Dans tout le problème, n et p désignent deux entiers vérifiant $1 \leq p \leq n$. On note $\mathscr{M}_{n,p}(\mathbb{R})$ l'espace vectoriel des matrices à n lignes et p colonnes, à coefficients réels. La transposée d'une matrice A de $\mathscr{M}_{n,p}(\mathbb{R})$ est notée ${}^tA$. Lorsqu'une matrice A est inversible, on note A^{-1} son inverse.

Dans tout le problème, on identifie les deux espaces vectoriels $\mathscr{M}_{n,1}(\mathbb{R})$ (respectivement $\mathscr{M}_{p,1}(\mathbb{R})$) et $\mathbb{R}^n$ (resp. $\mathbb{R}^p$), c'est-à-dire qu'on identifie un vecteur (point) de $\mathbb{R}^n$ (resp. $\mathbb{R}^p$) avec le vecteur-colonne de ses coordonnées dans la base canonique de $\mathbb{R}^n$ (resp. $\mathbb{R}^p$).

On munit $\mathbb{R}^n$ (resp. $\mathbb{R}^p$) de sa structure euclidienne canonique, et pour tous vecteurs u et v de $\mathbb{R}^n$ (resp. $\mathbb{R}^p$, on note $\langle u\,|v\,\rangle ={}^t uv$ leur produit scalaire, et $\|u\|$ la norme de u associée.

Pour tout i de $[\![1;n]\!]$, on note f_i une fonction définie sur $\mathbb{R}^p$ à valeurs réelles, et de classe C^2 sur $\mathbb{R}^p$. Soit F la fonction définie sur $\mathbb{R}^p$, à valeurs réelles, par : $F(x_1,x_2,\ldots,x_p) = \frac{1}{2}\sum_{i=1}^{n}[f_i(x_1,x_2,\ldots,x_p)]^2$.

Autrement dit si $X=(x_1,x_2,\ldots,x_p)$ est un point de $\mathbb{R}^p$ on a : $F(X)=\frac{1}{2}\sum_{i=1}^{n} f_i^2(X) = \frac{1}{2}\|f(X)\|^2$ en notant $f(X)$ le vecteur $\big(f_1(X),\ldots,f_n(X)\big)$.

Le problème a pour objet l'étude de quelques aspects mathématiques liés à la recherche du minimum de la fonction F.

Partie 1 - Gradient et hessienne

Pour tout point $X=(x_1,x_2,\ldots,x_p)$ de $\mathbb{R}^p$, on rappelle que

• le *gradient* de F au point X, noté $\nabla F(X)$, est le vecteur suivant :

$$\nabla F(X) = \big(\frac{\partial F}{\partial x_1}(X),\ldots,\frac{\partial F}{\partial x_p}(X)\big)$$

• la matrice hessienne de F au point X, notée $\nabla^2 F(X)$, est la matrice symétrique de $\mathscr{M}_p(\mathbb{R})$ suivante :

$$\nabla^2 F(X) = \Big(\frac{\partial^2 F}{\partial x_k \partial x_j}(X)\Big)_{1\leq k,j\leq p}$$

Pour tout point $X=(x_1,\ldots,x_p)$ de $\mathbb{R}^p$, on note $J(X)$ la matrice de $\mathscr{M}_{n,p}(\mathbb{R})$ définie par :

$$J(X) = \Big(\frac{\partial f_i}{\partial x_j}(X)\Big)_{\substack{1\leq i\leq n\\ 1\leq j\leq p}}$$

dans laquelle i désigne l'indice de ligne et j l'indice de colonne. On pose : $G(X)={}^t J(X)J(X)$.

Si X est un point de $\mathbb{R}^p$ vérifiant $\nabla F(X)\neq 0$, on dit qu'un vecteur h de $\mathbb{R}^p$ est une *direction de décroissance* de F en X, si on a : $\langle \nabla F(X)\,|h\,\rangle < 0$.

Dans les trois exemples suivants, on suppose que p est égal à 2.

(1). *Un premier exemple.*

On considère les deux fonctions f_1 et f_2 définies sur $\mathbb{R}^2$ par : $f_1(x_1,x_2)=x_1^2+x_2+1$ et $f_2(x_1,x_2)=x_1+x_2^2+1$.

(a). Justifier que F est de classe C^2 sur $\mathbb{R}^2$. Calculer, en tout point (x_1,x_2) de $\mathbb{R}^2$, le gradient $\nabla F(x_1,x_2)$.

(b). Montrer que le système d'équations qui permet de déterminer les éventuels points critiques de F, peut se mettre sous la forme suivante :

$$\begin{cases} 2x_1^3+2x_1x_2+3x_1+x_2^2+1=0 \\ (x_1-x_2)(2x_1^2+2x_1x_2+2x_2^2-x_1-x_2+3=0 \end{cases}$$

(c). Établir, pour tout (x_1,x_2) de $\mathbb{R}^2$, l'inégalité : $2x_1^2+2x_1x_2+2x_2^2-x_1-x_2+3>0$.

En déduire que l'unique point critique de F est $\left(-\frac{1}{2},-\frac{1}{2}\right)$.

(d). Déterminer, en tout point (x_1,x_2) de $\mathbb{R}^2$, la matrice hessienne $\nabla^2 F(x_1,x_2)$.

En déduire que F admet un minimum local en $\left(-\frac{1}{2},-\frac{1}{2}\right)$.

(e). On note pour tout point X de $\mathbb{R}^2$, $\nabla^2 f_1(X)$ et $\nabla^2 f_2(X)$ respectivement, les matrices hessiennes de f_1 et f_2 au point X. Préciser la matrice $J(X)$. Exprimer ${}^tJ(X)f(X)$ et $G(X)+\sum_{i=1}^{2} f_i(X)\nabla^2 f_i(X)$ en fonction de $\nabla F(X)$ et $\nabla^2 F(X)$ respectivement.

(2). *Un deuxième exemple.*

Soit $a=(a_1,\ldots,a_n)$, $b=(b_1,\ldots,b_n)$ et $c=(c_1,\ldots,c_n)$ trois vecteurs non nuls donnés de $\mathbb{R}^n$, tels que la famille (a,b) soit libre.

Pour tout i de $[\![1;n]\!]$, la fonction f_i est définie sur $\mathbb{R}^2$ par : $f_i(x_1,x_2)=a_ix_1+b_ix_2-c_i$.

(a). Exprimer, pour tout point (x_1,x_2) de $\mathbb{R}^2$, le gradient $\nabla F(x_1,x_2)$ à l'aide de x_1, x_2, $\|a\|$, $\|b\|$, $\langle a\,|b\,\rangle$, $\langle a\,|c\,\rangle$ et $\langle b\,|c\,\rangle$.

(b). Justifier l'inégalité : $\|a\|^2\times\|b\|^2-\langle a\,|b\,\rangle^2>0$. En déduire que la fonction F possède un unique point critique $(\widehat{x_1},\widehat{x_2})$.

Exprimer $\widehat{x_1}$ et $\widehat{x_2}$ en fonction de $\|a\|$, $\|b\|$, $\langle a\,|b\,\rangle$, $\langle a\,|c\,\rangle$ et $\langle b\,|c\,\rangle$.

(c). Calculer, en tout point (x_1,x_2) de $\mathbb{R}^2$, la matrice hessienne $\nabla^2 F(x_1,x_2)$; en déduire que F admet un minimum local en $(\widehat{x_1},\widehat{x_2})$.

(d). En utilisant la structure euclidienne de $\mathbb{R}^n$, montrer que F admet un minimum global en $(\widehat{x_1},\widehat{x_2})$.

(3). *Un troisième exemple.*

On suppose que $c_1,c_2,\ldots,c_n$ sont n réels donnés non tous égaux. On note $\overline{c}$ et s^2 respectivement, la moyenne arithmétique et la variance de la série statistique $(c_i)_{1\leq i\leq n}$.

Pour tout i de $[\![1;n]\!]$, la fonction f_i est définie sur $\mathbb{R}^2$ par : $f_i(x_1,x_2)=x_1+x_2-c_i$.

(a). Déterminer les points critiques de F.

(b). Soit $(\widehat{x_1},\widehat{x_2})$ un point critique de F. Exprimer $F(\widehat{x_1},\widehat{x_2})$ en fonction de s^2. Montrer, pour tout (x_1,x_2) de $\mathbb{R}^2$, l'égalité : $F(x_1,x_2)-F(\widehat{x_1},\widehat{x_2})=\frac{n}{2}(x_1+x_2-\overline{c})^2$.

(c). En déduire la nature des points critiques de F. Ce résultat était-il prévisible ?

(4). *Retour au cas général.*

Soit $X=(x_1,\ldots,x_p)$ un point de $\mathbb{R}^p$.

(a). Exprimer $\nabla F(X)$ en fonction de ${}^tJ(X)$ et de $f(X)$.

(b). Pour tout i de $[\![1;n]\!]$, on note $\nabla^2 f_i(X)$ la matrice hessienne de f_i au point X.

Établir la formule : $\nabla^2 F(X)=G(X)+\sum_{i=1}^{n} f_i(X)\nabla^2 f_i(X)$.

Partie II - Une approximation de F

Dans cette partie, on conserve les définitions et les notations de la partie **I**, et on suppose que X est un vecteur **fixé** de $\mathbb{R}^p$ vérifiant : $\nabla F(X)\neq 0$.

Pour tout vecteur $h=(h_1,h_2,\ldots,h_p)$ de $\mathbb{R}^p$, on pose : $\ell(h)=f(X)+J(X)h$ et $L(h)=\frac{1}{2}\|\ell(h)\|^2$.

(1). Établir, pour tout h de $\mathbb{R}^p$, l'égalité : $L(h)=F(X)+{}^t h\nabla F(X)+\frac{1}{2}{}^t hG(X)h$.

(2). Soit P une matrice symétrique de $\mathscr{M}_p(\mathbb{R})$.

(a). Justifier que P est diagonalisable.

(b). On note $\theta_1, \ldots, \theta_p$ les valeurs propres de P, et on pose : $\theta = \max\limits_{1 \leq j \leq p} |\theta_j|$. Montrer, pour tout vecteur h de $\mathbb{R}^p$, l'inégalité suivante : $\left|{}^t hPh\right| \leq \theta \|h\|^2$.

(3).(a). Écrire un développement limité à l'ordre 2 de la fonction F au point X.

(b). En déduire, à l'aide de la question **2.b**, que l'on a : $\lim\limits_{\|h\| \to 0} \dfrac{F(X+h) - L(h)}{\|h\|} = 0$

Pour X fixé de $\mathbb{R}^p$, on dit que $L(h)$ est une approximation à l'ordre 2 de $F(X+h)$ lorsque $\|h\|$ tend vers 0.

(4). On note : $G(X) = \left(g_{i,j}(X)\right)_{1 \leq i,j \leq p}$. Soit φ_1 et φ_2 deux fonctions définies sur $\mathbb{R}^p$ par : $\varphi_1(h) = {}^t h\nabla F(X)$ et $\varphi_2(h) = {}^t hG(X)h$.

(a). Montrer que pour tout j de $[\![1;p]\!]$, on a : $\dfrac{\partial \varphi_1}{\partial h_j}(h) = \dfrac{\partial F}{\partial x_j}(X)$ et $\dfrac{\partial \varphi_2}{\partial h_j}(h) = 2\sum\limits_{i=1}^{p} g_{i,j}(X)h_i$.

(b). En déduire que le gradient $\nabla L(h)$ de L en h, est donné par : $\nabla L(h) = \nabla F(X) + G(X)h$.

(c). Soit $\nabla^2 L(h)$ la matrice hessienne de L en h. Établir la formule : $\nabla^2 L(h) = G(X)$.

Soit J une matrice de $\mathscr{M}_{n,p}(\mathbb{R})$.

(5).(a). Montrer que la matrice ${}^t JJ$ est diagonalisable et que ses valeurs propres sont positives ou nulles.

(b). Montrer que lorsque la matrice ${}^t JJ$ est inversible, le rang de la matrice J est égal à p.

(6). Montrer que si la fonction L admet des points critiques $\widehat{h}$, alors ceux-ci vérifient l'inéquation : $\left\langle \widehat{h} \,|\nabla F(X) \right\rangle \leq 0$.

(7). On suppose que la matrice $G(X)$ est inversible.

(a). Montrer que L admet un unique point critique $\widehat{h}$ donné par : $\widehat{h} = -(G(X))^{-1} \times {}^t J(X)f(X)$.

(b). Établir que $\widehat{h}$ est une direction de décroissance de F en X. En déduire que L admet un minimum local en $\widehat{h}$.

Partie III - Une décomposition d'une matrice rectangulaire

Afin de réduire les inconvénients liés à l'inversion de la matrice $G(X)$, on remplace celle-ci par la matrice $G(X) + \mu I$, où μ désigne un paramètre réel strictement positif, et I la matrice identité d'ordre p. Certains résultats d'algèbre linéaire permettent alors de substituer à l'inversion d'une matrice, le calcul plus simple d'une somme de matrices.

Soit J une matrice non nulle de $\mathscr{M}_{n,p}(\mathbb{R})$.

(1). Montrer qu'il existe une matrice V orthogonale de $\mathscr{M}_p(\mathbb{R})$, un entier q tel que $1 \leq q \leq p$, et des réels $\lambda_1, \lambda_2, \ldots, \lambda_q$ tels que $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q > 0$, qui vérifient l'égalité : ${}^t V {}^t JJV = D$, où $D = (d_{i,j})_{1 \leq i,j \leq p}$ est définie par : $d_{i,i} = \lambda_i$ si $1 \leq i \leq q$, et $d_{i,j} = 0$ sinon. Si $q < p$, on pose : $\lambda_{q+1} = \cdots = \lambda_p = 0$.

Pour tout i de $[\![1;p]\!]$, on note V_i la i-ième colonne de V.

(2).(a). Montrer que le rang de ${}^t JJ$ est égal à q.

(b). Montrer que, pour tout i de $[\![1;q]\!]$, JV_i est un vecteur propre de la matrice $J{}^t J$ associé à la valeur propre λ_i. En déduire que les matrices ${}^t JJ$ et $J{}^t J$ ont les mêmes valeurs propres non nulles.

(c). Soit $(Y_1, \ldots, Y_r)$ une base du sous-espace propre de ${}^t JJ$ associée à une valeur propre λ non nulle. Montrer que la famille $(JY_1, \ldots, JY_r)$ est une famille libre de $\mathscr{M}_{n,1}(\mathbb{R})$.

(d). En déduire que les sous-espaces propres de ${}^t JJ$ et de $J{}^t J$ associés à la même valeur propre non nulle sont de même dimension, et que le rang de $J{}^t J$ est égal à q.

(3). On pose, pour tout i de $[\![1;q]\!]$: $U_i = \dfrac{1}{\sqrt{\lambda_i}} JV_i$.

(a). Montrer que la famille $(U_1, \ldots, U_q)$ est une famille orthonormée de vecteurs propres de $J{}^t J$.

(b). En déduire qu'il existe une base orthonormée $(U_1, \ldots, U_q, U_{q+1}, \ldots, U_n)$ de $\mathscr{M}_{n,1}(\mathbb{R})$, formée de vecteurs propres de $J{}^t J$.

(4). On note U la matrice de $\mathscr{M}_n(\mathbb{R})$ telle que, pour tout i de $[\![1;n]\!]$, la i-ième colonne de U est la matrice-colonne U_i de $\mathscr{M}_{n,1}(\mathbb{R})$.

Soit $S = (s_{i,j})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}}$ la matrice de $\mathscr{M}_{n,p}(\mathbb{R})$ définie par : $s_{i,i} = \sqrt{\lambda_i}$ si $1 \leq i \leq p$ et $s_{i,j} = 0$ sinon.

Établir l'égalité matricielle suivante : $S = {}^t UJV$. En déduire l'égalité : $J = US{}^t V$.

(5).(a). Montrer que la matrice $({}^t JJ + \mu I)$ est inversible.

(b). On note $R = (r_{i,j})_{\substack{1 \leq i \leq p \\ 1 \leq j \leq n}}$ la matrice de $\mathscr{M}_{p,n}(\mathbb{R})$ définie par : $r_{i,j} = \dfrac{\sqrt{\lambda_i}}{\lambda_i + \mu}$ si $1 \leq i \leq p$ et $r_{i,j} = 0$ sinon.

Établir la formule suivante : $({}^t JJ + \mu I)^{-1} \times {}^t J = VR{}^t U$.

(c). En déduire l'égalité : $({}^t JJ + \mu I)^{-1} \times {}^t J = \sum\limits_{i=1}^{q} \dfrac{\sqrt{\lambda_i}}{\lambda_i + \mu} V_i {}^t U_i$.

(6). Soit X un vecteur fixé de $\mathbb{R}^p$ vérifiant : $\nabla F(X) \neq 0$.

Pour tout vecteur h de $\mathbb{R}^p$, on pose : $M(h) = L(h) + \dfrac{\mu}{2} \|h\|^2$.

(a). Montrer que : $\lim\limits_{\|h\| \to 0} \dfrac{F(X+h) - M(h)}{\|h\|} = 0$.

(b). Calculer, pour tout h de $\mathbb{R}^p$, le gradient $\nabla M(h)$ et la matrice hessienne $\nabla^2 M(h)$ de M en h.

(c). En appliquant les résultats des questions précédentes à la matrice $J(X)$, montrer que M admet un unique point critique h^*. Donner une expression de h^* qui utilise les résultats de la question **5.c**.

(d). Montrer que M admet un minimum local en h^*.

À partir de ce minimum local h^* de M (ou du minimum local $\hat{h}$ de L), on pourrait utiliser une méthode algorithmique permettant, sous certaines conditions, d'approcher avec une précision donnée un minimum local de la fonction F

Éléments de correction

[pb]

Partie I. Gradient et hessienne

Dans les trois exemples suivants, on suppose $p = 2$.

(1). Un premier exemple

On considère f_1 et f_2 définies sur $\mathbb{R}^2$ par : $f_1(x_1, x_2) = x_1^2 + x_2 + 1$ et $f_2(x_1, x_2) = x_1 + x_2^2 + 1$.

(a). On sait que les applications coordonnées : $(x_1,x_2)\mapsto x_1$, $(x_1,x_2)\mapsto x_2$, ainsi que les applications constantes sont de classe C^2 sur $\mathbb{R}^2$, alors par produit et sommes, les fonctions f_1 et f_2 sont de classe C^2 sur $\mathbb{R}^2$ et à valeurs réelles. La fonction carrée étant de classe C^2 sur $\mathbb{R}$, par composée les fonctions f_1^2 et f_2^2 sont de classe C^2 sur $\mathbb{R}^2$ et la fonction F, combinaison linéaire de f_1^2 et f_2^2, est alors de classe C^2 sur $\mathbb{R}^2$. De plus, pour tout $A=(x_1,x_2)\in\mathbb{R}^2$,

$$\frac{\partial F}{\partial x_1}(A)=\frac{\partial f_1}{\partial x_1}(A)f_1(A)+\frac{\partial f_2}{\partial x_1}(A)f_2(A)$$
$$\frac{\partial F}{\partial x_2}(A)=\frac{\partial f_1}{\partial x_2}(A)f_1(A)+\frac{\partial f_2}{\partial x_2}(A)f_2(A)$$

On en déduit que :

$$\frac{\partial F}{\partial x_1}(A)=2x_1(x_1^2+x_2+1)+(x_1+x_2^2+1)=2x_1^3+2x_1x_2+3x_1+x_2^2+1$$
$$\frac{\partial F}{\partial x_2}(A)=2x_2^3+2x_1x_2+3x_2+x_1^2+1$$

La gradient de F en $A=(x_1,x_2)$ est donc :

$$\nabla F(x_1,x_2)=\left(2x_1^3+2x_1x_2+3x_1+x_2^2+1,2x_2^3+2x_1x_2+3x_2+x_1^2+1\right)$$

(b). $A=(x_1,x_2)$ est un point critique de F sur l'ouvert $\mathbb{R}^2$ si et seulement si $\nabla F(x_1,x_2)=0$, donc si et seulement si le système suivant est vérifié :

$$\left\{\begin{array}{c}2x_1^3+2x_1x_2+3x_1+x_2^2+1=0\\2x_2^3+2x_1x_2+3x_2+x_1^2+1=0\end{array}\right.$$

En effectuant l'opération $L_2\leftarrow L_2-L_1$, on trouve le système équivalent suivant :

$$\left\{\begin{array}{c}2x_1^3+2x_1x_2+3x_1+x_2^2+1=0\\2(x_2^3-x_1^3)+3(x_2-x_1)+x_1^2-x_2^2=0\end{array}\right.$$

La dernière équation se factorise par (x_2-x_1), et $A=(x_1,x_2)$ est un point critique de F si et seulement si le système suivant est vérifié :

$$\left\{\begin{array}{c}2x_1^3+2x_1x_2+3x_1+x_2^2+1=0\\(x_1-x_2)(2x_1^2+2x_1x_2+2x_2^2+3-x_1-x_2)=0\end{array}\right.$$

(c). On peut écrire :

$$2x_1^2+2x_1x_2+2x_2^2+3-x_1-x_2=(x_1^2+2x_1x_2+x_2^2)+(x_1^2-x_1)+(x_2^2-x_2)+3$$
$$2x_1^2+2x_1x_2+2x_2^2+3-x_1-x_2=(x_1+x_2)^2+\left(x_1-\frac{1}{2}\right)^2+\left(x_2-\frac{1}{2}\right)^2+\frac{5}{2}$$

On en déduit que $2x_1^2+2x_1x_2+2x_2^2+3-x_1-x_2>0$ pour tout (x_1,x_2) de $\mathbb{R}^2$.

On en déduit que $\nabla F(x_1,x_2)=0$ si et seulement si

$$\left\{\begin{array}{c}2x_1^3+2x_1x_2+3x_1+x_2^2+1=0\\x_1-x_2=0\end{array}\right.\Leftrightarrow\left\{\begin{array}{c}x_2=x_1\\2x_1^3+3x_1^2+3x_1+1=0\end{array}\right.$$

— <u>1ère méthode</u>

On remarque que si $x_1=-\dfrac{1}{2}$ alors $2x_1^3+3x_1^2+3x_1+1=0$, on en déduit que :

$$\nabla F(x_1,x_2)=0\Leftrightarrow\left\{\begin{array}{c}x_2=x_1\\2(x_1+\frac{1}{2})(x_1^2+x_1+1)=0\end{array}\right.$$

Or l'équation du second degré $x^2+x+1=0$ a pour discriminant $\Delta=1-4=-3$ qui est strictement négatif, donc l'équation n'a pas de solution réelle.

Le gradient de F ne s'annule donc qu'en $(x_1,x_2)=\left(-\dfrac{1}{2},-\dfrac{1}{2}\right)$ et F n'a donc qu'un unique point critique qui est $\left(-\dfrac{1}{2},-\dfrac{1}{2}\right)$.

— <u>2nd méthode</u>

$$\nabla F(x_1,x_2)=0\Leftrightarrow\left\{\begin{array}{c}x_2=x_1\\2x_1^3+3x_1^2+3x_1+1=0\end{array}\right.\Leftrightarrow\left\{\begin{array}{c}x_2=x_1\\x_1^3+(x_1+1)^3=0\end{array}\right.\Leftrightarrow\left\{\begin{array}{c}x_2=x_1\\(x_1+1)^3=\end{array}\right.$$

Or la fonction $x\mapsto x^3$ réalise une bijection de $\mathbb{R}$ dans $\mathbb{R}$ alors

$$\nabla F(x_1,x_2)=0\Leftrightarrow\left\{\begin{array}{c}x_2=x_1\\x_1+1=-x_1\end{array}\right.\Leftrightarrow\left\{\begin{array}{l}x_2=x_1\\x_1=-\frac{1}{2}\end{array}\right.$$

L'unique point qui annule le gradient de F, donc l'unique point critique de F sur $\mathbb{R}^2$, est donc le point $\left(-\dfrac{1}{2},-\dfrac{1}{2}\right)$.

(d). On rappelle que pour $X\in\mathbb{R}^2$, on a :

$$\frac{\partial F}{\partial x_1}(X)=2x_1^3+2x_1x_2+3x_1+x_2^2+1$$
$$\frac{\partial F}{\partial x_2}(X)=2x_2^3+2x_1x_2+3x_2+x_1^2+1$$

Par définition de matrice hessienne en un point X d'une fonction de classe C^2 sur $\mathbb{R}^2$, on a :

$$\nabla^2F(X)=\begin{pmatrix}6x_1^2+2x_2+3 & 2x_1+2x_2\\2x_1+2x_2 & 6x_2^2+2x_1+3\end{pmatrix}$$

Au point $X=\left(-\dfrac{1}{2},-\dfrac{1}{2}\right)$, la matrice hessienne de F est donc (avec les notations de Monge) :

$$\nabla^2F(X)=\begin{pmatrix}\frac{7}{2} & -2\\-2 & \frac{7}{2}\end{pmatrix}=\begin{pmatrix}r & s\\s & \end{pmatrix}$$

$rt-s^2=\dfrac{7}{2}.\dfrac{7}{2}-(-2)^2=\dfrac{33}{4}>0$ avec $r=\dfrac{7}{2}>0$, d'après le cours on sait qu'alors F admet un minimum local au point $X=\left(-\dfrac{1}{2},-\dfrac{1}{2}\right)$.

(e). On a immédiatement :

$$J(X)=\begin{pmatrix}2x_1 & 1\\1 & 2x_2\end{pmatrix}$$

On en déduit que J est une matrice symétrique et

$${}^tJ(X)f(X)=J(X)f(X)=\begin{pmatrix}2x_1 & 1\\1 & 2x_2\end{pmatrix}.\begin{pmatrix}x_1^2+x_2+1\\x_2^2+x_1+1\end{pmatrix}=\begin{pmatrix}2x_1^3+2x_1x_2+3x_1+x_2^2+1\\2x_2^3+2x_1x_2+3x_2+x_1^2+1\end{pmatrix}$$

On remarque que <u>${}^tJ(X).f(X)=\nabla F(X)$</u>.

Par définition de la hessienne en un point X de $\mathbb{R}^2$ d'une fonction de classe C^2, on a :

$$\nabla^2 f_1(X) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \qquad \nabla^2 f_2(X) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$$

$$G(X) = ({}^t J(X)).J(X) = \begin{pmatrix} 4x_1^2+1 & 2x_1+2x_2 \\ 2x_1+2x_2 & 1+4x_2^2 \end{pmatrix}$$

$$f_1(X)\nabla^2 f_1(X) = (x_1^2+x_2+1)\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 2x_1^2+2x_2+2 & 0 \\ 0 & 0 \end{pmatrix}$$

$$f_2(X)\nabla^2 f_2(X) = \begin{pmatrix} 0 & 0 \\ 0 & 2x_2^2+2x_1+2 \end{pmatrix}$$

On en déduit que

$$G(X) + \sum_{i=1}^{2} f_i(X)\nabla^2 f_i(X) = \begin{pmatrix} 6x_1^2+2x_2+3 & 2(x_1+x_2) \\ 2(x_1+x_2) & 6x_2^2+2x_1+3 \end{pmatrix}$$

On remarque que $G(X) + \sum_{i=1}^{2} f_i(X)\nabla^2 f_i(X) = \nabla^2 F(X)$.

(2). Un second exemple

Soit $a = (a_1, \dots, a_n)$, $b = (b_1, \dots, b_n)$ et $c = (c_1, \dots, c_n)$ trois vecteurs non nuls donnés de $\mathbb{R}^n$, tels que la famille (a, b) soit libre.
Pour tout i de $\{1, \dots, n\}$, la fonction f_i est définie sur $\mathbb{R}^2$ par : $f_i(x_1, x_2) = a_i x_1 + b_i x_2 - c_i$.

(a). D'après l'expression de F, on a en X :

$$\forall k \in \{1,2\}, \quad \frac{F}{\partial x_k}(X) = \sum_{i=1}^{n} f_i(X)\frac{\partial f_i}{\partial x_k}(X)$$

D'où le gradient de F en $X = (x_1, x_2)$:

$$\nabla F(X) = \left(\sum_{i=1}^{n} a_i(a_i x_1 + b_i x_2 - c_i), \sum_{i=1}^{n} b_i(a_i x_1 + b_i x_2 - c_i)\right)$$

On peut donc écrire :

$$\nabla F(X) = \begin{pmatrix} \|a\|^2 x_1 + \langle a\,|b\rangle x_2 - \langle a\,|c\rangle \\ \langle a\,|b\rangle x_1 + \|b\|^2 x_2 - \langle b\,|c\rangle \end{pmatrix}$$

(b). D'après l'inégalité de Cauchy-Schwarz pour le produit scalaire canonique sur $\mathbb{R}^n$, on a :

$$|\langle a\,|b\rangle| \leqslant \|a\|\,\|b\|$$

avec égalité si et seulement si les vecteurs a et b sont liés, donc ici, les vecteurs a et b formant une famille libre, on obtient (par stricte croissance de la fonction carré sur $\mathbb{R}^+$) :

$$\|a\|^2 . \|b\|^2 - \langle a\,|b\rangle^2 > 0$$

$X = (x_1, x_2)$ est un point critique de F si et seulement si $\nabla F(X) = 0$, or

$$\nabla F(X) = 0 \Leftrightarrow \begin{cases} \|a\|^2 x_1 + \langle a\,|b\rangle x_2 - \langle a\,|c\rangle = 0 \\ \langle a\,|b\rangle x_1 + \|b\|^2 x_2 - \langle b\,|c\rangle = 0 \end{cases}$$

a n'est pas nul donc $\|a\|^2 \neq 0$, on effectue l'opération $L_2 \leftarrow \|a\|^2 L_2 - \langle a\,|b\rangle L_1$, alors

$$\nabla F(X) = 0 \Leftrightarrow \begin{cases} \|a\|^2 x_1 + \langle a\,|b\rangle x_2 - \langle a\,|c\rangle = 0 \\ \left(\|a\|^2\|b\|^2 - \langle a\,|b\rangle^2\right) x_2 + \langle a\,|b\rangle\langle a\,|c\rangle - \langle b\,|c\rangle\|a\|^2 = 0 \end{cases}$$

Puisque $\|a\|^2 . \|b\|^2 - \langle a\,|b\rangle^2 > 0$, le système précédent n'a qu'une seule solution qui est :

$$\overline{x_2} = \frac{\|a\|^2\langle b\,|c\rangle - \langle a\,|b\rangle\langle a\,|c\rangle}{\|a\|^2 . \|b\|^2 - \langle a\,|b\rangle^2}$$

$$\overline{x_1} = \frac{\|b\|^2\langle a\,|c\rangle - \langle a\,|b\rangle\langle b\,|c\rangle}{\|a\|^2 . \|b\|^2 - \langle a\,|b\rangle^2}$$

F a donc pour unique point critique, le point $(\overline{x_1}, \overline{x_2})$ donné précédemment.

(c). D'après le gradient de F en $X = (x_1, x_2)$ on a immédiatement la hessienne de F en X :

$$\nabla^2 F(X) = \begin{pmatrix} \|a\|^2 & \langle a\,|b\rangle \\ \langle a\,|b\rangle & \|b\|^2 \end{pmatrix} = \begin{pmatrix} r & s \\ s & t \end{pmatrix}$$

$rt - s^2 = \|a\|^2\|b\|^2 - \langle a\,|b\rangle^2 > 0$ et $r = \|a\|^2 > 0$ alors F admet un minimum local au point critique $(\overline{x_1}, \overline{x_2})$.

(d). Avec la structure euclidienne de $\mathbb{R}^n$, en notant $A = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \\ a_n & b_n \end{pmatrix}$, $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ et $C = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$, on remarque que

$$F(X) = \frac{1}{2}\|AX - C\|^2$$

Les vecteurs a et b formant une famille libre, la matrice A, qui appartient à $\mathcal{M}_{n,2}(\mathbb{R})$, est de rang 2 alors d'après le cours, on sait que $\underset{X\in\mathbb{R}^2}{Min}\|AX - C\|$ existe et est obtenu pour X solution du système ${}^tA.AX = {}^tAC$.

$${}^tA.AX = {}^tAC \Leftrightarrow \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ b_1 & b_2 & \dots & b_n \end{pmatrix}\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \\ a_n & b_n \end{pmatrix}.X = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ b_1 & b_2 & \dots & b_n \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$

$${}^tA.AX = {}^tAC \Leftrightarrow \begin{pmatrix} \|a\|^2 & \langle a\,|b\rangle \\ \langle a\,|b\rangle & \|b\|^2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \langle a\,|c\rangle \\ \langle b\,|c\rangle \end{pmatrix} \Leftrightarrow \nabla F(X) = 0$$

On trouve donc que $\underset{X\in\mathbb{R}^2}{Min}\|AX - C\|$ est obtenu pour $X = (\overline{x_1}, \overline{x_2})$ unique point critique de F. On aura donc :

$$\underset{X\in\mathbb{R}^2}{Min} F(X) = F(\overline{X}) = F(\overline{x_1}, \overline{x_2})$$

F admet donc bien un minimum global en $(\overline{x_1}, \overline{x_2})$.

(3). Un troisième exemple

On suppose que $c_1, \dots, c_n$ sont n réels donnés non tous égaux. On note $\bar{c}$ et s^2 respectivement, la moyenne arithmétique et la variance de la série statistique $(c_i)_{1\leqslant i\leqslant n}$.

$$\bar{c} = \frac{1}{n}\sum_{i=1}^{n} c_i \qquad s^2 = \frac{1}{n}\sum_{i=1}^{n}(c_i - \bar{c})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} c_i^2\right) - \bar{c}^2$$

Pour tout i de $\{1, \dots, n\}$, la fonction f_i est définie sur $\mathbb{R}^2$ par : $f_i(x_1, x_2) = x_1 + x_2 - c_i$.

(a). On rappelle que : $\forall k \in \{1,2\}$, $\dfrac{F}{\partial x_k}(X) = \displaystyle\sum_{i=1}^{n} f_i(X)\frac{\partial f_i}{\partial x_k}(X)$, D'où

$$\forall k \in \{1,2\}, \quad \frac{F}{\partial x_k}(X) = \sum_{i=1}^{n}(x_1 + x_2 - c_i) = nx_1 + nx_2 - n\bar{c} = n(x_1 + x_2 - \bar{c})$$

On en déduit que $X = (x_1, x_2)$ est un point critique de F sur $\mathbb{R}^2$ si et seulement si $x_1 + x_2 = \bar{c}$.
On a donc une droite de points critiques.

(b). Soit $(\overline{x_1}, \overline{x_2})$ un point critique de F.

$$F(\overline{x_1}, \overline{x_2}) = \frac{1}{2}\sum_{i=1}^{n} f_i^2(\overline{x_1}, \overline{x_2}) = \frac{1}{2}\sum_{i=1}^{n}(\bar{c} - c_i)^2$$

$$F(\overline{x_1}, \overline{x_2}) = \frac{n}{2}s^2$$

On en déduit que pour tout $(x_1, x_2) \in \mathbb{R}^2$:

$$F(x_1, x_2) - F(\overline{x_1}, \overline{x_2}) = \frac{1}{2}\sum_{i=1}(x_1 + x_2 - c_i)^2 - \frac{n}{2}s^2$$

$$F(x_1, x_2) - F(\overline{x_1}, \overline{x_2}) = \frac{1}{2}\sum_{i=1}^{n}\left[(x_1 + x_2 - c_i)^2 - (\bar{c} - c_i)^2\right]$$

$$F(x_1, x_2) - F(\overline{x_1}, \overline{x_2}) = \frac{1}{2}\sum_{i=1}^{n}(x_1 + x_2 - c_i - \bar{c} + c_i)(x_1 + x_2 - c_i + \bar{c} - c_i)$$

$$F(x_1, x_2) - F(\overline{x_1}, \overline{x_2}) = \frac{(x_1 + x_2 - \bar{c})}{2}(nx_1 + nx_2 - 2n\bar{c} + n\bar{c}) = \frac{n}{2}(x_1 + x_2 - \bar{c})^2$$

(c). On déduit de l'égalité précédente que :

$$\forall (x_1, x_2) \in \mathbb{R}^2, \quad F(x_1, x_2) - F(\overline{x_1}, \overline{x_2}) \geqslant 0$$

F admet donc un minimum global en $(\overline{x_1}, \overline{x_2})$.

Pouvait-on prévoir ce résultat ?

— 1ère méthode

On peut noter $F(x_1, x_2) = g(\alpha) = \dfrac{1}{2}\displaystyle\sum_{i=1}^{n}(\alpha - c_i)^2$ avec $\alpha = x_1 + x_2$.

$$F(x_1, x_2) = g(\alpha) = \frac{n}{2}\alpha^2 - n\bar{c}\alpha + \frac{1}{2}\sum_{i=1}^{n} c_i^2$$

$$F(x_1, x_2) = g(\alpha) = \frac{n}{2}\alpha^2 - n\bar{c}\alpha + \frac{n}{2}s^2 + \frac{n}{2}\bar{c}^2 = \frac{n}{2}(\alpha - \bar{c})^2 + \frac{n}{2}s^2$$

On en déduit que $\forall \alpha \in \mathbb{R}$, $g(\alpha) \geqslant \dfrac{n}{2}s^2$ et qu'il y a égalité si et seulement si $\alpha = \bar{c}$. D'où l'existence d'un minimum global pour F, qui vaut $\dfrac{n}{2}s^2$, atteint en tout point (x_1, x_2) tel que $x_1 + x_2 = \bar{c}$.

— Autre méthode, plus longue

On remarque que :

$$\underset{(x_1,x_2)\in\mathbb{R}^2}{Min} F(x_1, x_2) = \frac{1}{2}\underset{\alpha\in\mathbb{R}}{Min}\,\|\alpha(1, \dots, 1) - (c_1, \dots, c_n)\|^2 = \frac{1}{2}\underset{v\in E}{Min}\,\|v - c\|^2$$

où $E = Vect < (1, \dots, 1) >$ et $c = (c_1, \dots, c_n)$.
On sait alors que le minimum cherché **existe et est obtenu pour le vecteur v de E égal au projeté orthogonal de c sur E.**
Donc F admet un minimum global qui est obtenu pour $(x_1, x_2) = p(c)$ où p est la projection orthogonale sur E.
Une base orthonormale de E est $\left\{u = \dfrac{1}{\sqrt{n}}(1, \dots, 1)\right\}$ alors le projeté orthogonal de c sur E est le vecteur

$$p(c) = \langle c\,|u\,\rangle\, u = \frac{1}{n}\left(\sum_{i=1}^{n} c_i\right)(1, \dots, 1) = (\bar{c}, \dots, \bar{c})$$

ce qui donne

$$\underset{(x_1,x_2)\in\mathbb{R}^2}{Min} F(x_1, x_2) = \frac{1}{2}\,\|c - p(c)\|^2 = \frac{1}{2}\sum_{i=1}^{n}(c_i - \bar{c})^2 = \frac{n}{2}s^2$$

et ce minimum est atteint pour (x_1, x_2) tel que $x_1 + x_2 = \bar{c}$.

(4). Retour au cas général
Soit $X = (x_1, \dots, x_p)$ un point de $\mathbb{R}^p$.

(a). On rappelle que $\forall i \in \{1, \dots, n\}$, $\dfrac{\partial F}{\partial x_i}(X) = \displaystyle\sum_{k=1}^{n} f_k(X)\frac{\partial f_k}{\partial x_i}(X)$. Avec la définition de la matrice $J(X)$ et la matrice-colonne $f(X)$, on obtient par produit matriciel :

$$\nabla F(X) = \left({}^t J(X)\right).f(X)$$

(b). D'après l'expression des dérivées partielles d'ordre 1 de F en X, $\forall i \in \{1, \dots, n\}$ et $\forall j \in \{1, \dots, n\}$:

$$\frac{\partial^2 F}{\partial x_i \partial x_j}(X) = \sum_{k=1}^{n}\frac{\partial f_k}{\partial x_i}(X).\frac{\partial f_k}{\partial x_j}(X) + \sum_{k=1}^{n} f_k(X).\frac{\partial^2 f_k}{\partial x_i \partial x_j}(X)$$

Par produit matriciel, $\displaystyle\sum_{k=1}^{n}\frac{\partial f_k}{\partial x_i}(X).\frac{\partial f_k}{\partial x_j}(X)$ est le coefficient d'indice (i, j) de la matrice $\left({}^t J(X).J(X)\right)$ et, pour $k \in \{1, \dots, n\}$, $\dfrac{\partial^2 f_k}{\partial x_i \partial x_j}(X)$ est le coefficient d'indice (i, j) de la matrice $\nabla^2 f_k(X)$, on en déduit l'égalité matricielle suivante :

$$\nabla^2 F(X) = G(X) + \sum_{k=1}^{n} f_k(X).\nabla^2 f_k(X)$$

Partie II : Une approximation de F

On conserve les notations de la partie I, et on suppose que X est un vecteur fixé de $\mathbb{R}^p$ vérifiant $\nabla F(X) \neq 0$.
Pour tout vecteur $h = (h_1, \dots, h_p)$ de $\mathbb{R}^p$, on pose $l(h) = f(X) + J(X)h$ et $L(h) = \dfrac{1}{2}\,\|l(h)\|^2$.

(1).

$$L(h) = \frac{1}{2}\langle f(X)+J(X)h\,|\,f(X)+J(X)h\rangle = \frac{1}{2}\|f(X)\|^2+\langle J(X)h\,|\,f(X)\rangle+\frac{1}{2}\langle J(X)h\,|\,J(X)h$$

Par écriture matricielle du produit scalaire dans $\mathbb{R}^p$, on obtient :

$$L(h) = F(X)+\left({}^th.{}^tJ(X)\right)f(X)+\frac{1}{2}\left({}^th.{}^tJ(X)\right)J(X)h = F(X)+{}^th\nabla F(X)+\frac{1}{2}\left({}^thG(X)h\right)$$

(2). Soit P une matrice symétrique de $\mathcal{M}_p(\mathbb{R})$.

(a). P étant une matrice symétrique réelle, d'après le théorème spectral, on sait qu'elle est diagonalisable.

(b). On note $\theta_1,\dots,\theta_p$ les valeurs propres de P, et on pose $\theta = \underset{1\leqslant j\leqslant p}{max}|\theta_j|$.

P est symétrique réelle alors, d'après le théoprème spectral, on sait qu'il existe une matrice T orthogonale telle que : $P = T.D.{}^tT$ avec $D = diag(\theta_1,\dots,\theta_p)$ matrice diagonale dont les coefficients de la diagonale sont les réels $\theta_1,\dots,\theta_p$ dans cet ordre. Soit h un vecteur de $\mathbb{R}^p$, on aura :

$${}^th.P.h = {}^th.T.D.{}^tT.h = {}^tu.D.u$$

avec $u = {}^tT.h$, notons alors $u = (u_1,\dots,u_p)$ on obtient :

$$\left|{}^th.P.h\right| = \left|{}^tu.D.u\right| = \left|\sum_{j=1}^p \theta_j u_j^2\right| \leqslant \sum_{j=1}^p |\theta_j|\,u_j^2 \leqslant \theta\sum_{j=1}^p u_j^2$$

$$\left|{}^th.P.h\right| \leqslant \theta.\|u\|^2$$

or $\|u\|^2 = \langle u\,|\,u\rangle = {}^tu.u = {}^th.T.{}^tT.h = {}^th.h = \|h\|^2$ alors

$$\left|{}^th.P.h\right| \leqslant \theta.\|h\|^2$$

(3).(a). F est de classe C^2 sur $\mathbb{R}^p$ alors le développement limité à l'ordre 2 de F au point X est :

$$F(X+h) = F(X)+\langle\nabla F(X)\,|\,h\rangle+\frac{1}{2}\left({}^th.\nabla^2F(X).h\right)+\|h\|^2\epsilon(h)$$

avec $\epsilon(0)=0$ et ϵ continue en 0.

(b). Puisque $\langle\nabla F(X)\,|\,h\rangle = {}^th.\nabla F(X)$, on obtient :

$$F(X+h) = L(h)+\frac{1}{2}\left[{}^th.\left(\nabla^2F(X)-G(X)\right).h\right]+\|h\|^2\epsilon(h)$$

Notons P la matrice $\nabla^2F(X)-G(X)$, cette matrice est symétrique, comme somme de deux matrices symétriques, alors en utilisant le résultat de la question précédente, on aura

$$\left|{}^th.\left(\nabla^2F(X)-G(X)\right).h\right| \leqslant \theta\,\|h\|^2$$

où θ a été défini dans la question précédente et ici ne dépend pas de h mais seulement de F et X. On aura donc

$$\frac{|F(X+h)-L(h)|}{\|h\|} \leqslant \frac{\theta}{2}\|h\|+\|h\|\,|\epsilon(h)|$$

Ce qui donne par théorème d'encadrement :

$$\lim_{\|h\|\to 0}\frac{|F(X+h)-L(h)|}{\|h\|} = 0$$

Pour X fixé de $\mathbb{R}^p$, on dit que $L(h)$ est une approximation d'ordre 2 de $F(X+h)$ lorsque $\|h\|$ tend vers 0.

(4). On note $G(X) = (g_{i,j}(X))_{1\leqslant i,j\leqslant p}$. Soit φ_1 et φ_2 deux fonctions définies sur $\mathbb{R}^p$ par : $\varphi_1(h) = {}^th.\nabla F(X)$ et $\varphi_2(h) = {}^th.G(X).h$.

(a). En écrivant $h = (h_1,\dots,h_p)$, on a :

$$\varphi_1(h) = \sum_{k=1}^p h_k\frac{\partial F}{\partial x_k}(X)$$

$$\varphi_2(h) = \sum_{k=1}^p\sum_{i=1}^p h_k g_{k,i} h_i$$

On en déduit que pour tout $j\in\{1,\dots,p\}$ on a :

$$\frac{\partial\varphi_1}{\partial h_j}(h) = \frac{\partial F}{\partial x_j}(X)$$

$$\frac{\partial\varphi_2}{\partial h_j}(h) = \sum_{i=1}^p g_{j,i}h_i+\sum_{k=1}^p h_k g_{k,j}$$

La matrice $G(X)$ étant symétrique, on a :

$$\frac{\partial\varphi_2}{\partial h_j}(h) = 2\sum_{i=1}^p g_{i,j}h_i = 2\sum_{i=1}^p g_{j,i}h_i$$

(b). On a : $L(h) = F(X)+\varphi_1(h)+\dfrac{1}{2}\varphi_2(h)$ alors

$$\forall j\in\{1,\dots,p\},\quad \frac{\partial L}{\partial h_j}(h) = \frac{\partial\varphi_1}{\partial h_j}(h)+\frac{1}{2}\frac{\partial\varphi_2}{\partial h_j}(h)$$

$$\forall j\in\{1,\dots,p\},\quad \frac{\partial L}{\partial h_j}(h) = \frac{\partial F}{\partial x_j}(X)+\sum_{i=1}^p g_{j,i}h_i$$

On en déduit, par formule du produit matriciel, que :

$$\nabla L(h) = \nabla F(X)+G(X).h$$

(c). D'après les égalités ci-dessus, on aura pour $k\in\{1,\dots,p\}$ et $j\in\{1,\dots,p\}$:

$$\frac{\partial^2 L}{\partial x_k\partial x_j}(h) = g_{j,k} = g_{k,j}$$

Et donc $\nabla^2L(h) = G(X)$.

(5). Soit J une matrice de $\mathcal{M}_{n,p}(\mathbb{R})$.

(a). La matrice $S = {}^tJ.J$ est une matrice carrée d'ordre p, symétrique réelle, alors elle est diagonalisable. Notons $\lambda_1,\dots,\lambda_p$ les valeurs propres de S, pour tout vecteur X de $\mathbb{R}^p$, on a :

$$\langle SX\,|\,X\rangle = {}^tX.S.X = {}^tX.{}^tJ.J.X = {}^t(JX).(JX) = \langle JX\,|\,JX\rangle = \|JX\|^2$$

d'où $\langle SX\,|\,X\rangle \geqslant 0$.
En particulier pour un vecteur propre X_i associé à la valeur propre λ_i de S, on obtient :

$$0\leqslant\langle SX_i\,|\,X_i\rangle = \lambda_i^tX_iX_i = \lambda_i\|X_i\|^2$$

et donc $\forall i\in\{1,\dots,p\}\quad \lambda_i\geqslant 0$.
La matrice ${}^tJ.J$ est donc diagonalisable avec des valeurs propres positives ou nulles.

(b). On suppose que $S = \left({}^tJ.J\right)$ est inversible, alors 0 n'est pas valeur propre de S, la forme quadratique associée à S est donc définie-positive, c'est-à-dire :

$$\forall X \neq 0, \quad \langle SX \,|X\,\rangle > 0$$

ce qui donne : $X \neq 0\mathbb{R}ightarrow \|JX\|^2 > 0$, et donc

$$X \neq 0\mathbb{R}ightarrow JX \neq 0$$

le noyau de J est donc réduit au vecteur nul, ce qui donne par le théorème du rang appliqué à J :

$$\dim \mathbb{R}^p = dimKer(J) + rg(J) = rg(J) = p$$

La matrice J est de rang égal à p.

(6). On suppose que L admet des points critiques $\hat{h}$, alors en ces points le gradient de L est nul :

$$0 = \nabla L(\hat{h}) = \nabla F(X) + G(X).\hat{h}$$

On aura donc :

$$\left\langle \hat{h} \,|\nabla F(X)\,\right\rangle = -{}^t\hat{h}G(X).\hat{h} = -{}^t\hat{h}.{}^tJ(X).J(X).\hat{h} = -\left\|J(X).\hat{h}\right\|^2 \leqslant 0$$

(7). On suppose que la matrice $G(X)$ est inversible.

(a). h est un point critique de L si et seulement si le gradient de L en h est nul. h est donc un point critique de L si et seulement si :

$$G(X).h = -\nabla F(X)$$

La matrice $G(X)$ étant inversible on obtient :
h est un point critique de L si et seulement si $h = -\left(G(X)\right)^{-1}.\nabla F(X)$. L admet donc un unique point critique $\hat{h}$ donné par la formule :

$$\hat{h} = -\left(G(X)\right)^{-1}.{}^tJ(X).f(X)$$

(b). Par définition, $\hat{h}$ est une direction décroissante de F en X lorsque $\left\langle \nabla F(X) \middle| \hat{h} \right\rangle < 0$.
On rappelle que si M est symétrique inversible alors M^{-1} aussi, d'où

$$\left\langle \hat{h} \,|\nabla F(X)\,\right\rangle ={}^t \hat{h}.\nabla F(X) = -{}^t\left(G^{-1}(X).\nabla F(X)\right).\nabla F(X) = -{}^t\nabla F(X).G^{-1}(X).\nabla F($$

La matrice $G(X) ={}^t J(X).J(X)$ est symétrique réelle, inversible, on peut lui appliquer les résultats de la question 5, on en déduit que ses valeurs propres sont toutes strictement positives, il en est alors de même de sa matrice inverse $G^{-1}(X)$. La forme quadratique q' associée à la matrice symétrique $G^{-1}(X)$ est donc aussi définie-positive et puisque le vecteur $\nabla F(X)$ n'est pas nul, on a $q'(\nabla F(X)) > 0$, ce qui donne :

$$\left\langle \hat{h} \,|\nabla F(X)\,\right\rangle = -q'(\nabla F(X)) < 0$$

$\hat{h}$ est bien une direction décroissante de F en X.

Par développement limité à l'ordre 2 de L en $\hat{h}$, on a :

$$L(\hat{h} + h) = L(\hat{h}) + \left\langle \nabla L(\hat{h}) \,|h\,\right\rangle + \frac{1}{2}{}^t h\nabla^2 L(\hat{h}).h + \|h\|^2\,\epsilon(h)$$

avec ici $\nabla L(\hat{h}) = 0$, $\nabla^2 L(\hat{h}) = G(X)$ et $\epsilon(h)$ tend vers 0 lorsque $\|h\|$ tend vers 0.
Par application des résultats de la question 5, on sait que la matrice hessienne de L en $\hat{h}$, qui est $G(X)$, n'a que des valeurs propres strictement positives, le cours indique alors que L admet un minimum local en $\hat{h}$.

Partie III. Une décomposition d'une matrice rectangulaire

Soit J une matrice non nulle de $\mathcal{M}_{n,p}(\mathbb{R})$.

(1). D'après les résultats de la question 5 partie II, on sait que ${}^tJ.J$ est une matrice symétrique réelle carrée d'ordre p diagonalisable dont les valeurs propres sont positives ou nulles. Notons alors $\lambda_1, \lambda_2, \ldots, \lambda_q$ ses valeurs propres non nulles rangées en ordre décroissantes : $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_q > 0$ et si $q < p$ on note $\lambda_{q+1} = \ldots = \lambda_p = 0$ les autres valeurs propres de ${}^tJ.J$.
${}^tJ.J$ étant symétrique réelle, on sait qu'il existe une matrice V carrée d'ordre p orthogonale telle que $D ={}^t V.{}^tJ.J.V$ soit une matrice diagonale carrée d'ordre p dont les coefficients diagonaux sont les p valeurs propres $\lambda_1, \lambda_2, \ldots, \lambda_p$ de ${}^tJ.J$.
D'après le cours, on sait que la matrice V, carrée d'ordre p, est la matrice de passage de la base canonique de $\mathcal{M}_{p,1}(\mathbb{R})$ (base orthonormale) à une base orthonormale de vecteurs propres de ${}^tJ.J$.
La i-ième colonne de V est un vecteur propre de ${}^tJ.J$ associé à la valeur propre λ_i, on la note V_i.

(2). (a). ${}^tJ.J$ est semblable à la matrice diagonale D précédente, cette matrice D est clairement de rang q, donc ${}^tJ.J$ aussi.

(b). On a rappelé que V_i est un vecteur propre de ${}^tJ.J$ associé à la valeur propre λ_i alors :

$$\left(J.{}^tJ\right).J.V_i = J.\left({}^tJ.JV_i\right) = \lambda_i J.V_i$$

Soit $i \in \{1, \ldots, q\}\,, V_i \neq 0$ et ${}^tJ.J.V_i = \lambda_i V_i \neq 0$ alors

$$JV_i \neq 0 \quad \text{et} \quad \left(J.{}^tJ\right).J.V_i = \lambda_i J.V_i$$

donc JV_i est un vecteur propre de $J.{}^tJ$ associé à la valeur propre λ_i.
Donc toute valeur propre non nulle de ${}^tJ.J$ est aussi valeur propre non nulle de $J.{}^tJ$.
Réciproquement si λ est une valeur propre non nulle de $J.{}^tJ$ et V est un vecteur propre associé, on aura $J.{}^tJ.V = \lambda.V \neq 0$ alors ${}^tJ.V \neq 0$ et

$$({}^tJ.J).{}^tJ.V ={}^t J.(\lambda V) = \lambda{}^tJ.V$$

et donc ${}^tJ.V$ est un vecteur propre de ${}^tJ.J$ associé à la valeur propre non nulle λ.

On en déduit que les matrices ${}^tJ.J$ et $J.{}^tJ$ ont les mêmes valeurs propres non nulles.

(c). Soit $(Y_1, \ldots, Y_r)$ une base du sous-espace propre de ${}^tJ.J$ associée à une valeur propre λ non nulle. Soit $(\alpha_1, \ldots, \alpha_r) \in \mathbb{R}^r$ tel que $\sum_{i=1}^{r} \alpha_i JY_i = 0$, on aura alors par multiplication à gauche par ${}^tJ$:

$${}^tJ.\sum_{i=1}^{r} \alpha_i JY_i = 0 = \sum_{i=1}^{r} \alpha_i({}^tJ.J.Y_i) = \sum_{i=1}^{r} \alpha_i \lambda Y_i$$

La famille $(Y_1, \ldots, Y_r)$ étant libre, on obtient : $\forall i \in \{1, \ldots, r\}\,, \quad \lambda\alpha_i = 0$ et puisque $\lambda \neq 0$, on a

$$\sum_{i=1}^{r} \alpha_i JY_i = 0 \to \forall i \in \{1, \ldots, r\}\,, \quad \alpha_i = 0$$

La famille $(JY_1, \ldots, JY_r)$ est donc libre.

(d). — Soit $(Y_1,\dots,Y_r)$ une base du sous-espace propre de ${}^tJ.J$ associée à une valeur propre λ non nulle. D'après ce qui a été vu dans les deux questions précédentes, la famille $(JY_1,\dots,JY_r)$ est une famille libre de vecteurs propres de $J.{}^tJ$ associés à la valeur propre λ, on en déduit que le sous-espace propre de $J.{}^tJ$ associé à la valeur propre λ est de dimension au moins égale à $r = dimker({}^tJ.J-\lambda Id)$: $dimKer(J.{}^tJ-\lambda I) \geqslant r$.

— Mais avec le même raisonnement que dans la question précédente (intervertir les rôles de J et ${}^tJ$), on peut montrer que si $(X_1,\dots,X_t)$ est une base du sous-espace propre de $J.{}^tJ$ associé à la valeur propre $\lambda \neq 0$ alors la famille $({}^tJX_1,\dots,{}^tJX_t)$ est une famille libre de vecteurs propres de ${}^tJ.J$ associés à la valeur propre λ. On en déduit que la dimension du sous-espace propre de ${}^tJ.J$ associée à une valeur propre λ est au moins égale à t, on a donc $t \leqslant r$.

— On a obtenu avec $r = dimKer({}^tJ.J-\lambda I)$ et $t = dimKer(J.{}^tJ-\lambda I)$: $t \geqslant r$ et $t \leqslant r$ donc $r = t$.

Les sous-espaces propres de ${}^tJ.J$ et $J.{}^tJ$ associés à la même valeur propre λ non nulle sont de même dimension.

Puisque $J.{}^tJ$ est symétrique réelle, elle est diagonalisable et ses valeurs propres non nulles sont celles de ${}^tJ.J$ c'est-à-dire : $\lambda_1,\dots,\lambda_q$.
On en déduit que $J.{}^tJ$ est semblable à une matrice diagonale (carrée d'ordre n) dont les coefficients non nuls de la diagonale sont les valeurs propres $\lambda_1,\dots,\lambda_q$.
Cette matrice diagonale est donc de rang q et la matrice $J.{}^tJ$ est aussi de rang q.

(3). On pose, pour tout i de $\{1,\dots,q\}$: $U_i = \dfrac{1}{\sqrt{\lambda_i}}JV_i$.

(a). On a déjà dit que la matrice orthognale V est une matrice de passage entre deux bases orthonormales. La famille $(V_1,\dots,V_q)$, des q premières colonnes de V, est donc une famille orthonormale.

$$\langle U_i\,|U_j\,\rangle = \frac{1}{\sqrt{\lambda_i}\sqrt{\lambda_j}}\langle JV_i\,|JV_j\,\rangle = \frac{1}{\sqrt{\lambda_i}\sqrt{\lambda_j}}{}^tV_i{}^tJ.JV_j = \frac{\lambda_j}{\sqrt{\lambda_i}\sqrt{\lambda_j}}{}^tV_i.V_j$$

$$\langle U_i\,|U_j\,\rangle = \sqrt{\frac{\lambda_j}{\lambda_i}}\langle V_i\,|V_j\,\rangle$$

On en déduit que si $i \neq j$ alors $\langle U_i\,|U_j\,\rangle = 0$ et $\langle U_i\,|U_i\,\rangle = \langle V_i\,|V_i\,\rangle = 1$.
La famille $(U_1,\dots,U_q)$ est une famille orthonormale formée de vecteurs propres de $J.{}^tJ$: d'après le résultat de la question 2b) U_i est associé à la valeur propre λ_i.

(b). La matrice carrée d'ordre n $J.{}^tJ$ étant symétrique réelle, avec pour valeurs propres non nulles λ_1,λ_q, est de rang q, on peut donc compléter la famille orthonormale $(U_1,\dots,U_q)$ par une base orthonormale du noyau de $J.{}^tJ$ pour obtenir une base orthonormale de $\mathcal{M}_{n,1}(\mathbb{R})$. Il existe donc bien une base orthonormée $(U_1,\dots,U_q,U_{q+1},\dots,U_n)$ de $\mathcal{M}_{n,1}(\mathbb{R})$ formée de vecteurs propres de $J.{}^tJ$.

(4). On note U la matrice de $\mathcal{M}_n(\mathbb{R})$ dont la i-ième colonne est la matrice U_i définie précédemment.
Soit $S = (s_{i,j}) \in \mathcal{M}_{n,p}(\mathbb{R})$ avec $s_{i,i} = \sqrt{\lambda_i}$ pour $i \in \{1,\dots,q\}$ et tous les autres coefficients sont nuls.

En notant $V = \begin{pmatrix}V_1 & V_1 & \dots & V_q & \dots & V_p\end{pmatrix}$, $U = \begin{pmatrix}U_1 & \dots & U_q & \dots & U_n\end{pmatrix}$, on aura :

$${}^tU.J.V = \begin{pmatrix}{}^tU_1\\ {}^tU_2\\ \vdots\\ {}^tU_n\end{pmatrix}\begin{pmatrix}JV_1 & JV_2 & \dots & JV_p\end{pmatrix} = \left({}^tU_i.J.V_j\right)_{1\leqslant i\leqslant n, 1\leqslant j\leqslant p}$$

— Pour $i \in \{1,\dots,q\}$ et $j \in \{1,\dots,q\}$, on a :

$${}^tU_i.J.V_j = \frac{1}{\sqrt{\lambda_i}}{}^tV_i.{}^tJ.J.V_j = \frac{\lambda_j}{\sqrt{\lambda_i}}{}^tV_i.V_j = \frac{\lambda_j}{\sqrt{\lambda_i}}\langle V_i\,|V_j\,\rangle$$

La famille $(V_1,\dots,V_p)$ est orthonormale, on en déduit que si $i \neq j$ alors ${}^tU_i.J.V_j = 0$ et ${}^tU_i.J.V_i = \sqrt{\lambda_i}$.

— Pour $j \in \{1,\dots,q\}$, JV_j est un vecteur propre de $J.{}^tJ$ associé à la valeur propre non nulle λ_j. On en déduit qu'il existe des scalaires $\alpha_1,\dots,\alpha_q$ tels que $JV_j = \sum_{k=1}^{q}\alpha_kU_k$

$$\forall i \in \{q+1,\dots,n\}, \forall j \in \{1,\dots,q\}_q uad{}^tU_i.J.V_j = {}^tU_i.\sum_{k=1}^{q}\alpha_kU_k = \sum_{k=1}^{q}\alpha_k\langle U_i\,|U_k\,\rangle = 0$$

— Pour $j \in \{q+1,\dots,p\}$, V_j est un vecteur du noyau de ${}^tJ.J$ alors

$$\|JV_j\|^2 = \langle JV_j\,|JV_j\,\rangle = {}^tV_j.{}^tJ.J.V_j = 0$$

ce qui entraine que $JV_j = 0$ pour $j \in \{q+1,\dots,p\}$ et donc

$$\forall i \in \{1,\dots,n\}, \quad \forall j \in \{q+1,\dots,p\}, \quad {}^tU_i.J.V_j = 0$$

Finalement par égalité des coefficients d'indice (i,j) des matrices S et ${}^tU.J.V$, on a bien : $S = {}^tU.J.V$.
Les matrice U et V étant orthogonales on a immédiatement : $U.S.{}^tV = J$.

(5). **(a).** μ est un réel strictement positif et ${}^tJ.J$ n'a pas de valeur propre strictement négative alors $(-\mu)$ n'est pas valeur propre de ${}^tJ.J$ et donc $\left({}^tJ.J+\mu I\right) = {}^tJ.J-(-\mu)I$ est inversible.

(b). On note $R = (r_{i,j})$ la matrice de $\mathcal{M}_{p,n}(\mathbb{R})$ définie par : $r_{i,i} = \dfrac{\sqrt{\lambda_i}}{\lambda_i+\mu}$ si $1 \leqslant i \leqslant q$ et $r_{i,j} = 0$ sinon.
On sait que ${}^tJ.J = V.D.{}^tV$ avec la matrice D diagonale dont les coefficients de la diagonale sont les valeurs propres $\lambda_1,\dots,\lambda_p$ de ${}^tJ.J$, on a alors :

$$\left({}^tJ.J+\mu I\right)^{-1}.\left({}^tJ\right) = V.R.{}^tU \Leftrightarrow \left({}^tJ\right) = \left({}^tJ.J+\mu I\right).V.R.{}^tU \Leftrightarrow \left({}^tJ\right) = \left(V.(D+\mu I){}^tV\right).V.R.{}^tU$$

$$\left({}^tJ.J+\mu I\right)^{-1}.\left({}^tJ\right) = V.R.{}^tU \Leftrightarrow V.(D+\mu I).R.{}^tU = {}^tJ \Leftrightarrow (D+\mu I).R = {}^tV.{}^tJ.U$$

$$\left({}^tJ.J+\mu I\right)^{-1}.\left({}^tJ\right) = V.R.{}^tU \Leftrightarrow (D+\mu I).R = {}^tV.{}^t\left(U.S.{}^tV\right).U = \left({}^tS\right)$$

$(D+\mu I)$ est une matrice carrée d'ordre p diagonale dont les coefficients de la diagonale sont $\lambda_1+\mu,\dots,\lambda_p+\mu$, on en déduit que $(D+\mu I).R = (\alpha_{i,j}) \in \mathcal{M}_{p,n}(\mathbb{R})$ avec

$$\forall i \in \{1,\dots,q\} \quad \alpha_{i,i} = \sqrt{\lambda_i} = s_{i,i} \text{ et } \alpha_{i,j} = 0 \text{ dans tous les autres cas}$$

On retrouve bien $(D+\mu I).R = \left({}^tS\right)$ donc l'égalité $\left({}^tJ.J+\mu I\right)^{-1}.\left({}^tJ\right) = V.R.{}^tU$ est vérifiée.

(c). En écrivant la matrice V en colonne : $V = \begin{pmatrix} V_1 & \dots & V_p \end{pmatrix}$ et la matrice $({}^tU)$ en ligne : $({}^tU) = \begin{pmatrix} {}^tU_1 \\ \vdots \\ {}^tU_n \end{pmatrix}$ et R n'ayant pour coefficients non nuls que les $r_{i,i}$ pour $i \in \{1, \dots, q\}$, on obtient :

$$\left({}^tJ.J + \mu I\right)^{-1}.\left({}^tJ\right) = V.R.{}^tU = \begin{pmatrix} V_1 & \dots & V_n \end{pmatrix} \begin{pmatrix} r_{1,1}\left({}^tU_1\right) \\ r_{2,2}\left({}^tU_2\right) \\ \vdots \\ r_{q,q}\left({}^tU_q\right) \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\left({}^tJ.J + \mu I\right)^{-1}.\left({}^tJ\right) = \sum_{i=1}^{q} r_{i,i} V_i.{}^tU_i$$

$$\left({}^tJ.J + \mu I\right)^{-1}.\left({}^tJ\right) = \sum_{i=1}^{q} \frac{\sqrt{\lambda_i}}{\lambda_i + \mu} V_i.{}^tU_i$$

(6). Soit X un vecteur fixé de $\mathbb{R}^p$ vérifiant $\nabla F(X) \neq 0$.
Pour tout vecteur h de $\mathbb{R}^p$, on pose $M(h) = L(h) + \frac{\mu}{2} \|h\|^2$.

(a).

$$F(X+h) - M(h) = (F(X+h) - L(h)) + (L(h) - M(h)) = (F(X+h) - L(h)) - \frac{\mu}{2} \|h\|^2$$

On avait déjà montré : $\lim_{\|h\| \to 0} \frac{|F(X+h) - L(h)|}{\|h\|} = 0$ alors on a immédiatement par somme de deux limites nulles :

$$\lim_{\|h\| \to 0} \frac{|F(X+h) - M(h)|}{\|h\|} = 0$$

(b). En notant $h = (h_1, \dots, h_p)$, on a : $M(h) = L(h) + \frac{\mu}{2} \sum_{i=1}^{n} h_i^2$. On en déduit que :

$$\forall k \in \{1, \dots, p\}, \quad \frac{\partial M}{\partial h_k}(h) = \frac{\partial L}{\partial h_k}(h) + \frac{\mu}{2}.2h_k$$
$$\nabla M(h) = \nabla L(h) + \mu h = \nabla F(X) + (G(X) + \mu I)h$$

On aura aussi :

$$\forall k \in \{1, \dots, p\}, \quad \frac{\partial^2 M}{\partial h_k^2}(h) = \frac{\partial^2 L}{\partial h_k^2}(h) + \mu$$

$$\forall k \in \{1, \dots, p\}, \forall i \in \{1, \dots, p\}, \quad \frac{\partial^2 M}{\partial h_i \partial h_k}(h) = \frac{\partial^2 L}{\partial h_i \partial h_k}(h)$$

On en déduit que :

$$\nabla^2 M(h) = \nabla^2 L(h) + \mu I = (G(X) + \mu I)$$

(c). h est un point critique de M si et seulement si $\nabla M(h) = 0$ c'est-à-dire si et seulement si $(G(X) + \mu I)h = -\nabla F(X)$.
En appliquant les résultats des questions précédentes à la matrice $J(X)$, puisque $G(X) = \left({}^tJ(X).J(X)\right)$, on sait que $(G(X) + \mu I)$ est inversible. M admet donc un unique point critique h^* et h^* est donné par la formule :

$$h^* = -(G(X) + \mu I)^{-1}.\nabla F(X) = -(G(X) + \mu I)^{-1} \left({}^tJ(X)\right) f(X)$$

$$h^* = -\left({}^tJ(X).J(X) + \mu I\right)^{-1}.\left({}^tJ(X)\right).f(X)$$

$$h^* = -\sum_{i=1}^{q} \frac{\sqrt{\lambda_i}}{\lambda_i + \mu} V_i.{}^tU_i f(X)$$

(d). La hessienne de M au point critique h^* est :

$$\nabla^2 M(h^*) = (G(X) + \mu I) = \left({}^tJ(X)\right).J(X) + \mu I$$

Cette matrice est symétrique réelle et ses valeurs propres sont les réels $\lambda_1 + \mu, \dots, \lambda_p + \mu$ qui sont strictement positifs (les $\lambda_1, \dots, \lambda_p$, valeurs propres de $G(X)$, sont positives et μ est strictement positif). On en déduit que M admet un minimum local en h^*.