

TD 5 – Automates à pile et algébricité

1 Automates à pile (« pushdown automata »)

Exercice 1. Pour chacun des langages suivants, donner un automate à pile (AP) le reconnaissant.

1. $\{w \in \{a, b\}^* \mid |w|_a = |w|_b\}$.
2. $\{va^n \mid v \in \{a, b\}^* \wedge |v|_a = n\}$.
3. $\{w \in \{a, b\}^* \mid w = w^R\}$ (langage des palindromes sur $\{a, b\}$).
4. $\{w \in \{a, b\}^* \mid w \neq w^R\}$ (langage des non palindromes sur $\{a, b\}$).
5. Pour $k \in \mathbb{N}_{>0}$ fixé, le langage des mots sur $A_k = \{a_1, \dots, a_k, \bar{a}_1, \dots, \bar{a}_k\}$ engendré par la grammaire algébrique $\mathcal{G} = (\{S, T\}, A_k, R, S)$ où R contient les règles :

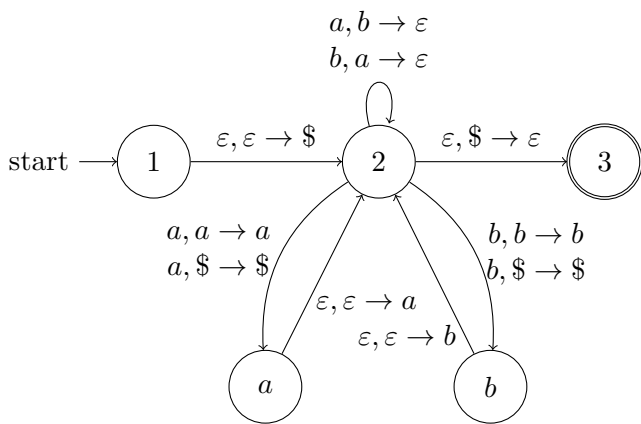
$$S \rightarrow TS \mid \varepsilon$$

$$T \rightarrow a_1 S \bar{a}_1 \mid \dots \mid a_k S \bar{a}_k .$$

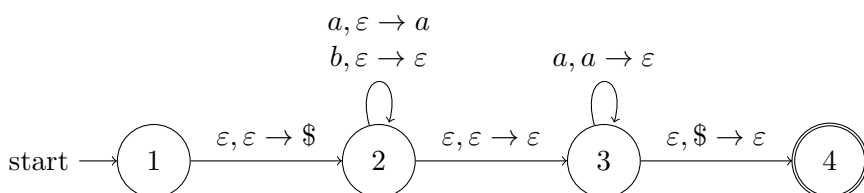
(C'est le langage de Dyck des mots bien parenthésés avec k types de parenthèses.)

Solution 1.

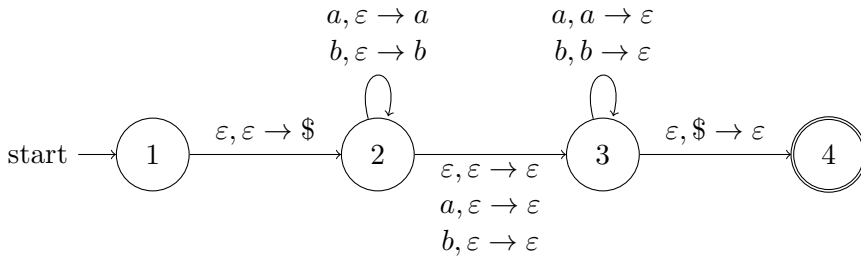
1. On construit l'AP suivant :



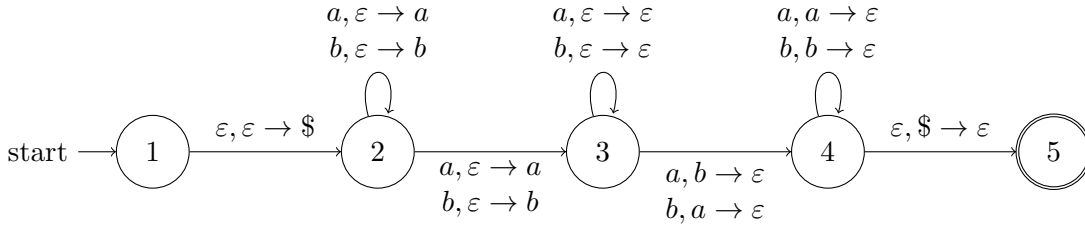
2. On construit l'AP suivant :



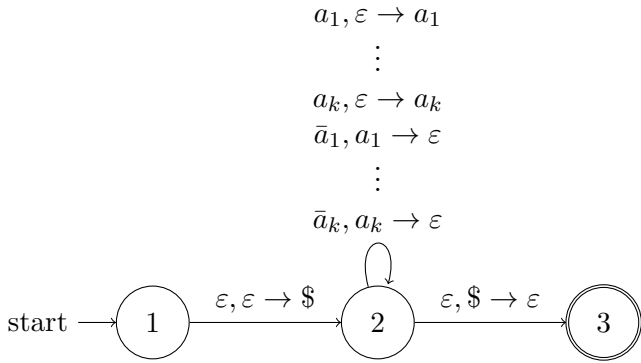
3. On construit l'AP suivant :



4. On construit l'AP suivant :



5. On construit l'AP suivant :



Exercice 2. Soit Σ un alphabet. Montrer que pour tous L, K langages sur Σ tels que L est rationnel et K est algébrique, $L \cap K$ est aussi un langage algébrique sur Σ .

Solution 2. Soit $\mathcal{A} = (Q_L, \Sigma, \delta_L, q_{L,0}, F_L)$ un AFD reconnaissant L et $\mathcal{P} = (Q_K, \Sigma, \Gamma, \delta_K, q_{K,0}, F_K)$ un AP reconnaissant K .

On construit l'AP $\mathcal{P}' = (Q_L \times Q_K, \Sigma, \Gamma, \delta', (q_{L,0}, q_{K,0}), F_L \times F_K)$ où $\delta' : (Q_L \times Q_K) \times (\Sigma \cup \{\varepsilon\}) \times (\Gamma \cup \{\varepsilon\}) \rightarrow \mathfrak{P}((Q_L \times Q_K) \times (\Gamma \cup \{\varepsilon\}))$ est telle que pour tous $q_L \in Q_L, q_K \in Q_K, a \in \Sigma$ et $b \in \Gamma \cup \{\varepsilon\}$,

$$\delta'((q_L, q_K), a, b) = \{((p, \delta_L(q_K, a)), c) \mid (p, c) \in \delta_K(q_K, a, b)\}$$

et

$$\delta'((q_L, q_K), \varepsilon, b) = \{((p, q_K), c) \mid (p, c) \in \delta_K(q_K, \varepsilon, b)\}.$$

On peut montrer que \mathcal{P}' reconnaît $L \cap K$.

2 Algébricité

Exercice 3. Montrer que le langage $\{a^n b^n c^n \mid n \in \mathbb{N}\}$ n'est pas algébrique.

Solution 3. Si ce langage était algébrique, d'après le lemme de pompage, pour p suffisamment grand, on aurait que $a^p b^p c^p$ dans le langage se décomposerait comme $a^p b^p c^p = uvxyz$ avec $vy \neq \varepsilon$, $|vxy| \leq p$ et $uv^i xy^i z$ dans le langage pour tout i . On aurait donc que vy contient au plus deux lettres différentes, ce qui impliquerait que le mot uxz ne contenant pas le même nombre de a , de b et de c est dans le langage.

Exercice 4. En déduire que la classe des langages algébriques n'est close ni par intersection, ni par complémentation.

Solution 4. On peut montrer que les langages $\{a^n b^n c^m \mid n, m \in \mathbb{N}\}$ et $\{a^m b^n c^n \mid n, m \in \mathbb{N}\}$ sont tous les deux algébriques. Or, leur intersection est égale à $\{a^n b^n c^n \mid n \in \mathbb{N}\}$, qui n'est pas un langage algébrique par ce qui a été montré dans l'exercice précédent.

Comme vu dans la feuille de TD précédente, le langage $\{a^i b^j c^k \mid i, j, k \in \mathbb{N}, i \neq j \vee j \neq k\}$ est algébrique. On montre que son complémentaire n'est pas algébrique, parce que s'il l'était, alors son intersection avec le langage rationnel $\mathcal{L}(a^* b^* c^*)$, qui est égale à $\{a^n b^n c^n \mid n \in \mathbb{N}\}$, serait elle aussi algébrique.

Exercice 5. Montrer que les langages suivants ne sont pas algébriques.

1. $\{a^n b^n c^m \mid n, m \in \mathbb{N}, n \leq m \leq 2n\}$.
2. $\{a^i b^j c^k \mid i, j, k \in \mathbb{N}, i < j < k\}$.
3. $\{a^{2^n} \mid n \in \mathbb{N}\}$.
4. $\{a^{P(n)} \mid n \in \mathbb{N}\}$ pour P un polynôme à coefficients dans \mathbb{N} de degré au moins 2.
5. $\{ww \mid w \in \{a, b\}^*\}$.
6. $\{w\#w \mid w \in \{a, b\}^*\}$.
7. $\{w\#w' \mid w, w' \in \{a, b\}^*, |w| = |w'| \wedge w \neq w'\}$.

Solution 5.

1. Si ce langage était algébrique, d'après le lemme de pompage, pour p suffisamment grand, on aurait que $a^p b^p c^{2p}$ dans le langage se décomposerait comme $a^p b^p c^{2p} = uvxyz$ avec $vy \neq \varepsilon$, $|vxy| \leq p$ et $uv^i xy^i z$ dans le langage pour tout i . On aurait donc que
 - soit vy contient uniquement des a ou des b , ce qui impliquerait que le mot uxz contenant soit un nombre de a et de b différent, soit un nombre $p' < p$ de a et b tel que $2p' < 2p$ est dans le langage ;
 - soit vy contient uniquement des b ou des c , ce qui impliquerait que le mot $uv^2 xy^2 z$ contenant un nombre de b strictement supérieur au nombre n de a ou un nombre de c strictement supérieur à $2n$ est dans le langage.
2. Si ce langage était algébrique, d'après le lemme de pompage, pour p suffisamment grand, on aurait que $a^p b^{p+1} c^{p+2}$ dans le langage se décomposerait comme $a^p b^{p+1} c^{p+2} = uvxyz$ avec $vy \neq \varepsilon$, $|vxy| \leq p$ et $uv^i xy^i z$ dans le langage pour tout i . On aurait donc que
 - soit vy contient au moins un a mais pas de c , ce qui impliquerait que le mot $uv^3 xy^3 z$ contenant strictement au moins autant de a que de c est dans le langage ;
 - soit vy contient au moins un b mais pas de a , ce qui impliquerait que le mot uxz contenant au plus autant de b que de a est dans le langage ;
 - soit vy contient uniquement des c , ce qui impliquerait que le mot uxz contenant au plus autant de c que de b est dans le langage.
3. Si ce langage était algébrique, d'après le lemme de pompage, pour p suffisamment grand, on aurait que a^{2^p} dans le langage se décomposerait comme $a^{2^p} = uvxyz$ avec $vy \neq \varepsilon$, $|vxy| \leq p$ et $uv^i xy^i z$ dans le langage pour tout i . On aurait donc $vy = a^d$ avec $0 < d \leq p$, ce qui impliquerait que $uv^2 xy^2 z = a^{2^p+d}$ est dans le langage, sachant que $2^{p+1} - 2^p = 2^p > p \geq d$.
4. Soit $k \in \mathbb{N}$ le degré de P . Si ce langage était algébrique, d'après le lemme de pompage, pour p suffisamment grand, on aurait que $a^{P(p)}$ dans le langage se décomposerait comme $a^{P(p)} = uvxyz$ avec $vy \neq \varepsilon$, $|vxy| \leq p$ et $uv^i xy^i z$ dans le langage pour tout i . On aurait donc $vy = a^d$

avec $0 < d \leq p$, ce qui impliquerait que $w^2xy^2z = a^{P(p)+d}$ est dans le langage, sachant que $P(p+1) - P(p) > p \geq d$.

5. Si ce langage était algébrique, d'après le lemme de pompage, pour p suffisamment grand, on aurait que $a^pb^pa^pb^p$ dans le langage se décomposerait comme $a^pb^pa^pb^p = uvxyz$ avec $vy \neq \varepsilon$, $|vxy| \leq p$ et uv^ixy^iz dans le langage pour tout i . On aurait donc que
 - soit vxy est un facteur du premier bloc a^pb^p , ce qui impliquerait que le mot uxz dans le langage est soit de longueur impaire, soit vérifie que la première moitié du mot se termine par un a alors que la seconde moitié du mot se termine par un b ;
 - soit vxy est un facteur du bloc central b^pa^p , ce qui impliquerait que le mot uxz dans le langage est soit de longueur impaire, soit vérifie que la première moitié du mot se termine par un a alors que la seconde moitié du mot se termine par un b , soit vérifie que la seconde moitié du mot commence par un b alors que la première moitié du mot commence par un a , soit est égal à $a^pb^da^db^p$ où $d < p$;
 - soit vxy est un facteur du second bloc a^pb^p , ce qui impliquerait que le mot uxz dans le langage est soit de longueur impaire, soit vérifie que la seconde moitié du mot commence par un b alors que la première moitié du mot commence par un a .
6. Similaire au précédent.
7. Si ce langage était algébrique, d'après le lemme de pompage, pour p suffisamment grand, on aurait que $a^{p^l+pb^p}\#a^pb^{p^l+p}$ dans le langage se décomposerait comme $a^{p^l+pb^p}\#a^pb^{p^l+p} = uvxyz$ avec $vy \neq \varepsilon$, $|vxy| \leq p$ et uv^ixy^iz dans le langage pour tout i . On aurait donc que
 - soit vxy est un facteur du premier bloc $a^{p^l+pb^p}$, ce qui impliquerait que le mot uxz dans le langage contient un bloc à gauche de $\#$ moins long que le bloc à droite de $\#$;
 - soit vxy est un facteur du bloc central $b^p\#a^p$ contenant $\#$, ce qui impliquerait soit que le mot uxz dans le langage ne contient pas $\#$, soit que $v = b^i$, $x = b^j\#a^k$ et $y = a^l$ et que donc, ou $i \neq l$ et donc le mot uxz dans le langage contient un bloc à gauche de $\#$ de longueur différente de celle du bloc à droite de $\#$, ou $i = l$ et donc le mot $a^{p^l+pb^p}\#a^pb^{p^l+p} = uv^{p^l/i+1}xy^{p^l/i+1}z$ est dans le langage ;
 - soit vxy est un facteur du second bloc $a^pb^{p^l+p}$, ce qui impliquerait que le mot uxz dans le langage contient un bloc à droite de $\#$ moins long que le bloc à gauche de $\#$.

3 Limitations du lemme de pompage et lemme d'Ogden

Soit $K = \{a^ib^jc^kd^l \mid i, j, k, l \in \mathbb{N}, i = 0 \vee j = k = l\}$ un langage sur $\{a, b, c, d\}$.

Exercice 6. Montrer que K vérifie le lemme de pompage.

Solution 6. Soit $p = 1$. Soit $s \in K$ tel que $|s| \geq 1$.

Si $s = b^jc^kd^l$ pour $j, k, l \in \mathbb{N}$, alors en décomposant s comme $s = uvxyz$ avec $u = x = y = \varepsilon$, v étant égal à la première lettre de s et z étant égal au reste, on a nécessairement que $vy \neq \varepsilon$, $|vxy| \leq 1 = p$ et pour tout $m \in \mathbb{N}$, $uv^mxy^mz \in K$, puisqu'il est toujours nécessairement de la forme $b^{j'}c^{k'}d^{l'}$ pour $j', k', l' \in \mathbb{N}$.

Sinon, si $s = a^ib^jc^jd^j$ pour $i \in \mathbb{N}_{>0}$, $j \in \mathbb{N}$, alors en décomposant s comme $s = uvxyz$ avec $u = x = y = \varepsilon$, $v = a$ et $z = a^{i-1}b^jc^jd^j$, on a nécessairement que $vy \neq \varepsilon$, $|vxy| \leq 1 = p$ et pour tout $m \in \mathbb{N}$, $uv^mxy^mz = a^ma^{i-1}b^jc^jd^j \in K$.

On a donc que K vérifie le lemme de pompage pour $p = 1$.

Lemme 3.1 (Ogden). Soit L un langage algébrique sur un alphabet Σ . Il existe $n \in \mathbb{N}_{>0}$ tel que pour tout $s \in L$ avec $|s| \geq n$ et tout sous-ensemble $D \subseteq \llbracket 1, |s| \rrbracket$ d'au moins n positions dans s dites distinguées, il est possible de décomposer s comme $s = uvxyz$ avec $u, v, x, y, z \in \Sigma^*$ de telle façon que :

1. au moins l'une des lettres dans le facteur v ou y corresponde à une position distinguée ;
2. le facteur vxy contienne au plus n lettres correspondant à des positions distinguées ;
3. pour tout $i \in \mathbb{N}$, $w^ixy^iz \in L$.

Solution 6. Soit $\mathcal{G} = (V, \Sigma, R, S)$ une grammaire algébrique engendrant L . On note $b = \max\{|v| \mid (X \rightarrow v) \in R\}$.

Si $b \leq 1$, le résultat s'obtient trivialement.

Autrement, $b \geq 2$. On peut montrer que pour tout arbre de dérivation de \mathcal{G} , dont la racine est étiquetée par un symbole quelconque de V et dont on a distingué un sous-ensemble D de ses feuilles, on a que pour tout $h \in \mathbb{N}$, si $b^h < |D|$, alors il existe un sous-arbre de dérivation T' de T avec D' l'ensemble des feuilles distinguées obtenu en restreignant D à T' vérifiant $b^h < |D'| \leq b^{h+1}$. (On montre ceci par récurrence sur la profondeur de T , en observant qu'il existe nécessairement un des sous-arbres fils de T ayant au moins $\frac{|D|}{b}$ feuilles distinguées.)

Étant donné un mot $s \in L$ avec $|s| \geq b^{|V|+1}$ et un sous-ensemble d'au moins $b^{|V|+1}$ positions distinguées dans s , on a un arbre de dérivation T pour s dans \mathcal{G} dont la racine est étiquetée par S et ayant un sous-ensemble D de feuilles distinguées de cardinal au moins $b^{|V|+1}$. D'après le résultat précédent, on peut donc construire une suite de sous-arbres de dérivation $T_0, \dots, T_{|V|}$ de T ayant pour ensemble de feuilles distinguées $D_0, \dots, D_{|V|}$ respectivement tels que T_i soit sous-arbre de T_{i+1} et $b^i < |D_i| \leq b^{i+1}$ pour tout $i \in \llbracket 0, |V| \rrbracket$. Il doit donc exister $i, j \in \llbracket 0, |V| \rrbracket, i < j$ tels que T_i et T_j aient la même racine. Si on note $x \in \Sigma^*$ le mot dérivé par T_i et vxy avec $v, y \in \Sigma^*$ le mot dérivé par T_j (qui est tel, puisque x est nécessairement un facteur du mot dérivé par T_j), on a que s se décompose comme $s = uvxyz$ avec $u, z \in \Sigma^*$. Il est direct de vérifier que, par construction, cette décomposition vérifie les conditions du lemme d'Ogden.

Exercice 7. L'utiliser pour montrer que K n'est pas algébrique.

Solution 7. Si K était algébrique, d'après le lemme d'Ogden, pour p suffisamment grand, on aurait que $ab^p c^p d^p \in K$, où toutes les positions en-dehors de la première seraient distinguées, se décomposerait comme $ab^p c^p d^p = uvxyz$ avec vy contenant au moins une lettre de $\{b, c, d\}$, vxy contenant au plus p lettres de $\{b, c, d\}$ et $uv^i xy^i z \in K$ pour tout $i \in \mathbb{N}$. On aurait donc que vy contient au plus deux lettres différentes de $\{b, c, d\}$, ce qui impliquerait que le mot $uv^2 xy^2 z$, commençant par un a mais ne contenant pas le même nombre de b , de c et de d est dans K .