Limites des langages rationnels

Est ce que tout langage est rationnel?
Existe-t-il des langages non rationnels?

Hypothèse: tout est rationnel

- Les langages rationnels sont reconnaissables : ils sont en bijection avec les automates.
- Un automate peut être décrit par un texte
- Ce texte correspond (codage ASCII) à un nombre binaire
- D'où que les langages reconnaissables sont en bijection avec $\mathbb{N}$.

Hypothèse: tout est rationnel

- On énumère les automates finis sur un alphabet à une lettre et on les ordonne dans une liste :
  - A_0 le premier automate
  - A_1 le second automate
  - …
- On énumère les langages rationnels sur un alphabet à un seul lettre :
  - L_0 reconnu par A_0
  - L_1 reconnu par A_1
  - …

Tableau mots/langages

	L_0	L_1	L_2	L_3
w_0	O	O	N	N
w_1	N	N	O	N
w_2	O	N	O	N
w_3	N	O	O	N

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

$w_i \in D \Leftrightarrow w_i \notin L_i$
D n'est pas dans T

D n'est pas dans T

- Si D était dans le tableau, il existerait j tel que $D=L_j$.
- Puisque $D=L_j$, si
  - $w_j \in L_j$ alors, par définition de D, $w_j \notin D \Rightarrow$ contradiction
  - $w_j \notin L_j$ alors, par définition de D, $w_j \in D \Rightarrow$ contradiction
- l'ensemble des langages est infini, mais non dénombrable

D n'est pas dans T

- Il existe des langages qui ne sont pas rationnels

Preuve par technique de diagonalisation due à Cantor.

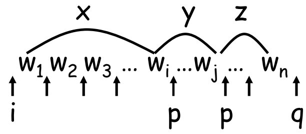
Très utile pour montrer qu'un ensemble infini n'est pas dénombrable.

Un langage qui n'est pas rationnel

7

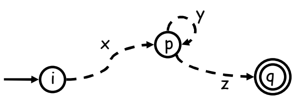
Le langage L= $\{0^k1^k \mid k \geq 0\}$

- Supposons L rationnel; il existe A un AFD à n états qui le reconnaît. Choisissons w un mot de L de longueur ≥ n (par exemple $w=0^n1^n$). Que se passe-t'il lors de la lecture de w?
- En lisant les n premiers 0, un état p de A est visité plusieurs fois (on passe par n+1 états pour lire n symboles).



8

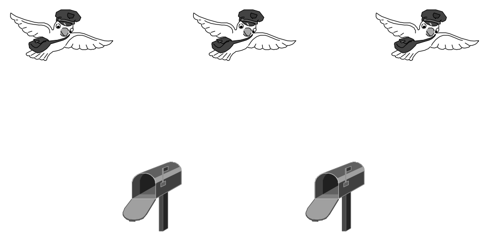
Le langage L= $\{0^k1^k \mid k \geq 0\}$



- Puisque w=xyz∈L, xz∈L ainsi que xyyz et, ∀i≥0, xy^iz. Mais chacun de ces mots possède plus ou moins de 0 que de 1, une contradiction.
- application simple
  - du principe des pigeons pour les anglo-saxons
  - des tiroirs de Dirichlet chez nous
  - ou aussi tiroirs et chaussettes …

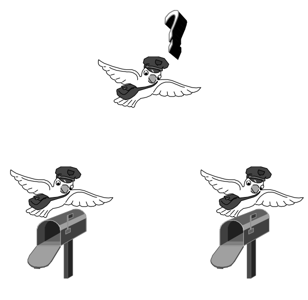
9

Principe des pigeons



10

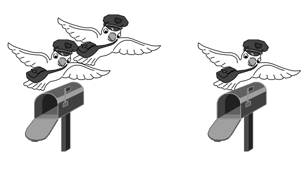
Principe des pigeons



11

Principe des pigeons

Si on cherche à mettre n pigeons dans m cages (n>m), une cage contiendra plus d'un pigeon



12

Ce qu'on vient de faire

- On a montré qu'il existe au moins un langage non rationnel, le langage

  $L = \{0^k 1^k : k \geq 0\}$
- **But** : trouver une technique pour montrer la non rationalité d'un langage, i.e. pour décider le problème :
  - Donnée : L un langage
  - Question : L est-il non rationnel?

13

Le lemme de la pompe

14

Technique de démonstration

- On utilise un résultat sur les langages rationnels: le lemme de la pompe.
- Il exprime une propriété particulière des rationnels.
- Si un langage ne possède pas cette propriété, il n'est pas rationnel.

- **Propriété:** *Tout mot (suffisamment long) d'un langage rationnel contient un facteur qui peut être itéré autant que l'on veut de telle sorte que le mot résultant est toujours dans la langage.*

15

Le lemme de la pompe

- Si L est rationnel, alors il existe un nombre n tel que pour tout mot w de L, $|w| \geq n$, w peut être factorisé en w=xyz de telle sorte que
  1. Pour tout $i \geq 0$, $xy^iz \in L$
  2. $|y|>0$
  3. $|xy| \leq n$
- Quand w est factorisé en xyz, soit x soit z peut être ε mais la condition 2 assure que $y \neq \varepsilon$.
- La condition 3 assure que le préfixe xy est de longueur au plus n. Cette condition est utile pour certains langages.

16

Exemple pour $L= \{0^k1^k \mid k \geq 0\}$

- Supposons L rationnel. Alors par le lemme, il existe n tel que pour tout mot w=xyz , $y \neq \varepsilon$, $|xy| \leq n$ et $\forall i$, $xy^iz \in L$.
- En particulier pour $w=0^n1^n$. Comme $|xy| \leq n$, y ne contient que des zéros. Alors pour i=0, le mot $xz \notin L$. Une contradiction
- L n'est pas rationnel

17

principe de la preuve

- On utilise le principe des pigeons entre le nombre de lettres d'un mot (de longueur au moins n) du langage L et n le nombre d'états d'un (hypothétique) AFD qui reconnaît L

- On montre alors que le mot peut être factorisé en vérifiant les 3 conditions

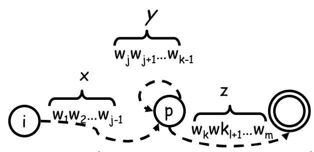
18

Preuve

- Soit A un AFD à n états reconnaissant L et soit $w=w_1w_2...w_m$ avec $m \geq n$.
- $r_0, r_1, ..., r_m$ est la suite des états pris par A lors de la lecture de w. $r_i=\delta(r_{i-1}, w_i)$ $1 \leq i \leq m$. La suite {r} est de longueur $m+1 \geq n+1$.
- Par le principe des pigeons, deux des n+1 premiers éléments de la suite {r} représentent un même état p : r_j et r_k avec $j<k$.
- Comme r_k est l'un des n+1 premiers éléments de la suite, $k \leq n$.

Preuve

- On pose
  - $x=w_1w_2...w_{j-1}$
  - $y=w_jw_{j+1}...w_{k-1}$
  - $z=w_kw_{k+1}...w_m$



- lire x mène de r_0 à r_j, y de r_j à $r_k=r_j$ et z de r_k à r_m , un état d'acceptation.
- A doit donc accepter xy^iz $\forall i \geq 0$
- $j \neq k$; donc $|y|>0$
- $k \leq n+1$ et donc $|xy| \leq n$
- Les conditions du lemme sont donc satisfaites

Remarques

- Observons que le lemme dit

  L rationnel $\Rightarrow$ L satisfait le lemme

- Mais on ne sait rien pour la réciproque:

- Si L satisfait le lemme, on ne sait pas si L est rationnel

Utilisation du lemme

1. On suppose que L est rationnel
2. Le lemme $\Rightarrow$ tout mot de longueur $\geq$n du langage peut être « gonflé »
3. Trouver w ($|w| \geq n$) qui ne peut pas être gonflé, **quelle que soit** sa factorisation.
4. Une contradiction pour chaque factorisation
5. L n'est donc pas rationnel

Point délicat

- Le point 3. est le plus délicat

3. Trouver w ($|w| \geq n$) qui ne peut pas être gonflé, **quelle que soit** sa factorisation.

Il faut :

trouver un mot qui, pour toute factorisation, permet de trouver une valeur de i (la valeur de répétition) qui nous mène à un mot qui n'est pas de L. On contredit ainsi le lemme

Pour toute factorisation

Pourquoi faut-il trouver un mot qui, pour toute factorisation, permet de trouver une valeur de *i* (la valeur de répétition) qui nous mène à un mot qui n'est pas dans L ?

Pour toute factorisation

- On fait un raisonnement par l'absurde :
  - On utilise le fait
    L rationnelle $\Rightarrow$ "L vérifie le lemme"
  - Équivalent à
    P $\equiv$ "L non rationnelle" $\vee$ "L vérifie le lemme"
  - Par l'absurde : il faut nier P

$\neg$P $\equiv$ "L rationnelle" $\wedge$ "L ne satisfait pas le lemme"

- Que veut dire que L ne satisfait pas le lemme ?

25

Présentation courte du lemme

$\forall$L

$\boxed{[(\exists n)(\forall w)\ w\in L \text{ et } |w|\geq n}$ A

$\Rightarrow$

$\boxed{(\exists\ x,y,z)(w=xyz,\ |xy|\leq n,\ |y|\geq 1 \text{ et } (\forall i)(xy^iz\in L))]}$ B

- L ne satisfait pas le lemme :
- A$\Rightarrow$B est faux

26

Montrer A$\Rightarrow$B est faux

- Il suffit de trouver un exemple pour lequel
  - A $\equiv(\exists n)(\forall w)\ w\in L$ et $|w|\geq n$ est **vrai** et
  - B $\equiv(\exists\ x,y,z)(w=xyz,\ |xy|\leq n,\ |y|\geq 1$ et $(\forall i)(xy^iz\in L))$ est **faux**
- B est **faux** $\equiv$ $\neg$B est **vrai**

$\neg B\equiv \neg[(\exists\ x,y,z)(w=xyz,\ |xy|\leq n,\ |y|\geq 1 \text{ et } (\forall i)(xy^iz\in L))]$

$\equiv (\forall\ x,y,z)\ \neg[\ (\underbrace{\boxed{w=xyz,\ |xy|\leq n,\ |y|\geq 1}}_{C} \text{ et } \underbrace{\boxed{(\forall i)(xy^iz\in L)}}_{D})\ ]$

  - On transforme $\neg$B en une nouvelle implication

27

Un peu de (rappel ?) de logique

$\neg(C\wedge D) = C \Rightarrow \neg D$

C	D	C $\wedge$ D	$\neg$(C $\wedge$ D)	$\neg$D	C $\Rightarrow$ $\neg$D
V	V	V	F	F	F
V	F	F	V	V	V
F	V	F	V	F	V
F	F	F	V	V	V

Montrer A$\Rightarrow$B est faux

- Il suffit de trouver un exemple pour lequel
  - A $\equiv(\exists n)(\forall w)\ w\in L$ et $|w|\geq n$ est **vrai** et
  - B $\equiv(\exists\ x,y,z)(w=xyz,\ |xy|\leq n,\ |y|\geq 1$ et $(\forall i)(xy^iz\in L))$ est **faux**
- B est **faux** $\equiv$ $\neg$B est **vrai**

$\neg B\equiv \neg[(\exists\ x,y,z)(w=xyz,\ |xy|\leq n,\ |y|\geq 1 \text{ et } (\forall i)(xy^iz\in L))]$

$\equiv (\forall\ x,y,z)\ \neg[\ (\underbrace{\boxed{w=xyz,\ |xy|\leq n,\ |y|\geq 1}}_{C} \text{ et } \underbrace{\boxed{(\forall i)(xy^iz\in L)}}_{D})\ ]$

  - On transforme $\neg$B en une nouvelle implication

$(\forall\ x,y,z)\ C \Rightarrow \neg D$

$(\forall\ x,y,z)\ (w=xyz,\ |xy|\leq n,\ |y|\geq 1 \Rightarrow (\exists i)(xy^iz\notin L))$

29

Montrer que B est faux

$(\forall\ x,y,z)\ (w=xyz,\ |xy|\leq n,\ |y|\geq 1 \Rightarrow (\exists i)(xy^iz\notin L))$

- Qui est vraie, sauf si
  - Le membre gauche de l'implication est vraie
    et
  - Le membre droit $(\exists i)(xy^iz\notin L)$ est faux.
- Il suffit de trouver un exemple pour lequel
  $(\forall\ x,y,z)$
  - $C \equiv (w=xyz,\ |xy|\leq n,\ |y|\geq 1)$ est vrai et
  - $\neg D \equiv (\exists i)\ (xy^iz\notin L)$ est vrai

$(\forall\ x,y,z)\ (w=xyz,\ |xy|\leq n,\ |y|\geq 1 \Rightarrow (\exists i)(xy^iz\notin L))$

30

Exemple $L=\{w\in\{0,1\}^*:|w|_0=|w|_1\}$

- Supposons L rationnel et soit n la valeur fixée par le lemme.
- On choisit $w=0^n1^n$. On peut alors factoriser w en accord avec le lemme

  $w=xyz$ avec $|xy|\leq n$ et $|y|>0$
- y ne contient que des 0 et xy^2z n'est plus dans le langage; une contradiction

31

Une présentation comme jeu

- L'utilisation du lemme peut être présenté comme un jeu entre deux joueurs (vous et un adversaire) :
  - Votre but est de prouver que L n'est pas rationnel.
  - Les correspondances des quantificateurs :

    vous ~ $\forall$ **et** adversaire ~ $\exists$

1. Vous choisissez L.
2. L'adversaire choisit n.
3. Vous choisissez $w\in L$, $|w|\geq n$.
4. L'adversaire choisit x,y,z. $w=xyz$, $|xy|\leq n$, $|y|\geq 1$.
5. Vous choisissez i tel que $xy^iz \notin L$.

- Chaque choix peut dépendre des précédents.

32

Une présentation comme jeu

Etape	Vous	Adversaire
1.	Choix de L	
2.		Choix de n
3.	Choix de $w\in L$, $\|w\|\geq n$	
4.		Choix de x,y,z t.q. $w=xyz$, $\|xy\|\leq n$, $\|y\|\geq 1$
5.	Choix de i t.q. $xy^iz \notin L$	

Une présentation comme jeu - exemple

1. $L=\{w\in(a+b)^* : |w|_a\leq|w|_b\}$
2. n
3. $w=a^nb^n$
4. $w=xyz$, $x=a^j$, $y=a^k$, $z=a^{n-j-k}b^n$, $j\geq 0$, $k>0$, $j+k\leq n$

   $(|xy|=j+k\leq n,\ |y|=k>0)$
5. i=2 :

   $$xy^2z=a^ja^ka^ka^{n-j-k}b^n=a^{n+k}b^n\notin L$$

Conclusion : L n'est pas rationnel !

34

Prouver la non rationalité

- Pour montrer que L n'est pas rationnel :

  on fait un raisonnement par l'absurde.
  - On utilise le raisonnement avec des AFD
  - On utilise le lemme de la pompe
- Autre méthode :

  on utilise les propriétés de clôture

Union	Intersection	Etoile	Concatenation	Substitution
Oui	Oui	Oui	Oui	Oui

35

Exemple $L=\{w\in\{0,1\}^*:|w|_0=|w|_1\}$

Autre méthode :

- Supposons $L=\{w\in\{0,1\}^*:|w|_0=|w|_1\}$ rationnel
- Par les propriétés de clôture, $L\cap 0^*1^*$ doit être rationnel (0^*1^* est rationnel)
- $L\cap 0^*1^*= \{0^n1^n : n\geq 0\}$
- Comme $\{0^n1^n : n\geq 0\}$ n'est pas rationnel, L ne peut être rationnel.

36