Langages formels

Paul Gastin

Paul.Gastin@lsv.fr

http://www.lsv.fr/~gastin/Langages/

L3 Informatique Cachan
2019-2020

Plan

Motivations

Définition :

1. Description et analyse (lexicale et syntaxique) des langages (programmation, naturels, ...)
2. Modèles de calcul
3. Abstractions mathématiques simples de phénomènes complexes dans le but de
   - Prouver des propriétés.
   - Concevoir des algorithmes permettant de tester des propriétés ou de résoudre des problèmes.
4. Types de données

Références

[7] Olivier Carton.
Langages formels, calculabilité et complexité.
Vuibert, 2008.

[9] John E. Hopcroft et Jeffrey D. Ullman.
Introduction to automata theory, languages and computation.
Addison-Wesley, 1979.

[10] Dexter C. Kozen.
Automata and Computability.
Springer, 1997.

[13] Jacques Sakarovitch.
Éléments de théorie des automates.
Vuibert informatique, 2003.

[8] Hubert Comon, Max Dauchet, Remi Gilleron, Florent Jacquemard, Denis Lugiez, Sophie Tison, Marc Tommasi.
Tree Automata Techniques and Applications.
http://www.grappa.univ-lille3.fr/tata/

Références

[1] Alfred V. Aho, Ravi Sethi et Jeffrey D. Ullman.
Compilers: principles, techniques and tools.
Addison-Wesley, 1986.

[2] Alfred V. Aho et Jeffrey D. Ullman.
The theory of parsing, translation, and compiling. Volume I: Parsing.
Prentice-Hall, 1972.

[3] Luc Albert, Paul Gastin, Bruno Petazzoni, Antoine Petit, Nicolas Puech et Pascal Weil.
Cours et exercices d'informatique.
Vuibert, 1998.

[4] Jean-Michel Autebert.
Théorie des langages et des automates.
Masson, 1994.

Références

[5] Jean-Michel Autebert, Jean Berstel et Luc Boasson.
Context-Free Languages and Pushdown Automata.
Handbook of Formal Languages, Vol. 1, Springer, 1997.

[6] Jean Berstel.
Transduction and context free languages.
Teubner, 1979.

[11] Jean-Éric Pin.
Automates finis et applications.
Polycopié du cours à l'École Polytechnique, 2004.

[12] Grzegorz Rozenberg et Arto Salomaa, éditeurs.
Handbook of Formal Languages,
Vol. 1, Word, Language, Grammar,
Springer, 1997.

[14] Jacques Stern.
Fondements mathématiques de l'informatique.
Mc Graw Hill, 1990.

Plan

Plan

Grammaires de type 0

Définition : Grammaires générales (type 0)

$G = (\Sigma, V, P, S)$ où

- Σ est l'alphabet terminal
- V est l'alphabet non terminal (variables)
- $S \in V$ est l'axiome (variable initiale)
- $P \subseteq (\Sigma \cup V)^* \times (\Sigma \cup V)^*$ est un ensemble **fini** de règles ou productions.

Exemple : Grammaire G_1 pour $\{a^{2^n} \mid n > 0\}$

$$1 : \quad S \to DXaF \qquad 3 : \; XF \to YF \qquad 5 : \; DY \to DX \qquad 7 : \; aZ \to Za$$
$$2 : \; Xa \to aaX \qquad 4 : \; aY \to Ya \qquad 6 : \; XF \to Z \qquad 8 : \; DZ \to \varepsilon$$

Définition : Dérivation

$\alpha \in (\Sigma \cup V)^*$ se dérive en $\beta \in (\Sigma \cup V)^*$, noté $\alpha \to \beta$, s'il existe $(\alpha_2, \beta_2) \in P$ tel que $\alpha = \alpha_1\alpha_2\alpha_3$ et $\beta = \alpha_1\beta_2\alpha_3$.
On note $\overset{*}{\to}$ la clôture réflexive et transitive de $\to$.

Grammaires de type 0

Définition : Langage engendré

Soit $G = (\Sigma, V, P, S)$ une grammaire et $\alpha \in (\Sigma \cup V)^*$.
Le langage engendré par α est $\mathcal{L}_G(\alpha) = \{u \in \Sigma^* \mid \alpha \xrightarrow{*} u\}$.
Le langage *élargi* engendré par α est $\widehat{\mathcal{L}}_G(\alpha) = \{\beta \in (\Sigma \cup V)^* \mid \alpha \xrightarrow{*} \beta\}$.
Le langage engendré par G est $L_G(S)$.
Un langage est *de type 0* s'il peut être engendré par une grammaire de type 0.

Théorème : Chomsky 1959, voir [9, Thm 9.3 & 9.4]

Un langage $L \subseteq \Sigma^*$ est de type 0 ssi il est récursivement énumérable.

Grammaires contextuelles

Définition : Grammaire contextuelle (type 1, context-sensitive)

Une grammaire $G = (\Sigma, V, P, S)$ est *contextuelle* si toute règle $(\alpha, \beta) \in P$ vérifie $|\alpha| \leq |\beta|$.
Un langage est de type 1 (ou contextuel) s'il peut être engendré par une grammaire contextuelle.

Les règles 6 et 8 de la grammaire G_1 ne sont pas 'non-décroissantes'.

Exemple : Grammaire G_2 contextuelle pour $\{a^{2^n} \mid n > 0\}$

$1 : S \to DTF$	$3 : T \to XT$	$5 : Xaa \to aaXa$	$7 : Daaa \to aaDaa$
$2 : S \to aa$	$4 : T \to aa$	$6 : XaF \to aaF$	$8 : DaaF \to aaaa$

Remarque :

Le langage engendré par une grammaire contextuelle est propre.
Si on veut engendrer le mot vide on peut ajouter $\hat{S} \to S + \varepsilon$.

Grammaires contextuelles

Définition : Forme normale (context-sensitive/contextuelle)

Une grammaire $G = (\Sigma, V, P, S)$ contextuelle est en forme normale si toute règle est de la forme $(\alpha_1 X \alpha_2, \alpha_1 \beta \alpha_2)$ avec $X \in V$ et $\beta \neq \varepsilon$.

Les règles 5, 7 et 8 de la grammaire G_2 ne sont pas en forme normale.

Théorème : Forme normale [4, Prop. 2, p. 156]

Tout langage de type 1 est engendré par une grammaire contextuelle en forme normale.

Exemple : Une grammaire contextuelle en FN pour $\{a^{2^n} \mid n > 0\}$

$1 : S \to aTa$	$3 : T \to XT$	$5 : XA \to XY$	$8 : Xa \to AAa$
$2 : S \to aa$	$4 : T \to AA$	$6 : XY \to ZY$	$9 : ZA \to AAA$
		$7 : ZY \to ZX$	$10 : aA \to aa$

Grammaires contextuelles

Théorème : Kuroda 1964, voir [9, Thm 9.5 & 9.6]

Un langage est de type 1 ssi il est accepté par une machine de Turing non déterministe en espace linéaire.
Les langages contextuels sont strictement inclus dans les langages récursifs.

Théorème : Problème du mot (Kuroda 1964 pour déterministe)

Étant donnés un mot w et une grammaire G, décider si $w \in L_G(S)$.
Le problème du mot est PSPACE-complet pour les grammaires de type 1.

Théorème : indécidabilité du vide

On ne peut pas décider si une grammaire contextuelle engendre un langage vide.

Exercices :

1. Montrer que $\{a^{n^2} \mid n > 0\}$ est contextuel.
2. Montrer que $\{ww \mid w \in \{a, b\}^+\}$ est contextuel.

Grammaires algébriques

Définition : Grammaire hors contexte ou algébrique ou de type 2

Une grammaire $G = (\Sigma, V, P, S)$ est *hors contexte ou algébrique* si $P \subseteq V \times (\Sigma \cup V)^*$ (sous ensemble *fini*).

Un langage est de type 2 (ou hors contexte ou algébrique) s'il peut être engendré par une grammaire hors contexte.

On note Alg la famille des langages algébriques.

Exemples :

1. Le langage $\{a^n b^n \mid n \geq 0\}$ est algébrique.
2. Expressions complètement parenthésées.

Lemme : fondamental

Soit $G = (\Sigma, V, P, S)$ une grammaire algébrique, $\alpha_1, \alpha_2, \beta \in (\Sigma \cup V)^*$ et $n \geq 0$.

$$\alpha_1\alpha_2 \xrightarrow{n} \beta \quad \Longleftrightarrow \quad \alpha_1 \xrightarrow{n_1} \beta_1, \alpha_2 \xrightarrow{n_2} \beta_2 \text{ avec } \beta = \beta_1\beta_2 \text{ et } n = n_1 + n_2$$

Langages de Dyck

Définition : D_n^*

Soit $\Sigma_n = \{a_1, \ldots, a_n\} \cup \{\bar{a}_1, \ldots, \bar{a}_n\}$ l'alphabet formé de n paires de parenthèses.
Soit $G_n = (\Sigma_n, V, P_n, S)$ la grammaire définie par $S \to a_1 S \bar{a}_1 S + \cdots + a_n S \bar{a}_n S + \varepsilon$.
Le langage $D_n^* = \mathcal{L}_{G_n}(S)$ est appelé langage de Dyck sur n paires de parenthèses

Exercices : Langages de Dyck

1. Montrer que
   $D_1^* = \{w \in \Sigma_1^* \mid |w|_{a_1} = |w|_{\bar{a}_1}$ et $|v|_{a_1} \geq |v|_{\bar{a}_1}$ pour tous $v \leq w\}$.
2. On considère le système de réécriture (type 0) $R_n = (\Sigma_n, P'_n)$ dont les règles sont $P'_n = \{(a_i \bar{a}_i, \varepsilon) \mid 1 \leq i \leq n\}$.
   Montrer que $D_n^* = \{w \in \Sigma_n^* \mid w \xrightarrow{*} \varepsilon$ dans $R_n\}$.
3. Soit Γ un alphabet disjoint de Σ_n, $\Sigma = \Sigma_n \cup \Gamma$ et $L \subseteq \Sigma^*$ un langage.

   On définit la clôture $\mathrm{clot}(L) = \{v \in \Sigma^* \mid \exists w \in L, w \xrightarrow{*} v$ dans $R_n\}$.
   Montrer que si L est reconnaissable, alors $\mathrm{clot}(L)$ aussi.

   On définit la réduction $\mathrm{red}(L) = \{v \in \mathrm{clot}(L) \mid v \not\to$ dans $R_n\}$.
   Montrer que si L est reconnaissable, alors $\mathrm{red}(L)$ aussi.

Grammaires linéaires

Définition : Grammaire linéaire

La grammaire $G = (\Sigma, V, P, S)$ est

- linéaire si $P \subseteq V \times (\Sigma^* \cup \Sigma^* V \Sigma^*)$,
- linéaire gauche si $P \subseteq V \times (\Sigma^* \cup V\Sigma^*)$,
- linéaire droite si $P \subseteq V \times (\Sigma^* \cup \Sigma^* V)$.

Un langage est linéaire s'il peut être engendré par une grammaire linéaire.

On note Lin la famille des langages linéaires.

Exemples :

- Le langage $\{a^n b^n \mid n \geq 0\}$ est linéaire.
- Le langage $\{a^n b^n c^p \mid n, p \geq 0\}$ est linéaire.

Proposition :

Un langage est rationnel si et seulement si il peut être engendré par une grammaire linéaire gauche (ou droite).

Hiérarchie de Chomsky

Théorème : Chomsky

1. Les langages réguliers (type 3) sont strictement contenus dans les langages linéaires.
2. Les langages linéaires sont strictement contenus dans les langages algébriques (type 2).
3. Les langages algébriques propres (type 2) sont strictement contenus dans les langages contextuels (type 1).
4. les langages contextuels (type 1) sont strictement contenus dans les langages récursifs.
5. les langages récursifs sont strictement contenus dans les langages récursivement énumérables (type 0).

Bibliographie

[4] Jean-Michel Autebert.
Théorie des langages et des automates.
Masson, 1994.

[5] Jean-Michel Autebert, Jean Berstel et Luc Boasson.
Context-Free Languages and Pushdown Automata.
Handbook of Formal Languages, Vol. 1, Springer, 1997.

[7] Olivier Carton.
Langages formels, calculabilité et complexité.
Vuibert, 2008.

[9] John E. Hopcroft et Jeffrey D. Ullman.
Introduction to automata theory, languages and computation.
Addison-Wesley, 1979.

[10] Dexter C. Kozen.
Automata and Computability.
Springer, 1997.

[14] Jacques Stern.
Fondements mathématiques de l'informatique.
Mc Graw Hill, 1990.

Plan

Arbres de dérivation

Définition :

Soit $G = (\Sigma, V, P, S)$ une grammaire.

Un arbre de dérivation pour G est un arbre t étiqueté dans $V \cup \Sigma \cup \{\varepsilon\}$ tel que chaque nœud interne u est étiqueté par une variable $x \in V$ et si les fils de u portent les étiquettes $\alpha_1, \ldots, \alpha_k$ alors $(x, \alpha_1 \cdots \alpha_k) \in P$.

De plus, si $k \neq 1$, on peut supposer $\alpha_1, \ldots, \alpha_k \neq \varepsilon$.

Exemple :

Arbres de dérivation pour les expressions.
Mise en évidence des priorités ou de l'associativité G ou D.

Définition : Ambigüité

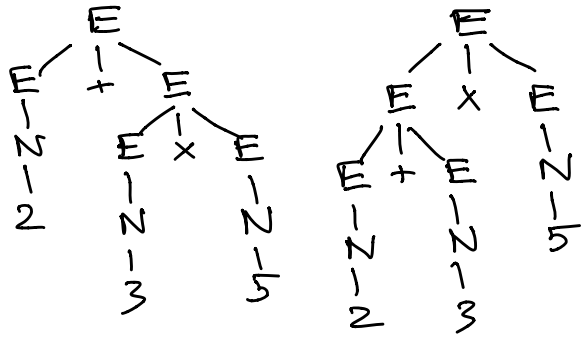
- Une grammaire est ambigüe s'il existe deux arbres de dérivations (distincts) de même racine et de même frontière.
- Un langage algébrique est *non ambigu* s'il existe une grammaire non ambigüe qui l'engendre. Dans le cas contraire, on dit qu'il est *inhéremment ambigu*.

Considérons la grammaire G_1

$E \to E+E \mid E \times E \mid (E) \mid N$

$N \to 0 \mid 1 \mid 2 \mid \cdots \mid 9$

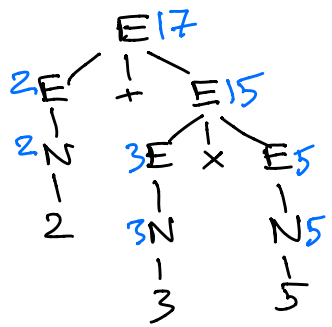
Le mot $w = 2+3\times5$ admet deux arbres de dérivation T_1 et T_2

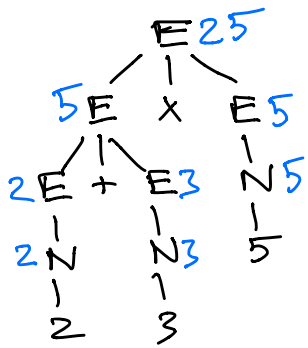


La grammaire G_1 est donc ambigüe.

L'arbre de dérivation est utile après l'analyse syntaxique pour la génération du code.

Par exemple pour l'évaluation de l'expression.





→ La grammaire doit être non-ambigüe et générer le bon arbre de dérivation.

Grammaire G_2

$E \to E+T \mid T$

$T \to T \times F \mid F$

$F \to N \mid (E)$

Grammaire G_3

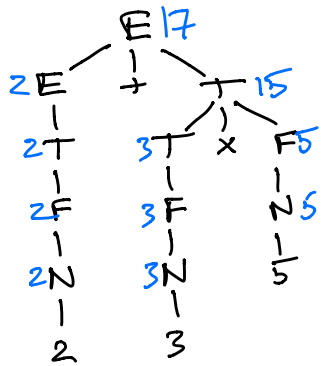
$E \to T+E \mid T$

$T \to F \times T \mid F$

$F \to N \mid (E)$

Ces grammaires sont non ambigües

L'arbre de dérivation correspond à la priorité de $\times$ sur $+$



Si on commence la dérivation par $E \to T$ alors pour engendrer le symbole $+$ il faudra "remonter" à E avec $F \to (E)$ et il y aura des parenthèses.

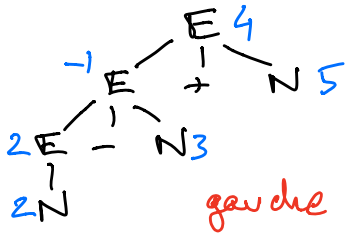
La différence entre les grammaires G_2 et G_3 se situe au niveau de l'associativité des opérateurs qui sont à un même niveau de priorité.

Exemple : analyser le mot $2-3+5$

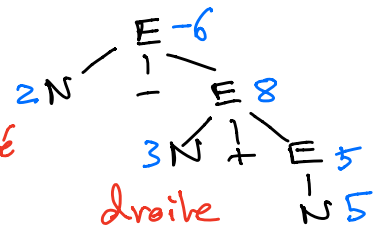
avec G_2 : ou G_3

$E \to E+N \mid E-N \mid N$

$E \to N+E \mid N-E \mid N$



associativité gauche



associativité droite

Arbres de dérivation

Lemme : Dérivations et arbres de dérivation

Soit $G = (\Sigma, V, P, S)$ une grammaire.

1. Si $x \xrightarrow{*} \alpha$ une dérivation de G alors il existe un arbre de dérivation t de G tel que $\mathrm{rac}(t) = x$ et $\mathrm{Fr}(t) = \alpha$.
2. Si t un arbre de dérivation de G alors il existe une dérivation $\mathrm{rac}(t) \xrightarrow{*} \mathrm{Fr}(t)$ dans G.
   Si $\mathrm{Fr}(t) \in \Sigma^*$ alors on peut faire une dérivation gauche $\mathrm{rac}(t) \xrightarrow{*}_g \mathrm{Fr}(t)$.

Une dérivation est *gauche* si on dérive toujours le non terminal le plus à gauche.

Remarques :

- 2 dérivations sont équivalentes si elles sont associées au même arbre de dérivation.
- Il y a bijection entre dérivations gauches terminales et arbres de dérivation ayant une frontière dans Σ^*.
- Si la grammaire est linéaire, il y a bijection entre dérivations et arbres de dérivations.

Preuve du Lemme : Arbres de dérivation

1. Récurrence sur n longueur de dérivation

$x \xrightarrow{n} \alpha \quad x \in V \quad \alpha \in (\Sigma \cup V)^*$

$\underline{n=0}$ $x = \alpha$: $t = x$ seulement la racine x

$\underline{n>0}$ $x \xrightarrow{1} \varepsilon = \alpha$ $t = x$ | ε

ou $x \xrightarrow{1} \beta_1 \cdots \beta_k \xrightarrow{n-1} \alpha$

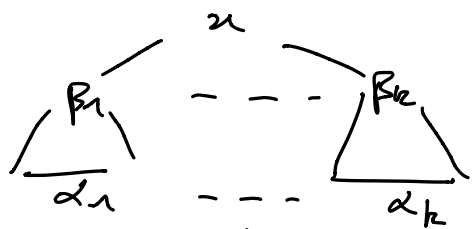
avec $\beta_i \in \Sigma \cup V$.

Lemme fondamental : $\alpha = \alpha_1 \cdots \alpha_k \quad \beta_i \xrightarrow{n_i} \alpha_i$

$n-1 = n_1 + \cdots + n_k$

HR arbre t_i avec $rac(t_i) = \beta_i \quad Fr(t_i) = \alpha_i$

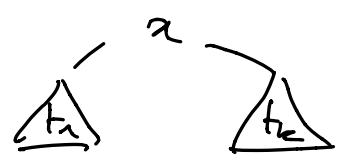
alors $t = x(t_1, \cdots, t_k)$



2. Récurrence sur $h(t)$ hauteur de l'arbre.

- $h(t) = 0$ alors $t = x \quad x \xrightarrow{0} x$
- $h(t) > 0$ alors $t = x(t_1, \cdots, t_k)$

$t_1, \cdots, t_k$ sous arbres de la racine



HR On a des dérivations

$\beta_i = rac(t_i) \xrightarrow{*} Fr(t_i) = \alpha_i \quad \forall\, 1 \le i \le k$

et $x \to \beta_1 \cdots \beta_k$ G P et $\alpha = \alpha_1 \cdots \alpha_k$

Donc $x \to \beta_1 \cdots \beta_k \xrightarrow{*} \alpha_1 \cdots \alpha_k = \alpha$.

Ambigüité

Exemples :

- La grammaire $S \to SS + aSb + \varepsilon$ est ambigüe mais elle engendre un langage non ambigu.
- La grammaire $E \to E + E \mid E \times E \mid a \mid b \mid c$ est ambigüe et engendre un langage rationnel.

Proposition : Tout langage rationnel peut être engendré par une grammaire linéaire droite non ambigüe.

Exercice : if then else

Montrer que la grammaire suivante est ambigüe.

$$S \to \texttt{if}\ c\ \texttt{then}\ S\ \texttt{else}\ S \mid \texttt{if}\ c\ \texttt{then}\ S \mid a$$

Montrer que le langage engendré n'est pas ambigu.

Grammaires et automates d'arbres

Théorème : (admis)

1. Soit L un langage d'arbres reconnaissable.
   Le langage $\mathrm{Fr}(L)$ des frontières des arbres de L est algébrique.
2. Soit L' un langage algébrique *propre* ($\varepsilon \notin L'$).
   Il existe un langage d'arbres reconnaissable L tel que $L' = \mathrm{Fr}(L)$.

Lemme d'itération

Théorème : Bar-Hillel, Perles, Shamir ou Lemme d'itération

Soit $L \in \mathrm{Alg}$, il existe $N \geq 0$ tel que pour tout $w \in L$,
si $|w| \geq N$ alors on peut trouver une factorisation $w = \alpha u \beta v \gamma$ avec
$|uv| > 0$ et $|u\beta v| < N$ et $\alpha u^n \beta v^n \gamma \in L$ pour tout $n \geq 0$.

Exemple :

Le langage $L_1 = \{a^n b^n c^n \mid n \geq 0\}$ n'est pas algébrique.

Corollaire :

Les familles Alg et Lin ne sont pas fermées par intersection ou complémentaire.

Lemme d'Ogden

Plus fort que le théorème de Bar-Hillel, Perles, Shamir.

Lemme : Ogden

Soit $G = (\Sigma, V, P, S)$ une grammaire. Il existe un entier $N \in \mathbb{N}$ tel que pour tout $x \in V$ et $w \in \widehat{L}_G(x)$ contenant au moins N lettres distinguées, il existe $y \in V$ et $\alpha, u, \beta, v, \gamma \in (\Sigma \cup V)^*$ tels que

- $w = \alpha u \beta v \gamma$,
- $x \xrightarrow{*} \alpha y \gamma$, $y \xrightarrow{*} uyv$, $y \xrightarrow{*} \beta$,
- $u\beta v$ contient moins de N lettres distinguées,
- soit α, u, β soit β, v, γ contiennent des lettres distiguées.

Lemme d'Ogden

Exercice :

Le langage $L_2 = \{a^n b^n c^p d^p \mid n, p \geq 0\}$ est algébrique mais pas linéaire.

Corollaire :

La famille Lin n'est pas fermée par concaténation ou itération.

Exercice :

Le langage $L_3 = \{a^n b^n c^p \mid n, p > 0\} \cup \{a^n b^p c^p \mid n, p > 0\}$ est linéaire et (inhéremment) ambigu.

Corollaire :

Les langages non ambigus ne sont pas fermés par union.

Grammaires quadratiques

Définition : Taille d'une grammaire

La taille d'une grammaire $G = (\Sigma, V, P, S)$ est $|G| = |\Sigma| + |V| + ||P||$
avec $||P|| = \sum_{x \to \alpha \in P} 1 + |\alpha|$.

Pour simplifier certains algorithmes ou obtenir une meilleure complexité, il est parfois utile de borner la taille des seconds membres des règles.

Définition : Forme normale quadratique faible (FNQF)

Une grammaire $G = (\Sigma, V, P, S)$ est en FNQF si $P \subseteq V \times (V \cup \Sigma)^{\leq 2}$.

Remarque: Pour une grammaire en FNQF on a $||P|| = \mathcal{O}(|P|)$.

Proposition :

Soit $G = (\Sigma, V, P, S)$ une grammaire.
On peut construire en temps $\mathcal{O}(|G|)$ une grammaire équivalente G' en FNQF.
En particulier, $|G'| = \mathcal{O}(|G|)$.

Forme normale quadratique faible

ex: $S \to aSbS \mid \varepsilon$ n'est pas en FNQF
le membre droit $aSbS$ est de longueur >2

$S \to \varepsilon \mid aS_1$ est une grammaire
$S_1 \to SS_2$ équivalente en FNQF
$S_2 \to bS$

Lemme Soit $G = (\Sigma, V, P, S)$
On peut construire en temps linéaire
G' équivalente en FNQF
En particulier $|G'| = O(|G|)$

$V' \leftarrow V$, $P' \leftarrow \emptyset$
$\forall$ règle $r = (x, \alpha) \in P$ faire $\quad O(\|P\|)$
  si $|\alpha| \leq 2$ alors
    ajouter $x \to \alpha$ à P'
  sinon $\alpha = \alpha_1 \alpha_2 \cdots \alpha_k$ $k \geq 3$ $\alpha_i \in \Sigma \cup V$
    ajouter $r_1, \ldots, r_{k-2}$ à V'
      où $r_1, \ldots, r_{k-2}$ sont des nouvelles variables
    ajouter à P
      $x \to \alpha_1 r_1$, $r_1 \to \alpha_2 r_2$, $\ldots$ $r_{k-2} \to \alpha_{k-1} \alpha_k$

Grammaires réduites

Définition : Grammaires réduites

La grammaire $G = (\Sigma, V, P, S)$ est réduite si toute variable $x \in V$ est

- productive : $\mathcal{L}_G(x) \neq \emptyset$, i.e., $\exists\ x \xrightarrow{*} u \in \Sigma^*$, et
- accessible : il existe une dérivation $S \xrightarrow{*} \alpha x \beta$ avec $\alpha, \beta \in (\Sigma \cup V)^*$.

Lemme : Soit $G = (\Sigma, V, P, S)$ une grammaire.

1. On peut calculer l'ensemble des variables productives de G $\mathcal{O}(|G|)$.
2. On peut décider si $\mathcal{L}_G(S) = \emptyset$ $\mathcal{O}(|G|)$.
3. On peut calculer l'ensemble des variables accessibles de G $\mathcal{O}(|G|)$.

Corollaire : Soit $G = (\Sigma, V, P, S)$ une grammaire

Si $\mathcal{L}_G(S) \neq \emptyset$, on peut construire une grammaire réduite équivalente $\mathcal{O}(|G|)$.

Preuve : Restreindre aux variables productives, puis aux variables accessibles.

Calcul des Variables productives en temps linéaire

$X_0 = \emptyset$ $\quad X_{n+1} = \{x \in V \mid \exists\, x \to \alpha \in P$ avec $\alpha \in (\Sigma \cup X_n)^*\}$

$X = \bigcup_{n \geq 0} X_n$

Rem : - $(X_n)_{n \geq 0}$ suite $\nearrow$

- $X_n = X_{n+1} \Rightarrow X = X_n$
- $X = X_{|V|}$

Lemme $\forall n \geq 0 \quad X_n = Y_n$ où

$Y_n = \{x \in V \mid \exists\, t$ arbre de dérivation avec $x = rac(t),\ Fr(t) \in \Sigma^*,\ h(t) \leq n\}$

Preuve : Récurrence sur n.

- **$n = 0$** si $x = rac(t)$ et $h(t) = 0$ alors $Fr(t) = x \notin \Sigma^*$. Donc $Y_0 = \emptyset$ ✓

- **$n \to n+1$:** $\boxed{\subseteq}$ Soit $x \in X_{n+1}$

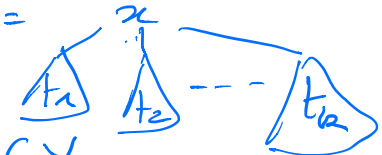
$\exists\, x \to \alpha \in P$ avec $\alpha \in (\Sigma \cup X_n)^*$

$\alpha = \alpha_1 \alpha_2 \cdots \alpha_k$

si $\alpha_i \in \Sigma$ soit $t_i = \alpha_i$

si $\alpha_i \in X_n$ soit t_i un AD de racine α_i, $Fr(t_i) \in \Sigma^*$ et $h(t_i) \leq n$ (HR),

Soit $t =$



t est 1 AD

$Fr(t) \in \Sigma^*$

$h(t) \leq n+1$

Donc $x \in Y_{n+1}$

⊇ Soit $x \in Y_{n+1}$ et t un AD tq ...
$Fr(t) \in \Sigma^+$ donc $h(t) \geq 1$

Alors $t =$ 

avec $\forall i$, t_i AD
$Fr(t_i) \in \Sigma^*$
$h(t_i) \leq n$

- soit $h(t_i) = 0$ et $\alpha_i = rac(t_i) \in \Sigma \ (\cup \{\varepsilon\})$
- Soit $\alpha_i = rac(t_i) \in Y_n \subseteq X_n$ par HR

Donc $x \to \alpha = \alpha_1 \dots \alpha_k \in P$ et $\alpha \in (\Sigma \cup X_n)^*$ □

Cor : X est l'ensemble des variables productives

Calcul naif de X : $O(|V| \cdot |G|)$
→ $|V|$ itérations car $X = X_{|V|}$
→ Chaque itération (calcul de X_{n+1}) se fait en $O(\|P\|)$

Réduction à l'évaluation de clauses de Horn.

Pour chaque règle $x \to \alpha \in P$
créer la clause $y_1, \dots, y_k \to x$ | $O(|G|)$
où $V \cap alph(\alpha) = \{y_1, \dots, y_k\}$

En particulier, une règle $x \to \alpha$ avec $\alpha \in \Sigma^*$ engendre la clause $\to x$

Une variable est productive ssi elle s'évalue à Vrai
⇒ On peut calculer les variables productive en temps linéaire.

Calcul de X en $\mathcal{O}(|G|)$: Algo direct

Pour chaque variable $y \in V$ on crée la liste Règles(y) des règles $x \to \alpha \in P$ tq $y \in \text{alph}(\alpha)$.

```
CréerFile(V) F ;
∀ x ∈ V faire | Prod[x] ← False          O(|V|)
              | Règles[x] ← ∅
Pour chaque r = (x, α) ∈ P faire          O(||P||)
    n[r] ← |V ∩ alph(α)|
    Si n[r] = 0 ∧ ¬Prod[x] alors
        Prod[x] ← True ; F.enfiler(x)     O(1+|α|)
    Sinon
        ∀ y ∈ V ∩ alph(α) faire
            ajouter r à Règles[y]

TQ F ≠ ∅ faire                            O(||P||)
    y ← F.défiler()
    ∀ r = (x, α) ∈ Règles[y] faire
        n[r] ← n[r] - 1
        si n[r] = 0 ∧ ¬Prod[x] alors
            Prod[x] ← True ; F.enfiler(x)
```

$\Rightarrow$ On peut décider en temps $\mathcal{O}(|G|)$ si $\mathcal{L}_G(S) = \emptyset$.

Calcul des variables accessibles en temps linéaire

Soit $G = (\Sigma, V, P, S)$

On construit un graphe

Sommets : V

arêtes : $E = \{(x, y) \mid \exists (x, \alpha) \in P$ avec $y \in \text{alphabet}(\alpha)\}$

Remarques :

1) On peut construire le graphe (V, E) en temps $\mathcal{O}(|G|)$

2) Une variable est accessible dans G ssi elle est accessible à partir de S dans le graphe (V, E).

3) On peut calculer l'ensemble Y des variables accessibles par un parcours (largeur ou profondeur) du graphe (V, E) à partir de S.

→ temps linéaire.

Réduction d'une grammaire. (Linéaire)

- Calculer l'ensemble X des var productives $O(|G|)$
  Calculer $P' = P \cap X \times (\Sigma \cup X)^*$ $O(|G|)$
  $G = (\Sigma, V, P, S)$ équivalente à $G' = (\Sigma, X, P', S)$

- Calculer l'ensemble Y des var accessibles de la grammaire G'. $O(|G|)$
  Calculer $P'' = P' \cap Y \times (\Sigma \cup Y)^*$ $O(|G|)$
  $G'' = (\Sigma, Y, P'', S)$ est réduite et équivalente à G.

⚠ Si on restreint aux variables accessibles puis aux variables productives la grammaire obtenue n'est pas forcément réduite.

Ex: $S \to a + S_1 \quad S_1 \to S_1 S_2 \quad S_2 \to b$
Toutes les variables sont accessibles
Mais S_1 n'est pas productive
Si on l'enlève il reste $S \to a \quad S_2 \to b$
et S_2 n'est plus accessible.

Grammaires propres

Définition : Grammaires propres

La grammaire $G = (\Sigma, V, P, S)$ est propre si $P \subseteq V \times ((\Sigma \cup V)^+ \setminus V)$,
i.e., elle ne contient pas de règle de la forme $x \to \varepsilon$ ou $x \to y$ avec $x, y \in V$.
Un langage $L \subseteq \Sigma^*$ est propre si $\varepsilon \notin L$.

Lemme :

Soit $G = (\Sigma, V, P, S)$ une grammaire.
On peut calculer l'ensemble des variables x telles que $\varepsilon \in \mathcal{L}_G(x)$ $\mathcal{O}(|G|)$.
On peut construire une grammaire équivalente sans ε-règle autre que $S \to \varepsilon$ et dans ce cas S n'apparaît dans aucun membre droit $\mathcal{O}(|G|)$.

Proposition :

Soit $G = (\Sigma, V, P, S)$ une grammaire.
On peut construire une grammaire propre G' qui engendre $\mathcal{L}_G(S) \setminus \{\varepsilon\}$ $\mathcal{O}(|G|^2)$.

Remarque : la réduction d'une grammaire propre est une grammaire propre.

Corollaire :

On peut décider si un mot $u \in \Sigma^*$ est engendré par une grammaire G.

Suppression des ε-règles : (Linéaire)

1) Calculer $Z = \{x \in V \mid \exists\, x \to^* \varepsilon\}$

Rem : Z est l'ensemble des variables productives avec les règles $P \cap V \times V^*$

Donc on peut calculer Z en $O(|G|)$

→ On peut décider en $O(|G|)$ si $\varepsilon \in \mathcal{L}_G(S)$

2) On suppose G en FNQF

$P' \leftarrow \emptyset$

$\forall$ règle $x \to \alpha \in P$ $\quad O(\|P\|)$

- Si $\alpha \neq \varepsilon$ alors ajouter $x \to \alpha$ à P'
- Si $\alpha = \alpha_1 \alpha_2 \wedge \alpha_1 \in Z$ alors ajouter $x \to \alpha_2$ à P'
- Si $\alpha = \alpha_1 \alpha_2 \wedge \alpha_2 \in Z$ alors ajouter $x \to \alpha_1$ à P'

$G' = (\Sigma, V, P')$

Rem Si $S \in Z$ on peut ajouter un nouvel axiome S' et les règles $S' \to S + \varepsilon$

Rem : . G' est encore quadratique

. Si $\mathcal{L}_G(x) = \{\varepsilon\}$ alors $\mathcal{L}_{G'}(x) = \emptyset$ (cf. lem ci-dessous)

Donc G' n'est pas forcément réduite, même si G est réduite

. La réduction d'une grammaire sans ε-règle est encore sans ε-règle.

Lemme $\forall x \in V \quad \mathcal{L}_{G'}(x) = \mathcal{L}_G(x) \setminus \{\varepsilon\}$

Preuve: $\subseteq$ facile.

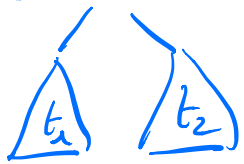
Si $x \to \alpha' \in P'$ alors $x \to^+ \alpha'$ dans G.

Donc $x \to^* w$ dans $G' \Rightarrow x \to^* w$ dans G.

De plus $\varepsilon \notin \mathcal{L}_{G'}(x)$ car il n'y a pas d'ε-règle dans G'.

$\supseteq$ Rec. sur hauteur arbre dérivation

Soit t AD avec $rac(t) = x$ et $w = Fr(t) \in \Sigma^+$

1er Cas $t =$ 

Soit $x_i = rac(t_i)$ et $w_i = Fr(t_i)$

Si $w_i \neq \varepsilon$ alors

(HR) $x_i \to^* w_i$ dans G'

- Si $w_1 \neq \varepsilon \neq w_2$ alors $x \to x_1 x_2 \in P'$ et $x \to^* w_1 w_2 = w$ dans G'
- Si $w_1 \neq \varepsilon = w_2$ alors $x \to x_1 \in P'$ et $x \to^* w_1 = w$ dans G'
- Si $w_1 = \varepsilon \neq w_2$ idem
- $w_1 = \varepsilon = w_2$ est impossible car $w_1 w_2 = w \neq \varepsilon$

2ème Cas $t =$ 

$w = Fr(t_1) \in \Sigma^+$

HR $rac(t_1) \to^* w$ dans G'

De plus $x \to rac(t_1) \in P'$ □

Suppression des règles $x \to y$ $\quad O(|V| \times |G|)$

Soit $G_1 = (\Sigma, V, P_1)$ une grammaire.

Soit $P_2 = \{ (x, \alpha) \mid \exists\, y \to \alpha \in P_1$ avec $\alpha \notin V$ et $x \to^* y$ dans $G_1 \}$

$G_2 = (\Sigma, V, P_2)$ n'a pas de règle $x \to y$

Lemme $\forall x \in V \quad \mathcal{L}_{G_1}(x) = \mathcal{L}_{G_2}(x)$

Preuve $\subseteq$ récurrence longueur dérivation.

- si $x \to w \in \Sigma^*$ dans G_1 alors $x \to w \in P_2$ ✓
- si $x \to x_1 \to \dots \to x_k \to \alpha \to^* w$ dans G_1 avec $\alpha \notin V$ alors $x \to^* x_k$ dans G_1 donc $x \to \alpha \in P_2$. HR $\Rightarrow$ $\alpha \to^* w$ dans G_2 ✓

$\supseteq$ récurrence longueur dérivation

$x \to \alpha \to^* w \in \Sigma^*$ dans G_2.

$\exists\, y \to \alpha \in P_1$ avec $x \to^* y$ dans G_1

De plus (HR) $\alpha \to^* w$ dans G_1

Donc $x \to^* w$ dans G_1 □

Rem : $\|P_2\| = |V| \times \|P_1\|$

- G_1 quadratique $\Rightarrow$ G_2 aussi
- G_1 sans ε-règle $\Rightarrow$ G_2 aussi
- G_2 peut avoir des variables non accessibles
- La réduction d'une grammaire propre (pas de $x \to \varepsilon$ ou $x \to y$) est encore propre

Calcul de P_2 en temps $O(|V| \times |G_1|)$

On suppose que G_1 n'a pas d'ε-règle !

1) $\forall y \in V$ calculer

$$W(y) = \{x \in V \mid x \xrightarrow{*} y \text{ dans } G_1\}$$

Algo • Construire le graphe (V, E) tq

$E = \{(y, x) \in V^2 \mid x \to y \in P_1\}$ $O(\|P_1\|)$

• $\forall y \in V$ faire un DFS dans (V, E) à partir de y pour obtenir $W(y)$

DFS(y) : $O(|E|) = O(\|P_1\|)$

$\forall y \in V$: $|V|$ fois $\to O(|V| \times \|P_1\|)$

2) $\forall y \to \alpha \in P_1$ avec $\alpha \notin V$ $O(|V| \times \|P_1\|)$

$\forall x \in W(y)$

Ajouter $x \to \alpha$ à P_2

Corollaire: le problème du mot est décidable

Soit G une grammaire et $w \in \Sigma^*$ un mot

- On a déjà vu comment décider si $\varepsilon \in L_G(S)$ en temps $O(|G|)$
- si $w \neq \varepsilon$, calculer G' équivalente propre et réduite. Si $x \xrightarrow{n} |w|$ alors $n < 2|w|$

car chaque dérivation ajoute une lettre terminale ou augmente strictement la longueur du mot □

Grammaires quadratiques

Définition : Forme normale de Chomsky (FNC)

Une grammaire $G = (\Sigma, V, P, S)$ est en forme normale de Chomsky si $P \subseteq \{(S, \varepsilon)\} \cup (V \times (V^2 \cup \Sigma))$ et si $(S, \varepsilon) \in P$ alors S n'apparaît dans aucun membre droit.

Proposition :

Soit $G = (\Sigma, V, P, S)$ une grammaire.
On peut construire en temps $\mathcal{O}(|G|^2)$ une grammaire équivalente en FN de Chomsky.

Remarque :

La réduction d'une grammaire en FNC est encore en FNC.

Forme normale de Chomsky FNC $O(|G|^2)$

Soit $G = (\Sigma, V, P, S)$

1) Mettre la grammaire en FNQF $O(|G|)$
2) Supprimer les règles $x \to \varepsilon$ $O(|G|)$
3) Introduire de nouvelles variables $O(|G|)$
   $(x_a)_{a \in \Sigma}$ et les règles $x_a \to a$
   Dans les règles $x \to \alpha_1 \alpha_2$, remplacer les lettres terminales par les variables correspondantes
   ex : $x \to ay$ devient $x \to x_a y$
   les règles sont maintenant
   $$P' \subseteq V' \times (\Sigma \cup V' \cup V'V')$$
4) Supprimer les règles $x \to y$ $O(|V| \cdot |G|)$

la grammaire est en FNC
on peut éventuellement la réduire.

Problèmes décidables

Proposition : Problème du mot : Cocke, Younger, Kasami [9, p. 139]

Soit G une grammaire algébrique.

On peut décider si un mot w est engendré par G en temps $\mathcal{O}(|w|^3)$.

Exercice :

Soit G une grammaire algébrique et $\mathcal{A}$ un automate fini.

Montrer que l'on peut décider en temps polynomial si $\mathcal{L}(G) \cap \mathcal{L}(\mathcal{A}) \neq \emptyset$.

Proposition : Vide et finitude

Soit G une grammaire algébrique.

On peut décider si le langage engendré par G est vide, fini ou infini (PTIME).

Soit $G = (\Sigma, V, P)$ une grammaire

$w = a_1 a_2 \cdots a_n \in \Sigma^+$ un mot.

On suppose G en FNGF et sans ε-règle

$$P \subseteq V \times (\Sigma \cup V \cup V^2)$$

Rem: On peut transformer en temps linéaire une grammaire G arbitraire en une grammaire G' en FNGF et sans ε-règle

Programmation dynamique

Pour $1 \le i \le j \le n$ $\quad w[i,j] = a_i \cdots a_j$

$\forall 1 \le i \le j \le n \quad f(i,j) = \{ x \in V \mid x \xrightarrow{*} w[i,j] \}$

Equations de récurrence

$$\forall 1 \le i \le n \quad f(i,i) = \{ x \in V \mid x \xrightarrow{*} a_i \}$$

On peut calculer $f(i,i)$ en temps $\quad O(|P|)$

Construire le graphe (V, E) où

$E = \{ (y, x) \mid x \to y \in P \}$ $\quad O(|P|)$

Calculer $U_i = \{ x \in V \mid x \to a_i \in P \}$ $\quad O(|P|)$

Calculer $\mathrm{reach}(U_i)$ dans (V, E) $\quad O(|P|)$

$\forall 1 \le i < j \le n$

$$f(i,j) = \mathrm{Reach}\{ x \in V \mid \exists\, x \to yz \in P \ \exists\, i \le k < j \ y \in f(i,k) \wedge z \in f(k+1,j) \}$$

$O(|P|) + O(|P| \cdot (j-i)) = O(|P||w|)$

On calcule $f(i,j)$ $\forall 1 \le i \le j \le n$ $\quad O(|P||w|^3)$

$f(1,n) = \{ x \in V \mid x \xrightarrow{*} w \}$

Problèmes indécidables

Proposition :

Soient L, L' deux langages algébriques et R un langage rationnel.
Les problèmes suivants sont indécidables :

- $L \cap L' = \emptyset$?
- $L = \Sigma^*$?
- $L = L'$?
- $L \subseteq L'$?
- $R \subseteq L$?
- L est-il rationnel ?
- L est-il déterministe ?
- L est-il ambigu ?
- $\overline{L}$ est-il algébrique ?
- $L \cap L'$ est-il algébrique ?

Problème de correspondance de Post : PCP

Données : 2 morphismes $f, g : A^* \to B^*$

Question : $\exists w \in A^+$ tq $f(w) = g(w)$?

Ce problème est indécidable.

Lemme 1: Le langage L_f est linéaire (et déterministe) :

$L_f = \{ \tilde{u} \# f(u) \mid u \in A^+ \}$ où $\tilde{u}$ miroir de u

Preuve: L_f est engendré par la grammaire linéaire

G_f : $F \to aFf(a)$ $\forall a \in A$ et $F \to \#$

exemple de dérivation

$S \to a_n F f(a_n) \to a_n a_{n-1} F f(a_{n-1}) f(a_n)$

$\cdots \to a_n \cdots a_1 F f(a_1) \cdots f(a_n)$

$\to a_n \cdots a_1 \# f(a_1 \cdots a_n)$ (f morphisme)

Corollaire 1 Le vide de l'intersection de langages linéaires est indécidable.

Preuve: PCP a une solution ssi $L_f \cap L_g \neq \emptyset$

Corollaire 2 l'ambiguité d'une grammaire est indécidable

Preuve: G_f et G_g sont non ambigües

Donc $G_f \cup G_g$ est ambigüe ssi $L_f \cap L_g \neq \emptyset$

Pour $G_f \cup G_g$ on ajoute l'axiome S et les règles

$S \to F$ $S \to G$

Lemme 2: Le langage $L_{\neq f}$ est linéaire (et déterministe):

$$L_{\neq f} = \{ \tilde{u} \# v \mid u \in A^*, v \in B^*, v \neq f(u) \}$$

Preuve On construit la grammaire linéaire $G_{\neq f}$

$S \to a S f(a) \quad \forall a \in A$

$S \to a S_1 w$ si $a \in A$, $|w| \leq |f(a)|$ mais w n'est pas un suffixe de $f(a)$

$S \to a S_2 w$ si $a \in A$ et $|w| < |f(a)|$

$S \to S_3$

avec des grammaires linéaires (droite) pour les langages rationnels

$$\mathcal{L}(S_1) = A^* \# B^*$$
$$\mathcal{L}(S_2) = A^* \#$$
$$\mathcal{L}(S_3) = \# B^+$$

Remarque: On utilise les règles

- $S \to a S_2 w$ pour engendrer un mot $\tilde{u} \# v$ avec v trop court : $f(u) \in B^+ v$
- $S \to S_3$ pour engendrer un mot $\tilde{u} \# v$ avec v trop long : $v \in B^+ f(u)$
- $S \to a S_1 w$ pour engendrer un mot $\tilde{u} \# v$ avec $v = v_2 b v_1$ $f(u) = v_3 b' v_1$ $b \neq b'$

Corollaire: L'universalité d'un langage linéaire est indécidable

Preuve: le langage $L = L_{\neg f} \cup L_{\neg g} \cup \overline{A^+ \# B^*}$ est linéaire.

Claim $L = \Sigma^*$ ssi PCP n'a pas de solution

1) Si PCP a une solution $u \in A^+$ tq $u = f(u) = g(u)$
alors $\tilde{u} \# u \in A^+ \# B^* \setminus (L_{\neg f} \cup L_{\neg g})$
donc $\tilde{u} \# u \notin L$ et $L \neq \Sigma^*$

2) Si $L \neq \Sigma^*$ $\exists w \in A^+ \# B^* \setminus (L_{\neg f} \cup L_{\neg g})$
alors $w = \tilde{u} \# v$ avec $u \in A^+$, $v = f(u) = g(u)$

Propriétés de clôture

Proposition :

1. La famille Alg est fermée par concaténation, itération.
2. La famille Alg est fermée par substitution algébrique.
3. Les familles Alg et Lin sont fermées par union et miroir.
4. Les familles Alg et Lin sont fermées par intersection avec un rationnel.
5. Les familles Alg et Lin sont fermées par morphisme.
6. Les familles Alg et Lin sont fermées par projection inverse.
7. Les familles Alg et Lin sont fermées par morphisme inverse.

Définition : Substitutions algébriques

Une substitution $\sigma : A \to \mathcal{P}(B^*)$ est algébrique si $\forall a \in A$, $\sigma(a) \in \mathrm{Alg}$

Définition : Projection

La projection de A sur $B \subseteq A$ est le morphisme $\pi : A^* \to B^*$ défini par

$$\pi(a) = \begin{cases} a & \text{si } a \in B \\ \varepsilon & \text{sinon.} \end{cases}$$

Transductions rationnelles

Définition : Transduction rationnelle

Une transduction rationnelle (TR) $\tau : A^* \to \mathcal{P}(B^*)$ est la composée d'un morphisme inverse, d'une intersection avec un rationnel et d'un morphisme.

Soient A, B, C trois alphabets, $K \in \mathrm{Rat}(C^*)$ et $\varphi : C^* \to A^*$ et $\psi : C^* \to B^*$ deux morphismes. L'application $\tau : A^* \to \mathcal{P}(B^*)$ définie par $\tau(w) = \psi(\varphi^{-1}(w) \cap K)$ est une TR.

Proposition :

Les familles Alg, Lin et Rat sont fermées par TR.

Transductions rationnelles

Théorème : Chomsky et Schützenberger

Les propositions suivantes sont équivalentes :

1. L est algébrique.
2. Il existe une TR τ telle que $L = \tau(D_2^*)$.
3. Il existe un entier n, un rationnel K et un morphisme alphabétique ψ tels que $L = \psi(D_n^* \cap K)$.

Corollaire :

Les langages non ambigus ne sont pas fermés par morphisme.

Théorème : Elgot et Mezei, 1965

La composée de deux TR est encore une TR.

Théorème : Nivat, 1968

Une application $\tau : A^* \to \mathcal{P}(B^*)$ est une TR si et seulement si son graphe

$$\{(u, v) \mid v \in \tau(u)\}$$

est une relation rationnelle (i.e., un langage rationnel de $A^* \times B^*$).

Forme normale de Greibach

Définition :

La grammaire $G = (\Sigma, V, P)$ est en

FNG (forme normale de Greibach)	si $P \subseteq V \times \Sigma V^*$
FNPG (presque Greibach)	si $P \subseteq V \times \Sigma(V \cup \Sigma)^*$
FNGQ (Greibach quadratique)	si $P \subseteq V \times (\Sigma \cup \Sigma V \cup \Sigma V^2)$

Remarque : on passe trivialement d'une FNPG(Q) à une FNG(Q).

Théorème :

Soit $G = (\Sigma, V, P)$ une grammaire propre.
On peut construire $G' = (\Sigma, V', P')$ en FNG équivalente à G,
i.e., $V \subseteq V'$ et $\mathcal{L}_G(x) = \mathcal{L}_{G'}(x)$ pour tout $x \in V$.

La difficulté est d'éliminer la récursivité gauche des règles.

Exemple : Mettre la grammaire suivante en FNPG

$$G_1 : \begin{cases} x_1 \to x_1 b + a \\ x_2 \to x_1 b + a x_2 x_1 \end{cases}$$

$$G_1 : \begin{cases} x_1 \longrightarrow x_1 b + a \\ x_2 \longrightarrow x_1 b + a x_2 x_1 \end{cases}$$

Méthode "à la main".

$L(x_1) = ab^* = a + ab^+$

On peut engendrer b^+ avec $x'_1 \to b + bx'_1$

Donc dans G'_1 on met

$$x_1 \longrightarrow a + a\,x'_1$$
$$x'_1 \longrightarrow b + b\,x'_1$$

On a $L_{G_1}(x_1) = L_{G'_1}(x_1)$

Reste à traiter $x_2 \longrightarrow x_1 b$

Il suffit de remplacer x_1 par $a + ax'_1$

On obtient dans G'_1 $x_2 \to ab + ax'_1 b + ax_2 x_1$

Forme normale de Greibach

Preuve

Soit $G = (\Sigma, V, P)$ une grammaire avec $V = \{x_1, \ldots, x_n\}$.

Pour $i, j \in \{1, \ldots, n\}$ on pose $\begin{cases} \alpha_{i,j} &= x_i^{-1} P(x_j) \subseteq (\Sigma \cup V)^* \\ \beta_j &= P(x_j) \cap (\Sigma \cdot (\Sigma \cup V)^* \cup \{\varepsilon\}) \end{cases}$

de sorte que les règles de G s'écrivent $x_j \to \sum_i x_i \alpha_{i,j} + \beta_j$ pour $1 \leq j \leq n$.

On peut écrire P vectoriellement : $X \to XA + B$

avec $X = (x_1, \ldots, x_n)$, $B = (\beta_1, \ldots, \beta_n)$ et $A = (\alpha_{i,j})_{1 \leq i,j \leq n}$.

On définit $G' = (\Sigma, V', P')$ par $V' = V \uplus \{y_{i,j} \mid 1 \leq i, j \leq n\}$ et

$$P' : \quad \begin{array}{lcl} X & \to & BY + B \\ Y & \to & AY + A \end{array} \qquad \text{i.e.} \qquad \begin{array}{lcl} x_j & \to & \sum_k \beta_k y_{k,j} + \beta_j \\ y_{i,j} & \to & \sum_k \alpha_{i,k} y_{k,j} + \alpha_{i,j} \end{array}$$

avec $Y = (y_{i,j})_{1 \leq i,j \leq n}$.

Proposition : Equivalence des grammaires

Les grammaires G et G' sont équivalentes, i.e., $\forall x \in V, \quad \mathcal{L}_G(x) = \mathcal{L}_{G'}(x)$.

On illustre la preuve sur

$$G_2 \left| \begin{array}{l} x_1 \to x_1x_1 + x_1x_2 + x_2a + b \\ x_2 \to x_1x_2 + x_2x_1 + a \end{array} \right.$$

$$A = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} = \begin{pmatrix} x_1+x_2 & x_2 \\ a & x_1 \end{pmatrix} \qquad B = (\beta_1, \beta_2) = (b, a) \qquad X = (x_1, x_2)$$

On a bien $X \to XA + B$

$$Y = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix}$$

$X \to BY + B$ s'écrit
$$x_1 \to b y_{11} + a y_{21} + b$$
$$x_2 \to b y_{12} + a y_{22} + a$$

Ces règles sont déjà en FNG

$Y \to AY + A$ s'écrit
$$y_{11} \to (x_1+x_2) y_{11} + x_2 y_{21} + x_1 + x_2$$
$$y_{12} \to (x_1+x_2) y_{12} + x_2 y_{22} + x_2$$
$$y_{21} \to a y_{11} + x_1 y_{21} + a$$
$$y_{22} \to a y_{12} + x_1 y_{22} + x_1$$

Ces règles ne sont pas en FNG.
Mais les variables à gauche sont x_1 ou x_2
→ substitution

$$y_{11} \to (b y_{11} + a y_{21} + b)(y_{11} + \varepsilon) + (b y_{12} + a y_{22} + a)(y_{11} + y_{21} + \varepsilon)$$
$$y_{12} \to (b y_{11} + a y_{21} + b) y_{12} + (b y_{12} + a y_{22} + a)(y_{12} + y_{22} + \varepsilon)$$
$$y_{21} \to a y_{11} + (b y_{11} + a y_{21} + b) y_{21} + a$$
$$y_{22} \to a y_{12} + (b y_{11} + a y_{21} + b)(y_{22} + \varepsilon)$$

la grammaire G_2'' est en FNG

Preuve de la Proposition

$\forall x \in V \quad \mathcal{L}_G(x) = \mathcal{L}_{G'}(x)$

Induction sur la longueur d'une dérivation gauche dans G pour $\subset$ et droite dans G' pour $\supset$

1) $\mathcal{L}_G(x) \subset \mathcal{L}_{G'}(x)$

Une dérivation gauche $x \xrightarrow{n}_g w$ dans G s'écrit

$x = x_{i_0} \to_g x_{i_1} v_1 \to_g x_{i_2} v_2 v_1 \dots \to_g x_{i_k} v_k \dots v_2 v_1 \to_g v_{k+1} v_k \dots v_2 v_1$

avec $v_1 \in \alpha_{i_1 i_0}$ $v_2 \in \alpha_{i_2 i_1}$ $\dots$ $v_k \in \alpha_{i_k i_{k-1}}$ et $v_{k+1} \in \beta_{i_k}$

puis $v_{k+1} v_k \dots v_2 v_1 \xrightarrow{n-k-1}_g w$

Lemme fondamental : $w = w_{k+1} w_k \dots w_2 w_1$

et $v_i \xrightarrow{n_i}_g w_i$ avec $n+k-1 = n_1 + \dots + n_{k+1}$

HR : dans G' on a $v_i \xrightarrow{*}_d w_i \quad \forall\, 1 \le i \le k+1$

De plus, on a dans G' la dérivation droite suivante :

$x = x_{i_0} \to_d v_{k+1} y_{i_k i_0} \to_d v_{k+1} v_k y_{i_{k-1} i_0} \dots \to_d v_{k+1} \dots v_2 y_{i_1 i_0} \to_d v_{k+1} \dots v_2 v_1$

$\xrightarrow{*}_d w_{k+1} w_k \dots w_2 w_1 = w$

2) $\mathcal{L}_G(x) \supset \mathcal{L}_{G'}(x)$: similaire.

Forme normale de Greibach

Exemple : Mettre la grammaire suivante en FNG(Q)

$$G_2 : \begin{cases} x_1 \to x_1x_1 + x_1x_2 + x_2a + b \\ x_2 \to x_1x_2 + x_2x_1 + a \end{cases}$$

Remarque : Grammaire propre

Si G est propre alors pour $1 \leq i, j \leq n$ on a

$$\alpha_{i,j} \subseteq (\Sigma \cup V)^+ \qquad \text{et} \qquad \beta_j \subseteq \Sigma \cdot (\Sigma \cup V)^*$$

donc les règles $X \to BY + B$ de G' sont en FNPG.

On définit G'' à partir de G' en remplaçant chaque variable x_ℓ en tête d'un mot de $\alpha_{i,j}$ par sa définition $\sum_k \beta_k y_{k,\ell} + \beta_\ell$.

Proposition : FNG et FNGQ

- Les grammaires G et G'' sont équivalentes.
- Si G est une grammaire propre alors G'' est en FNPG.
- Si G est propre et en FN de Chomsky, alors G'' est en FNGQ.

Plan

Automates à pile

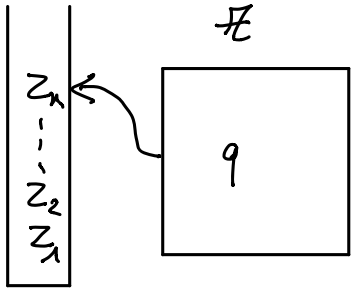
Définition : $\mathcal{A} = (Q, \Sigma, Z, T, q_0 z_0, F)$ où

- Q ensemble fini d'états
- Σ alphabet d'entrée
- Z alphabet de pile
- $T \subseteq QZ \times (\Sigma \cup \{\varepsilon\}) \times QZ^*$ ensemble fini de transitions
- $q_0 z_0 \in QZ$ configuration initiale
- $F \subseteq Q$ acceptation par état final.

De plus, $\mathcal{A}$ est *temps-réel* s'il n'a pas d'ε-transition.

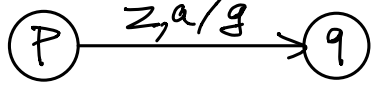
Définition : Système de transitions (infini) associé

- $\mathcal{T} = (QZ^*, T', q_0 z_0, FZ^*)$
- Une configuration de $\mathcal{A}$ est un état $ph \in QZ^*$ de $\mathcal{T}$
- Transitions de $\mathcal{T}$: $T' = \{pzh \xrightarrow{a} qgh \mid (pz, a, qg) \in T\}$.
- $\mathcal{L}(\mathcal{A}) = \{w \in \Sigma^* \mid \exists\, q_0 z_0 \xrightarrow{w} qh \in FZ^* \text{ dans } \mathcal{T}\}$.



Configuration $qz_n \cdots z_2 z_1$: automate $\mathcal{A}$ dans l'état q sommet de pile (tête de lecture) sur z_n

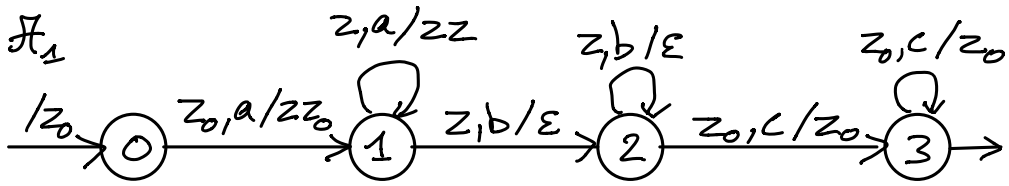
---

Une transition $(pz, a, qg) \in T$ se représente graphiquement par 

configuration initiale par 

Ex: $L_1 = \{ a^n b^n c^p \mid n, p > 0 \}$

$\mathcal{A}_1$



exemple de calcul acceptant pour $a^3 b^3 c^2$

$0 z_0 \xrightarrow{a} 1 z z_0 \xrightarrow{a} 1 z z z_0 \xrightarrow{a} 1 z^3 z_0 \xrightarrow{b} 2 z^2 z_0 \xrightarrow{b} 2 z z_0$
$\xrightarrow{b} 2 z_0 \xrightarrow{c} 3 z_0 \xrightarrow{c} 3 z_0$

Rem $\mathcal{A}_1$ est déterministe (D)
$\mathcal{A}_1$ est temps réel (TR)
$\mathcal{A}_1$ est "à fond de pile testable", i.e.,
la pile ne contient plus qu'un symbole ssi le sommet de pile est z_0

Automates à pile

Exemples :

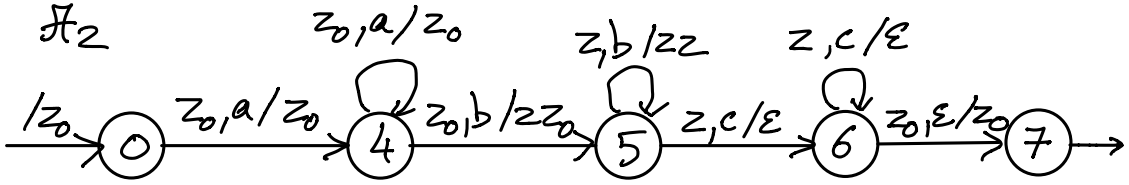
- $L_1 = \{a^n b^n c^p \mid n, p > 0\}$ et $L_2 = \{a^n b^p c^p \mid n, p > 0\}$
- $L = L_1 \cup L_2$ (non déterministe)

Exercices :

1. Montrer que le langage $\{w\tilde{w} \mid w \in \Sigma^*\}$ et son complémentaire peuvent être acceptés par un automate à pile.
2. Montrer que le complémentaire du langage $\{ww \mid w \in \Sigma^*\}$ peut être accepté par un automate à pile.
3. Soit $\mathcal{A} = (Q, \Sigma, Z, T, q_0 z_0, F)$ un automate à pile. Montrer qu'on peut construire un automate à pile équivalent $\mathcal{A}'$ tel que $T' \subseteq Q'Z \times (\Sigma \cup \{\varepsilon\}) \times Q'Z^{\leq 2}$.
4. Soit $\mathcal{A}$ un automate à pile. Montrer qu'on peut construire un automate à pile équivalent $\mathcal{A}'$ tel que les mouvements de la pile sont uniquement du type *push* ou *pop*.

Ex: $L_2 = \{ a^n b^p c^p \mid n,p > 0 \}$

$\mathcal{A}_2$



exemple de calcul acceptant pour $a b^2 c^2$

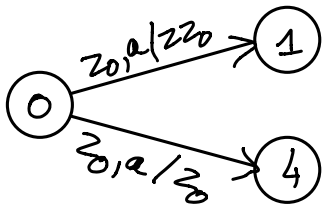
$0 z_0 \xrightarrow{a} 4 z_0 \xrightarrow{b} 5 z z_0 \xrightarrow{b} 5 z^2 z_0 \xrightarrow{c} 6 z z_0 \xrightarrow{c} 6 z_0 \xrightarrow{\varepsilon} 7 z_0$

L'automate $\mathcal{A}_2$ est D mais pas TR.

Ex: $L_3 = L_1 \cup L_2$

Il suffit de faire $\mathcal{A}_3 = \mathcal{A}_1 \cup \mathcal{A}_2$ en fusionnant l'état initial 0.

Mais $\mathcal{A}_3$ est non déterministe (ND) à cause de



Rem: pas d'automate D pour L_3 : $D \subsetneq ND$

Exercice: Construire un automate à pile déterministe pour

$L_4 = \{ a^n b^p c^n \mid n,p > 0 \} \cup \{ a^n b^p d^p \mid n,p > 0 \}$

Rem: Pas d'automate D+TR pour L_4

$D+TR \subsetneq D$

Propriétés fondamentales

Lemme : fondamental

Soient $\mathcal{A} = (Q, \Sigma, Z, T, q_0 z_0)$ un automate à pile, $p, r \in Q$, $g, h \in Z^*$, $w \in \Sigma^*$ et $n \geq 0$. Les conditions suivantes sont équivalentes:

1. $pgh \xrightarrow[n]{w} r$ est un calcul de $\mathcal{A}$,
2. il existe deux calculs $pg \xrightarrow[n_1]{w_1} q$ et $qh \xrightarrow[n_2]{w_2} r$ de $\mathcal{A}$ avec $q \in Q$, $w = w_1 w_2$ et $n = n_1 + n_2$.

Preuve

$\Longrightarrow$: Dans le calcul $pgh \xrightarrow[n]{w} r$, on considère la première fois que le contenu de la pile est h (possible car initialement gh, finalement ε). Soit qh la configuration correspondante après n_1 étapes. Le calcul s'écrit $pgh \xrightarrow[n_1]{w_1} qh \xrightarrow[n_2]{w_2} r$.

Si $p_0 g_0 h \xrightarrow{a_1} p_1 g_1 h \cdots \xrightarrow{a_k} p_k g_k h$ est un calcul de $\mathcal{A}$ tel que $g_i \neq \varepsilon$ pour $0 \leq i < n$ alors $p_0 g_0 \xrightarrow{a_1} p_1 g_1 \cdots \xrightarrow{a_k} p_k g_k$ est un calcul de $\mathcal{A}$.

$\Longleftarrow$: On utilise la remarque suivante: Si $sf \xrightarrow[k]{v} s'f'$ est un calcul de $\mathcal{A}$ avec $s \in Q$, $f, f', f'' \in Z^*$ alors $sff'' \xrightarrow[k]{v} s'f'f''$ est aussi un calcul de $\mathcal{A}$.

Acceptation généralisée

Définition :

Soit $\mathcal{A} = (Q, \Sigma, Z, T, q_0 z_0)$ un automate à pile et $K \subseteq QZ^*$ un langage reconnaissable. Le langage reconnu par $\mathcal{A}$ avec *acceptation généralisée* K est

$$\mathcal{L}_K(\mathcal{A}) = \{w \in \Sigma^* \mid \exists\ q_0 z_0 \xrightarrow{w} qh \in K \text{ dans } \mathcal{T}\}$$

Cas particuliers :

- $K = FZ^*$: acceptation classique par état final.
- $K = Q$: acceptation par pile vide.
- $K = F$: acceptation par pile vide et état final.
- $K = QZ'Z^*$ avec $Z' \subseteq Z$: acceptation par sommet de pile.

Exemple :

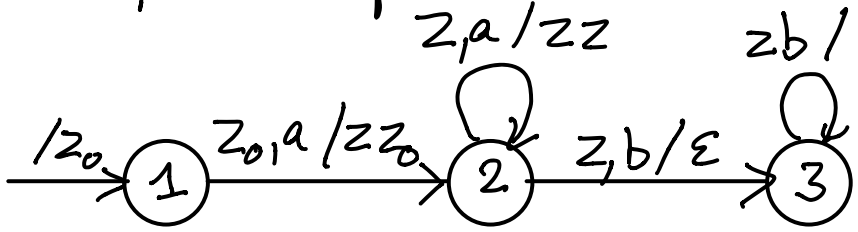
$L = \{a^n b^n \mid n \geq 0\}$ peut être accepté par pile vide ou par sommet de pile.

Proposition : Acceptation généralisée

Soit $\mathcal{A}$ un automate à pile avec acceptation généralisée K, on peut effectivement construire un automate à pile $\mathcal{A}'$ acceptant par état final tel que $\mathcal{L}_K(\mathcal{A}) = \mathcal{L}(\mathcal{A}')$. Donc, tous les modes d'acceptation ci-dessus sont équivalents.

Exemple : $L_5 = \{ a^n b^n \mid n \geq 0 \}$

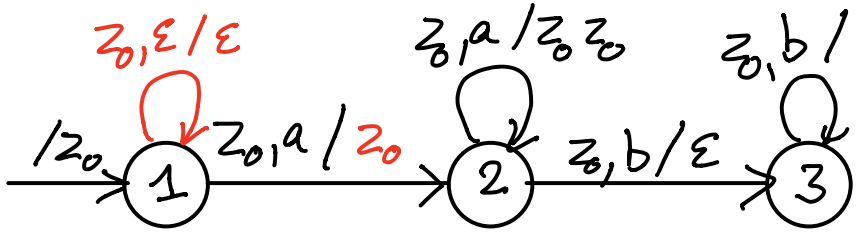
- acceptation par sommet de pile



$Z' = \{ z_0 \}$

déterministe

- Acceptation par pile vide



non déterministe

non ambigu

Acceptation généralisée

Preuve : Acceptation généralisée

Soit $\mathcal{A} = (Q, \Sigma, Z, T, q_0 z_0)$ un automate à pile et $K \subseteq QZ^*$ un langage reconnu par l'automate fini déterministe $\mathcal{B} = (P, Z \cup Q, \delta, p_0, F)$ avec $P \cap Q = \emptyset$.
Intuition: l'automate $\mathcal{A}'$ commence par simuler $\mathcal{A}$, puis (choix ND) lit la configuration atteinte $qh \in QZ^*$ dans l'automate $\mathcal{B}$ pour voir si elle est acceptante.
Soit $\mathcal{A}' = (Q', \Sigma, Z \uplus \{\bot\}, T', q_0' \bot, \{f\})$ avec $Q' = Q \uplus P \uplus \{q_0', f\}$, et

1. $q_0' \cdot \bot \xrightarrow{\varepsilon} q_0 \cdot z_0 \bot \in T'$, Initialisation
2. $T \subseteq T'$, Simulation
3. $q \cdot z \xrightarrow{\varepsilon} \delta(p_0, q) \cdot z \in T'$, si $q \in Q$ et $z \in Z \uplus \{\bot\}$, Acceptation
4. $p \cdot z \xrightarrow{\varepsilon} \delta(p, z) \cdot \varepsilon \in T'$, si $p \in P$ et $z \in Z$, Acceptation
5. $p \cdot \bot \xrightarrow{\varepsilon} f \cdot \varepsilon \in T'$, si $p \in F$, Acceptation

Exercice: Montrer que $\mathcal{L}(\mathcal{A}, K) = \mathcal{L}(\mathcal{A}')$.

Remarque: $\mathcal{A}'$ reconnaît aussi $\mathcal{L}(\mathcal{A}, K)$ par pile vide.

Exercice: Modifier $\mathcal{A}'$ pour qu'il reconnaisse $\mathcal{L}(\mathcal{A}, K)$ par sommet de pile.

Automates à pile et grammaires

Proposition : Automate à pile vers grammaire

Soit $\mathcal{A} = (Q, \Sigma, Z, T, q_0 z_0)$ un automate à pile reconnaissant par pile vide. On peut construire une grammaire G qui engendre $\mathcal{L}(\mathcal{A})$.

De plus, si $\mathcal{A}$ est *temps-réel* alors G est en FNG.

Proposition : Grammaire vers automate à pile

Soit $G = (\Sigma, V, P, S)$ une grammaire. On peut construire un automate à pile simple (un seul état) $\mathcal{A}$ qui accepte $L_G(S)$ par pile vide.

De plus, si G est en FNPG alors on peut construire un tel $\mathcal{A}$ *temps-réel*.

Si G est en FNGQ alors on peut construire un tel $\mathcal{A}$ *standardisé* ($T \subseteq Z \times \Sigma \times Z^{\leq 2}$).

Preuve Automate à Pile vers Grammaire

On construit $G=(\Sigma, V, P)$ avec

$V=\{ X_{p,z,q} \mid p,q \in Q \text{ et } z \in Z \}$

de telle sorte que l'on ait

Lemme: $\mathcal{L}_G(X_{p,z,q}) = \{ w \in \Sigma^* \mid pz \xrightarrow{w}_* q \text{ dans } \mathcal{A} \}$

Pour cela on définit les règles de G par

$P = \{ X_{p,z,q} \longrightarrow a X_{p_1,y_1,p_2} \cdots X_{p_n,y_n,q} \mid$

$pz \xrightarrow{a} p_1 y_1 \cdots y_n \in T$ (donc $a \in \Sigma \cup \{\varepsilon\}$, $n \geq 0$, $y_i \in Z$)

$p_1, p_2, \ldots, p_n \in Q \}$

Rem: si $n=0$ la règle est $X_{p,z,q} \to a$

Preuve du lemme:

C Récurrence sur longueur d'une dérivation.

$X_{p,z,q} \xrightarrow{m} w \in \Sigma^*$

- Si $m=1$ alors $w=a \in \Sigma \cup \{\varepsilon\}$ et $pz \xrightarrow{a} q \in T$: OK
- Si $m>1$ alors la dérivation s'écrit :

$X_{p,z,q} \longrightarrow a X_{p_1,y_1,p_2} \cdots X_{p_n,y_n,q} \xrightarrow{m-1} w$

avec $pz \xrightarrow{a} p_1 y_1 \cdots y_n \in T$

lemme fondamental : $X_{p_i,y_i,p_{i+1}} \xrightarrow{m_i} w_i$ dans G avec

$w = w_1 \cdots w_n$ et $m-1 = m_1 + \cdots + m_n$ $(p_{n+1}=q)$

HR : $p_i y_i \xrightarrow{w_i}_* p_{i+1}$ dans $\mathcal{A}$

On recolle en utilisant le LF pour $\mathcal{A}$

$pz \xrightarrow{a} p_1 y_1 \cdots y_n \xrightarrow{w_1}_* p_2 y_2 \cdots y_n \xrightarrow{w_2}_* p_3 y_3 \cdots y_n$

$\cdots p_n y_n \xrightarrow{w_n}_* p_{n+1} = q$

2 Récurrence sur longueur du calcul $px \xrightarrow{w}_m q$ dans $\mathcal{A}$

- $m=1$ $w=a \in \Sigma \cup \{\varepsilon\}$ et $px \xrightarrow{a} q \in T$
  donc $px \to aq \in P$: OK
- $m>1$ le calcul s'écrit
  $px \xrightarrow{a} p_1 y_1 \cdots y_n \xrightarrow[m-1]{v} q$ avec $n \geq 1$ et $w=av$

Donc $X_{pzq} \to a X_{p_1 y_1 p_2} \cdots X_{p_n y_n p_{n+1}} \in P$ et
LF des automates à Pile appliqué $n-1$ fois
$\exists p_2, \dots, p_n$ et des calculs $p_i y_i \xrightarrow[m_i]{v_i} p_{i+1}$ de $\mathcal{A}$
avec $m-1 = m_1 + \cdots + m_n$ et $v = v_1 \cdots v_n$ $(p_{n+1}=q)$
HR on a dans G les dérivations
$X_{p_i y_i p_{i+1}} \xrightarrow{*} v_i$
On recolle pour obtenir dans G

$$X_{pzq} \to a X_{p_1 y_1 p_2} \cdots X_{p_n y_n p_{n+1}}$$
$$\xrightarrow{*} a v_1 X_{p_2 y_2 p_3} \cdots X_{p_n y_n p_{n+1}}$$
$$\vdots$$
$$\xrightarrow{*} a v_1 \cdots v_{n-1} X_{p_n y_n p_{n+1}} \xrightarrow{*} a v_1 \cdots v_n = w \quad \square$$

Preuve Grammaire vers Automate à Pile

Définition : Automate LL ou expansion/vérification

Soit $G = (\Sigma, V, P, S)$ une grammaire réduite.
On construit l'automate à pile simple non déterministe qui accepte par pile vide :
$\mathcal{A} = (\Sigma, \Sigma \cup V, T, S)$ où les transitions de T sont des

- expansions : $\{(x, \varepsilon, \alpha) \mid (x, \alpha) \in P\}$ ou
- vérifications : $\{(a, a, \varepsilon) \mid a \in \Sigma\}$.

Remarque : sommet de pile à gauche.

Lemme $\forall \alpha \in (\Sigma \cup V)^* \quad \forall w \in \Sigma^*$

$\alpha \xrightarrow[g]{*} w$ dans G ssi $\alpha \xrightarrow[*]{w} \mathcal{E}$ dans $\mathcal{A}$

$\Rightarrow$ Rec. sur n dans $\alpha \xrightarrow[g]{n} w$ dérivation

- $n=0$ $\alpha = w \in \Sigma^*$

  Vérifications : $\alpha \xrightarrow[*]{w} \mathcal{E}$ dans $\mathcal{A}$

- $n>0$ $\alpha = u x \delta$ avec $u \in \Sigma^*, x \in V, \delta \in (\Sigma \cup V)^*$

On applique la règle $x \to \gamma \in P$: $\alpha \xrightarrow[g]{1} u\gamma\delta \xrightarrow[g]{n-1} w$

LF $w = uv$ et $\gamma\delta \xrightarrow[g]{n-1} v$

HR $\gamma\delta \xrightarrow[*]{v} \mathcal{E}$ dans $\mathcal{A}$

donc $\alpha \xrightarrow[*]{u} \gamma\delta \xrightarrow[*]{v} \mathcal{E}$ dans $\mathcal{A}$

(vérifications)

$\Leftarrow$ Rec sur n dans $\alpha \xrightarrow[n]{w} \mathcal{E}$ calcul de $\mathcal{A}$

$n=0$: $w = \mathcal{E} = \alpha$ donc $\alpha \xrightarrow[g]{0} w$

$n>0$: 1) si $\alpha = a\delta$ avec $a \in \Sigma$ alors le calcul est $\alpha \xrightarrow[1]{a} \delta \xrightarrow[n-1]{v} \mathcal{E}$ avec $w = av$

HR $\delta \xrightarrow[g]{*} v$ donc $\alpha = a\delta \xrightarrow[g]{*} av = w$

2) $\alpha = x\delta$ avec $x \in V$ alors le calcul est

$\alpha \xrightarrow[\mathcal{E}]{} \gamma\delta \xrightarrow[n-1]{w} \mathcal{E}$

HR $\gamma\delta \xrightarrow[g]{*} w$ Donc $\alpha = x\delta \xrightarrow[g]{1} \gamma\delta \xrightarrow[g]{*} w$ □

Remarque : Si G est en FNPG, on remplace les expansions par $\{(x, a, \alpha) \mid x \to a\alpha \in P\}$

L'automate obtenu est TR.

Configurations accessibles (mots de pile)

Proposition : Reconnaissabilité des configurations accessibles

Soit $\mathcal{A} = (Q, \Sigma, Z, T, q_0 z_0)$ un automate à pile.
Pour $p \in Q$ et $g \in Z^*$, on note

$$\mathcal{C}(pg) = \{qh \in QZ^* \mid \exists\ pg \not\Rightarrow qh \text{ dans } \mathcal{T}\}$$

l'ensemble des configurations accessibles à partir de pg.
On peut effectivement construire un automate fini $\mathcal{B}$ qui reconnaît $\mathcal{C}(pg)$.

Preuve : Voir plus loin.

Corollaire : Décidabilité du vide

Soit $\mathcal{A} = (Q, \Sigma, Z, T, q_0 z_0, F)$ un automate à pile.
On peut décider si $\mathcal{L}(\mathcal{A}) = \emptyset$.

Preuve

Il suffit de tester si $q_0 \in F$ ou $\mathcal{C}(q_0 z_0) \cap FZ^* \neq \emptyset$.

Calculs d'accessibilité

Corollaire :

Soit $\mathcal{A} = (Q, \Sigma, Z, T, q_0 z_0)$ un automate à pile.
On peut effectivement calculer les ensembles suivants :

1. $X = \{(p, x, q) \in Q \times Z \times Q \mid \exists\ px \overset{+}{\underset{\mathcal{T}}{\Rightarrow}} q \text{ dans } \mathcal{T}\}$
2. $Y = \{(p, x, q, y) \in Q \times Z \times Q \times Z \mid \exists\ px \overset{+}{\underset{\mathcal{T}}{\Rightarrow}} qy \text{ dans } \mathcal{T}\}$
3. $W = \{(p, x, q, y) \in Q \times Z \times Q \times Z \mid \exists\ px \overset{+}{\underset{\mathcal{T}}{\Rightarrow}} qyh \text{ dans } \mathcal{T}\}$
4. $X' = \{(p, x, q) \in Q \times Z \times Q \mid \exists\ px \overset{\varepsilon}{\underset{\mathcal{T}}{\Rightarrow}} q \text{ dans } \mathcal{T}\}$
5. $Y' = \{(p, x, q, y) \in Q \times Z \times Q \times Z \mid \exists\ px \overset{\varepsilon}{\underset{\mathcal{T}}{\Rightarrow}} qy \text{ dans } \mathcal{T}\}$
6. $W' = \{(p, x, q, y) \in Q \times Z \times Q \times Z \mid \exists\ px \overset{\varepsilon}{\underset{\mathcal{T}}{\Rightarrow}} qyh \text{ dans } \mathcal{T}\}$

Preuve

1. $(p, x, q) \in X$ ssi $q \in \mathcal{C}(px)$.
2. $(p, x, q, y) \in Y$ ssi $qy \in \mathcal{C}(px)$.
3. $(p, x, q, y) \in W$ ssi $\mathcal{C}(px) \cap qyZ^* \neq \emptyset$.

On applique ce qui précède à l'automate $\mathcal{A}'$ obtenu à partir de $\mathcal{A}$ en ne conservant que les ε-transitions.

Clôture et réduction.

Soit Γ un alphabet, $\overline{\Gamma} = \{\overline{a} \mid a \in \Gamma\}$ une copie de Γ et $\widetilde{\Gamma} = \Gamma \uplus \overline{\Gamma}$.
On définit la réduction sur $\widetilde{\Gamma}^*$ par $\overline{a}a \xrightarrow{\text{red}} \varepsilon$ pour $a \in \Gamma$.

Remarque : Soit $D = \{w \in \widetilde{\Gamma}^* \mid w \xrightarrow[*]{\text{red}} \varepsilon\}$.

On peut montrer que D est engendré par la grammaire $S \to \varepsilon + \sum_{a\in\Gamma} \overline{a}SaS$.
Il s'agit donc du langage de Dyck en considérant $\overline{a}$ comme une parenthèse ouvrante et a comme la parenthèse fermante correspondante.

Pour $L \subseteq \widetilde{\Gamma}^*$ on pose $\mathrm{Clot}(L) = \{w \in \widetilde{\Gamma}^* \mid \exists v \in L,\ v \xrightarrow[*]{\text{red}} w\}$.

Lemme : Reconnaissabilité de la clôture

Si $L \subseteq \widetilde{\Gamma}^*$ est un langage reconnaissable alors $\mathrm{Clot}(L) \subseteq \widetilde{\Gamma}^*$ aussi.
On peut construire un automate pour $\mathrm{Clot}(L)$ à partir d'un automate pour L (PTime).

Clôture et réduction.

Preuve : Reconnaissabilité de la clôture

Soit $\mathcal{A} = (Q, \widetilde{\Gamma}, T, I, F)$ un automate fini reconnaissant L.

Pour $p, q \in Q$, on note $\mathcal{A}_{p,q} = (Q, \widetilde{\Gamma}, T, p, q)$ et $L_{p,q} = \mathcal{L}(\mathcal{A}_{p,q})$.

On peut décider si $L_{p,q} \cap D \neq \emptyset$ en PTIME

On peut calculer (PTIME) l'ensemble $R = \{(p, q) \in Q^2 \mid L_{p,q} \cap D \neq \emptyset\}$.

On définit $\mathcal{A}' = (Q, \widetilde{\Gamma}, T', I, F)$ en ajoutant à $\mathcal{A}$ des ε-transitions:

$$T' = T \cup \{(p, \varepsilon, q) \mid (p, q) \in R\}.$$

Remarque: $R^2 = R$ et donc dans un calcul de $\mathcal{A}'$ il est inutile de prendre deux (ou plus) ε-transitions consécutives.

Lemme: $\mathcal{L}(\mathcal{A}') = \mathrm{Clot}(L)$.

$\subseteq$: Soit $w = a_1 \cdots a_n \in \mathcal{L}(\mathcal{A}')$. On a un calcul acceptant dans $\mathcal{A}'$

$$p_0 \xrightarrow{\varepsilon} q_0 \xrightarrow{a_1} p_1 \xrightarrow{\varepsilon} q_1 \cdots \xrightarrow{a_n} p_n \xrightarrow{\varepsilon} q_n$$

pour tout $0 \leq i \leq n$, il existe un mot $v_i \in L_{p_i,q_i} \cap D$. On obtient

$$v = v_0 a_1 v_1 \cdots a_n v_n \xrightarrow[*]{\mathrm{red}} w$$

Clôture et réduction.

Preuve : Reconnaissabilité de la clôture

$\supseteq$: Par induction sur le nombre n de réductions: Soit $v \xrightarrow[n]{\text{red}} w$ avec $v \in L$.

Si $n = 0$ alors $w = v \in L = \mathcal{L}(\mathcal{A}) \subseteq \mathcal{L}(\mathcal{A}')$.

Si $n > 0$ alors $v \xrightarrow[n-1]{\text{red}} w' \xrightarrow[1]{\text{red}} w$.

On peut écrire $w' = w_1 \overline{a} a w_2$ et $w = w_1 w_2$ avec $a \in \Gamma$.

Par induction, on a un calcul acceptant pour w' dans $\mathcal{A}'$. Ce calcul s'écrit

$$q_0 \xrightarrow{w_1} p \xrightarrow{\overline{a}} p' \xrightarrow{\varepsilon} q' \xrightarrow{a} q \xrightarrow{w_2} q_f .$$

$(p', q') \in R$ donc il existe $u \in L_{p',q'} \cap D$.

On en déduit $\overline{a} u a \in L_{p,q} \cap D$ et donc $(p, q) \in R$.

On obtient le calcul acceptant pour w dans $\mathcal{A}'$:

$$q_0 \xrightarrow{w_1} p \xrightarrow{\varepsilon} q \xrightarrow{w_2} q_f .$$

Configurations accessibles (Preuve)

Soit $\mathcal{A} = (Q, \Sigma, Z, T, q_0 z_0)$ un automate à pile.

On définit $\Gamma = Q \uplus Z$ et le langage fini $K = \{qh\overline{x}\,\overline{p} \mid \exists\ (px, a, qh) \in T\} \subseteq \widetilde{\Gamma}^+$.

Lemme : Soit $n \geq 0$

il existe un calcul $pg \xrightarrow[n]{} qh$ dans $\mathcal{A}$ ssi il existe $w \in K^n$ tel que $wpg \xrightarrow[2n]{\text{red}} qh$

Corollaire : $\mathcal{C}(pg) = \mathrm{Clot}(K^+pg) \cap QZ^*$.

Puisque K est un langage fini, le langage K^+pg est reconnaissable et on peut construire (PTIME) un automate $\mathcal{B}$ qui reconnaît $\mathrm{Clot}(K^+pg) \cap QZ^*$.

Preuve du Lemme

$\Longrightarrow$: par récurrence sur n. Soit $pg \xrightarrow[n]{} qh$ un calcul de $\mathcal{A}$.

Si $n = 0$ alors $qh = pg$ et on prend $w = \varepsilon \in K^0$.

Si $n > 0$ alors $g = xg'$ et le calcul s'écrit

$$pg = pxg' \xrightarrow[1]{a} p'h'g' \xrightarrow[n-1]{} qh \text{ avec } (px, a, p'h') \in T.$$

Par induction, il existe $w' \in K^{n-1}$ tel que $w'p'h'g' \xrightarrow[2(n-1)]{\text{red}} qh$.

Soit $w = w'(p'h'\overline{x}\,\overline{p}) \in K^n$. On a $wpg \xrightarrow[2n]{\text{red}} qh$.

Configurations accessibles (Preuve)

Preuve du Lemme

$\Longleftarrow$: par récurrence sur n. Soit $w \in K^n$ tel que $wpg \xrightarrow[2n]{\text{red}} qh$.

Si $n = 0$ alors $w = \varepsilon$ et $qh = pg$: on a $pg \xrightarrow[0]{} qh$ dans $\mathcal{A}$.

Si $n > 0$ alors $w = w'v$ avec $w' \in K^{n-1}$ et $v = p'h'\overline{x}\,\overline{r} \in K$.

Toutes les lettres barrées doivent être réduites à partir de $wpg = w'p'h'\overline{x}\,\overline{r}pg$.

On en déduit $r = p$ et $g = xg'$.

Donc, on a une transition $(px, a, p'h') \in T$ et une réduction $w'p'h'g' \xrightarrow[2(n-1)]{\text{red}} qh$.

Par induction, il existe un calcul $p'h'g' \xrightarrow[n-1]{} qh$ dans $\mathcal{A}$.

Finalement,

$$pg = pxg' \xrightarrow[1]{a} p'h'g' \xrightarrow[n-1]{} qh \text{ dans } \mathcal{A}.$$

Calculs d'accessibilité

Exercice : Calculs infinis

Soit $\mathcal{A} = (Q, \Sigma, Z, T, q_0 z_0)$ un automate à pile.
Montrer qu'on peut effectivement calculer les ensembles suivants :

1. $V = \{(p, x) \in Q \times Z \mid \exists\ px \xrightarrow[\omega]{}$ dans $\mathcal{T}\}$
2. $V' = \{(p, x) \in Q \times Z \mid \exists\ px \xrightarrow[\omega]{\varepsilon}$ dans $\mathcal{T}\}$

Preuve : Indications

On calcule l'ensemble V comme plus grand point fixe. Pour cela, on définit:

$$
\begin{aligned}
V_0 &= Q \times Z \\
V_{m+1} = \{(p, x) \in Q \times Z \mid\ & \exists (px, a, p_1 y_1 \cdots y_n) \in T, \\
& \exists 1 \leq k \leq n,\ \exists p_2, \ldots, p_k \in Q \text{ tels que} \\
& (p_k, y_k) \in V_m \text{ et } \forall 1 \leq i < k,\ (p_i, y_i, p_{i+1}) \in X\}
\end{aligned}
$$

Montrer alors que $(V_m)_{m \geq 0}$ est une suite décroissante et que

$$V = \bigcap_{m \geq 0} V_m$$

Langages déterministes

Définition : Automate à pile déterministe

$\mathcal{A} = (Q, \Sigma, Z, T, q_0 z_0, F)$ est *déterministe* si

- $\forall (pz, a) \in QZ \times (\Sigma \cup \{\varepsilon\}), \quad |T(pz, a)| \leq 1,$
- $\forall\, pz \in QZ, \quad T(pz, \varepsilon) \neq \emptyset \implies \forall\, a \in \Sigma,\ T(pz, a) = \emptyset$

Un langage $L \subseteq \Sigma^*$ est *déterministe* s'il existe un automate à pile déterministe qui accepte L par état final.

Exemples :

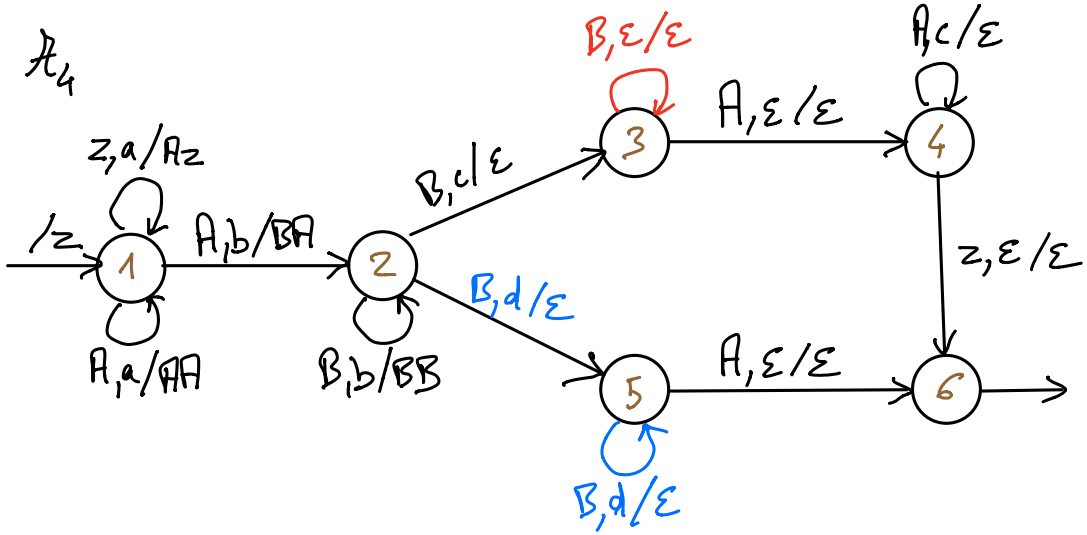
1. Le langage $L_4 = \{a^n b^p c^n \mid n, p > 0\} \cup \{a^n b^p d^p \mid n, p > 0\}$ est déterministe mais pas D+TR.
2. $L_5 = \{a^n b a^n \mid n > 0\}$ peut être accepté par un automate D+TR mais pas par un automate D+S car il n'est pas fermé par préfixe.

Exercices :

1. Montrer que D_n^* est D+TR mais pas D+S.
2. Montrer que le langage $\{a^n b^n \mid n > 0\} \cup \{a^n b^{2n} \mid n > 0\}$ est non ambigu mais pas déterministe.

Ex: $L_4 = \{ a^n b^p c^n \mid n,p>0 \} \cup \{ a^n b^p d^p \mid n,p>0 \}$

$\mathcal{A}_4$



Exemple de calcul

$1z \xrightarrow{a} 1Az \xrightarrow{a} 1AAz \xrightarrow{b} 2BAAz \xrightarrow{b} 2BBAAz \xrightarrow{c} 3BAAz$
$\xrightarrow{\varepsilon} 3AAz \xrightarrow{\varepsilon} 4Az \xrightarrow{c} 4z \xrightarrow{\varepsilon} 6$

$1z \xrightarrow{a} 1Az \xrightarrow{a} 1AAz \xrightarrow{b} 2BAAz \xrightarrow{b} 2BBAAz \xrightarrow{d} 5BAAz$
$\xrightarrow{d} 5AAz \xrightarrow{\varepsilon} 6Az$

L'automate $\mathcal{A}_4$ est déterministe mais pas temps-réel
On pourrait enlever les ε-transitions
$(3A, \varepsilon, 4)$, $(4z, \varepsilon, 6)$ et $(5A, \varepsilon, 6)$
mais pas $(3B, \varepsilon, 3)$
Il n'y a pas d'automate D+TR pour L_4
Donc D+TR $\subsetneq$ D

Acceptation par pile vide

Exemples :

1. Le langage $L_5 = \{a^n b a^n \mid n \geq 0\}$ peut être accepté par *pile vide* par un automate D+TR+S.
2. Le langage $L_4 = \{a^n b^p c^n \mid n, p > 0\} \cup \{a^n b^p d^p \mid n, p > 0\}$ peut être accepté par *pile vide* par un automate D.

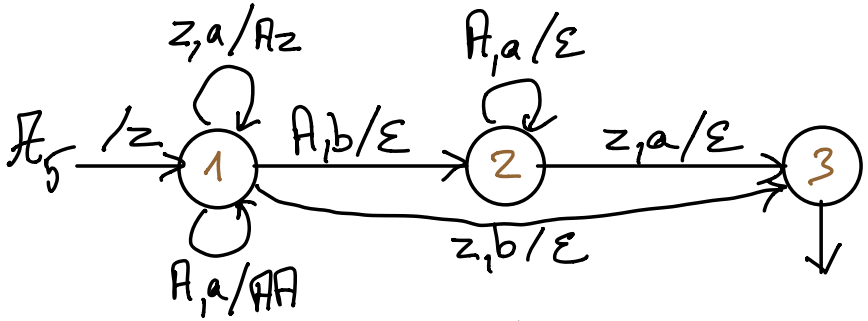
Exercices :

1. Montrer qu'un langage L est déterministe et préfixe ($L \cap L\Sigma^+ = \emptyset$) ssi il existe un automate déterministe qui accepte L par pile vide.
2. Montrer que pour les automates à pile déterministes, l'acceptation par pile vide est équivalente à l'acceptation par pile vide ET état final.

Exercice :

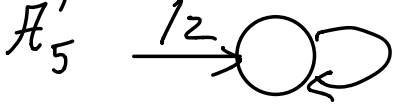
Montrer que D_n^* peut être accepté par sommet de pile par un automate D+TR+S.

Ex: $L_5 = \{ a^n b a^n \mid n \geq 0 \}$



$\mathcal{A}_5$ est D+TR

Mais pas d'automate simple avec acceptation par état final car L_5 n'est pas fermé par préfixe.



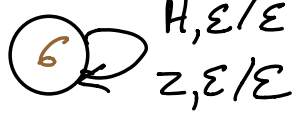
z,a/zA
z,b/ε
A,a/ε

$\mathcal{A}'_5$ est D+TR+S et accepte L_5 par pile vide.

Exemple de calcul sur $a^2 b a^2$

$z \xrightarrow{a} zA \xrightarrow{a} zAA \xrightarrow{a} zAAA \xrightarrow{b} AAA \xrightarrow{a} AA \xrightarrow{a} A \xrightarrow{a} \varepsilon$

Ex L_4 est reconnu par pile vide par l'automate $\mathcal{A}'_4$ obtenu en ajoutant à $\mathcal{A}_4$ les transitions $(6, A, \varepsilon, 6)$ et $(6, z, \varepsilon, 6)$



Complémentaire

Théorème : Les déterministes sont fermés par complémentaire.

Soit $\mathcal{A} = (Q, \Sigma, Z, T, q_0 z_0, F)$ un automate à pile déterministe, on peut effectivement construire un automate à pile déterministe $\mathcal{A}'$ qui reconnaît $\Sigma^* \setminus \mathcal{L}(\mathcal{A})$.

Il y a deux difficultés principales :

1. Un automate déterministe peut se bloquer (deadlock) ou entrer dans un ε-calcul infini (livelock). Dans ce cas il y a des mots qui n'admettent aucun calcul dans l'automate.
2. Même avec un automate déterministe, un mot peut avoir plusieurs calculs (ε-transitions à la fin) certains réussis et d'autres non.

Blocage

Définition : Blocage

Un automate à pile $\mathcal{A} = (Q, \Sigma, Z, T, q_0 z_0)$ est sans blocage si pour toute configuration accessible $p\alpha$ et pour toute lettre $a \in \Sigma$ il existe un calcul $p\alpha \xrightarrow[*]{\varepsilon} \xrightarrow{a}$.

Proposition : Critère d'absence de blocage

Un automate *déterministe* est sans blocage si et seulement si pour toute configuration accessible $p\alpha$ on a

1. $\alpha \neq \varepsilon$, et donc on peut écrire $\alpha = x\beta$ avec $x \in Z$,
2. $px \xrightarrow{\varepsilon}$ ou $\forall a \in \Sigma$, $px \xrightarrow{a}$,
3. $px \not\xrightarrow[\omega]{\varepsilon}$.

De plus, ce critère est décidable.

Remarque :

Si $\mathcal{A}$ est sans blocage alors chaque mot $w \in \Sigma^*$ a un unique calcul maximal (et fini) $q_0 z_0 \xrightarrow[*]{w} p\alpha \not\xrightarrow{\varepsilon}$ dans $\mathcal{A}$ (avec $\alpha \neq \varepsilon$).

Blocage

Preuve : Critère d'absence de blocage

Si $\mathcal{A}$ est sans blocage alors il satisfait clairement le critère.

Réciproquement, supposons que $\mathcal{A}$ satisfait le critère.

On calcule $X' = \{(r, x, s) \in Q \times Z \times Q \mid \exists\, rx \xrightarrow[+]{\varepsilon} s \text{ dans } \mathcal{T}\}$.

Soit $p\alpha$ une configuration accessible et $a \in \Sigma$.

On écrit $\alpha = x_1 x_2 \cdots x_n$ avec $x_i \in Z$.

Il n'existe pas de calcul $p\alpha \xrightarrow[+]{\varepsilon} p'$ car sinon la configuration p' est accessible ce qui contredit le point 1 du critère.

Donc il existe $1 \leq k \leq n$ et des états $p_2, \ldots, p_k$ tels que

$$(p, x_1, p_2), (p_2, x_3, p_3), \ldots, (p_{k-1}, x_{k-1}, p_k) \in X'$$

et il n'existe pas p_{k+1} avec $(p_k, x_k, p_{k+1}) \in X'$.

La configuration $p_k x_k \cdots x_n$ est accessible, donc $p_k x_k \not\xrightarrow[\omega]{\varepsilon}$ (critère point 3).

On en déduit $p_k x_k \xrightarrow[*]{\varepsilon} qy\beta$ avec $qy \not\xrightarrow{\varepsilon}$ et donc $qy \xrightarrow{a}$ (critère point 2).

Finalement, $$p\alpha \xrightarrow[*]{\varepsilon} p_k x_k \cdots x_n \xrightarrow[*]{\varepsilon} qy\beta x_{k+1} \cdots x_n \xrightarrow{a}$$

Décidabilité du critère: on vérifie que $\mathcal{C}(q_0 z_0) \cap Q = \emptyset$ (point 1) et pour chaque px tel que $px = q_0 z_0$ ou $\mathcal{C}(q_0 z_0) \cap pxZ^* \neq \emptyset$, on vérifie le point 2 du critère et que $(p, x) \notin V'$ (point 3).

Blocage

Proposition : Suppression des blocages

Soit $\mathcal{A} = (Q, \Sigma, Z, T, q_0 z_0, F)$ un automate à pile déterministe, on peut effectivement construire un automate à pile déterministe *sans blocage* $\mathcal{A}' = (Q', \Sigma, Z', T', q_0' z_0', F')$ qui reconnaît le même langage.

Preuve

$Q' = Q \uplus \{q_0', d, f\}$, $F' = F \uplus \{f\}$, $Z' = Z \uplus \{\bot\}$, $z_0' = \bot$ et pour $p \in Q$, $a \in \Sigma$ et $x \in Z$

1. $q_0' \bot \xrightarrow{\varepsilon} q_0 z_0 \bot$,
2. Si $px \xrightarrow{a} q\alpha \in T$ alors $px \xrightarrow{a} q\alpha \in T'$,
3. Si $px \overset{q}{\not\to}$ et $px \overset{\varepsilon}{\not\to}$ dans $\mathcal{A}$ alors $px \xrightarrow{a} dx \in T'$,
4. Si $px \xrightarrow[\omega]{\varepsilon}$ dans $\mathcal{A}$ et $px \xrightarrow{\varepsilon} q\alpha \in T$ alors $px \xrightarrow{\varepsilon} q\alpha \in T'$,
5. Si $px \xrightarrow[\omega]{\varepsilon}$ dans $\mathcal{A}$ et $\exists\ px \xrightarrow[*]{\varepsilon} q\alpha$ avec $q \in F$ alors $px \xrightarrow{\varepsilon} fx \in T'$,
6. Si $px \xrightarrow[\omega]{\varepsilon}$ dans $\mathcal{A}$ et $\forall\ px \xrightarrow[*]{\varepsilon} q\alpha$ on a $q \notin F$ alors $px \xrightarrow{\varepsilon} dx \in T'$,
7. $p\bot \xrightarrow{\varepsilon} d\bot$, $d\bot \xrightarrow{a} d\bot$, $dx \xrightarrow{a} dx$ et $fx \xrightarrow{a} dx$.

Cette construction est effective.

Blocage

Preuve : $\mathcal{A}'$ est déterministe, sans blocage, constructible en PTIME

Déterministe: il suffit de le vérifier.

Sans blocage: on vérifie les conditions du critère.

1. $\mathcal{C}'(q'_0\bot) \cap Q = \emptyset$: on ne peut pas vider entièrement la pile car le fond de pile $\bot$ n'est jamais effacé.

2. Soit $px\alpha$ une configuration accessible. Supposons $px \overset{\varepsilon}{\not\to}$ dans $\mathcal{A}'$.

Si $p \in Q$ alors $x \neq \bot$ (7) et donc (2,3) $\forall\, a \in \Sigma$, $px \xrightarrow{a}$

On a $p \neq q'_0$ et si $p \in \{d, f\}$ alors (7) $\forall\, a \in \Sigma$, $px \xrightarrow{a}$ (notons que $px \neq f\bot$).

3. Soit $px\alpha$ une configuration accessible. Supposons $px \xrightarrow[\omega]{\varepsilon}$ dans $\mathcal{A}'$.

Ce calcul ne peut pas utiliser les états d et f: il n'utilise que les transitions (1,4).

La transition (1) est utilisée au plus une fois.

Si $px \neq q'_0\bot$ alors $px \xrightarrow[\omega]{\varepsilon}$ dans $\mathcal{A}$ et on ne peut pas utiliser la transition (4).

Si $px = q'_0\bot$ alors $q_0 z_0 \xrightarrow[\omega]{\varepsilon}$ dans $\mathcal{A}$ et la transition (4) est impossible aussi.

Constructible en PTIME:

Pour les conditions des transitions (4,5,6) on utilise les ensembles V', X' et W'.

Blocage

Preuve : $\mathcal{L}(\mathcal{A}) \subseteq \mathcal{L}(\mathcal{A}')$

Soit $q_0 z_0 \xrightarrow{w} q\beta$ un calcul acceptant de $\mathcal{A}$ pour $w \in \Sigma^*$.

Si le calcul ne passe pas par une configuration $px\alpha$ avec $px \xrightarrow[\omega]{\varepsilon} ((p,x) \in V')$ alors

$$q'_0\bot \xrightarrow[(1)]{\varepsilon} q_0 z_0 \bot \xrightarrow[(2,4)]{w} q\beta\bot$$

est un calcul acceptant de $\mathcal{A}'$.

Sinon, le calcul s'écrit $q_0 z_0 \xrightarrow{w} px\gamma \xrightarrow[*]{\varepsilon} q\beta$ avec $px \xrightarrow[\omega]{\varepsilon}$

Donc $\beta = \alpha\gamma$ avec $px \xrightarrow[*]{\varepsilon} q\alpha$ et $q \in F$. On obtient un calcul acceptant dans $\mathcal{A}'$:

$$q'_0\bot \xrightarrow[(1)]{\varepsilon} q_0 z_0 \bot \xrightarrow[(2,4)]{w} px\gamma\bot \xrightarrow[(5)]{\varepsilon} fx\gamma\bot$$

Preuve : $\mathcal{L}(\mathcal{A}') \subseteq \mathcal{L}(\mathcal{A})$

Un calcul acceptant de $\mathcal{A}'$ ne peut pas atteindre l'état puits d. Deux possibilités:

Si le calcul accepte grâce à l'état f, il s'écrit:

$$q'_0\bot \xrightarrow[(1)]{\varepsilon} q_0 z_0 \bot \xrightarrow[(2,4)]{w} px\gamma\bot \xrightarrow[(5)]{\varepsilon} fx\gamma\bot$$

On en déduit $q_0 z_0 \xrightarrow{w} px\gamma$ dans $\mathcal{A}$ et il existe $px \xrightarrow[*]{\varepsilon} q\alpha$ dans $\mathcal{A}$ avec $q \in F$.

Donc w est accepté par le calcul $q_0 z_0 \xrightarrow{w} px\gamma \xrightarrow[*]{\varepsilon} q\alpha\gamma$ de $\mathcal{A}$.

Sinon, le calcul de $\mathcal{A}'$ s'écrit $q'_0\bot \xrightarrow[(1)]{\varepsilon} q_0 z_0 \bot \xrightarrow[(2,4)]{w} q\beta\bot$ avec $q \in F$

et on obtient le calcul acceptant $q_0 z_0 \xrightarrow{w} q\beta$ dans $\mathcal{A}$ pour w.

Complémentaire

Proposition :

Soit $\mathcal{A} = (Q, \Sigma, Z, T, q_0 z_0, F)$ un automate à pile déterministe, on peut effectivement construire un automate à pile déterministe $\mathcal{A}'$ qui reconnaît $\Sigma^* \setminus \mathcal{L}(\mathcal{A})$.

Proposition :

Soit $\mathcal{A} = (Q, \Sigma, Z, T, q_0 z_0, F)$ un automate à pile déterministe, on peut effectivement construire un automate à pile déterministe équivalent $\mathcal{A}'$ tel qu'on ne puisse pas faire d'ε-transition à partir d'un état final de $\mathcal{A}'$.

Exercice :

Montrer que tout langage déterministe est non ambigu.

Complémentaire

Preuve : Complémentaire

On suppose $\mathcal{A}$ déterministe et sans blocage. On pose $Q' = Q \times \{1,2,3,4\}$.
L'état initial est $q'_0 = (q_0, 1)$ si $q_0 \notin F$ et $q'_0 = (q_0, 2)$ sinon.

1. Si $px \xrightarrow{\varepsilon} q\alpha \in T$ et $i \in \{1,2\}$ alors
$$(p,i)x \xrightarrow{\varepsilon} (q,j)\alpha \in T' \text{ avec } j = \begin{cases} 1 & \text{si } i = 1 \wedge q \notin F \\ 2 & \text{sinon.} \end{cases}$$

2. Si $px \xrightarrow{a} q\alpha \in T$ et $i \in \{1,2\}$ alors $(p,i)x \xrightarrow{\varepsilon} (p,i+2)x \in T'$ et
$$(p,i+2)x \xrightarrow{a} (q,j)\alpha \in T' \text{ avec } j = \begin{cases} 1 & \text{si } q \notin F \\ 2 & \text{sinon.} \end{cases}$$

L'automate $\mathcal{A}'$ est déterministe et sans blocage.
Soit $w \in \Sigma^*$ et $q_0 z_0 \xrightarrow{w} p\alpha$ l'unique calcul *maximal* de $\mathcal{A}$ sur w. On a $p\alpha \not\xrightarrow{\varepsilon}$.
L'unique calcul *maximal* de $\mathcal{A}'$ sur w s'écrit $q'_0 z_0 \xrightarrow{w} (p,j)\alpha$ avec $j = 3$ si le calcul de $\mathcal{A}$ n'a pas visité F depuis la dernière transition visible ($\neq \varepsilon$), et $j = 4$ sinon.

- Avec $F' = Q \times \{3\}$ on obtient $\mathcal{L}(\mathcal{A}') = \overline{\mathcal{L}(\mathcal{A})}$.
- Avec $F' = Q \times \{4\}$ on obtient $\mathcal{L}(\mathcal{A}') = \mathcal{L}(\mathcal{A})$.

Dans les deux cas, à partir d'un état de F' il n'y a pas d'ε-transition.
De plus, chaque mot $w \in \mathcal{L}(\mathcal{A}')$ a un unique calcul acceptant dans $\mathcal{A}'$.

Langages déterministes

Exercice :

Soit $\mathcal{A} = (Q, \Sigma, Z, T, q_0 z_0, K)$ un automate à pile déterministe avec acceptation généralisée par le langage rationnel $K \subseteq QZ^*$.
Montrer qu'on peut effectivement construire un automate à pile déterministe équivalent reconnaissant par état final.

Exercice :

Soit $\mathcal{A}$ un automate à pile déterministe. Montrer qu'on peut effectivement construire un automate à pile déterministe qui reconnaît le même langage et dont les ε-transitions sont uniquement effaçantes : $px \xrightarrow{\varepsilon} q$.

Lemme d'itération pour les déterministes

Lemme : Itération

Soit $L \subseteq \Sigma^*$ un langage déterministe. Il existe un entier $N \in \mathbb{N}$ tel que tout mot $w \in L$ contenant au moins N lettres distinguées se factorise en $w = \alpha u \beta v \gamma$ avec

1. $\forall p \geq 0 : w = \alpha u^p \beta v^p \gamma \in \mathcal{L}(\mathcal{A})$,
2. $u\beta v$ contient moins de N lettres distinguées,
3. soit α, u, β soit β, v, γ contiennent des lettres distiguées,
4. pour tout $\gamma' \in \Sigma^*$,

$$\exists p : \alpha u^p \beta v^p \gamma' \in L \quad \Longrightarrow \quad \forall p : \alpha u^p \beta v^p \gamma' \in L$$

Preuve Admis. Voir [4]

Langages déterministes

Proposition : Décidabilité et indécidabilité

On ne peut pas décider si un langage algébrique est déterministe.

Soient L, L' deux langages déterministes et R un langage rationnel.
Les problèmes suivants sont décidables :

- $R \subseteq L$? $\qquad R \subseteq L \iff R \cap \overline{L} = \emptyset$
- $L = R$? $\qquad R = L \iff R \cap \overline{L} = \emptyset = \overline{R} \cap L$
- L est-il rationnel ?
- $L = L'$? $\qquad$ Géraud Sénizergues, prix Gödel 2002

Les problèmes suivants sont indécidables :

- $L \cap L' = \emptyset$?
- $L \subseteq L'$?
- $L \cap L'$ est-il algébrique ?
- $L \cap L'$ est-il déterministe ?
- $L \cup L'$ est-il déterministe ?

Plan

Plan

Plan