第四章 最优性理论

修贤超

https://xianchaoxiu.github.io

目录

最优化问题解的存在性

- 考虑优化问题

$$
\begin{aligned}
\min_{x\in\mathbb{R}^n} \quad & f(x) \\
\text{s.t.} \quad & x \in \mathcal{X}
\end{aligned}
$$

- 首先分析最优解的存在性, 然后考虑如何求出其最优解
- 回顾 Weierstrass 定理, 即定义在紧集上的连续函数一定存在最大 (最小) 值点
- 而在许多实际问题中, 定义域可能不是紧的, 目标函数也不一定连续

推广的 Weierstrass 定理

- 若函数 $f : \mathcal{X} \to (-\infty, +\infty]$ 适当且闭，且以下条件中任意一个成立
  - $\operatorname{dom} f = \{x \in \mathcal{X} : f(x) < +\infty\}$ 是有界的
  - 存在一个常数 $\bar{\gamma}$ 使得下水平集
    $$C_{\bar{\gamma}} = \{x \in \mathcal{X} \mid f(x) \le \bar{\gamma}\}$$
    是非空且有界的
  - f 是强制的，即对于任一满足 $\|x^k\| \to +\infty$ 的点列 $\{x^k\} \subset \mathcal{X}$，都有
    $$\lim_{k\to\infty} f(x^k) = +\infty,$$

  则函数 f 的最小值点集 $\{x \in \mathcal{X} \mid f(x) \le f(y), \forall y \in \mathcal{X}\}$ 非空且紧

注记

- 三个条件在本质上都是保证 $f(x)$ 的最小值不能在无穷远处取到
- 定理仅要求 $f(x)$ 为适当且闭的函数, 并不需要 $f(x)$ 的连续性
- **例子** 当定义域不是有界闭集时, 对于强制函数 $f(x) = x^2,\ x \in \mathcal{X} = \mathbb{R}$, 其全局最优解一定存在
- **例子** 对于适当且闭的函数 $f(x) = e^{-x},\ x \in \mathcal{X} = \mathbb{R}$, 不满足三个条件中任意一个, 因此不能断言其全局极小值点存在

目录

无约束可微问题的最优性理论

- 无约束可微优化问题通常表示为

$$\min_{x\in\mathbb{R}^n} \quad f(x)$$

- 验证一个点是否为极小值点，称其为最优性条件
  - 一阶最优性条件
  - 二阶最优性条件

下降方向

- 对于可微函数 f 和点 $x \in \mathbb{R}^n$, 如果存在向量 d 满足

$$\nabla f(x)^\top d < 0$$

那么称 d 为 f 在点 x 处的一个**下降方向**

- 一阶最优性条件是利用梯度 (一阶) 信息来判断给定点的最优性
- 如果 f 在点 x 处存在一个下降方向 d, 那么对于任意的 $T > 0$, 存在 $t \in (0, T]$, 使得

$$f(x + td) < f(x)$$

因此, 在局部最优点处不能有下降方向

一阶必要条件

- 假设 f 在全空间 $\mathbb{R}^n$ 可微. 如果 x^* 是一个局部极小点, 那么

$$\nabla f(x^*) = 0$$

证明 任取 $v \in \mathbb{R}^n$, 考虑 f 在点 $x = x^*$ 处的泰勒展开

$$f(x^* + tv) = f(x^*) + tv^\top \nabla f(x^*) + o(t)$$

整理得

$$\frac{f(x^* + tv) - f(x^*)}{t} = v^\top \nabla f(x^*) + o(1)$$

根据 x^* 的最优性, 在上式中分别对 t 取点 0 处的左、右极限可知

$$\lim_{t \to 0^+} \frac{f(x^* + tv) - f(x^*)}{t} = v^\top \nabla f(x^*) \geq 0$$

$$\lim_{t \to 0^-} \frac{f(x^* + tv) - f(x^*)}{t} = v^\top \nabla f(x^*) \leq 0$$

即对任意的 v 有 $v^\top \nabla f(x^*) = 0$, 由 v 的任意性知 $\nabla f(x^*) = 0$

二阶最优性条件

- 称满足 $\nabla f(x)=0$ 的点 x 为 f 的稳定点 (或驻点、临界点)
- 对于 $f(x)=x^3$, 满足 $f'(x)=0$ 的点为 $x^*=0$, 但其不是局部最优解, 因此仅仅是必要条件，还需要加一些额外的限制条件, 才能保证最优解的充分性
- 假设 f 在点 x 的一个开邻域内是二阶连续可微的, 考虑

$$f(x+d)=f(x)+\nabla f(x)^\top d+\frac{1}{2}d^\top\nabla^2 f(x)d+o(\|d\|^2)$$

- 当一阶必要条件满足时, 简化为

$$f(x+d)=f(x)+\frac{1}{2}d^\top\nabla^2 f(x)d+o(\|d\|^2)$$

二阶最优性条件

- 假设 f 在点 x 的一个开邻域内是二阶连续可微的，则以下最优性条件成立
  - 二阶必要条件 若x^* 是 f 的一个局部极小点, 则

$$\nabla f(x^*) = 0, \nabla^2 f(x^*) \succeq 0$$

  - 二阶充分条件 若满足

$$\nabla f(x^*) = 0, \nabla^2 f(x^*) \succ 0$$

  则 x^* 是 f 的一个局部极小点

- 对于给定点的全局最优性判断还需要借助实际问题的性质

二阶最优性条件

- **必要性** 若 $\nabla^2 f(x^*)$ 有负的特征值 $\lambda_- < 0$, 设 $\nabla^2 f(x^*)d = \lambda_- d$, 则

$$\frac{f(x^*+d)-f(x^*)}{\|d\|^2} = \frac{1}{2}\frac{d^\top}{\|d\|}\nabla^2 f(x^*)\frac{d}{\|d\|} + o(1) = \frac{1}{2}\lambda_- + o(1)$$

当 $\|d\|$ 充分小时, $f(x^*+d) < f(x^*)$, 这和点 x^* 的最优性矛盾

- **充分性** 由 $\nabla f(x^*) = 0$ 时的二阶展开,

$$\frac{f(x^*+d)-f(x^*)}{\|d\|^2} = \frac{d^\top \nabla^2 f(x^*)d + o(\|d\|^2)}{\|d\|^2} \geq \frac{1}{2}\lambda_{\min} + o(1).$$

当 $\|d\|$ 充分小时有 $f(x^*+d) \geq f(x^*)$, 即二阶充分条件成立

实例：实数情形的相位恢复

- **考虑**

$$\min_{x\in\mathbb{R}^n} \quad f(x) = \sum_{i=1}^{m} r_i^2(x)$$

其中 $r_i(x) = (a_i^\top x)^2 - b_i^2, i = 1, 2, \cdots, m$

- **计算梯度和的海瑟矩阵**

$$\nabla f(x) = 2\sum_{i=1}^{m} r_i(x)\nabla r_i(x) = 4\sum_{i=1}^{m}((a_i^\top x)^2 - b_i^2)(a_i^\top x)a_i$$

$$\nabla^2 f(x) = \sum_{i=1}^{m}(12(a_i^\top x)^2 - 4b_i^2)a_i a_i^\top$$

目录

无约束不可微问题的最优性理论

- 仍考虑问题

$$\min_{x\in\mathbb{R}^n} \quad f(x)$$

- $f(x)$ 是不可微函数，例如 $\|x\|_1$
- 目标函数可能不存在梯度和海瑟矩阵

凸优化问题一阶充要条件

- 假设 f 是适当且凸的函数, 则 x^* 为全局极小点当且仅当 $0 \in \partial f(x^*)$
  - **必要性** 因为 x^* 为全局极小点, 所以

$$f(y) \geq f(x^*) = f(x^*) + 0^\top (y - x^*), \quad \forall y \in \mathbb{R}^n$$

    因此 $0 \in \partial f(x^*)$
  - **充分性** 如果 $0 \in \partial f(x^*)$, 那么根据次梯度的定义

$$f(y) \geq f(x^*) + 0^\top (y - x^*) = f(x^*), \quad \forall y \in \mathbb{R}^n$$

    因而 x^* 为一个全局极小点

复合优化问题的一阶必要条件

- 考虑一般复合优化问题

$$\min_{x\in\mathbb{R}^n} \quad \psi(x) = f(x) + h(x)$$

其中 f 为光滑函数 (可能非凸), h 为凸函数 (可能非光滑)

- 定理 4.5　令 x^* 为复合优化问题的一个局部极小点, 那么

$$-\nabla f(x^*) \in \partial h(x^*)$$

其中 $\partial h(x^*)$ 为凸函数 h 在点 x^* 处的次梯度集合

- 由于目标函数可能是整体非凸的，因此一般没有一阶充分条件

实例: ℓ_1 范数优化问题

- 考虑

$$\min_{x\in\mathbb{R}^n} \quad \psi(x) = f(x) + \mu\|x\|_1$$

- $\|x\|_1$ 不是可微的, 但可以计算其次微分

$$\partial_i \|x\|_1 = \begin{cases} \{1\}, & x_i > 0 \\ [-1, 1], & x_i = 0 \\ \{-1\}, & x_i < 0 \end{cases}$$

- 若 x^* 是局部最优解, 则 $-\nabla f(x^*) \in \mu\partial\|x^*\|_1$, 即

$$\nabla_i f(x^*) = \begin{cases} -\mu, & x_i^* > 0 \\ a \in [-\mu, \mu], & x_i^* = 0 \\ \mu, & x_i^* < 0 \end{cases}$$

目录

对偶理论

- 一般的约束优化问题

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & c_i(x) \le 0, \; i \in \mathcal{I} \\ & c_i(x) = 0, \; i \in \mathcal{E} \end{aligned}$$

- 可行域定义为

$$\mathcal{X} = \{x \in \mathbb{R}^n \mid c_i(x) \le 0, \; i \in \mathcal{I} \text{ 且 } c_i(x) = 0, \; i \in \mathcal{E}\}$$

- 通过将 $\mathcal{X}$ 的示性函数加到目标函数中可以得到无约束优化问题，但是转化后问题的目标函数是不连续的、不可微的以及不是有限的

拉格朗日函数

- 拉格朗日函数 $L : \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^p \to \mathbb{R}$

$$L(x, \lambda, \nu) = f(x) + \sum_{i \in \mathcal{I}} \lambda_i c_i(x) + \sum_{i \in \mathcal{E}} \nu_i c_i(x)$$

  - λ_i 为第 i 个不等式约束对应的拉格朗日乘子
  - ν_i 为第 i 个等式约束对应的拉格朗日乘子

- 拉格朗日对偶函数 $g : \mathbb{R}^m_+ \times \mathbb{R}^p \to [-\infty, +\infty)$

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathbb{R}^n} \; L(x, \lambda, \nu) \\ &= \inf_{x \in \mathbb{R}^n} \left(f(x) + \sum_{i \in \mathcal{I}} \lambda_i c_i(x) + \sum_{i \in \mathcal{E}} \nu_i c_i(x)\right) \end{aligned}$$

拉格朗日对偶函数

- **引理 4.1** 若 $\lambda \geq 0$, 则 $g(\lambda, \nu) \leq p^*$

证明 若 $\tilde{x} \in \mathcal{X}$, 则

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) \leq L(\tilde{x}, \lambda, \nu) \leq f(\tilde{x})$$

对 $\tilde{x}$ 取下界得

$$g(\lambda, \nu) \leq \inf_{\tilde{x} \in \mathcal{X}} f(\tilde{x}) = p^*$$

拉格朗日对偶问题

- 拉格朗日对偶问题

$$\max_{\lambda \geq 0, \nu} g(\lambda, \nu) = \max_{\lambda \geq 0, \nu} \inf_{x \in \mathbb{R}^n} L(x, \lambda, \nu)$$

  - 称 λ 和 ν 为对偶变量，设最优值为 q^*
  - q^* 为 p^* 的最优下界，称 $p^* - q^*$ 为对偶间隙
  - 拉格朗日对偶问题是一个凸优化问题
  - $g = \{(\lambda, \nu) \mid \lambda \geq 0, g(\lambda, \nu) > -\infty\}$，称其元素为对偶可行解

例子 标准形式线性规划及其对偶

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned} \qquad\qquad \begin{aligned} \max \quad & -b^\top \nu \\ \text{s.t.} \quad & A^\top \nu + c \geq 0 \end{aligned}$$

弱对偶性与强对偶性

- 弱对偶性 $\quad d^* \le p^*$
  - 对凸问题与非凸问题都成立
  - 可导出复杂问题的非平凡下界, 例如, SDP 问题

$$\begin{array}{rl} \max & -\mathbf{1}^\top \nu \\ \text{s.t.} & W + \operatorname{diag}(\nu) \succeq 0 \end{array}$$

  给出了二路划分问题的一个下界

$$\min \quad x^\top W x \quad \text{s.t.} \quad x_i^2 = 1, \quad i = 1, \cdots, n$$

- 强对偶性 $\quad d^* = p^*$
  - 对一般问题而言通常不成立
  - (通常) 对凸问题成立
  - 称保证凸问题强对偶性成立的条件为约束品性

适当锥与广义不等式

- 称满足如下条件的锥 K 为适当锥
  - K 是凸锥
  - K 是闭集
  - K 是实心的 (solid), 即 $\mathrm{int}\ K \neq \emptyset$
  - K 是尖的 (pointed), 即对任意非零向量 x, 若 $x \in K$, 则 $-x \notin K$
- 适当锥 K 可以诱导出广义不等式, 它定义了全空间上的偏序关系

$$x \preceq_K y \iff y - x \in K$$

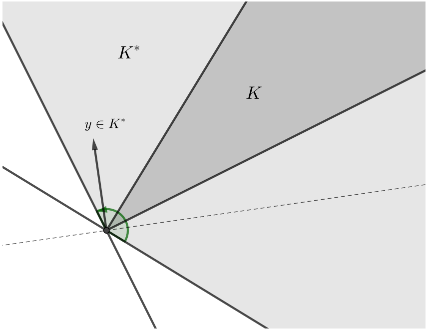
- 当 $K = \mathbb{R}^n_+$ 时, $x \preceq_K y$ 是我们之前经常使用的记号 $x \leq y$
- 当 $K = \mathcal{S}^n_+$ 时, $X \preceq_K Y$ 表示 $Y - X \succeq 0$, 即 $Y - X$ 是半正定矩阵

对偶锥与拉格朗日乘子

- 令 K 为全空间 Ω 的子集，称集合

$$K^* = \{y \in \Omega \mid \langle x, y\rangle \geq 0,\ \forall x \in K\}$$

为其**对偶锥**



注记

- 假设非负锥 $K=\mathbb{R}^n_+, \Omega=\mathbb{R}^n$, 定义 $\langle x,y\rangle = x^\top y$, 那么 $K^*=\mathbb{R}^n_+$
- 假设半正定锥 $K=\mathcal{S}^n_+, \Omega=\mathcal{S}^n$, 定义

  $$\langle X,Y\rangle = \mathrm{Tr}(XY^\top)$$

  可以证明

  $$\langle X,Y\rangle \geq 0,\ \forall X\in\mathcal{S}^n_+ \quad \Leftrightarrow \quad Y\in\mathcal{S}^n_+$$

  即半正定锥的对偶锥仍为半正定锥
- 称满足 $K=K^*$ 的锥 K 为自对偶锥

广义不等式约束优化问题拉格朗日函数的构造

- 广义不等式约束优化问题

$$
\begin{aligned}
\min_{x\in\mathbb{R}^n} \quad & f(x) \\
\text{s.t.} \quad & c_i(x) \preceq_{K_i} 0,\ i \in \mathcal{I} \\
& c_i(x) = 0,\ i \in \mathcal{E}
\end{aligned}
$$

- 拉格朗日函数

$$
L(x,\lambda,\nu) = f(x) + \sum_{i\in\mathcal{I}} c_i(x)\lambda_i + \sum_{i\in\mathcal{E}} \nu_i c_i(x), \quad \lambda_i \in K_i^*,\ \nu_i \in \mathbb{R}
$$

- 容易验证 $L(x,\lambda,\nu) \le f(x),\ \forall\, x \in \mathcal{X},\ \lambda_i \in K_i^*,\ \nu_i \in \mathbb{R}$
- 对偶函数 $g(\lambda,\nu) = \inf_{x\in\mathbb{R}^n}\ L(x,\lambda,\nu)$，对偶问题为

$$
\max_{\lambda_i\in K_i^*,\ \nu_i\in\mathbb{R}} \quad g(\lambda,\nu)
$$

实例：线性规划问题的对偶

- **考虑线性规划问题**

$$
\begin{aligned}
\min_{x} \quad & c^\top x \\
\text{s.t.} \quad & Ax = b \\
& x \geq 0
\end{aligned}
$$

- **拉格朗日函数**

$$
L(x, s, \nu) = c^\top x + \nu^\top (Ax - b) - s^\top x = -b^\top \nu + (A^\top \nu - s + c)^\top x
$$

- **对偶函数**

$$
g(s, \nu) = \inf_{x} \ L(x, s, \nu) = \begin{cases} -b^\top \nu, & A^\top \nu - s + c = 0 \\ -\infty, & \text{其他} \end{cases}
$$

实例：线性规划问题的对偶

- **对偶问题**

$$
\begin{array}{rl} \displaystyle\max_{s,\nu} & -b^\top \nu \\ \text{s.t.} & A^\top \nu - s + c = 0 \\ & s \ge 0 \end{array}
\quad \overset{y=-\nu}{\Longleftrightarrow} \quad
\begin{array}{rl} \displaystyle\max_{s,y} & b^\top y \\ \text{s.t.} & A^\top y + s = c \\ & s \ge 0 \end{array}
$$

- **若保留约束 $x \ge 0$, 则拉格朗日函数为**

$$L(x, y) = c^\top x - y^\top (Ax - b) = b^\top y + (c - A^\top y)^\top x$$

- **对偶问题需要将 $x \ge 0$ 添加到约束里**

$$
\max_{y} \left\{ \inf_{x}\ b^\top y + (c - A^\top y)^\top x \quad \text{s.t.} \quad x \ge 0 \right\} \quad \Rightarrow \quad
\begin{array}{rl} \displaystyle\max_{y} & b^\top y \\ \text{s.t.} & A^\top y \le c \end{array}
$$

实例：线性规划问题的对偶

- 将max $b^\top y$ 改写为 min $-b^\top y$，对偶问题的拉格朗日函数为

$$L(y,x) = -b^\top y + x^\top (A^\top y - c) = -c^\top x + (Ax-b)^\top y$$

- 得到对偶函数

$$g(x) = \inf_y \; L(y,x) = \begin{cases} -c^\top x, & Ax = b \\ -\infty, & \text{其他} \end{cases}$$

- 相应的对偶问题是

$$\begin{aligned} \max_x \quad & -c^\top x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

- 该问题与原始问题完全等价，表明线性规划问题与其对偶问题互为对偶

实例: ℓ_1 正则化问题的对偶

- 考虑

$$\min_{x\in\mathbb{R}^n} \quad \frac{1}{2}\|Ax-b\|^2+\mu\|x\|_1$$

- 令 $r=Ax-b$, 问题等价于

$$\begin{aligned}\min_{x\in\mathbb{R}^n} \quad & \frac{1}{2}\|r\|^2+\mu\|x\|_1 \\ \text{s.t.} \quad & r=Ax-b\end{aligned}$$

- 拉格朗日函数

$$\begin{aligned}L(x,r,\lambda) &= \frac{1}{2}\|r\|^2+\mu\|x\|_1-\langle\lambda, Ax-b-r\rangle \\ &= \frac{1}{2}\|r\|^2+\lambda^\top r+\mu\|x\|_1-(A^\top\lambda)^\top x+b^\top\lambda\end{aligned}$$

实例: ℓ_1 正则化问题的对偶

- **对偶函数**

$$g(\lambda) = \inf_{x,r} \ L(x, r, \lambda) = \begin{cases} b^\top \lambda - \frac{1}{2}\|\lambda\|^2, & \|A^\top \lambda\|_\infty \le \mu \\ -\infty, & \text{其他} \end{cases}$$

- **对偶问题**

$$\begin{aligned} \max_{\lambda} \quad & b^\top \lambda - \frac{1}{2}\|\lambda\|^2 \\ \text{s.t.} \quad & \|A^\top \lambda\|_\infty \le \mu \end{aligned}$$

实例：半定规划问题的对偶问题

- 考虑

$$
\begin{aligned}
\min_{X \in \mathcal{S}^n} \quad & \langle C, X \rangle \\
\text{s.t.} \quad & \langle A_i, X \rangle = b_i, \ i = 1, 2, \cdots, m \\
& X \succeq 0
\end{aligned}
$$

- 拉格朗日函数

$$
L(X, y, S) = \langle C, X \rangle - \sum_{i=1}^{m} y_i(\langle A_i, X \rangle - b_i) - \langle S, X \rangle, \quad S \succeq 0
$$

半定规划对偶问题的对偶问题

- 对偶函数

$$g(y,S)=\inf_{X}\ L(X,y,S)=\begin{cases} b^\top y, & \sum_{i=1}^{m} y_iA_i-C+S=0 \\ -\infty, & \text{其他} \end{cases}$$

- 对偶问题

$$\begin{aligned} \min_{y\in\mathbb{R}^m}\quad & -b^\top y \\ \text{s.t.}\quad & \sum_{i=1}^{m} y_iA_i-C+S=0 \\ & S\succeq 0 \end{aligned}$$

Q&A

Thank you!

感谢您的聆听和反馈