le cnam

Codes et Automates finis

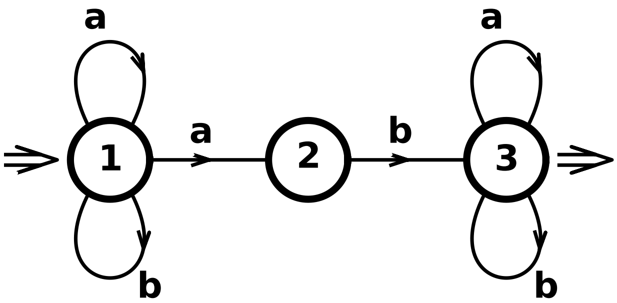
Cours 13 Automates non déterministes

- Définition

Un automate fini non déterministe $\mathscr{A}$ est la donnée d'un quintuple $\mathscr{A} = (Q, A, T, I, F)$. L'ensemble des états Q est fini, l'alphabet A est un ensemble fini, l'ensemble des états initiaux I est une partie de Q et l'ensemble F des états finals est aussi un sous-ensemble de Q : $I \subset Q$ et $F \subset Q$. L'ensemble T des transitions est défini maintenant de façon très souple : il regroupe tous les triplets (p, a, q) avec $p, q \in Q$ et $a \in A$ de sorte que $p \xrightarrow{a} q$ est une transition de l'automate. On a simplement $T \subset Q \times A \times Q$.
Un automate fini déterministe peut être vu comme un cas particulier d'automate fini non déterministe ! Dans le cas déterministe, pout tout $(p, a) \in Q \times Q$, il existe un unique $q = \delta(p, a)$ de sorte que $(p, a, \delta(p, a)) \in T$. On a une dépendance fonctionnelle.
Pour un automate non déterministe, on peut avoir plusieurs transitions $p \xrightarrow{a} q_1$ et $p \xrightarrow{a} q_2$ à partir d'un même couple (p, a) composé d'un état $p \in Q$ et d'une lettre de l'alphabet $a \in A$. On peut aussi ne pas avoir de transition à partir d'un état $p \in Q$ et d'une lettre a de l'alphabet.

- Exemple 1

On considère l'automate fini non déterministe décrit à la figure ci-dessous. Cet automate n'est pas déterministe puisque deux transitions sont issues de l'état 1 à l'aide de la lettre "a". De plus, aucune transition ne part de l'état 2 à l'aide de cette même lettre "a".



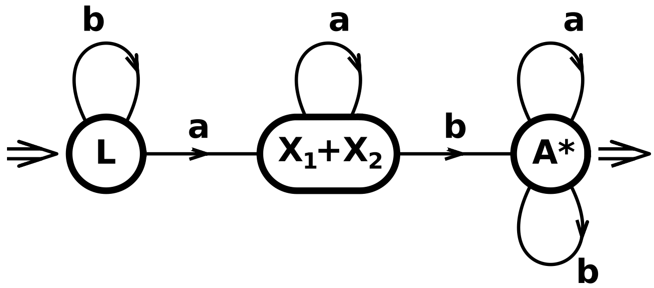
- Langages de départ

Comme dans le cas des automates déterministes, nous introduisons les langages de départ X_p des mots qui pilotent un chemin de l'automate partant de l'état fixé $p \in Q$ et allant jusqu'à un état final $f \in F$.
Le raisonnement fait pour le cas des automates finis déterministes est inchangé et on a encore l'ensemble de relations $X_p = \sum_{a \in A, q \in Q, (p,a,q) \in T} a.X_q$ si $p \notin F$ et
$X_p = \sum_{a \in A, q \in Q, (p,a,q) \in T} a.X_q + \varepsilon$ si $p \in F$.

François Dubois, mai 2024.

• Langages de départ pour l'exemple 1

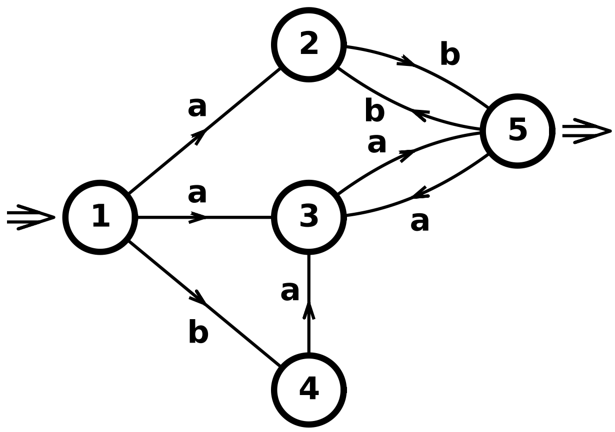
On peut donc écrire l'ensemble des équations pour les langages de départ de l'automate décrit à la figure ci-dessus. Il vient $X_1 = (a+b)X_1 + aX_2$, $X_2 = bX_3$, $X_3 = (a+b)X_3 + \varepsilon$.
La résolution de ce système d'équations linéaires n'offre pas de difficulté : on a grâce au lemme d'Arden $X_3 = A^*$, donc $X_2 = bA^*$. Puis à nouveau grâce au lemme d'Arden :
$X_1 = A^* aX_2 = A^* abA^*$. Et X_1 est exactement le langage L accepté par l'automate décrit à la page précédente.
La remarque importante est que ce langage est rationnel ; il s'obtient à l'aide d'un nombre fini d'opérations de réunions, concaténations et d'actions de l'opérateur "étoile". On sait alors qu'il possède un nombre fini de quotients à gauche que l'on peut déterminer sans difficulté. On a $L = X_1$ et $X_1 = a(X_1+X_2) + bX_1$, donc $a^{-1}L = X_1 + X_2$; $b^{-1}L = X_1 = L$. On a ensuite $a^{-1}(X_1+X_2) = a^{-1}X_1 + a^{-1}X_2 = X_1 + X_2 + \emptyset = X_1 + X_2$. Puis $b^{-1}(X_1+X_2) = b^{-1}X_1 + b^{-1}X_2 = X_1 + X_3$. Or $X_3 = A^*$ dans ce cas précis. Donc $X_1 + X_3 = A^*$ et $b^{-1}X_3 = A^*$. Les quotients à gauche du langage L sont les suivants : $L = X_1$, $X_1 + X_2$, $X_3 = A^*$. On vérifie facilement que ces quotients à gauche forment des langages tous différents.
L'automate minimal peut se construire graphiquement à partir de l'ensemble des relations entre les quotients à gauche, ainsi que proposé à la figure ci-dessous.



À l'aide de ce procédé, nous avons construit un automate déterministe $\mathscr{A}_d$ qui a le même langage des mots acceptés que l'automate non déterministe $\mathscr{A}$ initial : $\mathscr{L}(\mathscr{A}_d) = \mathscr{L}(\mathscr{A})$.

• Exemple 2

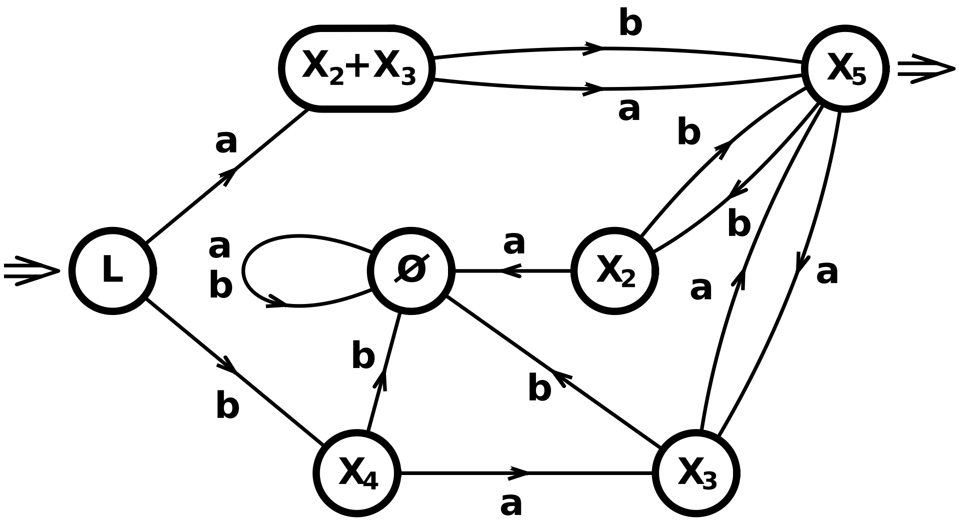
Cet exemple est décrit à la figure ci-dessous.



Cet automate fini n'est pas déterministe. D'une part, deux transitions avec le label "a" sont issues de l'état ①. D'autre part, il n'existe pas de transition avec le label "a" issue de l'état ② et il n'existe pas de transition avec le label "b" issue des états ③ et ④.
Nous allons appliquer la méthodologie suivie pour l'exemple précédent : écriture des équations de départ, calcul du langage rationnel $L = \mathscr{L}(\mathscr{A})$ des mots acceptés par l'automate, détermination des quotients à gauche de ce langage L et enfin explicitation de l'automate déterministe minimal $\mathscr{A}_d = \mathscr{A}(L)$ qui permet de transformer l'automate non déterministe décrit à la figure ci-dessus en un automate fini équivalent puisque $\mathscr{A}$ et $\mathscr{A}_d$ ont même langage L des mots acceptés.
Les cinq équations de départ s'écrivent sans difficulté : $X_1 = a(X_2 + X_3) + bX_4,\ X_2 = bX_5,$ $X_3 = aX_5,\ X_4 = aX_3,\ X_5 = aX_3 + bX_2 + \varepsilon$. On a $X_4 = aX_3 = a^2X_5$; on reporte les valeurs de X_2 et X_3 dans la dernière équation. Il vient $X_5 = a(aX_5) + b(bX_5) + \varepsilon = (a^2 + b^2)X_5 + \varepsilon$. Grâce au lemme d'Arden, $X_5 = (a^2 + b^2)^*$. On utilise les valeurs de X_2, X_3 et X_4 pour calculer le langage X_1 : $X_1 = a(bX_5) + a(aX_5) + b(a^2X_5) = (ab + a^2 + ba^2)X_5$ et
$L = X_1 = (ab + a^2 + ba^2)(a^2 + b^2)^*$.
Les quotients à gauche de ce langage L se déterminent en utilisant les équations qui précèdent : $a^{-1}L = a^{-1}X_1 = X_2 + X_3,\ b^{-1}L = b^{-1}X_1 = X_4,\ a^{-1}(X_2 + X_3) = a^{-1}X_2 + a^{-1}X_3 = \emptyset + X_5 = X_5,$ $b^{-1}(X_2 + X_3) = b^{-1}X_2 + b^{-1}X_3 = X_5 + \emptyset = X_5,\ a^{-1}X_4 = X_3,\ b^{-1}X_4 = \emptyset,\ a^{-1}X_5 = X_3,$
$b^{-1}X_5 = X_2$. Enfin, $a^{-1}X_3 = X_5$, $b^{-1}X_3 = \emptyset$ et $a^{-1}X_2 = \emptyset$, $b^{-1}X_2 = X_5$. On dispose donc d'un total de sept quotients à gauche : $L,\ X_2 + X_3,\ X_4,\ X_5,\ X_3,\ \emptyset,\ X_2$. On vérifie que ces langages sont tous différents en observant les mots de faible longueur : $L = ab + a^2 + ba^2 + aba^2 + \ldots,$ $X_5 = \varepsilon + a^2 + b^2 + a^2a^2 + a^2b^2 + b^2a^2 + b^2b^2 + \ldots,\ X_2 = b + ba^2 + b^3 + \ldots$ et des mots d'au moins cinq lettres, $X_3 = a + a^3 + ab^2 + \ldots$ et des mots d'au moins cinq lettres,
$X_2 + X_3 = a + b + ba^2 + b^3 + a^3 + ab^2 + \ldots,\ X_4 = a^2 + a^4 + a^2b^2 + a^6 + a^4b^2 + a^2b^2a^2 + a^2b^4 + \ldots$. Le langage vide ne contient aucun mot et est bien sûr différent des six autres langages qui en contiennent. Seuls les langages X_5 et L contiennent des mots de deux lettres. Mais X_5 contient le mot sans lettre, alors que ce n'est pas le cas pour L. Donc ces deux langages sont différents et différents des quatre autres. Le langage X_4 ne comporte pas de mot de trois lettres, alors que c'est le cas pour X_2, X_3 et $X_2 + X_3$. On a donc $X_2 \neq X_3$ car $b \in X_2$ et $b \notin X_3$ d'une part, $a \in X_3$ et $a \notin X_2$ d'autre part. Donc $X_2 \neq X_3$ et $X_2 + X_3 = X_2 \cup X_3$ est nécessairement différent de X_2 et de X_3. Les sept quotients à gauche sont différents.
L'explicitation de l'automate minimal demande de vérifier que pour $m \in L$, on a bien
$m^{-1}L \in X_5$. Donc le langage X_5 est un état de sortie. Ce n'est pas le cas des autres langages.
L'automate "déterminisé" de l'automate non déterministe initial, présenté à la figure ci-dessus, comporte plus d'états (ici 7 au lieu de 5), ce qui n'était pas le cas pour le premier exemple. La situation exposée pour ces deux cas particuliers est en fait générale et on a les deux théorèmes suivants.

- Théorème (Kleene)

Le langage $\mathscr{L}(\mathscr{A})$ des mots acceptés par un automate fini non déterministe $\mathscr{A}$ est un langage rationnel. En conséquence, il a un nombre fini de quotients à gauche.



• Théorème de déterminisation

Pour tout automate fini non déterministe $\mathscr{A}$, il existe un automate fini déterministe non unique $\mathscr{A}_d$ qui admet le même langage des mots acceptés : $\mathscr{L}(\mathscr{A}) = \mathscr{L}(\mathscr{A}_d)$.

• Application

Dans le système d'exploitation Unix, la commande "grep" permet la recherche d'une occurrence dans une expression rationnelle. Cette commande construit d'abord un automate non déterministe pour prendre en compte l'ordre donné puis cet automate est "déterminisé" afin de donner lieu à un algorithme non ambigu.

• Construction théorique de l'automate déterminisé

Soit $\mathscr{A} = (Q, A, T, I, F)$ un automate fini non déterministe ; les transitions $p \xrightarrow{a} q$ de l'automate sont décrites par un sous-ensemble T du produit cartésien $Q \times A \times Q$. L'automate $\widetilde{\mathscr{A}} = (\widetilde{Q}, \widetilde{A}, \widetilde{T}, \widetilde{I}, \widetilde{F})$ décrit ci-dessous admet le même langage des mots acceptés que l'automate initial $\mathscr{A}$. De plus, il est déterministe. On a $\widetilde{Q} = \mathscr{P}(Q)$, ensemble des parties de l'ensemble Q, $\widetilde{A} = A$, $\widetilde{I} = \{I\}$, $\widetilde{F} = \{P \in \widetilde{Q}, P \cap F \neq \emptyset\}$. Pour $P \in \widetilde{Q}$ et $a \in A$, on peut construire $R = \delta(P, a)$ de sorte que $(q \in R) \Longleftrightarrow (\exists\, p \in P, (p, a, q) \in T)$.

La construction précédente a l'avantage de donner une borne sur le nombre d'états de l'automate déterministe équivalent. Mais cette borne est exponentielle. En effet, si $\sharp Q$ est le nombre d'états de l'automate $\mathscr{A}$, [on a $\sharp Q = 5$ pour le second exemple], le choix proposé $\widetilde{Q} = \mathscr{P}(Q)$ montre que $\sharp \widetilde{Q} = 2^{\sharp Q}$, soit $2^5 = 32$ (!) pour l'exemple 2. En pratique, des algorithmes plus efficaces peuvent être mis en œuvre.

Exercices

• Un automate fini non déterministe

On s'intéresse à l'automate non déterministe décrit à la figure ci-dessous.

a) Expliquer pourquoi cet automate fini n'est pas déterministe.

b) Préciser les équations de départ de cet automate.

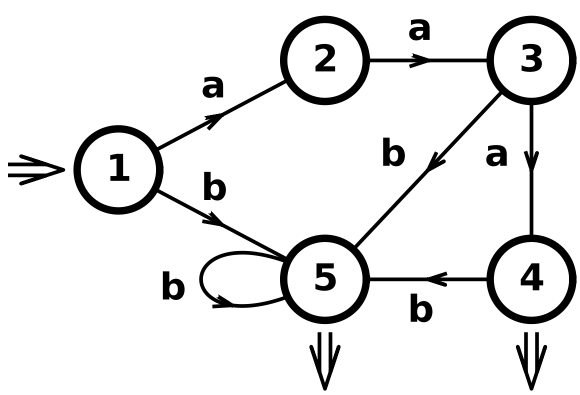
c) Déterminer le langage L des mots acceptés.

d) Quels sont les quotients à gauche du langage L ?

e) À l'aide des mots de faible longueur des divers quotients à gauche du langage L, montrer que ces quotients sont tous différents.

f) Construire l'automate minimal déterministe associé au langage L.

g) Vérifier que les mots acceptés par l'automate trouvé à la question f) sont exactement ceux du langage L.



• Un autre automate fini non déterministe

Reprendre les questions a) à g) de l'exercice précédent avec l'automate décrit à la figure ci-dessous.