



PRÉPASIMPLE

prepasimple.fr

HEC

Mathématiques 1

2008

Corrigé de mathématiques approfondies

Mathématiques approfondies — Prépa ECG

Annales de concours · Corrigé

Corrigé

Les produits scalaires sont les produits scalaires réels canoniques. Les vecteurs sont identifiés à des colonnes. Pour les extrema, on vérifie la régularité et les points critiques, puis on utilise la hessienne et les signes de ses valeurs propres, ou sa forme quadratique. Aucun déterminant n'est utilisé.

Partie I. Gradient et hessienne

1. (a) Les fonctions f_1, f_2 sont polynomiales. La fonction $F = \frac{1}{2}(f_1^2 + f_2^2)$ est donc polynomiale, de classe $\mathcal{C}^\infty$, en particulier $\mathcal{C}^2$, sur l'ouvert $\mathbb{R}^2$.

La dérivation des carrés donne

$$\frac{\partial F}{\partial x_1} = 2x_1 f_1 + f_2 = 2x_1^3 + 2x_1 x_2 + 3x_1 + x_2^2 + 1,$$

$$\frac{\partial F}{\partial x_2} = f_1 + 2x_2 f_2 = 2x_2^3 + 2x_1 x_2 + 3x_2 + x_1^2 + 1.$$

Le gradient est la colonne formée de ces deux expressions.

1. (b) Notons $A = \partial F / \partial x_1$ et $B = \partial F / \partial x_2$ dans cette question seulement. Un point est critique si et seulement si $A = B = 0$, ce qui équivaut à $A = 0$ et $A - B = 0$.

En soustrayant les formules de (a),

$$A - B = 2(x_1^3 - x_2^3) + 3(x_1 - x_2) + (x_2^2 - x_1^2).$$

Or $x_1^3 - x_2^3 = (x_1 - x_2)(x_1^2 + x_1 x_2 + x_2^2)$ et $x_2^2 - x_1^2 = -(x_1 - x_2)(x_1 + x_2)$. Donc

$$A - B = (x_1 - x_2)(2x_1^2 + 2x_1 x_2 + 2x_2^2 - x_1 - x_2 + 3).$$

On obtient exactement le système demandé; la transformation est réversible.

1. (c) Comme $(x_1 + x_2)^2 \geq 0$, $2x_1 x_2 \geq -x_1^2 - x_2^2$. Par conséquent,

$$2x_1^2 + 2x_1 x_2 + 2x_2^2 - x_1 - x_2 + 3 \geq x_1^2 + x_2^2 - x_1 - x_2 + 3,$$

et le membre de droite vaut $(x_1 - 1/2)^2 + (x_2 - 1/2)^2 + 5/2 > 0$.

La deuxième équation du système impose donc $x_1 = x_2 = t$. La première devient $2t^3 + 3t^2 + 3t + 1 = 0$.

Méthode 1 — Factorisation du polynôme

On vérifie

$$2t^3 + 3t^2 + 3t + 1 = (2t + 1)(t^2 + t + 1).$$

Comme $t^2 + t + 1 = (t + 1/2)^2 + 3/4 > 0$, l'unique racine réelle est $t = -1/2$.

Réciproquement, ce couple annule les deux composantes du gradient. L'unique point critique est donc $(-1/2, -1/2)$.

Méthode 2 — Monotonie du polynôme

Pour $g(t) = 2t^3 + 3t^2 + 3t + 1$, on a

$$g'(t) = 6t^2 + 6t + 3 = 6(t + 1/2)^2 + 3/2 > 0.$$

La fonction est strictement croissante sur $\mathbb{R}$ et s'annule en $-1/2$. Elle ne peut avoir d'autre racine. On obtient le même point critique. La factorisation est la méthode la plus courte ; la monotonie est utile lorsque la factorisation n'est pas repérée.

1. (d) La dérivation des expressions de (a) donne, en tout point,

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

Au point critique $X_0 = (-1/2, -1/2)$,

$$\nabla^2 F(X_0) = \begin{pmatrix} 7/2 & -2 \\ -2 & 7/2 \end{pmatrix}.$$

Les vecteurs $(1, 1)$ et $(1, -1)$ sont propres, de valeurs propres respectives $3/2$ et $11/2$, comme on le vérifie par multiplication. Ils forment une base de $\mathbb{R}^2$, donc ce sont toutes les valeurs propres.

La fonction est de classe $\mathcal{C}^2$ sur un ouvert contenant X_0 , son gradient s'y annule et les valeurs propres de sa hessienne sont strictement positives. Le critère du programme donne un minimum local strict en X_0 .

Sa valeur est $F(X_0) = \frac{1}{2}((3/4)^2 + (3/4)^2) = 9/16$. La question demande seulement un minimum local; cette conclusion ne repose sur aucun critère ancien utilisant les coefficients de la hessienne.

1. (e) La matrice des dérivées premières est

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

et les hessiennes des deux fonctions sont

$$\nabla^2 f_1 = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}, \quad \nabla^2 f_2 = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}.$$

En multipliant ${}^t(J(X))$ par ${}^t(f_1(X), f_2(X))$, on obtient ${}^t(2x_1 f_1 + f_2, f_1 + 2x_2 f_2)$, donc

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

D'autre part,

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

En ajoutant $f_1 \nabla^2 f_1 + f_2 \nabla^2 f_2$, les coefficients diagonaux deviennent $4x_1^2 + 1 + 2f_1 = 6x_1^2 + 2x_2 + 3$ et $4x_2^2 + 1 + 2f_2 = 6x_2^2 + 2x_1 + 3$. Les coefficients hors diagonale sont inchangés. Ainsi

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

2. (a) Ici $F(x_1, x_2) = \frac{1}{2} \|x_1 a + x_2 b - c\|^2$. Posons

$$A = \|a\|^2, \quad B = \|b\|^2, \quad C = \langle a, b \rangle, \quad d = \langle a, c \rangle, \quad e = \langle b, c \rangle.$$

Ces lettres sont des scalaires dans les questions 2(a)–(d). En dérivant la somme finie,

$$\frac{\partial F}{\partial x_1} = \langle a, x_1 a + x_2 b - c \rangle = Ax_1 + Cx_2 - d,$$

$$\frac{\partial F}{\partial x_2} = \langle b, x_1 a + x_2 b - c \rangle = Cx_1 + Bx_2 - e.$$

Ainsi $\nabla F = {}^t(Ax_1 + Cx_2 - d, Cx_1 + Bx_2 - e)$.

2. (b) La famille (a, b) est libre ; les deux vecteurs sont donc non nuls, et ils ne sont pas colinéaires. L'inégalité de Cauchy-Schwarz est stricte dans ce cas : $C^2 < AB$. Posons $\Delta = AB - C^2 > 0$; il s'agit ici d'une expression scalaire, sans utiliser de déterminant.

Le système critique est $Ax_1 + Cx_2 = d, Cx_1 + Bx_2 = e$. En multipliant la première équation par B et la seconde par C , puis en soustrayant, on obtient $\Delta x_1 = Bd - Ce$. En multipliant la seconde par A et la première par C , on obtient $\Delta x_2 = Ae - Cd$.

Comme $\Delta > 0$, les coordonnées sont nécessairement

$$\hat{x}_1 = \frac{Bd - Ce}{\Delta}, \quad \hat{x}_2 = \frac{Ae - Cd}{\Delta}.$$

Pour vérifier l'existence, la substitution donne

$$A\hat{x}_1 + C\hat{x}_2 = \frac{ABd - ACe + ACe - C^2d}{\Delta} = d,$$

$$C\hat{x}_1 + B\hat{x}_2 = \frac{BCd - C^2e + ABe - BCd}{\Delta} = e.$$

Ce point critique existe donc et il est unique.

2. (c) La fonction est polynomiale, donc de classe $\mathcal{C}^2$ sur $\mathbb{R}^2$. Sa hessienne, constante, est

$$H = \begin{pmatrix} A & C \\ C & B \end{pmatrix}.$$

Pour tout $u = (u_1, u_2) \neq 0$,

$${}^t u H u = Au_1^2 + 2Cu_1u_2 + Bu_2^2 = \|u_1a + u_2b\|^2 > 0,$$

car la liberté de (a, b) empêche cette combinaison d'être nulle. Cette étude de la forme quadratique montre que toutes les valeurs propres de H sont strictement positives : pour un vecteur propre non nul v de valeur propre λ , $\lambda\|v\|^2 = {}^t v H v > 0$. Le théorème spectral garantit que toutes les valeurs propres sont réelles et que l'on dispose d'une base propre.

Au point critique $(\hat{x}_1, \hat{x}_2)$, le critère de la hessienne donne donc un minimum local strict.

2. (d) Posons $E = \text{Vect}(a, b)$ et $\hat{u} = \hat{x}_1a + \hat{x}_2b$. Les équations critiques donnent

$$\langle a, \hat{u} - c \rangle = \langle b, \hat{u} - c \rangle = 0. \text{ Donc } \hat{u} - c \in E^\perp, \text{ et } \hat{u} \in E.$$

Ainsi $\hat{u}$ est la projection orthogonale de c sur E . Pour tout (x_1, x_2) , le vecteur $x_1a + x_2b - \hat{u}$ appartient à E et est orthogonal à $\hat{u} - c$. Par Pythagore,

$$F(x_1, x_2) = \frac{1}{2}\|x_1a + x_2b - \hat{u}\|^2 + \frac{1}{2}\|\hat{u} - c\|^2.$$

Le second terme vaut $F(\hat{x}_1, \hat{x}_2)$ et le premier est positif ou nul. Le point critique réalise donc un minimum global. L'égalité impose $(x_1 - \hat{x}_1)a + (x_2 - \hat{x}_2)b = 0$, puis, par liberté, $x_1 = \hat{x}_1$ et $x_2 = \hat{x}_2$. Le minimum global est unique.

3. (a) Posons $t = x_1 + x_2$. Alors

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

et chacune des deux dérivées partielles vaut

$$\sum_{i=1}^n (t - c_i) = n(t - \bar{c}).$$

Ainsi

$$\nabla F(x_1, x_2) = n(x_1 + x_2 - \bar{c}) \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Les points critiques sont exactement ceux de la droite $x_1 + x_2 = \bar{c}$.

3. (b) Pour un point critique, $\hat{x}_1 + \hat{x}_2 = \bar{c}$, d'où

$$F(\hat{x}_1, \hat{x}_2) = \frac{1}{2} \sum_{i=1}^n (\bar{c} - c_i)^2 = \frac{n}{2} s^2.$$

Pour tout t , écrivons $t - c_i = (t - \bar{c}) + (\bar{c} - c_i)$. En développant et en utilisant

$$\sum_i (\bar{c} - c_i) = 0,$$

$$\sum_i (t - c_i)^2 = n(t - \bar{c})^2 + \sum_i (\bar{c} - c_i)^2.$$

Ainsi

$$F(x_1, x_2) - F(\hat{x}_1, \hat{x}_2) = \frac{n}{2}(x_1 + x_2 - \bar{c})^2.$$

La variance statistique est définie avec le diviseur n . Elle est strictement positive, car les c_i ne sont pas tous égaux.

3. (c) La différence obtenue en (b) est positive ou nulle partout. Tous les points critiques sont donc des minima globaux, de valeur $ns^2/2$. Ils sont aussi des minima locaux, mais aucun n'est strict : chaque voisinage contient d'autres points de la droite critique donnant la même valeur.

Le résultat était prévisible puisque F ne dépend que de la somme $x_1 + x_2$ et reste constante sous la transformation $(x_1, x_2) \mapsto (x_1 + t, x_2 - t)$.

Pour préciser pourquoi une hessienne semi-définie ne suffisait pas à conclure seule, la

fonction est de classe $\mathcal{C}^2$ et sa hessienne vaut $n \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Ses valeurs propres sont $2n$ et 0 ,

associées à $(1, 1)$ et $(1, -1)$. Ce cas laisse le critère strict de la hessienne sans conclusion. C'est l'identité exacte de (b) qui établit les minima, sans déduire abusivement leur nature de la valeur propre nulle.

4. (a) Les f_i sont de classe $\mathcal{C}^2$; leurs carrés et leur somme finie le sont aussi. Pour chaque j ,

$$\frac{\partial F}{\partial x_j}(X) = \sum_{i=1}^n f_i(X) \frac{\partial f_i}{\partial x_j}(X).$$

Cette expression est le coefficient de ligne j de ${}^t(J(X))f(X)$. Donc

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

4. (b) En dérivant l'expression précédente par rapport à x_k ,

$$\frac{\partial^2 F}{\partial x_k \partial x_j}(X) = \sum_{i=1}^n \frac{\partial f_i}{\partial x_k}(X) \frac{\partial f_i}{\partial x_j}(X) + \sum_{i=1}^n f_i(X) \frac{\partial^2 f_i}{\partial x_k \partial x_j}(X).$$

La première somme est le coefficient (k, j) de ${}^t(J(X))J(X) = G(X)$. La seconde est celui de

$$\sum_i f_i(X) \nabla^2 f_i(X).$$

Ainsi

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

La régularité $\mathcal{C}^2$ permet d'utiliser Schwarz pour la symétrie des dérivées mixtes.

Partie II. Une approximation de F

Dans cette partie, X est fixé. Pour alléger les formules, notons $g = \nabla F(X)$ et $G = G(X)$, sans oublier l'hypothèse $g \neq 0$.

1. On développe le carré de norme :

$$\|f(X) + J(X)h\|^2 = \|f(X)\|^2 + 2{}^t(f(X))J(X)h + {}^tJ(X)J(X)h.$$

Par I.4(a), ${}^t(f(X))J(X)h = {}^thg$. En divisant par 2, on obtient

$$L(h) = F(X) + {}^thg + \frac{1}{2}{}^thGh.$$

2. (a) La matrice P est réelle et symétrique. Le théorème spectral donne une base orthonormale propre, donc une matrice Q telle que $Q^{-1} = {}^tQ$ et ${}^tQPQ = \text{diag}(\theta_1, \dots, \theta_p)$. En particulier, P est diagonalisable sur $\mathbb{R}$.
2. (b) Posons $z = {}^tQh$. Le changement de base orthonormale conserve la norme : $\|z\|^2 = \|h\|^2$. Alors

$${}^t_hPh = \sum_{j=1}^p \theta_j z_j^2.$$

L'inégalité triangulaire et $|\theta_j| \leq \theta$ donnent

$$|{}^t_hPh| \leq \sum_{j=1}^p |\theta_j| z_j^2 \leq \theta \sum_{j=1}^p z_j^2 = \theta \|h\|^2.$$

Le résultat couvre aussi $h = 0$ et le cas $P = 0$, où $\theta = 0$.

3. (a) La fonction F est de classe $\mathcal{C}^2$ sur l'ouvert $\mathbb{R}^p$. Le développement limité d'ordre 2 au point X est

$$F(X + h) = F(X) + {}^t_h \nabla F(X) + \frac{1}{2} {}^t_h \nabla^2 F(X) h + o(\|h\|^2),$$

lorsque $\|h\| \rightarrow 0$. Cette écriture signifie que le reste, divisé par $\|h\|^2$ pour $h \neq 0$, tend vers 0.

3. (b) Posons $E = \nabla^2 F(X) - G(X)$. Cette matrice est symétrique, comme différence de deux matrices symétriques. D'après II.1 et (a),

$$F(X + h) - L(h) = \frac{1}{2} {}^t_h E h + o(\|h\|^2).$$

La question 2(b) fournit une constante $C \geq 0$ telle que $|{}^t_h E h| \leq C \|h\|^2$. Pour $h \neq 0$, on en déduit

$$\left| \frac{F(X+h) - L(h)}{\|h\|} \right| \leq \frac{C}{2} \|h\| + \frac{|o(\|h\|^2)|}{\|h\|}.$$

Le dernier terme est $\|h\|$ multiplié par une quantité tendant vers 0. Les deux termes de droite tendent donc vers 0, ce qui prouve la limite demandée.

L'expression « approximation à l'ordre 2 » est ici une convention locale définie dans l'énoncé par cette limite. Le résultat établi est $F(X+h) - L(h) = o(\|h\|)$. Plus précisément, le développement précédent donne une erreur majorée par une constante fois $\|h\|^2$ pour h assez petit. Le modèle L est quadratique en h , mais sa hessienne $G(X)$ n'est pas nécessairement celle de F : il ne faut donc pas le confondre avec le développement limité d'ordre 2 de F .

Par exemple, dans I.1 au point $X = (0, 0)$, on a $\nabla F(X) = (1, 1) \neq 0$, $G(X) = I_2$ et $\nabla^2 F(X) = 3I_2$. La différence commence alors par $\|h\|^2$ et n'est pas $o(\|h\|^2)$. La terminologie locale ne permet donc pas d'affirmer que la différence soit négligeable devant $\|h\|^2$.

4. (a) La fonction

$$\varphi_1(h) = \sum_{i=1}^p h_i g_i$$

est linéaire ; donc $\partial \varphi_1 / \partial h_j = g_j = \partial F / \partial x_j(X)$.

On a

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

La dérivée par rapport à h_j comporte les termes où $k = j$ et ceux où $i = j$:

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

Comme $G(X)$ est symétrique, les deux sommes sont égales. Ainsi

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

Le terme diagonal est bien compté deux fois, conformément à la dérivée de h_j^2 .

4. (b) Puisque $L(h) = F(X) + \varphi_1(h) + \frac{1}{2}\varphi_2(h)$, la dérivée partielle d'indice j est

$$g_j + \sum_i g_{i,j}(X) h_i.$$

La symétrie de G identifie la seconde somme au coefficient d'indice j de Gh . Donc

$$\nabla L(h) = g + Gh.$$

4. (c) Les coefficients de G sont constants puisque X est fixé. En dérivant le coefficient d'indice j du gradient par rapport à h_k , on obtient $g_{j,k}(X)$. Ainsi $\nabla^2 L(h) = G(X)$ pour tout h .

5. (a) La matrice $H = {}^t J J$ est symétrique car ${}^t({}^t J J) = {}^t J J$. Le théorème spectral la rend diagonalisable sur $\mathbb{R}$.

Si $v \neq 0$ est propre pour une valeur propre λ , alors

$$\lambda \|v\|^2 = {}^t v H v = {}^t (J v) (J v) = \|J v\|^2 \geq 0.$$

Comme $\|v\|^2 > 0$, $\lambda \geq 0$. Toutes ses valeurs propres sont positives ou nulles.

5. (b) Si $J v = 0$, alors ${}^t J J v = 0$. Réciproquement, si ${}^t J J v = 0$, alors $0 = {}^t v {}^t J J v = \|J v\|^2$, donc $J v = 0$. Ainsi, pour toute matrice rectangulaire,

$$\text{Ker}({}^t J J) = \text{Ker}(J).$$

Lorsque la matrice carrée ${}^t J J$ est inversible, ce noyau est nul, et le théorème du rang

donne $\text{rg}(J) = p$. L'identité montre aussi la réciproque : si J est de rang p , alors tJJ est inversible. Elle sera utile dans la partie III.

6. Au point critique $\hat{h}$, la question 4(b) donne $g + G\hat{h} = 0$. En prenant le produit scalaire avec $\hat{h}$,

$$\langle \hat{h}, g \rangle = -{}^t\hat{h}G\hat{h} = -\|J(X)\hat{h}\|^2 \leq 0.$$

L'inégalité demandée est donc satisfaite, même si G n'est pas inversible. Sous l'hypothèse $g \neq 0$, elle est en fait stricte dès qu'un point critique existe : si le carré de norme était nul, $J(X)\hat{h} = 0$, puis $G\hat{h} = 0$ et l'équation critique donnerait $g = 0$, contradiction.

7. (a) L'équation critique $g + Gh = 0$ a une unique solution puisque G est inversible. Avec I.4(a),

$$\hat{h} = -G(X)^{-1}\nabla F(X) = -G(X)^{-1}({}^t(J(X))f(X)).$$

L'existence de l'inverse est ici une hypothèse explicite.

7. (b) Comme $g \neq 0$, le vecteur $\hat{h}$ de (a) est non nul. Par 5(b), $J(X)$ est de rang p , donc $J(X)\hat{h} \neq 0$. La question 6 donne

$$\langle g, \hat{h} \rangle = -\|J(X)\hat{h}\|^2 < 0.$$

Ainsi $\hat{h}$ est une direction de décroissance de F en X , au sens défini dans l'énoncé.

Méthode 1 — Hessienne du modèle

Pour conclure sur le minimum de L , on examine sa hessienne, et pas seulement la propriété de décroissance de F . La fonction L est polynomiale de degré au plus 2, donc de classe $\mathcal{C}^2$ sur $\mathbb{R}^p$. Le gradient s'annule en $\hat{h}$ et sa hessienne est G . Par 5(a), ses valeurs propres sont positives ou nulles ; l'inversibilité exclut la valeur 0. Elles sont donc toutes strictement positives. Le critère de la hessienne donne un minimum local strict de L en $\hat{h}$.

Méthode 2 — Identité exacte d'accroissements

On peut établir directement un minimum global strict par une identité exacte. Pour tout $k \in \mathbb{R}^p$, en développant le modèle quadratique et en utilisant $g + G\hat{h} = 0$,

$$L(\hat{h} + k) - L(\hat{h}) = \frac{1}{2} {}^t k G k = \frac{1}{2} \|J(X)k\|^2.$$

Cette différence est positive pour $k \neq 0$. La hessienne suffit pour la conclusion locale demandée ; l'identité explicite le caractère global du minimum du modèle.

Partie III. Une décomposition d'une matrice rectangulaire

Ici $J \neq 0$ et $\mu > 0$. La décomposition est construite à partir du théorème spectral appliqué aux matrices symétriques ; elle n'est pas invoquée comme un théorème extérieur au référentiel.

1. La matrice $H = {}^t J J$ est réelle symétrique et ses valeurs propres sont positives ou nulles par II.5(a). Par le théorème spectral, on peut choisir une base orthonormale de vecteurs propres et ranger les valeurs propres dans l'ordre décroissant.

Au moins une valeur propre est strictement positive. En effet, sinon, la diagonalisation donnerait $H = 0$. Alors $\|Jv\|^2 = {}^t v H v = 0$ pour tout v , ce qui entraînerait $J = 0$, contraire à l'hypothèse.

Notons q le nombre de valeurs propres strictement positives, comptées avec leurs répétitions dans une base propre. On a $1 \leq q \leq p$, et on peut les écrire $\lambda_1 \geq \dots \geq \lambda_q > 0$, suivies de zéros si $q < p$.

La matrice V dont les colonnes sont ces vecteurs propres orthonormaux satisfait $V^{-1} = {}^t V$ et ${}^t V H V = D$. Ceci établit toutes les propriétés demandées.

2. (a) Les matrices H et D sont semblables, donc elles ont le même rang. La matrice diagonale D a exactement q coefficients non nuls. Ainsi $\text{rg}({}^t J J) = q$.

2. (b) Pour $1 \leq i \leq q$, $HV_i = \lambda_i V_i$. En multipliant par J ,

$$(J^t J)(JV_i) = JHV_i = \lambda_i JV_i.$$

Il faut vérifier que JV_i est non nul : comme $\|V_i\| = 1$,

$$\|JV_i\|^2 = {}^t(V_i)HV_i = \lambda_i > 0.$$

Ainsi JV_i est bien un vecteur propre de $J^t J$ pour λ_i . Chaque valeur propre non nulle de H est donc une valeur propre de $J^t J$.

Réciproquement, soit $z \neq 0$ tel que $J^t Jz = \lambda z$, avec $\lambda \neq 0$. Si ${}^t Jz = 0$, cette équation donnerait $\lambda z = 0$, contradiction. Donc ${}^t Jz \neq 0$, et

$$H({}^t Jz) = {}^t J(J^t Jz) = \lambda {}^t Jz.$$

La valeur λ est donc aussi propre pour H . Les deux matrices ont les mêmes valeurs propres non nulles.

2. (c) Supposons

$$\sum_{j=1}^r a_j JY_j = 0.$$

Posons

$$Y = \sum_{j=1}^r a_j Y_j,$$

qui appartient à l'espace propre de H pour λ . L'égalité supposée donne $JY = 0$, puis $HY = {}^t JY = 0$.

Mais $HY = \lambda Y$ et $\lambda \neq 0$, donc $Y = 0$. La famille $(Y_1, \dots, Y_r)$ est libre, ce qui impose $a_j = 0$ pour chaque j . Les images $(JY_1, \dots, JY_r)$ forment donc une famille libre.

2. (d) Pour une valeur propre $\lambda \neq 0$, l'application $Y \mapsto JY$ envoie l'espace propre de H dans celui de $K = J^t J$ et y est injective, d'après (c). Donc $\dim E_\lambda(H) \leq \dim E_\lambda(K)$.

Le même raisonnement avec tJ donne une injection dans l'autre sens : si $z \in E_\lambda(K)$ et ${}^tJz = 0$, alors $Kz = 0 = \lambda z$, donc $z = 0$. Ainsi $\dim E_\lambda(K) \leq \dim E_\lambda(H)$. Les dimensions sont égales.

On peut aussi vérifier directement la surjectivité : pour $z \in E_\lambda(K)$, le vecteur $Y = {}^tJz/\lambda$ appartient à $E_\lambda(H)$ et $JY = z$.

La matrice K est symétrique réelle et diagonalisable. Dans une base propre, son rang est la somme des dimensions des espaces propres associés aux valeurs propres non nulles. Ces dimensions sont les mêmes que pour H ; leur somme vaut q par (a). Donc $\text{rg}(J^tJ) = q$.

En complément, l'identité de noyaux de II.5(b) et le théorème du rang donnent $\text{rg}(J) = \text{rg}(H) = q$.

3. (a) Les divisions par $\sqrt{\lambda_i}$ sont licites car $\lambda_i > 0$ pour $i \leq q$. Par 2(b), chaque U_i est propre pour $K = J^tJ$, associé à λ_i .

Pour $1 \leq i, j \leq q$,

$$\langle U_i, U_j \rangle = \frac{{}^t(V_i)HV_j}{\sqrt{\lambda_i}\sqrt{\lambda_j}} = \frac{\lambda_j}{\sqrt{\lambda_i}\sqrt{\lambda_j}} \langle V_i, V_j \rangle.$$

Si $i \neq j$, le produit scalaire est nul ; si $i = j$, il vaut 1. La famille est donc orthonormale, même lorsque certaines valeurs propres sont égales.

3. (b) Posons $E = \text{Vect}(U_1, \dots, U_q)$. Il est de dimension q . Tous ces vecteurs appartiennent à $\text{Im}(K)$ puisque $U_i = KU_i/\lambda_i$, et l'image de K a dimension q d'après 2(d). Donc $E = \text{Im}(K)$.

Comme K est symétrique, $E^\perp = \text{Ker}(K)$. Pour le vérifier directement, $z \perp \text{Im}(K)$ équivaut à $\langle z, Kv \rangle = 0$ pour tout v , soit $\langle Kz, v \rangle = 0$ pour tout v , donc $Kz = 0$.

Choisissons une base orthonormale $(U_{q+1}, \dots, U_n)$ du supplémentaire orthogonal $E^\perp$. Ces vecteurs sont propres pour la valeur 0. Avec les $U_1, \dots, U_q$, ils donnent une base orthonormale propre de tout $\mathbb{R}^n$.

Si $q = n$, le supplémentaire est nul et aucun vecteur n'est à ajouter. Ce cas est possible lorsque $n = p = q$.

4. Pour $1 \leq j \leq q$, la définition de U_j donne $JV_j = \sqrt{\lambda_j}U_j$. Pour $q < j \leq p$,

$$\|JV_j\|^2 = {}^t(V_j)HV_j = \lambda_j\|V_j\|^2 = 0,$$

donc $JV_j = 0$.

Le coefficient (i, j) de tUJV est $\langle U_i, JV_j \rangle$. Pour $j \leq q$, il vaut $\sqrt{\lambda_j}$ si $i = j$ et 0 sinon, par orthonormalité; pour $j > q$, il est nul. Ce sont exactement les coefficients de S . Ainsi $S = {}^tUJV$.

Les colonnes de U et de V sont des bases orthonormales, donc $U^{-1} = {}^tU$ et $V^{-1} = {}^tV$. En multipliant l'égalité à gauche par U et à droite par tV ,

$$J = US{}^tV.$$

Les dimensions sont $U : n \times n$, $S : n \times p$, $V : p \times p$, ce qui donne bien une matrice J de taille $n \times p$.

5. (a) Méthode 1 — Diagonalisation

On a $H + \mu I_p = V(D + \mu I_p){}^tV$. Les coefficients diagonaux $\lambda_i + \mu$ sont tous strictement positifs, y compris lorsque $\lambda_i = 0$. La matrice diagonale est inversible, donc $H + \mu I_p$ est inversible, avec

$$(H + \mu I_p)^{-1} = V(D + \mu I_p)^{-1}{}^tV.$$

Méthode 2 — Noyau et carré de norme

Si $(H + \mu I_p)v = 0$, alors

$$0 = {}^tv(H + \mu I_p)v = \|Jv\|^2 + \mu\|v\|^2.$$

Les deux termes sont positifs ou nuls et $\mu > 0$, donc $v = 0$. La matrice carrée a un noyau nul et est inversible. Cette preuve établit l'inversibilité sans connaître les valeurs propres; la première fournit en plus la formule utilisée en (b).

5. (b) Par 4, ${}^tJ = V^tS^tU$. En utilisant l'inverse obtenu en (a),

$$(H + \mu I_p)^{-1} {}^tJ = V(D + \mu I_p)^{-1} {}^tV V^t S^t U = V(D + \mu I_p)^{-1} {}^tS^t U.$$

La matrice $(D + \mu I_p)^{-1} {}^tS^t$ est de taille $p \times n$. Ses seuls coefficients éventuellement non nuls sont les coefficients diagonaux, qui valent $\sqrt{\lambda_i}/(\lambda_i + \mu)$ pour $1 \leq i \leq p$. Elle est donc égale à R . On obtient

$$({}^tJ J + \mu I_p)^{-1} {}^tJ = V R^t U.$$

Toutes les divisions sont justifiées par $\lambda_i + \mu > 0$.

5. (c) Les coefficients de R sont nuls hors des q premières positions diagonales, car $\sqrt{\lambda_i} = 0$ pour $i > q$. En développant le produit matriciel,

$$V R^t U = \sum_{i=1}^q \frac{\sqrt{\lambda_i}}{\lambda_i + \mu} V_i^t (U_i).$$

Chaque produit $V_i^t (U_i)$ est une matrice de taille $p \times n$, dont le coefficient (a, b) vaut $(V_i)_a (U_i)_b$. Cette somme finie possède donc la bonne taille et donne la formule demandée.

6. (a) Par II.3(b), $[F(X + h) - L(h)]/\|h\| \rightarrow 0$. Pour $h \neq 0$,

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

Le second terme tend vers 0, et le premier aussi. La limite demandée vaut donc 0.

6. (b) La fonction M est polynomiale en h , donc de classe $\mathcal{C}^2$ sur $\mathbb{R}^p$. Par II.4 et la dérivation de

$$\frac{\mu}{2} \sum_j h_j^2,$$

$$\nabla M(h) = \nabla F(X) + (G(X) + \mu I_p)h,$$

$$\nabla^2 M(h) = G(X) + \mu I_p.$$

6. (c) L'hypothèse $\nabla F(X) \neq 0$ et I.4(a) impliquent que $J(X) \neq 0$: si cette matrice était nulle, le gradient serait nul. Les résultats de cette partie s'appliquent donc à $J(X)$.

Par 5(a), $G(X) + \mu I_p$ est inversible. L'équation critique possède l'unique solution

$$h^* = -(G(X) + \mu I_p)^{-1} \nabla F(X) = -({}^t(J(X))J(X) + \mu I_p)^{-1} {}^t(J(X))f(X).$$

Notons q, λ_i, V_i, U_i les éléments construits pour la matrice $J(X)$. Par 5(c),

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

Le facteur ${}^t(U_i)f(X)$ est un scalaire ; chaque terme est donc un vecteur de $\mathbb{R}^p$. Cette expression remplace le calcul explicite de l'inverse par une somme après construction des bases.

6. (d) Méthode 1 — Critère de la hessienne

La fonction M est de classe $\mathcal{C}^2$ sur l'ouvert $\mathbb{R}^p$, et son gradient s'annule en h^* . Sa hessienne, constante, est $G(X) + \mu I_p$. Dans la base orthonormale propre de $G(X)$, ses valeurs propres sont $\lambda_i + \mu > 0$. Le critère de la hessienne donne un minimum local strict en h^* .

Méthode 2 — Différence de valeurs du modèle

Une identité exacte apporte une conclusion plus forte : pour tout $k \in \mathbb{R}^p$, les termes

linéaires s'annulent grâce à l'équation critique et

$$M(h^* + k) - M(h^*) = \frac{1}{2}k(G(X) + \mu I_p)k = \frac{1}{2}\|J(X)k\|^2 + \frac{\mu}{2}\|k\|^2.$$

Pour $k \neq 0$, le second terme est strictement positif. Ainsi h^* est même l'unique minimum global du modèle régularisé M , que $J(X)$ soit de rang plein ou non. Le critère de la hessienne est adapté à une étude locale générale ; l'identité exacte exploite la structure quadratique du modèle et fournit ici une conclusion globale.

Relevé des adaptations

Aucune question n'a nécessité de modification mathématique. Les critères d'extrema et la construction de la décomposition matricielle sont traités avec les outils du référentiel.