

第十一章 对偶理论

第 33 讲 Lagrange 对偶问题

黄定江

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

1 33.1 对偶问题与对偶约束

2 33.2 对偶解释

3 33.3 最优性条件

1 33.1 对偶问题与对偶约束

2 33.2 对偶解释

3 33.3 最优性条件

33.1.1 显式表达对偶约束

Lagrange 对偶问题

$$\begin{array}{ll}\text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \geq 0\end{array}\quad (1)$$

- 从 Lagrange 函数能够得到的**最好**下界
- 凸优化问题；最优点记为 d^* ；如果 $\lambda \geq 0, (\lambda, \nu) \in \text{dom } g$ ，则 (λ, ν) 是对偶可行的，称 (λ^*, ν^*) 是对偶最优解或者是最优 Lagrange 乘子
- 经常，通过使隐含约束变为显式约束，问题能够得到简化

标准形式线性规划的 Lagrange 对偶

例 1

考虑问题

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \quad x \geq 0 \end{array} \quad (2)$$

给出其对偶问题

标准形式线性规划的 Lagrange 对偶

对偶函数：

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

- 其对偶问题是在满足约束 $\lambda \geq 0$ 的条件下极大化对偶函数 g

对偶问题：

$$\begin{aligned} & \text{maximize} && g(\lambda, v) = \begin{cases} -b^T v & A^T v - \lambda + c = 0 \\ -\infty & \text{其他情况} \end{cases} \\ & \text{subject to} && \lambda \geq 0 \end{aligned} \tag{3}$$

标准形式线性规划的 Lagrange 对偶

- 当且仅当 $A^T v - \lambda + c = 0$ 时对偶函数 g 有界, 通过隐式约束变为显式约束, 得到等价问题

$$\begin{aligned} & \text{maximize} && -b^T v \\ & \text{subject to} && A^T v - \lambda + c = 0, \quad \lambda \geq 0 \end{aligned} \tag{4}$$

进一步, 此问题可以表述为

$$\begin{aligned} & \text{maximize} && -b^T v \\ & \text{subject to} && A^T v - \lambda + c \geq 0 \end{aligned} \tag{5}$$

这个是一个不等式形式的线性规划。标准形式线性规划 (2) 的 Lagrange 对偶问题是问题 (3), 等价于问题 (4) 和 (5)

不等式形式线性规划 (LP) 的 Lagrange 对偶

例 2

考虑问题

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b \end{array} \quad (6)$$

给出其对偶问题

不等式形式线性规划 (LP) 的 Lagrange 对偶

Lagrange 函数:

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

对偶函数:

$$g(\lambda) = \inf_x L(x, \lambda) = -b^T \lambda + \inf_x (A^T \lambda + c)^T x = \begin{cases} -b^T \lambda & A^T \lambda + c = 0 \\ -\infty & \text{其他情况} \end{cases} \quad (7)$$

称对偶变量 λ 是对偶可行的, 如果 $\lambda \geq 0$ 且 $A^T \lambda + c = 0$

不等式形式线性规划 (LP) 的 Lagrange 对偶

对偶问题

$$\begin{array}{ll}\text{maximize} & -b^T\lambda \\ \text{subject to} & A^T\lambda + c = 0, \quad \lambda \geq 0\end{array}\quad (8)$$

- 可发现：不等式形式线性规划的对偶问题是一个标准形式的线性规划
- 可自行证明对偶问题的对偶问题等价于原问题

二次规划的 Lagrange 对偶

例 3

考虑问题

$$\begin{array}{ll} \text{minimize} & x^T P x \\ \text{subject to} & A x \leq b \end{array} \quad (9)$$

求其对偶问题

二次规划的 Lagrange 对偶

对偶函数:

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

对偶问题:

$$\begin{aligned} & \text{maximize} && -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda \\ & \text{subject to} && \lambda \geq 0 \end{aligned} \tag{10}$$

33.1.2 弱对偶性

弱对偶性：

$$d^* \leq p^* \quad (11)$$

d^* 表示 Lagrange 对偶问题的最优值，根据定义，这是通过 Lagrange 函数得到的原问题最优值 p^* 的最好下界

- 不等式总是成立的（不论是凸问题还是非凸问题）
- 当原问题很难求解时，弱对偶不等式可以给出原问题最优值的一个下界

33.1.2 弱对偶性

例如，考虑双向划分问题，其对偶问题是一个半定规划问题

$$\begin{array}{ll}\text{maximize} & -\mathbf{1}^T \nu \\ \text{subject to} & W + \text{diag}(\nu) \succeq 0\end{array}$$

其中， $\nu \in R^n$ 。对偶问题的最优值给出了双向划分问题最优值的一个下界

33.1.3 强对偶性和 Slater 约束准则

强对偶性：

$$d^* = p^* \quad (12)$$

- 对于一般情况，强对偶性不成立
- (一般) 只对凸问题成立，可表述为

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m, \\ & && Ax = b, \end{aligned} \quad (13)$$

其中，函数 f_0, \dots, f_m 是凸函数

- 保证强对偶性在凸问题中成立的条件称为约束准则

33.1.3 强对偶性和 Slater 约束准则

一个简单的约束准则是 **Slater 条件**：存在一点 $x \in \text{relint } \mathcal{D}$ 使得

$$f_i(x) < 0, \quad i = 1, \dots, m, \quad Ax = b. \quad (14)$$

- 满足上述条件的点有时称为**严格可行**
- 当 Slater 条件成立（且原问题是凸问题）时，强对偶性成立

33.1.3 强对偶性和 Slater 约束准则

条件进一步改进：当不等式约束函数 f_i 中有一些仿射函数时，仿射不等式不需要严格成立，该条件为：存在一点 $x \in \text{relint } \mathcal{D}$ 使得

$$f_i(x) \leq 0, \quad i = 1, \dots, k, \quad f_i(x) < 0, \quad i = k+1, \dots, m, \quad Ax = b \quad (15)$$

- 当所有约束条件都是线性等式或不等式且 $\text{dom } f_0$ 是开集时，改进的 Slater 条件 (15) 就是可行性条件
- 若 Slater 条件满足，对于凸问题强对偶性成立，且 $d^* > -\infty$ 时对偶问题能够取得最优值，即存在一组对偶可行解 $(\lambda^*, \nu^*) = d^* = p^*$
- 有很多其他的约束准则

线性方程组的最小二乘解

例 4

考虑问题

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

求其对偶问题及对偶性

线性方程组的最小二乘解

考虑问题

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

对偶问题:

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

- 这是一个凹二次函数的无约束极大化问题
- Slater 条件此时是原问题的可行性条件, 所以如果 $b \in R(A)$, 即 $p^* < \infty$, 有 $p^* = d^*$
- 对于此问题, 强对偶性通常成立, 即使 $p^* = \infty$

二次约束二次规划 (QCQP) 的 Lagrange 对偶

例 5

考虑问题

$$\begin{aligned} & \text{minimize} \quad (1/2)x^T P_0 x + q_0^T x + r_0 \\ & \text{subject to} \quad (1/2)x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \dots, m \end{aligned} \tag{16}$$

其中, $P_0 \in S_{++}^n$, $P_i \in S_{++}^n$, $i = 1, \dots, m$ 。求其对偶问题及对偶性

二次约束二次规划 (QCQP) 的 Lagrange 对偶

考虑问题

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

其中, $P_0 \in S_{++}^n$, $P_i \in S_{++}^n$, $i = 1, \dots, m$

Lagrange 函数:

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

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

二次约束二次规划 (QCQP) 的 Lagrange 对偶

如果 $\lambda \geq 0$, 有 $P(\lambda) > 0$ 及

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

对偶问题:

$$\begin{aligned} & \text{maximize} && -(1/2)q(\lambda)^T P(\lambda)^{-1} q(\lambda) + r(\lambda) \\ & \text{subject to} && \lambda \geq 0 \end{aligned} \tag{17}$$

- 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

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

对偶问题 (17) 和原问题 (16) 之间强对偶性成立

熵的最大化

例 6

考虑问题

$$\begin{aligned} & \text{minimize} && \sum_{i=1}^n x_i \log x_i \\ & \text{subject to} && Ax \leq b, \quad \mathbf{1}^T x = 1 \end{aligned}$$

定义域为 $D = R_+^m$ 。求其对偶问题及对偶性

熵的最大化

对偶问题：

$$\begin{aligned} \text{maximize} \quad & -b^T \lambda - \nu - e^{-\nu-1} \sum_{i=1}^n e^{-a_i^T \lambda} \\ \text{subject to} \quad & \lambda \geq 0 \end{aligned} \tag{18}$$

其中对偶变量 $\lambda \in R^m$, $\nu \in R$ 。对于原问题，根据 (弱化的)Slater 条件，如果存在 $x > 0$ ，使得 $Ax \leq b$, $\mathbf{1}^T x = 1$ ，最优对偶间隙为零

熵的最大化

关于对偶变量 ν 解析求最大可以简化对偶问题 (18)。对于任意固定 λ ，当目标函数对 ν 的导数为零时，即

$$\nu = \log \sum_{i=1}^n e^{-a_i^T \lambda} - 1$$

目标函数取最大值。将 ν 的最优值代入对偶问题得

$$\begin{aligned} \text{maximize} \quad & -b^T \lambda - \log\left(\sum_{i=1}^n e^{-a_i^T \lambda}\right) \\ \text{subject to} \quad & \lambda \geq 0 \end{aligned}$$

这是一个非负约束的几何规划问题（凸优化问题）

1 33.1 对偶问题与对偶约束

2 33.2 对偶解释

3 33.3 最优性条件

33.2.1 通过函数值集合理解强弱对偶性

考虑标准形式的优化问题:

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

自变量 $x \in R^n$, 定义域 \mathcal{D} , 最优值 p^*
可以通过集合

$$\mathcal{G} = \{(f_1(x), \dots, f_m(x), h_1(x), \dots, h_p(x), f_0(x)) \in R^m \times R^p \times R \mid x \in \mathcal{D}\} \quad (19)$$

给出对偶函数的简单几何解释

33.2.1 通过函数值集合理解强弱对偶性

利用集合 \mathcal{G} , 表达最优值 p^*

$$p^* = \inf\{t \mid (u, v, t) \in \mathcal{G}, u \leq 0, v = 0\}$$

求以 (λ, ν) 为自变量的对偶函数, 在 $(u, v, t) \in \mathcal{G}$ 上极小化仿射函数

$$(\lambda, \nu, 1)^T(u, v, t) = \sum_{i=1}^m \lambda_i u_i + \sum_{i=1}^p \nu_i v_i + t$$

得到

$$g(\lambda, \nu) = \inf\{(\lambda, \nu, 1)^T(u, v, t) \mid (u, v, t) \in \mathcal{G}\}$$

如果下确界有限, 则不等式

$$(\lambda, \nu, 1)^T(u, v, t) \geq g(\lambda, \nu)$$

是集合 \mathcal{G} 的一个 (非竖直) 支撑超平面

33.2.1 通过函数值集合理解强弱对偶性

假设 $\lambda \geq 0$, 如果 $u \leq 0$ 且 $v = 0$, 则成立 $t \geq (\lambda, \nu, 1)^T(u, v, t)$, 有

$$\begin{aligned} p^* &= \inf\{t \mid (u, v, t) \in \mathcal{G}, u \leq 0, v = 0\} \\ &\geq \inf\{(\lambda, \nu, 1)^T(u, v, t) \mid (u, v, t) \in \mathcal{G}, u \leq 0, v = 0\} \\ &\geq \inf\{(\lambda, \nu, 1)^T(u, v, t) \mid (u, v, t) \in \mathcal{G}\} \\ &= g(\lambda, \nu) \end{aligned}$$

即弱对偶性成立。针对只有一个不等式约束的简单问题, 图1和2描述了上述几何意义

33.2.1 通过函数值集合理解强弱对偶性

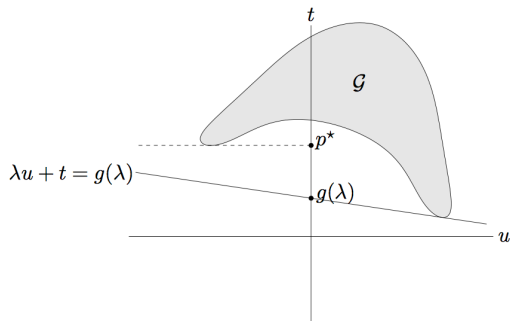


图 1: 针对只有一个不等式约束 $f_1(x) \leq 0$ 的简单问题, 对偶函数和下界 $g(\lambda) \leq p^*$ 的几何解释。给定 λ , 在集合 $\mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in \mathcal{D}\}$ 上极小化 $(\lambda, 1)^T(u, t)$, 得到斜率为 λ 的支撑超平面。支撑超平面与坐标轴 $u = 0$ 的交点即为 $g(\lambda)$ 。

33.2.1 通过函数值集合理解强弱对偶性

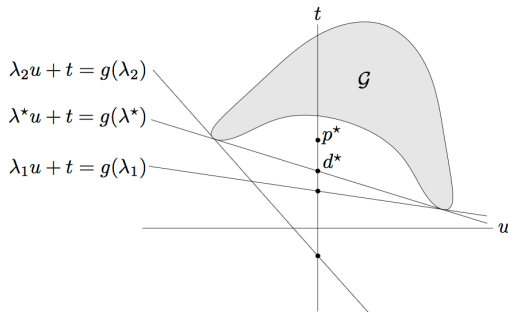


图 2: 对偶可行的三个 λ 值对应的支撑超平面, 这三个值中包含最优值 λ^* 。强对偶性此时不成立: 最优对偶间隙 $p^* - d^*$ 大于零。

上镜图变化

以另一种方式理解对偶性的几何意义，同样基于集合 \mathcal{G} ，解释为什么对（大部分）凸问题，强对偶性总是成立，定义集合 $\mathcal{A} \subseteq R^m \times R^p \times R$ 为

$$\mathcal{A} = \mathcal{G} + (R_+^m \times \{0\} \times R_+) \quad (20)$$

或者更明确的

$$\mathcal{A} = \{(u, v, t) \mid \exists x \in \mathcal{D}, f_i(x) \leq u_i, \ i = 1, \dots, m, \\ h_i(x) = v_i, \ i = 1, \dots, p, f_0(x) \leq t\}$$

上镜图变化

可以认为 \mathcal{A} 是 \mathcal{G} 的一种上镜图形式, 因为 \mathcal{A} 中包含 \mathcal{G} 中的所有点以及一些“较坏”的点, 即目标函数值或者不等式约束函数值较大的点。可以通过 \mathcal{A} 来描述最优值

$$p^* = \inf\{t \mid (0, 0, t) \in \mathcal{A}\}$$

为了得到当 $\lambda \geq 0$ 时关于 (λ, ν) 的对偶函数, 可以在 \mathcal{A} 上极小化仿射函数 $(\lambda, \nu, 1)^T(u, v, t)$: 如果 $\lambda \geq 0$, 则

$$g(\lambda, \nu) = \inf\{(\lambda, \nu, 1)^T(u, v, t) \mid (u, v, t) \in \mathcal{A}\}$$

如果下确界有限, 则

$$(\lambda, \nu, 1)^T(u, v, t) \geq g(\lambda, \nu)$$

定义了 \mathcal{A} 的一个非竖直的支撑超平面

上镜图变化

特别地, 因为 $(0, 0, p^*) \in \mathbf{bd}\mathcal{A}$, 有

$$p^* = (\lambda, \nu, 1)^T(0, 0, p^*) \geq g(\lambda, \nu) \quad (21)$$

即弱对偶性成立。强对偶性成立, 当且仅当存在某些对偶可行变量 (λ, ν) , 使得上式取等号, 即对集合 \mathcal{A} , 存在一个边界点 $(0, 0, p^*)$ 处的非竖直的支撑超平面。如图3

上镜图变化

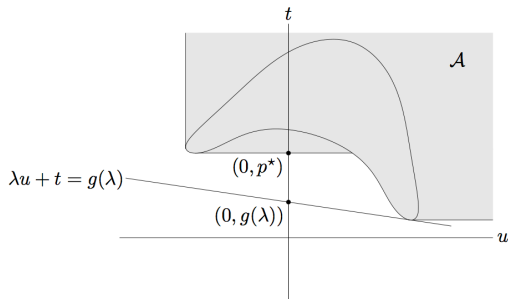


图 3: 对于具有一个 (不等式) 约束的问题, 对偶函数和下界 $g(\lambda) \leq p^*$ 的几何解释。给定 $\mathcal{A} = \{(u, t) \mid \exists x \in \mathcal{D}, f_0(x) \leq t, f_1(x) \leq u\}$ 上极小化 $(\lambda, 1)^T(u, t)$ 。这样可以得到斜率为 $-\lambda$ 的支撑超平面。此支撑超平面和坐标轴 $u = 0$ 的交点即为 $g(\lambda)$ 。

33.2.2 在约束准则下强对偶性成立的证明

考虑问题

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

其中函数 f_0, \dots, f_m 均为凸函数, 假设 Slater 条件满足: 即存在一点 $\tilde{x} \in \text{relint } \mathcal{D}$ 使得 $f_i(\tilde{x}) < 0, i = 1, \dots, m$ 且 $A\tilde{x} = b$ 。为简化证明, 附加两个假设条件:

- \mathcal{D} 的内点集不为空集 (即 $\text{relint } \mathcal{D} = \text{int } \mathcal{D}$)
- $\text{rank } A = p$ 。假设最优值 p^* 有限

33.2.2 在约束准则下强对偶性成立的证明

公式 (20) 中集合 \mathcal{A} 是凸集, 定义另一凸集 \mathcal{B} 为

$$\mathcal{B} = \{(0, 0, s) \in R^m \times R^p \times R | s < p^*\}$$

集合 \mathcal{A} 和集合 \mathcal{B} 不相交。设存在 $(u, v, t) \in \mathcal{A} \cap \mathcal{B}$ 。因为 $(u, v, t) \in \mathcal{B}$, 有 $u = 0$, $v = 0$, 以及 $f_0(x) \leq t \leq p^*$, 而这与 p^* 是原问题的最优值矛盾

根据超平面定理, 存在 $(\tilde{\lambda}, \tilde{\nu}, \mu) \neq 0$ 和 α 使得

$$(u, v, t) \in \mathcal{A} \Rightarrow \tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \geq \alpha \quad (22)$$

和

$$(u, v, t) \in \mathcal{B} \Rightarrow \tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \leq \alpha \quad (23)$$

根据式 (22), 有 $\tilde{\lambda} \geq 0$ 和 $\mu \geq 0$ 。式 (23) 意味着对所有 $t < p^*$ 有 $\mu t \leq \alpha$, 因此 $\mu p^* \leq \alpha$ 。结合式 (22), 对任意 $x \in \mathcal{D}$, 下式成立

$$\sum_{i=1}^m \tilde{\lambda}_i f_i(x) + \tilde{\nu}^T (Ax - b) + \mu f_0(x) \geq \alpha \geq \mu p^* \quad (24)$$

33.2.2 在约束准则下强对偶性成立的证明

设 $\mu > 0$, 式 (24) 两端除以 μ , 可得任意 $x \in \mathcal{D}$, 下式成立

$$L(x, \tilde{\lambda}/\mu, \tilde{\mu}/\mu) \geq p^*$$

定义

$$\lambda = \tilde{\lambda}/\mu, \quad \nu = \tilde{\nu}/\mu$$

对 x 求极小可以得到 $g(\lambda, \nu) \geq p^*$ 。根据强对偶性, 有 $g(\lambda, \nu) \leq p^*$, 因此 $g(\lambda, \nu) = p^*$ 。说明当 $\mu > 0$ 时强对偶性成立, 且对偶问题能达到最优值。

33.2.2 在约束准则下强对偶性成立的证明

考虑当 $\mu = 0$ 时的情形。根据式 (24)，对任意 $x \in \mathcal{D}$ ，有

$$\sum_{i=1}^m \tilde{\lambda}_i f_i(x) + \tilde{\nu}^T(Ax - b) \geq 0 \quad (25)$$

满足 Slater 条件的点 \tilde{x} 同样满足式 (25)，因此有 $\sum_{i=1}^m \tilde{\lambda}_i f_i(x) \geq 0$

- $f_i(\tilde{x}) < 0$ 且 $\tilde{\lambda}_i \geq 0$ ，有 $\tilde{\lambda} = 0$
- $(\tilde{\lambda}, \tilde{\nu}, \mu) \neq 0$ 且 $\tilde{\lambda} = 0$ ， $\mu = 0$ ，所以 $\tilde{\nu} \neq 0$
- 式 (25) 表明对任意 $x \in \mathcal{D}$ 有 $\tilde{\nu}^T(Ax - b) \geq 0$ 。又因为 \tilde{x} 满足 $\tilde{\nu}^T(A\tilde{x} - b) = 0$ ，且 $\tilde{x} \in \text{int } \mathcal{D}$ ，因此除了 $A^T \tilde{\nu} = 0$ 的情况，总存在 \mathcal{D} 中的点使得 $\tilde{\nu}^T(Ax - b) < 0$
- $A^T \tilde{\nu} = 0$ 显然与假设 $\text{rank } A = p$ 矛盾
- 上述证明的集合意义如图??

33.2.2 在约束准则下强对偶性成立的证明

33.2.3 强弱对偶性的极大极小描述

将原、对偶优化问题以一种更为对称的方式进行表达

$$\begin{aligned}\sup_{\lambda \geq 0} L(x, \lambda) &= \sup_{\lambda \geq 0} (f_0(x) + \sum_{i=1}^m \lambda_i f_i(x)) \\ &= \begin{cases} f_0(x) & f_i(x) \leq 0, \quad i = 1, \dots, m \\ \infty & \text{其他情况} \end{cases}\end{aligned}$$

假设 x 不可行，即存在某些 i 使得 $f_i(x) > 0$ 。选择 $\lambda_j = 0, j \neq i$ ，以及 $\lambda_i \rightarrow \infty$ 可以得出 $\sup_{\lambda \geq 0} L(x, \lambda) = \infty$ 。反过来，如果 x 可行，则有 $f_i \leq 0, i = 1, \dots, m$ ， λ 的最优选择为 $\lambda = 0, \sup_{\lambda \geq 0} L(x, \lambda) = f_0(x)$ 。这意味着可将原问题的最优值写成如下形式

$$p^* = \inf_x \sup_{\lambda \geq 0} L(x, \lambda)$$

33.2.3 强弱对偶性的极大极小描述

根据对偶函数的定义，有

$$d^* = \sup_{\lambda \geq 0} \inf_x L(x, \lambda)$$

因此弱对偶性可以表述为下述不等式

$$\sup_{\lambda \geq 0} \inf_x L(x, \lambda) = \inf_x \sup_{\lambda \geq 0} L(x, \lambda) \quad (26)$$

强对偶性可以表示为下面的不等式

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

强对偶性意味着对 x 求极小和对 $\lambda \geq 0$ 求极大可以互换而不影响结果

33.2.3 强弱对偶性的极大极小描述

事实上, 不等式 (26) 是否成立和 L 的性质无关: 对任意 $f: R^n \times R^m \rightarrow R$ (以及任意 $W \subseteq R^n$ 和 $Z \subseteq R^m$), 下式成立:

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) \leq \inf_{w \in W} \sup_{z \in Z} f(w, z) \quad (27)$$

这个一般性的等式称为极大极小不等式。若等式成立, 即

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) = \inf_{w \in W} \sup_{z \in Z} f(w, z) \quad (28)$$

称 f (以及 W 和 Z) 满足强极大极小性质或者鞍点性质。强极大极小性质只在特殊情形下成立, 例如, 函数 $f: R^n \times R^m \rightarrow R$ 是满足强对偶性问题的 Lagrange 函数, 而 $W = R^n$, $Z = R_+^m$

33.2.4 鞍点解释

称一对 $\tilde{w} \in W, \tilde{z} \in Z$ 是函数 f (以及 W 和 Z) 的鞍点, 如果对于任意 $w \in W$ 和 $z \in Z$ 下式成立

$$f(\tilde{w}, z) \leq f(\tilde{w}, \tilde{z}) \leq f(w, \tilde{z})$$

换言之, $f(w, \tilde{z})$ 在 \tilde{w} 处取得最小值 (关于变量 $w \in W$), $f(\tilde{w}, z)$ 在 \tilde{z} 处取得最大值 (关于变量 $z \in Z$) :

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

上式意味着强极大极小性质 (28) 成立, 且共同值为 $f(\tilde{w}, \tilde{z})$

- 如果 x^* 和 λ^* 分别是原问题和对偶问题的最优点, 且强对偶性成立, 则它们是 Lagrange 函数的一个鞍点。反过来同样成立

1 33.1 对偶问题与对偶约束

2 33.2 对偶解释

3 33.3 最优性条件

33.3.1 互补松弛性

设原问题和对偶问题的最优值都可以达到且相等（即强对偶性成立）。令 x^* 是原问题的最优解， (λ^*, ν^*) 是对偶问题的最优解，这表明

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

在上面的式子链中，两个不等式取等号，得出一些有意义的结论

33.3.1 互补松弛性

- $L(x, \lambda^*, \nu^*)$ 关于 x 求极小时在 x^* 处取得最小值

-

$$\sum_{i=1}^m \lambda_i^* f_i(x^*) = 0$$

求和项的每一项都非正，因此有

$$\lambda_i^* f_i(x^*) = 0, \quad i = 1, \dots, m.$$

称为**互补松弛性**：它对任意原问题最优解 x^* 以及对偶问题最优解 (λ^*, ν^*) 都成立（当强对偶性成立时）。可将其写成

$$\lambda_i^* > 0 \implies f_i(x^*) = 0$$

或者等价地

$$f_i(x^*) < 0 \implies \lambda_i^* = 0$$

33.3.2 Karush-Kuhn-Tucker(KKT) 最优性条件

非凸问题的 KKT 条件： 以下四个条件称为 KKT 条件 (f_i, h_i 可微)

- 原问题约束：

$$f_i(x^*) \leq 0, \quad i = 1, \dots, m$$

$$h_i(x^*) = 0, \quad i = 1, \dots, p$$

- 对偶约束： $\lambda_i^* \geq 0, \quad i = 1, \dots, m$
- 松弛互补： $\lambda_i^* f_i(x^*) = 0, \quad i = 1, \dots, m$
- Lagrange 关于 x 梯度为 0：

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

由上一页：如果强对偶性成立 x, λ, ν 是最优点，他们一定满足 KKT 条件

33.3.2 Karush-Kuhn-Tucker(KKT) 最优性条件

凸问题的 KKT 条件

如果 $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ 是满足凸问题的 KKT 条件的点, 那么它们一定是最优点:

- 由互补松弛: $f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
- 由最后一个条件 (和凸性): $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

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

如果满足 Slater 条件: x 是最优点当且仅当存在 λ, ν 满足 KKT 条件

- 回顾 Slater 蕴含强对偶性, 并且对偶最优点可取到
- 是无约束问题最优条件 $\nabla f_0(x) = 0$ 的推广

等式约束二次凸问题求极小

例 7

考虑问题

$$\begin{aligned} & \text{minimize} && (1/2)x^T Px + q^T x + r \\ & \text{subject to} && Ax = b \end{aligned}$$

其中 $P \in S_+^n$ 。求其 KKT 条件

等式约束二次凸问题求极小

此问题的 KKT 条件为

$$\mathbf{A}x^* = b, \quad Px^* + q + \mathbf{A}^T v^* = 0$$

可将其写成

$$H_x = \begin{bmatrix} P & \mathbf{A}^T \\ \mathbf{A} & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$$

求解变量 x^*, ν^* 的 $m+n$ 个方程，其中变量的维数为 $m+n$ ，可以得到原问题的最优原变量和对偶变量

33.3.3 通过解对偶问题求解原问题

假设强对偶性成立，对偶最优解 (λ^*, ν^*) 已知。假设 $L(x, \lambda^*, \nu^*)$ 的最小点，即下列问题的解

$$\text{minimize} \quad f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x)$$

唯一。

- 如果此问题的解是原问题可行解，那么它就是原问题最优解；如果不是，那么原问题不存在最优点，即原问题的最优解无法达到
- 当对偶问题比原问题更易求解时，比如说对偶问题可以解析求解或者有某些特殊的结构更易分析，上述方法很有意义

熵的最大化

例 8

考虑问题

$$\begin{aligned} & \text{minimize} && f_0(x) = \sum_{i=1}^n x_i \log x_i \\ & \text{subject to} && \mathbf{A}x \leq b, \quad \mathbf{1}^T x = \mathbf{1} \end{aligned}$$

其中定义域为 R_{++}^n 。求其最优解

熵的最大化

考虑问题

$$\begin{aligned} & \text{minimize} && f_0(x) = \sum_{i=1}^n x_i \log x_i \\ & \text{subject to} && \mathbf{A}x \leq b, \quad \mathbf{1}^T x = 1 \end{aligned}$$

对偶问题：

$$\begin{aligned} & \text{maximize} && -b^T \lambda - \nu - e^{-\nu-1} \sum_{i=1}^n e^{-a_i^T \lambda} \\ & \text{subject to} && \lambda \geq 0 \end{aligned}$$

- 假设 Slater 条件的弱化形式成立，即存在 $x > 0$ 使得 $\mathbf{A}x \leq b$ 以及 $\mathbf{1}^T x = 1$ ，因此强对偶性成立，存在一个对偶最优解 (λ^*, ν^*)

熵的最大化

设对偶问题已经解出。 (λ^*, ν^*) 处的 Lagrange 函数为

$$L(x, \lambda^*, \nu^*) = \sum_{i=1}^n x_i \log x_i + \lambda^{*T}(\mathbf{A}x - b) + \nu^*(\mathbf{1}^T x - 1)$$

它在 \mathcal{D} 上严格凸且有下界，因此有一个唯一解 x^* ，

$$x_i^* = 1 / \exp(a_i^T \lambda^* + \nu^* + 1), \quad i = 1, \dots, n$$

其中 a_i 是矩阵 A 的列向量。如果 x^* 数原问题可行解，则其必是原问题的最优解。如果 x^* 不是原问题可行解，则原问题的最优解不能达到

在等式约束下极小化可分函数

例 9

考虑问题

$$\begin{aligned} & \text{minimize} \quad f_0(x) = \sum_{i=1}^n f_i(x_i) \\ & \text{subject to} \quad a^T x = b \end{aligned}$$

其中 $a \in R^n, b \in R$, 函数 $f_i: R \rightarrow R$ 是可微函数, 严格凸。求其最优解

在等式约束下极小化可分函数

Lagrange 函数:

$$L(x, \nu) = \sum_{i=1}^n f_i(x_i) + \nu(a^T x - b) = -b\nu + \sum_{i=1}^n (f_i(x_i) + \nu a_i x_i)$$

对偶函数:

$$\begin{aligned} g(\nu) &= -b\nu + \inf_x \left(\sum_{i=1}^n (f_i(x_i) + \nu a_i x_i) \right) \\ &= -b\nu + \sum_{i=1}^n \inf_{x_i} (f_i(x_i) + \nu a_i x_i) \\ &= -b\nu - \sum_{i=1}^n f_i^*(-\nu a_i) \end{aligned}$$

在等式约束下极小化可分函数

对偶问题：

$$\text{maximize} \quad -b\nu - \sum_i^n f_i^*(-\nu a_i)$$

- 假设找到一个对偶最优解 ν^* 。 $L(x, \nu^*)$ 具有唯一的最小点 \tilde{x} ，有 $\tilde{x} = x^*$ 。可以通过求解 $\Delta_x L(x, \nu^*) = 0$ 得到 x^* ，即求解方程组 $f'_i(x_i^*) = -\nu^* a_i$ 。