Conjugada de Fenchel y cálculo de algunos operadores proximales

Def

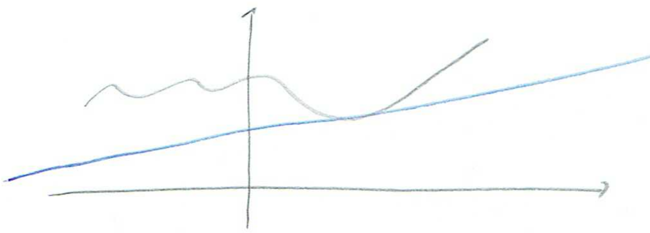
Dada $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convexa, la conjugada de Fenchel de f es:

$$f^*: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \qquad f^*(x^*) = \sup_x \{ \langle x, x^* \rangle - f(x) \}$$

Interpretación

Decimos que $\ell(x) = \langle x, x^* \rangle - \alpha$ es minorante afín si

$f(x) \geq \langle x, x^* \rangle - \alpha \quad \forall x.$



Debe ser entonces

$$\alpha \geq \langle x, x^* \rangle - f(x) \quad \forall x$$

El mejor valor de α que da una minorante es:

$$\bar{\alpha} = \sup_x \{ \langle x, x^* \rangle - f(x) \} \quad \text{que es} \quad f^*(x^*)$$

Prop

$$y \in \partial f(x) \iff x \in \partial f^*(y)$$

Descomposición de Moreau

Para todo $x \in \mathbb{R}^n$, vale $\quad x = \text{prox}_f(x) + \text{prox}_{f^*}(x)$

Dem

$$u = \text{prox}_f(x) \iff x - u \in \partial f(u) \iff u \in \partial f^*(x-u) \iff 0 \in \partial f^*(x-u) - u$$

$$\iff 0 \in \partial f^*(x-u) + x - u - x \iff x - u \text{ minimiza } f^*(z) + \frac{1}{2}\|z - x\|_2^2$$

$$\iff x - u = \text{prox}_{f^*}(x) \quad \Rightarrow \quad \text{prox}_f(x) + \text{prox}_{f^*}(x) = u + x - u = x$$

Obs

Se puede pensar como una generalización de $v = \Pi_L(v) + \Pi_{L^\perp}(v)$ donde L es un subespacio y $L^\perp$ es el complemento ortogonal.

(Hay que ver que $(I_L)^* = I_{L^\perp}$)

Operadores proximales de normas

Se puede ver que la conjugada de la función norma es la indicatriz de la bola unitaria con la norma dual. Esto es:

$$\text{Si} \quad f(x) = \|x\| \implies f^*(y) = I_B(y) \quad \text{con} \quad B = \{y \,/\, \|y\|_* \leq 1\}$$

Algunas normas y sus duales

La norma dual de $\|\cdot\|_2$ es ella misma

La norma dual de $\|\cdot\|_1$ es $\|\cdot\|_\infty$ y viceversa

En general, la norma dual de $\|\cdot\|_p$ es $\|\cdot\|_q$ con $\frac{1}{p} + \frac{1}{q} = 1$

Esto permite evaluar el operador proximal de una norma, si sabemos proyectar sobre la bola de la norma dual:

$$\text{prox}_f(x) = x - \text{prox}_{f^*}(x) = x - P_B(x)$$

o más en general:

$$\text{prox}_{\lambda f}(x) = x - \lambda P_B\left(\frac{x}{\lambda}\right)$$

Ejemplos:

La proyección a la bola unidad con la norma 2 es $P_B(v) = \begin{cases} \frac{v}{\|v\|_2} & \text{si } \|v\|_2 > 1 \\ v & \text{si } \|v\|_2 \leq 1 \end{cases}$

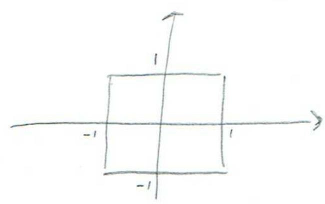
Por lo tanto, si $f(x) = \|x\|_2 \Rightarrow \operatorname{prox}_{\lambda f}(v) = \begin{cases} \left(1 - \frac{\lambda}{\|v\|_2}\right) v & \text{si } \|v\|_2 \geq \lambda \\ 0 & \text{si } \|v\|_2 < \lambda \end{cases}$

Este operador se suele llamar Block soft thresholding

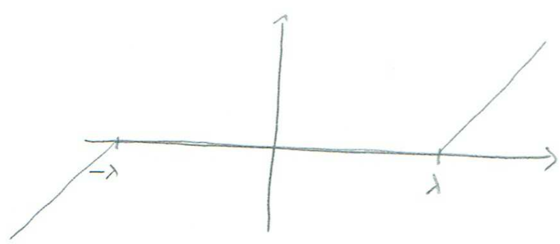
Tomemos ahora $f(x) = \|x\|_1$ nuevamente.

Si B es la bola unidad de la norma $\|\cdot\|_\infty$,

la proyección es: $(P_B(v))_i = \begin{cases} 1 & \text{si } v_i > 1 \\ v_i & |v_i| \leq 1 \\ -1 & v_i < -1 \end{cases}$



Entonces $(\operatorname{prox}_{\lambda f}(v))_i = \begin{cases} v_i - \lambda & \text{si } v_i \geq \lambda \\ 0 & |v_i| \leq \lambda \\ v_i + \lambda & v_i \leq -\lambda \end{cases}$



Convergencia del Método Proximal

Recordemos que tenemos $x^{k+1} = \text{prox}_f(x^k)$ y buscamos un punto fijo de prox_f.

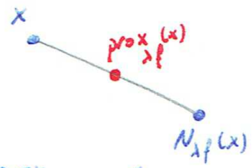
Def

Decimos que $T: \mathbb{R}^n \to \mathbb{R}^n$ es <u>no expansivo</u> sii $\|T(x) - T(y)\| \leq \|x - y\| \quad \forall x, y \in \mathbb{R}^n$

Consideremos el mapa $N_{\lambda f}: \mathbb{R}^n \to \mathbb{R}^n$ definido como $N_{\lambda f}(x) = 2\,\text{prox}_{\lambda f}(x) - x$

Obs:

- Interpretación geométrica:



- Si x^* es punto fijo de $\text{prox}_{\lambda f}$, entonces
$$N_{\lambda f}(x^*) = 2\,\text{prox}_{\lambda f}(x^*) - x^* = 2x^* - x^* = x^*$$

Veremos que $N_{\lambda f}$ es no expansivo, y que cualquier iteración interpolada
$$x^{k+1} = (1-\alpha)x^k + \alpha N_{\lambda f}(x^k), \quad \alpha \in (0,1)$$
converge a un punto fijo.

Prop

prox_f es firmemente no expansivo, es decir:
$$\|\text{prox}_f(x) - \text{prox}_f(x')\|^2 \leq (\text{prox}_f(x) - \text{prox}_f(x'))^T(x - x') \qquad \forall x, x' \in \mathbb{R}^n$$

Dem

Llamemos $z = \text{prox}_f(x)$, $z' = \text{prox}_f(x')$. De las condiciones de optimalidad:
$$x - z \in \partial f(z) \implies f(y) - f(z) \geq (x-z)^T(y-z) \quad \forall y \in \mathbb{R}^n \overset{y=z'}{\implies} f(z') - f(z) \geq (x-z)^T(z'-z)$$
$$x' - z' \in \partial f(z') \implies f(y) - f(z') \geq (x'-z')^T(y-z') \quad \forall y \in \mathbb{R}^n \overset{y=z}{\implies} f(z) - f(z') \geq (x'-z')^T(z-z')$$

Sumando: $\quad 0 \geq (x - z - x' + z')^T(z' - z) = \|z' - z\|^2 - (x' - x)^T(z' - z)$

Prop

N_f es no expansivo

Dem

$$\|N_f(x) - N_f(x')\|^2 = \|2\,\text{prox}_f(x) - x - 2\,\text{prox}_f(x') + x'\|^2 = \|2(\text{prox}_f(x) - \text{prox}_f(x')) - (x - x')\|^2$$
$$= \underbrace{4\|\text{prox}_f(x) - \text{prox}_f(x')\|^2 - 4(\text{prox}_f(x) - \text{prox}_f(x'))^T(x-x')}_{\leq 0 \text{ pues } \text{prox}_f \text{ es firmemente no expansivo}} + \|x - x'\|^2$$

$$\implies \|N_f(x) - N_f(x')\|^2 \leq \|x - x'\|^2$$

Teorema (Krasnosel'skiĭ - Mann)

Sea $N: \mathbb{R}^n \to \mathbb{R}^n$ no expansivo con al menos un punto fijo. Entonces la iteración

$$x^{k+1} = (1-\alpha^k)x^k + \alpha^k N(x^k) \quad \text{con} \quad \alpha^k \in [0,1] \ \forall k \quad y \quad \sum_k \alpha^k(1-\alpha^k) = +\infty$$

converge a un punto fijo de N.

Dem (usaremos la igualdad: $\|\alpha x + (1-\alpha)y\|^2 = \alpha\|x\|^2 + (1-\alpha)\|y\|^2 - \alpha(1-\alpha)\|x-y\|^2$, $\alpha \in [0,1]$)

Sea x^* un punto fijo de N. Entonces:

$$\|x^{k+1} - x^*\|^2 = \|(1-\alpha^k)x^k + \alpha^k N(x^k) - (1-\alpha^k+\alpha^k)x^*\|^2 = \|(1-\alpha^k)(x^k - x^*) + \alpha^k(N(x^k) - N(x^*))\|^2$$

$$= (1-\alpha^k)\|x^k - x^*\|^2 + \alpha^k \underbrace{\|N(x^k) - N(x^*)\|^2}_{\leq \|x^k - x^*\|^2} - \alpha^k(1-\alpha^k)\|N(x^k) - x^k\|^2 \leq \|x^k - x^*\|^2 - \alpha^k(1-\alpha^k)\|N(x^k) - x^k\|^2$$

O sea: $\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \alpha^k(1-\alpha^k)\|N(x^k) - x^k\|^2$

Sumando para todo k: $\displaystyle\sum_{k=0}^{+\infty} \alpha^k(1-\alpha^k)\|N(x^k) - x^k\|^2 \leq \|x^0 - x^*\|^2$

Como $\sum \alpha^k(1-\alpha^k) = +\infty \Rightarrow \exists$ subsucesión $(x^k)_K$ / $\displaystyle\lim_{\substack{k\in K \\ k\to\infty}} \|N(x^k) - x^k\| = 0$

Como $\{x^k\}_K$ es acotada $\Rightarrow \exists$ subsucesión $\{x^k\}_{\bar{K}}$ / $\displaystyle\lim_{\substack{k\in \bar{K} \\ k\to\infty}} x^k = \bar{x}$

Como N es no expansivo $\Rightarrow$ es continua $\Rightarrow$ $\displaystyle\lim_{\substack{k\in \bar{K} \\ k\to\infty}} N(x^k) = N(\bar{x})$

Como teníamos $\displaystyle\lim_{\substack{k\in \bar{K} \\ k\to\infty}} \|N(x^k) - x^k\| = 0 \Rightarrow \bar{x}$ es punto fijo de N

Como $\bar{x}$ es punto fijo, tenemos $\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2$

Entonces $\|x^{k+1} - \bar{x}\|$ es monótona decreciente, y como tiene una subsucesión convergente a $0 \Rightarrow$ toda la sucesión converge a $0 \Rightarrow x^k \to \bar{x}$

Corolario

Si x^* es mínimo de $f \Rightarrow x^{k+1} = \text{prox}_f(x^k)$ converge a x^*.

Dem Tomar $\alpha^k = \frac{1}{2}$