EE364b Spring 2020 Homework 2

Due Friday 4/24 at 11:59pm via Gradescope

2.1 (8 points, 1 point per question) Let f be a convex function with domain in $\mathbf{R}^n$. We fix $x \in \mathbf{int\ dom}\ f$ and $d \in \mathbf{R}^n$. Recall the definition of the directional derivative of f at x along the direction d

$$f'(x,d) = \lim_{t \downarrow 0} \frac{f(x+td) - f(x)}{t}\,.$$

In this question, we aim to show that $f'(x,d)$ exists and is finite, and that we have the following relationship between $\partial f(x)$ and $f'(x,d)$,

$$f'(x,d) = \sup_{g \in \partial f(x)} g^T d\,.$$

(a) Show that the ratio $\frac{f(x+td)-f(x)}{t}$ is a non-decreasing function of $t > 0$. Deduce that $f'(x,d)$ exists, and is either finite or equal to $-\infty$.

We know from the lectures that, since $x \in \mathbf{int\ dom}\ f$, the subdifferential set $\partial f(x)$ is non-empty, convex and compact.

(b) Let $g \in \partial f(x)$. Show that $f'(x,d) \geq g^T d$. Deduce that $f'(x,d)$ is finite and that $f'(x,d) \geq \sup_{g \in \partial f(x)} g^T d$.

In the remaining part of this question, we will establish the converse inequality $f'(x,d) \leq \sup_{g \in \partial f(x)} g^T d$, by showing the existence of a subgradient $g^* \in \partial f(x)$ such that $f'(x,d) \leq {g^*}^T d$. We introduce the two following sets

$$\begin{aligned} C_1 &= \{(z,t) \mid z \in \mathbf{dom}\ f,\ f(z) < t\} \\ C_2 &= \{(y,v) \mid y = x + \alpha d,\ v = f(x) + \alpha f'(x,d),\ \alpha \geq 0\}\ . \end{aligned}$$

(c) Prove that C_1 and C_2 are non-empty, convex and disjoint.

(d) Justify why there exists a nonzero vector $(a, \beta) \in \mathbf{R}^n \times \mathbf{R}$ such that

$$a^T(x + \alpha d) + \beta(f(x) + \alpha f'(x,d)) \leq a^T z + \beta w\,, \tag{1}$$

for all $\alpha \geq 0$, $z \in \mathbf{dom}\ f$ and $f(z) < w$.

(e) Prove that β must be strictly positive. Define $\tilde{a} = \frac{a}{\beta}$. Show that

$$\tilde{a}^T(x + \alpha d) + f(x) + \alpha f'(x,d) \leq \tilde{a}^T z + w \tag{2}$$

for all $\alpha \geq 0$, $z \in \mathbf{dom}\ f$ and $f(z) < w$.

(f) Prove that $-\tilde{a} \in \partial f(x)$.

(g) Prove that $-\tilde{a}^T d \geq f'(x,d)$.

We illustrate the above result with an example.

(h) Let $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$, $\lambda > 0$, and fix a direction $d \in \mathbf{R}^n$. Consider the function $f(x) = \frac{1}{2}\|Ax-b\|_2^2 + \lambda\|x\|_1$. Compute $f'(0,d)$. *Remark: you can either compute it directly by using the definition of the directional derivative, or, use the variational formula* $f'(0,d) = \sup_{g\in\partial f(0)} g^T d$.

2.2 (4 Points) In this question, we will show that a subgradient of the function $h(x) = \min_{z\in C}\|x-z\|_2$ is

$$g = \frac{x-z^*}{\|x-z^*\|_2},$$

where C is a compact convex set in $\mathbf{R}^n$, x is a given point in $\mathbf{R}^n$ which does not belong to C and $z^* = P_C(x) := \operatorname{argmin}_{z\in C}\|x-z\|_2$ denotes the Euclidean projection of x onto C (which exists and is unique).

(a) (0.5 point) Use the fact that $\|x-z\|_2 = \max_{u:\|u\|_2\leq 1} u^T(x-z)$ to transform the minimization problem $h(x) = \min_{z\in C}\|x-z\|_2$ into the following saddle point problem

$$\min_{z\in C}\max_{u:\|u\|_2\leq 1} u^T(x-z)\,.$$

(b) (2 points) Now, we will use (a simple version of) the Sion's minimax theorem, which can be stated as follows.

Let $X \subseteq \mathbf{R}^n$ and $Y \subseteq \mathbf{R}^n$ be compact and convex sets. Let f be a real valued function on $X \times Y$ such that

- *$f(x,\cdot)$ is continuous and concave on Y, $\forall x \in X$*
- *$f(\cdot,y)$ is continuous and convex on X, $\forall y \in Y$*

Then, we have

$$\min_{x\in X}\max_{y\in Y} f(x,y) = \max_{y\in Y}\min_{x\in X} f(x,y)\,.$$

Further, there exists a (saddle) point $(x^*,y^*) \in X \times Y$ such that

$$f(x^*,y^*) = \min_{x\in X} f(x,y^*) = \max_{y\in Y} f(x^*,y) = \min_{x\in X}\max_{y\in Y} f(x,y) = \max_{y\in Y}\min_{x\in X} f(x,y)\,.$$

Apply Sion's minimax theorem to conclude that

$$\min_{z\in C}\max_{u:\|u\|_2\leq 1} u^T(x-z) = \max_{u:\|u\|_2\leq 1}\min_{z\in C} u^T(x-z)\,.$$

Define $u^* = \frac{x-z^*}{\|x-z^*\|_2}$. Show that (z^*,u^*) is a saddle point of the above minimax problem.

(c) (1.5 points) Using the 'max-min' representation of $h(x)$, compute a subgradient of h at x.

2.3 (4 points) *For this question, you need to submit your code in addition to any description of your algorithm.* Let Σ be an $n \times n$ diagonal matrix with diagonal entries $\sigma_1 \geq \cdots \geq \sigma_n > 0$, and y a given vector in $\mathbf{R}^n$. Consider the compact convex sets $\mathcal{E} = \{z \in \mathbf{R}^n \mid \|\Sigma^{\frac{1}{2}} z\|_2 \leq 1\}$ and $B = \{z \in \mathbf{R}^n \mid \|z - y\|_\infty \leq 1\}$.

(a) (2 points) Formulate an optimization problem and propose an algorithm in order to find a point $x \in \mathcal{E} \cap B$. *You can assume that $\mathcal{E} \cap B$ is not empty.* Your algorithm must be provably converging (although you do not need to prove it and you can simply refer to the lectures' slides).

(b) (2 points) Implement your algorithm with the following data: $n = 2$, $y = (7/4, 0)$, $\sigma_1 = 1$, $\sigma_2 = 0.5$ and $x = (0, 4)$. Plot the objective value of your optimization problem versus the number of iterations.

2.4 *(4 points)* Consider the optimization problem

$$\text{minimize} \quad \left\{ f(x_1, \ldots, x_J) := \tfrac{1}{2}\|b - \textstyle\sum_{j=1}^J A_j x_j\|_2^2 + \lambda \cdot \textstyle\sum_{j=1}^J \|x_j\|_2 \right\},$$

with variable $x_1, \ldots, x_J \in \mathbf{R}^n$ and problem data $A_1, \ldots, A_J \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$ and $\lambda > 0$. We will apply the subgradient method.

(a) (2 points) Show that the subgradient method with Polyak's step length updates the current point to a point at which the first order (linear) approximation has value f^* (optimal value).

(b) (2 points) Let $J = 15$, $n = 10$, $m = 200$ and $\lambda = 1$. Generate random matrices $A_1, \ldots, A_J \in \mathbf{R}^{m \times n}$ with independent Gaussian entries with mean 0 and variance $1/m$, and, random vectors $x_1, \ldots, x_J \in \mathbf{R}^n$ with independent Gaussian with mean 0 and variance $1/n$, then set $b = \sum_{j=1}^J A x_j$. Plot convergence in terms of the objective $f(x_1^{(k)}, \ldots, x_J^{(k)})$. Try different step length schedules, including Polyak's step length.