第 5 章：对偶

（2）最优性条件

曹语

课程主页：https://yucaoyc.github.io/math3806

回顾

原始的优化问题为:

$$
\begin{aligned}
p^\star = \text{minimize} \quad & f_0(\boldsymbol{x}) \\
\text{subject to} \quad & f_i(\boldsymbol{x}) \leqslant 0, \quad i = 1, \cdots, m \\
& h_i(\boldsymbol{x}) = 0, \quad i = 1, \cdots, p
\end{aligned}
$$

其中优化变量为 $\boldsymbol{x} \in \mathbb{R}^n$

对偶问题为

$$
d^\star = \max_{\boldsymbol{\lambda} \succeq \mathbf{0}, \boldsymbol{\nu}} \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})
$$

其中

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

我们总是具有如下关系：$d^\star \leq p^\star$

目标：

- 强对偶成立的充分条件
- 强对偶和鞍点问题、博弈论的关联
- 强对偶成立的必要条件

练习：最大熵问题

$$
\begin{array}{ll}
\text{minimize} & \sum_{i=1}^{n} x_i \log x_i \\
\text{subject to} & \boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b} \\
& \mathbf{1}^\top \boldsymbol{x} = 1
\end{array}
$$

- 写出拉格朗日函数
- 计算拉格朗日对偶函数
- 写出拉格朗日对偶问题

答案

- 拉格朗日函数

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \nu) = \sum_{i=1}^{n} x_i \log x_i + \boldsymbol{\lambda}^\top (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) + \nu\left(\mathbf{1}^\top \boldsymbol{x} - 1\right)$$

- 拉格朗日对偶函数

$$\inf_{\boldsymbol{x}} \; L(\boldsymbol{x}, \boldsymbol{\lambda}, \nu) = -\boldsymbol{b}^\top \boldsymbol{\lambda} - \nu - e^{-\nu-1} \sum_{i=1}^{n} e^{-(\boldsymbol{A}^\top \boldsymbol{\lambda})_i}$$

- 拉格朗日对偶问题

$$\begin{array}{ll} \text{maximize} & -\boldsymbol{b}^\top \boldsymbol{\lambda} - \nu - e^{-\nu-1} \sum_{i=1}^{n} e^{-(\boldsymbol{A}^\top \boldsymbol{\lambda})_i} \\ \text{subject to} & \boldsymbol{\lambda} \succeq \mathbf{0} \end{array}$$

目录

(1) Slater 条件

定理 (Slater 定理)

若如下的条件成立，

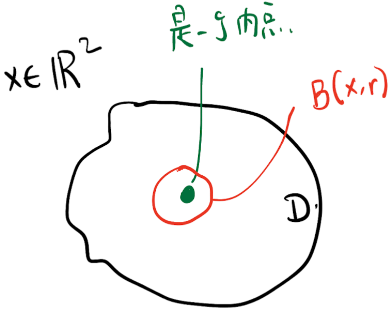
- 原问题是凸优化问题，且目标和约束函数的定义域为 $\mathcal{D}$;
- *(Slater 条件)* 存在相对内点 $\boldsymbol{x} \in \mathit{relint}\ \mathcal{D}$ 使得

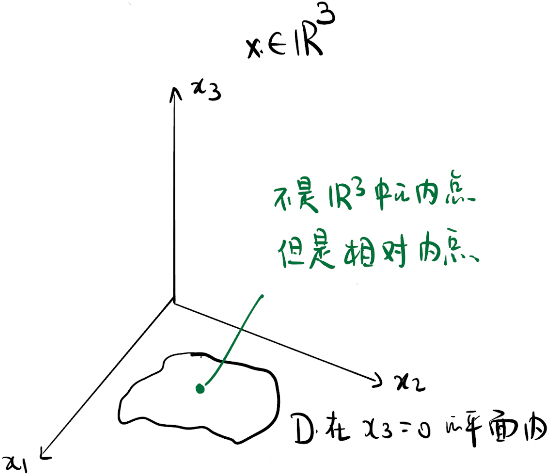
$$f_i(\boldsymbol{x}) < 0, \qquad i = 1, 2, \cdots, m, \qquad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

则强对偶性成立。

相对内部 $\text{relint}\ \mathcal{D} = \left\{ \boldsymbol{x} \in \mathcal{D} \mid B(\boldsymbol{x}, r) \cap \text{aff}\ \mathcal{D} \subset \mathcal{D} \right\}$ 的直观理解是：在仿射子空间内是内点。

相对内点的直观图：





【参见课本 2.1.3 节】

Slater 条件可改进为：如果 $f_1, f_2, \cdots, f_k$ 是仿射函数，且 $\exists\ \boldsymbol{x} \in \operatorname{relint} \mathcal{D}$ 使得

$$
\begin{aligned}
& f_i(\boldsymbol{x}) \leq 0, \qquad && i = 1, 2, \cdots, k \\
& f_i(\boldsymbol{x}) < 0, \qquad && i = k+1, \cdots, m, \\
& \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}
\end{aligned}
$$

(2) Slater 定理的应用实例

请应用 Slater 条件来判断如下例子是否有强对偶性：假设可行集非空

例子 1：

$$\begin{array}{ll} \text{minimize} & \boldsymbol{x}^\top \boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

例子 2：

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

例子 3： 最大熵问题，见第4页。

(3) Slater 定理的证明

考虑一个简化的版本：

$$
\begin{aligned}
p^\star = \text{minimize} \quad & f_0(\boldsymbol{x}) \\
\text{subject to} \quad & f_1(\boldsymbol{x}) \leqslant 0
\end{aligned}
$$

我们后续将证明，<u>当存在一点 $\boldsymbol{x}$ 满足 $f_1(\boldsymbol{x}) < 0$，则强对偶成立。</u>

证明： 考虑集合

$$\mathcal{G} = \left\{ \left(f_1(\boldsymbol{x}), f_0(\boldsymbol{x}) \right) : \ \boldsymbol{x} \in \mathcal{D} \right\}$$

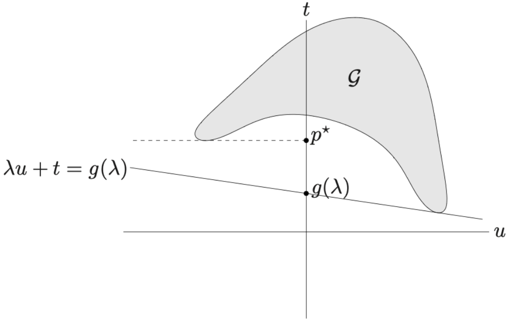
因此，优化问题为

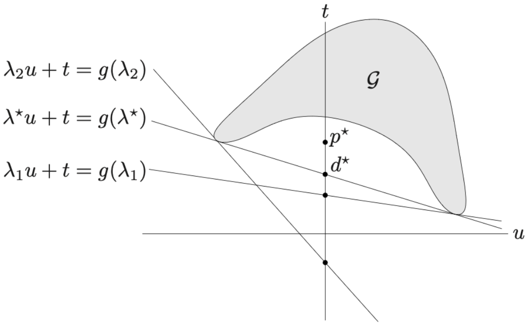
$$p^\star = \inf\{t : \ (u, t) \in \mathcal{G}, u \leq 0\}$$

Slater 定理的几何视角

拉格朗日对偶函数

$$g(\lambda) = \inf_{\boldsymbol{x}} \; f_0(\boldsymbol{x}) + \lambda f_1(\boldsymbol{x}) = \inf_{(u,t)\in\mathcal{G}} \; t + \lambda u \le \begin{bmatrix} \lambda & 1 \end{bmatrix}^\top \begin{bmatrix} u \\ t \end{bmatrix}$$





解释 $g(\lambda) \le p^\star$　　　　$d^\star \le p^\star$ 是弱对偶性

Slater 定理的证明

我们假设可行集非空，因此不妨假设 $p^\star$ 有限

【若 $p^\star = -\infty$，则必然有 $d^\star = -\infty$】

结论 1： 若原问题是凸优化，则如下集合 $\mathcal{A}$ 是凸集

$$\mathcal{A} := \mathcal{G} + \mathbb{R}_+ \times \mathbb{R}_+ = \{(u,t) : \ \exists\, \boldsymbol{x} \in \mathcal{D},\ f_1(\boldsymbol{x}) \le u, f_0(\boldsymbol{x}) \le t\}$$

证明： 可以直接通过凸集的定义来证明

结论 2： 定义凸集 $\mathcal{B} := \{(0,s) : s < p^\star\}$，凸集 $\mathcal{A}$ 和 $\mathcal{B}$ 不相交。

由超平面分离定理可知，存在 $(\widetilde{\lambda}, \widetilde{\mu}) \neq (0,0)$ 和 α，满足

$$(u,t) \in \mathcal{A} \implies \widetilde{\lambda} u + \widetilde{\mu} t \geq \alpha$$

$$(u,t) \in \mathcal{B} \implies \widetilde{\lambda} u + \widetilde{\mu} t \leq \alpha$$

$\Longrightarrow$

- 由于 $(u,t) \in \mathcal{A}$ 中 u 可以任意大，因此 $\tilde{\lambda} \geq 0$，类似可知 $\tilde{\mu} \geq 0$。
- 由于对于所有 $t < p^\star$，$\tilde{\mu} t \leq \alpha$，因此，$p^\star \tilde{\mu} \leq \alpha$。

因此，对于任意的 $\boldsymbol{x} \in \mathcal{D}$，

$$\tilde{\lambda} f_1(\boldsymbol{x}) + \tilde{\mu} f_0(\boldsymbol{x}) \geq \alpha \geq p^\star \tilde{\mu}$$

- 假设 $\tilde{\mu} > 0$，同时除 $\tilde{\mu}$ 可知，存在 $\lambda^\star = \frac{\tilde{\lambda}}{\tilde{\mu}} \geq 0$，满足对于任意的 $\boldsymbol{x} \in \mathcal{D}$

$$\lambda^\star f_1(\boldsymbol{x}) + f_0(\boldsymbol{x}) \geq p^\star$$

  因而，$g(\lambda^\star) \geq p^\star$；由弱对偶性可知 $g(\lambda^\star) \leq p^\star$，因此 $g(\lambda^\star) = p^\star$。因此我们得到强对偶性。

- 若 $\tilde{\mu} = 0$，则 $\tilde{\lambda} > 0$；上述结果变成了

$$\tilde{\lambda}\ f_1(\boldsymbol{x}) \geq 0, \qquad \forall \boldsymbol{x} \in \mathcal{D} \qquad \Longrightarrow \qquad f_1(\boldsymbol{x}) \geq 0, \qquad \forall \boldsymbol{x} \in \mathcal{D}.$$

  这和 Slater 条件矛盾。

目录

原优化问题可以改写成

$$\min_{\boldsymbol{x}} \max_{\boldsymbol{\lambda} \succeq \mathbf{0},\ \boldsymbol{\nu}} \quad f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i h_i(\boldsymbol{x})$$

而对偶问题为

$$\max_{\boldsymbol{\lambda} \succeq \mathbf{0},\ \boldsymbol{\nu}} \min_{\boldsymbol{x}} \quad f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i h_i(\boldsymbol{x})$$

因此原问题和对偶问题可以看成是不同优化顺序的选择。

更一般的，对于任意的函数 $f : W \times Z \to \mathbb{R}$，我们具有**极大极小不等式**

$$\sup_{z\in Z} \inf_{w\in W} f(w,z) \le \inf_{w\in W} \sup_{z\in Z} f(w,z) \tag{1}$$

如果等号成立，我们称 f 具有**强极大极小性质**，或者**鞍点性质**。

引理

假设公式(1)的左边的优化在 $z^\star$ 取到极大，右边的在 $w^\star$ 取到极大，因此我们有

$$\sup_{z\in Z} \inf_{w\in W} f(w,z) \le f(w^\star, z^\star) \le \inf_{w\in W} \sup_{z\in Z} f(w,z)$$

证明： 根据 $z^\star$ 的含义，$\sup_{z\in Z} \inf_{w\in W} f(w,z) = \inf_{w\in W} f(w,z^\star) \le f(w^\star,z^\star)$。类似我们可知 $f(w^\star,z^\star) \le \inf_{w\in W} \sup_{z\in Z} f(w,z)$。

- 若强极大极小性质成立，且令 $w^\star$ 和 $z^\star$ 分别是原问题和对偶问题的最优解，则

$$\inf_{w\in W} f(w,z^\star) = f(w^\star,z^\star) = \sup_{z\in Z} f(w^\star,z)$$

因此，

$$f(w^\star,z) \le f(w^\star,z^\star) \le f(w,z^\star) \tag{2}$$

该点 $(w^\star,z^\star)$ 的确是函数 f 的鞍点，故而等式

$$\sup_{z\in Z}\inf_{w\in W} f(w,z) = \inf_{w\in W}\sup_{z\in Z} f(w,z)$$

得名鞍点性质。

- 反之，若 $(w^\star,z^\star)$ 是鞍点，即满足公式(2)，则强极大极小性质成立。

例子： $f(w,z)=w^2-z^2$，点 $(0,0)$ 是该问题的鞍点。

案例：矩阵对策的混合策略

假设有两个玩家 1 和 2，1 具有选择 $k \in \{1, 2, \cdots, n\}$，2 具有决策 $l \in \{1, 2, \cdots, m\}$，玩家选择完后，玩家 1 支付玩家 2 金额 $P_{k,l}$。假设大家使用随机策略，即选择满足分布

$$\mathrm{Prob}(k = i) = w_i, \qquad \mathrm{Prob}(l = i) = z_i$$

期望的支付金额为 $f(\boldsymbol{z}, \boldsymbol{w}) = \sum_k \sum_l w_k z_l P_{k,l} = \boldsymbol{w}^\top \boldsymbol{P} \boldsymbol{z}$。我们将定义集合

$$W = \{\boldsymbol{w} : \ \boldsymbol{w} \succeq \mathbf{0}, \mathbf{1}^\top \boldsymbol{w} = 1\}$$
$$Z = \{\boldsymbol{z} : \ \boldsymbol{z} \succeq \mathbf{0}, \mathbf{1}^\top \boldsymbol{z} = 1\}$$

- **玩家 1 先选择，即玩家 1 的选择被玩家 2 可知：** 如果自己选择 $\boldsymbol{w}$，玩家 2 可以选择

  $$\max\left\{\boldsymbol{w}^\top \boldsymbol{P}\boldsymbol{z} \mid \boldsymbol{z} \in Z\right\}$$

  玩家 1 选择一个策略来最小化损失，因此

  $$p_1^\star = \min_{\boldsymbol{w}\in W} \max\left\{\boldsymbol{w}^\top \boldsymbol{P}\boldsymbol{z} \mid \boldsymbol{z} \in Z\right\} = \min_{\boldsymbol{w}\in W} \max_{\boldsymbol{z}\in Z} \boldsymbol{w}^\top \boldsymbol{P}\boldsymbol{z}$$

- 同理，**若玩家 2 先选择**，玩家 2 会选择一个策略来优化

  $$p_2^\star = \max_{\boldsymbol{z}\in Z} \min_{\boldsymbol{w}\in W} \boldsymbol{w}^\top \boldsymbol{P}\boldsymbol{z}$$

  问题： 直观能告诉我们 $p_1^\star$ 和 $p_2^\star$ 的关系吗?

- **玩家 1 先选择，即玩家 1 的选择被玩家 2 可知：** 如果自己选择 $\boldsymbol{w}$，玩家 2 可以选择

$$\max\left\{\boldsymbol{w}^\top \boldsymbol{P}\boldsymbol{z} \mid \boldsymbol{z} \in Z\right\}$$

玩家 1 选择一个策略来最小化损失，因此

$$p_1^\star = \min_{\boldsymbol{w}\in W} \max\left\{\boldsymbol{w}^\top \boldsymbol{P}\boldsymbol{z} \mid \boldsymbol{z} \in Z\right\} = \min_{\boldsymbol{w}\in W} \max_{\boldsymbol{z}\in Z} \boldsymbol{w}^\top \boldsymbol{P}\boldsymbol{z}$$

- 同理，**若玩家 2 先选择**，玩家 2 会选择一个策略来优化

$$p_2^\star = \max_{\boldsymbol{z}\in Z} \min_{\boldsymbol{w}\in W} \boldsymbol{w}^\top \boldsymbol{P}\boldsymbol{z}$$

问题：直观能告诉我们 $p_1^\star$ 和 $p_2^\star$ 的关系吗?

由于玩家 1 若先选择策略，对他一般总是不利，即 $p_1^\star \geq p_2^\star$。

- 对于该特定问题，我们可以通过 Slater 定理判断 $p_1^\star = p_2^\star$，即玩家 1 的优化问题的对偶问题等价于玩家 2 的。

  【具体步骤见课内或者课本 5.2.5】

- 强对偶意味着具有鞍点 $(\boldsymbol{w}^\star, \boldsymbol{z}^\star)$，在博弈论里它被称为**Nash equilibrium（纳什均衡）**；通过公式(2)（同时复制如下）

$$f(\boldsymbol{w}^\star, \boldsymbol{z}) \le f(\boldsymbol{w}^\star, \boldsymbol{z}^\star) \le f(\boldsymbol{w}, \boldsymbol{z}^\star).$$

  由此可见 Nash Equilibrium 意味着任何玩家修改策略只能对自己不利。

拉格朗日对偶可以看成是 $\boldsymbol{x}$ 和 $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ 之间的博弈。

目录

非凸优化的 KKT 条件

设定： 一般的优化形式；目标函数和约束函数可微

假设： 假设强对偶成立，且 $\boldsymbol{x}^\star$，$(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ 分别是原问题和对偶问题的最优解

由于对偶间隙为 0，

$$
\begin{aligned}
f_0(\boldsymbol{x}^\star) = g(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) &= \inf_{\boldsymbol{x}} \; f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i^\star f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i^\star h_i(\boldsymbol{x}) \\
&\le f_0(\boldsymbol{x}^\star) + \sum_{i=1}^{m} \lambda_i^\star f_i(\boldsymbol{x}^\star) + \sum_{i=1}^{p} \nu_i^\star h_i(\boldsymbol{x}^\star) \le f_0(\boldsymbol{x}^\star)
\end{aligned}
$$

由此可见，所有的不等号其实应该是等号，因此我们得到**互补松弛性条件**

$$
\lambda_i^\star f_0(\boldsymbol{x}^\star) = 0, \qquad i = 1, 2, \cdots, m
$$

类似公式(2)，可验证

$$L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = \inf_{\boldsymbol{x}} \; L(\boldsymbol{x}, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$$

由于 $\boldsymbol{x}$ 在拉格朗日函数中是无约束的，因此最优性条件可知，

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = \mathbf{0}$$

化简得到，

$$\nabla f_0(\boldsymbol{x}^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(\boldsymbol{x}^\star) + \sum_{i=1}^{p} \nu_i^\star \nabla h_i(\boldsymbol{x}^\star) = \mathbf{0}$$

KKT 条件

定理 (KKT 条件)

对于一般的优化问题，若目标函数和约束函数可微，则强对偶成立的必要条件为：

$$\begin{aligned} f_i(\boldsymbol{x}^\star) &\le 0, & i &= 1,2,\cdots,m \\ h_i(\boldsymbol{x}^\star) &= 0, & i &= 1,2,\cdots,p \\ \lambda_i^\star &\ge 0, & i &= 1,2,\cdots,m \\ \lambda_i^\star f_i(\boldsymbol{x}^\star) &= 0, & i &= 1,2,\cdots,m \end{aligned}$$

$$\nabla f_0(\boldsymbol{x}^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(\boldsymbol{x}^\star) + \sum_{i=1}^{p} \nu_i^\star \nabla h_i(\boldsymbol{x}^\star) = \mathbf{0}$$

该（不等式）方程组被称为**Karush-Kuhn-Tucker（KKT）条件**

凸优化和 KKT 条件

定理

对于凸优化问题，如果 $\boldsymbol{x}^\star$，$\boldsymbol{\lambda}^\star$，$\boldsymbol{\nu}^\star$ 是 *KKT* 条件的解，则 $\boldsymbol{x}^\star$，$(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ 分别是原问题和对偶问题的最优解，且对偶间隙为 0。

证明：对于凸优化，由于 h_i 是仿射，继而由 $\lambda_i^\star \geq 0$ 可得到 $L(\boldsymbol{x}, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ 是关于 $\boldsymbol{x}$ 的凸函数。主要由 KKT 条件的第 5 行得到，$g(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$，因此

$$\sup_{\boldsymbol{\lambda} \succeq \boldsymbol{0}, \boldsymbol{\nu}} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \geq L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star).$$

并且由于 $L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = f_0(\boldsymbol{x}^\star)$ 可知

$$L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) \geq \inf_{\boldsymbol{x}} f_0(\boldsymbol{x}) = \inf_{\boldsymbol{x}} \sup_{\boldsymbol{\lambda} \succeq \boldsymbol{0}, \boldsymbol{\nu}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$$

根据极大极小不等式可知，我们可知上述的不等号其实为等号，

$$\sup_{\boldsymbol{\lambda} \succeq \boldsymbol{0}, \boldsymbol{\nu}} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = \inf_{\boldsymbol{x}} \sup_{\boldsymbol{\lambda} \succeq \boldsymbol{0}, \boldsymbol{\nu}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$$

故而对偶间隙为 0。$\boldsymbol{x}^\star$ 是最优解等可以类似说明。

定理 (推论)

对于某个凸优化问题，目标函数和约束函数可微，且 *Slater* 条件满足，则 *KKT* 是最优性的充要条件。

证明： Slater 条件意味着对偶间隙为 0，因此 KKT 条件成立；反之，若 KKT 条件成立，则对偶间隙为 0，且 KKT 条件的解是最优解。

例子 1:

$$\begin{array}{ll} \text{minimize} & \boldsymbol{x}^\top \boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

对此，KKT 条件为

$$\boldsymbol{A}\boldsymbol{x}^\star = \boldsymbol{b}, \qquad 2\boldsymbol{x}^\star + \boldsymbol{A}^\top \boldsymbol{\nu}^\star = 0$$

可以写成等式

$$\begin{bmatrix} \boldsymbol{A} & 0 \\ 2\boldsymbol{I} & \boldsymbol{A}^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{x}^\star \\ \boldsymbol{\nu}^\star \end{bmatrix} = \begin{bmatrix} \boldsymbol{b} \\ 0 \end{bmatrix}$$

对此，优化问题变成了求解线性方程组。

目录

总结

- 知道 Slater 条件的含义，并且能应用；Slater 定理的证明不要求
- 知道极大极小不等式，鞍点性质的含义，对于对偶问题和博弈论关系有所了解
- 能够复述并使用 KKT 条件；知道第26，27, 28页的结论

阅读作业 & 参考资料:

- 课本第 5.2 - 5.5 章