Cours OPT 202

Optimisation Différentiable – Théorie et Algorithmes

Exercices de la séance 9
(conjugaison et sous-différentiabilité)

Ensembles convexes

1. Enveloppe convexe
2. Face exposée d'un convexe

Calcul de conjuguée

3. Fonction conjuguée d'une forme quadratique

Le sous-différentiel en dessin

4. Représentations du sous-différentiel.

Calcul de sous-différentiel (par la conjuguée)

5. Sous-différentiel d'une norme
6. Sous-différentiel de la valeur propre maximale
7. Sous-différentiel de la distance à un ensemble
8. Sous-différentiel de la fonction duale d'un problème quadratique.

Modélisation

9. Mécaniques lagrangienne et hamiltonienne sur un espace vectoriel

1 Enveloppe convexe

Soient $\mathbb{E}$ un espace vectoriel et $P \subset \mathbb{E}$. On rappelle que l'*enveloppe convexe* de P est l'intersection de tous les convexes contenant P; c'est donc le plus petit convexe contenant P (l'intersection de convexes est convexe). On la note

$$\operatorname{co} P := \bigcap \{C : C \text{ est un convexe contenant } P\}.$$

On appelle *combinaison convexe* de $\mathbb{E}$, un élément x de $\mathbb{E}$ de la forme

$$x = \sum_{i=1}^{m} t_i x_i,$$

où

$$m \in \mathbb{N}^*, \quad t_i \in \mathbb{R} \quad \text{avec} \quad \sum_{i=1}^{m} t_i = 1 \quad \text{et} \quad t_i \geqslant 0, \quad x_i \in \mathbb{E}.$$

Démontrez les affirmations suivantes.

1. Un ensemble est convexe si, et seulement si, il contient toutes les combinaisons convexes de ses éléments.
2. $\operatorname{co} P$ est l'ensemble des combinaisons convexes des éléments de P :
$$\operatorname{co} P = \left\{ \sum_{i=1}^{m} t_i x_i : m \in \mathbb{N}^*,\ t_i \in \mathbb{R} \text{ avec } \sum_{i=1}^{m} t_i = 1 \text{ et } t_i \geqslant 0,\ x_i \in P \right\}.$$
3. **Théorème de Carathéodory** (1911). Soit P une partie d'un espace vectoriel $\mathbb{E}$ de dimension n. Alors tout élément de $\operatorname{co} P$ peut s'écrire comme une combinaison convexe de $n+1$ éléments de P.

2 Face exposée d'un convexe

Soit C un convexe de $\mathbb{R}^n$. On dit que $F \subset C$ est une **face** de C si F est convexe et si tout segment $[x, y]$ de C dont l'intérieur relatif $]x, y[$ intersecte F est complètement dans F. Cette dernière condition peut aussi s'écrire

$$x, y \in C,\ t \in]0, 1[,\ (1-t)x + ty \in F \quad \Longrightarrow \quad x, y \in F$$

On dit qu'une partie E de C est **exposée** s'il existe $c \in \mathbb{R}^n$ tel que

$$E = \arg\min \{c^{\mathsf{T}} x : x \in C\}.$$

1. Montrez que toute partie exposée d'un convexe est une face, mais que la réciproque est généralement fausse.

Soit $P := \{x \in \mathbb{R}^n : Ax = a,\ Bx \leqslant b\}$ un polyèdre convexe de $\mathbb{R}^n$ (on suppose les dimensions consistantes).

2. Montrez que F est une face de P si, et seulement si, il existe un ensemble d'indices I tel que $F = \{x \in P : (Bx - b)_I = 0\}$.
3. Montrez que toute face d'un polyèdre convexe est exposée.

3 Fonction conjuguée d'une forme quadratique

Soient $\mathbb{E}$ un espace euclidien (produit scalaire $\langle\cdot,\cdot\rangle$), A un opérateur auto-adjoint défini positif, $b \in \mathbb{E}$ et $f : \mathbb{E} \to \mathbb{R}$ la fonction quadratique définie par

$$f(x) = \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle.$$

Montrez les affirmations suivantes.

1. Pour tout $x^* \in \mathbb{E}$,
$$f^*(x^*) = \frac{1}{2}\langle A^{-1}(x^*-b), (x^*-b)\rangle.$$
2. Si A est seulement semi-définie positive, on a $f^*(Ax+b) = \frac{1}{2}\langle Ax, x\rangle$ pour tout $x \in \mathbb{E}$ et $f^*(x^*) = +\infty$ si $x^* \notin \mathcal{R}(A) + b$.

4 Représentations du sous-différentiel

1. **Comportement d'une fonction convexe le long de ses sous-gradients**.

   Soient $\mathbb{E}$ un espace euclidien, $f \in \mathrm{Conv}(\mathbb{E})$ et $x^* \in \partial f(x) \setminus \{0\}$.

   (a) Montrez que f croît strictement en x le long de x^*, dans le sens où
   $$\forall\, t > 0 : \quad f(x + tx^*) > f(x).$$
   (b) Montrez, par un contre-exemple, que f peut croître en x le long de $-x^*$, dans le sens où
   $$\forall\, t > 0 : \quad f(x - tx^*) > f(x).$$
   Par conséquent, la situation est différente de celle du cas differentiable (convexe ou pas d'ailleurs), où l'on a $f(x - t\nabla f(x)) < f(x)$ pour tout $t > 0$ petit.

2. **Reconnaissance d'un sous-différentiel**.

   Soit $f : \mathbb{R}^2 \to \mathbb{R}$ une fonction *convexe*, dont on note $\partial f(x)$ le sous-différentiel en x pour le produit scalaire euclidien. La figure 11 donne quatre dessins : A, B, C et D. Dans

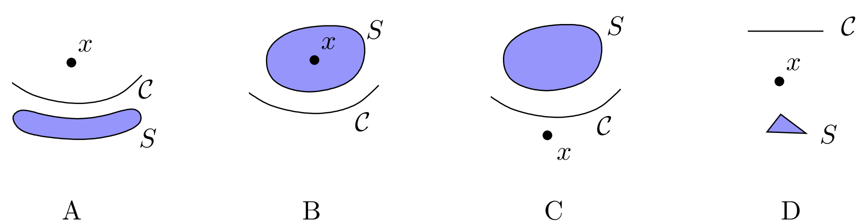

Fig. 11: Représentations du sous-différentiel d'une fonction convexe ?

   chacun d'eux, on a noté x un point de $\mathbb{R}^2$, $\mathcal{C}$ une *partie* d'une courbe de niveau de f (i.e., f est constante sur une courbe contenant $\mathcal{C}$) et S un ensemble. Pour chacun de ces dessins :

   (a) dites si S peut être l'ensemble $x + \partial f(x)$,
   (b) dans l'affirmative en (a) et si $S = x + \partial f(x)$, dites si x est un minimiseur de f,
   (c) dans l'affirmative en (a) et (b), et si $S = x+\partial f(x)$, dites si x est l'*unique* minimiseur de f.

5 Sous-différentiel d'une norme

Soient $f(\cdot) := \|\cdot\|$ une norme sur un espace vectoriel $\mathbb{E}$, $\langle \cdot, \cdot \rangle$ un produit scalaire et $\|\cdot\|_{\mathrm{D}}$ la norme duale de $\|\cdot\|$ par rapport à ce produit scalaire :

$$\|v\|_{\mathrm{D}} := \sup_{\|u\| \leqslant 1} \langle v, u \rangle. \tag{12.1}$$

On note $\bar{B}_{\mathrm{D}} := \{x \in \mathbb{E} : \|x\|_{\mathrm{D}} \leqslant 1\}$, la boule fermée unité pour la norme duale.

1. Montrez que $f \in \overline{\mathrm{Conv}}(\mathbb{E})$ et que f a une minorante affine.

2. Démontrez l'inégalité de Cauchy-Schwarz généralisée

$$\forall u, v \in \mathbb{E} : \quad |\langle u, v \rangle| \leqslant \|u\| \, \|v\|_{\mathrm{D}}.$$

3. Montrez que $f^* = \mathcal{I}_{\bar{B}_{\mathrm{D}}}$, c'est-à-dire

$$f^*(x^*) = \begin{cases} 0 & \text{si } \|x^*\|_{\mathrm{D}} \leqslant 1 \\ +\infty & \text{sinon.} \end{cases}$$

4. Montrez que

$$\partial f(x) = \{x^* \in \mathbb{E} : \|x^*\|_{\mathrm{D}} \leqslant 1 \text{ et } \langle x^*, x \rangle = \|x\|\}. \tag{12.2}$$

   En particulier,

   (a) $\partial f(0) = \bar{B}_{\mathrm{D}}$,

   (b) $\|x^*\|_{\mathrm{D}} = 1$ si $x^* \in \partial f(x)$ et $x \neq 0$.

   (c) Si $\mathbb{E} = \mathbb{R}^n$, $\langle \cdot, \cdot \rangle$ est le produit scalaire euclidien, $1 < p < \infty$ et $x \neq 0$, on a

$$\begin{aligned} \partial(\|\cdot\|_1)(x) = \{x^* \in \mathbb{R}^n : \; & x_i^* = -1 \text{ si } x_i < 0, \\ & x_i^* \in [-1, +1] \text{ si } x_i = 0, \\ & x_i^* = +1 \text{ si } x_i > 0\}, \end{aligned} \tag{12.3}$$

$$\nabla(\|\cdot\|_p)(x) = \{\mathrm{sgn}(x_i)(|x_i|/\|x\|_p)^{p-1}\}_{i \in [1:n]}, \tag{12.4}$$

$$\partial(\|\cdot\|_\infty)(x) = \mathrm{co}\{\mathrm{sgn}(x_i) e^i : i \in I\}, \tag{12.5}$$

   où $I := \{i \in [1:n] : |x_i| = \|x\|_\infty\}$.

5. On définit la *norme biduale* $\|\cdot\|_{\mathrm{DD}}$ comme la norme duale de la norme duale, c'est-à-dire

$$\|x\|_{\mathrm{DD}} := \sup_{\|y\|_{\mathrm{D}} \leqslant 1} \langle x, y \rangle. \tag{12.6}$$

   Montrez que

   (a) $\|\cdot\|_{\mathrm{DD}} = \|\cdot\|$;

   (b) $\forall x \in \mathbb{E}$, $\exists y \in \mathbb{E}$: $\|y\| = 1$ et $\langle y, x \rangle = \|x\|_{\mathrm{D}}$;
   $\forall x \in \mathbb{E}$, $\exists y \in \mathbb{E}$: $\|y\|_{\mathrm{D}} = 1$ et $\langle y, x \rangle = \|x\|$;

   (c) $\partial f(x) \neq \varnothing$ et $\partial f(x) = \arg\max_{\|y\|_{\mathrm{D}} \leqslant 1} \langle y, x \rangle$.

6 Sous-différentiel de la valeur propre maximale

On note $\mathcal{S}^n$ [resp. $\mathcal{S}^n_+$] l'ensemble des matrices d'ordre n symétriques [resp. symétriques semi-définies positives], que l'on munit du produit scalaire $\langle A, B\rangle = \operatorname{tr} AB$. On note $\lambda_{\max} : \mathcal{S}^n \to \mathbb{R}$ l'application valeur propre maximale.

1. Montrez que l'application $\lambda_{\max} : \mathcal{S}^n \to \mathbb{R}$ est convexe, fermée, propre et a une minorante affine.

2. Montrez que
$$\lambda^*_{\max}(A^*) = \begin{cases} 0 & \text{si } A^* \in \mathcal{S}^n_+ \text{ et } \operatorname{tr} A^* = 1 \\ +\infty & \text{sinon.} \end{cases}$$

3. (a) Montrez que
$$\partial\lambda_{\max}(A) = \operatorname{co}\{vv^\mathsf{T} : \|v\|_2 = 1,\ Av = \lambda_{\max}(A)v\}.$$

   (b) En déduire que $\lambda_{\max}(\cdot)$ est différentiable en A si, et seulement si, $\lambda_{\max}(A)$ est simple, et que dans ce cas $\nabla\lambda_{\max}(A) = vv^\mathsf{T}$, où $\pm v$ sont les uniques vecteurs propres unitaires correspondants à la valeur propre maximale.

   (c) Retrouvez ce dernier résultat en utilisant le théorème des fonctions implicites.

7 Sous-différentiel de la distance à un ensemble

Soient $\mathbb{E}$ un espace vectoriel muni d'un produit scalaire noté $\langle\cdot,\cdot\rangle$, $\|\cdot\|$ une norme sur $\mathbb{E}$ (qui n'est pas nécessairement la norme associée au produit scalaire) et C un ensemble convexe fermé non vide de $\mathbb{E}$. On considère la fonction $d_C : \mathbb{E} \to \mathbb{R}$, la *distance à* C, définie par
$$d_C(x) = \inf_{y\in C} \|x-y\|.$$
On sait qu'elle est convexe (TD1).

1. Montrez que la conjuguée d^*_C de d_C s'écrit
$$d^*_C = \mathcal{I}_{B_D} + \sigma_C,$$
où B_D est la boule-unité fermée pour la *norme duale*, $\mathcal{I}_P$ est l'*indicatrice* de l'ensemble P (vaut 0 sur P et $+\infty$ ailleurs) et σ_C est la *fonction d'appui* de C (elle vaut $\sup_{y\in C}\langle x^*, y\rangle$ en $x^* \in \mathbb{E}$, le supremum de la fonction linéaire $y \in \mathbb{E} \mapsto \langle x^*, y\rangle$ sur C).

2. Montrez que
$$\partial d_C(x) = \{x^* \in B_D \cap N_C(\bar{x}) : \langle x^*, x-\bar{x}\rangle = \|x-\bar{x}\|\}, \tag{12.7}$$
où
   - $\bar{x}$ est *une projection* de x sur C, c'est-à-dire une solution du problème $\inf\{\|x-y\| : y \in C\}$ (celle-ci n'est pas nécessairement unique car la norme n'est pas nécessairement associée à un produit scalaire) et
   - $N_C(\bar{x}) := \{\nu\colon \langle \nu, y-\bar{x}\rangle \leqslant 0, \text{ pour tout } y \in C\}$ est le *cône normal* à C en $\bar{x}$.

3. Retrouvez (12.7) à partir de la formule donnant le sous-différentiel d'une fonction marginale (c'est plus long).

On suppose dorénavant que la norme $\|\cdot\|$ est celle associée au produit scalaire $\langle\cdot,\cdot\rangle$.

4. Montrez que $\|\cdot\|_{\mathrm{D}} = \|\cdot\|$.
5. On considère les cas exclusifs suivants.
   (a) Montrez que si $x \notin C$, alors d_C est différentiable en x et
   $$\nabla d_C(x) = \frac{x-\bar{x}}{\|x-\bar{x}\|}.$$
   (b) Montrez que si $x \in C^\circ$ (l'intérieur de C), alors d_C est différentiable en x et
   $$\nabla d_C(x) = 0.$$
   (c) Montrez que si $x \in \partial C = C \backslash C^\circ$ (la frontière de C), alors
   $$\partial d_C(x) = B \cap N_C(x).$$

Soit P une partie non vide et fermée de $\mathbb{E}$, mais non nécessairement convexe.

6. Montrez par un contre-exemple que d_P n'est pas nécessairement différentiable sur le complémentaire de P.
7. Soit $x \in \mathbb{E}$ tel que d_P soit différentiable en x et $d'_P(x) \neq 0$. Montrez que
   (a) $x \notin P$,
   (b) $\|\nabla d_P(x)\| \leqslant 1$,
   (c) le problème $\inf_{y\in P} \|x-y\|$ a une solution (une telle solution s'appelle une *projection* de x sur P),
   (d) $\nabla d_P(x) = (x-\bar{x})/\|x-\bar{x}\|$, où $\bar{x}$ est une projection de x sur P,
   (e) x a une unique projection sur P.

8 Sous-différentiel de la fonction duale d'un problème d'optimisation quadratique convexe

Soient H une matrice d'ordre n symétrique semi-définie positive (qui n'est donc pas nécessairement inversible), $g \in \mathbb{R}^n$, A une matrice $m \times n$ et $b \in \mathbb{R}^m$. On s'intéresse au problème d'optimisation quadratique convexe suivant

$$(P) \qquad \left\{ \begin{array}{l} \inf_x \left(q(x) := \frac{1}{2}\, x^\mathsf{T} H x + g^\mathsf{T} x\right) \\ Ax = b. \end{array} \right.$$

On prend comme fonction duale, la fonction $\delta : \mathbb{R}^m \to \overline{\mathbb{R}}$ définie en $\lambda \in \mathbb{R}^m$ par

$$\delta(\lambda) := -\inf_x \left(q(x) + \lambda^\mathsf{T}(Ax-b)\right). \tag{12.8}$$

1. Montrez que le domaine de δ s'écrit[6]

$$\operatorname{dom}\delta = \{\lambda \in \mathbb{R}^m : A^\mathsf{T}\lambda + g \in \mathcal{R}(H)\}. \tag{12.9}$$

2. Soient $H^\dagger$ la matrice pseudo-inverse de H[7], $S := AH^\dagger A^\mathsf{T}$ et $v := AH^\dagger g$. Montrez que

$$\forall\, \lambda \in \operatorname{dom}\delta, \qquad \delta(\lambda) = \tfrac{1}{2}\,\lambda^\mathsf{T} S\lambda + (v+b)^\mathsf{T}\lambda + \tfrac{1}{2}\,g^\mathsf{T} H^\dagger g. \tag{12.10}$$

3. On note $\partial\delta(\lambda)$ le sous-différentiel de δ en $\lambda \in \mathbb{R}^m$. Montrez que $\lambda^* \in \partial\delta(\lambda)$ si, et seulement si, il existe un $x \in \mathbb{R}^n$ tel que (λ, x) résout

$$\left\{\begin{array}{l} \inf_{(\mu,y)}\ \delta(\mu) - (\lambda^*)^\mathsf{T}\mu \\ A^\mathsf{T}\mu + g = Hy. \end{array}\right. \tag{12.11}$$

   En déduire que

$$\forall\, \lambda \in \operatorname{dom}\delta, \qquad \partial\delta(\lambda) = S\lambda + v + b + A(\mathcal{N}(H)). \tag{12.12}$$

9 Mécaniques lagrangienne et hamiltonienne sur un espace vectoriel

En *mécanique lagrangienne*, les équations du mouvement décrivent l'évolution temporelle de variables $q \in \mathbb{R}^n$ (les paramètres décrivant le système considéré) et de leur vitesse $\dot q = \frac{dq}{dt} \in \mathbb{R}^n$ au cours du temps t à partir de la donnée d'un *lagrangien* $L(q, \dot q, t)$ et l'*équation d'Euler-Lagrange*

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot q}\right) - \frac{\partial L}{\partial q} = 0. \tag{12.13}$$

En *mécanique hamiltonienne*, les équations du mouvement s'obtiennent à partir de l'*hamiltonien* $H(p, q, t)$ défini comme la conjuguée en $p \in \mathbb{R}^n$ de $L(q, \cdot, t)$:

$$H(p, q, t) = \sup_{\dot q \in \mathbb{R}^n} \Big(p^\mathsf{T}\dot q - L(q, \dot q, t)\Big). \tag{12.14}$$

La variable p est appelée *moment*. La définition de H implique que cette fonction est convexe en p. Les équations du mouvement sont alors données par les *équations d'Hamilton*, qui sont les équations différentielles du premier ordre suivantes

$$\dot q = \frac{\partial H}{\partial p}(p, q, t) \qquad \text{et} \qquad \dot p = -\frac{\partial H}{\partial q}(p, q, t). \tag{12.15}$$

Nous allons voir que ces équations se déduisent l'équation d'Euler-Lagrange dans le cas où $L(q, \cdot, t)$ est différentiable et où le problème de maximisation en (12.14) a une solution.

Familiarisons-nous avec ces formalismes, dans le cas où les forces $F(q)$ agissant sur une particule de coordonnée $q = q(t) \in \mathbb{R}^n$ et de masse constante $m \in \mathbb{R}_{++}$, dont on veut décrire le mouvement, dérivent d'un potentiel $V = V(q)$,

$$F(q) = -\nabla V(q) \in \mathbb{R}^n,$$

et où le lagrangien s'écrit comme la différence entre l'énergie cinétique et ce potentiel :

$$L(q, \dot q) = \frac{m}{2}\|\dot q\|_2^2 - V(q).$$

[6] Cette question requiert le résultat de Frank et Wolfe.

[7] Rappelons que $H^\dagger$ est la matrice qui est définie par le fait que, quel que soit $b \in \mathbb{R}^n$, $H^\dagger b$ est la solution de norme minimale de $\inf \frac{1}{2}\|Hx - b\|_2$.

1. Montrez que l'équation d'Euler-Lagrange (12.13) se ramène à l'*équation de Newton* (on a noté $\ddot{q}$ la dérivée seconde de q par rapport à t) :

$$F(q) = m\ddot{q}.$$

2. Écrire l'hamiltonien et montrez que les équations d'Hamilton conduisent également à l'équation de Newton ci-dessus.

On considère à présent le cas général, en supposant toutefois que $L(q, \cdot, t)$ est différentiable et que le problème de maximisation en (12.14) a une solution. Notre objectif est de montrer que les équations d'Hamilton (12.15) se déduisent de l'équation d'Euler-Lagrange (12.13).

3. En utilisant l'équation d'Euler-Lagrange, montrez que si $\dot{q}$ est solution du problème de maximisation en (12.14), on a

$$p = \frac{\partial L}{\partial \dot{q}}(q, \dot{q}, t) \qquad \text{et} \qquad \dot{p} = \frac{\partial L}{\partial q}(q, \dot{q}, t). \tag{12.16}$$

4. En déduire les équations d'Hamilton.