACM Symposium on Theory of Computing,* Association for Computing Machinery, New York, 1977, 178–185.

[7] J. P. Mascle, Torsion Matrix Semigroups and Recognizable Transductions, in L. Kott (ed.), *Automata, Languages and Programming,* Lecture Notes in Computer Science, 226, Springer-Verlag, Berlin, 1986, 244–253.

[8] I. Simon, Limited Subsets of a Free Monoid, in *Proc. 19th Annual Symposium on Foundations of Computer Science,* Institute of Electrical and Electronics Engineers, Piscataway, N.J., 1978, 143–150.

[9] J. Simon, private communication.