Automates finis (III)

Montaigne 2023-2024
– mpi23@arrtes.net –

Stabilité des langages

Langages réguliers (par définition)

Soit un alphabet Σ. L'ensemble $\mathrm{Reg}(\Sigma)$ des **langages réguliers** est défini par induction.

- Base : $\emptyset \in \mathrm{Reg}(\Sigma)$; $\{\varepsilon\} \in \mathrm{Reg}(\Sigma)$; $\forall a \in \Sigma, \{a\} \in \mathrm{Reg}(\Sigma)$
- Induction - Si $L \in \mathrm{Reg}(\Sigma)$ et $L' \in \mathrm{Reg}(\Sigma)$ alors :

$$L \cup L' \in \mathrm{Reg}(\Sigma) \qquad L \cdot L' \in \mathrm{Reg}(\Sigma) \qquad L^* \in \mathrm{Reg}(\Sigma)$$

Les opérations d'**union**, de **concaténation** et d'**étoilé** préservent donc le caractère régulier des langages. Et dans le cas présent, cette propriété est liée à la définition même des langages réguliers. On dit que l'ensemble des langages réguliers est **stable** ou **clos** par ces opérations.

Stabilité et clôture

Soit E un ensemble et si $f : E \to F$ une application unaire, on dit que E est stable (ou clos) par f si pour tout $x \in E$, $f(x) \in E$.

Soit E un ensemble et si $g : E \to F$ une application binaire, on dit que E est stable (ou clos) par g si pour tout $(x, y) \in E^2$, $g(x, y) \in E$.

Une illustration simple de cette notion est celle de l'ensemble des nombres pairs et de l'opération d'addition. Si deux entiers sont pairs, leur addition l'est encore. L'ensemble des nombres pairs est stable pour l'addition sur les entiers.

Stabilité et clôture

La **clôture** est une **propriété algébrique forte**.

Définir un langage par composition.
Sur $\Sigma = \{a, b\}$, le langage des mots qui commencent soit par un a, soit par un b peut s'écrire $(a|b)\Sigma^*$. C'est un langage régulier, réunion des deux langages réguliers $a\Sigma^*$ et $b\Sigma^*$.

Montrer qu'un langage n'est pas régulier.
Nous verrons que les langages réguliers sont clos par intersection. Sachant que $L = \{a^n b^n, n \in \mathbb{N}\}$ n'est pas régulier, que $L_1 = a^* b^*$ est régulier et que $L = L_1 \cap L_2$ où $L_2 = \{w \in \Sigma^*, |w|_a = |w|_b\}$, on déduit que L_2 n'est pas un langage régulier.

Définir un langage comme une classe.
L'ensemble des langages réguliers est le plus petit ensemble clos par union, concaténation et étoile et qui contient le langage vide et les langages réduits à un seul mot d'une lettre.

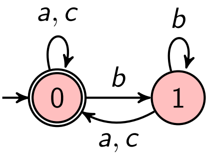
Complémentation

Théorème 1 (stabilité par complémentaire)

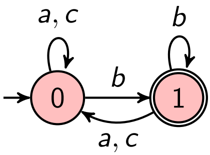
Soit L un langage régulier sur un alphabet Σ. Le langage $\overline{L}$ est régulier.

Complémentation - démonstration

Soit L un langage régulier sur un alphabet Σ. Il existe un automate **déterministe et complet** $\mathcal{A} = (Q, \Sigma, q_0, F, \delta)$ tel que $L = \mathcal{L}(\mathcal{A})$. On construit l'automate $\mathcal{A}' = (Q, \Sigma, q_0, Q \setminus F, \delta)$. Il est clair que $\mathcal{L}(\mathcal{A}') = \overline{\mathcal{L}(\mathcal{A})}$. En effet, tout chemin acceptant dans $\mathcal{A}$ ne l'est pas dans $\mathcal{A}'$ et vice-versa, car on a inversé les états acceptants et non acceptants.



$(a|b|c)^* b$



$(a|b|c)^* (a|c)$

Complémentation - démonstration

La construction de $\mathcal{A}'$ exige que $\mathcal{A}$ soit **complet** et **déterministe**.

- Si $\mathcal{A}$ n'est pas **complet** alors tous les mots rejetés par absence de chemin dans $\mathcal{A}$ ne seront pas reconnus dans $\mathcal{A}'$.
- Si $\mathcal{A}$ n'est pas **déterministe**, un mot n'est reconnu que s'il existe un chemin acceptant. Dès lors, pour ne pas accepter un mot, $\mathcal{A}'$ devrait refuser tous les chemins, ce qu'un automate non déterministe ne peut pas faire. Ici, la contrainte de déterminisme de $\mathcal{A}$ est liée au fait qu'il existe une quantification existentielle implicite lorsque l'on considère des automates non déterministes.

Prendre le complémentaire d'un langage régulier (donné par exemple par une expression régulière ou un automate non déterministe) est donc exponentiel en temps et en espace.

Intersection

Théorème 2 (stabilité par intersection)

Soit L_1 et L_2 deux langages réguliers sur un alphabet Σ. Le langage $L_1 \cap L_2$ est régulier.

Intersection - démonstration 1

Soit L_1 et L_2 deux langages réguliers sur le même alphabet Σ. Il existe $\mathcal{A}_1 = (Q_1, \Sigma, q_{10}, F_1, \delta_1)$ et $\mathcal{A}_2 = (Q_2, \Sigma, q_{20}, F_2, \delta_2)$ tels que $L_1 = \mathcal{L}(\mathcal{A}_1)$ et $L_2 = \mathcal{L}(\mathcal{A}_2)$. On peut supposer ces automates non déterministes (cas le plus général).

On construit l'automate :

$$\mathcal{A} = (Q_1 \times Q_2, \Sigma, (q_{10}, q_{20}), F_1 \times F_2, \delta)$$

avec :

$$\delta : \begin{cases} Q_1 \times Q_2 \times \Sigma \to \mathcal{P}(Q_1 \times Q_2) \\ ((q_1, q_2), a) \mapsto \{(p_1, p_2) \mid p_1 \in \delta_1(q_1, a), p_2 \in \delta_2(q_2, a)\} \end{cases}$$

On montre que $\mathcal{L}(\mathcal{A}) = L_1 \cap L_2$.

Intersection - démonstration

Montrons que $\mathcal{L}(\mathcal{A}) \subseteq L_1 \cap L_2$.

Si $v \in \mathcal{L}(\mathcal{A})$ alors il existe $(q_{10}, q_{20}) \xrightarrow{v}^*_{\mathcal{A}} (q_1, q_2) \in F_1 \times F_2$.

Par construction $q_{10} \xrightarrow{v}^*_{\mathcal{A}_1} q_1$ et $q_{20} \xrightarrow{v}^*_{\mathcal{A}_2} q_2$ sont acceptants.
Par conséquent $v \in \mathcal{L}(\mathcal{A}_1) = L_1$ et$v \in \mathcal{L}(\mathcal{A}_2) = L_2$.
Donc $v \in L_1 \cap L_2$.

Montrons que $L_1 \cap L_2 \subseteq \mathcal{L}(\mathcal{A})$.
Si $v \in L_1 \cap L_2$, alors $v \in L_1 = \mathcal{L}(\mathcal{A}_1)$ et $v \in L_2 = \mathcal{L}(\mathcal{A}_2)$.

Donc $q_{10} \xrightarrow{v}^*_{\mathcal{A}_1} q_1$, avec $q_1 \in F_1$ et $q_{20} \xrightarrow{v}^*_{\mathcal{A}_2} q_2$, avec $q_2 \in F_2$.

Par construction, $(q_{10}, q_{20}) \xrightarrow{v}^*_{\mathcal{A}} (q_1, q_2)$ et ainsi $v \in \mathcal{L}(\mathcal{A})$.

Intersection - démonstration 2

Une autre démonstration est possible à l'aide de la **lois de De Morgan** suivante appliquée aux ensembles.

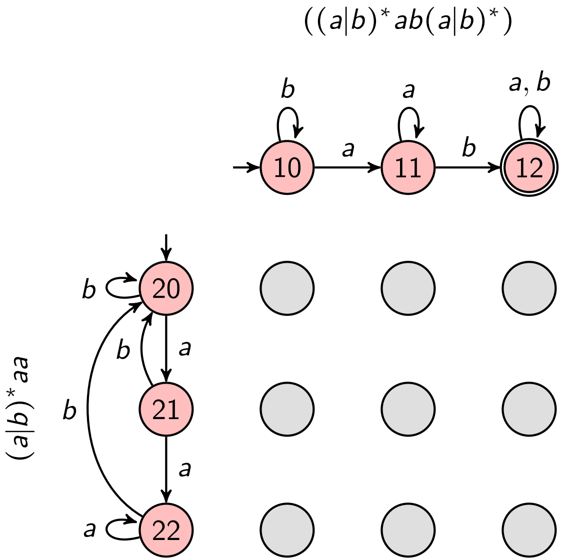
$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

Soient L_1 et L_2 deux langages réguliers.

- Par complémentation, $\overline{L_1}$ et $\overline{L_2}$ sont réguliers.
- L'ensemble des langages réguliers étant stable par union, $\overline{L_1} \cup \overline{L_2}$ est régulier.
- À nouveau par complémentation, $\overline{\overline{L_1} \cup \overline{L_2}}$ est régulier.
- Par la loi de De Morgan, $L_1 \cap L_2$ est régulier.

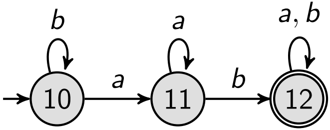
Ajoutons que la **différence ensembliste** de deux langages réguliers est un langage régulier puisque $L_1 \setminus L_2 \equiv L_1 \cap \overline{L_2}$.

Intersection - Exemple

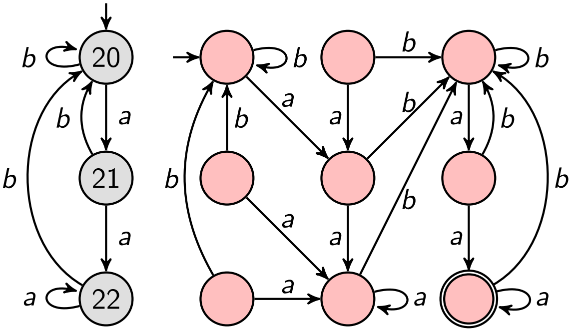


Intersection - Exemple

$((a|b)^*ab(a|b)^*)$



$(a|b)^*aa$



Miroir

Théorème 3 (stabilité par miroir)

Le langage miroir L^R d'un langage régulier L sur un alphabet Σ est régulier.

Miroir

L est régulier donc il existe $\mathcal{A} = (Q, \Sigma, q_0, F, \delta)$ tel que $L = \mathcal{L}(\mathcal{A})$.
On construit l'automate $\mathcal{A}_\mathsf{R} = (Q_\mathsf{R}, \Sigma, q_i, \{q_0\}, \delta_\mathsf{R})$ avec

- $Q_\mathsf{R} = Q \cup \{q_i\}$
- $\delta : \begin{cases} Q_\mathsf{R} \times \Sigma \to \mathcal{P}(Q_\mathsf{R}) \\ (q_i, \varepsilon) \mapsto F \\ (q, a) \mapsto \{p \mid q \in \delta(p, a), a \in \Sigma \cup \{\varepsilon\}\} \end{cases}$

L'automate $\mathcal{A}_\mathsf{R}$ est l'automate $\mathcal{A}$ dans lequel les transitions ont été inversées.
L'état initial q_0 de $\mathcal{A}$ devient l'état acceptant de $\mathcal{A}_\mathsf{R}$.
L'état q_i est nécessaire si F n'est pas un singleton ; sinon, il y aurait plusieurs états initiaux dans $\mathcal{A}_\mathsf{R}$.
$q_0 \xrightarrow{v}{}^*_{\mathcal{A}} q$ est un chemin acceptant pour v dans $\mathcal{A}$ si et seulement si $q_i \xrightarrow{\varepsilon} q \xrightarrow{v}{}^*_{\mathcal{A}_\mathsf{R}} q_0$ est un chemin acceptant pour v dans $\mathcal{A}_\mathsf{R}$.

Lemme de l'étoile

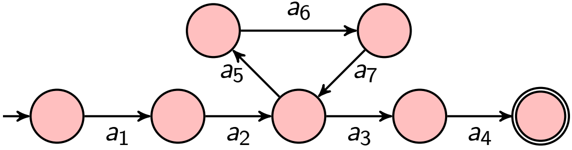
Idée intuitive du lemme de l'étoile

- Le **lemme de l'étoile** affirme que passé une certaine taille, tout mot d'un langage L peut être construit par répétition d'un facteur itérant y s'insérant au sein d'un mot xz de L.
- Ce résultat, évident pour tout **langage fini**, reste valable pour tout **langage de cardinal infini**.
- Ce lemme est aussi nommé **lemme de pompage** car le facteur itérant y peut être « pompé » autant de fois que nécessaire pour produire un mot de L.

Idée intuitive du lemme de l'étoile

Soit L un langage régulier reconnu par un automate à n états.

Lors de la lecture d'un mot de taille supérieure à n, l'exécution de l'automate passe deux fois par le même état.



Par exemple, avec $n = 7$, $w = a_1 a_2 (a_5 a_6 a_7) a_3 a_4 \in L$, on a $|w| \geqslant n$ et $a_1 a_2 (a_5 a_6 a_7)^* a_3 a_4 \in L$.

Formulation

Théorème 4 (Lemme de l'étoile)

Si L est un langage régulier, il existe $k \in \mathbb{N}$ tel que tout mot $w \in L$ vérifiant $|w| \geqslant k$ se factorise sous la forme $w = xyz$ avec :

- $|xy| \leqslant k$;
- $y \neq \varepsilon$;
- *pour tout entier naturel n,* $xy^n z \in L$.

L'entier k, appelé la **constante d'itération**, ne dépend que de L. y est appelé le **facteur itérant**.

La première condition $|xy| \leqslant k$ n'est pas indispensable mais elle permet généralement de raccourcir la preuve.

Démonstration

Démonstration

- Soit $\mathcal{A} = (\Sigma, Q, q_0, F, \delta)$ et $k = |Q|$.
- Soit w un mot de L tel que $|w| \geqslant k$.
- Le chemin $q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} q_2 \dots \xrightarrow{a_p} q_p$ reconnaissant w implique $p + 1$ états.
- Il passe nécessairement deux fois par le même état $q_i = q_j$ avec $1 \leqslant i < j \leqslant k$.

$$q_0 \xrightarrow{a_1} \dots \xrightarrow{a_i} q_i \dots \xrightarrow{a_j} q_j \dots \xrightarrow{a_p} q_p$$

- Posons $x = a_1 \dots a_i$, $y = a_{i+1} \dots a_j$ et $z = a_{j+1} \dots a_p$. Alors, pour tout entier naturel n, le chemin étiqueté par $xy^n z$ conduit à l'état acceptant q_p.
- Donc $xy^n z \in L$.

Méthodologie

En pratique, **prouver** qu'un langage L est **non régulier** se fait **par l'absurde**.

- Supposer que L est régulier.
- L possède un constante d'itération l et tous les mots suffisamment longs peuvent être pompés.
- Choisir alors un mot w de L.
- Considérer (si nécessaire) les façons de diviser w en xyz.
- Montrer que pour un certain entier i, $xy^iz \notin L$.
- w ne pouvant être pompé, aboutir à une contradiction et conclure que le langage n'est pas régulier.

Exemple 1

Soit $\Sigma = \{a, b\}$ un alphabet. Considérons le langage défini par :

$$L = \{a^p b^p \mid p \in \mathbb{N}\}$$

Montrer que L n'est pas régulier.

Exemple 1 - démonstration

Supposons L régulier. Il existe $k \in \mathbb{N}^*$ tel que tout mot w vérifiant $|w| \geqslant k$ satisfait le **lemme de l'étoile**. Prenons :

$$w = a^k b^k$$

Puisque $|w| = 2k$, w est assez long et $w \in L$. D'après le **lemme de l'étoile**, il existe x, y, z tels que $w = xyz$, avec :

- $|xy| \leqslant k$
- $y \neq \varepsilon$
- pour tout entier naturel n, $xy^n z \in L$.

Exemple 1 - démonstration

Puisque la condition $|xy| \leqslant k$ doit être vérifiée, y doit apparaître dans les k premiers caractères de w sous la forme a^i, pour un certain entier $i < k$.

Comme $y \neq \varepsilon$, alors $i \geqslant 1$.

On peut choisir $x = a^{k-i}$, $y = a^i$, $z = b^k$. On a bien :

$$xyz = a^{k-i}a^i b^k = a^k b^k = w$$

Exemple 1 - démonstration

Écrivons alors xy^2z (lemme avec $n = 2$). Ce mot est de la forme :

$$xy^2z = a^{k-i}a^{2i}b^k = a^{k+i}b^k$$

Puisque $i \geqslant 1$, $k + i \neq k$, ce mot n'appartient pas à L.

On vient de trouver un mot qui ne satisfait pas le lemme de l'étoile.

Donc L **n'est pas régulier**.

Exemple 2

Soit $\Sigma = \{), (\}$ un alphabet. Considérons le langage des expressions bien parenthésées défini par :

$$L = \{w \in \Sigma^* \mid w \text{ bien parenthésé}\}$$

Montrer que L n'est pas régulier.

Exemple 2 - démonstration

L contient des mots comme $(())((()())$ ou comme $(((())))$. Pour utiliser le **lemme de l'étoile**, l'idée est de considérer un mot aussi simple que possible. Par exemple, pour un entier naturel k :

$$w = (^k)^k$$

Puisque $|w| = 2k$ et $w \in L$, w doit satisfaire les conditions du lemme de l'étoile. Il doit dont exister x, y, z tels que $w = xyz$, , avec $|xy| \leqslant k$, $y \neq \varepsilon$, vérifiant :

$$\forall q \in \mathbb{N}, \quad xy^q z \in L$$

Puisque la condition $|xy| \leqslant k$ doit être vérifiée, y doit apparaître dans les k premiers caractères de w sous la forme $(^p$, pour un certain entier p. Comme $y \neq \varepsilon$, alors : $p \geqslant 1$.

Exemple 2 - démonstration

Puisque $q = 2$, alors $xy^q = ({}^{k+p}$. Puis :

$$xy^2 z = ({}^{k+p})^k$$

Ce mot doit appartenir à L, en vertu du lemme de l'étoile. Or ce mot est une expression mal parenthésée, qui n'est pas dans L.

Donc, L **n'est pas régulier**.