

Théorème de Kleene

MPI

Q. Fortier

Expressions régulières linéaires

Définition 1 : Expressions régulières linéaires

Une expression régulière est *linéaire* si chaque lettre y apparaît au plus une fois.

Expressions régulières linéaires

Définition 1 : Expressions régulières linéaires

Une expression régulière est *linéaire* si chaque lettre y apparaît au plus une fois.

Exemple : $ba \mid a$ n'est pas linéaire, mais a^*b est linéaire.

Définition 2 : Linéarisation

Linéariser une expression e consiste à remplacer chaque occurrence de lettre a par une nouvelle lettre $\varphi(a)$ différente.

Expressions régulières linéaires

Définition 1 : Expressions régulières linéaires

Une expression régulière est *linéaire* si chaque lettre y apparaît au plus une fois.

Exemple : $ba \mid a$ n'est pas linéaire, mais a^*b est linéaire.

Définition 2 : Linéarisation

Linéariser une expression e consiste à remplacer chaque occurrence de lettre a par une nouvelle lettre $\varphi(a)$ différente.

Exemple : $e = \varepsilon \mid b(a \mid bb)^*b$ peut se linéariser en $e' = \varepsilon \mid b_0(a_0 \mid b_1b_2)^*b_3$.

Définition 3 : $P(L)$, $S(L)$, $F(L)$, $N(L)$

Pour un langage $L \subset \Sigma^*$:

- $P(L) = \{a \in \Sigma \mid a\Sigma^* \cap L \neq \emptyset\}$: premières lettres.
- $S(L) = \{a \in \Sigma \mid \Sigma^*a \cap L \neq \emptyset\}$: dernières lettres.
- $F(L) = \{m \in \Sigma^2 \mid \Sigma^*m\Sigma^* \cap L \neq \emptyset\}$: facteurs de taille 2.
- $N(L) = \Sigma^2 \setminus F(L)$: facteurs de taille 2 interdits.

Définition 3 : $P(L)$, $S(L)$, $F(L)$, $N(L)$

Pour un langage $L \subset \Sigma^*$:

- $P(L) = \{a \in \Sigma \mid a\Sigma^* \cap L \neq \emptyset\}$: premières lettres.
- $S(L) = \{a \in \Sigma \mid \Sigma^*a \cap L \neq \emptyset\}$: dernières lettres.
- $F(L) = \{m \in \Sigma^2 \mid \Sigma^*m\Sigma^* \cap L \neq \emptyset\}$: facteurs de taille 2.
- $N(L) = \Sigma^2 \setminus F(L)$: facteurs de taille 2 interdits.

Exercice 1

Donner ces ensembles pour $L = a(a \mid b)^*$.

Définition 4 : Langage local

Soit $\text{Loc}(L)$ l'ensemble des mots dont la première lettre appartient à $P(L)$, la dernière à $S(L)$ et sans facteur de $N(L)$:

$$\text{Loc}(L) = (P(L)\Sigma^* \cap \Sigma^*S(L)) \setminus \Sigma^*N(L)\Sigma^*$$

On dit que L est *local* si $\text{Loc}(L) = L \setminus \{\varepsilon\}$, c'est-à-dire, comme $L \setminus \{\varepsilon\} \subset L$ est toujours vrai :

$$\underbrace{m_1 \in P(L) \wedge m_n \in S(L) \wedge (\forall k \in \llbracket 1, n-1 \rrbracket, m_k m_{k+1} \in F(L))}_{m \in \text{Loc}(L)} \implies m \in L$$

On note $\mathcal{L}_{\text{loc}}(\Sigma)$ l'ensemble des langages locaux sur Σ .

Exercice 2

Montrer que $\mathcal{L}_{\text{loc}}(\Sigma) \subset \mathcal{L}_{\text{reg}}(\Sigma)$.

Exercice 3

Les langages suivants sont-ils locaux ?

1. $L_1 = a^*$
2. $L_2 = (ab)^*$
3. $L_3 = a^* \mid (ab)^*$
4. $L_4 = a^*(ab)^*$

Théorème 1

- Si $L_1 \in \mathcal{L}_{\text{loc}}(\Sigma_1)$, $L_2 \in \mathcal{L}_{\text{loc}}(\Sigma_2)$ avec $\Sigma_1 \cap \Sigma_2 = \emptyset$ alors $L_1 \cup L_2 \in \mathcal{L}_{\text{loc}}(\Sigma_1 \cup \Sigma_2)$ et $L_1 L_2 \in \mathcal{L}_{\text{loc}}(\Sigma_1 \cup \Sigma_2)$.
- Si $L \in \mathcal{L}_{\text{loc}}(\Sigma)$ alors $L^* \in \mathcal{L}_{\text{loc}}(\Sigma)$.

Théorème 1

- Si $L_1 \in \mathcal{L}_{\text{loc}}(\Sigma_1)$, $L_2 \in \mathcal{L}_{\text{loc}}(\Sigma_2)$ avec $\Sigma_1 \cap \Sigma_2 = \emptyset$ alors $L_1 \cup L_2 \in \mathcal{L}_{\text{loc}}(\Sigma_1 \cup \Sigma_2)$ et $L_1 L_2 \in \mathcal{L}_{\text{loc}}(\Sigma_1 \cup \Sigma_2)$.
- Si $L \in \mathcal{L}_{\text{loc}}(\Sigma)$ alors $L^* \in \mathcal{L}_{\text{loc}}(\Sigma)$.

Exercice 4

1. La stabilité par union et concaténation reste-t-elle vraie si $\Sigma_1 \cap \Sigma_2 \neq \emptyset$?
2. $\mathcal{L}_{\text{loc}}(\Sigma)$ est-il stable par complémentaire ?

Théorème 2 : Linéaire $\implies$ local

Si e est une expression régulière linéaire alors $\mathcal{L}(e)$ est local.

Définition 5 : Automate local

Un AD $A = (\Sigma, Q, q_0, F, \delta)$ est *local* si toutes les transitions étiquetées par une même lettre aboutissent au même état :

$$\forall q_1, q_2 \in Q, a \in \Sigma, \delta(q_1, a) = \delta(q_2, a)$$

Automate local

Définition 5 : Automate local

Un AD $A = (\Sigma, Q, q_0, F, \delta)$ est *local* si toutes les transitions étiquetées par une même lettre aboutissent au même état :

$$\forall q_1, q_2 \in Q, a \in \Sigma, \delta(q_1, a) = \delta(q_2, a)$$

Théorème 3 ♡

$L \in \mathcal{L}_{\text{loc}}(\Sigma) \iff L$ est reconnu par un automate local.

Automate local

Définition 5 : Automate local

Un AD $A = (\Sigma, Q, q_0, F, \delta)$ est *local* si toutes les transitions étiquetées par une même lettre aboutissent au même état :

$$\forall q_1, q_2 \in Q, a \in \Sigma, \delta(q_1, a) = \delta(q_2, a)$$

Théorème 3 ♡

$L \in \mathcal{L}_{\text{loc}}(\Sigma) \iff L$ est reconnu par un automate local.

Exercice 5

Construire un automate local reconnaissant $a(a \mid b)^*$.

Algorithme de Berry-Sethi

Algorithme 1 : Berry-Sethi ♡

Entrée : une expression régulière e .

Sortie : un automate *de Glushkov* reconnaissant $\mathcal{L}(e)$.

Linéariser e en e' .

Calculer $P(\mathcal{L}(e'))$, $S(\mathcal{L}(e'))$ et $F(\mathcal{L}(e'))$, et tester $\varepsilon \in \mathcal{L}(e')$.

Construire l'automate local A' reconnaissant $\mathcal{L}(e')$ comme dans la preuve de Théorème 3.

Renvoyer l'automate A obtenu en remplaçant chaque étiquette a' de A' par sa lettre d'origine $\pi(a')$.

Algorithme de Berry-Sethi

Algorithme 1 : Berry-Sethi ♡

Entrée : une expression régulière e .

Sortie : un automate *de Glushkov* reconnaissant $\mathcal{L}(e)$.

Linéariser e en e' .

Calculer $P(\mathcal{L}(e'))$, $S(\mathcal{L}(e'))$ et $F(\mathcal{L}(e'))$, et tester $\varepsilon \in \mathcal{L}(e')$.

Construire l'automate local A' reconnaissant $\mathcal{L}(e')$ comme dans la preuve de Théorème 3.

Renvoyer l'automate A obtenu en remplaçant chaque étiquette a' de A' par sa lettre d'origine $\pi(a')$.

Exercice 6

Appliquer Algorithme : Berry-Sethi à $e = a(a \mid b)^*$.

$$(\Sigma \cup \{\varepsilon\}) \rightarrow \mathcal{P}(Q).$$

$$(\Sigma \cup \{\varepsilon\}) \rightarrow \mathcal{P}(Q).$$

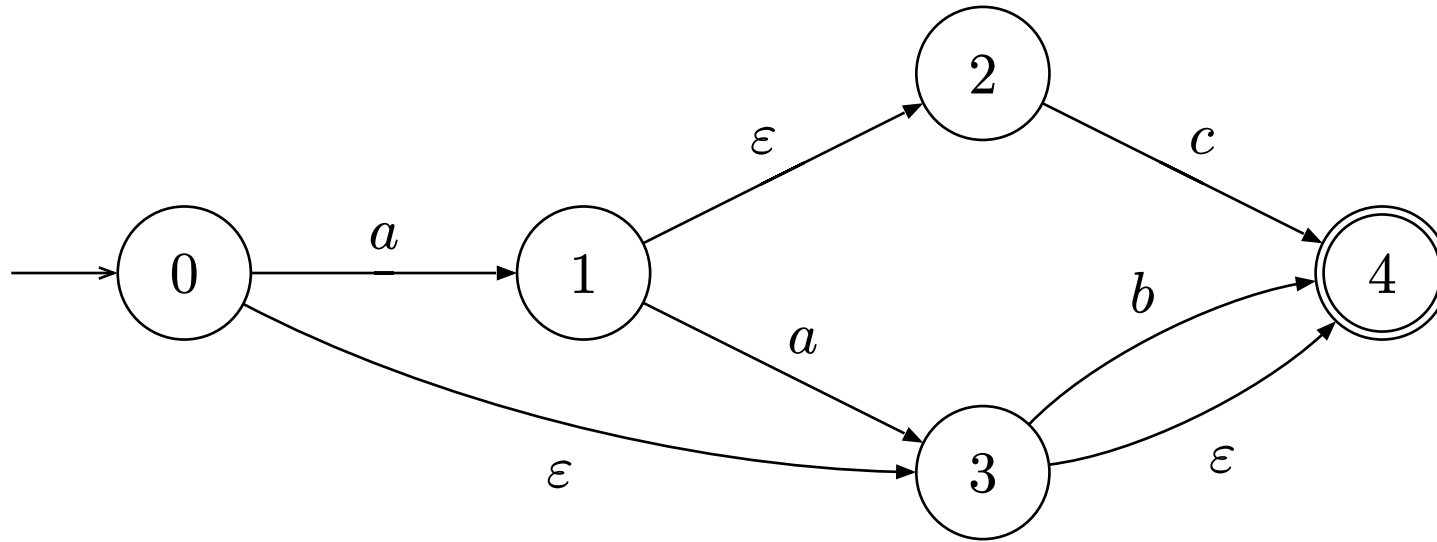


Fig. 1. – Automate à ε -transitions reconnaissant $\{\varepsilon, b, ac, aa, aab\}$.

$$\varepsilon(0) = \{0, 3, 4\} \text{ et } \varepsilon(1) = \{1, 2\}$$

Définition 7 : ε -fermeture

Pour un ensemble d'états $X \subset Q$, on note $\varepsilon(X)$ l' ε -fermeture de X , c'est-à-dire l'ensemble des états atteignables depuis X en utilisant uniquement des ε -transitions.

$$(\Sigma \cup \{\varepsilon\}) \rightarrow \mathcal{P}(Q).$$

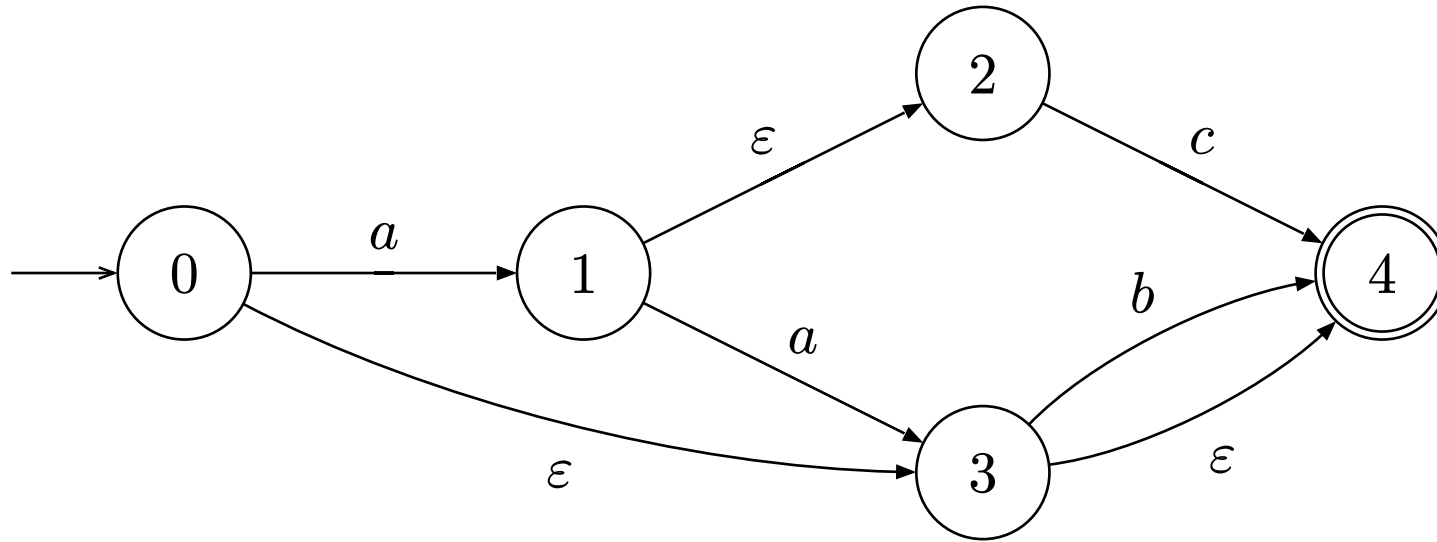


Fig. 1. – Automate à ε -transitions reconnaissant $\{\varepsilon, b, ac, aa, aab\}$.

$$\varepsilon(0) = \{0, 3, 4\} \text{ et } \varepsilon(1) = \{1, 2\}$$

Définition 7 : ε -fermeture

Pour un ensemble d'états $X \subset Q$, on note $\varepsilon(X)$ l' ε -fermeture de X , c'est-à-dire l'ensemble des états atteignables depuis X en utilisant uniquement des ε -transitions.

Théorème 4 : Suppression des ε -transitions

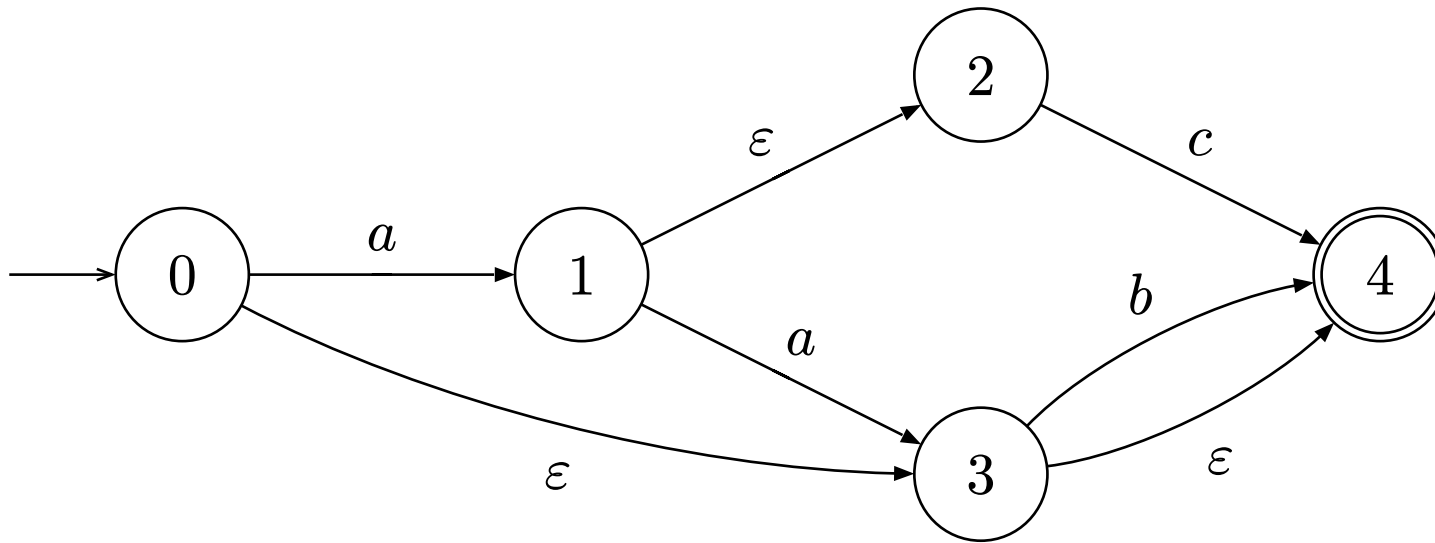
Tout automate avec ε -transitions est équivalent à un automate sans ε -transition.

Exercice 8

Donner un automate sans ϵ -transition équivalent à Fig. 1.

Exercice 8

Donner un automate sans ϵ -transition équivalent à Fig. 1.



Remarque : un automate avec ϵ -transitions n'est pas déterministe, mais il peut être déterminisé en adaptant l'algorithme de déterminisation ou en utilisant préalablement Théorème : Suppression des ϵ -transitions.

Automate de Thompson

La construction de Thompson est une alternative récursive à [Algorithme : Berry-Sethi](#) pour obtenir un automate à partir d'une expression régulière.

Exercice 9 : Algorithme de Thompson

1. Définir récursivement un automate $T(e)$ reconnaissant une expression régulière e .
2. Dessiner $T(a(ab \mid b)^*)$.
3. Comparer les tailles des automates de Thompson et de Glushkov.

Reconnaissable $\implies$ régulier

Lemme 1 : Normalisation

Tout automate est équivalent à un automate avec ε -transitions *normalisé*, c'est-à-dire avec :

- Un unique état initial sans transition entrante.
- Un unique état final sans transition sortante.

Reconnaissable $\implies$ régulier

Définition 8 : Automate généralisé

Un *automate généralisé* $A = (\mathcal{R}(\Sigma), Q, I, F, T)$ sur un alphabet Σ permet des transitions étiquetées par des expressions régulières avec $T \subset Q \times \mathcal{R}(\Sigma) \times Q$ fini.

Un mot m est *accepté* par A s'il existe un *chemin acceptant* de I à F $q_0 \xrightarrow{e_1} q_1 \xrightarrow{e_2} \dots \xrightarrow{e_n} q_n$ tel que $m \in \mathcal{L}(e_1) \dots \mathcal{L}(e_n)$.

Le *langage* $\mathcal{L}(A)$ *reconnu* par A est l'ensemble des mots acceptés par A .

Reconnaissable $\implies$ régulier

Définition 8 : Automate généralisé

Un *automate généralisé* $A = (\mathcal{R}(\Sigma), Q, I, F, T)$ sur un alphabet Σ permet des transitions étiquetées par des expressions régulières avec $T \subset Q \times \mathcal{R}(\Sigma) \times Q$ fini.

Un mot m est *accepté* par A s'il existe un *chemin acceptant* de I à F $q_0 \xrightarrow{e_1} q_1 \xrightarrow{e_2} \dots \xrightarrow{e_n} q_n$ tel que $m \in \mathcal{L}(e_1) \dots \mathcal{L}(e_n)$.

Le *langage* $\mathcal{L}(A)$ *reconnu* par A est l'ensemble des mots acceptés par A .

Remarques :

- Un automate avec ε -transitions est un cas particulier d'automate généralisé.
- Une transition de $p \xrightarrow{e} q$ peut être franchie en lisant un mot quelconque de $\mathcal{L}(e)$.

Théorème 5 HP

Tout automate généralisé est équivalent à un automate fini classique.

Théorème 5 HP

Tout automate généralisé est équivalent à un automate fini classique.

Exercice 10 : Substitution régulière

Soit $L \subset \Sigma^*$ un langage reconnaissable, Γ un alphabet et $s : \Sigma \rightarrow \mathcal{R}(\Gamma)$ qu'on étend à $s : \Sigma^* \rightarrow \mathcal{R}(\Gamma)$ par :

$$s(m_1 \dots m_n) = s(m_1) \dots s(m_n), \quad s(\varepsilon) = \varepsilon$$

Montrer que $s(L)$ est reconnaissable.

Théorème 5 HP

Tout automate généralisé est équivalent à un automate fini classique.

Exercice 10 : Substitution régulière

Soit $L \subset \Sigma^*$ un langage reconnaissable, Γ un alphabet et $s : \Sigma \rightarrow \mathcal{R}(\Gamma)$ qu'on étend à $s : \Sigma^* \rightarrow \mathcal{R}(\Gamma)$ par :

$$s(m_1 \dots m_n) = s(m_1) \dots s(m_n), \quad s(\varepsilon) = \varepsilon$$

Montrer que $s(L)$ est reconnaissable.

Lemme 2 : Élimination d'un état

Soit $A = (\mathcal{R}(\Sigma), Q, I, F, \delta)$ un automate généralisé et $q \in Q \setminus (I \cup F)$.

Alors A est équivalent à l'automate $A' = (\mathcal{R}(\Sigma), Q \setminus \{q\}, I, F, \delta')$ où, pour tout $p, r \in Q \setminus \{q\}$:

$$\delta'(p, r) = \delta(p, r) \mid \delta(p, q)\delta(q, q)^*\delta(q, r)$$

Théorème 5 HP

Tout automate généralisé est équivalent à un automate fini classique.

Exercice 10 : Substitution régulière

Soit $L \subset \Sigma^*$ un langage reconnaissable, Γ un alphabet et $s : \Sigma \rightarrow \mathcal{R}(\Gamma)$ qu'on étend à $s : \Sigma^* \rightarrow \mathcal{R}(\Gamma)$ par :

$$s(m_1 \dots m_n) = s(m_1) \dots s(m_n), \quad s(\varepsilon) = \varepsilon$$

Montrer que $s(L)$ est reconnaissable.

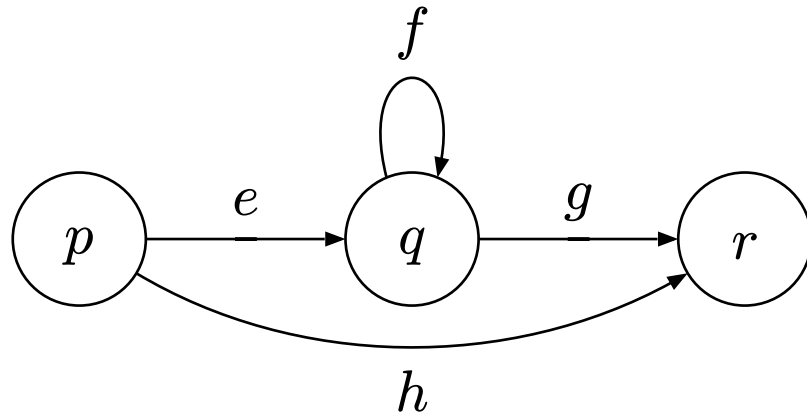
Lemme 2 : Élimination d'un état

Soit $A = (\mathcal{R}(\Sigma), Q, I, F, \delta)$ un automate généralisé et $q \in Q \setminus (I \cup F)$.

Alors A est équivalent à l'automate $A' = (\mathcal{R}(\Sigma), Q \setminus \{q\}, I, F, \delta')$ où, pour tout $p, r \in Q \setminus \{q\}$:

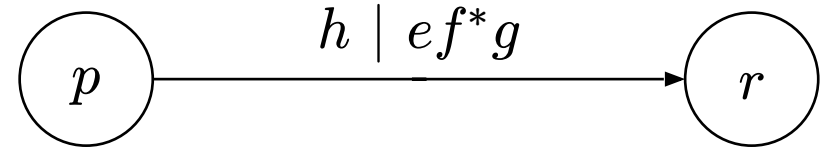
$$\delta'(p, r) = \delta(p, r) \mid \delta(p, q)\delta(q, q)^*\delta(q, r)$$

Reconnaissable $\implies$ régulier



Automate A

$\longrightarrow$



Automate A' après élimination de q

Reconnaissable $\implies$ régulier

Algorithme 2 : Élimination des états

Entrée : un automate normalisé $A = (\Sigma, Q, q_i, q_f, \delta)$.

Sortie : une expression régulière e telle que $\mathcal{L}(e) = \mathcal{L}(A)$.

Tant que $|Q| > 2$:

 Choisir un état $q \in Q \setminus \{q_i, q_f\}$.

 Éliminer q comme dans Lemme : Élimination d'un état.

Renvoyer l'étiquette de q_i à q_f , ou $\emptyset$ si pas de transition.

Reconnaissable $\implies$ régulier

Algorithme 2 : Élimination des états

Entrée : un automate normalisé $A = (\Sigma, Q, q_i, q_f, \delta)$.

Sortie : une expression régulière e telle que $\mathcal{L}(e) = \mathcal{L}(A)$.

Tant que $|Q| > 2$:

 Choisir un état $q \in Q \setminus \{q_i, q_f\}$.

 Éliminer q comme dans **Lemme : Élimination d'un état**.

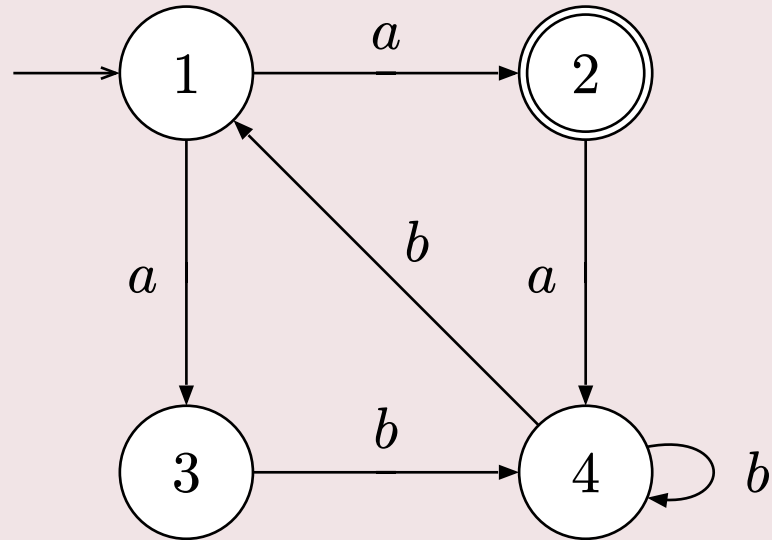
Renvoyer l'étiquette de q_i à q_f , ou $\emptyset$ si pas de transition.

Variant : $|Q|$ diminue à chaque itération.

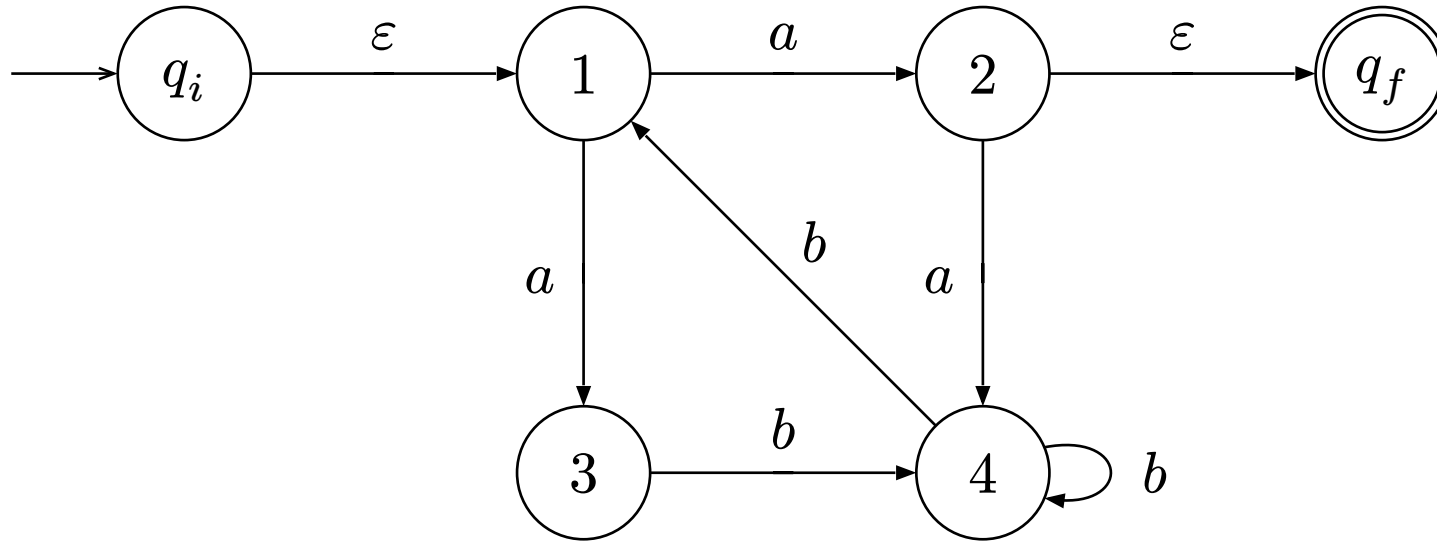
Invariant : l'automate reconnaît toujours $\mathcal{L}(A)$.

Exercice 11

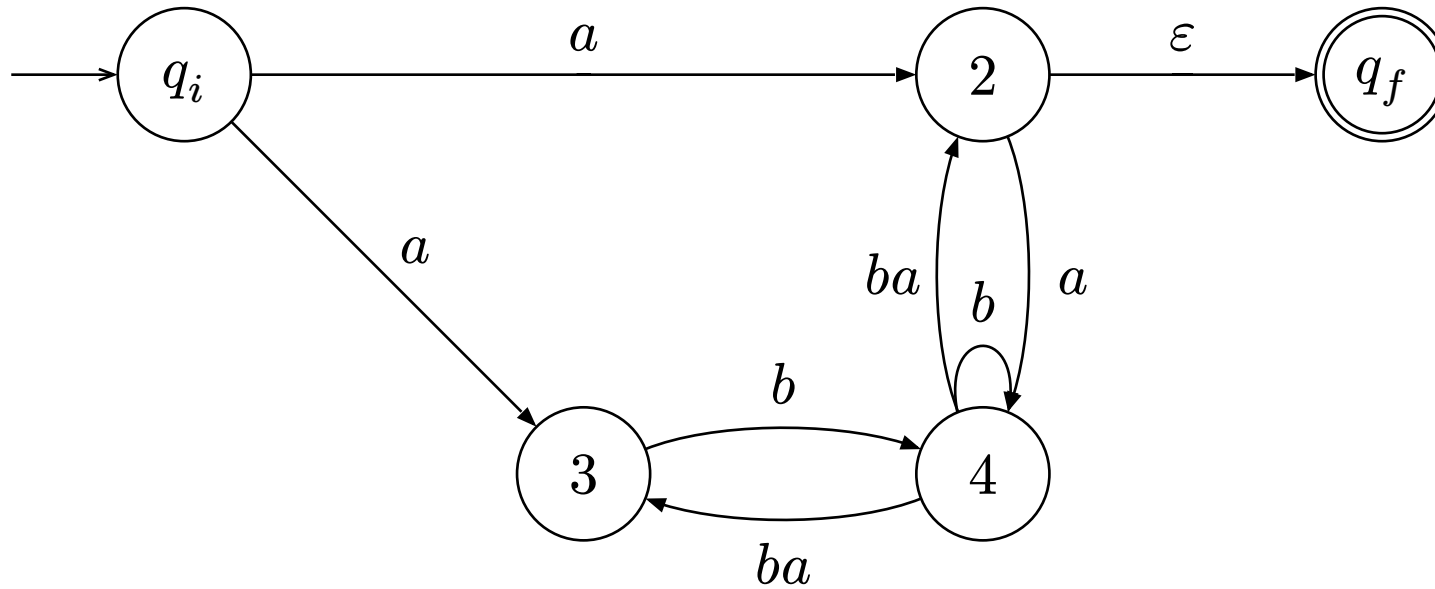
Appliquer l'algorithme d'élimination des états.



Normaliser l'exemple

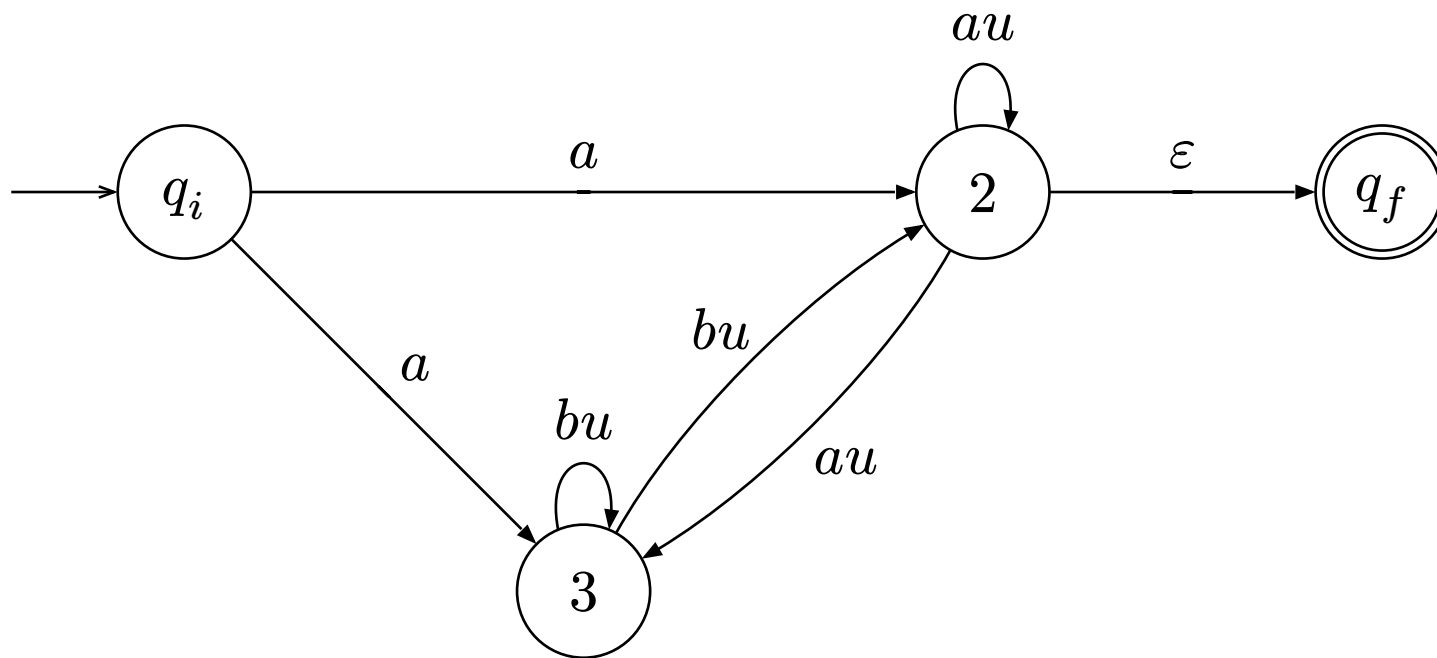


Supprimer l'état 1

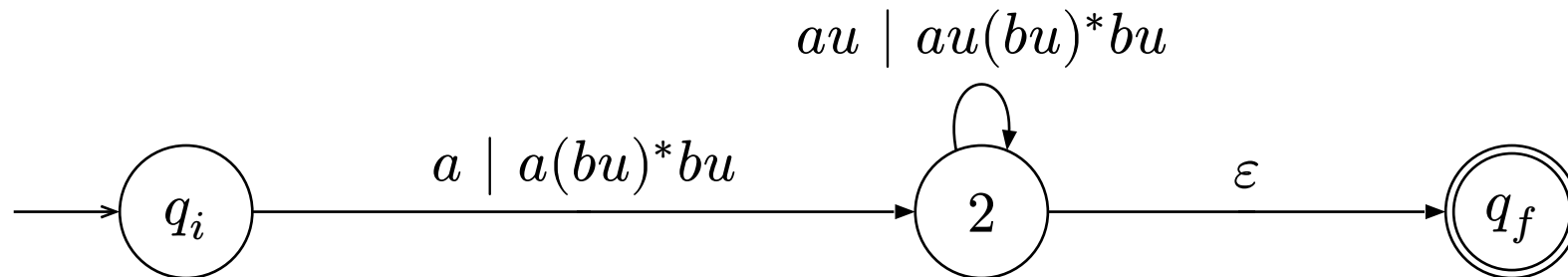


Supprimer l'état 4

Poser $u = b^*ba$.



Supprimer l'état 3



Supprimer l'état 2

On obtient, avec $u = b^*ba$:

$$(a \mid a(bu)^*bu)(au \mid au(bu)^*bu)^*.$$

Une expression équivalente plus courte est :

$$a(bu)^*(au(bu)^*)^*.$$

Exercice 12 : Analyse de Algorithme : Élimination des états

1. Quelle structure de données utiliser dans Algorithme : Élimination des états ?
2. Donner alors les complexités temporelle et spatiale, en fonction de $n = |Q|$.
3. Montrer que l'expression régulière renvoyée peut être de taille exponentielle en n . Est-ce en contradiction avec la question précédente ?

Théorème de Kleene

Théorème 6 : Théorème de Kleene ♡

Un langage est régulier si et seulement s'il est reconnaissable par un automate fini : $\mathcal{L}_{\text{rec}}(\Sigma) = \mathcal{L}_{\text{reg}}(\Sigma)$.

Théorème de Kleene

Corollaire 1 : Stabilités de $\mathcal{L}_{\text{rec}}$

Les langages réguliers sont stables par union finie, concaténation, étoile, intersection finie, complémentaire et différence.

Théorème de Kleene

Corollaire 1 : Stabilités de $\mathcal{L}_{\text{rec}}$

Les langages réguliers sont stables par union finie, concaténation, étoile, intersection finie, complémentaire et différence.

Attention : $\mathcal{L}_{\text{rec}}$ n'est pas stable par inclusion, ni union ou intersection infinie.