

拉格朗日对偶

张腾

2023 年 4 月 6 日

考虑一般形式的优化问题

$$\inf_x f_0(x), \quad \text{s.t. } f_i(x) \leq 0, \quad i \in [m]$$

其中自变量 $x \in \mathbb{R}^n$, 问题可行域 $\mathcal{D} = \cap_{i=0}^m \text{dom}(f_i)$ 非空, 最优值为 p^* 。注意这里并不要求问题是凸优化, 此外没有等式约束是因为它总可以写成两个不等式约束。

1 对偶函数

该问题对应的拉格朗日函数 $L: \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$ 定义如下:

$$L(x, \lambda) = f_0(x) + \sum_{i \in [m]} \lambda_i f_i(x)$$

其中 $\lambda = [\lambda_1, \dots, \lambda_m]$ 称为拉格朗日乘子 (向量)。对偶函数定义为拉格朗日函数关于 x 取极小, 即

$$g(\lambda) = \inf_{x \in \mathcal{D}} L(x, \lambda) = \inf_{x \in \mathcal{D}} \left\{ f_0(x) + \sum_{i \in [m]} \lambda_i f_i(x) \right\}$$

如果拉格朗日函数关于 x 无下界, 则对偶函数的值为 $-\infty$ 。注意

$$f_0(x) + \sum_{i \in [m]} \lambda_i f_i(x)$$

是关于 λ 的线性函数, 参数 x 的不同取值对应不同的线性函数, 因此 $g(\lambda)$ 是一族线性函数取下确界, 故对偶函数总是凹的, 即便原问题非凸。

1.1 最优值的下界

对任意 $\lambda \geq 0$, 设 $x \in \mathcal{D}$, 于是 $L(x, \lambda) = f_0(x) + \sum_{i \in [m]} \lambda_i f_i(x) \leq f_0(x)$, 从而有

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

由于上式对 $\forall x \in \mathcal{D}$ 都成立, 因此有 $g(\lambda) \leq p^*$, 即对偶函数构成了原问题最优值 p^* 的一个下界。

注. 虽然 $g(\lambda) \leq p^*$ 恒成立, 但当 $g(\lambda) = -\infty$ 时意义不大。只有当 $\mathbf{0} \leq \lambda \in \text{dom}(g)$ 时, 对偶函数才能给出 p^* 的一个非平凡下界, 满足 $\mathbf{0} \leq \lambda \in \text{dom}(g)$ 的 λ 称为是对偶可行的。

1.2 从线性逼近的角度来理解

原问题可以重写为一个无约束问题

$$\inf_{\mathbf{x}} f_0(\mathbf{x}) + \sum_{i \in [m]} I_-(f_i(\mathbf{x}))$$

其中 $I_- : \mathbb{R} \mapsto \mathbb{R}$ 是非正实数的指示函数:

$$I_-(u) = \begin{cases} 0, & u \leq 0 \\ \infty, & u > 0 \end{cases}$$

该函数可以理解为我们对约束函数值 $u = f_i(\mathbf{x})$ 的一种“不满”，只是这种不满很“强硬”，随着函数值 $f_i(\mathbf{x})$ 从非正数变为正数，不满瞬间从 0 升到 ∞ 。对比拉格朗日函数

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f_0(\mathbf{x}) + \sum_{i \in [m]} \lambda_i f_i(\mathbf{x})$$

可以看出就是用线性函数 $\lambda_i f_i(\mathbf{x})$ 代替了原本的指示函数 $I_-(f_i(\mathbf{x}))$ ，这可看作一种“妥协”或者“软化”，注意 $\lambda_i \geq 0$ 时有 $\lambda_i f_i(\mathbf{x}) \leq I_-(f_i(\mathbf{x}))$ ，即这个线性函数是指示函数的一个下界。

1.3 与共轭函数的关系

考虑只有线性不等式约束的优化问题

$$\inf_{\mathbf{x}} f(\mathbf{x}), \quad \text{s.t. } \mathbf{A}\mathbf{x} \leq \mathbf{b}$$

其对偶函数为

$$\begin{aligned} g(\boldsymbol{\lambda}) &= \inf_{\mathbf{x}} \{f(\mathbf{x}) + \boldsymbol{\lambda}^\top (\mathbf{A}\mathbf{x} - \mathbf{b})\} \\ &= -\mathbf{b}^\top \boldsymbol{\lambda} + \inf_{\mathbf{x}} \{f(\mathbf{x}) + \boldsymbol{\lambda}^\top \mathbf{A}\mathbf{x}\} \\ &= -\mathbf{b}^\top \boldsymbol{\lambda} - \sup_{\mathbf{x}} \{-\boldsymbol{\lambda}^\top \mathbf{A}\mathbf{x} - f(\mathbf{x})\} \\ &= -\mathbf{b}^\top \boldsymbol{\lambda} - f^*(-\mathbf{A}^\top \boldsymbol{\lambda}) \end{aligned}$$

故对偶函数 g 的定义域也可由共轭函数 f^* 的定义域得到。

2 对偶问题

对于 $\forall \boldsymbol{\lambda} \geq \mathbf{0}$ ，对偶函数 $g(\boldsymbol{\lambda})$ 给出了原问题最优值 p^* 的下界，一个自然的问题就是所有下界中最好的下界有多好，即考虑如下优化问题：

$$\sup_{\boldsymbol{\lambda}} g(\boldsymbol{\lambda}), \quad \text{s.t. } \boldsymbol{\lambda} \geq \mathbf{0}$$

该问题称为原问题的对偶问题。前面曾提到“对偶可行”的概念，称满足 $\mathbf{0} \leq \lambda \in \text{dom}(g)$ 的 λ 是对偶可行的，因为此时它是对偶问题的一个可行点。注意对偶问题是极大化一个凹函数，因此对偶问题总是凸优化，与原问题是否为凸优化无关。

原问题很难求解时，可以考虑通过对偶问题先估计个下界。

例 1 (双向划分问题).

$$\inf_{\mathbf{x}} \mathbf{x}^\top \mathbf{W} \mathbf{x}, \quad \text{s.t. } x_i^2 = 1, i \in [n]$$

其中 $\mathbf{W} \in \mathbb{S}^n$ 。注意 $x_i = \pm 1$ ，故这是一个整数规划，可行点的数量随着 n 呈指数增长。这个问题可以看成 n 个元素集合 $\{1, 2, \dots, n\}$ 的双向划分问题， $x_i = 1$ 的分在一组， $x_i = -1$ 的分在一组， $[\mathbf{W}]_{ij}$ 是 i 和 j 分在一组的成本， $-[\mathbf{W}]_{ij}$ 是 i 和 j 不分在一组的成本。拉格朗日函数为

$$\begin{aligned} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\theta}) &= \mathbf{x}^\top \mathbf{W} \mathbf{x} + \sum_{i \in [n]} \lambda_i (x_i^2 - 1) + \sum_{i \in [n]} \theta_i (1 - x_i^2) \\ &= \mathbf{x}^\top \mathbf{W} \mathbf{x} + \sum_{i \in [n]} (\lambda_i - \theta_i)(x_i^2 - 1) \\ &= \mathbf{x}^\top (\mathbf{W} + \text{diag}(\boldsymbol{\lambda} - \boldsymbol{\theta})) \mathbf{x} - \mathbf{e}^\top (\boldsymbol{\lambda} - \boldsymbol{\theta}) \end{aligned}$$

注意 $\boldsymbol{\lambda} - \boldsymbol{\theta}$ 总是整体出现，设 $\boldsymbol{\nu} = \boldsymbol{\lambda} - \boldsymbol{\theta}$ ，因为是两个非负向量相减，因此 $\boldsymbol{\nu}$ 无界，对偶函数为

$$g(\boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\nu}) = \begin{cases} -\mathbf{e}^\top \boldsymbol{\nu}, & \mathbf{W} + \text{diag}(\boldsymbol{\nu}) \succeq \mathbf{0} \\ -\infty, & \text{其它} \end{cases}$$

从而有对偶问题

$$\sup_{\boldsymbol{\nu}} -\mathbf{e}^\top \boldsymbol{\nu}, \quad \text{s.t. } \mathbf{W} + \text{diag}(\boldsymbol{\nu}) \succeq \mathbf{0}$$

ν_i 要尽可能的小，但不能无限的小，它还得使 $\mathbf{W} + \text{diag}(\boldsymbol{\nu})$ 半正定，取 $-\lambda_{\min}(\mathbf{W})$ 是可行的，这样就得到原问题最优值 p^* 的一个下界 $n\lambda_{\min}(\mathbf{W})$ 。

注. 从这个例子也可以看出当原问题有等式约束 $h_j(\mathbf{x}) = 0, j \in [l]$ 时，对偶函数为

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} \left\{ f_0(\mathbf{x}) + \sum_{i \in [m]} \lambda_i f_i(\mathbf{x}) + \sum_{j \in [l]} \nu_j h_j(\mathbf{x}) \right\}$$

对偶问题为

$$\sup_{\boldsymbol{\lambda}, \boldsymbol{\nu}} g(\boldsymbol{\lambda}, \boldsymbol{\nu}), \quad \text{s.t. } \boldsymbol{\lambda} \geq \mathbf{0}$$

2.1 弱对偶

设对偶问题的最优值为 d^* ，易知有 $d^* \leq p^*$ ，这称为弱对偶性。差值 $p^* - d^*$ 称为对偶间隙，它总是非负的。

即便当 d^* 和 p^* 为无穷时，弱对偶性也成立。若原问题无下界，即 $p^* = -\infty$ ，为了保证弱对偶性，必须有 $d^* = -\infty$ ，即对偶问题不可行。反过来，若对偶问题无上界，即 $d^* = \infty$ ，为了保证弱对偶性，必须有 $p^* = \infty$ ，即原问题不可行。

2.2 强对偶

如果有 $d^* = p^*$, 即对偶间隙为零, 则称强对偶性成立。

通常来讲, 强对偶性是不成立的, 但如果原问题是凸优化, 即 f_0, \dots, f_m 是凸函数, 则强对偶性通常(但不总是)成立, 有很多研究给出了除凸性条件外强对偶性成立还需的额外条件, 其中最有名的就是 Slater 条件: 存在一点 $\tilde{x} \in \text{relint}(\mathcal{D})$ 使得所有不等式约束严格成立。

注. Slater 条件有个弱化版本也可保证强对偶性, 它不要求线性不等式严格成立, 因此可以取等号。对于凸优化, 如果有等式约束, 则必是线性等式(否则不是凸优化), 其可对应地写成两个线性不等式, 根据弱化版 Slater 条件, 只需这两个线性不等式成立, 即等式约束成立即可。

下面我们给出非弱化版 Slater 条件的证明。

定理 2 (Slater 条件). 设原问题是凸优化, 可行域 \mathcal{D} 相对内部非空, 存在一点 $\tilde{x} \in \text{relint}(\mathcal{D})$ 使得所有不等式严格成立, 则强对偶性成立。

证明. 由于 \mathcal{D} 相对内部非空, 故原问题最优值 $p^* < \infty$, 若 $p^* = -\infty$, 则根据弱对偶性有 $d^* = -\infty$, 此时强对偶性已成立, 故只需考虑 p^* 有限的情况。

引入集合 $\mathcal{A} = \{(\mathbf{u}, t) \mid \exists \mathbf{x} \in \mathcal{D}, f_0(\mathbf{x}) \leq t, f_i(\mathbf{x}) \leq u_i, i \in [m]\}$, 首先我们说明这个集合是凸的, 设 $\forall (\mathbf{u}, t), (\mathbf{v}, s) \in \mathcal{A}$, 则存在 $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$ 使得

$$\begin{aligned} f_i(\mathbf{x}_1) &\leq u_i, f_i(\mathbf{x}_2) \leq v_i, i \in [m] \\ f_0(\mathbf{x}_1) &\leq t, f_0(\mathbf{x}_2) \leq s \end{aligned}$$

根据 f_0, \dots, f_m 的凸性可知对于 $\forall \alpha \in [0, 1]$ 有

$$\begin{aligned} f_i(\alpha \mathbf{x}_1 + (1 - \alpha) \mathbf{x}_2) &\leq \alpha f_i(\mathbf{x}_1) + (1 - \alpha) f_i(\mathbf{x}_2) \leq \alpha u_i + (1 - \alpha) v_i, i \in [m] \\ f_0(\alpha \mathbf{x}_1 + (1 - \alpha) \mathbf{x}_2) &\leq \alpha f_0(\mathbf{x}_1) + (1 - \alpha) f_0(\mathbf{x}_2) \leq \alpha t + (1 - \alpha) s \end{aligned}$$

由于原问题是凸优化, 因此 \mathcal{D} 是凸集, 故 $\alpha \mathbf{x}_1 + (1 - \alpha) \mathbf{x}_2 \in \mathcal{D}$, 从而 $\alpha(\mathbf{u}, t) + (1 - \alpha)(\mathbf{v}, s) \in \mathcal{A}$, 这就证明了 \mathcal{A} 是凸集。

易知 $(\mathbf{0}, p^*) \notin \text{relint}(\mathcal{A})$, 否则存在 $\epsilon > 0$ 使得 $(\mathbf{0}, p^* - \epsilon) \in \text{relint}(\mathcal{A})$, 这与 p^* 是最优解矛盾, 因此存在超平面分离 $(\mathbf{0}, p^*)$ 和 \mathcal{A} , 即存在 $(\lambda, \mu) \neq \mathbf{0}$ 使得

$$\lambda^\top \mathbf{u} + \mu t \geq \lambda^\top \mathbf{0} + \mu p^* = \mu p^*, \quad \forall (\mathbf{u}, t) \in \mathcal{A}$$

注意根据集合 \mathcal{A} 的定义, \mathbf{u} 和 t 均无上界, 因此必须有 $\lambda \geq \mathbf{0}$ 和 $\mu \geq 0$ 。

若 $\mu = 0$, 则 λ 必不为零且

$$\lambda^\top \mathbf{u} \geq 0, \quad \forall (\mathbf{u}, t) \in \mathcal{A}$$

显然满足 Slater 条件的点 \tilde{x} 也满足上式, 即

$$\sum_{i \in [m]} \lambda_i f_i(\tilde{x}) \geq 0$$

注意所有的 $f_i(\tilde{x})$ 都严格小于零, 因此只能 $\lambda = \mathbf{0}$, 矛盾。故 $\mu > 0$, 两边同除以 μ 可得

$$\frac{1}{\mu} \lambda^\top \mathbf{u} + t \geq p^*, \quad \forall (\mathbf{u}, t) \in \mathcal{A}$$

从而

$$g\left(\frac{\lambda}{\mu}\right) = \inf_{x \in \mathcal{D}} L\left(x, \frac{\lambda}{\mu}\right) = \inf_{x \in \mathcal{D}} \left\{ f_0(x) + \frac{1}{\mu} \sum_{i \in [m]} \lambda_i f_i(x) \right\} = \inf_{(\mathbf{u}, t) \in \mathcal{A}} \left\{ \frac{1}{\mu} \lambda^\top \mathbf{u} + t \right\} \geq p^*$$

于是

$$d^* = \sup_{\lambda \geq 0} g\left(\frac{\lambda}{\mu}\right) \geq p^*$$

结合弱对偶性可知强对偶性成立。 ♣

注意 Slater 条件只是一个充分条件, 下面进一步证明若原问题最优值 p^* 有限, 则对偶问题的最优解集合有界。

证明. 由强对偶性, $d^* = p^*$ 也有限, 从而最优解存在。设 $\tilde{\lambda} \geq \mathbf{0}$ 是一个最优解, 则

$$d^* = g(\tilde{\lambda}) = \inf_{x \in \mathcal{D}} \left\{ f_0(x) + \sum_{i \in [m]} \tilde{\lambda}_i f_i(x) \right\} \leq f_0(\tilde{x}) + \sum_{i \in [m]} \tilde{\lambda}_i f_i(\tilde{x}) \leq f_0(\tilde{x}) + \left\{ \max_{i \in [m]} f_i(\tilde{x}) \right\} \sum_{i \in [m]} \tilde{\lambda}_i$$

移项整理得

$$\left\{ \min_{i \in [m]} -f_i(\tilde{x}) \right\} \sum_{i \in [m]} \tilde{\lambda}_i \leq f_0(\tilde{x}) - d^*$$

于是

$$\|\tilde{\lambda}\|_2 \leq \sum_{i \in [m]} \tilde{\lambda}_i \leq \frac{f_0(\tilde{x}) - d^*}{\min_{i \in [m]} -f_i(\tilde{x})}$$

第一个不等号是因为 ℓ_2 范数不超过 ℓ_1 范数, 第二个不等号利用了 \tilde{x} 是满足 Slater 条件的点。 ♣

3 几种解释

拉格朗日对偶有许多不同的解释。

3.1 几何解释

参考图 1, 引入集合

$$\mathcal{G} = \{(f_1(x), \dots, f_m(x), f_0(x)) \in \mathbb{R}^m \times \mathbb{R} \mid x \in \mathcal{D}\}$$

上面强对偶证明中引入的集合 \mathcal{A} 可以看作集合 \mathcal{G} 的非负扩展, 即 $\mathcal{A} = \mathcal{G} + \{\mathbb{R}_+^m \times \mathbb{R}_+\}$ 。

易知原问题最优值 $p^* = \inf\{t \mid (\mathbf{u}, t) \in \mathcal{G}, \mathbf{u} \leq \mathbf{0}\}$, 即 (\mathbf{u}, t) 平面上的左半部分与 \mathcal{G} 的交集中的点的纵坐标的最小值。此外对偶函数

$$g(\lambda) = \inf_{x \in \mathcal{D}} L(x, \lambda) = \inf_{x \in \mathcal{D}} \left\{ f_0(x) + \sum_{i \in [m]} \lambda_i f_i(x) \right\} = \inf_{(\mathbf{u}, t) \in \mathcal{G}} \{t + \lambda^\top \mathbf{u}\}$$

注意 (\mathbf{u}, t) 平面上的直线 $t = -\lambda^\top \mathbf{u} + g$ 的截距为 g , 因此 $g(\lambda)$ 是以 $-\lambda$ 为斜率且与 \mathcal{G} 相交的直线族的最小截距, 对偶问题最优值 $d^* = \sup_{\lambda \geq 0} g(\lambda)$ 就是这些最小截距中的最大者。

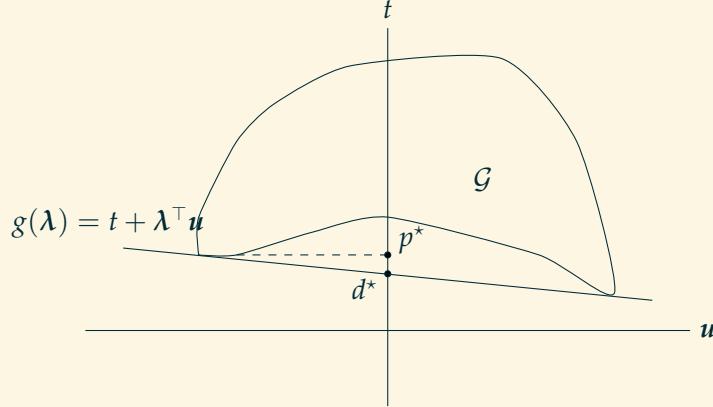


图 1: 拉格朗日对偶的几何解释

类似地, Slater 条件也可以如此解释, 参见图 2。存在一点 $x \in \text{relint}(\mathcal{D})$ 使得所有不等式约束严格成立, 意味着凸集 \mathcal{A} 与左半平面 (不包含纵轴) 交集非空, 此时存在过 $(\mathbf{0}, p^*)$ 的非竖直超平面分离 $(\mathbf{0}, p^*)$ 和 \mathcal{A} , 从而强对偶性成立。

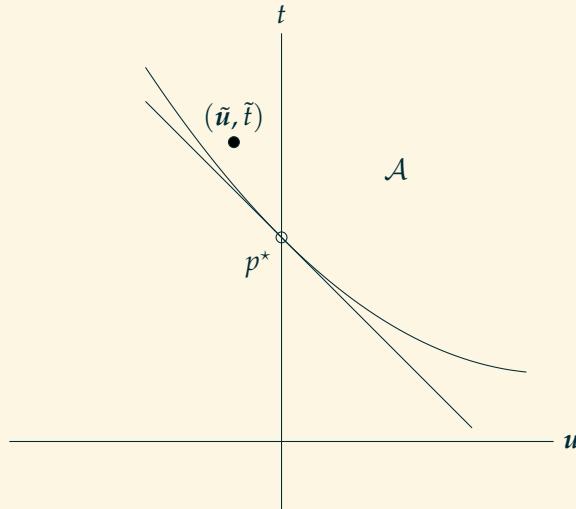


图 2: Slater 条件的解释

3.2 极大极小

设函数 $f : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$, $\mathcal{X} \in \mathbb{R}^n$, $\mathcal{Y} \in \mathbb{R}^m$, 易知有

$$f(x, y) \leq \sup_{y \in \mathcal{Y}} f(x, y), \quad \forall x \in \mathcal{X}, y \in \mathcal{Y}$$

两边同时对 $x \in \mathcal{X}$ 取下确界有

$$\inf_{x \in \mathcal{X}} f(x, y) \leq \inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} f(x, y), \quad \forall y \in \mathcal{Y}$$

左边对 $y \in \mathcal{Y}$ 取上确界有

$$\sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} f(x, y) \leq \inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} f(x, y)$$

这称为极大极小不等式。若等号成立，则称 f 满足极大极小性质或鞍点性质，此时求极小和求极大互换顺序不影响结果。

若 $(\tilde{x}, \tilde{y}) \in \mathcal{X} \times \mathcal{Y}$ 满足

$$f(\tilde{x}, y) \leq f(\tilde{x}, \tilde{y}) \leq f(x, \tilde{y}), \quad \forall x \in \mathcal{X}, y \in \mathcal{Y}$$

即

$$f(\tilde{x}, \tilde{y}) = \inf_{x \in \mathcal{X}} f(x, \tilde{y}) = \sup_{y \in \mathcal{Y}} f(\tilde{x}, y)$$

则称 (\tilde{x}, \tilde{y}) 是 f 的鞍点。此时有

$$\inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} f(x, y) \leq \sup_{y \in \mathcal{Y}} f(\tilde{x}, y) = f(\tilde{x}, \tilde{y}) = \inf_{x \in \mathcal{X}} f(x, \tilde{y}) \leq \sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} f(x, y)$$

结合极大极小不等式知鞍点性质满足。

回到拉格朗日对偶，易知

$$\sup_{\lambda \geq 0} L(x, \lambda) = \sup_{\lambda \geq 0} \left\{ f_0(x) + \sum_{i \in [m]} \lambda_i f_i(x) \right\} = \begin{cases} f_0(x) & f_i(x) \leq 0, i \in [m] \\ \infty & \text{o.w.} \end{cases}$$

于是

$$p^* = \inf_{x \in \mathcal{D}} \sup_{\lambda \geq 0} L(x, \lambda)$$

而

$$d^* = \sup_{\lambda \geq 0} g(\lambda) = \sup_{\lambda \geq 0} \inf_{x \in \mathcal{D}} L(x, \lambda)$$

故极大极小不等式对应的就是弱对偶性，鞍点性质对应的就是强对偶性。

设 x^* 和 λ^* 分别是原问题和对偶问题的最优解, 则

$$\inf_{x \in \mathcal{D}} \sup_{\lambda \geq 0} L(x, \lambda) = \sup_{\lambda \geq 0} L(x^*, \lambda), \quad \sup_{\lambda \geq 0} \inf_{x \in \mathcal{D}} L(x, \lambda) = \inf_{x \in \mathcal{D}} L(x, \lambda^*)$$

于是

$$\sup_{\lambda \geq 0} \inf_{x \in \mathcal{D}} L(x, \lambda) = \inf_{x \in \mathcal{D}} L(x, \lambda^*) \leq L(x^*, \lambda^*) \leq \sup_{\lambda \geq 0} L(x^*, \lambda) = \inf_{x \in \mathcal{D}} \sup_{\lambda \geq 0} L(x, \lambda)$$

若强对偶性成立, 则上式中的不等号全部取等号, 从而有

$$\inf_{x \in \mathcal{D}} L(x, \lambda^*) = L(x^*, \lambda^*) = \sup_{\lambda \geq 0} L(x^*, \lambda)$$

即 (x^*, λ^*) 是拉格朗日函数 L 的鞍点。

反过来, 若 (x^*, λ^*) 是拉格朗日函数 L 的鞍点, 则

$$\inf_{x \in \mathcal{D}} \sup_{\lambda \geq 0} L(x, \lambda) \leq \sup_{\lambda \geq 0} L(x^*, \lambda) = L(x^*, \lambda^*) = \inf_{x \in \mathcal{D}} L(x, \lambda^*) \leq \sup_{\lambda \geq 0} \inf_{x \in \mathcal{D}} L(x, \lambda)$$

结合极大极小不等式知上式中的不等号全部取等号, 从而 x^* 和 λ^* 分别是原问题和对偶问题的最优解, 且强对偶性成立。

3.3 博弈论

甲和乙两个人进行博弈, 若甲选择 $x \in \mathcal{D}$, 乙选择 $\lambda \geq 0$, 则甲支付给乙的金额为 $L(x, \lambda)$ 。

设甲先做出选择, 乙在知道甲的选择后再选择, 那么乙面对的问题就是

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

甲为了让自己支付的金额尽可能的少, 只能考虑最坏情况, 即

$$\inf_{x \in \mathcal{D}} \sup_{\lambda \geq 0} L(x, \lambda)$$

反过来, 若乙先做出选择, 甲在知道乙的选择后再选择, 那么甲面对的问题就是

$$\inf_{x \in \mathcal{D}} L(x, \lambda)$$

乙为了让自己获得的金额尽可能的多, 只能考虑最坏情况, 即

$$\sup_{\lambda \geq 0} \inf_{x \in \mathcal{D}} L(x, \lambda)$$

弱对偶性表明, 甲先选乙后选时甲需支付的金额大于等于乙先选甲后选时需支付的金额, 因此后选是有优势的。当鞍点性质满足, 即强对偶性成立时, 先选后选对最后支付的金额没有影响。

3.4 影子价格

拉格朗日对偶在经济学上也有一个有趣的解释。设变量 \mathbf{x} 表示公司的某种经营策略, $f_0(\mathbf{x})$ 表示采取策略 \mathbf{x} 时的成本, 约束 $f_i(\mathbf{x})$ 表示一些资源的限制, 比如仓库容量、劳动力、环境等等, 在满足这些限制的情况下最小化成本就是求解如下优化问题:

$$\inf_{\mathbf{x}} f_0(\mathbf{x}), \quad \text{s.t. } f_i(\mathbf{x}) \leq 0, \quad i \in [m]$$

现在假设约束可以被违背, 但需付出一些额外的成本, 成本与违背的量呈线性关系, 即 $\lambda_i f_i(\mathbf{x})$, 其中系数 $\lambda_i \geq 0$ 为第 i 种资源的单位成本, 当然如果 $f_i(\mathbf{x}) < 0$, 公司也会相应得到收益, 因此最终的总成本表达式为

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f_0(\mathbf{x}) + \sum_{i \in [m]} \lambda_i f_i(\mathbf{x})$$

在资源价格为 $\boldsymbol{\lambda}$ 时, 公司的最优总成本就是 $\inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}) = g(\boldsymbol{\lambda})$, 对偶问题最优值 d^* 就是最坏情况下的最优总成本。

称约束不可违背为情形 1, 约束可以违背为情形 2。弱对偶性表明, 情形 2 下的最优总成本不超过情形 1 下的最优总成本。如果 \mathbf{x}^* 是情形 1 下的最优策略, 那么情形 2 下以 \mathbf{x}^* 为经营策略时, 其成本不超过 $f_0(\mathbf{x}^*)$, 这是因为可以通过违背一些约束(付出成本)使得其他约束不紧(获得收益), 差价就产生了对偶间隙。

若强对偶性成立, 对偶问题有最优解 $\boldsymbol{\lambda}^*$ 。当资源价格为 $\boldsymbol{\lambda}^*$ 时, 公司允许约束被违背相比不可违背并不能带来任何优势。基于此, 对偶问题最优解 $\boldsymbol{\lambda}^*$ 有时也称为原问题的影子价格。

4 最优性条件

设 \mathbf{x}^* 和 $\boldsymbol{\lambda}^*$ 分别是原问题和对偶问题的最优解且强对偶性成立, 则

$$f_0(\mathbf{x}^*) = g(\boldsymbol{\lambda}^*) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}^*) = \inf_{\mathbf{x} \in \mathcal{D}} \left\{ f_0(\mathbf{x}) + \sum_{i \in [m]} \lambda_i^* f_i(\mathbf{x}) \right\} \leq f_0(\mathbf{x}^*) + \sum_{i \in [m]} \lambda_i^* f_i(\mathbf{x}^*) \leq f_0(\mathbf{x}^*)$$

于是上式中的不等号全部取等号。第一个不等号取等号表明 $\inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}^*) = L(\mathbf{x}^*, \boldsymbol{\lambda}^*)$, 这在前面鞍点解释中就证明过, 此处不再赘述。第二个不等号取等号表明

$$\sum_{i \in [m]} \lambda_i^* f_i(\mathbf{x}^*) = 0$$

注意 $\lambda_i^* \geq 0$, $f_i(\mathbf{x}^*) \leq 0$, 即每一个求和项都是非正的, 因此只可能是

$$\lambda_i^* f_i(\mathbf{x}^*) = 0, \quad i \in [m]$$

这称为互补松弛条件。

4.1 KKT 条件

假设 f_0, \dots, f_m 可微 (因此定义域是开集), 但不假设它们是凸函数。 x^* 和 λ^* 分别是原问题和对偶问题的最优解且强对偶性成立, 因为 $L(x, \lambda^*)$ 在 x^* 处取得最小值, 故

$$\nabla f_0(x^*) + \sum_{i \in [m]} \lambda_i^* \nabla f_i(x^*) = \mathbf{0}$$

因此

$$\begin{aligned} f_i(x^*) &\leq 0, i \in [m] \\ \lambda_i^* &\geq 0, i \in [m] \\ \lambda_i^* f_i(x^*) &= 0, i \in [m] \\ \nabla f_0(x^*) + \sum_{i \in [m]} \lambda_i^* \nabla f_i(x^*) &= \mathbf{0}, \end{aligned}$$

这称之为 KKT 条件。综上, 对于目标函数和约束函数可微的任意优化问题, 如果强对偶性成立, 那么任意一对原问题和对偶问题的最优解必须满足 KKT 条件, 即这是一个必要条件。

当原问题为凸优化, 即 f_0, \dots, f_m 为凸函数时, KKT 条件就成了充要条件。假设 $\tilde{x}, \tilde{\lambda}$ 是任意满足 KKT 条件的点:

$$\begin{aligned} f_i(\tilde{x}) &\leq 0, i \in [m] \\ \tilde{\lambda}_i &\geq 0, i \in [m] \\ \tilde{\lambda}_i f_i(\tilde{x}) &= 0, i \in [m] \\ \nabla f_0(\tilde{x}) + \sum_{i \in [m]} \tilde{\lambda}_i \nabla f_i(\tilde{x}) &= \mathbf{0}, \end{aligned}$$

其中第一个条件表明 \tilde{x} 是原问题的可行解, 最后一个条件表明 \tilde{x} 是拉格朗日 (凸) 函数 $L(x, \tilde{\lambda})$ 的驻点, 因此 $L(x, \tilde{\lambda})$ 在 \tilde{x} 处取得最小值 (若非凸这里就不成立了), 从而

$$d^* \geq g(\tilde{\lambda}) = \inf_{x \in \mathcal{D}} L(x, \tilde{\lambda}) = f_0(\tilde{x}) + \sum_{i \in [m]} \tilde{\lambda}_i f_i(\tilde{x}) = f_0(\tilde{x}) \geq p^*$$

结合弱对偶性可知上式中的不等号全部取等号, 因此 $\tilde{x}, \tilde{\lambda}$ 分别是原问题和对偶问题的最优解。

综上, 可微凸优化若满足 Slater 条件, 那么 KKT 条件就是最优性的充要条件。Slater 条件意味着强对偶性成立, 从而 KKT 条件是必要条件; 原问题是可微凸优化, 意味着 KKT 条件也是充分条件。此时 x^* 是原问题最优解, 当且仅当存在 λ^* 使得二者满足 KKT 条件。

KKT 条件在优化领域有着重要的作用, 有些时候是可以通过解析求解 KKT 条件来求解优化问题的。

5 择一定理

易知求解不等式组

$$f_i(x) \leq 0, i \in [m] \tag{1}$$

与求解优化问题

$$\inf_{\mathbf{x}} 0, \quad \text{s.t. } f_i(\mathbf{x}) \leq 0, \quad i \in [m]$$

是等价的，因为

$$p^* = \begin{cases} 0 & (1) \text{ feasible} \\ \infty & (1) \text{ infeasible} \end{cases}$$

对偶函数 $g(\lambda) = \inf_{\mathbf{x} \in \mathcal{D}} \sum_{i \in [m]} \lambda_i f_i(\mathbf{x})$ ，显然零是可以取到的，只需令 $\lambda = \mathbf{0}$ 即可；若

$$\lambda \geq \mathbf{0}, \quad g(\lambda) > 0 \tag{2}$$

可行，则 $g(\lambda)$ 无上界，因为 $g(\lambda)$ 关于 λ 是正齐次的，即当 $\alpha > 0$ 时有 $g(\alpha\lambda) = \alpha g(\lambda)$ ，故对偶问题最优值为

$$d^* = \begin{cases} \infty & (2) \text{ feasible} \\ 0 & (2) \text{ infeasible} \end{cases}$$

若不等式组 (1) 可行，则 $p^* = 0$ ，由弱对偶性知 $d^* \leq 0$ ，故不等式组 (2) 不可行。若不等式组 (2) 可行，则 $d^* = \infty$ ，由弱对偶性知 $p^* = \infty$ ，故不等式组 (1) 不可行。综上，不等式组 (1) 和不等式组 (2) 最多只有一个可行，这称为弱择一。注意这与不等式组 (1) 是否是凸的无关，此外不等式组 (2) 一定是凸的。

5.1 严格不等式的情形

考虑严格不等式组

$$f_i(\mathbf{x}) < 0, \quad i \in [m] \tag{3}$$

同样依靠对偶函数可得不等式组

$$\lambda \geq \mathbf{0}, \quad \lambda \neq \mathbf{0}, \quad g(\lambda) \geq 0 \tag{4}$$

这两个不等式组也是弱择一的，假设存在 $\tilde{\mathbf{x}}$ 满足不等式组 (3)，那么对任意 $\lambda \geq \mathbf{0}$ 且 $\lambda \neq \mathbf{0}$ 有

$$\lambda_1 f_1(\tilde{\mathbf{x}}) + \dots + \lambda_m f_m(\tilde{\mathbf{x}}) < 0$$

从而

$$g(\lambda) = \inf_{\mathbf{x} \in \mathcal{D}} \sum_{i \in [m]} \lambda_i f_i(\mathbf{x}) \leq \sum_{i \in [m]} \lambda_i f_i(\tilde{\mathbf{x}}) < 0$$

所以这两个不等式组是弱择一的。

5.2 强择一

依然考虑严格不等式组 (3) 和不等式组 (4)，现假设不等式组 (3) 是凸的，即 f_i 都是凸函数，且存在一点 $\tilde{x} \in \text{relint}(\mathcal{D})$ ，下面证明不等式组 (3) 和不等式组 (4) 有且仅有一个可行，这称为强择一。

易知不等式组 (3) 可行当且仅当优化问题

$$\inf_{x \in \mathcal{D}, s} s, \quad \text{s.t. } f_i(x) - s \leq 0, \quad i \in [m] \quad (5)$$

最优值 $p^* < 0$ 。对偶函数为

$$\inf_{x \in \mathcal{D}, s} \left\{ s + \sum_{i \in [m]} \lambda_i (f_i(x) - s) \right\} = \inf_{x \in \mathcal{D}, s} \left\{ s(1 - e^\top \lambda) + \sum_{i \in [m]} \lambda_i f_i(x) \right\} = \begin{cases} g(\lambda) & e^\top \lambda = 1 \\ -\infty & \text{o.w.} \end{cases}$$

故对偶问题为

$$\sup_{\lambda} g(\lambda), \quad \text{s.t. } \lambda \geq 0, \quad e^\top \lambda = 1$$

下面说明优化问题 (5) 满足 Slater 条件，由于存在一点 $\tilde{x} \in \text{relint}(\mathcal{D})$ ，任选 $\tilde{s} > \max_i f_i(\tilde{x})$ ，则点 (\tilde{x}, \tilde{s}) 是优化问题 (5) 的严格可行点，因此强对偶性成立，且对偶最优值 d^* 可以取到 ($d^* = p^* \neq -\infty$)，故存在 λ^* 使得

$$g(\lambda^*) = p^*, \quad \lambda^* \geq 0, \quad e^\top \lambda^* = 1$$

若不等式组 (3) 不可行，则 $g(\lambda^*) = p^* \geq 0$ ，从而不等式组 (4) 可行。若不等式组 (4) 可行，则 $p^* = d^* \geq 0$ ，从而不等式组 (3) 不可行。综上，这两个不等式组是强择一的。

下面考虑非严格不等式组

$$f_i(x) \leq 0, \quad i \in [m] \quad (6)$$

和不等式组

$$\lambda \geq 0, \quad g(\lambda) \geq 0 \quad (7)$$

假设存在一点 $\tilde{x} \in \text{relint}(\mathcal{D})$ 且优化问题 (5) 最优值 p^* 可以取到 ($p^* \neq \infty$)，则强对偶性成立。若不等式组 (6) 不可行，这意味着 $d^* = p^* > 0$ ，从而不等式组 (7) 可行。若不等式组 (7) 可行，则 $p^* = d^* \geq 0$ ，从而不等式组 (6) 不可行。

5.3 Farkas 引理

考虑线性规划

$$\inf_x c^\top x, \quad \text{s.t. } Ax \leq 0$$

其中 $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, 其对偶问题为

$$\sup_{\lambda} 0, \quad \text{s.t. } A^\top \lambda + c = 0, \quad \lambda \geq 0$$

显然 Slater 条件满足，强对偶性成立。若对偶问题有解，即不等式组

$$\mathbf{A}^\top \lambda + \mathbf{c} = \mathbf{0}, \quad \lambda \geq \mathbf{0} \quad (8)$$

可行，则 $\mathbf{c}^\top \mathbf{x}^* = 0$ 。综上不等式组 (8) 与不等式组

$$\mathbf{c}^\top \mathbf{x} < 0, \quad \mathbf{A}\mathbf{x} \leq \mathbf{0}$$

是强择一的。

6 广义不等式约束

考虑带广义不等式约束的优化问题

$$\inf_{\mathbf{x}} f_0(\mathbf{x}), \quad \text{s.t. } f_i(\mathbf{x}) \preceq_{K_i} \mathbf{0}, \quad i \in [m]$$

其中 $f_i : \mathbb{R}^n \mapsto \mathbb{R}^{k_i}$, $K_i \in \mathbb{R}^{k_i}$ 是正常锥，其余同前，同样这里并不要求问题是凸优化。

6.1 弱对偶与强对偶

引入拉格朗日乘子 $\lambda_i \in \mathbb{R}^{k_i}$, 并定义拉格朗日函数如下：

$$L(\mathbf{x}, \lambda_i) = f_0(\mathbf{x}) + \sum_{i \in [m]} \lambda_i^\top f_i(\mathbf{x})$$

之前数值不等式情形下要求拉格朗日乘子 $\lambda_i \geq 0$, 这里需要替换成

$$\lambda_i \succeq_{K_i^*} \mathbf{0}, \quad i \in [m]$$

其中 K_i^* 是 K_i 的对偶锥。设 $\tilde{\mathbf{x}}$ 是原问题一个可行点，即 $f_i(\tilde{\mathbf{x}}) \preceq_{K_i} \mathbf{0}$, 于是由对偶锥的定义可知

$$L(\tilde{\mathbf{x}}, \lambda_i) = f_0(\tilde{\mathbf{x}}) + \sum_{i \in [m]} \lambda_i^\top f_i(\tilde{\mathbf{x}}) \leq f_0(\tilde{\mathbf{x}})$$

从而同前有

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

由于上式对任意可行点 $\tilde{\mathbf{x}}$ 都成立，因此有 $g(\lambda_i) \leq p^*$, 即对偶函数构成了原问题最优值 p^* 的一个下界，于是弱对偶性成立

$$d^* = \sup_{\lambda_i} g(\lambda_i) \leq p^*$$

同前若 $d^* = p^*$, 则称强对偶性成立。在原问题是凸优化的情况下，即 f_0 是凸函数， f_i 是 K_i -凸函数，有广义 Slater 条件：若存在一点 $\mathbf{x} \in \text{relint}(\mathcal{D})$ 使得所有不等式约束严格成立，即

$$f_i(\mathbf{x}) \prec_{K_i} \mathbf{0}, \quad i \in [m]$$

那么强对偶性成立（且可以取到对偶最优）。

下面看几个例子：

例 3 (半定规划).

$$\inf_x \mathbf{c}^\top \mathbf{x}, \quad \text{s.t. } x_1 \mathbf{A}_1 + \cdots + x_m \mathbf{A}_m + \mathbf{G} \preceq \mathbf{0}$$

其中 $\mathbf{A}_1, \dots, \mathbf{A}_m, \mathbf{G} \in \mathbb{S}^k$ 是给定对称矩阵, 这里 \preceq 指的是半正定锥。

注意半正定锥是自对偶锥, 所以拉格朗日乘子是属于半正定锥的矩阵 \mathbf{Z} , 拉格朗日函数为

$$\begin{aligned} L(\mathbf{x}, \mathbf{Z}) &= \mathbf{c}^\top \mathbf{x} + \langle x_1 \mathbf{A}_1 + \cdots + x_m \mathbf{A}_m + \mathbf{G}, \mathbf{Z} \rangle \\ &= \langle \mathbf{G}, \mathbf{Z} \rangle + \mathbf{c}^\top \mathbf{x} + \sum_{i \in [m]} x_i \langle \mathbf{A}_i, \mathbf{Z} \rangle \\ &= \langle \mathbf{G}, \mathbf{Z} \rangle + \sum_{i \in [m]} x_i (\langle \mathbf{A}_i, \mathbf{Z} \rangle + c_i) \end{aligned}$$

从而对偶问题是

$$\sup_{\mathbf{Z}} \langle \mathbf{G}, \mathbf{Z} \rangle, \quad \text{s.t. } \mathbf{Z} \succeq \mathbf{0}, \langle \mathbf{A}_i, \mathbf{Z} \rangle + c_i = 0, i \in [m]$$

若原问题是严格可行的, 即 $\exists \mathbf{x}$ 满足 $x_1 \mathbf{A}_1 + \cdots + x_n \mathbf{A}_n + \mathbf{G} \prec \mathbf{0}$, 则强对偶性成立。

例 4 (标准形式的锥规划).

$$\inf_x \mathbf{c}^\top \mathbf{x}, \quad \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \succeq_K \mathbf{0}$$

其中 $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $K \subseteq \mathbb{R}^n$ 是一个正常锥。

引入拉格朗日乘子 $\nu \in \mathbb{R}^m$ 和 $\lambda \succeq_{K^*} \mathbf{0}$, 拉格朗日函数为

$$L(\mathbf{x}, \nu, \lambda) = \mathbf{c}^\top \mathbf{x} - \lambda^\top \mathbf{x} + \nu^\top (\mathbf{A}\mathbf{x} - \mathbf{b}) = -\mathbf{b}^\top \nu + (\mathbf{A}^\top \nu - \lambda + \mathbf{c})^\top \mathbf{x}$$

从而对偶问题是

$$\sup_{\nu, \lambda} -\mathbf{b}^\top \nu, \quad \text{s.t. } \lambda \succeq_{K^*} \mathbf{0}, \mathbf{A}^\top \nu + \mathbf{c} = \lambda$$

令 $y = -\nu$ 并消去 λ 可得

$$\sup_y \mathbf{b}^\top y, \quad \text{s.t. } \mathbf{A}^\top y \preceq_{K^*} \mathbf{c}$$

若 Slater 条件成立, 即 $\exists \mathbf{x}$ 满足 $\mathbf{x} \succ_K \mathbf{0}$ 且 $\mathbf{A}\mathbf{x} = \mathbf{b}$, 则强对偶性成立。

同理不难证明不等式形式的锥规划 $\inf_{\mathbf{A}\mathbf{x} \leq_K \mathbf{b}} \mathbf{c}^\top \mathbf{x}$ 的对偶问题为

$$\sup_\lambda -\mathbf{b}^\top \lambda, \quad \text{s.t. } \mathbf{A}^\top \lambda + \mathbf{c} = \mathbf{0}, \lambda \succeq_{K^*} \mathbf{0}$$

例 5 (二阶锥规划).

$$\inf_x \mathbf{f}^\top \mathbf{x}, \quad \text{s.t. } \|\mathbf{A}_i \mathbf{x} + \mathbf{b}_i\|_2 \leq \mathbf{c}_i^\top \mathbf{x} + d_i, i \in [m]$$

其中 $\mathbf{f} \in \mathbb{R}^n$, $\mathbf{A}_i \in \mathbb{R}^{n_i \times n}$, $\mathbf{b}_i \in \mathbb{R}^{n_i}$, $\mathbf{c}_i \subseteq \mathbb{R}^n$, $d_i \in \mathbb{R}$ 。

显然原问题等价于

$$\inf_{\mathbf{x}} \mathbf{f}^\top \mathbf{x}, \quad \text{s.t. } \|\mathbf{y}_i\|_2 \leq z_i, \quad \mathbf{A}_i \mathbf{x} + \mathbf{b}_i = \mathbf{y}_i, \quad \mathbf{c}_i^\top \mathbf{x} + d_i = z_i, \quad i \in [m]$$

引入拉格朗日乘子 $\lambda_i \geq 0$, μ_i 和 ν_i , 可得对偶函数

$$\begin{aligned} g(\lambda_i, \mu_i, \nu_i) &= \inf_{\mathbf{x}, \mathbf{y}_i, z_i} L(\mathbf{x}, \mathbf{y}_i, z_i, \lambda_i, \mu_i, \nu_i) \\ &= \inf_{\mathbf{x}, \mathbf{y}_i, z_i} \left\{ \mathbf{f}^\top \mathbf{x} + \sum_{i \in [m]} \lambda_i (\|\mathbf{y}_i\|_2 - z_i) - \sum_{i \in [m]} \mu_i^\top (\mathbf{y}_i - \mathbf{A}_i \mathbf{x} - \mathbf{b}_i) + \sum_{i \in [m]} \nu_i (z_i - \mathbf{c}_i^\top \mathbf{x} - d_i) \right\} \\ &= \inf_{\mathbf{x}, \mathbf{y}_i, z_i} \left\{ \left(\mathbf{f} + \sum_{i \in [m]} (\mathbf{A}_i^\top \mu_i - \nu_i \mathbf{c}_i) \right)^\top \mathbf{x} + \sum_{i \in [m]} (\lambda_i \|\mathbf{y}_i\|_2 - \mu_i^\top \mathbf{y}_i) + \sum_{i \in [m]} (\nu_i - \lambda_i) z_i \right. \\ &\quad \left. + \sum_{i \in [m]} (\mu_i^\top \mathbf{b}_i - \nu_i d_i) \right\} \end{aligned}$$

关于 \mathbf{x} 极小化可得 $\mathbf{f} + \sum_{i \in [m]} (\mathbf{A}_i^\top \mu_i - \nu_i \mathbf{c}_i) = \mathbf{0}$; 关于 z_i 极小化可得 $\nu_i - \lambda_i = 0$; 关于 \mathbf{y}_i 极小化, 由于 $\lambda_i \|\mathbf{y}_i\|_2 - \mu_i^\top \mathbf{y}_i \geq \lambda_i \|\mathbf{y}_i\|_2 - \|\mu_i\|_2 \|\mathbf{y}_i\|_2 = (\lambda_i - \|\mu_i\|_2) \|\mathbf{y}_i\|_2$, 故

$$\inf_{\mathbf{y}} \{\lambda_i \|\mathbf{y}_i\|_2 - \mu_i^\top \mathbf{y}_i\} = \inf_{\mathbf{y}} \{(\lambda_i - \|\mu_i\|_2) \|\mathbf{y}_i\|_2\} = \begin{cases} 0 & \|\mu_i\|_2 \leq \lambda_i \\ -\infty & \text{o.w.} \end{cases}$$

综上, 对偶问题是

$$\sup_{\lambda_i, \mu_i, \nu_i} \sum_{i \in [m]} (\mu_i^\top \mathbf{b}_i - \nu_i d_i), \quad \text{s.t. } \mathbf{f} + \sum_{i \in [m]} (\mathbf{A}_i^\top \mu_i - \nu_i \mathbf{c}_i) = \mathbf{0}, \quad \nu_i - \lambda_i = 0, \quad \|\mu_i\|_2 \leq \lambda_i, \quad i \in [m]$$

消去 λ_i 可得

$$\sup_{\mu_i, \nu_i} \sum_{i \in [m]} (\mathbf{b}_i^\top \mu_i - d_i \nu_i), \quad \text{s.t. } \sum_{i \in [m]} (\mathbf{A}_i^\top \mu_i - \nu_i \mathbf{c}_i) + \mathbf{f} = \mathbf{0}, \quad \|\mu_i\|_2 \leq \nu_i, \quad i \in [m]$$

6.2 最优性条件

设 \mathbf{x}^* 和 $\lambda_1^*, \dots, \lambda_m^*$ 分别是原问题和对偶问题的最优解且强对偶性成立, 同前有

$$f_0(\mathbf{x}^*) = g(\lambda_i^*) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \lambda_i^*) \leq f_0(\mathbf{x}^*) + \sum_{i \in [m]} \lambda_i^* f_i(\mathbf{x}^*) \leq f_0(\mathbf{x}^*)$$

于是有互补松弛条件

$$\lambda_i^* f_i(\mathbf{x}^*) = 0, \quad i \in [m]$$

注意与数值不等式不同的是, 可能存在 $\lambda_i^* \neq \mathbf{0}$ 和 $f_i(\mathbf{x}^*) \neq \mathbf{0}$ 满足此条件。

假设 f_0, f_1, \dots, f_m 可微, 同前有 KKT 条件

$$f_i(\mathbf{x}^*) \preceq_{K_i} \mathbf{0}, \quad i \in [m]$$

$$\begin{aligned}\lambda_i^* &\succeq_{K_i^*} \mathbf{0}, i \in [m] \\ \lambda_i^* f_i(\mathbf{x}^*) &= 0, i \in [m] \\ \nabla f_0(\mathbf{x}^*) + \sum_{i \in [m]} Df_i(\mathbf{x}^*)^\top \lambda_i^* &= \mathbf{0},\end{aligned}$$

其中 $Df_i(\mathbf{x}) \in \mathbb{R}^{k_i \times n}$ 是 $f(\mathbf{x})$ 在 \mathbf{x} 处的 Jacobian 矩阵。当原问题为凸优化时，KKT 条件变成最优性的充要条件。

6.3 择一定理

通过优化问题

$$\inf_{\mathbf{x}} 0, \quad \text{s.t. } f_i(\mathbf{x}) \preceq_{K_i} \mathbf{0}, i \in [m]$$

不难证明不等式组

$$f_i(\mathbf{x}) \preceq_{K_i} \mathbf{0}, i \in [m]$$

与不等式组

$$\lambda_i \succeq_{K_i^*} \mathbf{0}, i \in [m], \quad g(\lambda_i) > 0$$

是弱择一的。

严格不等式组

$$f_i(\mathbf{x}) \prec_{K_i} \mathbf{0}, i \in [m] \tag{9}$$

与不等式组

$$\lambda_i \succeq_{K_i^*} \mathbf{0}, i \in [m], \quad [\lambda_1, \dots, \lambda_m] \neq \mathbf{0}, \quad g(\lambda_i) \geq 0 \tag{10}$$

是弱择一的。

若不等式组 (9) 是凸的，即 f_i 都是 K_i -凸函数，且存在一点 $\tilde{\mathbf{x}} \in \text{relint}(\mathcal{D})$ ，通过优化问题

$$\inf_{\mathbf{x} \in \mathcal{D}, s} s, \quad \text{s.t. } f_i(\mathbf{x}) \preceq s e_i, i \in [m] \tag{11}$$

其中 $e_i \succ_{K_i} \mathbf{0}$ 是任意固定向量，可知不等式组 (9) 与不等式组 (10) 也是强择一的。

此外，若优化问题 (11) 的最优值还能够取到，则非严格不等式组

$$f_i(\mathbf{x}) \preceq_{K_i} \mathbf{0}, i \in [m]$$

和不等式组

$$\lambda_i \succeq_{K_i^*} \mathbf{0}, i \in [m], \quad g(\lambda_i) > 0$$

是强择一的。

例 6 (线性矩阵不等式的强择一). 设 $\mathbf{A}_1, \dots, \mathbf{A}_m, \mathbf{G} \in \mathbb{S}^k$, 严格线性矩阵不等式

$$x_1\mathbf{A}_1 + \dots + x_m\mathbf{A}_m + \mathbf{G} \prec \mathbf{0}$$

和不等式组

$$\mathbf{Z} \succeq \mathbf{0}, \quad \mathbf{Z} \neq \mathbf{0}, \quad \langle \mathbf{G}, \mathbf{Z} \rangle \geq 0, \quad \langle \mathbf{A}_i, \mathbf{Z} \rangle = 0, \quad i \in [m]$$

是强择一的, 对比不等式组 (9) 和不等式组 (10) 易知。

对于非严格不等式的情形要复杂一些, 需要额外条件:

$$\sum_{i \in [m]} v_i \mathbf{A}_i \succeq \mathbf{0} \implies \sum_{i \in [m]} v_i \mathbf{A}_i = \mathbf{0}$$

才能保证非严格线性矩阵不等式

$$x_1\mathbf{A}_1 + \dots + x_m\mathbf{A}_m + \mathbf{G} \preceq \mathbf{0}$$

和不等式组

$$\mathbf{Z} \succeq \mathbf{0}, \quad \langle \mathbf{G}, \mathbf{Z} \rangle > 0, \quad \langle \mathbf{A}_i, \mathbf{Z} \rangle = 0, \quad i \in [m]$$

是强择一的。

只需证明优化问题

$$\inf_{x,s} s, \quad \text{s.t. } x_1\mathbf{A}_1 + \dots + x_m\mathbf{A}_m + \mathbf{G} \preceq s\mathbf{I}$$

的最优值能够取到。

显然 $x_1\mathbf{A}_1 + \dots + x_m\mathbf{A}_m + \mathbf{G}$ 的最大特征值有限, 因此 s 有限。用反证法, 假设最优值 s^* 有限但取不到, 于是存在序列 $\{\mathbf{x}^{(k)}\}, \{s^{(k)}\} \rightarrow s^*$ 满足 $x_1^{(k)}\mathbf{A}_1 + \dots + x_m^{(k)}\mathbf{A}_m + \mathbf{G} \preceq s^{(k)}\mathbf{I}$, 两边同除以 $\|\mathbf{x}^{(k)}\|_2$ 可得

$$\frac{x_1^{(k)}}{\|\mathbf{x}^{(k)}\|_2}\mathbf{A}_1 + \dots + \frac{x_m^{(k)}}{\|\mathbf{x}^{(k)}\|_2}\mathbf{A}_m + \frac{1}{\|\mathbf{x}^{(k)}\|_2}\mathbf{G} \preceq \frac{s^{(k)}}{\|\mathbf{x}^{(k)}\|_2}\mathbf{I}$$

若 $\|\mathbf{x}^{(k)}\|_2 \rightarrow \infty$, $1/\|\mathbf{x}^{(k)}\|_2 \rightarrow 0$, $s^{(k)}/\|\mathbf{x}^{(k)}\|_2 \rightarrow 0$, 又 $\{x_i^{(k)}/\|\mathbf{x}^{(k)}\|_2\}$ 有界, 因此存在收敛的子序列, 设该子序列的极限为 v_i , 则有 $v_1\mathbf{A}_1 + \dots + v_m\mathbf{A}_m \preceq \mathbf{0}$, 由条件可知 $v_1 = \dots = v_m = 0$, 这与 $\|\mathbf{x}^{(k)}\|_2 \rightarrow \infty$ 矛盾, 故 $\|\mathbf{x}^{(k)}\|_2$ 有界, 从而存在收敛的子序列, 设该子序列的极限为 \tilde{x} , 于是

$$\tilde{x}_1\mathbf{A}_1 + \dots + \tilde{x}_m\mathbf{A}_m + \mathbf{G} \preceq s^*\mathbf{I}$$

从而最优解可以取到。

最优值能够取到这个条件是必需的, 看一个反例, 设 $m = 1$, 且

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad x_1\mathbf{A}_1 + \mathbf{G} = \begin{bmatrix} 0 & 1 \\ 1 & x_1 \end{bmatrix}$$

易知

$$|x_1\mathbf{A}_1 + \mathbf{G} - \lambda\mathbf{I}| = \begin{vmatrix} -\lambda & 1 \\ 1 & x_1 - \lambda \end{vmatrix} = \lambda^2 - x_1\lambda - 1$$

故 $x_1\mathbf{A}_1 + \mathbf{G}$ 的两个特征值的乘积为 -1 , 从而异号, 因此 $x_1\mathbf{A}_1 + \mathbf{G} \preceq \mathbf{0}$ 无解。

另一个不等式组为

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \succeq \mathbf{0}, \quad z_{12} = z_{21}, \quad \langle \mathbf{G}, \mathbf{Z} \rangle = \text{tr} \left(\begin{bmatrix} z_{21} & z_{22} \\ z_{11} & z_{12} \end{bmatrix} \right) > 0, \quad \langle \mathbf{A}_1, \mathbf{Z} \rangle = \text{tr} \left(\begin{bmatrix} 0 & 0 \\ z_{21} & z_{22} \end{bmatrix} \right) = 0$$

其中第二三两个式子等价于 $z_{12} > 0$, 第四个式子等价于 $z_{22} = 0$, 回代入第一个式子易知有

$$|\mathbf{Z} - \lambda\mathbf{I}| = \begin{vmatrix} z_{11} - \lambda & z_{12} \\ z_{12} & -\lambda \end{vmatrix} = \lambda^2 - z_{11}\lambda - z_{12}^2$$

由 $z_{12} > 0$ 知 \mathbf{Z} 的两个特征值的乘积为负数, 从而异号, 因此 $\mathbf{Z} \succeq \mathbf{0}$ 无解。