Théorie des Langages 1

Cours 7 : Propriétés de fermeture

L. Rieg

Grenoble INP - Ensimag, 1re année

Année 2022-2023

Stabilité des langages réguliers

Théorème

La classe des langages réguliers est fermée par :

- *union, concaténation et concaténation itérée (cf. cours 3)*
- *substitution régulière et homomorphisme*
- *complémentation*
- *intersection, différence*

Attention : les inclusions ne donnent rien !

L régulier et $L \subset M \not\Longrightarrow M$ régulier — penser à $L = \emptyset$

M régulier et $L \subset M \not\Longrightarrow L$ régulier — penser à $M = V^*$

Substitution régulière

Définition

Soit V et W deux vocabulaires.

Une substitution régulière est une fonction $s : V \to \mathcal{P}(W^*)$ qui à tout $a \in V$ associe un langage régulier $s(a) \subseteq W^*$.

On étend s aux mots par induction : $s^* : V^* \to \mathcal{P}(W^*)$

- $s^*(\varepsilon) = \{\varepsilon\}$
- $s^*(aw) = s(a).s^*(w)$

On étend s^* aux langages : $\overline{s} : \mathcal{P}(V^*) \to \mathcal{P}(W^*)$.

$$\forall L \subseteq V^*,\ \overline{s}(L) = \bigcup_{w \in L} s^*(w)$$

On pourra noter s au lieu de s^* ou $\overline{s}$.

Propriété de fermeture

Exemple

Soient $V = \{a, b\}$ et $W = \{c, d\}$. On pose :

$$\begin{aligned} L &= \{ab^i \mid i \geq 0\} \\ s(a) &= \{c^i \mid i \geq 0\} \\ s(b) &= \{cd\} \end{aligned}$$

Alors $s(L) =$

Théorème

La classe des langages réguliers est fermée par substitution régulière.

Autrement dit, si L est un langage régulier et s est une substitution régulière, alors $s(L)$ est un langage régulier.

Lemme intermédiaire sur les mots

Proposition

Soit s une substitution régulière. Pour tous $u, v \in V^*$, on a

$$s^*(u.v) = s^*(u).s^*(v) \ .$$

Preuve : Soit $v \in V^*$. On prouve que pour tout $u \in V^*$ on a $s^*(u.v) = s^*(u).s^*(v)$ par induction structurelle sur u.

Lemme intermédiaire sur les ER

Lemme

Soit s une substitution régulière, et soient E et E' des expressions régulières sur V. Alors

$$\begin{array}{lcl} s(E.E') & = & s(E).s(E') \\ s(E+E') & = & s(E) \cup s(E') \\ s(E^*) & = & s(E)^* \end{array}$$

Preuve.

Preuve du théorème

Étapes :

1. On étend les substitutions régulières aux expressions régulières :
   $$\forall \text{ ER } E \text{ sur } V,\ s(E) \stackrel{\text{def}}{=} s(\mathcal{L}(E)).$$
2. On prouve que si E est une expression régulière sur V, alors il existe une expression régulière E' sur W telle que $s(E) = \mathcal{L}(E')$.

Preuve du théorème de l'étape 2.
Par induction structurelle :

Base: Pour $E \in \{\varnothing, \epsilon, a\}$, OK : $\emptyset$, $\{\varepsilon\}$ et $s(a)$ ($s(a)$ régulier donc représentable par ER)

Induction: Pour $E \in \{E_1.E_2,\ E_1 + E_2,\ {E_1}^*\}$, cf lemme précédent

Homomorphismes

Définition

Une substitution régulière qui à tout $a \in V$ associe un singleton est un homomorphisme.

Exemple

$$\begin{array}{rcl} L & = & \{ab^i \mid i \geq 0\} = ab^* \\ s(a) & = & \{cdc\} = cdc \\ s(b) & = & \{dc\} = dc \end{array}$$

Alors $s(L) =$

Corollaire

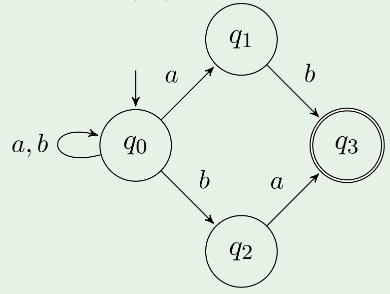
La classe des langages réguliers est fermée par homomorphisme.
(et par homomorphisme inverse, voir poly §2.3)

Complémentation

Question : si L est un langage régulier, peut-on construire un automate qui reconnaît $\overline{L} = V^* \setminus L$?

Exemple

$L = (a+b)^*(ab+ba)$

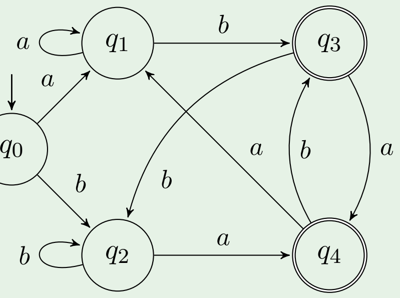


Complémentation (suite)

Problème : on ne peut pas intervertir F et $Q \setminus F$ car
il peut y avoir deux chemins de même trace dans un AFND,
et si l'un mène en F mais pas l'autre on accepte...
Solution : On déterminise...

Exemple

$L = (a+b)^*(ab+ba)$



Remarque : même problème et même solution s'il manque des chemins

Complémentation (fin)

Proposition

La classe des langages réguliers est fermée par complémentation.

Preuve.

Langages non-réguliers

Question : comment prouver qu'un langage n'est pas régulier ?
Autrement dit : Étant donné un langage L, comment prouver que pour tout automate fini A, $\mathcal{L}(A) \neq L$?

Idée : Utiliser des propriétés de fermeture

Supposons donné un langage M dont on connaît la non-régularité.

Pour prouver que L n'est pas régulier, on peut procéder par l'absurde :

1. On suppose que L est régulier.
2. On exhibe une série de transformations qui préservent la régularité et qui permettent de passer de L à M.
3. Donc M devrait être régulier.
   Contradiction : l'hypothèse que L était régulier est fausse.

Cette technique de preuve est appelée réduction : on réduit le problème de la régularité de L à celle de M. (*cf.* TL2)

Exercice

On admet que $M = \{0^p 1^p \mid p \geq 0\}$ n'est pas régulier.

Montrer que $L = \{wcw' \mid w, w' \in \{a,b\}^*,\ |w|_a = |w'|_b\}$ n'est pas régulier.

Un premier langage non-régulier

Pour utiliser les propriétés de fermeture, il faut connaître au moins un langage M non-régulier.

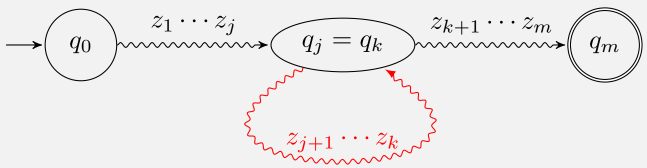
Question : comment prouver que M n'est pas régulier ?

- En se servant d'une condition nécessaire sur les langages réguliers qui permettra de refaire un raisonnement par l'absurde
- On va supposer qu'il existe un automate fini A tel que $\mathcal{L}(A) = M$ et tenter d'aboutir à une contradiction
- La condition nécessaire la plus standard est donnée par le lemme de l'étoile

  (lemme de pompage, de la pompe, *pumping lemma*)

Principe du lemme de l'étoile

Soit $A = \langle Q, V, \delta, I, F\rangle$ un automate *sans ε-transition*, avec $|Q| = n \geq 1$.
Soit $z = z_1 \cdots z_m$ un mot sur V de longueur $m \geq n$ reconnu par A.
Il existe donc un chemin $(q_0, z_1, q_1) \cdots (q_{m-1}, z_m, q_m)$ dans A,
avec $q_0 \in I$ et $q_m \in F$.
Il y a $m+1$ états dans ce chemin et n états dans Q, et $m+1 > n$
donc $\exists j, k$ tels que $0 \leq j < k \leq n$ et $q_j = q_k$



$$z_1 \cdots z_j (z_{j+1} \cdots z_k)^2 z_{k+1} \cdots z_m \in \mathcal{L}(A)$$

Énoncé du lemme de l'étoile

Lemme

Soit L un langage régulier. Alors il existe $n \geq 1$ tel que pour tout mot z, si $z \in L$ et $|z| \geq n$, alors z est de la forme uvw avec :

- $|uv| \leq n$
- $|v| \geq 1$
- $\forall i \geq 0, uv^i w \in L \qquad \iff uv^*w \subseteq L$

Attention

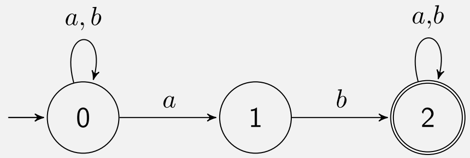
Le lemme de l'étoile est une condition nécessaire mais non suffisante des langages réguliers : il existe des langages non-réguliers qui le satisfont.

Remarques

- *Que se passe-t-il pour les langages finis ?*
- *Des lemmes de l'étoile existent aussi pour d'autres classes de langages.*

Illustrations

- Pour le langage V^*abV^* (les mots qui contiennent ab) :



On a $n = 3$ (nombre d'états de l'automate).

Pour $z = abb$, le chemin acceptant dans l'automate donne $u = ab$, $v = b$, $w = \varepsilon$.

Pour $z = aab$, le chemin acceptant dans l'automate donne $u = \varepsilon$, $v = a$, $w = ab$.

Et pour $z = aabb$?

- Pour le langage $\{ab\}$: même automate, moins les boucles.
On a $n = 3$ et il n'existe pas de mot $z \in \{ab\}$ tel que $|z| \geq 3$.
Le lemme de l'étoile est vrai par vacuité.

Comment se servir du lemme de l'étoile ?

On procède par l'absurde : on suppose que L est régulier et satisfait donc le lemme de l'étoile.

- On considère l'entier n du lemme. (on ne sait rien de sa valeur)
- On choisit un mot $z \in L$ de longueur au moins n. (z dépendra de n)
- Le mot z est décomposé en uvw, où $|uv| \leq n$ et $|v| \geq 1$.
(on ne contrôle pas la façon dont z est décomposé, hormis la contrainte sur les longueurs)
- On choisit une valeur de i telle que $uv^iw \notin L$.

On obtient une contradiction : L ne peut pas être régulier.

Remarque

On utilise en fait la contraposée du lemme de l'étoile :

$$L \text{ régulier} \Rightarrow \exists n \geq 1, \forall z \in L, \ldots$$
$$\forall n \geq 1, \exists z \in L, \ldots \Rightarrow L \text{ non régulier}$$

Exemple

Montrer que $M = \{0^p1^p \mid p \geq 0\}$ n'est pas régulier.