

Optimization Theory

Lecture 02

Fudan University

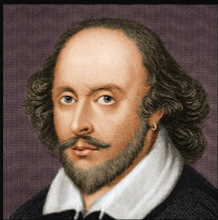
luoluo@fudan.edu.cn

Outline

- 1 Convex Set
- 2 Convex Function
- 3 Convex Optimization

Outline

- 1 Convex Set
- 2 Convex Function
- 3 Convex Optimization



**To quit, or not to quit, that
is the question.**

~Students

You can make the decision after the sections of convex analysis.

We say a set $\mathcal{C} \subseteq \mathbb{R}^n$ is convex if for all $\mathbf{x}, \mathbf{y} \in \mathcal{C}$ and $\alpha \in [0, 1]$, it holds that

$$\alpha \mathbf{x} + (1 - \alpha) \mathbf{y} \in \mathcal{C}.$$

Geometrically, a set \mathcal{C} is convex means that the line-segment connecting any two points in \mathcal{C} also belongs to \mathcal{C} .

Given any collection of convex sets (finite, countable or uncountable), their intersection is itself a convex set.

Projection

Given a closed and convex set $\mathcal{C} \subseteq \mathbb{R}^n$ and any point $\mathbf{y} \in \mathbb{R}^d$, we define the projection of \mathbf{y} onto \mathcal{C} in Euclidean norm as the point in \mathcal{C} that is closest to \mathbf{y} as

$$\text{proj}_{\mathcal{C}}(\mathbf{y}) = \arg \min_{\mathbf{x} \in \mathcal{C}} \|\mathbf{y} - \mathbf{x}\|_2^2.$$

Projection

Some properties of the projection:

- ① The projection $\text{proj}_{\mathcal{C}}(\mathbf{y})$ is uniquely defined.
- ② If $\mathbf{y} \notin \mathcal{C}$, then $\mathbf{z} = \text{proj}_{\mathcal{C}}(\mathbf{y})$ lies on the boundary of \mathcal{C} . The hyperplane

$$\{\mathbf{x} : \langle \mathbf{y} - \mathbf{z}, \mathbf{x} - \mathbf{z} \rangle = 0\}$$

separates \mathbf{y} and \mathcal{C} in that they lie on different sides, that is

$$\langle \mathbf{y} - \mathbf{z}, \mathbf{y} - \mathbf{z} \rangle > 0 \quad \text{and} \quad \langle \mathbf{y} - \mathbf{z}, \mathbf{x} - \mathbf{z} \rangle \leq 0$$

for any $\mathbf{x} \in \mathcal{C}$. It implies

$$\|\mathbf{x} - \mathbf{z}\|_2^2 \leq \|\mathbf{x} - \mathbf{y}\|_2^2$$

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

Outline

- 1 Convex Set
- 2 Convex Function
- 3 Convex Optimization

Convex Function

A function $f : \mathcal{C} \rightarrow \mathbb{R}$, defined on a convex set \mathcal{C} , is convex if it holds

$$f(\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}) \leq \alpha f(\mathbf{x}) + (1 - \alpha) f(\mathbf{y})$$

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

Epigraph

The epigraph of a function $f : \mathcal{C} \rightarrow \mathbb{R}$ is defined as the set

$$\text{epi } f \triangleq \{(\mathbf{x}, u) \in \mathcal{C} \times \mathbb{R} : f(\mathbf{x}) \leq u\}.$$

We say a function $f(\mathbf{x})$ is closed if its epigraph is closed.

Theorem

A function $f(\mathbf{x})$ is convex if and only if its epigraph is a convex set.

Extended Arithmetic Operations

We shall define convex function with possibly infinite values, which leads to arithmetic calculations involving $+\infty$ and $-\infty$:

- $-(-\infty) = +\infty$
- $\alpha \pm (+\infty) = (+\infty) \pm \alpha = +\infty$ for $\alpha \in \mathbb{R}$,
- $\alpha \pm (-\infty) = (-\infty) \pm \alpha = -\infty$ for $\alpha \in \mathbb{R}$,
- $\alpha \cdot (\pm\infty) = (\pm\infty) \cdot \alpha = \pm\infty$ for $\alpha \in (0, +\infty)$
- $\alpha \cdot (\pm\infty) = (\pm\infty) \cdot \alpha = \mp\infty$ for $\alpha \in (-\infty, 0)$
- $\alpha/(\pm\infty) = 0$ for $\alpha \in (-\infty, +\infty)$
- $(\pm\infty)/\alpha = \pm\infty$ for $\alpha \in (0, +\infty)$
- $(\pm\infty)/\alpha = \mp\infty$ for $\alpha \in (-\infty, 0)$
- $\inf \emptyset = \infty, \sup \emptyset = -\infty$

The extended real number system $\overline{\mathbb{R}}$, defined as

$$[-\infty, +\infty] \quad \text{or} \quad \mathbb{R} \cup \{-\infty, +\infty\}.$$

Extended Arithmetic Operations

The expressions

$$(+\infty) - (+\infty), \quad (-\infty) + (+\infty), \quad \frac{+\infty}{-\infty} \quad \text{and} \quad \frac{-\infty}{+\infty}.$$

are undefined and are avoided.

In the context of convex analysis, we also define

$$0 \cdot \infty = \infty \cdot 0 = 0 \quad \text{and} \quad 0 \cdot (-\infty) = (-\infty) \cdot 0 = 0.$$

Proper Convex Function

One may extend a convex function with domain $\mathcal{C} \subset \mathbb{R}^d$ to a proper convex function

$$f_{\mathcal{C}}(\mathbf{x}) = \begin{cases} f(\mathbf{x}), & \text{if } \mathbf{x} \in \mathcal{C}, \\ +\infty, & \text{if } \mathbf{x} \notin \mathcal{C}. \end{cases}$$

We define

$$\text{dom } f \triangleq \{\mathbf{x} : f(\mathbf{x}) < +\infty\}.$$

We say a convex function is proper if its domain is non-empty and its values are all larger than $-\infty$.

We say a function $f(\mathbf{x})$ on \mathbb{R}^d is concave if $-f(\mathbf{x})$ is convex. Linear functions are both convex and concave.

Some properties of convex function:

- ① Given any $\mathbf{g} : \mathbb{R}^d \rightarrow \mathbb{R}^k$ such that each component $g_j(\mathbf{x})$ is convex, then the set $\mathcal{C} = \{\mathbf{x} : \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\}$ is convex.
- ② The supremum over a family of convex functions is convex.
- ③ The positively weighted sum of convex functions is convex.
- ④ The partial minimization of a convex function is convex.
- ⑤ The composition of convex functions may not preserve convexity.

Indicator Function

Given a closed convex set $\mathcal{C} \in \mathbb{R}^d$, we can define a convex function $\mathbb{1}_{\mathcal{C}}(\mathbf{x})$ on \mathbb{R}^d , called the indicator function of \mathcal{C} on \mathbb{R}^d , as

$$\mathbb{1}_{\mathcal{C}}(\mathbf{x}) \triangleq \begin{cases} 0, & \text{if } \mathbf{x} \in \mathcal{C}, \\ +\infty, & \text{if } \mathbf{x} \notin \mathcal{C}. \end{cases}$$

We may write $f_{\mathcal{C}}(\mathbf{x}) = f(\mathbf{x}) + \mathbb{1}_{\mathcal{C}}(\mathbf{x})$ and the problem

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

is equivalent to

$$\min_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x}) + \mathbb{1}_{\mathcal{C}}(\mathbf{x}).$$

Closed Convex Function

We shall focus on closed functions in convex optimization.

- ① All convex functions can be made closed by taking the closure of its epigraph.
- ② In some pessimistic case, a closed convex function may not be continuous at the boundary of its domain. Consider the function

$$f(x, y) = \begin{cases} \frac{x^2}{y}, & (x, y) \neq (0, 0) \\ 0, & (x, y) = (0, 0) \end{cases}$$

with domain $\{(x, y) : y > 0\} \cup \{(0, 0)\}$.

- ③ We will only consider problems where the optimal solution can be achieved at a point that is continuous.

Outline

- 1 Convex Set
- 2 Convex Function
- 3 Convex Optimization

Why do we love convex optimization?

Theorem

Let $f(\mathbf{x})$ be a convex function defined on a convex set \mathcal{C} and \mathbf{x}^ be a local solution of*

$$\min_{\mathbf{x} \in \mathcal{C}} f(\mathbf{x}). \quad (1)$$

That is, there exist some $\delta > 0$ such that any $\hat{\mathbf{x}} \in \mathcal{B}_\delta(\mathbf{x}^) \cap \mathcal{C}$ holds*

$$f(\mathbf{x}^*) \leq f(\hat{\mathbf{x}}).$$

Then the local solution \mathbf{x}^ is a global solution of problem (1).*

Theorem

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

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle$$

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

However, the gradient may not exist in general case.

Subgradient and Subdifferential

We say a vector $\mathbf{g} \in \mathbb{R}^d$ is a subgradient of a proper convex function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ at $\mathbf{x} \in \text{dom } f$ if

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle$$

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

The set of subgradients at $\mathbf{x} \in \text{dom } f$ is called the subdifferential of f at \mathbf{x} , defined as

$$\partial f(\mathbf{x}) \triangleq \{ \mathbf{g} : f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle \text{ holds for any } \mathbf{y} \in \mathbb{R}^d \}.$$

Examples of Subdifferential

- ① The subdifferential of $f(x) = |x|$ at 0 is the set

$$\partial f(x) = [-1, 1].$$

What about the general norm?

- ② The subdifferential of an indicator function $\mathbb{1}_{\mathcal{C}}(\mathbf{x})$ is

$$\partial \mathbb{1}_{\mathcal{C}}(\mathbf{x}) = \mathcal{N}_{\mathcal{C}}(\mathbf{x}),$$

where

$$\mathcal{N}_{\mathcal{C}}(\mathbf{x}) = \{\mathbf{g} \in \mathbb{R}^d : \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle \leq 0 \text{ for all } \mathbf{y} \in \mathcal{C}\}$$

is called the normal cone of \mathcal{C} at \mathbf{x} .

- ③ If a convex function f is differentiable at \mathbf{x} , then

$$\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}.$$

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

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

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

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

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

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

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

$$\partial h(\mathbf{x}) \supseteq \mathbf{A}^\top \partial f_1(\mathbf{Ax} + \mathbf{b}).$$

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

$$\partial h(\mathbf{x}) = \mathbf{A}^\top \partial f_1(\mathbf{Ax} + \mathbf{b}).$$

Theorem

Consider proper closed convex function f and closed convex set $\mathcal{C} \subset (\text{dom } f)^\circ$. 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

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

Regularity Conditions

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

- 1 We say that a function $f : \mathcal{C} \rightarrow \mathbb{R}$ 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 $f : \mathbb{R}^d \rightarrow \mathbb{R}$ 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 $f : \mathcal{C} \rightarrow \mathbb{R}$ defined on convex set \mathcal{C} is μ -strongly-convex if and only if

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

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

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

holds 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 $\text{dom } f$ if

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

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

Theorem

A function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ 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 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$.

Theorem

A function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ 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 $f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{1}{2L} \|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|_2^2 \leq f(\mathbf{y})$
- ③ $\frac{1}{L} \|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|_2^2 \leq \langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle$

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

Twice Differentiable Function