



Examen – Optimisation - EDP

1 Introduction

- Les deux parties sont à rédiger sur des feuilles séparées ;
- Documents autorisés : 1 page A4 recto verso manuscrite ;
- Le barème est donné à titre indicatif.
- Un corrigé sera mis sous le **GitLab** dans la journée.

2 Partie I

▷ **Exercice 1.** (3 points) On cherche à prévoir pour le lendemain la valeur d'un indice noté PO de pollution à l'ozone en exploitant 2 prédicteurs potentiels constitués de prévisions d'un modèle météorologique à l'échéance de 24 heures des variables :

- T : la température de l'air à 2 mètres en $^{\circ}C$;
- FF : la force du vent à 10 mètres en m/s ;

On dispose de $n = 80$ valeurs de chacune des variables T, FF et PO (cf. la table 1).

i	1	2	...	80
PO	PO_1	PO_2	...	PO_{80}
T	T_1	T_2	...	T_{80}
FF	FF_1	FF_2	...	FF_{80}

TABLE 1 – Données.

On considère le modèle $PO(\beta, T, FF) = \beta_0 + \beta_1 T + \beta_2 FF + \beta_3 T \cdot FF$.

1.1. Écrire le problème aux moindres carrés permettant d'estimer les paramètres $\beta_0, \beta_1, \beta_2$ et β_3 . Le problème est-il un problème aux moindres carrés linéaire ? Si oui on donnera le vecteur \mathbf{y} et la matrice \mathbf{X} permettant d'écrire ce problème sous la forme

$$(P) \begin{cases} \min \frac{1}{2} \|\mathbf{y} - \mathbf{X}\beta\|^2 \\ \beta \in \mathbb{R}^p. \end{cases}$$

► Il s'agit bien d'un problème aux moindres carrés linéaire. La fonction résidu s'écrit :

$$r: \mathbb{R}^4 \longrightarrow \mathbb{R}^{80}$$

$$\beta \longmapsto r(\beta) = \begin{pmatrix} r_1(\beta) \\ \vdots \\ r_{80}(\beta) \end{pmatrix} = \mathbf{y} - \mathbf{X}\beta \ .$$

avec

$$\mathbf{y} = \begin{pmatrix} PO_1 \\ \vdots \\ PO_{80} \end{pmatrix} \quad \text{et} \quad \mathbf{X} = \begin{pmatrix} 1 & T_1 & FF_1 & T_1 FF_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & T_{80} & FF_{80} & T_{80} FF_{80} \end{pmatrix}$$

▷ **Exercice 2.** (6 points) On considère le problème d'optimisation suivant

$$(P) \begin{cases} \min f(x) = x_1^4 + x_2^2 - 2x_1^2 x_2 + 2x_2 \\ x \in \mathbb{R}^2. \end{cases}$$

2.1. Calculer $\nabla f(x)$ et $\nabla^2 f(x)$.

►

—

$$\nabla f(x) = \begin{pmatrix} 4x_1^3 - 4x_1 x_2 \\ 2x_2 - 2x_1^2 + 2 \end{pmatrix};$$

—

$$\nabla^2 f(x) = \begin{pmatrix} 12x_1^2 - 4x_2 & -4x_1 \\ -4x_1 & 2 \end{pmatrix}.$$

2.2. Donner le ou les points qui vérifient la condition nécessaire de solution du premier ordre.

► $\nabla f(x) = 0 \iff x_1 = 0$ et $x_2 = -1$. Le seul point qui vérifie la condition nécessaire du premier ordre est donc $x^* = (0, -1)$.

2.3. Déterminer le/les minima locaux.

► La matrice hessienne en x^* est

$$\nabla^2 f(0, -1) = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}.$$

Elle est définie positive. La condition suffisante du deuxième ordre de solution est donc vérifiée. Par suite le point $x^* = (0, -1)$ est un minimum local.

2.4. La fonction f est-elle convexe ?

►

$$\nabla^2 f(0, 1) = \begin{pmatrix} -4 & 0 \\ 0 & 2 \end{pmatrix}.$$

n'est pas semi-définie positive. Par suite f n'est pas convexe.

▷ **Exercice 3.** (5 points) Soit λ est un réel strictement positif, on considère le problème suivant

$$(P) \begin{cases} \min f(\beta) = \frac{1}{2} \|\mathbf{y} - \mathbf{X}\beta\|^2 + \frac{\lambda}{2} \|\beta\|^2 \\ \beta \in \mathbb{R}^p. \end{cases}$$

3.1. Donner $\nabla f(\beta)$ et $\nabla^2 f(\beta)$.

►

- $\nabla f(\beta) = \mathbf{X}^T \mathbf{X} \beta - \mathbf{X}^T y + \lambda \beta$;
- $\nabla^2 f(\beta) = \mathbf{X}^T \mathbf{X} + \lambda I$

3.2. Le problème est-il convexe ?

► \mathbb{R}^p est convexe et $\nabla^2 f(\beta)$ est définie positive pour tout β : $\mathbf{X}^T \mathbf{X}$ est semi-définie positive et $\lambda > 0$, donc les valeurs propres de $\mathbf{X}^T \mathbf{X} + \lambda I$ qui sont les valeurs propres de $\mathbf{X}^T \mathbf{X} + \lambda$ sont strictement positives. Par suite f est strictement convexe et (P) est un problème convexe.

3.3. Caractérisez la solution. Est-elle unique ?

► Le problème est défini sur un ouvert convexe. C'est un problème convexe et différentiable. Par suite β est une solution si et seulement si

$$\nabla f(\beta) = (\mathbf{X}^T \mathbf{X} + \lambda I) \beta - \mathbf{X}^T y = 0,$$

Soit

$$(\mathbf{X}^T \mathbf{X} + \lambda I) \beta = \mathbf{X}^T y.$$

La matrice $(\mathbf{X}^T \mathbf{X} + \lambda I)$ étant symétrique et définie positive, elle est inversible. Par suite il y a une solution et une seule.

3 Partie II

▷ **Exercice 4.** (3 points) On suppose que la fonction $u : \mathbb{R} \rightarrow \mathbb{R}$, est une fonction \mathcal{C}^4 sur le segment $[x - h_0, x + h_0]$, avec $h_0 > 0$.

4.1. En utilisant les développements limités de Taylor-Lagrange de $u(x + h)$ et de $u(x - h)$ démontrez qu'il existe

$$C \geq 0 \text{ t.q. } \forall h \in]0, h_0] \mid \left| \frac{u(x + h) - 2u(x) + u(x - h)}{h^2} - u^{(2)}(x) \right| \leq Ch^2$$

► Le développement de Taylor-Lagrange donne

$$\begin{aligned} u(x + h) &= u(x) + hu'(x) + \frac{h^2}{2}u''(x) + \frac{h^3}{6}u^{(3)}(x) + \frac{h^4}{24}u^{(4)}(\xi^+) \\ u(x - h) &= u(x) - hu'(x) + \frac{h^2}{2}u''(x) - \frac{h^3}{6}u^{(3)}(x) + \frac{h^4}{24}u^{(4)}(\xi^-) \end{aligned}$$

où ξ^+ et ξ^- sont des points dans les intervalles $]x, x + h[$ et $]x - h, x[$. On en déduit alors que

$$\frac{u(x + h) + u(x - h) - 2u(x)}{h^2} - u''(x) = \frac{h^2}{12} \left(\frac{1}{2}(u^{(4)}(\xi^+) + u^{(4)}(\xi^-)) \right).$$

Le théorème des valeurs intermédiaires nous dit alors qu'il existe $\xi \in]x - h, x + h[$ tel que

$$u^{(4)}(\xi) = \frac{1}{2}(u^{(4)}(\xi^+) + u^{(4)}(\xi^-)).$$

Il suffit alors de prendre

$$C = \frac{1}{12} \sup_{y \in [x-h_0, x+h_0]} u^{(4)}(y)$$

▷ **Exercice 5.** (5 points) Soit σ est une fonction donnée de \mathbb{R} à valeurs dans \mathbb{R} . Pour une matrice W on note w_i la i^e ligne de cette matrice.

Un modèle à une couche (constituée de m neurones) d'un réseau de neurones est défini par

$$y: \quad \mathbb{R}^n \times \mathbb{R}^{(n+1)m} \quad \longrightarrow \quad \mathbb{R}^m$$

$$(x, \beta) = (x, \begin{pmatrix} w_1^T \\ \vdots \\ w_m^T \\ b \end{pmatrix}) \quad \longmapsto \quad y(x, \beta) = \begin{pmatrix} \sigma(w_1 x + b_1) \\ \vdots \\ \sigma(w_m x + b_m) \end{pmatrix} = \begin{pmatrix} \sigma(\langle w_1^T, x \rangle + b_1) \\ \vdots \\ \sigma(\langle w_m^T, x \rangle + b_m) \end{pmatrix}.$$

5.1. Soit x^k fixé, calculer en fonction de la dérivée σ' la matrice jacobienne en β de la fonction $g_1(\beta) = y(x^k, \beta)$.

►

Remarque 1. On désire, à partir de données $(x^k, y^k) \in \mathbb{R}^n \times \mathbb{R}^m, k = 1, \dots, K$ estimer par la méthode des moindres carrés les paramètres β . Pour cela il nous faut donc résoudre le problème d'optimisation

$$(P) \begin{cases} \min f(\beta) = \frac{1}{2} \sum_{k=1}^K \|y^k - y(x^k, \beta)\|^2 \\ \beta \in \mathbb{R}^{(n+1)m}. \end{cases}$$

Pour résoudre ce problème on a besoin de calculer les dérivés partielles

$$\frac{\partial y(x^k, \beta)}{\partial \beta}.$$

$$J_{g_1}(\beta) = \begin{pmatrix} \sigma'(w_1 x^k + b_1)(x^k)^T & 0 & \dots & 0 & \sigma'(w_1 x^k + b_1) & 0 & \dots & 0 \\ 0 & \sigma'(w_2 x^k + b_2)(x^k)^T & \dots & 0 & 0 & \sigma'(w_2 x^k + b_2) & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \sigma'(w_m x^k + b_m)(x^k)^T & 0 & 0 & \dots & \sigma'(w_m x^k + b_m) \end{pmatrix}$$

Remarque 2. Cette matrice jacobienne est la matrice jacobienne de la dérivée partielle :

$$\frac{\partial y(x^k, \beta)}{\partial \beta}.$$

On considère maintenant l'application, toujours noté y

$$y: \quad \mathbb{R}^n \times \mathcal{M}_{m,n}(\mathbb{R}) \times \mathbb{R}^m \quad \longrightarrow \quad \mathbb{R}^m$$

$$(x, W, b) \quad \longmapsto \quad y(x, W, b) = \sigma(Wx + b),$$

avec

$$\begin{aligned}\sigma: \mathbb{R}^m &\longrightarrow \mathbb{R}^m \\ z &\longmapsto \sigma(z) = \begin{pmatrix} \sigma(z_1) \\ \vdots \\ \sigma(z_m) \end{pmatrix}.\end{aligned}$$

5.2. Soit x^k fixé et soit la fonction

$$\begin{aligned}g_2: \mathcal{M}_{m,n}(\mathbb{R}) \times \mathbb{R}^m &\longrightarrow \mathbb{R}^m \\ (W, b) &\longmapsto g_2(W, b) = Wx^k + b.\end{aligned}$$

Calculer $g_2'(W, b) \cdot (H, h)$ (on précisera les espaces auxquels appartiennent H et h).

► H est dans $\mathcal{M}_{m,n}(\mathbb{R})$ et h est dans \mathbb{R}^m . Revenons à la définition

$$g_2(W + H, b + h) = (W + H)x^k + b + h = g_2(W, b) + (Hx^k + h) + 0.$$

Par suite $g_2'(W, b) \cdot (H, h) = Hx^k + h$

Remarque 3. g_2 est une application linéaire et continue. Par suite $g_2'(W, b) \cdot (H, h) = g_2(H, h)$!

5.3. En déduire, en fonction de σ' ,

$$\frac{\partial y(x^k, \beta)}{\partial \beta} \cdot (H, h).$$

► Calculons maintenant

$$\sigma(z) \cdot \phi = \begin{pmatrix} \sigma'(z_1)\phi_1 \\ \sigma'(z_2)\phi_2 \\ \vdots \\ \sigma'(z_m)\phi_m \end{pmatrix}$$

Par suite, en appliquant le théorème des fonctions composées on obtient

$$\frac{\partial y(x^k, \beta)}{\partial \beta} \cdot (H, h) = \begin{pmatrix} \sigma'(w_1.x^k + b_1)(H_1.x^k + h_1) \\ \vdots \\ \sigma'(w_m.x^k + b_m)(H_m.x^k + h_m) \end{pmatrix}$$

5.4. Quel lien pouvez-vous faire entre les questions 5.1 et 5.3

► On a

$$\frac{\partial y(x^k, \beta)}{\partial \beta} \cdot (H, h) = J_{g_1}(\beta) \begin{pmatrix} H_1^T \\ \vdots \\ H_m^T \\ h \end{pmatrix}$$