Informatique théorique III

Série 5 solution

Prof. Nestmann, 2004

1. Langage non régulier et propriétés de stabilité

Soit $\Sigma = \{a, b\}$. On admet dans cet exercice que $L \triangleq \{a^n b^n \mid n \in \mathbb{N}\}$ n'est pas régulier.

Montrer que les langages suivants ne sont pas réguliers :

1. $L_1 \triangleq \{w \in \Sigma^* \mid |w|_a = |w|_b\}$
2. $L_2 \triangleq \{a^i b^j \mid i, j \in \mathbb{N}, i \neq j\}$
3. $L_3 \triangleq \{b^{2n} a^n \mid n \in \mathbb{N}\}$

On pourra raisonner par l'absurde et exprimer L en fonction du langage à considérer et d'opérations qui conservent la propriété de régularité d'un langage.

Solution.

1. On a $L = L_1 \cap a^* b^*$. Si L_1 était régulier, alors comme $a^* b^*$ est régulier et l'intersection de deux langages réguliers est régulier, on en déduirait que L est régulier, ce qui n'est pas. Donc L_1 n'est pas régulier.
2. On a $L = a^* b^* \setminus L_2$. Si L_2 était régulier, alors comme $a^* b^*$ est régulier et la différence de deux langages réguliers est régulier, on aurait que L est régulier, ce qui n'est pas. Donc L_2 n'est pas régulier.
3. Soit $h : \Sigma^* \to \Sigma^*$ l'homomorphisme de mot défini par $h(a) \triangleq bb$ et $h(b) \triangleq a$. On a $L = h^R(L_3)$. Si L_3 était régulier, alors comme l'image réciproque d'un langage régulier par un homomorphisme de mot est régulier, on aurait L régulier, ce qui n'est pas. Donc L_3 n'est pas régulier.

■

2. Propriétés de stabilité

Soit Σ un alphabet.

Soit $u \in \Sigma^*$. On dit que $v \in \Sigma^*$ est un *préfixe propre* de u si v est un préfixe de u (i.e. s'il existe $w \in \Sigma^*$ tel que $u = v \cdot w$) et $v \neq u$.

On a donc l'équivalence suivante :

$$v \text{ est un préfixe propre de } u \Leftrightarrow \exists w \in \Sigma^* \setminus \{\epsilon\} : u = v \cdot w$$

Soit L un langage régulier.

Montrer que $\min(L) = \{u \in L \mid u \text{ ne possède pas de préfixe propre dans } L\}$ est un langage régulier en exprimant $\min(L)$ en fonction de L et d'autres langages réguliers et en utilisant les propriétés de stabilité des langages réguliers.

Solution. Par définition $\min(L) = \{u \in L \mid \forall v \in L : v \text{ n'est pas un préfixe propre de } u\}$.

Soit $u \in \Sigma^*$,

$$\begin{aligned}
u \in \min(L) &\Leftrightarrow u \in L \wedge \forall v \in L : v \text{ n'est pas un préfixe propre de } u \\
&\Leftrightarrow u \in L \wedge \forall v \in L : \neg(\exists w \in \Sigma^* \setminus \epsilon : u = v \cdot w) \\
&\Leftrightarrow u \in L \wedge \neg(\exists v \in L : \exists w \in \Sigma^* \setminus \epsilon : u = v \cdot w) \\
&\Leftrightarrow u \in L \wedge \neg(u \in L \cdot (\Sigma^* \setminus \{\epsilon\}) \\
&\Leftrightarrow u \in L \setminus (L \cdot (\Sigma^* \setminus \{\epsilon\}))
\end{aligned}$$

On a donc montré que $\min(L) = L \setminus (L \cdot (\Sigma^* \setminus \{\epsilon\}))$.

Si L est régulier, alors $L \cdot (\Sigma^* \setminus \{\epsilon\})$ est régulier car $\Sigma^* \setminus \{\epsilon\}$ est régulier et par propriété de stabilité. D'où, par propriété de stabilité, $\min(L) = L \setminus (L \cdot (\Sigma^* \setminus \{\epsilon\}))$ est régulier. ■

3. Propriétés de stabilité (artisanal)

Soit Σ un alphabet et L un langage régulier sur Σ. Montrer que les langages suivants sont réguliers :

1. $\mathrm{mirror}(L) = \{u \in \Sigma^* \mid \overline{u} \in L\}$ (où $\overline{u}$ a été défini dans la série 1)
2. $\frac{1}{2}L = \{u \in \Sigma^* \mid \exists v \in \Sigma^* : |u| = |v| \wedge u \cdot v \in L\}$

Indication : Prendre un AFD M reconnaissant L et construire un AFD, AFN ou AFN$_{\mathbf{e}}$ (défini en fonction de M) reconnaissant $f(L)$.

Solution.
Avertissement : Dans cet exercice, certains résultats sont énoncés sans preuve. Bien entendu, vous devez être capables de prouver ces résultats. Néanmoins, le but de cet exercice était plutôt de stimuler votre créativité que de vous faire faire des preuves.

Soit $M = (Q, \Sigma, \delta, s, F)$ un AFD reconnaissant L, i.e. tel que $\mathrm{L}(M) = L$.

1. L'idée est simplement d'inverser (retourner) toutes les transitions de M, de prendre un nouvel état initial $s_{\mathbf{e}}$ qui va sur l'ensemble des états finaux de M par $\mathbf{e}$-transition et de prendre comme état final l'état initial de M.

   On définit $M' = (Q', \Sigma, \Delta', s_{\mathbf{e}}, F')$ l'AFN$_{\mathbf{e}}$ suivant :

$$Q' \triangleq Q \cup \{s_{\mathbf{e}}\} \qquad s_{\mathbf{e}} \text{ étant un nouvel état, i.e. } s_{\mathbf{e}} \notin Q$$
$$F' \triangleq \{s\}$$

   et

$$\begin{aligned} \Delta' : Q' \times (\Sigma \cup \{\,\mathbf{e}\,\}) &\to \mathcal{P}(Q') \\ (q, a) &\mapsto \begin{cases} \{q' \in Q \mid \delta(q', a) = q\} & \text{si } q \neq s_{\mathbf{e}} \text{ et } a \in \Sigma \\ F & \text{si } q = s_{\mathbf{e}} \text{ et } a = \mathbf{e} \\ \emptyset & \text{sinon} \end{cases} \end{aligned}$$

   On peut montrer (par induction sur $w \in \Sigma^*$) que

$$\forall w \in \Sigma^*, q \in Q : \widehat{\Delta}(q, w) = \left\{q' \in Q \mid \widehat{\delta}(q', \overline{w}) = q\right\}$$

   Et donc, soit $w \in \Sigma^*$,

$$\begin{aligned} w \in \mathrm{L}(M') &\Leftrightarrow s \in \widehat{\Delta}(s_{\mathbf{e}}, w) \\ &\Leftrightarrow s \in \bigcup_{q \in F} \widehat{\Delta}(q, w) \\ &\Leftrightarrow \exists q \in F : s \in \widehat{\Delta}(q, w) \\ &\Leftrightarrow \exists q \in F : \widehat{\delta}(s, \overline{w}) = q \\ &\Leftrightarrow \widehat{\delta}(s, \overline{w}) \in F \\ &\Leftrightarrow \overline{w} \in \mathrm{L}(M) \\ &\Leftrightarrow w \in \mathrm{mirror}(L) \end{aligned}$$

   On conclut que $\mathrm{mirror}(L)$ est régulier.

2. L'idée ici est d'exécuter M dans les deux sens à la fois.
   On définit $M' = (Q', \Sigma, \delta', s', F')$ l'AFD suivant :

$$\begin{aligned} Q' &\triangleq Q \times \mathcal{P}(Q) \\ s' &\triangleq (s, F) \\ F' &\triangleq \{(q, P) \mid q \in P\} \end{aligned}$$

   et

$$\begin{aligned} \delta' : \quad & Q' \times \Sigma \to Q' \\ & ((q, P), a) \mapsto (\delta(q, a), \{q' \in Q \mid \exists b \in \Sigma : \delta(q', b) \in P\}) \end{aligned}$$

   On peut montrer (par induction sur $w \in \Sigma^*$) que

$$\forall w \in \Sigma^* : \widehat{\delta'}(s', w) = (q, P) \Leftrightarrow \widehat{\delta}(s, w) = q \wedge P = \left\{q' \in Q \mid \exists w' \in \Sigma^* : |w| = |w'| \wedge \widehat{\delta}(q', w') \in F\right\}$$

   On peut alors montrer que $\mathrm{L}(M') = \frac{1}{2}\,\mathrm{L}(M)$.
   En effet, soit $w \in \Sigma^*$,

$$\begin{aligned} w \in \frac{1}{2}L &\Leftrightarrow w \in \frac{1}{2}\,\mathrm{L}(M) \\ &\Leftrightarrow \exists w' \in \Sigma^* : |w| = |w'| \wedge w \cdot w' \in \mathrm{L}(M) \\ &\Leftrightarrow \exists w' \in \Sigma^* : |w| = |w'| \wedge \widehat{\delta}(s, w \cdot w') \in F \\ &\Leftrightarrow \exists w' \in \Sigma^* : |w| = |w'| \wedge \widehat{\delta}(\widehat{\delta}(s, w), w') \in F \\ &\Leftrightarrow \widehat{\delta}(s, w) \in \left\{q \in Q \mid \exists w' \in \Sigma^* : |w| = |w'| \wedge \widehat{\delta}(q, w') \in F\right\} \\ &\Leftrightarrow \widehat{\delta'}(s', w) \in F' \\ &\Leftrightarrow w \in \mathrm{L}(M') \end{aligned}$$

   On conclut que $\frac{1}{2}L$ est régulier.

■

4. Langages finis et langages réguliers

Un langage est *fini* s'il contient un nombre fini de mots. Il est *infini* dans le cas contraire. Démontrez le théorème suivant.

Théorème 4.1 *Tous les langages finis sont réguliers.* □

Preuve. Soit L un langage fini quelconque. Nous avons les deux cas suivant,

1. $L = \emptyset$.
   Prenons l'expression régulière $\mathbf{x} = \boldsymbol{\emptyset}$, nous avons $L(\mathbf{x}) = \emptyset = L$. Il existe donc une expression régulière $\mathbf{x}$ telle que $\mathrm{L}(\mathbf{x}) = L$. L est donc régulier.
2. $L \neq \emptyset$
   L étant fini, nous pouvons énumérer ses mots. Supposons que $L = \{w_1, w_2, \ldots, w_n\}$ avec $n = \#(L)$. Construisons, l'expression régulière suivante,

$$\mathbf{w_1} + \mathbf{w_2} + \ldots + \mathbf{w_n}$$

   Nous avons $L(\mathbf{w_1} + \mathbf{w_2} + \ldots + \mathbf{w_n}) = \bigcup_{i=1}^{n} L(\mathbf{w_i}) = \bigcup_{i=1}^{n} \{w_i\} = L$. Il existe donc une expression régulière $\mathbf{x}$ telle que $\mathrm{L}(\mathbf{x}) = L$. L est donc régulier.

De (1) et (2) et en généralisant sur L, nous déduisons le théorème. ■

5. Langages non réguliers

Montrez à l'aide du lemme de gonflement la non-régularité des langages suivants.

1. $L = \{w \in \{a,b\}^* \mid |w|_a = |w|_b\}$
2. Un *palindrome* est un mot sur un alphabet Σ que l'on peut lire de manière identique dans un sens ou dans l'autre[1]. En d'autre termes on a :

$$\mathrm{Pal}(w) \Leftrightarrow \forall i \in [1, |w|] : (w)_i = (w)_{|w|+1-i}$$

   $L = \{w \in \{a,b\}^* \mid \mathrm{Pal}(w)\}$

3. Soit $\Sigma = \{(,), a\}$, un mot w sur Σ dont les parenthèses sont *équilibrées* satisfait la propriété suivante :

$$\mathrm{Equil}(w) \Leftrightarrow (|w|_(= |w|_)) \wedge (\forall u, v \in \Sigma^* : w = uv \Rightarrow |u|_(\geq |u|_))$$

   Par exemple les mots $a(a()(((a)))a)$ et $a()()$ sont équilibrés, mais les mots $(a((()))$ et $))(($ ne le sont pas.

   $L = \{w \in \Sigma \mid \mathrm{Equil}(w)\}$ □

Solution. Pour démontrer qu'un langage n'est pas régulier, nous pouvons utiliser le lemme de gonflement. Par sa forme, ce lemme énonce une propriété des langages réguliers,

$$L \text{ régulier} \Rightarrow \mathbf{G}(L)$$

où $\mathbf{G}(()L)$ est la propriété. Cette propriété est nécessaire et par les règles de la logique propositionnelle nous savons que $(B \Rightarrow A) \Leftrightarrow (\neg A \Rightarrow \neg B)$. Dès lors, la contraposée de la proposition ci-dessus est :

$$\neg\,\mathbf{G}(L) \Rightarrow \neg(L \text{ régulier})$$

La propriété est donc *nécessaire* dans le sens où si elle n'est pas satisfaite, le langage n'est pas régulier. Notez bien que si la propriété est satisfaite pour un langage L, cela ne nous dit *rien* sur la régularité de celui-ci. Dans notre cas, la proposition $\mathbf{G}(L)$ est,

$$\begin{aligned}&\exists n \in \mathbb{N} : \forall w \in L : \\ &|w| \geq n \Rightarrow \big(\exists x, y, z : (w = xyz) \wedge (y \neq \epsilon) \wedge (|xy| \leq n) \wedge (\forall k \in \mathbb{N} : xy^k z \in L)\big)\end{aligned}$$

Sa négation $\neg\,\mathbf{G}(L)$ est donc[2],

$$\begin{aligned}&\forall n \in \mathbb{N} : \exists w \in L : \\ &(|w| \geq n) \wedge \big(\forall x, y, z : \neg\big((w = xyz) \wedge (y \neq \epsilon) \wedge (|xy| \leq n)\big) \vee \neg(\forall k \geq 0 : xy^k z \in L)\big) \\ &\Leftrightarrow \\ &\forall n \in \mathbb{N} : \exists w \in L : \\ &(|w| \geq n) \wedge \big(\forall x, y, z : (w = xyz) \wedge (y \neq \epsilon) \wedge (|xy| \leq n) \Rightarrow (\exists k \geq 0 : xy^k z \notin L)\big)\end{aligned}$$

Une stratégie pour démontrer $\neg\,\mathbf{G}(L)$ — et donc qu'un langage L n'est pas régulier — est la suivante.

[1] « Ésope reste élu par cette crapule et se repose » est un palindrome modulo les accents et les espaces. Un grand palindrome écrit en français (environ 1500 mots) a été publié par Georges Perec sous le titre 9691 dans Oulipo, *La littérature potentielle*, Gallimard 1973 (seconde édition Folio, 1988).

[2] Rappelez vous $(A \Rightarrow B) \Leftrightarrow (\neg A \vee B)$ et la dualité des quantificateurs et connecteurs logiques.

1. Soit $n \in \mathbb{N}$. Construire un $w \in L$ *particulier* dont la taille est au moins n (votre construction doit donc dépendre de n).
2. Soit $w = xyz$ une décomposition de w qui satisfait $y \neq \epsilon$ et $|xy| \leq n$. Montrer qu'il existe un k tel que $xy^kz \notin L$.

Attention. Vous ne pouvez pas choisir un n particulier car vous devez montrer la propriété *pour tout* n. De même vous ne pouvez pas choisir une décomposition particulière qui vous arrange, vous devez montrer cela *pour toute* décomposition qui satisfait uniquement les deux propositions données ci-dessus.

Nous utilisons cette stratégie pour chacun des langages proposés dans la donnée de l'exercice.

1. Soit $n \in \mathbb{N}$, construisons le mot $w = a^nb^n \in L$.

   Soit $w = xyz$ une décomposition de w telle que $y \neq \epsilon$ et $|xy| \leq n$. Sachant que $w = a^nb^n$, nous avons $xy = a^i$ avec $|y|_a > 0$

   Prenons $k = 0$ et formons $xy^0z = xz = a^{n-i}b^n$, nous montrons que xz n'est pas dans L. Nous avons :
   (a) $|xz|_a = |xyz|_a - |y|_a$
   (b) $|xz|_b = |xyz|_b$ (car $|xy|_b = |a^i|_b = 0$).
   Étant donné que $|y|_a > 0$ nous avons $|xz|_a \neq |xz|_b$ et donc $xz \notin L$.

   L n'est donc pas régulier.

2. Soit $n \in \mathbb{N}$, construisons le mot $w = b^naab^n \in L$.

   Soit $w = xyz$ une décomposition de w telle que $y \neq \epsilon$ et $|xy| \leq n$. Sachant que $w = b^naab^n$, nous avons $y = b^i$ avec $i > 0$.

   Prenons $k = 0$ et formons $xy^0z = xz = b^{n-i}aab^n$, nous montrons que xz n'est pas dans L. Nous avons $(xz)_{n-i+1} = a$ cependant $(xz)_{(2n+2-i)+1-(n-i+1)} = (xz)_{n+2} = b$ car $i > 0$. Donc $\neg\,\mathrm{Pal}(w)$ et donc $xz \notin L$.

   L n'est donc pas régulier.

3. Soit $n \in \mathbb{N}$, construisons le mot $w = (^n)^n \in L$.

   Soit $w = xyz$ une décomposition de w telle que $y \neq \epsilon$ et $|xy| \leq n$. Sachant que $w = (^n)^n$, nous avons $xy = (^i$ avec $|y|_(> 0$.

   Prenons $k = 0$ et formons $xy^0z = xz = (^{n-i})^n$, nous montrons que xz n'est pas dans L. Nous avons :
   (a) $|xz|_(= |xyz|_(- |y|_($
   (b) $|xz|_) = |xyz|_)$ (car $|xy|_) = |(^i|_) = 0$).
   Etant donné que $|y|_(> 0$ nous avons $|xz|_(\neq |xz|_)$, d'où $\neg\,\mathrm{Equil}(xz)$ et donc $xz \notin L$.

   L n'est donc pas régulier.

Remarque. Prenez le temps de comparer le langage L du point 1 ci-dessus avec le langage régulier du point 1.1(e) de la série 4.

Intuitivement, dans le langage L nous devons « compter » un nombre arbitrairement grand de a et de b pour vérifier la propriété $|w|_a = |w|_b$ — pensez par exemple au mot a^nb^n avec n arbitraire. Ces nombres étant arbitrairement grands il n'est pas possible de « mémoriser » leur valeur avec un nombre *fini* d'états. Un AFD n'est donc pas suffisant pour reconnaître L.

Ceci n'est pas le cas dans le langage de la série 4 car la propriété additionnelle sur les préfixes satisfaite par les mots du langage permet de vérifier $|w|_a = |w|_b$ en vérifiant l'égalité sur les sous-mots de longueur 2 (c.-à-d. fixe et finie) des mots. Bien évidemment ce langage là n'est pas du tout égal à L. ■

6. Langages réguliers infinis

L'algorithme suivant permet de décider la finitude d'un langage régulier L sur un alphabet Σ. L'algorithme *accepte* un langage L si et seulement si il est fini.

1. *Trouver un nombre n satisfaisant la condition du lemme de gonflement $\mathbf{G}(L)$ (le plus petit par exemple).*
2. *Si pour tout $w \in \Sigma^*$ tel que $n \leq |w| \leq 2n-1$ on a $w \notin L$ alors L est accepté.*
   Sinon L est rejeté.

Démontrer la validité de l'algorithme, c'est à dire que pour tout langage régulier L, L est accepté ssi L est fini. Pour cela :

1. Démontrer que pour tout langage régulier L, le point 1 peut être satisfait.
2. Démontrer la proposition suivante qui est implicitement contenue dans le point 2 de l'algorithme.

   Soit L un langage régulier et n un nombre satisfaisant la condition du lemme de gonflement $\mathbf{G}(L)$. Alors :

   $$\exists w \in L : n \leq |w| \leq 2n-1 \Rightarrow L \text{ infini} \tag{a}$$
   $$\neg\exists w \in L : n \leq |w| \leq 2n-1 \Rightarrow L \text{ fini} \tag{b}$$

   N.B. Un langage L est fini ssi il existe une borne finie sur la longueur de ses mots. □

Preuve.

1. L étant supposé régulier, le lemme de gonflement nous garantit l'existence de n.
2. (a) Supposons qu'il existe un $w \in L$ tel que $n \leq |w| \leq 2n-1$.
   L étant régulier, n étant le nombre du lemme de gonflement et $|w| \geq n$ nous savons, par le lemme de gonflement, que nous pouvons décomposer w en xyz tel que $y \neq \epsilon$, $|xy| \leq n$ et que — ce qui nous intéresse vraiment — $\forall k \in \mathbb{N} : xy^kz \in L$. Cette dernière proposition nous indique qu'il n'y a pas de borne finie sur la longueur des mots appartenenant à L. Ce langage est donc infini.

   (b) Supposons$^{(*)}$ qu'il n'existe pas de $w \in L$ tel que $n \leq |w| \leq 2n-1$.

   Pour démontrer que L est fini nous démontrons que $2n$ est une borne (finie) sur la longueur des mots de L.

   Par l'absurde supposons$^{(**)}$ qu'il existe des mots de L dont la longueur est plus grande ou égale à $2n$. Soit S l'ensemble de ces mots et w l'un des plus petits éléments de S, dans le sens où $\forall w' \in S : |w| \leq |w'|$.

   L étant régulier, n étant le nombre du lemme de gonflement et $|w| \geq n$, nous savons, par le lemme de gonflement que nous pouvons décomposer w en $w = xyz$ avec $y \neq \epsilon$ et $|xy| \leq n$ et $xy^0z = xz \in L$.

   Sachant que $|xy| \leq n$ et que $|xyz| \geq 2n$ nous avons$^{(i)}$ $n \leq |xz|$.
   Sachant que $y \neq \epsilon$ on a $|xz| < |w|$ et w étant l'un des plus petits éléments de S, nous savons que$^{(ii)}$ $|xz| < 2n$.

   Dès lors par (i) et (ii) nous avons $n < |xz| < 2n$ avec $xz \in L$, ce qui contredit $(*)$. L'hypothèse $(**)$ est fausse et donc L est fini.

■