LF – Théorie des langages formels

Sylvain Brandel

2023 – 2024

sylvain.brandel@univ-lyon1.fr

CM 9

AUTOMATES À PILE ALGÉBRICITÉ

Automates à pile

- Un automate à pile est un sextuplet $M = (K, \Sigma, \Gamma, \delta, s, F)$ avec :
  - K : ensemble fini d'états
  - Σ : ensemble fini de symboles d'entrée (alphabet)
  - Γ : ensemble fini de symboles de la pile
  - $s \in K$: état initial
  - $F \subset K$: ensemble des états finaux (acceptants)
  - $\Delta \subset (K \times (\Sigma \cup \{\varepsilon\}) \times (\Gamma \cup \{\varepsilon\})) \times (K \times (\Gamma \cup \{\varepsilon\}))$: relation de transition

Automates à pile

- Une transition $((q, \sigma, \gamma), (q', \gamma')) \in \Delta$ où :
  - q est l'état courant
  - σ est le symbole d'entrée courant
  - γ est le symbole sommet de la pile
  - q' est le nouvel état
  - γ' est le nouveau symbole en sommet de pile

  a pour effet :

  (1) De passer de l'état q à l'état q'

  (2) D'avancer la tête de lecture après σ

  (3) De dépiler γ du sommet de la pile

  (4) D'empiler γ' sur la pile

Automates à pile

- Soit $M = (K, \Sigma, \Gamma, \delta, s, F)$ un automate à pile. Une configuration de M est définie par un triplet $(q, w, z) \in K \times \Sigma^* \times \Gamma^*$ où :
  - q est l'état courant de M
  - w est la partie de la chaîne restant à analyser
  - z est le contenu de la pile

- Soient (q, w, z) et (q', w', z') deux configurations d'un automate à pile M. On dit qu'on passe de (q, w, z) à (q', w', z') en une étape

  ssi $\exists\ \sigma \in \Gamma \cup \{\varepsilon\}$ et $\gamma, \gamma' \in \Gamma \cup \{\varepsilon\}$

  $\text{tels que}\ \ w = \sigma w'$

  et $z = \gamma z'', z' = \gamma' z''$ avec $z'' \in \Gamma^*$

  et $((q, \sigma, \gamma), (q', \gamma')) \in \Delta$

  On note $(q, w, z) \vdash_M (q', w', z')$

Automates à pile

- La relation $\vdash_M^*$ est la fermeture réflexive transitive de $\vdash_M$

- Soit $M = (K, \Sigma, \Gamma, \delta, s, F)$ un automate à pile

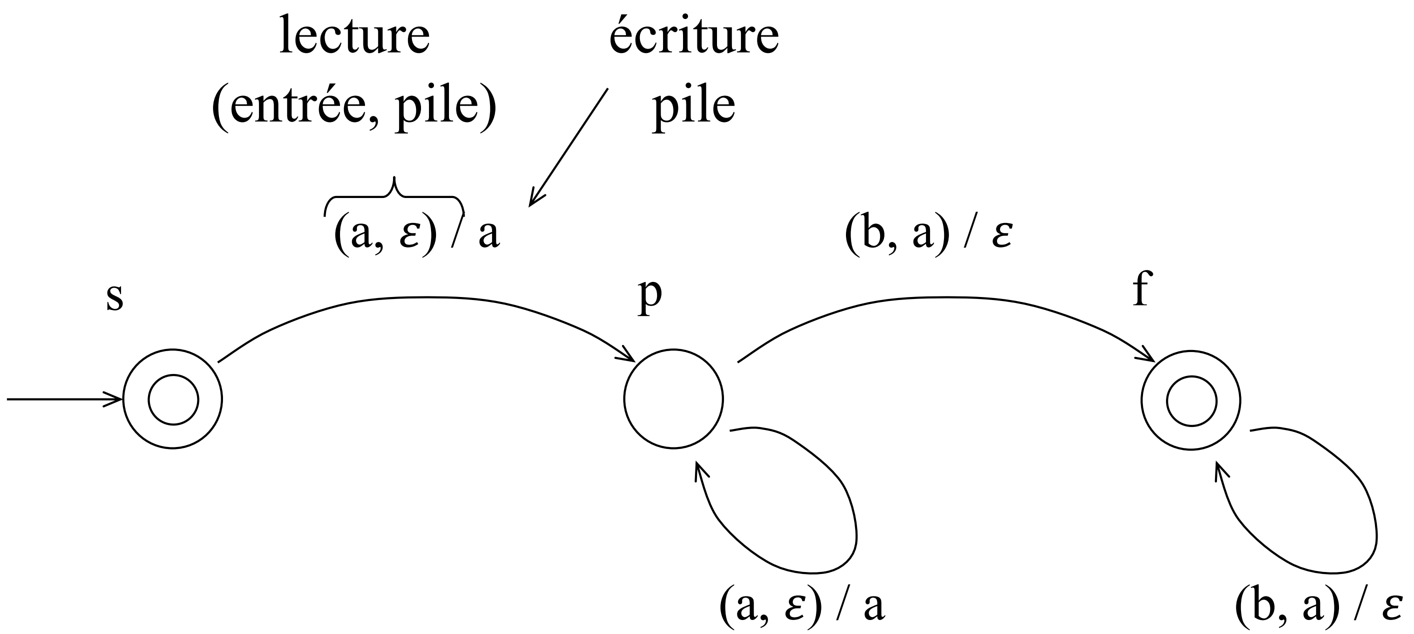
> Un mot $w \in \Sigma^*$ est accepté par M ssi $(s, w, \varepsilon) \vdash_M^* (f, \varepsilon, \varepsilon)$ avec $f \in F$

- Le langage accepté par M, noté $L(M)$, est l'ensemble des mots acceptés par M

Automates à pile

- Exemple : Soit $M = (K, \Sigma, \Gamma, \delta, s, F)$ avec :
  - K = {s, p, f}
  - Σ = {a, b}
  - Γ = {a, b}
  - F = {s, f}

Δ = { ((s, a, ε), (p, a)),
((p, a, ε), (p, a)),
((p, b, a), (f, ε)),
((f, b, a), (f, ε)) }



Automates à pile

- Un automate à pile est déterministe s'il y a au plus une transition applicable pour tout triplet de la forme

  (État courant, symbole d'entrée, sommet de pile).

- Les automates à pile non déterministes reconnaissent plus de langages que les automates à pile déterministes

Automates à pile et grammaires algébriques

- Théorème

  La classe des langages acceptés par les automates à pile est égale à la classe des langages engendrés par les grammaires algébriques

- Un automate à pile est dit simple ssi quelle que soit la transition $((q, \sigma, \gamma), (q', \gamma')) \in \Delta$, on a :

  $\gamma \in \Gamma$ (sauf pour $q = s$ où on ne dépile rien)

  et $|\gamma'| \leq 2$

- Proposition

  On peut transformer tout automate à pile en un automate simple équivalent

Propriétés des langages algébriques

Preuve d'algébricité

- Pour montrer qu'un langage est algébrique, on peut :
  - Soit définir une grammaire algébrique qui engendre ce langage
  - Soit définir un automate à pile qui l'accepte

- Il est également possible d'utiliser les propriétés de stabilité de la classe des langages algébriques

Propriétés des langages algébriques
Propriétés de stabilité

- Théorème

  La classe des langages algébriques est stable par les opérations d'union, de concaténation et d'étoile de Kleene

- Preuve

  Soient deux grammaires $G_1 = (V_1, \Sigma_1, R_1, S_1)$ et $G_2 = (V_2, \Sigma_2, R_2, S_2)$, avec $V_1 \cap V_2 = \varnothing$ (on renomme éventuellement les non-terminaux)

  La preuve (constructive) consiste à :

  - Construire une grammaire G à partir de G_1 et G_2 validant les propriétés de stabilité
  - Montrer que $L(G) = L(G_1)$ op $L(G_2)$ (op $\in \{\cup, .\}$) et $L(G) = L(G_1)^*$

Propriétés des langages algébriques

Propriétés de stabilité

- Preuve

(a) Union

Soit $G = (V, \Sigma, R, S)$ avec :

- $V = V_1 \cup V_2 \cup \{S\}$ où $S \notin V_1 \cup V_2$ (renommage éventuel)
- $\Sigma = \Sigma_1 \cup \Sigma_2$
- $R = R_1 \cup R_2 \cup \{S \rightarrow S_1 \mid S_2\}$

(b) Concaténation

Soit $G = (V, \Sigma, R, S)$ avec :

- $V = V_1 \cup V_2 \cup \{S\}$ où $S \notin V_1 \cup V_2$ (renommage éventuel)
- $\Sigma = \Sigma_1 \cup \Sigma_2$
- $R = R_1 \cup R_2 \cup \{S \rightarrow S_1S_2\}$

(c) Opération étoile

Soit $G = (V, \Sigma, R, S)$ avec :

- $V = V_1 \cup \{S\}$ où $S \notin V_1$ (renommage éventuel)
- $\Sigma = \Sigma_1$
- $R = R_1 \cup \{S \rightarrow S_1S \mid \varepsilon\}$

Propriétés des langages algébriques
Propriétés de stabilité

- Contrairement à la classe des langages rationnels, la classe des langages algébriques n'est pas stable par intersection et complémentation

- Théorème

  L'intersection d'un langage rationnel et d'un langage algébrique est algébrique

Propriétés des langages algébriques

Lemme de l'étoile pour les langages algébriques

- Définition

  Une grammaire algébrique $G = (V, \Sigma, R, S)$ est sous forme normale de Chomsky si chaque règle est de la forme :

  $A \rightarrow BC$ avec $B, C \in V - \{S\}$

  ou $A \rightarrow \sigma$ avec $\sigma \in \Sigma$

  ou $A \rightarrow e$

- Théorème

  Pour toute grammaire algébrique, il existe une grammaire sous forme normale de Chomsky équivalente

Propriétés des langages algébriques

Lemme de l'étoile pour les langages algébriques

- Théorème (lemme de la double étoile)

  Soit L un langage algébrique

  Il existe un nombre k, dépendant de L, tel que tout mot $z \in L$, $|z| \geq k$, peut être décomposé en $z = uvwxy$ avec :

  (*i*) $|vwx| \leq k$

  (*ii*) $|v| + |x| > 0$ (ie. $v \neq \varepsilon$ ou $x \neq \varepsilon$)

  (*iii*) $uv^iwx^iy \in L$, $\forall\ i \geq 0$

  (d'où l'appellation de double étoile : v^i et x^i = v^* et x^*)

Propriétés des langages algébriques
Lemme de l'étoile pour les langages algébriques

- Lemme

  Soit G = (V, Σ, R, S) une grammaire algébrique sous forme normale de Chomsky

  Soit $S \Rightarrow_G^* w$ une dérivation de $w \in \Sigma^*$ dont l'arbre de dérivation est noté T

  Si la hauteur de T est n alors $|w| \leq 2^{n-1}$

- Corollaire

  Soit G = (V, Σ, R, S) une grammaire algébrique sous forme normale de Chomsky

  Soit $S \Rightarrow_G^* w$ une dérivation de $w \in L(G)$

  Si $|w| \geq 2^n$ alors l'arbre de dérivation est de hauteur $\geq n+1$

Propriétés des langages algébriques

Lemme de l'étoile pour les langages algébriques

- Exemple

Montrons que $L = \{ a^n b^n c^n \mid n \geq 0\}$ est non algébrique

Supposons que L est algébrique

D'après le lemme de la double étoile, il existe une constante k, dépendant de L, telle que :

$\forall$ $z \in L$, $|z| \geq k$, z peut être décomposé en $z = uvwxy$ avec :

(*i*) $|vwx| \leq k$

(*ii*) $|v| + |x| > 0$ (au moins un des deux n'est pas le mot vide)

(*iii*) $uv^iwx^iy \in L$, $\forall\, i \geq 0$

Propriétés des langages algébriques

Lemme de l'étoile pour les langages algébriques

- Exemple

  Considérons la chaîne particulière $z_0 = a^k b^k c^k$.
  On a bien $z_0 \in L$ et $|z_0| = 3k \geq k$.
  Les décompositions de $z_0 = uvwxy$ satisfaisant $|vwx| \leq k$ et $|v| + |x| > 0$ sont telles que :

  – Soit l'une des sous-chaînes v ou x contient plus d'un type de symbole, de la forme a^+b^+ ou b^+c^+.
    uv^iwx^iy avec $i > 1$ contient un a après un b ou un b après un c.
    (par exemple uv^2wx^2y = u aabb aabb w x x y, si v = aabb)
    donc la chaîne uv^iwx^iy n'est plus de la forme $a^pb^pc^p$ avec $p \geq 0$,
    donc $uv^iwx^iy \notin L$ pour $i > 1$.

  – Soit v et x sont des sous-chaînes de a^k ou de b^k ou de c^k.
    Comme au plus une des chaînes v ou x est vide, toute chaîne de la forme uv^iwx^iy avec $i > 1$ est caractérisée par une augmentation de un ($v = \varepsilon$ ou $x = \varepsilon$) ou deux ($v \neq \varepsilon$ et $x \neq \varepsilon$) des trois types de terminaux.
    donc pour $i > 1$, la chaîne uv^iwx^iy est de la forme $a^pb^qc^r$ mais avec $p \neq q$ ou $q \neq r$.
    donc $uv^iwx^iy \notin L$ pour $i > 1$.

  – Pas d'autres possibilités pour v et x, les autres sous-chaines u, w et y n'influencent pas.

  Pour toutes les décompositions possibles de la chaîne z_0 il y a une contradiction.

  Donc l'hypothèse est fausse $\Rightarrow$ L non algébrique.

Propriétés des langages algébriques

Preuve de non algébricité

- Pour montrer qu'un langage est non algébrique, on peut utiliser :
  - Le lemme de la double étoile
  - Les propriétés de stabilité de la classe des langages algébriques
  - Le théorème qui dit que l'intersection d'un langage algébrique et d'un langage rationnel est algébrique

Notion de décidabilité

- Une question est décidable s'il existe un algorithme (c'est-à-dire un processus déterministe) qui s'arrête avec une réponse (oui ou non) pour chaque entrée

- Une question est indécidable si un tel algorithme n'existe pas

Problèmes indécidables pour les langages algébriques

- Théorème

  Les questions suivantes sont décidables :

  – Étant donnés une grammaire algébrique G et un mot w est-ce que $w \in L(G)$?

  – Étant donnée une grammaire algébrique G, est-ce que $L(G) = \emptyset$?

  Les questions suivantes sont indécidables :

  – Soit G une grammaire algébrique. Est-ce que $L(G) = \Sigma^*$?

  – Soient G_1 et G_2 deux grammaires algébriques. Est-ce que $L(G_1) = L(G_2)$?

  – Soient M_1 et M_2 deux automates à pile. Est-ce que $L(M_1) = L(M_2)$?

  – Soit M un automate à pile. Trouver un automate à pile équivalent minimal en nombre d'états.