Cours OPT 201

Optimisation Différentiable – Théorie et Algorithmes

Exercices de la séance 1
(calcul différentiel, analyse convexe, existence et unicité de solution)

Calcul différentiel

1. Gradient et hessien
2. Autres exemples de dérivation
3. Direction du gradient
4. Méthode de l'état adjoint

Analyse convexe

5. Courbes de niveau d'une fonction convexe
6. Les fonctions det et ld
7. Distance à un ensemble convexe, fonction marginale

Existence et unicité de solution

8. Fonction semi-continue inférieurement
9. Existence et unicité de solution d'un problème quadratique
10. Existence et unicité de solution d'un problème avec contraintes
11. Problèmes d'optimisation sans solution

1 Gradient et hessien

Soit A une matrice d'ordre n et b un vecteur de $\mathbb{R}^n$. On considère la fonction *quadratique* :

$$f(x) = \frac{1}{2} x^\mathsf{T} A x - b^\mathsf{T} x.$$

1. Calculer le gradient et le hessien de f pour le produit scalaire euclidien.

2. Que deviennent le gradient et le hessien pour un produit scalaire quelconque.

3. On suppose que A est symétrique. Montrez que[1]

$$\begin{aligned} f \text{ est convexe} \quad &\Longleftrightarrow \quad A \succcurlyeq 0, \\ f \text{ est strictement convexe} \quad &\Longleftrightarrow \quad A \succ 0. \end{aligned}$$

On suppose à présent que $f : \mathbb{R}^n \to \mathbb{R}$ est une fonction non linéaire quelconque, mais régulière. Soit $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ un difféomorphisme tel que $\varphi'(\tilde{x})$ soit inversible pour tout $\tilde{x} \in \mathbb{R}^n$. On considère le changement de variables

$$\tilde{x} = \varphi^{-1}(x)$$

qui transforme f en une fonction $\tilde{f} : \mathbb{R}^n \to \mathbb{R}$ définie par

$$\tilde{f}(\tilde{x}) = f(\varphi(\tilde{x})) = (f \circ \varphi)(\tilde{x}),$$

si bien que $\tilde{f} = f \circ \varphi$ et $\tilde{f}(\tilde{x}) = f(x)$ pour tout $x \in \mathbb{R}^n$ pourvu que $\tilde{x} = \varphi^{-1}(x)$.

4. Exprimez le gradient et le hessien de $\tilde{f}$ par rapport au gradient et au hessien de f.

5. Montrez que f et $\tilde{f}$ ont leurs points stationnaires et minima locaux et globaux qui se correspondent.

2 Autres exemples de dérivation

1. Soient $\mathbb{E}_1$, $\mathbb{E}_2$ et $\mathbb{F}$ des espaces normés et $b : \mathbb{E}_1 \times \mathbb{E}_2 \to \mathbb{F}$ une application bilinéaire continue. Montrez que b est différentiable en tout point (x_1, x_2) de $\mathbb{E}_1 \times \mathbb{E}_2$ et que pour tout (h_1, h_2) et $(k_1, k_2) \in \mathbb{E}_1 \times \mathbb{E}_2$, on a

$$\begin{aligned} b'(x_1, x_2) \cdot (h_1, h_2) &= b(h_1, x_2) + b(x_1, h_2), \\ b''(x_1, x_2) \cdot ((h_1, h_2), (k_1, k_2)) &= b(h_1, k_2) + b(k_1, h_2). \end{aligned}$$

2. On note $\mathrm{GL}_n(\mathbb{R})$ l'ensemble des matrices d'ordre n inversibles (ou *groupe général linéaire de degré* n), qui est donc un *ouvert* de l'ensemble des matrices réelles d'ordre n. Calculez la dérivée première et seconde de la fonction

$$i : M \in \mathrm{GL}_n(\mathbb{R}) \to i(M) = M^{-1} \in \mathrm{GL}_n(\mathbb{R}).$$

[1] L'écriture « $A \succcurlyeq 0$ » signifie que A est symétrique semi-définie positive et l'écriture « $A \succ 0$ » signifie que A est symétrique définie positive.

3. Soit $\mathcal{S}^n$ l'ensemble des matrices d'ordre n symétriques, muni du produit scalaire euclidien $(A, B) \in \mathcal{S}^n \times \mathcal{S}^n \to \langle A, B\rangle = \operatorname{tr}(AB) \in \mathbb{R}$ (la trace du produit AB, c'est-à-dire la somme des éléments diagonaux de AB), dont la norme associée est notée (elle porte le nom de *norme de Frobenius*) :

$$\|\cdot\|_F = \langle\cdot,\cdot\rangle^{1/2}.$$

Soit $M \in \mathcal{S}^n$. On considère l'application

$$\varphi : x \in \mathbb{R}^n \mapsto \frac{1}{2}\|M - xx^\mathsf{T}\|_F^2.$$

Calculez le gradient et le hessien de φ en $x \in \mathbb{R}^n$ pour le produit scalaire euclidien.

3 Direction du gradient

Soient $\langle\cdot,\cdot\rangle$ un produit scalaire sur $\mathbb{R}^n$ et f une application d'un ouvert ω de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$, de classe C^1. On suppose qu'en $x_0 \in \omega$, le gradient $\nabla f(x_0)$ de f est non nul. Montrez que

1. $\nabla f(x_0)$ est orthogonal au plan tangent en x_0 à l'ensemble de niveau de f passant par x_0, défini par $\{x \in \omega : f(x) = f(x_0)\}$;
2. $\nabla f(x_0)$ est dirigé vers les valeurs croissantes de f.

4 Méthode de l'état adjoint

On se place dans la situation suivante. Soit $J : \mathbb{R}^n \to \mathbb{R} : x \mapsto J(x)$, une fonction. On suppose que la variable x est partitionnée en $x = (y, u)$, où $y \in \mathbb{R}^m$, $1 \leqslant m < n$, est appelée *variable d'état* et $u \in \mathbb{R}^{n-m}$ est appelée *variable de commande*. Ces variables sont liées par l'équation non linéaire

$$F(y, u) = 0,$$

dite *équation d'état* du problème. On suppose que $F'_y(y_0, u_0)$ (dérivée de F par rapport à y en (y_0, u_0)) est inversible, si bien qu'il existe une fonction implicite $u \mapsto y(u)$ telle que On cherche à calculer, de manière efficace, le gradient de la fonction f définie par

$$f(u) = J(y(u), u).$$

1. **Méthode directe**. Montrez que

$$\nabla f(u) = y'(u)^\mathsf{T}\nabla_y J(y, u) + \nabla_u J(y, u),$$

où $y'(u)$ est solution de

$$F'_y(y, u)y'(u) = -F'_u(y, u). \tag{1.1}$$

2. **Méthode de l'état adjoint**. Montrez que si l'on définit $p \in \mathbb{R}^m$ comme solution de l'équation suivante (dite *équation adjointe*)

$$F'_y(y, u)^\mathsf{T} p = -\nabla_y J(y, u), \tag{1.2}$$

on a

$$\nabla f(u) = \nabla_u J(y, u) + F'_u(y, u)^\mathsf{T} p.$$

Remarque. La méthode directe est beaucoup plus coûteuse en temps de calcul. Elle demande en effet d'évaluer $y'(u)$, matrice de dimension $m \times (n-m)$, donnant la dérivée de la fonction implicite. Il faut donc résoudre les $n-m$ systèmes linéaires de l'équation (1.1). Dans la méthode de l'état adjoint, il ne faut résoudre qu'un seul système linéaire, celui de l'équation adjointe (1.2).

5 Courbes de niveau d'une fonction convexe

Chaque dessin A, B et C de la figure 1 représente 3 courbes de niveau (iso-valeurs) d'une

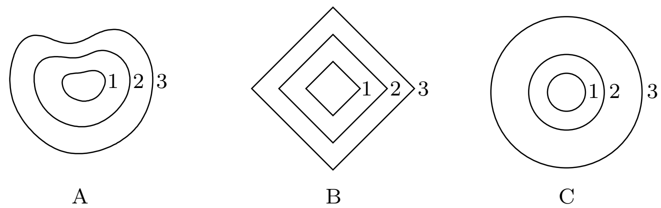


Fig. 1: Courbes de niveau d'une fonction convexe ?

fonction f définie sur $\mathbb{R}^2$ à valeurs dans $\mathbb{R}$. De manière plus précise, la courbe étiquetée « i » ($i = 1, 2, 3$) représente l'ensemble $\{x \in \mathbb{R}^2 : f(x) = i\}$. Ces fonctions *peuvent*-elles être convexes ? Justifiez vos réponses.

6 Les fonctions det et ld

Soit $\mathbb{R}^{n\times n}$ l'espace vectoriel des matrices d'ordre n. On munit $\mathbb{R}^{n\times n}$ du produit scalaire $\langle A, B\rangle = \operatorname{tr} A^{\mathsf{T}} B$ (la trace du produit matriciel $A^{\mathsf{T}} B$), qui en fait un espace euclidien. On s'intéresse à l'application *déterminant* $\det : \mathbb{R}^{n\times n} \to \mathbb{R}$.

1. Montrez que $\det'(I) \cdot H = \operatorname{tr} H$.

2. Montrez que, si A est inversible :

$$\nabla \det(A) = (\det A) A^{-\mathsf{T}} = \operatorname{cof}(A),$$

   où $\operatorname{cof}(A)$ est la matrice des cofacteurs de A (son élément (i, j) est $(-1)^{i+j}$ fois le déterminant de $A_{(i)}^{(j)}$, la matrice A dont on a ôté la ligne i et la colonne j).

Soit $\mathcal{S}^n$ l'espace vectoriel des matrices d'ordre n symétriques. On munit $\mathcal{S}^n$ du produit scalaire $\langle A, B\rangle = \operatorname{tr} AB$ (la trace du produit AB), qui en fait un espace euclidien. On note $\mathcal{S}^n_{++}$ le cône des matrices symétriques définies positives. On considère l'application *log-déterminant* $\operatorname{ld} : \mathcal{S}^n \to \mathbb{R} \cup \{+\infty\}$, définie en $A \in \mathcal{S}^n$ par

$$\operatorname{ld}(A) = \begin{cases} -\log\det A & \text{si } A \in \mathcal{S}^n_{++} \\ +\infty & \text{sinon} \end{cases}$$

3. Montrez que $\operatorname{ld} \in \overline{\operatorname{Conv}}(\mathcal{S}^n)$ (elle est propre convexe et fermée).

4. Montrez que ld est strictement convexe sur $\mathcal{S}^n_{++}$.

7 Distance à un ensemble convexe, fonction marginale

Soit C un ensemble convexe d'un espace vectoriel normé $\mathbb{E}$ (on note $\|\cdot\|$ la norme). La fonction distance $d_C(x)$ d'un point $x \in \mathbb{E}$ à l'ensemble C est définie par

$$d_C(x) = \inf_{y \in C} \|x - y\|.$$

Montrez que

1. d_C est convexe,
2. d_C est non expansive (= contractante, mais non strictement), i.e.,

$$\forall\, x_0,\, x_1 \in \mathbb{E} : \quad |d_C(x_0) - d_C(x_1)| \leqslant \|x_0 - x_1\|.$$

La manière avec laquelle la fonction $d_C(\cdot)$ a été construite est un cas particulier du procédé général que voici. Soient $\mathbb{E}$ et $\mathbb{F}$ deux espaces vectoriels et $\varphi : \mathbb{E} \times \mathbb{F} \to \overline{\mathbb{R}}$ une fonction. La *fonction marginale* de φ est la fonction $f : \mathbb{E} \to \overline{\mathbb{R}}$ définie par :

$$f(x) = \inf_{y \in \mathbb{F}} \varphi(x, y).$$

3. Montrez que φ est convexe $\Longrightarrow$ f est convexe.
4. Retrouvez le résultat du point 1 à partir du point 3.

8 Fonction semi-continue inférieurement

Soit $\mathbb{E}$ un espace topologique. Montrez que les trois propriétés suivantes sont équivalentes.

(i) La fonction $f : \mathbb{E} \to \overline{\mathbb{R}}$ est *semi-continue inférieurement* (s.c.i. en abrégé), c'est-à-dire que pour tout $x \in \mathbb{E}$ et pour toute suite $\{x_k\}_{k \geqslant 1}$ convergeant vers x, on a

$$f(x) \leqslant \liminf_{k \to \infty} f(x_k).$$

(ii) La fonction f est *fermée*, c'est-à-dire que son épigraphe est fermé.

(iii) Les ensembles de sous-niveau de f sont fermés : pour tout $\nu \in \mathbb{R}$,

$$N_\nu(f) := \{x \in \mathbb{E} : f(x) \leqslant \nu\} \text{ est fermé.}$$

9 Existence et unicité de solution d'un problème quadratique

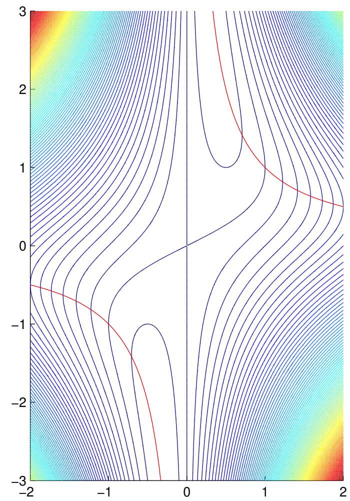
Soient $A \in \mathcal{S}^n$ et $b \in \mathbb{R}^n$. On considère le problème $\min\{f(x) : x \in \mathbb{R}^n\}$, où

$$f(x) = \frac{1}{2} x^\mathsf{T} A x + b^\mathsf{T} x.$$

1. Montrez que les propriétés suivantes sont équivalentes
   (a) le problème a une solution,
   (b) f est bornée inférieurement,
   (c) $A \succcurlyeq 0$ et $b \in \mathcal{R}(A)$,

   et que dans ce cas,
   – l'ensemble des solutions est de la forme $x_0 + \mathcal{N}(A)$, où x_0 est une solution particulière du problème, et
   – la valeur optimale vaut $-\frac{1}{2} b^\mathsf{T} A^\dagger b$, où $A^\dagger$ est le pseudo-inverse de A.
2. Montrez que le problème a une solution unique si, et seulement si, $A \succ 0$.

Remarque. L'implication $(c) \Rightarrow (a)$ du point 1 ne peut s'étendre à des polynômes de degré strictement supérieur à 2. Par exemple, le *polynôme quartique* $x \in \mathbb{R}^2 \mapsto x_1^2 + (1 - x_1 x_2)^2$ est strictement positif sur $\mathbb{R}^2$ (ses courbes de niveau sont données dans la figure ci-jointe). Il n'atteint donc pas sa borne inférieure qui est nulle (on l'approche en prenant $x_1 \to 0$, $x_1 \neq 0$ et $x_2 = 1/x_1$, le long des courbes rouges dans la figure).



Rappels. 1) Les propriétés suivantes sont équivalentes :

(i) A est semi-définie positive (i.e., $x^\mathsf{T} A x \geqslant 0$, pour tout $x \in \mathbb{R}^n$),

(ii) toutes les valeurs propres de A sont positives,

(iii) tous les mineurs principaux de A (les $\det A_{I,I}$, où $I \subset \{1, \ldots, n\}$) sont positifs.

2) De manière similaire, les propriétés suivantes sont équivalentes :

(i) A est définie positive (i.e., $x^\mathsf{T} A x > 0$, pour tout $x \in \mathbb{R}^n \setminus \{0\}$),

(ii) toutes les valeurs propres de A sont strictement positives,

(iii) tous les mineurs principaux de tête de A (les $\det A_{I,I}$, où $I = \{1, \ldots, j\}$, avec $j = 1, \ldots, n$) sont strictement positifs.

3) Le pseudo-inverse d'une matrice A de type $m \times n$ est l'unique matrice $A^\dagger$ de type $n \times m$ qui vérifie

$$AA^\dagger A = A, \quad A^\dagger A A^\dagger = A^\dagger, \quad (AA^\dagger)^\mathsf{T} = AA^\dagger \quad \text{et} \quad (A^\dagger A)^\mathsf{T} = A^\dagger A. \tag{1.3}$$

Cette notion sera revue lors l'étude des problèmes de moindres-carrés linéaires.

10 Existence et unicité de solution d'un problème avec contraintes

On considère le problème dans $\mathbb{R}^2$

$$\begin{cases} \min x_1^2 + x_2^2 - \dfrac{9}{4} x_1 x_2 \\ x_1 \geqslant 2x_2 \geqslant 0. \end{cases}$$

Montrez que ce problème admet une solution et une seule.

11 Problèmes d'optimisation sans solution

On considère les situations suivantes.

1. Montrez que le problème sur $\mathbb{R}$,
$$\min e^x,$$
n'a pas de solution.

2. Soient $A_0 \in \mathbb{R}^{m \times n}$ une matrice et $b_0 \in \mathbb{R}^m$ un vecteur. On cherche $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ et $x \in \mathbb{R}^n$ solution du problème d'optimisation (appelé « *Problème de moindres carrés total* »)
$$\begin{cases} \min \ \dfrac{1}{2}\|A - A_0\|^2 + \dfrac{1}{2}\|b - b_0\|^2 \\ Ax = b. \end{cases}$$

On a noté de la même manière, par $\|\cdot\|$, des normes matricielle et vectorielle quelconques.

Montrez que ce problème n'a pas nécessairement une solution.

Qu'en est-il si on ajoute à la fonction-coût le terme $\frac{1}{2}\|x\|^2$?

3. Soit $F : \mathbb{R}^n \to \mathbb{R}^m$, $X \subset \mathbb{R}^n$ et $\|\cdot\|$ une norme sur $\mathbb{R}^m$. On considère le problème

$$\min\{\|F(x)\| : x \in X\}.$$

On suppose que $0 \notin F(X)$ et que $F(X)$ est un ouvert. Montrez que :

(a) ce problème n'a pas de solution,

(b) si F est continue et $\{x_k\}$ est une suite minimisante,

- soit $\|x_k\| \to \infty$,
- soit $\liminf d_{X^c}(x_k) = 0$.

On a noté $d_{X^c}(x_k)$ la distance de x_k au complémentaire de X, qui est supposée infinie si $X = \mathbb{R}^n$.