553.665 Introduction to Convexity, Fall 2022
Section 7

Hongyu Cheng

October 28, 2022

In lecture notes (Theorem 3.4.5) we presented an important result that $\partial(f+g)(\mathbf{x}) = \partial f(\mathbf{x}) + \partial g(\mathbf{x})$ for convex functions $f, g : \mathbb{R}^d \to \mathbb{R}$. We will prove this in today's section. More precisely, let $E, Y \subseteq \mathbb{R}^d$, for any convex functions $f\colon E \to \mathbb{R} \cup \{+\infty\}$ and $g\colon Y \to \mathbb{R} \cup \{+\infty\}$ and linear map $A\colon E \to Y$ with $0 \in \operatorname{int}(\operatorname{dom} g - A\operatorname{dom} f)$, then $\partial(f + g \circ A)(\mathbf{x}) = \partial f(\mathbf{x}) + A^* \partial g(A\mathbf{x})$.

Exercise 1. Subdifferential Calculus.

(a) Let $E, Y \subseteq \mathbb{R}^d$, then for any convex functions $f\colon E \to \mathbb{R} \cup \{+\infty\}$ and $g\colon Y \to \mathbb{R} \cup \{+\infty\}$ and linear map $A\colon E \to Y$, show that the following perturbed value function is convex:

$$h(\mathbf{u}) = \inf_{\mathbf{x} \in E} \{f(\mathbf{x}) + g(A\mathbf{x} + \mathbf{u})\}.$$

(b) For any f and g, use the definition of subgradients to show the following partial calculus rule

$$\partial(f + g \circ A)(\mathbf{x}) \supseteq \partial f(\mathbf{x}) + A^* \partial g(A\mathbf{x}) .$$

(Note: In our case, here $A^* = A^\top$.)

(c) Show that equality holds above whenever f and g are convex with $0 \in \operatorname{int}(\operatorname{dom} g - A\operatorname{dom} f)$.

(a) *Proof.* $\forall \mathbf{u}_1, \mathbf{u}_2 \in \operatorname{dom} g - A\operatorname{dom} f$, $\forall \lambda \in [0,1]$. Then $\forall \varepsilon > 0$, by the definition of $h(\mathbf{u}_1), h(\mathbf{u}_2)$, there exists $\mathbf{x}_1, \mathbf{x}_2 \in E$ such that

$$f(\mathbf{x}_1) + g(A\mathbf{x}_1 + \mathbf{u}_1) < h(\mathbf{u}_1) + \varepsilon \tag{1}$$
$$f(\mathbf{x}_2) + g(A\mathbf{x}_2 + \mathbf{u}_2) < h(\mathbf{u}_2) + \varepsilon \tag{2}$$

Then we consider $\mathbf{x}_0 = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$:

$$\begin{aligned}
h(\lambda\mathbf{u}_1 + (1-\lambda)\mathbf{u}_2) &= \inf_{x \in E}\{f(\mathbf{x}) + g(A\mathbf{x} + \lambda\mathbf{u}_1 + (1-\lambda)\mathbf{u}_2)\} \\
&\le f(\mathbf{x}_0) + g(A\mathbf{x}_0 + \lambda\mathbf{u}_1 + (1-\lambda)\mathbf{u}_2) \\
&= f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) + g(\lambda(A\mathbf{x}_1 + \mathbf{u}_1) + (1-\lambda)(A\mathbf{x}_2 + \mathbf{u}_2)) \\
&\le \lambda f(\mathbf{x}_1) + (1-\lambda)f(\mathbf{x}_2) + \lambda g(A\mathbf{x}_1 + \mathbf{u}_1) + (1-\lambda)g(A\mathbf{x}_2 + \mathbf{u}_2) \\
&\overset{(1),(2)}{<} \lambda h(\mathbf{u}_1) + \lambda\varepsilon + (1-\lambda)h(\mathbf{u}_2) + (1-\lambda)\varepsilon \\
&= \lambda h(\mathbf{u}_1) + (1-\lambda)h(\mathbf{u}_2) + \varepsilon
\end{aligned}$$

let $\varepsilon \to 0$, we have

$$h(\lambda \mathbf{u}_1 + (1-\lambda)\mathbf{u}_2) \le \lambda h(\mathbf{u}_1) + (1-\lambda) h(\mathbf{u}_2),$$

which implies $h(\mathbf{u})$ is convex. $\square$

(b) *Proof.* $\forall \Phi \in \partial f(\mathbf{x}) + A^* \partial g(A\mathbf{x}) \implies \Phi = \phi + A^*\psi$, where $\phi \in \partial f(\mathbf{x}), \psi \in \partial g(A\mathbf{x})$. Then, $\forall \mathbf{y} \in \mathrm{dom}(g \circ A) \cap \mathrm{dom} f$, we have

$$\begin{aligned} g(A\mathbf{y}) &\ge g(A\mathbf{x}) + \langle A^*\psi, \mathbf{y} - \mathbf{x} \rangle \\ f(\mathbf{y}) &\ge f(\mathbf{x}) + \langle \phi, \mathbf{y} - \mathbf{x} \rangle \end{aligned}$$

hence,

$$f(\mathbf{y}) + g(A\mathbf{y}) \ge g(A\mathbf{x}) + f(\mathbf{x}) + \langle \phi + A^*\psi, \mathbf{y} - \mathbf{x} \rangle, \ \forall \mathbf{y} \in \mathrm{dom}(g \circ A) \cap \mathrm{dom} f.$$

Thus, by definition of subgradient, $\mathbf{y} = A^*\psi + \phi \in \partial(f + g \circ A)(\mathbf{x})$. This proves $\partial(f + g \circ A)(\mathbf{x}) \supseteq \partial f(\mathbf{x}) + A^* \partial g(A\mathbf{x})$. $\square$

(c) *Proof.* $\forall \Phi \in \partial(f + g \circ A)(\mathbf{x}) \implies 0 \in \partial(f + g \circ A - \langle \Phi, \cdot \rangle)(\mathbf{x})$, that is, $\mathbf{x}$ minimizes $f(\mathbf{y}) + g(A\mathbf{y}) - \langle \Phi, \mathbf{y} \rangle$. Let

$$H(\mathbf{u}) = \inf_{\mathbf{y} \in E} \{ f(\mathbf{y}) + g(A\mathbf{y} + \mathbf{u}) - \langle \Phi, \mathbf{y} \rangle \}.$$

Exercise 1 (a) yields that $H(\mathbf{u})$ is convex. Also, $0 \in \mathrm{int\ dom} H$ since $0 \in \mathrm{int}(\mathrm{dom} g - A\mathrm{dom}\ f)$, so by Max Formula $-\Psi \in \partial H(0)$ exists. Then by definition of subgradient,

$$H(0) \le H(\mathbf{u}) + \langle \Psi, \mathbf{u} \rangle. \tag{3}$$

Recall that $H(\mathbf{u}) = \inf_{\mathbf{y} \in E} \{ f(\mathbf{y}) + g(A\mathbf{y} + \mathbf{u}) - \langle \Phi, \mathbf{y} \rangle \}$ and $\mathbf{x}$ minimizes $f(\mathbf{y}) + g(A\mathbf{y}) - \langle \Phi, \mathbf{y} \rangle$, hence $\forall \mathbf{y}, \forall \mathbf{u}$ we have,

$$\begin{aligned} \underbrace{f(\mathbf{x}) + g(A\mathbf{x}) - \langle \Phi, \mathbf{x} \rangle}_{H(0)} &\overset{(3)}{\le} \underbrace{\inf_{\mathbf{y} \in E} \{ f(\mathbf{y}) + g(A\mathbf{y} + \mathbf{u}) - \langle \Phi, \mathbf{y} \rangle \}}_{H(\mathbf{u})} + \langle \Psi, \mathbf{u} \rangle \\ &\le f(\mathbf{y}) + g(A\mathbf{y} + \mathbf{u}) - \langle \Phi, \mathbf{y} \rangle + \langle \Psi, \mathbf{u} \rangle \end{aligned} \tag{4}$$

Take $\mathbf{y} = \mathbf{x}$ in Equation (4), we have

$$\begin{aligned} g(A\mathbf{x}) &\le g(A\mathbf{x} + \mathbf{u}) + \langle \Psi, \mathbf{u} \rangle, \ \forall \mathbf{u}, \\ \implies g(A\mathbf{x} + \mathbf{u}) &\ge g(A\mathbf{x}) + \langle -\Psi, (A\mathbf{x} + \mathbf{u}) - A\mathbf{x} \rangle, \ \forall \mathbf{u}, \\ \implies g(\mathbf{z}) &\ge g(A\mathbf{x}) + \langle -\Psi, \mathbf{z} - A\mathbf{x} \rangle, \ \forall \mathbf{z}. \end{aligned}$$

This proves $-\Psi \in \partial g(A\mathbf{x})$.

Take $\mathbf{u} = A(\mathbf{x} - \mathbf{y}) \in \mathrm{dom}\ g - A\mathrm{dom}\ f$ in Equation (4), we have

$$\begin{aligned} f(\mathbf{x}) + g(A\mathbf{x}) - \langle \Phi, \mathbf{x} \rangle &\le f(\mathbf{y}) + g(A\mathbf{x}) - \langle \Phi, \mathbf{y} \rangle + \langle \Psi, A(\mathbf{x} - \mathbf{y}) \rangle, \ \forall \mathbf{y} \\ \implies f(\mathbf{x}) &\le f(\mathbf{y}) + \langle \Phi, \mathbf{x} - \mathbf{y} \rangle + \langle A^*\Psi, \mathbf{x} - \mathbf{y} \rangle, \ \forall \mathbf{y} \\ \implies f(\mathbf{y}) &\ge f(\mathbf{x}) + \langle \Phi + A^*\Psi, \mathbf{y} - \mathbf{x} \rangle, \ \forall \mathbf{y} \end{aligned}$$

Therefore, by the definition of subgradient, $\Phi + A^*\Psi \in \partial f(\mathbf{x})$.

Thus,

$$\Phi = \Phi + A^*\Psi - A^*\Psi = \underbrace{\Phi + A^*\Psi}_{\in \partial f(\mathbf{x})} + A^* \underbrace{(-\Psi)}_{\in \partial g(A\mathbf{x})} \in \partial f(\mathbf{x}) + A^* \partial g(A\mathbf{x}),$$

which implies $\partial(f + g \circ A)(\mathbf{x}) \subseteq \partial f(\mathbf{x}) + A^* \partial g(A\mathbf{x})$, then completes the proof. □

Exercise 2. **Max Formula.** Let $E \subseteq \mathbb{R}^d$, then for any convex function $f : E \to \mathbb{R} \cup \{+\infty\}$ and $\bar{\mathbf{x}} \in \operatorname{int} \operatorname{dom} f$, $\mathbf{r} \in E$, the following exists (is finite):

$$f'(\bar{\mathbf{x}}; \mathbf{r}) = \sup\{\langle \phi, \mathbf{r} \rangle : \phi \in \partial f(\bar{\mathbf{x}})\}.$$

In particular, $\partial f(\bar{\mathbf{x}}) \neq \varnothing$.

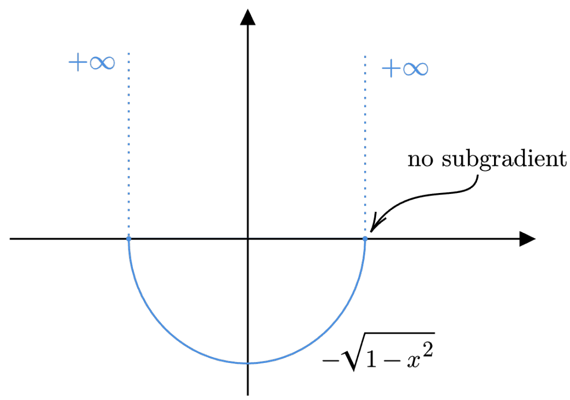

Figure 1: On the boundary of dom f, there may not be any $\phi \in \partial f(x)$.

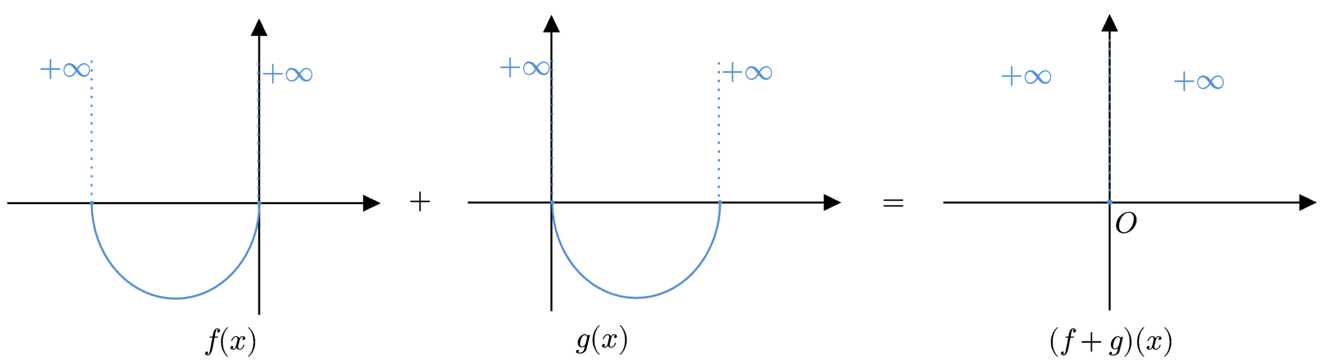

Figure 2: No sum rule in general.