第十章 最优性条件和对偶理论

第 35 讲 Lagrange 对偶

黄定江

DaSE @ ECNU
djhuang@dase.ecnu.edu.cn

1. 10.1 Lagrange 对偶函数
   - Lagrange
   - Lagrange 对偶函数
   - 最优值的下界
   - 通过线性逼近来理解
   - Lagrange 对偶函数和共轭函数
2. 10.2 常见的对偶函数
   - 线性方程组的最小二乘解
   - 标准形式的线性规划
   - 双向划分问题
   - 等式约束条件下的范数极小

10.1.1 Lagrange

考虑标准形式的优化问题 (不一定是凸的)

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

自变量 $x \in R^n$，**定义域** $\mathcal{D} = \bigcap_{i=1}^{m} \mathbf{dom}\, f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\, h_i$，**最优值** p^*

10.1.1 Lagrange

Lagrange 函数：$L : \mathcal{D} \times R^m \times R^p \to R$ 为

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

其中定义域为 $\mathbf{dom}\, L = \mathcal{D} \times R^m \times R^p$

- Lagrange 对偶的基本思想：添加约束条件的加权和，得到增广的目标函数
- λ_i 称为第 i 个不等式约束 $f_i(x) \le 0$ 对应的 **Lagrange 乘子**
- ν_i 称为第 i 个等式约束 $h_i(x) = 0$ 对应的 **Lagrange 乘子**
- 向量 λ 和 ν 称为对偶变量或者问题 (1) 的 **Lagrange 乘子向量**

10.1.2 Lagrange 对偶函数

Lagrange 对偶函数 (或对偶函数) $g: R^m \times R^p \to R$ 为 Lagrange 函数关于 x 取得的最小值：即对 $\lambda \in R^m, \nu \in R^p$，有

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

- 函数关于 x 无下界，则对偶函数取值为 $-\infty$
- 对偶函数是一族关于 (λ, ν) 的仿射函数的逐点下确界，所以即使原问题 (1) 非凸，对偶函数也是凹函数

10.1.3 最优值的下界

对偶问题构成了原问题 (1) 最优值 p^* 的下界：即对任意 $\lambda \geq 0$ 和 ν，下式成立

$$g(\lambda, \nu) \leq p^* \tag{2}$$

证明.

设 $\tilde{x}$ 是原问题 (1) 的一个可行点，即 $f_i(\tilde{x}) \leq 0$ 且 $h_i(\tilde{x} = 0)$。根据假设，$\lambda \geq 0$，有

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

左边第一项非正，第二项为零 □

10.1.3 最优值的下界

Proof. Cont.

根据上述不等式，有

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

因此

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

由于每一个可行点 $\tilde{x}$ **都满足** $g(\lambda, \nu) \leq f_0(\bar{x})$**，因此** $g(\lambda, \nu) \leq p^*$ **成立** □

10.1.3 最优值的下界

- 针对 $x \in R$ 和具有一个不等式约束的某简单问题，图 (1) 描述了式 (2) 所给出的下界
- 虽然不等式 (2) 成立，但当 $g(\lambda, v) = -\infty$ 时其意义不大
- 只有当 $\lambda \geq 0$ 且 $((\lambda, v)) \in \mathbf{dom}\, g$ 时，$g(\lambda, v)$ 是对偶可行的

10.1.3 最优值的下界

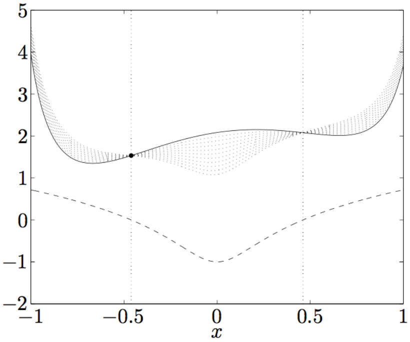

图 1: **对偶可行点给出的下界**。实线表示目标函数 f_0，虚线表示约束函数 f_1。可行集 $[-0.46, 0.46]$，如图中两条垂直点线所示。最优点 $x^* = -0.46$，最优值 $p^* = 1.54$。点线表示一系列 Lagrange 函数 $L(x, \lambda)$，其中 $\lambda = 0.1, 0.2, \cdots, 1.0$。每个 Lagrange 函数都有一个极小值，均小于原问题最优目标值 p^*，这是因为在可行集上有 $L(x, \lambda) \leq f_0(x)$。

10.1.3 最优值的下界

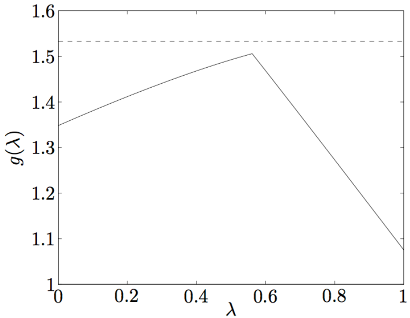

图 2: **图1中问题的对偶函数** g。函数 f_0 和 f_1 都不是凸函数，但是对偶函数是凹函数。水平虚线是原问题的最优函数值 p^*。

10.1.4 通过线性逼近来理解

通过对集合 $\{0\}$ 和 $-R_+$ 的示性函数进行线性逼近来理解 Lagrange 函数和其给出下界的性质。将原问题 (1) 重新描述为一个无约束问题：

$$\text{minimize} \quad f_0(x) + \sum_{i=1}^{m} I\_(f_i(x)) + \sum_{i=1}^{p} I_0(h_i(x)) \tag{3}$$

其中，$I\_ : R \to R$ 是非正实数集的示性函数

$$I\_(u) = \begin{cases} 0 & u \leq 0 \\ \infty & u > 0 \end{cases}$$

- 在表达式 (3) 中，函数 $I\_(u)$ 可以理解为对约束函数值 $u = f_i(x)$ 的一种恼怒或不满：如果 $f_i(x) \leq 0, I\_(u)$ 为零，如果 $f_i(x) > 0$，$I\_(u)$ 为 ∞
- 可以认为函数 $I\_$ 是一个“砖墙式”或“无限强硬”的不满意方程；即随着函数 $f_i(x)$ 从非正数变为正数，不满意度从 0 升到无穷大

10.1.4 通过线性逼近来理解

设在表达式 (3) 中，用线性函数 $\lambda_i u$ 替代函数 $I\_(u)$，其中 $\lambda_i \geq 0$，用函数 $v_i u$ 替代 $I_0(u)$。则目标函数变为 Lagrange 函数 $L(x, \lambda, v)$，且对偶函数值 $g(\lambda, v)$ 是问题 (4) 的最优值

$$\text{minimize} \quad L(x, \lambda, v) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} v_i h_i(x) \tag{4}$$

- 表达式 (4) 中，用线性或者“软”的不满意函数替换了函数 $I\_$ 和 I_0，如果 $f_i(x) = 0$，不满意度为 0，当 $f_i(x) > 0$ 时，不满意度大于 0（假设 $\lambda_i > 0$），随着约束“越来越被违背”，我们越来越不满意
- 用线性函数 $\lambda_i u$ 去逼近 $I\_(u)$ 远远不够，但线性函数可看成是示性函数的一个下估计，因为对任意 u，有 $\lambda_i u \leq I\_(u)$ 和 $v_i u \leq I_0(u)$，随之可得到，对偶函数是原问题最有函数值的一个下界

10.1.5 Lagrange 对偶函数和共轭函数

回忆函数 $f : R^n \to R$ 的共轭函数 f^* 为

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} \left(y^{\mathrm{T}} x - f(x)\right)$$

考虑问题

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax \leq b, \quad Cx = d \end{array} \tag{5}$$

利用函数 f_0 的共轭函数，其对偶函数为

$$\begin{aligned} g(\lambda, \nu) &= \inf_x \left(f_0(x) + \lambda^{\mathrm{T}}(Ax - b) + \nu^{\mathrm{T}}(Cx - d)\right) \\ &= -b^{\mathrm{T}}\lambda - d^{\mathrm{T}}\nu + \inf_x \left(f_0(x) + (A^{\mathrm{T}}\lambda + C^{\mathrm{T}}\nu)^{\mathrm{T}} x\right) \\ &= -b^{\mathrm{T}}\lambda - d^{\mathrm{T}}\nu - f_0^*(-A^{\mathrm{T}}\lambda - C^{\mathrm{T}}\nu) \end{aligned} \tag{6}$$

函数 g 的定义域 $\mathbf{dom}\, g = \{(\lambda, v) | -A^{\mathrm{T}}\lambda - C^{\mathrm{T}}v \in \mathbf{dom}\, f_0^*\}$

等式约束条件下的范数极小化

例 1

考虑问题

$$\begin{aligned} \textit{minimize} \quad & \|x\| \\ \textit{subject to} \quad & Ax = b \end{aligned} \tag{7}$$

其中 $\|\cdot\|$ 是任意范数。求其对偶函数

等式约束条件下的范数极小化

$f_0 = \|\cdot\|$ 的共轭函数为

$$f_0^*(y) = \begin{cases} 0 & \|y\|_* \le 1 \\ \infty & \text{其他情况} \end{cases}$$

可以看出此函数是对偶范数单位球的示性函数。利用结论 (6) 得到问题 (7) 的对偶函数为

$$g(v) = -b^{\mathrm{T}}v - f_0^*(-A^{\mathrm{T}}v) = \begin{cases} -b^{\mathrm{T}}v & \|A^{\mathrm{T}}v\|_* \le 1 \\ -\infty & \text{其他情况} \end{cases}$$

熵的最大化

例 2

考虑问题

$$\begin{array}{ll} \textit{minimize} & f_0(x) = \sum_{i=1}^{n} x_i \log x_i \\ \textit{subject to} & Ax \leq b, \quad \mathbf{1}^{\mathrm{T}} x = 1 \end{array} \tag{8}$$

其中 $\mathbf{dom}\, f_0 = R_{++}^n$。求其对偶函数

熵的最大化

关于实变量 u 的负熵函数 $u\log u$ 的共轭函数是 e^{v-1}。由于函数 f_0 是不同变量的负熵函数的和，其共轭函数为

$$f_0^*(y) = \sum_{i=1}^{n} e^{y_i-1}$$

其定义域为 $\mathbf{dom}\, f_0^* = R^n$。根据前述结论 (6)，问题 (8) 的对偶函数为

$$g(\lambda, v) = -b^{\mathrm{T}}\lambda - v - \sum_{i=1}^{n} e^{-a_i^{\mathrm{T}}\lambda - v - 1} = -b^{\mathrm{T}}\lambda - v - e^{-v-1}\sum_{i=1}^{n} e^{-a_i^{\mathrm{T}}\lambda}$$

其中 a_i 是矩阵 A 的第 i 列向量

10.2.1 线性方程组的最小二乘解

例 3

考虑问题

$$\begin{aligned} \textit{minimize} \quad & x^{\mathrm{T}} x \\ \textit{subject to} \quad & Ax = b \end{aligned} \tag{9}$$

其中 $A \in R^{p \times n}$。**求其对偶函数**

10.2.1 线性方程组的最小二乘解

此问题无不等式约束，有 p 个 (线性) 等式约束，其 Lagrange 函数为

$$L(x,\nu) = x^{\mathrm{T}}x + \nu^{\mathrm{T}}(Ax - b)$$

定义域为 $R^n \times R^p$。对偶函数为

$$g(\nu) = \inf_x L(x,\nu)$$

10.2.1 线性方程组的最小二乘解

因为 $L(x, v)$ 是 x 的二次凸函数，可通过求解最优性条件得到函数的最小值

- L 关于 x 取极小值，梯度为零：

$$\nabla_x L(x,\nu) = 2x + A^{\mathrm{T}}\nu = 0 \implies x = -(1/2)A^{\mathrm{T}}\nu$$

- 代入到 L 得到函数 g：

$$g(\nu) = L((-1/2)A^{\mathrm{T}}\nu, \nu) = -(1/4)\nu^{\mathrm{T}}AA^{\mathrm{T}}\nu - b^{\mathrm{T}}\nu$$

是关于 ν 的二次凹函数，定义域为 R^p

最优值下界：根据性质 (2)，对 $\forall \nu \in R^p$

$$p^* \geq -(1/4)\nu^{\mathrm{T}}AA^{\mathrm{T}}\nu - b^{\mathrm{T}}\nu$$

10.2.2 标准形式的线性规划

例 4

考虑问题

$$\begin{aligned} \textit{minimize} \quad & c^{\mathrm{T}} x \\ \textit{subject to} \quad & Ax = b, \quad x \geq 0 \end{aligned} \tag{10}$$

其中，不等式约束函数为 $f_i(x) = -x_i, i = 1, \cdots, n$。求其对偶函数

10.2.2 标准形式的线性规划

为了推导 Lagrange 函数，对 n 个不等式约束引入 Lagrange 乘子 λ_i，对等式约束引入 Lagrange 乘子 ν_i，得到

$$
\begin{aligned}
L(x, \lambda, \nu) &= c^{\mathrm{T}} x - \sum_{i=1}^{n} \lambda_i x_i + \nu^{\mathrm{T}}(Ax - b) \\
&= -b^{\mathrm{T}} \nu + (c + A^{\mathrm{T}} \nu - \lambda)^{\mathrm{T}} x
\end{aligned}
$$

L 关于 x 线性，因此对偶函数为

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

g 在仿射域 $\{(\lambda, \nu) | A^{\mathrm{T}} \nu - \lambda + c = 0\}$ 上是线性函数，因此总体是凹函数
最优值下界：如果$\mathrm{A}^{\mathrm{T}} \nu + c \geq 0$，$p^* \geq -b^{\mathrm{T}} \nu$

10.2.3 双向划分问题

例 5

考虑问题

$$\begin{array}{ll} \textit{minimize} & x^{\mathrm{T}} W x \\ \textit{subject to} & x_i^2 = 1, \quad i = 1, \cdots, n \end{array} \tag{11}$$

其中，$W \in S^n$。求其对偶函数

10.2.3 双向划分问题

- 这是一个非凸问题；可行集包含 2^n 个离散点
- 解释：可将 $\{1,\cdots,n\}$ 划分为两个集合：

$$\{1,\cdots,n\}=\{i\,|x_i=-1\}\cup\{i\,|x_i=1\}$$

矩阵系数 W_{ij} 可以看成分量 i,j 在同一分区的成本，$-W_{ij}$ 可以看成分量 i,j 在不同分区内的成本。原问题中的目标函数是考虑分量间所有配件的成本，即寻找使得总成本最小的划分

10.2.3 双向划分问题

Lagrange 函数：

$$
\begin{aligned}
L(x,\nu) &= x^{\mathrm{T}} W x + \sum_{i=1}^{n} \nu_i (x_i^2 - 1) \\
&= x^{\mathrm{T}} (W + \mathrm{diag}(\nu)) x - \mathbf{1}^{\mathrm{T}} \nu
\end{aligned}
$$

对偶函数:

$$
\begin{aligned}
g(\nu) = \inf_x (x^{\mathrm{T}} W x + \sum_i \nu_i (x_i^2 - 1)) &= \inf_x x^{\mathrm{T}} (W + \mathrm{diag}(\nu)) x - \mathbf{1}^{\mathrm{T}} \nu \\
&= \begin{cases} -\mathbf{1}^{\mathrm{T}} \nu & W + \mathrm{diag}(\nu) \succeq 0 \\ -\infty & \text{其他情况} \end{cases}
\end{aligned}
$$

最优值下界：如果 $W + \mathrm{diag}(\nu) \succeq 0$，$p^* \geq -\mathbf{1}^{\mathrm{T}} \nu = n\lambda_{\min}(W)$

10.2.4 等式约束条件下的范数极小化

例 6

考虑问题

$$\begin{aligned} \textit{minimize} \quad & \|x\| \\ \textit{subject to} \quad & Ax = b \end{aligned} \tag{12}$$

求其对偶函数

10.2.4 等式约束条件下的范数极小化

对偶函数：

$$g(\nu) = \inf_x(\|x\| - \nu^{\mathrm{T}}Ax + b^{\mathrm{T}}\nu) = \begin{cases} b^{\mathrm{T}}\nu & \|A^{\mathrm{T}}\nu\|_* \leq 1 \\ -\infty & \text{其他情况} \end{cases}$$

其中 $\|v\|_* = \sup_{\|u\| \leq 1} u^{\mathrm{T}}v$ **是** $\|\cdot\|$ **的对偶范数。**

最优值下界： **如果** $\|A^{\mathrm{T}}\nu\|_* \leq 1$， $p^* \geq b^{\mathrm{T}}\nu$

10.2.4 等式约束条件下的范数极小化

证明.

对全部的 x，当 $\|y\|_* \le 1$，$\inf_x(\|x\| - y^{\mathrm{T}} x \ge 0)$，否则为 $-\infty$

- 如果 $\|y\|_* \le 1$，那么对所有 x，$\|x\| - y^{\mathrm{T}} x \ge 0$，当 $x = 0$ 时取等
- 如果 $\|y\|_* > 1$，令 $x = tu$ 其中 $\|u\| \le 1, u^{\mathrm{T}} y = \|y\|_* > 1$：

$$\|x\| - y^{\mathrm{T}} x = t(\|u\| - \|y\|_*) \to -\infty \quad \text{当} t \to \infty$$

□