ÉCOLES NORMALES SUPÉRIEURES

CONCOURS D'ADMISSION 2015 — FILIÈRE MP – SPÉCIALITÉ INFO

COMPOSITION D'INFORMATIQUE-MATHÉMATIQUES (ULCR)

(Durée : 4 heures)

L'utilisation des calculatrices **n'est pas autorisée** pour cette épreuve.
Le langage de programmation choisi par le candidat doit être spécifié en tête de la copie.

⋆⋆⋆

Automates et matrices pondérés

Ce sujet comporte six pages et quatre parties. Il porte sur l'étude d'automates pondérés, de leur modélisation mathématique et de certains de leurs comportements asymptotiques.

La partie I introduit la notion d'**automate pondéré**, de langage pondéré (à chaque mot est associé un poids) et établit leurs premières propriétés. La partie II s'intéresse au poids d'un mot choisi aléatoirement. La partie III fait le lien entre un automate et sa représentation matricielle, et enfin, la partie IV s'intéresse au comportement asymptotique des automates et des matrices pondérés.

Les notions introduites en partie I sont utilisées dans les parties II et III. La partie IV s'appuie sur certains résultats de la partie III. Les résultats d'une question pourront être admis dans la suite du sujet.

Partie I. Automates pondérés

On appelle **automate pondéré** un automate fini dont les arêtes ont un poids. Plus précisément, c'est un sextuplet $\mathcal{A} = (Q, \Sigma, I, F, E, W)$, où

— Q est l'ensemble fini des **états** ;
— Σ est un alphabet fini ;
— $I, F \subseteq Q$ sont respectivement les états initiaux et terminaux de l'automate ;
— $E \subseteq Q \times \Sigma \times Q$ sont les transitions de l'automate ;
— $W : E \to \mathbb{R}$ est la **fonction de pondération** qui à chaque transition associe un nombre réel.

Un **chemin** dans l'automate $\mathcal{A}$ est une suite finie de transitions $s = (e_1, e_2, \dots, e_k)$ avec $k \in \mathbb{N}$, $e_i = (p_i, a_i, q_i) \in E$ et tels que $\forall i \in \{1 \dots, k-1\}$, $q_i = p_{i+1}$. Ce chemin est un **cycle** si $p_1 = q_k$ et c'est un **cycle élémentaire** s'il n'existe pas $1 \leq i < j \leq k$ tel que $p_i = p_j$. Enfin, ce chemin est **sans cycle** s'il n'existe pas $1 \leq i \leq j \leq k$ tel que $(e_i, \dots, e_j)$ est un cycle.

L'entier k est la **longueur de ce chemin** ; le mot $u = a_1 \dots a_k$ est le **mot (ou étiquette) de ce chemin** ; et le **poids de ce chemin** est

$$W(s) = \sum_{i=1}^{k} W(e_i), \qquad \text{avec la convention} \sum_{i=1}^{0} W(e_i) = 0.$$

Ce chemin est dit **acceptant** si $q_1 \in I$ et $q_k \in F$. Si un mot $u \in \Sigma^*$ est **accepté** par l'automate s'il existe un chemin acceptant de mot u. Le poids $T_{\mathcal{A}}(u)$ d'un mot u dans l'automate est le poids maximal d'un chemin acceptant de mot u. Si un tel chemin n'existe pas, alors par convention, le poids du mot est $-\infty$:

$$T_{\mathcal{A}}(u) = \max \{W(s) \mid s \text{ chemin acceptant de mot } u \text{ dans } \mathcal{A}\}.$$

Un **langage pondéré** L sur un alphabet fini Σ^* est un ensemble tel qu'il existe $w : \Sigma^* \to \mathbb{R} \cup \{-\infty\}$ tel que $L = \{(u, w(u)) \mid u \in \Sigma^*\}$. Le poids de $u \in \Sigma^*$ dans L est $L(u) = w(u)$, et on note $Supp(L) = \{u \in \Sigma^* \mid L(u) \neq -\infty\}$ le support du langage.

Le langage $L_{\mathcal{A}} = \{(u, T_{\mathcal{A}}(u)) \mid u \in \Sigma^*\}$ est le langage pondéré reconnu par l'automate : pour tout mot $u \in \Sigma^*$, $L_{\mathcal{A}}(u) = T_{\mathcal{A}}(u)$.

On supposera dans tout le problème que tous les états des automates pondérés sont utiles : pour tout état $q \in Q$, il existe $i \in I$ et $f \in F$ tels qu'il existe un chemin de i à q et un chemin de q à f.

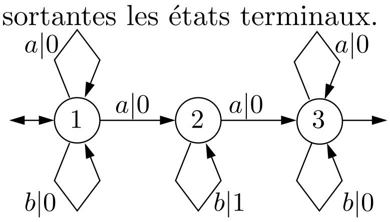
Question 1

a. Montrer que le support du langage d'un automate pondéré est un langage reconnu par un automate fini.

b. Réciproquement, montrer que si L est un langage reconnu par un automate fini, il est possible de trouver un automate pondéré $\mathcal{A}$ de support L tel que $u \in L \Leftrightarrow L_{\mathcal{A}}(u) = 0$.

Question 2 On pose $\Sigma = \{a, b\}$.

a. Quel est le langage pondéré sur Σ^* reconnu par l'automate pondéré suivant ? La notation "$a|w$" signifie que la transition est étiquetée par a et de poids w. Les flèches entrantes désignent les états initiaux et les flèches sortantes les états terminaux.



b. Donner un automate pondéré sur Σ^* reconnaissant le langage $\{(u, \max(|u|_a, |u|_b)), u \in \Sigma^*\}$, où $|u|_a$ (respectivement $|u|_b$) est le nombre d'occurrences de a (respectivement b) dans u.

On fixe $c \in \mathbb{R}$. On s'intéresse au problème de savoir si tous les poids des mots reconnus par un automate pondéré sont inférieurs à c.

On note L le langage pondéré reconnu par l'automate pondéré $\mathcal{A}$.

Question 3

a. Montrer que pour tout mot u, $L(u) \leq c$ si et seulement si tous les chemins acceptants de $\mathcal{A}$ d'étiquette u sont de poids inférieur ou égal à c.

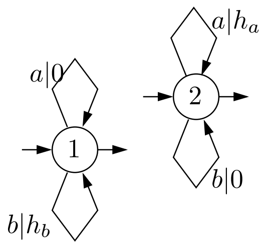
b. Montrer que l'on a l'équivalence entre les deux assertions :

(i) $\forall u \in \Sigma^*$, $L(u) \leq c$.

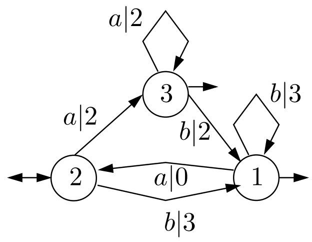
(ii) tous les cycles élémentaires de l'automate pondéré sont négatifs ou nuls et tous les chemins acceptants sans cycle sont de poids inférieur ou égal à c.

Partie II. Poids asymptotique d'un mot choisi aléatoirement

Dans cette partie on s'intéresse aux deux automates $\mathcal{A}_1$ et $\mathcal{A}_2$ suivants, où $h_a, h_b \in \mathbb{N}$.



$\mathcal{A}_1$



$\mathcal{A}_2$

Soient $\Sigma = \{a, b\}$ et $k \in \mathbb{N} \setminus \{0\}$. Considérons l'espace probabilisé $(\Sigma^k, \mathcal{P}(\Sigma^k), \mathbf{P})$ tel que pour tout $i \in \{0, \ldots, k-1\}$, $\mathbf{P}(\Sigma^i a \Sigma^{k-i-1}) = p$ et $\mathbf{P}(\Sigma^i b \Sigma^{k-i-1}) = 1 - p = q$ et tel que les événements $E_i = \Sigma^i a \Sigma^{k-i-1}$ forment une famille mutuellement indépendante. Si X est une variable aléatoire sur $\mathbb{N}$, on note $\mathbf{E}[X]$ son espérance et $\mathbf{V}\mathrm{ar}(X)$ sa variance.

Question 4 Soit $u \in \Sigma^k$. Calculer $\mathbf{P}(u)$.

Soit X_k une variable aléatoire sur Σ^k.

Question 5 On s'intéresse dans cette question à l'automate $\mathcal{A}_1$. On pose $\alpha = \max(p\, h_a, q\, h_b)$.

a. Calculer $\mathbf{E}[|X_k|_a]$ et $\mathbf{V}\mathrm{ar}(|X_k|_a)$. (On rappelle que $|X_k|_a$ est le nombre d'occurrences de a dans X_k.)

b. Montrer que pour tout $\beta > p$, il existe $\delta_1 > 0$ tel que $\mathbf{P}(|X_k|_a \geq \beta k) \leq \frac{\delta_1}{k}$ et que pour tout $\gamma < p$, il existe $\delta_2 > 0$ tel que $\mathbf{P}(|X_k|_a \leq \gamma k) \leq \frac{\delta_2}{k}$.

c. Quel est le langage pondéré reconnu par l'automate $\mathcal{A}_1$?

d. Déduire des questions précédentes que $\lim_{k\to\infty} \frac{\mathbf{E}[L_{\mathcal{A}_1}(X_k)]}{k} = \alpha$ (Indication : on pourra d'abord montrer que pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que $\mathbf{P}(|L_{\mathcal{A}_1}(X_k) - \alpha k| \geq \epsilon k) \leq \frac{\delta}{k}$).

Question 6 On s'intéresse maintenant à l'automate $\mathcal{A}_2$. On note Y_i l'état obtenu après lecture du mot $X[i]$, le préfixe de X_k de longueur i, (ici, on a donc $Y_0 = 2$).

a. Pour tout $i \in \{1, \ldots, k\}$, calculer $\mathbf{P}(Y_i = 1)$.

b. En déduire $\mathbf{P}(Y_i = 2)$ et $\mathbf{P}(Y_i = 3)$.

c. En déduire $\lim_{k\to\infty} \frac{\mathbf{E}[L_{\mathcal{A}_2}(X_k)]}{k}$. (Indication : on pourra s'intéresser à $\mathbf{E}[L_{\mathcal{A}_2}(X[i]) - L_{\mathcal{A}_2}(X[i-1])]$).

Partie III. Matrices pondérées

Soit $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\}$. On appelle **matrice pondérée** une matrice à coefficients dans $\overline{\mathbb{R}}$. Soient $B, C \in \overline{\mathbb{R}}^{n,n}$ deux matrices pondérées. On définit les opérations suivantes :
- $B \oplus C$ par $\forall i, j \in \{1, \ldots, n\}$, $(B \oplus C)_{i,j} = \max(B_{i,j}, C_{i,j})$;
- $B \otimes C$ par $\forall i, j \in \{1, \ldots, n\}$, $(B \otimes C)_{i,j} = \max_{k\in\{1,\ldots,n\}}(B_{i,k} + C_{k,j})$.

Un **vecteur pondéré** est un élément de $\overline{\mathbb{R}}^{1,n}$ (vecteur ligne) ou $\overline{\mathbb{R}}^{n,1}$ (vecteur colonne). On définit de la même manière
- pour $B \in \overline{\mathbb{R}}^{n,n}$ et $u \in \overline{\mathbb{R}}^{n,1}$, $B \otimes u$ par $\forall i \in \{1, \ldots, n\}$, $(B \otimes u)_i = \max_{j\in\{1,\ldots,n\}}(B_{i,j} + u_j)$;
- pour $B \in \overline{\mathbb{R}}^{n,n}$ et $u \in \overline{\mathbb{R}}^{1,n}$, $u \otimes B$ par $\forall i \in \{1, \ldots, n\}$, $(u \otimes B)_i = \max_{j\in\{1,\ldots,n\}}(u_j + B_{j,i})$.

Question 7

a. Montrer que l'opération $\oplus$ est associative : pour tous $B, C, D \in \overline{\mathbb{R}}^{n,n}$, $(B \oplus C) \oplus D = B \oplus (C \oplus D)$.

b. Montrer que l'opération $\oplus$ est commutative : pour tous $B, C \in \overline{\mathbb{R}}^{n,n}$, $B \oplus C = C \oplus B$.

c. Montrer que l'opération $\otimes$ est associative : pour tous $B, C, D \in \overline{\mathbb{R}}^{n,n}$, $(B \otimes C) \otimes D = B \otimes (C \otimes D)$.

d. Montrer que l'opération $\otimes$ est distributive sur $\oplus$: $B, C, D \in \overline{\mathbb{R}}^{n,n}$, $B \otimes (C \oplus D) = (B \otimes C) \oplus (B \otimes D)$ et $(C \oplus D) \otimes B = (C \otimes B) \oplus (D \otimes B)$.

e. Donner une matrice $U \in \overline{\mathbb{R}}^{n,n}$ telle que pour tout $B \in \overline{\mathbb{R}}^{n,n}$, $B \otimes U = U \otimes B = B$.

Soit $M \in \overline{\mathbb{R}}^{n,n}$. On pose $M^0 = U$ et pour tout $k \in \mathbb{N}$, $M^{k+1} = M \otimes M^k$.

Une matrice $M \in \overline{\mathbb{R}}^{n,n}$ peut être représentée par un graphe orienté et pondéré $\mathcal{G}(M) = (S, A, w)$ dont l'ensemble des sommets est $S = \{1, \ldots, n\}$, l'ensemble des arcs est $A = \{(i,j) \mid M_{i,j} \neq -\infty\}$, et la fonction de pondération est $w : A \to \mathbb{R}; (i,j) \mapsto M_{i,j}$.

Question 8 Montrer que $(M^k)_{i,j}$ est égal au poids maximal d'un chemin de longueur k de i vers j dans $\mathcal{G}(M)$ (on utilisera la convention qu'un chemin de longueur 0 est de poids nul et que s'il n'y a pas de chemin de i vers j de longueur k, ce poids est $-\infty$).

Soit $\mathcal{A} = (Q, \Sigma, I, F, E, W)$ un automate pondéré avec $Q = \{1, \ldots, n\}$. Pour tout $a \in \Sigma$, soit $\mathcal{M}(a)$ la matrice du graphe pondéré obtenu en ne gardant que les transitions d'étiquette

a : $\mathcal{G}(\mathcal{M}(a)) = (Q, A, w)$, avec $A = \{(p, q) \mid (p, a, q) \in E\}$ et $w : (p, q) \mapsto W(p, a, q)$. Pour tout $u = a_1 \dots a_k \in \Sigma^*$, on pose $\mathcal{M}(u) = \mathcal{M}(a_1) \otimes \cdots \otimes \mathcal{M}(a_k)$ avec la convention $\mathcal{M}(u) = U$ si u est le mot vide.

Question 9

a. Soient $p, q \in Q$. Montrer que le poids maximum d'un chemin d'étiquette $u \in \Sigma^*$ de p vers q est égal à $\mathcal{M}(u)_{p,q}$.

b. En déduire une expression matricielle pour $L_{\mathcal{A}}(u)$, c'est-à-dire de la forme $M_1 \otimes \cdots \otimes M_\ell$ où M_i sont des matrices ou des vecteurs.

On suppose dans la question suivante que tous les cycles dans l'automate sont de poids négatif ou nul.

Question 10

a. Considérons la matrice $M = \oplus_{a\in\Sigma}\mathcal{M}(a)$ et $M^* = \oplus_{k\in\mathbb{N}} M^k$. Montrer que M^* est bien définie (c'est-à-dire que ses coefficients sont dans $\overline{\mathbb{R}}$).

b. Donner une expression matricielle pour le poids du mot le plus lourd reconnu par l'automate $\mathcal{A}$.

c. Soit $M_{i,j}^{\leq p}$ le poids maximal d'un chemin de i vers j passant uniquement par des sommets intermédiaires q tels que $q \leq p$ (i et j non compris). On pose $M^{\leq p} = (M_{i,j}^{\leq p})_{i,j\leq n}$. Exprimer $M_{i,j}^{\leq p}$ en fonction de $M^{\leq p-1}$. En déduire un algorithme par programmation dynamique qui calcule M^*.

d. Adapter cet algorithme pour qu'il réponde `vrai` si tous les mots reconnus par l'automate sont de poids inférieur ou égal à c et `faux` sinon, sans faire d'hypothèse a priori sur les poids des cycles de l'automate.

Partie IV. Croissance asymptotique du poids des mots

On dit qu'une matrice pondérée est **irréductible** si le graphe orienté qui lui est associé est fortement connexe. Soit M une matrice pondérée irréductible. On suppose dans les questions 11 à 14 que le cycle de poids maximum dans ce graphe est **de poids exactement égal à zéro**.

On appelle vecteur propre de M associé à la valeur propre $\lambda \in \mathbb{R}$ un vecteur $v \neq (-\infty, \dots, -\infty)$ tel que $M \otimes v = \lambda \otimes v$, où $\lambda \otimes v$ est le vecteur $(\lambda + v_1, \dots, \lambda + v_n)$.

Soient $v, v' \in \overline{\mathbb{R}}^{n,1}$. On note $v \leq v'$ si pour tout $i \in \{1, \dots, n\}$, $v_i \leq v'_i$ et de la même manière, on note $v \geq v'$ si pour tout $i \in \{1, \dots, n\}$, $v_i \geq v'_i$.

Question 11 Soient $v \neq (-\infty, \dots, -\infty)$ et $\lambda \in \mathbb{R}$ tels que $M \otimes v \leq \lambda \otimes v$.

a. Montrer que pour tout $k \in \mathbb{N}$, $M^k \otimes v \leq (k\lambda) \otimes v$ puis que $\forall i \in \{1, \dots, n\}$, $v_i \neq -\infty$.

b. En déduire que $\lambda \geq 0$.

Question 12 Soient $v \neq (-\infty, \dots, -\infty)$ et $\lambda \in \mathbb{R}$ tels que $M \otimes v \geq \lambda \otimes v$.

a. Montrer que pour tout $i \in \{1, \ldots, n\}$ il existe $j \in \{1, \ldots, n\}$ tel que $\lambda + v_i \leq M_{i,j} + v_j$.
b. En déduire que $\lambda \leq 0$.

Soit I^* l'ensemble des sommets de $\mathcal{G}(M)$ qui sont sur un cycle de poids 0.

Question 13 Montrer que pour $i \in I^*$, le i-ème vecteur colonne de M^* est un vecteur propre pour M. Quel est l'ensemble des valeurs propres de M ?

Pour $i, j \in \{1, \ldots, n\}$ et $i^* \in I^*$, on note $(M^k)_{i,i^*,j} = \max_{k_1+k_2=k}(M^{k_1})_{i,i^*} + (M^{k_2})_{i^*,j}$ le poids du chemin le plus lourd de i vers j de longueur k et passant par i^*.

Question 14
a. Montrer qu'il existe $d \in \mathbb{N} \setminus \{0\}$ tel que pour tout $i^* \in I^*$, $(M^d)_{i^*,i^*} = 0$.
b. Soient $i, j \in \{1, \ldots, n\}$ et $i^* \in I^*$. Montrer qu'il existe $K > 0$ tel que pour tout $k \geq K$, $(M^{(k+1)d})_{i,i^*,j} = (M^{kd})_{i,i^*,j}$.
c. Soient $i, j \in \{1, \ldots, n\}$. Montrer qu'il existe K' tel que pour tout $k \geq K'$, $M^{kd}_{i,j} = \max_{i^* \in I^*}(M^{kd})_{i,i^*,j}$.
d. En déduire qu'il existe K_0 et d tels que pour tout $k \geq K_0$, $M^{k+d} = M^k$.

On se place maintenant dans le cas où le cycle de poids maximal n'est plus égal à 0, mais la matrice M est toujours irréductible. On pose

$$\rho(M) = \max_{k \leq n-1} \frac{\max_{i \in \{1,\ldots,n\}}(M^k)_{i,i}}{k}.$$

Question 15
a. Donner une interprétation de $\rho(M)$ en terme de graphe.
b. Soit $\lambda \in \mathbb{R}$. Montrer que λ est une valeur propre de M si et seulement si 0 est valeur propre de la matrice $(-\lambda) \otimes M$, où $(-\lambda) \otimes M = (M_{i,j} - \lambda)_{i,j \in \{1,\ldots,n\}}$.
c. En déduire que $\rho(M)$ est l'unique valeur propre de M.
d. Montrer qu'il existe K et d tels que pour tout $k \geq K$, $M^{k+d} = d\rho(M) \otimes M^k$.

Question 16 Montrer que $\rho(M)$ est valeur propre de M même si M n'est pas irréductible.

Soit $\mathcal{A}$ un automate pondéré dont tous les états sont terminaux. Pour toute suite $(u_k)_{k\geq 1} \in \Sigma^{\mathbb{N}}$, on note $u[k] = u_1 \cdots u_k$.

Question 17 Montrer qu'il existe λ tel que :
(i) pour toute suite $(u_k)_{k\geq 1}$, $\lim_{k\to\infty} \frac{L_{\mathcal{A}}(u[k])}{k} \leq \lambda$ (si cette limite existe) ;
(ii) il existe une suite $(u_k)_{k\geq 1}$ telle que $\lim_{k\to\infty} \frac{L_{\mathcal{A}}(u[k])}{k} = \lambda$.

* *
*