Centrale Maths 1 PSI 2021 — Corrigé

Ce corrigé est proposé par Julie Gauthier (professeur agrégé) ; il a été relu par Angèle Niclas (ENS Lyon) et Benjamin Monmege (enseignant-chercheur à l'université).

Comment fonctionne Google ? Ce sujet décrit l'algorithme PageRank qui permet d'identifier les pages web les plus populaires. Pour cela, on réduit le web à un graphe orienté sur lequel on effectue une marche aléatoire. L'outil central est la notion de matrice stochastique, c'est-à-dire dont tous les coefficients sont positifs et dont la somme des termes d'une ligne vaut toujours 1.

- Dans la partie I, on établit des propriétés générales d'une marche aléatoire sur un graphe en situation d'équiprobabilité. Puis on applique ce modèle à deux exemples géométriques en explicitant les matrices de transition utilisées. On s'intéresse en particulier à la stabilité du comportement sur le temps long.
- La partie suivante introduit la notion de matrice stochastique, dont les matrices de transition vues à la partie I constituent des cas particuliers. Après avoir donné une condition nécessaire et suffisante pour qu'une matrice soit stochastique, et avoir constaté des résultats de stabilité de l'ensemble des matrices stochastiques, on s'intéresse à leur spectre. Pour finir, on étudie la suite constituée des puissances d'une matrice stochastique. En établissant des inégalités sur ses coefficients, on obtient des résultats de convergence.
- Dans la partie III, on travaille sur le graphe du web puis sur l'algorithme PageRank. Cette partie contient des questions de programmation Python et de complexité algorithmique.

Ce sujet porte principalement sur l'algèbre linéaire. Il donne l'occasion de manipuler des suites de matrices. De manière inhabituelle, on multiplie le plus souvent des matrices par des vecteurs ligne à gauche plutôt que par des vecteurs colonne à droite.

Indications

Partie I

1 Utiliser la formule des probabilités totales.

2 Revenir à la définition de la matrice T et utiliser la formule des probabilités composées.

3 Raisonner par récurrence et utiliser le résultat de la question 2.

4 Passer à la limite dans les formules des questions 1 et 2.

6 La matrice de transition T de la question 5 est symétrique réelle. Trouver ses valeurs propres et les dimensions des sous-espaces propres associés pour conclure.

7 La matrice Q étant orthogonale, utiliser la formule de la question 6 pour calculer T^k puis passer à la limite pour k tendant vers $+\infty$. Reconnaître alors la limite obtenue comme une projection sur un sous-espace propre de T.

8 Appliquer le résultat de la question 3 dans cette situation puis passer à la limite en utilisant la limite de la suite $(\mathrm{T}^k)_{k\in\mathbb{N}}$ obtenue à la question 7 ainsi que le résultat de la question 4.

11 Raisonner par l'absurde.

Partie II

12 Exprimer comme une somme chaque coefficient de $\mathrm{M}\,(1,\,\dots,\,1)^{\mathrm{T}}$ puis revenir à la définition d'une matrice stochastique.

14 Écrire explicitement les coefficients du vecteur ligne XM comme des sommes. Sommer ensuite tous les coefficients de XM. Conclure en permutant les sommes finies.

15 Il y a deux méthodes possibles :

- appliquer le résultat de la question 14 à chacune des lignes de M ;
- utiliser la caractérisation de la question 12. Il faudra dans ce cas justifier la positivité des coefficients de MN.

17 Écrire $(\lambda - m_{h,h})u_h$ comme une somme. Majorer cette somme en module en utilisant la définition de u_h. Pour conclure, il faudra utiliser le fait que M est stochastique. L'inégalité $|\lambda| \leqslant 1$ s'obtient à partir de celle sur $|\lambda - m_{h,h}|$. Pour cela, il suffit de décomposer $\lambda = \lambda - m_{h,h} + m_{h,h}$.

18 Utiliser une astuce similaire à celle de la question 17 pour obtenir l'égalité souhaitée. Faire un dessin de la situation en utilisant un disque de centre d'affixe δ. Pour la dernière partie de la question, noter que si tous les termes diagonaux de M sont strictement positifs alors $\delta > 0$ et se ramener au cas d'égalité de l'inégalité triangulaire.

20 Noter que la transposée d'une distribution de probabilité invariante est un vecteur propre de la transposée de M associé à la valeur propre 1. Que dire de la dimension de $\mathrm{Ker}\,(\mathrm{M}^{\mathrm{T}} - \mathrm{I}_n)$ en utilisant la question 19 ?

21 Écrire explicitement les coefficients de $\mathrm{M}^{k+1} = \mathrm{M}\,\mathrm{M}^k$.

22 Définir i_0 à partir de $\alpha_j^{(k)}$ et j_0 à partir de $\beta_j^{(k)}$.

24 Faire apparaître $\beta_j^{(k)}$ et $\alpha_j^{(k)}$ puis utiliser les inégalités obtenues dans les questions 22 et 23.

25 Traiter à part le cas $n = 1$. Dans l'autre cas, utiliser l'inégalité obtenue à la question 24 pour montrer que la suite des différences $\left(\beta_j^{(k)} - \alpha_j^{(k)}\right)_{k\in\mathbb{N}}$ converge vers 0. Conclure en utilisant la définition des $\beta_j^{(k)}$ et $\alpha_j^{(k)}$. Penser à montrer que la limite obtenue est une matrice stochastique.

26 Noter que $(\alpha_j^k)_{k\in\mathbb{N}}$ est croissante. On a supposé que les coefficients de M sont tous strictement positifs. Que peut-on en déduire pour $\alpha_j^{(1)}$?

27 Appliquer le résultat de la question 25.

28 Utiliser les résultats des questions 20 et 27.

Partie III

30 Penser à appliquer les résultats des questions 13 et 16.

32 Appliquer les résultats de la partie II.C.2, en particulier ceux des questions 27 et 28, à la matrice B. Une des difficultés de cette question réside dans la compréhension des propriétés (i) et (ii). En voici une interprétation : pour tout couple d'entiers $(i, j) \in [\![1\,;n]\!]^2$ tel que la page i pointe vers la page j,

(i) si μ_i augmente alors μ_j augmente ;

(ii) si λ_i augmente alors la contribution de la page i dans μ_j diminue.

35 Pour la complexité, on peut raisonner par récurrence. En particulier, il faut montrer que les bornes proposées sont atteintes dans le pire et dans le meilleur des cas pour l'algorithme d'exponentiation rapide.

I. Marche aléatoire sur un graphe

1 Soit $k \in \mathbb{N}$. Par définition, pour tout $i \in [\![1\,;n]\!]$, $p_i^{(k)}$ représente la probabilité que le point soit sur le sommet i à l'étape de rang k. Comme il y a exactement n sommets numérotés de 1 à n, l'ensemble constitué des événements

$$\{\text{Le point est sur le sommet } i\}$$

pour $i \in [\![1\,;n]\!]$ forme un système complet d'événements. Dès lors, la formule des probabilités totales assure que

$$\boxed{\forall k \in \mathbb{N} \qquad \sum_{i=1}^{n} p_i^{(k)} = 1}$$

2 Soit $k \in \mathbb{N}$. Le vecteur ligne $\mathrm{P}^{(k)} = (p_i^{(k)})_{i \in [\![1\,;n]\!]}$ est constitué des probabilités que le point soit sur le sommet i à l'étape de rang k. Pour $i \in [\![1\,;n]\!]$, $p_i^{(k+1)}$ est la probabilité que le point soit sur le sommet i à l'étape de rang $k+1$. Or, pour tout $j \in [\![1\,;n]\!]$, $t_{j,i}$ représente la probabilité que le point soit sur le sommet i à l'étape $k+1$ sachant qu'il était sur le sommet j à l'étape k.

La formule des probabilités composées couplée à la formule des probabilités totales, dans le système complet d'événements $\{\text{Le point se situe sur le sommet } j\}$ pour $j \in [\![1\,;n]\!]$, assure que la probabilité que le point soit sur le sommet i à l'étape de rang $k+1$ s'écrit

$$p_i^{(k+1)} = \sum_{j=1}^{n} p_j^{(k)} t_{j,i}$$

Ceci donne matriciellement

$$\boxed{\forall k \in \mathbb{N} \qquad \mathrm{P}^{(k+1)} = \mathrm{P}^{(k)}\,\mathrm{T}}$$

3 Montrons par récurrence que la propriété

$$\mathscr{P}(k): \quad \mathrm{P}^{(k)} = \mathrm{P}^{(0)}\,\mathrm{T}^k$$

est vraie pour tout $k \in \mathbb{N}$.

- $\underline{\mathscr{P}(0)}$ est vraie car $\mathrm{T}^0 = \mathrm{I}_n$.
- $\underline{\mathscr{P}(k) \Longrightarrow \mathscr{P}(k+1)}$: Soit $k \in \mathbb{N}$. On suppose que $\mathrm{P}^{(k)} = \mathrm{P}^{(0)}\,\mathrm{T}^k$. D'après le résultat de la question 2 et $\mathscr{P}(k)$,

$$\mathrm{P}^{(k+1)} = \mathrm{P}^{(k)}\,\mathrm{T} = \mathrm{P}^{(0)}\,\mathrm{T}^k\,\mathrm{T} = \mathrm{P}^{(0)}\,\mathrm{T}^{k+1}$$

ce qui prouve que $\mathscr{P}(k+1)$ est vraie.

- <u>Conclusion</u> : $\boxed{\forall k \in \mathbb{N} \qquad \mathrm{P}^{(k)} = \mathrm{P}^{(0)}\,\mathrm{T}^k}$

4 On passe à la limite pour k tendant vers $+\infty$ dans la formule de récurrence obtenue à la question 2. Par continuité du produit matriciel, il vient

$$\boxed{\mathrm{P} = \mathrm{P}\,\mathrm{T}}$$

De plus, par définition de $\mathrm{P}^{(k)}$, pour tout $k \in \mathbb{N}$, pour tout $i \in [\![1\,;n]\!]$, on sait que $p_i^{(k)} \geqslant 0$. Par passage à la limite quand k tend vers $+\infty$,

$$\boxed{\forall i \in [\![1\,;n]\!] \qquad p_i \geqslant 0}$$