Algèbre Linéaire et Analyse de Données

Devoir Personnel
Licence 2 MIASHS (2025-2026)

Guillaume Metzler
Université de Lyon, Université Lumière Lyon 2
Laboratoire ERIC UR 3083, Lyon, France
guillaume.metzler@univ-lyon2.fr

Résumé

Pour rappel, ce devoir est facultatif et permettra de modifier votre note en fonction de ce que vous avez pu traiter dans ce devoir.
Il se compose de deux exercices et d'un problème qui sont indépendants.
La qualité de la rédaction et la clarté des explications sera prise en compte dans l'évaluation de votre travail.
Bien évidemment, vous avez le droit d'utiliser tous les résultats qui ont été démontrés en cours sans démonstration, sauf si on vous demande explicitement de le montrer.

Le sujet a pour but de faire le lien entre votre cours d'algèbre linéaire et le cours sur la régression linéaire avec ce point de vue vectoriel/matriciel. Il nécessite également quelques connaissances en statistiques et en analyse de fonctions, mais pour des fonctions à plusieurs variables (ici deux) mais les éléments nécessaires à la résolution vous sont présentées.

Objectif du problème et articulations entre ses différentes parties

L'objectif du problème est d'établir une expression matricielle donnant la solution du problème de minimisation des moindres carrés classiquement utilisée en régression linéaire multiple. Cette expression est utilisée pour établir les formules classiques de la régression linéaire simple.

La partie A donne une description de l'orthogonal de l'image d'une matrice. Elle permet de construire dans la partie B , la matrice de la projection orthogonale sur un sous-espace de $\mathbb{R}^p$. La partie C utilise cette approche dans le contexte de la régression linaire simple. Elle utilise donc des résultats de la partie B qui elle-même utilise des résultats de la partie A.

Notations et Rappels

Lorsque n et p sont deux entiers naturels non nuls, $\mathscr{M}_{n,p}(\mathbb{R})$ désigne l'ensemble des matrices à n lignes et p colonnes à coefficients réels et $\mathscr{M}_n(\mathbb{R})$ l'ensemble des matrices carrées d'ordre n à coefficients réels.

Par exemple :

$$\mathbb{R}^n = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \ \forall i \in \{1, \dots, n\}, \ x_i \in \mathbb{R} \right\}$$

Pour toute matrice $\mathbf{M} \in \mathscr{M}_{n,p}(\mathbb{R})$, on note $\mathbf{M}^\top$ sa transposée.
On rappelle que : $\forall (\mathbf{M}, \mathbf{N}) \in (\mathscr{M}_{n,p}(\mathbb{R}))^2, \ (\mathbf{MN})^\top = \mathbf{N}^\top \mathbf{M}^\top$.
On rappelle la définition du noyau et la définition de l'image d'une matrice $\mathbf{A} \in \mathscr{M}_{n,p}(\mathbb{R})$

$$\mathrm{Ker}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^p \text{ vérifiant } \mathbf{Ax} = 0\} \ ; \ \mathrm{Im}(\mathbf{A}) = \{\mathbf{Ax} \mid \mathbf{x} \in \mathbb{R}^p\}$$

Le produit scalaire usuel sur $\mathbb{R}^n$ est défini pour tout $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ et $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$ par : $\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + \cdots + x_n y_n = \mathbf{x}^\top \mathbf{y}$.

Pour tout $\mathbf{x} \in \mathbb{R}^n$, on note $\|\mathbf{x}\|$ la norme de $\mathbf{x}$ associée à ce produit scalaire, définie par :

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\sum_{i=1}^n x_i^2} = \sqrt{\mathbf{x}^\top \mathbf{x}}$$

L'orthogonal d'un sous-espace F de $\mathbb{R}^n$ pour ce produit scalaire est défini par :

$$F^\perp = \{\mathbf{x} \in \mathbb{R}^n \text{ tel que : } \forall \mathbf{y} \in F, \langle \mathbf{x} \mid \mathbf{y} \rangle = 0\}$$

Partie A : L'égalité $\mathrm{Im}(\mathbf{A})^\perp = \mathrm{Ker}\left(\mathbf{A}^\top\right)$

1. Soient $\mathbf{c}_1, \dots, \mathbf{c}_p$ les vecteurs colonnes d'une matrice $A \in \mathscr{M}_{n,p}(\mathbb{R})$.
   Pour tout $j \in \{1, \dots, p\}$, déterminer $E_j \in \mathbb{R}^n$ tel que $\mathbf{c}_j = \mathbf{Ae}_j$.
   Montrer que la famille $(\mathbf{c}_1, \dots, \mathbf{c}_p$) est une famille génératrice de $\mathrm{Im}(\mathbf{A})$.

1. Un premier exemple

Dans cette première partie, on prend $\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \in \mathscr{M}_{3,4}(\mathbb{R})$.

2. Déterminer une base de $\mathrm{Im}(A)$.
3. Montrer que $(\mathrm{Im}(\mathbf{A}))^\perp = \mathrm{Ker}\left(\mathbf{A}^\top\right)$.

2. Une démonstration

4. Justifier que pour tout $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^n$, on a $\mathbf{x}^\top \mathbf{y} = \mathbf{y}^\top \mathbf{x}$.
5. Soit $\mathbf{x} \in \mathbb{R}^n$. Montrer que si pour tout $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{x}^\top \mathbf{y} = 0$ alors $\mathbf{x}$ est nul.

Soit $\mathbf{A} \in \mathscr{M}_{n,p}(\mathbb{R})$.

6. Montrer que pour tout $\mathbf{x} \in \mathbb{R}^n$ et pour tout $\mathbf{z} \in \mathbb{R}^p$, on a $\mathbf{x}^\top \mathbf{A}\mathbf{z} = \mathbf{z}^\top \mathbf{A}^\top \mathbf{x}$.
7. En déduire que : $\mathbf{A}^\top \mathbf{x} = 0 \Longleftrightarrow \left(\forall \mathbf{z} \in \mathbb{R}^p, \mathbf{x}^\top \mathbf{A}\mathbf{z} = 0\right)$.
8. En déduire que $(\mathrm{Im}(\mathbf{A}))^\perp = \mathrm{Ker}\left(\mathbf{A}^\top\right)$.

3. Une application

Dans cette partie $\mathbf{A} \in \mathscr{M}_n(\mathbb{R})$ c'est-à-dire $\mathbf{A}$ est une matrice carrée d'ordre n à coefficients réels.

9. Justifier que les sous-espaces $\mathrm{Ker}\left(\mathbf{A}^\top\right)$ et $\mathrm{Im}(\mathbf{A})$ sont supplémentaires dans $\mathbb{R}^n$.
10. Justifier que les sous-espaces $\mathrm{Ker}(\mathbf{A})$ et $\mathrm{Im}\left(\mathbf{A}^\top\right)$ sont supplémentaires dans $\mathbb{R}^n$.
11. En déduire que pour tout $\mathbf{x} \in \mathbb{R}^n$, il existe $(\mathbf{x}', \mathbf{x}'') \in \mathrm{Ker}\left(\mathbf{A}^\top\right) \times \mathrm{Im}\left(\mathbf{A}^\top\right)$ tel que $\mathbf{x} = \mathbf{x}' + \mathbf{A}\mathbf{x}''$. On admet que le couple $(\mathbf{x}', \mathbf{x}'')$ ainsi associé à $\mathbf{x}$ est unique. On note alors u l'application $u : \mathbf{x} \mapsto \mathbf{x}''$.
12. Montrer que u est un endomorphisme de $\mathbb{R}^n$.

On note $\mathbf{B}$ la matrice de u dans la base canonique de $\mathbb{R}^n$.

13. Soit $\mathbf{x} \in \mathbb{R}^n$, et $(\mathbf{x}', \mathbf{x}'')$ défini de manière unique en à la question 11. A l'aide de la définition d'une projection orthogonale et de la question 8 montrer que $\mathbf{A}\mathbf{x}''$ est le projeté orthogonal de $\mathbf{x}$ sur $\mathrm{Im}(\mathbf{A})$. En déduire que $\mathbf{AB}$ est la matrice de la projection orthogonale sur $\mathrm{Im}(\mathbf{A})$.
14. Montrer que $\mathbf{BA}$ est la matrice de la projection orthogonale sur $\mathrm{Im}\left(\mathbf{A}^\top\right)$.

4. Un deuxième exemple

Jusqu'à la fin de cette partie, on prend $\mathbf{A} = \begin{pmatrix} 0 & 0 & \dots & \dots & \dots & 0 \\ 1 & \ddots & \ddots & & & \vdots \\ 0 & \frac{1}{2} & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & \frac{1}{n-1} & 0 \end{pmatrix} \in \mathscr{M}_n(\mathbb{R})$ avec $n \geqslant 2$.

On note $\mathcal{B}_c = (\mathbf{e}_1, \dots, \mathbf{e}_n)$ la base canonique de $\mathbb{R}^n$.

15. Déterminer le rang de $\mathbf{A}$ et le rang de $\mathbf{A}^\top$.

16. Déterminer une base de chacun des espaces $\text{Im}(\mathbf{A})^{\perp}$ et $\text{Ker}\left(\mathbf{A}^{\top}\right)$.
On exprimera chaque base à l'aide des vecteurs de $\mathcal{B}_c$.
17. Vérifier que l'endomorphisme u de $\mathbb{R}^n$ construit en question 11, est l'application :

$$u : X = \sum_{k=1}^{n} x_k \mathbf{e}_k = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} \longmapsto X'' = \sum_{k=1}^{n-1} k x_{k+1} \mathbf{e}_k = \begin{pmatrix} x_2 \\ 2x_3 \\ \vdots \\ (n-1)x_n \\ 0 \end{pmatrix}$$

18. Construire la matrice $\mathbf{B}$ de u dans $\mathcal{B}_c$.

19. Vérifier le résultat de la question 14.

Partie B : Une expression de la matrice d'un projecteur orthogonal dans une base quelconque

L'espace $\mathbb{R}^p$ est muni de son produit scalaire canonique rappelé en préambule. Soient F un sous-espace vectoriel de $\mathbb{R}^p$ de dimension k et $\mathcal{B}_F = (\mathbf{f}_1, \dots, \mathbf{f}_k)$ une base de F. On note p_F la projection orthogonale sur F et $\mathbf{A}$ la matrice de $\mathcal{B}_F$ dans la base canonique de $\mathbb{R}^p$. La matrice $\mathbf{A}$ est donc dans $\mathscr{M}_{p,k}(\mathbb{R})$ et ses colonnes sont les vecteurs $\mathbf{f}_1, \cdots, \mathbf{f}_k$ exprimés dans la base canonique de $\mathbb{R}^p$. Le but de cette partie est de démontrer que la matrice $\mathbf{A}^{\top}\mathbf{A}$ est inversible et que $\mathbf{A}\left(\mathbf{A}^{\top}\mathbf{A}\right)^{-1}\mathbf{A}^{\top}$ est la matrice de p_F dans la base canonique de $\mathbb{R}^p$.

20. Justifier que $\text{Ker}(\mathbf{A}) = \{\mathbf{0}\}$.
21. Montrer que : $\mathbf{x} \in \text{Ker}\left(\mathbf{A}^{\top}\mathbf{A}\right) \Rightarrow \|\mathbf{A}\mathbf{x}\| = 0 \Rightarrow \mathbf{x} = \mathbf{0}$.
22. Montrer que $\mathbf{A}^{\top}\mathbf{A}$ est inversible.
23. Montrer que $F = \text{Im}(\mathbf{A})$.
24. Montrer que pour tout $\mathbf{x} \in \mathbb{R}^p$, il existe $\mathbf{y} \in \mathscr{M}_{k,1}(\mathbb{R})$ tel que : $p_F(\mathbf{x}) = \mathbf{A}\mathbf{y}$ et $\mathbf{x} - \mathbf{A}\mathbf{y} \in \text{Ker}\left(\mathbf{A}^{\top}\right)$.
25. En déduire que la matrice $\mathbf{A}\left(\mathbf{A}^{\top}\mathbf{A}\right)^{-1}\mathbf{A}^{\top}$ est la matrice de p_F dans la base canonique de $\mathbb{R}^p$.

Partie C : Régression linéaire simple

On pose $\mathbf{u} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n$ le vecteur colonne à n lignes contenant que des 1. On note $\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{R}^n$ un vecteur colonne donné et non colinéaire à u. S oit $\mathbf{A}$ la matrice : $\mathbf{A} = \begin{pmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \vdots \\ 1 & a_n \end{pmatrix} \in \mathscr{M}_{n,2}(\mathbb{R})$.

Lors d'un processus expérimental d'entrée $a_1 \in \mathbb{R}$, on recueille l'observation $b_1 \in \mathbb{R}$. On répète ce même processus avec l'entrée $a_2 \in \mathbb{R}$, on recueille l'observation $b_2 \in \mathbb{R}$. Ainsi de suite jusqu'au dernier processus d'entrée $a_n \in \mathbb{R}$, où on recueille l'observation $b_n \in \mathbb{R}$.

On note le vecteur $\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \in \mathbb{R}^n$ contenant donc les n observations après les n processus expérimentaux d'entrées $a_1, \ldots, a_n$.
On pose $g : \mathbb{R}^2 \to \mathbb{R}$ la fonction définie par : $\forall \mathbf{x} = (x_1, x_2) \in \mathbb{R}^2,\ g(x_1, x_2) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$.

1. Minimisation de g sur un exemple

On prend dans cette sous partie, $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{pmatrix}$ et $\mathbf{b} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

26. Expliciter la fonction g et justifier qu'elle est dérivable.

27. Calculer les dérivées partielles de g. Cela consiste à dériver la fonction g une fois par rapport à la variable x_1 (en considérant x_2 comme une constante), cette dérivée est appelée dérivée partielle et est notée $\dfrac{\partial g}{\partial x_1}$. On calcule ensuite une autre dérivée de la fonction la fonction g une fois par rapport à la variable x_2 (en considérant x_1 comme une constante), cette dérivée est appelée dérivée partielle et est notée $\dfrac{\partial g}{\partial x_2}$.
Il s'agit en fait de regarder les variations de la fonction g par rapport à chacune des variables x_1 et x_2.
28. En déduire que g admet un unique point critique dans $\mathbb{R}^2$ que l'on précisera.
Un point critique de g est un point (x_1, x_2) solution du système suivant :
$$\begin{cases} \dfrac{\partial g}{\partial x_1}(x_1, x_2) & = & 0, \\ \dfrac{\partial g}{\partial x_2}(x_1, x_2) & = & 0. \end{cases}$$
29. Montrer qu'en ce point g atteint son minimum. Pour tout $(x_1, x_2) \in \mathbb{R}^2$, on écrira $g(x_1, x_2)$ sous la forme $\beta_1 (x_1 - \alpha_1)^2 + \beta_2 (x_2 - \alpha_2)^2 + m$ où $\beta_1 > 0$ et $\beta_2 > 0$.

2. Minimisation de g via une projection orthogonale

On se place dans les conditions énoncées dans le préambule de la partie C .

30. Démontrer que : $\forall\ \mathbf{x} = (x_1, x_2) \in \mathbb{R}^2,\quad g(x_1, x_2) = \sum_{i=1}^n b_i^2 - 2\sum_{i=1}^n (x_1 + a_i x_2) b_i + \sum_{i=1}^n (x_1 + a_i x_2)^2$. Justifier que g est dérivable sur $\mathbb{R}^2$.
31. Calculer le gradient de g en $(x_1, x_2) \in \mathbb{R}^2$ noté $\nabla g(x_1, x_2)$.
La gradient est un vecteur qui contient les dérivées partielles de la fonction g (voir question 27).
32. Montrer que $\forall\, \mathbf{x}\,(x_1, x_2) \in \mathbb{R}^2,\ \nabla g(x_1, x_2) = -2\mathbf{A}^\top b + 2\mathbf{A}^\top\mathbf{A}\binom{x_1}{x_2}$.
33. Justifier à l'aide de la partie B que la matrice $\mathbf{A}^\top\mathbf{A}$ est inversible.
34. Montrer que si g admet un extremum en (α_1, α_2) alors $(\alpha_1, \alpha_2 = \left(\mathbf{A}^\top\mathbf{A}\right)^{-1}\mathbf{A}^\top b$. On note $F = \mathrm{Im}(\mathbf{A})$ et p_F le projecteur orthogonal sur F.

35. Justifier que $p_F(\mathbf{b}) = \mathbf{A}\binom{\alpha_1}{\alpha_2}$.

36. En déduire que $g(\alpha_1, \alpha_2)$ est le minimum de g sur $\mathbb{R}^2$.

3. Une autre expression de la solution du problème de minimisation de g

On note :

$$m_a = \frac{1}{n}\sum_{i=1}^{n} a_i \text{ et } m_b = \frac{1}{n}\sum_{i=1}^{n} b_i$$

les valeurs moyennes et

$$S_{a,b} = \sum_{i=1}^{n} (a_i - m_a)(b_i - m_b)$$

37. Montrer que $S_{a,b} = \sum_{i=1}^{n} (a_i \times b_i) - n \times m_a \times m_b$.

38. Montrer que $S_{a,a} = \|\mathbf{a} - m_a \cdot \mathbf{u}\|^2$ où $\mathbf{u} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n$.

39. Montrer que $S_{a,a} \neq 0$.

40. En déduire que la matrice $\mathbf{A}^\top \mathbf{A}$ est inversible et montrer que

$$\left(\mathbf{A}^\top \mathbf{A}\right)^{-1} = \frac{1}{n \times S_{a,a}} \begin{pmatrix} \sum_{i=1}^{n} a_i^2 & -\sum_{i=1}^{n} a_i \\ -\sum_{i=1}^{n} a_i & n \end{pmatrix}$$

.

41. En déduire que : $\alpha_2 = \frac{S_{a,b}}{S_{a,a}}$ et $\alpha_1 = m_b - \alpha_2 \times m_a$

42. Étudions cela sur un exemple et considérons la fonction g définie sur $\mathbb{R}^2$ par :

$$g : (x_1, x_2) \longmapsto (x_1 + 2x_2 - 1)^2 + (x_1 + x_2 - 2)^2 + (x_1 + x_2 - 1)^2 + (x_1 - x_2)^2 + (x_1 - 3x_2 - 1)^2$$

Déterminer en quel point le minimum de g sur $\mathbb{R}^2$ est atteint.

4. Espérance des estimateurs de l'argmin

On suppose que chaque b_i, pour $i \in \{1, \cdots, n\}$, est une observation d'une variable aléatoire réelle B_i.
On suppose que les variables aléatoires B_i suivent toutes la même loi et sont deux à deux indépendantes.
On modélise le processus expérimental à l'aide du modèle linéaire, c'est à dire on suppose que :

- Pour $i \in \{1, \cdots, n\}$, $B_i = x_1 + x_2 a_i + \mathcal{E}_i$ où $\mathcal{E}_i$ est une variable aléatoire réelle.
- On suppose que les variables $\mathcal{E}_i$ suivent toutes la même loi, sont centrées et d'écart type noté $\sigma > 0$, et sont deux à deux indépendantes.
- On note $\mathbf{B} = \begin{pmatrix} B_1 \\ \vdots \\ B_n \end{pmatrix}$ le vecteur aléatoire dont les coordonnées sont les variables aléatoires B_i.
- De même, on note $\mathcal{E} = \begin{pmatrix} \mathcal{E}_1 \\ \vdots \\ \mathcal{E}_n \end{pmatrix}$ le vecteur aléatoire dont les coordonnées sont les variables aléatoires $\mathcal{E}_i$.

43. Déterminer l'espérance de B_i en fonction des nombres réels x_1, x_2 et a_i.
44. Déterminer la variance de B_i en fonction de σ.

On pose (X_1, X_2) le vecteur aléatoire $\left(\mathbf{A}^\top \mathbf{A}\right)^{-1} \mathbf{A}^\top \mathbf{B}$. Ainsi définies, X_1 et X_2 sont des variables aléatoires réelles appelées estimateurs des coefficients par la méthode des moindre carrés.
On note : $m_a = \frac{1}{n} \sum_{i=1}^n a_i \in \mathbb{R}$ et on définit les variables aléatoires

$$m_B = \frac{1}{n} \sum_{i=1}^n B_i \text{ et } S_{a,B} = \sum_{i=1}^n (a_i - m_a)(B_i - m_B)$$

45. Déterminer l'espérance de m_B en fonction des nombres réels x_1, x_2 et m_a.
46. Justifier que $X_2 = \dfrac{S_{a,B}}{S_{a,a}}$ et que $X_1 = m_B - X_2 m_a$.
47. Déterminer l'espérance de X_2.