Homework 5 Solutions

Due May 8th, 2019 by noon

This homework is not graded. The purpose of this homework assignment is to help you train for the midterm exam.

"Reading" Assignment

Review the mathematical derivations covered in classes. Do not passively read the handouts.

Do it yourself, several times, without notes

Exercise 1

In this exercise, you will compute the subdifferentials of several functions.

- $f(x) = \|x\|_1^2$

  Assume that $x \in \mathbb{R}^d$. Observe that

$$\begin{aligned} f(x) &= \left(\sum_{i=1}^{d} |x_i|\right)^2 \\ &= \sum_{i=1}^{d} |x_i|^2 + \sum_{i,j:i\neq j} |x_i||x_j| \\ &= \sum_{i=1}^{d} x_i^2 + 2\sum_{j>i}\sum_{i=1}^{d-1} |x_i||x_j|. \end{aligned}$$

  Define $g(x) = \sum_{i=1}^{d} x_i^2$ and $h_{ij}(x) = |x_i||x_j|$. Then we can write

$$f(x) = g(x) + 2\sum_{j>i}\sum_{i=1}^{d-1} h_{ij}(x).$$

  From properties of subdifferentials we know that

$$\partial f(x) = \partial g(x) + 2\sum_{j>i}\sum_{i=1}^{d-1} \partial h_{ij}(x).$$

Moreover, we know that $\partial g(x) = \nabla g(x) = 2x$. Next, observe that $\partial h_{ij}(x)$ is zero in every coordinate except the ith and jth coordinates. The ith coordinate (when $i \neq j$, which is the case we care about) is given by

$$\begin{aligned}
[\partial h_{ij}(x)]_i &= |x_j|\partial|x_i| \\
&= \begin{cases} |x_j|, & \text{if } x_i > 0 \\ -|x_j|, & \text{if } x_i < 0 \\ [-|x_j|, |x_j|], & \text{if } x_i = 0 \end{cases}
\end{aligned}$$

and similarly for the jth coordinate.

Adding everything up, we find

$$\begin{aligned}
\partial f(x) &= \left\{ v \in \mathbb{R}^d : \text{For all } i = 1, \dots, d,\ v_i \in \begin{cases} \{2x_i + 2\sum_{j \neq i} \operatorname{sign}(x_i)|x_j|\} & \text{if } x_i \neq 0 \\ 2x_i + 2\sum_{j \neq i}[-|x_j|, |x_j|] & \text{if } x_i = 0 \end{cases} \right\} \\
&= \left\{ v \in \mathbb{R}^d : \text{For all } i = 1, \dots, d,\ v_i \in \begin{cases} \{2\sum_{j=1}^d \operatorname{sign}(x_i)|x_j|\} & \text{if } x_i \neq 0 \\ 2\sum_{j=1}^d[-|x_j|, |x_j|] & \text{if } x_i = 0 \end{cases} \right\}
\end{aligned}$$

- $f(x) = \|x\|_2^1$

  I claim that

$$\partial f(x) = \begin{cases} \frac{x}{\|x\|_2}, x \neq 0 \\ B_2,\ x = 0, \end{cases}$$

  where $B_2 = \{v \in \mathbb{R}^d : \|v\|_2 \leq 1\}$.

  To see this, assume $x \in \mathbb{R}^d$ and note that

$$f(x) = \sqrt{\sum_{i=1}^d x_i^2}.$$

  First let's consider the case where $x \neq 0$. In this case, the function is differentiable at x and the partial derivatives are, for all $i = 1, \dots, d$,

$$\begin{aligned}
\frac{\partial f}{\partial x_i} &= \frac{1}{2}\left(\sum_{i=1}^d x_i^2\right)^{-1/2} (2x_i) \\
&= \frac{x_i}{\sqrt{\sum_{i=1}^n x_i^2}} \\
&= \frac{x_i}{\|x\|_2}.
\end{aligned}$$

Hence, we have that for $x \neq 0$,

$$\partial f(x) = \nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_d} \end{bmatrix} = \frac{1}{\|x\|_2} \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} = \frac{x}{\|x\|_2}.$$

Now consider the case where $x = 0$. Then by definition of a subgradient, each subgradient must satisfy the following for all $y \in \mathbb{R}^d$:

$$\begin{aligned} & f(y) \geq f(0) + g^T(y - 0) \\ \iff & \|y\|_2 \geq g^T y. \end{aligned}$$

First suppose $\|g\|_2 > 1$. Then taking $y = g$, we find that $\|y\|_2 \geq \|y\|_2^2$, contradicting the fact that $\|g\|_2 = \|y\|_2 > 1$. Hence, any g with $\|g\|_2 > 1$ cannot be a subgradient of f at 0.

Next, suppose $\|g\|_2 \leq 1$. Then by the Cauchy-Schwarz inequality we have that

$$g^T y \leq \|g\|_2 \|y\|_2 \leq \|y\|_2.$$

Therefore, any g such that $\|g\|_2 \leq 1$ is a subgradient of f at 0. We will denote the set of all such vectors by B_2.

Hence,

$$\partial f(x) = \begin{cases} \frac{x}{\|x\|_2}, x \neq 0 \\ B_2, \ x = 0. \end{cases}$$

- $f(x) = \left(\max(0, x^T\beta)\right)^4$
  This function is differentiable, so the subdifferential is just the gradient:

$$\partial f(x) = \nabla f(x) = 4\beta \left(\max(0, x^T\beta)\right)^3.$$