Convex and Nonsmooth Optimization: Homework 7
Muzhi Allyn Xu
March 2020

Problem 1

For fixed $x \in \mathbb{R}^n$, define the sets

$$I = \{i \in [n] : x_i = f(x)\}, \quad J = [n] \setminus I = \{j \in [n] : x_j < f(x)\}$$

i.e. I are the coordinates which maximize $\max_{1 \leq i \leq n} f(x)$ (here $[n] = \{1, ..., n\}$). We note that it follows that $f(x) = \max_{i \in I} x_i = x_k$ for any $k \in I$ and that $f(x) > \max_{j \in J} x_j$. We claim that the general formula for subgradients $\partial f(x)$ is

$$\partial f(x) = \{y \in \mathbb{R}^n : y_i \geq 0 \ \forall i \in I, \ y_j = 0 \ \forall j \in J, \ \sum_{i \in I} y_i = 1\} = \operatorname{conv}\{e_i \mid i \in I\}$$

which we will prove as follows. (The equality of $\operatorname{conv}\{e_i \mid i \in I\}$ follows from the definition of convex combination.)

We first note that since f convex, we can relax the requirement for y to be a subgradient; namely rather than requiring $f(x+z) \geq f(x) + y^T z$ hold for all $z \in \mathbb{R}^n$, we can require it to hold only for sufficiently small z. More precisely, $y \in \partial f(x)$ if and only if $\exists \, \delta > 0$ s.t. $f(x+z) \geq f(x) + y^T z$ for all $||z||_\infty \leq \delta$. This is because if such a δ exists, then by convexity $f(x+z)$ must be greater than $f(x) + y^T z$ for all $z \in \mathbb{R}^n$ (or else we could linearly interpolate back to the point $(x, f(x))$ and find violating points arbitrarily close to x).

In light of the above, we define $\delta = \frac{1}{2}(f(x) - \max_{j \in J} x_j)$ so that $f(x+z)$ is always maximized by some coordinate $k \in I$ whenever $||z||_\infty \leq \delta$. We will prove the subgradient condition holds for this specific δ, i.e. for all z with $||z||_\infty \leq \delta$. For such a z we have $f(x+z) = \max_{i \in I}(x_i + z_i) = f(x) + \max_{i \in I} z_i$ and so the subgradient condition is further simplified to simply requiring $\max_{i \in I} z_i \geq y^T z$.

We now prove necessity of each condition of $y \in \partial f(x)$. If $\max_{i \in I} z_i \geq y^T z$ for every $||z||_\infty \leq \delta$, then we can also replace z with $-z$. This results in the following necessary inequalities for $y \in \partial f(x)$:

$$\min_{i \in I} z_i \leq y^T z \leq \max_{i \in I} z_i$$

for every z with $||z||_\infty \leq \delta$. Now taking $z = \delta e_j$ for $j \in J$, we find immediately find that $0 \leq \delta y_j \leq 0$. Thus $y_j = 0$ for every $j \in J$. We can also take $z = \delta e_i$ for which we have $0 \leq \delta y_i \leq \delta$. Thus $y_i \geq 0$ for every $i \in I$. Lastly, taking $z = \sum_{i=1}^n \delta e_i$, we have $\delta \leq \delta \sum_{i=1}^n y \leq \delta$. Thus $\sum_{i \in I} y_i = 1$ since $y_j = 0$ for $j \in I$. Hence we see that all the conditions for $y \in \partial f(x)$ above are necessary.

Conversely, if $y \in \mathbb{R}^n$ satisfies $y_i \geq 0$ for $i \in I$, $y_j = 0$ for $j \in J$ and $\sum_{i \in I} y_i = 1$, then for any $||z||_\infty \leq \delta$:

$$y^T z = \sum_{i \in I} y_i z_i \leq \max_{i \in I} z_i \sum_{i \in I} y_i = \max_{i \in I} z_i$$

which demonstrates that $y \in \partial f(x)$. This completes our proof.

Problem 2

We have for any $x \in \mathbb{R}^n$,

$$f^*(y) = \sup_{x \in \mathbb{R}^n} (x^T y - f(x)) \geq x^T y - f(x)$$

Rearranging gives the Fenchel-Young inequality $f(x) + f^*(y) \geq x^T y$. Equality holds if and only if $y \in \partial f(x)$ because,

$$\begin{aligned}
f^*(y) + f(x) = x^T y &\iff \sup_{z \in \mathbb{R}^n} (z^T y - f(z)) + f(x) = x^T y \\
&\iff \sup_{z \in \mathbb{R}^n} ((x+z)^T y - f(x+z)) + f(x) = x^T y \\
&\iff \sup_{z \in \mathbb{R}^n} (y^T z - f(x+z) + f(x)) = 0 \\
&\iff y^T z - f(x+z) + f(x) \leq 0 \text{ for all } z \in \mathbb{R}^n \\
&\iff f(x+z) \geq f(x) + y^T z \text{ for all } z \in \mathbb{R}^n \\
&\iff y \in \partial f(x)
\end{aligned}$$

Note that in the fourth equivalence where we convert the supremum equaling zero into an inequality, we are using the fact that we can take $z = 0$ so the supremum is bounded below by zero.

Problem 3

Suppose $y \in \partial f(x)$. Then $f(x+z) \geq f(x) + y^T z$ for $z \in \mathbb{R}^n$. Let $d \in \mathbb{R}^n$ be any direction. Then,

$$f'(x; d) = \lim_{t \searrow 0} \frac{f(x+td) - f(x)}{t} \geq \lim_{t \searrow 0} \frac{t \cdot y^T d}{t} = y^T d$$

Conversely, suppose $y \notin \partial f(x)$. Then there exists $d \in \mathbb{R}^n$ such that $f(x+d) < f(x) + y^T d$. Note that $d \neq 0$ or else we cannot have strict inequality. Since f is convex, then for any $0 < t < 1$, we have,

$$f(x+td) - f(x) = f(t(x+d) + (1-t)x) - f(x) \leq t f(x+d) + (1-t) f(x) - f(x) = t(f(x+d) - f(x))$$

Dividing the above by t and taking the limit as $t \searrow 0$, we find

$$f'(x; d) = f(x+d) - f(x) < y^T d.$$