Option Informatique en Spé MP et MP*
Déterminisation d'un automate fini reconnaissant un langage fini
Le corrigé

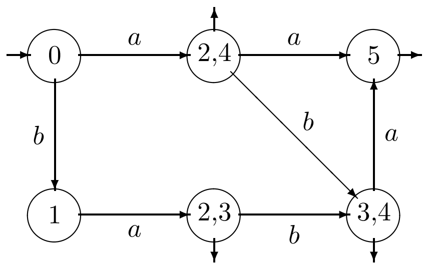
Question 1 • Nous remarquons qu'il n'y a pas de boucle dans le graphe de cet automate : il ne reconnaît donc qu'un nombre fini de mots. Nous énumérons ceux-ci en effectuant un parcours en largeur de ce graphe. Le langage reconnu est donc :

$$\{a, aa, ab, aba, ba, bab, baba\}$$

Question 2 • Voici la table obtenue en construisant l'automate des parties ; pour des raisons typographiques évidentes, la lecture se fait par colonnes et non par lignes :

	0	2,4	1	5	3,4	2,3
a	2,4	5	2,3	$\emptyset$	5	$\emptyset$
b	1	3,4	$\emptyset$	$\emptyset$	$\emptyset$	3,4

Et voici la représentation graphique de l'automate :



Le lecteur averti aura certainement noté que cet automate est minimal.

Question 3 • Procédons par induction structurelle. Le langage vide est reconnu par un automate ayant un seul état, initial mais non final ; en rendant final cet unique état, nous obtenons un automate qui reconnaît le langage $\{\varepsilon\}$; notons que dans les deux cas le graphe de l'automate est un arbre binaire réduit à sa racine. Considérons maintenant un langage L non réduit au mot vide : notons $L_a = \{u \in X^* \mid au \in L\}$ et de même $L_b = \{u \in X^* \mid bu \in L\}$. Construisons alors l'automate dont le graphe est un arbre binaire, dont le sous-arbre gauche (resp. droit) est l'automate qui reconnaît L_a (resp. L_b). L'état initial est la racine, qui est aussi un état final ssi $\varepsilon \in L$. La hauteur de cet arbre est la longueur maximale d'un mot de L.

Question 4 • Raisonnons par l'absurde : supposons qu'il existe un calcul de $\mathcal{A}$, d'étiquette $v \neq \varepsilon$, commençant et finissant dans l'état s. Comme cet état est accessible et coaccessible, il existe des mots u et w tels que $s \in \delta^*(I, u)$ et $\delta^*(S, w) \cap F \neq \emptyset$. Alors $\delta^*(I, uv^k w) \cap F \neq \emptyset$ quel que soit $k \in \mathbb{N}$, ce qui contredit la finitude du langage reconnu par $\mathcal{A}$. Nous avons utilisé la technique qui sert également à établir le lemme de l'étoile.

Question 5 • La longueur d'un mot de L est au plus $n - 1$. En effet, tout calcul de $\mathcal{A}$ ayant pour étiquette un mot u de longueur au moins n comporte un cycle, ce qui est interdit d'après la question précédente. Notons que ce maximum peut être atteint, par exemple par l'automate reconnaissant le mot a^{n-1} et lui seul.

Question 6 • Raisonnons par l'absurde, en supposant $\mathbf{S} = \bigcup_{1 \leqslant j \leqslant s} S_j$ contenu dans $\mathbf{T} = \bigcup_{1 \leqslant k \leqslant t} T_k$. Soit p_0 un élément de $\mathbf{S}$; p_0 appartient à l'un au moins des membres de la famille $\mathcal{T}$, que nous noterons T_1 ; il existe alors un indice j_1 et un mot $u_1 \in X^+$ tels que $\Delta(S_{j_1}, u_1) = T_1$; donc il existe un état $p_1 \in S_{j_1}$ tel que $p_0 \in \delta(p_1, u_1)$.

La construction se poursuit par récurrence : une fois obtenus des états $p_0, \dots, p_t$ de $\mathcal{A}$ appartenant chacun à un membre au moins de la famille $\mathcal{S}$, et des mots $u_1, \dots, u_t$ appartenant tous à X^+ et tels que $p_k \in \delta(p_{k-1}, u_k)$ pour tout $k \in [\![1, t]\!]$, nous pouvons affirmer que p_t appartient à au moins un membre T_{t+1} de la famille $\mathcal{T}$; du coup, il existe un élément $S_{j_{t+1}}$ de $\mathcal{S}$ et un mot u_{t+1} non vide tels que $\Delta(S_{j_{t+1}}, u_{t+1}) = T_{t+1}$; il se trouve donc un état $p_{t+1} \in S_{j_{t+1}}$ tel que $p_t \in \delta(p_{t+1}, u_{t+1})$.

La suite d'états $(p_t)_{t \geqslant 0}$ prenant ses valeurs dans l'ensemble fini Q, il existe nécessairement des indices k et ℓ tels que $0 \leqslant k < \ell$ et $p_k = p_\ell$. Alors $p_k \in \delta^*(p_k, u_{k+1}u_{k+2} \dots u_\ell)$; comme $|u_{k+1}u_{k+2} \dots u_\ell| \neq 0$, nous mettons en évidence un cycle dans le graphe de $\mathcal{A}$, ce qui termine la preuve.

Question 7 • Il suffit d'appliquer le résultat précédent en prenant pour $\mathcal{S}$ la famille Q'_j des états de niveau j, et pour $\mathcal{T}$ la famille $\displaystyle\bigcup_{k>j} Q'_k$ des états de niveau strictement supérieur à j : ces deux parties de Q' sont disjointes par définition. D'autre part, une récurrence immédiate montre que tout état de niveau $k > j$ peut être atteint depuis au moins un état de niveau j, par lecture d'un mot de longueur $k - j > 0$.

Question 8 • Notons Q_j l'ensemble des états de $\mathcal{A}$ qui appartiennent à au moins un état de $\mathcal{A}'$ de niveau supérieur ou égal à j. Par définition, Q_j est une partie de Q qui contient Q_{j+1}, et nous venons de montrer que cette inclusion est stricte. Comme $|Q_0| \leqslant |Q| = n$, nous en déduisons par récurrence $|Q_j| \leqslant n - j$. Un état de $\mathcal{A}'$ de niveau supérieur ou égal à j apparaît ainsi comme une partie de Q_j ; le nombre de ces états est donc majoré par le cardinal de $\mathcal{P}(Q_k)$, d'où $\displaystyle\Big|\bigcup_{k\geqslant j} Q'_k\Big| \leqslant 2^{n-j}$.

Question 9 • D'après la question précédente, nous avons $|Q'_j| \leqslant 2^{n-j}$ puisque Q'_j est contenu dans $\displaystyle\bigcup_{k\geqslant j} Q'_j$.

• Par ailleurs, il y a au plus 2^j états de niveau j, pour tout $j \in \mathbb{N}$. En effet, soit $\mathcal{Q} \in Q'_j$; comme l'automate $\mathcal{A}'$ est émondé, il existe au moins un mot $u \in X^j$ tel que $\Delta^*(I, u) = \mathcal{Q}$ (ceci a été vu à la question 7). Choisissons alors, parmi les mots u de longueur j tels que $\Delta^*(I, u) = \mathcal{Q}$ celui qui est le plus petit pour l'ordre lexicographique, et notons-le $\Phi(\mathcal{Q})$. Comme $\mathcal{A}'$ est déterministe, la fonction Φ est injective ; donc le cardinal de Q'_j est majoré par celui de X^j, soit 2^j puisque $|X| = 2$.

Question 10 • Considérons d'abord le cas $n = 2p$: notons que $\min(2^j, 2^{n-j})$ est maximal pour $j = n - j$, soit $j = p$, et vaut alors 2^p. Nous en déduisons :

$$|Q'| = \sum_{0\leqslant j<p} |Q'_j| + \Big|\bigcup_{j\geqslant p} Q'_j\Big| \leqslant \sum_{0\leqslant j<p} 2^j + 2^{n-p} = 2^p - 1 + 2^p = 2^{p+1} - 1$$

• Examinons maintenant le cas $n = 2p + 1$; cette fois, $\min(2^j, 2^{n-j})$ est maximal pour $j \in \{p, p+1\}$ et vaut alors 2^{p+1}. Nous en déduisons :

$$|Q'| = \sum_{0\leqslant j\leqslant p} |Q'_j| + \Big|\bigcup_{j>p} Q'_j\Big| \leqslant \sum_{0\leqslant j\leqslant p} 2^i + 2^{n-(p+1)} = 2^{p+1} - 1 + 2^p = 3 \cdot 2^p - 1$$

Question 11 • La preuve la plus simple consiste à remarquer que $\equiv_L$ est la relation d'équivalence canoniquement associée à l'application qui, à un mot u, associe $u^{-1}L = \{w \in X^* \mid uw \in L\}$.

Question 12 • Soit $w \in X^*$. Si $uw \in L$, alors $\delta^*(i, uw) \in F$. Du coup :

$$\delta^*(i, vw) = \delta^*(\delta^*(i, v), w) = \delta^*(\delta^*(i, u), w) = \delta^*(i, uw)$$

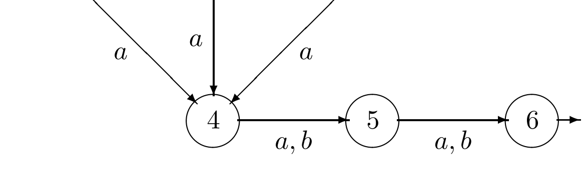
Ainsi $\delta^*(i, vw) \in F$, et donc $vw \in L$. Par raison de symétrie, $vw \in L \Rightarrow uw \in L$; comme ceci vaut quel que soit $w \in X^*$, nous en déduisons que $u \equiv_L v$.

Question 13 • Considérons la fonction φ qui, à une classe $\mathbf{u}$ modulo $\equiv_L$, associe $\delta^*(i, u)$ où u est le plus petit élément de $\mathbf{u}$, pour l'ordre lexicographique. Supposons $\varphi(\mathbf{u}) = \varphi(\mathbf{v})$; alors $\delta^*(i, u) = \delta^*(i, v)$, donc $u \equiv_L v$ soit $\mathbf{u} = \mathbf{v}$. Ceci montre que φ est injective, et par suite que $|Q|$ est au moins égal au nombre de classes modulo $\equiv_L$.

Question 14 • Considérons l'automate fini (non déterministe) ayant $Q = [\![1, n]\!]$ pour ensemble d'états, 1 pour état initial et n pour unique état final, et dont la table des transitions δ est la réunion des trois ensembles suivants :

$$\begin{aligned}
\delta_1 &= \big\{(i, a, i+1) \mid 1 \leqslant i < p\big\} \bigcup \big\{(i, b, i+1) \mid 1 \leqslant i < p\big\} \\
\delta_2 &= \big\{(i, a, p) \mid 1 \leqslant i < p\big\} \\
\delta_3 &= \big\{(i, a, i+1) \mid p+1 \leqslant i < n\big\} \bigcup \big\{(i, b, i+1) \mid p+1 \leqslant i < n\big\}
\end{aligned}$$

Voici par exemple l'automate qui correspond à $n = 6$:



Nous vérifions que cet automate reconnaît exactement les mots de la forme uav avec $|u| < p$ et $|v| = p - 1$, lesquels forment précisément le langage L : un tel mot est reconnu en suivant $|u|$ transitions appartenant à δ_1, qui mènent dans l'état $|u| + 1$; puis l'unique transition partant de cet état et appartenant à δ_2, laquelle mène dans l'état $p + 1$; et enfin $p - 1$ transitions appartenant à δ_2, menant de l'état $p + 1$ à l'état $n = 2p$. Réciproquement, considérons un mot w reconnu par notre automate : notant u le préfixe lu avant de passer dans l'état $p + 1$, nous aurons certainement $|u| < p$ et $w = uav$ avec $|v| = p - 1$.

Question 15 • Considérons deux mots x et y distincts, de longueurs respectives i et j, avec $0 \leqslant i \leqslant j \leqslant p$. Supposons $i < j$ dans un premier temps, et notons w le mot a^{n-i-1} ; comme $p > i$, nous pouvons écrire :

$$xw = xa^{n-i-1} = xa^{2p-i-1} = xa^{p-i-1}aa^{p-1}$$

Ceci prouve que xw appartient à L. Mais yw n'appartient pas à L, puisque $|yw| = j + n - i - 1 > n - 1$. Nous n'avons donc pas $x \equiv_L y$.

• Si $i = j$, nous pouvons supposer sans perte de généralité $x = uav$ et $y = ubv'$; notons cette fois $w = a^{p-|v|-1}$, ce qui est licite puisque $|v| < |uav| = |x| = i \leqslant p$; le mot $xw = uavw$ appartient à L, puisque $|vw| = p - 1$ et :

$$|xw| = i + p - |v| - 1 < i + p \leqslant 2p = n$$

En revanche, le mot $yw = ubv'w$ n'appartient pas à L puisque $|v'w| = p - 1$. Dans ce cas encore, nous n'avons pas $x \equiv_L y$.

• Résumons : nous avons prouvé que deux mots x et y distincts, de longueur p au plus, ne sont pas équivalents modulo $\equiv_L$.

Question 16 • $|x| \geqslant n$ implique $x \equiv_L b^p$; en effet, quel que soit $w \in X^*$, aucun des mots xw et $b^p w$ n'appartient à L.

• Il reste à examiner le cas $p < |x| < n$; remarquons que nous avons nécessairement $p \geqslant 2$. Pour alléger le discours, nous noterons $\rho(t, k)$ la k-ième lettre du mot t, en partant de la droite ; ainsi $t \in L$ ssi $|t| < n$ et $\rho(t, p) = a$. Notons $d = n - |x|$; alors $d < p$, et $p - d \geqslant 1$. Soient u, v et w les mots définis par $x = uvw$, $|u| = n - p$, $|v| = d$ et donc $|w| = p - d$. Alors x est équivalent modulo $\equiv_L$ à vb^{p-d}, qui est de longueur p. En effet, soit z un mot quelconque. Si $|z| \geqslant d$, alors $|xz| \geqslant n$, donc $xz \notin L$; quant au mot $vb^{p-d}z$, il n'appartient pas non plus à L, car ou bien $|z| \geqslant p$, auquel cas $|vb^{p-d}z| \geqslant 2p = n$, ou bien $d \leqslant |z| < p$, mais alors $\rho(vb^{p-d}z, p) = \rho(vb^{p-d}, p - |z|) = b$ car $p - |z| \geqslant p - d$. Si maintenant $|z| < d$, alors $|xz| < n$ et $|vb^{p-d}z| = p + |z| < 2p = n$; donc :

$$\begin{aligned} \rho(xz, p) &= \rho(uvwz, p) = \rho(uv, p - |z| - |w|) = \rho(uv, d - |z|) \\ \rho(vb^{p-d}z, p) &= \rho(v, p - |z| - p + d) = \rho(v, d - |z|) \end{aligned}$$

Ceci prouve que x et vb^{p-d} sont équivalents modulo $\equiv_L$.

Résumons : nous avons montré que tout mot de longueur strictement supérieure à p est équivalent modulo $\equiv_L$ à un mot de longueur inférieure ou égale à p.

Question 17 • À la question 15, nous avons établi qu'il existait au moins autant de classes modulo $\equiv_L$ que de mots de longueur inférieure ou égale à p ; nous avons ensuite montré à la question 16 que tout mot de longueur strictement supérieure à p appartenait à la classe d'un mot de longueur p au plus. Il existe donc exactement autant de classes modulo $\equiv_L$ que de mots de longueur p au plus, soit :

$$\sum_{0 \leqslant j \leqslant p} 2^j = 2^{p+1} - 1$$

Ceci prouve que tout automate reconnaissant L possède au moins $2^{p+1} - 1$ états.

Question 18 • Prenons comme langage l'ensemble L des mots de la forme uav avec $|uav| < n$ et $|av| = p$. Ce langage est reconnu par un automate fini (non déterministe) à n états construit comme celui de la question 14. Comptons les classes modulo $\equiv_L$. L'argumentation utilisée plus haut concernant les mots de longueur p au plus s'applique à nouveau, nous donnant déjà $2^{p+1} - 1$ classes. Soient x et y deux mots de longueur $p + 1$ et z un mot de longueur p au plus ; $xa^p \notin L$ et $za^p \in L$ donc x et z ne sont pas équivalents modulo $\equiv_L$; par ailleurs, $x \equiv_L y$ ssi x et y sont égaux ou ne diffèrent que par leur première lettre. Ceci nous fournit 2^p nouvelles classes modulo $\equiv_L$. Enfin, tout mot x de longueur au moins égale à n est dans la classe de b^p, déjà répertoriée. Il y a donc au total $2^{p+1} - 1 + 2^p = 3 \cdot 2^p - 1$ classes modulo $\equiv_L$; ainsi, tout automate fini déterministe complet reconnaissant L possède au moins $3 \cdot 2^p - 1$ états. Ceci prouve l'optimalité de la majoration établie, pour le cas où n est impair.

Références bibliographiques

▶ Une préversion de l'analyse faite par Kai SALOMAA et Sheng YU est disponible sur le Web, à l'URL :
`http://www.csd.uwo.ca/tech-reports/493/finite.ps`

▶ *NFA to DFA Transformation for Finite Languages*, Kai SALOMAA and Sheng YU. Automata Implementation, Proceedings of WIA'96. London, Ontario : August 1996. Lecture Notes in Computer Science, Vol. 1260. Springer-Verlag. 149-158.

▶ *NFA to DFA Transformation for Finite Languages over Arbitrary Alphabets*, Kai SALOMAA and Sheng YU. Journal of Automata, Languages and Combinatorics, 2 (1997) 3, 177-186.

▶ *A Maxmin Problem on Finite Automata*, Jean-Marc CHAMPARNAUD and Jean-Éric PIN. Discrete Applied Mathematics, 23 (1989), 91-96.

▶ *Minimisation of acyclic deterministic automata in linear time*, Daniel REVUZ. Theoretical Computer Science, Volume 92, Issue 1 (January 1992).

▶ Ce sujet a été traité par les étudiants au cours des vacances de Noël 1998.

▶ *Remarque* : plusieurs étudiants ont proposé le langage L formé des mots de la forme uav avec $|uav| < n$ et $|v| = p$. Cet exemple ne répond pas à la question. Rappelons que le *résiduel* d'un mot x vis-à-vis d'un langage L est $u^{-1}L = \{w \in X^* \mid uw \in L\}$; dire que deux mots u et v sont équivalents modulo $\equiv_L$ revient donc à dire que $u^{-1}L = v^{-1}L$. Considérons alors le cas $p = 1$; on a $L = \{aa, ab\}$ si bien que les classes modulo $\equiv_L$ sont au nombre de 4 seulement : la classe de ε, dont le résiduel est L ; la classe de a, dont le résiduel est $\{a, b\}$; la classe commune de a et b, dont le résiduel est $\{\varepsilon\}$; et la classe de tous les autres mots, dont le résiduel est vide.

▶ Un grand merci à Jean-Éric PIN et Kai SALOMAA.

FIN