Calculatrice autorisée, documents autorisés : 3 feuilles recto-verso.

Exercice 1 (11 points environ)

Étant donné une matrice réelle X de dimension $p \times n$, on appelle *décomposition en valeurs singulières (SVD)* de X toute factorisation de X qui est de la forme $X = PLQ'$ et vérifiant :

(i) P et Q sont des matrices de dimensions respectives $(p \times r)$ et $(n \times r)$ telles que : $P'P = Q'Q = I_r$, où I_r désigne la matrice identité d'ordre r.

(ii) L est une matrice diagonale de dimension $(r \times r)$ dont les termes diagonaux $\ell_1, \dots, \ell_r$ sont non nuls, positifs et rangés par ordre décroissant.

(iii) Le rang de X est égal à r.

Les $\alpha^{\text{ème}}$ $(1 \leq \alpha \leq r)$ colonnes de P et Q seront notées respectivement u_α et v_α.

Dans ce qui suit, X désigne une matrice qui est réelle, de dimension $p \times n$ et de rang r.

1. *Unicité de la SVD.*
Dans cette question, nous supposons que $X = PLQ'$ est une SVD de X.

1a.† Montrer que $PL^2P' = \displaystyle\sum_{\alpha=1}^{r} \ell_\alpha^2 \, u_\alpha u'_\alpha$.

1b. Montrer que u_α est vecteur propre de XX' et que sa valeur propre associée est ℓ_α^2.

1c. Montrer que v_α est vecteur propre de $X'X$ et que sa valeur propre associée est ℓ_α^2.

1d. Montrer que si la $\alpha^{\text{ème}}$ valeur propre de $X'X$ est simple, alors $u_\alpha = cXv_\alpha$, où c est une quantité que l'on exprimera uniquement en fonction de ℓ_α.

1e. Par un argument de symétrie et sans donner de démonstration, indiquer l'expression de v_α en fonction de X, u_α et ℓ_α.

1f. Déduire de ce qui précède une condition suffisante pour que la SVD de X soit unique, au signe près des vecteurs colonnes de P et Q.

2. *Existence de la SVD.*
On note $a_1, \dots, a_n$ une base orthonormée de $\mathbb{R}^n$ formée de vecteurs propres de $X'X$, et $A = \begin{pmatrix} a_1 & \dots & a_r \end{pmatrix}$ la matrice de dimension $n \times r$ dont les colonnes sont les vecteurs propres de $X'X$ associés aux valeurs propres $\lambda_1, \dots, \lambda_r$ qui sont non nulles et rangées en ordre décroissant. On note également $D = \mathrm{Diag}\big(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_r}\big)$, et U la matrice de dimension $p \times r$ dont les colonnes u_α sont définies par $u_\alpha = \dfrac{1}{\sqrt{\lambda_\alpha}} X a_\alpha$, avec $1 \leq \alpha \leq r$.

2a. Montrer que l'on a $\mathrm{Ker}(X'X) = \mathrm{Ker} X$.

2b. Montrer que $UDA' = \displaystyle\sum_{\alpha=1}^{n} X a_\alpha a'_\alpha$.

---

†. On pourra utiliser le résultat suivant : si A et B sont deux matrices possédant respectivement m colonnes notées a_α $(1 \leq \alpha \leq m)$, et m lignes notées b'_α $(1 \leq \alpha \leq m)$, en d'autres termes si l'on a $A = \begin{pmatrix} a_1 & \dots & a_m \end{pmatrix}$ et $B = \begin{pmatrix} b'_1 \\ \vdots \\ b'_m \end{pmatrix}$, alors $AB = \displaystyle\sum_{\alpha=1}^{m} a_\alpha \, b'_\alpha$

2c. En déduire que $X = UDA'$ et vérifier que cette décomposition est une SVD de X.

3. *Relations entre SVD et ACP.*
On suppose que X est un tableau de données où n individus (en colonnes) sont décrits par p variables (en lignes). De plus, X est supposé centré et $X = PLQ'$ désigne une SVD de X pour laquelle on définit les notations u_α, v_α et ℓ_α comme indiqué au début de l'énoncé. On effectue l'ACP de X sur matrice variance[§]. Pour tout $\alpha \in \{1, \dots, r\}$, on note w^α le $\alpha^{\text{ème}}$ vecteur axial factoriel, Ψ_α la $\alpha^{\text{ème}}$ composante principale et λ_α la variance de Ψ_α.
3a. Exprimer w^α, Ψ_α et λ_α en fonction de u_α, v_α et ℓ_α

3b. Montrer que $x_i^j = \displaystyle\sum_{\alpha=1}^{r} w_i^\alpha \Psi_\alpha^j$, où x_i^j désigne le terme général de la matrice X.

Exercice 2 (9 points environ)

On considère le tableau X de données suivant :

$$X = \begin{pmatrix} 1 & 2 & 0 & -1 \\ 1 & 0 & -2 & -1 \\ 1 & 1 & -1 & -1 \end{pmatrix}.$$

On rappelle que le terme général de X, noté x_i^j, indique la valeur prise par la $i^{\text{ème}}$ variable (avec $i \in \{1, 2, 3\}$) pour le $j^{\text{ème}}$ individu (avec $j \in \{1, 2, 3, 4\}$). Dans ce qui suit, on examine les résultats de l'ACP sur matrice variance[§] du tableau X.

1. Déterminer la valeur de c telle que l'on ait $x_1^j + x_2^j + c\, x_3^j = 0$ pour tout $j \in \{1, 2, 3, 4\}$.

2. Calculer la matrice variance V du tableau X.

3. Déduire de la question **1.** l'existence d'un axe factoriel trivial (i.e. associé à la valeur propre nulle) du nuage des individus, et déterminer un vecteur directeur de cet axe.

4. Montrer que le vecteur $v = \big(1, 1, 1\big)'$ dirige un axe factoriel non trivial du nuage des individus, associé à une valeur propre λ dont on précisera la valeur. Expliquer pourquoi cet axe est le premier axe factoriel.

5. Calculer un vecteur directeur du second axe factoriel du nuage des individus et déterminer l'inertie projetée sur cet axe.

6. Soit Ψ_1 la première composante principale du nuage des individus. Calculer Ψ_1^j pour tout $j \in \{1, 2, 3, 4\}$.

7. Calculer les coordonnées des trois variables sur le premier axe factoriel de l'espace des variables.

8. Calculer la qualité de représentation de chaque individu sur le premier axe factoriel.

9. Calculer la contribution relative de chaque individu à l'inertie du premier axe factoriel.

§. On rappelle que l'ACP sur matrice variance utilise la métrique identité et des poids égaux à $1/n$