Cours SOD314: TD2

Mardi 11 février 2020

Ex. 1: Soit $A \in \mathbb{R}^{n\times n}$ une matrice symétrique semi-définie positive, $b \in \mathbb{R}^n$ un vecteur et $c \in \mathbb{R}$ une constante. On définit la fonction

$$h : x \to \frac{1}{2}x^\top Ax + b^\top x + c \ . \tag{1}$$

- Montrer que $\text{prox}_{th}(x) = (I + tA)^{-1}(x - tb)$ pour tout $t > 0$.
- En déduire les opérateurs proximaux des fonctions suivantes:
  - $f(x) = b^\top x + c$.
  - $f(x) = c$.
  - $f(x) = \frac{1}{2}\|x\|_2^2$.

Correction:

i) On a par définition $\text{prox}_{th(}(y) = \arg\min_{x\in\mathbb{R}^n} \frac{1}{2}x^\top Ax + b^\top x + c + \frac{1}{2t}\|x - y\|_2$. *En écrivant les conditions d'optimalité au premier ordre du problème précédent, on obtient:* $Ax + b + \frac{1}{t}(x - y) = 0$, *soit* $(tAx + x) = y - tb$. *A est symétrique semi-définie positive, donc $tA + I$ est symétrique définie positive et a fortiori inversible. On en déduit alors* $\text{prox}_{th(}(y) = (I + tA)^{-1}(y - tb)$, *d'où le résultat.*

ii) On en déduit les cas particuliers:

- *Si $A = 0$,* $\text{prox}_{th(}(x) = x - tb$
- *Si $A = 0$ et $b = 0$,* $\text{prox}_{th(}(x) = x$.
- *Si $b = 0$, $c = 0$ et $A = I$, on a* $\text{prox}_{th(}(x) = \dfrac{1}{1+t}x$.

Ex. 2.:

- Soit $f : x \in \mathbb{R} \to |x|$. Montrer que

$$\text{prox}_{\lambda f}(v) = \begin{cases} v - \lambda & \text{si } v \geq \lambda \\ 0 & \text{si } |v| < \lambda \\ v + \lambda & \text{si } v \leq -\lambda \ . \end{cases} \tag{2}$$

  Tracer la fonction.

- En déduire $\text{prox}_{\lambda g}$, où $g = \|\cdot\|_1$.

Correction: *i)* $\text{prox}_{\lambda f}$ *est le minimiseur de la fonction*

$$h(y) = \begin{cases} \lambda y + \dfrac{1}{2}(y - x)^2 & si\ y > 0 \\ -\lambda y + \dfrac{1}{2}(y - x)^2 & si\ y \leq 0 \ . \end{cases}$$

On obtient dès lors la disjonction suivante. Si le minimiseur est atteint en $y > 0$, alors $y^\sharp = x - \lambda$. On en déduit que si $x > \lambda$, alors $\text{prox}_{\lambda f}(x) = x - \lambda$ *si $x > \lambda$. En suivant le même raisonnement, on montre que* $\text{prox}_{\lambda f}(x) = x + \lambda$ *si $x < -\lambda$. Enfin, si $|x| \leq \lambda$, alors $y^\sharp$ est atteint en un point de non-différentiabilité, donc nécessairement $y^\sharp = 0$.*

ii) On a $g(x) = \sum_{i=1}^n |x_i|$ fonction additive. On a alors

$$prox_{\lambda g}(x) = prox_{\lambda f}(x_1) \times \cdots \times prox_{\lambda f}(x_n) \ ,$$

où $f(x_i) = |x_i|$ est la fonction valeur absolue (voir Proposition 12, Chapitre 2).

Notons qu'en utilisant la décomposition de Moreau et l'Exercice 4, on obtient directement que

$$prox_{\lambda f}(x) = x - prox_{\lambda f^\star}(x) \ ,$$

avec $prox_{\lambda f^\star}(x) = \lambda \mathrm{proj}_{B_\infty}\left(\frac{x}{\lambda}\right)$ et où $B_\infty = \left\{x \in \mathbb{R}^n \ : \ |x_i| \leq 1 \ , \ \ \forall i = 1, \cdots, n\right\}$.

Ex. 3.: Soit $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ une fonction convexe. Calculer l'opérateur proximal de f dans les cas où:

- $f(x) = \|x\|_1 + \frac{\gamma}{2}\|x\|_2^2$.
- $f(x) = -\sum_{i=1}^n \log(x_i)$.

Correction: *ii) Calculons l'opérateur proximal de $g(x_i) = -\log(x_i)$, l'opérateur proximal de f s'en déduisant par formule d'additivité. $prox_{\lambda g}(x)$ réalise le minimum de la fonction*

$$y \mapsto -\log(y) + \frac{1}{2}(y-x)^2 \ ,$$

les conditions d'optimalité au premier ordre nous donnant alors $\dfrac{1}{y^\sharp} = y^\sharp - x$, ce qui se réécrit comme le polynôme de deuxième ordre $(y^\sharp)^2 - xy^\sharp - 1 = 0$. On en déduit alors que $prox_{\lambda g}$ est la racine positive de ce polynôme, soit $prox_{\lambda g}(x) = \dfrac{x + \sqrt{x^2+4}}{2}$.

Ex. 4.: Soit C un ensemble convexe, avec fonction support $\sigma_C(x) = \sup_{y \in C} y^\top x$ et fonction indicatrice χ_C.

1. Montrer que $\mathrm{prox}_{\chi_C}(x) = \mathrm{proj}_C(x)$.
2. En déduire que $\mathrm{prox}_{\sigma_C}(x) = x - \mathrm{proj}_C(x)$.

Correction:

i) On a

$$\begin{aligned} prox_{\chi_C}(x) &= \operatorname*{arg\,min}_{y \in \mathbb{R}^n} \ \chi_C(y) + \frac{1}{2}\|y-x\|_2^2 \\ &= \operatorname*{arg\,min}_{y \in C} \frac{1}{2}\|y-x\|_2^2 \ , \end{aligned}$$

qui est exactement la définition de l'opérateur projection sur l'ensemble C.

ii) Si on prend la transformée de Fenchel de χ_C:

$$\chi_C^\star(x) = \sup_{y \in \mathbb{R}^n} p^\top y - \chi_C(y) = \sup_{y \in C} p^\top y \ ,$$

on obtient exactement la fonction support σ_C. On en déduit $\chi_C^\star = \sigma_C$. Or, on sait d'après la décomposition de Moreau que $prox_f(x) + prox_{f^\star}(x) = x$, soit

$$prox_{\sigma_C}(x) = x - prox_{\chi_C}(x) = x - \mathrm{proj}_C(x) \ . \tag{3}$$

D'où le résultat.

Ex. 5.: Soit $f(x) = \|x\|$ une norme définie sur $\mathbb{R}^n$. On admet que la transformée de Fenchel de f satisfait

$$f^\star(y) = \begin{cases} 0 & \text{si } \|y\|_* \leq 1 \\ +\infty & \text{sinon} \ , \end{cases} \tag{4}$$

où $\|y\|_* = \sup_{x \in \mathbb{R}^n} y^\top x$ s.t. $\|x\| \leq 1$ est la norme duale de $\|\cdot\|$. En déduire que $\mathrm{prox}_f(x) = x - \mathrm{proj}_{B_*}(x)$ où $B_* = \left\{y \in \mathbb{R}^n \ : \ \|y\|_* \leq 1\right\}$.

Correction: *La transformée de Fenchel de f est la fonction indicatrice de B_*. En appliquant le résultat de l'exercice précedent, on en déduit directement que*

$$prox_f(x) = x - \mathrm{proj}_{B_*}(x) \ .$$

Ex. 6.: Calculer l'opérateur proximal de $f(x) = \max_i\{x_i\}$. (indice: utiliser que $\max_i\{x_i\} = \max_{y \in \Delta_n} y^\top x$ où $\Delta_n = \{z \in \mathbb{R}^n \mid \forall i \ z_i \geq 0 \ , \ \sum_{i=1}^n z_i = 1\}$).

Correction: *On a $f(x) = \max_i x_i = \max_{y \in \Delta_n} y^\top x$. La fonction f est donc la fonction support de Δ_n. Donc $f^\star = \chi_{\Delta_n}$, fonction indicatrice du simplexe en dimension n. On en déduit alors, par théorème de décomposition de Moreau:*

$$prox_f(x) = x - \mathrm{proj}_{\Delta_n}(x) \ .$$