Troisième devoir surveillé

[CS21]
4 heures
Calculatrices autorisées

Ce sujet comprend trois parties. Les résultats généraux du I.A sont utilisés dans la partie II et dans la sous-partie III.A. Le III.B.2 peut être traité indépendamment du III.B.1.

Notations

$\mathbb{N}$ désigne l'ensemble des entiers naturels et $\mathbb{R}$ l'ensemble des nombres réels.

Pour m et n deux entiers naturels, $[\![m, n]\!]$ désigne l'ensemble des entiers k tels que $m \leqslant k \leqslant n$.

Pour n un entier naturel non nul, $\mathcal{M}_n(\mathbb{R})$ désigne l'ensemble des matrices carrées d'ordre n à coefficients réels et $\mathcal{O}_n(\mathbb{R})$ désigne l'ensemble des matrices orthogonales d'ordre n. On note I_n la matrice identité d'ordre n.

Pour une matrice A de $\mathcal{M}_n(\mathbb{R})$, on note $A^\top$ sa transposée.

Le module d'un nombre complexe z est noté $|z|$.

Définitions

Un *graphe orienté* G est un ensemble $G = (S, A)$ où S est un ensemble fini dont les éléments s'appellent les *sommets* du graphe G et où $A \subset S^2$. Les éléments de A s'appellent les *arêtes orientées* du graphe G.

Si $a = (s, s') \in A$, a est l'arête orientée reliant le sommet s au sommet s'. Si $(s, s') \in S^2$, on dit que s' est un *sommet voisin* de s s'il existe une arête orientée reliant s à s', c'est-à-dire si $(s, s') \in A$.

On note que s peut être un sommet voisin de lui-même (si $(s, s) \in A$) et que s' peut être un voisin de s alors que s n'est pas un voisin de s' (si $(s, s') \in A$ alors que $(s', s) \notin A$).

I Marche aléatoire sur un graphe

On considère un graphe orienté fini dont les sommets sont numérotés de 1 à n.

Un point se déplace aléatoirement d'un sommet à un autre de ce graphe en suivant les arêtes orientées du graphe. Le nombre d'étapes de cette marche aléatoire peut tendre vers l'infini. À chaque étape, le point se déplace du sommet où il se trouve vers l'un de ses sommets voisins de façon équiprobable. Ceci entraine notamment que la probabilité de passer du sommet i au sommet j ne dépend pas du rang de l'étape.

Pour $1 \leqslant i, j \leqslant n$, on note $t_{i,j}$, la probabilité que le point passe du sommet i au sommet j ; en particulier, s'il n'y a pas d'arête reliant i à j, $t_{i,j} = 0$. La matrice dont le coefficient de la ligne i et de la colonne j est égal à $t_{i,j}$ est notée T. Cette matrice s'appelle la *matrice de transition du graphe*.

Pour $k \in \mathbb{N}$, on note $P^{(k)}$ le vecteur ligne $(p_1^{(k)}, p_2^{(k)}, \ldots, p_{n-1}^{(k)}, p_n^{(k)})$, où, pour $1 \leqslant i \leqslant n$, $p_i^{(k)}$ est la probabilité que le point soit sur le sommet i à l'étape de rang k.

I.A – Résultats généraux

Q 1. Justifier que, pour tout entier naturel k,

$$p_1^{(k)} + \cdots + p_n^{(k)} = 1.$$

Q 2. Montrer que, pour tout tout entier naturel k,

$$P^{(k+1)} = P^{(k)} T.$$

Q 3. En déduire, pour tout entier naturel k, une expression de $P^{(k)}$ en fonction de T, k et $P^{(0)}$.

Q 4. On suppose que la suite de vecteurs $(P^{(k)})_{k\in\mathbb{N}}$ converge vers un vecteur $P = (p_1, \ldots, p_n)$. Montrer que $PT = P$, que pour tout $i \in [\![1, n]\!]$, $p_i \geqslant 0$ et que $p_1 + \cdots + p_n = 1$.

I.B – Marche aléatoire sur un tétraèdre

Dans cette sous-partie, on considère le graphe orienté $G = (S, A)$ où

$$\begin{cases} S = \{1, 2, 3, 4\}, \\ A = \{(1,2), (2,1), (1,3), (3,1), (1,4), (4,1), \\ \qquad (2,3), (3,2), (2,4), (4,2), (3,4), (4,3)\}. \end{cases}$$

La figure 1 représente ce graphe en version complète à gauche et en version simplifiée à droite. Les sommets sont représentés par des cercles et l'arête orientée reliant le sommet s au sommet s', par une flèche de s vers s'.

Par la suite, nous utiliserons la représentation simplifiée dans laquelle si le graphe comporte les deux arêtes orientées (s, s') et (s', s) elles sont représentées par un seul trait avec une flèche à chaque extrémité.

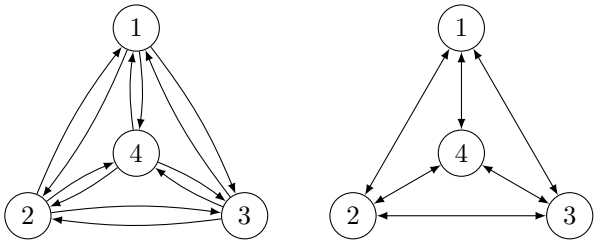

Figure 1

On suppose que, lorsque le point est sur l'un des sommets du graphe, il a la même probabilité de se rendre sur chacun des trois autres sommets du graphe.

On pose

$$J_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Q 5. Exprimer la matrice de transition T en fonction de J_4 et I_4.

Q 6.* Démontrer qu'il existe une matrice $Q \in \mathcal{O}_4(\mathbb{R})$ telle que

$$T = \frac{1}{3} Q \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} Q^\top.$$

Q 7. Montrer que la suite de matrices $(T^k)_{k\in\mathbb{N}}$ converge et identifier géométriquement l'endomorphisme canoniquement associé à la matrice limite.

Q 8. Montrer que, quel que soit le vecteur ligne $P^{(0)} = (p_1^{(0)}, p_2^{(0)}, p_3^{(0)}, p_4^{(0)})$, où pour $1 \leqslant i \leqslant 4$, $p_i^{(0)}$ est la probabilité que le point soit au départ sur le sommet i, la suite $(P^{(k)})_{k\in\mathbb{N}}$ converge vers le vecteur ligne (1/4, 1/4, 1/4, 1/4).

I.C – Marche aléatoire sur une pyramide tronquée à base carrée

Dans cette question, on suppose que G est le graphe représenté figure 2.

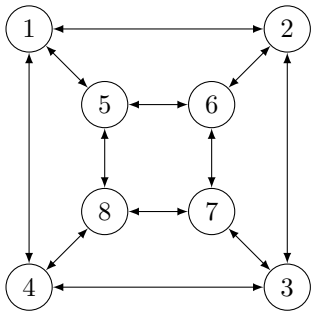

Figure 2

On rappelle que, lorsque le point est sur l'un des sommets du graphe, il a la même probabilité de se rendre sur chacun des sommets à qui il est relié. On suppose qu'au départ, le point est sur le sommet 1, de sorte que

$$P^{(0)} = (1, 0, 0, 0, 0, 0, 0, 0).$$

On note $S_1 = \{1, 3, 6, 8\}$ et $S_2 = \{2, 4, 5, 7\}$.

Q 9. Donner la matrice de transition T de ce graphe et calculer

$$(1, 1, 1, 1, 1, 1, 1, 1)T.$$

Q 10. Montrer que, si le point se trouve sur un sommet de la partie S_1 à une étape donnée, il se trouvera sur un sommet de la partie S_2 à l'étape suivante et que, s'il se trouve sur un sommet de S_2 à une étape donnée, il se trouvera sur un sommet de S_1 à l'étape suivante.

Q 11. La suite de vecteurs $(P^{(k)})_{k\in\mathbb{N}}$ est-elle convergente ?

II Matrices stochastiques et distributions de probabilité

Définitions

Soit $X = (x_1, \ldots, x_n)$ un vecteur de $\mathbb{R}^n$. On dit que X est une distribution de probabilité si pour tout $i \in [\![1, n]\!]$, $x_i \geqslant 0$ et $x_1 + \cdots + x_n = 1$.

Soit M une matrice de $\mathcal{M}_n(\mathbb{R})$. On dit que M est une matrice stochastique si chaque ligne de M est une distribution de probabilité.

II.A –

Q 12. Soit $M \in \mathcal{M}_n(\mathbb{R})$, dont tous les coefficients sont positifs ou nuls. Montrer que M est une matrice stochastique si et seulement si

$$M \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Q 13. Montrer que la matrice de transition d'un graphe, définie dans la partie I, est une matrice stochastique et que, pour tout entier naturel k, le vecteur $P^{(k)}$, lui aussi défini dans la partie I, est une distribution de probabilité.

II.B –

Soient $M \in \mathcal{M}_n(\mathbb{R})$ et $N \in \mathcal{M}_n(\mathbb{R})$ deux matrices stochastiques, $X \in \mathbb{R}^n$ une distribution de probabilité et $\alpha \in [0, 1]$.

Q 14. Montrer que XM est une distribution de probabilité.

Q 15. Montrer que MN est une matrice stochastique.

Q 16. Montrer que $\alpha M + (1 - \alpha) N$ est une matrice stochastique.

II.C –

Soit $M = (m_{i,j})$ une matrice stochastique de $\mathcal{M}_n(\mathbb{R})$ et λ une valeur propre (réelle ou complexe) de M. On note $(u_1, \ldots, u_n)$ les composantes (réelles ou complexes), dans la base canonique, d'un vecteur propre u associé à λ.

II.C.1)

Q 17. Soit $h \in \{1, \ldots, n\}$ tel que $|u_h| = \max_{1\leqslant i\leqslant n} |u_i|$. Montrer que $|\lambda - m_{h,h}| \leqslant 1 - m_{h,h}$. En déduire que $|\lambda| \leqslant 1$.

Q 18. Soit $\delta = \min_{1\leqslant i\leqslant n} m_{i,i}$. Montrer que $|\lambda - \delta| \leqslant 1 - \delta$. Donner une interprétation géométrique de ce résultat et montrer que, si tous les termes diagonaux de M sont strictement positifs, alors 1 est la seule valeur propre de M de module 1.

II.C.2)

On suppose désormais que tous les coefficients $m_{i,j}$ $(1 \leqslant i, j \leqslant n)$ de la matrice stochastique M sont strictement positifs.

Q 19. Démontrer que $\dim(\ker(M - I_n)) = 1$.

Si $(u_1, \ldots, u_n)$ désigne les composantes (réelles) dans la base canonique d'un vecteur de $\ker(M - I_n)$, on pourra utiliser $\min_{1\leqslant i\leqslant n} u_i$.

Q 20. En déduire qu'il existe au plus une distribution de probabilité X invariante par M, c'est-à-dire vérifiant $XM = X$.

On pose $\varepsilon = \min_{1\leqslant i,j\leqslant n} m_{i,j}$.

On s'intéresse à la suite $(M^k)_{k\in\mathbb{N}}$ des puissances de M. On note $m_{i,j}^{(k)}$ le coefficient de la matrice M^k situé à la ligne i et la colonne j.

Pour tout $j \in [\![1,n]\!]$, on pose

$$\begin{cases} \alpha_j^{(k)} = \min\limits_{1\leqslant i\leqslant n} m_{i,j}^{(k)}, \\ \beta_j^{(k)} = \max\limits_{1\leqslant i\leqslant n} m_{i,j}^{(k)}. \end{cases}$$

Dans les quatre questions suivantes, j est un entier fixé dans $[\![1,n]\!]$ et k est fixé dans $\mathbb{N}$.

Q 21. Démontrer les inégalités

$$\alpha_j^{(k)} \leqslant \alpha_j^{(k+1)} \leqslant \beta_j^{(k+1)} \leqslant \beta_j^{(k)}.$$

Q 22. Démontrer qu'il existe un couple $(i_0, j_0) \in [\![1,n]\!]^2$ tel que

$$\alpha_j^{(k+1)} - \alpha_j^{(k)} \geqslant m_{i_0,j_0}\,(\beta_j^{(k)} - \alpha_j^{(k)}).$$

Q 23. Démontrer qu'il existe un couple $(i_1, j_1) \in [\![1,n]\!]^2$ tel que

$$\beta_j^{(k)} - \beta_j^{(k+1)} \geqslant m_{i_1,j_1}\,(\beta_j^{(k)} - \alpha_j^{(k)}).$$

Q 24. En déduire que

$$\beta_j^{(k+1)} - \alpha_j^{(k+1)} \leqslant (1-2\varepsilon)\,(\beta_j^{(k)} - \alpha_j^{(k)}).$$

Q 25. Démontrer que la suite (M^k) converge vers une matrice stochastique

$$B = \begin{pmatrix} b_1 & \cdots & b_n \\ b_1 & \cdots & b_n \\ b_1 & \cdots & b_n \end{pmatrix}$$

dont toutes les lignes sont égales.

On note P^∞ la ligne $(b_1, \ldots, b_n)$.

Q 26. Démontrer que,

$$\forall i \in [\![1,n]\!],\ b_i > 0.$$

Q 27. Démontrer que la suite $(P^{(k)})_{k\in\mathbb{N}} = (P^{(0)}M^k)_{k\in\mathbb{N}}$ converge vers P^∞, quelle que soit la distribution de probabilité initiale $P^{(0)}$.

Q 28. Démontrer que P^∞ est l'unique distribution de probabilité P invariante par M, c'est-à-dire vérifiant $PM = P$.

III Le graphe du web

On modélise le web par un graphe orienté à n sommets représentant chacun une page du web et dont les arêtes orientées représentent les liens hypertextes entre celles-ci. Lorsque la page i contient au moins un lien vers la page j, on dit que la page i *pointe vers* la page j. Cette situation est modélisée par l'existence de l'arête orientée de i vers j, notée $i \to j$. On dit que $i \to j$ est une *arête sortante* de i et une *arête entrante* de j. Si aucun lien de la page i ne pointe vers la page j, on note $i \not\to j$.

Pour tout entier $i \in [\![1,n]\!]$, λ_i désigne le nombre d'arêtes sortantes de la page i, c'est-à-dire le nombre de pages vers laquelle elle pointe. On suppose qu'aucune page ne pointe vers elle-même.

Les moteurs de recherche effectuent un classement entre les différentes pages du web à partir d'une mesure d'importance attribuée à chacune d'elles. Cette mesure d'importance s'appelle la *pertinence* de la page. On se propose d'étudier deux algorithmes permettant de mesurer la pertinence de chaque page du web.

On appelle pertinences des pages du web les éléments de toute suite finie $(\mu_j)_{1\leqslant j\leqslant n}$ vérifiant les deux conditions suivantes :

(i) la pertinence μ_j de la page j est une fonction croissante de la pertinence de chacune des pages qui pointent vers elle ;

(ii) la contribution de la page i dans la pertinence de chacune des pages vers lesquelles elle pointe est une fonction décroissante de λ_i (voir définition ci-dessus).

III.A – Premier modèle de navigation sur le web

On suppose qu'un surfeur navigue sur le web de la manière suivante : lorsqu'il se trouve sur la page i,

— si la page i pointe vers d'autres pages, il se dirige au hasard, de manière équiprobable, vers l'une de ces pages ;

— si la page i ne pointe vers aucune page, il reste sur la page i.

Q 29. Vérifier que la matrice de transition associée à ce modèle de navigation est la matrice $A = (a_{i,j})_{1\leqslant i,j\leqslant n}$ avec

$$\begin{cases} a_{i,i} = \begin{cases} 1 & \text{si la page } i \text{ ne pointe vers} \\ & \text{aucune autre page} \\ 0 & \text{sinon} \end{cases} \\ a_{i,j} = \begin{cases} 0 & \text{si } i \not\to j \\ 1/\lambda_i & \text{si } i \to j \end{cases} \quad \text{pour } i \neq j \end{cases}$$

III.B – L'algorithme PageRank

Selon le modèle précédent, lorsque le surfeur arrive sur une page ne comportant aucun lien vers d'autres pages, il lui est impossible de quitter cette page. Ce modèle n'étant pas conforme à la réalité, on décide de remplacer la matrice A par la matrice

$$B = (1-\alpha)\,A + \frac{\alpha}{n}\,J_n$$

où J_n est la matrice de $\mathcal{M}_n(\mathbb{R})$ dont tous les coefficients sont égaux à 1, A est la matrice stochastique décrite à la question 29 et α est un réel de $]0,1[$, appelé *facteur d'amortissement*.

III.B.1)

Q 30. Montrer que B est une matrice stochastique dont tous les coefficients sont strictement positifs.

Q 31. Dans le modèle de navigation admettant B pour matrice de transition, donner la probabilité de quitter une page ne contenant aucun lien vers une autre page.

Soit Q une distribution de probabilité. On définit la suite $(Q^{(k)})_{k\in\mathbb{N}}$ par $Q^{(k)} = Q\,B^k$ pour tout entier naturel k.

Q 32. Démontrer que la suite $(Q^{(k)})_{k\in\mathbb{N}}$ converge et que sa limite Q^∞ vérifie les conditons (i) et (ii) décrites dans l'introduction de cette partie. Elle fournit donc des pertinences pour les n pages du web. On exprimera la pertinence de chaque page j en fonction de celles des pages qui pointent vers elle en distinguant les pages qui pointent vers une autre page et les autres.

Connu sous le nom de PageRank cet algorithme de calcul des pertinences a été inventé par Sergey Brin et Larry Page, fondateurs de Google.

III.B.2) Calcul de B^k

Dans la pratique, on se contente d'approcher Q^∞ par $Q^{(k)} = Q\,B^k$ pour une valeur de k suffisamment grande.

On suppose que le module `numpy` de Python a été importé à l'aide de l'instruction `import numpy as np`. Ceci donne ainsi accès aux fonctions suivantes :

— `np.identity(n)` crée I_n, la matrice identité d'ordre `n` ;

— `A.shape` donne, sous forme de couple, la taille du tableau `A`, par exemple `np.identity(5).shape` → `(5, 5)` ;

— `np.dot(A, B)` calcule le produit matriciel de `A` et `B` si `A` et `B` sont deux tableaux à deux dimensions compatibles avec le produit matriciel.

L'algorithme naïf de calcul de B^k repose sur sa définition, à savoir :

$$\begin{cases} B^0 = I_n, \\ B^k = B\,B^{k-1} \quad \forall k \geqslant 1. \end{cases}$$

Q 33. Écrire une fonction Python `puissance1(B, k)` qui prend en argument une matrice carrée `B` et un entier naturel `k` et renvoie la matrice B^k calculée en utilisant l'algorithme naïf.

L'algorithme d'exponentiation rapide repose sur le principe suivant :

$$\begin{cases} B^0 = I_n, \\ B^k = (B^2)^{\frac{k}{2}} & \text{si } k \text{ est pair,} \\ B^k = B\,(B^2)^{\frac{k-1}{2}} & \text{si } k \text{ est impair.} \end{cases}$$

Q 34. Écrire une fonction Python `puissance2(B, k)` qui prend en argument une matrice carrée `B` et un entier naturel `k` et renvoie la matrice B^k calculée à partir de l'algorithme d'exponentiation rapide.

Q 35. On suppose que $k \geqslant 2$ et on note p l'unique entier tel que $2^p \leqslant k < 2^{p+1}$. Pour chacun des appels `puissance1(B, k)` et `puissance2(B, k)`, calculer le nombre d'appels à la fonction `np.dot`, dans le pire et dans le meilleur des cas.

• • • FIN • • •