Special Topics in Management Science 26:711:685

Convex Analysis and Optimization

Fall 2023 Rutgers University Prof. Eckstein

Solutions to Homework 4

1. Take any $x \in U$. Consider any $d \in U^\perp$. For any $x' \in U$, we have $x' - x \in U$ and thus $\langle d, x' - x \rangle = 0$. Thus, one has $\langle d, x' - x \rangle \leq 0$ for all $x' \in U$ and so $d \in N_U(x)$. By the arbitrary choice of d, it follows that $U^\perp \subseteq N_U(x)$. Conversely, take any $d \in N_U(x)$. Now take an arbitrary $u \in U$. Since U is a linear subspace, we have $x + u \in U$ and $x - u \in U$. Since $d \in N_U(x)$, we then have

$$\begin{array}{lclcl} \langle d, (x+u) - x \rangle \geq 0 & & & \Leftrightarrow & \langle d, u \rangle \leq 0 \\ \langle d, (x-u) - x \rangle \geq 0 & \Leftrightarrow & \langle d, -u \rangle \leq 0 & \Leftrightarrow & \langle d, u \rangle \geq 0, \end{array}$$

so we conclude $\langle d, u \rangle = 0$. Since $u \in U$ was arbitrary, $d \in U^\perp$. Since $d \in N_U(x)$ was arbitrary, we have $N_U(x) \subseteq U^\perp$, and so we conclude $N_U(x) = U^\perp$.

2. (a) Here, we simply follow the logic used in class in the proof of the existence of subgradients: $(x, f(x) + 1) \in \operatorname{epi} f$, so having $(x, f(x)) \in \operatorname{ri} \operatorname{epi} f$ would require existence of that $\delta > 0$ such that

$$(x, f(x)) + \delta\big((x, f(x)) - (x, f(x) + 1)\big) = (x, f(x) - \delta) \in \operatorname{epi} f,$$

which is clearly impossible since we have to have $f(x) - \delta < f(x)$, which means that $(x, f(x) - \delta) \notin \operatorname{epi} f$. So, we conclude that $(x, f(x)) \notin \operatorname{ri} \operatorname{epi} f$.

(b) Consider any point $(x, z) \in \operatorname{epi} f$ with $x \notin \operatorname{ri} \operatorname{dom} f$. Then the prolongation principle implies that there exists $y \in \operatorname{dom} f$ such that for all $\delta > 0$, one has $x + \delta(x - y) \notin \operatorname{dom} f$, meaning that $f(x + \delta(x - y)) = \infty$ for all $\delta > 0$. Now, $(y, f(y)) \in \operatorname{epi} f$, so consider, for any $\delta > 0$,

$$(x, z) + \delta\big((x, z) - (y, f(y))\big) = \big(x + \delta(x - y), z + \delta(z - f(y))\big).$$

Since $f(x + \delta(x - y)) = \infty$, no such point can be in $\operatorname{epi} f$. So, the prolongation principle shows that $(x, z) \notin \operatorname{ri} \operatorname{epi} f$.

We have now shown that if (x, z) violates either condition defining R, it cannot be in $\operatorname{ri} \operatorname{epi} f$, and so $\operatorname{ri} \operatorname{epi} f \subseteq R$. Note that this inclusion is all that is actually needed to prove the Rockafellar-Moreau theorem; however, it is also possible to prove the opposite inclusion, as follows:

(c) To establish that in fact $\operatorname{ri} \operatorname{epi} f = R$, we will take an arbitrary element $(x, z) \in R$ and show that it is $\operatorname{ri} \operatorname{epi} f$. To do so by the prolongation principle, we need to establish that given any $(x', z') \in \operatorname{epi} f$, there exists $\delta > 0$ such that

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

First, because $x \in \operatorname{ri}\operatorname{dom} f$ and we must have $x' \in \operatorname{dom} f$, we know there exists $\delta_1 > 0$ such that $x + \delta_1(x - x') \in \operatorname{dom} f$ — and by convexity this remains true if we replace δ_1 by any $\delta \in (0, \delta_1]$.

Since $z > f(x)$, one has $\big(z - f(x)\big)/2 > 0$. Since f is continuous relative to $\operatorname{dom} f$ on the relative interior of its domain, it follows that there exists a $\delta_2 \leq \delta_1$ such that

$$\begin{aligned} \delta \leq \delta_2 \qquad &\Rightarrow \qquad f\big(x + \delta(x - x')\big) - f(x) \leq \frac{z - f(x)}{2} \\ &\Leftrightarrow \qquad f\big(x + \delta(x - x')\big) \leq f(x) + \frac{z - f(x)}{2} = \frac{z + f(x)}{2}. \end{aligned}$$

Furthermore, one can easily devise a $\delta_3 > 0$ such that

$$\delta \leq \delta_3 \qquad \Rightarrow \qquad z + \delta(z - z') \geq z - \frac{z - f(x)}{2} = \frac{z + f(x)}{2};$$

specifically, if $z' \leq z$ then any value of δ_3 is possible, and if $z' > z$ then $\delta_3 \leq \big(z - f(x)\big)/2(z' - z)$ can easily shown to be valid.

Then, for any $\delta \leq \min\{\delta_2, \delta_3\}$, one has

$$f\big(x + \delta(x - x')\big) \leq \frac{z + f(x)}{2} \leq z + \delta(z - z'),$$

meaning that

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

By the prolongation principle and the arbitrary choice of $(x', z') \in \operatorname{epi} f$, this establishes that $(x, z) \in \operatorname{ri}\operatorname{epi} f$.

3. (a) Consider any $d \in \partial f(Ax)$. Then, for any $x' \in \mathbb{R}^n$, we have

$$\begin{aligned} g(x') = f(Ax') &\geq f(Ax) + \langle d, Ax' - Ax \rangle \\ &= f(Ax) + \langle d, A(x' - x) \rangle \\ &= g(x) + \langle A^\top d, x' - x \rangle \end{aligned}$$

Since this holds for any $x' \in \mathbb{R}^n$, if follows that $A^\top d \in \partial g(x)$. Since $d \in \partial f(Ax)$ was arbitrary, $A^\top \partial f(x) \subseteq \partial g(x)$.

(b) First, consider F_1. We have

$$\begin{aligned} \operatorname{epi} F_1 &= \{(x, z, w) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \mid f(z) \geq w\} \\ &= \{(x, z, w) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \mid (z, w) \in \operatorname{epi} f\} \\ &= \mathbb{R}^n \times \operatorname{epi} f. \end{aligned}$$

Since f is convex, $\operatorname{epi} f$ is convex. Therefore, $\mathbb{R}^n \times \operatorname{epi} f = \operatorname{epi} F_1$ is also a convex set, so F_1 is convex. The linear subspace U is a convex set, hence its indicator function $\delta_U = F_2$ is convex. Finally, since it is the sum of two other convex functions F_1 and F_2, we see that F has to be convex. For the last assertion, we note that

$$\begin{array}{rlr}
 & d \in \partial g(x) & \\
\Leftrightarrow & f(Ay) \geq f(Ax) + \langle d, y - x\rangle & \forall\, y \in \mathbb{R}^n \\
\Leftrightarrow & F(y, Ay) \geq F(x, Ax) + \langle d, y - x\rangle & \forall\, y \in \mathbb{R}^n \\
\Leftrightarrow & F(y, z) \geq F(x, Ax) + \langle d, y - x\rangle & \forall\, y \in \mathbb{R}^n,\ z \in \mathbb{R}^m \quad (*) \\
\Leftrightarrow & F(y, z) \geq F(x, Ax) + \langle d, y - x\rangle + \langle 0, z - Ax\rangle & \forall\, y \in \mathbb{R}^n,\ z \in \mathbb{R}^m \\
\Leftrightarrow & (d, 0) \in \partial F(x, Ax). &
\end{array}$$

The justification for the step marked "$(*)$" is that if $z \neq Ay$, we have $F(y, z) = \infty$.

(c) Take any $(v, u) \in \partial F_1(x, z)$, in which case we must have $f(z) = F_1(x, z) < \infty$. Then, applying the subgradient inequality at the point $(v + x, z)$, we have

$$\begin{aligned}
f(z) = F_1(v + x, z) &\geq F_1(x, z) + \langle v, (v + x) - x\rangle + \langle u, z - z\rangle \\
&= f(z) + \langle v, v\rangle + \langle u, 0\rangle \\
&= f(z) + \|v\|^2 .
\end{aligned}$$

Condensing this chain of reasoning, we have $f(z) \geq f(z) + \|v\|^2$. Since $f(z) < \infty$, we must have $v = 0$. Thus, only vectors of the form $(0, u)$ can be members of $\partial F_1(x, z)$. Next, we note that

$$\begin{array}{rlr}
 & u \in \partial f(z) & \\
\Leftrightarrow & f(z') \geq f(z) + \langle u, z' - z\rangle & \forall\, z' \in \mathbb{R}^m \\
\Leftrightarrow & F_1(x', z') \geq F(x, z) + \langle 0, x' - x\rangle + \langle u, z - z'\rangle & \forall\, x' \in \mathbb{R}^n,\ z' \in \mathbb{R}^m \\
\Leftrightarrow & (0, u) \in \partial F_1(x, z). &
\end{array}$$

Thus, we must have

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

We now turn our attention to F_2. Since F_2 is just the indicator function of the subspace U, problem 1 tells us that $\partial F_2(x, z) = N_U(x, z) = U^\perp$ whenever $(x, z) \in U$, and $\partial F_2(x, z) = N_U(x, z) = \emptyset$ otherwise. Note that U consists of all vectors (x, z) satisfying $Ax - z = 0$, that is

$$[A \ \ {-I}] \begin{bmatrix} x \\ z \end{bmatrix} = 0.$$

Therefore, $U^\perp$ consists of all vectors of the form

$$[A \ \ {-I}]^\top w = (A^\top w, -w) \qquad w \in \mathbb{R}^m.$$

(d) We next note that

$$\begin{aligned}
\operatorname{ri} \operatorname{dom} F_1 &= \operatorname{ri}(\mathbb{R}^n \times \operatorname{dom} f) = \mathbb{R}^n \times \operatorname{ri} \operatorname{dom} f \\
\operatorname{ri} \operatorname{dom} F_2 &= \operatorname{ri} U = U = \{(x, Ax) \mid x \in \mathbb{R}^n\}
\end{aligned}$$

(it is easily seen that $\operatorname{aff} V = V$ and hence $\operatorname{ri} V = V$ for any linear subspace V). From the assumption $\operatorname{ri}\operatorname{dom} f \cap \operatorname{im} A \neq \emptyset$, we know there exists some $\bar{x} \in \mathbb{R}^n$ with $A\bar{x} \in \operatorname{ri}\operatorname{dom} f$. So, $(\bar{x}, A\bar{x})$ is in both $\operatorname{ri}\operatorname{dom} F_1$ and $\operatorname{ri}\operatorname{dom} F_2$.

(e) We can then use the Rockafellar-Moreau theorem to conclude that, for any $x \in \mathbb{R}^n$

$$\begin{aligned}
\partial F(x, Ax) &= \partial(F_1 + F_2)(x, Ax) \\
&= \partial F_1(x, Ax) + \partial F_2(x, Ax) \\
&= \big(\{0\} \times \partial f(Ax)\big) + U^\perp \\
&= \{(0, u) \mid u \in \partial f(Ax)\} + \{(A^\top w, -w) \mid w \in \mathbb{R}^m\} \\
&= \{(A^\top w, u - w) \mid u \in \partial f(Ax),\, w \in \mathbb{R}^m\}
\end{aligned}$$

Since the existence of $(\bar{x}, A\bar{x}) \in \operatorname{ri}\operatorname{dom} F_1 \cap \operatorname{ri}\operatorname{dom} F_2$ shows that $F = F_1 + F_2$ is proper, we have $\partial F(x, z) = \emptyset$ whenever $z \neq Ax$. Thus, a full expression for $\partial F(x, z)$ is

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

(f) Take any $x \in \mathbb{R}^n$. From part (a), we know that if $d \in \partial g(x)$, then we must have $(d, 0) \in \partial F(x, Ax)$, which, in view of the formula obtained for $\partial F(x, z)$ in part (e), means that there exist $u \in \partial f(Ax)$ and $w \in \mathbb{R}^m$ such that $d = A^\top w$ and $u - w = 0$. The second of these equations just means $w = u$, and so $d = A^\top u$ for $u \in \partial f(Ax)$. Thus, every $d \in \partial g(x)$ is expressible as $d = A^\top u$ for $u \in \partial f(Ax)$, meaning that $\partial g(x) \subseteq A^\top \partial f(Ax)$. Since we already established the opposite inclusion, we have proved the desired equality.