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

3. 例子:强对偶与弱对偶

大 纲

1. 拉格朗日对偶问题

2. 弱对偶和强对偶

3. 例子:强对偶与弱对偶

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	$f_i(x) \le 0,\ i = 1, 2, \dots, m$
	$h_i(x) = 0,\ \ i = 1, 2, \dots, p$

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

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

- 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
- ν_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**： $g\colon\ \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

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

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

- g 是凹函数

- **下界性质**：如果 $\lambda \succeq 0$, 那么 $g(\lambda, \nu) \le p^*$

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

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

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

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$L(x,\nu) = x^T x + \nu^T(Ax - b)$
2. 要在 x 上最小化 L， 设梯度等于零:

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

代入 L 中得到 g :

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

ν 的凹函数

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

1 拉格朗日对偶问题

■ 标准形式 LP

最小化	$c^T x$
约束条件	$Ax = b, \quad x \geq 0$

1. 拉格朗日函数是

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

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 $\{(\lambda,\nu) \mid A^T \nu - \lambda + c = 0\}$ 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:

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

其中 $\|\nu\|_* = \sup_{\|u\| \le 1} u^T \nu$ 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 $\inf_x \left(\|x\| - y^T x\right)$

首先考虑情况（1） $x \neq 0$, 则$\inf_x \left(\|x\| - y^T x\right) = \inf_x \left(1 - y^T x/\|x\|\right) \|x\|$. $x/\|x\|$可以看作为范数恒为1的向量. 则 $\inf_x \left(1 - y^T x/\|x\|\right) \|x\| = \inf_x \left(1 - \sup_{\|u\|=1} y^T u\right) \|x\| = \inf_x (1 - \|y\|_*) \|x\|$. 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:** $p_* \geq b^T \nu$ 若 $\left\|A^T \nu\right\|_* \leq 1$

1 拉格朗日对偶问题

■ 两向划分

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

1. 非凸问题；可行集包含2^n个离散点
2. 解释：将$\{1, \cdots, n\}$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：

$$g(\nu) = \inf_x \left((x^T W x + \sum_i \nu_i (x_i^2 - 1)) \right)$$

$$= \inf_x x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu$$

1 拉格朗日对偶问题

■ 两向划分(续)

□ 对偶方程:

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

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

□ 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

最小化	$f_0(x)$
约束条件	$Ax \preceq b,\ Cx = d$

□ 对偶方程:

$$g(\nu) = \inf_{x \in \mathbf{dom}\, f_0} \left(f_0(x) + \left(A^T\lambda + C^T\nu\right)^T x - b^T\lambda - d^T\nu \right)$$

$$= - \sup_{x \in \mathbf{dom}\, f_0} \left(-f_0(x) - \left(A^T\lambda + C^T\nu\right)^T x \right) - b^T\lambda - d^T\nu$$

$$= - f_0^*(-A^T\lambda - C^T\nu) - b^T\lambda - d^T\nu$$

1. 回顾共轭的定义 $f^*(y) = \sup_{x \in \mathbf{dom}\, f} \left(y^T x - f(x)\right)$
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

最小化	$g(\lambda, \nu)$
约束条件	$\lambda \succeq 0$

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 $\lambda \succeq 0$, $(\lambda, \nu) \in \textbf{dom}\ g$
4. 通常通过将隐式约束 $(\lambda, \nu) \in \textbf{dom}\ g$ 显式化来简化问题

□ 例 1：标准形式 LP 及其对偶

最小化	$c^T x$	最大化	$-b^T x$
约束条件	$Ax = b,$	约束条件	$A^T \nu + c \succeq 0$
	$x \succeq 0$		

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 $\phi\big(f_0(x)\big)$ 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:**

$$\text{最小化} \quad f_0(Ax+b)$$

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	$b^T\nu - f_0^*(\nu)$
约束条件	$Ax+b-y=0$	约束条件	$A^T\nu = 0$

由此得对偶函数:

$$g(\nu) = \inf_{x,y}\left(f_0(y) - \nu^T y + \nu^T Ax + b^T\nu\right)$$

$$= \begin{cases} -f_0^*(\nu) + b^T\nu & A^T\nu = 0 \\ -\infty & \text{其他情况} \end{cases}$$

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

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

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**

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

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

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

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

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**

最小化 $f_0(x)=\begin{cases} c^T x & -\mathbf{1}\preceq x\preceq \mathbf{1} \\ \infty & \text{其他情况} \end{cases}$

约束条件 $Ax=b$

■ 对偶函数

$$g(\nu)=\inf_{-\mathbf{1}\preceq x\preceq \mathbf{1}}\left(c^T x+\nu^T(Ax-b)\right)$$
$$=-b^T\nu-\left\|A^T\nu+c\right\|_1$$

■ **对偶问题:**

最大化 $-b^T\nu-\left\|A^T\nu+c\right\|_1$

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:

$$\text{最小化} \qquad f(x) = \max_{i=1,\cdots,m} \left(a_i^T x + b_i\right)$$

■ **解:** LP 问题是

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

具有拉格朗日函数:

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

$$= (1 - \sum_{i=1}^{m} \lambda_i) t + \left(\sum_{i=1}^{m} \lambda_i a_i^T\right) x + \sum_{i=1}^{m} b_i^T \lambda_i$$

1 拉格朗日对偶问题

■ **解(续):** 那么

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

$$= \sum_{i=1}^{m} b_i^T\lambda_i + \inf_{t}\left(1 - \sum_{i=1}^{m}\lambda_i\right)t + \inf_{x}\left(\sum_{i=1}^{m}\lambda_i a_i^T\right)x$$

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:

$$\begin{array}{ll} \text{最大化} & \sum_{i=1}^{m} b_i^T\lambda_i \\ \text{约束条件} & \sum_{i=1}^{m}\lambda_i a_i^T = 0 \\ & \sum_{i=1}^{m}\lambda_i = 1 \\ & \lambda_i \geq 0,\ i = 1,\cdots,m \end{array}$$

大 纲

1. 拉格朗日对偶问题

2. 弱对偶和强对偶

3. 例子:强对偶与弱对偶

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

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

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

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

如果它是严格可行的, 即

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

强对偶性成立

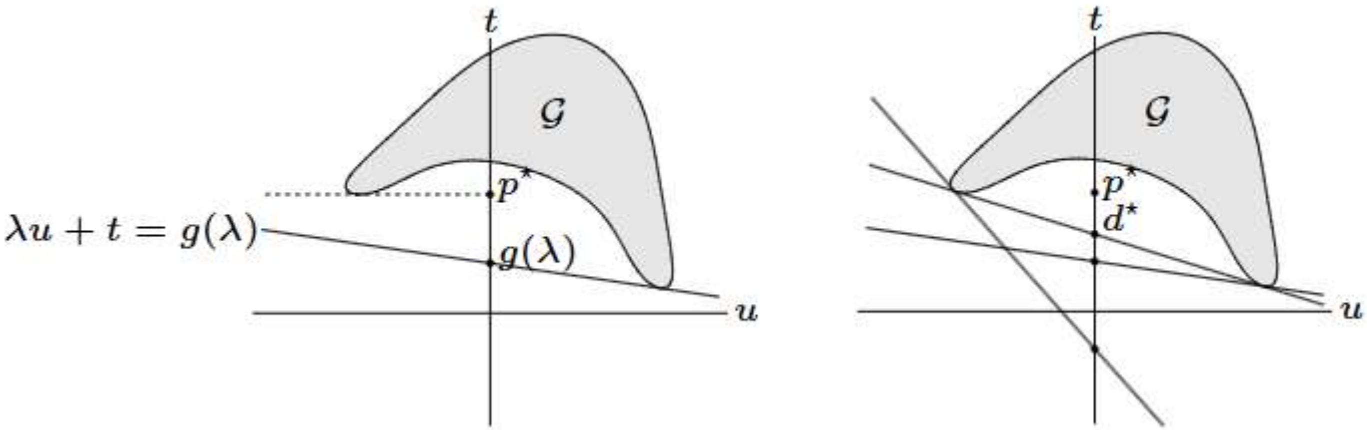
- 存在许多其他类型的约束准则

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：**

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



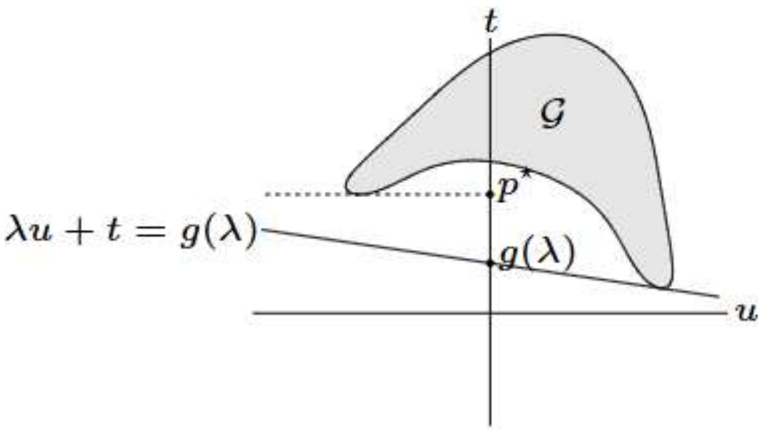
1. $\lambda u + t = g(\lambda)$ 是$\mathcal{G}$的（非垂直）支撑超平面
2. 超平面与 t 轴相交于 $t = g(\lambda)$

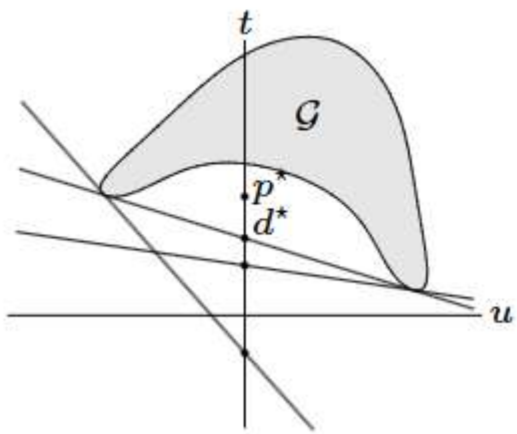
2 弱对偶和强对偶

为简化起见, 考虑具有一个约束条件 $f_1(x) \le 0$ 的问题

■ **对偶函数的几何解释:**

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



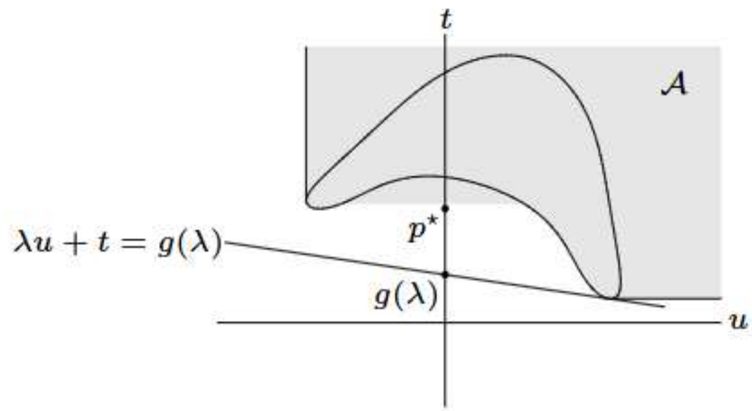


1. $g(\lambda)$一定在p^*下方
2. d^*一定在p^*下方

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）

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



- **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. **rank** $A = p$ （假设）. 定义以下两个互斥的集合:

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

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

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

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

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

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 $x \in \mathcal{D}$

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

2 弱对偶和强对偶

情况 1: $\mu > 0$

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

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

情况 2: $\mu = 0$

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

假设 $\tilde{x}$ 满足Slater约束条件. 那么

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

2 弱对偶和强对偶

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

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$\tilde{v}^T(Ax - b) < 0$ $(\mathbf{rank}\,A = p, \tilde{v}$ 是 p 维的),矛盾

2 弱对偶和强对偶

- **例** 5: (使用对偶形式求解原问题)使用等式约束最小化求和函数:

最小化 $\quad f_0(x) = \sum_{i=1}^{n} f_i(x_i)$

约束条件 $\quad a^T x = b$

f_i 严格凸且可微

- **解:** 拉格朗日函数是:

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

$$= -b\nu + \sum_{i=1}^{n} \left(f_i(x_i) + \nu a_i x_i\right)$$

$g(\nu) = -b\nu - \sum_{i=1}^{n} f_i^*(-\nu a_i)$, 对偶问题是:

最大化 $\quad -b\nu - \sum_{i=1}^{n} f_i^*(-\nu a_i)$

大 纲

1. 拉格朗日对偶问题

2. 弱对偶和强对偶

3. 例子:强对偶与弱对偶

3 例子：强对偶与弱对偶

- **例 2:不等式形式 LP**

- **原问题:**

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

- **对偶函数:**

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

3 例子：强对偶与弱对偶

- 例 2: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 $A\tilde{x} \prec b$ 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

4 例子：强对偶与弱对偶

- 例 3:二次规划

- 原问题:

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

- 对偶函数:

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

3 例子：强对偶与弱对偶

- **例 3:二次规划**

- **对偶问题:**

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

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

3 例子：强对偶与弱对偶

- 例 4:一个具有强对偶性的非凸问题*

- **原问题:**

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

$A \not\succeq 0$, 所以非凸

- **对偶函数:**

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

1. 若 $A + \lambda I \not\succeq 0$ 或若 $A + \lambda I \succeq 0$ 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: $g(\lambda) = -b^T (A + \lambda I)^{\dagger} b - \lambda$

3 例子：强对偶与弱对偶

- **例 4:一个具有强对偶性的非凸问题:**

- **对偶问题**与等价 SDP**:**

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

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

强对偶性, 尽管原始问题不是凸的(不容易证明)

谢 谢 ！