第 5 章：对偶

(1)：Lagrange 对偶

授课教师：曹语

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

从无约束到有约束

背景： 我们已经学习了无约束的凸优化问题，其中的重点是

- 优化问题的最优性条件
- 梯度法和牛顿法

然而工程实际问题往往具有约束：

- 结构设计中，重量、强度、位移都有允许范围；
- 通信与控制中，功率、带宽、稳定性都不能随意违反；
- 生产计划中，资源、需求、预算形成线性或非线性约束。

下一阶段的目标： 学习如何理解并求解带约束的优化问题。

回顾练习

逐点上确界： 如果对于任意 $\boldsymbol{y} \in \mathcal{A}$ ，函数 $g(\boldsymbol{x}, \boldsymbol{y})$ 关于 $\boldsymbol{x}$ 都是凸的，则函数 $f(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in \mathcal{A}} g(\boldsymbol{x}, \boldsymbol{y})$ 是关于 $\boldsymbol{x}$ 的凸函数。

练习： 判断下面的函数是否是凸/凹的？并说明为什么？

- $f(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in \mathcal{A}} \boldsymbol{x}^\top \boldsymbol{y} + h(\boldsymbol{y})$，其中 $h : \mathbb{R}^n \to \mathbb{R}$
- $f(\boldsymbol{x}) = \inf_{\boldsymbol{y} \in \mathcal{A}} \boldsymbol{x}^\top \boldsymbol{y} + h(\boldsymbol{y})$，其中 $h : \mathbb{R}^n \to \mathbb{R}$

回顾练习

逐点上确界： 如果对于任意 $\boldsymbol{y} \in \mathcal{A}$，函数 $g(\boldsymbol{x}, \boldsymbol{y})$ 关于 $\boldsymbol{x}$ 都是凸的，则函数 $f(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in \mathcal{A}} g(\boldsymbol{x}, \boldsymbol{y})$ 是关于 $\boldsymbol{x}$ 的凸函数。

练习： 判断下面的函数是否是凸/凹的？并说明为什么？

- $f(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in \mathcal{A}} \boldsymbol{x}^\top \boldsymbol{y} + h(\boldsymbol{y})$，其中 $h : \mathbb{R}^n \to \mathbb{R}$
- $f(\boldsymbol{x}) = \inf_{\boldsymbol{y} \in \mathcal{A}} \boldsymbol{x}^\top \boldsymbol{y} + h(\boldsymbol{y})$，其中 $h : \mathbb{R}^n \to \mathbb{R}$

答案： 第一项是仿射函数族的逐点上确界，因此是凸函数；第二项是仿射函数族的逐点下确界，因此是凹函数。

目录

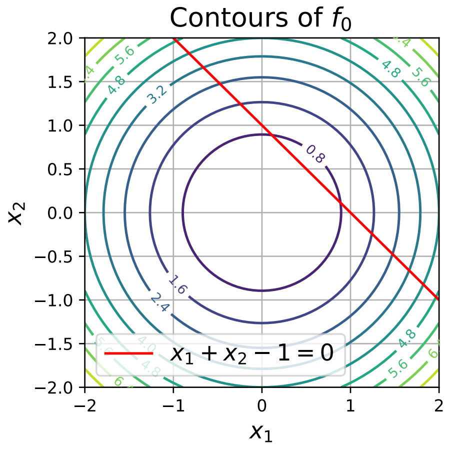
我们考虑优化如下问题:

$$
\begin{aligned}
\text{minimize} \quad & f_0(\boldsymbol{x}) \\
\text{subject to} \quad & h(\boldsymbol{x}) = 0 \\
& \boldsymbol{x} \in \mathbb{R}^2
\end{aligned}
$$

右侧的例子为:

$$f_0(x_1, x_2) = x_1^2 + x_2^2$$

$$h(x_1, x_2) = x_1 + x_2 - 1$$



Lagrange 乘子的几何理解

- 如果我们找到了最优点 $\boldsymbol{x}^*$，则在这点附近，沿任意可行方向 $\boldsymbol{v}$ 满足：

$$\nabla f_0(\boldsymbol{x}^*)^\top \boldsymbol{v} \geq 0, \qquad \boldsymbol{v}\text{ 可行}$$

对于等式约束，沿切线的正反两个方向通常都可行，因此

$$\nabla f_0(\boldsymbol{x}^*)^\top \boldsymbol{v} = 0, \qquad \boldsymbol{v}\text{ 可行}$$

- 对于等式约束，若 $\nabla h(\boldsymbol{x}^*) \neq \mathbf{0}$，可行方向的一阶近似是约束曲线的切线方向，因此

$$\nabla h(\boldsymbol{x}^*)^\top \boldsymbol{v} = 0$$

问题：目标函数等高线和约束曲线有什么关系？

因此我们可以看到：$\nabla f_0(\boldsymbol{x}^*)$ 平行于 $\nabla h(\boldsymbol{x}^*)$，即存在 ν 满足

$$\nabla f_0(\boldsymbol{x}^*) + \nu \nabla h(\boldsymbol{x}^*) = 0.$$

此处的 ν 被称为**Lagrange 乘子**

并且我们的问题变成了求解 $(\boldsymbol{x}^*, \nu)$ 来满足

$$\nabla f_0(\boldsymbol{x}^*) + \nu \nabla h(\boldsymbol{x}^*) = 0, \qquad h(\boldsymbol{x}^*) = 0$$

这其实就是知名的**Karush-Kuhn-Tucker（KKT）条件**对该特定问题的形式。

Joseph-Louis Lagrange (1736 - 1813，18 世纪知名科学家，图片来自 wikipedia）

内容概要

- 这个形式似乎建议我们可以考虑如下的函数

$$L(\boldsymbol{x}, \nu) = f_0(\boldsymbol{x}) + \nu h(\boldsymbol{x})$$

- 如果有多个函数 $h_i(\boldsymbol{x})$，$i = 1, 2, \cdots, p$，我们考虑

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

本节目标

本节课的目标：

- 对于一般的凸优化问题，拓展并建立 Lagrange 对偶函数
  （即上述 L 的更一般形式）
- 介绍 Lagrange 对偶问题
  （即关于 L，先优化 $\boldsymbol{x}$，再优化 $\boldsymbol{\nu}$；还是顺序倒过来）

目录

Lagrange 函数

考虑标准的优化问题：

$$
\begin{array}{ll}
\text{minimize} & f_0(\boldsymbol{x}) \\
\text{subject to} & f_i(\boldsymbol{x}) \leqslant 0, \quad i = 1, \cdots, m \\
& h_i(\boldsymbol{x}) = 0, \quad i = 1, \cdots, p \\
& \boldsymbol{x} \in \mathcal{D}
\end{array}
$$

其中 $\mathcal{D}$ 是变量的定义域，优化变量为 $\boldsymbol{x} \in \mathcal{D} \subseteq \mathbb{R}^n$。

定义该问题的**Lagrange 函数** $L : \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ 为

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

其中 λ_i, ν_i 被称为**Lagrange 乘子**；$\boldsymbol{\lambda}$, $\boldsymbol{\nu}$ 被称为**Lagrange 乘子向量**。

什么时候是凸优化问题

对于上页的标准形式，如果

- $\mathcal{D}$ 是凸集；
- $f_0, f_1, \cdots, f_m$ 都是凸函数；
- 等式约束 $h_i(\boldsymbol{x}) = 0$ 都是仿射约束，例如 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$；

则该问题称为**凸优化问题**。

- $f_i(\boldsymbol{x}) \leq 0$ 可理解为第 i 个工程限制：功率不能超标、库存不能为负等。
- 不等式约束对应的乘子要满足 $\lambda_i \geq 0$；等式约束的乘子 ν_i 可正可负。
- λ_i 可以理解为这个限制的**约束敏感度**：如果该限制稍微放宽，最优值大约会改善多少。ν_i 也具有类似理解。【具体含义在后面解释】

Lagrange 对偶函数

Lagrange 对偶函数（或对偶函数） $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R} \cup \{-\infty\}$ 为

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

- 这里的 $\inf_{\boldsymbol{x} \in \mathcal{D}}$ 是只在定义域内取下确界；
- 不等式和等式约束已经被搬进 L 中，不再单独要求 $\boldsymbol{x}$ 满足这些约束。

对偶函数为什么是凹的

定理

无论问题是否是凸的，该函数 g 一定是凹函数。

答案：固定 $\boldsymbol{x}$ 时，$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ 关于 $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ 是仿射函数；g 是这些仿射函数的逐点下确界，所以是凹函数。

计算对偶函数的三步

对一个具体问题，计算 $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ 通常按如下步骤：

1. 写出 $L = f_0 + \sum_i \lambda_i f_i + \sum_i \nu_i h_i$；
2. 把 $\boldsymbol{\lambda}, \boldsymbol{\nu}$ 暂时看成常数，对 $\boldsymbol{x}$ 求 $\inf L$；
3. 如果 $\inf L = -\infty$，这组乘子没有给出有效下界；否则得到一个关于乘子的函数 g。

例子 1：一维不等式约束

例子 1: 考虑最简单的一维问题

$$\begin{array}{ll} \text{minimize} & x^2 \\ \text{subject to} & 1 - x \leq 0. \end{array}$$

Lagrange 函数为

$$L(x, \lambda) = x^2 + \lambda(1 - x).$$

对 x 求下确界：

$$\frac{\partial L}{\partial x} = 2x - \lambda = 0 \quad \Rightarrow \quad x = \frac{\lambda}{2}, \qquad g(\lambda) = \lambda - \frac{\lambda^2}{4}.$$

例子 2：二次目标加等式约束

例子 2: 考虑优化问题

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

Lagrange 函数为

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

对 $\boldsymbol{x}$ 求最小值:

$$\nabla_x L = 2\boldsymbol{x} + \boldsymbol{A}^\top \boldsymbol{\nu} = 0 \quad \Rightarrow \quad \boldsymbol{x}^* = -\frac{1}{2}\boldsymbol{A}^\top \boldsymbol{\nu}, \qquad g(\boldsymbol{\nu}) = -\frac{1}{4}\boldsymbol{\nu}^\top \boldsymbol{A}\boldsymbol{A}^\top \boldsymbol{\nu} - \boldsymbol{b}^\top \boldsymbol{\nu}.$$

例子 3：线性规划

例子 3: 考虑线性规划问题

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

把 $\boldsymbol{x} \succeq \boldsymbol{0}$ 写成 $-\boldsymbol{x} \preceq \boldsymbol{0}$，则

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\nu}^\top (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) - \boldsymbol{\lambda}^\top \boldsymbol{x} = \left(\boldsymbol{c} + \boldsymbol{A}^\top \boldsymbol{\nu} - \boldsymbol{\lambda}\right)^\top \boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{\nu}.$$

若 $\boldsymbol{x}$ 前的系数不为 $\boldsymbol{0}$，则 $\inf_{\boldsymbol{x}} L = -\infty$；否则得到

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

目录

弱对偶性：为什么是下界

定理

对于任意的 $\boldsymbol{\lambda} \succeq 0$，以及任意的 $\boldsymbol{\nu}$

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

其中 p^* 是原问题的最优值。

证明思路：对任意原问题可行点 $\boldsymbol{x}$，

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

因此 $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\boldsymbol{z} \in \mathcal{D}} L(\boldsymbol{z}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f_0(\boldsymbol{x})$. 对所有可行 $\boldsymbol{x}$ 取下确界，即得到 $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*$。

对偶可行性

我们称满足 $\boldsymbol{\lambda} \succeq 0$，$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > -\infty$ 的 $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ 是**对偶可行**的。

此时 $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ 给出了原问题最优值 p^* 的一个下界。

Lagrange 对偶问题

由于

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \le p^*$$

因此，我们可以考虑最好的下界：

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

该问题被称为**Lagrange 对偶问题**。该问题的最优解 $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ 被称为**对偶最优解**或**最优 Lagrange 乘子**。对偶问题的最优值记为 d^*，因此

$$d^* \le p^*$$

该性质被称为**弱对偶性**。

例子 1：一维问题的对偶问题

例子 1： 原问题

$$\begin{array}{ll} \text{minimize} & x^2 \\ \text{subject to} & 1 - x \le 0 \end{array}$$

前面已经算得

$$g(\lambda) = \lambda - \frac{\lambda^2}{4}.$$

因此对偶问题是

$$\max_{\lambda \ge 0} \left(\lambda - \frac{\lambda^2}{4} \right).$$

令导数为零：

$$1 - \frac{\lambda}{2} = 0 \quad \Rightarrow \quad \lambda^\star = 2, \qquad d^\star = 1.$$

例子 1：为什么最优乘子像约束价格?

将约束稍微放宽为

$$1 - x \le u \qquad \Longleftrightarrow \qquad x \ge 1 - u.$$

当 $u \approx 0$ 时，新的最优解和最优值为

$$x^\star(u) = 1 - u, \qquad p^\star(u) = (1-u)^2 \approx 1 - 2u.$$

由于上一页得到 $\lambda^\star = 2$，该系数 $\lambda^\star$ 其实就是线性近似部分的系数，

$$p^\star(u) \approx p^\star(0) - \lambda^\star u.$$

所以 $\lambda^\star = 2$ 的含义是：约束放宽很小的 u 个单位时，最优值大约下降 $\lambda^\star u$。这就是把最优乘子理解为**边际价值**的来源。更严格的说明将在第二部分学完 KKT 条件后再回顾。

例子 2：等式约束二次问题的对偶

例子 2： 原问题

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

的对偶问题是

$$\text{maximize} \quad -\frac{1}{4}\boldsymbol{\nu}^\top \boldsymbol{A}\boldsymbol{A}^\top \boldsymbol{\nu} - \boldsymbol{b}^\top \boldsymbol{\nu}$$

练习： 假设 $\boldsymbol{A}\boldsymbol{A}^\top$ 可逆，最优乘子 $\boldsymbol{\nu}^\star$ 和对偶最优值 $d^\star$ 是什么？

例子 2：等式约束二次问题的对偶

例子 2： 原问题

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

的对偶问题是

$$\text{maximize} \quad -\frac{1}{4}\boldsymbol{\nu}^\top \boldsymbol{A}\boldsymbol{A}^\top \boldsymbol{\nu} - \boldsymbol{b}^\top \boldsymbol{\nu}$$

练习： 假设 $\boldsymbol{A}\boldsymbol{A}^\top$ 可逆，最优乘子 $\boldsymbol{\nu}^\star$ 和对偶最优值 $d^\star$ 是什么？

答案： 令 $\boldsymbol{M} = \boldsymbol{A}\boldsymbol{A}^\top$。由一阶条件 $-\frac{1}{2}\boldsymbol{M}\boldsymbol{\nu} - \boldsymbol{b} = \boldsymbol{0}$ 可得

$$\boldsymbol{\nu}^\star = -2\boldsymbol{M}^{-1}\boldsymbol{b}, \qquad d^\star = \boldsymbol{b}^\top \boldsymbol{M}^{-1}\boldsymbol{b}.$$

例子 3：线性规划的对偶问题

例子 3：原问题为

$$\min \ \boldsymbol{c}^\top \boldsymbol{x} \qquad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \quad \boldsymbol{x} \succeq \boldsymbol{0}.$$

前面已经算得

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

因此对偶问题可写成

$$\begin{array}{ll} \text{maximize} & -\boldsymbol{b}^\top \boldsymbol{\nu} \\ \text{subject to} & \boldsymbol{A}^\top \boldsymbol{\nu} - \boldsymbol{\lambda} + \boldsymbol{c} = \boldsymbol{0} \\ & \boldsymbol{\lambda} \succeq \boldsymbol{0}. \end{array}$$

例子 3：消去 $\boldsymbol{\lambda}$

由等式约束

$$\boldsymbol{A}^\top \boldsymbol{\nu} - \boldsymbol{\lambda} + \boldsymbol{c} = \mathbf{0} \qquad \Rightarrow \qquad \boldsymbol{\lambda} = \boldsymbol{A}^\top \boldsymbol{\nu} + \boldsymbol{c}.$$

再结合 $\boldsymbol{\lambda} \succeq \mathbf{0}$，得到

$$\boldsymbol{A}^\top \boldsymbol{\nu} + \boldsymbol{c} \succeq \mathbf{0}.$$

所以线性规划的对偶问题等价于

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

课堂快速检查

练习： 考虑一维优化问题

$$\begin{array}{ll}\text{minimize} & x^2 \\ \text{subject to} & x - 1 \leq 0.\end{array}$$

1. 写出 Lagrange 函数 $L(x, \lambda)$；
2. 计算对偶函数 $g(\lambda)$；
3. 求最优乘子 $\lambda^\star$，并从边际价值角度解释为什么它是 0。

课堂快速检查

练习： 考虑一维优化问题

$$\begin{array}{ll} \text{minimize} & x^2 \\ \text{subject to} & x - 1 \leq 0. \end{array}$$

1. 写出 Lagrange 函数 $L(x, \lambda)$；
2. 计算对偶函数 $g(\lambda)$；
3. 求最优乘子 $\lambda^\star$，并从边际价值角度解释为什么它是 0。

答案： $L(x, \lambda) = x^2 + \lambda(x - 1)$；$g(\lambda) = -\lambda - \lambda^2/4$；对偶问题为 $\max_{\lambda \geq 0} -\lambda - \lambda^2/4$，最优解 $\lambda^\star = 0$。原问题最优解是 $x^\star = 0$，约束 $x^\star - 1 < 0$ 不紧，因此它没有边际价值。

对偶间隙

差值 $p^* - d^*$ 为**最优对偶间隙**；如果

$$d^* = p^*,$$

则我们称之为**强对偶性**。

将在本章第 2 部分学习何时具有强对偶性，以及介绍 KKT 条件。

目录

总结

- 知道 Lagrange 函数和 Lagrange 对偶函数的定义，并能解释对偶函数为什么是凹函数
- 理解弱对偶性：任意对偶可行点都给出原问题最优值的下界
- 知道 Lagrange 对偶问题的含义，并知道相关的术语名词

阅读作业 & 参考资料：

- 课本第 3.3，5.1，5.2 章