

MTH 8302 - Modèles de Régression et d'Analyse de Variance

Clarifications Concernant les 4 Premières Questions du Problème 1 de la Série d'Exercices 1 et le Calcul de la Hessienne avec le Produit Extérieur

Polytechnique Montréal - Hiver 2025

Chiheb Trabelsi

9 Février 2025

POLYTECHNIQUE
MONTREAL
UNIVERSITÉ
D'INGENIERIE



Table des Matières

1 Problème 1 : Gradients et Hessiennes

Problème 1 : Gradients et Hessiennes

Problème 1 : Gradients et Hessiennes

• Rappel :

- Une matrice $\mathbf{A} \in \mathbb{R}^{n \times n}$ est symétrique si $\mathbf{A}^T = \mathbf{A}$.
- Le gradient $\nabla f(\mathbf{x})$ d'une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est un vecteur contenant les dérivées partielles :

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix} = \left(\frac{\partial f}{\partial \mathbf{x}} \right).$$

- La Hessienne $\nabla^2 f(\mathbf{x})$ est une matrice symétrique $n \times n$ contenant les dérivées secondes :

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}.$$

Règles de Dérivées Partielles :

- **Dérivée d'une forme quadratique** : Si \mathbf{x} est un vecteur et \mathbf{A} est une matrice symétrique, la dérivée de la forme quadratique $\mathbf{x}^T \mathbf{A} \mathbf{x}$ par rapport à \mathbf{x} est donnée par :

$$\frac{\partial}{d\mathbf{x}}(\mathbf{x}^T \mathbf{A} \mathbf{x}) = 2\mathbf{A}\mathbf{x}$$

- **Dérivée d'une forme linéaire** : Pour un vecteur \mathbf{x} et une matrice constante \mathbf{A} , la dérivée de $\mathbf{A}\mathbf{x}$ par rapport à \mathbf{x} est :

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{A}\mathbf{x}) = \mathbf{A}$$

où \mathbf{A} est considéré comme une constante par rapport à \mathbf{x} .

- **Dérivée d'un produit de type vecteur-matrice-vecteur** : Si \mathbf{b} et \mathbf{x} sont des vecteurs et \mathbf{A} est une matrice, alors la dérivée de $\mathbf{b}^T \mathbf{A} \mathbf{x}$ par rapport à \mathbf{x} est :

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{b}^T \mathbf{A} \mathbf{x}) = \mathbf{A}^T \mathbf{b}$$

Règles de Dérivation :

- **Dérivée d'une matrice dépendant d'un vecteur** : Si \mathbf{A} est une matrice qui ne dépend pas du vecteur \mathbf{x} et \mathbf{A} n'est pas nécessairement symétrique, alors la dérivée de $\mathbf{x}^T \mathbf{A} \mathbf{x}$ est :

$$\frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^T \mathbf{A} \mathbf{x}) = (\mathbf{A} + \mathbf{A}^T) \mathbf{x}$$

si \mathbf{A} n'est pas nécessairement symétrique.

- **Dérivée du produit de deux matrices** : Si $\mathbf{A}(\mathbf{x})$ et $\mathbf{B}(\mathbf{x})$ sont des matrices dépendant du vecteur \mathbf{x} , alors la dérivée de leur produit $\mathbf{A}(\mathbf{x})\mathbf{B}(\mathbf{x})$ est donnée par :

$$\frac{\partial}{\partial \mathbf{x}} (\mathbf{A}(\mathbf{x})\mathbf{B}(\mathbf{x})) = \frac{\partial \mathbf{A}}{\partial \mathbf{x}} \mathbf{B} + \mathbf{A} \frac{\partial \mathbf{B}}{\partial \mathbf{x}}$$

Problème 1 : Question 1

- **Fonction** : $g(\mathbf{x}) = \log(1 + \mathbf{x}^\top \mathbf{C}\mathbf{x}) + \mathbf{d}^\top \mathbf{x}$, où \mathbf{C} symétrique.
- **Objectif** : Trouver $\nabla_{\mathbf{x}}g(\mathbf{x})$
- **Solution** :
- **Justification Détaillée** :

Problème 1 : Question 1 (Solution)

- Voici la fonction mathématique décrite :

$$f(\mathbf{x}) = \log(1 + \mathbf{x}^T \mathbf{C} \mathbf{x}) + \mathbf{d}^T \mathbf{x}$$

- Dans cette fonction :
 - \mathbf{x} est un vecteur.
 - \mathbf{C} est une matrice (On considérera tout d'abord le cas où \mathbf{C} n'est pas nécessairement symétrique).
 - \mathbf{d} est un vecteur.
 - \mathbf{x}^T indique la transposée de \mathbf{x} .
 - \mathbf{d}^T indique la transposée de \mathbf{d} .
 - L'expression $\mathbf{x}^T \mathbf{C} \mathbf{x}$ est une forme quadratique.
 - La fonction logarithmique $\log(1 + \mathbf{x}^T \mathbf{C} \mathbf{x})$ opère sur le résultat scalaire de $1 + \mathbf{x}^T \mathbf{C} \mathbf{x}$.
 - Le produit scalaire $\mathbf{d}^T \mathbf{x}$ résulte en un scalaire.
- Cette fonction combine un terme logarithmique impliquant une forme quadratique et un terme linéaire.

Problème 1 : Question 1 (Solution)

- **Expression de la Fonction** : $f(\mathbf{x}) = \log(1 + \mathbf{x}^\top \mathbf{C}\mathbf{x})$
- **Application de la Règle de la Chaîne** : Pour calculer le gradient de $f(\mathbf{x})$, la règle de la chaîne est utilisée pour la dérivation du terme logarithmique par rapport à une fonction quadratique interne. La règle de la chaîne stipule que la dérivée d'une fonction composée, $\log(g(\mathbf{x}))$, est donnée par:

$$\frac{\partial}{\partial \mathbf{x}} \log(g(\mathbf{x})) = \frac{1}{g(\mathbf{x})} \cdot \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}}$$

où $g(\mathbf{x}) = 1 + \mathbf{x}^\top \mathbf{C}\mathbf{x}$ dans notre cas.

- **Dérivée de $\mathbf{x}^\top \mathbf{C}\mathbf{x}$ par rapport à \mathbf{x}** : La dérivée de $\mathbf{x}^\top \mathbf{C}\mathbf{x}$ par rapport à \mathbf{x} est $(\mathbf{C} + \mathbf{C}^\top)\mathbf{x}$. Ce résultat utilise la propriété de la dérivée des formes quadratiques:

$$\frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^\top \mathbf{A}\mathbf{x}) = (\mathbf{A} + \mathbf{A}^\top)\mathbf{x}$$

où \mathbf{A} est une matrice, et \mathbf{A}^\top représente la transposée de \mathbf{A} .

Problème 1 : Question 1 (Solution)

- **Gradient de la fonction logarithmique** : En appliquant la règle de la chaîne:

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \frac{1}{1 + \mathbf{x}^T \mathbf{C} \mathbf{x}} (\mathbf{C} + \mathbf{C}^T) \mathbf{x}$$

Ce calcul intègre la contribution des termes linéaires et non-linéaires de la fonction original.

- **Calcul de la Hessienne** : Pour trouver la Hessienne, nous différencions $\nabla_{\mathbf{x}} f(\mathbf{x})$ par rapport à \mathbf{x} encore une fois, en appliquant la règle du produit :
 - Le gradient $\nabla_{\mathbf{x}}(1 + \mathbf{x}^T \mathbf{C} \mathbf{x})$ est $(\mathbf{C} + \mathbf{C}^T) \mathbf{x}$.
 - La dérivée de ce gradient par rapport à \mathbf{x} implique l'utilisation de la règle du produit et du quotient. Le terme $(\mathbf{C} + \mathbf{C}^T) \mathbf{x}$ donne $\mathbf{C} + \mathbf{C}^T$ (transformation linéaire).
 - Dans le cas où \mathbf{C} est symétrique, $\mathbf{C} = \mathbf{C}^T \Rightarrow (\mathbf{C} + \mathbf{C}^T) \mathbf{x} = 2\mathbf{C} \mathbf{x}$.

Problème 1 : Question 2

- **Fonction** : $g(\mathbf{x}) = \log(1 + \mathbf{x}^\top \mathbf{C}\mathbf{x}) + \mathbf{d}^\top \mathbf{x}$, où \mathbf{C} symétrique.
- **Objectif** : Trouver $\nabla_{\mathbf{x}}^2 g(\mathbf{x})$
- **Solution** :
- **Justification Détaillée** :

Problème 1 : Question 2 (Solution)

- Fonction considérée et son gradient:

$$f(\mathbf{x}) = \log(1 + \mathbf{x}^\top \mathbf{C}\mathbf{x}) \quad \text{et} \quad \nabla_{\mathbf{x}} f(\mathbf{x}) = \frac{(\mathbf{C} + \mathbf{C}^\top)\mathbf{x}}{1 + \mathbf{x}^\top \mathbf{C}\mathbf{x}}.$$

- Identification des terms : Soient $u(\mathbf{x})$ et $v(\mathbf{x})$ deux fonctions différentiables d'un vecteur \mathbf{x} .

- $u(\mathbf{x}) = (\mathbf{C} + \mathbf{C}^\top)\mathbf{x}$.
- $v(\mathbf{x}) = 1 + \mathbf{x}^\top \mathbf{C}\mathbf{x}$.

- La règle du quotient s'écrit :

$$\nabla_{\mathbf{x}} \left(\frac{u(\mathbf{x})}{v(\mathbf{x})} \right) = \frac{v(\mathbf{x})\nabla_{\mathbf{x}}u(\mathbf{x}) - u(\mathbf{x}) \otimes \nabla_{\mathbf{x}}v(\mathbf{x})}{(v(\mathbf{x}))^2}$$

- Définition des termes :

- $u(\mathbf{x}) \in \mathbb{R}^n$ est le numérateur.
- $v(\mathbf{x}) \in \mathbb{R}$ est le dénominateur.
- $\nabla_{\mathbf{x}}u(\mathbf{x}) \in \mathbb{R}^n$ est le gradient du numérateur.
- $\nabla_{\mathbf{x}}v(\mathbf{x}) \in \mathbb{R}^n$ est le gradient du dénominateur.
- \otimes est le produit extérieur (outer product) : $\mathbf{x} \otimes \mathbf{x} = \mathbf{x}\mathbf{x}^\top \in \mathbb{R}^{n \times n}$.

Problème 1 : Question 2 (Solution)

- **Produit Extérieur $\mathbf{x} \otimes \mathbf{x}$:**

- Cette opération prend un vecteur \mathbf{x} de dimension n et produit une matrice $n \times n$.
- Chaque élément (i, j) de cette matrice est donné par $x_i x_j$, représentant le produit des éléments i et j du vecteur \mathbf{x} .

- Si \mathbf{x} est le vecteur $[x_1, x_2, x_3]$, alors le produit extérieur $\mathbf{x} \otimes \mathbf{x}$ sera :

$$\mathbf{x} \otimes \mathbf{x} = \mathbf{xx}^T \begin{bmatrix} x_1 \cdot x_1 & x_1 \cdot x_2 & x_1 \cdot x_3 \\ x_2 \cdot x_1 & x_2 \cdot x_2 & x_2 \cdot x_3 \\ x_3 \cdot x_1 & x_3 \cdot x_2 & x_3 \cdot x_3 \end{bmatrix}$$

- Dans le contexte de la Hessienne, nous utilisons cette opération pour former une matrice $n \times n$ où chaque terme est le produit des dérivées appropriées.

Problème 1 : Question 2 (Solution)

- En appliquant la règle du quotient :

$$\nabla_{\mathbf{x}} \left(\frac{u(\mathbf{x})}{v(\mathbf{x})} \right) = \frac{v(\mathbf{x})\nabla_{\mathbf{x}}u(\mathbf{x}) - u(\mathbf{x}) \otimes \nabla_{\mathbf{x}}v(\mathbf{x})}{(v(\mathbf{x}))^2}$$

$$\Rightarrow \mathbf{H} = \frac{(1 + \mathbf{x}^T \mathbf{C} \mathbf{x})(\mathbf{C} + \mathbf{C}^T) - (\mathbf{C} + \mathbf{C}^T)\mathbf{x} \otimes (\mathbf{C} + \mathbf{C}^T)\mathbf{x}}{(1 + \mathbf{x}^T \mathbf{C} \mathbf{x})^2}.$$

- Le produit extérieur \otimes est utilisé ici pour générer une matrice carrée $n \times n$.
- Chaque terme correspond au produit croisé des composantes des vecteurs (formant des produits extérieurs entre les vecteurs).
- Il permet d'inclure toutes les dérivées partielles secondes dans la Hessienne :

$$(\mathbf{C} + \mathbf{C}^T)\mathbf{x} \otimes (\mathbf{C} + \mathbf{C}^T)\mathbf{x} = (\mathbf{C} + \mathbf{C}^T)\mathbf{x}\mathbf{x}^T(\mathbf{C}^T + \mathbf{C})$$

- Dans le cas où \mathbf{C} est symétrique, cela donne $2\mathbf{C}\mathbf{x}\mathbf{x}^T 2\mathbf{C} = 4\mathbf{C}\mathbf{x}\mathbf{x}^T \mathbf{C}$.

Problème 1 : Question 2 (Solution)

- Expression finale de la Hessienne :

$$\mathbf{H} = \frac{(\mathbf{C} + \mathbf{C}^\top)}{1 + \mathbf{x}^\top \mathbf{C} \mathbf{x}} - \frac{(\mathbf{C} + \mathbf{C}^\top) \mathbf{x} \otimes (\mathbf{C} + \mathbf{C}^\top) \mathbf{x}}{(1 + \mathbf{x}^\top \mathbf{C} \mathbf{x})^2}.$$

- Si \mathbf{C} symétrique, c'est à dire, $\mathbf{C} = \mathbf{C}^\top$:

$$\mathbf{H} = \frac{2\mathbf{C}}{1 + \mathbf{x}^\top \mathbf{C} \mathbf{x}} - \frac{4\mathbf{C} \mathbf{x} \mathbf{x}^\top \mathbf{C}}{(1 + \mathbf{x}^\top \mathbf{C} \mathbf{x})^2}.$$

Problème 1 : Question 3

- **Fonction :** $h(\mathbf{x}) = e^{-\mathbf{d}^\top \mathbf{x}} + \mathbf{x}^\top \mathbf{B}\mathbf{x}$
- **Objectif :** Trouver $\nabla_{\mathbf{x}} h(\mathbf{x})$
- **Solution :**
- **Justification détaillée:**

Problème 1 : Question 3 (Solution)

- Voici la fonction mathématique décrite :

$$h(\mathbf{x}) = e^{-\mathbf{d}^T \mathbf{x}} + \mathbf{x}^T \mathbf{B} \mathbf{x}$$

- Dans cette fonction :
 - \mathbf{x} est un vecteur.
 - \mathbf{B} est une matrice.
 - \mathbf{d} est un vecteur.
 - \mathbf{x}^T indique la transposée de \mathbf{x} .
 - $\mathbf{B}\mathbf{x}$ est une transformation linéaire de \mathbf{x} .
 - $\mathbf{x}^T \mathbf{B} \mathbf{x}$ est une forme quadratique.
 - L'exponentielle $e^{-\mathbf{d}^T \mathbf{x}}$ affecte le terme linéaire.
- Cette fonction combine un terme exponentiel et un terme quadratique.

Problème 1 : Question 3 (Solution)

- Considérons une fonction composée :

$$f(g(\mathbf{x}))$$

où :

- $f : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction scalaire.
- $g : \mathbb{R}^n \rightarrow \mathbb{R}$ est une fonction scalaire prenant un vecteur en entrée.
- $\mathbf{x} \in \mathbb{R}^n$ est un vecteur.
- La dérivée en chaîne s'écrit :

$$\nabla_{\mathbf{x}} (f(g(\mathbf{x}))) = \frac{df(g(\mathbf{x}))}{dg(\mathbf{x})} \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} = f'(g(\mathbf{x})) \cdot \nabla_{\mathbf{x}} g(\mathbf{x})$$

où :

- $f'(g(\mathbf{x})) = \frac{d}{dg} f(g) \in \mathbb{R}$ est la dérivée de f par rapport à son argument scalaire g . Le résultat est un scalaire.
- $\nabla_{\mathbf{x}} g(\mathbf{x}) \in \mathbb{R}^n$ est le gradient de g par rapport à \mathbf{x} .
- Cas particulier : Si $f(g(\mathbf{x})) = e^{g(\mathbf{x})}$, alors $f'(g) = e^g$, d'où :

$$\nabla_{\mathbf{x}} e^{g(\mathbf{x})} = e^{g(\mathbf{x})} \cdot \nabla_{\mathbf{x}} g(\mathbf{x}).$$

Problème 1 : Question 3 (Solution)

- **Expression de la Fonction** : $h(\mathbf{x}) = e^{-\mathbf{d}^\top \mathbf{x}} + \mathbf{x}^\top \mathbf{B} \mathbf{x}$
- **Application de la Règle de la Chaîne** : Pour calculer le gradient de $h(\mathbf{x})$, nous appliquons la règle de la chaîne sur le terme exponentiel.

$$\frac{\partial}{\partial \mathbf{x}} e^{-\mathbf{d}^\top \mathbf{x}} = -e^{-\mathbf{d}^\top \mathbf{x}} \mathbf{d}$$

- **Dérivée de $\mathbf{x}^\top \mathbf{B} \mathbf{x}$ par rapport à \mathbf{x}** : La dérivée de $\mathbf{x}^\top \mathbf{B} \mathbf{x}$ par rapport à \mathbf{x} est $(\mathbf{B} + \mathbf{B}^\top) \mathbf{x}$, en utilisant la règle :

$$\frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^\top \mathbf{A} \mathbf{x}) = (\mathbf{A} + \mathbf{A}^\top) \mathbf{x}$$

où \mathbf{A} est une matrice.

Problème 1 : Question 3 (Solution)

- **Expression de la Fonction :**

$$h(\mathbf{x}) = e^{-\mathbf{d}^\top \mathbf{x}} + \mathbf{x}^\top \mathbf{B} \mathbf{x}$$

- **Application de la Règle de la Chaîne :** Nous considérons le terme exponentiel séparément. Notons $g(\mathbf{x}) = -\mathbf{d}^\top \mathbf{x}$, alors :

$$\nabla_{\mathbf{x}} e^{g(\mathbf{x})} = e^{g(\mathbf{x})} \nabla_{\mathbf{x}} g(\mathbf{x}).$$

- **Calcul du gradient de $g(\mathbf{x})$:**

- Nous avons $g(\mathbf{x}) = -\mathbf{d}^\top \mathbf{x}$.
- La dérivée de $\mathbf{d}^\top \mathbf{x}$ par rapport à \mathbf{x} est simplement \mathbf{d} car :

$$\frac{\partial}{\partial \mathbf{x}} (\mathbf{d}^\top \mathbf{x}) = \mathbf{d}.$$

- Ceci est dû au fait que $\mathbf{d}^\top \mathbf{x}$ est un produit scalaire et suit la règle générale : $\frac{\partial}{\partial \mathbf{x}} (\mathbf{a}^\top \mathbf{x}) = \frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^\top \mathbf{a}) = \mathbf{a}$.

Problème 1 : Question 3 (Solution)

- **Pourquoi nous obtenons \mathbf{d} et non \mathbf{d}^\top :**

- Nous voulons que $\nabla_{\mathbf{x}}h(\mathbf{x})$ soit un vecteur colonne $\in \mathbb{R}^n$.
- Si nous prenions \mathbf{d}^\top , nous obtiendrions un vecteur ligne, ce qui n'est pas conforme à la convention de notation pour les gradients.
- Par conséquent, nous utilisons directement \mathbf{d} comme la dérivée du produit scalaire, garantissant ainsi que le gradient est bien un vecteur colonne.

- **Gradient final du terme exponentiel :**

$$\nabla_{\mathbf{x}}e^{-\mathbf{d}^\top\mathbf{x}} = -\mathbf{d}e^{-\mathbf{d}^\top\mathbf{x}} = -e^{-\mathbf{d}^\top\mathbf{x}}\mathbf{d} \quad (\text{car } -e^{-\mathbf{d}^\top\mathbf{x}} \text{ est un scalaire}).$$

- **Gradient de la fonction :**

$$\nabla_{\mathbf{x}}h(\mathbf{x}) = -e^{-\mathbf{d}^\top\mathbf{x}}\mathbf{d} + (\mathbf{B} + \mathbf{B}^\top)\mathbf{x}$$

- Si \mathbf{B} est symétrique, càd, $\mathbf{B} = \mathbf{B}^\top$, alors :

$$\nabla_{\mathbf{x}}h(\mathbf{x}) = -e^{-\mathbf{d}^\top\mathbf{x}}\mathbf{d} + 2\mathbf{B}\mathbf{x}$$

Problème 1 : Question 4

- **Fonction** : Identique à la question 3.
- **Objectif** : Trouver $\nabla_{\mathbf{x}}^2 h(\mathbf{x})$
- **Solution** :
- **Justification détaillée**:

Problème 1 : Question 4 (Solution)

- **Calcul de la Hessienne** : Nous différencions $\nabla_{\mathbf{x}}h(\mathbf{x})$ par rapport à \mathbf{x} , en appliquant la règle de la dérivée en chaîne:
 - La dérivée du terme exponentiel :

$$\nabla_{\mathbf{x}}(-e^{-\mathbf{d}^T \mathbf{x}} \mathbf{d}) = e^{-\mathbf{d}^T \mathbf{x}} \mathbf{d} \otimes \mathbf{d}$$

- La dérivée du terme quadratique :

$$\nabla_{\mathbf{x}}((\mathbf{B} + \mathbf{B}^T)\mathbf{x}) = \mathbf{B} + \mathbf{B}^T$$

Problème 1 : Question 4 (Solution)

- **Produit Extérieur $\mathbf{x} \otimes \mathbf{x}$:**
 - Cette opération prend un vecteur \mathbf{x} de dimension n et produit une matrice $n \times n$.
 - Chaque élément (i, j) de cette matrice est donné par $x_i x_j$, représentant le produit des éléments i et j du vecteur \mathbf{x} .
- La Hessienne inclut un produit extérieur pour capturer tous les termes mixtes des dérivées :

$$e^{-\mathbf{d}^\top \mathbf{x}} \mathbf{d} \otimes \mathbf{d} = e^{-\mathbf{d}^\top \mathbf{x}} \mathbf{d} \mathbf{d}^\top = \mathbf{d} \mathbf{d}^\top e^{-\mathbf{d}^\top \mathbf{x}} \quad (\text{puisque } e^{-\mathbf{d}^\top \mathbf{x}} \text{ est un scalaire}).$$

- **Expression finale :**

$$\nabla_{\mathbf{x}}^2 h(\mathbf{x}) = e^{-\mathbf{d}^\top \mathbf{x}} \mathbf{d} \mathbf{d}^\top + (\mathbf{B} + \mathbf{B}^\top)$$

- **Cas particulier :** Si \mathbf{B} est symétrique ($\mathbf{B} = \mathbf{B}^\top$), alors :

$$\nabla_{\mathbf{x}}^2 h(\mathbf{x}) = e^{-\mathbf{d}^\top \mathbf{x}} \mathbf{d} \mathbf{d}^\top + 2\mathbf{B}$$