Cours OPT 202

Optimisation Différentiable – Théorie et Algorithmes

Exercices de la séance 13
(consolidation)

Analyse convexe

1. Cône dual d'une intersection de deux cônes

Dualité

2. Dualisation lagrangienne de la projection sur un polyèdre convexe
3. Poursuite de base ou recouvrement ℓ_1

Conjugaison et sous-différentiabilité

4. Convexification du compteur de composantes non nulles
5. Ensemble saillant de minimiseurs
6. Sous-différentiel de la fonction duale
7. Sous-différentiel de la fonction valeur

1 Cône dual d'une intersection de deux cônes

Soient $\mathbb{E}$ une espace euclidien et K_1 et K_2 deux cônes convexes fermés non vides de $\mathbb{E}$. Montrez que

$$(K_1 \cap K_2)^+ = \overline{K_1^+ + K_2^+}. \tag{13.1}$$

2 Dualisation lagrangienne de la projection sur un polyèdre convexe

On considère le problème de la projection sur un polyèdre convexe, qui peut s'écrire de la manière suivante

$$\begin{cases} \inf_{x\in\mathbb{R}^n} \frac{1}{2}\|x-z\|^2 \\ Ax = b \\ x \geqslant 0, \end{cases} \tag{13.2}$$

où $z \in \mathbb{R}^n$ est donné, $\|\cdot\|$ est la norme euclidienne, $A \in \mathbb{R}^{m\times n}$ et $b \in \mathbb{R}^m$. On suppose que l'ensemble admissible du problème est non vide : il existe un $x_0 \geqslant 0$ tel que $Ax_0 = b$.

1) Dans quel sens peut on dire que le dual lagrangien du problème (13.2) est le problème suivant:

$$\sup_{y\in\mathbb{R}^m} \left(-\frac{1}{2}\|(A^\mathsf{T} y + z)^+\|^2 + b^\mathsf{T} y + \frac{1}{2}\|z\|^2 \right). \tag{13.3}$$

2) Montrez qu'il n'y a pas de saut de dualité entre (13.2) et (13.3).

3) Montrez que les propriétés suivantes sont équivalentes :

   (a) l'ensemble des solutions de (13.3) est borné,
   (b) $\forall\, p \neq 0$ tel que $A^\mathsf{T} p \leqslant 0$, on a $b^\mathsf{T} p < 0$,
   (c) A est surjective et il existe un $x > 0$ tel que $Ax = b$.

4) Le but de ce numéro est d'examiner un dual de (13.2) de la forme (13.3) mais avec y restreint à l'image de A.

   4.1) Montrez que l'on peut aussi prendre comme dual de (13.2) le problème

$$\sup_{y\in\mathcal{R}(A)} \left(-\frac{1}{2}\|(A^\mathsf{T} y + z)^+\|^2 + b^\mathsf{T} y + \frac{1}{2}\|z\|^2 \right). \tag{13.4}$$

   4.2) Montrez qu'il n'y a pas non plus de saut de dualité entre (13.2) et (13.4).
   4.3) Montrez que les propriétés suivantes sont équivalentes :
      (a) l'ensemble des solutions de (13.4) est borné,
      (b) $\forall\, p \in \mathcal{R}(A) \setminus \{0\}$ tel que $A^\mathsf{T} p \leqslant 0$, on a $b^\mathsf{T} p < 0$,
      (c) il existe un $x > 0$ tel que $Ax = b$.

   Remarque. Si l'on compare les points 3.c et 4.3.c, on voit l'intérêt de la dualisation (13.4) qui est celui de restreindre l'ensemble des solutions duales. En effet, l'ensemble des solutions duales est borné sous des hypothèses plus faibles (pas besoin d'imposer la surjectivité de A).

3 Poursuite de base ou recouvrement ℓ_1

On considère le problème qui consiste à recouvrer un vecteur $x \in \mathbb{R}^n$ satisfaisant une contrainte affine et de *norme* ℓ_1 minimale. Il s'écrit donc

$$(P_1) \qquad \begin{cases} \inf_{x\in\mathbb{R}^n} \ \|x\|_1 \\ Ax = b, \end{cases}$$

où $\|x\|_1 = \sum_i |x_i|$, $A : \mathbb{R}^n \to \mathbb{R}^m$ est linéaire et $b \in \mathbb{R}^m$. On munit $\mathbb{R}^n$ et $\mathbb{R}^m$ du produit scalaire euclidien. On note $\mathcal{R}(A)$ l'image de A.

Le problème (P_1) s'utilise comme palliatif, plus aisé à résoudre (on peut le transformer en problème d'optimisation linéaire, voir la question 2 ci-dessous[1]), du problème NP-ardu, qui consiste à minimiser sur le même sous-espace affine le *compteur de composantes non nulles* de x :

$$\|x\|_0 := |\{i \in [1:n] : x_i \neq 0\}|.$$

Malgré sa notation, cette fonction $\|\cdot\|_0$ n'est pas une norme, par manque d'homogénéité positive. On obtient alors les solutions du système linéaire $Ax = b$ avec le plus de zéros possibles (utilisation en *acquisition comprimée* ou « *compressed sensing* » en *traitement du signal* [3, 2, 4, 1], dans les problèmes d'*apprentissage* [5], dans les problèmes à *données massives* ou « *big data* »). Les raisons pour lesquelles la norme ℓ_1 approche bien le compteur ℓ_0 sont les suivantes :

- la norme ℓ_1 est la plus grande fonction convexe (donc rendant le problème convexe, qui est plus facile à résoudre) qui minore le compteur ℓ_0 sur la boule unité (raison qui satisfait le spécialiste en optimisation),
- la norme ℓ_1 a ses sommets sur les axes de coordonnées (toutes les composantes y sont nulles sauf une) et la « projection » (au sens de la norme ℓ_1) de 0 sur le sous-espace affine $\{x : Ax = b\}$ a tendance à se trouver sur un sommet de la boule $\operatorname{val}(P_1)\bar{B}_1$ (raison qui satisfait le praticien).

Venons-en qux questions.

1. Montrez que le problème (P_1) a une solution si, et seulement si, $b \in \mathcal{R}(A)$.

2. Montrez comment on peut transformer (P_1) en un problème d'optimisation linéaire.

3. Montrez que, pour $y \in \mathbb{R}^n$, on a
$$\sup_{x\in\mathbb{R}^n} \left(y^\mathsf{T} x - \|x\|_1\right) = \mathcal{I}_{B_\infty}(y), \qquad (13.5)$$
où $B_\infty := \{y \in \mathbb{R}^n : \|y\|_\infty \leqslant 1\}$ est la boule unité fermée pour la norme $y \mapsto \|y\|_\infty := \max_i |y_i|$ et $\mathcal{I}_{B_\infty}$ est la fonction indicatrice de B_∞ (c'est-à-dire, $\mathcal{I}_{B_\infty}(y) = 0$ si $y \in B_\infty$ et $\mathcal{I}_{B_\infty}(y) = +\infty$ si $y \notin B_\infty$).

4. Montrez que le dual lagrangien du problème (P_1) s'écrit
$$(D_1) \qquad \begin{cases} \sup_{y\in\mathbb{R}^m} \ b^\mathsf{T} y \\ \|A^\mathsf{T} y\|_\infty \leqslant 1. \end{cases}$$

5. Montrez que le problème dual (D_1) a une solution si, et seulement si, $b \in \mathcal{R}(A)$.

[1] C'est pour cette raison que le problème (P_1) s'appelle parfois le problème de *poursuite de bases* (« *basis pursuit* » [3 ; 1998]), car l'algorithme du simplexe utilisé pour résoudre ce problème passe de « base » en « base » pour trouver une « solution basique » (beaucoup de jargon lié à l'algorithme du simplexe).

4 Convexification du compteur de composantes non nulles

On note $\|\cdot\|_0$ le *compteur de composante non nulle* d'un vecteur ; donc pour $x \in \mathbb{R}^n$, on a

$$\|x\|_0 := |\{i \in [1:n] : x_i \neq 0\}|$$

Dans cette définition, $|\cdot|$ est utilisé pour désigner le cardinal d'un ensemble. Malgré la notation, $\|\cdot\|_0$ n'est pas une norme, par manque de positivité homogène. On note également $\|x\|_p$ la norme ℓ_p de $x \in \mathbb{R}^n$ ($p = 1$ ou ∞) et $B_1 := \{x \in \mathbb{R}^n : \|x\|_1 \leqslant 1\}$ la boule unité fermée de $\mathbb{R}^n$ pour la norme ℓ_1. On considère la fonction $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ définie en $x \in \mathbb{R}^n$ par

$$f(x) = \|x\|_0 + \mathcal{I}_{B_1}(x),$$

où $\mathcal{I}_{B_1}$ est l'indicatrice de B_1 ($\mathcal{I}_{B_1}(x) = 0$ si $x \in B_1$ et $\mathcal{I}_{B_1}(x) = +\infty$ si $x \notin B_1$).

On note $f^* : \mathbb{R}^n \to \overline{\mathbb{R}}$ et $f^{**} : \mathbb{R}^n \to \overline{\mathbb{R}}$ les conjuguée et biconjuguée de f pour le produit scalaire euclidien de $\mathbb{R}^n$.

1. Montrez que $f^*(x^*) \geqslant 0$ pour tout $x^* \in \mathbb{R}^n$.
2. Montrez que $f^*(x^*) \leqslant (\|x^*\|_\infty - 1)^+$ pour tout $x^* \in \mathbb{R}^n$, où, pour un scalaire t, on a noté $t^+ := \max(0, t)$.
3. Montrez que, pour tout $x^* \in \mathbb{R}^n$, on a
$$f^*(x^*) = (\|x^*\|_\infty - 1)^+. \tag{13.6}$$
4. Montrez que, pour tout $x \in \mathbb{R}^n$, on a
$$f^{**}(x) = \|x\|_1 + \mathcal{I}_{B_1}(x). \tag{13.7}$$

Conclusion : on a montré que, sur la boule unité de la norme ℓ_1, $\|\cdot\|_1$ est la plus grande fonction convexe qui minore la fonction non convexe $\|\cdot\|_0$ (alors que sur $\mathbb{R}^n$ cette plus grande fonction convexe est nulle).

5 Ensemble saillant de minimiseurs

Soient $\mathbb{E}$ un espace euclidien de produit scalaire noté $\langle\cdot,\cdot\rangle$ (norme associée $\|\cdot\|$), $B := \{x \in \mathbb{E} : \|x\| \leqslant 1\}$ la boule unité fermée de $\mathbb{E}$, $f \in \overline{\operatorname{Conv}}(\mathbb{E})$ (ce qui signifie que f est une fonction convexe, propre et fermée) et $f_{\min} := \inf\{f(x) : x \in \mathbb{E}\}$. On suppose que $f_{\min}$ est fini et que f a un ensemble de minimiseurs $S := \{x \in \mathbb{E} : f(x) = f_{\min}\}$ non vide.

1. Montrez que, quel que soit $x \in \mathbb{E}$, le problème $\inf\{\|x - y\| : y \in S\}$ a une solution et une seule. On la note $\mathrm{P}_S(x)$.

On note $\operatorname{dist}(x, S) := \|x - \mathrm{P}_S(x)\|$ la distance de x à S, pour la norme $\|\cdot\|$. On dit que l'ensemble des minimiseurs S est *saillant* s'il existe $\alpha > 0$ tel que

$$\forall x \in \mathbb{E} : \quad f(x) \geqslant f_{\min} + \alpha \operatorname{dist}(x, S). \tag{13.8}$$

Dans une première partie, on se propose de montrer que cette propriété est *équivalente* à l'inclusion suivante (pour le même $\alpha > 0$)

$$S + \alpha B \subset \bigcup_{x \in S} \Big(x + \partial f(x)\Big), \tag{13.9}$$

où le sous-différentiel $\partial f(x)$ de f en $x \in \mathbb{E}$ est calculé pour le produit scalaire $\langle \cdot, \cdot \rangle$.

Supposons dans un premier temps que (13.8) ait lieu et l'on se donne pour objectif de démontrer (13.9). Soient $\bar{x}_0 \in S$ et $u \in B$. Il s'agit de montrer que $\bar{x}_0 + \alpha u$ peut s'écrire de la forme $\bar{x} + \bar{g}$, avec $\bar{x} \in S$ et $\bar{g} \in \partial f(\bar{x})$. On tente sa chance en prenant

$$\bar{x} := \mathrm{P}_S(\bar{x}_0 + \alpha u) \qquad \text{et} \qquad \bar{g} := \bar{x}_0 + \alpha u - \bar{x}.$$

2. Montrez que $\|\bar{g}\| \leqslant \alpha$.
3. Montrez que $\forall x \in \mathbb{E}$, $f(x) \geqslant f(\bar{x}) + \langle \bar{g}, x - \bar{x} \rangle$.
4. Montrez que (13.9) est vérifiée.

Supposons à présent que (13.9) ait lieu et l'on se donne pour objectif de démontrer (13.8). Soient $x \in \mathbb{E} \backslash S$ et $\bar{x} := \mathrm{P}_S(x)$.

5. Montrez qu'il existe un $\bar{x}_1 \in S$ et $g_1 \in \partial f(\bar{x}_1)$ tels que

$$\bar{x} + \alpha \frac{x - \bar{x}}{\|x - \bar{x}\|} = \bar{x}_1 + g_1.$$

6. Montrez que $f(x) \geqslant f_{\min} + \langle g_1, x - \bar{x}_1 \rangle$ et en déduire (13.8).

Dans une seconde partie, on considère l'algorithme de minimisation de f, qui calcule en $x \in \mathbb{E}$ le nouvel itéré x_+ comme solution de

$$\inf_{y \in \mathbb{E}} \left(f(y) + \frac{1}{2}\|y - x\|^2 \right). \tag{13.10}$$

7. Sachant que les fonctions de $\overline{\mathrm{Conv}}(\mathbb{E})$ ont une minorante affine, montrez que le problème (13.10) a une solution et une seule.
8. Supposons que f ait un ensemble saillant de minimiseurs. Montrez que si x est suffisamment proche de S, alors $x_+ \in S$.

6 Sous-différentiel de la fonction duale

Proposition 13.25 ou notes manuscrites du cours.

7 Sous-différentiel de la fonction valeur

1. Sous-différentiel de la fonciton marginale : proposition 3.68 ou notes manuscrites du cours.

2. Sous-différentiel de la fonction valeur : proposition 4.48 ou notes manuscrites du cours.