Name: Uzair Akbar ID: 03697290

Convex Optimization for ML & CV

Weekly Exercises 3 – Subdifferential

Exercise 1

We have to show that for $p \in \partial J(u)$, for $J(u)$ differentiable at $u \in \mathrm{int}(\mathrm{dom}(J))$ (i.e. $\nabla J(u)$ exists), then $p = \nabla J(u)$.

The definition of subdifferential is given as:

$$\partial J(u) := \{ p \in \mathbb{R}^n \mid J(v) \geq J(u) + \langle p, v-u \rangle, \ \forall v \in \mathrm{dom}(J) \}$$

For any $u \in \mathrm{int}(\mathrm{dom}(J))$, since interior is open

$$\Rightarrow \exists \ r > \varepsilon > 0, \quad u \pm \varepsilon w \in B(u, r)$$

for radius r of open ball $B(u,r)$ with center u such that $B(u,r) \subset \mathrm{int}(\mathrm{dom}(J)) \subset \mathrm{dom}(J)$.

$$\therefore u \pm \varepsilon w \in \mathrm{dom}(J), \quad \forall w \in \mathbb{R}^n$$

Therefore, the definition of the subdifferential implies that

$$J(u \pm \varepsilon w) \geq J(u) \pm \varepsilon \langle p, w \rangle$$

[Substituting $v = u \pm \varepsilon w$ in the definition stated above.]

$$\Rightarrow J(u) - J(u-\varepsilon w) \leq \varepsilon \langle p, w \rangle \leq J(u-\varepsilon w) - J(u)$$

$$\Rightarrow \frac{J(u) - J(u-\varepsilon w)}{\varepsilon} \leq \langle p, w \rangle \leq \frac{J(u-\varepsilon w) - J(u)}{\varepsilon}$$

Since this holds for all $0 < \varepsilon < r$, it should also hold for $\lim \varepsilon \to 0$:

$$\Rightarrow \lim_{\varepsilon \to 0} \frac{J(u) - J(u-\varepsilon w)}{\varepsilon} \leq \langle p, w \rangle \leq \lim_{\varepsilon \to 0} \frac{J(u-\varepsilon w) - J(u)}{\varepsilon}$$

Now, because

$$\because \partial_w J(u) := \lim_{\varepsilon \to 0} \frac{J(u+\varepsilon w) - J(u)}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{J(u) - J(u-\varepsilon w)}{\varepsilon}$$

$$= \langle \nabla J(u), w \rangle \qquad [\text{Given}]$$

Therefore:

$$\Rightarrow \langle \nabla J(u), w \rangle \leq \langle p, w \rangle \leq \langle \nabla J(u), w \rangle$$

$$\therefore \langle \nabla J(u), w \rangle = \langle p, w \rangle, \quad \forall\, w \in \mathbb{R}^n$$

[Sandwich/Squeeze theorem]

$$\Rightarrow \langle \nabla J(u) - p, w \rangle = 0, \quad \forall\, w \in \mathbb{R}^n$$

$$\Rightarrow \nabla J(u) - p = \underline{0} \qquad [\underline{0} \text{ denotes zero vector}]$$

$$\Rightarrow \nabla J(u) = p$$

But $p \in \partial J(u)$. [initial assumption]

Hence, $\partial J(u) = \{\nabla J(u)\}$ proved.

Exercise 2

Part A

We have to show that

$$\langle p, x \rangle = \|x\|, \ \|p\|_* \leq 1 \iff p \in \partial \|\cdot\|(x)$$

for $p \in E$ with norm $\|\cdot\|$ and dual norm $\|p\|_* := \sup_{\|x\| \leq 1} \langle p, x \rangle$.

Proof $\Rightarrow$

Assuming that $p \in E$ such that

$$\langle p, x \rangle = \|x\|, \ \|p\|_* \leq 1$$

$$\Rightarrow \langle p, y - x \rangle + \|x\| = \langle p, y \rangle - \langle p, x \rangle + \|x\|$$

$= \langle p, y \rangle$ [$\because \langle p, x \rangle = \|x\|$ assumption]

$\leq \|p\|_* \|y\|$ [Cauchy-Schwarz given $y \neq 0$]

$\leq \|y\|$ [$\because \|p\|_* \leq 1$ assumed]

$\therefore \langle p, y - x \rangle + \|x\| \leq \|y\|$, which is the definition of subdifferential. Hence, $p \in \partial \|\cdot\|(x)$ proved.

Scanned by CamScanner

④

Proof $\Leftarrow$

Assume $p \in \partial\|\cdot\|(x)$, then

$\Rightarrow \langle p, y-x\rangle + \|x\| \leq \|y\|$, $\forall y \in \mathbb{E}$ [Definition of Subdifferential]

$\Rightarrow \langle p, y\rangle - \langle p, x\rangle + \|x\| \leq \|y\|$

$\Rightarrow \langle p, y\rangle - \|y\| \leq \langle p, x\rangle - \|x\|$, $\forall y \in \mathbb{E}$

$\Rightarrow \sup_y \left(\langle p, y\rangle - \|y\|\right) \leq \langle p, x\rangle - \|x\|$ [$\because$ the inequality holds for all $y \in \mathbb{E}$]

Now, for $\|p\|_* \leq 1$

$\because \langle p, y\rangle - \|y\| \leq \|y\|(\|p\|_* - 1)$ [Cauchy-Schwarz]

$\Rightarrow \langle p, y\rangle - \|y\| \leq 0$, $\forall y \in \mathbb{E}$ [$\because \|p\|_* \leq 1$]

$\therefore \sup_y \left(\langle p, y\rangle - \|y\|\right) = 0$

Hence,

$$0 \leq \langle p, x\rangle - \|x\|$$

$$\Rightarrow \quad 0 \leq \langle p, x\rangle - \|x\| \leq (\|p_*\| - 1)\|x\|$$ [Cauchy-Schwarz]

$$\leq 0$$ [$\because \|p\|_* \leq 1$]

$\Rightarrow \langle p, x\rangle - \|x\| = 0$ [Sandwich theorem]

$\therefore \langle p, x\rangle = \|x\|$ for $\|p\|_* \leq 1$, $p \in \mathbb{E}$

Hence, proved.

Also note that for $\|p\|_* > 1$, $\sup_y\{\langle p, y\rangle - \|y\|\}$ is unbounded/undefined.

Part B

First we derive the duals of $\|u\|_1$, $\|u\|_2$ and $\|u\|_\infty$.

$\|u\|_1$ dual:

$$\|u\|_* := \sup_{\|x\|_1 \leq 1} \langle u, x \rangle \quad , \; p, u \in \mathbb{R}^n$$

But
$$\begin{aligned}
\langle u, x \rangle &= \sum_{i=1}^n u_i x_i \\
&\leq \left| \sum_{i=1}^n u_i x_i \right| \\
&\leq \sum_{i=1}^n |u_i x_i| && \text{[Triangle inequality]} \\
&= \sum_{i=1}^n |u_i|\,|x_i| && \text{[Multiplicativity: } |ab| = |a|\cdot|b| \text{]} \\
&\leq \sum_{i=1}^n \max_j \{|u_j|\}\, |x_i| \\
&= \max_j \{|u_j|\} \sum_{i=1}^n |x_i| \\
&\leq \max_j \{|u_j|\} && \left[\because \|x\|_1 = \sum_{i=1}^n |x_i| \leq 1 \text{ assumed.}\right] \\
&= \|u\|_\infty
\end{aligned}$$

$\therefore \langle u, x \rangle \leq \|u\|_\infty$ for $\|x\|_1 \leq 1$

$\Rightarrow \sup_{\|x\|_1 \leq 1} \langle u, x \rangle = \|u\|_\infty$

$\therefore \|u\|_* = \|u\|_\infty$

⑥

$\|u\|_2$ dual:

$$\|u\|_* := \sup_{\|x\|_2 \leq 1} \langle u, x \rangle \quad , p, u \in \mathbb{R}^n$$

But $\langle u, x \rangle = \|u\|_2 \cdot \|x\|_2 \cdot \cos\theta$

$$\leq \|u\|_2 \cdot \|x\|_2 \quad \cancel{[\text{for } u \parallel x]} \ [\because \cos\theta \leq 1]$$

$$\leq \|u\|_2 \quad [\because \|x\|_2 \leq 1 \text{ assumed}]$$

$$\therefore \langle u, x \rangle \leq \|u\|_2 \quad \text{for } \|x\|_2 \leq 1$$

$$\Rightarrow \sup_{\|x\|_2 \leq 1} \langle u, x \rangle = \|u\|_2$$

$$\therefore \|u\|_* = \|u\|_2$$

$\|u\|_\infty$ dual:

We can show that $\|u\|_* = \|u\|_1$ similar to before:

$$\because \langle u, x \rangle = \sum_{i=1}^{n} u_i x_i \leq \sum_{i=1}^{n} |u_i| \cdot |x_i|$$

$$\leq \max_j \{|x_j|\} \sum_{i=1}^{n} |u_i|$$

$$= \|x\|_\infty \sum_{i=1}^{n} |u_i|$$

$$\leq \sum_{i=1}^{n} |u_i| \quad [\because \|x\|_\infty \leq 1]$$

$$= \|u\|_1 \quad \text{for } \|x\|_\infty \leq 1$$

$$\Rightarrow \sup_{\|x\|_\infty \leq 1} \langle u, x \rangle = \|u\|_1$$

$$\therefore \|u\|_* = \|u\|_1$$

Now we derive the subdifferentials for $\|u\|_i$, $i \in \{1, 2, \infty\}$ using their respective duals.

$\|u\|_1$ subdifferential:

$$\partial\|\cdot\|_1(u) = \{p \in \mathbb{R}^n \mid \|p\|_\infty \leq 1, \langle p, u\rangle = \|u\|_1\} \quad [\because \|p\|_* = \|p\|_\infty]$$

$$= \left\{ p \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} |p_i| \leq 1 & , u_i = 0 \\ p_i = +1 & , u_i > 0 \\ p_i = -1 & , u_i < 0 \end{array} \right\}$$

$\|u\|_2$ subdifferential:

$$\partial\|\cdot\|_2(u) = \{p \in \mathbb{R}^n \mid \|p\|_2 \leq 1, \langle p, u\rangle = \|u\|_2\} \quad [\because \|p\|_* = \|p\|_2]$$

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

[$\underline{0}$ denotes zero vector
$\hat{u}$ is unit vector along u.]

$\|u\|_\infty$ subdifferential:

$$\partial\|\cdot\|_\infty(u) = \{p \in \mathbb{R}^n \mid \|p\|_1 \leq 1, \langle p, u\rangle = \|u\|_\infty\} \quad [\because \|p\|_* = \|p\|_1]$$

Exercise 3

By definition of norm-cone:

$N_C(u) := \{p \in \mathbb{R}^n \mid \langle p, v-u \rangle \leq 0, \ \forall v \in C\}, u \in C.$

Given that $C = \{u \in \mathbb{R}^n \mid Au \leq b\}$ is a bounded/closed polyhedra, we define

$C' := \{A^T \lambda \mid \lambda \geq 0, \ \lambda_i = 0 \text{ if } (Au-b)_i < 0\}$

Now we need to show that

$$p \in C' \iff p \in N_C(u)$$

Proof $\Rightarrow$

Let $p = A^T\lambda \in C'$, then

$$\langle p, v-u \rangle = p^T(v-u)$$
$$= \lambda^T A(v-u) \quad \text{— (I)}$$

Now, for the i^{th} element in vector $A(v-u)$ in (I):

if $\underbrace{(Au-b)_i = 0}_{\text{from definition of } C'}$ when $\lambda_i > 0 \Rightarrow \lambda_i (A(v-u))_i$

$\Rightarrow (Au)_i = b_i$

$$= \lambda_i (Av_i - b_i)$$
$$\leq 0 \qquad [\because Av_i \leq b_i \text{ for } v \in C]$$

And $\underbrace{\lambda_i = 0 \text{ otherwise}}_{\text{definition of } C'} \Rightarrow \lambda_i (A(v-u))_i = 0$

Therefore (I) becomes:

$$\langle p, v-u \rangle = \lambda^T A(v-u)$$

$$\leq 0 \qquad \forall v \in C$$

and hence $p \in N_c(u)$ [$\langle p, v-u \rangle \leq 0$ is the definition of norm-cone]

Proof $\Leftarrow$