Optimization Theory

Lecture 02

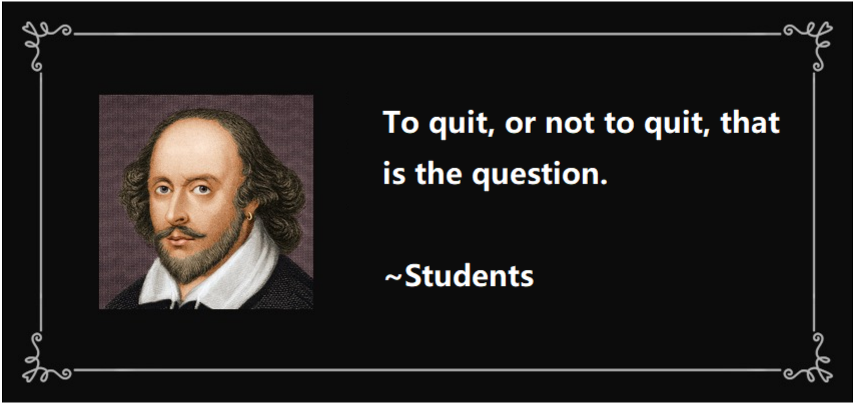
Fudan University

luoluo@fudan.edu.cn

Outline

Outline

Convex Analysis



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

Convex Set

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

$$\mathrm{proj}_{\mathcal{C}}(\mathbf{y}) = \operatorname*{arg\,min}_{\mathbf{x} \in \mathcal{C}} \|\mathbf{y} - \mathbf{x}\|_2^2 .$$

Projection

Some properties of the prjection:

1. The projection $\mathrm{proj}_{\mathcal{C}}(\mathbf{y})$ is uniquely defined.
2. If $\mathbf{y} \notin \mathcal{C}$, then $\mathbf{z} = \mathrm{proj}_{\mathcal{C}}(\mathbf{y})$ 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

Convex Function

A function $f : \mathcal{C} \to \mathbb{R}$, 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 $f : \mathcal{C} \to \mathbb{R}$ is defined as the set

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

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

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

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

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

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

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

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.

Convex Function

Some properties of convex function:

1. Given any $\mathbf{g} : \mathbb{R}^d \to \mathbb{R}^k$ 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

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

   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

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*

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

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

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

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

First-Order Condition

Theorem

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

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

hols 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 $f : \mathbb{R}^d \to \mathbb{R}$ at $\mathbf{x} \in \operatorname{dom} f$ 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 $\mathbf{x} \in \operatorname{dom} f$ is called the subdifferential of f at $\mathbf{x}$, defined as

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

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
$$\mathcal{N}_{\mathcal{C}}(\mathbf{x}) = \left\{\mathbf{g} \in \mathbb{R}^d : \langle \mathbf{g}, \mathbf{y} - \mathbf{x}\rangle \leq 0 \text{ for all } \mathbf{y} \in \mathcal{C}\right\}$$
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 $\mathrm{ri}(\mathrm{dom}\, f_1)$ 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 $\mathcal{C}$.

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

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

Subdifferential Calculus

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

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

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

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

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

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

Optimal Condition

Theorem

Consider proper closed convex function f and closed convex set $\mathcal{C} \subset (\mathrm{dom}\, f)^\circ$. 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 $y \in \mathcal{C}$ 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 $f : \mathcal{C} \to \mathbb{R}$ 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 $\operatorname{dom} f$ if

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

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

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

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

Second-Order Characterization

Theorem (Optimality)

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a twice differentiable function. Suppose that $\nabla^2 f(\cdot)$ is continuous in an open neighborhood of $\mathbf{x}^* \in \mathbb{R}^d$.

1. *If $\mathbf{x}^*$ is a local minimizer of $f(\cdot)$, then it holds that*

$$\nabla f(\mathbf{x}^*) = \mathbf{0} \qquad \text{and} \qquad \nabla^2 f(\mathbf{x}^*) \succeq \mathbf{0}.$$

2. *If it holds that*

$$\nabla f(\mathbf{x}^*) = \mathbf{0} \qquad \text{and} \qquad \nabla^2 f(\mathbf{x}^*) \succ \mathbf{0},$$

   then the point $\mathbf{x}^*$ is a strict local minimizer of $f(\cdot)$.

Second-Order Characterization

Theorem (Smoothness and Convexity)

Let $f(\cdot)$ be a twice differentiable function defined on $\mathbb{R}^d$

1. *It is L-smooth if and only if $-L\mathbf{I} \preceq \nabla^2 f(\mathbf{x}) \preceq L\mathbf{I}$ for all $\mathbf{x} \in \mathbb{R}^d$.*
2. *It is convex if and only if $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ for all $\mathbf{x} \in \mathbb{R}^d$.*
3. *It is μ-strongly-convex if and only if $\nabla^2 f(\mathbf{x}) \succeq \mu\mathbf{I}$ for all $\mathbf{x} \in \mathbb{R}^d$.*

We also say $f(\cdot)$ is ℓ-weakly convex if

$$\nabla^2 f(\mathbf{x}) \succeq -\ell\mathbf{I}.$$

for some $\ell > 0$.

Second-Order Characterization

Some examples:

1. For unconstrained quadratic problem

$$\min_{\mathbf{x}\in\mathbb{R}^d} f(\mathbf{x}) \triangleq \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} - \mathbf{b}^\top \mathbf{x},$$

   where $\mathbf{A} \in \mathbb{R}^{d\times d}$. We have

$$\nabla^2 f(\mathbf{x}) = \mathbf{A}.$$

2. For regularized generalized linear model

$$\min_{\mathbf{x}\in\mathbb{R}^d} f(\mathbf{x}) \triangleq \frac{1}{n}\sum_{i=1}^{n}\phi_i(\mathbf{a}^\top \mathbf{x}) + \frac{\lambda}{2}\|\mathbf{x}\|_2^2 .$$

   where $\phi_i : \mathbb{R}^d \to \mathbb{R}$ is twice differentiable. We have

$$\nabla f(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\phi_i'(\mathbf{a}_i^\top \mathbf{x})\mathbf{a} + \lambda\mathbf{x} \quad \text{and} \quad \nabla^2 f(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\phi_i''(\mathbf{a}_i^\top \mathbf{x})\mathbf{a}_i\mathbf{a}_i^\top + \lambda\mathbf{I}.$$