# 第四章 最优性理论

修贤超

https://xianchaoxiu.github.io

# 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

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

- 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 $L: \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^p \to \mathbb{R}$

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

  - $\lambda_i$ 为第 $i$ 个不等式约束对应的拉格朗日乘子
  - $\nu_i$ 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

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

- 设 $p^*$ 是原始问题的最优解， $q^*$ 是对偶问题的最优解
- 弱对偶性 $q^*\leq p^*$
  - 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
- 强对偶性 $q^*=p^*$
  - (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

# 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**

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

- **若保留约束 $x \geq 0$, 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 $x \geq 0$ 添加到约束里**

$$\max_y \; \{\inf_x \; b^\top y + (c - A^\top y)^\top x \quad \text{s.t.} \quad x \geq 0\} \quad \Rightarrow \quad \begin{aligned} \max_y \quad & b^\top y \\ \text{s.t.} \quad & A^\top y \leq c \end{aligned}$$

# 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: $\ell_1$ 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

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

- 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: $\ell_1$ 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

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

## 实例：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**

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

# 目录

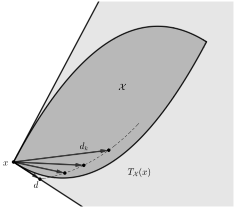
# 切锥

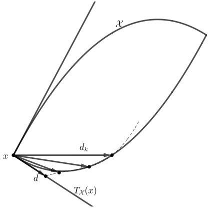
- 给定可行域 $\mathcal{X}$ 及 $x \in \mathcal{X}$, 若存在序列 $\{z_k\}_{k=1}^{\infty} \subset \mathcal{X}, \lim_{k\to\infty} z_k = x$ 以及正标量序列 $\{t_k\}_{k=1}^{\infty}$, $t_k \to 0$ 满足

$$\lim_{k\to\infty} \frac{z_k - x}{t_k} = d$$

  则称向量 $d$ 为 $\mathcal{X}$ 在点 $x$ 处的一个**切向量**

- 所有点 $x$ 处的切向量构成的集合称为**切锥**, 用 $T_{\mathcal{X}}(x)$ 表示



# 几何最优性条件

- **一般优化问题**

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & c_i(x) \le 0,\ i \in \mathcal{I} \\ & c_i(x) = 0,\ i \in \mathcal{E} \end{aligned} \tag{1}$$

- **定理 4.6** 假设可行点 $x^*$ 是问题 (1) 的一个局部极小点. 如果 $f(x)$ 和 $c_i(x)$, $i \in \mathcal{I} \cup \mathcal{E}$ 在点 $x^*$ 处是可微的, 那么

$$d^\top \nabla f(x^*) \ge 0, \quad \forall d \in T_{\mathcal{X}}(x^*)$$

等价于

$$T_{\mathcal{X}}(x^*) \cap \{d \mid \nabla f(x^*)^\top d < 0\} = \emptyset$$

# 线性化可行锥

- 定义 4.6 对于可行点$x \in \mathcal{X}$, 定义积极集 $\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} \mid c_i(x) = 0\}$, 点 $x$ 处的线性化可行方向锥定义为

$$\mathcal{F}(x) = \left\{ d \;\middle|\; \begin{array}{l} d^\top \nabla c_i(x) = 0, \ \forall\ i \in \mathcal{E} \\ d^\top \nabla c_i(x) \le 0, \ \forall\ i \in \mathcal{A}(x) \cap \mathcal{I} \end{array} \right\}$$

- 命题 4.1 设 $c_i(x), i \in \mathcal{E} \cup \mathcal{I}$ 一阶连续可微, 则对任意可行点 $x$ 有

$$T_{\mathcal{X}}(x) \subseteq \mathcal{F}(x)$$

- 线性可行化方向锥容易计算, 但不能反映可行域 $\mathcal{X}$ 的本质特征
- 切锥能反映可行域 $\mathcal{X}$ 的本质特征, 但不容易计算
- 引入约束品性, 确保 $T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$, 从而用 $\mathcal{F}(x)$ 取代 $T_{\mathcal{X}}(x)$

# KKT 条件

- **定理 4.7** 假设 $x^*$ 是一般优化问题 (1) 的一个局部最优点

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

如果 $T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$ 成立, 那么存在拉格朗日乘子 $\lambda_i^*$ 使得

稳定性条件 $\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) + \sum_{i\in\mathcal{I}\cup\mathcal{E}} \lambda_i^* \nabla c_i(x^*) = 0$

原始可行性条件 $c_i(x^*) = 0,\ \forall i \in \mathcal{E}$

原始可行性条件 $c_i(x^*) \le 0,\ \forall i \in \mathcal{I}$

对偶可行性条件 $\lambda_i^* \ge 0,\ \forall i \in \mathcal{I}$

互补松弛条件 $\lambda_i^* c_i(x^*) = 0,\ \forall i \in \mathcal{I}$

# 临界锥

- 若 $x^*$ 是满足 KKT 条件的点, 假设 $T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$, 则 $\forall d \in \mathcal{F}(x^*)$ 有

$$d^\top \nabla f(x^*) = -\sum_{i\in\mathcal{E}} \underbrace{\lambda_i^* d^\top \nabla c_i(x^*)}_{=0} - \sum_{i\in\mathcal{A}(x^*)\cap\mathcal{I}} \underbrace{\lambda_i^* d^\top \nabla c_i(x^*)}_{\leq 0} \geq 0$$

- 定义 4.10 设$(x^*, \lambda^*)$ 是满足 KKT 条件的 KKT 对, 定义临界锥为

$$\mathcal{C}(x^*, \lambda^*) = \{d \in \mathcal{F}(x^*) \mid d^\top \nabla c_i(x^*) = 0,\ \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ 且 } \lambda_i^* > 0\}$$

- 当 $d \in \mathcal{C}(x^*, \lambda^*)$ 时, $\forall i \in \mathcal{E} \cup \mathcal{I}$ 有 $\lambda_i^* \nabla c_i(x^*)^\top d = 0$, 故

$$d^\top \nabla f(x^*) = \sum_{i\in\mathcal{E}\cup\mathcal{I}} \lambda_i^* d^\top \nabla c_i(x^*) = 0$$

# 二阶最优性条件

- **定理 4.8 (二阶必要条件)** 假设 $x^*$ 是一个局部最优解, 且 $T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$. 令 $\lambda^*$ 为相应的拉格朗日乘子, 即 $(x^*, \lambda^*)$ 满足 KKT 条件, 那么

$$d^\top \nabla^2_{xx} L(x^*, \lambda^*) d \geq 0, \quad \forall d \in \mathcal{C}(x^*, \lambda^*)$$

- **定理 4.9 (二阶充分条件)** 假设在可行点 $x^*$ 处, 存在一个拉格朗日乘子 $\lambda^*$, 使得 $(x^*, \lambda^*)$ 满足 KKT 条件. 如果

$$d^\top \nabla^2_{xx} L(x^*, \lambda^*) d > 0, \quad \forall d \in \mathcal{C}(x^*, \lambda^*), \ d \neq 0$$

  那么 $x^*$ 为一个严格局部极小解

- 回顾无约束优化问题的二阶最优性条件

# 例子

- 考虑
$$\min \quad x_1^2 + x_2^2 \quad \text{s.t.} \quad \frac{x_1^2}{4} + x_2^2 - 1 = 0$$

- 拉格朗日函数为
$$L(x,\lambda) = x_1^2 + x_2^2 + \lambda(\frac{x_1^2}{4} + x_2^2 - 1)$$

- 该问题可行域在任意一点 $x = (x_1, x_2)^\top$ 处的线性化可行方向锥为
$$\mathcal{F}(x) = \{(d_1, d_2) \mid \frac{x_1}{4} d_1 + x_2 d_2 = 0\}$$

- 根据 $\mathcal{C}(x,\lambda) = \mathcal{F}(x)$，计算出 4 个 KKT 对
$$(x^\top, \lambda) = (2, 0, -4),\ (-2, 0, -4),\ (0, 1, -1),\ (0, -1, -1)$$

# 例子

- 第一个 KKT 对 $y = (2, 0, -4)$, 计算可得

$$\nabla^2_{xx} L(y) = \begin{bmatrix} 0 & 0 \\ 0 & -6 \end{bmatrix}, \quad \mathcal{C}(y) = \{(d_1, d_2) \mid d_1 = 0\}$$

取 $d = (0, 1)$, 则 $d^\top \nabla^2_{xx} L(y) d = -6 < 0$, 因此 $y$ 不是局部最优点

- 第三个 KKT 对 $z = (0, 1, -1)$, 计算可得

$$\nabla^2_{xx} L(z) = \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathcal{C}(z) = \{(d_1, d_2) \mid d_2 = 0\}$$

对于任意的 $d = (d_1, 0)$ 且 $d_1 \neq 0$, 有 $d^\top \nabla^2_{xx} L(z) d = \frac{3}{2} d_1^2 > 0$, 因此 $z$ 是一个严格局部最优点

# 目录

- 4.1 最优化问题解的存在性
- 4.2 无约束可微问题的最优性理论
- 4.3 无约束不可微问题的最优性理论
- 4.4 对偶理论
- 4.5 一般约束优化问题的最优性理论
- 4.6 带约束凸优化问题的最优性理论

# 带约束凸优化问题

- 考虑带约束的凸优化问题

$$\begin{aligned} \min_{x \in \mathcal{D}} \quad & f(x) \\ \text{s.t.} \quad & c_i(x) \leqslant 0, \quad i = 1, 2, \cdots, m \\ & Ax = b \end{aligned} \tag{2}$$

  - $f(x)$ 为适当的凸函数
  - $c_i(x)$ 是凸函数且 $\mathrm{dom}\ c_i = \mathbb{R}^n$
  - 集合 $\mathcal{D}$ 表示自变量 $x$ 的定义域，即 $\mathcal{D} = \{x \mid f(x) < +\infty\}$

# Slater 约束品性与强对偶原理

- **定义 4.11** 集合 $\mathcal{D}$ 的相对内点集定义为

$$\mathrm{relint}\ \mathcal{D} = \{x \in \mathcal{D} \mid \exists\ r > 0,\ \text{使得}\ B(x,r) \cap \mathrm{affine}\ \mathcal{D} \subseteq \mathcal{D}\}$$

- **定义 4.12** 若对凸优化问题 (2) 存在 $x \in \mathrm{relint}\ \mathcal{D}$ 满足

$$c_i(x) < 0, \quad i = 1, 2, \cdots, m, \quad Ax = b$$

  则称满足 Slater 约束条件

- **定理 4.10** 若凸优化问题满足 Slater 条件, 则强对偶原理成立

# 一阶充要条件

- **定理 4.11** 对于凸优化问题 (2), 如果 Slater 条件成立, 那么 $x^*, \lambda^*$ 分别是原始、对偶全局最优解当且仅当

$$
\begin{aligned}
\text{稳定性条件} \quad & 0 \in \partial f(x^*) + \sum_{i \in \mathcal{I}} \lambda_i^* \partial c_i(x^*) + \sum_{i \in \mathcal{E}} \lambda_i^* a_i \\
\text{原始可行性条件} \quad & Ax^* = b, \ \forall i \in \mathcal{E} \\
\text{原始可行性条件} \quad & c_i(x^*) \le 0, \ \forall i \in \mathcal{I} \\
\text{对偶可行性条件} \quad & \lambda_i^* \ge 0, \ \forall i \in \mathcal{I} \\
\text{互补松弛条件} \quad & \lambda_i^* c_i(x^*) = 0, \ \forall i \in \mathcal{I}
\end{aligned}
$$

# 实例：仿射空间的投影问题

- 考虑

$$\min_{x\in\mathbb{R}^n} \quad \frac{1}{2}\|x-y\|_2^2 \qquad \text{s.t.} \quad Ax=b$$

- 拉格朗日函数 $L(x,\lambda)=\frac{1}{2}\|x-y\|^2+\lambda^\top(Ax-b)$
- KKT 条件

$$\begin{cases} x^*-y+A^\top\lambda^*=0 \\ Ax^*=b \end{cases}$$

- 第一式左右两边同时左乘 $A$ 可得

$$Ax^*-Ay+AA^\top\lambda=0 \quad \Rightarrow \quad \lambda^*=(AA^\top)^{-1}(Ay-b)$$

- 将 $\lambda^*$ 代回第一式可知

$$x^*=y-A^\top(AA^\top)^{-1}(Ay-b)$$

# 总结

- 无约束优化问题及其最优性条件

| 问题 | 一阶条件 | 二阶条件 |
| --- | --- | --- |
| 可微问题 | $\nabla f(x^*) = 0$ (必要) | 必要/充分 |
| 凸问题 | $0 \in \partial f(x^*)$ (充要) | – |
| 复合优化问题 | $-\nabla f(x^*) \in \partial h(x^*)$ (必要) | – |
| 非凸非光滑 | $0 \in \partial f(x^*)$ (必要) | – |

- 约束优化问题的最优性条件和相应约束品性

| 问题 | 一阶条件 | 二阶条件 | 约束品性 |
| --- | --- | --- | --- |
| 一般问题 | KKT 条件 (必要) | 必要/充分 | LICQ |
| 凸问题 | KKT 条件 (充要) | – | Slater |

# *Q&A*

# *Thank you!*

感谢您的聆听和反馈