Cours SOD314: TD1

Mardi 4 février 2020

Ex. 1: Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction continue différentiable. Soit $x \in \mathbb{R}^n$.

- Résoudre le problème

$$\begin{aligned} &\min_{d \in \mathbb{R}^n} \; \nabla f(x)^\top d \\ &\text{s.t. } \|d\|_2 = 1 \; , \end{aligned} \tag{1}$$

  et montrer que la direction de pente maximale est $-\nabla f(x)$.

- On suppose maintenant que le gradient de f est L-Lipschitzien:

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \; . \tag{2}$$

  On rappelle que f satisfait alors l'inégalité:

$$f(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2}\|y - x\|_2^2 \; . \tag{3}$$

  Montrer que le pas optimal de descente est $\frac{1}{L}$.

Correction: *i) On réécrit la contrainte $\|d\|_2^2 = 1$ pour obtenir un problème équivalent. On dualise alors la contrainte pour obtenir le Lagrangien $\mathcal{L}(d, \mu) = \nabla f(x)^\top d + \mu(\|d\|_2^2 - 1)$. Les conditions d'optimalité au premier ordre donne: $\nabla_d \mathcal{L}(d, \mu) = 0$, soit $\nabla f(x) + 2\mu d = 0$, soit $d = -\dfrac{\nabla f(x)}{2\mu}$. Par ailleurs, on a $\|d\|_2 = 1$, d'où $\mu = \frac{\|\nabla f(x)\|}{2}$, puis $d = -\dfrac{\nabla f(x)}{\|\nabla f(x)\|}$.*

ii) Soit $x_{k+1} = x_k - \alpha \nabla f(x_k)$. On a alors

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + \nabla f(x_k)^\top (x_{k+1} - x_k) + \frac{L}{2}\|x_{k+1} - x_k\|_2^2 \\ &= f(x_k) - \alpha \nabla f(x_k)^\top \nabla f(x_k) + \frac{L}{2}\|\alpha \nabla f(x_k)\|_2^2 \\ &= f(x_k) + (\frac{L}{2}\alpha^2 - \alpha)\|\nabla f(x_k)\|_2^2 \; . \end{aligned}$$

En prenant alors $\alpha = \frac{1}{L}$ on minimise le terme de droite, qui est lui-même une borne supérieure pour $f(x_{k+1})$. On assure alors que l'objectif décroît au moins de $-\frac{1}{2L}\|\nabla f(x_k)\|_2^2$ entre deux itérations.

Ex. 2:(Transformée de Fenchel) Calculer la transformée de Fenchel des fonctions suivantes:

- $f(x) = |x|$.
- $f(x) = \frac{1}{2}\|x\|_2^2$.
- $f(x) = \frac{1}{2}x^\top Q x + b^\top x$ où $Q \in \mathbb{R}^{n \times n}$ est une matrice symétrique définie positive.
- $f(x) = \log(1 + \exp(x))$.

Correction:

i) Notons que $f^\star(p) = \sup_{x \in \mathbb{R}}(px - |x|) = \sup\left\{ \sup_{x \geq 0}(px - x), \sup_{x<0}(px + x)\right\}$. On a alors

- *Si* $p < -1$, *alors* $\sup_{x<0}((p+1)x) = +\infty$ *et donc* $f^\star(p) = +\infty$.
- *Si* $p > 1$ *alors* $\sup_{x\geq 0}((p-1)x) = +\infty$ *et donc* $f^\star(p) = +\infty$.
- *Si* $|p| \leq 1$, *alors* $\sup_{x<0}((p+1)x) = 0$ *et* $\sup_{x\geq 0}((p-1)x) = 0$. *D'où* $f^\star(p) = 0$.

On en déduit alors

$$f^\star(p) = \begin{cases} 0 & si\ |p| \leq 1 \\ +\infty & sinon \end{cases}$$

ii) On a par définition $f^\star(p) = \sup_{x\in\mathbb{R}^n} p^\top x - \frac{1}{2}\|x\|_2^2$. *Les conditions d'optimalité au premier ordre nous donne que le mimimum satisfait* $x^\sharp = p$, *d'où* $f^\star(p) = \frac{1}{2}\|p\|_2^2$.

iii) On a $f^\star(p) = \sup_x p^\top x - \frac{1}{2}x^\top Qx - b^\top x$. *Les conditions d'optimalité au premier ordre donnent*

$$p - Qx - b = 0$$

soit $x = Q^{-1}(p-b)$. *En remplaçant dans le problème d'optimisation précédent, on obtient* $f^\star(p) = \frac{1}{2}(p-b)^\top Q^{-1}(p-b)$.

iv) On a $f^\star(p) = \sup_x p^\top x - \log(1+\exp(x))$. *Les conditions d'optimalité au premier ordre nous donnent, si existence de la solution:* $p = \dfrac{\exp x}{1+\exp x}$. *On en déduit:*

- *Si* $p \in]0,1[$, *alors* $x = \log(\dfrac{p}{1-p})$. *En remplaçant on obtient alors* $f^\star(p) = p\log(\dfrac{p}{1-p}) + \log(1-p) = (1-p)\log(1-p) + p\log(p)$.
- *Si* $p \in \{0,1\}$, $f^\star(p) = 0$. *Dans ce cas, le sup n'est pas atteint.*
- *Sinon* $f^\star(p) = +\infty$.

Ex. 3: Soit $f : \mathbb{R}^n \to]-\infty,+\infty]$ et $g : \mathbb{R}^n \to]-\infty,+\infty]$ deux fonctions propres, convexes, s.c.i. On définit l'opération d'*inf-convolution* comme

$$(f \,\square\, g)(x) = \inf_{y\in\mathbb{R}^n} f(y) + g(x-y) \ . \tag{4}$$

Montrer que $(f+g)^\star = f^\star \,\square\, g^\star$.

Correction:

On a

$$\begin{aligned}(f \,\square\, g)^\star(p) &= \sup_x\ p^\top x - \inf_y(f(y) + g(x-y)) \\ &= \sup_{x,y}\ p^\top x - f(y) - g(x-y) \\ &= \sup_{x,y}\ p^\top y + p^\top(x-y) - f(y) - g(x-y) \\ &= \sup_{y,z}\ p^\top y + p^\top z - f(y) - g(z) \\ &= \sup_y(p^\top y - f(y)) + \sup_z(p^\top z - g(z)) \\ &= f^\star(p) + g^\star(p)\end{aligned}$$

On en déduit $(f \,\square\, g)^\star = f^\star + g^\star$. *Si on prend* $f' = f^\star$ *et* $g' = g^\star$, *le résultat se réécrit* $(f^\star \,\square\, g^\star)^\star = f + g$. *En effet,* f *et* g *étant convexes propres s.c.i, on a que* $f^{\star\star} = f$ *et* $g^{\star\star} = g$ *(Chapitre 1, proposition 5). D'où le résultat final en prenant les transformées de Fenchel des deux membres de l'égalité précédente.*

Ex. 4: On étudie un problème de régression avec une régularisation en norme ℓ_1. On suppose disponible un ensemble de labels $y_1, \cdots y_n \in \mathbb{R}$ correspondant à des observations $x_1, \cdots, x_n \in \mathbb{R}^p$. Le problème de régression cherche à trouver le paramètre $\theta \in \mathbb{R}^p$ solution du problème d'optimisation

$$\min_{\theta\in\mathbb{R}^p} \sum_{i=1}^n l(x_i^\top\theta, y_i) + \lambda\|\theta\|_1 \ , \tag{5}$$

où $l : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ est une fonction de pénalisation et $\lambda \in \mathbb{R}$ un paramètre de régularisation. Le terme $\|\cdot\|_1$ est non-differentiable. Pour le prendre en compte, on réécrit le problème sous la forme:

$$\min_{\theta\in\mathbb{R}^p} F(\theta) + G(\theta) \ , \tag{6}$$

avec $G : x \to \|x\|_1$.

- Calculer le sous-différentiel de G.
- Calculer la transformée de Fenchel de G. Faire un dessin.
- Exprimer $(F+G)^\star$ en fonction de $F^\star$ et $G^\star$.
- Expliquer l'intérêt de la formulation duale du problème.

Correction:

i) On a $G(x) = \|x\|_1 = \sum_{i=1}^n |x_i|$ qui est une fonction additive.

ii) On sait que la transformée de Fenchel de $h(x) = \|x\|_1$ vaut $h^\star(p) = \chi_{[-1,1]}(p)$, où $\chi_{[-1,1]}$ désigne la fonction indicatrice de l'intervalle $[-1,1]$. En utilisant l'additivité de G, on en déduit $G^\star(p) = \chi_B(p)$, avec $B = p \in \mathbb{R}^n \ : \ |x_i| \leq 1 \ , \ \forall i = 1, \cdots, n$.

iii) En utilisant l'exercice 3, on obtient $(F+G)^\star = F^\star \,\square\, G^\star$.

iv) Si la transformée de Fenchel de F est facile à calculer, la formulation duale du problème $(F+G)^\star$ présente l'avantage de ne plus être non-lisse, la fonction G se réécrivant dans le dual comme une fonction indicatrice, induisant des contraintes sur les bornes de $p \in \mathbb{R}^n$ Ces contraintes de borne sont habituellement faciles à traiter pour les solveurs d'optimisation.