Introductory Lectures on Optimization

General Convex Problem (2)

Hui Qian
qianhui@zju.edu.cn

College of Computer Science, Zhejiang University

November 20, 2024

Outline

Part I
Continuity and Differentiability

Continuity

In this section we will see that the structure of convex functions in the interior of its domain is very simple.

Lemma 16 (Lemma 3.1.2)
Let function f be convex and $\boldsymbol{x}_0 \in \text{int}(\text{dom } f)$. Then f is locally upper bounded at $\boldsymbol{x}_0$.

Proof. Let us choose some $\epsilon > 0$ such that $\boldsymbol{x}_0 \pm \epsilon \boldsymbol{e}_i \in \text{int}(\text{dom } f)$, $i = 1 \dots n$, where $\boldsymbol{e}_i$ are the coordinate vectors of $\mathbb{R}^n$. Denote

$$\Delta = \text{Conv}\{\boldsymbol{x}_0 \pm \epsilon \boldsymbol{e}_i, i = 1 \dots n\}.$$

Let us show that $\boxed{\Delta \supset B_2(\boldsymbol{x}_0, \bar{\epsilon})}$ with $\bar{\epsilon} = \frac{\epsilon}{\sqrt{n}}$.

Continuity

Proof. (Continued.) Indeed, consider ($\boldsymbol{x} \in B_2$)

$$\boldsymbol{x} = \boldsymbol{x}_0 + \sum_{i=1}^{n} h_i \boldsymbol{e}_i, \quad \text{and let } \|\boldsymbol{x} - \boldsymbol{x}_0\|_2^2 = \sum_{i=1}^{n} (h_i)^2 \leq \bar{\epsilon}^2 = \frac{\epsilon^2}{n}.$$

We can assume that $h_i \geq 0$ (otherwise, in the above representation we can choose $-\boldsymbol{e}_i$ instead of $\boldsymbol{e}_i$). Then

$$\beta \equiv \sum_{i=1}^{n} h_i = \langle \mathbf{1},\ \boldsymbol{h} \rangle \leq \|\mathbf{1}\|_2 \, \|\boldsymbol{h}\|_2 = \sqrt{n} \sqrt{\sum_{i=1}^{n} (h_i)^2} \leq \epsilon.$$

Continuity

Proof. (Continued.) Therefore for $\bar{h}_i = \frac{1}{\beta} h_i$, we have

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{x}_0 + \beta \sum_{i=1}^{n} \bar{h}_i \boldsymbol{e}_i = \boldsymbol{x}_0 + \frac{\beta}{\epsilon} \sum_{i=1}^{n} \bar{h}_i \epsilon \boldsymbol{e}_i \\ &= \left(1 - \frac{\beta}{\epsilon}\right) \boldsymbol{x}_0 + \frac{\beta}{\epsilon} \sum_{i=1}^{n} \bar{h}_i (\boldsymbol{x}_0 + \epsilon \boldsymbol{e}_i) \in \Delta. \end{aligned}$$

Therefore, we have $\Delta \supset B_2(\boldsymbol{x}_0, \bar{\epsilon})$.

Continuity

Proof. (Continued.) Thus, using Corollary (3.1.2), we obtain

$$M \equiv \max_{\boldsymbol{x} \in B_2(\boldsymbol{x}_0, \bar{\epsilon})} f(\boldsymbol{x}) \leq \max_{\boldsymbol{x} \in \Delta} f(\boldsymbol{x}) = \max_{1 \leq i \leq n} f(\boldsymbol{x}_0 \pm \epsilon \boldsymbol{e}_i).$$

$\square$

Corollary (3.1.2): Let

$$\Delta = \mathrm{Conv}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\} \equiv \{\boldsymbol{x} = \sum_{i=1}^{m} \alpha_i \boldsymbol{x}_i | \alpha_i \geq 0, \sum_{i=1}^{m} \alpha_i = 1\},$$

then $\max_{\boldsymbol{x} \in \Delta} f(\boldsymbol{x}) = \max_{1 \leq i \leq n} f(\boldsymbol{x}_i)$.

Continuity

The above result implies continuity of a convex function at any interior point of its domain.

Theorem 17 (Theorem 3.1.8)
Let f be convex and $\boldsymbol{x}_0 \in \operatorname{int}(\operatorname{dom} f)$. Then f is locally Lipschitz continuous at $\boldsymbol{x}_0$.

Proof. Let $B_2(\boldsymbol{x}_0, \epsilon) \subset \operatorname{dom} f$ and $\sup\{f(\boldsymbol{x}) | \boldsymbol{x} \in B_2(\boldsymbol{x}_0, \epsilon)\} \leq M$ (in view of Lemma 16, M is finite). Consider $\boldsymbol{y} \in B_2(\boldsymbol{x}_0, \epsilon), \boldsymbol{y} \neq \boldsymbol{x}_0$. Denote

$$\alpha = \frac{1}{\epsilon} \|\boldsymbol{y} - \boldsymbol{x}_0\|, \quad z = \boldsymbol{x}_0 + \frac{1}{\alpha}(\boldsymbol{y} - \boldsymbol{x}_0).$$

Clearly, $\|z - \boldsymbol{x}_0\| = \frac{1}{\alpha} \|\boldsymbol{y} - \boldsymbol{x}_0\| = \epsilon$.
Remark. 构造的 z 没有超出邻域。$\boldsymbol{y}$ 是 $\boldsymbol{x}_0$ 和 z 的凸组合。

Continuity

Proof. (Continued.) Therefore $\alpha \le 1$ and $\boldsymbol{y} = \alpha z + (1-\alpha)\boldsymbol{x}_0$. Hence,

$$\begin{aligned} f(\boldsymbol{y}) &\le \alpha f(z) + (1-\alpha) f(\boldsymbol{x}_0) \\ &= f(\boldsymbol{x}_0) + \alpha(f(z) - f(\boldsymbol{x}_0)) \\ &\le f(\boldsymbol{x}_0) + \alpha(M - f(\boldsymbol{x}_0)) \\ &= \boxed{f(\boldsymbol{x}_0) + \frac{M - f(\boldsymbol{x}_0)}{\epsilon} \|\boldsymbol{y} - \boldsymbol{x}_0\|}. \end{aligned}$$

Continuity

Proof. (Continued.) Further, denote $u = \boldsymbol{x}_0 + \frac{1}{\alpha}(\boldsymbol{x}_0 - \boldsymbol{y})$. Then $\|u - \boldsymbol{x}_0\| = \epsilon$ and $\boldsymbol{y} = \boldsymbol{x}_0 + \alpha(\boldsymbol{x}_0 - u)$. Therefore, in view of Theorem 3.1.1, we have

$$\begin{aligned} f(\boldsymbol{y}) &\geq f(\boldsymbol{x}_0) + \alpha(f(\boldsymbol{x}_0) - f(u)) \\ &\geq f(\boldsymbol{x}_0) - \alpha(M - f(\boldsymbol{x}_0)) \\ &= \boxed{f(\boldsymbol{x}_0) - \frac{M - f(\boldsymbol{x}_0)}{\epsilon}\|\boldsymbol{y} - \boldsymbol{x}_0\|}. \end{aligned}$$

Thus, $|f(\boldsymbol{y}) - f(\boldsymbol{x}_0)| \leq \frac{M - f(\boldsymbol{x}_0)}{\epsilon}\|\boldsymbol{y} - \boldsymbol{x}_0\|$. □

Differentiability of Convex Function

The convex functions possess a property, which is very close to differentiability.

Definition 18 (Definition 3.1.3)

Let $\boldsymbol{x} \in \mathrm{dom}\ f$. We call f differentiable in a direction p at point $\boldsymbol{x}$ if the following limit exists:

$$f'(\boldsymbol{x};\boldsymbol{p}) = \lim_{a \downarrow 0} \frac{1}{\alpha}[f(\boldsymbol{x}+\alpha\boldsymbol{p}) - f(\boldsymbol{x})]. \tag{5}$$

The value $f'(\boldsymbol{x};\boldsymbol{p})$ is called the directional derivative of f at $\boldsymbol{x}$.

Differentiability of Convex Function

Theorem 19 (Theorem 3.1.9)

Convex function f is differentiable in any direction at any interior point of its domain.

Proof. let $\boldsymbol{x} \in \text{int}(\text{dom } f)$. Consider the function

$$\phi(\alpha) = \frac{1}{\alpha}[f(\boldsymbol{x} + \alpha \boldsymbol{p}) - f(\boldsymbol{x})], \quad \alpha > 0,$$

(prove that ϕ is descending.) Let $\beta \in (0, 1]$ and $\alpha \in (0, \epsilon]$, ϵ be small enough to have $\boldsymbol{x} + \epsilon \boldsymbol{p} \in \text{dom } f$. Then

$$f(\boldsymbol{x} + \alpha\beta \boldsymbol{p}) = f\left((1 - \beta)\boldsymbol{x} + \beta(\boldsymbol{x} + \alpha \boldsymbol{p})\right) \leq (1 - \beta) f(\boldsymbol{x}) + \beta f(\boldsymbol{x} + \alpha \boldsymbol{p}). \tag{6}$$

Differentiability of Convex Function

Proof. (Continued.) Therefore,

$$\phi(\alpha\beta) = \frac{1}{\alpha\beta}[f(\boldsymbol{x}+\alpha\beta\boldsymbol{p}) - f(\boldsymbol{x})] \underbrace{\leq}_{\text{by (6)}} \frac{1}{\alpha}[f(\boldsymbol{x}+\alpha\boldsymbol{p}) - f(\boldsymbol{x})] = \phi(\alpha).$$

Thus $\phi(\alpha)$ decreases as $\alpha \downarrow 0$. Let us choose $\gamma > 0$ small enough to have $\boldsymbol{x} - \gamma\boldsymbol{p} \in \operatorname{dom} f$. Then, in view of (3.1.3), we have(see the remark in next page)

$$\phi(\alpha) \geq \frac{1}{\gamma}[f(\boldsymbol{x}) - f(\boldsymbol{x}-\gamma\boldsymbol{p})].$$

Hence, the limit in (5) exists. (Monotonically decreasing and bounded, the limit exists) $\square$

Differentiability of Convex Function

Remarks.

In view of (3:1:3), we have

$$f(\boldsymbol{x}+\beta(\boldsymbol{x}-\boldsymbol{y}))-f(\boldsymbol{x}) \geq \beta(f(\boldsymbol{x})-f(\boldsymbol{y})),$$
$$\frac{1}{\alpha}\{f(\boldsymbol{x}+\beta(\boldsymbol{x}-\boldsymbol{y}))-f(\boldsymbol{x})\} \geq \frac{\beta}{\alpha}(f(\boldsymbol{x})-f(\boldsymbol{y})).$$

Let $\gamma=\frac{\alpha}{\beta}$ and $\boldsymbol{y}=\boldsymbol{x}-\gamma\boldsymbol{p}$. We obtain

$$\frac{1}{\alpha}\{f(\boldsymbol{x}+\beta\gamma\boldsymbol{p})-f(\boldsymbol{x})\} \geq \frac{1}{\gamma}(f(\boldsymbol{x})-f(\boldsymbol{x}-\gamma\boldsymbol{p})).$$

Differentiability of Convex Function

The directional derivative provides us a global lower support of the convex function.

Lemma 20 (Lemma 3.1.3)

Let f be a convex function and $\boldsymbol{x} \in \operatorname{int}(\operatorname{dom} f)$. Then $f'(\boldsymbol{x};\boldsymbol{p})$ is a convex function of $\boldsymbol{p}$, which is homogeneaus of degree 1. For any $\boldsymbol{y} \in \operatorname{dom} f$ we have

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + f'(\boldsymbol{x};\boldsymbol{y}-\boldsymbol{x}). \tag{7}$$

Proof. (Homogeneous:) Indeed, for $\boldsymbol{p} \in \mathbb{R}^n$ 和 $\tau > 0$, we have

$$\begin{aligned} f'(\boldsymbol{x};\tau\boldsymbol{p}) &= \lim_{\alpha\downarrow 0} \frac{1}{\alpha}[f(\boldsymbol{x}+\tau\alpha\boldsymbol{p}) - f(\boldsymbol{x})] \\ &= \tau \lim_{\beta\downarrow 0} \frac{1}{\beta}[f(\boldsymbol{x}+\beta\boldsymbol{p}) - f(\boldsymbol{x})] = \tau f'(\boldsymbol{x};\boldsymbol{p}). \end{aligned}$$

Differentiability of Convex Function

Proof. (Continued.) Further, for any $\boldsymbol{p}_1, \boldsymbol{p}_2 \in \mathbb{R}^n$ and $\beta \in [0,1]$ we obtain

$$
\begin{aligned}
f'(\boldsymbol{x}; \beta \boldsymbol{p}_1 + (1-\beta)\boldsymbol{p}_2) &= \lim_{\alpha \downarrow 0} \frac{1}{\alpha} [f(\boldsymbol{x} + \alpha(\beta \boldsymbol{p}_1 + (1-\beta)\boldsymbol{p}_2)) - f(\boldsymbol{x})] \\
&\le \lim_{\alpha \downarrow 0} \frac{1}{\alpha} \{\beta[f(\boldsymbol{x} + \alpha \boldsymbol{p}_1) - f(\boldsymbol{x})] \\
&\qquad + (1-\beta)[f(\boldsymbol{x} + \alpha \boldsymbol{p}_2) - f(\boldsymbol{x})]\} \\
&= \beta f'(\boldsymbol{x}; \boldsymbol{p}_1) + (1-\beta) f'(\boldsymbol{x}; \boldsymbol{p}_2).
\end{aligned}
$$

Thus, $f'(\boldsymbol{x}; \boldsymbol{p})$ is convex in p.

$$
\begin{aligned}
& f\left(\boldsymbol{x} + \alpha(\beta \boldsymbol{p}_1 + (1-\beta)\boldsymbol{p}_2)\right) - f(\boldsymbol{x}) \\
&= f\left(\beta \boldsymbol{x} + (1-\beta)\boldsymbol{x} + \alpha(\beta \boldsymbol{p}_1 + (1-\beta)\boldsymbol{p}_2)\right) - \beta f(\boldsymbol{x}) - (1-\beta) f(\boldsymbol{x})
\end{aligned}
$$

Differentiability of Convex Function

Proof. (Continued.) Finally, let $\alpha \in (0, 1]$, $\boldsymbol{y} \in \operatorname{dom} f$, and $\boldsymbol{y}_\alpha = \boldsymbol{x} + \alpha(\boldsymbol{y} - \boldsymbol{x})$. Then in view of Theorem 3.1.1, we have

$$
\begin{aligned}
f(\boldsymbol{y}) &= f\left(\boldsymbol{y}_\alpha + \frac{1}{\alpha}(1-\alpha)(\boldsymbol{y}_\alpha - \boldsymbol{x})\right) \\
&\geq f(\boldsymbol{y}_\alpha) + \frac{1}{\alpha}(1-\alpha)[f(\boldsymbol{y}_\alpha) - f(\boldsymbol{x})], \\
&= f(\boldsymbol{x} + \alpha(\boldsymbol{y} - \boldsymbol{x})) + (1-\alpha)\left(\frac{1}{\alpha}[f(\boldsymbol{x} + \alpha(\boldsymbol{y} - \boldsymbol{x})) - f(\boldsymbol{x})]\right),
\end{aligned}
$$

and we get (7) taking the limit in $\alpha \downarrow 0$. $\square$

Part II
Separation Theorems

Hyperplane

Definition 21 (Definition 3.1.4)
Let Q be a convex set. We say that hyperplane

$$\mathcal{H}(g,\gamma) = \{\boldsymbol{x} \in \mathbb{R}^n | \langle g,\ \boldsymbol{x}\rangle = \gamma\}, \quad g \neq 0,$$

is supporting to Q if any $\boldsymbol{x} \in Q$ satisfies inequality $\langle g,\ \boldsymbol{x}\rangle \leq \gamma$.

We say that the hyperplane $\mathcal{H}(g,\gamma)$ separates a point $\boldsymbol{x}_0$ from Q if

$$\langle g,\ \boldsymbol{x}\rangle \leq \gamma \leq \langle g,\ \boldsymbol{x}_0\rangle. \tag{8}$$

for all $\boldsymbol{x} \in Q$. If the second inequality in (8) is strict, we call the separation strict.

Projection

Definition 22 (Definition 3.1.5)

Let Q be a closed set and $\boldsymbol{x}_0 \in \mathbb{R}^n$. Denote

$$\pi_Q(\boldsymbol{x}_0) = \operatorname{argmin}\{\|\boldsymbol{x} - \boldsymbol{x}_0\| : \boldsymbol{x} \in Q\}.$$

We call $\pi_Q(\boldsymbol{x}_0)$ the projection of point $\boldsymbol{x}_0$ onto the set Q.

Projection

Theorem 23 (Theorem 3.1.10)
If Q is a convex set, then there exists a unique projection $\pi_Q(\boldsymbol{x}_0)$.

Proof. Indeed, $\pi_Q(\boldsymbol{x}_0) = \operatorname{argmin}\{\phi(\boldsymbol{x}) | \boldsymbol{x} \in Q\}$, where the function $\phi(\boldsymbol{x}) = \frac{1}{2}\left\|\boldsymbol{x} - \boldsymbol{x}_0\right\|^2$ belongs to $\mathcal{S}_{1,1}^{1,1}(\mathbb{R}^n)$. Therefore, $\pi_Q(\boldsymbol{x}_0)$ is unique and well defined in view of Theorem 2.2.6 $\square$

Remark. It is clear that $\pi_Q(\boldsymbol{x}_0) = \boldsymbol{x}_0$ if and only if $\boldsymbol{x}_0 \in Q$.

Projection

Lemma 24 (Lemma 3.1.4)

Let Q be a closed and convex set and $\boldsymbol{x}_0 \notin Q$. Then for any $\boldsymbol{x} \in Q$ we have

$$\langle \pi_Q(\boldsymbol{x}_0) - \boldsymbol{x}_0,\ \boldsymbol{x} - \pi_Q(\boldsymbol{x}_0) \rangle \geq 0. \tag{9}$$

Proof. Note that $\pi_Q(\boldsymbol{x}_0)$ is a solution to the minimization problem $\min_{\boldsymbol{x}\in Q} \phi(\boldsymbol{x})$ with $\phi(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}_0\|^2$. Therefore, in view of Theorem 2.2.5 we have

$$\langle \nabla\phi(\pi_Q(\boldsymbol{x}_0)),\ \boldsymbol{x} - \pi_Q(\boldsymbol{x}_0) \rangle \geq 0,$$

for all $\boldsymbol{x} \in Q$. It remains to note that $\nabla\phi(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{x}_0$. □

定理 2.2.5: 从最小点出发，都是增长方向。

Projection

Lemma 25 (Lemma 3.1.5)

For any $\boldsymbol{x} \in Q$ we have

$$\|\boldsymbol{x} - \pi_Q(\boldsymbol{x}_0)\|^2 + \|\pi_Q(\boldsymbol{x}_0) - \boldsymbol{x}_0\|^2 \leq \|\boldsymbol{x} - \boldsymbol{x}_0\|^2 .$$

Proof. Indeed, in view of (9) we have

$$\begin{aligned}\|\boldsymbol{x} - \pi_Q(\boldsymbol{x}_0)\|^2 - \|\boldsymbol{x} - \boldsymbol{x}_0\|^2 &= \langle \boldsymbol{x}_0 - \pi_Q(\boldsymbol{x}_0),\ 2\boldsymbol{x} - \pi_Q(\boldsymbol{x}_0) - \boldsymbol{x}_0\rangle \\ &= \langle \boldsymbol{x}_0 - \pi_Q(\boldsymbol{x}_0),\ 2\boldsymbol{x} - 2\pi_Q(\boldsymbol{x}_0) + \pi_Q(\boldsymbol{x}_0) - \boldsymbol{x}_0\rangle \\ &\leq -\|\boldsymbol{x}_0 - \pi_Q(\boldsymbol{x}_0)\|^2 .(\text{ from (9)})\end{aligned}$$

□

Remark. $\langle \boldsymbol{x}_0 - \pi_Q(\boldsymbol{x}_0),\ 2\boldsymbol{x} - 2\pi_Q(\boldsymbol{x}_0)\rangle + \langle \boldsymbol{x}_0 - \pi_Q(\boldsymbol{x}_0),\ \pi_Q(\boldsymbol{x}_0) - \boldsymbol{x}_0\rangle$

Main Theorems

Theorem 26 (Theorem 3.1.11)

Let Q be a closed convex set and $\boldsymbol{x}_0 \notin Q$. Then there exists a hyperplane $\mathcal{H}(\boldsymbol{g}, \gamma)$, which strictly separates $\boldsymbol{x}_0$ from Q. Namely, we can take

$$\boldsymbol{g} = \boldsymbol{x}_0 - \pi_Q(\boldsymbol{x}_0) \neq 0, \quad \gamma = \langle \boldsymbol{x}_0 - \pi_Q(\boldsymbol{x}_0),\ \pi_Q(\boldsymbol{x}_0) \rangle.$$

Proof. Indeed, in view of (9), for any $\boldsymbol{x} \in Q$ we have

$$\begin{aligned}
\underbrace{\langle \boldsymbol{x}_0 - \pi_Q(\boldsymbol{x}_0),\ \boldsymbol{x} \rangle}_{\langle g,\ \boldsymbol{x} \rangle} &\leq \underbrace{\langle \boldsymbol{x}_0 - \pi_Q(\boldsymbol{x}_0),\ \pi_Q(\boldsymbol{x}_0) \rangle}_{\gamma} = \underbrace{\langle \boldsymbol{x}_0 - \pi_Q(\boldsymbol{x}_0),\ \boldsymbol{x}_0 \rangle}_{\langle g,\ \boldsymbol{x}_0 \rangle} - \|\boldsymbol{x}_0 - \pi_Q(\boldsymbol{x}_0)\|^2 \\
&\leq \underbrace{\langle \boldsymbol{x}_0 - \pi_Q(\boldsymbol{x}_0),\ \boldsymbol{x}_0 \rangle}_{\langle g,\ \boldsymbol{x}_0 \rangle}
\end{aligned}$$

Main Theorems

$$\psi_Q(\boldsymbol{g}) = \sup\{\langle \boldsymbol{g},\ \boldsymbol{x}\rangle | \boldsymbol{x} \in Q\}.$$

Let us give an example of an application of the above theorem.

Corollary 27 (Corollary 3.1.3)

Let Q_1 and Q_2 be two closed convex sets.

1. If for any $\boldsymbol{g} \in \mathrm{dom}\ \psi_{Q_2}$ we have $\psi_{Q_1}(\boldsymbol{g}) \leq \psi_{Q_2}(\boldsymbol{g})$, then $Q_1 \subseteq Q_2$.
2. Let $\mathrm{dom}\ \psi_{Q_1} = \mathrm{dom}\ \psi_{Q_2}$ and for any $\boldsymbol{g} \in \mathrm{dom}\ \psi_{Q_1}$, we have $\psi_{Q_1}(\boldsymbol{g}) = \psi_{Q_2}(\boldsymbol{g})$. Then $Q_1 \equiv Q_2$.

Main Theorems

Theorem 28 (Theorem 3.1.12)

Let Q be a closed and convex set, and $\boldsymbol{x}_0$ belong to the boundary of set Q. Then there exists a hyperplane $\mathcal{H}(\boldsymbol{g}, \gamma)$, supporting to Q and passing through $\boldsymbol{x}_0$.
(Such a vector $\boldsymbol{g}$ is called supporting to Q at $\boldsymbol{x}_0$)

Proof. Consider a sequence $\{\boldsymbol{y}_k\}$ such that $\boldsymbol{y}_k \neq Q$ and $\boldsymbol{y}_k \to \boldsymbol{x}_0$. Denote

$$\boldsymbol{g}_k = \frac{\boldsymbol{y}_k - \pi_Q(\boldsymbol{y}_k)}{\|\boldsymbol{y}_k - \pi_Q(\boldsymbol{y}_k)\|}, \quad \gamma_k = \langle \boldsymbol{g}_k,\ \pi_Q(\boldsymbol{y}_k)\rangle.$$

In view of Theorem 26, for all $\boldsymbol{x} \in Q$ we have

$$\langle \boldsymbol{g}_k,\ \boldsymbol{x}\rangle \leq \gamma_k \leq \langle \boldsymbol{g}_k,\ \boldsymbol{y}_k\rangle. \tag{10}$$

Main Theorems

Proof. (Continued.)
However, $\|\boldsymbol{g}_k\| = 1$ and the sequence $\{\gamma_k\}$ is bounded:

$$\begin{aligned} |\gamma_k| &= |\langle \boldsymbol{g}_k, \pi_Q(\boldsymbol{y}_k) - \boldsymbol{x}_0 \rangle + \langle \boldsymbol{g}_k, \boldsymbol{x}_0 \rangle| \\ \text{(Lemma 3.1.5)} \quad &\le \|\pi_Q(\boldsymbol{y}_k) - \boldsymbol{x}_0\| + \|\boldsymbol{x}_0\| \le \|\boldsymbol{y}_k - \boldsymbol{x}_0\| + \|\boldsymbol{x}_0\| . \end{aligned}$$

Therefore, without loss of generality we can assume that there exist $\boldsymbol{g}^* = \lim_{k\to\infty} \boldsymbol{g}_k$ and $\gamma^* = \lim_{k\to\infty} \gamma_k$. It remains to take the limit in (10). □

Remark.

$$\langle \boldsymbol{g}^*, \boldsymbol{x} \rangle \le \gamma^* \le \langle \boldsymbol{g}^*, \boldsymbol{x}_0 \rangle.$$

Main Theorems

Rmark. From Lemma 3.1.5, we have

$$\|\boldsymbol{x} - \pi_Q(\boldsymbol{x}_0)\|^2 + \|\pi_Q(\boldsymbol{x}_0) - \boldsymbol{x}_0\|^2 \leq \|\boldsymbol{x} - \boldsymbol{x}_0\|^2 .$$

That is

$$\|\boldsymbol{x} - \pi_Q(\boldsymbol{x}_0)\|^2 \leq \|\boldsymbol{x} - \boldsymbol{x}_0\|^2 .$$

In this context, $\boldsymbol{x} \equiv \boldsymbol{x}_0$ and $\boldsymbol{x}_0 \equiv \boldsymbol{y}_k$. Thus we arrive at

$$\|\boldsymbol{x}_0 - \pi_Q(\boldsymbol{y}_k)\|^2 \leq \|\boldsymbol{x}_0 - \boldsymbol{y}_k\|^2 .$$

Also

$$\|\pi_Q(\boldsymbol{y}_k) - \boldsymbol{x}_0\| \leq \|\boldsymbol{y}_k - \boldsymbol{x}_0\| .$$

Part III
Subdifferential and Subgradient

Definition

Definition 29 (Definition 3.1.6)

Let f be a convex function. A vector g is called a subgradient of function f at point $\boldsymbol{x}_0 \in \operatorname{dom} f$ if for any $\boldsymbol{x} \in \operatorname{dom} f$ we have

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}_0) + \langle \boldsymbol{g}, \boldsymbol{x} - \boldsymbol{x}_0 \rangle. \tag{11}$$

The set of all subgradients of f at $\boldsymbol{x}_0$, is called the subdifferential of function f at point $\boldsymbol{x}_0$.

Subgradient and Convexity

The subdifferentiability of a function implies convexity.

Lemma 30 (Lemma 3.1.6)

Let for any $\boldsymbol{x} \in \operatorname{dom} f$ subdifferential $\partial f(\boldsymbol{x})$ be nonempty. Then f is a convex function.

Proof. Indeed, let $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom} f, \alpha \in [0,1]$. Consider $\boldsymbol{y}_\alpha = \boldsymbol{x} + \alpha(\boldsymbol{y} - \boldsymbol{x})$. Let $\boldsymbol{g} \in \partial f(\boldsymbol{y}_\alpha)$. Then

$$
\begin{aligned}
f(\boldsymbol{y}) &\geq f(\boldsymbol{y}_\alpha) + \langle \boldsymbol{g}, \boldsymbol{y} - \boldsymbol{y}_\alpha \rangle = f(\boldsymbol{y}_\alpha) + (1-\alpha)\langle \boldsymbol{g}, \boldsymbol{y} - \boldsymbol{x} \rangle, \\
f(\boldsymbol{x}) &\geq f(\boldsymbol{y}_\alpha) + \langle \boldsymbol{g}, \boldsymbol{x} - \boldsymbol{y}_\alpha \rangle = f(\boldsymbol{y}_\alpha) - \alpha\langle \boldsymbol{g}, \boldsymbol{y} - \boldsymbol{x} \rangle,
\end{aligned}
$$

Adding these inequalities multiplied by α and $(1-\alpha)$ respectively, we get

$$
\alpha f(\boldsymbol{y}) + (1-\alpha) f(\boldsymbol{x}) \geq f(\boldsymbol{y}_\alpha).
$$

Subgradient and Convexity

Theorem 31 (Theorem 3.1.13)

Let f be closed and convex and $\boldsymbol{x}_0 \in \operatorname{int}(\operatorname{dom} f)$. Then $\partial f(\boldsymbol{x}_0)$ is a nonempty bounded set.

Proof. Note that the point $(f(\boldsymbol{x}_0), \boldsymbol{x}_0)$ belongs to the boundary of $\operatorname{epi}(f)$. Hence, in view of Theorem 3.1.12, there exists a hyperplane supporting to $\operatorname{epi}(f)$ at $(f(\boldsymbol{x}_0), \boldsymbol{x}_0)$:

$$\langle(-\alpha, \boldsymbol{d}),\ (\tau, \boldsymbol{x})\rangle \leq \langle(-\alpha, \boldsymbol{d}),\ (f(\boldsymbol{x}_0), \boldsymbol{x}_0)\rangle.$$

$$\text{That is, } -\alpha\tau + \langle \boldsymbol{d},\ \boldsymbol{x}\rangle \leq -\alpha f(\boldsymbol{x}_0) + \langle \boldsymbol{d},\ \boldsymbol{x}_0\rangle, \tag{12}$$

for all $(\tau, \boldsymbol{x}) \in \operatorname{epi}(f)$. Note that we can take

$$\|(-\alpha, \boldsymbol{d})\|^2 = \|\boldsymbol{d}\|^2 + \alpha^2 = 1. \tag{13}$$

注：可以归一化。

Subgradient and Convexity

Proof. (Continued.) Since for all $\tau \geq f(\boldsymbol{x}_0)$ the point $(\tau, \boldsymbol{x}_0)$ belongs to $\mathrm{epi}(f)$, we conclude that $\alpha \geq 0$.

Remark. since $-\alpha\tau + \langle \boldsymbol{d},\ \boldsymbol{x}_0\rangle \leq -\alpha f(\boldsymbol{x}_0) + \langle \boldsymbol{d},\ \boldsymbol{x}_0\rangle$, we have $-\alpha\tau \leq -\alpha f(\boldsymbol{x}_0)$, that is $\alpha(\tau - f(\boldsymbol{x}_0)) \geq 0$.

Now, we prove $\alpha > 0$: Recall, that a convex function is locally Lipschitz continuous at the interior of its domain (Theorem 17). This means that there exist some $\epsilon > 0$ and $M > 0$ such that $B_2(\boldsymbol{x}_0, \epsilon) \subseteq \mathrm{dom}\ f$ and

$$f(\boldsymbol{x}) - f(\boldsymbol{x}_0) \leq M \left\|\boldsymbol{x} - \boldsymbol{x}_0\right\|.$$

for all $\boldsymbol{x} \in B_2(\boldsymbol{x}_0, \epsilon)$.

Subgradient and Convexity

Proof. (Continued.) Therefore, in view of (12), for any $\boldsymbol{x}$ from this ball we have

$$\langle \boldsymbol{d},\ \boldsymbol{x} - \boldsymbol{x}_0 \rangle \leq \alpha(f(\boldsymbol{x}) - f(\boldsymbol{x}_0)) \leq \alpha M \|\boldsymbol{x} - \boldsymbol{x}_0\| .$$

Remark. The first inequality comes from the fact that $(f(\boldsymbol{x}), \boldsymbol{x})$ also belongs to $\mathrm{epi}(f)$. We can obtain it from (12).

Choosing $\boldsymbol{x} = \boldsymbol{x}_0 + \epsilon \boldsymbol{d}$ we get $\|\boldsymbol{d}\|^2 \leq M\alpha \|\boldsymbol{d}\|$. Thus, in view of the normalizing condition (13) we obtain

$$\alpha \geq \frac{1}{\sqrt{1 + M^2}} > 0.$$

Subgradient and Convexity

Proof. (Continued.)

Prove the non-empty: Choosing $\boldsymbol{g} = \boldsymbol{d}/\alpha$ we get

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}_0) + \langle \boldsymbol{g}, \ \boldsymbol{x} - \boldsymbol{x}_0 \rangle.$$

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

Prove the bounded: Finally, if $\boldsymbol{g} \in \partial f(\boldsymbol{x}_0), \boldsymbol{g} \neq 0$, then choosing $\boldsymbol{x} = \boldsymbol{x}_0 + \epsilon \boldsymbol{g} / \|g\|$ we obtain

$$\epsilon \|\boldsymbol{g}\| = \langle \boldsymbol{g}, \ \boldsymbol{x} - \boldsymbol{x}_0 \rangle \leq f(\boldsymbol{x}) - f(\boldsymbol{x}_0) \leq M \|\boldsymbol{x} - \boldsymbol{x}_0\| = M\epsilon.$$

Thus, $\partial f(\boldsymbol{x}_0)$ is bounded. $\square$

Subgradient and Convexity

Let us show that the conditions of the above theorem cannot be relaxed.

Example 32 (Example 3.1.4)

Consider the function $f(\boldsymbol{x}) = -\sqrt{\boldsymbol{x}}$ with the domain $\{\boldsymbol{x} \in \mathbb{R}^1 | \boldsymbol{x} \geq 0\}$. This function is convex and closed, but the subdifferential does not exist at $\boldsymbol{x} = 0$.

Remark.

1. $\nabla^2 f(\boldsymbol{x}) = \frac{1}{4}\boldsymbol{x}^{-\frac{3}{2}}$. Therefore when $\boldsymbol{x} \geq 0$, it is convex. Since $f(\boldsymbol{x})$ is continuous, it is closed.
2. At $\boldsymbol{x}_0 = 0$, if $-\sqrt{\boldsymbol{x}} \geq 0 + \boldsymbol{g}\boldsymbol{x}$, we have $g \leq -\frac{1}{\sqrt{\boldsymbol{x}}}, \boldsymbol{x} \geq 0$. Thus, we arrive at $g \leq -\infty$. Therefore, subgradient can not be obtained.

Thus, even function is closed and convex, the sub-differential may be empty at non-interior points.

Subgradient and Convexity

Let us deterine an important relation between the subdifferential and the directional derivative of convex function.

Theorem 33 (Theorem 3.1.14)

Let f be a closed convex function. For any $\boldsymbol{x}_0 \in \operatorname{int}(\operatorname{dom} f)$ and $\boldsymbol{p} \in \mathbb{R}^n$ we have

$$f'(\boldsymbol{x}_0; \boldsymbol{p}) = \max\{\langle \boldsymbol{g},\ \boldsymbol{p}\rangle \,\big|\, \boldsymbol{g} \in \partial f(\boldsymbol{x}_0)\}.$$

$$f'(\boldsymbol{x}; \boldsymbol{p}) = \lim_{a \downarrow 0} \frac{1}{\alpha}[f(\boldsymbol{x} + \alpha \boldsymbol{p}) - f(\boldsymbol{x})].$$

Properties

Theorem 34 (Theorem 3.1.15)

We have $f(\boldsymbol{x}^*) = \min_{\boldsymbol{x} \in \operatorname{dom} f} f(\boldsymbol{x})$ if and only if

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

Proof.

Indeed, if $0 \in \partial f(\boldsymbol{x}^*)$, then $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \langle 0, \boldsymbol{x} - \boldsymbol{x}^* \rangle = f(\boldsymbol{x}^*)$ for all $\boldsymbol{x} \in \operatorname{dom} f$.

On the other hand, if $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$ for all $\boldsymbol{x} \in \operatorname{dom} f$, then $0 \in \partial f(\boldsymbol{x}^*)$ in view of Definition 29. $\square$

Properties

Theorem 35 (Theorem 3.1.16)

For any $\boldsymbol{x}_0 \in \operatorname{dom} f$, all vectors $\boldsymbol{g} \in \partial f(\boldsymbol{x}_0)$ are supporting to the level set $\mathcal{L}_f(f(\boldsymbol{x}_0))$:

$$\langle \boldsymbol{g}, \ \boldsymbol{x}_0 - \boldsymbol{x} \rangle \geq 0, \quad \forall \boldsymbol{x} \in \mathcal{L}_f(f(\boldsymbol{x}_0)) \equiv \{\boldsymbol{x} \in \operatorname{dom} f : f(\boldsymbol{x}) \leq f(\boldsymbol{x}_0)\}.$$

Corollary 36 (Corollary 3.1.4)

Let $Q \subseteq \operatorname{dom} f$ be a closed convex set, $\boldsymbol{x}_0 \in Q$ and

$$\boldsymbol{x}^* = \operatorname{argmin}\{f(\boldsymbol{x}) | \boldsymbol{x} \in Q\}.$$

Then for any $g \in \partial f(\boldsymbol{x}_0)$ we have $\langle \boldsymbol{g}, \ \boldsymbol{x}_0 - \boldsymbol{x}^* \rangle \geq 0$.

Rules for computing

Lemma 37 (Lemma 3.1.7)

Let f be closed and convex. Assume that it is differentiable on its domain. Then $\partial f(\boldsymbol{x}) = \{\nabla f(\boldsymbol{x})\}$ for any $\boldsymbol{x} \in \operatorname{int}(\operatorname{dom} f)$.

Lemma 38 (Lemma 3.1.8)

Let function $f(y)$ be closed and convex with $\operatorname{dom} f \subseteq \mathbb{R}^m$. Consider a linear operator

$$\mathcal{A}(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{b} : \quad \mathbb{R}^n \to \mathbb{R}^m.$$

Then $\phi(\boldsymbol{x}) = f(\mathcal{A}(\boldsymbol{x}))$ is a closed convex function with domain $\operatorname{dom} \phi = \{\boldsymbol{x} \mid \mathcal{A}(\boldsymbol{x}) \in \operatorname{dom} f\}$. For any $\boldsymbol{x} \in \operatorname{int}(\operatorname{dom} \phi)$ we have

$$\partial \phi(\boldsymbol{x}) = A^\top \partial f(\mathcal{A}(\boldsymbol{x})).$$

Rules for Computing

Lemma 39 (Lemma 3.1.9)

Let $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$ be closed convex functions and $\alpha_1, \alpha_2 \geq 0$. Then function $f(\boldsymbol{x}) = \alpha_1 f_1(\boldsymbol{x}) + \alpha_2 f_2(\boldsymbol{x})$ is closed and convex and

$$\partial f(\boldsymbol{x}) = \alpha_1 \partial f_1(\boldsymbol{x}) + \alpha_2 \partial f_2(\boldsymbol{x}). \tag{14}$$

for any x from $\text{int}(\text{dom } f) = \text{int}(\text{dom } f_1) \cap \text{int}(\text{dom } f_2)$.

Rules for Computing

Lemma 40 (Lemma 3.1.10)

Let functions $f_i(\boldsymbol{x}), i = 1 \ldots m$ be closed and convex. Then function

$$f(\boldsymbol{x}) = \max_{1 \leq i \leq m} f_i(\boldsymbol{x})$$

is also closed and convex. For any $\boldsymbol{x} \in \text{int}(\text{dom } f) = \cap_{i=1}^{m} \text{int}(\text{dom } f_i)$ we have

$$\partial f(\boldsymbol{x}) = \text{Conv}\{\partial f_i(\boldsymbol{x}) | i \in I(\boldsymbol{x})\}, \tag{15}$$

where $I(\boldsymbol{x}) = \{i : f_i(\boldsymbol{x}) = f(\boldsymbol{x})\}$.

Rules for Computing

Lemma 41 (Lemma 3.1.11)

Let Δ be a set and $f(\boldsymbol{x}) = \sup\{\phi(\boldsymbol{y}, \boldsymbol{x}) | \boldsymbol{y} \in \Delta\}$. Suppose that for any fixed $\boldsymbol{y} \in \Delta$ the function $\phi(\boldsymbol{y}, \boldsymbol{x})$ is closed and convex in $\boldsymbol{x}$. Then $f(\boldsymbol{x})$ is closed convex.
Moreover, for any $\boldsymbol{x}$ from

$$\mathrm{dom}\ f = \{\boldsymbol{x} \in \mathbb{R}^n | \exists \gamma : \phi(\boldsymbol{y}, \boldsymbol{x}) \leq \gamma, \ \ \forall \boldsymbol{y} \in \Delta\}$$

we have

$$\partial f(\boldsymbol{x}) \supseteq \mathrm{Conv}\{\partial \phi_{\boldsymbol{x}}(\boldsymbol{y}, \boldsymbol{x}) | \boldsymbol{y} \in I(\boldsymbol{x})\},$$

where $I(\boldsymbol{x}) = \{\boldsymbol{y} | \phi(\boldsymbol{y}, \boldsymbol{x}) = f(\boldsymbol{x})\}$.

Rules for Computing

Theorem 42

Let $\|\cdot\|$ be a vector norm in $\mathbb{R}^n$, then

$$\partial \|\cdot\| = \left\{ V(\boldsymbol{x}) \triangleq \{ \boldsymbol{v} \in \mathbb{R}^n | \langle \boldsymbol{v}, \ \boldsymbol{x} \rangle = \|\boldsymbol{x}\|, \|\boldsymbol{v}\|_* \leq 1 \} \right\},$$

where $\|\boldsymbol{v}\|_*$ is the dual norm of $\|\cdot\|$, defined as

$$\|\boldsymbol{v}\|_* \triangleq \sup_{\|\boldsymbol{u}\| \leq 1} \langle \boldsymbol{v}, \ \boldsymbol{u} \rangle.$$

Proof.

We prove that $V(\boldsymbol{x}) \subset \partial \|\boldsymbol{x}\|$, and $\partial \|\boldsymbol{x}\| \subset V(\boldsymbol{x})$. We remark here notation $\|\cdot\|^*$ is the Fenchel conjugate of the norm $\|\cdot\|$, and $\|\cdot\|_*$ is the dual norm of the norm $\|\cdot\|$.

Rules for Computing

Proof. (Continued.) ($V(\boldsymbol{x}) \subset \partial \|\boldsymbol{x}\|$)
Let $\boldsymbol{v} \in V(\boldsymbol{x})$, and $\boldsymbol{y}$ be an arbitrary vector. Then

$$\begin{aligned} \|\boldsymbol{x}\| + \langle \boldsymbol{v},\ \boldsymbol{y} - \boldsymbol{x}\rangle &= \|\boldsymbol{x}\| + \langle \boldsymbol{v},\ \boldsymbol{y}\rangle - \langle \boldsymbol{v},\ \boldsymbol{x}\rangle \\ &= \|\boldsymbol{x}\| + \langle \boldsymbol{v},\ \boldsymbol{y}\rangle - \|\boldsymbol{x}\| \quad \because \boldsymbol{v} \in V(\boldsymbol{x}) \\ &= \langle \boldsymbol{v},\ \boldsymbol{y}\rangle \\ &\le \|\boldsymbol{y}\| \cdot \|\boldsymbol{v}\|_* \quad \text{Holder's inequality} \\ &\le \|\boldsymbol{y}\|. \quad \because \boldsymbol{v} \in V(\boldsymbol{x}) \end{aligned}$$

That is $\|\boldsymbol{y}\| \ge \|\boldsymbol{x}\| + \langle \boldsymbol{v},\ \boldsymbol{y} - \boldsymbol{x}\rangle$, for any $\boldsymbol{y}$, which is the definition of sub-gradient of $\|\cdot\|$ and $\boldsymbol{v} \in \partial \|\boldsymbol{x}\|$.

That implies $V(\boldsymbol{x}) \subset \partial \|\boldsymbol{x}\|$.

Rules for Computing

Proof. (Continued.) ($\partial \|\boldsymbol{x}\| \subset V(\boldsymbol{x})$)

Let $\boldsymbol{v} \in \partial \|\boldsymbol{x}\|$. Thus we have

$$\|\boldsymbol{y}\| \geq \|\boldsymbol{x}\| + \langle \boldsymbol{v},\ \boldsymbol{y} - \boldsymbol{x} \rangle,$$

for any $\boldsymbol{y}$. That is

$$\langle \boldsymbol{v},\ \boldsymbol{y} \rangle - \|\boldsymbol{y}\| \leq \langle \boldsymbol{v},\ \boldsymbol{x} \rangle - \|\boldsymbol{x}\|, \forall \boldsymbol{y}.$$

$$\Rightarrow \underbrace{\sup_{y} \{\langle \boldsymbol{v},\ \boldsymbol{y} \rangle - \|\boldsymbol{y}\|\}}_{\triangleq \|\boldsymbol{v}\|^* \text{ Fenchel conjugate}} \leq \langle \boldsymbol{v},\ \boldsymbol{x} \rangle - \|\boldsymbol{x}\|.$$

$\|\boldsymbol{v}\|^*$ is the Fenchel conjugate of norm at the point $\boldsymbol{v}$, which is the indicator function on the unit ball of dual norm.

Rules for Computing

Proof. (Continued.) ($\partial \|\boldsymbol{x}\| \subset V(\boldsymbol{x})$)
That is

$$\|\boldsymbol{v}\|^* = \begin{cases} 0 & \|\boldsymbol{v}\|_* \leq 1, \\ +\infty & \|\boldsymbol{v}\|_* > 1. \end{cases}$$

Since the case $\|\boldsymbol{v}\|_* > 1$ is impossible as $\langle \boldsymbol{v},\ \boldsymbol{x} \rangle - \|\boldsymbol{x}\|$ is always finite, we have $\|\boldsymbol{v}\|_* \leq 1$.
That is

$$\begin{aligned} 0 &\leq \langle \boldsymbol{v},\ \boldsymbol{x} \rangle - \|\boldsymbol{x}\| \\ &\leq \|\boldsymbol{x}\| \cdot \|\boldsymbol{v}\|_* - \|\boldsymbol{x}\| \\ &\leq 0. \quad \because \|\boldsymbol{v}\|_* \leq 1 \end{aligned}$$

So $\langle \boldsymbol{v},\ \boldsymbol{x} \rangle = \|\boldsymbol{x}\|$. That implies $\partial \|\boldsymbol{x}\| \subset V(\boldsymbol{x})$. $\square$

Examples

Example 43 (Example 3.1.5)

1. $f(\boldsymbol{x}) = |\boldsymbol{x}|, \boldsymbol{x} \in \mathbb{R}^1$.
2. $f(\boldsymbol{x}) = \sum_{i=1}^{m} |\langle \boldsymbol{a}_i, \ \boldsymbol{x} \rangle - \boldsymbol{b}_i|$.
3. $f(\boldsymbol{x}) = \max_{1 \le i \le n} \boldsymbol{x}^{(i)}$.
4. $f(\boldsymbol{x}) = \|\boldsymbol{x}\|$.
5. $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |\boldsymbol{x}^{(i)}|$.

References I

Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

Yurii Nesterov. *Lectures on convex optimization*, volume 137. Springer, 2018.

Sébastien Bubeck et al. Convex optimization: Algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3-4):231–357, 2015.

Thank You!

Email:qianhui@zju.edu.cn