Exercices sur les automates et les langages formels

J. Berstel and L. Boasson

6 février 2011 11 h 39

Table des matières

1 Mots

Exercice 1.1 Soient u et v deux mots non vides. Montrer que les conditions suivantes sont équivalentes :

(i) $uv = vu$;
(ii) il existe des entiers $n, m \geq 1$ tels que $u^n = v^m$;
(iii) il existe des entiers $k, \ell \geq 1$ et un mot w tels que $u = w^k$ et $v = w^\ell$.
(iv) les mots infinis u^ω et v^ω ont un préfixe commun de longeur $|u| + |v|$.
(v) il existe des entiers $r, s \geq 1$ et des mots $x_1, x_2, \ldots, x_r, y_1, y_2, \ldots, y_s$, tous égaux à u ou à v, avec $x_1 \neq y_1$, tels que $x_1 x_2 \cdots x_r = y_1 y_2 \cdots y_s$.

Exercice 1.2 (Application du précédent. Soient u et v deux mots non vides. Montrer que si $u^k v^k = v^k u^k$ pour un entier $k \geq 2$, alors $uv = vu$.

Exercice 1.3 Soient x, y, z trois mots non vides. Montrer que si $xy = zx$ et $yx = xz$, alors $y = z$.

Exercice 1.4 Soient x, y, u, v des mots non vides.

1. Montrer que si $xuy = yxu = vyx$, alors $u = v$ et ils sont tous les quatres puissances d'un même mot.

2. Montrer que si $xvy = yxu = vyx$, alors $u = y$ et ils sont tous les quatres puissances d'un même mot.

Exercice 1.5 Deux mots x et y sont *conjugués* s'il exist des mots u, v tels que $x = uv$ et $y = vu$. Pour tout mot $w = a_0a_1 \cdots a_{n-1}$, où $a_1, a_2, \ldots, a_{n-1}$ sont des lettres, on pose

$$w^{[k]} = a_k a_{k+1} \cdots a_{n-1} a_0 a_1 \cdots a_{k-1}\,, \quad (0 \leq k < n)\,.$$

Ce sont les conjugués de w. Pour deux mots x, y de longueur n, on définit

$$D(x, y) = \{k \mid \exists j \,:\, y^{[j]} < x^{[k]}\}\,.$$

(L'ordre est l'ordre lexicographique.) Montrer que x, y sont conjugués si et seulement si $D(x, y) \neq \{0, \ldots, n-1\}$ et $D(y, x) \neq \{0, \ldots, n-1\}$.

Exercice 1.6 ("Factorisation de Catalan") Soit D le langage de Dyck sur $A = \{a, b\}$, solution de l'équation $D = \varepsilon + aDbD$. Montrer que

$$A^* = (Db)^* D (aD)^*\,.$$

Exercice 1.7 Soient A et E deux alphabets. On dit qu'un mot $e \in E^+$ apparaît dans $w \in A^*$ s'il existe un morphisme non effaçant ϕ de E^* dans A^* tel que $\phi(e)$ est facteur de w. Un mot $e \in E^+$ est inévitable s'il apparaît dans tout mot assez long. Par exemple, le mot xx est inévitable sur 2 lettres, mais évitable sur trois lettres (voir exercice suivant).

1. Les mots de Zimin z_n sont définis comme suit : z_0 est une lettre, et $z_{n+1} = z_n a z_n$, où a est une lettre qui ne figure pas dans z_n. Par exemple, $z_1 = aba$, $z_2 = abacaba$. Montrer que les mots de Zimin sont inévitables.

2. Démontrer que si E a n lettres, tout mot e de E^* de longueur $\geq 2^n$ est évitable sur un alphabet de taille assez grande.

2 Langages particuliers

Exercice 2.1 Un langage L de A^* est dit *local* s'il existe deux parties $P, S \subset A$ et une partie $N \subset A^2$ telles que

$$L \setminus \{\varepsilon\} = (PA^* \cap A^*S) \setminus A^*NA^*\,.$$

La terminologie s'explique ainsi : pour tester si un mot appartient à L, il suffit de vérifier que sa première lettre est dans P, sa dernière lettre est dans S, et que ses facteurs de longueur 2 ne sont pas dans N. Toutes ces vérifications sont "locales".

1. Montrer, en donnant les ensembles P, S, N appropriés, que $(abc)^*$ est un langage local sur $A = \{a, b, c\}$.

2. Montrer qu'un langage local L sur A est reconnu par un automate fini ayant un ensemble Q d'états vérifiant $\mathrm{Card}(Q) = 1 + \mathrm{Card}(A)$ et tel que $\{q \cdot a \mid q \in Q\}$ est un singleton pour tout $a \in A$.

3. Montrer que si L est local, alors L^* est local.

4. Montrer que si L_1 et L_2 sont deux langages locaux sur A_1 et A_2 respectivement, avec $A_1 \cap A_2 = \emptyset$, alors $L_1 \cup L_2$ et $L_1 L_2$ sont des langages locaux.

5. Montrer qu'un langage L sur A est local si et seulement si, pour tout $a \in A$, l'ensemble $\{(ua)^{-1}L \mid u \in A^*\}$ contient au plus un élément.

6. On appelle automate *local* un automate fini déterministe (mais pas néssairement complet) $\mathcal{A} = (Q, i, T)$ sur A tel que pour toute lettre a, l'ensemble $\{q{\cdot}a \mid q \in Q\}$ contient au plus un élément. Montrer qu'un langage reconnaissable est local ssi il est reconnu par un automate local.

7. Montrer que pour tout langage reconnaissable $X \subset A^*$, il existe un alphabet B, un langage local K sur B et un morphisme littéral $f : B^* \to A^*$ tel que $f(K) = X$.

Exercice 2.2 1. Donner un exemple d'une transduction rationnelle $\tau : A^* \to A^*$ telle que l'ensemble $\{u \in A^* \mid u \in \tau(u)\}$ n'est pas un langage rationnel.

2. Donner un exemple d'une transduction rationnelle $\tau : A^* \to A^*$ telle que l'ensemble $\{u \in A^* \mid \tau(u) = A^*\}$ n'est pas un langage rationnel.

3. Démontrer que pour toute transduction rationnelle $\tau : A^* \to A^*$, l'ensemble $\{u \in A^* \mid \tau(u)\text{infini }\}$ est un langage rationnel.

3 Itération

Exercice 3.1 Soient $A = \{a, b, c, d\}$ et $K, L, M \subset A^*$ définis par

$$M = A^2 \setminus \{ab, ba, bc, cd, dc\}\,, \quad K = A^* M A^*\,, \quad L = \{('ab)^n [cd)^n \mid n \geq 1\}\,.$$

Montrer que le langage $K \cup L$ vérife le lemme de l'étoile, et qu'il n'est pas rationnel.

Exercice 3.2 Un langage $K \subset A^*$ est *local* s'il existe trois ensembles $U, V \subset A$ et $W \subset A^2$ tels que

$$K \setminus \varepsilon = (UA^* \cap A^*V) \setminus A^*WA^*\,.$$

Soit K un langage local. Montrer que, su $xyz, xy^2z \in K$ pour trois mots x, y, z, alors $xy^+z \subset K$.

Exercice 3.3 Soit $K \subset A^*$ un langage rationnel. Montrer qu'il existe deux entiers n, m tels que si $xy^kz \in K$ pour un entier $k \geq n$, alors $xy^k(y^m)^*z \subset K$.

Exercice 3.4 Soit $K \subset A^*$ un langage rationnel qui vérifie la propriété suivante : pour tout entier $n > 0$, il existe un mot dans K dont la longueur est un multiple de n. Montrer qu'il existe dex mots non vides x, y, z tels que $xy^*z \subset K$ et $|xz| = |y|$.

Exercice 3.5 Soit $L \subset A^*$ un langage algébrique vérifiant la même propriété, à savoir : pour tout entier $n > 0$, il existe un mot dans K dont la longueur est un multiple de n. Montrer qu'il existe dex mots non vides x, u, y, v, z tels que $xu^kyv^kz \in L$ pour tout $k \geq 0$ et $|xyz| = |uv|$.

Note Les deux derniers exercices sont de (Kobayashi, 1969, Lemme 1, page 104).

Exercice 3.6 Soit L un langage linéaire (engendré par une grammaire linéaire).

1. Montrer qu'il existe un entier N tel que tout mot w de L de longueur au moins N possède une factorization $w = xuyvz$ telle que $uv \neq \varepsilon$, $xu^kyv^kz \in L$ pour $k \geq 0$ et $|xuvz| \leq N$.

2. Montrer que, pour $L = \{a^nb^n \mid n > 0\}$, le langage L^2 n'est pas linéaire.

4 Décidabilité

Exercice 4.1 Soit L un langage rationnel sur A. On se propose de montrer qu'il est décidable s'il existe un mot $w \in A^+$, et des entiers i, j distincts tels que $w^i \in L$ et $w^j \in L$.

1. Montrer qu'il existe un mot $w \in A^+$, et des entiers i, j distincts tels que $w^i \in L$ et $w^j \in L$ si et seulement s'il existe un mot infini uv^ω tel que le chemin partant de l'état initial et d'étiquette uv^ω passe par au moins deux fois par un état terminal.

2. Montrer que le chemin partant de l'état initial et d'étiquette uv^ω est lui-même ultimement périodique, et qu'il est décidable s'il passe au moins deux fois par un état terminal. Conclure.

5 Opérations sur les langages

Exercice 5.1 Le but de cet exercice est de montrer que la clôture par conjugaison d'un langage algébrique est encore algébrique, en utilisant le théorème de Chomsky-Schützenberger.

Soit A un alphabet fini. Deux mots x, y sur A sont *conjugués* s'il existe dex mots u, v tels que $x = uv$ et $y = vu$. La relation de conjugaison est une relation d'équivalence. Pour tout mot x, on note $\gamma(x)$ l'ensemble des mots de A^* qui sont conjugués à x. Si $L \subset A^*$, on pose $\gamma(L) = \bigcup_{x \in L} \gamma[x)$.

1. Montrer que si K est un langage rationnel local, alors $\gamma(K)$ est un langage rationnel.

2. Montrer que pour tout morphisme alphabetique $h : A^* \to B^*$ et tout langage $L \subset A^*$, on a
$$\gamma(h(L)) = h(\gamma(L)) .$$
En déduire que si L est rationnel, alors $\gamma(L)$ est rationnel.

3. Soient L, M deux langages sur A. Montrer que $\gamma(L \cap M) \subset \gamma(L) \cap \gamma(M)$ et donner un exemple qui montre que l'inclusion peut être stricte.

4. Soit $\#$ un lettre qui n'est pas dans A. Montrer que $\gamma(\#L \cap \#M) = \gamma(\#L) \cap \gamma(\#M)$.

5. On rappelle que le *langage de Dyck* D^* sur $A \cup \bar{A}$, où $\bar{A}$ est une copie disjointe de A est engendré par la grammaire dont les règles sont
$$S \to \sum_{a \in A} aS\bar{a}S + \varepsilon .$$
Donner une description de $\gamma(\#D^*)$. Montre que c'est un langage algébrique.

6. Le théorème Chomsky-Schützenberger s'énonce comme suit. L est algébrique si est seulement si il est l'image, par un morphisme alphabétique, d'un langage de Dyck et d'un langage rationnel. Déduire de ce qui précède que si L est algébrique, alors $\gamma(L)$ l'est aussi.

Exercice 5.2 Soit K un langage. On partage chaque mot w de K de longueur paire en deux parties $w = w_0 w_1$ de même longueur, et on pose $K_0 = \{w_0 \mid w \in K\}$. Si K est rationnel, alors K_0 est rationnel. Si K est algébrique, alors K_0 n'est pas nécessairement algébrique.

Exercice 5.3 (Suite du précédent) On fixe un entier $N \geq 2$ et deux entiers d, f, avec $0 \leq d < f \leq N$. Soit K un langage. On partage chaque mot w de K de longueur multiple de N en N parts de longueurs égales : $w = w_0 w_1 \cdots w_{N-1}$, $|w_0| = |w_1| = \cdots = |w_{N-1}|$, et on conserve le facteur "central" $w_P = w_d w_{d+1} \cdots w_{f-1}$. On considère le langage $K_P = \{w_P \mid w \in K\}$. Si K est rationnel, alors K_P est rationnel.

Exercice 5.4 (Suite du précédent) On fixe un entier $N \geq 2$ et une partie P de $\{0, 1, \ldots, N-1\}$, soit $P = \{i_1 < i_2 < \cdots < i_p\}$ Soit K un langage. On partage chaque mot w de K de longueur multiple de N en N parts de longueurs égales : $w = w_0 w_1 \cdots w_{N-1}$, et on conserve le facteur $w_P = w_{i_1} w_{i_2} \cdots w_{i_p}$. Même si K est un langage rationnel, le langage $K_P = \{w_P \mid w \in K\}$ n'est pas nécessairement rationnel.

Exercice 5.5 Soit K un langage. On pose $s(K) = \{w \mid ww \in K\}$. Si K est un langage rationnel, alors $s(K)$ est rationnel. Si K est algébrique, $s(K)$ n'est pas nécessairement algébrique.

Exercice 5.6 Soit K un langage. On pose $m(K) = \{w \mid w\widetilde{w} \in K\}$. Si K est un langage rationnel, alors $m(K)$ est rationnel. Si K est algébrique, $m(K)$ n'est pas nécessairement algébrique.

Exercice 5.7 (suite de précédent) Soit K un langage. Pour un entier $n \geq 2$, on pose $s(K, n) = \{w \mid w^n \in K\}$. Si K est un langage rationnel, alors $s(K, n)$ est rationnel.

Exercice 5.8 (suite de précédent) Soit K un langage. On considère le langage $\mathrm{Rac}(K) = \{w \mid \exists n \geq 2, w^n \in K\}$ Si K est un langage rationnel, alors $\mathrm{Rac}(K)$ est rationnel.

Exercice 5.9 (suite de précédent) Soit K un langage. On considère le langage $\mathrm{Exp}(K) = \{w^n \mid n \geq 1, w \in K\}$. Ce langage n'est pas rationnel, si K est rationnel.

Exercice 5.10 Soient $A = \{a, b\}$, $B = \{a, b, c\}$. Soient $n \geq 1$ un entier fixé, et $u_1, \ldots, u_n$, $v_1, \ldots, v_n$ des mots de A^+. On définit les trois langages suivants :

$$\begin{array}{rcl} U & = & \{ba^{i_k} \cdots ba^{i_1} c u_{i_1} \cdots u_{i_k} \mid k \geq 1,\ 1 \leq i_1, \ldots, i_k \leq n\} \\ V & = & \{ba^{i_k} \cdots ba^{i_1} c v_{i_1} \cdots v_{i_k} \mid k \geq 1,\ 1 \leq i_1, \ldots, i_k \leq n\} \\ X & = & Uc\widetilde{V} \end{array}$$

(On rappelle que pour $w = a_1 \cdots a_n$, on a $\widetilde{w} = a_n \cdots a_1$, et que $\widetilde{K} = \{\widetilde{w} \mid w \in K\}$.) Par ailleurs, on pose

$$M = \{xcyc\widetilde{y}c\widetilde{x} \mid x, y \in A^*\}, \qquad K = X \cap M\,.$$

1. Prouver que U, V et M sont des langages algébriques.
2. Montrer que K est vide ou infini.
3. Montrer que K ne contient pas de langage algébrique infini.
4. Montrer que $B^* \setminus K$ est algébrique.

Exercice 5.11 Soit G la grammaire dont les règles sont

$$S \to 11 \mid 1001 \mid S0 \mid SS\,.$$

1. Montrer que tout mot engendré par G est la représentation en base 2 d'un entier multiple de 3.
2. Montrer que G n'engendre pas tous les multiples de 3 en base 2.
3. Donne une grammaire que engendre toutes les représentations des multiples de 3 en base 2.

Exercice 5.12 On cherche à construire un langage algébrique dont le complémentaire a $n^{1/3}$ mots de longeur n.

On considère le langage

$$K = \{a^n\#a^pb^q \mid p + q = n, n, p, q \in \mathbb{N}\}$$

Ce langage a exactement $n+1$ mots de longueur $2n+1$. Ainsi, pour $n = 2$, il contient les mots $aa\#aa$, $aa\#ab$ et $aa\#bb$. On laisse les b inchangés, et on substitue au i-ième a du préfixe a^n le mot $u_i aba^2b \cdots a^ibc$. Ainsi, $aaaa\#abbb$ devient $abcaba^2bcaba^2ba^3bcaba^2ba^3ba^4bc\#abbb$. Le langage L cherché est le langage obtenu en faisant les substitutions dans les mots de K.

Le mot u_i est de longueur $k_i = 1 + i + (1 + 2 + \cdots + i) = (1+i)(2+i)/2$. Le mot obtenu à partir de $a^n\#a^pb^q$ a longueur $k_1 + \cdots k_n + n + 1$, et ce nombre est clairement en n^3. Ainsi, L a environ n mots de longueur n^3.

Prouver que le complémentaire de L est algébrique.

6 Problèmes

6.1 Index rationnel

On note $\mathsf{Rat}_n(A^*)$ l'ensemble des langages rationnels sur A reconnus par un automate fini (nondéterministe) à n états. Par ailleurs, on pose, pour $L \subset A^*$, $L \neq \emptyset$,

$$\|L\| = \min\{|w| \,:\, w \in L\}\,, \quad \mu(L)\{w \in L \,:\, |w| = \|L\|\}\,,$$

et $\|\emptyset\| = 0$, $\mu(\emptyset) = \emptyset$. Ainsi, $\mu(L)$ est l'ensemble des mots de L de longueur minimale. Pour $M \subset A^*$, on définit

$$g_M(n) = \max\{\|M \cap K\| \,:\, K \in \mathsf{Rat}_n(A^*)\}\,.$$

Question 1. Soit $L \subset A^*$. Montrer que pour $n \leq m$, on a $g_L(n) \leq g_L(m)$.

Question 2. On considère, sur l'alphabet $A = \{a, b\}$, le langage $X = a^*b$, et pour $n \geq 1$, le langages $K_n = (a^{n-1}b)^*$.

a) Vérifier que $n = \|X \cap K_n\|$ et $2n = \|X^2 \cap K_n\|$. En déduire que $g_X(n) \geq n$ et $g_{X^2} \geq 2n$.

b) Prouver que $g_X(n) \leq n$.

Question 3. Vérifier que pour $L, M \subset A^*$, on a $g_{L \cup M}(n) = \max(g_L(n), g_M(n))$.

Question 4. Démontrer que pour $L, M \subset A^*$, on a $g_{LM}(n) = g_L(n) + g_M(n)$. En déduire que pour $X = X = a^*b$, on a $g_{X^2}(n) = 2n$.

Question 5. Démontrer que pour $X \subset A^*$, on a $g_{X^+}(n) \leq n \cdot g_X(n)$.

Question 6. Soit $K \subset A^*$ un langage rationnel, et $L \subset A$. Démontrer qu'il existe un entier p tel que $g_{L \cap K}(n) \leq g_L(pn)$ pour tout n.

Question 7. Soit $\phi : A^* \to B^*$ un morphisme alphabétique ($\phi(A) \subset B \cup \varepsilon$). Soient $L \subset A^*$ et $M \subset B^*$. Démontrer que $g_{\phi(L)}(n) \leq g_L(n)$ et $g_{\phi^{-1}(M)}(n) \leq n \cdot g_M(n)$.

Question 8. Soit $\tau : A^* \to B^*$ une transduction rationnelle, soit $L \subset A^*$ et posons $M = \tau(L)$. Déduire des questions précédentes un majoration de g_M au moyen de g_L.

On considère l'alphabet $A = \{a, b, c\}$ et la grammaire G dont les règles sont

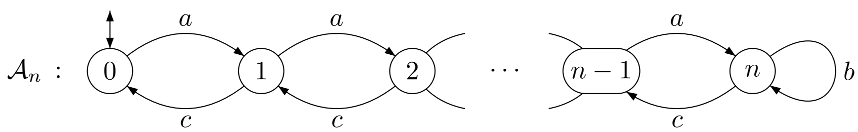
$$S \to a\,S\,S\,c\,, \quad S \to b\,.$$

Soit E le langage engenré par cette grammaire.

Question 9. Montrer que $w \in E$ si et seulement si $w = b$ ou il existe $u, v \in E$ tels que $w = auvc$.

Question 10. Montrer que E est préfixe. (On rappelle qu'un ensemble X est préfixe si aucun mot de X n'est préfixe d'un autre mot de X.)

On pose $K_0 = b^*$ et on définit, pour $n \geq 1$, K_n comme le langage reconnu par l'automate $\mathcal{A}_n$ représenté par la figure suivante.



Question 11. Montrer que $K_n = (aK_{n-1}c)^*$ pour $n \geq 1$.

On définit une suite de mots (w_n) par $w_0 = b$ et $w_n = aw_{n-1}w_{n-1}c$ pour $n \geq 1$.

Question 12. Montrer que, pour $n \geq 0$, on a $w_n \in K_n \cap E$.

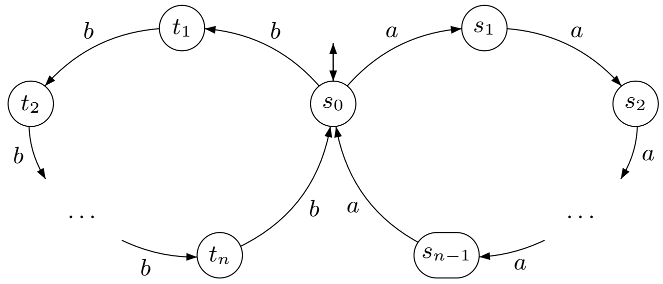
On définit, pour $i, j \leq n$ les langages $L_n^{i,j}$ par : $w \in L_n^{i,j}$ si et seulement si w est l'étiquette d'un chemin de i à j dans l'automat $\mathcal{A}_n$.

Question 13. Montrer que pour $i \neq j$, on a $L_n^{i,j} \cap E = \emptyset$, et que $w \in L_n^{i,i} \cap E$ implique $w \in K_{n-i} \cap E$.

Question 14. En déduire que $E \cap K_n = \{w_n\}$.

Question 15. Déduire de ce qui précède que $g_E(n) \geq 3\,2^n - 2$.

On considère le langage $P = \{a^n b^n \mid n \geq 1\}$. Pour $n \geq 1$, soit R_n le langage rationnel reconnu par l'automate suivant.



Question 16. Vérifier que $\mu(P \cap R_n) = \{a^{n(n+1)}b^{n(n+1)}\}$.

Question 17. En déduire que $g_P(n) > n^2/2$.

Question 18. En calculant une grammaire pour $P \cap R$, avec $R \in \mathsf{Rat}_n(A^*)$, montrer que $g_P(n) \leq 2n^2$.

Question 19. (difficile) Montrer que pour tout langage algébrique L, il existe $\alpha > 0$ tel que $g_L(n) \leq 2^{\alpha n^2}$.

Note La fonction g_L est l'*index rationnel* du langage L. L'index rationnel a été défini et étudié dans Boasson et al. (1981).

6.2 Langages métalinéaires

Une grammaire $G = (V, P, S)$ sur A est dite *métalinéaire* si elle vérifie les conditions suivantes :

– l'axiome S ne figure dans aucun membre droit de règle ;
– les seules règles non linéaires ont S pour membre gauche.

Un langage algébrique est métalinéaire s'il existe une grammaire métalinéaire qui l'engendre. On désigne par Mlin la famille des langages métalinéaires.

Question 1. Montrer que la famille des langages linéaires est strictement incluse dans Mlin.

Question 2. Montrer que Mlin est fermé par union et produit.

Question 3.
a) Montrer que Mlin est fermé pzr intersection rationnelle et par morphisme.
b) Montrer que Mlin est fermé par morphisme alphabétique inverse.
c) En déduire que Mlin est un cône rationnel.

On désigne par Lin_k la fermeture par union de $(\mathsf{Lin})^k$.

Question 4.
a) Montrer que Lin_k est un cône rationnel principal.
b) Montrer que

$$\mathsf{Mlin} = \bigcup_{k \geq 1} \mathsf{Lin}_k \ .$$

Question 5. Soit D le langage $\{a^n b^n \mid n \geq 1\}$.
a) Montrer que $D^k \in \mathsf{Lin}_k$ et $D^k \notin \mathsf{Lin}_{k-1}$.
b) Montrer que $D^* \notin \mathsf{Mlin}$. En déduire que Mlin n'est pas fermé par étoile.

6.3 Approximants du langage de Dyck

Soit A un alphabet fixe. Pour deux parties X, Y de A^*, on pose

$$v(X,Y) = \min\{|w| \mid w \in (X \setminus Y) \cup (Y \setminus X)\}\,.$$

Par convention, $v(X,X) = +\infty$. On défini la distance $d(X,Y)$ par

$$d(X,Y) = 2^{-v(X,Y)}$$

Question 1. Vérifier que d est une distance ultramétrique, c'est-à-dire que :
(1) $d(X,Y) = 0$ si et seulement si $X = Y$;
(2) $d(X,Y) \leq \max(d(X,Z), d(Z,Y))$ pour tous $X, Y, Z \subset A^*$.

Étant donné une suite $(X_n)_{n\geq 0}$ de parties de A^*, on écrira $\lim_{n\to\infty} X_n = X$ lorsque $\lim_{n\to\infty} d(X_n, X) = 0$.

Question 2. Soit $L \subset A^*$. Pour tout entier $p \geq 1$, soit R_p l'union des langages rationnels contenus dans L et dont l'automate minimal a au plus p états.
a) Montrer que R_p est un langage rationnel.
b) Montrer que $\lim_{p\to\infty} R_p = L$.
c) Pour tout langage rationnel $R \subset L$, il existe un entier p tel que $R \subset R_p$.

Soit maintenant $A = \{a, b\}$ et soit D le langage de Dyck sur A, défini comme suit : $w \in D$ si et seulement si $|w|_a = |w|_b$ et $|w'|_a \geq |w'|_b$ pour tout préfixe w' de w.

On appelle *hauteur* d'un mot w le nombre $h(w) = \max\{|w'|_a - |w'|_b : w' \text{ préfixe de } w\}$. Pour $L \subset A^*$, on pose $h(L) = \max\{h(w) \mid w \in L\}$.

Question 3. Soit R un langage rationnel contenu dans D, et soit k le nombre d'états de l'automate minimal complet reconnaissant R. Montrer que si $k \geq 2$, alors $h(R) \leq k - 2$.

Pour $p \geq 2$, on pose $Q_p = \{w \in D \mid h(w) \leq p - 2\}$.

Question 4. Construire un automate reconnaissant Q_p/

Question 5. Montrer que $Q_2 = \{e\}$ et $Q_{p+1} = aQ_pbQ_p \cup \{e\}$ pour $p \geq 2$.

Question 6. a) Montrer que, pour tout $X \subset A^*$, l'application f_X de l'ensemble des parties de A^* dans lui-même qui à Y associe $f_X(Y) = XY$ est continue.
b) Montrer que $\lim_{p\to\infty} Q_p = D$, et en déduire que le langage D vérifie l'équation $D = aDbD \cup \{e\}$.

6.4 Centre d'un langage

Soit A un alphabet fini. On écrit $x \leq y$ lorsque x est préfixe de y, et on note $P(X)$ l'ensemble des préfixes des mots d'un ensemble X.

Soit $L \subset A^*$. On appelle *centre* de L, et on note $c(L)$, l'ensemble des mots infiniment prolongeables de L, c'est-à-dire :

$$c(L) = \{w \in A^* \mid \forall n \in \mathbb{N}, \exists x \in L,\, |x| \geq n,\, w \leq x\}$$

En d'autre termes, $w \in c(L)$ si et seulement si $wA^* \cap L$ est infini.

Question 1. a) Calculer $c(a^*)$, $c(a^*b)$, $c(\{a^nb^n \mid n \geq 0\})$.
b) Montrer que $c(L_1 \cup L_2) = c(L_1) \cup c(L_2)$ et que $c(L) \subset P(L)$.
c) Montrer que $P(c(L)) = c(L)$.

d) Montrer que $c(L)$ est infini si et seulement L est infini.

Question 2. a) Montrer que pour tout $w \in c(L)$, il exist $a \in A$ tel que $wa \in C(L)$.

b) Montrer que $c(c(L)) = c(L)$.

Question 3. Montrer que si $h : A^* \to B^*$ est un morphisme littéral (c'est-à-dire vérifiant $h(A) \subset B$), alors $h(c(L)) = c(h(L))$.

Soit R un langage local de la forme $R = DA^* \cap A^*F \setminus A^*TA^*$, ave $D, F \subset A$ et $T \subset A^2$.

Question 4. On pose $B_0 = F$ et $B_{i+1} = B_i \cup \{a \in A \mid \exists b \in B_i \,:\, ab \notin T\}$. Soient enfin $B = \bigcup_{i\geq 0} B_i$, $D' = D \cap B$, $T' = T \cap B^2$.

a) Montrer que B est effectivement calculable.

b) Montrer que $P(R) = D'B^* \setminus A^*TA^* = D'B^* \setminus B^*T'B^*$.

Question 5. Soit $Z_0 = \{b \in B \mid \exists x \in B^* \,:\, bxb \in B^* \setminus B^*T'B^*\}$. Montrer que l'on peut calculer Z_0 effectivement.

Question 6. Montrer que tout mot $w \in B^* \setminus B^*T'B^*$ de la forme $w = axb$, avec $a \in D'$, $x \in B^*$ et $b \in Z_0$, on a $w \in C(R)$.

On pose $Z_{i+1} = Z_i \cup \{a \in B \mid \exists b \in B_i \,:\, ab \notin T\}$, et soient $Z = \bigcup_{i\geq 0} Z_i$.

Question 7. Montrer que Z est effectivement calculable.

Question 8. Déduire des questions précédentes que le centre d'un langage rationnel est encore un langage rationnel.

Question 9. Soit L un langage algébrique engendré par une grammaire G. Donner une construction qui permet d'obtenir, à partir de G, une grammaire engendrant $P(L)$. En déduire que le centre d'un langage algébrique est un langage algébrique.

6.5 Langages très simples

Une grammaire $G = (V, S, P)$ sur A est *très simple* si elle vérifie les deux conditions suivantes :

(i) Pour toute règle $X \to \alpha$, on a $\alpha \in AV^*$;

(ii) Pour tout lettre $a \in A$, il existe exactement une règle dont le membre droit commence par a.

Question 1. Quelles sont les grammaires très simples parmi les grammaires suivantes :

– $S \to aSS \mid b$;

– $S \to aSbS \mid aSb \mid abS \mid ab$;

– $S \to aSTSU \mid d$; $T \to b$; $U \to c$.

Question 2. Montrer qu'une grammaire très simple est inambiguë.

On considère désormais une grammaire très simple fixée $G = (V, S, P)$ sur A.

Question 3. Soient $\alpha \in V^*$, $u, v \in A^*$ tels que $\alpha \xrightarrow{*} uv$. Montrer qu'il un mot $\beta \in V^*$ unique tel que

$$\alpha \xrightarrow{*} u\beta \quad \text{et} \quad \beta \xrightarrow{*} v\,.$$

Question 4. Soient $\alpha, \beta, \beta' \in V^*$ et $u \in A^*$. Montrer l'implication

$$\alpha \xrightarrow{*} u\beta,\ \alpha \xrightarrow{*} u\beta' \implies \beta = \beta'\,.$$

Question 5. En déduire que pour tout α, le langage $L_G(\alpha)$ est préfixe (c'est-à-dire vérifie $u, uv \in L_G(\alpha) \implies v = \varepsilon$).

Question 6. Soient $\alpha, \beta \in V^*$ et $u \in A^*$. Montrer que

$$\alpha \xrightarrow{*} u,\ \beta \xrightarrow{*} u \implies \alpha = \beta\,.$$

On suppose désormais que $L_G(X)$ n'est pas vide pour toute variable X, et qu'il existe une dérivation $S \xrightarrow{*} uXv$ pour des mots u, v (toute variable est "accessible"). On dit que G a un *facteur itérant* s'il existe $\alpha \in V^+$, $u \in A^+$ tels que $\alpha \xrightarrow{+} u\alpha$.

Question 7. Montrer que si G est sans facteur itérant, alors $L_G(\alpha)$ est suffixe pour tout $\alpha \in V^*$.

Question 8. Montrer réciproquement que si $L_G(\alpha)$ est suffixe pour tout $\alpha \in V^*$, alors G est sans facteur itérant.

Question 9. Montrer que G est sans facteur itérant si et seulement si $L_G(S)$ ne contient pas de langage rationnel infini.

7 Solution

2.2 On considère la transduction τ qui à $a^n b^p$ associe $a^+ b^n$. On a $a^q b^n \in \tau(a^q b^n)$ si et seulement si $q = n$.

On considère ensuite la relation rationnelle $S = \{(u, v) \mid u \neq v\} \cup T$, où T est la relation associée à τ. On a $\sigma(u) = A^*$ si et seulement si $u = a^n b^n$ (le tout modulo $a^* b^*$).

Pour la dernière question, on ne conserve, dans le transducteur associé, que les chemins dont l'étiquette de sortie contient le mot vide sur un circuit.

3.4 Soit Q le nombre d'états de l'automate, soit $w \in K$ avec $|w| \equiv 0 \bmod Q!$, et soit $w = xyz$, $q = |y| \leq Q$, avec $xy^*z \subset K$, et soit k tel $|w| = kQ!$. On a

$$|xy^j y^n z| = kQ! + (j + n - 1)q\,,$$

donc $|xy^j z| = |y^n|$ si et seulement si $kQ! + jq - q = nq$, donc si et seulement si $kQ! = (n - j + 1)q$. On pose $T = kQ!/q$. C'est un entier. Soient n, j des entiers positifs tels que $T = n - j + 1$. Posons enfin

$$X = xy^j\ ,\ Z = z\ ,\ Y = y^n\ .$$

Alors $|XZ| = |xy^j z| = |y^n| = |Y|$, et évidemment $XY^*Z \subset K$.

5.2 On note λ le morphisme littéral qui envoie toutes les lettres sur une même lettre. Soient X, Y deux parties de A^*. Alors $H(X, Y) = \lambda^{-1}(\lambda(X) \cap \lambda(Y))$ est l'ensemble des mots w de A^* tels qu'il existe des mots $x \in X$ et $y \in Y$ avec $|w| = |x| = |y|$. Si X, Y sont rationnel, et plus généralement si $\lambda(X)$ et $\lambda(Y)$ sont rationnels, alors $H(X, Y)$ est rationnel. En particulier, $X \cap H(X, Y)$ est l'ensemble des mots $x \in X$ tels qu'il existe $y \in Y$ avec $|x| = |y|$.

Soit maintenant $\mathcal{A} = (Q, I, T)$ un automate fini reconnaissant K. On pose ${}_pK_q = \{w \in A^* \mid p \xrightarrow{w} q\}$. Pour tout $i \in I$, $t \in T$, on pose

$$ {}_iP_t = \bigcup_{q \in Q} {}_iK_q \cap H({}_iK_q, {}_qK_t) $$

On a $K_0 = \bigcup_{i \in I, t \in T} {}_iP_t$. D'où le résultat.

Pour la deuxième partie, considérons le langage

$$ L = \{a^n b^n c^{p-1} \bar{c} d^{3p} \mid n, p > 0\} $$

et l'ensemble $L_0 \cap a^* b^* c^* \bar{c}$. Comme $2n + p = 3p$ si un mot est dans cet ensemble, on a $L_0 \cap a^* b^* c^* \bar{c} = \{a^n b^n c^{n-1} \bar{c}\}$.

5.3 On pose

$$ H(X_0, X_1, \ldots, X_{N-1}) = \lambda^{-1}(\lambda(X_0) \cap \lambda(X_0) \cap \cdots \cap \lambda(X_{N-1})) $$

et, pour une suite $q = (q_0, q_1, \ldots, q_{N-1}, q_N)$ d'états, et un entier n avec $0 \le n < N$

$$ R(n, q) = ({}_{q_n}K_{q_{n+1}} \cap H({}_{q_0}K_{q_1}, {}_{q_1}K_{q_2}, \ldots {}_{q_{N-1}}qK_{q_N})) $$

Un mot x est dans $R(n, q)$ si et seulement s'il existe des mots $x_0, \ldots, x_{n-1}$, et $x_{n+1}, \ldots, x_{N-1}$, tous de la même longueur que x, de sorte que $x_j \in {}_{q_j}K_{q_{j+1}}$ et $x \in {}_{q_n}K_{q_{n+1}}$. Soit

$$ {}_iP_t = \bigcup_q (R(d, q) \cap R(d+1, q) \cap \cdots \cap R(f-1, q)) $$

où l'union est sur toutes les suites $q = (q_0, q_1, \ldots q_{N-1}, q_N)$ telles que $q_0 = i$ et $q_N = t$.

Alors $K_P = \bigcup_{i \in I, t \in T} {}_iP_t$. D'où le résultat.

5.4 On considère le langage $K = a^+ \bar{a} b^+ \bar{c} c^+$, et $N = 3$, $P = \{0, 2\}$. K_P est donc obtenu en supprimant, dans chaque mot de longueur multiple de 3 de K, le tiers du milieux et en concaténant la partie gauche avec la partie droite. Soit $L = K_P \cap a^+ \bar{a} \bar{c} c^+$. Comme les lettres barrées sont consécutives, et que $\bar{a}$ termine les a, et $\bar{c}$ précède les c, ce sont exactement les b qui ont été supprimés. Soit n les nombre de b supprimés. Alors le mot d'origine est $a^{n-1} \bar{a} b^n \bar{c} c^{n-1}$ et $L = \{a^n \bar{a} \bar{c} c^n \mid n \ge 0\}$.

5.5 On considère un automate fini $\mathcal{A} = (Q, I, T)$ reconnaissant K. On pose ${}_pK_q = \{w \in A^* \mid p \xrightarrow{w} q\}$, pour $p, q \in Q$. On a

$$ s(K) = \bigcup_{i \in I, t \in T} \bigcup_{q \in Q} ({}_iK_q \cap {}_qK_t)\,. $$

Par ailleurs, considérons le langage algébrique $L = \{a^n b^p c^q a^p b^r c^n\}$ Un mot dans $s(L) \cap a^+ b^+ c^+$ doit vérifier $n = p$, $q = n$, donc $s(L) \cap a^+ b^+ c^+ = \{a^n b^n c^n \mid n > 0\}$.

5.6 On a

$$ m(K) = \bigcup_{i \in I, t \in T} \bigcup_{q \in Q} ({}_iK_q \cap ({}_qK_t)^\sim)\,. $$

Par ailleurs, le langage $L = \{a^n b^p c^q c^p b^n a^r\}$ fait l'affaire.

5.7 On a

$$s(K) = \bigcup_{i \in I, t \in T} \bigcup_{q_1, \ldots, q_{n-1} \in Q} ({}_i K_{q_1} \cap {}_{q_1} K_{q_2} \cap \cdots \cap {}_{q_{n-1}} K_t) \,.$$

5.8 On a $\mathrm{Rac}(K) = \bigcup_{n \geq 2} s(K, n)$. Si K est rationnel, il existe un entier N tel que $\bigcup_{n \geq 2} s(K, n) = \bigcup_{2 \leq n \leq N} s(K, n)$.

5.9 Il suffit de considérer par exemple $K = a^+ b$.

Références

Luc Boasson, Bruno Courcelle, and Maurice Nivat. The rational index : A complexity measure for languages. *SIAM J. Comput.*, 10(2) :284–296, 1981.

Kojiro Kobayashi. Classification of formal languages by functional binary transductions. *Information and Control*, 15(1) :95–109, 1969.