优化理论与算法

第六章

对偶理论

郭加熠|助理教授

目录

拉格朗日函数（Lagrangian）

标准形式问题（不要求是凸的）：

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

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 $\operatorname{dom} L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$，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 $f_i(x) \leq 0$ 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 $(\lambda, \nu) \in \mathrm{dom}(g)$ 且 $\lambda \geq 0$，那么 (λ, ν) 是对偶问题**可行解**
- 常常将隐含约束 $(\lambda, \nu) \in \mathrm{dom}(g)$ 显现写出

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

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：

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

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）：**

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

- **多个对偶最优解（Multiple Dual Optimal）：**

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

- **无对偶最优解（No Dual Optimal）：**

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

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

一个具有强对偶的非凸问题

$$\begin{aligned} \min \quad & x^T A x + 2b^T x \\ \text{s.t.} \quad & x^T x \le 1 \end{aligned}$$

由于 $A \not\succeq 0$，该问题**非凸**。**对偶函数（Dual Function）**：

$$g(\lambda) = \inf_x \left(x^T (A + \lambda I) x + 2b^T x - \lambda \right)$$

- 若 $A + \lambda I \not\succeq 0$ 或 $A + \lambda I \succeq 0$ 且 $b \notin \mathcal{R}(A + \lambda I)$，则 $g(\lambda) = -\infty$
- 否则，最小值由 $x = -(A + \lambda I)^\dagger b$ 达到，且 $g(\lambda) = -b^T (A + \lambda I)^\dagger b - \lambda$

对偶问题（Dual Problem）及等价 SDP 表达：

$$\begin{aligned} \max \quad & -b^T (A + \lambda I)^\dagger b - \lambda \\ \text{s.t.} \quad & A + \lambda I \succeq 0, \; b \in \mathcal{R}(A + \lambda I) \end{aligned} \qquad \begin{aligned} \max \quad & -t - \lambda \\ \text{s.t.} \quad & \begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \succeq 0 \end{aligned}$$

即使原始问题**不是凸的**，仍然可以实现**强对偶**（证明并不容易）。

目录

对偶与问题重构

- 一个问题的等价形式可能会导出非常不同的对偶问题；
- 当对偶问题难以推导或不具启发性时，重构原始问题可能会非常有用；

常见的重构方式

- 引入新变量和等式约束；
- 将显式约束转为隐式，或反之；
- 变换目标函数或约束函数，

  例如：将 $f_0(x)$ 替换为 $\phi(f_0(x))$，其中 ϕ 是凸且单调递增的函数。

表示问题（Representation Issue）

$$\begin{aligned} \text{minimize} \quad & -x_2 \\ \text{subject to} \quad & \|x\| \leq x_1 \\ & x \in D, \quad D = \{(x_1, x_2) \mid x_2 \geq 0\} \end{aligned}$$

- 放松不等式约束会得到一个对偶问题，其中 $d^* = -\infty$，而 $p^* = 0$。
- 但更仔细查看约束集合发现：$C = \{(x_1, x_2) \mid x_1 \geq 0,\, x_2 = 0\}$，写出等价问题

$$\begin{aligned} \text{minimize} \quad & -x_2 \\ \text{subject to} \quad & x_1 \geq 0, \quad x_2 = 0 \\ & x \in \mathbb{R}^2 \end{aligned}$$

这个问题是没有对偶间隙的！

对偶间隙与约束的“表示形式”（约束的建模方式）密切相关

引入新变量与等式约束

$$\text{minimize} \quad f_0(Ax+b)$$

- 对偶函数是常数：$g = \inf_x L(x) = \inf_x f_0(Ax+b) = p^*$
- 虽然满足强对偶性，但此时对偶问题几乎无用

重写后的问题及其对偶问题：

$$\begin{array}{ll} \text{minimize} & f_0(y) \\ \text{subject to} & Ax+b-y=0 \end{array} \qquad \begin{array}{ll} \text{maximize} & b^T\nu - f_0^*(\nu) \\ \text{subject to} & A^T\nu = 0 \end{array}$$

对偶函数推导如下:

$$g(\nu) = \inf_{x,y}\left(f_0(y) - \nu^T y + \nu^T Ax + b^T\nu\right) = \begin{cases} -f_0^*(\nu) + b^T\nu & \text{if } A^T\nu = 0 \\ -\infty & \text{Otherwise} \end{cases}$$

范数逼近问题的对偶问题

范数逼近问题： 最小化 $\|Ax-b\|$ 转成最小化 $\|Ax-b\|^2$

原问题：

$$\begin{aligned} \text{minimize} \quad & \|y\|^2 \\ \text{subject to} \quad & y = Ax - b \end{aligned}$$

对偶问题：

$$\begin{aligned} \text{maximize} \quad & b^T\nu - \frac{\|\nu\|^2}{4} \\ \text{subject to} \quad & A^T\nu = 0. \end{aligned}$$

推导：

$$g(\nu) = \inf_{x,y}\left(\|y\|^2 + \nu^T y - \nu^T Ax + b^T\nu\right)$$

$$= \begin{cases} b^T\nu + \inf_y(\|y\|^2 + \nu^T y) & \text{if } A^T\nu = 0 \\ -\infty & \text{o.w.} \end{cases} = \begin{cases} b^T\nu - \frac{\|\nu\|^2}{4} & \text{if } A^T\nu = 0 \\ -\infty & \text{o.w.} \end{cases}$$

隐式约束

带有盒约束（box constraints）的线性规划：原始问题与对偶问题

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & -1 \preceq x \preceq 1 \end{array} \qquad \begin{array}{ll} \text{maximize} & -b^T\nu - \mathbf{1}^T\lambda_1 - \mathbf{1}^T\lambda_2 \\ \text{subject to} & c + A^T\nu + \lambda_1 - \lambda_2 = 0 \\ & \lambda_1 \succeq 0, \ \lambda_2 \succeq 0 \end{array}$$

将盒约束隐式化后的重构问题：

$$\begin{array}{ll} \text{minimize} & f_0(x) = \begin{cases} c^T x & -1 \preceq x \preceq 1 \\ \infty & \text{其他情况} \end{cases} \\ \text{subject to} & Ax = b \end{array}$$

对偶函数：$g(\nu) = \inf_{-1 \preceq x \preceq 1} \left(c^T x + \nu^T (Ax - b) \right) = -b^T\nu - \|c + A^T\nu\|_1$

对偶问题： maximize $\quad -b^T\nu - \|c + A^T\nu\|_1$

目录

互补松弛性（Complementary Slackness）

假设强对偶成立，$x^\star$ 为原始问题最优解，$(\lambda^\star, \nu^\star)$ 为对偶问题最优解：

$$
\begin{aligned}
f_0(x^\star) = g(\lambda^\star, \nu^\star) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) \\
&\le f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star) \\
&\le f_0(x^\star)
\end{aligned}
$$

因此，两个不等式必须同时取等：

- $x^\star$ 使拉格朗日函数 $L(x, \lambda^\star, \nu^\star)$ 取最小值；
- 对于所有 $i = 1, \ldots, m$，有：$\lambda_i^\star f_i(x^\star) = 0$，称为**互补松弛性**。

$$
\lambda_i^\star > 0 \Rightarrow f_i(x^\star) = 0, \qquad f_i(x^\star) < 0 \Rightarrow \lambda_i^\star = 0
$$

Karush-Kuhn-Tucker（KKT）条件

以下四个条件统称为 KKT 条件（适用于 f_i, h_i 可微的优化问题）：

1. 原始约束（Primal constraints）： $f_i(x) \leq 0,\ i = 1, \ldots, m,$

   $h_i(x) = 0,\ i = 1, \ldots, p$

2. 对偶可行性（Dual constraints）： $\lambda \geq 0$

3. 互补松弛性（Complementary slackness）： $\lambda_i f_i(x) = 0,\ i = 1, \ldots, m$

4. 拉格朗日函数关于 x 的梯度为零（Stationarity）：

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$

若强对偶成立且 x, λ, ν 为最优解，则它们必须满足 KKT 条件。

凸问题的 KKT 条件（KKT Conditions for Convex Problem）

如果 $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ 满足一个凸优化问题的 KKT 条件，则它们是最优解：

- 根据互补松弛性：$f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
- 根据第 4 条条件和凸性：$g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

因此，有：

$$f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$$

若满足 Slater 条件，则有：x 是最优的当且仅当存在 λ, ν 满足 KKT 条件。

- 回顾：Slater 条件意味着强对偶成立，且对偶最优值是可达到的
- KKT 条件可视为无约束问题中 $\nabla f_0(x) = 0$ 的推广

示例：Water-filling 算法（假设 $\alpha_i > 0$）

$$\begin{aligned} \min \quad & -\sum_{i=1}^{n} \log(x_i + \alpha_i) \\ \text{s.t.} \quad & x \geq 0, \quad \mathbf{1}^T x = 1 \end{aligned}$$

x 为最优当且仅当 $x \geq 0$, $\mathbf{1}^T x = 1$，且存在 $\lambda \in \mathbb{R}^n, \nu \in \mathbb{R}$ 满足：

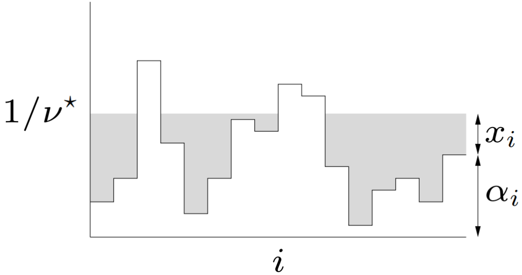
$$\lambda \geq 0, \quad \lambda_i x_i = 0, \quad \frac{1}{x_i + \alpha_i} + \lambda_i = \nu$$

- 若 $\nu < 1/\alpha_i$： $\lambda_i = 0$, $x_i = 1/\nu - \alpha_i$
- 若 $\nu \geq 1/\alpha_i$： $\lambda_i = \nu - 1/\alpha_i$, $x_i = 0$
- ν 由 $\mathbf{1}^T x = \sum_{i=1}^n \max\{0, \ 1/\nu - \alpha_i\} = 1$ 确定

示例：Water-filling 算法（假设 $\alpha_i > 0$）

解释（Interpretation）：满足 $\sum_{i=1}^{n} \max\{0,\ 1/\nu - \alpha_i\} = 1$ 关系的 ν

- 总共 n 个槽位，第 i 个槽底高度为 α_i
- 用单位水量“灌入”该区域
- 最终水面高度为 $1/\nu^\star$



目录

扰动与敏感性分析（Perturbation and Sensitivity Analysis）

未扰动问题：

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{array}$$

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

扰动问题：

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le u_i, \quad i = 1, \dots, m \\ & h_i(x) = v_i, \quad i = 1, \dots, p \end{array}$$

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) - u^T\lambda - v^T\nu \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- x 是原始变量，u, v 是扰动参数；
- $p^\star(u, v)$ 表示扰动参数下的最优值；
- 我们关注的是：能否从未扰动问题及其对偶解中，得到关于 $p^\star(u, v)$ 的有用信息。

全局敏感性分析结果（Global Sensitivity Result）

假设未扰动问题满足强对偶性，且 $\lambda^\star, \nu^\star$ 是其对偶最优解。应用弱对偶性到扰动问题有：

$$
\begin{aligned}
p^\star(u, v) &\geq g(\lambda^\star, \nu^\star) - u^T\lambda^\star - v^T\nu^\star \\
\text{(假设原问题强对偶成立)} &= p^\star(0, 0) - u^T\lambda^\star - v^T\nu^\star
\end{aligned}
$$

敏感性解释（Sensitivity Interpretation）：

- 若 $\lambda_i^\star$ 较大：当收紧约束 i（即 $u_i < 0$）时，$p^\star$ 将显著上升
  若 $\lambda_i^\star$ 较小：当放松约束 i（即 $u_i > 0$）时，$p^\star$ 不会显著下降
- 若 $\nu_i^\star$ 较大且为正：当取 $v_i < 0$ 时，$p^\star$ 显著上升；
  若 $\nu_i^\star$ 较大且为负：当取 $v_i > 0$ 时，$p^\star$ 显著上升
- 若 $\nu_i^\star$ 较小且为正：当取 $v_i > 0$ 时，$p^\star$ 不会显著下降；
  若 $\nu_i^\star$ 较小且为负：当取 $v_i < 0$ 时，$p^\star$ 不会显著下降

局部敏感性（Local Sensitivity）

如果 $p^\star(u,v)$ 在 $(0,0)$ 处可微，则有：

$$\lambda_i^\star = -\frac{\partial p^\star(0,0)}{\partial u_i}, \quad \nu_i^\star = -\frac{\partial p^\star(0,0)}{\partial v_i}$$

证明（以 $\lambda_i^\star$ 为例）来自全局敏感性结果：

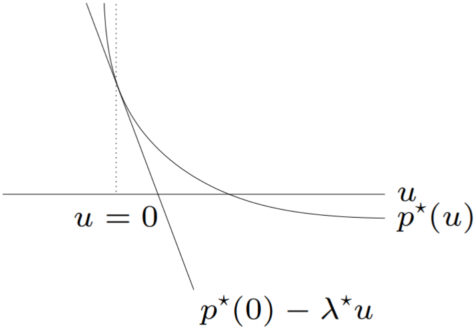
$$\frac{\partial p^\star(0,0)}{\partial u_i} = \lim_{t\searrow 0} \frac{p^\star(te_i,0) - p^\star(0,0)}{t} \geq -\lambda_i^\star$$

$$\frac{\partial p^\star(0,0)}{\partial u_i} = \lim_{t\nearrow 0} \frac{p^\star(te_i,0) - p^\star(0,0)}{t} \leq -\lambda_i^\star$$

因此，两者相等，得证。

局部敏感性（Local Sensitivity）

图示说明： $p^\star(u)$ 的函数图像（单个不等式约束）：



图中展示了：$p^\star(u)$ 与其切线 $p^\star(0) - \lambda^\star u$，体现 $\lambda^\star$ 的几何意义。

示例：利润最大化问题

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & a^T x = b \end{aligned}$$

- 该问题可以表示为一个公司在资源使用水平为 b 的情况下使利润 $-f(x)$ 最大的问题;
- 公司希望研究在资源水平从 b 变为 $b+u$ 时 ($u>0$ 使用更多资源，$u<0$ 使用更少资源), 最优利润会如何变化。

形式假设： 存在无间隙的最优原始解 $x^\star$ 和一个对偶解 $\lambda^\star \in \mathbb{R}$，且 f 是连续可微函数.

我们关注的是**因水平 b 的微小变化而导致的最优值 $f(x^\star)$ 的变化**。

示例：利润的敏感性分析

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & a^T x = b + u \end{aligned}$$

- 设 $x^\star(u)$ 为扰动问题的最优解；设 $\Delta x^\star = x^\star(u) - x^\star$ 为最优解的变化；设 $\Delta f^\star = f(x^\star(u)) - f(x^\star)$ 为最优值的变化；
- 根据 KKT 条件，有 $\nabla f(x^\star) = -\nu^\star a$；
- 使用泰勒一阶展开可得：$\Delta f^\star = \nabla f(x^\star)^T \Delta x^\star + o(\|\Delta x^\star\|) \approx -\nu^\star a^T \Delta x^\star$
- 由 $a^T x^\star = b$ 且 $a^T x^\star(u) = b + u$ 得：$a^T \Delta x^\star = u \Rightarrow \Delta f^\star \approx -\nu^\star u$

微小扰动误差说明

$$\frac{\text{最优利润的变化}}{\text{资源水平的变化}} = \frac{\Delta f^\star}{u} = -\nu^\star$$

注： 拉格朗日乘子 $\nu^\star$ 表示单位资源水平变化对最优利润的影响率。