Langages algébriques, ambiguïté, indécidabilité

Devoir à rendre au plus tard le 25 avril

Le but de ce devoir est de démontrer un théorème permettant, de façon générique, d'obtenir de nombreux résultats d'indécidabilité sur les langages algébriques. Il s'agit du résultat de la question 3.4, dont nous verrons quelques applications à la question 3.5. Les deux premières parties peuvent être traitées indépendamment, mais seront toutes deux nécessaires pour la partie finale.

1 Langages bornés

Définition 1 (Langage borné)**.** Un langage $L \subseteq \Sigma^*$ est *borné* si il existe $k \geq 0$ et des mots $w_1, \ldots, w_k \in \Sigma^*$ tel que :

$$L \subseteq w_1^* \cdots w_k^*$$

1. Soit $L = a^*b^*c^*$ et $L' = b^*c^*a^*$. Montrer que $L \cup L'$ est borné. Plus généralement, montrer que l'union maintient le caractère borné des langages.
2. Donner un exemple de langage non borné.
3. Soit $\mathcal{A} = \langle Q, \delta, q_0, F \rangle$ un automate fini déterministe, et $L = L(\mathcal{A})$. Montrer que, si L n'est pas borné, alors il existe deux symboles distincts $a, b \in \Sigma$, des mots $w_1, w_2, w_3, w_4 \in \Sigma^*$ et un état $q \in Q$ tels que :

$$\delta(q_0, w_1) = q \tag{1}$$
$$\delta(q, aw_2) = q \tag{2}$$
$$\delta(q, bw_3) = q \tag{3}$$
$$\delta(q, w_4) \in F \tag{4}$$

4. Soit L un langage régulier sur l'alphabet $\Sigma = \{a, b\}$. Montrer que les trois conditions suivantes sont équivalentes :
   (1) L n'est pas borné.
   (2) Il existe des mots $w_1, w_2, w_3, w_4 \in \Sigma^*$ tels que :
   $$w_1\{aw_2, bw_3\}^*w_4 \subseteq L$$
   (3) Il existe des mots $w_1, w_2, w_3, w_4 \in \Sigma^*$ tels que :
   $$w_1\{w_2a, w_3b\}^*w_4 \subseteq L$$

2 Ambiguïté

Définition 2 (Degré d'ambiguïté (grammaire))**.** Soit G une grammaire algébrique et k un entier strictement positif. On dit que G est de *degré d'ambiguïté* k si tous les mots de $L(G)$ ont au plus k arbres de dérivation distincts dans G.

On dit que G est de *degré d'ambiguïté infini* si G n'est pas de degré d'ambiguïté k, pour tout k.

Définition 3 (Degré d'ambiguïté (langage))**.** Soit L un langage algébrique et k un entier strictement positif. On dit que L est de *degré d'ambiguïté inhérente* k si L ne peut pas être engendré par une grammaire de degré d'ambiguïté inférieur à k, mais qu'il existe une grammaire de degré d'ambiguïté k qui engendre L.

Un langage de degré d'ambiguïté inhérente k, pour un certain k positif, est dit de *degré d'ambiguïté inhérente fini.*

On dit que L est de *degré d'ambiguïté inhérente infini* si toutes les grammaires qui engendrent L sont de degré d'ambiguïté infini.

Remarquons qu'une grammaire de degré d'ambiguïté 1 est en fait une grammaire non ambiguë. De même, un langage de degré d'ambiguïté inhérente 1 est un langage non ambigu. Ainsi, un langage de degré d'ambiguïté inhérente au moins 2 est inhéremment ambigu.

On pose $L_0 = \{a^m b^m c^n \mid m, n \geq 0\}$ et $L_1 = \{a^m b^n c^n \mid m, n \geq 0\}$.

1. Rappeler une démonstration de l'existence de langages inhéremment ambigus.
2. Peut-on trouver des grammaires de degré d'ambiguïté infini ne contenant qu'un nombre fini de mots ?
3. Soit L_0 et L_1 deux langages algébriques. Soit $n \geq 1$. Montrer que $(L_0 \cup L_1)^n$ est de degré d'ambiguïté inhérente 2^n.
4. Montrer que $(L_0 \cup L_1)^*$ est de degré d'ambiguïté inhérente infini.
5. Soit L un langage algébrique. Montrer que L est de degré d'ambiguïté inhérente fini si et seulement si il existe un automate à pile $\mathcal{A}$ acceptant par pile vide tel que $\mathcal{A}$ reconnait L, et $\mathcal{A}$ a un nombre borné d'exécutions acceptantes sur tout mot de L.
6. Montrer que les langages de degré d'ambiguïté inhérente fini sont clos par quotients droit et gauche par un mot. Autrement dit, si L est de degré d'ambiguïté inhérente fini et que w est un mot de Σ^*, alors $w^{-1}L$ et Lw^{-1} sont de degré d'ambiguïté inhérente fini.
7. Montrer que les langages de degré d'ambiguïté inhérente fini sont clos par morphisme inverse. Autrement dit, si L est de degré d'ambiguïté inhérente fini et que φ est un morphisme, alors $\varphi^{-1}(L)$ est de degré d'ambiguïté inhérente fini.

3 Conclusion

Dans toute cette partie, on fixe un alphabet $\Sigma = \{a, b\}$. Soit w_1, w_2, w_3, w_4 des mots de Σ^*. On définit les morphismes φ et ψ par :

$$\varphi : \begin{cases} \varphi(a) = aw_2 \\ \varphi(b) = bw_3 \end{cases}$$

$$\psi : \begin{cases} \psi(a) = aw_2aw_2 \\ \psi(b) = bw_3bw_3 \end{cases}$$

1. Soit L_1 et L_2 deux langages algébriques. Soit $u \notin L_2$. On pose :
$$w = \psi(u)aw_2bw_3$$
$$K_1 = \psi(\Sigma^*)aw_2bw_3\varphi(L_1)$$
$$K_2 = \psi(L_2)aw_2bw_3\varphi(\Sigma^*)$$
Montrer que $w^{-1}K_1 = \varphi(L_1)$ et $w^{-1}K_2 = \emptyset$.
2. Soit L_0 un langage algébrique tel que $\{aw_2, bw_3\}^* \subseteq L_0$. Montrer qu'il existe un langage régulier K_3 tel que, si on pose :
$$K = K_1 \cup K_2 \cup (L_0 \cap K_3)$$
alors, si $L_2 = \Sigma^*$, on a $K = L_0$, et si $L_2 \neq \Sigma^*$, alors il existe $w \in \{aw_2, bw_3\}^+$ tel que $L_1 = \varphi^{-1}(w^{-1}K)$. Montrer de plus que l'on peut effectivement construire une grammaire algébrique reconnaissant K à partir de grammaires algébriques reconnaissant L_0, L_1 et L_2.

Soit $\mathcal{S}$ un sous-ensemble des langages algébriques sur Σ. On pose :

$$\mathcal{L}(\mathcal{S}, w_1, w_2, w_3, w_4) = \{L \mid \exists L' \in \mathcal{S}, L = \varphi^{-1}(w^{-1}(w_1^{-1}L'w_4^{-1})), w \in \{aw_2, bw_3\}^+\}$$

3. Montrer que s'il existe un langage algébrique L_0 tel que :
(1) $L_0 \in \mathcal{S}$.
(2) $w_1\{aw_2, bw_3\}^*w_4 \subseteq L_0$.
(3) $\mathcal{L}(\mathcal{S}, w_1, w_2, w_3, w_4)$ est un sous-ensemble strict des langages algébriques sur Σ.
Alors l'appartenance d'un langage algébrique à $\mathcal{S}$ est indécidable.
4. On suppose que $\mathcal{S}$ est un sous-ensemble des langages algébriques de degré d'ambiguïté inhérente fini sur $\Sigma = \{a, b\}$ et qu'il existe dans $\mathcal{S}$ un langage L_0 tel que L_0 contient un langage régulier non borné. Montrer alors que l'appartenance d'un langage algébrique à $\mathcal{S}$ est indécidable.
5. Utiliser ce résultat pour démontrer que l'appartenance d'un langage algébrique aux familles suivantes est indécidable :
   - Les langages de degré d'ambiguïté inhérente fini.
   - Les langages de degré d'ambiguïté inhérente k, pour un certain k donné.
   - Les langages algébriques non ambigus.
   - Les langages réguliers.
   - Les langages algébriques déterministes.