RAIRO. INFORMATIQUE THÉORIQUE

PHILIPPE GOHON

Automates de coût borné sur un alphabet à une lettre

RAIRO. Informatique théorique, tome 19, n° 4 (1985), p. 351-357
<http://www.numdam.org/item?id=ITA_1985__19_4_351_0>

R.A.I.R.O. Informatique théorique/Theoretical Informatics
(vol. 19, n° 4, 1985, p. 351 à 357)

AUTOMATES DE COÛT BORNÉ SUR UN ALPHABET A UNE LETTRE (*)

par Philippe GOHON ([1])

Communiqué par J.-E. PIN

Résumé. — *Soit A un automate à n états sur un alphabet à une seule lettre et muni d'une fonction de coût (appelée « distance function » par K. Hashiguchi). Si son coût* $c(A)$ *est borné, alors* $c(A) < n^2$. *De plus, cette borne est asymptotiquement optimale en ce sens qu'il existe une famille* $\{A_n\}$ *d'automates, où* A_n *possède n états et telle que* $\lim_{n \to +\infty} c(A_n)/n^2 = 1$.

Abstract. — *Let A be an automaton with n states on a single letter alphabet and having a cost function (called "distance function" by K. Hashiguchi). If its cost* $c(A)$ *is bounded, then* $c(A) < n^2$. *Moreover, this bound is asymptotically optimal in the sense that there exists a family* $\{A_n\}$ *of automata, where* A_n *has n states and such that* $\lim_{n \to +\infty} c(A_n)/n^2 = 1$.

INTRODUCTION

Dans son article « Limitedness theorem on finite automata with distance functions », K. Hashiguchi a montré que pour un automate A muni d'une fonction de coût c, si $c(A)$ est fini, alors $c(A) \leqq m(m+1)^{3m^2} . 2^{m(6m^2+1)}$ où $m = 2^n$, n étant le nombre d'états de l'automate. Le but du présent article est de prouver que si l'alphabet sur lequel est construit l'automate ne possède qu'une seule lettre, alors $c(A) < \infty$ implique $c(A) < n^2$. De plus, cette borne est asymptotiquement optimale en ce sens qu'il existe une famille $\{A_n\}$ d'automates, où A_n possède n états et telle que $\lim_{n \to +\infty} c(A_n)/n^2 = 1$.

(*) Reçu février 1984, révisé janvier 1985.

([1]) Université de Haute-Normandie, Faculté des Sciences et des Techniques, Laboratoire d'Informatique, B.P. n° 67, 76130 Mont-Saint-Aignan.

0399-0540/85/04 351 07/$ 2.70/

I. DÉFINITIONS ET NOTATIONS

Nous supposons connue la notion d'automate fini sur un alphabet. Toutefois, comme nous ne travaillerons qu'avec des automates sur un alphabet à une seule lettre, nous allons, pour simplifier, modifier légèrement certaines définitions et notations.

Nous appellerons *automate* un quadruplet $A=(Q, I, T, F)$ où Q est un ensemble fini, I et T deux parties de Q et F une partie de $Q \times Q$. Les éléments de Q, I, T et F sont appelés respectivement états, états initiaux, états terminaux et flèches.

Un *chemin* de A est une suite finie $(q_0, q_1, \ldots, q_m)$ d'états telle que (q_i, q_{i+1}) appartient à F pour tout i $(0 \leqq i \leqq m-1)$. Le nombre m est la *longueur* du chemin. Si q_0 appartient à I et q_m à T, on dit que le chemin est *réussi*. Un entier positif ou nul m est dit *reconnu* par A s'il existe un chemin réussi de longueur m. On note $L(A)$ l'ensemble des entiers reconnus par A; on sait que $L(A)$ est une partie reconnaissable de $\mathbb{N}$. Pour une partie Q' de Q, on dit que le chemin $(q_0, q_1, \ldots, q_m)$ *passe* par Q', s'il existe $i\,(0 \leqq i \leqq m)$ tel que q_i appartienne à Q'. Si $x=(q_0, q_1, \ldots, q_m)$ et $y=(q_m, q_{m+1}, \ldots, q_{m+m'})$, alors xy notera le chemin $(q_0, q_1, \ldots, q_{m+m'})$.

Une *fonction de coût* c définie sur A est une application de $Q \times Q$ dans $\mathbb{N} \cup \{\infty\}$ vérifiant :

— si $(q_1, q_2) \in F$, $c(q_1, q_2) \in \{0, 1\}$;

— si $(q_1, q_2) \notin F$, $c(q_1, q_2) = \infty$.

Cela permet de définir le coût d'un chemin quelconque par : $c(q_0, q_1, \ldots, q_m) = \sum_{i=0}^{m-1} c(q_i, q_{i+1})$, (avec évidemment $a+\infty=\infty+a=\infty$ pour tout a de $\mathbb{N} \cup \{\infty\}$). On posera comme étant nul le coût de tout chemin de longueur zéro.

Pour un entier m reconnu par l'automate A, le coût de m, noté $c(m)$, est égal à :

$$\inf \{c(q_0, q_1, \ldots, q_m)/q_0 \in I, q_m \in T, q_i \in Q \; (1 \leqq i \leqq m-1)\}.$$

On posera : $c(A) = \sup \{c(m)/m \in L(A)\}$. On dira que A est de coût borné lorsque $c(A) < \infty$. Sur les dessins, la flèche (p, q) sera notée $(p) \rightarrow (q)$ si $c(p, q)=0$ *et* $(p) \Rightarrow (q)$ si $c(p, q)=1$.

II. PROPRIÉTÉS ÉLÉMENTAIRES DES PARTIES RECONNAISSABLES DE $\mathbb{N}$

On rappelle qu'un langage L défini sur un monoïde libre Σ^* est reconnaissable et qu'il est reconnu par un monoïde fini M s'il existe un morphisme φ de Σ^* dans M tel que $L = \bar{\varphi}^1(\varphi(L))$.

Notons $Z(p, s)$ le monoïde fini à un seul générateur τ, possédant $(s+p-1)$ éléments et tel que $\tau^{p+s}=\tau^s$.

On sait (*voir* [1], p. 100-103) qu'une partie de $\mathbb{N}$ est reconnaissable si et seulement si elle est reconnue par un monoïde $Z(p, s)$. La propriété : L reconnu par $M \Leftrightarrow L=\bar{\varphi}^1(\varphi(L))$ se traduit alors par : L reconnu par $Z(p, s) \Leftrightarrow (\forall m \geqq s,\ m \in L \Leftrightarrow m+p \in L)$.

De plus, on a les deux propriétés suivantes :

PROPRIÉTÉ 1 : Soit L un langage fini sur $\mathbb{N}$ reconnu par le monoïde $Z(p, s)$. On a alors : $\forall m \in L,\ m \leqq s-1$.

Preuve : Si on avait $m \in L$ avec $m \geqq s$, on aurait $(m+\lambda p) \in L$ pour tout entier positif λ, d'où L infini, ce qui est contraire à l'hypothèse.

PROPRIÉTÉ 2 : Soient L_1 et L_2 deux langages reconnaissables de $\mathbb{N}$, reconnus respectivement par les monoïdes $Z(p_1, s_1)$ et $Z(p_2, s_2)$. Alors, $L_1 \cup L_2$ et $L_1 \setminus L_2$ sont reconnus par le monoïde $Z(p_1 p_2, \sup(s_1, s_2))$.

Preuve : Remarquons d'abord que si un langage L est reconnu par le monoïde $Z(p, s)$, il l'est également par $Z(\mu p, s')$ avec $s' \geqq s$ et μ entier strictement positif. En effet, comme L est reconnu par $Z(p, s)$, on a : $\forall m \geqq s$, $m \in L \Leftrightarrow m+p \in L$. On en déduit immédiatement : $\forall m \geqq s'$, $m \in L \Leftrightarrow m+\mu p \in L$, et donc L reconnu par $Z(\mu p, s')$.

Les langages L_1 et L_2 sont donc tous les deux reconnus par $Z(p_1 p_2, \sup(s_1, s_2))$. Il est alors aisé de vérifier que :

$$\forall m \geqq \sup(s_1, s_2), \quad m \in L_1 \cup L_2 \quad \Leftrightarrow \quad m+p_1 p_2 \in L_1 \cup L_2,$$

et que :

$$\forall m \geqq \sup(s_1, s_2), \quad m \in L_1 \setminus L_2 \quad \Leftrightarrow \quad m+p_1 p_2 \in L_1 \setminus L_2.$$

III. ÉTUDE DE L'ENSEMBLE DES COÛTS ASSOCIÉS A CERTAINS CHEMINS

Donnons d'abord une définition. Soit A un automate. Nous appellerons cycle tout chemin $(q_0, q_1, \ldots, q_p)$ tel que :

- $q_0=q_p$;
- $\forall i\,(1 \leqq i \leqq p),\ \forall j\,(1 \leqq j \leqq p)\ i \neq j \Rightarrow q_i \neq q_j$.

Si A est muni d'une fonction de coût, le cycle $(q_0, q_1, \ldots, q_p)$ est dit de *coût nul* si : $\forall i\,(0 \leqq i \leqq p-1)\ c(q_i, q_{i+1})=0$.

Remarque : Nous ne ferons pas de distinction entre les cycles $(q_0, q_1, \ldots, q_p)$ et $(q_i, q_{i+1}, \ldots, q_p, q_1, q_2, \ldots, q_{i-1}, q_i)$.

Le but de ce paragraphe est de prouver la proposition suivante :

PROPOSITION 3 : *Soient* A *un automate à* n *états, muni d'une fonction de coût* c *et* C *un cycle de* A, *de coût nul et de longueur* p. *Alors, tout entier* m *reconnu par* A *et dont au moins un chemin réussi passe par* C *est tel que :* $c(m)<n^2$.

Pour prouver cela, nous aurons besoin du lemme 4 et de la proposition 5.

LEMME 4 : *Soient* A *un automate à* n *états et* C *un cycle de* A *de longueur* p. *Alors, pour tout chemin* $w=(q_0, q_1, \ldots, q_m)$ *de longueur* m *supérieure ou égale à* $p(n-1)+n$ *et passant par* C, *il existe un chemin* $w'=(q'_0, q'_1, \ldots, q'_m)$ *de longueur* m *et tel que :*

— $q'_0=q_0$ *et* $q'_m=q_m$;

— *il existe trois chemins* x, y *et* z *tels que :* $w'=xyz$, x *ou* z *passe par* C *et* y *passe* $(p+1)$ *fois par un même état.*

Preuve : On a $m+1\geqq p(n-1)+n+1=p^2+(p+1)(n-p)+1$. Il existe alors un état q qui vérifie l'une des deux affirmations suivantes :

(1) q appartient à C et w passe (au moins) $(p+1)$ fois par q;

(2) q n'appartient pas à C et w passe (au moins) $(p+2)$ fois par q.

Si (1) est vérifiée, on a : $w=uv_1 v_2 \ldots v_p u'$, les extrémités des chemins v_i $(1\leqq i\leqq p)$ étant égales à q. Il suffit alors de prendre $x=u$, $y=v_1 v_2 \ldots v_p$ et $z=u'$.

Si (2) est vérifiée, on a : $w=uv_1 v_2 \ldots v_{p+1} u'$, les extrémités des $v_i (1\leqq i\leqq p+1)$ étant égales à q. Si u ou $v_{p+1} u'$ passe par C, il suffit de prendre $x=u$, $y=v_1 v_2 \ldots v_p$ et $z=v_{p+1} u'$. Sinon, il existe un indice $j\,(1\leqq j\leqq p)$ tel que v_j passe par C; on prend alors $x=u$, $y=v_1 v_2 \ldots v_{j-1} v_{j+1} \ldots v_{p+1}$ et $z=v_j u'$.

PROPOSITION 5 (Mêmes hypothèses que le lemme 4) : *Pour tout chemin réussi de longueur* m *supérieure ou égale à* $p(n-1)+n$ *et passant par* C, *il existe un chemin réussi de longueur* m' *strictement inférieure à* $p(n-1)+n$, *passant par* C *et tel que* $m-m'\equiv 0$ *(mod* p*).*

Preuve : D'après le lemme 4, il existe un chemin réussi $w'=xyz$ de longueur m tel que x ou z passe par C et y passe $(p+1)$ fois par un même état q. Posons $w'=(q_0, q_1, \ldots, q_m)$ et appelons $i_1, i_2, \ldots, i_{p+1} (i_1<i_2<\ldots<i_{p+1})$ les entiers vérifiant : $q_{i_1}=q_{i_2}=\ldots=q_{i_{p+1}}=q$. Considérons les $(p+1)$ restes obtenus en divisant par p les entiers $i_1-i_1, i_2-i_1, \ldots, i_{p+1}-i_1$. Il y en a nécessairement deux égaux; il existe donc deux entiers α et $\beta\,(\alpha<\beta)$ tels que $i_\beta-i_\alpha=\lambda p$ (λ entier). Le chemin $w''=(q_0, q_1, \ldots, q_{i_\alpha}, q_{i_\beta+1}, \ldots, q_m)$ est de longueur $(m-\lambda p)$ et passe par C. Si $(m-\lambda p)$ est strictement inférieur à $p(n-1)+n$, c'est fini, sinon on recommence le processus en partant de w''. On itère ainsi jusqu'à obtention d'un chemin vérifiant les propriétés énoncées.

On est maintenant en mesure d'effectuer la preuve suivante :

Preuve de la proposition 3 : Comme p est inférieur ou égal à n, il va nous suffire de montrer que $c(m)<p(n-1)+n$. Si $m<p(n-1)+n$, le résultat est évident. Supposons donc $m \geqq p(n-1)+n$. Il existe, d'après la proposition 5, un entier μ et un chemin réussi t passant par C et de longueur m' telle que : $m'<p(n-1)+n$ et $m-m'=\mu p$. Le chemin obtenu en insérant dans t le cycle C parcouru μ fois, est de longueur m et de coût strictement inférieur à $p(n-1)+n$. Donc $c(m)<p(n-1)+n$.

IV. PROPRIÉTÉS DES CYCLES DE COÛT NUL

Avant d'aborder le théorème principal, prouvons deux propriétés relatives aux cycles de coût nul.

PROPOSITION 6 : *Soient A un automate à n états, muni d'une fonction de coût c et C un cycle de A, de coût nul et de longueur p. Appelons $D(C)$ l'ensemble des entiers dont au moins un chemin réussi passe par C. $D(C)$ est un langage reconnaissable et il est reconnu par le monoïde $Z(p, n^2-n)$.*

Preuve : Posons $A=(Q, I, T, F)$. On a :

$$D(C)=\bigcup_{q \in C}[L(Q, I, q, F).L(Q, q, T, F)],$$

d'où $D(C)$ reconnaissable. Montrons maintenant que pour $m \geqq p(n-2)+n$, on a :

$$m \in D(C) \Leftrightarrow m+p \in D(C).$$

L'implication $m \in D(C) \Rightarrow m+p \in D(C)$ étant évidente, montrons sa réciproque. On a : $m+p \geqq p(n-2)+n+p=p(n-1)+n$; il existe donc, d'après la proposition 5, un chemin réussi w' de longueur m$'$ et passant par C, avec $m+p-m' \equiv 0 \pmod p$ et $m'<p(n-1)+n \leqq m+p$. En insérant dans w' le cycle C parcouru un nombre adéquat de fois (éventuellement nul), on obtient un chemin passant par C et de longueur m. D'où $m \in D(C)$. Comme $1 \leqq p \leqq n$, on vérifie aisément que $p(n-2)+n \leqq n^2-n$, d'où la conclusion.

PROPOSITION 7 : *Soit A un automate muni d'une fonction de coût c. Si $L(A)$ est infini, et si A ne contient pas de cycle de coût nul, alors $c(A)=\infty$.*

Preuve : Soient n le nombre d'états de l'automate, N un entier quelconque et m un élément de $L(A)$ tel que $m>Nn$. Comme A ne contient pas de cycle de coût nul, tout chemin de longueur supérieure ou égale à n est de coût strictement positif. Donc tout chemin de longueur supérieure à Nn est de coût supérieur à N, et donc $c(m) \geqq N$, d'où $c(A) = \infty$.

V. THÉORÈME PRINCIPAL

Prouvons maintenant le théorème principal.

THÉORÈME : *Soit A un automate à n états et muni d'une fonction de coût c telle que $c(A) < \infty$. On a alors $c(A) < n^2$. De plus, il existe une famille $\{A_n\}$ d'automates, où A_n possède n états, telle que* $\lim_{n \to +\infty} c(A_n)/n^2 = 1$.

Preuve : Si $L(A)$ est fini, on a : $\forall m \in L(A)\ m \leqq n-1$ d'où $c(A) \leqq n-1 < n^2$. Supposons donc $L(A)$ infini; considérons $K = \{C_1, C_2, \ldots, C_k\}$ l'ensemble des cycles de coût nul de A. La proposition 7 permet d'affirmer que $K \neq \varnothing$. Pour chaque cycle $C_i\,(1 \leqq i \leqq k)$, considérons l'ensemble $D(C_i)$ des entiers dont au moins un chemin réussi passe par C_i. Posons $D_0 = \bigcup_{i=1}^{k} D(C_i)$. D'après la proposition 3, on a : $\forall m \in D_0,\ c(m) < n^2$.

Si $L(A) = D_0$, alors c'est fini. Sinon, considérons l'automate A_∞ obtenu en enlevant de A les états (et les flèches qui s'y rattachent) appartenant aux éléments de K. On a :

$$L(A) = D_0 \cup (L(A_\infty) \setminus D_0).$$

D'après la proposition 7, on a :

$$\forall m \in (L(A_\infty) \setminus D_0), \quad m > N(n-1) \quad \Rightarrow \quad c(m) \geqq N.$$

$L(A_\infty) \setminus D_0$ est donc un ensemble fini. En utilisant la propriété 2 et la proposition 6, on montre que D_0 est reconnu par un monoïde $Z(p, n^2 - n)$. (p est un entier dont la valeur nous importe peu ici.) Un résultat connu (dont on peut trouver la preuve dans [3]) indique que si un automate possède r états, alors le langage qu'il reconnaît est reconnu par un monoïde $Z(p', s)$ avec $s \leqq (r-1)^2 + 1$. Comme A_∞ possède au plus $(n-1)$ états, on en déduit à l'aide de la proposition 6, que $L(A_\infty) \setminus D_0$ est reconnu par un monoïde $Z(p'', s')$ avec $s' \leqq \sup(n^2 - n, (n-2)^2 + 1) \leqq n^2$. Il s'en suit (propriété 1) que : $\forall m \in (L(A_\infty) \setminus D_0)\ m \leqq n^2 - 1$, d'où $c(m) < n^2$. Conclusion : $c(A) < n^2$.

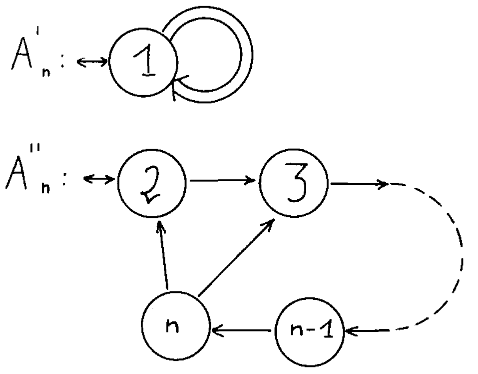
Pour la deuxième partie du théorème, supposons $n \geqq 4$ et considérons l'automate A_n (voir *fig.* 1), union des deux automates A'_n et A''_n.

On a :

$$L(A'_n) = \mathbb{N} \qquad \text{et} \qquad \forall m \in \mathbb{N}, \quad c_{A'_n}(m) = m$$

et :

$$L(A''_n) = \{0\} \cup [(n-1) + M(n-2, n-1)] \qquad \text{et} \qquad \forall m \in L(A''_n), \quad c_{A''_n}(m) = 0.$$



L'automate A_n.

[$M(n-2, n-1)$ désigne le sous-monoïde de $(\mathbb{N}, +)$ engendré par $(n-2)$ et $(n-1)$].

Or, il est prouvé, dans [4] par exemple, que le nombre : $(n-2-1)(n-1-1)-1=n^2-5n+5$ n'appartient pas à $M(n-2, n-1)$, et que : $\forall m \geqq n^2-5n+6 \; m \in M(n-2, n-1)$. On en déduit :

$$n^2-4n+4 \notin L(A_n''), \qquad \text{d'où} \quad c(n^2-4n+4)=n^2-4n+4,$$

et :

$$\forall m \geqq n^2-4n+5 \; m \in L(A_n''), \qquad \text{d'où} \quad \forall m \geqq n^2-4n+5 \; c(m)=0.$$

Donc $n^2-4n+4 \leqq c(A) < \infty$.

Terminons en précisant qu'après la soumission de cet article pour publication, nous avons pu déterminer la borne optimale. Pour information, il s'agit de : $\sup(|n-2|(n-2), [(n^2/4)+n-1])$, borne atteinte pour tout $n \geqq 1$.

BIBLIOGRAPHIE

1. S. Eilenberg, *Automata, Languages and Machines*, Academic Press, Vol. A, 1974.
2. K. Hashiguchi, *Limitedness Theorem on Finite Automata with Distance Functions*, J. Comp. Syst. Sc., Vol. 24, 1982, p. 233-244.
3. G. Markowsky, *Bounds on the Index and Period of a Binary Relation on a finite Set*, Semigroup Forum, Vol. 13, 1977, p. 253-259.
4. E. Selmer, *On the Linear Diophantine Problem of Frobenius*, J. reine angew. Math., Vol. 293/294, 1977, p. 1-17.