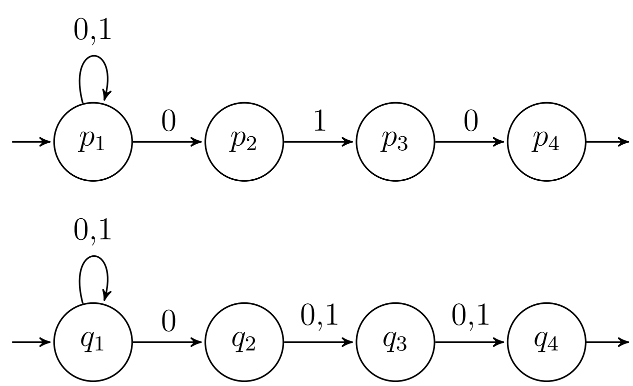
Langages Formels

David Baelde
<baelde@lsv.ens-cachan.fr>

Exercice 1 (Déterminisation)
Un automate fini non-déterministe sur l'alphabet Σ est donné par (Q, δ, I, F) où Q est un ensemble fini d'états, $\delta : Q \times \Sigma \to 2^Q$ est la fonction de transition, I est l'ensemble des états initiaux et F celui des états finaux. L'automate est déterministe si $\delta(q, a)$ est de cardinal au plus un pour tout $q \in Q$ et $a \in \Sigma$. Les notions de mot et langage reconnus sont standard. Deux automates sont dits équivalents s'ils acceptent le même langage.

1. Pour chacun des automates suivants, déterminer le langage reconnu et donner un automate déterministe équivalent :



2. Si $\Sigma = \{0, 1\}$, montrer que le langage $\Sigma^* 0 \Sigma^{n-1}$ ne peut être reconnu par un automate de moins de 2^n états.
3. Conclure que tout automate non-déterministe à n états admet un automate déterministe équivalent à au plus 2^n états, et que cette borne est précise.

Exercice 2 (Expressions rationnelles)
On définit inductivement les expressions rationnelles : $\underline{\emptyset}$, $\underline{\epsilon}$ et $\underline{a}$ (pour $a \in \Sigma$)

sont des expressions rationnelles ; si e_1 et e_2 sont des expressions rationnelles alors $e_1 \cdot e_2$, $e_1 + e_2$ et e_1^* en sont aussi.

Les expressions s'interprètent naturellement comme des langages. Un langage est dit rationnel s'il est l'interprétation d'une expression. On sait que tout langage rationnel est aussi régulier.

1. Montrer que l'appartenance du mot vide ϵ à un langage rationnel est calculable.
2. Étant donnés deux langages rationnels montrer que leur union, intersection et différence sont encore des langages rationnels.
3. Étant donné un langage rationnel L, montrer que l'ensemble des suffixes des mots de L est encore rationnel.
4. Étant donné un langage rationnel L non vide, montrer que l'on peut calculer le plus long préfixe commun à tous les mots de L.

Exercice 3 (Construction d'automates, ou pas)
Pour chacun des langages suivants sur $\Sigma = \{0, 1\}$, dire s'il est régulier (reconnaissable par un automate fini) ou pas. Justifier en donnant un automate ou un argument.

1. $\{ u \ : \ |u|_0 = |u|_1 \}$ où $|u|_a$ est le nombre d'occurrences de a dans u ;
2. $\{ 0^p 1^q \ : \ p \geq q \}$;
3. $\{ 0^p 1^q \ : \ p \geq q, \ q \leq 1337 \}$;
4. $\{ u \ : \ \overline{u}^2 = 0 \mod 3\}$ où $\overline{u}^2$ est l'interprétation de u en binaire, avec le bit de poids fort à droite.

Exercice 4 (Dérivées partielles, automate d'Antimirov)
La dérivée partielle est aux expressions rationnelles ce que le résiduel est aux langages : on va définir une opération $\partial_a(E)$ qui va correspondre (modulo interprétation) à $a^{-1}\mathcal{L}(E)$.

Formellement, pour une expression e et une lettre a, on définit l'*ensemble d'expressions* $\partial_a(e)$ comme suit, où l'opération de concaténation est naturellement étendue sur les ensembles d'expressions :

$$
\begin{array}{rcl}
\partial_a(\underline{\emptyset}) & = & \partial_a(\underline{\epsilon}) \;=\; \emptyset \\
\partial_a(\underline{b}) & = & \left\{ \begin{array}{ll} \emptyset & \text{si } a \neq b \\ \{\underline{\epsilon}\} & \end{array} \right. \\
\partial_a(e + e') & = & \partial_a(e) \cup \partial_a(e') \\
\partial_a(e^*) & = & \partial_a(e) \cdot e^* \\
\partial_a(e \cdot e') & = & \left\{ \begin{array}{ll} \partial_a(e) \cdot e' & \text{si } \epsilon \notin e \\ (\partial_a(e) \cdot e') \cup \partial_a(e') & \text{sinon} \end{array} \right.
\end{array}
$$

On définit ensuite $\partial_w(e)$ pour un mot w quelconque en posant $\partial_\epsilon(e) = \{e\}$ et $\partial_{wa}(e) = \partial_a(\partial_w(e))$, où $\partial_w(S) = \bigcup_{e \in S} \partial_w(e)$ quand S est un ensemble d'expressions.

1. Calculer les dérivées partielles de $(ab + b)^* ba$ par a et b.
2. Montrer que $\mathcal{L}(\partial_w(e)) = w^{-1}\mathcal{L}(e)$.
3. Utiliser cette construction pour traduire une expression rationnelle en un automate acceptant le même langage, en ne s'inquiétant pas de la finitude de cet automate pour l'instant.
4. On définit l'ensemble des suffixes propres d'un mot :

   $$\mathrm{Suf}(w) = \{\ v \in \Sigma^+ \ :\ \exists u, w = uv\ \}$$

   Montrer les résultats suivants pour $w \in \Sigma^+$:

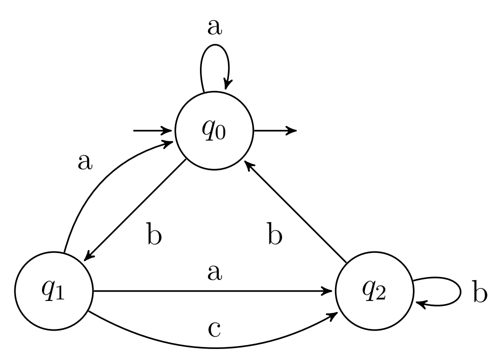
   $$
   \begin{array}{rcl}
   \partial_w(e + e') & = & \partial_w(e) \cup \partial_w(e') \\
   \partial_w(e \cdot e') & \subseteq & (\partial_w(e) \cdot e') \cup \displaystyle\bigcup_{v \in \mathrm{Suf}(w)} \partial_v(e') \\
   \partial_w(e^*) & \subseteq & \displaystyle\bigcup_{v \in \mathrm{Suf}(w)} \partial_v(e) \cdot e^*
   \end{array}
   $$

5. Soit $\|e\|$ le nombre d'occurrences de lettres de Σ dans e. Montrer que l'ensemble des dérivées partielles différentes de e contient au plus $\|e\| + 1$ éléments.

On pourra aussi se demander si cette méthode s'adapte pour produire directement un automate déterministe.

Exercice 5 (Algorithme de Brzozowski-McCluskey)
L'objectif de cet exercice est de traduire un automate fini en une expression rationnelle. Nous allons procéder par transformations successives de l'automate, en utilisant une généralisation du type d'automate considéré : la fonction de transition sera de type $Q \times 2^{\Sigma^*} \to 2^Q$. Une exécution d'un tel automate reconnait la concaténation des langages des transitions, et le langage reconnu par l'automate est l'union de ces exécutions.

1. Montrer que tout automate généralisé est équivalent à un automate généralisé pour lequel il existe exactement une transition entre chaque paire d'états : $q' \in \delta(q, L)$ et $q' \in \delta(q, L')$ implique $L = L'$.
2. Soit un automate généralisé $\mathcal{A}$ avec un unique état initial i, un unique état final f. Soit $q \in Q_{\mathcal{A}}$, $q \notin \{i, f\}$. Montrer qu'il existe un automate équivalent à $\mathcal{A}$ avec pour ensemble d'états $Q_{\mathcal{A}} \setminus \{q\}$.
3. En déduire que si L est reconnu par un automate généralisé $\mathcal{A}$, alors L appartient à la clôture rationnelle des étiquettes des transitions de $\mathcal{A}$.
4. Appliquer la construction pour calculer une expression rationnelle correspondant à l'automate suivant :



5. On considère l'alphabet $\Sigma_n = [1; n] \times [1; n]$, et on définit :

$$L = \{ \ (a_1, a_2)(a_2, a_3) \dots (a_m, a_{m+1}) \ : \ m \geq 1, \ a_1 = 1, \ a_{m+1} = n \ \}$$

   (a) Donner un automate de taille quadratique reconnaissant L.
   (b) Quelle est la taille de l'expression obtenue par la construction précédente sur cet automate ?