Devoir maison n°8

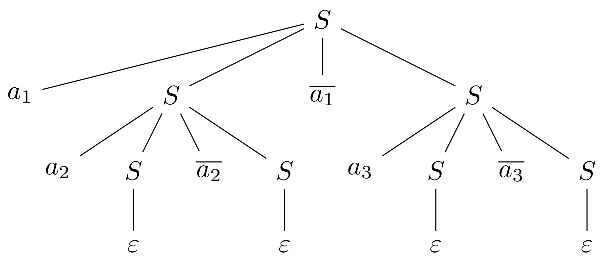
Corrigé

⋆⋆⋆

Théorème de Chomsky-Schützenberger

1 Langage de Dyck

Question 1



Question 2 On note $L = \{u \in \Sigma_1 \mid \forall v \in \mathrm{Pref}(u), |v|_{a_1} \geqslant |v|_{\overline{a_1}}\}$ et on montre par double inclusion que $L = D_1$.

– $L \subseteq D_1$: soit $u = a_1 a_2 \ldots a_n \in L$. Montrons par récurrence sur n que $u \in D_1$:
  - ⋆ si $n = 0$, alors $S \Rightarrow u$ est une dérivation de u, donc $u \in D_1$;
  - ⋆ supposons le résultat établi pour $|u| < n$, $n \in \mathbb{N}^*$ fixé. Comme $u \in L$, $a_1 = a$, sinon $v = a_1$ est un préfixe de u tel que $|v|_a < |v|_{\overline{a}}$. Notons i le plus petit indice > 1 tel que $|a_1 \ldots a_i|_a = |a_1 \ldots a_i|_{\overline{a}}$. Un tel i existe par définition de L, et de plus, $a_i = \overline{a}$, sinon $|a_1 \ldots a_{i-1}|_a < |a_1 \ldots a_{i-1}|_{\overline{a}}$. Posons $v = a_2 \ldots a_{i-1}$ et $w = a_{i+1} \ldots a_n$. Montrons que $v \in L$ et $w \in L$. En effet :
    - — $|v|_a = |a_1 \ldots a_{i-1}|_a - 1 = |a_1 \ldots a_{i-1}|_{\overline{a}} - 1 = |v|_{\overline{a}}$. De plus, par minimalité de i, pour $0 < j < i$, $|a_1 \ldots a_j|_a > |a_1 \ldots a_j|_{\overline{a}}$.
    - — $|w|_a = |u|_a - |v|_a - 1 = |u|_{\overline{a}} - |v|_{\overline{a}} - 1 = |w|_{\overline{a}}$. De plus, si w' est un préfixe de w, alors $av\overline{a}w'$ est un préfixe de u, d'où l'inégalité voulue.

    Dès lors, par hypothèse de récurrence, $S \Rightarrow^* v$ et $S \Rightarrow^* w$. On en déduit que $S \Rightarrow aS\overline{a}S \Rightarrow^* av\overline{a}w = u$, donc $u \in D_1$.

  On conclut par récurrence.

– $D_1 \subseteq L$: montrons par récurrence sur n que si $S \Rightarrow_g^n \alpha$, alors $|\alpha|_a = |\alpha|_{\overline{a}}$ et pour tout préfixe β de α, $|\beta|_a \geqslant |\beta|_{\overline{a}}$:
  - ⋆ pour $n = 0$, $\alpha = S$ et toutes ces quantités sont nulles ;
  - ⋆ supposons le résultat établi pour α, et soit $S \Rightarrow_g^n \alpha \Rightarrow_g \alpha'$ une dérivation de α'. Alors α' s'obtient en remplaçant le S le plus à gauche de α par $aS\overline{a}S$ ou par ε. Comme ces deux mots sont dans L, on en déduit que α' vérifie bien les propriétés voulues.

  On conclut par récurrence.

Question 3 Supposons par l'absurde que D_1 est rationnel et soit n sa longueur de pompage. On pose $u = a^n \overline{a}^n$. Par le lemme de pompage, il existe une décomposition $u = xyz$ telle que :

- $|xy| \leqslant n$;
- $|y| > 0$;
- pour $k \in \mathbb{N}, xy^k z \in D_1$.

Mais avec les deux premières hypothèses, on a $y = a^i$ pour un certain $i \in [\![1, n]\!]$. Dès lors, $xz = a^{n-i}\overline{a}^n \notin D_1$. On conclut par l'absurde que D_1 n'est pas rationnel.

Question 4 Ce résultat est faux. En effet, $u = a_1 a_2 \overline{a_1} \overline{a_2}$ est un mot du langage décrit, mais il n'est pas dans D_2. En effet, en raisonnant sur le nombre de symboles terminaux, une dérivation d'un tel mot u serait de taille 5, et dans une telle dérivation, 3 dérivations immédiates serait une règle $S \to \varepsilon$, une serait $S \to a_1 S \overline{a_1} S$ et une serait $S \to a_2 S \overline{a_2} S$. Aucun des arrangements possibles ne peut cependant donner le mot u.

Question 5 On utilise une pile initialement vide et on lit le mot lettre à lettre selon le principe suivant :

- si on lit un a_i, on l'empile dans la pile ;
- si on lit un $\overline{a_i}$, on dépile l'élément du haut de la pile. Si c'est a_i, on continue, sinon on renvoie faux.

Lorsqu'on arrive à la fin du mot, on vérifie que la pile est vide.

```
let dyck u =
    let rec verif pile u = match pile, u with
        | [], []            -> true
        | _, Ouv i :: v     -> verif (i :: pile) v
        | i :: q, Fer j :: -> i = j && verif q v
        | _                 -> false in
    verif [] u;;
```

2 Propriétés sur les langages algébriques

Question 6 On remarque que ε est l'unique mot u qui vérifie $u = u^2$. Dès lors, $\varphi(\varepsilon) = \varphi(\varepsilon\varepsilon) = \varphi(\varepsilon)\varphi(\varepsilon)$, donc $\varphi(\varepsilon) = \varepsilon$.

Question 7 Soit $G = (\Sigma, V, P, S)$ une grammaire qui engendre L. On pose $G' = (\Sigma', V', P', S)$ où :

- $V' = V \cup \{X_a \mid a \in \Sigma\}$;
- On définit $\psi : \Sigma \cup V \to V'$ un morphisme de mot par $\psi(X) = X$ si $X \in V$ et $\psi(a) = X_a$ si $a \in \Sigma$. Dès lors, P' contient les règles de production :
  * $X \to \psi(\alpha)$ si $X \to \alpha \in P$;
  * $X_a \to \varphi(a)$ pour tout $a \in \Sigma$.

Montrons que $L(G') = \varphi(L(G))$.

- Soit $u = a_1 \dots a_n \in L(G)$. On peut montrer par induction que si $S \Rightarrow_G^* u$, alors $S \Rightarrow_{G'}^* \psi(u)$. De plus, l'existence des règles $X_a \to \varphi(a)$ garantit que $\psi(u) \Rightarrow_{G'}^* \varphi(a_1)\varphi(a_2)\dots\varphi(a_n) = \varphi(u)$. On conclut que $\varphi(u) \in L(G')$.
- Soit $v = b_1 \dots b_n \in L(G')$ et A un arbre de dérivation de v dans G'. Comme les règles $X_a \to \varphi(a)$ sont les seules qui peuvent faire apparaître une lettre de Σ', on en déduit que chaque feuille $\neq \varepsilon$ a pour père une variable X_a. De plus, les variables X_a n'apparaissant pas du côté gauche d'une règle ailleurs que dans une règle $X_a \to \varphi(a)$, on en déduit que l'arbre A' obtenu en remplaçant un nœud $N(X_a, F(\varphi(a)))$ par $F(a)$ est un arbre de dérivation d'un mot $u \in L(G)$. Par construction, $v = \varphi(u)$.

Question 8 Le langage engendré est $\{a^n b^n \mid n \in \mathbb{N}^*\}$. En effet, on peut remplacer les règles $S \to AX$ et $X \to SB$ par $S \to ASB$. On reconnaît une grammaire déjà étudiée.

Question 9 Soit $G = (\Sigma, V, P, S)$ une grammaire hors-contexte en forme normale de Chomsky et $A = (Q, \Sigma, \delta, q_0, F)$ un AFD tels que $L \setminus \{\varepsilon\} = L(G)$ et $R = L(A)$.

On fait en sorte que (p, X, q) génère tous les mots $u \in \Sigma^*$ tels que $\delta^*(p, u) = q$ et $X \Rightarrow^* u$ (dans G). Pour cela, on pose P' contenant les règles :

- pour $q \in F$, $S' \to (q_0, S, q)$;
- pour $X \to a \in P$ et $q \in Q$, $(q, X, \delta(q, a)) \to a$;
- pour $X \to YZ \in P$ et $q_1, q_2, q_3 \in Q$, $(q_1, X, q_3) \to (q_1, Y, q_2)(q_2, Z, q_3)$.

Montrons que $L(G') = L(G) \cap L(A)$.

- soit $u \in L(G) \cap L(A)$. Alors $\delta^*(q_0, u) \in F$ et $S \Rightarrow^* u$. On en déduit que $(q_0, S, \delta^*(q_0, u)) \Rightarrow^* u$, donc $u \in L(G')$, car $S' \to (q_0, S, \delta^*(q_0, u))$ est une règle de production.
- soit $u \in L(G')$. Si $u = \varepsilon$, alors $q_0 \in F$ et $S \to \varepsilon \in P$, donc $u \in L \cap R$. Sinon, $S' \Rightarrow (q_0, S, q) \Rightarrow^* u$ est une dérivation de u dans G', avec $q \in F$. Cela implique que $S \Rightarrow^* u$ est une dérivation de u dans G et $\delta^*(q_0, u) = q$. Soit finalement $u \in L \cap R$.

Dès lors, si $\varepsilon \in L \cap R$, il suffit de rajouter une règle $S' \to \varepsilon$ pour obtenir la grammaire voulue.

Question 10 On pose $L = \{a^i b^i c^j \mid i, j \in \mathbb{N}\}$ et $L' = \{a^i b^j c^j \mid i, j \in \mathbb{N}\}$. Ces deux langages sont algébriques, par exemple L peut être engendré par la grammaire :

- $S \to XC$;
- $X \to aXb \mid \varepsilon$;
- $C \to cC \mid \varepsilon$.

Pourtant, $L \cap L' = \{a^i b^i c^i \mid i \in \mathbb{N}\}$ n'est pas un langage algébrique.

3 Théorème de Chomsky-Schützenberger

Question 11 On suppose que $L = \varphi(D_n \cap R)$ pour $n \in \mathbb{N}^*$ et $R \in \mathrm{Rat}(\Sigma)$. Par la question 9, $D_n \cap R$ est algébrique. Par la question 7, $L = \varphi(D_n \cap R)$ est algébrique.

Question 12 On obtient la grammaire :

- $S \to a_1 b_1 A \overline{b_1} c_1 X \overline{c_1 a_1} \mid a_2 b_2 A \overline{b_2} c_2 B \overline{c_2 a_2}$;
- $X \to a_3 b_3 S \overline{b_3} c_3 B \overline{c_3 a_3}$;
- $A \to a\overline{a}$;
- $B \to b\overline{b}$.

Question 13 Avec les notations de l'énoncé, on pose $n = 3k + |\Sigma|$ et $\Sigma_n = \Sigma'$. Montrons un résultat plus fort, c'est-à-dire que pour $X \in V$, si $X \Rightarrow^\ell u$ avec $u \in \Sigma'^*$, alors $u \in D_n$, par récurrence sur la taille des dérivations ℓ :

- si $\ell = 1$, alors la dérivation est de la forme $X \to a\overline{a}$, avec $a \in \Sigma$. Dès lors, $S_n \Rightarrow aS_n\overline{a}S_n \Rightarrow a\overline{a}S_n \Rightarrow a\overline{a}$ est une dérivation de u dans G_n (on a renommé le symbole de départ en S_n pour ne pas confondre avec S) ;
- si on suppose le résultat vrai pour toute dérivation de taille $< \ell$, avec $\ell > 1$ fixé, alors la première dérivation immédiate est de la forme : $X \Rightarrow a_i b_i Y \overline{b_i} c_i Z \overline{c_i a_i}$. De plus, il existe v et w tels que $u = a_i b_i v \overline{b_i} c_i w \overline{c_i a_i}$ et $Y \Rightarrow^* v$, $Z \Rightarrow^* w$. Par hypothèse de récurrence, v et w sont dans D_n. Dès lors, $S_n \Rightarrow a_i S_n \overline{a_i} S_n \Rightarrow^* a_i b_i S_n \overline{b_i} S_n \overline{a_i} c_i S_n \overline{c_i} S_n \overline{a_i} S_n \Rightarrow^* a_i b_i v \overline{b_i}\,\overline{a_i} c_i w \overline{c_i a_i} = u$.

On conclut par récurrence.

Question 14 On remarque que pour toute règle $X \to \alpha \in P'$, alors $X \to \varphi(\alpha) \in P$. Un raisonnement par récurrence et la définition des morphismes de mots permet donc de conclure que $\varphi(L(G')) \subseteq L(G)$. Réciproquement, pour toute règle $X \to \alpha \in P$, il existe β tel que $X \to \beta \in P'$ et $\alpha = \varphi(\beta)$. De la même manière, cela donne l'inclusion réciproque.

Question 15 $P(L(G'))$ et $N(L(G'))$ sont rationnels car finis. Comme la concaténation et le complémentaire préservent la rationalité, $R \in \mathrm{Rat}(\Sigma')$.

Question 16 Montrons que $L(G') = D_n \cap R$, où D_n est le langage défini à la question 13 et R le langage défini à la question précédente. On a déjà montré $L(G') \subseteq D_n$. De plus, $L(G') \subseteq R$, car tout mot de $L(G')$ commence par une première lettre d'un mot de $L(G')$ et ne contient aucun facteur de taille 2 de $N(L(G'))$. Cela montre la première inclusion.

Pour $X \in V$, si on note $L(G', X)$ le langage engendré par (Σ', V, P', X) et $R_X = P(L(G', X))\Sigma'^* \setminus \Sigma'^* N(L(G', X))\Sigma'^*$, alors on peut montrer par récurrence sur la taille des mots que $D_n \cap R_X \subseteq L(G', X)$. Dès lors, $D_n \cap R = D_n \cap R_S \subseteq L(G', S) = L(G')$.

Finalement, par la question 14, $L(G) = \varphi(L(G')) = \varphi(D_n \cap R)$.

Pour conclure complètement, on remarque que si L est un langage algébrique, alors il existe G une grammaire en forme normale de Chomsky telle que $L \setminus \{\varepsilon\} = L(G)$. Dès lors quitte à considérer $R \cup \{\varepsilon\}$ si $\varepsilon \in L$, on obtient le résultat voulu.

⋆⋆⋆