优化理论与算法

第六章

对偶理论

郭加熠|助理教授

目录

对偶理论

- 对优化理论至关重要
- 其理论与实践中的影响深远
- 一种强大的工具，提供：
  - 优化算法丰富类别发展的基础
  - 一种系统的泛化方法，用于构建上下界策略（适用于连续与离散优化）
  - 敏感性分析的基础

对偶理论的基本思想与核心问题

- 将一个“等价的对偶问题”与给定的（原始）问题关联起来
- 该方法适用于一般约束优化问题
- 研究的问题包括：
  - 原始问题与其对偶问题之间是否存在一般性的关系？
  - 在什么条件下，原始问题与对偶问题具有相同的最优值？
  - 在什么条件下，原始与对偶问题的最优解存在？
  - 原始与对偶问题的最优解之间有何关系？
  - 对偶最优解可以提供关于原始问题哪些方面的信息？

拉格朗日函数（Lagrangian）

标准形式问题（不要求是凸的）：

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{aligned}$$

变量 $x \in \mathbb{R}^n$，定义域 $\mathcal{D}$，最优值为 $p^\star$

拉格朗日函数：$L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$，定义域为 $\mathrm{dom}\, L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$，其形式为：

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- 目标函数与约束函数的加权和
- λ_i, ν_i 分别是不等式约束 $f_i(x) \le 0$ 与等式约束 $h_i(x) = 0$ 对应的**拉格朗日乘子**

拉格朗日对偶函数（Lagrange Dual Function）

拉格朗日对偶函数：$g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$，定义为：

$$
\begin{aligned}
g(\lambda, \nu) &= \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \\
&= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)
\end{aligned}
$$

- 函数 g 是**凹函数**，但在某些 λ, ν 情况下可能为 $-\infty$
- **下界性质（Lower bound property）**：若 $\lambda \geq 0$，则 $g(\lambda, \nu) \leq p^\star$

证明：若 $\tilde{x}$ 是可行解，且 $\lambda \geq 0$，则有

$$
f_0(\tilde{x}) \geq L(\tilde{x}, \lambda, \nu) \geq \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = g(\lambda, \nu)
$$

对所有可行的 $\tilde{x}$ 取最小值，可得：$p^\star \geq g(\lambda, \nu)$

标准形式线性规划（Standard Form LP）

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

拉格朗日函数：$L(x, \lambda, \nu) = c^T x + \nu^T (Ax - b) - \lambda^T x = -b^T \nu + (c + A^T \nu - \lambda)^T x$

L 是关于 x 的仿射函数，因此**拉格朗日对偶函数**：

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} -b^T \nu, & c + A^T \nu - \lambda = 0 \\ -\infty, & \text{否则} \end{cases}$$

- g 是定义在 $\{(\lambda, \nu) \mid c + A^T \nu - \lambda = 0\}$ 上的线性函数，因此是凹函数
- 下界性质（lower bound property）：$p^\star \geq -b^T \nu \quad \text{if} \;\; c + A^T \nu \geq 0$

对偶问题

拉格朗日对偶问题：

$$
\begin{aligned}
\max \quad & g(\lambda, \nu) \\
\text{s.t.} \quad & \lambda \geq 0
\end{aligned}
$$

- 利用拉格朗日函数，找到 p^* 最好的下界
- 对偶问题是**凸问题**，最优值用 d^* 表示
- 如果 $(\lambda, \nu) \in \mathsf{dom}(g)$ 且 $\lambda \geq 0$，那么 (λ, ν) 是对偶问题**可行解**
- 常常将隐含约束 $(\lambda, \nu) \in \mathsf{dom}(g)$ 显现写出

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

线性方程组的最小范数解（Least-norm Solution）

$$\begin{aligned} \min \quad & x^T x \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

拉格朗日函数：$L(x,\nu) = x^T x + \nu^T (Ax - b)$

对 x 最小化 L，令梯度为零：$\nabla_x L(x,\nu) = 2x + A^T \nu = 0 \quad \Rightarrow \quad x = -\frac{1}{2} A^T \nu$

将 x 代入 L 得到**拉格朗日对偶函数** g：

$$g(\nu) = L\left(-\frac{1}{2}A^T\nu, \nu\right) = -\frac{1}{4}\nu^T A A^T \nu - b^T \nu$$

- $g(\nu)$ 是 ν 的一个凹函数
- 下界性质（lower bound property）：

  $p^\star \geq -\frac{1}{4}\nu^T A A^T \nu - b^T \nu, \quad$ 对所有ν 成立

二次规划（Quadratic Program）

原始问题（Primal Problem）（假设 $P \in \mathbb{S}^n_{++}$）：

$$\begin{aligned} \min \quad & x^T P x \\ \text{s.t.} \quad & Ax \leq b \end{aligned}$$

对偶函数（Dual Function）：

$$g(\lambda) = \inf_x \left(x^T P x + \lambda^T (Ax - b) \right) = -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

对偶问题（Dual Problem）：

$$\begin{aligned} \max \quad & -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda \\ \text{s.t.} \quad & \lambda \geq 0 \end{aligned}$$

拉格朗日对偶与共轭函数

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & Ax \leq b, \quad Cx = d \end{aligned}$$

拉格朗日对偶函数: $$g(\lambda,\nu) = \inf_{x\in \mathrm{dom} f_0} \left(f_0(x) + (A^T\lambda + C^T\nu)^T x - b^T\lambda - d^T\nu \right)$$
$$= -f_0^*(-A^T\lambda - C^T\nu) - b^T\lambda - d^T\nu$$

- 回顾共轭函数定义： $f^*(y) = \sup_{x\in \mathrm{dom}\ f} \left(y^T x - f(x)\right)$
- 如果已知 f_0 的共轭函数，共轭函数可简化对偶函数的推导

例子：最大熵问题（Entropy Maximization）：

$$f_0(x) = \sum_{i=1}^n x_i \log x_i, \quad f_0^*(y) = \sum_{i=1}^n e^{y_i - 1}$$

两分划分问题（Two-way Partitioning）

$$\begin{aligned} \min \quad & x^T W x \\ \text{s.t.} \quad & x_i^2 = 1, \quad i = 1, \dots, n \end{aligned}$$

- 非凸问题：将集合 $\{1, \dots, n\}$ 分两组
- W 表示分到同一组与不同组的代价

拉格朗日对偶： $g(\nu) = \inf_x \left(x^T W x + \sum_i \nu_i (x_i^2 - 1) \right) = \inf_x \left(x^T (W + \mathrm{diag}(\nu)) x \right) - \mathbf{1}^T \nu$

$$= \begin{cases} -\mathbf{1}^T \nu, & \text{if} \quad W + \mathrm{diag}(\nu) \succeq 0 \\ -\infty, & \text{otherwise} \end{cases}$$

下界性质（lower bound property）：$p^\star \geq -\mathbf{1}^T \nu \quad \text{if} \quad W + \mathrm{diag}(\nu) \succeq 0$

举例：令 $\nu = -\lambda_{\min}(W)\mathbf{1}$，则有：$p^\star \geq n\lambda_{\min}(W)$

目录

弱对偶与强对偶（Weak and Strong Duality）

弱对偶（weak duality）：$d^\star \le p^\star$

- 总是成立（无论问题是凸的还是非凸的）
- 可用于为困难问题提供非平凡的下界

- 例如，求解如下 SDP 问题：

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

强对偶（strong duality）：$d^\star = p^\star$

- 一般不成立
- 对于凸优化问题通常成立
- 保证强对偶成立的条件被称为：**约束资格条件（constraint qualifications）**

强对偶的示例（Examples with Strong Duality）

- **唯一的对偶最优解（Unique Dual Optimal）：**

$$
\begin{aligned}
\min \quad & \frac{1}{2}(x_1^2 + x_2^2) \\
\text{s.t.} \quad & x_1 \leq 1
\end{aligned}
$$

- **多个对偶最优解（Multiple Dual Optimal）：**

$$
\min \quad |x_1| - x_2 \quad \text{s.t.} \quad x_1 \leq 0,\ x_2 = 0
$$

- **无对偶最优解（No Dual Optimal）：**

$$
\min \quad x \quad \text{s.t.} \quad x^2 \leq 0
$$

类似地，$d^\star = p^\star$ 与原始问题是否有最优解无关。

目录

几何解释

原始问题（不一定是凸的）且最优值记为 $p^* > -\infty$：

$$\min \quad f(x)$$

$$\text{s.t.} \quad f_1(x) \le 0, \quad x \in D$$

几何框架：定义集合 V 如下：

$$V = \{(u, w) \mid \exists x \in X \ \text{ such that } \ f_1(x) \le u, \ f(x) \le w\}$$

几何原始问题：确定集合 V 与 w 轴的最小截距：

$$\min \quad w$$

$$\text{s.t.} \quad (0, w) \in V$$

最小截距值记为 p^*，即：$p^* = \inf_{(0,w)\in V} w$

几何意义

几何框架: 定义集合 V 如下:

$$V = \{(u, w) \mid \exists x \in X \ \text{ such that } \ f_1(x) \le u,\ f(x) \le w\}$$

对偶函数:

$$q(\lambda) = \inf_{x \in D} \{f(x) + \lambda f_1(x)\} = \inf_{(u,w) \in V} \{w + \lambda u\}, \quad \lambda \succeq 0$$

对偶问题:

$$\begin{aligned} \max \quad & q(\lambda) \\ \text{s.t.} \quad & \lambda \succeq 0 \end{aligned}$$

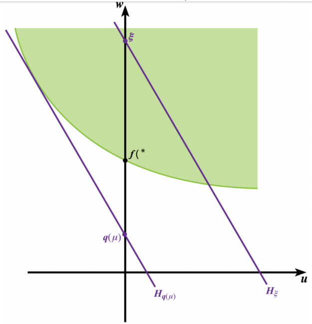
非垂直超平面（Nonvertical Hyperplanes）

- 记 H 为 $\mathbb{R}^m \times \mathbb{R}$ 中的一个**非垂直超平面**，定义为：

$$H = \{(u, w) \mid \lambda u + w = q(\lambda)\}$$

其中 $q(\lambda) = \inf_{(u,w)\in V}\{w + \lambda u\}, \quad \lambda \succeq 0$

- 超平面 H 是集合 V 的一个**支撑超平面**，与 w 轴的交点为 $(0, q(\lambda))$



观察与几何解释

原始问题（Primal）：

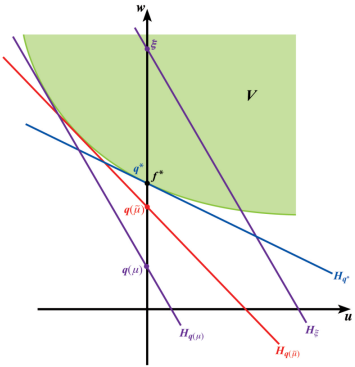
$$\min \quad w$$

$$\text{s.t.} \quad (0, w) \in V$$

对偶问题（Dual）：

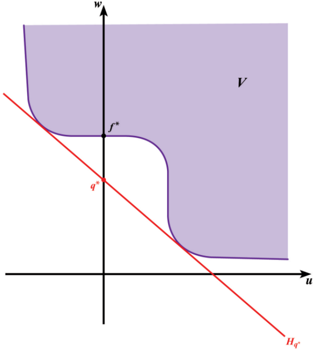
$$\max \quad q(\mu)$$

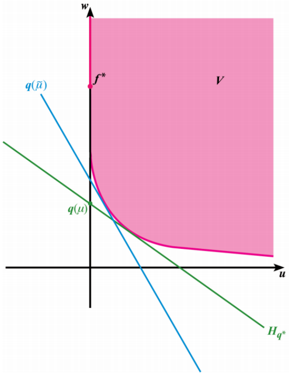
$$\text{s.t.} \quad \mu \succeq 0$$



- 对偶值 $q(\mu)$ 总是小于或等于任意满足 $(0, w) \in V$ 的 w，也即不超过 $f^\star$
- 对偶最优值 $q^\star$ 永远不超过原始最优值 $f^\star$： $q^\star \leq f^\star$ **（弱对偶）**
- 弱对偶可能是严格的，即 $q^\star < f^\star$，此时存在**对偶间隙（Duality Gap）**

对偶间隙示意图（Duality Gap Illustrations）





考虑以下凸问题：

$$\min \quad e^{-x}$$

$$\text{s.t} \quad x^2/y \le 0, \quad D = \{(x, y) \in \mathbb{R}^2 \mid y > 0\}$$

目录

Slater 约束资格条件（Slater's Constraint Qualification）

$$
\begin{aligned}
\min \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x) \le 0, \quad i = 1, \dots, m \\
& Ax = b
\end{aligned}
$$

对于上述凸优化问题，如果 **Slater 条件成立（该问题是严格可行的）**，即满足

$$
\exists\, x \in \mathrm{int}\, \mathcal{D} \quad \text{such that} \quad f_i(x) < 0, \forall i = 1, \dots, m, \quad \text{and } Ax = b
$$

那么该问题**强对偶成立**

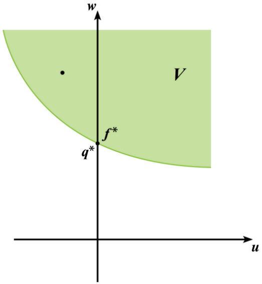
- 如果 $p^\star > -\infty$，那么上述条件同意保证**对偶最优解存在**
- 可进一步强化。例如，将 int $\mathcal{D}$ 替换为 relint $\mathcal{D}$，即相对内点，即线性不等式可不必严格满足
- 还有很多其他类型的约束资格条件也可保证强对偶

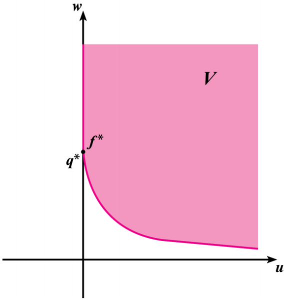
Slater 条件的图示证明（Proof）

考虑集合 $V \subset \mathbb{R}^m \times \mathbb{R}$，定义为：

$$V = \{(u, w) \mid f_i(x) \le u,\ f_0(x) \le w,\ x \in X\}$$

Slater 条件如下图所示：





不等式形式的线性规划（Inequality Form LP）

原始问题:

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax \leq b \end{aligned}$$

对偶函数:

$$g(\lambda) = \inf_x \left((c + A^T\lambda)^T x - b^T\lambda\right) = \begin{cases} -b^T\lambda, & \text{若} A^T\lambda + c = 0 \\ -\infty, & \text{否则} \end{cases}$$

对偶问题:

$$\begin{aligned} \max \quad & -b^T\lambda \\ \text{s.t.} \quad & A^T\lambda + c = 0, \quad \lambda \geq 0 \end{aligned}$$

- 根据 Slater 条件：若存在 $\tilde{x}$ 使得 $A\tilde{x} < b$，则 $p^\star = d^\star$
- 实际上，除非原始问题与对偶问题都不可行，一般都有 $p^\star = d^\star$

二次规划（Quadratic Program）

原始问题（Primal Problem）（假设 $P \in \mathbb{S}^n_{++}$）:

$$\begin{aligned} \min \quad & x^T P x \\ \text{s.t.} \quad & Ax \leq b \end{aligned}$$

对偶函数: $g(\lambda) = \inf_x \left(x^T P x + \lambda^T (Ax - b) \right) = -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T \lambda$

对偶问题:

$$\begin{aligned} \max \quad & -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T \lambda \\ \text{s.t.} \quad & \lambda \geq 0 \end{aligned}$$

- 根据 Slater 条件：若存在 $\tilde{x}$ 满足 $A\tilde{x} < b$，则 $p^\star = d^\star$
- 实际上，对于该问题，**总有** $p^\star = d^\star$