The University of Texas at Austin
Department of Electrical and Computer Engineering

ECE 381V: Large-Scale Optimization II — Spring 2022

Lecture 10

Caramanis & Mokhtari Monday, February 21, 2022

Goal: In this lecture, we talk about nonsmooth (nondifferentiable) functions and nonsmooth optimization. More precisely, we talk about the definitions of subgradient and subdifferential and their properties for convex functions. We further state the optimality conditions for convex nondifferentiable optimization problems in both constrained and unconstrained settings.

Disclaimer. The figures are borrowed from Lieven Vandenberghe's notes.

1 Nonsmooth Optimization

Consider a general constrained convex optimization problem where the constraint set is a simple convex set with no functional constraints:

$$\begin{aligned} &\min f(\mathbf{x}) \\ &\text{s.t } \mathbf{x} \in \mathcal{Q}, \end{aligned} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is convex on the convex set $\mathcal{Q}$. So far we talked about the cases that the function f is differentiable and its gradient ∇f is well-defined. However, in several settings, the objective function could be nondifferentiable. This setting is indeed more general than the setting that we have seen so far. Hence, we need a more general version of gradients to deal with this class of problems.

2 Subgradients

To start, let us first recap the most basic property of gradients for convex functions. If a function is convex and differentiable then we have

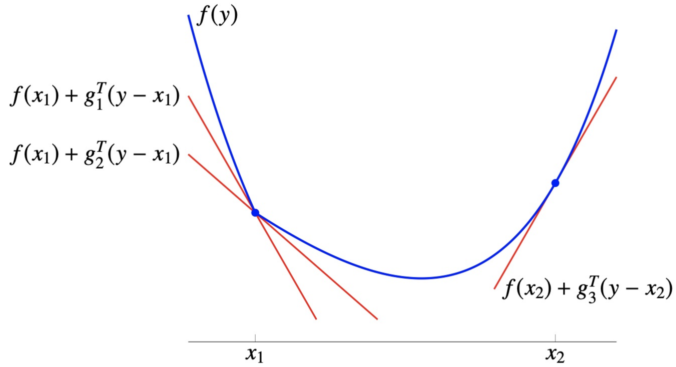
$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x})$$

This means that the first-order approximation of f at $\mathbf{x}$ is a global lower bound for $f(\mathbf{y})$. Based on this observation we define subgradients which is a more general version of gradient with a similar property for nondifferentiable functions.

Definition 1. *A vector $\mathbf{g}$ is called a subgradient of the function f at point $\mathbf{x}_0 \in dom(f)$, if for any $\mathbf{y} \in dom(f)$ we have*

$$f(\mathbf{y}) \geq f(\mathbf{x}_0) + \mathbf{g}^\top (\mathbf{y} - \mathbf{x}_0).$$

Note that at a specific point, the function f may not have a unique subgradient and it could possibly have multiple subgradients.



g_1, g_2 are subgradients at x_1; g_3 is a subgradient at x_2

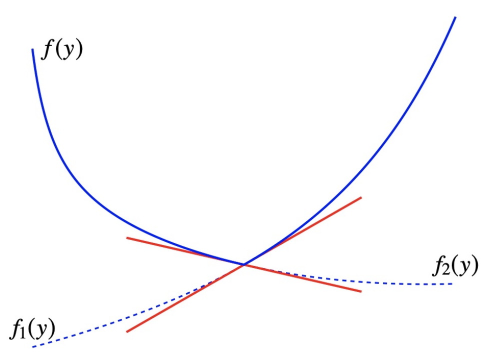
Definition 2. *The set of all subgradients of f $\mathbf{x}_0$ is called the subdifferential of f at $\mathbf{x}_0$ and is denoted by $\partial f(\mathbf{x}_0)$, i.e.,*

$$\partial f(\mathbf{x}_0) := \{\mathbf{g} \in \mathbb{R}^n \mid \mathbf{g}^\top(\mathbf{y} - \mathbf{x}_0) \leq f(\mathbf{y}) - f(\mathbf{x}_0), \text{ for all } \mathbf{y} \in dom(f)\}$$

Note that each subgradient can be interpreted as a halfspace. Hence, subdifferential which is an intersection of halfspaces is a closed convex set.

Example. Consider the function $f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x})\}$ where f_1 and f_2 are convex and differentiable. Indeed, this function is not differentiable and its subdifferential at $\mathbf{x}_0$ is:

1. If we have $f_1(\mathbf{x}_0) > f_2(\mathbf{x}_0)$, then the subdifferential of f at $\mathbf{x}_0$ is $\nabla f_1(\mathbf{x}_0)$.
2. If we have $f_1(\mathbf{x}_0) < f_2(\mathbf{x}_0)$, then the subdifferential of f at $\mathbf{x}_0$ is $\nabla f_2(\mathbf{x}_0)$.
3. If we have $f_1(\mathbf{x}_0) = f_2(\mathbf{x}_0)$, then the subdifferential of f at $\mathbf{x}_0$ is the line segment $[\nabla f_1(\mathbf{x}_0), \nabla f_2(\mathbf{x}_0)]$.



Example. Consider the function $f(x) = -\sqrt{x}$ for $x \geq 0$. It can be easily verified that the subdifferential of this function at $x = 0$ is an empty set. In other words, there is no a such that

$$-\sqrt{y} \geq ay$$

for any $y \geq 0$.

Theorem 1. *If the function f is convex, and $\mathbf{x}_0 \in$ int $(dom(f))$, then $\partial f(\mathbf{x}_0)$ is a nonempty bounded set.*

Perhaps one of the most important properties of subgradients is the monotonicity property for convex functions.

Theorem 2. *The subdifferential of a convex function is a monotone operator, i.e.,*

$$(\mathbf{g}_1 - \mathbf{g}_2)^\top (\mathbf{x}_1 - \mathbf{x}_2) \geq 0 \qquad \text{for all } \mathbf{x}_1, \mathbf{x}_2 \in dom(f),\ \mathbf{g}_1 \in \partial f(\mathbf{x}_1),\ \mathbf{g}_2 \in \partial f(\mathbf{x}_2)$$

This property is quite similar to the monotonicity of gradients for differentiable functions. It can be easily proved by adding the following expressions:

$$f(\mathbf{y}) \geq f(\mathbf{x}_1) + \mathbf{g}_1^\top (\mathbf{y} - \mathbf{x}_1)$$

and

$$f(\mathbf{y}) \geq f(\mathbf{x}_2) + \mathbf{g}_2^\top (\mathbf{y} - \mathbf{x}_2)$$

The following property of subgradients is a crucial property that we leverage in convex optimization. Before stating this result, let us revisit the notion of function sublevel set:

$$\mathcal{S}_f(a) = \{\mathbf{x} \in \text{dom}(f) \mid f(\mathbf{x}) \leq a\}$$

Theorem 3. *For any $\mathbf{x}_0 \in dom(f)$, all vectors $\mathbf{g} \in \partial f(\mathbf{x}_0)$ are supporting to the level set $\mathcal{S}_f(f(\mathbf{x}_0))$, i.e.,*

$$\mathbf{g}^\top (\mathbf{x}_0 - \mathbf{x}) \geq 0 \qquad \text{for all } \mathbf{x} \in \mathcal{S}_f(f(\mathbf{x}_0)).$$

Proof. Note that as $\mathbf{g} \in \partial f(\mathbf{x}_0)$, we have

$$f(\mathbf{x}) \geq f(\mathbf{x}_0) + \mathbf{g}^\top (\mathbf{x} - \mathbf{x}_0), \qquad \text{for all } \mathbf{x} \in \text{dom}(f)$$

Now for any $\mathbf{x} \in \mathcal{S}_f(f(\mathbf{x}_0))$ in addition we have $f(\mathbf{x}) \leq f(\mathbf{x}_0)$. Hence, we can conclude that for any $\mathbf{x} \in \mathcal{S}_f(f(\mathbf{x}_0))$ we have

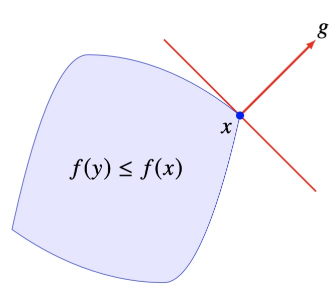
$$f(\mathbf{x}_0) \geq f(\mathbf{x}_0) + \mathbf{g}^\top (\mathbf{x} - \mathbf{x}_0)$$

which leads to the claim. □

An immediate corollary of this result is the following:

Corollary 1. *If $\mathbf{g}$ is a subgradient of f at $\mathbf{x}$, then*

$$f(\mathbf{y}) \leq f(\mathbf{x}) \quad \Longrightarrow \quad \mathbf{g}^\top (\mathbf{y} - \mathbf{x}) \leq 0$$



Corollary 2. *If $\mathbf{g}$ is a subgradient of f at $\mathbf{x}$, and $\mathbf{x}^*$ is a minimizer of the problem, i.e., $\mathbf{x}^* = \text{argmin}_{\mathbf{x} \in \mathcal{Q}} f(\mathbf{x})$, then we have*

$$\mathbf{g}^\top (\mathbf{x}^* - \mathbf{x}) \leq 0$$

3 Optimality Conditions

Previously, we noticed that there is a close connection between the gradient of a convex objective function and the optimality condition. For instance, in the convex unconstrained setting, we showed that $\nabla f(\mathbf{x}^*) = \mathbf{0} \iff \mathbf{x}^* \in \operatorname{argmin}_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$. Now, we aim to provide a similar guarantee for the nondifferentiable setting.

Theorem 4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then,*

$$\mathbf{0} \in \partial f(\mathbf{x}^*) \qquad \iff \qquad \mathbf{x}^* \in \operatorname*{argmin}_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$$

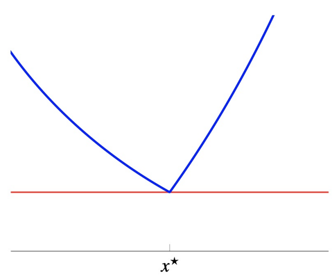
Proof. If $\mathbf{0} \in \partial f(\mathbf{x}^*)$, then we have for any $\mathbf{y} \in \mathbb{R}^n$

$$f(\mathbf{y}) \geq f(\mathbf{x}^*) + \mathbf{0}^\top(\mathbf{y} - \mathbf{x}^*) \geq f(\mathbf{x}^*)$$

hence $\mathbf{x}^*$ is an optimal solution.
If $\mathbf{x}^*$ is an optimal solution, then $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for any $\mathbf{x} \in \mathbb{R}^n$. Further, for any $\mathbf{g} \in \partial f(\mathbf{x}^*)$

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \mathbf{g}^\top(\mathbf{x} - \mathbf{x}^*)$$

Indeed, in this case we can have $\mathbf{g} = \mathbf{0}$ which implies that $\mathbf{0} \in \partial f(\mathbf{x}^*)$. $\square$



Now, to extend the above we need to define the notion of normal cone for a convex set.

Definition 3. *Given a set $\mathcal{S} \subseteq \mathbb{R}^n$ and a point $\mathbf{x} \in \mathcal{S}$, the normal cone of $\mathcal{S}$ at $\mathbf{x}$ is defined as*

$$\mathcal{N}_{\mathcal{S}}(\mathbf{x}) := \{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{y}^\top(\mathbf{z} - \mathbf{x}) \leq 0, \text{for all } \mathbf{z} \in \mathcal{S}\}$$

Important example follows:

Example. Consider the indicator function of a convex set $\mathcal{Q}$, denoted by $\delta_{\mathcal{Q}}(\mathbf{x})$, where $\delta_{\mathcal{Q}}(\mathbf{x}) = 0$ if $\mathbf{x} \in \mathcal{Q}$, and $\delta_{\mathcal{Q}}(\mathbf{x}) = \infty$ if $\mathbf{x} \notin \mathcal{Q}$. In this case the set of subdifferential of $\delta_{\mathcal{Q}}$ at $\mathbf{x} \in \mathcal{Q}$ is defined as

$$\partial\delta(\mathbf{x}) := \{\mathbf{g} \in \mathbb{R}^n \mid \mathbf{g}^\top(\mathbf{z} - \mathbf{x}) \leq \delta(\mathbf{z}) - \delta(\mathbf{x}), \text{ for all } \mathbf{z} \in \mathcal{Q}\}$$

which can be simplified as

$$\partial\delta(\mathbf{x}) := \{\mathbf{g} \in \mathbb{R}^n \mid \mathbf{g}^\top(\mathbf{z} - \mathbf{x}) \leq 0, \text{ for all } \mathbf{z} \in \mathcal{Q}\}$$

Hence, in this case we obviously for any $\mathbf{x} \in \mathcal{Q}$ we have

$$\partial\delta(\mathbf{x}) = \mathcal{N}_{\mathcal{Q}}(\mathbf{x})$$

Theorem 5. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and $\mathcal{Q}$ be a closed convex set. Then,*

$$\exists\ \mathbf{g} \in \partial f(\mathbf{x}^*) \ \ s.t. - \mathbf{g} \in \mathcal{N}_{\mathcal{Q}}(\mathbf{x}^*) \qquad \Longleftrightarrow \qquad \mathbf{x}^* \in \operatorname*{argmin}_{\mathbf{x} \in \mathcal{Q}} f(\mathbf{x})$$

Proof. Note that the problem of $\min_{\mathbf{x} \in \mathcal{Q}} f(\mathbf{x})$ can be written as

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) + \delta_{\mathcal{Q}}(\mathbf{x})$$

where $\delta_{\mathcal{Q}}(\mathbf{x})$ is the indicator function of the set $\mathcal{Q}$. Form the previous result we have

$$\mathbf{0} \in \partial(f(\mathbf{x}^*) + \delta_{\mathcal{Q}}(\mathbf{x}^*)) \qquad \Longleftrightarrow \qquad \mathbf{x}^* \in \operatorname*{argmin}_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) + \delta_{\mathcal{Q}}(\mathbf{x}) \qquad \Longleftrightarrow \qquad \mathbf{x}^* \in \operatorname*{argmin}_{\mathbf{x} \in \mathcal{Q}} f(\mathbf{x})$$

From the fact that $\partial \delta_{\mathcal{Q}}(\mathbf{x}) = \mathcal{N}_{\mathcal{Q}}(\mathbf{x})$ we obtain that

$$\mathbf{0} \in \partial f(\mathbf{x}^*) + \mathcal{N}_{\mathcal{Q}}(\mathbf{x}^*) \qquad \Longleftrightarrow \qquad \mathbf{x}^* \in \operatorname*{argmin}_{\mathbf{x} \in \mathcal{Q}} f(\mathbf{x})$$

and the claim follows. $\square$

Using the definition of normal cone we obtain the following necessary and sufficient condition for optimality.

Corollary 3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and $\mathcal{Q}$ be a closed convex set. Then,*

$$\exists\ \mathbf{g} \in \partial f(\mathbf{x}^*) \ \ s.t.\ \mathbf{g}^\top(\mathbf{z} - \mathbf{x}^*) \geq 0, \text{for all } \mathbf{z} \in \mathcal{S} \qquad \Longleftrightarrow \qquad \mathbf{x}^* \in \operatorname*{argmin}_{\mathbf{x} \in \mathcal{Q}} f(\mathbf{x})$$