Langages Formels

Licence Informatique – ENS Paris-Saclay

Partiel du 18 mars 2021

Toutes les réponses devront être correctement justifiées.

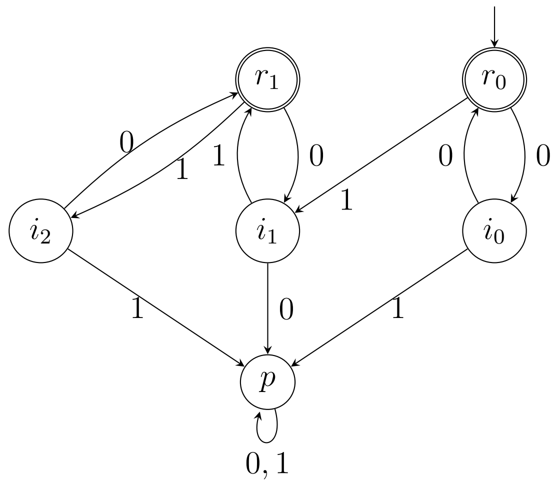
1 Automates et arithmétique

Soit $\Sigma = \{0, 1\}$. On interprète un mot $w \in \Sigma^*$ de longueur paire comme une paire d'entiers, c'est à dire si $w = a_0 b_0 a_1 b_1 \cdots a_n b_n$, on note $[w] = \langle a, b \rangle$, où $a = \sum_{k=0}^{n} a_k \cdot 2^k$ et b analogue. Par convention $[\varepsilon] = \langle 0, 0 \rangle$. D'ailleurs on note $A(w) = a$ et $B(w) = b$ si $[w] = \langle a, b \rangle$.

(a) Construire un automate quelconque pour

$$L := \{\, w \in \Sigma^{2n} \mid n \geq 0,\ A(w) + B(w) = 0 \pmod{2^n} \,\}.$$

Solution : On construit un automate qui simule l'addition habituelle de $A(w)$ et $B(w)$. Les états acceptants exigent que le nombre de lettres est pair et que la somme donne 0 sur les n paires de positions évaluées. Les états r_0, r_1 mémorisent la retenue après toute paire de chiffres, en démarrant avec r_0. Les états i_0, i_1, i_2 mémorisent la somme intermédiaire de la retenue et le prochain bit de $A(w)$. Si la somme de la retenue et de deux bits est impaire, on utilise l'état poubelle p.



(b) En construire l'automate déterministe complet minimal pour L (ou bien justifier pourquoi votre automate dans (a) est cet automate).

Solution : L'automate dans (a) est bien détermiste et complet, et il est aussi minimal. Pour montrer ce dernier, on montre que tout état tombe dans une classe différente de Nérode :
— r_0 est le seul état à accepter le mot 00 ;
— r_1 est le seul état à accepter le mot 10 ;
— i_0 est le seul état à accepter le mot 000 ;
— i_1 est le seul état à accepter le mot 1 ;
— i_2 est le seul état à accepter le mot 001 ;
— p est le seul état à n'accepter aucun mot.

(c) Soient $k, \ell \in \mathbb{N}^+$. Montrer que le langage

$$L_{k,\ell} := \{\, w \in \Sigma^{2n} \mid n \geq 0,\ k \cdot A(w) + \ell \cdot B(w) = 0 \pmod{2^n} \,\}$$

est accepté par un automate avec au plus $3 \cdot (k + \ell)$ états.

Solution : On généralise le principe de l'automate de (a), sauf que les bits de $A(w), B(w)$ possèdent les poids k, ℓ, respectivement. Pour chaque pair de lettres, on calcule la somme s de la retenue et ces deux poids, on exige que s soit paire, et la nouvelle retenue devient $s/2$. La retenue est désormais entre 0 et $k + \ell - 1$, en effet, elle n'exède jamais cette borne car $(k + \ell - 1) + k + \ell < 2(k + \ell)$. $L_{k,\ell}$ est donc accepté par l'automate DC $\mathcal{A}_{k,\ell} = \langle R \cup I \cup \{p\}, \delta, r_0, R\rangle$, où
— $R = \{r_0, \ldots, r_{k+\ell-1}\}$;
— $I = \{i_0, \ldots, i_{2k+\ell-1}\}$;
— pour $b = 0, 1$, $j = 0, \ldots, k + \ell - 1$, $j' = 0, \ldots, 2k + \ell - 1$:
  — $\delta(r_j, b) = i_{j+kb}$;
  — $\delta(i_{j'}, b) = \begin{cases} i_{(j'+\ell b)/2} & \text{si } j' + \ell b \text{ pair} \\ p & \text{sinon} \end{cases}$;
  — $\delta(p, b) = p$.

On vérifie que $\mathcal{A}_{k,\ell}$ possède $3k + 2\ell + 1 \leq 3(k + \ell)$ états (avec $l \geq 1$).

Encore mots de longueur paire, si $w = a_0 \cdots a_n b_0 \cdots b_n$, notons $[\![w]\!] = \langle a, b\rangle$, avec a, b comme avant, et $A'(w) = a$, $B'(w) = b$ si $[\![w]\!] = \langle a, b\rangle$. On définit, pour $k \in \mathbb{N}$:

$$L' := \{\, w \in \Sigma^{2n} \mid n \geq 0,\ A'(w) + B'(w) = 0 \pmod{2^n} \,\}$$

$$L_k := \{\, w \in \Sigma^{2n} \mid n \geq 0,\ A'(w) + B'(w) = 0 \pmod{2^k} \,\}$$

(d) L', est-il reconnaissable ? Et L_k, pour k quelconque ?

Solution : Supposons que L' est reconnaissable, et soit N la constante du lemme de l'étoile. On considère $w = 0^N 1 0^N 1$, alors $|w| = 2(N+1)$ et $A'(w) = B'(w) = 2^N$ et $A'(w) + B'(w) = 0 \pmod{2^{N+1}}$, du coup $w \in L'$. Alors par le lemme de l'étoile il existe un k t.q. $w' = 0^{N+k} 1 0^N 1 \in L'$, et k doit être pair. Mais $A'(w') = 0$ et $B'(w') > 0$, alors $w' \notin L'$, ce qui est absurde. Du coup, L' n'est pas reconnaissable.
Certains ont remarqué que L_0 contient tous les mots de longueur paire, ce qui est reconnaissable (j'aurais voulu écrire $k \in \mathbb{N}^+ \ldots$).

Sinon, par un raisonnement similaire, L_k n'est pas reconnaissable pour $k \geq 1$. On considère $w = 10^{k-1}0^N 1^k 0^N$, alors $A'(w) = 1$, $B'(w) = 2^k - 1$, du coup $w \in L_k$. Par le lemme de l'étoile il existe un k (pair) t.q. $w' = 10^{k-1}0^{N+k}1^k 0^N \in L_k$. Mais $A'(w') = 1$ est impair et $B'(w')$ est pair, du coup $w' \notin L_k$, ce qui est absurde.

2 Mélange de mots et monoïdes

Soient $u, v \in \Sigma^*$, pour un alphabet Σ quelconque, et $K, L \subseteq \Sigma^*$ des langages, et $\phi\colon \Sigma^* \to M$, $\psi\colon \Sigma^* \to M'$ des morphismes reconnaissant K resp. L.

Pour rappel, on dit que u est *sous-mot* de v, noté $u \sqsubseteq v$, si les lettres de u apparaissent dans v dans l'ordre, c'est à dire $v = v_0 u_1 v_1 u_2 \cdots u_n v_n$ avec $u_i, v_i \in \Sigma^*$ et $u = u_1 u_2 \cdots u_n$.

(a) En construire un morphisme reconnaissant le langage

$$K' := \{\, v \mid \exists u \in K : u \sqsubseteq v \,\}.$$

Solution : Pour $w \in \Sigma^*$, définissons $[\![w]\!] := \{\, v \mid v \sqsubseteq w \,\}$. On voit bien que $[\![wa]\!] = [\![w]\!] \cup \{\, va \mid v \sqsubseteq w \,\}$, pour $a \in \Sigma$.
On va montrer que K' est accepté par le morphisme $\phi' : \Sigma^* \to 2^M$ engendré par $\phi'(a) = \{\phi(\varepsilon), \phi(a)\}$, pour $a \in \Sigma$, avec l'opération $A \cdot B = \{\, ab \mid a \in A, b \in B \,\}$. Cette opération est bien associatif car l'opération sur M l'est, et son élément neutre est $\{\phi(\varepsilon)\} =: \phi'(\varepsilon)$.
On prouvera par récurrence que pour tout mot $w \in \Sigma^*$, $\phi'(w) = \phi([\![w]\!])$. Effectivement, c'est vrai pour tout mot de longueur 0 ou 1 par définition. Soit alors w un mot satisfaisant cette propriété. Alors

$$\begin{aligned} \phi'(wa) = \phi'(w) \cdot \phi'(a) &= \phi([\![w]\!]) \cdot \{\phi(\varepsilon), \phi(a)\} \\ &= \phi([\![w]\!]) \cup \{\, \phi(va) \mid v \sqsubseteq w \,\} = \phi([\![wa]\!]) \end{aligned}$$

Pour conclure, montrons que $w \in K'$ ssi $\phi'(w) \cap \phi(K) \neq \emptyset$. Effectivement, ceci est le cas ssi $\exists u \in \phi'(w)$ t.q. $\phi(u) \in \phi(K)$ ssi $u \sqsubseteq w$ et $u \in K$.

Le *mélange* $u \sqcup\!\sqcup\, v$ désigne l'ensemble des mots qu'on obtient en intercalant u et v :

$$u \sqcup\!\sqcup\, v = \{\, u_1 v_1 \cdots u_k v_k \mid k \geq 0,\ \forall i : u_i, v_i \in \Sigma^*,\ u_1 \cdots u_k = u,\ v_1 \cdots v_k = v \,\}$$

On étend cette notion aux langages avec $K \sqcup\!\sqcup\, L = \bigcup \{\, u \sqcup\!\sqcup\, v \mid u \in K,\ v \in L \,\}$.

(b) En construire un morphisme reconnaissant le langage $K \sqcup\!\sqcup\, L$.

Solution : Le raisonnement est similaire à celui de (a) – en effet on pourrait obtenir le résultat de (a) comme un cas spécial de (b).
Définissons alors $[\![w]\!] = \{\, \langle u, v \rangle \mid w \in u \sqcup\!\sqcup\, v \,\}$. On voit bien $[\![\varepsilon]\!] = \{\langle \varepsilon, \varepsilon \rangle\}$ et $[\![wa]\!] = \{\, \langle ua, v \rangle, \langle u, va \rangle \mid \langle u, v \rangle \in [\![w]\!] \,\}$.

Construisons $\chi : \Sigma^* \to 2^{M \times M'}$ engendré par $\chi(a) := \{\langle \phi(a), \psi(\varepsilon)\rangle, \langle \phi(\varepsilon), \psi(a)\rangle\}$ avec l'opération $A \cdot B := \{\, \langle ab, a'b'\rangle \mid \langle a, a'\rangle \in A, \langle b, b'\rangle \in B \,\}$. De nouveau, l'associativité suit de celle de M et M', et son élément neutre est $\{\langle \phi(\varepsilon), \psi(\varepsilon)\rangle\} =: \chi(\varepsilon)$. Prouvons par récurrence que $\chi(w) = \{\, \langle \phi(u), \psi(v)\rangle \mid \langle u, v\rangle \in [\![w]\!] \,\}$. Encore c'est vrai par définition pour les motsde longueur au plus 1. Soit w un mot satisfaisant la propriété, alors

$$\begin{aligned} \chi(wa) = \chi(w) \cdot \chi(a) &= \chi(w) \cdot \{\langle \phi(a), \psi(\varepsilon)\rangle\} \cup \chi(w) \cdot \{\langle \phi(\varepsilon), \psi(a)\rangle\} \\ &= \{\, \langle \phi(ua), \psi(v)\rangle, \langle \phi(u), \psi(va)\rangle \mid \langle u, v\rangle \in [\![w]\!] \,\} \\ &= \{\, \langle \phi(u'), \psi(v')\rangle \mid \langle u', v'\rangle \in [\![wa]\!] \,\}. \end{aligned}$$

Pour montrer que χ reconnait $K \sqcup\!\sqcup L$, on montre que $w \in K \sqcup\!\sqcup L$ ssi $\chi(w) \cap (\phi(K) \times \psi(L)) \neq \emptyset$. Effectivement, ceci est le cas ssi $\exists u, v$ t.q. $\langle \phi(u), \psi(v)\rangle \in \chi(w)$ et $u \in \phi(K), v \in \psi(L)$ ssi $w \in u \sqcup\!\sqcup v$ et $u \in K$, $v \in L$.

Si $\Sigma = \{a_1, \ldots, a_k\}$, on associe à un mot u le k-vecteur indiquant le nombre d'occurrences de toute lettre :

$$\chi(u) := \langle |u|_{a_1}, \ldots, |u|_{a_k}\rangle$$

Par exemple, $\chi(abad) = \langle 2, 1, 0, 1\rangle$ si $\Sigma = \{a, b, c, d\}$. Pour un langage K, on note $\chi(K) = \{\, \chi(u) \mid u \in K \,\}$. On utilise l'addition et multiplication scalaire habituelle entre vecteurs. Un vecteur de base est 1 dans une seule dimension et 0 dans les autres.

Un ensemble $S \subseteq \mathbb{N}^k$ de vecteurs est *linéaire* s'il existe des vecteurs $x_0, x_1, \ldots, x_n$ t.q.

$$S = \{\, x_0 + b_1 x_1 + \cdots + b_n x_n \mid b_1, \ldots, b_n \in \mathbb{N} \,\}$$

D'ailleurs, S est *ultralinéaire* si $x_0, x_1, \ldots, x_n$ sont des vecteurs de base et *semilinéaire* si S est une union fini d'ensembles linéaires.

(c) Soit S un ensemble (i) ultralinéaire (ii) linéaire (iii) semilinéaie. En général, $\chi^{-1}(S)$ est-il reconnaissable ?

Solution :

(i) Ultralinéaire : (J'aurais voulu écrire que $x_1, \ldots, x_n$ sont des multiples de vecteurs de base, sans contrainte sur x_0, mais ma cervelle ne fonctionnait plus. Heureusement, ce ne fait pas une grande différence.)

Soit alors S ultralinéaire et soient $a_{j_0}, \ldots, a_{j_n}$ les lettres de Σ que x_i égale 1 dans sa j_i-ème dimension, pour $i = 0, \ldots, n$. Alors tout mot dans $\chi^{-1}(S)$ contient au moins une occurrence de a_{j_0} et des nombres arbitraires des lettres das $\{a_{j_1}, \ldots, a_{j_n}\}$ (le fait que plusieurs lettres dans cet ensemble peuvent être identiques ne change rien). Du coup $\chi^{-1}(S)$ est le langage de l'expression régulière $\sum_{i=1}^{n} a_{j_i} \cdot a_{j_0} \cdot \sum_{i=1}^{n} a_{j_i}$ et est donc reconnaissable.

(ii) Linéaire : Considérons le cas $k = 2$. L'ensemble $S = \{\, \langle 0, 0\rangle + b_1 \cdot \langle 1, 1\rangle \mid b_1 \in \mathbb{N} \,\}$ est linéaire, et $\chi^{-1}(S) \cap a^*b^* = \{\, a^n b^n \mid n \geq 0 \,\}$. Comme ce dernier n'est pas reconnaissable mais a^*b^* l'est, alors $\chi^{-1}(S)$ ne peut pas être reconnaissable.

(iii) Les ensembles linéaires étant un cas spécial des semilinéaires, le résultat négatif de (ii) s'étend à ce cas aussi.

(d) Montrer que pour tout ensemble semilinéaire S il existe un langage reconnaissable L tel que $\chi(L) = S$.

Solution : Pour un vecteur $x = \langle x_1, \ldots, x_k \rangle \in \mathbb{N}^k$, définissions le mot $\alpha(x) = a_1^{x_1} \cdots a_k^{x_k}$. Regardons d'abord le cas où S est linéaire, de la forme $S = \{ x_0 + b_1 x_1 + \cdots + b_n x_n \mid b_1, \ldots, b_n \in \mathbb{N} \}$. Soit $L := \mathcal{L}(\alpha(x_0) \cdot \alpha(x_1)^* \cdots \alpha(x_n)^*)$, alors $\chi(L) = S$. Pour un ensemble semilinéaire, le résultat suit du fait que les langages reconnaissables sont clos sous l'union fini.