

# Convex Optimization Theory and Applications

## Topic 4 - Duality

Li Li and Shuning Wang

Department of Automation  
Tsinghua University

Fall, 2009-2021.

## 4.0. Outline

4.1. 线性规划问题的对偶理论

4.2. Definition and Examples 基本定义和例子

4.3. Slater 条件与 KKT 条件

4.4. 灵敏度与择一理论

4.5. 广义不等式约束问题

4.6. 对偶问题和问题变形

## 4.1. 线性规划问题的对偶理论

对于一个线性规划问题（原问题），我们可以找到一个对偶问题

原问题

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax \geq b \\ & x \geq 0 \end{aligned}$$

对偶问题

$$\begin{aligned} \max \quad & b^T u \\ \text{s.t.} \quad & A^T u \leq c^T \\ & u \geq 0 \end{aligned}$$

弱对偶性：原问题任何可行目标值都是对偶问题最优目标值的界（推论：原对偶问题目标值相等的一对可行解是各自的最优解）

强对偶性：原对偶问题只要有一个有有界最优解，另一个就有最优解，并且最优目标值相等

## 4.1. 线性规划问题的对偶理论

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax \geq b \\ & x \geq 0 \end{aligned}$$

$$\begin{aligned} \max \quad & b^T u \\ \text{s.t.} \quad & A^T u \leq c^T \\ & u \geq 0 \end{aligned}$$

### Theorem (Weak Duality Theorem)

*For the canonical form LPP, if  $\mathbf{x}$  is a feasible solution (not necessarily basic) of the primal problem and  $\mathbf{u}$  is a feasible solution (not necessarily basic) of the dual problem, then*

$$\mathbf{c}^T \mathbf{x} \geq \mathbf{b}^T \mathbf{u}$$

### Proof.

Because  $\mathbf{x}$  is a feasible solution of the primal problem, we have  $A\mathbf{x} \geq \mathbf{b}$ . So, for any  $\mathbf{u} \geq \mathbf{0}$ , we have

## 4.1. 线性规划问题的对偶理论

$$\mathbf{u}^T A \mathbf{x} \geq \mathbf{u}^T \mathbf{b} = \mathbf{b}^T \mathbf{u}$$

Because  $\mathbf{u}$  is a feasible solution to the dual problem, we have  $A^T \mathbf{u} \leq \mathbf{c}$ . So, for any  $\mathbf{x} \geq \mathbf{0}$ , we have

$$\mathbf{x}^T A^T \mathbf{u} \leq \mathbf{x}^T \mathbf{c}$$

Combining these two inequalities, we have  $\mathbf{c}^T \mathbf{x} \geq \mathbf{u}^T A \mathbf{x} \geq \mathbf{b}^T \mathbf{u}$ .

## 4.1. 线性规划问题的对偶理论

为了证明线性规划问题的强对偶性，我们需要介绍 Farkas 引理，证明过程也是我们凸集分离定理的一个应用

### Theorem (Farkas' Lemma)

*Given  $A \in \mathbb{R}^{m \times n}$  is an  $m \times n$  matrix,  $\mathbf{b} \in \mathbb{R}^m$  is an  $m$ -dimensional column vector. Exactly one of the following linear system is feasible:*

*I. There exists an  $\mathbf{x} = [x_1, \dots, x_n]^T \in \mathbb{R}^n$  such that  $A\mathbf{x} = \mathbf{b}$  and  $\mathbf{x} \geq \mathbf{0}$ .*

*II. There exists a  $\mathbf{y} = [y_1, \dots, y_m]^T \in \mathbb{R}^m$  such that  $A^T \mathbf{y} \geq \mathbf{0}$  and  $\mathbf{b}^T \mathbf{y} < 0$ .*

## 4.1. 线性规划问题的对偶理论

Proof.

First, we use contradiction method to show that both systems cannot simultaneously have feasible solutions.

If both system are simultaneously feasible,  $\mathbf{b}^T \mathbf{y} < 0$  implies  $\mathbf{y} \neq \mathbf{0}$  and  $\mathbf{b} \neq \mathbf{0}$ .

Meanwhile, if  $\mathbf{b} \neq \mathbf{0}$ ,  $A\mathbf{x} = \mathbf{b}$  implies  $\mathbf{x} \neq \mathbf{0}$ . If both systems holds, then we have

$$\mathbf{b}^T \mathbf{y} = (A\mathbf{x})^T \mathbf{y} = \mathbf{x}^T (A^T \mathbf{y}) \geq 0 \quad (16)$$

which contradicts  $\mathbf{b}^T \mathbf{y} < 0$ .

## 4.1. 线性规划问题的对偶理论

Second, we show that at least one of them has a feasible solution. If System (I) is feasible, we can finish right here. Otherwise, System (I) is infeasible, we have  $\Omega = \{A\mathbf{x}, \mathbf{x} \geq \mathbf{0}\}$  is a closed convex set. Moreover,  $\mathbf{b} \notin \Omega$ .

According to Separating Hyperplane Theorem, there exists a hyperplane  $\mathbf{y}^T \mathbf{x} = z$  that separates  $\mathbf{b}$  from  $\Omega$ , where  $\mathbf{y} = [y_1, \dots, y_m]^T \in \mathbb{R}^m$  is an  $m$ -dimensional column vector. That is,  $\mathbf{y}^T \mathbf{b} < z$  and  $\forall \mathbf{s} \in \Omega, \mathbf{y}^T \mathbf{s} \geq z$ .

Since  $\mathbf{0} \in \Omega$ , we have  $z \leq 0$ . As a result,  $\mathbf{y}^T \mathbf{b} < 0$ .

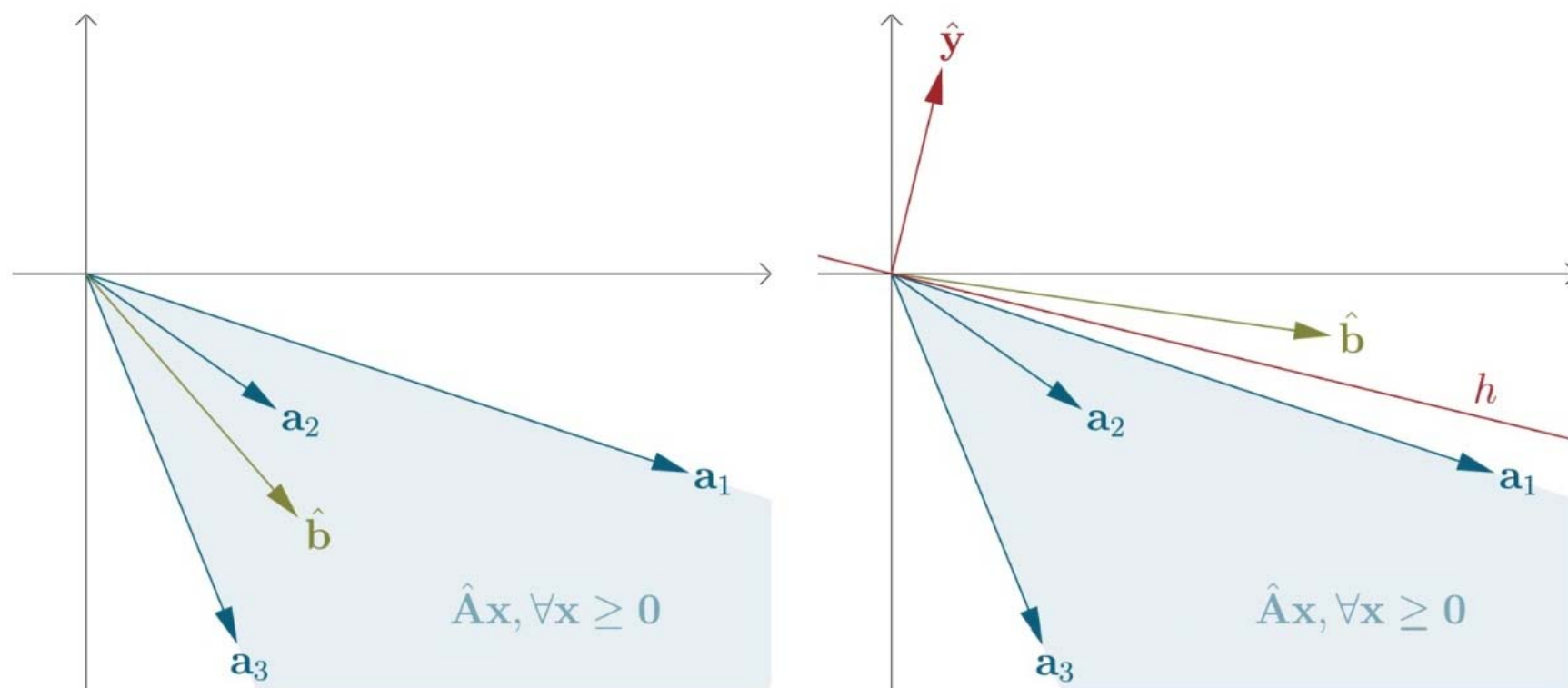
On the other hand, since  $\mathbf{y}^T A\mathbf{x} > 0$  for all  $\mathbf{x} \in \Omega$ , we can see that  $\mathbf{y}^T A > \mathbf{0}$ , since each element of  $\mathbf{x}$  can be arbitrarily large.

Therefore, we prove the whole statement.



## 4.1. 线性规划问题的对偶理论

我们把矩阵  $A$  的列空间写出来，其实就是  $n$  个  $m$  维的向量，这些向量前面加权非负系数组合出来的点构成的集合就是一个凸锥。左为  $b$  在凸锥内的情况；右图为  $b$  在凸锥外的情况，如果是右图的情况，总能找到过原点的超平面（二维情况下为直线，法向量为  $y$ ），把  $b$  和凸锥分开。



## 4.1. 线性规划问题的对偶理论

### Theorem (Strong Duality Theorem)

*For the canonical form LPP, a feasible solution  $\mathbf{x}^*$  to the primal problem is optimal if and only if there exists a feasible solution  $\mathbf{u}^*$  to the dual such that*

$$\mathbf{c}^T \mathbf{x}^* = \mathbf{b}^T \mathbf{u}^*$$

*Meanwhile,  $\mathbf{u}^*$  is an optimal solution to the dual.*

### Proof.

First, We prove the sufficiency.

Based on weak duality theorem, for any feasible solution  $\mathbf{x}$  of the primal problem, we have

$$\mathbf{c}^T \mathbf{x} \geq \mathbf{b}^T \mathbf{u}^* = \mathbf{c}^T \mathbf{x}^*$$

which shows that  $\mathbf{x}^*$  is also the optimal solution of the primal problem.

## 4.1. 线性规划问题的对偶理论

Similarly, for any feasible solution  $\mathbf{u}$  of the dual problem, we have

$$\mathbf{b}^T \mathbf{u} \leq \mathbf{c}^T \mathbf{x}^* = \mathbf{b}^T \mathbf{u}^*$$

which shows that  $\mathbf{u}^*$  is also the optimal solution of the dual problem.

Next, we prove the necessariness based on Farkas' Lemma, since we do not introduce the simplex algorithm here.

Suppose  $\mathbf{x}^*$  is an optimal solution. We will show that there exists a dual feasible solution  $\mathbf{u}$  with  $\mathbf{b}^T \mathbf{u} = \mathbf{c}^T \mathbf{x}^*$ .

Let us define  $I$  as the set of constraint index that active at  $\mathbf{x}^*$ . That is,

$$\begin{aligned} \mathbf{a}_i^T \mathbf{x}^* &= b_i, & i \in I \\ \mathbf{a}_i^T \mathbf{x}^* &> b_i, & i \notin I \end{aligned}$$

## 4.1. 线性规划问题的对偶理论

$\mathbf{x}^*$  implies that, for any  $\mathbf{d} \in \mathbb{R}^n$ , the following set

$$a_i^T \mathbf{d} \geq 0, \mathbf{c}^T \mathbf{d} < 0, i \in I$$

is infeasible. Otherwise, we would have a small enough  $\epsilon > 0$  such that

$$a_i^T (\mathbf{x}^* + \epsilon \mathbf{d}) \geq b_i, \mathbf{c}^T (\mathbf{x}^* + \epsilon \mathbf{d}) < \mathbf{c}^T \mathbf{x}^*, i = 1, \dots, m$$

According to Farkas' Lemma, we know that the above inequality is infeasible if and only if there exists  $\lambda_i, i \in I$  that

$$\lambda_i \geq 0, \sum_{i \in I} \lambda_i a_i = \mathbf{c}$$

## 4.1. 线性规划问题的对偶理论

This yields a dual feasible solution  $\mathbf{u}$  satisfying

$$\begin{aligned} u_i &= \lambda_i, & i \in I \\ u_i &= 0, & i \notin I \end{aligned}$$

Finally, we show that  $\mathbf{u}$  is the optimal solution for the dual problem. Indeed, we have

$$\mathbf{b}^T \mathbf{u} = \sum_{i \in I} b_i u_i = \sum_{i \in I} (a_i^T \mathbf{x}_i^*) u_i = \mathbf{u}^* A \mathbf{x}^* = \mathbf{c}^T \mathbf{x}^*$$

Based on Weak Duality Theorem, we see  $\mathbf{u}$  is the optimal solution for the dual problem. Thus comes our statement according to strong duality.

## 4.1. 线性规划问题的对偶理论

Based on weak and strong duality theorems, we can get the co-feasibility relationship between the primal and dual problems as follows

### Theorem

*For the canonical form LPP, the co-feasibility relationship between the primal and dual problems can be determined as*

<i>Dual \ Primal</i>	<i>Infeasible</i>	<i>Optimal</i>	<i>Unbounded</i>
<i>Infeasible</i>	✓	×	✓
<i>Optimal</i>	×	✓	×
<i>Unbounded</i>	✓	×	×

对于一般原/对偶问题也成立

## 4.2. Definition and Examples

让我们从寻找一个优化问题（原问题）的下界入手，考虑

$$\min\{f_0(x): f_i(x) \leq 0, i = 1, 2, \dots, m\} \quad (4.1)$$

现在的问题是如何找到问题(4.1)最优值的一个最好的下界？首先我们知道若方程组

$$\begin{cases} f_0(x) < v \\ f_i(x) \leq 0, i = 1, 2, \dots, m \end{cases} \quad (4.2)$$

无解，则 $v$ 是问题(1)的一个下界。注意到方程组(4.2)有解可以推出对于任意的 $\lambda \geq 0$ ，以下方程

$$f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) < v \quad (4.3)$$

有解。因此根据逆否命题，方程组(4.2)无解的充分条件是存在 $\lambda \geq 0$ ，让方程(4.3)无解。

## 4.2. Definition and Examples

而方程(4.3)无解的充要条件是

$$\min_x f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) \geq v \quad (4.4)$$

因为我们要找最好的下界，所以这个时候的 $v$ 和 $\lambda$ 应该取

$$v = \max_{\lambda \geq 0} \min_x f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) \quad (4.5)$$

由此引入了 dual problem。证明逻辑是根据式(4.5)取 $v$ 和 $\lambda$ ，则(4.4)成立，从而导出(4.3)无解，然后可以知道(4.2)无解，因此 $v$ 是问题(1)的下界。



## 4.2. Definition and Examples

对于一般性优化问题

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{aligned}$$

称以下函数为其 **Lagrange 函数**

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

其中  $\lambda_i, \mu_i$  称为对应的不等式、等式约束的 **Lagrange 乘子** Lagrange multipliers,  $\lambda \in R^m, \mu \in R^p$  称为 **Lagrange 乘子向量** 或者对偶变量 **dual variables**, 可以被视为违反不同约束所带来的负面影响的权重

## 4.2. Definition and Examples

让我们进一步观察一下 Lagrange 函数的形式特征

$$f_0(x) + \max_{\lambda, \mu: \lambda_i \geq 0, \forall i} \left[ \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \mu_i h_i(x) \right]$$

令  $A = \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \mu_i h_i(x)$ ，对  $A$  进行观察，有如下结论：

- 如果存在任何  $f_i(x) > 0$ ，那么将它对应的参数  $\lambda_i$  设置为极大就能最大化  $A$  的值；而如果  $f_i(x) \leq 0$ ，由于参数存在约束  $\lambda_i \geq 0$ ，因此只有当  $\lambda_i = 0$  时， $A$  才能取得极大值，并且极大值为 0
- 类似的，如果存在任何  $h_i(x) \neq 0$ ，那么将它对应的参数  $\mu_i$  设置为与  $h_i(x)$  同号，并设置  $|\mu_i|$  为极大就能最大化  $A$  的值；如果  $h_i(x) = 0$ ，则  $A$  最大值为 0 并且与参数  $\mu_i$  的取值无关

## 4.2. Definition and Examples

对上述观察进行整理可知，如果  $x$  原始可行，那么  $A$  的极大值为 0，而如果任意约束条件不成立，则  $A$  的极大值为  $\infty$ ，即目标函数为

$$f_0(x) + \begin{cases} 0, & \text{if } x \text{ is primal feasible} \\ \infty, & \text{if } x \text{ is primal infeasible} \end{cases}$$

从直觉上来说， $A$  这一部分可以视作是一类“阻拦”函数，防止我们将不可行解作为原凸优化问题的解。

由此可以看出，Lagrange 函数可以视作最初凸优化问题的一个改进版本，它将最初凸优化问题中的约束变为了式子的一部分。两者之间的区别在于，Lagrange 函数的不可行解会导致原始目标的值为  $\infty$ 。原始问题的最优解  $x^*$  即为最初凸优化问题的最优解，原始目标的最优值  $p^*$  就是全局最优值。

## 4.2. Definition and Examples

称以下函数为上述优化问题的**对偶函数**（一定是凹函数）

$$g(\lambda, \mu) = \inf_{x \in D} f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \mu_i h_i(x)$$

其中  $D = \bigcap_{i=0,1,\dots,m} \text{dom } f_i \cap \bigcap_{i=0,1,\dots,p} \text{dom } h_i$  是以上问题的优化变量的定义域

$$\text{dom } g = \{(\lambda, \mu) \mid g(\lambda, \mu) > -\infty\}$$

称以下问题为上述优化问题的对偶问题

$$\max \{g(\lambda, \mu) \mid \text{s.t. } \lambda \geq 0, (\lambda, \mu) \in \text{dom } g\}$$

称以下集合为上述对偶问题可行集

$$\{(\lambda, \mu) \mid \lambda \geq 0, (\lambda, \mu) \in \text{dom } g\}$$

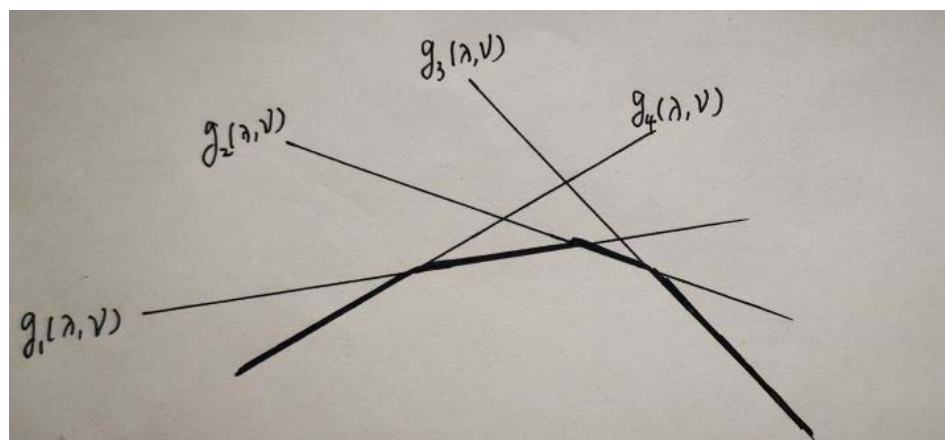
## 4.2. Definition and Examples

对偶函数一定是凹函数，其凹性与原目标函数和约束函数凹凸与否无关。

证明：  $L(x, \lambda, \mu)$  可以看作是一个无限的函数集，这个函数集中每个元素是  $L(x_i, \lambda, \mu)$ ， $x$  取遍其在定义域上的所有值得到不同的  $x_i$ 。针对不同的  $x_i$ ， $L(x_i, \lambda, \mu)$  的表达式不一样，由于这个表达式是只关于  $\lambda$  和  $\mu$  的，故用  $g_i(\lambda, \mu)$  来表示。所以

$$g(\lambda, \mu) = \inf\{g_1(\lambda, \mu), g_2(\lambda, \mu), \dots, g_\infty(\lambda, \mu)\}$$

当  $L$  看成是关于  $\lambda$  或  $\mu$  的函数时， $L$  是一个仿射函数，亦即， $g_i(\lambda, \mu)$  是仿射函数，对仿射函数集取下确界，得到的函数是凹函数，如下图所示：



## 4.2. Definition and Examples

**证明：**要证对偶函数一定是凹函数，根据凹函数的定义，就是要证

$$g(\theta\lambda_1 + (1 - \theta)\lambda_2, \theta\nu_1 + (1 - \theta)\nu_2) \geq \theta g(\lambda_1, \nu_1) + (1 - \theta)g(\lambda_2, \nu_2), \quad \theta \in R \quad (\text{公式3})$$

根据对偶函数的定义可知，对偶函数是拉格朗日函数在把 $\lambda$ 和 $\nu$ 当做常量， $x$ 变化时的最小值，如果拉格朗日函数没有最小值（可以认为最小值为 $-\infty$ ），则对偶函数取值为 $-\infty$ ，所以，可以把对偶函数按照下面的方式表达：

$$g(\lambda, \nu) = \min\{L(x_1, \lambda, \nu), L(x_2, \lambda, \nu), \dots, L(x_n, \lambda, \nu)\}, \quad n = +\infty \quad (\text{公式4})$$

即无穷多个 $x$ 变化时，拉格朗日函数的最小值。

另外，由于把 $\lambda$ 和 $\nu$ 分开来写，式子太长了，为了简便，记 $\gamma = (\lambda, \nu)$ ，接下来证明（公式3）：

$$g(\theta\gamma_1 + (1 - \theta)\gamma_2) = \min\{L(x_1, \theta\gamma_1 + (1 - \theta)\gamma_2), L(x_2, \theta\gamma_1 + (1 - \theta)\gamma_2), \dots, L(x_n, \theta\gamma_1 + (1 - \theta)\gamma_2)\} \quad (\text{公式5})$$

$$\geq \min\{\theta L(x_1, \gamma_1) + (1 - \theta)L(x_1, \gamma_2), \theta L(x_2, \gamma_1) + (1 - \theta)L(x_2, \gamma_2), \dots, \theta L(x_n, \gamma_1) + (1 - \theta)L(x_n, \gamma_2)\} \quad (\text{公式6})$$

$$\geq \theta \min\{L(x_1, \gamma_1), L(x_2, \gamma_1), \dots, L(x_n, \gamma_1)\} + (1 - \theta) \min\{L(x_1, \gamma_2), L(x_2, \gamma_2), \dots, L(x_n, \gamma_2)\} \quad (\text{公式7})$$

$$= \theta g(\gamma_1) + (1 - \theta)g(\gamma_2) \quad (\text{公式8})$$

至此，（公式3）得证，所以原命题得证。

## 4.2. Definition and Examples

### Lagrange dual problem

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

- finds best lower bound on  $p^*$ , obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted  $d^*$
- $\lambda, \nu$  are dual feasible if  $\lambda \succeq 0, (\lambda, \nu) \in \mathbf{dom} g$
- often simplified by making implicit constraint  $(\lambda, \nu) \in \mathbf{dom} g$  explicit

**example:** standard form LP and its dual (page 5–5)

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

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

## 4.2. Definition and Examples

弱对偶性

$$d^* = \sup_{\lambda \geq 0} g(\lambda, \mu), \quad p^* = \inf \left\{ f_0(x) \mid \text{s.t. } f_i(x) \leq 0, 1 \leq i \leq m, h_i(x) = 0, 1 \leq i \leq p \right\}$$

设  $(x, \lambda, \mu)$  是任意的原对偶可行对（指  $x$  是原问题可行解， $\lambda, \mu$  是对偶问题可行解），则总是成立

$$g(\lambda, \mu) \leq d^* \leq p^* \leq f_0(x)$$

Proof: if  $\tilde{x}$  is feasible and  $\lambda \geq 0$ , then

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

minimizing over all feasible  $\tilde{x}$  gives  $p^* \geq g(\lambda, \mu)$ .



## 4.2. Definition and Examples

推论：设  $(x, \lambda, \mu)$  是原对偶可行对，称  $\delta(x, \lambda, \mu) = f_0(x) - g(\lambda, \mu)$  为其**对偶间隙**，则成立

$$0 \leq f_0(x) - p^* \leq \delta(x, \lambda, \mu), \quad 0 \leq d^* - g(\lambda, \mu) \leq \delta_0(x, \lambda, \mu)$$

如果存在原对偶可行对  $(x, \lambda, \mu)$  满足  $g(\lambda, \mu) = f_0(x)$ ，那么  $x$  是原问题最优解， $\lambda, \mu$  是对偶问题最优解，此时  $d^* = p^*$ ，称原对偶问题满足**强对偶性**

## 4.2. Definition and Examples

一般形式线性规划问题  $\min \{c^T x \mid \text{s.t. } Gx \leq h, Ax = b\}$

Lagrange 函数  $L(x, \lambda, \mu) = c^T x + \lambda^T (Gx - h) + \mu^T (Ax - b)$

对偶函数  $g(\lambda, \mu) = \begin{cases} -h^T \lambda - b^T \mu & \text{if } c + G^T \lambda + A^T \mu = 0 \\ -\infty & \text{otherwise} \end{cases}$

对偶问题  $\max \{-h^T \lambda - b^T \mu \mid \text{s.t. } c + G^T \lambda + A^T \mu = 0, \lambda \geq 0\}$

对偶问题可行集  $\{(\lambda, \mu) \mid \lambda \geq 0, c + G^T \lambda + A^T \mu = 0\}$

## 4.2. Definition and Examples

线性规划问题满足强对偶性

原问题  $\min \{c^T x \mid \text{s.t. } Gx \leq h, Ax = b\}$

对偶问题  $\max \{-h^T \lambda - b^T \mu \mid \text{s.t. } c + G^T \lambda + A^T \mu = 0, \lambda \geq 0\}$

如果  $p^* = -\infty$ ，对偶问题无可行解，可视为  $d^* = -\infty$ ；如果  $d^* = \infty$ ，原问题无可行解，可视为  $p^* = \infty$ ；下面考虑  $-\infty < p^* < \infty$  的情况

$x$  是线性规划最优解的充要条件：存在  $\lambda, \mu$  一起满足

$$c + G^T \lambda + A^T \mu = 0, \lambda \geq 0; \lambda^T (Gx - h) = 0; Gx \leq h, Ax = b$$

$$\Rightarrow \lambda, \mu \text{ 是对偶问题可行解, } c^T x = -\lambda^T Gx - \mu^T Ax = -h^T \lambda - b^T \mu$$

## 4.2. Definition and Examples

### Least-norm solution of linear equations

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

#### dual function

- Lagrangian is  $L(x, \nu) = x^T x + \nu^T (Ax - b)$
- to minimize  $L$  over  $x$ , set gradient equal to zero:

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

- plug in in  $L$  to obtain  $g$ :

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

a concave function of  $\nu$

**lower bound property:**  $p^* \geq -(1/4)\nu^T AA^T \nu - b^T \nu$  for all  $\nu$

## 4.2. Definition and Examples

### Equality constrained norm minimization

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

#### dual function

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

where  $\|v\|_* = \sup_{\|u\| \leq 1} u^T v$  is dual norm of  $\|\cdot\|$

proof: follows from  $\inf_x (\|x\| - y^T x) = 0$  if  $\|y\|_* \leq 1$ ,  $-\infty$  otherwise

- if  $\|y\|_* \leq 1$ , then  $\|x\| - y^T x \geq 0$  for all  $x$ , with equality if  $x = 0$
- if  $\|y\|_* > 1$ , choose  $x = tu$  where  $\|u\| \leq 1$ ,  $u^T y = \|y\|_* > 1$ :

$$\|x\| - y^T x = t(\|u\| - \|y\|_*) \rightarrow -\infty \quad \text{as } t \rightarrow \infty$$

**lower bound property:**  $p^* \geq b^T \nu$  if  $\|A^T \nu\|_* \leq 1$

## 4.2. Definition and Examples

Dual norm: Let  $\|\cdot\|$  be any norm. Its dual norm is defined as

$$\|x\|_* = \max_{\|y\| \leq 1} x^T y, \quad \text{s.t. } \|y\| \leq 1$$

You can think of this as the operator norm of  $x^T$ . The dual norm is indeed a norm. The first two properties are straightforward to prove. The triangle inequality can be shown in the following way:

$$\|x + z\|_* = \max_{\|y\| \leq 1} (x^T y + z^T y) \leq \max_{\|y\| \leq 1} x^T y + \max_{\|y\| \leq 1} z^T y = \|x\|_* + \|z\|_*$$

对偶范数的等价定义

$$\|z\|_* = \sup_{\|x\| \leq 1} z^T x = \sup_{\|x\|=1} z^T x = \sup_{x \neq 0} \frac{z^T x}{\|x\|}$$

If  $\|x\|$  is a norm and  $\|x\|_*$  is the dual norm of it,

$$\|z^T x\| \leq \|z\|_* \cdot \|x\|_* \text{ holds.}$$

## 4.2. Definition and Examples

由霍尔德(Hölder)不等式可以直接得出: $l_p$ -范数的对偶范数是 $l_q$ -范数, 其中  $\frac{1}{p} + \frac{1}{q} = 1$

$$z^T x \leq \|x\|_p \|z\|_q \Rightarrow \|z\|_* = \sup_{x \neq 0} \frac{z^T x}{\|x\|_p} = \|z\|_q$$

霍尔德(Hölder)不等式

设  $p, q > 0, \frac{1}{p} + \frac{1}{q} = 1$ . 则

$$x^{\frac{1}{p}} y^{\frac{1}{q}} \leq \frac{x}{p} + \frac{y}{q}, \forall x, y \geq 0$$

等号仅当  $x = y$  时成立。

## 4.2. Definition and Examples

证明：考察对数函数  $\log(x)$  显然是一个凹函数：

$$\log(\theta x + (1-\theta)y) \geq \theta \log(x) + (1-\theta) \log(y)$$

取  $\theta = \frac{1}{p}$ ，则  $1-\theta = \frac{1}{q}$ ，故

$$\log\left(\frac{1}{p}x + \frac{1}{q}y\right) \geq \frac{1}{p}\log(x) + \frac{1}{q}\log(y)$$

两边同时取指数，得

$$\frac{x}{p} + \frac{y}{q} \geq x^{\frac{1}{p}} y^{\frac{1}{q}}$$

对引理中的不等式，做如下替换

$$x_i = \frac{a_i^p}{\sum_{j=1}^n a_j^p}, y_i = \frac{b_i^q}{\sum_{j=1}^n b_j^q}$$



## 4.2. Definition and Examples

得到  $n$  个不等式:

$$\frac{a_i b_i}{\left(\sum_{j=1}^n a_j^p\right)^{\frac{1}{p}} \left(\sum_{j=1}^n b_j^q\right)^{\frac{1}{q}}} \leq \frac{1}{p} \frac{a_i^p}{\sum_{j=1}^n a_j^p} + \frac{1}{q} \frac{b_i^q}{\sum_{j=1}^n b_j^q}$$

将上式两边对  $i = 1, 2, \dots, n$  求和, 就得到

$$\begin{aligned} \frac{\sum_{i=1}^n a_i b_i}{\left(\sum_{j=1}^n a_j^p\right)^{\frac{1}{p}} \left(\sum_{j=1}^n b_j^q\right)^{\frac{1}{q}}} &\leq \frac{1}{p} + \frac{1}{q} = 1 \\ \Rightarrow \sum_{i=1}^n a_i b_i &\leq \left(\sum_{j=1}^n a_j^p\right)^{\frac{1}{p}} \left(\sum_{j=1}^n b_j^q\right)^{\frac{1}{q}} \end{aligned}$$

上式要求  $a_i, b_i \geq 0$ 。否则, 需要给等式右端的  $a_i, b_i$  加上绝对值, 得到如下不等式

$$a^T b \leq \|a\|_p \|b\|_q$$

事实上  $\|\cdot\|_q$  正是  $\|\cdot\|_p$  的对偶范数。

## 4.2. Definition and Examples

Please prove that Dual norm of  $l_1$  is  $l_\infty$ .

## 4.2. Definition and Examples

Please prove that Dual norm of  $l_1$  is  $l_\infty$ .

Let  $\|z\|_D = \sup\{z^T x \mid \|x\|_1 \leq 1\}$ , then

$$z^T x \leq \sum_{i=1}^n |z_i x_i| = \sum_{i=1}^n |z_i| |x_i| \leq \left(\max_{i=1}^n |z_i|\right) \sum_{i=1}^n |x_i|$$

Since  $\|x\|_1 \leq 1$ , we have  $z^T x \leq \max_{i=1}^n |z_i|$ . With these, we are able to show that  $l_\infty$  norm of  $z$  is an upper bound of  $z^T x$  when  $\|x\|_1 \leq 1$ .

We just have to pick at element of  $x$  that attains it.

Given a  $z$ , we check it's component and look for the one with maximum norm. Say it is component  $z_i$ . Then we pick  $x = \text{sign}(z_i) e_i$ . We have  $\|x\|_1 = 1$ . Also,

$$z^T x = z^T \text{sign}(z_i) e_i = \text{sign}(z_i) z_i = |z_i| = \|z\|_\infty$$

## 4.2. Definition and Examples

The dual norm of a dual norm of a vector is original norm.

The fact that dual to dual norm is equal to the original norm in case of finite-dimensional spaces is equivalent to the fact that the corresponding Banach space is reflexive. By James' theorem, a Banach space  $B$  is reflexive if and only if every continuous linear functional on  $B$  attains its maximum on the closed unit ball in  $B$ . That is surely true for finite-dimensional spaces with any norms, thus dual to dual norm must be equivalent to the original norm.

## 4.2. Definition and Examples

### Two-way partitioning

$$\begin{array}{ll}\text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \dots, n\end{array}$$

- a nonconvex problem; feasible set contains  $2^n$  discrete points
- interpretation: partition  $\{1, \dots, n\}$  in two sets;  $W_{ij}$  is cost of assigning  $i, j$  to the same set;  $-W_{ij}$  is cost of assigning to different sets

### dual function

$$\begin{aligned}g(\nu) &= \inf_x (x^T W x + \sum_i \nu_i (x_i^2 - 1)) = \inf_x x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu \\ &= \begin{cases} -\mathbf{1}^T \nu & W + \mathbf{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise} \end{cases}\end{aligned}$$

**lower bound property:**  $p^* \geq -\mathbf{1}^T \nu$  if  $W + \mathbf{diag}(\nu) \succeq 0$

example:  $\nu = -\lambda_{\min}(W)\mathbf{1}$  gives bound  $p^* \geq n\lambda_{\min}(W)$

## 4.2. Definition and Examples

**weak duality:**  $d^* \leq p^*$

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems  
for example, solving the SDP

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

gives a lower bound for the two-way partitioning problem

**strong duality:**  $d^* = p^*$

- does not hold in general
- (usually) holds for convex problems
- conditions that guarantee strong duality in convex problems are called **constraint qualifications**

## 4.2. Definition and Examples

我们关心对偶问题的原因是：

1. 对偶问题总是凸优化问题，即便原问题非凸优化问题，对偶问题仍然是凸优化的。
2. 很多情况下对偶问题比原问题要简单求解，当原问题不太好求解时，我们想通过解对偶问题得到对原问题的解的最好的近似。当原问题为凸问题，且 Slater 条件成立时，strong duality holds，对偶问题的解就是原问题的解，这个时候我们称 duality gap 为零，否则的话，对偶问题的解始终是原问题的解的下界，这个时候 duality gap 不等于零。

## 4.3. Slater 条件与 KKT 条件

Slater Condition 是凸优化问题满足强对偶性的充分条件  
strong duality holds for a convex problem

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

if it is strictly feasible, *i.e.*,

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

相对内部 Relative Interior  $\text{Relint} D = \{x \in D \mid B(x, r) \cap \text{aff} D \in D \quad \exists r \in D$

- also guarantees that the dual optimum is attained (if  $p^* > -\infty$ )
- can be sharpened: *e.g.*, can replace  $\mathbf{int} \mathcal{D}$  with  $\mathbf{relint} \mathcal{D}$  (interior relative to affine hull); linear inequalities do not need to hold with strict inequality, . . .
- there exist many other types of constraint qualifications



## 4.3. Slater 条件与 KKT 条件

**Slater 条件（约束品性）：**

对于凸优化问题  $\min \{f_0(x) \mid \text{s.t. } f_i(x) \leq 0, 1 \leq i \leq m, Ax = b\}$ ，如果存在  $D = \bigcap_{i=0,1,\dots,m} \text{dom } f_i$  的内点  $\hat{x}$ ，满足  $f_i(\hat{x}) < 0, 1 \leq i \leq m, A\hat{x} = b$ ，则称该问题满足 Slater 约束品性。

**Slater 定理：**

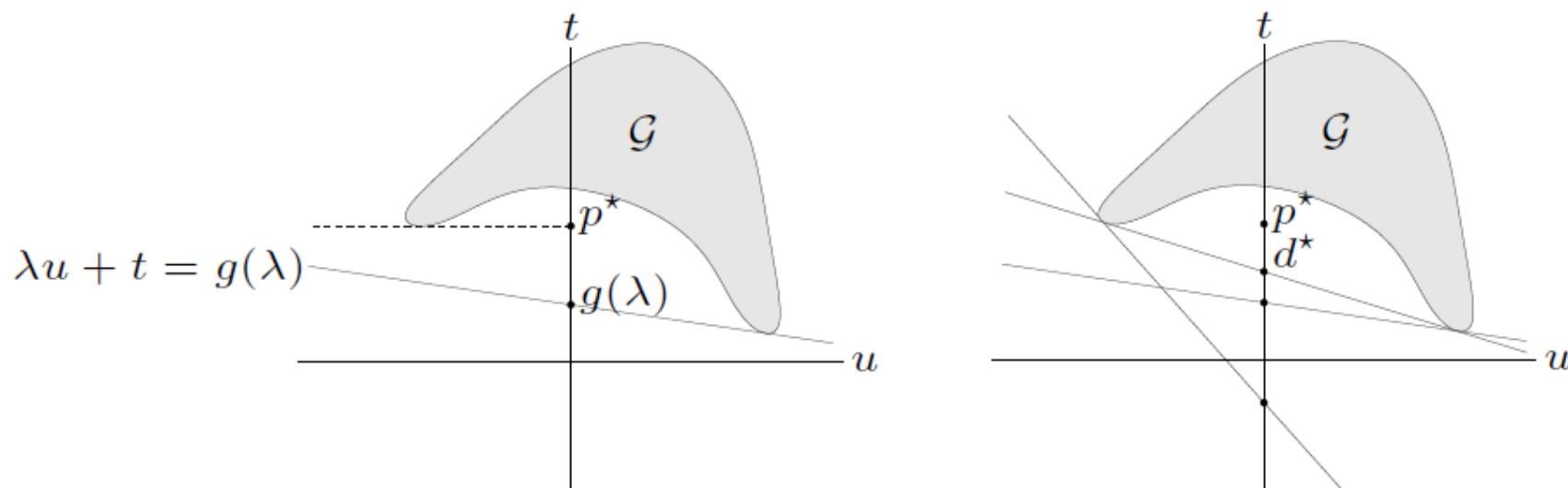
对于凸优化问题，Slater 约束品性可保证强对偶性成立

## 4.3. Slater 条件与 KKT 条件

for simplicity, consider problem with one constraint  $f_1(x) \leq 0$

**interpretation of dual function:**

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



- $\lambda u + t = g(\lambda)$  is (non-vertical) supporting hyperplane to  $\mathcal{G}$
- hyperplane intersects  $t$ -axis at  $t = g(\lambda)$

## 4.3. Slater 条件与 KKT 条件

我们的 Lagrange 函数是  $L(x, \lambda) = f_0(x) + \lambda f_1(x)$ ，Lagrange 对偶函数是  $g(\lambda) = \inf_{u, t} (t + \lambda u)$ ， $G = \{(f_0(x), f_1(x)) \mid x \in D\}$ 。

显然，对于给定的  $\lambda$ ， $g(\lambda)$  为关于变量  $(u, t)$  仿射函数  $(\lambda, 1)^T (u, t) = \lambda u + t$ ，注意我们需要升 1 维来表示这类仿射函数，或者说超平面。这里横轴为约束，竖轴为目标。

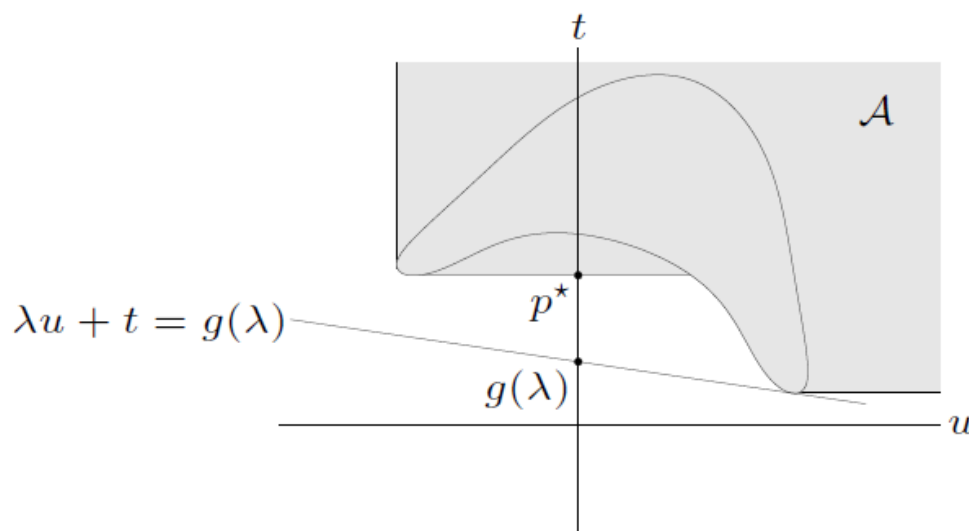
给定  $\lambda$ ，对偶函数的最优解为法线方向  $(\lambda, 1)$  的超平面与集合  $G$  的支撑超平面，此时，找到的解为  $t$ -截距的下确界。

进一步变化不同的  $\lambda$ ，对偶问题的最优解为：找到对应的所有支撑超平面中， $t$  值最大的那个

## 4.3. Slater 条件与 KKT 条件

**epigraph variation:** same interpretation if  $\mathcal{G}$  is replaced with

$$\mathcal{A} = \{(u, t) \mid f_1(x) \leq u, f_0(x) \leq t \text{ for some } x \in \mathcal{D}\}$$

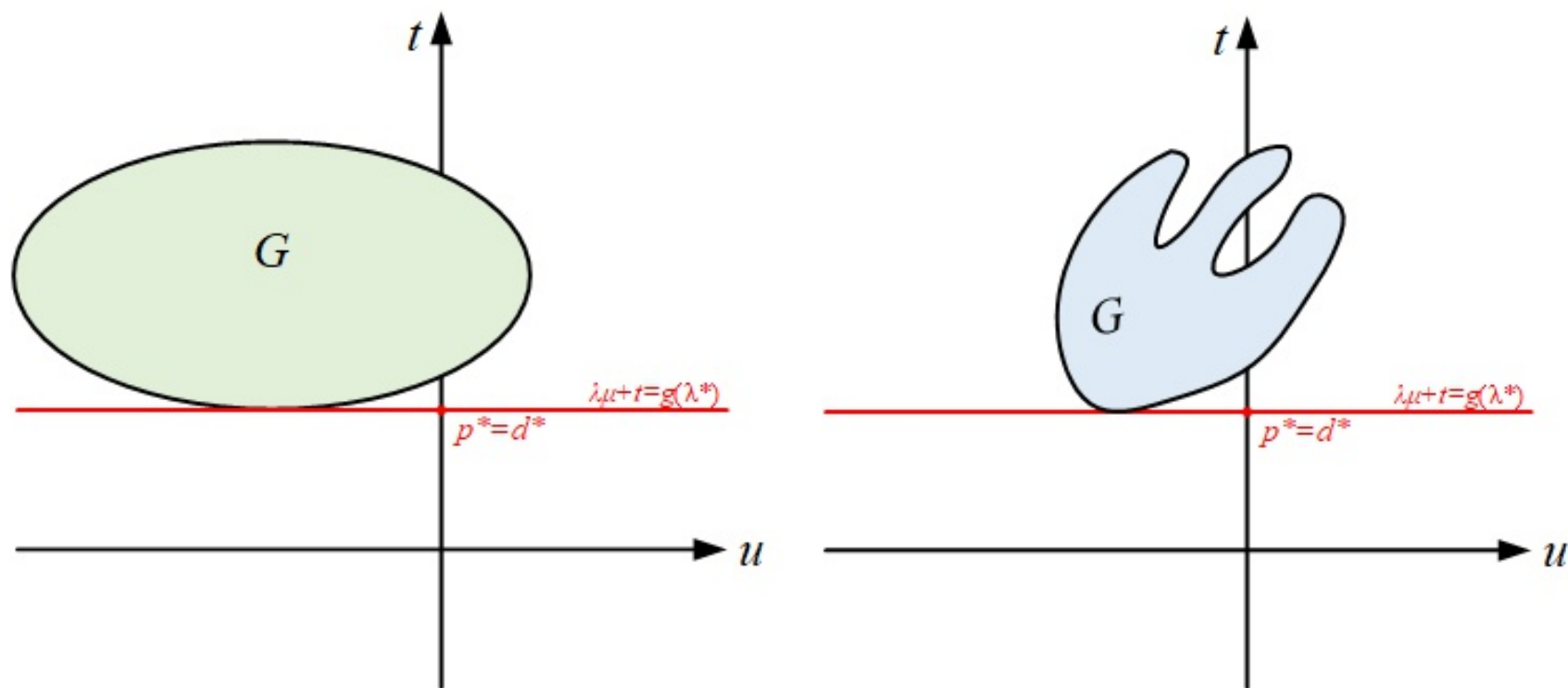


**strong duality**

- holds if there is a non-vertical supporting hyperplane to  $\mathcal{A}$  at  $(0, p^*)$
- for convex problem,  $\mathcal{A}$  is convex, hence has supp. hyperplane at  $(0, p^*)$
- Slater's condition: if there exist  $(\tilde{u}, \tilde{t}) \in \mathcal{A}$  with  $\tilde{u} < 0$ , then supporting hyperplanes at  $(0, p^*)$  must be non-vertical

## 4.3. Slater 条件与 KKT 条件

$G$  是凸集的情况下，最优对偶间隙为 0，成为强对偶



强对偶不一定要要求  $G$  是凸集

## 4.3. Slater 条件与 KKT 条件

Inequality form LP: Slater 条件本质说的是没有 duality gap, 但原/对偶问题能否同时取有意义的值, 还需要对偶问题有解

**primal problem**

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

**dual function**

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

**dual problem**

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

- from Slater's condition:  $p^* = d^*$  if  $A\tilde{x} \prec b$  for some  $\tilde{x}$
- in fact,  $p^* = d^*$  except when primal and dual are infeasible

注意!

## 4.3. Slater 条件与 KKT 条件

### Quadratic program

**primal problem** (assume  $P \in \mathbf{S}_{++}^n$ )

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

**dual function**

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

**dual problem**

$$\begin{array}{ll}\text{maximize} & -(1/4) \lambda^T A P^{-1} A^T \lambda - b^T \lambda \\ \text{subject to} & \lambda \succeq 0\end{array}$$

- from Slater's condition:  $p^* = d^*$  if  $A\tilde{x} \prec b$  for some  $\tilde{x}$
- in fact,  $p^* = d^*$  always

## 4.3. Slater 条件与 KKT 条件

A nonconvex problem with strong duality

$$\begin{array}{ll}\text{minimize} & x^T A x + 2b^T x \\ \text{subject to} & x^T x \leq 1\end{array}$$

$A \not\succeq 0$ , hence nonconvex

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

- unbounded below if  $A + \lambda I \not\succeq 0$  or if  $A + \lambda I \succeq 0$  and  $b \notin \mathcal{R}(A + \lambda I)$
- minimized by  $x = -(A + \lambda I)^\dagger b$  otherwise:  $g(\lambda) = -b^T (A + \lambda I)^\dagger b - \lambda$

**dual problem** and equivalent SDP:

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

$$\begin{array}{ll}\text{maximize} & -t - \lambda \\ \text{subject to} & \begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \succeq 0\end{array}$$

strong duality although primal problem is not convex (not easy to show)



## 4.3. Slater 条件与 KKT 条件

凸问题不满足强对偶性的例子

$$\min \left\{ e^{-x} \mid \text{s.t. } \frac{x^2}{y} \leq 0 \right\}, \quad D = \{(x, y) \mid y > 0\}$$

按定义计算或作图  $G = \left\{ (u, t) \mid u = \frac{x^2}{y}, t = e^{-x}, y > 0 \right\}$  均可得

$$d^* = 0, \quad p^* = 1$$

Lagrange 函数  $L(x, y, \lambda_1, \lambda_2) = e^{-x} + \lambda_1 \frac{x^2}{y} - \lambda_2 y$

## 4.3. Slater 条件与 KKT 条件

如果  $\lambda_2 > 0$  ,  $\lim_{x=\text{constant}, y \rightarrow +\infty} L(x, y, \lambda_1, \lambda_2) = \lim_{y \rightarrow +\infty} -\lambda_2 y = -\infty$

如果  $\lambda_2 < 0$  ,  $\lim_{x=\text{constant}, y \rightarrow -\infty} L(x, y, \lambda_1, \lambda_2) = \lim_{y \rightarrow -\infty} -\lambda_2 y = -\infty$

因此, 有意义的下确界只能是  $\lambda_2 = 0$  时取得

如果  $\lambda_2 = 0$  ,  $\lambda_1 > 0$  ,  $\lim_{x=\text{constant}, y \rightarrow 0^-} L(x, y, \lambda_1, \lambda_2) = -\infty$

如果  $\lambda_2 = 0$  ,  $\lambda_1 < 0$  ,  $\lim_{x=\text{constant}, y \rightarrow 0^+} L(x, y, \lambda_1, \lambda_2) = -\infty$

$$g(\lambda_1, \lambda_2) = \inf_{x, y} L(x, y, \lambda_1, \lambda_2)$$

因此, 对偶函数  $= \begin{cases} \inf_{x, y} e^{-x} = \lim_{x \rightarrow +\infty} e^{-x} = 0 & \text{if } \lambda_1 = \lambda_2 = 0 \\ -\infty & \text{otherwise} \end{cases}$

## 4.3. Slater 条件与 KKT 条件

非凸问题满足强对偶性的例子

$$\min \left\{ x^2 - y^2 \mid \text{s.t. } x^2 + y^2 \leq 1 \right\}$$

按定义计算或作图  $G = \{(u, t) \mid u = x^2 + y^2 - 1, t = x^2 - y^2\}$  均可得

$$d^* = -1, \quad p^* = -1$$

Lagrange 函数  $L(x, y, \lambda) = x^2 - y^2 + \lambda(x^2 + y^2 - 1)$

$$\begin{aligned} g(\lambda) &= \inf_{x, y} L(x, y, \lambda) = \inf_{x, y} \left[ (1 + \lambda)x^2 + (\lambda - 1)y^2 - \lambda \right] \\ &= \inf_{x=0, y} \left[ (\lambda - 1)y^2 - \lambda \right] \end{aligned}$$

对偶函数  $= \begin{cases} -1 & \text{if } \lambda - 1 \leq 0, \text{ then } y \text{ should be } \pm 1 \\ -\infty & \text{otherwise} \end{cases}$

## 4.3. Slater 条件与 KKT 条件

简化的 Slater 条件证明：第一，假设内部非空，而 Slater 条件只假设点在相对内部中；第二，优化目标没有仿射不等式约束。

若  $p^* = -\infty$ ，由弱对偶性知结论成立，以下假定  $p^* > -\infty$

用  $\bar{A}x = \bar{b}$  表示在  $Ax = b$  中删除冗余等式后的等式约束，这意味着两者具有相同的可行集，并且  $\bar{A}$  的行向量线性无关（如果  $A$  的行向量线性无关，则有  $\bar{A} = A$ ， $\bar{b} = b$ ）

定义两个集合（容易验证都是凸集）

$$\Omega_1 = \bigcup_{x \in D} \left\{ (u, v, t) \in R^{m \times p \times 1} \mid f_i(x) \leq u_i, i = 1, \dots, m, \bar{A}x - \bar{b} = v, f_0(x) \leq t \right\}$$

$$\Omega_2 = \left\{ (0, 0, s) \in R^{m \times p \times 1} \mid s < p^* \right\}$$

## 4.3. Slater 条件与 KKT 条件

$p^*$  是原问题可行目标的下确界  $\Rightarrow \Omega_1 \cap \Omega_2$  为空集

我们用反证法说明这点。如果我们假设  $(u, v, t) \in \Omega_1 \cap \Omega_2$ ，则  $(u, v, t) \in \Omega_2$  可以推出  $u = v = 0$ ， $t \leq p^*$ 。

同时  $(u, v, t) \in \Omega_1$ ，则存在  $x$  使得  $f_i(x) \leq 0$ ， $\bar{A}x - \bar{b} = 0$ ， $f_0(x) \leq t \leq p^*$ 。这和  $p^*$  是原问题可行目标的下确界相矛盾。

根据凸集分离定理，我们知道存在一个超平面的法线方向  $(\tilde{\lambda}, \tilde{v}, \mu) \neq 0$ ，和截距参数  $\alpha$  使得

$$\tilde{\lambda}^T u + \tilde{v}^T v + \mu t \geq \alpha \quad \forall (u, v, t) \in \Omega_1 \quad (4.6)$$

### 4.3. Slater 条件与 KKT 条件

$$\tilde{\lambda}^T u + \tilde{v}^T v + \mu t < \alpha \quad \forall (u, v, t) \in \Omega_2 \quad (4.7)$$

从(4.6)可以得知  $\tilde{\lambda} \geq 0, \mu \geq 0$ 。否则  $\tilde{\lambda}^T u + \mu t$  在  $\Omega_1$  集合中可以趋向  $-\infty$ ，这和其有下界  $\alpha$  矛盾。同时，(4.7) 表明  $\mu t < \alpha \quad \forall t < p^*$ ，所以  $\mu p^* \leq \alpha$ 。

因此，对于  $(f_1(x), \dots, f_m(x), \bar{A}x - \bar{b}, f_0(x)) \in \Omega_1, \forall x \in D$ ，我们可以得到

$$\sum_{1 \leq i \leq m} \tilde{\lambda}_i f_i(x) + \tilde{v}^T (\bar{A}x - \bar{b}) + \mu f_0(x) \geq \alpha \geq \mu p^*, \forall x \in D \quad (4.8)$$

下面说明  $\mu \neq 0$ ，因此一定有  $\mu > 0$ ！

如果  $\mu = 0$ ，代入(4.8)两端，我们可以得到

$$\sum_{1 \leq i \leq m} \tilde{\lambda}_i f_i(x) + \tilde{v}^T (\bar{A}x - \bar{b}) \geq 0, \quad \forall x \in D \quad (4.9)$$

对于满足 Slater 的点  $\hat{x} \in D$ ,  $\bar{A}\hat{x} = \bar{b}$ ，我们可以从(4.9)得出

$$\sum_{1 \leq i \leq m} \tilde{\lambda}_i f_i(\hat{x}) \geq 0。而此时  $\tilde{\lambda}_i \geq 0, f_i(\hat{x}) < 0, 1 \leq i \leq m$ ，因此  $\tilde{\lambda} = 0$ 。$$

而前述  $(\tilde{\lambda}, \tilde{v}, \mu) \neq 0$ ，且  $\tilde{\lambda} = 0, \mu = 0$ ，则必有  $\tilde{v} \neq 0$ 。代入(4.9)，对于  $\forall x \in D$ ，我们有  $\tilde{v}^T (\bar{A}x - \bar{b}) \geq 0$ 。但同时对于满足 Slater 的点  $\hat{x} \in D$ ,  $\bar{A}\hat{x} = \bar{b}$ ，则  $\tilde{v}^T (\bar{A}\hat{x} - \bar{b}) = 0$ 。而由于  $\hat{x} \in \text{int } D, \bar{A}\hat{x} = \bar{b}$ ，必然有  $\hat{x} \in D$  满足  $\tilde{v}^T (\bar{A}\hat{x} - \bar{b}) < 0$ ，除非  $\tilde{v}^T \bar{A} = 0$ 。而我们假设的  $\bar{A}$  行向量线性无关，也就是  $\tilde{v} = 0$ 。这和  $(\tilde{\lambda}, \tilde{v}, \mu) = 0$  矛盾！

### 4.3. Slater 条件与 KKT 条件

由  $\mu > 0$  可得 
$$\sum_{1 \leq i \leq m} \frac{\tilde{\lambda}_i}{\mu} f_i(x) + \frac{\tilde{v}^T}{\mu} (\bar{A}x - \bar{b}) + f_0(x) \geq p^*, \quad \forall x \in D$$

令  $\bar{v}$  为对  $\tilde{v}$ （在和  $A$  中被删除的行向量对应的位置）补充了 0 分量后的向量，则有

$$\sum_{1 \leq i \leq m} \frac{\tilde{\lambda}_i}{\mu} f_i(x) + \frac{\bar{v}^T}{\mu} (Ax - b) + f_0(x) \geq p^*, \quad \forall x \in D$$

进一步，令  $\lambda = \frac{\tilde{\lambda}}{\mu}$ ， $v = \frac{\bar{v}}{\mu}$ ，我们可以达到

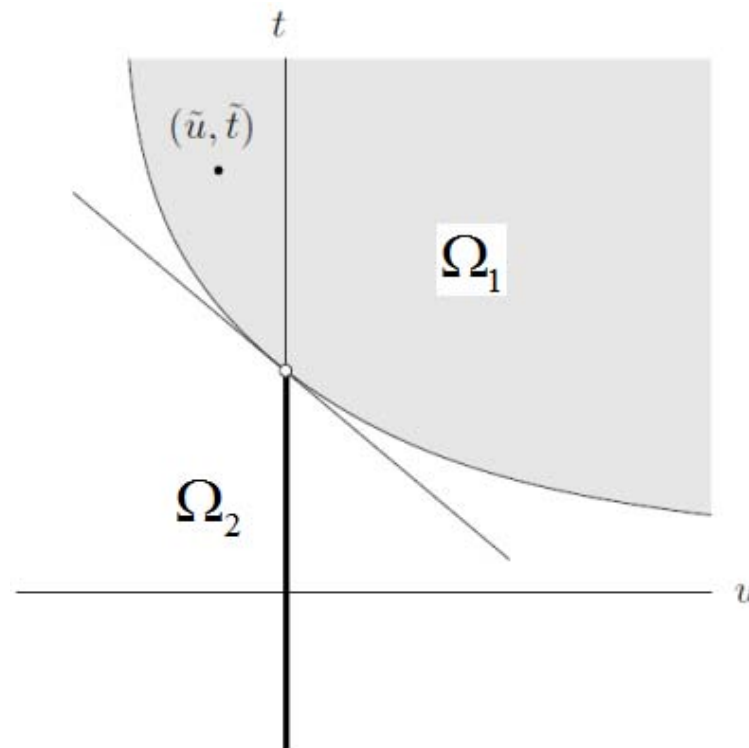
$$d^* \geq g(\lambda, v) = \inf_{x \in D} L(x, \lambda, v) \geq p^* \quad (4.9)$$

结合弱对偶性可知强对偶性成立。



## 4.3. Slater 条件与 KKT 条件

对于凸优化问题，Slater's constraint qualification guarantees that any separating hyperplane must be nonvertical, since it must pass to the left of the point  $(\tilde{u}, \tilde{v}) = (f_1(\tilde{x}), f_0(\tilde{x}))$ , where  $\tilde{x}$  is strictly feasible.



## 4.3. Slater 条件与 KKT 条件

### Slater 定理的第一个改进

如果 Slater 约束品性的  $\hat{x}$  是  $D$  的相对内点, 结论仍然成立

证明:  $\hat{x}$  是  $D$  的相对内点意味着  $D$  的仿射包  $\text{aff } D$  不等于全空间, 因此, 存在  $Q \in R^{n \times q}$  (对应线性空间的基矩阵) 满足

$$\text{aff } D = \{x \mid x = \hat{x} + Qy, y \in R^q\}$$

令  $\bar{f}_i(y) = f_i(\hat{x} + Qy)$ ,  $\bar{D} = \{y \in R^q \mid \hat{x} + Qy \in D\}$ , 则  $\hat{y} = 0 \in R^q$  是  $\bar{D}$  的内点, 满足  $\bar{f}_i(\hat{y}) < 0, i = 1, 2, \dots, m$ , 并且,  $p^*$  是以下问题的最优值

$$\min_{y \in \bar{D}} \left\{ \bar{f}_0(y) \mid \text{s.t. } \bar{f}_i(y) \leq 0, 1 \leq i \leq m, \bar{A}y = 0 \right\}$$

其中  $\bar{A} = AQ$ 。本质可以转为对行满秩的  $\bar{A}$  来证明, 过程类前。

## 4.3. Slater 条件与 KKT 条件

用  $\bar{d}$  表示上述问题的对偶问题的最大值，根据前面证明的结论，可知  $\bar{d} = p^*$ 。

$$\text{令 } \bar{g}(\lambda, \nu) = \inf_{y \in \bar{D}} \bar{f}_0(y) + \sum_{1 \leq i \leq m} \bar{f}_i(y) \lambda_i + \nu^T \bar{A}y$$

$$\begin{aligned} \Rightarrow \bar{g}(\lambda, \nu) &= \inf_{\hat{x} + Qy \in D} f_0(\hat{x} + Qy) + \sum_{1 \leq i \leq m} f_i(\hat{x} + Qy) \lambda_i + \nu^T (A(\hat{x} + Qy) - b) \\ &= \inf_{x \in D} f_0(x) + \sum_{1 \leq i \leq m} f_i(x) \lambda_i + \nu^T (Ax - b) \\ &= g(\lambda, \nu) \end{aligned}$$

$$\Rightarrow d^* = \sup_{\lambda \geq 0} g(\lambda, \nu) = \sup_{\lambda \geq 0} \bar{g}(\lambda, \nu) = \bar{d} = p^*$$

结论成立。

## 4.3. Slater 条件与 KKT 条件

### Slater 定理的第二个改进

如果 Slater 约束品性的  $\hat{x}$  只对非线性不等式成为严格不等式，即线性不等式可以是等式，结论仍然成立

证明：用  $f_i(x) = \bar{a}_i^T x - \bar{b}_i, i = 1, \dots, \bar{m} \leq m$  表示在  $\hat{x}$  处起作用，即  $\bar{a}_i^T \hat{x} - \bar{b}_i = 0$  的线性不等式约束。考虑等式和不等式方程组

$$\bar{a}_i^T d < 0, i = 1, \dots, \bar{m}, Ad = 0$$

如果  $d$  是该方程组的解，取充分小的  $t > 0$ ，令  $\hat{x}' = \hat{x} + td$ ，容易验证， $\hat{x}'$  是满足 Slater 条件的可行解，结论成立。

## 4.3. Slater 条件与 KKT 条件

如果以上方程组无解，定义

$$C = \left\{ (\bar{y}, \hat{y}) \mid \bar{y}_i = \bar{a}_i^T d, i = 1, \dots, \bar{m}, \hat{y} = Ad, d \in R^n \right\}$$

$$D = \left\{ (\bar{z}, \hat{z}) \mid \bar{z}_i < 0, i = 1, \dots, \bar{m}, \hat{z} = 0 \right\}$$

显然，这两个集合是凸集，此时无交点，根据凸集分离定理，存在不全为零的  $(\lambda, \nu)$  满足

$$\left( \sum_{i=1}^{\bar{m}} \lambda_i \bar{a}_i^T + \nu^T A \right) d \leq \sum_{i=1}^{\bar{m}} \lambda_i \bar{z}_i, \forall d \in R^n, \bar{z}_i < 0$$

由上式可推出  $\sum_{i=1}^{\bar{m}} \lambda_i \bar{a}_i^T + \nu^T A = 0, \lambda_i \leq 0, \forall i$ 。由于可假设  $A$  行

满秩，必有某些  $\lambda_i \neq 0$ ，用  $I_+$  表示它们的集合，上式成为

$$\sum_{i \in I_+} \lambda_i \bar{a}_i^T + \nu^T A = 0, \lambda_i < 0, \forall i \quad \circ$$

## 4.3. Slater 条件与 KKT 条件

任取不等于  $\hat{x}$  的可行解  $x$ ，利用上式（分别乘  $\hat{x}$  和  $x$  再相减）又可得到

$$\sum_{i \in I_+} \lambda_i (\bar{a}_i^T x - \bar{b}_i) = 0, \quad \lambda_i < 0, \forall i$$

由此可知，对任意可行解  $x$  均成立  $\bar{a}_i^T x = \bar{b}_i$ ， $\forall i \in I_+$ 。于是，可以把这些不等式约束视为等式约束。

上述过程表明，我们总可做到，或者获得一个满足 Slater 条件的可行解，或者减少不满足 Slater 条件的不等式的数目，如此继续，有限递降，最终一定可以满足 Slater 条件，完成证明。

推论：任何有可行解的线性规划问题都满足强对偶性

## 4.3. Slater 条件与 KKT 条件

### Slater 定理的第三个改进

矩阵  $A$  行满秩的假设可以去掉，只需要证明如果不是行满秩（存在冗余等式），那么可以通过一个线性变换，使得约束行满秩，且经过变换之后 Lagrange 函数的最优值不变，即强对偶性不变。

现在考虑前面假设不成立的情况： $A$  的行向量线性相关

由于  $Ax = b$  有解  $\hat{x}$ ，可以将其分为两组， $A_1x = b_1$ ， $A_2x = b_2$ ，满足以下条件：1)  $A_1$  行向量线性无关；2)  $(A_2, b_2)$  的每行向量都可以表示成  $(A_1, b_1)$  的行向量的线性组合，即有矩阵  $P$  使得  $A_2 = PA_1$ ， $b_2 = Pb_1$ ，此时，原问题等价于满足前面假设的问题

$$\min \left\{ f_0(x) \mid \text{s.t. } f_i(x) \leq 0, 1 \leq i \leq m, A_1x = b_1 \right\}$$

该问题的最小值显然等于原问题最小值  $p^*$ ，下面说明，其对偶问题的最大值，记为  $\bar{d}$ ，也等于  $d^*$ ，从而完成定理的证明

$$\text{令 } \bar{g}(\lambda, \bar{\mu}) = \inf_{x \in D} f_0(x) + \sum_{1 \leq i \leq m} f_i(x) \lambda_i + \bar{\mu}^T (A_1 x - b_1)$$

$$\begin{aligned} \Rightarrow g(\lambda, \mu) &= \inf_{x \in D} f_0(x) + \sum_{1 \leq i \leq m} f_i(x) \lambda_i + \mu^T (Ax - b) \\ &= \inf_{x \in D} f_0(x) + \sum_{1 \leq i \leq m} f_i(x) \lambda_i + \mu_1^T (A_1 x - b_1) + \mu_2^T (A_2 x - b_2) \\ &= \inf_{x \in D} f_0(x) + \sum_{1 \leq i \leq m} f_i(x) \lambda_i + (\mu_1^T + \mu_2^T P)(A_1 x - b_1) \\ &= \bar{g}(\lambda, \mu_1 + P^T \mu_2), \quad \text{其中 } \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \end{aligned}$$

$$\Rightarrow d^* = \sup_{\lambda \geq 0} g(\lambda, \mu) = \sup_{\lambda \geq 0} \bar{g}(\lambda, \mu_1 + P\mu_2) = \sup_{\lambda \geq 0} \bar{g}(\lambda, \bar{\mu}) = \bar{d}$$

推论：Slater 条件成立时，如果原问题有最优解，则对偶问题也有最优解。



## 4.3. Slater 条件与 KKT 条件

**Karush-Kuhn-Tucker (KKT) 条件**进一步刻画了可微约束优化问题（不一定凸优化问题）最优解应该满足的必要条件，而 Slater 条件是凸优化问题强对偶性的充分条件

考虑一般性（可微）优化问题

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{aligned}$$

假设：1) 原问题有最优解  $x^*$ ，2) 原对偶问题满足强对偶性，于是，存在  $\lambda^* \geq 0, \mu^*$  和  $x^*$  一起满足

$$-\infty < f_0(x^*) = p^* = d^* = g(\lambda^*, \mu^*)$$

隐含了原/对偶问题都有有界解

## 4.3. Slater 条件与 KKT 条件

此时可进行如下推导：

$$\begin{aligned}\Rightarrow \quad g(\lambda^*, \mu^*) &\leq L(x^*, \lambda^*, \mu^*) \\ &= f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \mu_i^* h_i(x^*) \\ &\leq f_0(x^*) \\ &= g(\lambda^*, \mu^*)\end{aligned}$$

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

$$L(x^*, \lambda^*, \mu^*) = g(\lambda^*, \mu^*) = \inf_{x \in D} f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \mu_i^* h_i(x)$$

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

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

## 4.3. Slater 条件与 KKT 条件

结论：如果强对偶性成立，原对偶最优对  $(x^*, \lambda^*, \mu^*)$  必须满足以下 4 个等式不等式方程称为 **Karush-Kuhn-Tucker (KKT) 条件**，是满足强对偶性的原对偶最优对的**必要条件**：

原问题可行条件：

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

对偶问题可行条件：

$$\lambda_i^* \geq 0, \quad i = 1, \dots, m$$

互补松弛 Complementary slackness：

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

拉格朗日不动性：

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

## 4.3. Slater 条件与 KKT 条件

### KKT 反向结论

如果  $\frac{\partial L(x^*, \lambda^*, \mu^*)}{\partial x} = 0$  可以保证  $L(x^*, \lambda^*, \mu^*) = \inf_{x \in D} L(x, \lambda^*, \mu^*)$ , 那么

KKT 条件是  $(x^*, \lambda^*, \mu^*)$  为原对偶最优对的充分条件

理由: KKT 条件的最后一个方程可保证  $g(\lambda^*, \mu^*) = L(x^*, \lambda^*, \mu^*)$ , 再结合其它条件可得  $g(\lambda^*, \mu^*) = f_0(x^*)$ , 由弱对偶性可得结论

推论: 如果  $\nabla f_0(x) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x) + \sum_{i=1}^p \mu_i^* \nabla h_i(x) = 0$  的解唯一, 或者原

问题是凸问题, 那么 KKT 条件是原对偶最优对的充分条件

## 4.3. Slater 条件与 KKT 条件

其它推论:

If  $\tilde{x}, \tilde{\lambda}, \tilde{\mu}$  satisfy KKT for a convex problem, then they are optimal:

- from complementary slackness:  $f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\mu})$
  - from 4<sup>th</sup> condition (and convexity):  $g(\tilde{\lambda}, \tilde{\mu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\mu})$
- hence,  $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\mu})$

If **Slater's condition** is satisfied:  $x$  is optimal if and only if there exist  $\lambda, \nu$  that satisfy KKT conditions

- recall that Slater implies strong duality, and dual optimum is attained
- generalizes optimality condition  $\nabla f_0(x) = 0$  for unconstrained problem

## 4.3. Slater 条件与 KKT 条件

对于一般性优化问题，如果在最优解  $x^*$  处起作用的不等式约束（ $f_i(x^*)=0$  的不等式约束）和等式约束的梯度一起线性无关，那么一定有对偶向量  $(\lambda^*, \mu^*)$  和  $x^*$  一起满足 KKT 条件

对于仅含不等式约束的问题，利用 Gordan 定理可得此结论

对于一般性优化问题，再利用隐函数定理可得此结论

结论：KKT 条件是非常广泛的优化问题最优解的必要条件

## 4.3. Slater 条件与 KKT 条件

**example: water-filling** (assume  $\alpha_i > 0$ )

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^n \log(x_i + \alpha_i) \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1 \end{array}$$

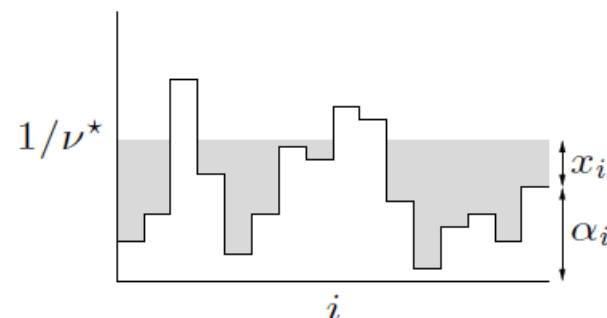
$x$  is optimal iff  $x \succeq 0$ ,  $\mathbf{1}^T x = 1$ , and there exist  $\lambda \in \mathbf{R}^n$ ,  $\nu \in \mathbf{R}$  such that

$$\lambda \succeq 0, \quad \lambda_i x_i = 0, \quad \frac{1}{x_i + \alpha_i} + \lambda_i = \nu$$

- if  $\nu < 1/\alpha_i$ :  $\lambda_i = 0$  and  $x_i = 1/\nu - \alpha_i$
- if  $\nu \geq 1/\alpha_i$ :  $\lambda_i = \nu - 1/\alpha_i$  and  $x_i = 0$
- determine  $\nu$  from  $\mathbf{1}^T x = \sum_{i=1}^n \max\{0, 1/\nu - \alpha_i\} = 1$

**interpretation**

- $n$  patches; level of patch  $i$  is at height  $\alpha_i$
- flood area with unit amount of water
- resulting level is  $1/\nu^*$



## 4.3. Slater 条件与 KKT 条件

已知  $a, b > 0$ ,  $a + b = 1$ , 求  $\frac{1}{a} + \frac{4}{b}$  的最大值



### 4.3. Slater 条件与 KKT 条件

**解** 附加条件可化为： $a + b - 1 = 0$ ，则

$$z = f(a, b) = \frac{1}{a} + \frac{4}{b}$$

$$\varphi(a, b) = a + b - 1 = 0$$

$$F(a, b, \lambda) = f(x, y) + \lambda \varphi(x, y)$$

$$= \frac{1}{a} + \frac{4}{b} + \lambda(a + b - 1)$$

### 4.3. Slater 条件与 KKT 条件

对  $a$  、  $b$  求导得

$$F'_a(a, b, \lambda) = -\frac{1}{a^2} + \lambda = 0 \quad (1)$$

$$F'_b(a, b, \lambda) = -\frac{4}{b^2} + \lambda = 0 \quad (2)$$

$$\varphi(a, b) = a + b - 1 = 0 \quad (3)$$

联立方程组得  $\lambda = \frac{1}{a^2} = \frac{4}{b^2}$  ,  $b = 2a$

又因为  $a + b - 1 = 0$  , 解得  $a = \frac{1}{3}$  ,  $b = \frac{2}{3}$

所以  $z_{min} = f(\frac{1}{3}, \frac{2}{3}) = 9$

## 4.4. 灵敏度与择一理论

用  $p^*(u, v)$  表示下述扰动问题的最优值， $p^*(0, 0)$  就是前面的  $p^*$

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq u_i, \quad i = 1, \dots, m \\ & h_i(x) = v_i, \quad i = 1, \dots, p \end{aligned}$$

若无扰动问题满足强对偶性：  $p^*(0, 0) = g(\lambda^*, \mu^*)$

$$\Rightarrow \quad p^*(u, v) \geq p^*(0, 0) - (\lambda^*)^T u - (\mu^*)^T v$$

再加上  $p^*(u, v)$  在  $(0, 0)$  处可导

$$\Rightarrow \quad \frac{\partial p^*(0, 0)}{\partial u} = -\lambda^*, \quad \frac{\partial p^*(0, 0)}{\partial v} = -\mu^* \quad (\text{影子价格})$$

## 4.4. 灵敏度与择一理论

**local sensitivity:** if (in addition)  $p^*(u, v)$  is differentiable at  $(0, 0)$ , then

$$\lambda_i^* = -\frac{\partial p^*(0, 0)}{\partial u_i}, \quad \nu_i^* = -\frac{\partial p^*(0, 0)}{\partial v_i}$$

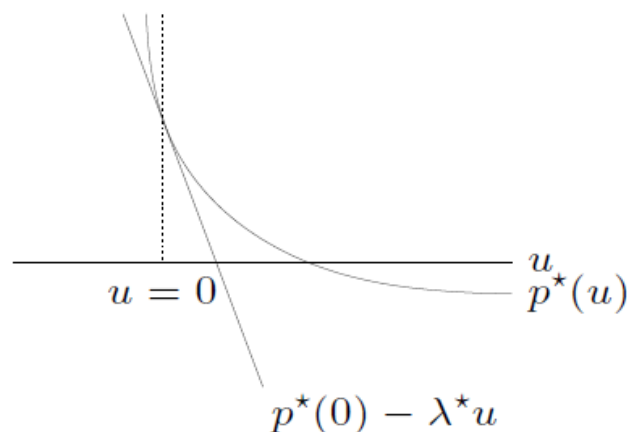
proof (for  $\lambda_i^*$ ): from global sensitivity result,

$$\frac{\partial p^*(0, 0)}{\partial u_i} = \lim_{t \searrow 0} \frac{p^*(te_i, 0) - p^*(0, 0)}{t} \geq -\lambda_i^*$$

$$\frac{\partial p^*(0, 0)}{\partial u_i} = \lim_{t \nearrow 0} \frac{p^*(te_i, 0) - p^*(0, 0)}{t} \leq -\lambda_i^*$$

hence, equality

$p^*(u)$  for a problem with one (inequality)  
constraint:



## 4.4. 灵敏度与择一理论

择一理论，类似 Farkas 引理，可视为其非线性化版本

**弱择一定理**（两组方程至多一组有解）

$f_i(x) \leq 0, 1 \leq i \leq m, h_i(x) = 0, 1 \leq i \leq p$  和  $\lambda \geq 0, g(\lambda, \mu) > 0$  弱择一

$f_i(x) < 0, 1 \leq i \leq m, h_i(x) = 0, 1 \leq i \leq p$  和  $\lambda \geq 0, \lambda \neq 0, g(\lambda, \mu) \geq 0$  弱择一

其中  $g(\lambda, \mu) = \inf_{x \in D} \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \mu_i h_i(x)$  是以下问题的对偶函数

$$\min \{0 \mid \text{s.t. } f_i(x) \leq 0, i = 1, \dots, m, h_i(x) = 0, i = 1, \dots, p\}$$

## 4.4. 灵敏度与择一理论

对凸问题，可得强择一定理（两组方程正好一组有解）

$f_i(x) \leq 0, i = 1, \dots, m, Ax = b$  和  $\lambda \geq 0, g(\lambda, \mu) > 0$  强择一

$f_i(x) < 0, i = 1, \dots, m, Ax = b$  和  $\lambda \geq 0, \lambda \neq 0, g(\lambda, \mu) \geq 0$  强择一

其中所有  $f_i$  是凸函数，且存在  $D$  的相对内点  $\tilde{x}$  使  $A\tilde{x} = b$

## 4.5. 对偶问题和问题变形

进一步讨论为什么要研究对偶问题

- equivalent formulations of a problem can lead to very different duals
- reformulating the primal problem can be useful when the dual is difficult to derive, or uninteresting

常用对偶问题变形技巧

- introduce new variables and equality constraints
- make explicit constraints implicit or vice-versa
- transform objective or constraint functions  
e.g. replace  $f_0(x)$  by  $\phi(f_0(x))$  with  $\phi$  convex, increasing

## 4.5. 对偶问题和问题变形

Introducing new variables and equality constraints

$$\text{minimize } f_0(Ax + b)$$

- dual function is constant:  $g = \inf_x L(x) = \inf_x f_0(Ax + b) = p^*$
- we have strong duality, but dual is quite useless

**reformulated problem and its dual**

$$\begin{array}{ll} \text{minimize} & f_0(y) \\ \text{subject to} & Ax + b - y = 0 \end{array}$$

$$\begin{array}{ll} \text{maximize} & b^T \nu - f_0^*(\nu) \\ \text{subject to} & A^T \nu = 0 \end{array}$$

dual function follows from

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



## 4.5. 对偶问题和问题变形

**norm approximation problem:** minimize  $\|Ax - b\|$

$$\begin{array}{ll}\text{minimize} & \|y\| \\ \text{subject to} & y = Ax - b\end{array}$$

can look up conjugate of  $\|\cdot\|$ , or derive dual directly

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

(see page 5-4)

**dual of norm approximation problem**

$$\begin{array}{ll}\text{maximize} & b^T \nu \\ \text{subject to} & A^T \nu = 0, \quad \|\nu\|_* \leq 1\end{array}$$

## 4.5. 对偶问题和问题变形

### Implicit constraints

**LP with box constraints:** primal and dual problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & -\mathbf{1} \preceq x \preceq \mathbf{1} \end{array} \qquad \begin{array}{ll} \text{maximize} & -b^T \nu - \mathbf{1}^T \lambda_1 - \mathbf{1}^T \lambda_2 \\ \text{subject to} & c + A^T \nu + \lambda_1 - \lambda_2 = 0 \\ & \lambda_1 \succeq 0, \quad \lambda_2 \succeq 0 \end{array}$$

**reformulation with box constraints made implicit**

$$\begin{array}{ll} \text{minimize} & f_0(x) = \begin{cases} c^T x & -\mathbf{1} \preceq x \preceq \mathbf{1} \\ \infty & \text{otherwise} \end{cases} \\ \text{subject to} & Ax = b \end{array}$$

dual function

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

**dual problem:** maximize  $-b^T \nu - \|A^T \nu + c\|_1$

## 4.6. 广义不等式约束问题

对于广义不等式约束优化问题

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq_{K_i} 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{aligned}$$

定义其 Lagrange 函数（要求  $\lambda_i \geq_{K_i^*} 0, \quad i = 1, \dots, m$ ）

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

由此得到对偶函数

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

然后可得到同普通不等式约束对应的所有的结论

## 4.6. 广义不等式约束问题

**lower bound property:** if  $\lambda_i \succeq_{K_i^*} 0$ , then  $g(\lambda_1, \dots, \lambda_m, \mu) \leq p^*$

proof: if  $\tilde{x}$  is feasible and  $\lambda_i \succeq_{K_i^*} 0$ , then

$$\begin{aligned} f_0(\tilde{x}) &\geq f_0(\tilde{x}) + \sum_{i=1}^m \lambda_i^T f_i(\tilde{x}) + \sum_{i=1}^p \mu_i h_i(\tilde{x}) \\ &\geq \inf_{x \in D} L(x, \lambda_1, \dots, \lambda_m, \mu) \\ &= g(\lambda_1, \dots, \lambda_m, \mu) \end{aligned}$$

minimizing over all feasible  $\tilde{x}$  gives  $p^* \geq g(\lambda_1, \dots, \lambda_m, \mu)$

**dual problem**                      maximize  $g(\lambda_1, \dots, \lambda_m, \mu)$   
   subject to  $\lambda_i \succeq_{K_i^*} 0, \quad i = 1, \dots, m$

- weak duality:  $p^* \geq d^*$  always
- strong duality:  $p^* = d^*$  for convex problem with constraint qualification (for example, Slater's: primal problem is strictly feasible)

## 4.6. 广义不等式约束问题

### Semidefinite program

**primal SDP** ( $F_i, G \in \mathbf{S}^k$ )

$$\begin{array}{ll}\text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + \cdots + x_n F_n \preceq G\end{array}$$

- Lagrange multiplier is matrix  $Z \in \mathbf{S}^k$
- Lagrangian  $L(x, Z) = c^T x + \text{tr}(Z(x_1 F_1 + \cdots + x_n F_n - G))$
- dual function

$$g(Z) = \inf_x L(x, Z) = \begin{cases} -\text{tr}(GZ) & \text{tr}(F_i Z) + c_i = 0, \quad i = 1, \dots, n \\ -\infty & \text{otherwise} \end{cases}$$

**dual SDP**

$$\begin{array}{ll}\text{maximize} & -\text{tr}(GZ) \\ \text{subject to} & Z \succeq 0, \quad \text{tr}(F_i Z) + c_i = 0, \quad i = 1, \dots, n\end{array}$$

$p^* = d^*$  if primal SDP is strictly feasible ( $\exists x$  with  $x_1 F_1 + \cdots + x_n F_n \prec G$ )

## 4.7. References

- [1] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004. <http://www.stanford.edu/~boyd/cvxbook/>  
<http://www.ee.ucla.edu/~vandenbe/cvxbook>
- [2] S. Artstein-Avidan, V. Milman, "The concept of duality in convex analysis, and the characterization of the Legendre transform," *Annals of Mathematics*, vol. 169, pp. 661-674, 2009.