Convex Optimization Theory and Applications

Topic 4 - Duality

Li Li

Department of Automation
Tsinghua University

Fall, 2009-2021.

4.0. Outline

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

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

对偶问题

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

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

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

Theorem (Weak Duality Theorem)

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

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

Proof.

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

4.1. 线性规划问题的对偶理论

$$\boldsymbol{u}^T Ax \geq \boldsymbol{u}^T \boldsymbol{b} = \boldsymbol{b}^T \boldsymbol{u}$$

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

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

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

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, $\boldsymbol{b} \in \mathbb{R}^m$ is an m-dimensional column vector. Exactly one of the following linear system is feasible:

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

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

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, $\boldsymbol{b}^T\boldsymbol{y} < 0$ implies $\boldsymbol{y} \neq \mathbf{0}$ and $\boldsymbol{b} \neq \mathbf{0}$.

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

$$\boldsymbol{b}^T\boldsymbol{y} = (A\boldsymbol{x})^T\boldsymbol{y} = \boldsymbol{x}^T\left(A^T\boldsymbol{y}\right) \geq 0 \tag{16}$$

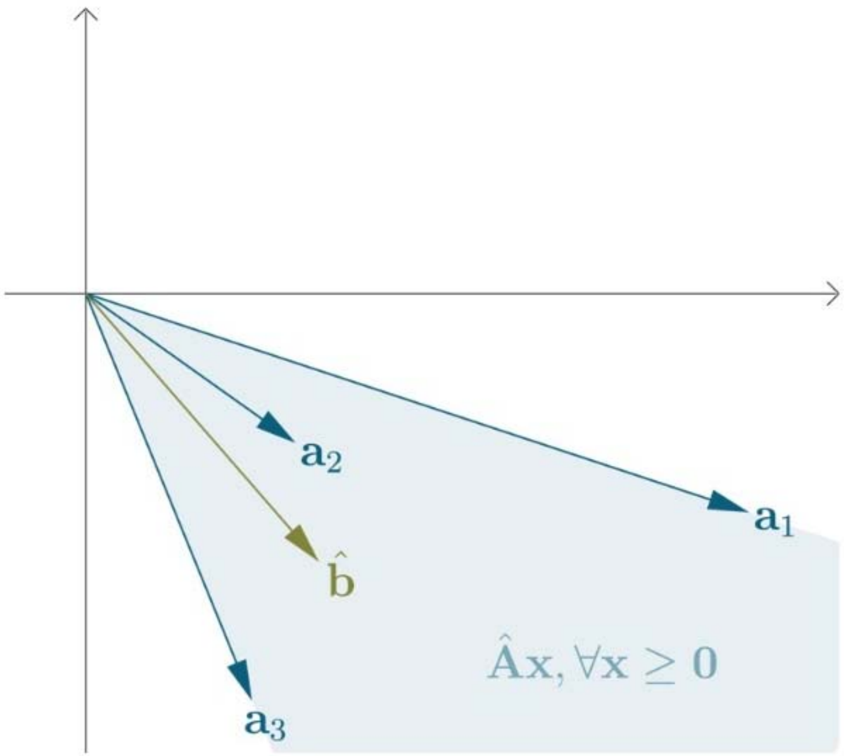
which contradicts $\boldsymbol{b}^T\boldsymbol{y} < 0$.

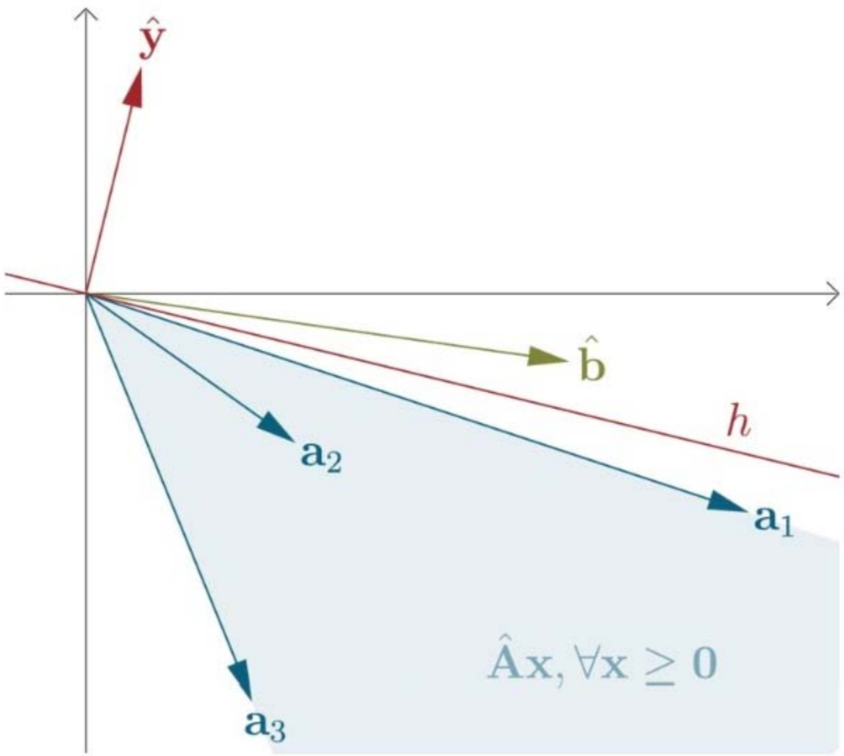
4.1. 线性规划问题的对偶理论

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 $\Omega = \{A\boldsymbol{x},\ \boldsymbol{x} \geq \mathbf{0}\}$ is a closed convex set. Moreover, $\boldsymbol{b} \notin \Omega$.
According to Separating Hyperplane Theorem, there exists a hyperplane $\boldsymbol{y}^T\boldsymbol{x} = z$ that separates $\boldsymbol{b}$ from Ω, where $\boldsymbol{y} = [y_1, \ldots, y_m]^T \in \mathbb{R}^m$ is an n-dimensional column vector. That is, $\boldsymbol{y}^T\boldsymbol{b} < z$ and $\forall\ \boldsymbol{s} \in \Omega$, $\boldsymbol{y}^T\boldsymbol{s} \geq z$.
Since $\mathbf{0} \in \Omega$, we have $z \leq 0$. As a result, $\boldsymbol{y}^T\boldsymbol{b} < 0$.
On the other hand, since $\boldsymbol{y}^T A\boldsymbol{x} > 0$ for all $\boldsymbol{x} \in \Omega$, we can see that $\boldsymbol{y}^T A > \mathbf{0}$, since each element of $\boldsymbol{x}$ can be arbitrarily large.
Therefore, we prove the whole statement. $\square$

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 $\boldsymbol{x}^*$ *to the primal problem is optimal if and only if there exists a feasible solution* $\boldsymbol{u}^*$ *to the dual such that*

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

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

Proof.

First, We prove the sufficiency.

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

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

which shows that $\boldsymbol{x}^*$ 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 $\boldsymbol{u}$ of the dual problem, we have

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

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

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

That is,

$$\begin{aligned} a_i^T\boldsymbol{x}^* &= b_i, \quad i \in I \\ a_i^T\boldsymbol{x}^* &> b_i, \quad i \notin I \end{aligned}$$

4.1. 线性规划问题的对偶理论

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

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

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

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

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

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

4.1. 线性规划问题的对偶理论

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

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

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

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

Based on Weak Duality Theorem, we see $\boldsymbol{u}$ 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

Dual \ Primal	*Infeasible*	*Optimal*	*Unbounded*
Infeasible	√	×	√
Optimal	×	√	×
Unbounded	√	×	×

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

$$\min\{f_0(x): f_i(x)\le 0, i=1,2,\ldots,m\} \tag{4.1}$$

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

$$\begin{cases} f_0(x)<v \\ f_i(x)\le 0, i=1,2,\ldots,m \end{cases} \tag{4.2}$$

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$\lambda\ge 0$，以下方程

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

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$\lambda\ge 0$，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

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

因为我们要找最好的下界，所以这个时候的v和λ应该取

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

由此引入了 dual problem。证明逻辑是根据式(4.5)取v和λ，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

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

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 λ_i, μ_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

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

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 λ_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 μ_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 μ_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∞，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∞。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 $D=\bigcap_{i=0,1,\ldots,m}\operatorname{dom} f_i\cap\bigcap_{i=0,1,\ldots,p}\operatorname{dom} h_i$ 是以上问题的优化变量的定义域

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

称以下问题为上述优化问题的对偶问题

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

称以下集合为上述对偶问题可行集

$$\{(\lambda,\mu)\mid \lambda\ge 0,(\lambda,\mu)\in\operatorname{dom} g\}$$

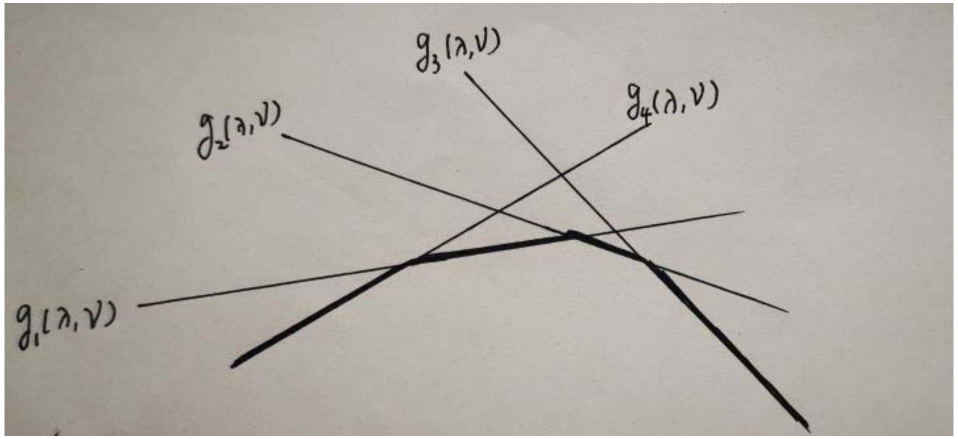
4.2. Definition and Examples

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λ和μ的，故用$g_i(\lambda,\mu)$来表示。所以

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

当L看成是关于λ或μ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λ和ν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：

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

即无穷多个x变化时，拉格朗日函数的最小值。
另外，由于把λ和ν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）：

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

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

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

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

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 $p^\star$, obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted $d^\star$
- λ, ν 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)

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

4.2. Definition and Examples

弱对偶性

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

设(x, λ, μ)是任意的**原对偶可行对**（指x是原问题可行解，λ, μ是对偶问题可行解），则总是成立

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

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

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

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

4.2. Definition and Examples

推论：设(x,λ,μ)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

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

如果存在原对偶可行对(x,λ,μ)满足$g(\lambda,\mu)=f_0(x)$，那么x是原问题最优解，λ,μ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 $\min \left\{ c^T x \,\middle|\, \text{s.t. } Gx \le h,\ Ax = b \right\}$

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 $\max \left\{ -h^T\lambda - b^T\mu \,\middle|\, \text{s.t. } c + G^T\lambda + A^T\mu = 0,\ \lambda \ge 0 \right\}$

对偶问题可行集 $\left\{ (\lambda,\mu) \,\middle|\, \lambda \ge 0,\ c + G^T\lambda + A^T\mu = 0 \right\}$

4.2. Definition and Examples

线性规划问题满足强对偶性

原问题 $\min\ \left\{ c^T x \middle| \text{ s.t. } Gx \le h,\ Ax = b \right\}$

对偶问题 $\max \left\{ -h^T \lambda - b^T \mu \middle| \text{ s.t. } c + G^T \lambda + A^T \mu = 0,\ \lambda \ge 0 \right\}$

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 λ, μ 一起满足

$$c + G^T \lambda + A^T \mu = 0,\ \ \lambda \ge 0;\ \ \lambda^T \left(Gx - h \right) = 0;\ \ Gx \le h,\ Ax = b$$

$\Rightarrow\ \ \lambda, \mu$ 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:

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

- plug in in L to obtain g:

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

a concave function of ν

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

4.2. Definition and Examples

Equality constrained norm minimization

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

dual function

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

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

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

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

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

lower bound property: $p^\star \ge b^T \nu$ if $\|A^T \nu\|_* \le 1$

4.2. Definition and Examples

Definition 5 (Dual norm). *Let* $||.||$ *be any norm. Its dual norm is defined as*

$$||x||_* = \max x^T y \quad s.t. \; ||y|| \le 1.$$

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:

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

对偶范数的等价定义

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

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

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

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$\frac{1}{p}+\frac{1}{q}=1$：

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

霍尔德（Hölder）不等式

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

$$x^{\frac{1}{p}} y^{\frac{1}{q}} \le \frac{x}{p} + \frac{y}{q},\ \forall\ x, y \ge 0,$$

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

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

两边同时去指数，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

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

得到 n 个不等式：

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

将上式两边对 $i=1,2,\cdots,n$ 求和，就得到

$$\frac{\sum_{i=1}^n a_ib_i}{(\sum_{j=1}^n a_j^p)^{\frac{1}{p}}(\sum_{j=1}^n b_j^q)^{\frac{1}{q}}}\le\frac{1}{p}+\frac{1}{q}=1,$$

$$\Rightarrow\sum_{i=1}^n a_ib_i\le(\sum_{j=1}^n a_j^p)^{\frac{1}{p}}(\sum_{j=1}^n b_j^q)^{\frac{1}{q}}$$

上式要求 $a_i,b_i\ge0$。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：

$$a^Tb\le||a||_p||b||_q$$

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

4.2. Definition and Examples

Please prove Dual norm of l_1 of is l_∞

4.2. Definition and Examples

Please prove Dual norm of l_1 of is l_∞

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

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

Finally since $||x||_1 \leq 1$, we have $z^T x \leq \max_{i=1}^n |z_i|$.

With these, I am able to show that l_∞ 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 $x = sign(z_i)e_i$. We have $\|x\| = 1$. Also,

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

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

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

- a nonconvex problem; feasible set contains 2^n discrete points
- interpretation: partition $\{1, \ldots, n\}$ 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

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

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

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

4.2. Definition and Examples

weak duality: $d^\star \le p^\star$

- 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: $d^\star = p^\star$

- 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

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

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

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

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 $p^\star > -\infty$)
- 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 $\min\ \left\{ f_0(x) \middle| \text{ s.t. } f_i(x) \le 0,\ 1 \le i \le m,\ Ax = b \right\}$，如果存在 $D = \bigcap_{i=0,1,\dots,m} \operatorname{dom} f_i$ 的内点 $\hat{x}$，满足 $f_i(\hat{x}) < 0,\ 1 \le i \le m,\ A\hat{x} = b$，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

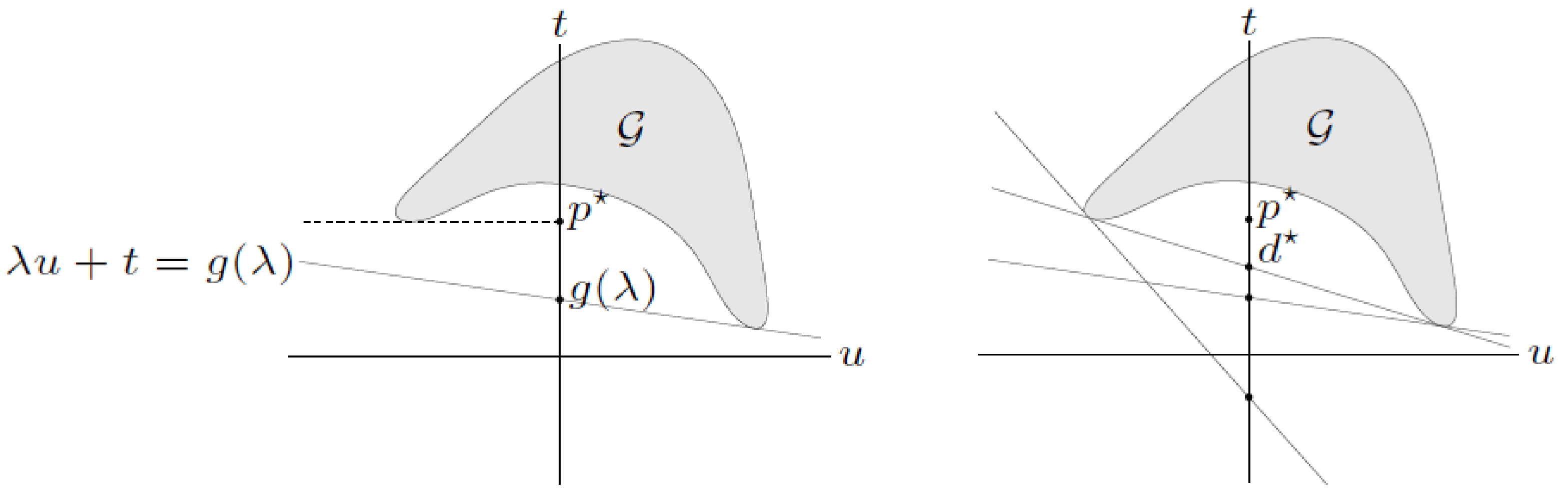
4.3. Slater 条件与 KKT 条件

对偶函数的几何解释：法线方向$(\lambda, u, 1)$与集合G相交的超平面的t-截距的下确界

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

interpretation of dual function:

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

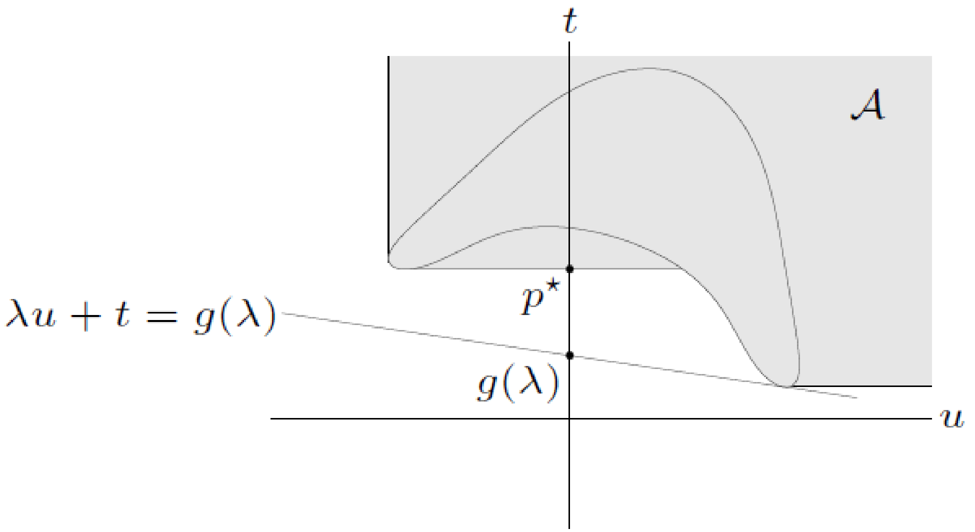


- $\lambda u + t = g(\lambda)$ is (non-vertical) supporting hyperplane to $\mathcal{G}$
- hyperplane intersects t-axis at $t = g(\lambda)$

4.3. Slater 条件与 KKT 条件

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

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

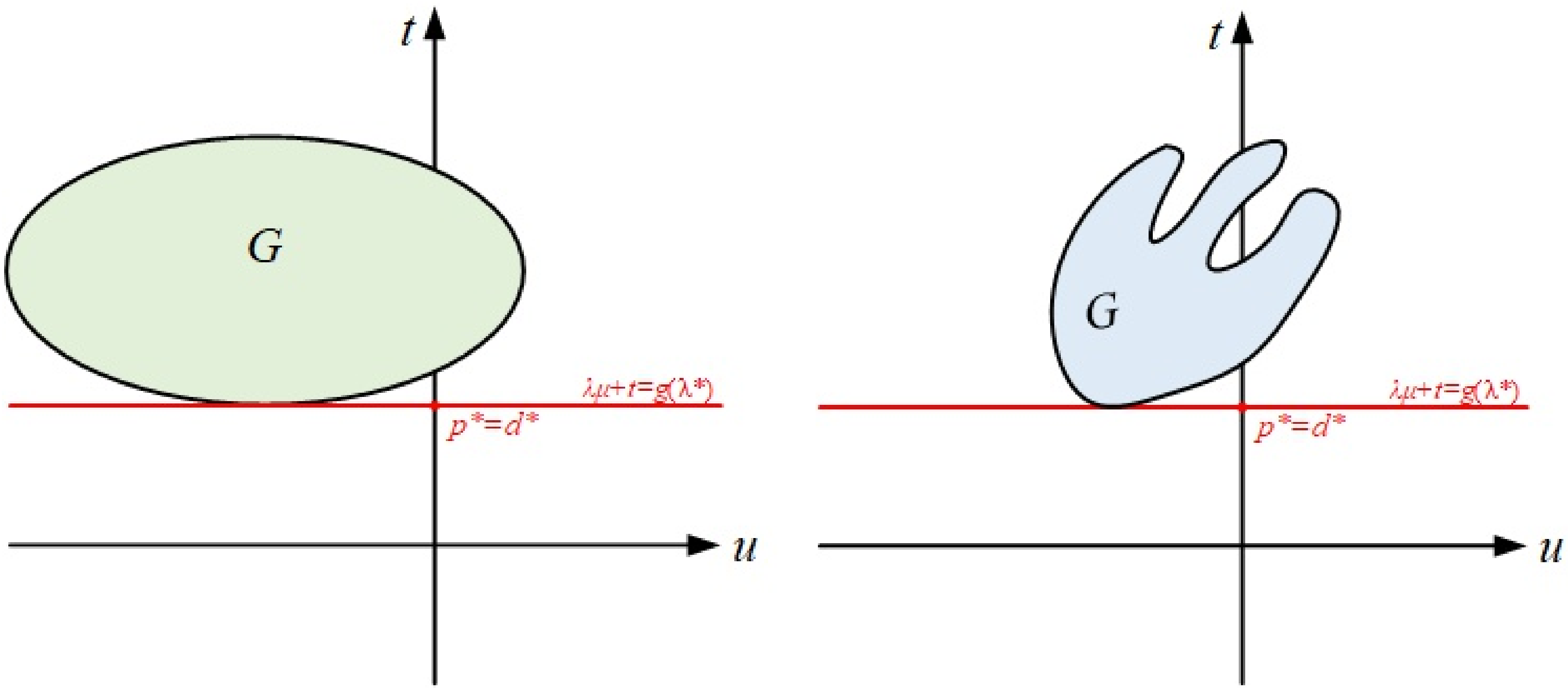


strong duality

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

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

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: $p^\star = d^\star$ if $A\tilde{x} \prec b$ for some $\tilde{x}$
- in fact, $p^\star = d^\star$ 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

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

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: $p^\star = d^\star$ if $A\tilde{x} \prec b$ for some $\tilde{x}$
- in fact, $p^\star = d^\star$ 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:

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

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

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

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

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

非凸问题满足强对偶性的例子

$$\min \left\{ x^2 - y^2 \,\middle|\, \text{s.t. } x^2 + y^2 \le 1 \right\}$$

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

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

4.3. Slater 条件与 KKT 条件

首先证明一个简化后的 Slater 条件，更复杂的证明类似：

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）

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

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

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 $u = v = 0$， $t \le p^*$。

同时 $(u,v,t) \in \Omega_1$，则存在 x 使得 $f_i(x) \le 0$， $\bar{A}x - \bar{b} = 0$， $f_0(x) \le t \le p^*$。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 $\left(\tilde{\lambda}, \tilde{\upsilon}, \mu\right) \ne 0$，和截距参数 α 使得

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

4.3. Slater 条件与 KKT 条件

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

从(4.6)可以得知 $\tilde{\lambda} \geq 0, \mu \geq 0$。否则 $\tilde{\lambda}^T u + \mu t$ 在 Ω_1 集合中可以趋向 $-\infty$，这和其有下界 α 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 $\left(f_1(x), \cdots, f_m(x), \bar{A}x - \bar{b}, f_0(x) \right) \in \Omega_1, \forall x \in D$，我们可以得到

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

4.3. Slater 条件与 KKT 条件

下面说明 $\mu \neq 0$，因此一定有 $\mu > 0$！

如果 $\mu = 0$ $\quad \Rightarrow \quad \sum_{1\le i\le m} \tilde{\lambda}_i f_i(x) + \tilde{\upsilon}^T\left(\bar{A}x - \bar{b}\right) \geq 0,\ \forall x \in D$

$$\hat{x} \in D,\ \bar{A}\hat{x} = \bar{b} \quad \Rightarrow \quad \sum_{1\le i\le m} \tilde{\lambda}_i f_i(\hat{x}) \geq 0$$

$$\tilde{\lambda}_i \geq 0,\ f_i(\hat{x}) < 0, 1 \leq i \leq m \quad \Rightarrow \quad \tilde{\lambda}_i = 0$$

$$\Rightarrow \quad \tilde{\upsilon}^T\left(\bar{A}x - \bar{b}\right) \geq 0, \forall x \in D$$

$$\hat{x} \in \operatorname{int} D, \bar{A}\hat{x} = \bar{b} \quad \Rightarrow \quad \tilde{\upsilon}^T \bar{A} = 0$$

$\bar{A}$ 行向量线性无关 $\quad \Rightarrow \quad \tilde{\upsilon} = 0 \quad \Rightarrow \quad \left(\tilde{\lambda}, \tilde{\upsilon}, \mu\right) = 0$ 矛盾！

4.3. Slater 条件与 KKT 条件

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

令 $\bar{\upsilon}$ 为对 $\tilde{\upsilon}$（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

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

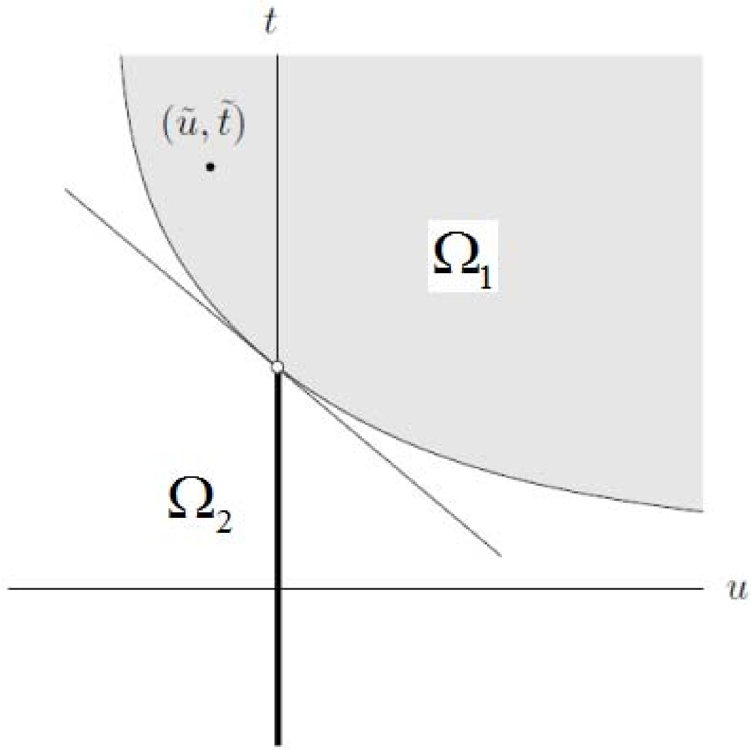
进一步，令 $\lambda = \frac{\tilde{\lambda}}{\mu}$， $\upsilon = \frac{\bar{\upsilon}}{\mu}$，我们可以达到

$$d^* \ge g\left(\lambda, \upsilon\right) = \inf_{x \in D} L\left(x, \lambda, \upsilon\right) \ge p^* \tag{4.9}$$

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 $\operatorname{aff} D$ 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

$$\operatorname{aff} D = \left\{ x \,\middle|\, x = \hat{x} + Qy,\ y \in R^q \right\}$$

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

$$\min_{y \in \overline{D}} \left\{ \overline{f_0}(y) \,\middle|\, \text{s.t. } \overline{f_i}(y) \le 0,\ 1 \le i \le m,\ \overline{A}y = 0 \right\}$$

其中 $\overline{A} = AQ$。

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 $\bar{d} = p^*$ 。

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

$$\Rightarrow \bar{g}(\lambda,\upsilon) = \inf_{\hat{x}+Qy\in D} f_0(\hat{x}+Qy) + \sum_{1\le i\le m} f_i(\hat{x}+Qy)\lambda_i + \upsilon^T\left(A\left(\hat{x}+Qy\right)-b\right)$$

$$= \inf_{x\in D} f_0(x) + \sum_{1\le i\le m} f_i(x)\lambda_i + \upsilon^T\left(Ax-b\right)$$

$$= g(\lambda,\upsilon)$$

$$\Rightarrow \quad d^* = \sup_{\lambda\ge 0} g(\lambda,\upsilon) = \sup_{\lambda\ge 0} \bar{g}(\lambda,\upsilon) = \bar{d} = p^*$$

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

$$\overline{a}_i^T d < 0, i = 1, \ldots, \overline{m}, \; Ad = 0$$

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

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

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

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(λ,μ)满足

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

由上式可推出$\sum_{i=1}^{\bar{m}}\lambda_i\bar{a}_i^T+\upsilon^T A=0,\quad \lambda_i\le 0,\forall i$。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$\sum_{i\in I_+}\lambda_i\bar{a}_i^T+\upsilon^T A=0,\quad \lambda_i<0,\forall i$。

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

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

由此可知，对任意可行解 x 均成立 $\overline{a}_i^T x = \overline{b}_i$， $\forall i \in I_+$。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

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

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

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 $\lambda^* \ge 0, \mu^*$ 和 x^* 一起满足

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

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：

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

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

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

$$\Rightarrow \quad \lambda_i^* f_i\left(x^*\right) = 0, \; i = 1, \ldots, m$$

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

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$\left(x^*,\lambda^*,\mu^*\right)$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：
$$f_i\left(x^*\right)\le 0,\ i=1,\dots,m$$
$$h_i\left(x^*\right)=0,\ i=1,\dots,p$$

对偶问题可行条件：$\lambda_i^*\ge 0,\ i=1,\dots,m$

互补松弛 Complementary slackness：$\lambda_i^* f_i\left(x^*\right)=0,\ i=1,\dots,m$

拉格朗日不动性：
$$\nabla f_0\left(x^*\right)+\sum_{i=1}^{m}\lambda_i^*\nabla f_i\left(x^*\right)+\sum_{i=1}^{p}\mu_i^*\nabla h_i\left(x^*\right)=0$$

4.3. Slater 条件与 KKT 条件

KKT 反向结论

如果 $\dfrac{\partial L\left(x^*,\lambda^*,\mu^*\right)}{\partial x}=0$ 可以保证 $L\left(x^*,\lambda^*,\mu^*\right)=\inf\limits_{x\in D} L\left(x,\lambda^*,\mu^*\right)$，那么 KKT 条件是 $\left(x^*,\lambda^*,\mu^*\right)$ 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 $g\left(\lambda^*,\mu^*\right)=L\left(x^*,\lambda^*,\mu^*\right)$，再结合其它条件可得 $g\left(\lambda^*,\mu^*\right)=f_0\left(x^*\right)$，由弱对偶性可得结论

推论：如果 $\nabla f_0\left(x\right)+\sum\limits_{i=1}^{m}\lambda_i^*\nabla f_i\left(x\right)+\sum\limits_{i=1}^{p}\mu_i^*\nabla h_i\left(x\right)=0$ 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 4th 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 λ, v 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（$f_i\left(x^*\right)=0$ 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 $\left(\lambda^*, \mu^*\right)$ 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 $\alpha_i > 0$)

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

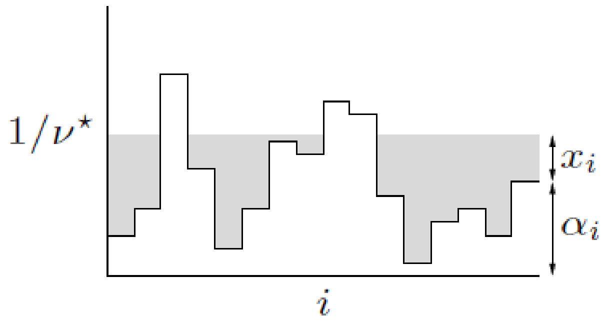
x is optimal iff $x \succeq 0$, $\mathbf{1}^T x = 1$, and there exist $\lambda \in \mathbf{R}^n$, $\nu \in \mathbf{R}$ such that

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

- 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 ν 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 α_i
- flood area with unit amount of water
- resulting level is $1/\nu^\star$



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^*

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

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

$$\Rightarrow \qquad p^*(u,v) \ge p^*(0,0) - \left(\lambda^*\right)^T u - \left(\mu^*\right)^T v$$

再加上 $p^*(u,v)$ 在 $(0,0)$ 处可导

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

4.4. 灵敏度与择一理论

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

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

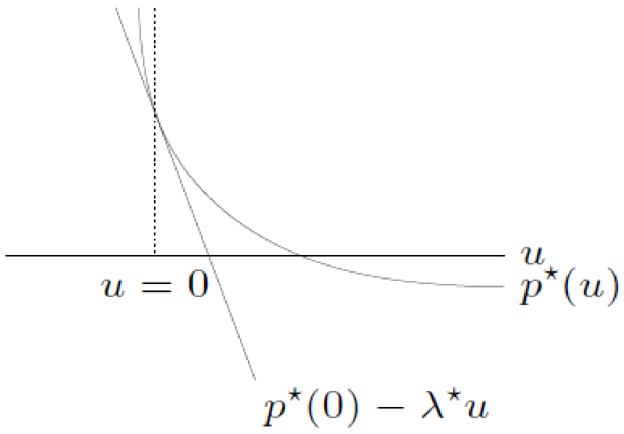
proof (for $\lambda_i^\star$): from global sensitivity result,

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

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

hence, equality

$p^\star(u)$ 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）

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

$f_i(x)<0,\ 1\le i\le m,\ h_i(x)=0,\ 1\le i\le p$ 和 $\lambda\ge 0, \lambda\ne 0, g(\lambda,\mu)\ge 0$ 弱择一

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

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

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）

$f_i(x) \le 0,\ i = 1, \ldots, m,\quad Ax = b$ 和 $\lambda \ge 0,\ \ g(\lambda, \mu) > 0$ 强择一

$f_i(x) < 0,\ i = 1, \ldots, m,\quad Ax = b$ 和 $\lambda \ge 0, \lambda \ne 0,\ \ g(\lambda, \mu) \ge 0$ 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 uninsteresting

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 ϕ convex, increasing

4.5. 对偶问题和问题变形

Introducing new variables and equality constraints

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

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

reformulated problem and its dual

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

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

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

(see page 5–4)

dual of norm approximation problem

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

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

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

定义其 Lagrange 函数（要求 $\lambda_i \succeq_{K_i^*} 0, \; i=1,\ldots,m$ ）

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

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 $g(\lambda_1,\dots,\lambda_m,\mu) \le p^*$

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

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

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

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: $p^* \ge d^*$ 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 $L(x, Z) = c^T x + \mathbf{tr}\left(Z(x_1 F_1 + \cdots + x_n F_n - G)\right)$
- dual function

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

dual SDP

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

$p^\star = d^\star$ 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.