ORIE 6300 Mathematical Programming I	November 17, 2014

Lecture 24

Lecturer: Damek Davis — *Scribe: Rundong Wu*

1 Last Time

1. **Definition 1** *Let $g : \mathbb{R}^n \to \bar{\mathbb{R}}$ be a function. Then the set-valued operators*

$$
\begin{aligned}
\partial g(x) &:= \{v \in \mathbb{R}^n \mid \forall y,\ g(y) \geq g(x) + \langle v, y - x\rangle\} \\
\partial g^{\infty}(x) &:= \left\{v \in \mathbb{R}^n \mid \begin{bmatrix} v \\ 0 \end{bmatrix} \in N_{\mathrm{epi}(g)}(x, g(x))\right\}
\end{aligned}
$$

are called the subdifferential and horizontal subdifferential respectively. We left it as an exercise to prove that $\partial^{\infty} g(x) = N_{\mathrm{dom}(g)}(x)$.

2. **Theorem 1** *Let $g : \mathbb{R}^n \to \bar{\mathbb{R}}$ be closed, convex. Then $\forall x \in \mathrm{dom}(g)$, we have*

$$
N_{\mathrm{epi}(g)}(x, g(x)) = \left\{\lambda \begin{bmatrix} v \\ -1 \end{bmatrix} \mid v \in \partial g(x), \lambda > 0\right\} \cup \left\{\begin{bmatrix} v \\ 0 \end{bmatrix} \mid v \in \partial^{\infty} g(x)\right\}
$$

3. **Corollary 2** *Let $g : \mathbb{R}^n \to \bar{\mathbb{R}}$ be closed and convex. Then the following holds.*
   1. $(\forall x \in \mathrm{dom}(g))\ \partial g(x) \cup (\partial g^{\infty}(x)\backslash\{0\}) \neq \emptyset$.
   2. $(\forall x \in \mathrm{int}(\mathrm{dom}(g)))\ \partial g(x) \neq \emptyset$.

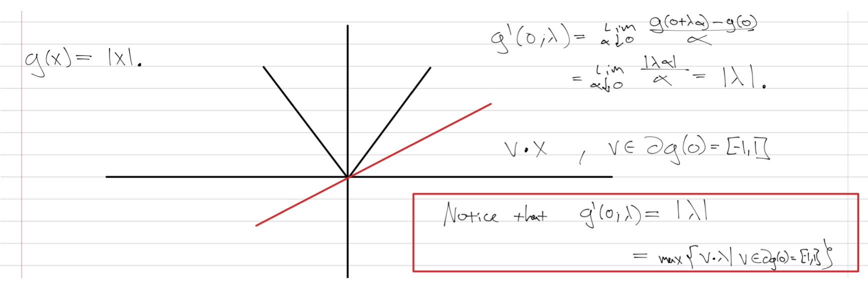
2 Today

Assumption: All functions considered in this lecture are closed, proper, convex. With this assumption, the set $\partial g(x)$ is closed and convex. This fact will be needed later and we leave the proof as an exercise.
This will be a very technical lecture. However, it will be one of the most useful "fact sheets" for those of you who study optimization.

To compute with subdifferentials, it's extremely useful to relate subdifferentials to directional derivatives.

Definition 2 *Let $x \in \mathrm{dom}(g)$, we call g differentiable at x in the direction $p \in \mathbb{R}^n$ if $g'(x; p) = \lim_{\alpha \downarrow 0} \frac{g(x+\alpha p) - g(x)}{\alpha}$ exists.*

Example 1



Lemma 3 *Let g be closed, convex. Let $x \in \operatorname{int}(\operatorname{dom}(g))$, then $\forall p \in \mathbb{R}^n$, $g'(x;p)$ exists.*

Proof: Let $r(\alpha) = \frac{1}{\alpha}[g(x+\alpha p) - g(x)]$ and $z = x + \beta p$, where $0 < \alpha \le \beta < \infty$ and $\lambda = \frac{\alpha}{\beta}$. Then

$$\begin{aligned} r(\alpha) &= \frac{1}{\alpha}[g(x+\alpha p) - g(x)] \\ &= \frac{1}{\alpha}[g(\lambda z + (1-\lambda)x) - g(x)] \\ &\le \frac{1}{\alpha}[\lambda g(z) + (1-\lambda)g(x) - g(x)] \\ &= \frac{1}{\alpha}[\lambda g(x+\beta p) - \lambda g(x)] \\ &= r(\beta) \end{aligned}$$

Thus $r(\alpha)$ is increasing. Further, $\forall v \in \partial g(x)$ (such v must exist because $x \in \operatorname{int}(\operatorname{dom}(g))$),

$$\begin{aligned} r(\alpha) &= \frac{1}{\alpha}[g(x+\alpha p) - g(x)] \\ &\ge \frac{1}{\alpha}\langle v, \alpha p\rangle \\ &= \langle v, p\rangle \end{aligned}$$

which means $r(\alpha)$ is bounded below. Therefore the limit must exist. $\square$

Proposition 4 *Let g be closed, convex. If $x \in \operatorname{int}(\operatorname{dom}(g))$, then the following holds.*
1. $g'(x;\cdot)$ is convex, homogeneous function of degree one, which means $g'(x;\tau p) = \tau g'(x;p)$.
2. $(\forall y \in \mathbb{R}^n)$ $g(y) \ge g(x) + g'(x; y-x)$.

Proof:
1. We leave this as an exercise.
2. Define $g_\alpha = (1-\alpha)x + \alpha y$. Then

$$\begin{aligned} & g(y_\alpha) \le (1-\alpha)g(x) + \alpha g(y) \\ \Rightarrow\quad & \tfrac{1}{\alpha}(1-\alpha)[g(y_\alpha) - g(x)] + g(y_\alpha) \le g(y) \\ \Rightarrow\quad & g(y) \ge \liminf_{\alpha\downarrow 0}\{g(y_\alpha) + \tfrac{1-\alpha}{\alpha}[g(y_\alpha) - g(x)]\} \ge g(x) + g'(x; y-x) \end{aligned}$$

$\square$

Finally, we find the following exact relation between $g'(x;p)$ and $\partial g(x)$.

Theorem 5 *(Max formula) Let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be closed, convex. Suppose $x \in \operatorname{int}(\operatorname{dom}(g))$, then $\forall p \in \mathbb{R}^n$, $g'(x;p) = \sup\{\langle v, p\rangle \mid v \in \partial g(x)\}$.*

Proof: First we prove the claim that $\partial g(x) = \partial_p g'(x;0)$.
"$\subseteq$": $\forall v \in \partial g(x)$,

$$\begin{aligned} g'(x;p) &= \lim_{\alpha\downarrow 0}\frac{1}{\alpha}[g(x+\alpha p) - g(x)] \\ &\ge \frac{1}{\alpha}\langle v, \alpha p\rangle \quad (*) \\ &= g'(x;0) + \langle v, p\rangle \\ \Rightarrow v &\in \partial_p g'(x;0) \end{aligned}$$

"$\supseteq$": $\forall y \in \mathbb{R}^n$,

$$\begin{aligned} g(y) &\ge g(x) + g'(x; y-x) \\ &\ge g(x) + \langle w, y-x\rangle \end{aligned}$$

where $w \in \partial_p g'(x;0)$. Thus $w \in \partial g(x)$. Therefore, $\partial g(x) = \partial_p g'(x;0)$.
Let $w \in \partial_p g'(x;p)$, then $\forall \bar{p} \in \mathbb{R}^n$, $\tau > 0$,

$$\tau g'(x;\bar{p}) = g'(x;\tau\bar{p}) \geq g'(x;p) + \langle w, \tau\bar{p} - p\rangle$$

Take $\tau \to \infty$, we get

$$\begin{aligned} g'(x;\bar{p}) &\geq \langle w, \bar{p}\rangle = g(x;0) + \langle w, \bar{p}\rangle \\ \Rightarrow w &\in \partial_p g'(x;0) = \partial g(x) \end{aligned}$$

Take $\tau \to 0$, we get

$$g'(x;p) \leq \langle w, p\rangle$$

From $(*)$ we have $g'(x;p) \geq \langle w, p\rangle$. Thus $g'(x;p) = \langle w, p\rangle = \sup\{\langle v, p\rangle \mid v \in \partial g(x)\}$. □

Why is the max formula useful?

Definition 3 *Given a closed conves set $C \subseteq \mathbb{R}^n$, the support function is defined as*

$$\sigma_C(v) = \sup\{\langle y, v\rangle \mid y \in C\}.$$

Support functions "find" points $y \in C$, s.t. $v \in N_C(y)$, because $N_C(y) = \{v \in \mathbb{R}^n \mid y$ maxizes $\langle x, v\rangle$ over all $x \in C\}$
The max formula shows that $g'(x;v) = \sigma_{\partial g(x)}(v)$.

Support functions completely characterize convex sets.

Proposition 6 *Let $C_1, C_2 \subseteq \mathbb{R}^n$ be closed convex sets. Suppose that $\sigma_{C_1} = \sigma_{C_2}$, then $C_1 = C_2$.*

Proof: Let $y \in C_1$, and suppose that $y \notin C_2$. Then by the separating hyperplane theorem, $\exists v \in \mathbb{R}^n, b \in \mathbb{R}$, such that

$$\sigma_{C_1}(v) \geq \langle v, y\rangle > b \geq \sup_{x \in C_2} \langle v, x\rangle = \sigma_{C_2}(v)$$

Thus we've reached a contradiction and $\sigma_{C_1} = \sigma_{C_2}$. □

3 Calculus

Convex subdifferentials satisfy several calculus rules. We start with differentiable functions.

Proposition 7 *Let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be differentiable on its domain and let $x \in \mathrm{int}(\mathrm{dom}(g))$, then $\partial g(x) = \{\nabla g(x)\}$.*

Proof: By the max formula we have $\forall p \in \mathbb{R}^n$,

$$\sigma_{\{\nabla g(x)\}}(p) = \langle \nabla g(x), p\rangle = g'(x;p) = \sigma_{\partial g(x)}(p).$$

□

Functions of the form $g(c(x))$, where $g : \mathbb{R}^n \to \overline{\mathbb{R}}$, $c : \mathbb{R}^m \to \mathbb{R}^n$ and g, c are both smooth, satisfy the chain rule $\nabla(g \circ c)(x) = \nabla c(x)^T \nabla g(c(x))$. Nonsmooth functions also satisfy a chain rule. But, in general, $g \circ c$ is NOT convex, unless C is linear.

Proposition 8 *(Chain rule) Let $A \in \mathbb{R}^{n\times m}$ and $b \in \mathbb{R}^n$. Let $f(y) = g(Ay - b)$ and $x \in \text{int}(\text{dom}(f))$. Then $\partial f(x) = A^T \partial g(Ax - b)$.*

Proof: $\forall p \in \mathbb{R}^n$, we have

$$\begin{aligned}
f'(x;p) &= \lim_{\alpha\downarrow 0} \frac{f(x+\alpha p) - f(x)}{\alpha} \\
&= \lim_{\alpha\downarrow 0} \frac{g(A(x+\alpha p) - b) - g(Ax-b)}{\alpha} \\
&= \lim_{\alpha\downarrow 0} \frac{g(z+\alpha Ap) - g(z)}{\alpha} \quad (z := Ax - b) \\
&= g'(Ax-b; Ap) \\
&= \max\{\langle v, Ap\rangle \mid v \in \partial g(Ax-b)\} \\
&= \max\{\langle \bar{v}, p\rangle \mid \bar{v} \in A^T \partial g(Ax-b)\}
\end{aligned}$$

Thus $\forall p \in \mathbb{R}^n$, $\max\{\langle v,p\rangle \mid v \in \partial f(x)\} = f'(x,p) = \max\{\langle v,p\rangle \mid v \in A^T\partial g(Ax-b)\}$, *i.e.* $\partial f(x) = A^T \partial g(Ax - b)$. □

Exercise: Given $g_1 : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $g_2 : \mathbb{R}^m \to \overline{\mathbb{R}}$. Let $g(x,y) = g_1(x) + g_2(y)$, then $\partial g(x,y) = \partial g_1(x) \times \partial g_2(y)$.

Corollary 9 *(Sum rule) Let $g_1, g_2 : \mathbb{R}^n \to \overline{\mathbb{R}}$ and define $g := g_1 + g_2$. Suppose that $x \in \text{int}(\text{dom}(g)) = \text{int}(\text{dom}(g_1)) \cap \text{int}(\text{dom}(g_2))$. Then $\partial g(x) = \partial g_1(x) + \partial g_2(x)$.*

Proof: Define $f(x_1, x_2) = g_1(x_1) + g_2(x_2)$ and let $g = f \circ \begin{bmatrix} I \\ I \end{bmatrix}$. Then using the chain rule, we get

$$\begin{aligned}
\partial g(x) &= \begin{bmatrix} I & I \end{bmatrix} \partial f\left(\begin{bmatrix} I \\ I \end{bmatrix} x\right) \\
&= \begin{bmatrix} I & I \end{bmatrix} \partial g_1(x) \times \partial g_2(x) \\
&= \partial g_1(x) + \partial g_2(x)
\end{aligned}$$

□

The product of convex functions is usually not convex so we shouldn't expect to see a "product rule".

We do have one final rule for maximums of convex functions.

Lemma 10 *Let $g_1, ..., g_m : \mathbb{R}^n \to \bar{\mathbb{R}}$, define $g(x) := \max\{g_1(x), ..., g_m(x)\}$. Assume $x \in \text{int}(\text{dom}(g)) = \cap_{i=1}^m \text{int}(\text{dom}(g_i))$. Then $\partial g(x) = \text{conv}\{\partial g_i(x) \mid g_i(x) = g(x)\}$.*

Proof: Without loss of generality, assume $\{i \mid g_i(x) = g(x)\} = \{1, ..., k\}$. Then $\forall p \in \mathbb{R}^n$, we have

$$\begin{aligned}
g'(x;p) &= \max_{i=1,\ldots,k} g_i'(x;p) \qquad (\textit{Exercise}) \\
&= \max_{i=1,\ldots,k} \{\langle v_i, p\rangle \mid v_i \in \partial g_i(x)\} \quad (*)
\end{aligned}$$

Define $\Delta_k = \{\lambda \in [0,1]^k \mid \sum_{i=1}^k \lambda_i = 1\}$. Then

$$\begin{aligned}
(*) &= \max_{\lambda \in \Delta_k} \{\sum_{i=1}^k \lambda_i \max\{\langle v_i, p\rangle \mid v_i \in \partial g_i(x)\}\} \\
&= \max\{\sum_{i=1}^k \langle \lambda_i v_i, p\rangle \mid v_i \in \partial g_i(x), \lambda \in \Delta_k\} \\
&= \max\{\langle v, p\rangle \mid v \in \sum_{i=1}^k \lambda_i v_i, v_i \in \partial g_i(x), \lambda \in \Delta_k\} \\
&= \max\{\langle v, p\rangle \mid v \in \operatorname{conv}\{\partial g_i(x) \mid g_i(x) = g(x)\}\} \\
&= \sigma_{\operatorname{conv}\{\partial g_i(x) \mid g_i(x) = g(x)\}}
\end{aligned}$$

□

Example 2 *Consider the maximal value function $v(u) := \max\{c^T x \mid Ax \leq b+u\}$. Suppose that $P(A^T, c) = \{y \mid A^T y = c, y \geq 0\}$ is bounded. Then*

$$\begin{aligned}
v(u) &= \min\{(b+u)^T y \mid y \in P(A^T, c)\} \\
&= \min\{(b+u)^T y_i \mid i = 1, ..., k\}
\end{aligned}$$

where $y_1, ..., y_k$ are the vertices of $P(A^T, c)$.

Then $-V(u)$ is convex and $\partial[-V](u) = \operatorname{conv}\{-y_i \mid (b+u)^T y_i = V(v)\}$.

We previously showed that $-\partial[-V](u) = \operatorname{argmin}\{(b+u)^T y \mid y \in P(A^T, c)\}$. Thus, every solution to the dual is a convex combination of vertices.

Next time we will examine properties of the subdifferential that immediately lead to algorithms.