

第十三讲：对偶

优化问题的对偶性

杨 林

大 纲

1. 拉格朗日对偶问题

2. 弱对偶和强对偶

大 纲

1. 拉格朗日对偶问题

2. 弱对偶和强对偶

1 拉格朗日对偶问题

- 拉格朗日函数：给定标准形式问题（不一定是凸的）

最小化	$f_0(x)$
约束条件	$f_i(x) \leq 0, \quad i = 1, 2, \dots, m$
	$h_i(x) = 0, \quad i = 1, 2, \dots, p$

变量 $x \in \mathbb{R}^n$, 定义域 \mathcal{D} , 最优值 p^*

1 拉格朗日对偶问题

■ 拉格朗日函数 (Lagrange) :

$L: \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$, 其中 $\text{dom } L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$

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

- 目标函数和约束函数的加权和
- λ_i 是与 $f_i(x) \leq 0$ 相关的拉格朗日乘子
- v_i 是与 $h_i(x) = 0$ 相关的拉格朗日乘子

1 拉格朗日对偶问题

■ **拉格朗日对偶函数**: $g: \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$

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

$$= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p v_i h_i(x) \right)$$

■ g 是凹函数

■ **下界性质**: 如果 $\lambda \geq 0$, 那么 $g(\lambda, v) \leq p^*$

□ **证明**: 如果 \tilde{x} 是可行的且 $\lambda \geq 0$, 那么

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

对所有的可行的 \tilde{x} 进行最小化得到 $p^* \geq g(\lambda, v)$

1 拉格朗日对偶问题

■ 线性方程的最小范数解

最小化	$x^T x$
约束条件	$Ax = b$

□ 对偶方程:

1. 拉格朗日函数是 $L(x, v) = x^T x + v^T (Ax - b)$
2. 要在 x 上最小化 L , 设梯度等于零:

$$\nabla_x L(x, v) = 2x + A^T v = 0 \Rightarrow x = -(1/2)A^T v$$

代入 L 中得到 g :

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

v 的凹函数

□ 下界性质: $p^* \geq -(1/4)v^T A A^T v - b^T v$ 对于所有 v

1 拉格朗日对偶问题

■ 标准形式 LP

$$\begin{array}{ll} \text{最小化} & c^T x \\ \text{约束条件} & Ax = b, \quad x \geq 0 \end{array}$$

1. 拉格朗日函数是

$$\begin{aligned} L(x, \lambda, \nu) &= c^T x + \nu^T (Ax - b) - \lambda^T x \\ &= -b^T \nu + (c + A^T \nu - \lambda)^T x \end{aligned}$$

2. L 是 x 的仿射函数, 因此

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

g 在仿射域 $\{(\lambda, \nu) | A^T \nu - \lambda + c = 0\}$ 上是线性的, 因此是凹的

□ 下界性质: $p^* \geq -b^T \nu$ 对于所有 ν 若 $A^T \nu + c \geq 0$

1 拉格朗日对偶问题

■ 等式约束范数最小化

最小化	$\ x\ $
约束条件	$Ax = b$

□ 对偶方程:

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

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

1 拉格朗日对偶问题

■ 等式约束范数最小化(续)

□ 证明: 分析 $\inf_x (\|x\| - y^T x)$

首先考虑情况 (1) $x \neq 0$, 则 $\inf_x (\|x\| - y^T x) = \inf_x (1 - y^T x / \|x\|) \|x\|$. $x / \|x\|$ 可以看作范数恒为1的向量. 则 $\inf_x (1 - y^T x / \|x\|) \|x\| = \inf_x (1 - \sup_{\|u\|=1} y^T u) \|x\| = \inf_x (1 - \|y\|_*) \|x\|$. 如果 $\|y\|_* \leq 1$, 极限为0, 否则为 $-\infty$

(2) 如果 $x = 0$, 不影响 (1) 中的情况

□ 下界性质: $p_* \geq b^T v$ 若 $\|A^T v\|_* \leq 1$

1 拉格朗日对偶问题

■ 两向划分

$$\begin{array}{ll} \text{最小化} & x^T W x \\ \text{约束条件} & x_i^2 = 1, i = 1, \dots, n \end{array}$$

1. 非凸问题;可行集包含 2^n 个离散点
2. 解释: 将 $\{1, \dots, n\}$ 划分为两个集合; W_{ij} 是 i 和 j 分配到同一集合的成本; $-W_{ij}$ 是分配到不同集合的成本

□ 对偶方程:

$$\begin{aligned} g(v) &= \inf_x \left(x^T W x + \sum_i v_i (x_i^2 - 1) \right) \\ &= \inf_x x^T (W + \mathbf{diag}(v)) x - \mathbf{1}^T v \end{aligned}$$

1 拉格朗日对偶问题

■ 两向划分(续)

□ 对偶方程:

$$g(\nu) = \begin{cases} -\mathbf{1}^T \nu & W + \mathbf{diag}(\nu) \geq 0 \\ -\infty & \text{其他情况} \end{cases}$$

□ 下界性质: $p^* \geq -\mathbf{1}^T \nu$ 若 $W + \mathbf{diag}(\nu) \geq 0$

□ 例: $\nu = -\lambda_{\min}(W)\mathbf{1}$ 给出界限 $p^* \geq n\lambda_{\min}(W)$

1 拉格朗日对偶问题

■ 拉格朗日对偶与共轭函数

$$\begin{array}{ll} \text{最小化} & f_0(x) \\ \text{约束条件} & Ax \leq b, \quad Cx = d \end{array}$$

□ 对偶方程:

$$\begin{aligned} g(v) &= \inf_{x \in \text{dom } f_0} \left(f_0(x) + (A^T \lambda + C^T v)^T x - b^T \lambda - d^T v \right) \\ &= - \sup_{x \in \text{dom } f_0} \left(-f_0(x) - (A^T \lambda + C^T v)^T x \right) - b^T \lambda - d^T v \\ &= -f_0^*(-A^T \lambda - C^T v) - b^T \lambda - d^T v \end{aligned}$$

1. 回顾共轭的定义 $f^*(y) = \sup_{x \in \text{dom } f} (y^T x - f(x))$
2. 若已知 f_0 的共轭, 可简化对偶的推导

1 拉格朗日对偶问题

■ 对偶问题

$$\begin{array}{ll} \text{最小化} & g(\lambda, \nu) \\ \text{约束条件} & \lambda \geq 0 \end{array}$$

1. 找到 p^* 的最佳下界, 该下界来自拉格朗日对偶函数
2. 一个凸优化问题; 最优值用 d^* 表示
3. λ, ν 是**对偶可行**的, 如果 $\lambda \geq 0, (\lambda, \nu) \in \mathbf{dom} \, g$
4. 通常通过将隐式约束 $(\lambda, \nu) \in \mathbf{dom} \, g$ 显式化来简化问题

□ 例 1: 标准形式 LP 及其对偶

$$\begin{array}{ll} \text{最小化} & c^T x \\ \text{约束条件} & Ax = b, \\ & x \geq 0 \end{array}$$

$$\begin{array}{ll} \text{最大化} & -b^T x \\ \text{约束条件} & A^T \nu + c \geq 0 \end{array}$$

1 拉格朗日对偶问题

■ 问题的对偶表述:

1. 一个问题的等价表述可能导致非常不同的对偶
2. 当对偶难以推导或无意义时, 重新表述原始问题可能得出有用的对偶问题

■ 常见的重新表述:

1. 引入新的变量和等式约束
2. 将显式约束隐式化或反之
3. 转换目标函数或约束函数, 例如用凸增函数 $\phi(f_0(x))$ 代替 $f_0(x)$

1 拉格朗日对偶问题

□ 例 1: 引入新变量和等式约束:

$$\text{最小化 } f_0(Ax + b)$$

1. 对偶函数是常数: $g = \inf_x L(x) = \inf_x f_0(Ax + b) = p^*$
2. 具有强对偶性, 但对偶问题相当无用

□ 重新表述的问题及其对偶问题:

最小化	$f_0(y)$	最大化	$b^T v - f_0^*(v)$
约束条件	$Ax + b - y = 0$	约束条件	$A^T v = 0$

由此得对偶函数:

$$\begin{aligned} g(v) &= \inf_{x,y} (f_0(y) - v^T y + v^T Ax + b^T v) \\ &= \begin{cases} -f_0^*(v) + b^T v & A^T v = 0 \\ -\infty & \text{其他情况} \end{cases} \end{aligned}$$

1 拉格朗日对偶问题

□ 例 2:范数逼近问题: 最小化 $\|Ax - b\|$

$$\begin{array}{ll} \text{最小化} & \|y\| \\ \text{约束条件} & y = Ax - b \end{array}$$

可以查找范数 $\|\cdot\|$ 的共轭, 或直接推导对偶

$$\begin{aligned} g(v) &= \inf_{x,y} (\|y\| - v^T y + v^T Ax + b^T v) \\ &= \begin{cases} -b^T v + \inf_y (\|y\| + v^T y) & A^T v = 0 \\ -\infty & \text{其他情况} \end{cases} \\ &= \begin{cases} -b^T v & A^T v = 0, \|v\|_* \leq 1 \\ -\infty & \text{其他情况} \end{cases} \end{aligned}$$

1 拉格朗日对偶问题

□ 范数逼近问题: 最小化 $\|Ax - b\|$

■ 范数逼近问题的对偶

$$\begin{array}{ll} \text{最小化} & b^T v \\ \text{约束条件} & A^T v = 0, \|v\|_* \leq 1 \end{array}$$

1 拉格朗日对偶问题

□ 例 3:变换隐式约束

■ 带框约束 LP 问题: 原问题与对偶问题

$$\begin{array}{ll}\text{最小化} & c^T x \\ \text{约束条件} & Ax + b = 0 \\ & -\mathbf{1} \leq x \leq \mathbf{1}\end{array}$$

$$\begin{array}{ll}\text{最大化} & -b^T x - \mathbf{1}^T \lambda_1 - \mathbf{1}^T \lambda_2 \\ \text{约束条件} & c + A^T v + \lambda_1 - \lambda_2 = 0 \\ & \lambda_1 \geq 0, \lambda_2 \geq 0\end{array}$$

1 拉格朗日对偶问题

□ 变换隐式约束(续)

■ 用框约束隐式化改写

$$\begin{array}{ll} \text{最小化} & f_0(x) = \begin{cases} c^T x & -\mathbf{1} \leq x \leq \mathbf{1} \\ \infty & \text{其他情况} \end{cases} \\ \text{约束条件} & Ax = b \end{array}$$

■ 对偶函数

$$\begin{aligned} g(v) &= \inf_{-\mathbf{1} \leq x \leq \mathbf{1}} (c^T x + v^T (Ax - b)) \\ &= -b^T v - \|A^T v + c\|_1 \end{aligned}$$

■ 对偶问题:

$$\begin{array}{ll} \text{最大化} & -b^T v - \|A^T v + c\|_1 \end{array}$$

1 拉格朗日对偶问题

□ 例 4 找出以下分段线性最小化问题的 LP 公式和对偶 LP:

$$\text{最小化} \quad f(x) = \max_{i=1, \dots, m} (a_i^T x + b_i)$$

■ 解: LP 问题是

$$\text{最小化 } t \quad \text{约束条件 } a_i^T x + b_i \leq t, i = 1, \dots, m$$

具有拉格朗日函数:

$$\begin{aligned} L(t, x, \lambda) &= t + \sum_{i=1}^m \lambda_i (a_i^T x + b_i - t) \\ &= (1 - \sum_{i=1}^m \lambda_i) t + \left(\sum_{i=1}^m \lambda_i a_i^T \right) x + \sum_{i=1}^m b_i^T \lambda_i \end{aligned}$$

1 拉格朗日对偶问题

■ 解(续): 那么

$$\begin{aligned} g(\lambda) &= \inf_{t,x} \left(1 - \sum_{i=1}^m \lambda_i \right) t + \left(\sum_{i=1}^m \lambda_i a_i^T \right) x + \sum_{i=1}^m b_i^T \lambda_i \\ &= \sum_{i=1}^m b_i^T \lambda_i + \inf_t \left(1 - \sum_{i=1}^m \lambda_i \right) t + \inf_x \left(\sum_{i=1}^m \lambda_i a_i^T \right) x \end{aligned}$$

若 $\sum_{i=1}^m \lambda_i a_i^T \neq 0$, $g(\lambda) = -\infty$; 若 $1 - \sum_{i=1}^m \lambda_i \neq 0$, $g(\lambda) = -\infty$. 那么, 对偶问题是:

最大化

约束条件

$$\begin{aligned} &\sum_{i=1}^m b_i^T \lambda_i \\ &\sum_{i=1}^m \lambda_i a_i^T = 0 \\ &\sum_{i=1}^m \lambda_i = 1 \\ &\lambda_i \geq 0, \quad i = 1, \dots, m \end{aligned}$$

大 纲

1. 拉格朗日对偶问题

2. 弱对偶和强对偶

2 弱对偶和强对偶

■ 弱对偶: $d^* \leq p^*$

1. 始终成立(对凸和非凸问题)
2. 可用于为困难问题找到非平凡的下界, 例如求解

$$\begin{array}{ll} \text{最大化} & -\mathbf{1}^T \boldsymbol{v} \\ \text{约束条件} & W + \mathbf{diag}(\boldsymbol{v}) \geq 0 \end{array}$$

给出双向划分问题的下界

■ 强对偶: $d^* = p^*$

1. 一般不成立
2. (通常)适用于凸问题
3. 保证凸问题强对偶性的条件被称为**约束准则**

2 弱对偶和强对偶

■ Slater约束准则

对于以下凸问题

$$\begin{array}{ll}\text{最小化} & f_0(x) \\ \text{约束条件} & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & Ax = b\end{array}$$

如果它是严格可行的, 即

$$\exists x \in \text{int } \mathcal{D} : f_i(x) < 0, i = 1, \dots, m, Ax = b$$

强对偶性成立

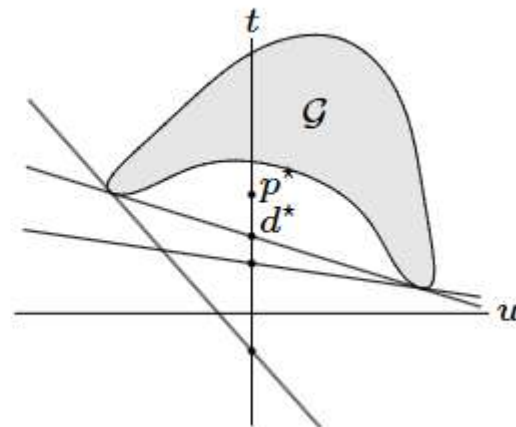
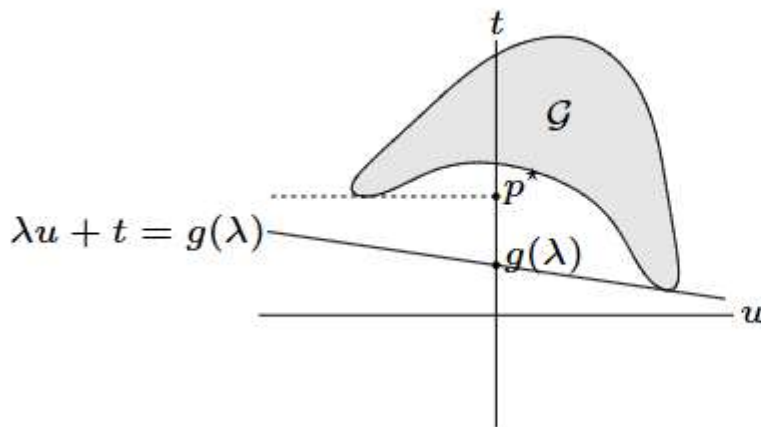
■ 存在许多其他类型的约束准则

2 弱对偶和强对偶

为简化起见, 考虑具有一个约束条件 $f_1(x) \leq 0$ 的问题

■ 对偶函数的几何解释:

$$g(\lambda) = \inf_{(u,t) \in \mathcal{G}} (t + \lambda u), \mathcal{G} = \{(f_1(x), f_0(x)) | x \in \mathcal{D}\}$$



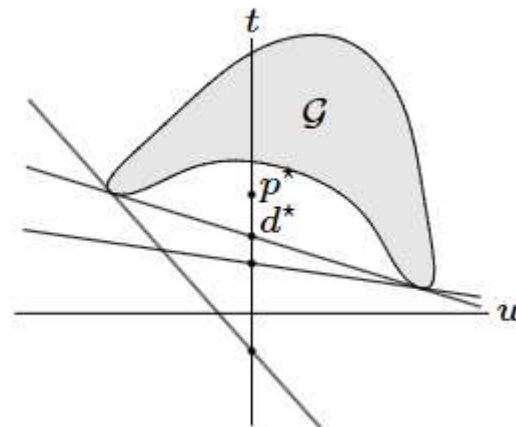
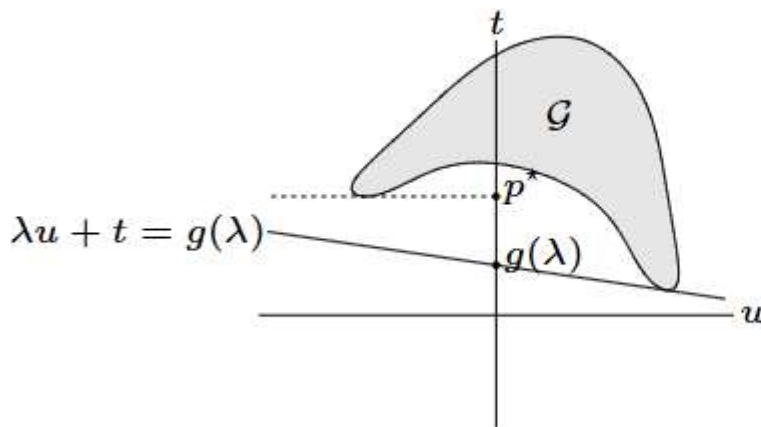
1. $\lambda u + t = g(\lambda)$ 是 \mathcal{G} 的 (非垂直) 支撑超平面
2. 超平面与 t 轴相交于 $t = g(\lambda)$

2 弱对偶和强对偶

为简化起见, 考虑具有一个约束条件 $f_1(x) \leq 0$ 的问题

■ 对偶函数的几何解释:

$$g(\lambda) = \inf_{(u,t) \in \mathcal{G}} (t + \lambda u), \mathcal{G} = \{(f_1(x), f_0(x)) | x \in \mathcal{D}\}$$

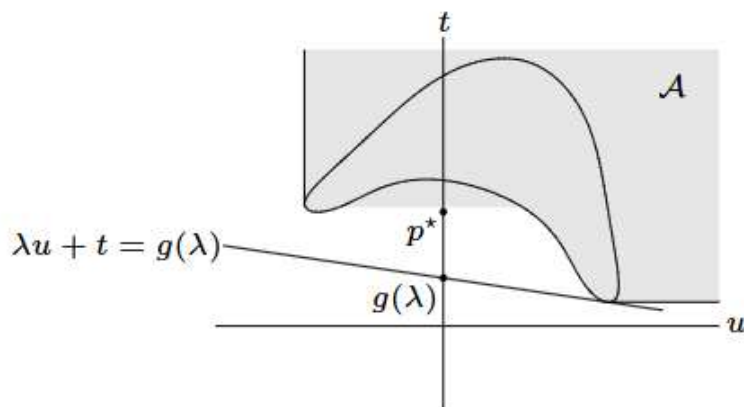


1. $g(\lambda)$ 一定在 p^* 下方
2. d^* 一定在 p^* 下方

2 弱对偶和强对偶

- 上镜图解释：如果用上镜图 \mathcal{A} 替换 \mathcal{G} ，解释相同(λ 非负)

$$\mathcal{A} = \{(u, t) | f_1(x) \leq u, f_0(x) \leq t \text{ 对于某些 } x \in \mathcal{D}\}$$



- 强对偶性:

1. 当 $(0, p^*)$ 处存在 \mathcal{A} 的非垂直支撑超平面时强对偶性才会成立
2. 对于凸问题, \mathcal{A} 是凸的, 因此在 $(0, p^*)$ 处有支撑超平面
3. 斯莱特条件: 如果 $(\tilde{u}, \tilde{t}) \in \mathcal{A}$ 其中 $\tilde{u} < 0$, 那么在 $(0, p^*)$ 处的支持超平面必须是非垂直的

2 弱对偶和强对偶

■ 数学证明:

1. 可行集包含非空内部点;

2. $\text{rank } A = p$ (假设). 定义以下两个互斥的集合:

$$\mathcal{A} = \{(u, v, t) | \exists x \in \mathcal{D}, f_i(x) \leq u_i, h_i(x) = v_i, f_0(x) \leq t\}$$

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

$\exists (\tilde{\lambda}, \tilde{v}, \mu) \neq 0$ 和 α , 使得 (分离超平面定理)

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

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

我们有: 1. $\tilde{\lambda} \geq 0$ 以及 $\mu \geq 0$ (由于上镜图的特性); 2. $\mu p^* \leq \alpha$

换句话说, 存在可行解 $x \in \mathcal{D}$

$$\sum \tilde{\lambda}_i f_i(x) + \tilde{v}^T (Ax - b) + \mu f_0(x) \geq \alpha \geq \mu p^*$$

2 弱对偶和强对偶

情况 1: $\mu > 0$

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

$g(\tilde{\lambda}/\mu, \tilde{v}/\mu) = p^*$, 证毕

情况 2: $\mu = 0$

$$\sum \tilde{\lambda}_i f_i(x) + \tilde{v}^T (Ax - b) \geq 0, \forall x \in \mathcal{D}$$

假设 \tilde{x} 满足Slater约束条件. 那么

$$A\tilde{x} - b = 0, \sum \tilde{\lambda}_i f_i(\tilde{x}) \geq 0 \implies \tilde{\lambda} = 0$$

2 弱对偶和强对偶

因此, $\tilde{v} \neq 0$ (超平面的非零参数), $\tilde{v}^T(Ax - b) \geq 0, \forall x \in \mathcal{D}$

考虑以下事实:

1. $\tilde{v}^T(A\tilde{x} - b) = 0$

2. $\tilde{x} \in \mathbf{int} \mathcal{D}$

除非 $A^T \tilde{v} = 0$, 否则存在 x , 使得 $\tilde{v}^T(Ax - b) < 0$ (**rank** $A = p$, \tilde{v} 是 p 维的), 矛盾

2 弱对偶和强对偶

□ 例 5: 不等式形式 LP

■ 原问题:

$$\begin{array}{ll} \text{最小化} & c^T x \\ \text{约束条件} & Ax \leq b \end{array}$$

■ 对偶函数:

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

2 强对偶与弱对偶

□ 例 5:不等式形式 LP

■ 对偶问题:

$$\begin{array}{ll} \text{最小化} & -b^T \lambda \\ \text{约束条件} & A^T \lambda + c = 0, \lambda \geq 0 \end{array}$$

1. 根据斯莱特条件: $p^* = d^*$,如果 存在 $A\tilde{x} < b$ 对于某些 \tilde{x}
2. 实际上, $p^* = d^*$ 除非原始问题和对偶问题都是不可行的

2 强对偶与弱对偶

□ 例 6: 二次规划

■ 原问题:

$$\begin{array}{ll} \text{最小化} & x^T P x \\ \text{约束条件} & A x \preceq b \end{array}$$

■ 对偶函数:

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

2 弱对偶和强对偶

□ 例 6: 二次规划

■ 对偶问题:

$$\begin{array}{ll} \text{最小化} & -\frac{1}{4}\lambda^T AP^{-1}A^T\lambda - b^T\lambda \\ \text{约束条件} & \lambda \geq 0 \end{array}$$

1. 根据斯莱特条件: $p^* = d^*$,如果 $A\tilde{x} < b$ 对于某些 \tilde{x}
2. 实际上, $p^* = d^*$ 总是成立

2 弱对偶和强对偶

□ 例 7: 一个具有强对偶性的非凸问题*

■ 原问题:

$$\begin{array}{ll} \text{最小化} & x^T A x + 2b^T x \\ \text{约束条件} & x^T x \leq 1 \end{array}$$

$A \not\geq 0$, 所以非凸

■ 对偶函数:

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

1. 若 $A + \lambda I \not\geq 0$ 或若 $A + \lambda I \geq 0$ 且 $b \notin \mathcal{R}(A + \lambda I)$, 则无界下界(以下无界)
2. 由 $x = -(A + \lambda I)^\dagger b$ 最小化, 否则: $g(\lambda) = -b^T (A + \lambda I)^\dagger b - \lambda$

2 弱对偶和强对偶

□ 例 7: 一个具有强对偶性的非凸问题*

■ 对偶问题与等价 SDP :

$$\begin{array}{ll} \text{最小化} & -b^T(A + \lambda I)^\dagger b - \lambda \\ \text{满足} & A + \lambda I \geq 0 \\ & b \in \mathcal{R}(A + \lambda I) \end{array}$$

$$\begin{array}{ll} \text{最大化} & -t - \lambda \\ \text{满足} & \begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \end{array}$$

强对偶性, 尽管原始问题不是凸的(不容易证明)

谢谢！