Optimization Theory

Lecture 03

Fudan University

luoluo@fudan.edu.cn

Outline

Outline

Subdifferential Calculus

Let f_1 and f_2 be proper convex functions on $\mathbb{R}^d$, then

$$\partial(f_1 + f_2)(\mathbf{x}) \supseteq \partial f_1(\mathbf{x}) + \partial f_2(\mathbf{x}).$$

If the sets $\mathrm{ri}(\mathrm{dom}\, f_1)$ and $\mathrm{ri}(\mathrm{dom}\, f_2)$ have a point in common (overlap sufficiently), we have

$$\partial(f_1 + f_2)(\mathbf{x}) = f_1(\mathbf{x}) + \partial f_2(\mathbf{x}).$$

We define the relative interior $\mathrm{ri}(\mathcal{C})$ for convex $\mathcal{C} \subseteq \mathbb{R}^d$ as

$$\begin{aligned}\mathrm{ri}(\mathcal{C}) = \{\mathbf{z} \in \mathcal{C} : \; & \text{for every } \mathbf{x} \in \mathcal{C} \text{ such that} \\ & \text{there exist a } \mu > 1 \text{ such that } (1-\mu)\mathbf{x} + \mu\mathbf{z} \in \mathcal{C}\}.\end{aligned}$$

It means every line segment in $\mathcal{C}$ having $\mathbf{z}$ as one endpoint can be prolonged beyond $\mathbf{z}$ without leaving $\mathcal{C}$.

Subdifferential Calculus

Nonempty subdifferential and convexity:

1. If any $\mathbf{x} \in \operatorname{dom} f$ satisfies $\partial f(\mathbf{x}) \neq \emptyset$, then f is convex.
2. If $f : \mathbb{R}^d \to \mathbb{R}$ is convex and $\mathbf{x}$ belongs to the interior of $\operatorname{dom} f$, then $\partial f(\mathbf{x}) \neq \emptyset$.

Theorem (Hyperplane Separation Theorem)

Let $\mathcal{X} \subseteq \mathbb{R}^d$ is a convex set and $\mathbf{x}_0$ belongs to its boundary. Then, there exists a nonzero vector $\mathbf{w} \in \mathbb{R}^d$ such that

$$\langle \mathbf{w}, \mathbf{x} \rangle \leq \langle \mathbf{w}, \mathbf{x}_0 \rangle.$$

The subgradient of a convex function may not exist at a boundary point of the domain.

As an example, consider the function

$$f(x) = -\sqrt{x}$$

defined on $[0, +\infty)$, where we have $\partial f(0) = \emptyset$.

Subdifferential Calculus

Given matrix $\mathbf{A} \in \mathbb{R}^{d\times m}$ and vector $\mathbf{b} \in \mathbb{R}^d$, define

$$h(\mathbf{x}) = f(\mathbf{A}\mathbf{x} + \mathbf{b}),$$

where f is a proper convex on $\mathbb{R}^d$. Then $h(\mathbf{x})$ is convex and

$$\partial h(\mathbf{x}) \supseteq \mathbf{A}^\top \partial f(\mathbf{A}\mathbf{x} + \mathbf{b}).$$

If the range of $\mathbf{A}$ contains a point of $\mathrm{ri}(\mathrm{dom}\, h)$, then

$$\partial h(\mathbf{x}) = \mathbf{A}^\top \partial f(\mathbf{A}\mathbf{x} + \mathbf{b}).$$

Optimal Condition

Theorem

Consider proper closed convex function f and closed convex set $\mathcal{C} \subseteq (\mathrm{dom}\, f)^\circ$. A point $\mathbf{x}^* \in \mathcal{C}$ is a solution of convex optimization problem

$$\min_{\mathbf{x} \in \mathcal{C}} f(\mathbf{x})$$

if and only if

$$\mathbf{0} \in \partial(f(\mathbf{x}^*) + \mathbb{1}_{\mathcal{C}}(\mathbf{x}^*)).$$

Equivalently, there exists a subgradient $\mathbf{g}^* \in \partial f(\mathbf{x}^*)$, such that any $\mathbf{y} \in \mathcal{C}$ satisfies

$$\langle \mathbf{g}^*, \mathbf{y} - \mathbf{x}^* \rangle \geq 0.$$

In particular, the point $\mathbf{x}^*$ is the solution of the problem in unconstrained case if

$$\mathbf{0} \in \partial f(\mathbf{x}^*).$$

Outline

Regularity Conditions

The following regularity conditions are useful in the convergence analysis of convex optimization problems.

1. We say that a function $f : \mathcal{C} \to \mathbb{R}$ is G-Lipschitz continuous if for all $\mathbf{x}, \mathbf{y} \in \mathcal{C}$, we have

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq G \|\mathbf{x} - \mathbf{y}\|_2 .$$

2. We say a differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ is L-smooth if it has L-Lipschitz continuous gradient. That is, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we have

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 \leq L \|\mathbf{x} - \mathbf{y}\|_2 .$$

3. If the function

$$g(\mathbf{x}) = f(\mathbf{x}) - \frac{\mu}{2} \|\mathbf{x}\|_2^2$$

is convex for some $\mu > 0$, we say f is μ-strongly convex.

Strong Convexity

Theorem

The function $f : \mathcal{C} \to \mathbb{R}$ defined on convex set $\mathcal{C}$ is μ-strongly-convex if and only if

$$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \le \alpha f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}) - \frac{\mu\alpha(1-\alpha)}{2}\|\mathbf{x}-\mathbf{y}\|_2^2$$

for all $\mathbf{x}, \mathbf{y} \in \mathcal{C}$ and $\alpha \in [0,1]$.

Theorem

If a function f is differentiable on open set $\mathcal{C}$, then it is μ-strongly convex on $\mathcal{C}$ if and only if

$$f(\mathbf{y}) \ge f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle + \frac{\mu}{2}\|\mathbf{y}-\mathbf{x}\|_2^2$$

hols for any $\mathbf{x}, \mathbf{y} \in \mathcal{C}$.

Strong Convexity

If there exists some

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathcal{C}} f(\mathbf{x}),$$

then it is the unique minimizer.

Moreover, the solution is stable such that any approximate solution $\hat{\mathbf{x}}$ satisfying

$$f(\mathbf{x}) \leq f(\mathbf{x}^*) + \epsilon$$

leads to

$$\|\mathbf{x}^* - \hat{\mathbf{x}}\|_2^2 \leq \frac{2\epsilon}{\mu}.$$

Lipschitz Continuity and Smoothness

Theorem

A convex function f is G-Lipschitz continuous on $\mathrm{dom}\, f$ if

$$\max_{\mathbf{g}\in\partial f(\mathbf{x})} \{\|\mathbf{g}\|_2\} \leq G$$

for all $\mathbf{x} \in \mathrm{dom}\, f$.

Theorem

A function $f : \mathbb{R}^d \to \mathbb{R}$ is L-smooth (possibly nonconvex), then it holds

$$|f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle| \leq \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|_2^2$$

holds for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$.

Smoothness and Convexity

Theorem

A function $f : \mathbb{R}^d \to \mathbb{R}$ is convex and L-smooth, then we have

1. $0 \leq f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \leq \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|_2^2$
2. $f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{1}{2L} \|\nabla f(\mathbf{y}) - \nabla(\mathbf{x})\|_2^2 \leq f(\mathbf{y})$
3. $\frac{1}{L} \|\nabla f(\mathbf{y}) - \nabla(\mathbf{x})\|_2^2 \leq \langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle$

for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$.