Special Topics in Operations Research 16:711:611

Convex Analysis and Optimization

Spring 2009 Rutgers University Prof. Eckstein

Solutions to Homework 5

1. Let $Z = \{x \in \mathbb{R}^n \mid Ax = b\}$, and define the following convex funtions

$$
\begin{aligned}
f_0(x) &= f(x) \\
f_j(x) &= \delta_{L(0,h_j)}(x) = \begin{cases} 0, & \text{if } h_j(x) \le 0 \\ +\infty, & \text{if } h_j(x) > 0 \end{cases} \qquad j = 1, \ldots, r \\
f_{r+1}(x) &= \delta_Z(x) = \begin{cases} 0, & \text{if } Ax = b \\ +\infty, & \text{if } Ax \ne b \end{cases} \\
f_{r+2}(x) &= \delta_X(x) = \begin{cases} 0, & \text{if } x \in X \\ +\infty, & \text{if } x \notin X, \end{cases}
\end{aligned}
$$

Note that $(f_0 + \cdots + f_{r+2})(x) = +\infty$ if x violates any of the constraints in (1), and otherwise $(f_0 + \cdots + f_{r+2})(x) = f(x)$. Therefore, solving (1) is equivalent to minimizing $f_0 + \cdots + f_{r+2}$ over $\mathbb{R}^n$, or equivalently solving

$$0 \in \partial(f_0 + \cdots + f_{r+2})(x).$$

Note that

$$
\begin{aligned}
\operatorname{ri}\operatorname{dom} f_0 &= \operatorname{ri}\operatorname{dom} f \\
\operatorname{ri}\operatorname{dom} f_j &= \{x \in \mathbb{R}^n \mid h_j(x) < 0\} \qquad j = 1, \ldots, r \qquad \text{(as proved in class)} \\
\operatorname{ri}\operatorname{dom} f_{r+1} &= \operatorname{ri} Z = Z \qquad \text{(since } Z = \operatorname{aff} Z\text{)} \\
\operatorname{ri}\operatorname{dom} f_{r+2} &= \operatorname{ri} X.
\end{aligned}
$$

The point $\overline{x}$ stipulated in the assumption lies in all these sets, and so we have

$$\operatorname{ri}\operatorname{dom} f_0 \cap \operatorname{ri}\operatorname{dom} f_1 \cap \cdots \cap \operatorname{ri}\operatorname{dom} f_{r+2} \ne \emptyset,$$

The Rockefellar-Moreau theorem then guarantees that

$$\partial(f_0 + \cdots + f_{r+2})(x) = \partial f_0(x) + \cdots + \partial f_{r+2}(x)$$

for all $x \in \mathbb{R}^n$, and so a necessary and sufficient condition for x^* to be optimal is

$$0 \in \partial f_0(x^*) + \cdots + \partial f_{r+2}(x^*).$$

We also proved in class that

$$
\begin{aligned}
\partial f_j(x) &= \begin{cases} \{0\}, & \text{if } h_j(x) < 0 \\ \{\mu \nabla f(x) \mid \mu \ge 0\}, & \text{if } h_j(x) = 0 \\ \emptyset, & \text{if } h_j(x) > 0 \end{cases} \\
\partial f_{r+1}(x) &= \begin{cases} \{A^\top \lambda \mid \lambda \in \mathbb{R}^m\}, & \text{if } Ax = b \\ \emptyset & \text{if } Ax \ne b. \end{cases}
\end{aligned}
$$

Finally, we know from the previous homework that $\partial f_{r+2}(x) = \partial \delta_X(x) = N_X(x)$. Rewriting the expression for $\partial f_j(x)$ as

$$\partial f_j(x) = \{\mu_j \nabla f(x) \mid \mu_j \geq 0,\ \mu_j h_j(x) = 0\},$$

we have that $0 \in \partial f_0(x) + \cdots + \partial f_{r+2}(x^*)$ if and only if it satisfies all the constraints in (1) and

$$0 \in \partial f(x^*) + \sum_{j=1}^{r} \{\mu_j \nabla f(x^*) \mid \mu_j \geq 0,\ \mu_j h_j(x^*) = 0\} + \{A^\top \lambda \mid \lambda \in \mathbb{R}^m\} + N_X(x).$$

This means in turn that there must exist $\mu^* \in \mathbb{R}^r$ and $\lambda^* \in \mathbb{R}^m$ with $\mu^* \geq 0$, $\mu_j^* h_j(x) = 0$ for $j = 1 \ldots, r$, and

$$0 \in \partial f(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla f(x^*) + A^\top \lambda^* + N_X(x).$$

This condition, coupled with feasibility with respect to all the constraints, is exactly what was to be proved.

2. (a) Suppose $x \in K$ and take any $y \in N_K(x)$. Since K is a convex cone, we have $x + z \in K$ for all $z \in K$. By the definition of $N_K(x)$, we must then have

$$\langle y, (x+z) - x \rangle = \langle y, z \rangle \leq 0 \qquad \forall\, z \in K.$$

Thus, we must have $y \in K^*$. Since K is a cone, we also have $0 \in K$, and so, again by the definition of $N_K(x)$, we also have $\langle y, 0 - x \rangle = -\langle x, y \rangle \leq 0$, that is, $\langle x, y \rangle \geq 0$. Since $y \in K^*$, we also have $\langle x, y \rangle \leq 0$ and thus $\langle x, y \rangle = 0$. Since $y \in N_K(x)$ was arbitrary, we have $N_K(x) \subseteq \{y \in K^* \mid \langle x, y \rangle = 0\}$.

Conversely, take any $y \in K^*$ with $\langle x, y \rangle = 0$. Then, for any $z \in K$, we have

$$\begin{aligned}
\langle y, z - x \rangle &= \langle y, z \rangle - \langle y, x \rangle \\
&= \langle y, z \rangle && \text{(because } \langle x, y \rangle = 0) \\
&\leq 0 && \text{(because } y \in K^* \text{ and } z \in K)
\end{aligned}$$

Therefore, $y \in N_K(x)$ and we have $\{y \in K^* \mid \langle x, y \rangle = 0\} \subseteq N_K(x)$. In view of the opposite inclusion proved above, $N_K(x) = \{y \in K^* \mid \langle x, y \rangle = 0\}$.

Next we observe that since K is convex, $T_K(x) = [N_K(x)]^*$. Also, we can write $N_K(x)$ as the intersection of two cones, specifically $N_K(x) = K^* \cap L$, where we let $L = \{y \in \mathbb{R}^n \mid \langle x, y \rangle = 0\}$, which is in fact a linear subspace. Note that $L^* = L^\perp = \{\alpha x \mid \alpha \in \mathbb{R}\}$.

In homework 3, we proved that for two nonempty cones $C_1, C_2 \subseteq \mathbb{R}^n$, we have $(C_1 + C_2)^* = C_1^* \cap C_2^*$. If C_1 and C_2 are convex, we then have

$$\begin{aligned}
(C_1^* \cap C_2^*)^* &= (C_1 + C_2)^{**} && \text{(taking the polar of both sides above)} \\
&= \operatorname{cl}\operatorname{conv}(C_1 + C_2) && \text{(by the polar cone theorem)} \\
&= \operatorname{cl}(C_1 + C_2) && \text{(because } C_1 \text{ and } C_2 \text{ are convex).}
\end{aligned}$$

Substituting $C_1 = K$ and $C_2 = L^\perp$, whence $C_2^* = (L^\perp)^* = (L^\perp)^\perp = L$, we obtain

$$T_K(x) = [N_K(x)]^* = (K^* \cap L)^* = (C_1^* \cap C_2^*)^* = \mathrm{cl}(C_1 + C_2) = \mathrm{cl}(K + L^\perp).$$

Thus,

$$\begin{aligned} T_K(x) &= \mathrm{cl}(K + L^\perp) \\ &= \mathrm{cl}(K + \{\alpha x \mid \alpha \in \mathbb{R}\}) \\ &= \mathrm{cl}(K + \{\alpha x \mid \alpha \le 0\}) \qquad \text{(positive multiples of } x \text{ are already in } K) \\ &= \mathrm{cl}(K - \{\alpha x \mid \alpha \ge 0\}). \end{aligned}$$

(b) We would like to compute the cone of feasible directions $F_Z(x)$ to Z at some $x \in Z$. The members of $F_Z(x)$ are vectors $d \in \mathbb{R}^n$ such that for all sufficiently small $\delta > 0$, $x + \delta d \in Z$, that is,

$$A(x + \delta d) - b \in K \qquad \Leftrightarrow \qquad (Ax - b) + \delta A d \in K.$$

In other works, $Ad \in F_K(Ax - b)$. Since Z is convex,

$$\begin{aligned} N_Z(x) &= [F_Z(x)]^* \\ &= \{d \in \mathbb{R}^n \mid Ad \in F_K(Ax - b\}^* \\ &= \mathrm{cl}\,\{A^\top \lambda \mid \lambda \in [F_K(Ax - b)]^*\} \qquad \text{(by homework 3, problem 4(b))} \\ &= \mathrm{cl}\,\{A^\top \lambda \mid \lambda \in N_K(Ax - b)\} \qquad \text{(since } K \text{ is convex)} \\ &= \mathrm{cl}\,\{A^\top \lambda \mid \lambda \in K^*,\ \langle Ax - b, \lambda\rangle = 0\} \qquad \text{(by part (a)).} \end{aligned}$$

Note that the formula we obtain generalizes those obtained for the cones $K = \{0\}$ and $K = \{y \in \mathbb{R}^m \mid y \le 0\}$ in class: for $K = \{0\}$, we obtain $K^* = \mathbb{R}^m$ and $Ax - b = 0$, so there are no constraints on λ. If $K = \{y \in \mathbb{R}^m \mid y \le 0\}$, then $K^* = \{d \in \mathbb{R}^m \mid d \ge 0\}$, and so $\lambda \ge 0$, and we also have the complementary slackness constraint $\langle Ax - b, \lambda\rangle = 0$.

(c) Using the notation of part (b), the problem is simply to minimize $f(x)$ over $x \in Z$. Therefore, a necessary condition for a local optimum is

$$0 \in \nabla f(x^*) + [T_Z(x^*)]^* = \nabla f(x^*) + N_Z(x^*).$$

Plugging in the above formula for $N_Z(x)$, and dropping the "cl" operation as assumed, we obtain the generalized Karush-Kuhn-Tucker conditions

$$\nabla f(x^*) + A^\top \lambda^* = 0 \qquad \lambda^* \in K^* \qquad \langle Ax^* - b, \lambda^*\rangle = 0.$$

3. (a) $\widehat{f}(y) = \sup_{x \in \mathbb{R}}\{xy - \frac{1}{2}x^2\}$. The supremand is a differentiable concave function, so we just set $\frac{\partial}{\partial x}(xy - \frac{1}{2}x^2) = 0$ obtaining $y - x = 0$, that is, $x = y$. Substituing $x = y$ into $xy - \frac{1}{2}x^2$, we have $\widehat{f}(y) = y^2 - \frac{1}{2}y^2 = \frac{1}{2}y^2$. Thus, in this case, $\widehat{f} = f$.

(b) In this case, we have

$$\begin{aligned}
\widehat{f}(y) &= \sup_{x\in[a,b]} \{xy - 0\} \\
&= \begin{cases} ax, & \text{if } x < 0 \\ 0, & \text{if } x = 0 \\ bx, & \text{if } x > 0 \end{cases}
\end{aligned}$$

(actually, all three formulas agree when $x = 0$).

(c) Here, we have $\widehat{f}(y) = \sup_{x\in\mathbb{R}}\{xy - e^x\}$. Note that if $y < 0$, then we have $\lim_{x\to-\infty}(xy - e^x) = +\infty$, and so the supremum is $+\infty$. For $y = 0$, we have $\widehat{f}(0) = \sup_{x\in\mathbb{R}}\{-e^x\} = 0$ (although this sup is not attained). For $y > 0$, we note that we are maximizing a differentiable concave function, and so try setting the derivative of the supremand to 0:

$$\frac{\partial}{\partial x}(xy - e^x) = 0 \quad\Leftrightarrow\quad y - e^x = 0 \quad\Leftrightarrow\quad y = e^x \quad\Leftrightarrow\quad x = \log y.$$

Substituting $x = \log y$ into $xy - e^x$, we obtain $\widehat{f}(y) = (\log y)y - e^{\log y} = y\log y - y$. Summarizing,

$$\widehat{f}(y) = \begin{cases} y\log y - y, & y > 0, \\ 0, & y = 0, \\ +\infty, & y < 0. \end{cases}$$

Note (not part of the assignment): by using L'Hôpital's rule, it is possible to show that $\lim_{y\downarrow 0}(y\log y - y) = 0$, so $\widehat{f}$ is right-continuous at 0; however, it turns out that $\widehat{f}$ is not differentiable at 0, nor does it have any subgradients there.