Special Topics in Operations Research 16:711:611

Convex Analysis and Optimization

Spring 2009 Rutgers University Prof. Eckstein

Solutions to Homework 4

1. To use a less confusing notation, we would like to show that whenever $d \in \partial f(x)$ and $d' \in \partial f(x')$, then $\langle x - x', d - d' \rangle \geq 0$. Then

$$\begin{aligned} f(x') &\geq f(x) + \langle d, x' - x \rangle && \text{because } d \in \partial f(x) \\ f(x) &\geq f(x') + \langle d', x - x' \rangle && \text{because } d' \in \partial f(x). \end{aligned}$$

Rearranging each of these inequalities slightly, we obtain

$$\begin{aligned} f(x') - f(x) &\geq \langle d, x' - x \rangle \\ f(x) - f(x') &\geq \langle d', x - x' \rangle. \end{aligned}$$

Adding this second pair of inequalities produces

$$0 \geq \langle d, x' - x \rangle + \langle d', x - x' \rangle = \langle -d, x - x' \rangle + \langle d', x - x' \rangle = \langle d' - d, x - x' \rangle.$$

Negating $0 \geq \langle d' - d, x - x' \rangle$ yields $\langle x - x', d - d' \rangle \geq 0$.

2. (a) As shown in class, if f is proper, then $\partial f(x)$ must be empty whenever $x \notin \operatorname{dom} f$. We clearly have $\operatorname{dom} \delta_C = C$, and thus $\partial \delta_C(x) = \emptyset = N_C(x)$ whenever $x \notin C$. It remains to show that the $\partial \delta_C(x) = N_C(x)$ for $x \in C$. For $x \in C$, we have $\delta_C(x) = 0$, and thus

$$\begin{aligned} d \in \partial \delta_C(x) \quad &\Leftrightarrow \quad \delta_C(y) \geq \delta_C(x) + \langle d, y - x \rangle && \forall y \in \mathbb{R}^n \\ &\Leftrightarrow \quad \delta_C(y) \geq \langle d, y - x \rangle && \forall y \in \mathbb{R}^n \\ &\Leftrightarrow \quad 0 \geq \langle d, y - x \rangle && \forall y \in C, \end{aligned}$$

the last equivalence holding because $\delta_C(y) = \infty$ for $y \notin C$, so $\delta_C(y) \geq \langle d, y - x \rangle$ always holds whenever $y \notin C$, for any $d \in \mathbb{R}^n$. The last condition is precisely $d \in N_C(x)$.

(b) We want to show that for all $y \in \mathbb{R}^n$,

$$f(y) \geq f(x) + \langle w + d, y - x \rangle.$$

For $y \notin \operatorname{dom} f$, we have $f(y) = \infty$, and so the above inequality has to hold; only the $y \in \operatorname{dom} f$ case remains. In that case, we have

$$\begin{aligned} f(y) &\geq f(x) + \langle d, y - x \rangle && \text{because } d \in \partial f(x) \\ 0 &\geq \langle w, y - x \rangle && \text{because } w \in N_{\operatorname{dom} f}(x) \text{ and } y \in \operatorname{dom} f. \end{aligned}$$

If we add these inequalities, we obtain $f(y) \geq f(x) + \langle w + d, y - x \rangle$, as required.

(c) We showed in class that $\partial f(x)$ has to be closed, so it is sufficient to show that $\partial f(x)$ is bounded if and only if $x \in \operatorname{int} \operatorname{dom} f$. First, we prove the "if" part, which is similar to a proof I gave in class. If $x \in \operatorname{int} \operatorname{dom} f$, then there exists an $\epsilon > 0$ such that $B(x, \epsilon) \subseteq \operatorname{dom} f$, where $B(x, \epsilon)$ denotes the open ball of radius ϵ around the point x. Suppose that $\partial f(x)$ is unbounded; then there would have to exist some sequence $\{g^k\} \subset \partial f(x)$ with $\|g^k\| \to \infty$. Without loss of generality, we can assume that $g^k \neq 0$ for all k. Consider the sequence $\{d^k\}$ defined by $d^k = g^k/\|g^k\|$. This sequence has a subsequence $\{d^k\}_{k \in \mathcal{K}}$ that converges to some vector d with $\|d\| = 1$. Consider the sequence

$$x + (\epsilon/2)d^k \underset{\mathcal{K}}{\to} x + (\epsilon/2)d \in B(x, \epsilon).$$

Since g^k is a subgradient of f at x, we should have

$$\begin{aligned} f(x + (\epsilon/2)d^k) &\geq f(x) + \langle g^k, x + (\epsilon/2)d^k - x \rangle \\ &= f(x) + \langle g^k, (\epsilon/2)d^k \rangle \\ &= f(x) + (\epsilon/2)\langle g^k, d^k \rangle \\ &= f(x) + (\epsilon/2)\langle g^k, (1/\|g^k\|)g^k \rangle \\ &= f(x) + (\epsilon/2\|g^k\|)\langle g^k, g^k \rangle \\ &= f(x) + (\epsilon/2\|g^k\|) \left\|g^k\right\|^2 \\ &= f(x) + (\epsilon/2) \left\|g^k\right\| \end{aligned}$$

Thus, we have $f(x+(\epsilon/2)d^k) \to \infty$, since $\|g^k\| \to \infty$. Since $\operatorname{int} \operatorname{dom} f$ is nonempty, $\operatorname{ri} \operatorname{dom} f = \operatorname{int} \operatorname{dom} f \supseteq B(x, \epsilon)$. Since convex functions must be continuous on the relative interior of their domain, f has to be continuous on $B(x, \epsilon)$. Since $\|d^k\| = 1$ for all k, we have $x + (\epsilon/2)d^k \in B(x, \epsilon)$ for all k. Hence, f being continuous throughout $B(x, \epsilon)$, we should have $f(x + (\epsilon/2)d^k) \to_{\mathcal{K}} f(x + (\epsilon/2)d)$, which should be finite since $x + (\epsilon/2)d \in B(x, \epsilon) \subseteq \operatorname{dom} f$. This contradicts $f(x + (\epsilon/2)d^k) \to \infty$. We conclude that $\partial f(x)$ has to be bounded.

We prove the converse by showing that $\partial f(x)$ must be unbounded if it is nonempty and we have $x \notin \operatorname{int} \operatorname{dom} f$. In this case, the supporting hyperplane theorem asserts that there exists some $a \in \mathbb{R}^n \backslash \{0\}$ such that $\langle a, x' \rangle \leq \langle a, x \rangle$ for all $x' \in C$, that is, $\langle a, x' - x \rangle \leq 0$ for all $x' \in C$, and thus $a \in N_C(x)$. Since $N_C(x)$ is a cone, it must contain the set of vectors $\{\alpha a \mid \alpha \geq 0\}$. We know that $\partial f(x) \neq \emptyset$, so take any $d \in \partial f(x)$. From part (b), we know that we must also have $d + \alpha a \in \partial f(x)$ for any $\alpha \geq 0$, since $\alpha a \in N_C(x)$. Since $a \neq 0$ and we can take α arbitrarily large, it follows that $\partial f(x)$ must be unbounded.

(d) Take any $x \in U$. Consider any $d \in U^\perp$. For any $x' \in U$, we have $x' - x \in U$ and thus $\langle d, x' - x \rangle = 0$. So, we have $d \in N_U(x)$, and 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

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

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 $N_U(x) = U^\perp$.

3. In $\mathbb{R}^2$, define two convex sets
$$C_1 = \left\{x \in \mathbb{R}^2 \mid x_1 \le 0\right\} \qquad C_2 = \left\{x \in \mathbb{R}^2 \mid \|x - (1,0)\| \le 1\right\},$$
and let $f_1 = \delta_{C_1}$ and $f_2 = \delta_{C_2}$, whence $\partial f_1(x) = N_{C_1}(x)$ and $\partial f_2(x) = N_{C_2}(x)$ by problem 2(a). Note that $C_1 \cap C_2 = \{0\}$. Some straightforward calculations (omitted here) show that
$$N_{C_1}(0) = \{(d_1, d_2) \mid d_1 \ge 0,\, d_2 = 0\} \qquad N_{C_2}(0) = \{(d_1, d_2) \mid d_1 \le 0,\, d_2 = 0\},$$
From which we surmise that
$$\partial f_1(0) + \partial f_2(0) = N_{C_1}(0) + N_{C_2}(0) = \{(y_1, y_2) \mid y_2 = 0\}.$$
However, for all x,
$$\begin{aligned} f_1(x) + f_2(x) &= \delta_{C_1}(x) + \delta_{C_2}(x) \\ &= \begin{cases} 0, & \text{if } x \in C_1 \text{ and } x \in C_2 \\ +\infty, & \text{otherwise} \end{cases} \\ &= \delta_{C_1 \cap C_2}(x). \end{aligned}$$
Thus, we have $\partial(f_1 + f_2)(0) = N_{\{0\}}(0) = \mathbb{R}^2$, which strictly contains $\partial f_1(0) + \partial f_2(0)$ as computed above.

4. Consider the convex function
$$(f + \delta_C)(x) = f(x) + \delta_C(x) = \begin{cases} f(x), & \text{if } x \in C \\ +\infty, & \text{if } x \notin C. \end{cases}$$
Minimizing f over C is equivalent to minimizing $f + \delta_C$ over $\mathbb{R}^n$. A necessary and sufficient condition for x^* to minimize $f + \delta_C$ over $\mathbb{R}^n$ is $0 \in \partial(f + \delta_C)(x^*)$. Note that $\operatorname{ri}\operatorname{dom}\delta_C = \operatorname{ri} C$, and we therefore have $\operatorname{ri}\operatorname{dom} f \cap \operatorname{ri}\operatorname{dom}\delta_C \neq \emptyset$ because of the assumption $\operatorname{ri}\operatorname{dom} f \cap \operatorname{ri} C \neq \emptyset$. The Rockafellar-Moreau theorem then assures us that for any $x \in \mathbb{R}^n$, $\partial(f + \delta_C)(x) = \partial f(x) + \partial\delta_C(x)$; we also know from problem 2(a) that $\partial\delta_C(x) = N_C(x)$. Therefore, our necessary and sufficient condition $0 \in \partial(f + \delta_C)(x^*)$ can be written $0 \in \partial f(x) + N_C(x)$.

   Finally, $0 \in \partial f(x) + N_C(x)$ is equivalent to the existence of some $d \in \partial f(x^*)$ with $-d \in N_C(x^*)$. Having $-d \in N_C(x^*)$ means $\langle -d, y - x\rangle \le 0$ for all $y \in C$, that is, $\langle d, y - x\rangle \ge 0$ for all $y \in C$.

5. (a) Consider any $d \in \partial f(Ax)$. Then, for any $x' \in \mathbb{R}^n$, we have
$$\begin{aligned} g(x') = f(Ax') &\ge f(Ax) + \langle d, Ax' - Ax\rangle \\ &= f(Ax) + \langle d, A(x' - x)\rangle \\ &= g(x) + \langle A^\top d, x' - x\rangle \end{aligned}$$
   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 by problem 4(b) of homework 2. 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

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

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

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

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, parts (a) and (d) of problem 2 tell us that $\partial F_2(x,z) = U^\perp$ whenever $(x,z) \in U$, and $\partial F_2(x,z) = \emptyset$ 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

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

We next note that $\operatorname{ri}\operatorname{dom} F_1 = \mathbb{R}^n \times \operatorname{ri}\operatorname{dom} f$ and $\operatorname{ri}\operatorname{dom} F_2 = \operatorname{ri} U = U$. 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 $\operatorname{ri}\operatorname{dom} F_2$. 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

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

(d) Take any $x \in \mathbb{R}^n$. From part (b), 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 $\partial g(x) \subseteq A^\top \partial f(Ax)$. In view of part (a), we have proved the desired equality.