

第 5 章：对偶

(1): Lagrange 对偶

曹语

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

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

然而实际问题往往具有约束

下一阶段的目标：学习如何求解带约束的优化问题，及其相关经典的算法

回顾练习

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

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

目录

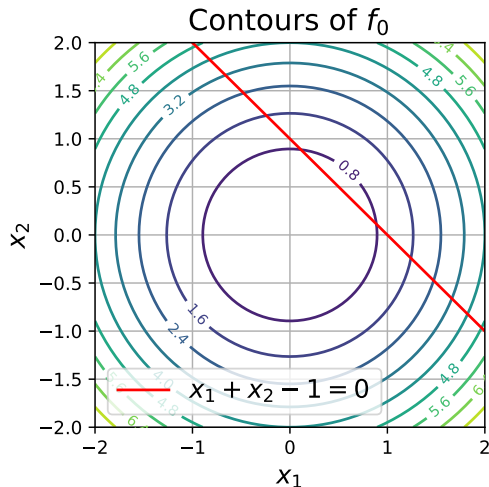
1. 前置背景
2. 共轭函数
3. Lagrange 对偶函数
4. Lagrange 对偶问题
5. 总结

我们考虑优化如下问题：

$$\begin{array}{ll}\text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & h(\mathbf{x}) = 0 \\ & \mathbf{x} \in \mathbb{R}^2\end{array}$$

右侧的例子为：

$$\begin{aligned}f_0(x_1, x_2) &= x_1^2 + x_2^2 \\ h(x_1, x_2) &= x_1 + x_2 - 1\end{aligned}$$



Lagrange 乘子的几何理解

- ▶ 如果我们找到了最优点 $\mathbf{x}^*$ ，则在这点附近，我们需要满足

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

如果反方向 $-\mathbf{v}$ 也可行，我们得到

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

- ▶ 又因为可行方向 $\mathbf{v}$ 具有一定限制，因此

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

问题：在几何意义中，这个意味着什么？

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

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

此处的 ν 被称为 **Lagrange 乘子**

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

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

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

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

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

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

本节课的目标:

- ▶ 先讨论共轭函数
(若 $p = n$, $h_i(\mathbf{x}) = -\mathbf{x}_i$, 并且最小化 $\mathbf{x}$, 我们得到共轭函数。)
- ▶ 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 $\mathbf{x}$, 再优化 $\boldsymbol{\nu}$; 还是顺序倒过来)

目录

1. 前置背景
2. 共轭函数
3. Lagrange 对偶函数
4. Lagrange 对偶问题
5. 总结

共轭函数

定义 (共轭函数)

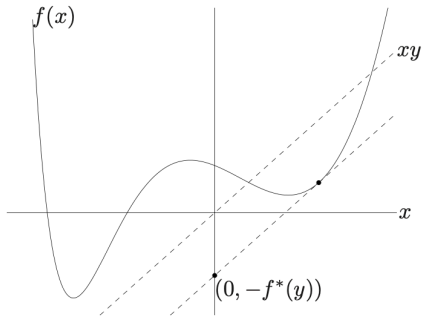
设函数 $f: \Omega \rightarrow \mathbb{R}$ ，定义函数 $f^*: \Omega^* \rightarrow \mathbb{R}$ 为

$$f^*(\mathbf{y}) = \sup_{\mathbf{x} \in \Omega} (\mathbf{y}^\top \mathbf{x} - f(\mathbf{x}))$$

f^* 被称为函数 f 的**共轭函数**； Ω^* 是指所有能满足右端非 ∞ 的 $\mathbf{y}$

定理 (共轭函数是凸函数)

对于任意函数 f ， f^* 为凸函数。



例子

例子 1: 若 $f(x) = ax + b$,

则 $\Omega^* = \{a\}$, 且 $f^*(a) = -b$

例子 2: 若 $f(x) = -\log x$, $\Omega = (0, \infty)$,

则 $\Omega^* = (-\infty, 0)$, $f^*(y) = -\log(-y) - 1$

例子 3: 若 $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x}$, $\mathbf{Q} \in \mathbf{S}_{++}^n$,

则 $\Omega^* = \mathbb{R}^n$, $f^*(\mathbf{y}) = \frac{1}{2}\mathbf{y}^\top \mathbf{Q}^{-1}\mathbf{y}$

练习: 若 $f(x) = e^x$, $\Omega = \mathbb{R}$,

请写出 $\Omega^* = \underline{\hspace{2cm}}$, $f^*(y) = \underline{\hspace{2cm}}$

(1) Fenchel 不等式

Fenchel 不等式： 对于任意的 $\mathbf{x} \in \Omega$, $\mathbf{y} \in \Omega^*$,

$$f(\mathbf{x}) + f^*(\mathbf{y}) \geq \mathbf{x}^\top \mathbf{y}$$

例子： 若 $f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^\top \mathbf{Q} \mathbf{x}$, 其中 $\mathbf{Q} \in \mathbf{S}_{++}^n$,

$$2\mathbf{x}^\top \mathbf{y} \leq \mathbf{x}^\top \mathbf{Q} \mathbf{x} + \mathbf{y}^\top \mathbf{Q}^{-1} \mathbf{y}$$

(2) 复合仿射变换

复合仿射变换

- 若 $a > 0, b \in \mathbb{R}$, 则 $g(x) = a f(x) + b$ 的共轭函数为

$$g^*(y) = a f^*(y/a) - b$$

- 若 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 非奇异, $\mathbf{b} \in \mathbb{R}^n$, 则 $g(\mathbf{x}) = f(\mathbf{Ax} + \mathbf{b})$ 的共轭函数为

$$g^*(\mathbf{y}) = f^*(\mathbf{A}^{-\top} \mathbf{y}) - \mathbf{b}^\top \mathbf{A}^{-\top} \mathbf{y}$$

(3) Legendre 变换

Legendre 变换/可微函数的共轭:

若 f 可微, 则 f^* 被称为 f 的 Legendre 变换

Adrien-Marie Legendre (1752 - 1833), 知名法国数学家。

Legendre 变换的物理应用（不要求）

拉格朗日力学

$$\mathcal{L} = \mathcal{T} - \mathcal{V}$$

其中 $\mathcal{T}$ 是动能， $\mathcal{V}$ 是势能。

对于一条轨迹 $\gamma = \{\mathbf{x}(t)\}_{t \in [t_1, t_2]}$ ，我们可以考虑如下

$$\mathcal{L}(\mathbf{x}, \dot{\mathbf{x}}) = \frac{1}{2}m\|\dot{\mathbf{x}}\|^2 - V(\mathbf{x}) \quad (1)$$

整条路径的作用量（action）被定义为

$$\mathcal{A}(\mathbf{x}) = \int_{t_1}^{t_2} \mathcal{L}(\mathbf{x}(t), \dot{\mathbf{x}}(t)) \, dt$$

经典物理中，物体的行为总是最小化作用量，我们可以得到物体的轨迹为（即 Euler-Lagrange 方程）

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}}(\mathbf{x}, \dot{\mathbf{x}}) = \frac{\partial \mathcal{L}}{\partial \mathbf{x}}(\mathbf{x}, \dot{\mathbf{x}})$$

代入公式(1)可知

$$m\ddot{\mathbf{x}} = -\nabla V(\mathbf{x}),$$

即牛顿力学。

对于公式(1)，固定 $\mathbf{x}$ ，我们可以计算其共轭函数：

$$\mathcal{L}^*(\mathbf{x}, \mathbf{p}) := \sup_{\mathbf{v}} \mathbf{p}^\top \mathbf{v} - \left(\frac{1}{2} m \|\mathbf{v}\|^2 - V(\mathbf{x}) \right) = \frac{\|\mathbf{p}\|^2}{2m} + V(\mathbf{x})$$

该共轭函数即为哈密顿量，是 Hamiltonian mechanics 的基础。

目录

1. 前置背景
2. 共轭函数
3. Lagrange 对偶函数
4. Lagrange 对偶问题
5. 总结

Lagrange 函数

考虑标准的优化问题：

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

其中优化变量为 $\mathbf{x} \in \mathbb{R}^n$

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

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

其中

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

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

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

定理

无论问题是否是凸的，该函数 g 一定是凹函数。

例子 1: 考虑优化问题

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

它对应的 $g(\boldsymbol{\nu}) = -\frac{1}{4}\boldsymbol{\nu}^\top \mathbf{AA}^\top \boldsymbol{\nu} - \mathbf{b}^\top \boldsymbol{\nu}$

例子 2: 考虑线性规划问题

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

它对应的

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

目录

1. 前置背景
2. 共轭函数
3. Lagrange 对偶函数
4. Lagrange 对偶问题
5. 总结

定理

对于任意的 $\lambda \succeq 0$ ，以及任意的 ν

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

其中 p^* 是原问题的最优值。

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

由于

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

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

$$d^* \leq p^*$$

该性质被称为**弱对偶性**。

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。

例子 1: 原问题

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

的对偶问题是

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

例子 2: 原问题

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

的对偶问题是

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

它又等价于

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

目录

1. 前置背景
2. 共轭函数
3. Lagrange 对偶函数
4. Lagrange 对偶问题
5. 总结

总结

- ▶ 掌握共轭函数的定义，共轭函数是凸函数，对于简单例子能计算
- ▶ 知道 Fenchel 不等式；知道 Legendre 变换的含义
- ▶ 知道 Lagrange 函数，和 Lagrange 对偶函数的定义，知道 Lagrange 对偶函数是凹函数，以及 Lagrange 对偶函数是最优值的一个下界
- ▶ 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