Limites des langages algébriques

Tout est algébrique?

- On connaît bien les langages algébriques
  - Plusieurs façons de les caractériser
  - Les opérations qui préservent l'algébricité
    - Union, concaténation et étoile
- Tout langage est-il algébrique ?
  - Comme pour les rationnels, il existe des langages non algébriques.
  - Pour le prouver, on utilise un argument de diagonalisation.

Hypothèse: tout est algébrique

- algébriques = reconnaissables; en bijection avec AP et **N**.
- On énumère les AP sur un alphabet à une lettre et on les ordonne dans une liste; L_0 est reconnu par le 1er AP de la liste, L_1 par 2e…
- Si M était dans T, il existerait un indice j tel que $M=L_j$. Puisque $M=L_j$, si
  - $j \in L_j$ alors, par définition de M, $j \notin M$
  - $j \notin L_j$ alors, par définition de M, $j \in M$
- Une contradiction dans les deux cas

L'ensemble des algébriques est infini mais dénombrable

	L_0	L_1	L_2	L_3
mot_0	O	N	N	O
mot_1	N	N	O	N
mot_2	O	N	O	O

T[i,j]=Oui si $i \in L_j$
Non sinon

$i \in M \Leftrightarrow i \notin L_i$
M n'est pas dans T

Question

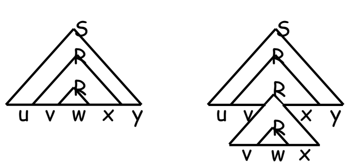
- Comment montrer qu'un langage L donné **n'est pas** algébrique.
- **Problème**

  Donnée : L un langage

  Question : L est-il non algébrique?
- Pour les rationnels, on a utilisé le principe des tiroirs sur les états parcourus pour montrer qu'on passe plusieurs fois par le même état

Question

- Pour les algébriques, on utilise le principe des tiroirs sur l'arbre syntaxique qui doit contenir plusieurs fois le même sous-arbre car les variables sont en nombre fini.



Lemme de la pompe

Lemme : Soit L un langage algébrique. Il existe une constante n (qui ne dépend que de L) telle que si $z \in L$, $|z| \geq n$, z se factorise en z=uvwxy tel que

i. $|vx|>0$ et

ii. $|vwx| \leq n$ et

iii. $\forall i \geq 0,\ uv^iwx^iy \in L$

Utilisation du Lemme de la pompe

- Comme pour les rationnels, ce lemme ne sert qu'à montrer la **non algébricité** d'un langage.
- On utilise la contraposée, en supposant L algébrique et on cherche une contradiction
- Si, pour un $w \in L$ quelconque de longueur suffisante $\forall$ décomposition w=uvwxy vérifiant
  1) $|vx|>0$ et
  2) $|vwx| \leq n$ alors
  3) $\exists i \geq 0$, $uv^iwx^iy \notin L$

On conclut que L n'est pas algébrique

Exemple $L=\{a^ib^ic^i : i>0\}$

- On suppose L algébrique et on fixe n.
- Soit $z=a^nb^nc^n$=uvwxy
- $|vwx| \leq n \Rightarrow$ vx ne peut avoir à la fois des a et c
  - v et x ne contiennent que des a $\Rightarrow$ uwy manque de a
  - v et x ne contiennent que des b $\Rightarrow$ uwy manque de b
  - v et x ne contiennent que des c $\Rightarrow$ uwy manque de b
  - vx contient des a et b $\Rightarrow$ uwy manque de a et de b
  - vx contient des b et c $\Rightarrow$ uwy manque de b et de c
- Pour chaque factorisation, on aboutit à une contradiction (le mot uwy n'est pas dans le langage). On en déduit donc que L n'est pas algébrique.

Autre exemple : $L=\{a^ib^jc^id^j : i,j \geq 1\}$

- On suppose L algébrique; soit n la constante du lemme
- On choisit $z=a^nb^nc^nd^n$
  1) $|vx|>0$ et
  2) $|vwx| \leq n$ et
  3) $\exists i \geq 0$, $uv^iwx^iy \notin L$
- Par la condition 2, vx contient soit
  - Qu'une seule lettre a ou b ou c ou d
    uwy n'est pas dans le langage, car cette lettre manque
  - Que des a et des b
  - Que des b et des c
  - Que des c et des d

  } cas analogues

  uwy n'est pas dans le langage, car deux lettres manquent

Mais…

- Ca ne marche pas à tous les coups! Exemple de

$L=\{a^ib^jc^kd^l : i=0 \text{ ou } j=k=l\}$

Qui n'est pas algébrique mais, en choisissant

$z=b^jc^kd^l$ ou $z=a^ib^jc^jd^j$

et en factorisant z=uvwxy, on peut toujours trouver une décomposition pour laquelle $uv^mwx^my \in L$ (c.a.d. vwx ne contient que des b ou que des a)

- Le lemme n'est pas utilisable avec un langage de cette forme
- Comment faire?

Lemme d'Ogden

- Il faut une version plus forte du lemme de la pompe.
- C'est le lemme d'Ogden qui permet de marquer des positions et d'ajouter à v et x des conditions sur le marquage
- **Lemme** : Soit L un langage algébrique. Il existe une constante n qui ne dépend que de L telle que $\forall z \in L$ avec au moins n lettres marquées, z se factorise en z=uvwxy tel que
  1) v et x ont ensemble au moins une position marquée et
  2) vwx a au plus n positions marquées et
  3) $\forall i \geq 0$, $uv^iwx^iy \in L$

Exemple $L=\{a^ib^jc^k : i \neq j,\ j \neq k \text{ et } k \neq l\}$

- On suppose L algébrique; soit n la constante du lemme d 'Ogden
- On choisit le mot $z=\underline{a}^nb^{n+n!}c^{n+2n!}$ avec les a marqués
  1) vx a au moins une position marquée et
  2) vwx a au plus n positions marquées et
  3) $\exists i \geq 0$, $uv^iwx^iy \notin L$
- Cas 1 : v (ou x) a deux symboles distincts (a et b). Pour i=2, $uv^iwx^iy \notin L$ car on a (ab)(ab)

Exemple $L=\{a^ib^jc^k: i\neq j,\ j\neq k \text{ et } k\neq l\}$

- v ou x doit contenir des a. (condition 1).
- Si $x\in b^*$ ou c^*, $v\in a^+$; si $x\in a^+$, $v\in a^*$
- Cas 2 : $x\in b^*$ et $v=a^p$ $0<p<n+1$ donc p divise n!
  - Soit q t.q. pq=n!

    $z'=uv^{2q+1}wx^{2q+1}y$ doit être dans L
  - mais $v^{2q+1}=a^{2pq+p}=a^{2n!+p}$.
  - Comme $|uwy|_a=(n-p)$, $|z'|_a= 2n!+n$
  - Ni v ni x ne contiennent de c, donc $|z'|_c= |z|_c= n+2n!$
  - Mais dans ce cas, $|z'|_a= |z'|_c$, une contradiction
- Le autres cas sont traités de la même manière.

Propriétés de clôture

L'union

Lemme : Les langages algébriques sont clos pour l'union

- Soient $G_1=(N_1,T_1,S_1,R_1)$ et $G_2=(N_2,T_2,S_2,R_2)$ qui engendrent resp. L_1 et L_2
- On construit G_3 qui engendre $L_1\cup L_2$.

  $G_3=(N_1\cup N_2\cup\{S_3\},T_1\cup T_2,S_3,R_1\cup R_2\cup\{S_3\to S_1|S_2\})$
- Pour que tout marche bien, il faut ajouter comme hypothèse que $N_1\cap N_2=\emptyset$ et que $\{S_3\}\notin N_1\cup N_2$
- On ajoute la règle $S_3\to S_1|S_2$
  - Par la dérivation $S_3\to S_1$, on engendre L_1
  - Par la dérivation $S_3\to S_2$, on engendre L_2

Exemple

$L=\{a^ib^j: i,j \geq 0, i=j \text{ ou } i>j\}$

- On découpe L en deux langages:

  $L_1=\{a^ib^j: i,j \geq 0, i=j\}$

  $L_2=\{a^ib^j: i,j\geq 0, i>j\}$

Dont on va faire l'union : $S_1\to a\ S_1b|\varepsilon$

$S_2\to a\ S_2b|a|aS_2$

On ajoute : $S_3\to S_1|\ S_2$ pour obtenir l'union

Preuve

- On montre $L_1\cup L_2=L(G_3)$
- $L_1\cup L_2 \subseteq L(G_3)$
  - $w\in L_1$, la dérivation $S_3\to S_1\to^* w$ est une dérivation de G_3 puisque toute règle de G_1 est aussi une règle de G_3.
  - $w\in L_2$, la dérivation $S_3\to S_2\to^* w$ est une dérivation de G_3
- $L(G_3)\subseteq L_1\cup L_2$
  - Soit $w\in L(G_3)$. Alors la dérivation $S_3\to^* w$ commence soit par $S_3\to S_1\to^* w$ soit par $S_3\to S_2\to^* w$
  - Dans le premier cas, seules des variables de G_1 peuvent apparaître car $N_1\cap N_2=\emptyset$ et seules les règles de G_1 sont utilisées. L'autre cas est pareil.

La concaténation

Lemme : Les langages algébriques sont clos pour la concaténation

- $G_1=(N_1,T_1,S_1,R_1)$ et $G_2=(N_2,T_2,S_2,R_2)$ engendrent respectivement L_1 et L_2
- On construit G_4 qui engendre $L_1.L_2$

  $G_4=(N_1N_2\cup\{S_4\},T_1\cup T_2,S_4, R_1\cup R_2\cup\{S_4\to S_1S_2\})$
- Pour que tout marche, il faut l'hypothèse $N_1\cap N_2=\emptyset$ et $\{S_4\}\notin N_1\cup N_2$
- En ajoutant la règle $S_4\to S_1S_2$
  - à partir de S_1 (resp S_2), on engendre L_1 (resp L_2)

Exemple

$L=\{a^i b^j c^k : i,j,k \geq 0,\ i=j\}$

- On découpe L en deux langages:

  $L_1=\{a^i b^j : i,j \geq 0,\ i=j\}$

  $L_2=\{c^k : k \geq 0\}$

Dont on va faire la concaténation

$S_1 \to a\ S_1 b \mid \varepsilon$

$S_2 \to c S_2 \mid \varepsilon$

On ajoute la règle $S_4 \to S_1 S_2$ pour la concaténation

19

Étoile de Kleene

Lemme : Les langages algébriques sont clos pour l'étoile de Kleene

- $G_1=(N_1,T_1,S_1,R_1)$ engendre L_1
- On construit G_5 qui engendre $(L_1)^*$

  $G_5=(N_1 \cup \{S_5\},\ T_1, S_5,\ R_1 \cup \{S_5 \to \varepsilon \mid S_1 S_5\})$

Pour que tout marche, il faut l'hypothèse $\{S_5\} \notin N_1$

- En ajoutant la règle $S_5 \to \varepsilon \mid S_1 S_5$
  - Par la dérivation $S_5 \to S_1$, on engendre L_1
  - Par la dérivation $S_5 \to S_5$, on peut recommencer
- On engendre donc bien l'étoile de Kleene

20

Exemple

$L=\{(a^n b^n)^k : k,n \geq 0\}=(a^n b^n)^*$

- Soit

  $G_1=(N_1,T_1,\ S_1,\ \{S_1 \to \varepsilon \mid a\ S_1 b\})$

- Qui engendre $L_1=\{a^n b^n : n \geq 0\}$
- Par la règle $S_5 \to \varepsilon \mid S_1 S_5$ on engendre $(L_1)^*$

21

Théorème

Les langages algébriques sont clos

- Pour l'union
- Pour la concaténation
- Pour l'étoile de Kleene

- Cela nous donne une manière de construire une grammaire algébrique en « découpant » astucieusement le langage à engendrer

22

$L=\{a^i b^j c^k : i,j,k \geq 0,\ i=j \text{ ou } j=k\}$

- Se décompose comme

  $\{a^i b^j c^k : i,j,k \geq 0,\ i=j\} \cup \{a^i b^j c^k : i,j,k \geq 0,\ j=k\}$

  - $\{a^i b^j c^k : i,j,k \geq 0,\ i=j\} = \{a^i b^j : i,j \geq 0,\ i=j\}.\{c^k : k \geq 0\}$
    - $\{a^i b^j : i,j \geq 0,\ i=j\}$ engendré par $S_1 \to \varepsilon \mid a S_1 b$
    - $\{c^k : k \geq 0\}$ engendré par $S_2 \to \varepsilon \mid c S_2$ (c^*)
    - Leur concaténation : $S_5 \to S_1 S_2$
  - $\{a^i b^j c^k : i,j,k \geq 0,\ j=k\} = \{a^i : i \geq 0\}.\{b^j c^k : j,k \geq 0,\ j=k\}$
    - $\{a^i : i \geq 0\}$ engendré par $S_3 \to \varepsilon \mid a S_3$
    - $\{b^j c^k : j,k \geq 0,\ j=k\}$ engendré par $S_4 \to \varepsilon \mid b S_4 c$
    - Leur concaténation : $S_6 \to S_3 S_4$
  - L'union : $S \to S_5 \mid S_6$

23

Clôture par intersection

- Si L et M sont deux langages algébriques, alors on ne peut en déduire que $L \cap M$ est algébrique

  $L=\{a^i b^i c^i : i>0\}$ n'est pas algébrique

  - $L_1=\{a^i b^i c^j : i>0,\ j>0\}$ est algébrique
    - Concaténation de $L=\{a^i b^i : i>0\}$ et $M=\{c^j : j>0\}$
  - $L_2=\{a^i b^j c^j : i>0,\ j>0\}$ est algébrique
    - Concaténation de $N=\{a^i : i>0\}$ et $P=\{b^j c^j : j>0\}$
  - $L_1 \cap L_2 = L$ n'est pas algébrique
- **Remarque** :On ne peut pas dire que l'intersection de deux langages algébriques n'est jamais algébrique
- Il suffit de prendre L=M avec L et M algébriques pour que $L \cap M = L$ soit algébrique

24

Corollaire

- On en déduit que les langages algébriques ne sont pas clos par complémentation
- Sinon, par les lois de Morgan, on aurait

$$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$$

et les algébriques seraient clos par intersection

Intersection avec rationnels

- **Théorème**: Si L est algébrique et M rationnel, alors $L \cap M$ est algébrique
- L'idée est de faire fonctionner en parallèle un AP pour L et un AF pour M

Intersection avec rationnels

- **Construction** :
  - On a A =$(Q_1,\Sigma,\Gamma,\delta_1,q_1,Z,F_1)$ un AP qui accepte L par EF
  - Et B =$(Q_2, \Sigma, \delta_2, q_2,F_2)$ un AF qui accepte M
  - On construit un automate à pile produit

$$(Q_1 \times Q_2, \Sigma,\Gamma, \delta,(q_1, q_2),Z, F_1 \times F_2)$$

  - Transition normale : Si $\delta_1(q,a,X)=(q',\gamma)$ et $\delta_2(p,a)=p'$, alors
    - $\delta((q,p),a,X)=((q',p'),\gamma)$
  - ε-transition : Si $\delta_1(q, \varepsilon,X)=(q',\gamma)$, on ne s'intéresse pas à N:
    - $\delta((q,p),\varepsilon,X)=((q',p),\gamma)$

Utilité de ces propriétés

- Comme pour les langages rationnels, les propriétés de clôture sont utiles
  - Pour construire des langages algébriques
  - Pour montrer qu'un langage donné n'est pas algébrique

Moralité

- permet de simplifier les preuves de non-algébricité de certains langages
- Par exemple pour un langage qui conduit à une « explosion » de cas à considérer :

$$D=\{ww:w\in\{a,b\}^*\}$$

  - On suppose D algébrique
  - Soit $L=a^+b^+a^+b^+$ un langage rationnel
  - Alors, $D\cap L$ est algébrique
  - Mais $D\cap L =\{a^ib^ja^ib^j:i,j>0\}$ réputé non algébrique (analogue à $\{a^ib^jc^id^j:i,j>0\}$)
  - Donc D n'est pas algébrique