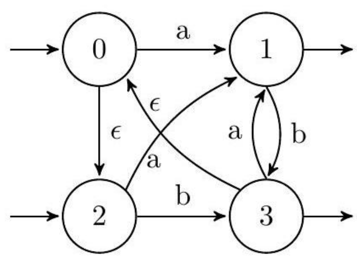
I Langage local, linéaire et automate de Glushkov

1. Construire l'automate de Glushkov reconnaissant $(ab|b)^*b$.
2. Montrer que si L_1 et L_2 sont locaux alors $L_1 \cap L_2$ est local (même si les alphabets ne sont pas disjoints, contrairement à l'union et concaténation vue en cours).
3. Montrer qu'il existe un nombre fini de langages locaux sur un alphabet fixé.

> Solution : Tout langage local est de la forme $(P\Sigma^* \cap \Sigma^* S) \setminus \Sigma^* N \Sigma^*$ où $P \subseteq \Sigma$, $S \subseteq \Sigma$, $N \subseteq \Sigma \times \Sigma$, en rajoutant éventuellement ε. Si $n = |\Sigma|$, il y a au plus 2^n choix pour P, 2^n pour S et 2^{2n} pour N (car $|\Sigma \times \Sigma| = |\Sigma||\Sigma| = 2^{2n}$). D'où au plus $2^n \times 2^n \times 2^{2n}$ valeurs différentes pour l'ensemble $(P\Sigma^* \cap \Sigma^* S) \setminus \Sigma^* N \Sigma^*$.

4. Soient $\mathcal{L}_{\text{rat}}$, $\mathcal{L}_{\text{loc}}$, $\mathcal{L}_{\text{lin}}$ l'ensemble des langages réguliers, locaux et linéaires (c'est à dire décrits par une expression régulière linéaire). Quelles sont les inclusions entre ces 3 ensembles ? Ces inclusions sont-elles strictes ?

II Exercice CCP (élimination des états)



1. Déterminiser cet automate.
2. Construire une expression régulière dénotant le langage reconnu par cet automate, à l'aide de la méthode d'élimination des états.
3. Décrire simplement avec des mots le langage reconnu par cet automate.

III Stabilité des langages réguliers

Pour chacune des propositions suivantes, expliquer pourquoi elle est vraie ou donner un contre-exemple :

1. Si L est régulier et $L \subseteq L'$ alors L' est régulier.
2. Si L' est régulier et $L \subseteq L'$ alors L est régulier.
3. Si L est régulier alors L^* est régulier.
4. Si L^* est régulier alors L est régulier.
5. Si L est régulier sur un alphabet Σ alors $\Sigma^* \setminus L$ (complémentaire de L) est régulier.
6. Une union finie de langages réguliers est régulier (L_1, ..., L_n réguliers $\implies \displaystyle\bigcup_{k=1}^{n} L_k$ régulier).
7. Une union dénombrable de langages réguliers est régulier (L_1, ..., L_n, ... réguliers $\implies \displaystyle\bigcup_{k=1}^{\infty} L_k$ régulier).
8. Une intersection finie de langages réguliers est régulier (L_1, ..., L_n réguliers $\implies \displaystyle\bigcap_{k=1}^{n} L_k$ régulier).
9. Une intersection dénombrable de langages réguliers est régulier (L_1, ..., L_n, ... réguliers $\implies \displaystyle\bigcap_{k=1}^{\infty} L_k$ régulier).

IV Reconnaissable $\implies$ régulier par programmation dynamique ($\approx$ Floyd-Warshall)

Cet exercice est une alternative à la méthode d'élimination des états pour obtenir une expression régulière à partir d'un automate.

Soit $(\Sigma, Q, 0, F, \delta)$ un automate déterministe tel que $Q = \{0, ..., n-1\}$.
Soit $L(i, j, k)$ le langage des étiquettes des chemins de i à j n'utilisant que des états intermédiaires strictement inférieurs à k.

1. Montrer que $L(i, j, 0)$ est régulier, pour tous les états i, j.

> Solution : Soit i et j deux états (éventuellement égaux) et $a_1, ..., a_p$ toutes les étiquettes des transitions de i vers j. Alors $L(i, j, 0) = \{a_1, ..., a_p\}$ est fini donc régulier.

2. Donner une équation de récurrence sur $L(i, j, k)$, en la démontrant.

> Solution : Soit $u \in L(i, j, k)$ l'étiquette d'un chemin C de i vers j n'utilisant que des états intermédiaires strictement inférieurs à k. Si C n'utilise pas l'état k alors $u \in L(i, j, k-1)$. Sinon C est la concaténation d'un chemin de i à k, d'un certain nombre de cycles de k à k, puis d'un chemin de k à j. On en déduit :
>
> $$L(i, j, k) = L(i, j, k-1) \cup (L(i, k, k-1)L(k, k, k-1)^* L(k, j, k-1))$$

3. En déduire, par récurrence, que tout langage reconnaissable est régulier.

> Solution : Montrons par récurrence sur $k \in \{0, ..., n\}$, $P(k)$: pour tous états i, j, $L(i, j, k)$ est régulier ».
> $P(0)$ est vraie d'après la question 1.
> Supposons $P(k)$ vraie pour un $k > 0$. Soient i, j deux états. Alors $L(i, j, k-1)$, $L(i, k, k-1)$, $L(k, k, k-1)$, $L(k, j, k-1)$ sont réguliers par hypothèse de récurrence donc $L(i, j, k) = L(i, j, k-1) \cup (L(i, k, k-1)L(k, k, k-1)^* L(k, j, k-1))$ est régulier comme concaténation et étoile de langages réguliers.

4. Écrire le pseudo-code d'un algorithme ayant pour entrée un automate déterministe et renvoyant une expression régulière dénotant le langage reconnu par cet automate. Préciser la complexité.

V Résiduels et minimisation d'automate

Si $u \in \Sigma^\star$ et L un langage sur Σ, on appelle résiduel de L par rapport à u le langage $u^{-1}L$ tel que :

$$u^{-1}L = \{m \in \Sigma^\star \mid um \in L\}$$

Intuitivement le résiduel de L par rapport à u est l'ensemble des mots avec lesquels on peut compléter u pour obtenir un mot de L. L'ensemble des résiduels de L est noté $\mathcal{R}_L = \{u^{-1}L \mid u \in \Sigma^\star\}$.

1. Si L est un langage, quel est le résiduel de L par rapport à ε ?

> Solution : $\varepsilon^{-1}L = L$. En effet, si $m \in \Sigma^\star$, $m \in \varepsilon^{-1}L \Leftrightarrow \varepsilon m = m \in L$.

2. Montrer que pour tout langage L sur Σ et tous mots $u, v \in \Sigma^\star$, $u^{-1}(v^{-1}L) = (vu)^{-1}L$.

> Solution : Soit L un langage et $u, v \in \Sigma^\star$. Soit $m \in \Sigma^\star$. Alors :
>
> $$m \in u^{-1}(v^{-1}L) \Leftrightarrow um \in v^{-1}L \Leftrightarrow vum \in L \Leftrightarrow m \in (vu)^{-1}L$$

3. a) Soit $L_1 = (a|b)^\star ab(a|b)^\star$. Déterminer $(bab)^{-1}L_1$ et $(aaaba)^{-1}L_1$. Que constate-t-on ?

> Solution : Le langage L_1 est l'ensemble des mots contenant ab. Les deux mots proposés contiennent déjà le motif ab, on peut donc les compléter par n'importe quoi et rester dans L_1. On en déduit que $(bab)^{-1}L = (aaaba)^{-1}L = (a|b)^\star$.

   b) Expliciter $\mathcal{R}_{L_1}$.

Solution : $u^{-1}L_1 = \begin{cases} (a|b)^\star \text{ si } u \in L_1 \\ L|b(a|b)^\star \text{ si } u \notin L_1 \text{ et finit par un } a \\ L \text{ si } u \notin L_1 \text{ et finit par un } b \end{cases}$ D'où : $\mathcal{R}_{L_1} = \{(a|b)^\star, L_1|b(a|b)^\star, L_1\}$

On veut montrer qu'un langage L est reconnaissable si et seulement s'il admet un nombre fini de résiduels.

4. Soit L un langage reconnaissable par un automate $A = (\Sigma, Q, q_0, F, \delta)$ qu'on peut supposer déterministe, complet et dont tous les états sont accessibles sans perte de généralité.

   a) Montrer que la fonction $\varphi : \begin{cases} Q \longrightarrow \mathcal{R}_L \\ q \longmapsto u^{-1}L \text{ où } u \text{ est l'un des mots tels que } \delta^\star(q_0, u) = q \end{cases}$ est correctement définie.

   Solution : Soit $q \in Q$. Comme A est accessible, il existe un mot u permettant d'atteindre l'état q depuis q_0. Par ailleurs, le déterminisme de A assure que $\delta^\star(q_0, u)$ est bien réduit à un seul état : q. En particulier, il existe bien un mot u tel que $\delta^\star(q_0, u) = q$ ce qui permet de construire l'image de q par φ.

   Il reste à montrer que cette image ne dépend pas du choix du mot u. Pour ce faire, considérons deux mots u et v tels que $\delta^\star(q_0, u) = \delta^\star(q_0, v)$. Montrons qu'alors $u^{-1}L = v^{-1}L$. Si $m \in \Sigma^\star$, on a :

$$\begin{aligned} m \in u^{-1}L &\Leftrightarrow um \in L \text{ par définition d'un résiduel} \\ &\Leftrightarrow \delta^\star(q_0, um) \in F \text{ puisque } A \text{ reconnaît } L \\ &\Leftrightarrow \delta^\star(\delta^\star(q_0, u), m) \in F \\ &\Leftrightarrow \delta^\star(\delta^\star(q_0, v), m) \in F \text{ par hypothèse sur } u \text{ et } v \\ &\Leftrightarrow \delta^\star(q_0, vm) \in F \\ &\Leftrightarrow vm \in L \Leftrightarrow m \in v^{-1}L \end{aligned}$$

   Ainsi le résiduel associé à l'état q par φ est indépendant du choix du mot u.

   b) Montrer que L admet un nombre fini de résiduels.

   Solution : Montrons que φ est surjective. Soit $u^{-1}L$ un résiduel de L. Comme A est complet, le mot u peut être lu à partir de q_0 et il existe donc un état $q \in Q$ tel que $\delta^\star(q_0, u) = q$. En particulier, $u^{-1}L = \varphi(q)$.

   Ceci garantit que $|Q| \geq |\{u^{-1}L \,|\, u \in \Sigma^\star\}|$. Mais l'automate A est un automate fini donc $|Q|$ est fini ce qui garantit que le langage L a un nombre fini de résiduels.

5. Soit à présent L un langage ayant un nombre fini de résiduels. On définit alors un automate $M(L) = (\Sigma, Q, I, F, \delta)$, appelé automate des résiduels associé à L, où :
   - $Q = \mathcal{R}_L$.
   - $I = \{\varepsilon^{-1}L\}$.
   - $F = \{u^{-1}L \,|\, u \in L\}$.
   - pour toute lettre $a \in \Sigma$ et tout résiduel $u^{-1}L \in Q$, $\delta(u^{-1}L, a) = (ua)^{-1}L$.

   a) Montrer que la fonction δ est correctement définie, c'est-à-dire que l'image donnée par δ d'un résiduel $u^{-1}L$ ne dépend que du résiduel et pas de u.

   b) Montrer que $M(L)$ est un automate déterministe et complet.

   c) Montrer que $M(L)$ reconnaît le langage L.

   Solution :

   a) Soit $u, v \in \Sigma^\star$ tels que $u^{-1}L = v^{-1}L$ et $a \in \Sigma$. Montrons que le résiduel $\delta(u^{-1}L, a)$ et le résiduel $\delta(v^{-1}L, a)$ sont des langages égaux par équivalences. Si $m \in \Sigma^\star$, on a :

$$\begin{aligned} m \in \delta(u^{-1}L, a) &\Leftrightarrow m \in a^{-1}(u^{-1}L) \text{ par définition de } \delta \\ &\Leftrightarrow m \in (ua)^{-1}L \text{ d'après la question 2} \\ &\Leftrightarrow uam \in L \text{ par définition d'un résiduel} \\ &\Leftrightarrow am \in u^{-1}L \\ &\Leftrightarrow am \in v^{-1}L \text{ par égalité des deux résiduels} \\ &\Leftrightarrow m \in a^{-1}(v^{-1}L) = \delta(v^{-1}L, a) \end{aligned}$$

L'égalité de ces deux langages est maintenant démontrée ce qui garantit que l'état atteint en lisant a depuis l'état étiqueté par le résiduel $u^{-1}L$ ne dépend que de ce résiduel et non du choix de u.

b) Par hypothèse, l'ensemble des résiduels de L — qui est aussi l'ensemble des états de $M(L)$ — est fini donc $M(L)$ est un automate fini. Son caractère déterministe et complet est immédiat par construction de sa fonction de transition et le fait qu'il ait un seul état initial.

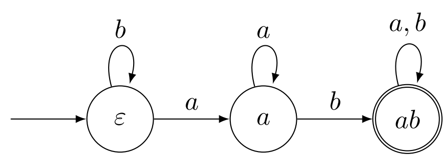
c) Procédons pour ce faire par double inclusion.

Si $m \in L$, $\delta^\star(\varepsilon^{-1}L, m) = m^{-1}L$ en lisant le mot m lettre à lettre et en utilisant la question 2. Or $m^{-1}L \in F$ puisque $m \in L$. On en déduit que lire m depuis l'état initial de $M(L)$ conduit à un état final de $M(L)$, c'est-à-dire que m est reconnu par $M(L)$.

Réciproquement, si m est reconnu par $M(L)$ alors il existe $u \in L$ tel que $m^{-1}L = \delta^\star(\varepsilon^{-1}L, m) = u^{-1}L$. Comme $u \in L$, $\varepsilon \in u^{-1}L$ et donc on a aussi $\varepsilon \in m^{-1}L$ ce qui garantit que $m \in L$.

6. A l'aide des questions précédentes, déterminer l'automate des résiduels associé au langage L_1.

Solution :



Un automate reconnaissant un langage L est dit minimal s'il est déterministe, complet et possède le plus petit nombre d'états possible parmi les automates déterministes et complets reconnaissant L.

7. Montrer que l'automate des résiduels est un automate minimal.

On peut montrer que cet automate minimal est en fait unique (à renommage des états près) ce qui fournit une façon de déterminer si deux automates reconnaissent le même langage : il suffit de les minimiser et comparer les deux automates obtenus (une autre méthode consistant à tester si $L_1 \Delta L_2 = \emptyset$: cf TP 2).
Il est possible de déterminer l'automate minimal équivalent à un automate donné grâce à l'algorithme de Moore.

VI Racine d'un langage

Soit L un langage reconnaissable sur Σ. Montrer que $\sqrt{L} = \{m \in \Sigma^* \mid m^2 \in L\}$ est un langage reconnaissable.

Solution : Soit $A = (\Sigma, Q, i, F, \delta)$ un automate déterministe reconnaissant L dont les états sont des entiers 0, ..., $n-1$.
Soit $L_{j,k}$ le langage des étiquettes des chemins de l'état j à l'état k dans A. $L_{j,k}$ est reconnaissable, par $(\Sigma, Q, j, \{k\}, \delta)$.
De plus, on montre que $\sqrt{L} = \displaystyle\bigcup_{q \in Q} \bigcup_{f \in F} L_{i,q} \cap L_{q,f}$:

- Soit $m \in \sqrt{L}$. Soit $q_m = \delta^*(i, m)$ et $f_m = \delta^*(q_m, m)$. Alors $m \in L_{i,q_m}$ et $m \in L_{q_m,f_m}$ et $f_m \in F$ (car $m^2 \in L$). Donc $m \in L_{i,q_m} \cap L_{q_m,f_m} \subseteq \displaystyle\bigcup_{q \in Q} \bigcup_{f \in F} L_{i,q} \cap L_{q,f}$.

- Soit $m \in \displaystyle\bigcup_{q \in Q} \bigcup_{f \in F} L_{i,q} \cap L_{q,f}$: il existe alors $q \in Q$ et $f \in F$ tels que $m \in L_{i,q} \cap L_{q,f}$. Alors $\delta^*(i, m^2) = \delta^*(q, m) = f \in F$.

D'où $m^2 \in L$, donc $m \in \sqrt{L}$.

Ainsi $\sqrt{L} = \bigcup\limits_{q\in Q} \bigcup\limits_{f\in F} L_{i,q} \cap L_{q,f}$ est régulier comme union et intersection de langages réguliers (une union de langages réguliers est régulier par définition, et une intersection de langages réguliers est régulier en utilisant par exemple l'automate produit du cours).