

Homework 4

Kailong Wang

November 16, 2023

Q1

Grade:

Recall That N_C denotes the normal cone map of the set C . Show that if U is a linear subspace of \mathbb{R}^n , then $N_U(x) = U^\perp$ for all $x \in U$, where U^\perp denotes the subspace orthogonal to U (by definition, $N_U(x) = \emptyset$ if $x \notin U$).

Solution

Proof. To show that $N_U(x) = U^\perp$, we need to show that $N_U(x) \subseteq U^\perp$ and $U^\perp \subseteq N_U(x)$.

- $N_U(x) \subseteq U^\perp$: Let $y \in N_U(x)$, then we have $y^\top(x - u) \leq 0$ for all $u \in U$. Since U is a linear subspace, we have $0 \in U$. Thus, $y^\top(x - 0) \leq 0$, which implies $y^\top x \leq 0$. Since $y^\top x \leq 0$ for all $y \in N_U(x)$, we have $x \in U^\perp$. Thus, $N_U(x) \subseteq U^\perp$.
- $U^\perp \subseteq N_U(x)$: Let $y \in U^\perp$, then we have $y^\top u = 0$ for all $u \in U$. Since U is a linear subspace, we have $x - u \in U$. Thus, $y^\top(x - u) \leq 0$ for all $u \in U$. Thus, $y \in N_U(x)$. Thus, $U^\perp \subseteq N_U(x)$.

□

Q2

Grade:

In the proof of the existence of subgradients and of the Rockafellar-Moreau theorem, we used portions of the following result: for a proper convex function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, one has

$$\text{ri epi } f = \{ (x, z) \mid x \in \text{ri dom } f, z > f(x) \}.$$

In this problem, we will prove this result, using the prolongation principle. Let R denote the set on the right-hand side of the above equation. Note that you can use some form of the prolongation principle in each of the three parts of this question.

- Show that for any $x \in \text{dom } f$, then $(x, f(x))$ cannot be in $\text{ri epi } f$.
- Show that a point $(x, z) \in \text{epi } f$ that has $x \notin \text{ri dom } f$ cannot be in $\text{ri epi } f \subseteq \mathbb{R}$. Together with the previous result, this allows us to conclude that $\text{ri epi } f \subseteq \mathbb{R}$.
- Show that any $(x, z) \in \mathbb{R}$ is also in $\text{ri epi } f$, and hence, in view of the previous results, that $\text{ri epi } f = \mathbb{R}$. This may be done by showing that for any $(x', z') \in \text{epi } f$, there exists $\delta > 0$ such that $(x, z) + \delta((x, z) - (x', z')) \in \text{epi } f$. Hint: you should need to use another fact we proved earlier, that a convex function is continuous relative to $\text{dom } f$ at all points of $\text{ri dom } f$, that is, if $x \in \text{ri dom } f$, then for any $\tau > 0$, there exists an $\epsilon > 0$ such that $x' \in \text{dom } f$ and $\|x' - x\| < \epsilon$ together imply $|f(x') - f(x)| < \tau$. For example, it should be possible to show that for small enough δ , one has $z + \delta(z - z') > (z + f(x))/2$ but $f(x + \delta(x - x')) < (z + f(x))/2$.

Solution

- (a) *Proof.* Assume $(x, f(x)) \in \text{ri epi } f$, then there exists $\epsilon > 0$ such that $B((x, f(x)), \epsilon) \subseteq \text{epi } f$. Since f is a proper convex function, we have $f(x) \neq \infty$. Consider the point $(x, f(x) - \frac{\epsilon}{2})$, though it is within the ball, it is clearly not in $\text{epi } f$ since the second component is strictly less than $f(x)$. This is contradicted to the original assumption. Thus, $(x, f(x)) \notin \text{epi } f$. \square
- (b) *Proof.* If $x \notin \text{ri dom } f$, then by the prolongation principle, there is a direction $d \in \mathbb{R}^n$ such that $x + \lambda d \notin \text{dom } f$ for all $\lambda > 0$. Therefore, for any z' and arbitrary small $\lambda > 0$, $(x + \lambda d, z') \notin \text{epi } f$, implying (x, z) is not in the relative interior of $\text{epi } f$. \square
- (c) *Proof.* Let $(x, z) \in R$, by definition, we know $x \in \text{ri dom } f$ and $z > f(x)$.
To prove $(x, z) \in \text{ri epi } f$, we must show that for every $(x', z') \in \text{epi } f$, there exists a $\delta > 0$ such that

$$(x, z) + \delta((x, z) - (x', z')) \in \text{epi } f.$$

The point halfway between (x, z) and (x', z') is

$$\left(\frac{x+x'}{2}, \frac{z+z'}{2} \right).$$

Due to the convexity of f , this lies strictly above the graph of f at x .

For $x \in \text{ri dom } f$, by continuity of convex functions, for any $\tau > 0$, there exists $\epsilon > 0$ such that if $\|x' - x\| < \epsilon$ and $x' \in \text{dom } f$, then $|f(x') - f(x)| < \tau$.

Choose $\tau = \frac{z-f(x)}{2}$. By continuity, there exists $\epsilon > 0$ ensuring that

$$f(x') < f(x) + \tau$$

whenever $\|x' - x\| < \epsilon$. Given our choice of τ , this means

$$f(x') < \frac{z+f(x)}{2}$$

for $\|x' - x\| < \epsilon$.

Choose δ small enough that the point

$$(x, z) + \delta((x, z) - (x', z'))$$

is within an ϵ -distance from x in its first coordinate, and lies below the midway point of (x, z) and (x', z') in its second coordinate. This ensures that this point lies strictly above the graph of f .

Thus, for any $(x', z') \in \text{epi } f$, there exists a $\delta > 0$ such that $(x, z) + \delta((x, z) - (x', z')) \in \text{epi } f$, proving that any $(x, z) \in R$ is also in $\text{ri epi } f$. \square

Q3**Grade:**

In this problem, we will prove the following "almost industrial strength" generalization of Proposition 4.2.5(a): let $\mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper convex function and let A be an $m \times n$ matrix. Define $g(x) = f(Ax)$, which is also a convex function. Then, for all $x \in \mathbb{R}^n$,

$$\partial g(x) \supseteq A^\top \partial f(Ax). \quad (1)$$

Furthermore, if $\text{ri dom } f \cap \text{im } A \neq \emptyset$, that is, there exists some point in $\bar{z} \in \text{ri dom } f$ that may be expressed as $\bar{z} = A\bar{x}$ for some $\bar{x} \in \mathbb{R}^n$, then for any $x \in \mathbb{R}^n$,

$$\partial g(x) = A^\top \partial f(Ax). \quad (2)$$

- (a) Prove eq. (1).

- (b) Define $U = \{ (x, z) \in \mathbb{R}^n \times \mathbb{R}^m \mid z = Ax \}$, which is a linear subspace of $\mathbb{R}^n \times \mathbb{R}^m$, along with the following functions $\mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$:

$$F_1(x, z) = f(z)$$

$$F_2(x, z) = \begin{cases} 0, & z = Ax \\ +\infty, & z \neq Ax \end{cases}$$

$$F(x, z) = F_1(x, z) + F_2(x, z) = \begin{cases} f(z), & z = Ax \\ +\infty, & z \neq Ax \end{cases}$$

Show that F_1, F_2 and F defined in this manner are convex and that $d \in \partial g(x)$ implies $(d, 0) \in \partial F(x, Ax)$.

- (c) Show that

$$\partial F_1(x, z) = \{0\} \times \partial f(z)$$

$$\partial F_2(x, z) = \begin{cases} \{(A^\top w, -w) \mid w \in \mathbb{R}^m\}, & z = Ax \\ \emptyset, & z \neq Ax \end{cases}$$

You may use the elementary linear-algebra fact that for any $p \times q$ matrix M , the subspace orthogonal to the subspace $\{y \in \mathbb{R}^q \mid My = 0\}$ is $\{M^\top w \mid w \in \mathbb{R}^q\}$.

- (d) For the remainder of this problem, assume $\text{ri dom } f \cap \text{im } A \neq \emptyset$. Show that, in this case, $\text{ri dom } F_1$ and $\text{ri dom } F_2$ must intersect.
- (e) Find an expression for $\partial F(x, z) = \partial(F_1 + F_2)(x, z)$. You may use version of the Moreau-Rockafellar theorem, which asserts that if $\text{ri dom } f_1 \cap \text{ri dom } f_2 \neq \emptyset$, then $\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x)$ for all $x \in \mathbb{R}^n$.
- (f) Combine the above results to show that $\partial g(x) = A^\top \partial f(Ax)$.

Solution

- (a) *Proof.* Let's take any $v \in \partial f(Ax)$. By the definition of subgradients, we have

$$f(y) \geq f(Ax) + v^\top (y - Ax) \quad \forall y \in \mathbb{R}^m.$$

Let $y = Ax + Az$, then we have

$$f(Ax + Az) \geq f(Ax) + v^\top (Ax + Az - Ax) = f(Ax) + v^\top (Az).$$

Notice that the left side is $g(x + z)$, and the function on the right involves z which is the perturbation in x .

$$g(x + z) \geq g(x) + v^\top Az \quad \forall z \in \mathbb{R}^n.$$

This is the definition of the subgradients of g at x , thus, we have $A^\top v \in \partial g(x)$. Since v is arbitrary, we have $\partial g(x) \supseteq A^\top \partial f(Ax)$. \square

- (b) *Proof.* The proof is as follow:

- **F_1 is convex:** $F_1(x, z)$ is convex since f is given to be a proper convex function.
- **F_2 is convex:** Consider any two points (x_1, z_1) and (x_2, z_2) in $\mathbb{R}^n \times \mathbb{R}^m$ and any $\lambda \in (0, 1)$.
 - If $z_1 = Ax_1$ and $z_2 = Ax_2$, then the line segment between (x_1, Ax_1) and (x_2, Ax_2) is entirely contained in the set $\{(x, z) \mid z = Ax\}$, and hence F_2 is zero along this segment.
 - If either $z_1 \neq Ax_1$ or $z_2 \neq Ax_2$, then F_2 takes the value $+\infty$ at one or both of these points, and it is trivially convex as $\infty \leq \infty$.

- **F is convex:** F is the sum of F_1 and F_2 , and the sum of two convex functions is also convex.
- $d \in \partial g(x)$ **implies** $(d, 0) \in \partial F(x, Ax)$: By the definition of subgradients and function g , we have:

$$g(x+h) \geq g(x) + d^\top h \text{ for all } h \in \mathbb{R}^n.$$

Given $g(x) = f(Ax)$, this can be rewritten as:

$$f(A(x+h)) \geq f(Ax) + d^\top h.$$

Considering the definition of F , we can express this inequality as:

$$F(x+h, A(x+h)) \geq F(x, Ax) + d^\top h.$$

Given the definition of the subgradients for functions of two variables, this means:

$$(d, 0) \in \partial F(x, Ax)$$

□

(c) *Proof.* The proof is as follow:

- Since F_1 is only dependent on z , its subgradients with respect to x will simply be 0. With respect to z , the subgradients will be the same as the subgradients of f at z . Thus, we have $\partial F_1(x, z) = \{0\} \times \partial f(z)$.
- – When $z = Ax$: To find the subgradients of F_2 , we want to find all vectors (d, w) such that:

$$F_2(x+h, z+k) \geq F_2(x, z) + \langle d, h \rangle + \langle w, k \rangle$$

for all (h, k) . Given that $F_2(x, z) = 0$ for $z = Ax$, the inequality becomes

$$F_2(x+h, z+k) \geq \langle d, h \rangle + \langle w, k \rangle.$$

Considering perturbing z slightly by some k such that $z+k \neq A(x+h)$. In this case $F_2(x+h, z+k) = +\infty$, thus, the inequality holds for all (d, w) . Thus, only need to deal with $z+k = A(x+h)$. Now the inequality becomes

$$0 \geq \langle d, h \rangle + \langle w, k \rangle.$$

Given $k = A(x+h) - Ax = Ah$, the inequality can be written as:

$$0 \geq \langle d, h \rangle + \langle w, Ah \rangle.$$

For this to hold for all h , d must be orthogonal to A and w must be orthogonal to the nullspace of A^\top . Using the hint, we have

$$d = A^\top w$$

for some $w \in \mathbb{R}^m$. Next for any h :

$$k = Ah \Rightarrow -k = -Ah.$$

Thus w should be the negative of any vector in \mathbb{R}^m to ensure the orthogonality condition. Thus, we have $\partial F_2(x, z) = \{ (A^\top w, -w) \mid w \in \mathbb{R}^m \}$.

- When $z \neq Ax$: In this case, $F_2(x, z) = +\infty$. Thus, the subgradients are empty.

□

(d) *Proof.* $\text{ri dom } F_1$ is the set of all z such that $f(z) < +\infty$, which means it is the relative interior of the domain of f , i.e. $\text{ri dom } f$. $\text{ri dom } F_2$ is the set of all z such that $z = Ax$ for some x , which means it is the image of A , i.e. $\text{im } A$. Given that $\text{ri dom } f \cap \text{im } A \neq \emptyset$, there exists some $\bar{z} \in \text{ri dom } f$ such that $\bar{z} = A\bar{x}$ for some $\bar{x} \in \mathbb{R}^n$. Therefore, \bar{z} belongs to both $\text{ri dom } F_1$ and $\text{ri dom } F_2$, which means $\text{ri dom } F_1$ and $\text{ri dom } F_2$ must intersect. □

(e) *Proof.* The Moreau-Rockafellar theorem states that if $\text{ri dom } f_1 \cap \text{ri dom } f_2 \neq \emptyset$, then $\partial(f_1+f_2)(x) = \partial f_1(x) + \partial f_2(x)$ for all $x \in \mathbb{R}^n$. Since $\text{ri dom } F_1$ and $\text{ri dom } F_2$ intersect, we have $\partial F(x, z) = \partial(F_1+F_2)(x, z) = \partial F_1(x, z) + \partial F_2(x, z)$. Thus, we have

- When $z = Ax$:

$$\begin{aligned}\partial F(x, z) &= \{0\} \times \partial f(z) + \{(A^\top w, -w) \mid w \in \mathbb{R}^m\} \\ &= \{(A^\top w, v - w) \mid w \in \mathbb{R}^m, v \in \partial f(z)\}.\end{aligned}$$

- When $z \neq Ax$:

$$\begin{aligned}\partial F(x, z) &= \emptyset + \emptyset \\ &= \emptyset.\end{aligned}$$

□

(f) *Proof.* To find $\partial g(x)$, we use the property that any d in $\partial g(x)$ must satisfy $(d, 0) \in \partial F(x, Ax)$. Given $F(x, z) = F_1(x, z) + F_2(x, z)$, $F(x, Ax) = F_1(x, Ax) + F_2(x, Ax) = f(Ax) = g(x)$. Thus, for any $d \in \partial g(x)$, the corresponding $(d, 0) \in \partial F(x, Ax)$ must have the form $(A^\top w, v - w)$ where $w \in \mathbb{R}^m$ and $v \in \partial f(Ax)$. But the second coordinate is 0, which implies $v = w \Rightarrow (d, 0) = (A^\top w, 0)$. This means $d = A^\top w$ for some $w \in \partial f(Ax)$. In other words:

$$d \in A^\top \partial f(Ax).$$

This is $\partial g(x) \subseteq A^\top \partial f(Ax)$. Combining the results from (a), we have $\partial g(x) = A^\top \partial f(Ax)$. □