最优化方法　第六讲

Weiwen Wang(王伟文)

暨南大学

2025 年 4 月 22 日

目录

本讲考虑带约束优化问题 (P)

$$\min_{\boldsymbol{x}} f(x)$$
$$\text{s.t. } \boldsymbol{x} \in C$$

其中 $f:\mathbb{R}^d \to \mathbb{R}$ 为连续可微函数, C 为闭凸集.

定义 1 (带约束优化问题的驻点)

设 f 为闭凸集 C 上的连续可微函数，称 $\boldsymbol{x}^* \in C$ 为优化问题 (P) 的驻点，若 $\nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x}-\boldsymbol{x}^*) \geq 0$ 对任意 $\boldsymbol{x} \in C$ 成立.

定理 2 (最优性必要条件)

若 $\boldsymbol{x}^* \in C$ 为优化问题 (P) 的最优解, 则 $\boldsymbol{x}^*$ 为优化问题 (P) 的驻点.

证明.

若 $\boldsymbol{x}^*$ 不是 P 的驻点，则存在 $\boldsymbol{x}' \in C$ 使得

$$\nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x}' - \boldsymbol{x}^*) < 0.$$

即

$$\lim_{t\to 0^+} \frac{f[\boldsymbol{x}^* + t(\boldsymbol{x}' - \boldsymbol{x}^*)] - f(\boldsymbol{x}^*)}{t} = \nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x}' - \boldsymbol{x}^*) < 0$$

故存在 $t' \in (0,1)$ 使得

$$f[\boldsymbol{x}^* + t(\boldsymbol{x}' - \boldsymbol{x}^*)] < f(\boldsymbol{x}^*).$$

又因为 $\boldsymbol{x}^* + t(\boldsymbol{x}' - \boldsymbol{x}^*) \in C$，与 $\boldsymbol{x}^*$ 的最优性矛盾. □

定理 3

设 f 为闭凸集 C 上的连续可微凸函数, $\boldsymbol{x}^* \in C$ 为优化问题 (P) 的最优解当且仅当 $\boldsymbol{x}^*$ 为 (P) 的驻点, 即

$$\nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x}-\boldsymbol{x}^*) \geq 0 \quad \forall \boldsymbol{x} \in C.$$

定义 4 (非空闭凸集的正交投影)

给定非空闭凸集 $C \subset \mathbb{R}^d$, 任意一点 $\boldsymbol{x} \in \mathbb{R}^d$ 到 C 上的正交投影定义为

$$\Pi_C(\boldsymbol{x}) = \arg\min_{\boldsymbol{z} \in C} \|\boldsymbol{x} - \boldsymbol{z}\|^2$$

定理 5 (非空闭凸集的正交投影存在性与唯一性)

给定非空闭凸集 $C \subset \mathbb{R}^d$, 对任意一点 $\boldsymbol{x} \in \mathbb{R}^d$, 其到 C 上的正交投影定义 $\Pi_C(\boldsymbol{x})$ 存在且唯一.

证明.

定义函数 $h(\boldsymbol{z}) = \|\boldsymbol{x} - \boldsymbol{z}\|^2$, $h(\boldsymbol{z})$ 为 coercive 函数, 故由 Lecture 1 定理 1.0.26 知 $\arg\min_{\boldsymbol{z}\in C} h(\boldsymbol{z})$ 非空, 即 $\Pi_C(\boldsymbol{x})$ 存在.

证明.

定义函数 $h(\boldsymbol{z}) = \|\boldsymbol{x} - \boldsymbol{z}\|^2$, $h(\boldsymbol{z})$ 为 coercive 函数, 故由 Lecture 1 定理 1.0.26 知 $\underset{\boldsymbol{z}\in C}{\arg\min}\, h(\boldsymbol{z})$ 非空, 即 $\Pi_C(\boldsymbol{x})$ 存在.

下证唯一性, 设 $\boldsymbol{u}, \boldsymbol{v}$ 均为 $\boldsymbol{x}$ 到集合 C 上的正交投影, $h(\boldsymbol{z})$ 为连续可微凸函数, 故由定理 3, 有

$$2\langle \boldsymbol{u} - \boldsymbol{x}, \boldsymbol{v} - \boldsymbol{u}\rangle \geq 0$$

$$2\langle \boldsymbol{v} - \boldsymbol{x}, \boldsymbol{u} - \boldsymbol{v}\rangle \geq 0$$

综上得到 $\|\boldsymbol{u} - \boldsymbol{v}\|^2 \leq 0$, 故 $\boldsymbol{u} = \boldsymbol{v}$. □

定理 6

给定非空闭凸集 $C \subset \mathbb{R}^d$, 对任意一点 $\boldsymbol{x} \in \mathbb{R}^d$, $\boldsymbol{u} = \Pi_C(\boldsymbol{x})$ 当且仅当

$$\langle \boldsymbol{x} - \boldsymbol{u}, \boldsymbol{z} - \boldsymbol{u} \rangle \leq 0$$

对任意 $\boldsymbol{z} \in C$ 成立.

定理 7

设集合 C 为闭凸集，则 $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$，

- $$\langle \Pi_C(\boldsymbol{x}) - \Pi_C(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \rangle \geq \|\Pi_C(\boldsymbol{x}) - \Pi_C(\boldsymbol{y})\|^2$$
- $$\|\Pi_C(\boldsymbol{x}) - \Pi_C(\boldsymbol{y})\| \leq \|\boldsymbol{x} - \boldsymbol{y}\|$$

证明.

对任意 $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d$，$\Pi_C(\boldsymbol{x})\in C,\Pi_C(\boldsymbol{y})\in C$，故由定理 6，

$$\langle \boldsymbol{x}-\Pi_C(\boldsymbol{x}),\Pi_C(\boldsymbol{y})-\Pi_C(\boldsymbol{x})\rangle\le 0$$

$$\langle \boldsymbol{y}-\Pi_C(\boldsymbol{y}),\Pi_C(\boldsymbol{x})-\Pi_C(\boldsymbol{y})\rangle\le 0$$

证明.

对任意 $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d$, $\Pi_C(\boldsymbol{x})\in C, \Pi_C(\boldsymbol{y})\in C$, 故由定理 6,

$$\langle \boldsymbol{x}-\Pi_C(\boldsymbol{x}), \Pi_C(\boldsymbol{y})-\Pi_C(\boldsymbol{x})\rangle \le 0$$

$$\langle \boldsymbol{y}-\Pi_C(\boldsymbol{y}), \Pi_C(\boldsymbol{x})-\Pi_C(\boldsymbol{y})\rangle \le 0$$

上述两项不等式相加, 整理得到

$$\langle \boldsymbol{x}-\boldsymbol{y}+(\Pi_C(\boldsymbol{y})-\Pi_C(\boldsymbol{x})), \Pi_C(\boldsymbol{y})-\Pi_C(\boldsymbol{x})\rangle \le 0.$$

因此

$$\|\Pi_C(\boldsymbol{x})-\Pi_C(\boldsymbol{y})\|^2 \le \langle \boldsymbol{x}-\boldsymbol{y}, \Pi_C(\boldsymbol{x})-\Pi_C(\boldsymbol{y})\rangle.$$

证明.

对任意 $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d$, $\Pi_C(\boldsymbol{x})\in C,\Pi_C(\boldsymbol{y})\in C$, 故由定理 6,

$$\langle \boldsymbol{x}-\Pi_C(\boldsymbol{x}),\Pi_C(\boldsymbol{y})-\Pi_C(\boldsymbol{x})\rangle\leq 0$$

$$\langle \boldsymbol{y}-\Pi_C(\boldsymbol{y}),\Pi_C(\boldsymbol{x})-\Pi_C(\boldsymbol{y})\rangle\leq 0$$

上述两项不等式相加, 整理得到

$$\langle \boldsymbol{x}-\boldsymbol{y}+(\Pi_C(\boldsymbol{y})-\Pi_C(\boldsymbol{x})),\Pi_C(\boldsymbol{y})-\Pi_C(\boldsymbol{x})\rangle\leq 0.$$

因此

$$\|\Pi_C(\boldsymbol{x})-\Pi_C(\boldsymbol{y})\|^2\leq\langle \boldsymbol{x}-\boldsymbol{y},\Pi_C(\boldsymbol{x})-\Pi_C(\boldsymbol{y})\rangle.$$

由 Cauchy-Schwartz 不等式, 即有

$$\|\Pi_C(\boldsymbol{x})-\Pi_C(\boldsymbol{y})\|\leq\|\boldsymbol{x}-\boldsymbol{y}\|.$$

设 $C = \mathbb{R}_+^d$, 计算任意 $\boldsymbol{x} \in \mathbb{R}^d$ 到 C 的投影即求解优化问题

$$\min \sum_{i=1}^{d} (z_i - x_i)^2$$
$$\text{s.t. } z_i \geq 0, z_2 \geq 0, \ldots, z_d \geq 0.$$

设 $C=\mathbb{R}_+^d$，计算任意 $\boldsymbol{x}\in\mathbb{R}^d$ 到 C 的投影即求解优化问题

$$\begin{aligned} \min \ & \sum_{i=1}^{d}(z_i-x_i)^2 \\ \text{s.t. } & z_i\geq 0, z_2\geq 0, \ldots, z_d\geq 0. \end{aligned}$$

因为目标函数求和项可分离，$\Pi_C(\boldsymbol{x})$ 的每一个坐标分量可单独求解

$$\min_{z_i}\{(z_i-x_i)^2 : z_i\geq 0\}$$

得到，其最优解为 $z_i^*=\max(x_i,0)\triangleq[x_i]_+$.

因此

$$\Pi_{\mathbb{R}^d_+}(\boldsymbol{x}) = ([x_1]_+, [x_2]_+, \cdots, [x_d]_+)^T$$

设有长方体

$$\mathsf{B} = \{\boldsymbol{z} \in \mathbb{R}^d : l_i \leq z_i \leq u_i\}$$

类似非负象限投影, 对任意 $\boldsymbol{x} \in \mathbb{R}^d$ 有

$$[\Pi_{\mathsf{B}}(\boldsymbol{x})]_i = \begin{cases} u_i, & x_i > u_i \\ x_i, & l_i \leq x_i \leq u_i \\ l_i, & x_i < l_i \end{cases}$$

设 $C \triangleq \{\boldsymbol{z} \in \mathbb{R}^d : \|\boldsymbol{z}\| \le r\}$, 计算任意 $\boldsymbol{x} \in \overline{C}$ 的投影 $\Pi_C(\boldsymbol{x})$ 即求解优化问题

$$\min_{\boldsymbol{z}} \{\|\boldsymbol{z} - \boldsymbol{x}\|^2 : \|\boldsymbol{z}\| \le r\}$$

等价于

$$\min_{\boldsymbol{z}} \{\|\boldsymbol{z} - \boldsymbol{x}\|^2 : \|\boldsymbol{z}\| = r\}$$

展开为

$$\min_{\boldsymbol{z}} \{r^2 - 2\langle \boldsymbol{z}, \boldsymbol{x}\rangle + \|\boldsymbol{x}\|^2 : \|\boldsymbol{z}\| = r\}$$

又因为 $\langle \boldsymbol{z}, \boldsymbol{x}\rangle \le \|\boldsymbol{z}\|\|\boldsymbol{x}\| = r\|\boldsymbol{x}\|$, 当 $\boldsymbol{z} = \frac{r\|\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$ 时等号成立.

$$\Pi_C(\boldsymbol{x}) = \begin{cases} \boldsymbol{x}, & \boldsymbol{x} \in C \\ \dfrac{r\|\boldsymbol{x}\|}{\|\boldsymbol{x}\|}, & \boldsymbol{x} \notin C \end{cases}$$

定理 8

设 $\eta > 0$, $\boldsymbol{x}^*$ 为优化问题 (P) 的驻点当且仅当

$$\boldsymbol{x}^* = \Pi_C(\boldsymbol{x}^* - \eta \nabla f(\boldsymbol{x}^*))$$

证明.

(⇒) 简记 $\boldsymbol{x}^+ = \boldsymbol{x}^* - \eta\nabla f(\boldsymbol{x}^*)$, 由定理 6,

$$\langle \boldsymbol{x}^* - \eta\nabla f(\boldsymbol{x}^*) - \Pi_C(\boldsymbol{x}^+), \boldsymbol{z} - \Pi_C(\boldsymbol{x}^+)\rangle \le 0 \quad \forall \boldsymbol{z} \in C$$

取 $\boldsymbol{z} = \boldsymbol{x}^*$, 上式可整理为

$$\|\boldsymbol{x}^* - \Pi_C(\boldsymbol{x}^+)\|^2 - \langle \eta\nabla f(\boldsymbol{x}^*), \boldsymbol{x}^* - \Pi_C(\boldsymbol{x}^+)\rangle \le 0.$$

又因为 $\boldsymbol{x}^*$ 为优化问题 (P) 的驻点, 故

$$\langle \nabla f(\boldsymbol{x}^*), \boldsymbol{x}^* - \Pi_C(\boldsymbol{x}^+)\rangle \le 0,$$

即

$$\|\boldsymbol{x}^* - \Pi_C(\boldsymbol{x}^+)\|^2 \le 0 \Longrightarrow \boldsymbol{x}^* = \Pi_C(\boldsymbol{x}^+).$$

□

证明.

(⇐) 因为 $\boldsymbol{x}^* = \Pi_C(\boldsymbol{x}^* - \eta\nabla f(\boldsymbol{x}^*))$, 由定理 6,

$$\langle \boldsymbol{x}^* - \eta\nabla f(\boldsymbol{x}^*) - \boldsymbol{x}^*, \boldsymbol{z} - \boldsymbol{x}^* \rangle \le 0 \quad \forall \boldsymbol{z} \in C.$$

即

$$\langle \nabla f(\boldsymbol{x}^*), \boldsymbol{z} - \boldsymbol{x}^* \rangle \ge 0 \quad \forall \boldsymbol{z} \in C.$$

故 $\boldsymbol{x}^*$ 为优化问题 (P) 的驻点. □

迭代格式

- 线搜索确定步长 η_k(回溯法或常数步长)
- $\boldsymbol{x}_{k+1} = \Pi_C(\boldsymbol{x}_k - \eta_k \nabla f(\boldsymbol{x}_k))$

定理 9

若 f 为 *L-smooth*, 则对任意 $\boldsymbol{x} \in C$ 及 $\eta \in (0, \frac{2}{L})$, 有不等式

$$f(\boldsymbol{x}) - f(\Pi_C(\boldsymbol{x} - \eta \nabla f(\boldsymbol{x}))) \geq \eta \left(1 - \frac{L\eta}{2}\right) \left\| \frac{1}{\eta}(\boldsymbol{x} - \Pi_C(\boldsymbol{x} - \eta \nabla f(\boldsymbol{x}))) \right\|^2$$

成立.

证明.

简记 $\boldsymbol{x}^+ = \Pi_C(\boldsymbol{x} - \eta\nabla f(\boldsymbol{x}))$, 由 L-smooth 性质

$$f(\boldsymbol{x}^+) \le f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{x}^+ - \boldsymbol{x}\rangle + \frac{L}{2}\|\boldsymbol{x} - \boldsymbol{x}^+\|^2.$$

由定理 6

$$\langle \boldsymbol{x} - \eta\nabla f(\boldsymbol{x}) - \boldsymbol{x}^+, \boldsymbol{x} - \boldsymbol{x}^+\rangle \le 0.$$

整理得到

$$\langle \nabla f(\boldsymbol{x}), \boldsymbol{x} - \boldsymbol{x}^+\rangle \ge \frac{1}{\eta}\|\boldsymbol{x} - \boldsymbol{x}^+\|^2$$

□

续.

整理得到

$$\langle \nabla f(\boldsymbol{x}), \boldsymbol{x}-\boldsymbol{x}^{+} \rangle \geq \frac{1}{\eta}\|\boldsymbol{x}-\boldsymbol{x}^{+}\|^{2}$$

因此

$$\begin{aligned} f(\boldsymbol{x})-f(\boldsymbol{x}^{+}) &\geq \nabla f(\boldsymbol{x}), \boldsymbol{x}-\boldsymbol{x}^{+} \rangle - \frac{L}{2}\|\boldsymbol{x}-\boldsymbol{x}^{+}\|^{2} \\ &\geq \eta \|\frac{1}{\eta}(\boldsymbol{x}-\boldsymbol{x}^{+})\|^{2} - \frac{L\eta^{2}}{2}\|\frac{1}{\eta}(\boldsymbol{x}-\boldsymbol{x}^{+})\|^{2} \\ &\geq \eta\left(1-\frac{L\eta}{2}\right)\left\|\frac{1}{\eta}(\boldsymbol{x}-\boldsymbol{x}^{+})\right\|^{2} \end{aligned}$$

□

定义 10 (Gradient Mapping)

定义 Gradient Mapping 为

$$G_\eta(\boldsymbol{x}) = \frac{1}{\eta}\left[\boldsymbol{x} - \Pi_C\left(\boldsymbol{x} - \eta\nabla f(\boldsymbol{x})\right)\right]$$

- 若 $\Pi_C\left(\boldsymbol{x} - \eta\nabla f(\boldsymbol{x})\right) = \boldsymbol{x} - \eta\nabla f(\boldsymbol{x})$, $G_\eta(\boldsymbol{x}) = \nabla f(\boldsymbol{x})$.
- 定理 8 表明, 若 $G_\eta(\boldsymbol{x}) = 0$, 则 $\boldsymbol{x}$ 为优化问题 (P) 的最优解.

引理 11

设 $\eta_1 \le \eta_2$, 则对任意 $\boldsymbol{x} \in \mathbb{R}^d$

$$\|G_{\eta_1}(\boldsymbol{x})\| \ge \|G_{\eta_2}(\boldsymbol{x})\|,$$

且

$$\eta_1 \|G_{\eta_1}(\boldsymbol{x})\| \le \eta_2 \|G_{\eta_2}(\boldsymbol{x})\|.$$

证明.

简记 $\boldsymbol{u} = \boldsymbol{x} - \eta_1 \nabla f(\boldsymbol{x})$, $\boldsymbol{v} = \boldsymbol{x} - \eta_2 \nabla f(\boldsymbol{x})$ 由定理 6

$$\langle \boldsymbol{u} - \Pi_C(\boldsymbol{u}), \Pi_C(\boldsymbol{u}) - \Pi_C(\boldsymbol{v}) \rangle \geq 0$$

即

$$\langle \eta_1 G_{\eta_1}(\boldsymbol{x}) - \eta_1 \nabla f(\boldsymbol{x}), \eta_1 G_{\eta_1}(\boldsymbol{x}) - \eta_2 G_{\eta_2}(\boldsymbol{x}) \rangle \geq 0. \quad (\circ)$$

证明.

简记 $\boldsymbol{u} = \boldsymbol{x} - \eta_1 \nabla f(\boldsymbol{x})$, $\boldsymbol{v} = \boldsymbol{x} - \eta_2 \nabla f(\boldsymbol{x})$ 由定理 6

$$\langle \boldsymbol{u} - \Pi_C(\boldsymbol{u}), \Pi_C(\boldsymbol{u}) - \Pi_C(\boldsymbol{v}) \rangle \geq 0$$

即

$$\langle \eta_1 G_{\eta_1}(\boldsymbol{x}) - \eta_1 \nabla f(\boldsymbol{x}), \eta_1 G_{\eta_1}(\boldsymbol{x}) - \eta_2 G_{\eta_2}(\boldsymbol{x}) \rangle \geq 0. \quad (\circ)$$

同理,

$$\langle \boldsymbol{v} - \Pi_C(\boldsymbol{v}), \Pi_C(\boldsymbol{v}) - \Pi_C(\boldsymbol{u}) \rangle \geq 0$$

$$\langle \eta_2 G_{\eta_2}(\boldsymbol{x}) - \eta_2 \nabla f(\boldsymbol{x}), \eta_2 G_{\eta_2}(\boldsymbol{x}) - \eta_1 G_{\eta_1}(\boldsymbol{x}) \rangle \geq 0. \quad (\Box)$$

$\Box$

续.

$[\eta_2\cdot(\circ)+\eta_1\cdot(\Box)]/(\eta_1\eta_2)$ 得到

$$\langle G_{\eta_1}(\boldsymbol{x})-G_{\eta_2}(\boldsymbol{x}),\eta_1 G_{\eta_1}(\boldsymbol{x})-\eta_2 G_{\eta_2}(\boldsymbol{x})\rangle\geq 0$$

将内积展开

$$\eta_1\|G_{\eta_1}(\boldsymbol{x})\|^2+\eta_2\|G_{\eta_2}(\boldsymbol{x})\|^2\leq\left(\eta_1+\eta_2\right)G_{\eta_1}(\boldsymbol{x})^T G_{\eta_2}(\boldsymbol{x})$$

应用 Cauchy-Schwartz 不等式

$$\eta_1\|G_{\eta_1}(\boldsymbol{x})\|^2+\eta_2\|G_{\eta_2}(\boldsymbol{x})\|^2\leq\left(\eta_1+\eta_2\right)\|G_{\eta_1}(\boldsymbol{x})\|\|G_{\eta_2}(\boldsymbol{x})\|$$

□

续.

若 $G_{\eta_2}(\boldsymbol{x})=0$, 引理的结论显然成立, 考虑 $G_{\eta_2}(\boldsymbol{x})\neq 0$. 不等式

$$\eta_1\|G_{\eta_1}(\boldsymbol{x})\|^2+\eta_2\|G_{\eta_2}(\boldsymbol{x})\|^2\leq(\eta_1+\eta_2)\|G_{\eta_1}(\boldsymbol{x})\|\|G_{\eta_2}(\boldsymbol{x})\|$$

左右两端同时除去 $\|G_{\eta_2}(\boldsymbol{x})\|^2$, 整理得到

$$\eta_1\frac{\|G_{\eta_1}(\boldsymbol{x})\|^2}{\|G_{\eta_2}(\boldsymbol{x})\|^2}-(\eta_1+\eta_2)\frac{\|G_{\eta_1}(\boldsymbol{x})\|}{\|G_{\eta_2}(\boldsymbol{x})\|}+\eta_2\leq 0.$$

记 $t=\frac{\|G_{\eta_1}(\boldsymbol{x})\|}{\|G_{\eta_2}(\boldsymbol{x})\|}$, 有二次不等式

$$\eta_1t^2-(\eta_1+\eta_2)t+\eta_2\leq 0,$$

解得 $1\leq t\leq\frac{\eta_2}{\eta_1}$, 即 $1\leq\frac{\|G_{\eta_1}(\boldsymbol{x})\|}{\|G_{\eta_2}(\boldsymbol{x})\|}\leq\frac{\eta_2}{\eta_1}$. □

回溯法

给定参数 (s,α,β), 其中 $s>0$, $\alpha\in(0,1)$, $\beta\in(0,1)$.

- η_k 初始化为 s.
- 若
$$f(\boldsymbol{x}_k)-f(\Pi_C(\boldsymbol{x}_k-\eta_k\nabla f(\boldsymbol{x}_k)))<\alpha\eta_k\|G_{\eta_k}(\boldsymbol{x}_k)\|^2,$$
取 $\eta_k=\beta\eta_k$.
- 直到 η_k 首次 满足
$$f(\boldsymbol{x}_k)-f(\Pi_C(\boldsymbol{x}_k-\eta_k\nabla f(\boldsymbol{x}_k)))\geq\alpha\eta_k\|G_{\eta_k}(\boldsymbol{x}_k)\|^2.$$

引理 12

设 f 为 L-*smooth*, 在投影梯度下降法中，以 (s,α,β) 参数的回溯法能在有限步内结束，且得到的步长 η_k 满足

$$\eta_k \geq \min\left\{s, \frac{2(1-\alpha)}{\beta}\right\}.$$

证明.

由充分下降引理, $\forall \eta \in (0, \frac{2}{L})$

$$f(\boldsymbol{x}) - f(\Pi_C(\boldsymbol{x} - \eta \nabla f(\boldsymbol{x}))) \geq \eta \left(1 - \frac{L\eta}{2}\right) \left\| G_\eta(\boldsymbol{x}) \right\|^2$$

若

$$\eta \left(1 - \frac{L\eta}{2}\right) \left\| G_\eta(\boldsymbol{x}) \right\|^2 \geq \alpha\eta \left\| G_\eta(\boldsymbol{x}) \right\|^2$$

即 $\eta \leq \frac{2(1-\alpha)}{L}$, 此时回溯法停止. □

续.

若 $\eta_k < s$,

$$\frac{\eta_k}{\beta}\left(1-\frac{L\eta_k}{2\beta}\right) \le f(\boldsymbol{x}_k) - f\left(\Pi_C(\boldsymbol{x}_k - \frac{\eta_k}{\beta}\nabla f(\boldsymbol{x}_k))\right) < \alpha\frac{\eta_k}{\beta}\left\|G_{\frac{\eta_k}{\beta}}(\boldsymbol{x}_k)\right\|^2$$

解得

$$\eta_k > \frac{2(1-\alpha)\beta}{L}.$$

或 $\eta_k = s$.

综上有

$$\eta_k \ge \min\left\{s, \frac{2(1-\alpha)}{\beta}\right\}$$

□

引理 13

设 f 为 L-smooth, 在投影梯度下降法中, 使用常数步长 $\eta \in (0, \frac{2}{L})$, 或使用以 (s, α, β) 参数的回溯法确定步长 η_k, 迭代过程中产生的序列 $\{\boldsymbol{x}_k\}$ 满足

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1}) \geq M \|G_d(\boldsymbol{x}_k)\|^2,$$

其中

$$M = \begin{cases} \eta(1 - \frac{L\eta}{2}), & \eta \in (0, \frac{2}{L}) \\ \alpha \min\left\{s, \frac{2(1-\alpha)}{\beta}\right\}, & \text{回溯法} \end{cases}$$

$$d = \begin{cases} \eta, & \eta \in (0, \frac{2}{L}) \\ s, & \text{回溯法} \end{cases}$$

定理 14

设 f 为 L-smooth, f 在非空闭凸集上 C 有最小值点 $\boldsymbol{x}^*$. 在投影梯度下降法中, 使用常数步长 $\eta \in (0, \frac{2}{L})$, 或使用以 (s, α, β) 参数的回溯法确定步长 η_k, $\{\boldsymbol{x}_k\}$ 为投影梯度法产生的序列, 则有

$$\min_{k \in \{0,1,\ldots,K\}} \|G_d(\boldsymbol{x}_k)\| \le \sqrt{\frac{f(\boldsymbol{x}_0) - f(\boldsymbol{x}^*)}{M(K+1)}}$$

其中

$$M = \begin{cases} \eta(1 - \dfrac{L\eta}{2}), & \eta \in (0, \dfrac{2}{L}) \\ \alpha \min\left\{s, \dfrac{2(1-\alpha)}{\beta}\right\}, & \text{回溯法} \end{cases}$$

$$d = \begin{cases} \eta, & \eta \in (0, \dfrac{2}{L}) \\ s, & \text{回溯法} \end{cases}$$

定理 15

设 f 为 *L-smooth* 且为凸函数, f 在非空闭凸集上 C 有最小值点 $\boldsymbol{x}^*$. 在投影梯度下降法中, 使用常数步长 $\eta \in (0, \frac{1}{L}]$, $\{\boldsymbol{x}_k\}$ 为产生的序列, 则

(1) $2\eta(f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}^*)) \leq \|\boldsymbol{x}_k - \boldsymbol{x}^*\|^2 - \|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|, \quad \forall k \geq 0.$

(2) $f(\boldsymbol{x}_K) - f(\boldsymbol{x}^*) \leq \frac{\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|^2}{2K\eta}, \quad \forall K \geq 0.$

证明.

1. 因为 $\boldsymbol{x}_{k+1} = \Pi_C(\boldsymbol{x}_k - \eta\nabla f(\boldsymbol{x}_k))$, 由定理 6

$$\langle \boldsymbol{x}_k - \eta\nabla f(\boldsymbol{x}_k) - \boldsymbol{x}_{k+1}, \boldsymbol{x}^* - \boldsymbol{x}_{k+1}\rangle \le 0$$

即

$$\langle \boldsymbol{x}_k - \boldsymbol{x}_{k+1}, \boldsymbol{x}^* - \boldsymbol{x}_{k+1}\rangle - \langle \eta\nabla f(\boldsymbol{x}_k), \boldsymbol{x}^* - \boldsymbol{x}_{k+1}\rangle \le 0$$

$$\langle \boldsymbol{x}_k - \boldsymbol{x}_{k+1}, \boldsymbol{x}^* - \boldsymbol{x}_{k+1}\rangle - \langle \eta\nabla f(\boldsymbol{x}_k), \boldsymbol{x}^* - \boldsymbol{x}_k + \boldsymbol{x}_k - \boldsymbol{x}_{k+1}\rangle \le 0$$

整理得到

$$\langle \eta\nabla f(\boldsymbol{x}_k), \boldsymbol{x}_k - \boldsymbol{x}^*\rangle \le \langle \boldsymbol{x}_k - \boldsymbol{x}_{k+1}, \boldsymbol{x}_{k+1} - \boldsymbol{x}^*\rangle + \langle \eta\nabla f(\boldsymbol{x}_k), \boldsymbol{x}_k - \boldsymbol{x}_{k+1}\rangle$$

□

续.

由 L-smooth 有

$$f(\boldsymbol{x}_{k+1}) \le f(\boldsymbol{x}_k) + \langle \nabla f(\boldsymbol{x}_k), \boldsymbol{x}_{k+1} - \boldsymbol{x}_k \rangle + \frac{L}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2$$

故

$$\langle \eta \nabla f(\boldsymbol{x}_k), \boldsymbol{x}_k - \boldsymbol{x}_{k+1} \rangle \le \eta(f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1})) + \frac{L\eta}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2$$

因为 f 为凸函数

$$f(\boldsymbol{x}^*) \ge f(\boldsymbol{x}_k) + \langle \nabla f(\boldsymbol{x}_k), \boldsymbol{x}^* - \boldsymbol{x}_k \rangle$$

即

$$\langle \eta \nabla f(\boldsymbol{x}_k), \boldsymbol{x}_k - \boldsymbol{x}_{k+1} \rangle \ge \eta(f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*))$$

□

续.

故不等式

$$\langle \eta \nabla f(\boldsymbol{x}_k), \boldsymbol{x}_k - \boldsymbol{x}^* \rangle \le \langle \boldsymbol{x}_k - \boldsymbol{x}_{k+1}, \boldsymbol{x}_{k+1} - \boldsymbol{x}^* \rangle + \langle \eta \nabla f(\boldsymbol{x}_k), \boldsymbol{x}_k - \boldsymbol{x}_{k+1} \rangle$$

可约化为

$$\begin{aligned} \eta(f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)) \le & \frac{1}{2}\|\boldsymbol{x}_k - \boldsymbol{x}^*\|^2 - \frac{1}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|^2 - \frac{1}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2 \\ & + \eta(f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1})) + \frac{L\eta}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2 \end{aligned}$$

整理得到

$$\begin{aligned} 2\eta(f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}_k)) \le & \|\boldsymbol{x}_k - \boldsymbol{x}^*\|^2 - \|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|^2 - (1 - L\eta)\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2 \\ \le & \|\boldsymbol{x}_k - \boldsymbol{x}^*\|^2 - \|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|^2 \end{aligned}$$

□

续.

2. 由结论 (1)K 项求和, 及 $\{f(\boldsymbol{x}_k)\}$ 单调非增得到. $\square$

定理 16

在定理 *15* 的条件下，投影梯度法产生的序列 $\{\boldsymbol{x}_k\}$ 收敛到优化问题 (P) 的最优解.

定理 16

在定理 *15* 的条件下，投影梯度法产生的序列 $\{\boldsymbol{x}_k\}$ 收敛到优化问题 (P) 的最优解.

证明.

由定理 15 结论 (1) 及 $f(\boldsymbol{x}_{k+1}) \geq f(\boldsymbol{x}^*)$) 有

$$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\| \leq \|\boldsymbol{x}_k - \boldsymbol{x}^*\| \quad \forall k \geq 0.$$

即序列 $\{\boldsymbol{x}_k\}$ 满足 Fejér 单调性. 因为

$$\|\boldsymbol{x}_k - \boldsymbol{x}^*\| \leq \|\boldsymbol{x}_0 - \boldsymbol{x}^*\| \quad \forall k \geq 1,$$

序列 $\{\boldsymbol{x}_k\}$ 有界，故存在收敛子列 $\{\boldsymbol{x}_{k_j}\}_{j=0}^{\infty}$ □

续.

设收敛子列 $\{\boldsymbol{x}_{k_j}\}_{j=0}^{\infty}$ 的极限为 $\hat{\boldsymbol{x}}$, 因为 f 为连续函数,

$$\lim_{k_j\to\infty} f(\boldsymbol{x}_{k_j}) = f(\hat{\boldsymbol{x}}) = f(\boldsymbol{x}^*)$$

这表明 $\hat{\boldsymbol{x}}$ 为优化问题 (P) 的最优解.

再次利用 Fejér 单调性.

$$\|\boldsymbol{x}_k - \hat{\boldsymbol{x}}\| \le \|\boldsymbol{x}_{k_j} - \hat{\boldsymbol{x}}\| \qquad \forall k \ge k_j$$

令 $k_j \to \infty$, 因而 $k \to \infty$, 故

$$\lim_{k\to\infty} \|\boldsymbol{x}_k - \hat{\boldsymbol{x}}\| = 0 \Longrightarrow \boldsymbol{x}_k \to \hat{\boldsymbol{x}}(k\to\infty).$$

□