TD Option informatique *n*°5 – Automates ambigus

Le sujet porte sur les automates finis, dont la définition est rappelée dans le préambule. Le sujet s'intéresse à la propriété d'ambiguïté des automates finis, propriété également définie dans le préambule. La première partie lie les notions de déterminisme et d'ambiguïté d'un automate fini en utilisant la notion de miroir d'un automate. La deuxième partie amène à définir un algorithme qui teste si un automate est ambigu et vous demande de déterminer sa complexité asymptotique. La troisième et la quatrième partie étudient la concision des automates non ambigus et ambigus.
Les parties 1 et 2 sont indépendantes.

Préambule

On note Σ un *alphabet* fini, c'est à dire un ensemble fini de symboles appelés *lettres*. Un *mot* sur Σ est une suite finie de lettres. On note Σ^n l'ensemble des mots de longueur n sur Σ et Σ^* l'ensemble de tous les mots sur Σ. Soient u, v deux mots sur Σ, on note $u \cdot v$ la *concaténation* de u et v. Un *langage* sur Σ est un sous-ensemble de Σ^*.

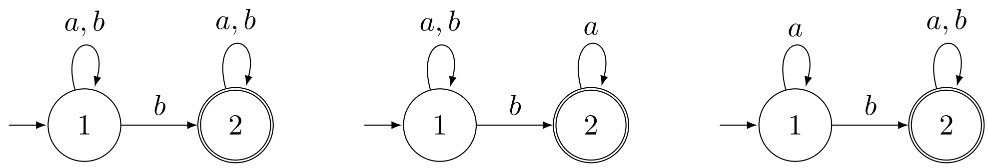
Un automate sur Σ est un tuple $\mathcal{A} = (Q, T, I, F)$ où :
- Q est un ensemble fini de symboles appelés *états* ;
- $T \subseteq Q \times \Sigma \times Q$ est appelé ensemble des *transitions* ;
- $I \subseteq$ est l'ensemble des états *initiaux* ;
- $F \subseteq Q$ est l'ensemble des états *finaux*.

La Figure 1 représente graphiquement trois automates. Les symboles dans Q sont encerclés, avec deux cercles pour les symboles dans F. Une transition $(q, a, q') \in T$ est représentée par une flèche étiquetée par $a \in \Sigma$, allant de l'état source q à l'état destination q'. Les états initiaux sont indiqués par une flèche sans état source.

Un calcul de $\mathcal{A}$ sur un mot $w = a_0 \dots a_{n-1} \in \Sigma^*$ est une suite finie d'états $q_0 \dots q_n \in Q^*$ telle que $q_0 \in I$ et pour tout $i < n$, $(q_i, a_i, q_{i+1}) \in T$. Un tel calcul est dit *acceptant* si $q_n \in F$. On dit alors que $\mathcal{A}$ accepte w. Le langage de $\mathcal{A}$, noté $L(\mathcal{A})$, est l'ensemble des mots acceptés par $\mathcal{A}$. On admettra (résultat peut-être vu en cours d'ici à ce que vous traitiez les questions qui l'utilisent) qu'un tel langage est toujours régulier.

Un automate $\mathcal{A}$ est dit *déterministe* si $|I| \leq 1$ et pour tout $(q, a) \in Q \times \Sigma, |\{q' | (q, a, q') \in T\}| \leq 1$.

Un automate $\mathcal{A}$ est dit *complet* si $|I| \leq 1$ et pour tout $(q, a) \in Q \times \Sigma, |\{q' | (q, a, q') \in T\}| \geq 1$.



$(a)\mathcal{A}_1$ $(b)\mathcal{A}_2$ $(c)\mathcal{A}_3$

FIGURE 1 - Trois automates reconnaissant le même langage

Soient $w \in \Sigma^*$ et $\mathcal{A}$ un automate. Le mot w est dit *ambigu* pour $\mathcal{A}$ s'il existe deux calculs acceptants ρ et ρ' de $\mathcal{A}$ sur w avec $\rho \neq \rho'$. On définit $d_{\mathcal{A}}(w)$, le *degré d'ambiguïté* de w dans $\mathcal{A}$, comme étant le nombre de calculs acceptants différents de $\mathcal{A}$ sur w. Ainsi, w est ambigu pour $\mathcal{A}$ si et seulement si $d_{\mathcal{A}}(w) > 1$. On note $\text{Amb}(\mathcal{A})$ l'ensemble des mots ambigus pour $\mathcal{A}$. L'automate $\mathcal{A}$ est dit *ambigu* si $\text{Amb}(\mathcal{A}) \neq 0$.

Ce sujet s'intéresse tout particulièrement aux automates *non ambigus.*

1 Déterminisme et ambiguïté

Question 1 Les trois automates de la Figure 1 acceptent le même langage. Donnez-en, sans justification, une description intuitive.

Question 2 Pour chacun des automates de la Figure 1, dites s'il est déterministe ou non déterministe. Justifiez vos affirmations.

Question 3 Pour l'automate $\mathcal{A}_1$ de la Figure 1, calculez $\text{Amb}(\mathcal{A}_1)$ et déduisez en si $\mathcal{A}_1$ est ambigu ou non. Faites de même pour les automates $\mathcal{A}_2$ et $\mathcal{A}_3$.

Soit $\mathcal{A}$ un automate. On note $\tilde{\mathcal{A}} = (\tilde{Q}, \tilde{T}, \tilde{I}, \tilde{F})$ l'*automate miroir* de $\mathcal{A}$, défini par :
- $\tilde{Q} = Q$;
- $\tilde{I} = F$;
- $\tilde{F} = I$;
- $\tilde{T} = \{(t, a, s) | (s, a, t) \in T\}$.

Un automate est dit *co-déterministe* si son automate miroir est déterministe.

Question 4 Soit $\mathcal{A}$ un automate, $w \in L(\mathcal{A})$ un mot accepté par $\mathcal{A}$ et $q_0 \dots q_n$ un calcul acceptant de $\mathcal{A}$ sur w. Montrez qu'il existe un mot $\tilde{w}$ tel que $q_n ... q_0$ soit un calcul acceptant de $\tilde{A}$ sur $\tilde{w}$.

Question 5 Montrez qu'un automate $\mathcal{A}$ est ambigu si et seulement si $\tilde{A}$ est ambigu.

Question 6 Montrez que si un automate $\mathcal{A}$ est déterministe, alors $\mathcal{A}$ n'est pas ambigu.

Question 7 Montrez que si un automate $\mathcal{A}$ est co-déterministe, alors $\mathcal{A}$ n'est pas ambigu.

Question 8 Pour chacune des questions suivantes, donnez un automate $\mathcal{A}$ ayant au plus 4 états et respectant les propriétés demandées. Justifiez vos réponses.

(*i*) $\mathcal{A}$ est non ambigu mais ni déterministe, ni co-déterministe ;
(*ii*) $L(\mathcal{A})$ est infini et $\text{Amb}(\mathcal{A}) = L(\mathcal{A})$;
(*iii*) $\text{Amb}(\mathcal{A})$ est infini, et $\text{Amb}(\mathcal{A}) \neq L(\mathcal{A})$.

2 Test d'ambiguïté

Le but de cette partie est d'obtenir un algorithme qui teste si un automate est ambigu et de déterminer sa complexité asymptotique.

Dans cette partie, $\mathcal{A}$ est un automate (Q, T, I, F) tel que $L(\mathcal{A}) \neq \emptyset$.

2.1 Une construction utile

En utilisant $\mathcal{A}$, on définit l'automate $\widehat{\mathcal{A}} = (\widehat{Q}, \widehat{T}, \widehat{I}, \widehat{F})$ comme suit :

— $\widehat{Q} = Q \times Q \times \{0, 1\}$;
— $\widehat{I} = \{(i, i, 0) | i \in I\} \cup \{(i, i', 1) | i \in I, i' \in I, i \neq i'\}$;
— $\widehat{F} = \{(f, f', 1) | f \in F, f' \in F\}$;
— $\widehat{T} = T_1 \cup T_2 \cup T_3$ avec :
  - $T_1 = \{((s, s, 0), a, (t, t, 0)) | (s, a, t) \in T\}$;
  - $T_2 = \{((s, s, 0), a, (t, t', 1)) | (s, a, t) \in T, (s, a, t') \in T, t \neq t'\}$;
  - $T_3 = \{((s, s', 1), a, (t, t', 1)) | (s, a, t) \in T, (s', a, t') \in T\}$.

Question 9 Soit $\mathcal{A}_1$ le premier automate de la figure Figure 1. Construisez l'automate $\widehat{\mathcal{A}_1}$. Il est inutile de faire figurer les états qui ne sont pas accessibles à partir d'un état initial. Donnez, sans justification, le langage $L(\widehat{\mathcal{A}_1})$.

Question 10 Soient $w \in \Sigma^*$ un mot et $\rho = (q_0, q'_0, b_0) \ldots (q_n, q'_n, b_n)$ un calcul de $\widehat{\mathcal{A}}$ sur w. On pose $\mu = q_0 \ldots q_n$ et $\mu' = q'_0 \ldots q'_n$.

(i) Montrez que μ et μ' sont des calculs de $\mathcal{A}$ sur w.
(ii) Montrez que $b_n = 0$ si et seulement si $\mu = \mu'$.
(iii) Montrez que, si ρ est acceptant, alors μ et μ' sont acceptants.
(iv) Montrez qu'il existe un calcul ρ tel que μ et μ' sont acceptants mais ρ ne l'est pas.

Question 11 Montrez que $L(\widehat{\mathcal{A}}) = \mathrm{Amb}(\mathcal{A})$.

2.2 L'algorithme

Question 12 Pour chaque ensemble $\widehat{Q}, \widehat{T}, \widehat{I}$ et $\widehat{F}$, donnez une borne asymptotique au nombre d'éléments qu'il contient en fonction des tailles de Q, T, I et F.

Question 13 Donnez un algorithme qui a comme entrée $\mathcal{A}$ et comme sortie $\widehat{\mathcal{A}}$ en précisant :

(i) quelle structure de données classique (matrice d'adjacence ou liste d'adjacence) est utilisée pour représenter $\mathcal{A}$ et $\widehat{\mathcal{A}}$,
(ii) une borne asymptotique de la complexité en temps d'exécution de cet algorithme en fonction de la somme des tailles des ensembles Q, T, I et F.

Question 14 Décrivez une méthode permettant de tester si $\mathcal{A}$ est ambigu en utilisant l'automate $\widehat{\mathcal{A}}$. Donnez une borne asymptotique de la complexité en temps de cette méthode en fonction de la somme des tailles des ensembles Q, T, I et F.

2.3 Généralisation

Soit k un entier strictement positif.

Question 15 Soit $w \in L(\mathcal{A})$ un mot de longueur k.

(i) Donnez, en fonction de k et $|Q|$, une borne supérieure sur le degré d'ambiguïté de w dans $\mathcal{A}$.

(ii) Donnez un automate $\mathcal{A}$ et un mot de longueur k pour lesquels la borne supérieure indiquée ci-dessus est atteinte.

Question 16 On pose $\text{Amb}_{\geq k}(\mathcal{A}) = \{w \in L(\mathcal{A}) | d_{\mathcal{A}}(w) \geq k\}$. Montrez que $\text{Amb}_{\geq k}(\mathcal{A})$ est régulier.

Question 17 On pose $\text{Amb}_{k}(\mathcal{A}) = \{w \in L(\mathcal{A}) | d_{\mathcal{A}}(w) = k\}$. Montrez que $\text{Amb}_{k}(\mathcal{A})$ est régulier.