AMATH 515: Homework 6
Sid Meka

1. Let $f'(x; v)$ for $x, v \in \mathbb{R}^n$ denote the directional derivative of a function $f : \mathbb{R}^n \to \mathbb{R}$ at x and in the direction of v defined as
$$f'(x; v) := \lim_{\alpha \downarrow 0} \frac{f(x + \alpha v) - f(x)}{\alpha}.$$
Prove the following statements:

   (a) If f is convex then for $0 < \alpha_1, < \alpha_2$ it holds that
$$\frac{f(x + \alpha_1 v) - f(x)}{\alpha_1} \leq \frac{f(x + \alpha_2 v) - f(x)}{\alpha_2}$$
Proof. Since f is convex, we apply the standard convexity inequality:
$$f(\lambda x_2 + (1 - \lambda)x) \leq \lambda f(x_2) + (1 - \lambda) f(x)$$
and
$$x_1 = \lambda x_2 + (1 - \lambda)x$$
for any $\lambda \in [0, 1]$ and points $x_1 = x + \alpha_1 v$ and $x_2 = x + \alpha_2 v$. Note that $\lambda = \frac{\alpha_1}{\alpha_2}$, so this actually means in this case, $\lambda \in (0, 1)$ as $\frac{\alpha_1}{\alpha_2} \in (0, 1)$ from the way α_1 and α_2 are defined.

   Applying convexity,
$$f(x_1) \leq \lambda f(x_2) + (1 - \lambda) f(x)$$
Rearranging, we get:
$$f(x_1) - f(x) \leq \lambda(f(x_2) - f(x))$$
Dividing both sides by α_1 and using $\lambda = \frac{\alpha_1}{\alpha_2}$, we obtain:
$$\frac{f(x_1) - f(x)}{\alpha_1} \leq \frac{\frac{\alpha_1}{\alpha_2}(f(x_1) - f(x)}{\alpha_1}$$
$$\frac{f(x_1) - f(x)}{\alpha_1} \leq \frac{f(x_1) - f(x)}{\alpha_2}$$
Now substituting $x_1 = x + \alpha_1 v$ and $x_2 = x + \alpha_2 v$ we have
$$\frac{f(x + \alpha_1 v) - f(x)}{\alpha_1} \leq \frac{f(x + \alpha_2 v) - f(x)}{\alpha_2}$$
as desired. □

   (b) Use the previous result to show that the directional derivative, for convex f, can be equivalently defined as
$$f'(x; v) := \inf_{\alpha > 0} \frac{f(x + \alpha v) - f(x)}{\alpha}.$$
Applying the previous part, consider:
$$\frac{f(x + \alpha_1 v) - f(x)}{\alpha_1} \leq \frac{f(x + \alpha_2 v) - f(x)}{\alpha_2} \leq \frac{f(x + \alpha_3 v) - f(x)}{\alpha_3} \leq \ldots \leq \frac{f(x + \alpha_n v) - f(x)}{\alpha_n}$$
where $\alpha_1 < \alpha_2 < \alpha_3 < \ldots < \alpha_n$. Notice how this pattern shows that we get $\frac{f(x+\alpha v)-f(x)}{\alpha}$ as nonincreasing as α decreases. Now, applying this instantaneously as $\alpha \to 0^+$, we get to the Greatest Lower Bound as we get to its limit as $\alpha \to 0^+$, or in other words, the infimum of $\frac{f(x+\alpha v)-f(x)}{\alpha}$ as long as $\alpha > 0$.

(c) Show that the directional derivative is additive, that is, $(f_1 + f_2)'(x; v) = f_1'(x; v) + f_2'(x; v)$.

$$
\begin{gathered}
(f_1 + f_2)'(x; v) = \\
\lim_{\alpha \downarrow 0} \frac{(f_1 + f_2)(x + \alpha v) - (f_1 + f_2)(x)}{\alpha} = \\
\lim_{\alpha \downarrow 0} \frac{f_1(x + \alpha v) + f_2(x + \alpha v) - f_1(x) - f_2(x)}{\alpha} = \\
\lim_{\alpha \downarrow 0} \frac{(f_1(x + \alpha v) - f_1(x)) + (f_2(x + \alpha v) - f_2(x))}{\alpha} = \\
\lim_{\alpha \downarrow 0} \frac{f_1(x + \alpha v) - f_1(x)}{\alpha} + \lim_{\alpha \downarrow 0} \frac{f_2(x + \alpha v) - f_2(x)}{\alpha} = \\
f_1'(x; v) + f_2'(x; v)
\end{gathered}
$$

(d) Show that the directional derivative is homogeneous with respect to the direction v, i.e., $f'(x; \lambda v) = \lambda f'(x; v)$ for all $\lambda > 0$.

$$
\begin{gathered}
f'(x; \lambda v) = \\
\lim_{\alpha \downarrow 0} \frac{f(x + \alpha(\lambda v)) - f(x)}{\alpha} = \\
\lim_{\alpha \downarrow 0} \frac{f(x + (\alpha\lambda) v) - f(x)}{\alpha} = \\
\lim_{\alpha' \downarrow 0} \frac{f(x + \alpha' v) - f(x)}{\alpha'} \cdot \frac{\alpha'}{\alpha} = \qquad \text{(where we define } \alpha' = \alpha\lambda \text{, so } \alpha' \to 0 \text{ as } \alpha \to 0) \\
\lim_{\alpha' \downarrow 0} \frac{f(x + \alpha' v) - f(x)}{\alpha'} \cdot \lambda = \qquad (\text{As } \lambda = \tfrac{\alpha'}{\alpha}) \\
\lambda \lim_{\alpha' \downarrow 0} \frac{f(x + \alpha' v) - f(x)}{\alpha'} = \\
\lambda f'(x; v)
\end{gathered}
$$

2. Recall that a function $x \mapsto \|x\|$ for $x \in \mathbb{R}^n$ is a norm if it satisfies:

(a) $\|x\| = 0$ iff $x = 0$

(b) $\|\lambda x\| = |\lambda| \cdot \|x\|$ for all $\lambda \in \mathbb{R}$

(c) $\|x + y\| \leq \|x\| + \|y\|$. Show for any norm $f(x) = \|x\|$, that f is not differentiable at $x = 0$ but $0 \in \partial f(0)$

Does this imply that 0 is a global minimizer?

We consider the function $f(x) = \|x\|$ and show it is not differentiable at $x = 0$, but that $0 \in \partial f(0)$. If $x = 0$, the directional derivative is:

$$
f'(0; v) = \lim_{\alpha \to 0^+} \frac{\|\alpha v\| - \|0\|}{\alpha}
$$

or more simply,

$$
f'(0; v) = \lim_{\alpha \to 0^+} \frac{\|\alpha v\|}{\alpha}
$$

$$
f'(0; v) = \lim_{\alpha \to 0^+} \frac{\alpha \|v\|}{\alpha}
$$

$$f'(0; v) = \lim_{\alpha \to 0^+} \|v\|$$

The subdifferential of f at $x = 0$ is given by:

$$\partial f(0) = \{g \in \mathbb{R}^n \mid g^T x \leq \|x\|, \forall x \in \mathbb{R}^n\}$$

Since f is a norm, $\partial f(0)$ contains all unit vectors in $\mathbb{R}^n$, meaning $0 \in \partial f(0)$. Thus, 0 is a minimizer because for all x, $\|x\| \geq 0$ and $\|0\| = 0$ meaning 0 is the global minimum.

3. Consider the function $f(x) = \|x\|_1 = \sum_{i=1}^{n} |x_i|$ for $x \in \mathbb{R}^n$. Identify the subdifferential of f.

From Homework 1 Question 2 Part c, we have shown that any norm is convex. We will use that fact when solving this problem. Because $\|x\|_1$ is convex as it is a norm, we can use the subdifferential.

We first start by providing the definition of the subdifferential. The subdifferential of f at x, denoted by $\partial f(x)$, is given by:

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

Now we consider the cases for each component. If $x_i > 0$, $|x_i| = x_i$, so $\frac{\partial f}{\partial x_i} = 1$ and the subdifferential contains only $g_i = 1$. If $x_i < 0$, $|x_i| = -x_i$, so $\frac{\partial f}{\partial x_i} = -1$ and the subdifferential contains only $g_i = -1$. If $x_i = 0$, we actually have a nondifferentiable $|x_i|$; here, the one sided derivatives are ± 1, which suggests that any g_i in the interval $[-1, 1]$ satisfies the subgradient condition.

Thus, we have:

$$\partial f(x) = \left\{ g \in \mathbb{R}^n \mid g_i = \begin{cases} 1 & \text{if } x_i > 0 \\ -1 & \text{if } x_i < 0 \\ s_i & \text{if } x_i = 0 \end{cases} \right\}$$

Note that s_i represents a value in the interval $[-1, 1]$ that is s_i is just a placeholder for a real number in $[-1, 1]$, meaning that for indices where $x_i = 0$, the subdifferential allows a range of possible values for g_i.

4. Consider the constrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad c_i(x) = 0, \quad i = 1, \ldots, m,$$

with continuous $f : \mathbb{R}^n \to \mathbb{R}$ and $c_i : \mathbb{R}^n \to \mathbb{R}$, and let $x^\star$ denote a global solution. Further define the penalized/relaxed approximations

$$\min_{x \in \mathbb{R}^n} f(x) + \frac{1}{2\mu_k} \sum_{i=1}^{m} c_i^2(x)$$

with positive penalty parameters $\mu_k \downarrow 0$ as $k \to \infty$, and denote their global minimizers as x^k. Your goal in this problem is to show that every limit point of the sequence $\{x^k\}_{k=1}^{\infty}$ is a global solution of the constrained problem. Note that the proof will not rely on convexity assumptions and holds for non-convex optimization problems as well.

(a) Use the optimality of $x^\star$ and the x^k to infer that

$$f(x^k) + \frac{1}{2\mu_k} \sum_{i=1}^{m} c_i^2(x^k) \leq f(x^\star),$$

and obtain a bound on $\sum_{i=1}^{m} c_i^2(x^k)$.

Since x^k is a global minimizer of the penalized function, we have:

$$f(x^k) + \frac{1}{2\mu_k}\sum_{i=1}^{m} c_i^2(x^k) \leq f(x^\star) + \frac{1}{2\mu_k}\sum_{i=1}^{m} c_i^2(x^\star)$$

Since $x^\star$ is a global minimizer of the original constrained problem, it satisfies the feasibility conditions:

$$c_i(x^\star) = 0\,, \quad \forall i = 1, 2, \ldots, m$$

Thus, evaluating the penalized objective at $x^\star$, we obtain:

$$f(x^k) + \frac{1}{2\mu_k}\sum_{i=1}^{m} c_i^2(x^k) \leq f(x^\star) + \frac{1}{2\mu_k}\sum_{i=1}^{m} (0^2)$$

Or more simply:

$$f(x^k) + \frac{1}{2\mu_k}\sum_{i=1}^{m} c_i^2(x^k) \leq f(x^\star)$$

To find the bound on $\sum_{i=1}^{m} c_i^2(x^k)$, we just work from:

$$f(x^k) + \frac{1}{2\mu_k}\sum_{i=1}^{m} c_i^2(x^k) \leq f(x^\star)$$

$$\frac{1}{2\mu_k}\sum_{i=1}^{m} c_i^2(x^k) \leq f(x^\star) - f(x^k)$$

$$\sum_{i=1}^{m} c_i^2(x^k) \leq 2\mu_k\left(f(x^\star) - f(x^k)\right)$$

Thus, our explicit upper bound is: $2\mu_k\left(f(x^\star) - f(x^k)\right)$.

(b) Now take $\overline{x}$ to be any limit point of the $\{x^k\}_{k=1}^{\infty}$, that is, $\overline{x}$ is the limit of some convergent subsequence of the x^k; you may assume such a subsequence exists. Show that $c_i(\overline{x}) = 0$ for $i = 1, \ldots, m$ and so $\overline{x}$ is feasible for the original constrained problem.

Let $\overline{x}$ be a limit point of $\{x^k\}$, meaning there exists a subsequence $\{x^{k_j}\}$ such that:

$$x^{k_j} \to \overline{x} \quad \text{as} \quad j \to \infty$$

Since we already established that $c_i(x^k) \to 0$, it follows from continuity of $c_i(x)$ that:

$$c_i(\overline{x}) = \lim_{j\to\infty} c_i(x^{k_j}) = 0 \text{ for such a } k_j$$

Also, remember that we are solving the constrained optimization problem:

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{subject to} \quad c_i(x) = 0\,, \quad i = 1, \ldots, m$$

That means we have:

$$c_i(\overline{x}) = \lim_{j\to\infty} c_i(x^{k_j}) = 0 \quad \forall i = 1, \ldots, m$$

Because $\overline{x}$ satisfies all constraints here, $\overline{x}$ is feasible for the original constrained problem.

(c) Use the non-negativity of the μ_k and the $c_i^2(x^k)$ to infer that $f(\overline{x}) \leq f(x^\star)$. Further use this result to argue that $\overline{x}$ is a global solution.

Recall that we have established

$$f(x^k) + \frac{1}{2\mu_k}\sum_{i=1}^{m} c_i^2(x^k) \leq f(x^\star)$$

Also, since μ_k is nonnegative, it follows that $\frac{1}{2\mu_k}\sum_{i=1}^{m} c_i^2(x^k)$ is nonnegative as $c_i : \mathbb{R}^n \to \mathbb{R}$ meaning that $c_i^2 : \mathbb{R}^n \to \mathbb{R}^+ \cup \{0\}$. Thus, because $\frac{1}{2\mu_k}\sum_{i=1}^{m} c_i^2(x^k)$ is nonnegative,

$$f(x^k) \leq f(x^k) + \frac{1}{2\mu_k}\sum_{i=1}^{m} c_i^2(x^k)$$

Thus,

$$f(x^k) \leq f(x^k) + \frac{1}{2\mu_k}\sum_{i=1}^{m} c_i^2(x^k) \leq f(x^\star)$$

Which means:

$$f(x^k) \leq f(x^\star)$$

We now consider the limit point of x^k, which leads to $\overline{x}$ as $\lim_{j\to\infty} f(x^{k_j}) \leq f(x^\star)$ for some subsequence $\{x^{k_j}\}$. Thus, we combine this knowledge along with the fact that f is continuous with $f : \mathbb{R}^n \to \mathbb{R}$ to state that

$$f(\overline{x}) \leq f(x^\star)$$

Since $x^\star$ is defined as a minimizer, we also have $f(x^\star) \leq f(\overline{x})$. Because both $f(\overline{x}) \leq f(x^\star)$ and $f(x^\star) \leq f(\overline{x})$, it must follow that $f(\overline{x}) = f(x^\star)$. Thus, because $f(\overline{x}) = f(x^\star)$, $\overline{x}$ attains the optimal value, making it a global solution to the original problem.