Convex Optimization for ML & CV

Name: Uzair Akbar ID: 03697290

Weekly Exercise 6 - Proximal Operator

Exercise 1

We have to show that,

$$0 \in \partial J(u^*) \iff u^* = \text{prox}_J(u^*)$$

Recall the following definitions:

$$\partial J(u) := \{ p \in E \mid J(v) \geq J(u) + \langle p, v-u \rangle, \forall v \in E \}$$

$$\text{prox}_J(v) := \arg\min_u J(u) + \frac{1}{2}\|u-v\|^2$$

Proof $\Rightarrow$

Assume $0 \in \partial J(u^*)$

$$\Rightarrow J(v) \geq J(u^*), \quad \text{for } p=0, \ \forall v \in E$$

$$\Rightarrow J(v) + \frac{1}{2}\|v-u^*\|^2 \geq J(u^*) \qquad \left[\because \frac{1}{2}\|v-u^*\|^2 \geq 0 \right]$$

$$\Rightarrow J(v) + \frac{1}{2}\|v-u^*\|^2 \geq J(u^*) + \frac{1}{2}\|u^*-u^*\|^2$$

$$\Rightarrow \min_v J(v) + \frac{1}{2}\|v-u^*\|^2 = J(u^*) + \frac{1}{2}\|u^*-u^*\|^2$$

$$\Rightarrow \min_v P_{u^*}(v) = P_{u^*}(u^*)$$

$$\left[\text{for } P_u(v) := J(v) + \frac{1}{2}\|v-u\|^2\right]$$

$$\Rightarrow \arg\min_v P_{u^*}(v) = u^*$$

$$\therefore \operatorname{prox}_J(u^*) = u^* \qquad \left[\text{Definition of } \operatorname{prox}_J(v)\right]$$

Proof $\Leftarrow$

Assume $\quad u^* = \operatorname{prox}_J(u^*)$

Now, define $\quad P_u(v) := J(v) + \frac{1}{2}\|v-u\|^2$

Then, $\quad u^* = \arg\min_v P_{u^*}(v)$

$$\Rightarrow 0 \in \partial P_{u^*}(u^*)$$

$$\Rightarrow 0 \in \partial\left\{J(u^*) + \frac{1}{2}\|u^*-u^*\|^2\right\}$$

$$\Rightarrow 0 \in \partial\left\{J(u^*)\right\}$$

$$\therefore 0 \in \partial J(u^*)$$

Exercise 2

Part A

We have to show that:

$$g(u) = \alpha f(u) + b, \ \alpha > 0$$

$$\implies \operatorname{prox}_{\lambda g}(v) = \operatorname{prox}_{\alpha\lambda f}(v)$$

From the definition of the proximal operator:

$$\because \operatorname{prox}_{\lambda g}(v) := \arg\min_u g(u) + \frac{1}{2\lambda}\|u-v\|_2^2$$

$$= \arg\min_u \alpha f(u) + b + \frac{1}{2\lambda}\|u-v\|_2^2$$

$$= \arg\min_u \alpha f(u) + \frac{1}{2\lambda}\|u-v\|_2^2$$

$$= \arg\min_u \frac{1}{\alpha}\left\{\alpha f(u) + \frac{1}{2\lambda}\|u-v\|_2^2\right\}$$

$$\left[\because \arg\min_x f(x) = \arg\min_x c f(x) \text{ for } c > 0\right]$$

$$= \arg\min_u f(u) + \frac{1}{2\alpha\lambda}\|u-v\|_2^2$$

$$\therefore \operatorname{prox}_{\lambda g}(v) = \operatorname{prox}_{\alpha\lambda f}(v)$$

Part B

We have to show that

$$J(u) = f(\mathcal{Q}u) \Rightarrow \operatorname{prox}_{\lambda J}(v) = \mathcal{Q}^T \cdot \operatorname{prox}_{\lambda f}(\mathcal{Q}v)$$

From the definition of the proximal operator:

$$\because \operatorname{prox}_{\lambda J}(v) := \arg\min_u J(u) + \frac{1}{2\lambda}\|u-v\|_2^2$$

$$= \arg\min_u f(\mathcal{Q}u) + \frac{1}{2\lambda}\|u-v\|_2^2$$

$$= \arg\min_u f(\mathcal{Q}u) + \frac{1}{2\lambda}\|\mathcal{Q}(u-v)\|_2^2$$

$[\because \|Au\|_2 = \|u\|_2,$ A orthonormal$]$

$$= \arg\min_u f(\mathcal{Q}u) + \frac{1}{2\lambda}\|\mathcal{Q}u - \mathcal{Q}v\|_2^2$$

$$= \mathcal{Q}^T \arg\min_w f(w) + \frac{1}{2\lambda}\|w - \mathcal{Q}v\|_2^2$$

$[w := \mathcal{Q}u]$

$$\therefore \operatorname{prox}_{\lambda J}(v) = \mathcal{Q}^T \cdot \operatorname{prox}_{\lambda f}(\mathcal{Q}v)$$

Exercise 3

$$\because J(u) = \|u\|_1 = \sum_{i=1}^{n} |u_i| \quad , u \in \mathbb{R}^n$$

$$\Rightarrow \text{prox}_{\lambda J}(v) = \arg\min_u J(u) + \frac{1}{2\lambda}\|u-v\|_2^2$$

$$\Rightarrow \quad = \arg\min_u \sum_{i=1}^{n} |u_i| + \frac{1}{2\lambda}\|u-v\|_2^2$$

$$\Rightarrow \quad = \arg\min_u \sum_{i=1}^{n} |u_i| + \frac{1}{2\lambda}\langle u-v, u-v\rangle$$

$$\Rightarrow \text{prox}_{\lambda J}(v) = \arg\min_u \sum_{i=1}^{n} |u_i| + \frac{1}{2\lambda}\sum_{i=1}^{n}(u_i - v_i)^2$$

Now, we solve the above minimization elementwise by first defining:

$$P_{v_i}(u_i) := |u_i| + \frac{1}{2\lambda}(u_i - v_i)^2$$

Then, for u_i^* to be a minimizer of $P_{v_i}(u_i)$,

$$\Rightarrow 0 \in \partial P_{v_i}(u_i^*) = \begin{cases} 1/\lambda\,(u_i^* - v_i) + \text{sign}(u_i^*) & , u_i^* \neq 0 \\ 1/\lambda\,(u_i^* - v_i) + \varepsilon , & u_i^* = 0, |\varepsilon| \leq 1 \end{cases}$$

$$\Rightarrow -\frac{1}{\lambda}(u_i^* - v_i) \in \begin{cases} \text{sign}(u_i^*) & , u_i^* \neq 0 \\ \varepsilon & , u_i^* = 0, |\varepsilon| \leq 1 \end{cases}$$

$$\Rightarrow v_i - u_i^* \in \begin{cases} \lambda \,\text{sign}(u_i^*) & , u_i^* \neq 0 \\ \varepsilon & , u_i^* = 0, |\varepsilon| \leq \lambda \end{cases}$$

$$\Rightarrow v_i \in \begin{cases} u_i^* + \lambda.\text{sign}(u_i^*) & , u_i^* \neq 0 \\ + \varepsilon & , u_i^* = 0, |\varepsilon| \leq \lambda \end{cases}$$

Now, for $v_i < -\lambda$,

$\because v_i = u_i^* + \lambda.\text{sign}(u_i^*)$ [$\because$ case 2 isn't possible as $|\varepsilon| \leq \lambda$]

$\Rightarrow -\lambda > u_i^* + \lambda\, \text{sign}(u_i^*)$ [$\because v_i < -\lambda$]

$\Rightarrow \text{sign}(u_i^*) = -1 \quad \Rightarrow v_i = u_i^* - \lambda$

$\therefore u_i^* = v_i + \lambda$, for $v_i < -\lambda$ — (i)

Similarly, for $v_i > \lambda$, $u_i^* = v_i - \lambda$. — (ii)

Lastly, for $v \in \varepsilon = [-\lambda, \lambda]$, $u_i^* = 0$. — (iii)

Putting (i), (ii) & (iii) together:

$$\Rightarrow \text{prox}_{\lambda J}(v) = u \in \mathbb{R}^n, \; u_i = \begin{cases} v_i + \lambda & , v_i < -\lambda \\ 0 & , v_i \in [-\lambda, \lambda] \\ v_i - \lambda & , v_i > \lambda. \end{cases}$$

Scanned by CamScanner

Exercise 4

$$J(X) := \|X\|_{1,2} = \sum_{i}^{m} \|X_i\|_2 \quad , X \in \mathbb{R}^{m \times n}, \quad X_i \in \mathbb{R}^n$$

Now,

$$\because \text{prox}_{\tau J}(X) := \arg\min_{Y} J(Y) + \frac{1}{2\tau}\|Y - X\|_2^2$$

let us define

$$P_x(Y) := J(Y) + \frac{1}{2\tau}\|Y - X\|_2^2$$

Then,

$$0 \in \partial P_x(Y^*) \quad , \quad \text{for } Y^* = \text{prox}_{\tau J}(X)$$

$$\Rightarrow 0 \in \partial J(Y^*) + \frac{1}{\tau}(Y^* - X)$$

$$\Rightarrow 0 \in \partial \|Y^*\|_{1,2} + \frac{1}{\tau}(Y^* - X)$$

$$\Rightarrow 0 \in \partial \|Y_i^*\|_2 + \frac{1}{\tau}(Y_i^* - X_i) \quad \left[\because \partial\|X\|_{1,2} = \begin{bmatrix} \partial\|X_1\|_2 \\ \vdots \\ \partial\|X_i\|_2 \\ \vdots \\ \partial\|X_m\|_2 \end{bmatrix} \right] \quad \text{—— (i)}$$

Now, as we already know:

$$\because \partial\|u\|_2 := \left\{ p \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} \|p\|_2 \leq 1 & , u = 0 \\ p = u/\|u\|_2 & , u \neq 0 \end{array} \right\}$$

$$= \left\{ p \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} p \in B_1(0) & , u = 0 \\ p = u/\|u\|_2 & , u \neq 0 \end{array} \right\}, \quad \left[B_1(0) \text{ is a closed unit ball.} \right]$$

Now, for $Y_i^* = 0$, (i) gives

$$\Rightarrow 0 \in B_1(0) + \frac{1}{\tau}(0 - X_i)$$ [$B_1(0)$ is a closed unit ball]

$$\Rightarrow 0 \in B_1(0) - \frac{X_i}{\tau}$$

$$\Rightarrow 0 \in B_\tau(0) - X_i$$

$$\Rightarrow X_i \in B_\tau(0)$$

$$\therefore \|X_i\|_2 \leq \tau \quad , \quad Y_i^* = 0 \quad \text{--- (ii)}$$

~~$\Rightarrow \operatorname{prox}_{\tau g}(X) = Y_i^* = 0$ for~~

Now, for $Y_i^* \neq 0$, (i) gives

$$\Rightarrow 0 = \frac{Y_i^*}{\|Y_i^*\|_2} + \frac{1}{\tau}(Y_i^* - X_i)$$

$$\Rightarrow X_i = \tau \frac{Y_i^*}{\|Y_i^*\|_2} + Y_i^* = (\tau + \|Y_i^*\|_2) \frac{Y_i^*}{\|Y_i^*\|_2}$$

$$\Rightarrow \|X_i\|_2 \frac{X_i}{\|X_i\|_2} = (\tau + \|Y_i^*\|_2) \frac{Y_i^*}{\|Y_i^*\|_2}$$

$$\Rightarrow \|X_i\|_2 = (\tau + \|Y_i^*\|_2)$$ $\left[\because X_i/\|X_i\| = Y_i^*/\|Y_i^*\|\right]$

$$\therefore \|X_i\|_2 > \tau \quad \text{--- (iii)}$$ $\left[\because Y_i^* \neq 0 \Rightarrow \|Y_i^*\|_2 > 0\right]$

Also, for $Y_i^* \neq 0$,

$$\Rightarrow 0 = \frac{Y_i^*}{\|Y_i^*\|_2} + \frac{1}{\tau}(Y_i^* - X_i)$$

$$\Rightarrow 0 = \frac{X_i}{\|X_i\|_2} + \frac{1}{\tau}(Y_i^* - X_i)$$

$\left[\because X_i/\|X_i\| = Y_i^*/\|Y_i^*\| \text{ as established before.}\right]$

$$\Rightarrow Y_i^* = X_i - \tau \frac{X_i}{\|X_i\|_2} \quad \text{--- (iv)}$$

Therefore, from (ii) & (iii) + (iv),

$$\Rightarrow \mathrm{prox}_{\tau J}(X) = Y^* = \left\{ Y \in \mathbb{R}^{m \times n} \,\middle|\, Y_i = \begin{cases} 0 &, \|X_i\|_2 \leq \tau \\ X_i - \tau \dfrac{X_i}{\|X_i\|_2} &, \|X_i\|_2 > \tau \end{cases} \right\}$$