Homework 2 Vũ Trung Hiếu 22022515

I) PROBLEM 1: properties and examples of ~~st~~ (sub) gradients

[a)] Show that $\partial f(x)$ is a closed and convex set for any function f (not neccessarily convex) and any point at x in its domain

Ta có

$$\partial f(x) = \{ g \in \mathbb{R}^n : f(y) \geq f(x) + g^T(y-x) \ \forall y \}$$

⊕ Chứng minh $\partial f(x)$ là tập lồi

Chọn $g_1, g_2 \in \partial f(x) \Rightarrow \begin{cases} f(y) \geq f(x) + g_1^T(y-x) \\ f(y) \geq f(x) + g_2^T(y-x) \end{cases}$

Xét $t g_1 + (1-t) g_2$

$f(x) + \cancel{t g_1} (t g_1 + (1-t) g_2)^T (y-x)$

$= t(f(x) + g_1^T(y-x)) + (1-t)(f(x) + g_2^T(y-x))$

$\leq \cancel{tt}\ t f(y) + (1-t) f(y) = f(y)$

$\Rightarrow t g_1 + (1-t) g_2 \in \partial f(x)$

$\Rightarrow \partial f(x)$ là tập lồi

⊕ Chứng minh $\partial f(x)$ là tập đóng

Đặt $S_y = \{ g \in \mathbb{R}^n : f(y) \geq f(x) + g^T(y-x) \}$

$\Rightarrow S_y$ là halfspace $\Rightarrow S_y$ là tập đóng

Mà $\partial f(x) = \bigcap_y S_y \Rightarrow \partial f(x)$ là tập đóng

[b)]

Để chứng minh $g \in \partial f(x)$, ta cần chứng minh $f(y) \geq f(x) + g^T(y-x) \quad \forall y \notin U$

Chọn điểm $z \notin U$

Do tập U mở

$\Rightarrow \exists\, 0 < t < 1 : y = tx + (1-t)z \in U$

$\Leftrightarrow f(y) \geq f(x) + g^T(y-x) \quad \forall y \in U$

$\Rightarrow f(tx + (1-t)z) \geq f(x) + g^T[tx + (1-t)z - x]$

mà f là hàm lồi

$\Rightarrow f(tx + (1-t)z) \leq tf(x) + (1-t)f(z)$

$\Rightarrow tf(x) + (1-t)f(z) \geq f(x) + g^T[tx + (1-t)z - x]$

$\Leftrightarrow (t-1)f(x) + (1-t)f(z) \geq g^T[tx + (1-t)z - x]$

$\Leftrightarrow$ ~~(t-1) f(x) + ...~~

$\Leftrightarrow (1-t)[f(z) - f(x)] \geq (1-t)g^T(z-x)$

$\Leftrightarrow f(z) \geq f(x) + g^T(z-x) \quad \forall z \notin U$

$\Rightarrow g \in \partial f(x) \quad (\square)$

$\boxed{c)}$ ~~For a con~~

Ta có

$g_x \in \partial f(x) \Rightarrow f(y) \geq f(x) + g_x^T(y-x)$

$g_y \in \partial f(y) \Rightarrow f(x) \geq f(y) + g_y^T(y-x)$

$\Rightarrow (g_x - g_y)^T(x-y) \geq 0$

$\boxed{d)}$

Ta có

với $x \neq 0 \Rightarrow \nabla f(x) = \dfrac{x}{\|x\|_2} \Rightarrow g \in \left\{ \dfrac{x}{\|x\|_2} \right\}$

Với $x = 0$

$\Rightarrow g \in \partial f(0) = \{ v : f(y) \geq f(0) + v^T(y-0)\ \forall y \}$

$\Rightarrow \|y\|_2 \geq v^T y$

$\Rightarrow \|v\|_2 \leq 1$

$\Rightarrow g \in \{ v : \|v\|_2 \leq 1 \}$

$\boxed{e)}$ Ta có

Xét $f(x) = \max_{s \in S} f_s(x)$

$\Rightarrow f(y) \geq f_s(y) \geq f_s(x) + g^T(y-x) = f(x) + g^T(y-x)$

$\Rightarrow \forall g \in \partial f_s(x) : g \in \partial f(x)$

$\Rightarrow \bigcup_{s : f_s(x) = f(x)} \partial f_s(x) \subseteq \partial f(x)$

Mà $\partial f(x)$ là tập lồi

$\Rightarrow \text{conv}\left(\bigcup_{s : f_s(x) = f(x)} \partial f_s(x) \right) \subseteq \partial f(x)$

II) Properties and examples of proximal operators

a) Để chứng minh $\text{prox}_{h,t}$ là hàm well defined

Ta cần chứng minh $f(z) = \frac{1}{2t} \| x - z \|_2^2 + h(z)$ là hàm lồi chặt

Với $0 < \alpha < 1$ ta có

$f(\alpha z_1 + (1-\alpha) z_2)$

$= \frac{1}{2t} \| x - (\alpha z_1 + (1-\alpha) z_2) \|_2^2 + h(\alpha z_1 + (1-\alpha) z_2)$

$= \frac{1}{2t} \| \alpha(x - z_1) + (1-\alpha)(x - z_2) \|_2^2 + h(\alpha z_1 + (1-\alpha) z_2)$

$\le \alpha\left(\frac{1}{2t}\|x - z_1\|_2^2 + h(z_1)\right) + (1-\alpha)\left(\frac{1}{2t}\|x - z_2\|_2^2 + h(z_2)\right)$

$- \frac{1}{2t}\alpha(1-\alpha)\|z_1 - z_2\|_2^2 + (1-\alpha)\left(\frac{1}{2t}\|x - z_2\|_2^2 + h(z_2)\right)$

$- \frac{1}{2t}\alpha(1-\alpha)\|z_1 - z_2\|_2^2$

$< \alpha f(z_1) + (1-\alpha) f(z_2)$ t do z_1, z_2 là 2 điểm phân biệt

$\Rightarrow f(z)$ là hàm lồi chặt

$\Rightarrow \arg\min_z f(z)$ chỉ có ~~dy~~ duy nhất 1 nghiệm

$\Rightarrow prox_{h,t}$ là hàm well-defined

b)

Ta có

$prox_{h,t}(x) = \arg\min_z \frac{1}{2t}\|x - z\|_2^2 + h(z) - u$

$\Leftrightarrow \frac{1}{2t}\|x - u\|_2^2 + h(u) = \min_y \frac{1}{2t}\|x - y\|_2^2 + h(y) \ \forall y$

$\Leftrightarrow \frac{1}{2t}\|x - u\|_2^2 + h(u) = \min_y \frac{1}{2t}\|x - y\|_2^2 + h(y)$

$\Leftrightarrow 0 \in \{-\frac{1}{t}(x - u)\} + \partial h(u)$

$\Leftrightarrow \frac{1}{t}(x - u) \in \partial h(u)$

$\Leftrightarrow h(y) \ge h(u) + \cancel{\frac{1}{t}(x-u)}\ \frac{1}{t}(x - u)^T(y - u) \ \forall y$

c)

Tính x^+:

$x^+ = prox_{h,t}(x) = \arg\min_z \left(\frac{1}{2t}\|x - z\|_2^2 + h(z)\right)$

$$= \arg\min_z \left(\frac{1}{2t} \|x - z\|_2^2 + \frac{1}{2} z^T A z - b^T z \right)$$

$$= \left(\frac{1}{t} I + A \right)^{-1} \left(\frac{1}{t} x + b \right)$$

(c) Chứng minh $x^+ = x + (A + \varepsilon I)^{-1} (b - Ax)$

$$x^+ = \left(\frac{1}{t} I + \frac{1}{2} (A + A^T) \right)^{-1} \left(\frac{1}{t} x + b \right)$$

$$= \left(\frac{1}{t} I + A \right)^{-1} \left(\frac{1}{t} x + Ax + b - Ax \right)$$

$$= \left(\frac{1}{t} I + A \right)^{-1} . \left(\frac{1}{t} I + A \right) . x + \left(\frac{1}{t} I + A \right)^{-1} . (b - Ax)$$

$$= x + \left(\frac{1}{t} I + A \right)^{-1} . (b - Ax)$$

Đặt $\varepsilon = \frac{1}{t}$ $(\varepsilon > 0)$

$$\Rightarrow x^+ = x + (A + \varepsilon I)^{-1} (b - Ax)$$

d)

Đặt $\text{prox}_{h,t}(X) = M$

$\Rightarrow M \in X - t\partial \|M\|_{tr}$

$\Rightarrow M \in U \Sigma V^T - t \{ UV^T + W : \|W\|_{op} \leq 1, U^T W = 0, WV = 0 \}$

$\Rightarrow M = U \Sigma_t V^T$ với $(\Sigma_t)_{ii} = \max \{ \Sigma_{ii} - t, 0 \}$

$\Rightarrow X_{ij} = \max \{ X_{ij}, 0 \}$

.... /........ / ....

III) PROBLEM 3 : Group lasso logistic regression

a)

Ta có

$$\text{prox}_{h,t}(\beta) = \underset{z}{\text{argmin}} \ \frac{1}{2t}\|B - z\|_2^2 + \lambda \sum_{j=1}^{J} w_j \|z_{(j)}\|_2$$

$$\text{prox}_{h,t}(\beta)_{(j)} = \text{prox}_{tw_j\lambda\|\cdot\|_2}(\beta_{(j)})$$

$$= \beta_{(j)} - tw_j\lambda \cdot \text{proj}_{B\|\cdot\|_2}\left(\frac{B}{tw_j\lambda}\right)_{(j)}$$

$$= \begin{cases} B_{(j)} - tw_j\lambda \cdot \dfrac{\frac{B_{(j)}}{tw_j\lambda}}{\left\|\frac{B_{(j)}}{tw_j\lambda}\right\|_2} & \text{với } \left\|\frac{B_{(j)}}{tw_j\lambda}\right\|_2 > 1 \\ B_{(j)} - tw_j\lambda \cdot \dfrac{B_{(j)}}{tw_j\lambda} & \text{với } \left\|\frac{B_{(j)}}{tw_j\lambda}\right\|_2 < 1 \end{cases}$$

$$= \cancel{\max\left(0, 1 - \frac{tw_j\lambda}{\|\beta_{(j)}\|_2}\right)}$$

$$= \max\left(0, 1 - \frac{tw_j\lambda}{\|B_{(j)}\|_2}\right) B_{(j)}$$

........./........./......................

b)

$$(\nabla g(\beta))_k = \left[\sum_{i=1}^{n} x_{ik}\left(\frac{e^{(X\beta)_i}}{1+e^{(X\beta)_i}} - y_i\right)\right] \text{ với } k = 1,2 \ldots p$$

$$\Rightarrow \nabla g(\beta) = \left(\frac{e^{(X\beta)}}{1+e^{(X\beta)}} - y\right)^T X$$

$$x^+_{(j)} = \operatorname{prox}_{h,t}(x - t\nabla g(x))_{(j)}$$

$$= \max\left(0, 1 - \frac{t w_j \lambda}{\|x - t\nabla g(x))_{(j)}\|_2}\right)(x - t\nabla g(x))_{(j)}$$