# 第十一章 最优性条件和对偶理论

## 第 32 讲 最优性条件和对偶理论

黄定江

DaSE @ ECNU

djhuang@dase.ecnu.edu.cn

## 关于最优性条件

为什么需要考虑优化问题的最优性条件?

- 方便验证一个点是否为极小值点
- 通过最优性条件获得解析解求解的表达式
- 通过最优性条件获得数值解求解的迭代方法

本小节主要考虑如下问题的最优性理论:

- 无约束可微问题
- 无约束不可微问题

## 32.1.1 无约束可微问题的最优性理论

先考虑如下无约束可微优化问题：

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}), \tag{1}$$

其中 $f(\boldsymbol{x})$ 是连续可微函数.

# 1. 一阶最优性条件

根据多元微积分的知识：

**定理 1**

假设 $f$ 在全空间 $\mathbb{R}^n$ 可微。若 $\boldsymbol{x}^*$ 是一个局部极小解，那么

$$\nabla f(\boldsymbol{x}^*) = 0 \tag{2}$$

Proof.

任取 $\boldsymbol{v} \in \mathbb{R}^n$, 考虑 $f$ 在点 $\boldsymbol{x} = \boldsymbol{x}^*$ 处的泰勒展开

$$f(\boldsymbol{x}^* + t\boldsymbol{v}) = f(\boldsymbol{x}^*) + t\boldsymbol{v}^{\mathrm{T}}\nabla f(\boldsymbol{x}^*) + o(t),$$

整理得

$$\frac{f(\boldsymbol{x}^* + t\boldsymbol{v}) - f(\boldsymbol{x}^*)}{t} = \boldsymbol{v}^{\mathrm{T}}\nabla f(\boldsymbol{x}^*) + o(1).$$

根据 $\boldsymbol{x}^*$ 的最优性, 在上式中分别对 $t$ 取点 0 处的左、右极限可知

$$\lim_{t\to 0^+} \frac{f(\boldsymbol{x}^* + t\boldsymbol{v}) - f(\boldsymbol{x}^*)}{t} = \boldsymbol{v}^{\mathrm{T}}\nabla f(\boldsymbol{x}^*) \geqslant 0,$$

$$\lim_{t\to 0^-} \frac{f(\boldsymbol{x}^* + t\boldsymbol{v}) - f(\boldsymbol{x}^*)}{t} = \boldsymbol{v}^{\mathrm{T}}\nabla f(\boldsymbol{x}^*) \leqslant 0,$$

即对任意的 $\boldsymbol{v}$ 有 $\boldsymbol{v}^{\mathrm{T}}\nabla f(\boldsymbol{x}^*) = 0$, 由 $\boldsymbol{v}$ 的任意性知 $\nabla f(\boldsymbol{x}^*) = 0$.

## 2. 二阶最优性条件

如果一阶必要条件满足，我们仍然不能确定当前点是否是一个局部极小点．这里考虑使用二阶信息来进一步判断给定点的最优性。

**定理 2**

假设 $f$ 在点 $\boldsymbol{x}^*$ 的一个开邻域内是二阶连续可微的，则以下最优性条件成立：

- 二阶必要条件：如果 $\boldsymbol{x}^*$ 是 $f$ 的一个局部极小点，那么

$$\nabla f(\boldsymbol{x}^*) = 0, \quad \nabla^2 f(\boldsymbol{x}^*) \succeq 0.$$

- 二阶充分条件：如果在点 $\boldsymbol{x}^*$ 处，有

$$\nabla f(\boldsymbol{x}^*) = 0, \quad \nabla^2 f(\boldsymbol{x}^*) \succ 0$$

  成立，那么 $\boldsymbol{x}^*$ 是 $f$ 的一个局部极小点。

Proof.

考虑 $f(\boldsymbol{x})$ 在点 $\boldsymbol{x}^*$ 处的二阶泰勒展开，

$$f(\boldsymbol{x}+\boldsymbol{d}) = f(\boldsymbol{x}) + \frac{1}{2}\boldsymbol{d}^{\mathrm{T}}\nabla^2 f(\boldsymbol{x})\boldsymbol{d} + o\left(\|\boldsymbol{d}\|^2\right)$$

这里因为一阶必要条件成立，所以 $\nabla f(\boldsymbol{x}^*) = 0$. 反设 $\nabla^2 f(\boldsymbol{x}^*) \succeq \mathbf{0}$ 不成立，即 $\nabla^2 f(\boldsymbol{x}^*)$ 有负的特征值. 取 $\boldsymbol{d}$ 为其负特征值 $\lambda_-$ 对应的特征向量，通过对上式变形得到

$$\frac{f(\boldsymbol{x}^*+\boldsymbol{d}) - f(\boldsymbol{x}^*)}{\|\boldsymbol{d}\|^2} = \frac{1}{2}\frac{\boldsymbol{d}^{\mathrm{T}}}{\|\boldsymbol{d}\|}\nabla^2 f(\boldsymbol{x}^*)\frac{\boldsymbol{d}}{\|\boldsymbol{d}\|} + o(1).$$

这里注意 $\frac{\boldsymbol{d}}{\|\boldsymbol{d}\|}$ 是 $\boldsymbol{d}$ 的单位化， $\square$

Proof.

（续）因此

$$\frac{f(\boldsymbol{x}^*+\boldsymbol{d})-f(\boldsymbol{x}^*)}{\|\boldsymbol{d}\|^2}=\frac{1}{2}\lambda_-+o(1).$$

当 $\|\boldsymbol{d}\|$ 充分小时, $f(\boldsymbol{x}^*+\boldsymbol{d})<f(\boldsymbol{x}^*)$, 这和点 $\boldsymbol{x}^*$ 的最优性矛盾. 因此二阶必要条件成立. 当 $\nabla^2 f(\boldsymbol{x})\succ\mathbf{0}$ 时, 对任意的 $\boldsymbol{d}\neq\mathbf{0}$ 有 $\boldsymbol{d}^{\mathrm{T}}\nabla^2 f(\boldsymbol{x}^*)\,\boldsymbol{d}\geqslant\lambda_{\min}\|\boldsymbol{d}\|^2>0$, 这里 $\lambda_{\min}>0$ 是 $\nabla^2 f(\boldsymbol{x}^*)$ 的最小特征值. 因此我们有

$$\frac{f(\boldsymbol{x}^*+\boldsymbol{d})-f(\boldsymbol{x}^*)}{\|\boldsymbol{d}\|^2}\geqslant\frac{1}{2}\lambda_{\min}+o(1).$$

当 $\|\boldsymbol{d}\|$ 充分小时有 $f(\boldsymbol{x}^*+\boldsymbol{d})\geqslant f(\boldsymbol{x}^*)$, 即二阶充分条件成立. □

我们以线性最小二乘问题为例来说明其最优性条件的具体形式.

例 1

线性最小二乘问题可以表示为

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \quad f(\boldsymbol{x}) \overset{def}{=} \frac{1}{2}\|\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}\|_2^2,$$

其中 $\boldsymbol{A}\in\mathbb{R}^{m\times n}, \boldsymbol{b}\in\mathbb{R}^m$ 分别是给定的矩阵和向量. 易知 $f(\boldsymbol{x})$ 是可微且凸的, 因此, $\boldsymbol{x}^*$ 为一个全局最优解当且仅当

$$\nabla f(\boldsymbol{x}^*) = \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b}) = 0.$$

因此, 线性最小二乘问题本质上等于求解线性方程组, 可以利用数值代数知识对其有效求解.

## 32.1.2 无约束不可微问题的最优性条件

仍考虑问题:

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}),$$

但其中 $f(\boldsymbol{x})$ 为不可微函数. 很多实际问题的目标函数不是光滑的。

例如 $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_1$. 对于此类问题, 由于目标函数可能不存在梯度和海瑟矩阵, 因此上一小节中的一阶和二阶条件不适用. 此时我们必须使用其他最优性条件来判断不可微问题的最优点。

# 1. 凸优化问题一阶最优性条件

对于目标函数是凸函数的情形，我们已经引入了次梯度的概念并给出了其计算法则. 一个自然的问题是：可以利用次梯度代替梯度来构造最优性条件吗？实际上有如下定理:

**定理 3**

假设 $f$ 是适当且凸的函数，则 $\boldsymbol{x}^*$ 为无约束优化问题的一个全局极小点当且仅当

$$0 \in \partial f(\boldsymbol{x}^*)$$

Proof.

先证必要性. 因为 $\boldsymbol{x}^*$ 为全局极小点, 所以

$$f(\boldsymbol{y}) \geqslant f(\boldsymbol{x}^*) = f(\boldsymbol{x}^*) + 0^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}^*), \quad \forall \boldsymbol{y} \in \mathbb{R}^n.$$

因此, $0 \in \partial f(\boldsymbol{x}^*)$. 再证充分性. 如果 $0 \in \partial f(\boldsymbol{x}^*)$, 那么根据次梯度的定义

$$f(\boldsymbol{y}) \geqslant f(\boldsymbol{x}^*) + 0^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}^*) = f(\boldsymbol{x}^*), \quad \forall \boldsymbol{y} \in \mathbb{R}^n.$$

因而 $\boldsymbol{x}^*$ 为一个全局极小点. □

这说明条件 $0 \in \partial f(\boldsymbol{x}^*)$ 是 $\boldsymbol{x}^*$ 为全局最优解的充要条件. 这个结论比前面的一阶条件要强, 其原因是凸问题有非常好的性质。

## 2. 复合优化问题的一阶最优性条件

在实际问题中, 目标函数不一定是凸函数, 但它可以写成一个光滑函数与一个非光滑凸函数的和。在压缩感知中，我们使用 $\ell_1$ 范数来获得信号的稀疏性；再比如在机器学习中使用 $\ell_1$ 正则化；还有经典的 LASSO 回归问题。这时我们需要考虑复合优化问题

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \psi(\boldsymbol{x}) \stackrel{\text{def}}{=} f(\boldsymbol{x}) + h(\boldsymbol{x}), \tag{3}$$

其中 $f$ 为光滑函数 (可能非凸), $h$ 为凸函数 (可能非光滑).

## 不可微复合优化一阶最优性条件

对于这种情形, 我们给出如下一阶必要条件:

**定理 4**

令 $\boldsymbol{x}^*$ 为问题 (3) 的一个局部极小点, 那么

$$-\nabla f(\boldsymbol{x}^*) \in \partial h(\boldsymbol{x}^*),$$

其中 $\partial h(\boldsymbol{x}^*)$ 为凸函数 $h$ 在点 $\boldsymbol{x}^*$ 处的次梯度集合.

Proof.

因为 $\boldsymbol{x}^*$ 为一个局部极小点，所以对于任意单位向量 $\boldsymbol{d} \in \mathbb{R}^n$ 和足够小的 $t > 0$,

$$f(\boldsymbol{x}^* + t\boldsymbol{d}) + h(\boldsymbol{x}^* + t\boldsymbol{d}) \geqslant f(\boldsymbol{x}^*) + h(\boldsymbol{x}^*).$$

给定任一方向 $\boldsymbol{d} \in \mathbb{R}^n$, 其中 $\|\boldsymbol{d}\| = 1$. 因为对光滑函数和凸函数都可以考虑方向导数，根据方向导数的定义,

$$\begin{aligned}
\psi'(\boldsymbol{x}^*; \boldsymbol{d}) &= \lim_{t \to 0+} \frac{\psi(\boldsymbol{x}^* + t\boldsymbol{d}) - \psi(\boldsymbol{x}^*)}{t} \\
&= \nabla f(\boldsymbol{x}^*)^{\mathrm{T}} \boldsymbol{d} + \partial h(\boldsymbol{x}^*; \boldsymbol{d}) \\
&= \nabla f(\boldsymbol{x}^*)^{\mathrm{T}} \boldsymbol{d} + \sup_{\theta \in \partial h(\boldsymbol{x}^*)} \theta^{\mathrm{T}} \boldsymbol{d},
\end{aligned}$$

其中 $\partial h(\boldsymbol{x}^*; \boldsymbol{d})$ 表示凸函数 $h(\boldsymbol{x})$ 在点 $\boldsymbol{x}^*$ 处的方向导数，最后一个等式利用了凸函数方向导数和次梯度的关系。 □

Proof.

（续）现在用反证法证明我们所需要的结论. 反设 $-\nabla f(\boldsymbol{x}^*) \notin \partial h(\boldsymbol{x}^*)$, 根据次梯度的性质可知 $\partial h(\boldsymbol{x}^*)$ 是有界闭凸集, 又根据严格分离定理, 存在 $\boldsymbol{d} \in \mathbb{R}^n$ 以及常数 $b$ 使得

$$\theta^{\mathrm{T}} \boldsymbol{d} < b < -\nabla f(\boldsymbol{x}^*)^{\mathrm{T}} \boldsymbol{d}, \quad \forall \theta \in \partial h(\boldsymbol{x}^*).$$

根据 $\partial h(\boldsymbol{x}^*)$ 是有界闭集可知对此方向 $\boldsymbol{d}$，

$$\psi'(\boldsymbol{x}^*; \boldsymbol{d}) = \nabla f(\boldsymbol{x}^*)^{\mathrm{T}} \boldsymbol{d} + \sup_{\theta \in \partial h(\boldsymbol{x}^*)} \theta^{\mathrm{T}} \boldsymbol{d} < 0.$$

这说明对充分小的非负实数 $t$,

$$\psi(\boldsymbol{x}^* + t\boldsymbol{d}) < \psi(\boldsymbol{x}^*).$$

这与 $\boldsymbol{x}^*$ 的局部极小性矛盾. 因此 $-\nabla f(\boldsymbol{x}^*) \in \partial h(\boldsymbol{x}^*)$. □

## 例 2

我们以 $\ell_1$ 范数正则化的优化问题为例，给出其最优解的最优性条件。前面我们已经介绍其一般形式可以写成

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \psi(\boldsymbol{x}) \stackrel{\text{def}}{=} f(\boldsymbol{x}) + \mu\|\boldsymbol{x}\|_1,$$

其中 $f(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$ 为光滑函数，正则系数 $\mu > 0$ 用来调节解的稀疏度.

尽管 $\|\boldsymbol{x}\|_1$ 不是可微的，但我们可以计算其次微分，在次梯度计算的例子中，我们已经计算出

$$\partial_i\|\boldsymbol{x}\|_1 = \begin{cases} \{1\}, & x_i > 0, \\ [-1,1], & x_i = 0, \\ \{-1\}, & x_i < 0. \end{cases}$$

## 例2 续

因此，如果 $\boldsymbol{x}^*$ 是优化问题的一个局部最优解，那么其满足

$$-\nabla f(\boldsymbol{x}^*) \in \mu\partial \|\boldsymbol{x}^*\|_1 ,$$

即

$$\nabla_i f(\boldsymbol{x}^*) = \begin{cases} -\mu, & x_i^* > 0, \\ a \in [-\mu, \mu], & x_i^* = 0, \\ \mu, & x_i^* < 0. \end{cases}$$

进一步地, 如果 $f(\boldsymbol{x})$ 是凸的（比如在 LASSO 回归中 $f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2$），那么满足上式的 $\boldsymbol{x}^*$ 就是全局最优解。

## 32.2.1 约束优化问题

现考虑标准形式的约束优化问题 (不一定是凸的)

$$
\begin{aligned}
\min \quad & f_0(\boldsymbol{x}) \\
\text{s.t.} \quad & f_i(\boldsymbol{x}) \leq 0, \quad i = 1, \cdots, m \\
& h_j(\boldsymbol{x}) = 0, \quad j = 1, \cdots, p
\end{aligned}
\tag{4}
$$

其中自变量 $\boldsymbol{x} \in \mathbb{R}^n$，假设定义域 $\mathcal{D} = \bigcap_{i=0}^{m} \textbf{dom}\, f_i \cap \bigcap_{j=1}^{p} \textbf{dom}\, h_j$ 非空。

# 约束优化问题

例 3

最大割问题：

$$\begin{aligned} \min \quad & \boldsymbol{x}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{x} \\ \textsf{\textit{s.t.}} \quad & x_i^2 = 1, \quad i = 1, \cdots, n \end{aligned}$$

其中，$\boldsymbol{W} \in \mathcal{S}^n$。

# 约束优化问题

**例 4**

支持向量机:

$$\begin{aligned} \min_{\boldsymbol{w},b} \quad & \frac{1}{2}\|\boldsymbol{w}\|^2 \\ s.t. \quad & y_i(\boldsymbol{w}^{\mathrm{T}} x_i + b) \geq 1, \quad i = 1, 2, \cdots, N \end{aligned}$$

## 32.2.2 引入：为什么需要对偶

- 约束优化问题的最优性理论分析
- 非凸的优化问题可以转化为凸优化问题便于求解
- 凸的计算复杂度高的问题转化为凸的计算复杂度低的问题
- 帮助证明原问题无解
- 便于敏感性分析

## 拉格朗日函数

**拉格朗日函数**：$L: \mathcal{D} \times \mathbb{R}_+^m \times \mathbb{R}^p \to \mathbb{R}$ 为

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

其中定义域为 $\mathbf{dom}\, L = \mathcal{D} \times \mathbb{R}_+^m \times \mathbb{R}^p$

- 拉格朗日函数：添加约束条件的加权和，得到增广的目标函数
- $\lambda_i$ 称为第 $i$ 个不等式约束 $f_i(\boldsymbol{x}) \leq 0$ 对应的**拉格朗日乘子**
- $\nu_j$ 称为第 $j$ 个等式约束 $h_j(\boldsymbol{x}) = 0$ 对应的**拉格朗日乘子**
- 向量 $\boldsymbol{\lambda}$ 和 $\boldsymbol{\nu}$ 称为对偶变量或者问题 (4) 的**拉格朗日乘子向量**

## 拉格朗日对偶函数

**定义 1**

**拉格朗日对偶函数 (或简称对偶函数)** $g:\mathbb{R}_+^m\times\mathbb{R}^p\to\mathbb{R}$ 是拉格朗日函数关于 $\boldsymbol{x}$ 取得的下确界：即对 $\boldsymbol{\lambda}\in\mathbb{R}_+^m,\boldsymbol{\nu}\in\mathbb{R}^p$，有

$$g(\boldsymbol{\lambda},\boldsymbol{\nu})=\inf_{\boldsymbol{x}\in\mathcal{D}}L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\nu})=\inf_{\boldsymbol{x}\in\mathcal{D}}\left(f_0(\boldsymbol{x})+\sum_{i=1}^m\lambda_i f_i(\boldsymbol{x})+\sum_{j=1}^p\nu_j h_j(\boldsymbol{x})\right)$$

- 如果函数关于 $\boldsymbol{x}$ 无下界，则对偶函数取值为 $-\infty$
- 对偶函数是一族关于 $(\boldsymbol{\lambda},\boldsymbol{\nu})$ 的仿射函数的逐点下确界，所以即使原问题 (4) 非凸，对偶函数也是凹函数

## 32.2.3 最优值的下界

**引理 1**

对偶函数构成了原问题 (4) 最优值 $p^*$ 的下界：即对任意 $\boldsymbol{\lambda} \geq \mathbf{0}$ 和 $\boldsymbol{\nu}$，下式成立

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^* \tag{5}$$

Proof.

设 $\tilde{\boldsymbol{x}}$ 是原问题 (4) 的一个可行点，即 $f_i(\tilde{\boldsymbol{x}}) \leq 0$ 且 $h_j(\tilde{\boldsymbol{x}}) = 0$。根据假设，$\boldsymbol{\lambda} \geq 0$，有

$$\sum_{i=1}^{m} \lambda_i f_i(\tilde{\boldsymbol{x}}) + \sum_{j=1}^{p} \nu_j h_j(\tilde{\boldsymbol{x}}) \leq 0$$

左边第一项非正，第二项为零。 □

## 最优值的下界

证明（续）.

根据上述不等式，有

$$L(\tilde{\boldsymbol{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\tilde{\boldsymbol{x}}) + \sum_{i=1}^{m} \lambda_i f_i(\tilde{\boldsymbol{x}}) + \sum_{j=1}^{p} \nu_j h_j(\tilde{\boldsymbol{x}}) \le f_0(\tilde{\boldsymbol{x}}).$$

因此

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \le L(\tilde{\boldsymbol{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \le f_0(\tilde{\boldsymbol{x}})$$

由于每一个可行点 $\tilde{\boldsymbol{x}}$ 都满足 $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \le f_0(\tilde{\boldsymbol{x}})$，因此 $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \le p^*$ 成立 □

## 32.2.4 求拉格朗日对偶函数的解析表达式

### 线性方程组的最小二乘解

例 5

考虑问题

$$\begin{aligned} \min \quad & \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \\ s.t. \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned} \tag{6}$$

其中 $\boldsymbol{A} \in \mathbb{R}^{p\times n}$，求其对偶函数。

解

此问题无不等式约束，有 $p$ 个 (线性) 等式约束，其拉格朗日函数为

$$L(\boldsymbol{x}, \boldsymbol{\nu}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$$

定义域为 $\mathbb{R}^n \times \mathbb{R}^p$。对偶函数为

$$g(\boldsymbol{\nu}) = \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\nu})$$

解

（续）因为 $L(\boldsymbol{x},\boldsymbol{\nu})$ 是 $\boldsymbol{x}$ 的二次凸函数，可通过求解最优性条件得到函数的最小值

- $L$ 关于 $\boldsymbol{x}$ 取极小值，梯度为零：

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{\nu}) = 2\boldsymbol{x} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\nu} = \boldsymbol{0} \implies \boldsymbol{x} = -(1/2)\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\nu}$$

- 代入到 $L$ 得到**对偶函数**$g$:

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

  是关于 $\boldsymbol{\nu}$ 的二次凹函数，定义域为 $\mathbb{R}^p$

**最优值下界**：根据引理 1，对 $\forall \boldsymbol{\nu} \in \mathbb{R}^p$

$$p^* \geq -(1/4)\boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\nu} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\nu}$$

# 标准形式的线性规划

例 6

考虑问题

$$\begin{aligned} \min \quad & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ s.t. \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{aligned} \tag{7}$$

求其对偶函数。

解

构造拉格朗日函数：

$$\begin{aligned} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) &= \boldsymbol{c}^{\mathrm{T}} \boldsymbol{x} - \sum_{i=1}^{n} \lambda_i x_i + \boldsymbol{\nu}^{\mathrm{T}} (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) \\ &= -\boldsymbol{b}^{\mathrm{T}} \boldsymbol{\nu} + (\boldsymbol{c} + \boldsymbol{A}^{\mathrm{T}} \boldsymbol{\nu} - \boldsymbol{\lambda})^{\mathrm{T}} \boldsymbol{x} \end{aligned}$$

$L$ 关于 $\boldsymbol{x}$ 线性，因此对偶函数为

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{cases} -\boldsymbol{b}^{\mathrm{T}} \boldsymbol{\nu} & \boldsymbol{A}^{\mathrm{T}} \boldsymbol{\nu} - \boldsymbol{\lambda} + \boldsymbol{c} = \boldsymbol{0} \\ -\infty & \text{其他情况} \end{cases}$$

最优值下界：如果 $\boldsymbol{A}^{\mathrm{T}} \boldsymbol{\nu} + \boldsymbol{c} \geq \boldsymbol{0}$，$p^* \geq -\boldsymbol{b}^{\mathrm{T}} \boldsymbol{\nu}$

## 最大割问题（双向划分问题）

**例 7**

考虑问题

$$\begin{aligned} \min \quad & \boldsymbol{x}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{x} \\ \text{s.t.} \quad & x_i^2 = 1, \quad i = 1, \cdots, n \end{aligned} \tag{8}$$

其中，$\boldsymbol{W} \in S^n$，求其对偶函数。

解

**拉格朗日函数：**

$$
\begin{aligned}
L(\boldsymbol{x}, \boldsymbol{\nu}) =& \boldsymbol{x}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{x} + \sum_{i=1}^{n} \nu_i (x_i^2 - 1) \\
=& \boldsymbol{x}^{\mathrm{T}} (\boldsymbol{W} + \operatorname{diag}(\boldsymbol{\nu})) \boldsymbol{x} - \mathbf{1}^{\mathrm{T}} \boldsymbol{\nu}
\end{aligned}
$$

**对偶函数:**

$$
\begin{aligned}
g(\boldsymbol{\nu}) = \inf_{\boldsymbol{x}} (\boldsymbol{x}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{x} + \sum_{i} \nu_i (x_i^2 - 1)) &= \inf_{\boldsymbol{x}} \boldsymbol{x}^{\mathrm{T}} (\boldsymbol{W} + \operatorname{diag}(\boldsymbol{\nu})) \boldsymbol{x} - \mathbf{1}^{\mathrm{T}} \boldsymbol{\nu} \\
&= \begin{cases} -\mathbf{1}^{\mathrm{T}} \boldsymbol{\nu} & \boldsymbol{W} + \operatorname{diag}(\boldsymbol{\nu}) \succeq \mathbf{0} \\ -\infty & \text{其他情况} \end{cases}
\end{aligned}
$$

**最优值下界：** 若取 $\boldsymbol{\nu} = -\lambda_{\min}(\boldsymbol{W}) \mathbf{1}$，则 $\boldsymbol{W} + \operatorname{diag}(\boldsymbol{\nu}) \succeq \mathbf{0}$，可得 $p^*$ 的一个下界：
$p^* \geq -\mathbf{1}^{\mathrm{T}} \boldsymbol{\nu} = n \lambda_{\min}(\boldsymbol{W})$

## 32.2.5 拉格朗日对偶函数与共轭函数的关系

回忆函数 $f: \mathbb{R}^n \to \mathbb{R}$ 的共轭函数 $f^*$ 为

$$f^*(\boldsymbol{y}) = \sup_{\boldsymbol{x} \in \mathbf{dom}\, f} (\boldsymbol{y}^{\mathrm{T}} \boldsymbol{x} - f(\boldsymbol{x}))$$

例 8

考虑问题

$$\begin{aligned} \min \quad & f_0(\boldsymbol{x}) \\ s.t. \quad & \boldsymbol{x} = \boldsymbol{0} \end{aligned} \tag{9}$$

该问题的拉格朗日函数为 $L(\boldsymbol{x}, \boldsymbol{\nu}) = f_0(\boldsymbol{x}) + \boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{x}$，对偶函数为

$$g(\boldsymbol{\nu}) = \inf_{\boldsymbol{x}}(f_0(\boldsymbol{x}) + \boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{x}) = -\sup_{\boldsymbol{x}}((-\boldsymbol{\nu})^{\mathrm{T}}\boldsymbol{x} - f_0(\boldsymbol{x})) = -f_0^*(-\boldsymbol{\nu})$$

## 例 9

考虑问题

$$
\begin{aligned}
\min \quad & f_0(\boldsymbol{x}) \\
s.t. \quad & \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b} \\
& \boldsymbol{C}\boldsymbol{x} = \boldsymbol{d}
\end{aligned} \tag{10}
$$

利用函数 $f_0$ 的共轭函数，其对偶函数为

$$
\begin{aligned}
g(\lambda, \nu) &= \inf_{\boldsymbol{x}} \left( f_0(\boldsymbol{x}) + \lambda^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) + \boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{C}\boldsymbol{x} - \boldsymbol{d}) \right) \\
&= -\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{d}^{\mathrm{T}}\boldsymbol{\nu} + \inf_{\boldsymbol{x}} \left( f_0(\boldsymbol{x}) + (\boldsymbol{A}^{\mathrm{T}}\lambda + \boldsymbol{C}^{\mathrm{T}}\boldsymbol{\nu})^{\mathrm{T}}\boldsymbol{x} \right) \\
&= -\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{d}^{\mathrm{T}}\boldsymbol{\nu} - f_0^*(-\boldsymbol{A}^{\mathrm{T}}\lambda - \boldsymbol{C}^{\mathrm{T}}\boldsymbol{\nu})
\end{aligned} \tag{11}
$$

函数 $g$ 的定义域　$\mathbf{dom}\, g = \{(\boldsymbol{\lambda}, \boldsymbol{\nu}) | -\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{C}^{\mathrm{T}}\boldsymbol{\nu} \in \mathbf{dom}\, f_0^*\}$

- 这些例子告诉我们对偶函数与共轭函数之间的转换关系
- 在第 10 章的内容中，已经计算过许多函数的共轭函数
- 因此，可以利用前面章节共轭函数的结论，以及它与对偶函数的关系，直接求得对偶函数

## 32.2.6 用共轭函数求解拉格朗日对偶函数

### 等式约束条件下的范数极小化

**例 10**

考虑问题

$$\begin{aligned} \min \quad & f_0(\boldsymbol{x}) = \|\boldsymbol{x}\| \\ s.t. \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned} \tag{12}$$

其中 $\|\cdot\|$ 是任意范数，求其对偶函数。

## 等式约束条件下的范数极小化

解

$f_0 = \|\cdot\|$ 的共轭函数为

$$f_0^*(\boldsymbol{y}) = \begin{cases} 0 & \|\boldsymbol{y}\|_* \le 1 \\ \infty & \text{其他情况} \end{cases}$$

可以看出此函数是对偶范数单位球的示性函数。利用结论 (11) 得到问题 (12) 的对偶函数为

$$g(\boldsymbol{\nu}) = -\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\nu} - f_0^*(-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\nu}) = \begin{cases} -\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\nu} & \|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\nu}\|_* \le 1 \\ -\infty & \text{其他情况} \end{cases}$$

# 熵的最大化

例 11

考虑问题

$$\begin{aligned} \min \quad & f_0(\boldsymbol{x}) = \sum_{i=1}^{n} x_i \log x_i \\ s.t. \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \\ & \mathbf{1}^{\mathrm{T}}\boldsymbol{x} = 1 \end{aligned} \tag{13}$$

其中 $\mathbf{dom}\, f_0 = \mathbb{R}^n_{++}$，求其对偶函数。

# 熵的最大化

解

关于实变量 $x$ 的负熵函数 $x\log x$ 的共轭函数是 $e^{y-1}$。由于函数 $f_0$ 是不同变量的负熵函数的和，其共轭函数为

$$f_0^*(\boldsymbol{y}) = \sum_{i=1}^{n} e^{y_i-1}$$

其定义域为 $\mathbf{dom}\, f_0^* = \mathbb{R}^n$。根据前述结论 *(11)*，问题 *(13)* 的对偶函数为

$$g(\boldsymbol{\lambda}, \nu) = -\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda} - \nu - \sum_{i=1}^{n} e^{-\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{\lambda}-\nu-1} = -\boldsymbol{b}^{\mathrm{T}}\lambda - \nu - e^{-\nu-1}\sum_{i=1}^{n} e^{-\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{\lambda}}$$

其中 $\boldsymbol{a}_i$ 是矩阵 $\boldsymbol{A}$ 的第 $i$ 列向量。

## 32.3.1 显式表达对偶约束

**拉格朗日对偶问题**

$$\begin{aligned} \max \quad & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{s.t.} \quad & \boldsymbol{\lambda} \geq \mathbf{0} \end{aligned} \tag{14}$$

- 从拉格朗日函数能够得到的**最好**下界
- 凸优化问题；最优值记为 $d^*$；如果 $\boldsymbol{\lambda} \geq \mathbf{0}, (\boldsymbol{\lambda}, \boldsymbol{\nu}) \in \mathbf{dom}\, g$，则 $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ 是对偶可行的，称 $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ 是**对偶最优解**或者是**最优拉格朗日乘子**
- 经常，通过使隐含约束变为显式约束，问题能够得到简化

## 标准形式线性规划的拉格朗日对偶

**例 12**

考虑问题

$$\begin{aligned} \min \quad & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ s.t. \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \quad \boldsymbol{x} \geq \mathbf{0} \end{aligned} \tag{15}$$

给出其对偶问题。

## 标准形式线性规划的拉格朗日对偶

解

**对偶函数：**

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{cases} -\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\nu} & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\nu} - \boldsymbol{\lambda} + \boldsymbol{c} = \boldsymbol{0} \\ -\infty & \text{其他情况} \end{cases}$$

- 其对偶问题是在满足约束 $\boldsymbol{\lambda} \geq \boldsymbol{0}$ 的条件下极大化对偶函数 $g$

**对偶问题：**

$$\max \quad g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{cases} -\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\nu} & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\nu} - \boldsymbol{\lambda} + \boldsymbol{c} = \boldsymbol{0} \\ -\infty & \text{其他情况} \end{cases} \tag{16}$$

$$s.t. \quad \boldsymbol{\lambda} \geq \boldsymbol{0}$$

## 标准形式线性规划的拉格朗日对偶

解

（续）当且仅当 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\nu}-\boldsymbol{\lambda}+\boldsymbol{c}=\mathbf{0}$ 时对偶函数 $g$ 有界，通过隐式约束变为显式约束，得到等价问题

$$\begin{aligned} \max \quad & -\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\nu} \\ s.t. \quad & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\nu}-\boldsymbol{\lambda}+\boldsymbol{c}=\mathbf{0} \\ & \boldsymbol{\lambda}\geq\mathbf{0} \end{aligned} \tag{17}$$

进一步，此问题可以表述为

$$\begin{aligned} \max \quad & -\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\nu} \\ s.t. \quad & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\nu}+\boldsymbol{c}\geq\mathbf{0} \end{aligned} \tag{18}$$

这个是一个不等式形式的线性规划。

## 不等式形式线性规划 (LP) 的拉格朗日对偶

**例 13**

考虑问题

$$\begin{aligned} \min \quad & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ s.t. \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \end{aligned} \tag{19}$$

给出其对偶问题。

## 不等式形式线性规划 (LP) 的拉格朗日对偶

解

**拉格朗日函数：**

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

**对偶函数：**

$$g(\boldsymbol{\lambda}) = \inf_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{\lambda}) = -\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda} + \inf_{\boldsymbol{x}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}+\boldsymbol{c})^{\mathrm{T}}\boldsymbol{x} = \begin{cases} -\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda} & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}+\boldsymbol{c}=\boldsymbol{0} \\ -\infty & \text{其他情况} \end{cases} \tag{20}$$

称对偶变量 $\boldsymbol{\lambda}$ 是对偶可行的，如果 $\boldsymbol{\lambda} \geq \boldsymbol{0}$ 且 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}+\boldsymbol{c}=\boldsymbol{0}$

## 不等式形式线性规划 (LP) 的拉格朗日对偶

解

（续）对偶问题

$$\begin{aligned} \max \quad & -\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda} \\ \text{s.t.} \quad & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} + \boldsymbol{c} = \boldsymbol{0} \\ & \boldsymbol{\lambda} \geq \boldsymbol{0} \end{aligned} \tag{21}$$

- 可发现：不等式形式线性规划的对偶问题是一个标准形式的线性规划
- 可自行证明对偶问题的对偶问题等价于原问题

## 二次规划的拉格朗日对偶

例 14

考虑问题

$$\begin{aligned} \min \quad & \boldsymbol{x}^{\mathrm{T}} \boldsymbol{P} \boldsymbol{x} \\ s.t. \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \end{aligned} \tag{22}$$

其中 $\boldsymbol{P} \in S_{+}^{n}$，求其对偶问题。

# 二次规划的拉格朗日对偶

解

**对偶函数：**

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

**对偶问题：**

$$\begin{aligned} \max \quad & -\frac{1}{4}\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda} \\ s.t. \quad & \boldsymbol{\lambda} \geq \mathbf{0} \end{aligned} \tag{23}$$

## 32.4.1 弱对偶性

用 $p^*$ 标记原问题的最优值，$d^*$ 标记 Lagrange 对偶问题的最优值。则如下定理成立：

**定理 5**

不等式

$$d^* \leq p^* \tag{24}$$

成立。这个性质称为**弱对偶性**。

- 不等式总是成立的（不论是凸问题还是非凸问题）
- $(p^* - d^*)$ 被称为最优对偶间隙
- 当原问题很难求解时，弱对偶不等式可以给出原问题最优值的一个下界

## 弱对偶性

例如，考虑双向划分问题，其对偶问题是一个半定规划问题

$$\begin{aligned} \max \quad & -\mathbf{1}^{\mathrm{T}}\boldsymbol{\nu} \\ \text{s.t.} \quad & \boldsymbol{W}+\mathrm{diag}(\boldsymbol{\nu}) \succeq \mathbf{0} \end{aligned}$$

其中，$\boldsymbol{\nu} \in \mathbb{R}^n$。对偶问题能够被有效地求解，而且对偶问题的最优值给出了双向划分问题最优值的一个下界。

## 32.4.2 强对偶性和 Slater 约束准则

一个很自然的问题：$p^*$ 和 $d^*$ 在什么条件下相等？也即，所谓的**强对偶性**：

**定义 2**

如果原问题和对偶问题的最优值相等，即等式

$$p^* = d^* \tag{25}$$

成立，最优对偶间隙为零，那么，称它们满足**强对偶性**。

- 对于一般情况，强对偶性不成立
- (一般) 只对凸问题成立
- 保证强对偶性在凸问题中成立的条件称为**约束准则**

## 强对偶性和 Slater 约束准则

现在，考虑如下形式的凸优化问题：

$$\begin{aligned} \min \quad & f_0(\boldsymbol{x}) \\ \text{s.t.} \quad & f_i(\boldsymbol{x}) \leq 0, \quad i = 1, \cdots, m, \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{aligned} \tag{26}$$

其中，函数 $f_0, \cdots, f_m$ 是凸函数。我们将证明当它的约束满足一定准则时，强对偶性将成立。

## 强对偶性和 Slater 约束准则

一个简单的约束准则是 **Slater 条件**：存在一点 $\boldsymbol{x} \in \mathbf{relint}\,\mathcal{D}$[1]使得

$$f_i(\boldsymbol{x}) < 0, \quad i = 1, \cdots, m, \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \tag{27}$$

- 满足上述条件的点有时称为**严格可行**

[1]给定集合 $\mathcal{D}$，记其仿射包 $\mathbf{affine}\,\mathcal{D}$。则：

$$\mathbf{relint}\,\mathcal{D} = \{\boldsymbol{x} \in \mathcal{D} | \exists r > 0, \text{使得 } B(\boldsymbol{x}, r) \cap \mathbf{affine}\,\mathcal{D} \subset \mathcal{D}\}$$

## 强对偶性和 Slater 约束准则

当 Slater 条件成立（且原问题是凸问题）时，强对偶性成立：

**定理 6**

**（强对偶性定理）**假设函数 $f_0, f_1, \cdots, f_m$ 以及 $h_1, \cdots, h_p$ 均为凸函数，而且满足 *Slater* 条件，那么

$$\sup_{\boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\nu}} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\boldsymbol{x} \in K} f(\boldsymbol{x}).$$

即对偶间隙为零。

# 强对偶性和 Slater 约束准则

改进的 Slater 条件：当不等式约束函数 $f_i$ 中有一些仿射函数时，仿射不等式不需要严格成立。该条件为：存在一点 $\boldsymbol{x} \in \textbf{relint}\, \mathcal{D}$ 使得

$$f_i(\boldsymbol{x}) \leq 0, \quad i = 1, \cdots, k, \quad f_i(\boldsymbol{x}) < 0, \quad i = k+1, \cdots, m, \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{28}$$

- 当所有约束条件都是线性等式或不等式且 $\textbf{dom}\, f_0$ 是开集时，改进的 Slater 条件 (28) 就是可行性条件

## 32.4.3 强对偶性成立的一些示例

### 线性方程组的最小二乘解

**例 15**

考虑问题

$$\begin{aligned} \min \quad & \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \\ s.t. \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

求其对偶问题及讨论强对偶性。

## 线性方程组的最小二乘解

解

**原问题：**

$$\begin{aligned} \min \quad & \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \\ s.t. \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

**对偶问题：**

$$\max \quad -(1/4)\boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\nu} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\nu}$$

- *Slater* 条件此时是原问题的可行性条件，强对偶性通常成立。

# 二次约束二次规划 (QCQP)

**例 16**

考虑问题

$$\begin{aligned} \min \quad & (1/2)\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_0\boldsymbol{x} + \boldsymbol{q}_0^{\mathrm{T}}\boldsymbol{x} + r_0 \\ s.t. \quad & (1/2)\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_i\boldsymbol{x} + \boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{x} + r_i \leq 0, \quad i = 1, \cdots, m \end{aligned} \tag{29}$$

其中，$\boldsymbol{P}_0 \in S_{++}^n$，$\boldsymbol{P}_i \in S_{++}^n$，$i = 1, \cdots, m$。求其对偶问题及讨论强对偶性。

## 二次约束二次规划 (QCQP)

解

**原问题**

$$\begin{aligned} \min \quad & (1/2)\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_0\boldsymbol{x} + \boldsymbol{q}_0^{\mathrm{T}}\boldsymbol{x} + r_0 \\ s.t. \quad & (1/2)\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_i\boldsymbol{x} + \boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{x} + r_i \le 0, \quad i = 1, \cdots, m \end{aligned}$$

其中，$\boldsymbol{P}_0 \in S_{++}^n$，$\boldsymbol{P}_i \in S_{++}^n$，$i = 1, \cdots, m$

*Lagrange* 函数：

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

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

## 二次约束二次规划 (QCQP)

解

（续）如果 $\boldsymbol{\lambda} \geq 0$，有 $\boldsymbol{P}(\boldsymbol{\lambda}) \succ \mathbf{0}$ 及

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

**对偶问题：**

$$\begin{aligned} \max \quad & -(1/2)\boldsymbol{q}(\boldsymbol{\lambda})^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{\lambda})^{-1}\boldsymbol{q}(\boldsymbol{\lambda}) + r(\boldsymbol{\lambda}) \\ \text{s.t.} \quad & \boldsymbol{\lambda} \geq 0 \end{aligned} \tag{30}$$

- 根据 *Slater* 条件，当二次不等式约束严格成立时，即存在一点 $\boldsymbol{x}$ 使得

$$(1/2)\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_i\boldsymbol{x} + \boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{x} + r_i < 0, \quad i = 1, \cdots, m$$

对偶问题 *(30)* 和原问题 *(29)* 之间强对偶性成立。

## 32.5.1 互补松弛性

设强对偶性成立。令 $\boldsymbol{x}^*$ 是原问题的最优解，$(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ 是对偶问题的最优解，这表明

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

因此，两个不等式取等号。这样便得出一些有意义的结论。

## 互补松弛性

$L(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ 在 $\boldsymbol{x}^*$ 处取得最小值

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

求和项的每一项都非正，因此有

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

被称为**互补松弛性**，即对每个 $i$，$\lambda_i^*$ 与 $f_i(\boldsymbol{x}^*)$ 至少有一个为 0。

## 32.5.2 Karush-Kuhn-Tucker(KKT) 最优性条件

将前面的讨论汇在一起，便形成以下 KKT 条件

**定义 3**

令 $\boldsymbol{x}^*$ 和 $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ 分别是原问题和对偶问题的最优解，其对偶间隙为零。因此有

- 原始约束：
$$\begin{aligned} f_i(\boldsymbol{x}^*) &\leq 0, \quad i = 1, \cdots, m \\ h_j(\boldsymbol{x}^*) &= 0, \quad j = 1, \cdots, p \end{aligned}$$
- 对偶约束：$\lambda_i^* \geq 0, \quad i = 1, \cdots, m$
- 互补松弛：$\lambda_i^* f_i(\boldsymbol{x}^*) = 0, \quad i = 1, \cdots, m$
- 稳定性条件：$\nabla f_0(\boldsymbol{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(\boldsymbol{x}^*) + \sum_{j=1}^p \nu_j^* \nabla h_j(\boldsymbol{x}^*) = 0$

这些公式被称为 ***KKT***（***Karush-Kuhn-Tucker***）条件（$f_i, h_j$ 可微）。

# Karush-Kuhn-Tucker(KKT) 最优性条件

## 凸问题的最优性条件

**定理 7**

*对于凸优化问题（26），如果 Slater 条件成立，那么 $\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*$ 分别是原始、对偶全局最优解当且仅当它们满足 KKT 条件。*

- 对于非凸优化问题，通常 KKT 条件只是局部最优解的一个必要条件。
- 它是无约束优化问题一阶最优性条件 $\nabla f_0(\boldsymbol{x}) = 0$ 的一个推广。
- 有时可以直接利用 KKT 条件，求得优化问题的解。

## 等式约束二次凸问题求极小

**例 17**

考虑问题

$$\begin{aligned} \min \quad & (1/2)\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{x} + \boldsymbol{q}^{\mathrm{T}}\boldsymbol{x} + r \\ s.t. \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

其中 $\boldsymbol{P} \in S_{+}^{n}$。求其 KKT 条件。

## 等式约束二次凸问题求极小

**解**

此问题的 *KKT* 条件为

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

可将其写成

$$\boldsymbol{H}_{\boldsymbol{x}} = \begin{bmatrix} \boldsymbol{P} & \boldsymbol{A}^{\mathrm{T}} \\ \boldsymbol{A} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}^* \\ \boldsymbol{\nu}^* \end{bmatrix} = \begin{bmatrix} -\boldsymbol{q} \\ \boldsymbol{b} \end{bmatrix}$$

求解变量 $\boldsymbol{x}^*, \boldsymbol{\nu}^*$ 的 $m+n$ 个方程，其中变量的维数为 $m+n$，可以得到原问题和对偶问题的最优解。

### 32.5.3 通过解对偶问题求解原问题

在原问题不易求解，且强对偶性成立时，可先将其转化为对偶问题再求解。具体方式：

- 求得对偶最优解 $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$
- 然后求下列问题的解

$$\min \quad f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}) + \sum_{j=1}^{p} \nu_j^* h_j(\boldsymbol{x})$$

- 如果此问题的解是原问题可行解，那么它就是原问题最优解
- 如果不是，那么原问题不存在最优点，即原问题的最优解无法达到

# 熵的最大化

**例 18**

考虑问题

$$\begin{aligned} \min \quad & f_0(\boldsymbol{x}) = \sum_{i=1}^{n} x_i \log x_i \\ \textit{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \\ & \mathbf{1}^{\mathrm{T}}\boldsymbol{x} = \mathbf{1} \end{aligned}$$

其中定义域为 $\mathbb{R}^n_{++}$，求其最优解。

## 熵的最大化

解

**对偶问题：**

$$\begin{aligned} \max \quad & -\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda} - \nu - e^{-\nu-1}\sum_{i=1}^{n} e^{-\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{\lambda}} \\ s.t. \quad & \boldsymbol{\lambda} \geq \mathbf{0} \end{aligned}$$

- 假设 *Slater* 条件的弱化形式成立，即存在 $\boldsymbol{x} > 0$ 使得 $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}$ 以及 $\mathbf{1}^{\mathrm{T}}\boldsymbol{x} = 1$，因此强对偶性成立，存在一个对偶最优解 $(\boldsymbol{\lambda}^*, \nu^*)$

## 熵的最大化

**解**

（续）设对偶问题已经解出。$(\boldsymbol{\lambda}^*, \nu^*)$ 处的拉格朗日函数为

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

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

$$x_i^* = 1/\exp(\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{\lambda}^* + \nu^* + 1), \quad i = 1, \cdots, n$$

其中 $\boldsymbol{a}_i$ 是矩阵 $\boldsymbol{A}$ 的列向量。如果 $\boldsymbol{x}^*$ 是原问题可行解，则其必是原问题的最优解。如果 $\boldsymbol{x}^*$ 不是原问题可行解，则原问题的最优解不能达到。

# 在等式约束下极小化可分函数

**例 19**

考虑问题

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

$$s.t. \quad \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} = b$$

其中 $\boldsymbol{a} \in \mathbb{R}^n$，$b \in \mathbb{R}$，函数 $f_i : \mathbb{R} \to \mathbb{R}$ 是可微函数，严格凸，求其最优解。

# 在等式约束下极小化可分函数

解

**拉格朗日函数：**

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

**对偶函数：**

$$\begin{aligned} g(\nu) &= -b\nu + \inf_{\boldsymbol{x}} \left( \sum_{i=1}^{n}(f_i(x_i) + \nu a_i x_i) \right) \\ &= -b\nu + \sum_{i=1}^{n} \inf_{x_i}(f_i(x_i) + \nu a_i x_i) \\ &= -b\nu - \sum_{i=1}^{n} f_i^*(-\nu a_i) \end{aligned}$$

# 在等式约束下极小化可分函数

解

（续）对偶问题：

$$\max \quad -b\nu - \sum_{i}^{n} f_i^*(-\nu a_i)$$

- 假设找到一个对偶最优解 $\nu^*$。$L(\boldsymbol{x}, \nu^*)$ 具有唯一的最小点 $\tilde{\boldsymbol{x}}$，有 $\tilde{\boldsymbol{x}} = \boldsymbol{x}^*$。这可以通过求解 $\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \nu^*) = 0$ 得到 $\boldsymbol{x}^*$，即求解方程组

$$f_i'(x_i^*) = -\nu^* a_i, \quad i = 1, \cdots, n.$$

## 32.6.1 线性可分支持向量机

分类学习最基本的想法就是基于训练集 $D$ 在样本空间中找到一个划分超平面 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b=0$，将不同类别的样本分开。但对于线性可分数据集，能将训练样本分开的划分超平面可能有很多，我们应该去找到哪一个呢?

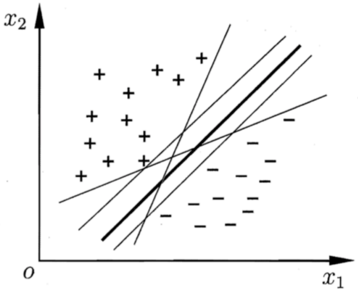


图 1: 存在多个划分超平面将两类训练样本分开

# 线性可分支持向量机

因此，期望找到具有“最大间隔”(maximun margin) 的划分超平面，将两类样本尽可能地分开。如下所示：

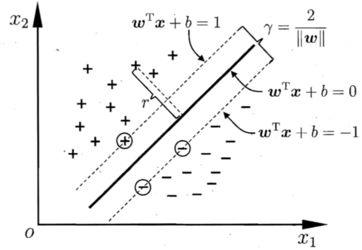


图 2: 支持向量与最大间隔

此时最大间隔为 $\gamma = \frac{2}{\|\boldsymbol{w}\|}$。

# 线性可分支持向量机

因此，该问题为要找到能分类正确，使得 $\gamma$ 最大的参数 $\boldsymbol{w}$ 和 $b$，即解如下优化问题：

$$\begin{aligned} \max_{\boldsymbol{w},b} \quad & \frac{2}{\|\boldsymbol{w}\|} \\ s.t. \quad & y_i(\boldsymbol{w}^{\mathrm{T}} x_i + b) \geq 1, \quad i = 1, 2, \cdots, N \end{aligned}$$

这等价于

$$\begin{aligned} \min_{\boldsymbol{w},b} \quad & \frac{1}{2}\|\boldsymbol{w}\|^2 \\ s.t. \quad & y_i(\boldsymbol{w}^{\mathrm{T}} x_i + b) \geq 1, \quad i = 1, 2, \cdots, N \end{aligned}$$

这便是线性可分支持向量机（Support vector Machine，简称 SVM）的模型。下面讨论其对偶问题。

## 线性可分支持向量机对偶问题

- 设拉格朗日乘子 $\boldsymbol{\alpha} \geq \mathbf{0}$，则该问题的拉格朗日函数可写为

$$L(\boldsymbol{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{w}\|^2 + \sum_{i=1}^{N} \alpha_i (1 - y_i(\boldsymbol{w}^{\mathrm{T}} x_i + b))$$

- 令 $L(\boldsymbol{w}, b, \boldsymbol{\alpha})$ 对 $\boldsymbol{w}$ 和 $b$ 的偏导为零可得

$$\boldsymbol{w} = \sum_{i=1}^{N} \alpha_i y_i x_i, \qquad 0 = \sum_{i=1}^{N} \alpha_i y_i$$

- 回代可得**对偶问题**

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^{\mathrm{T}} x_j \\ s.t. \quad & \sum_{i=1}^{N} \alpha_i y_i = 0, \quad \alpha_i \geq 0, \quad i = 1, 2, \cdots, N \end{aligned}$$

## 线性可分支持向量机对偶问题

假设通过对偶问题求得 $\boldsymbol{\alpha}$，则最终模型：

$$f(\boldsymbol{x}) = \mathbf{sign}(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b) = \mathbf{sign}\left(\sum_{i=1}^{N} \alpha_i y_i x_i^{\mathrm{T}} x + b\right)$$

KKT 条件

$$\begin{cases} \alpha_i \geq 0 \\ y_i f(x_i) - 1 \geq 0 \\ \alpha_i(y_i f(x_i) - 1) = 0 \end{cases}$$

总有 $\alpha_i = 0$ 或 $y_i f(x_i) = 1$

解的稀疏性：训练完成后, 大部分的训练样本都不需保留，最终模型仅与支持向量有关，支持向量机因此而得名。

## 32.6.2 线性支持向量机

现实中的训练样本集，通常是线性不可分的，或者是近似线性可分。对于线性近似可分数据集，在第六章已经介绍，通过引入松弛变量的方式使其“可分”。因此得到如下优化问题：

$$
\begin{aligned}
\min_{\boldsymbol{w},b,\xi} \quad & \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\xi_i \\
s.t. \quad & y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \geq 1 - \xi_i, i = 1, 2, \ldots, N \\
& \xi_i \geq 0, i = 1, 2, \ldots, N
\end{aligned}
$$

# 线性支持向量机对偶问题

- 设拉格朗日乘子 $\boldsymbol{\alpha} \geq \mathbf{0}$，$\boldsymbol{\mu} \geq \mathbf{0}$，则该问题的拉格朗日函数可写为

$$L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\mu}) = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\xi_i + \sum_{i=1}^{N}\alpha_i(1 - \xi_i - y_i(\boldsymbol{w}^{\mathrm{T}}x_i + b)) - \sum_{i=1}^{N}(\mu_i\xi_i)$$

- 令 $L(\boldsymbol{w}, b, \boldsymbol{\alpha})$ 对 $\boldsymbol{w}$、$b$ 和 $\boldsymbol{\xi}$ 的偏导为零可得

$$\boldsymbol{w} = \sum_{i=1}^{N}\alpha_i y_i x_i, \quad 0 = \sum_{i=1}^{N}\alpha_i y_i, \quad \boldsymbol{\alpha} + \boldsymbol{\mu} = C\mathbf{1}$$

## 线性支持向量机对偶问题

- 回代可得**对偶问题**

$$
\begin{aligned}
\max_{\boldsymbol{\alpha}} \quad & \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^{\mathrm{T}} x_j \\
s.t. \quad & \sum_{i=1}^{N} \alpha_i y_i = 0 \\
& 0 \le \alpha_i \le C, \quad i = 1, 2, \cdots, N
\end{aligned}
$$

同样地，求解对偶问题，然后获得最终模型。

## 本讲小结

### 无约束优化的最优性条件

- 无约束可微优化问题：一阶条件、二阶条件
- 无约束不可微优化问题：凸优化的一阶条件
- 无约束不可微优化问题：复合优化的一阶条件

### 约束优化问题的最优性条件

- 拉格朗日对偶函数与拉格朗日对偶问题
- 弱对偶性、Slater 约束准则与强对偶原理
- KKT 最优性条件：原始约束、对偶约束、稳定性条件、互补松弛

这些最优性条件将用于指导我们设计优化问题的求解算法。