一、拉格朗日对偶函数

1、定义

□ The Lagrangian $L: \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \mapsto \mathbf{R}$

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

- $\operatorname{dom} L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$

> 向量λ，v是对偶变量

□ $g: \mathbf{R}^m \times \mathbf{R}^p \mapsto \mathbf{R}$

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

- 不论原问题凸不凸，g都是凹函数，且无约束：逐点下确界

> 很容易构造出这样的对偶函数g

2、下界

- 对偶函数用于估计下界

例1：线性方程组的最小二乘

考虑问题

$$\begin{array}{ll} \text{minimize} & x^T x \\ \text{subject to} & Ax = b, \end{array} \tag{5.5}$$

其中 $A \in \mathbf{R}^{p\times n}$。这个问题没有不等式约束，有 p 个（线性）等式约束。其 Lagrange 函数是 $L(x,\nu) = x^T x + \nu^T(Ax-b)$，定义域为 $\mathbf{R}^n \times \mathbf{R}^p$。对偶函数是 $g(\nu) = \inf_x L(x,\nu)$。因为 $L(x,\nu)$ 是 x 的二次凸函数，可以通过求解如下最优性条件得到函数的最小值，

$$\nabla_x L(x,\nu) = 2x + A^T\nu = 0,$$

在点 $x = -(1/2)A^T\nu$ 处 Lagrange 函数达到最小值。因此对偶函数为

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

它是一个二次凹函数，定义域为 $\mathbf{R}^p$。根据对偶函数给出原问题下界的性质 (5.2)，对任意 $\nu \in \mathbf{R}^p$，有

$$-(1/4)\nu^T A A^T \nu - b^T \nu \leqslant \inf\{x^T x \mid Ax = b\}.$$

其相应的对偶问题为

$$\text{maximize} \quad -(1/4)\nu^T AA^T\nu - b^T\nu,$$

它是一个凹二次函数的无约束极大化问题。

Slater 条件此时即是原问题的可行性条件，所以如果 $b \in \mathcal{R}(A)$，即 $p^\star < \infty$，有 $p^\star = d^\star$。事实上，对于此问题，强对偶性通常成立，即使 $p^\star = \infty$ 亦如此。当 $p^\star = \infty$ 时，$b \notin \mathcal{R}(A)$，所以存在 z 使得 $A^Tz = 0$，$b^Tz \neq 0$。因此，对偶函数在直线 $\{tz \mid t \in \mathbf{R}\}$ 上无界，即对偶问题最优值也无界，$d^\star = \infty$。

例2：标准线性规划

$$\begin{array}{ll} \text{minimize} & c^Tx \\ \text{subject to} & Ax = b \\ & x \succeq 0, \end{array}$$

其中，不等式约束函数为 $f_i(x) = -x_i$，$i = 1, \cdots, n$。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，我们得到

$$L(x,\lambda,\nu) = c^Tx - \sum_{i=1}^n \lambda_i x_i + \nu^T(Ax-b) = -b^T\nu + (c + A^T\nu - \lambda)^T x.$$

对偶函数为

$$g(\lambda,\nu) = \inf_x L(x,\lambda,\nu) = -b^T\nu + \inf_x (c + A^T\nu - \lambda)^T x,$$

可以很容易确定对偶函数的解析表达式，因为线性函数只有恒为零时才有下界。因此，当 $c + A^T\nu - \lambda = 0$ 时，$g(\lambda,\nu) = -b^T\nu$，其余情况下 $g(\lambda,\nu) = -\infty$，即

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

注意到对偶函数 g 只有在 $\mathbf{R}^m \times \mathbf{R}^p$ 上的一个正常仿射子集上才是有限值。后面我们将会看到这是一种常见的情况。

只有当 λ 和 ν 满足 $\lambda \succeq 0$ 和 $A^T\nu - \lambda + c = 0$ 时，下界性质 (5.2) 才是非平凡的。在此情形下，$-b^T\nu$ 给出了线性规划问题 (5.6) 最优值的一个下界。

- 对偶问题

$$\begin{array}{ll} \text{maximize} & g(\lambda,\nu) = \begin{cases} -b^T\nu & A^T\nu - \lambda + c = 0 \\ -\infty & \text{其他情况} \end{cases} \\ \text{subject to} & \lambda \succeq 0. \end{array}$$

- 最简对偶形式

$$\begin{array}{ll} \text{maximize} & -b^T\nu \\ \text{subject to} & A^T\nu + c \succeq 0. \end{array}$$

3、共轭转化

模板1

$$= \inf_x \left(f(x)+\nu^T x\right) = -\sup_x \left((-\nu)^T x - f(x)\right) = -f^*(-\nu)$$

模板2

更一般地（也更有用地），考虑一个优化问题，其具有线性不等式以及等式约束，

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax \preceq b \\ & Cx = d. \end{array} \tag{5.10}$$

利用函数 f_0 的共轭函数，我们可以将问题 (5.10) 的对偶函数表述为

$$\begin{aligned} g(\lambda,\nu) &= \inf_x \left(f_0(x)+\lambda^T(Ax-b)+\nu^T(Cx-d)\right) \\ &= -b^T\lambda - d^T\nu + \inf_x \left(f_0(x)+(A^T\lambda+C^T\nu)^T x\right) \\ &= -b^T\lambda - d^T\nu - f_0^*(-A^T\lambda - C^T\nu). \end{aligned} \tag{5.11}$$

函数 g 的定义域也可以由函数 f_0^* 的定义域得到，

$$\mathbf{dom}\, g = \{(\lambda,\nu) \mid -A^T\lambda - C^T\nu \in \mathbf{dom}\, f_0^*\}.$$

例3：min范数--等式约束

> 范数的共轭函数是对偶范数单位球的示性函数

$$\begin{array}{ll} \min & \|x\| \\ \text{s.t.} & Ax = b \end{array}$$

- Conjugate of $f_0 = \|\cdot\|$

$$f_0^*(y) = \begin{cases} 0 & \|y\|_* \le 1, \\ \infty & \text{otherwise.} \end{cases}$$

- The Dual Function

$$g(\nu) = -b^\top\nu - f_0^*(-A^\top\nu) = \begin{cases} -b^\top\nu & \|A^\top\nu\|_* \le 1, \\ -\infty & \text{otherwise.} \end{cases}$$

例4：max熵

$$\begin{aligned} \min \quad & f_0(x) = \sum_{i=1}^{n} x_i \log x_i \\ \text{s.t.} \quad & Ax \preccurlyeq b \\ & \mathbf{1}^\top x = 1 \end{aligned}$$

□ Conjugate of f_0

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

□ The Dual Function

$$\begin{aligned} g(\lambda, \nu) &= -b^\top \lambda - \nu - f_0^*(-A^\top \lambda - \nu \mathbf{1}) \\ &= -b^\top \lambda - \nu - \textstyle\sum_{i=1}^{n} e^{-a_i^\top \lambda - \nu - 1} \\ &= -b^\top \lambda - \nu - e^{-\nu - 1} \textstyle\sum_{i=1}^{n} e^{-a_i^\top \lambda} \end{aligned}$$

4、拉格朗日对偶问题

题是：从 Lagrange 函数能够得到的最好下界是什么？

可以将这个问题表述为优化问题

$$\begin{aligned} \text{maximize} \quad & g(\lambda, \nu) \\ \text{subject to} \quad & \lambda \succeq 0. \end{aligned} \tag{5.16}$$

上述问题称为问题 (5.1) 的**Lagrange 对偶问题**。在本书中，原始问题 (5.1) 有时被称为**原问题**。前面提到的**对偶可行**的概念，即描述满足 $\lambda \succeq 0$ 和 $g(\lambda, \nu) > -\infty$ 的一组 (λ, ν)，此时具有意义。它意味着，这样的一组 (λ, ν) 是对偶问题 (5.16) 的一个可行解。称解 $(\lambda^\star, \nu^\star)$ 是**对偶最优解**或者是**最优 Lagrange 乘子**，如果它是对偶问题 (5.16) 的最优解。

Lagrange 对偶问题 (5.16) 是一个凸优化问题，这是因为极大化的目标函数是凹函数，且约束集合是凸集。因此，对偶问题的凸性和原问题 (5.1) 是否是凸优化问题无关。

例5：二次约束二次规划

$$\begin{aligned} \text{minimize} \quad & (1/2)x^T P_0 x + q_0^T x + r_0 \\ \text{subject to} \quad & (1/2)x^T P_i x + q_i^T x + r_i \leqslant 0, \quad i = 1, \cdots, m, \end{aligned} \tag{5.28}$$

其中，$P_0 \in \mathbf{S}_{++}^n$，$P_i \in \mathbf{S}_+^n$，$i = 1, \cdots, m$。其 Lagrange 函数为

$$L(x, \lambda) = (1/2)x^T P(\lambda) x + q(\lambda)^T x + r(\lambda),$$

其中

$$P(\lambda) = P_0 + \sum_{i=1}^{m} \lambda_i P_i, \qquad q(\lambda) = q_0 + \sum_{i=1}^{m} \lambda_i q_i, \qquad r(\lambda) = r_0 + \sum_{i=1}^{m} \lambda_i r_i.$$

对于一般的 λ 可以得到对偶函数 $g(\lambda)$ 的表达式，但是这样的推导非常复杂。但是，如果 $\lambda \succeq 0$，我们有 $P(\lambda) \succ 0$ 以及

$$g(\lambda) = \inf_x L(x, \lambda) = -(1/2)q(\lambda)^T P(\lambda)^{-1} q(\lambda) + r(\lambda).$$

因此，对偶问题可以表述为

$$\begin{array}{ll} \text{maximize} & -(1/2)q(\lambda)^T P(\lambda)^{-1} q(\lambda) + r(\lambda) \\ \text{subject to} & \lambda \succeq 0. \end{array} \tag{5.29}$$

根据 Slater 条件，当二次不等式约束严格成立时，即存在一点 x 使得

$$(1/2)x^T P_i x + q_i^T x + r_i < 0, \quad i = 1, \cdots, m.$$

问题 (5.29) 和 (5.28) 之间强对偶性成立。

5、强弱对偶性

弱对偶性

Lagrange 对偶问题的最优值，我们用 $d^\star$ 表示，根据定义，这是通过 Lagrange 函数得到的原问题最优值 $p^\star$ 的最好下界。特别地，我们有下面简单但是非常重要的不等式

$$d^\star \leqslant p^\star, \tag{5.23}$$

即使原问题不是凸问题，上述不等式亦成立。这个性质称为弱对偶性。

即使当 $d^\star$ 和 $p^\star$ 无限时，弱对偶性不等式 (5.23) 也成立。例如，如果原问题无下界，即 $p^\star = -\infty$，为了保证弱对偶性成立，必须有 $d^\star = -\infty$，即 Lagrange 对偶问题不可行。反过来，若对偶问题无上界，即 $d^\star = \infty$，为了保证弱对偶性成立，必须有 $p^\star = \infty$，即原问题不可行。

强对偶性

> 对偶间隙为0

如果等式

$$d^\star = p^\star \tag{5.24}$$

成立，即最优对偶间隙为零，那么强对偶性成立。这说明从 Lagrange 对偶函数得到的最好下界是紧的。

对于一般情况，强对偶性不成立。但是，如果原问题 (5.1) 是凸问题，即可以表述为如下形式

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \leqslant 0, & i = 1, \cdots, m, \\ & Ax = b, & \end{array} \tag{5.25}$$

其中，函数 $f_0, \cdots, f_m$ 是凸函数，强对偶性通常（但不总是）成立。有很多研究成果给

Slater约束准则

> 强对偶性成立条件：满足所有不等约束（不能=0）和等式约束的点

一个简单的约束准则是 Slater 条件：存在一点 $x \in \textbf{relint}\,\mathcal{D}$ 使得下式成立

$$f_i(x) < 0, \quad i = 1, \cdots, m, \qquad Ax = b. \tag{5.26}$$

满足上述条件的点有时称为**严格可行**，这是因为不等式约束严格成立。Slater 定理说明，当 Slater 条件成立（且原问题是凸问题）时，强对偶性成立。

当不等式约束函数 f_i 中有一些是仿射函数时，Slater 条件可以进一步改进。如果最前面的 k 个约束函数 $f_1, \cdots, f_k$ 是仿射的，则若下列弱化的条件成立，强对偶性成立。该条件为：存在一点 $x \in \textbf{relint}\,\mathcal{D}$ 使得

$$f_i(x) \leqslant 0, \quad i = 1, \cdots, k, \qquad f_i(x) < 0, \quad i = k+1, \cdots, m, \qquad Ax = b. \tag{5.27}$$

换言之，仿射不等式不需要严格成立。注意到当所有约束条件都是线性等式或不等式且 $\textbf{dom}\, f_0$ 是开集时，改进的 Slater 条件 (5.27) 就是可行性条件。

强弱对偶性的最值特征

> 左边是g

- 弱

> 先求最小再求最大，更小

$$\sup_{\lambda \geqslant 0} \inf_x L(x, \lambda) \leq \inf_x \sup_{\lambda \geqslant 0} L(x, \lambda)$$

- 强

$$\sup_{\lambda \geqslant 0} \inf_x L(x, \lambda) = \inf_x \sup_{\lambda \geqslant 0} L(x, \lambda)$$

- 普遍形式

☐ Max-min Inequality

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) \leq \inf_{w \in W} \sup_{z \in Z} f(w, z)$$

- For any $f: \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ and any $W \subseteq \mathbf{R}^n, Z \subseteq \mathbf{R}^m$

☐ Strong Max-min Property

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) = \inf_{w \in W} \sup_{z \in Z} f(w, z)$$

- Hold only in special cases

6、鞍点

二、最优性条件

1、次优解认证和终止准则

2、互补松弛性

3、KKT最优性条件

若某个凸优化问题具有可微的目标函数和约束函数，且其满足 Slater 条件，那么 KKT 条件是最优性的充要条件：Slater 条件意味着最优对偶间隙为零且对偶最优解可以达到，因此 x 是原问题最优解，当且仅当存在 (λ, ν)，二者满足 KKT 条件。

> 若目标和约束函数可微：原问题和对偶问题最优解一定满足KKT条件
>
> 解很多凸优化问题就是解KKT

- 先Slater验证后才能用
- 非凸问题

和前面一样，令 $x^\star$ 和 $(\lambda^\star, \nu^\star)$ 分别是原问题和对偶问题的某对最优解，对偶间隙为零。因为 $L(x, \lambda^\star, \nu^\star)$ 关于 x 求极小在 $x^\star$ 处取得最小值，因此函数在 $x^\star$ 处的导数必须为零，即，

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

因此，我们有

$$\begin{aligned} f_i(x^\star) &\leqslant 0, \quad i = 1, \cdots, m \\ h_i(x^\star) &= 0, \quad i = 1, \cdots, p \\ \lambda_i^\star &\geqslant 0, \quad i = 1, \cdots, m \\ \lambda_i^\star f_i(x^\star) &= 0, \quad i = 1, \cdots, m \\ \nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^{p} \nu_i^\star \nabla h_i(x^\star) &= 0, \end{aligned} \tag{5.49}$$

- 凸问题

当原问题是凸问题时，满足 KKT 条件的点也是原、对偶最优解。换言之，如果函数 f_i 是凸函数，h_i 是仿射函数，$\tilde{x}$，$\tilde{\lambda}$，$\tilde{\nu}$ 是任意满足 KKT 条件的点，

$$\begin{aligned} f_i(\tilde{x}) &\leqslant 0, \quad i = 1, \cdots, m \\ h_i(\tilde{x}) &= 0, \quad i = 1, \cdots, p \\ \tilde{\lambda}_i &\geqslant 0, \quad i = 1, \cdots, m \\ \tilde{\lambda}_i f_i(\tilde{x}) &= 0, \quad i = 1, \cdots, m \\ \nabla f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\nu}_i \nabla h_i(\tilde{x}) &= 0, \end{aligned}$$

· 236 · 5 对偶

那么 $\tilde{x}$ 和 $(\tilde{\lambda}, \tilde{\nu})$ 分别是原问题和对偶问题的最优解，对偶间隙为零。

例6：二次凸--等式约束

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Ax = b, \end{array} \tag{5.50}$$

其中 $P \in \mathbf{S}_+^n$。此问题的 KKT 条件为

$$Ax^\star = b, \qquad Px^\star + q + A^T \nu^\star = 0,$$

我们可以将其写成

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ \nu^\star \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}.$$

求解变量 $x^\star$，$\nu^\star$ 的 $m+n$ 个方程，其中变量的维数为 $m+n$，可以得到优化问题 (5.50) 的最优原变量和对偶变量。

4、通过对偶问题求解原问题