优化模型与软件工具

主讲教师：董庆兴

华中师范大学　信息管理学院
qxdong@mail.ccnu.edu.cn

2017 年 12 月 26 日

大纲

1. Lagrange 对偶问题

2. 鞍点解释

3. 最优性条件

约束最优化问题

- 考虑标准形式的约束优化问题：

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, i = 1, \cdots, m. \\ & h_j(\mathbf{x}) = 0, j = 1, \cdots, p. \end{aligned} \tag{1}$$

  其中 $\mathbf{x} \in \mathbb{R}^n$。
- 设问题的定义域 $\mathcal{D} = \bigcap_{i=0}^{m} \textbf{dom}\ f_i \cap \bigcap_{j=1}^{p} \textbf{dom}\ h_j$ 是非空集合
- 优化问题的最优值为 p^*
- 我们目前并没有假设优化问题 (1) 是凸优化问题

Lagrange 对偶问题

- 对于任意一组 $(\boldsymbol{\lambda}, \boldsymbol{\mu})$，其中 $\boldsymbol{\lambda} \succeq \mathbf{0}$，Lagrange 对偶函数给出了原问题 (1) 的最优值 p^* 的一个下界，一个很自然的问题是：从 Lagrange 函数能够得到的最好下界是什么？可将上述问题表述为优化问题：

$$\begin{array}{ll} \text{maximize} & g(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{subject to} & \boldsymbol{\lambda} \succeq \mathbf{0} \end{array} \tag{2}$$

- 上式称为原问题 (1) 的Lagrange 对偶问题. 满足 $\boldsymbol{\lambda} \succeq \mathbf{0}$ 和 $g(\boldsymbol{\lambda}, \boldsymbol{\mu}) > -\infty$ 的一组 $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ 是对偶问题 (2) 的一个可行解，称作对偶可行
- 称解 $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ 是对偶最优解或者是最优 Lagrange 乘子
- Lagrange 对偶问题 (2) 是一个凸优化问题，因为极大化目标函数是凹函数，且约束集合为凸集。其凸性与原问题 (1) 是否是凸优化问题无关

标准线性规划的 Lagrange 对偶

- 前面提到的例子说到了对偶函数的定义域 $\textbf{dom}\ \ g = \{(\boldsymbol{\lambda}, \boldsymbol{\mu}) | g(\boldsymbol{\lambda}, \boldsymbol{\mu}) > -\infty\}$ 的维数一般都小于 $m + p$，事实上很多情况下我们可以求出 $\textbf{dom}\ \ g$ 并将其表示为一系列等式约束
- 对于标准形式的线性规划问题

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

- 我们之前已经求得了其 Lagrange 对偶函数，因此我们有

$$
\begin{array}{ll}
\text{maximize} & g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \begin{cases} -\mathbf{b}^{\mathrm{T}}\boldsymbol{\mu}, & \mathbf{c} + \mathbf{A}^{\mathrm{T}}\boldsymbol{\mu} - \boldsymbol{\lambda} = 0 \\ -\infty, & \text{otherwise} \end{cases} \\
\text{subject to} & \boldsymbol{\lambda} \succeq \mathbf{0}
\end{array} \tag{3}
$$

显式表达对偶约束：标准线性规划

- 我们可知当且仅当 $\mathbf{c}+\mathbf{A}^{\mathrm{T}}\boldsymbol{\mu}-\boldsymbol{\lambda}=0$ 时(3)的目标函数有界，因此我们可将此隐含的约束显式化得到

$$\begin{aligned} \text{maximize} \quad & -\mathbf{b}^{\mathrm{T}}\boldsymbol{\mu} \\ \text{subject to} \quad & \mathbf{c}+\mathbf{A}^{\mathrm{T}}\boldsymbol{\mu}-\boldsymbol{\lambda}=0 \\ & \boldsymbol{\lambda}\succeq\mathbf{0} \end{aligned} \tag{4}$$

- 进一步地，上式可表示为

$$\begin{aligned} \text{maximize} \quad & -\mathbf{b}^{\mathrm{T}}\boldsymbol{\mu} \\ \text{subject to} \quad & \mathbf{c}+\mathbf{A}^{\mathrm{T}}\boldsymbol{\mu}\succeq\mathbf{0} \end{aligned} \tag{5}$$

- 从而得到了一个不等式线性规划。也就是标准式线性规划的 Lagrange 对偶问题是(3)，同时问题(3)等价于(4)和(5)

不等式线性规划的 Lagrange 对偶

- 对于不等式形式的线性规划问题

$$\begin{array}{ll} \text{minimize} & \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ \text{subject to} & \mathbf{A}\mathbf{x} \preceq \mathbf{b} \end{array} \tag{6}$$

- 其 Lagrange 函数为

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{c}^{\mathrm{T}}\mathbf{x} + \boldsymbol{\lambda}^{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{b}) = -\mathbf{b}^{\mathrm{T}}\boldsymbol{\lambda} + \left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} + \mathbf{c}\right)^{\mathrm{T}}\mathbf{x}$$

- 所以对偶函数为

$$g(\boldsymbol{\lambda}) = -\mathbf{b}^{\mathrm{T}}\boldsymbol{\lambda} + \inf_{\mathbf{x}} \left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} + \mathbf{c}\right)^{\mathrm{T}}\mathbf{x}$$

- 如果线性函数不是恒定值，则线性函数的下确界是 $-\infty$，有

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \begin{cases} -\mathbf{b}^{\mathrm{T}}\boldsymbol{\lambda}, & \mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} + \mathbf{c} = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

不等式线性规划的 Lagrange 对偶规划

- 称对偶变量 $\boldsymbol{\lambda}$ 是对偶可行的，如果有 $\boldsymbol{\lambda} \succeq \mathbf{0}$ 且 $\mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} + \mathbf{c} = 0$
- 通过显示表达对偶可行条件，可得

$$\begin{array}{ll} \text{maximize} & -\mathbf{b}^{\mathrm{T}}\boldsymbol{\lambda} \\ \text{subject to} & \mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} + \mathbf{c} = 0 \\ & \boldsymbol{\lambda} \succeq \mathbf{0} \end{array} \tag{7}$$

- 我们注意到标准形式的线性规划和不等式形式的线性规划以及它们之间的对偶问题的对称性：标准形式线性规划的对偶问题是不等式形式的，反之亦然

弱对偶性

- Lagrange 对偶问题的最优值用 d^* 来表示，根据定义，这是通过 Lagrange 函数得到的原问题最优值 p^* 的最好下界。也就是

$$d^* \leq p^* \tag{8}$$

- 即使当 d^* 和值 p^* 无限时，式(8)也成立：
  - 如果原问题无下界，即 $p^* = -\infty$，则必有 $d^* = -\infty$，也就是 Lagrange 对偶问题不可行
  - 如果对偶问题无上界，即 $d^* = \infty$，则必有 $p^* = \infty$，也就是原问题不可行
- 定义差值 $p^* - d^*$ 为原问题的最优对偶间隙。它给出了原问题最优值和通过 Lagrange 对偶函数所能得到的最好下界之间的差值。最优对偶间隙总是非负

强对偶性

- 如果
$$d^* = p^* \tag{9}$$
成立，即最优对偶间隙为零，那么强对偶性成立。这说明从 Lagrange 对偶函数得到的最好下界是紧的
- 对于一般情况，强对偶性是不成立的，但是如果原问题(1)是凸问题，即可以表述为如下形式

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \le 0, i = 1, \cdots, m. \\ & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned} \tag{10}$$

其中 $f_0, f_1, \cdots, f_m$ 为凸函数，强对偶性通常（但并不总是）成立

Slater 条件

- 有很多研究给出了强对偶性成立的条件（除了凸性条件之外），这些条件被称作约束准则
- 一个简单的约束准则叫做Slater 条件：存在一点 $\mathbf{x} \in \mathbf{relint}\ \mathcal{D}$ 使得

$$f_i(\mathbf{x}) < 0, i = 1, \cdots, m, \quad \mathbf{A}\mathbf{x} = \mathbf{b} \tag{11}$$

  成立。满足上述条件的点有时称作严格可行，因为不等式约束是严格成立的
- 当原问题为凸问题且 Slater 条件成立时，强对偶性成立

弱化 Slater 条件

- 当不等式约束 f_i 中有些是仿射函数时，Slater 条件可以进一步改进，不失一般性，设前 k 个约束是仿射的，则：存在一点 $\mathbf{x}$ 使得

$$f_i(\mathbf{x}) \leq 0, i = 1, \cdots, k, \quad f_j(\mathbf{x}) < 0, j = k+1, \cdots, m, \quad \mathbf{Ax} = \mathbf{b} \tag{12}$$

成立，则结合凸性条件即可判断强对偶性成立
- 换言之，仿射不等式不需要严格成立。尤其当所有约束条件都是线性等式或不等式且 $\mathbf{dom}\ f_0$ 是开集的时候，改进的 Slater 条件(12)就是可行性条件
- 若 Slater 条件以及其弱化形式得到满足，不但对于凸问题强对偶性成立，也意味着当 $d^* > -\infty$ 时对偶问题能够取得最优值，也就是存在一组对偶可行解 $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ 使得 $g(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = d^* = p^*$

线性方程组的最小二乘问题对偶

- 再次考虑

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

- 其相应的对偶问题应为

$$\text{maximize} \quad -\frac{1}{4}\boldsymbol{\mu}^{\mathrm{T}}\mathbf{A}\mathbf{A}^{\mathrm{T}}\boldsymbol{\mu} - \mathbf{b}^{\mathrm{T}}\boldsymbol{\mu}$$

  这是一个凹二次函数的无约束极大化问题

- Slater 条件事实上是原问题的可行条件，所以如果有 $\mathbf{b} \in \mathcal{R}(\mathbf{A})$，即 $p^* < \infty$，则 $p^* = d^*$

线性规划的 Lagrange 对偶

- 考虑问题：

$$\begin{aligned} \text{minimize} \quad & \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \succeq \mathbf{0} \end{aligned}$$

- 其相应的对偶问题应为

$$\begin{aligned} \text{maximize} \quad & -\mathbf{b}^{\mathrm{T}}\boldsymbol{\mu} \\ \text{subject to} \quad & \mathbf{c} + \mathbf{A}^{\mathrm{T}}\boldsymbol{\mu} \succeq \mathbf{0} \end{aligned}$$

- 根据 Slater 条件的弱化形式，对于任意的线性规划问题，无论是标准形式还是不等式形式，只要原问题可行，强对偶性均成立

二次约束二次规划的 Lagrange 对偶

- 考虑问题:

$$
\begin{array}{ll}
\text{minimize} & \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{P}_0\mathbf{x}+\mathbf{q}_0^{\mathrm{T}}\mathbf{x}+r_0 \\
\text{subject to} & \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{P}_i\mathbf{x}+\mathbf{q}_i^{\mathrm{T}}\mathbf{x}+r_i \le 0, i=1,\cdots,m
\end{array} \tag{13}
$$

- 对应的 Lagrange 函数为

$$
\begin{aligned}
L(\mathbf{x},\boldsymbol{\lambda}) &= \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{P}_0\mathbf{x}+\mathbf{q}_0^{\mathrm{T}}\mathbf{x}+r_0+\sum_{i=1}^{m}\lambda_i\left(\frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{P}_i\mathbf{x}+\mathbf{q}_i^{\mathrm{T}}\mathbf{x}+r_i\right) \\
&= \frac{1}{2}\mathbf{x}^{\mathrm{T}}\left(\mathbf{P}_0+\sum_{i=1}^{m}\lambda_i\mathbf{P}_i\right)\mathbf{x}+\left(\mathbf{q}_0+\sum_{i=1}^{m}\lambda_i\mathbf{q}_i\right)^{\mathrm{T}}\mathbf{x}+\left(r_0+\sum_{i=1}^{m}\lambda_i r_i\right) \\
&= \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{P}(\boldsymbol{\lambda})\mathbf{x}+\mathbf{q}(\boldsymbol{\lambda})^{\mathrm{T}}\mathbf{x}+r(\boldsymbol{\lambda})
\end{aligned}
$$

其中

$\mathbf{P}(\boldsymbol{\lambda})=\mathbf{P}_0+\sum_{i=1}^{m}\lambda_i\mathbf{P}_i,\ \mathbf{q}(\boldsymbol{\lambda})=\mathbf{q}_0+\sum_{i=1}^{m}\lambda_i\mathbf{q}_i,\ r(\boldsymbol{\lambda})=r_0+\sum_{i=1}^{m}\lambda_i r_i$

二次约束二次规划的 Lagrange 对偶

- 由于 $\lambda_i \geq 0$，可知

$$g(\boldsymbol{\lambda}) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = \inf_{\mathbf{x}} \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{P}(\boldsymbol{\lambda})\mathbf{x} + \mathbf{q}(\boldsymbol{\lambda})^{\mathrm{T}}\mathbf{x} + r(\boldsymbol{\lambda})$$

- 对上式求导即可得 $\mathbf{x}^* = -\mathbf{P}(\boldsymbol{\lambda})^{-1}\mathbf{q}(\boldsymbol{\lambda})$，可知

$$g(\boldsymbol{\lambda}) = -\frac{1}{2}\mathbf{q}(\boldsymbol{\lambda})^{\mathrm{T}}\mathbf{P}(\boldsymbol{\lambda})^{-1}\mathbf{q}(\boldsymbol{\lambda}) + r(\boldsymbol{\lambda})$$

- 因此，QCQP 的对偶规划为

$$\begin{aligned} \text{maximize} \quad & -\frac{1}{2}\mathbf{q}(\boldsymbol{\lambda})^{\mathrm{T}}\mathbf{P}(\boldsymbol{\lambda})^{-1}\mathbf{q}(\boldsymbol{\lambda}) + r(\boldsymbol{\lambda}) \\ \text{subject to} \quad & \boldsymbol{\lambda} \geq 0 \end{aligned} \tag{14}$$

- 由 Slater 条件，当二次不等式约束严格成立时，也就是存在 $\mathbf{x}$ 使得 $\frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{P}_i\mathbf{x} + \mathbf{q}_i^{\mathrm{T}}\mathbf{x} + r_i < 0, i = 1, \cdots, m$ 成立，则强对偶性成立

最大熵问题的对偶函数

$$\begin{aligned} \text{minimize} \quad & f_0(x) = \sum_{i=1}^{n} x_i \log x_i \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \succeq \mathbf{b} \\ & \mathbf{1}^{\mathrm{T}}\mathbf{x} = 1 \end{aligned}$$

其对偶函数为

$$\begin{aligned} \text{maximize} \quad & -\mathbf{b}^{\mathrm{T}}\boldsymbol{\lambda} - \mu - e^{-\mu-1}\sum_{i=1}^{n} e^{-\mathbf{a}_i^{\mathrm{T}}\boldsymbol{\lambda}} \\ \text{subject to} \quad & \boldsymbol{\lambda} \geq 0 \end{aligned}$$

针对 μ 解析求最大，可以得到对 μ 的导数为零时，有

$$\mu = \log \sum_{i=1}^{n} e^{-\mathbf{a}_i^{\mathrm{T}}\boldsymbol{\lambda}} - 1$$

最大熵问题的对偶函数

从而对偶问题可简化为

$$\begin{array}{ll} \text{maximize} & -\mathbf{b}^{\mathrm{T}}\boldsymbol{\lambda} - \log\left(\sum_{i=1}^{n} e^{-\mathbf{a}_i^{\mathrm{T}}\boldsymbol{\lambda}}\right) \\ \text{subject to} & \boldsymbol{\lambda} \geq 0 \end{array}$$

从而我们得到一个非负约束的几何规划问题（凸优化）

强弱对偶性的极大极小描述

我们可以将原、对偶优化问题以一种更为对称的方式进行表达。为简化讨论，假设没有等式约束，我们有

$$
\begin{array}{ll}
\text{minimize} & f_0(\mathbf{x}) \\
\text{subject to} & f_i(\mathbf{x}) \leq 0, i = 1, \cdots, m.
\end{array}
$$

首先我们注意到

$$
\begin{aligned}
\sup_{\boldsymbol{\lambda} \succeq 0} L(\mathbf{x}, \boldsymbol{\lambda}) &= \sup_{\lambda \succeq 0} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) \right) \\
&= \begin{cases} f_0(\mathbf{x}), & f_i(\mathbf{x}) \leq 0, i = 1, \cdots, m \\ \infty, & \text{otherwise} \end{cases}
\end{aligned}
$$

此时有

$$
p^* = \inf_{\mathbf{x}} \sup_{\boldsymbol{\lambda} \succeq 0} L(\mathbf{x}, \boldsymbol{\lambda})
$$

强弱对偶性的极大极小描述

而根据对偶函数的定义我们有

$$d^* = \sup_{\boldsymbol{\lambda} \succeq 0} g(\boldsymbol{\lambda}) = \sup_{\boldsymbol{\lambda} \succeq 0} \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda})$$

因此弱对偶性可以表述为

$$\sup_{\boldsymbol{\lambda} \succeq 0} \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) \leq \inf_{\mathbf{x}} \sup_{\boldsymbol{\lambda} \succeq 0} L(\mathbf{x}, \boldsymbol{\lambda}) \tag{15}$$

而强对偶性可以表示为

$$\sup_{\boldsymbol{\lambda} \succeq 0} \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = \inf_{\mathbf{x}} \sup_{\boldsymbol{\lambda} \succeq 0} L(\mathbf{x}, \boldsymbol{\lambda})$$

强对偶性意味着对 $\mathbf{x}$ 求极小和对 $\boldsymbol{\lambda} \succeq 0$ 求极大可以互换而不影响结果

极小极大不等式

事实上，不等式(15)是否成立与 L 的性质无关：对于任意 $f: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $W \in \mathbb{R}^n$, $Z \in \mathbb{R}^m$, 有

$$\sup_{\mathbf{z}\in Z} \inf_{\mathbf{w}\in W} f(\mathbf{w},\mathbf{z}) \leq \inf_{\mathbf{w}\in W} \sup_{\mathbf{z}\in Z} f(\mathbf{w},\mathbf{z}) \tag{16}$$

这一极大极小不等式成立。若等式成立，即

$$\sup_{\mathbf{z}\in Z} \inf_{\mathbf{w}\in W} f(\mathbf{w},\mathbf{z}) = \inf_{\mathbf{w}\in W} \sup_{\mathbf{z}\in Z} f(\mathbf{w},\mathbf{z}) \tag{17}$$

我们称 f（以及 W 和 Z）满足强极大极小性质或者鞍点性质

鞍点解释

鞍点

设有 $f:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$, $W\in\mathbb{R}^n, Z\in\mathbb{R}^m$, 有 $\tilde{\mathbf{w}}\in W,\tilde{\mathbf{z}}\in Z$ 对任意 $\mathbf{w}\in W,\mathbf{z}\in Z$ 有

$$f(\tilde{\mathbf{w}},z)\le f(\tilde{\mathbf{w}},\tilde{\mathbf{z}})\le f(\mathbf{w},\tilde{\mathbf{z}})$$

成立，则称 $(\tilde{\mathbf{w}},\tilde{\mathbf{z}})$ 为 f 的鞍点

也就是说，$f(\mathbf{w},\tilde{\mathbf{z}})$ 在 $\tilde{\mathbf{w}}$ 处取得最小值（关于变量 $\mathbf{w}\in W$），$f(\tilde{\mathbf{w}},\mathbf{z})$ 在 $\tilde{\mathbf{z}}$ 处取得最大值（关于变量 $\mathbf{z}\in Z$）：

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

最优与次优

最优解与 ϵ 次优

对于 $X_{opt} = \{\mathbf{x} | f_i(\mathbf{x}) \le 0, i = 1, \cdots, m, h_j(\mathbf{x}) = 0, j = 1, \cdots, p, f_0(\mathbf{x}) = p^*\}$. 如果问题(1)存在最优解，我们称这种最优值是可达的或者可得的，问题可解。如果 $X_{opt} = \emptyset$，则称最优值是不可得或不可达的。满足 $f_0(\bar{\mathbf{x}}) \le p^* + \epsilon, \epsilon > 0$ 的可行解 $\bar{\mathbf{x}}$ 称为$\epsilon-$ 次优

- 如果能够找到一个对偶可行解 $(\boldsymbol{\lambda}, \boldsymbol{\mu})$，就对原问题的最优值建立了一个下界：$p^* \ge g((\boldsymbol{\lambda}, \boldsymbol{\mu}))$。因此对偶可行点 $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ 为表达式 $p^* \ge g(\boldsymbol{\lambda}, \boldsymbol{\mu})$ 的成立提供了一个证明。强对偶性意味着存在着任意好的认证
- 如果 $\bar{\mathbf{x}}$ 是原问题的可行解且 $(\boldsymbol{\lambda}', \boldsymbol{\mu}')$ 对偶可行，那么

$$f_0(\bar{\mathbf{x}}) - p^* \le f_0(\bar{\mathbf{x}}) - g(\boldsymbol{\lambda}', \boldsymbol{\mu}')$$

- 说明 $\bar{\mathbf{x}}$ 是 $\epsilon-$ 次优，其中 $\epsilon = f_0(\bar{\mathbf{x}}) - g(\boldsymbol{\lambda}', \boldsymbol{\mu}')$。同样也说明对对偶问题 $(\boldsymbol{\lambda}', \boldsymbol{\mu}')$ 是 $\epsilon-$ 次优

次优解认证

- 定义原问题和对偶问题目标函数的差值 $f_0(\bar{\mathbf{x}}) - g(\boldsymbol{\lambda}', \boldsymbol{\mu}')$ 为原问题可行解 $\bar{\mathbf{x}}$ 与对偶问题可行解 $(\boldsymbol{\lambda}', \boldsymbol{\mu}')$ 之间的对偶间隙。一对原对偶问题的可行点将原问题（对偶问题）的最优值限制在一个区间上

$$p^* \in [g(\boldsymbol{\lambda}', \boldsymbol{\mu}'), f_0(\bar{\mathbf{x}})], \quad d^* \in [g(\boldsymbol{\lambda}', \boldsymbol{\mu}'), f_0(\bar{\mathbf{x}})]$$

  区间长度即为对偶间隙
- 如果原问题可行解 $\bar{\mathbf{x}}$ 与对偶问题可行解 $(\boldsymbol{\lambda}', \boldsymbol{\mu}')$ 之间的对偶间隙为零，即 $f_0(\bar{\mathbf{x}}) = g(\boldsymbol{\lambda}', \boldsymbol{\mu}')$，那么 $\bar{\mathbf{x}}$ 和 $(\boldsymbol{\lambda}', \boldsymbol{\mu}')$ 是一对对偶问题的最优解
- 对偶间隙为零时，$(\boldsymbol{\lambda}', \boldsymbol{\mu}')$ 是证明 $\bar{\mathbf{x}}$ 为最优解的一个认证（$\bar{\mathbf{x}}$ 是证明 $(\boldsymbol{\lambda}', \boldsymbol{\mu}')$ 为对偶最优的一个认证）

互补松弛

设原问题和对偶问题的最优值都可以达到并且相等（强对偶性成立），令 $\mathbf{x}^*, (\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ 分别为原问题和对偶问题的最优解，这表明

$$\begin{aligned} f_0(\mathbf{x}^*) &= g(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \\ &= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i^* f_i(\mathbf{x}) + \sum_{j=1}^p \mu_j^* h_j(\mathbf{x}) \right) \\ &\le f_0(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* f_i(\mathbf{x}^*) + \sum_{j=1}^p \mu_j^* h_j(\mathbf{x}^*) \\ &\le f_0(\mathbf{x}^*) \end{aligned}$$

第一行的等号说明最优对偶间隙为零,第二行的等号是对偶函数的定义
第三行不等式是根据 Lagrange 函数关于 $\mathbf{x}$ 求下确界必然小于等于其在 $\mathbf{x} = \mathbf{x}^*$ 处的值得来。最后一个不等式成立是因为
$\lambda_i^* \ge 0, f_i(\mathbf{x}^*) \le 0, i = 1, \cdots, m$

互补松弛

可知上式中的不等号取等号，即

$$
\begin{aligned}
f_0(\mathbf{x}^*) &= f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j^* h_j(\mathbf{x}^*) \\
&= f_0(\mathbf{x}^*)
\end{aligned}
$$

从而有

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

事实上，求和的每一项都是非正的，因此有

$$
\lambda_i^* f_i(\mathbf{x}^*) = 0, \quad i = 1, \cdots, m. \tag{18}
$$

上述条件称作互补松弛性

互补松弛性推广

- 互补松弛性对任意原问题最优解 $\mathbf{x}^*$ 和对偶问题最优解 $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ 都成立（当强对偶性成立时）
- 我们可以将互补松弛性条件写成

$$\lambda_i^* > 0 \Longrightarrow f_i(\mathbf{x}^*) = 0 \tag{19}$$

$$f_i(\mathbf{x}^*) < 0 \Longrightarrow \lambda_i^* = 0 \tag{20}$$

- 上式意味着在最优点处，除了第 i 个约束起作用的情况（也就是取等号），最优 Lagrange 乘子的第 i 项都为零

KKT 条件

现假设函数 $f_i(\mathbf{x}), h_j(\mathbf{x}), i = 0, 1, \cdots, m, j = 1, \cdots, p$ 可微，令 $\mathbf{x}^*$ 和 $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ 为一对对偶问题最优解，对偶间隙为零。因为 $L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ 关于 $\mathbf{x}$ 求极小，在 $\mathbf{x}^*$ 处取得最小值，因此有

$$\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j^* \nabla h_j(\mathbf{x}^*) = 0$$

因此我们有

$$\begin{cases} f_i(\mathbf{x}^*) \leq 0, & i = 1, \cdots, m \\ h_j(\mathbf{x}^*) = 0, & j = 1, \cdots, p \\ \lambda_i \geq 0, & i = 1, \cdots, m \\ \lambda_i^* f_i(\mathbf{x}^*) = 0, & i = 1, \cdots, m \\ \nabla f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j^* \nabla h_j(\mathbf{x}^*) = 0, & \end{cases} \tag{21}$$

式(21)被称作Karush-Kuhn-Tucker(KKT) 条件。如果强对偶性成立，那么任何一对原问题最优解和对偶问题最优解必须满足 KKT 条件

凸问题的 KKT 条件

当原问题是凸问题时，满足 KKT 条件的点也是原、对偶问题的最优解。换言之，如果函数 $f_i(\mathbf{x})$ 是凸函数，$h_j(\mathbf{x})$ 是仿射函数，，令 $\mathbf{x}'$ 和 $(\boldsymbol{\lambda}', \boldsymbol{\mu}')$ 为任意满足 KKT 条件的点：

$$\begin{cases} f_i(\mathbf{x}') \leq 0, & i = 1, \cdots, m \\ h_j(\mathbf{x}') = 0, & j = 1, \cdots, p \\ \lambda_i \geq 0, & i = 1, \cdots, m \\ \lambda_i' f_i(\mathbf{x}') = 0, & i = 1, \cdots, m \\ \nabla f_0(\mathbf{x}') + \sum_{i=1}^{m} \lambda_i' \nabla f_i(\mathbf{x}') + \sum_{j=1}^{p} \mu_j' \nabla h_j(\mathbf{x}') = 0, & \end{cases}$$

那么 $\mathbf{x}'$ 和 $(\boldsymbol{\lambda}', \boldsymbol{\mu}')$ 分别为原问题和对偶问题的最优解，对偶间隙为零。前两个条件说明 $\mathbf{x}'$ 是原问题的可行解；因为 $\lambda_i \geq 0$ 可知 $L(\mathbf{x}, \boldsymbol{\lambda}', \boldsymbol{\mu}')$ 是 $\mathbf{x}$ 的凸函数；最后一个 KKT 条件说明 L 在 $\mathbf{x} = \mathbf{x}'$ 处导数为零。因此有

$$g(\boldsymbol{\lambda}', \boldsymbol{\mu}') = L(\mathbf{x}', \boldsymbol{\lambda}', \boldsymbol{\mu}') \quad = f_0(\mathbf{x}') + \sum_{i=1}^{m} \lambda_i' f_i(\mathbf{x}') + \sum_{j=1}^{p} \mu_j' h_j(\mathbf{x}') = f_0(\mathbf{x}')$$

凸问题的 KKT 条件

- 因此，对目标函数和约束函数可微的任意凸优化问题，任意满足 KKT 条件的点分别是原对偶问题的最优解，对偶间隙为零
- 若某个凸优化问题具有可微的目标函数和约束函数，且满足 Slater 条件，那么 KKT 条件就是最优性的充要条件：Slater 条件意味着最优对偶间隙为零且对偶最优解可达；因此 $\mathbf{x}$ 是原问题的最优解，当且仅当存在 $(\boldsymbol{\lambda}, \boldsymbol{\mu})$，二者满足 KKT 条件
- KKT 条件在最优化领域有重要作用。在某些特殊情况下，是可以解析求解 KKT 条件的
- 更一般地，很多求解凸优化问题的方法可以认为或者理解为求解 KKT 条件的方法

等式约束二次凸问题求极小

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{P}\mathbf{x} + \mathbf{q}^{\mathrm{T}}\mathbf{x} + \mathbf{r} \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \end{array} \tag{22}$$

此问题的 KKT 条件为

$$\mathbf{A}\mathbf{x}^* = \mathbf{b}, \quad \mathbf{P}\mathbf{x}^* + \mathbf{q} + \mathbf{A}^{\mathrm{T}}\boldsymbol{\mu}^* = 0$$

可以写成

$$\begin{pmatrix} \mathbf{P} & \mathbf{A}^{\mathrm{T}} \\ \mathbf{A} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{x}^* \\ \boldsymbol{\mu}^* \end{pmatrix} = \begin{pmatrix} -\mathbf{q} \\ \mathbf{b} \end{pmatrix}$$

求解变量 $\mathbf{x}^*, \boldsymbol{\mu}^*$ 的 $m+n$ 个方程，其中变量维数为 $m+n$，可以得到(22)对偶问题和原问题的最优解

通过对偶问题求原问题

- 如果强对偶性成立且存在一个对偶最优解 $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$，那么任意原问题最优点也是 $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ 的最优解
- 假设强对偶性成立，对偶最优解 $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ 已知，假设 $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ 的最优解，即

$$\text{minimize} \quad f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j^* h_j(\mathbf{x}) \tag{23}$$

  的解唯一（对于凸问题，如果 L 是 $\mathbf{x}$ 的严格凸函数就会出现这种情况），那么如果问题(23)的解是原问题的可行解，就是原问题的最优解。如果它不是原问题可行解，那么原问题最优解不可达
- 当对偶问题可以解析求解或者具有某些更易分析的结构时，上述方法很有意义

最大熵问题的对偶函数

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) = \sum_{i=1}^{n} x_i \log x_i \\
\text{subject to} \quad & \mathbf{A}\mathbf{x} \succeq \mathbf{b} \\
& \mathbf{1}^{\mathrm{T}}\mathbf{x} = 1
\end{aligned}
$$

从而对偶问题可简化为

$$
\begin{aligned}
\text{maximize} \quad & -\mathbf{b}^{\mathrm{T}}\boldsymbol{\lambda} - \log\left(\sum_{i=1}^{n} e^{-\mathbf{a}_i^{\mathrm{T}}\boldsymbol{\lambda}}\right) \\
\text{subject to} \quad & \boldsymbol{\lambda} \geq 0
\end{aligned}
$$

假设 Slater 条件的弱化形式成立：也就是存在 $\mathbf{x} \succ 0$ 使得 $\mathbf{A}\mathbf{x} \succeq \mathbf{b}, \mathbf{1}^{\mathrm{T}}\mathbf{x} = 1$。因此强对偶性成立，存在一个最优解 $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$

最大熵问题的对偶函数

设对偶问题已经解出，$(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ 处的 Lagrange 函数为

$$L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \sum_{i=1}^{n} x_i \log x_i + (\boldsymbol{\lambda}^*)^{\mathrm{T}} (\mathbf{A}\mathbf{x} - \mathbf{b}) + \boldsymbol{\mu}^* (\mathbf{1}^{\mathrm{T}} \mathbf{x} - 1)$$

它在 $\mathcal{D}$ 上严格凸且有下界，因此有一个唯一解 $\mathbf{x}^*$，

$$x_i^* = \frac{1}{e^{(a_i^{\mathrm{T}} \boldsymbol{\lambda}^* + \boldsymbol{\mu}^* + 1)}}, \quad i = 1, \cdots, n,$$

如果 $\mathbf{x}^*$ 是原问题可行解，则必是原问题的最优解，如果不是原问题可行解，则可以说原问题的最优解不能达到