第四章：凸优化进阶之对偶理论

——优化论五部曲

Jason 博士

网易微专业 x 稀牛学院

人工智能数学基础微专业

凸优化问题

对偶

问题案例

凸优化问题标准形式

- 凸优化问题

$$
\begin{aligned}
\text{minimize} \quad & f_0(\mathbf{x}) \\
\text{subject to} \quad & f_i(\mathbf{x}) \le 0 \text{ for } i = 1, 2, \cdots, m \\
& h_i(\mathbf{x}) = 0 \text{ for } i = 1, 2, \cdots, p
\end{aligned} \tag{1}
$$

- 则有 $f_0(\mathbf{x})$ 是凸函数，可行域是凸集
  - 目标函数是凸的
  - 不等式约束函数必须是凸的
  - 等式约束函数必须是仿射的
- 在一个凸集上极小化一个凸的目标函数
- 最优值 (目标函数在可行域上的最小值)：

$p* = \min\{f_0(\mathbf{x}) : f_i(\mathbf{x}) \le 0, h_i(\mathbf{x}) = 0\}$
  - $p* = +\infty$ 不可行（可行域是空集）
  - $p* = -\infty$，unbounded below （存在可行点使得 $f_0(x) \to -\infty$）
  - $f_0(\mathbf{x}^*) = p*$

无约束凸优化问题

- 自变量为矢量的函数 $f: \mathbb{R}^n \to \mathbb{R}$

$$\min\ f(\mathbf{x}) \quad \mathbf{x} \in \mathbb{R}^n$$

- 全局最优解常规解法

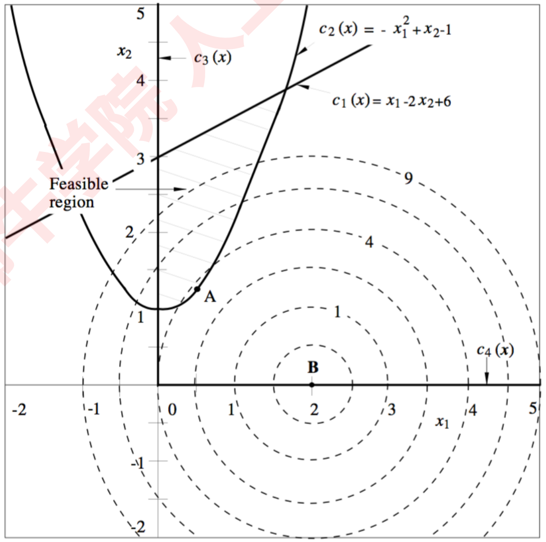
- 直接法
    - 梯度等于 0，求得驻点

- 迭代法
    - 梯度下降法
    - 牛顿法
    - 拟牛顿法

有约束凸优化问题

- 考虑以下约束优化问题

$$
\begin{aligned}
\text{minimize} \quad & f(\mathbf{x}) = x_1^2 + x_2^2 - 4x_1 + 4 = (x_1 - 2)^2 + x_2^2 \\
\text{subject to} \quad & c_1(\mathbf{x}) = x_1 - 2x_2 + 6 \geq 0 \\
& c_2(\mathbf{x}) = -x_1^2 + x_2 - 1 \geq 0,\ c_3(\mathbf{x}) = x_1 \geq 0,\ c_4(\mathbf{x}) = x_2 \geq 0
\end{aligned}
$$



凸优化问题

对偶

问题案例

一般优化问题

$$
\begin{aligned}
\text{minimize} \quad & f_0(\mathbf{x}) \\
\text{subject to} \quad & f_i(\mathbf{x}) \leq 0 \text{ for } i = 1, 2, \cdots, m \\
& h_i(\mathbf{x}) = 0 \text{ for } i = 1, 2, \cdots, p
\end{aligned} \tag{2}
$$

问题定义域 $\mathcal{D} = (\cap_{i=0}^{m} \text{dom } f_i) \cap (\cap_{i=0}^{p} \text{dom } h_i)$，注意与可行域区分.

Lagrangian

Lagrangian(合并目标函数和约束)

- Lagrangian

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

- 主变量：$\mathbf{x}$
- 对偶变量：$\boldsymbol{\lambda} \geq 0, \boldsymbol{v}$
- 意义解释

主问题

- Lagrangian

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

- 主问题

$$p^* = \min_{\mathbf{x}} \left(\max_{\boldsymbol{\lambda}, \boldsymbol{v}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{v}) \right)$$

- 分开看

$$\max_{\boldsymbol{\lambda}, \boldsymbol{v}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{v}) = f_0(\mathbf{x}) + \max_{\boldsymbol{\lambda}, \boldsymbol{v}} \left(\sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} v_i h_i(\mathbf{x}) \right)$$

对偶

Lagrrange 对偶函数

- Lagrangian

$$g(\boldsymbol{\lambda}, \boldsymbol{v}) = \min_{\mathbf{x}\in\mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{v}) = \min_{\mathbf{x}\in\mathcal{D}} \left\{ f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} v_i h_i(\mathbf{x}) \right\}$$

- 回忆：逐点最大：$f_1, \cdots, f_m$ 凸，则 $f(\mathbf{x}) = \max\{f_1(\mathbf{x}), \cdots, f_m(\mathbf{x})\}$ 凸. $f(\mathbf{x}, \mathbf{y})$ 对于每个 $\mathbf{y} \in \mathcal{A}$ 凸，则 $\max_{\mathbf{y}\in\mathcal{A}} f(\mathbf{x}, \mathbf{y})$ 凸. g 是 $(\boldsymbol{\lambda}, \boldsymbol{v})$ 的仿射函数的逐点下确界 (min)，总是凹的，即使原问题非凸.

对偶

Lagrrange 对偶函数

- Lagrangian

$$g(\boldsymbol{\lambda}, \boldsymbol{v}) = \min_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{v}) = \min_{\mathbf{x} \in \mathcal{D}} \left\{ f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} v_i h_i(\mathbf{x}) \right\}$$

- 如果 $\tilde{\mathbf{x}}$ 是一个可行点，则

$$g(\boldsymbol{\lambda}, \boldsymbol{v}) = \min_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{v}) \leq L(\tilde{\mathbf{x}}, \boldsymbol{\lambda}, \boldsymbol{v})$$

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

- 因此 $g(\boldsymbol{\lambda}, \boldsymbol{v}) \leq f_0(\mathbf{x}^*) = p^*$

Lagrange 对偶问题

Lagrange 对偶问题

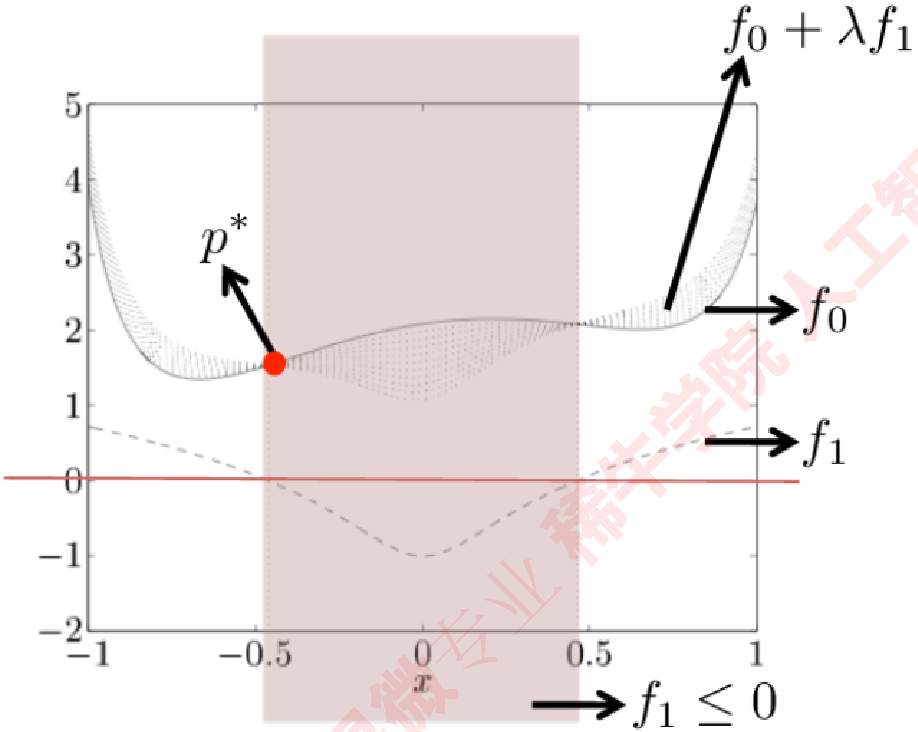
- Lagrange 对偶问题

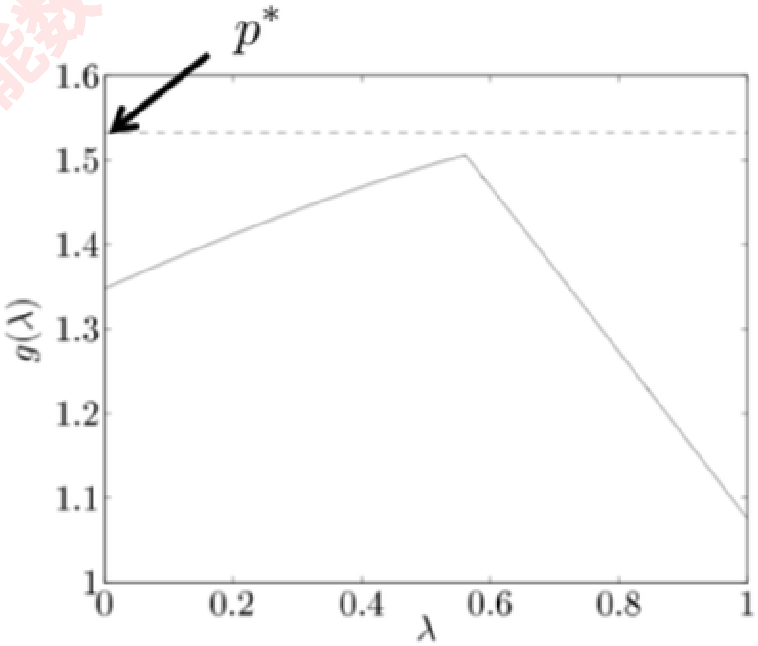
$$\begin{aligned} \text{maximize} \quad & g(\boldsymbol{\lambda}, \boldsymbol{v}) \\ \text{subject to} \quad & \boldsymbol{\lambda} \geq \mathbf{0} \end{aligned}$$

- 目标函数：$\max_{\boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{v}} \min_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{v})$
- 凹函数在凸集上最大化，这是一凸优化问题，设最优值为 d^*, 对应的极值点是 $\boldsymbol{\lambda}^*, \boldsymbol{v}^*$

对偶几何解释

- $L(\mathbf{x}, \boldsymbol{\lambda}) = f_0(\mathbf{x}) + \lambda f_1(\mathbf{x})$





强弱对偶解释

- 弱对偶： $d^* \le p^*$，无论凸或非凸问题总成立
- 强对偶： $d^* = p^*$
  - 该条件通常不成立
  - 对于凸优化问题通常 (usually) 成立
  - 凸优化问题可以写为

$$\begin{array}{rl} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \le 0 \text{ for } i = 1, 2, \cdots, m \\ & \mathbf{Ax}=\mathbf{b} \end{array}$$

  - 满足 Slater 条件 (constraint qualifications)：存在内点 $\mathbf{x}$，使得 $f_i(\mathbf{x}) < 0$ for $i = 1, 2, \cdots, m$ 均成立

从对偶问题解主问题

- 假定强对偶成立，$(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{v}^*)$ 是主对最优解，有

$$
\begin{aligned}
p^* = f_0(\mathbf{x}^*) = d^* &= g(\boldsymbol{\lambda}^*, \boldsymbol{v}^*) \\
&= \min_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}) + \sum_{i=1}^{p} v_i^* h_i(\mathbf{x}) \right) \\
&\le f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) + \sum_{i=1}^{p} v_i^* h_i(\mathbf{x}^*) \\
&\le f_0(\mathbf{x}^*)
\end{aligned}
$$

- 结论 1：$\lambda_i^* f_i(\mathbf{x}^*) = 0$
- 结论 2：$L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{v}^*)$ 关于 $\mathbf{x}^*$ 处取极小值，有 $\nabla_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{v}^*) = \mathbf{0}$
  - $g(\boldsymbol{\lambda}^*, \boldsymbol{v}^*) = L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{v}^*)$
  - $g(\boldsymbol{\lambda}^*, \boldsymbol{v}^*) = \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{v}^*)$

KKT 条件

- 凸优化问题

$$
\begin{aligned}
\text{minimize} \quad & f_0(\mathbf{x}) \\
\text{subject to} \quad & f_i(\mathbf{x}) \le 0 \text{ for } i = 1, 2, \cdots, m \\
& h_i(\mathbf{x}) = 0 \text{ for } i = 1, 2, \cdots, p
\end{aligned}
$$

- Lagrangian

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

凸优化问题强对偶成立，充要条件

- $f_i(\mathbf{x}^*) \le 0,\ i = 1, \cdots, m\ \left(\nabla_{\boldsymbol{\lambda}} L(\mathbf{x}^*, \boldsymbol{\lambda}, \boldsymbol{v}) \le \mathbf{0}\right)$
- $h_i(\mathbf{x}^*) = 0,\ i = 1, \cdots, p\ \left(\nabla_{\mathbf{v}} L(\mathbf{x}^*, \boldsymbol{\lambda}, \boldsymbol{v}) = \mathbf{0}\right)$
- $\lambda_i^* \ge 0,\ i = 1, \cdots, m$
- $\lambda_i^* f_i(\mathbf{x}^*) = 0,\ i = 1, \cdots, m$
- $\nabla f_0(\mathbf{x}^*) + \sum_i \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_i v_i^* \nabla h_i(\mathbf{x}^*) = 0\ \left(\nabla_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{v}^*) = \mathbf{0}\right)$

KKT 条件

- 非凸问题弱对偶，KKT 局部极小解的一阶必要条件
- 求解 KKT
- 在某些情况下可以得到闭合解
- 求解凸优化问题方法可以认为求解 KKT

主对问题思考

- 主问题

$$p^* = \min_{\mathbf{x}} \left(\max_{\boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{v}} L\left(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{v}\right) \right)$$

- 对偶问题

$$d^* = \max_{\boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{v}} \left(\min_{\mathbf{x}} L\left(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{v}\right) \right)$$

主对关系（强对偶成立 max min 可以交换）

$$\max_{\boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{v}} \left(\min_{\mathbf{x}} L\left(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{v}\right) \right) \leq \min_{\mathbf{x}} \left(\max_{\boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{v}} L\left(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{v}\right) \right)$$

凸优化问题

对偶

问题案例

SVM 前奏

最小 2 范数解问题

- 求解下列问题

$$p^* = \min_{\mathbf{x}} \ ||\mathbf{x}||_2^2$$
$$s.t. \ \mathbf{A}\mathbf{x} = \mathbf{b}$$

- Lagrangian

$$L(\mathbf{x}, \mathbf{v}) = \mathbf{x}^T\mathbf{x} + \mathbf{v}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

- 对偶函数

$$g(\mathbf{v}) = \min_{\mathbf{x}} \ L(\mathbf{x}, \mathbf{v})$$

- $\nabla_{\mathbf{x}} L(\mathbf{x}, \mathbf{v}) = \mathbf{0} \Rightarrow 2\mathbf{x} + \mathbf{A}^T\mathbf{v} = \mathbf{0} \Rightarrow \mathbf{x}^*(\mathbf{v}) = -\frac{1}{2}\mathbf{A}^T\mathbf{v}$
- $g(\mathbf{v}) = L(\mathbf{x}^*(\mathbf{v}), \mathbf{v}) = -\frac{1}{4}\mathbf{v}^T(\mathbf{A}\mathbf{A}^T)\mathbf{v} - \mathbf{v}^T\mathbf{b}$

SVM 前奏

最小 2 范数解问题

- 对偶问题

$$d^* = \max_{\mathbf{v}} -\frac{1}{4}\mathbf{v}^T\left(\mathbf{A}\mathbf{A}^T\right)\mathbf{v} - \mathbf{v}^T\mathbf{b}$$

- 可求得

$$\mathbf{v}* = -2\left(\mathbf{A}\mathbf{A}^T\right)^{-1}\mathbf{b}$$

$$d^* = \mathbf{b}^T\left(\mathbf{A}\mathbf{A}^T\right)^{-1}\mathbf{b}$$

- 由于 $p^* = d^*$，则

$$\mathbf{x}^* = \mathbf{x}^*\left(\mathbf{v}^*\right) = \mathbf{A}^T\left(\mathbf{A}\mathbf{A}^T\right)^{-1}\mathbf{b}$$

- Lower bound

$$p^* \geq -\frac{1}{4}\mathbf{v}^T\left(\mathbf{A}\mathbf{A}^T\right)\mathbf{v} - \mathbf{v}^T\mathbf{b} \quad \text{for all } \mathbf{v}$$

SVM 前奏

LP

- 求解下列问题

$$p^* = \min_{\mathbf{x}} \ \mathbf{c}^T \mathbf{x}$$
$$s.t. \ \mathbf{A}\mathbf{x} \leq \mathbf{b}$$

- Lagrangian

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

- 对偶函数

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

SVM 前奏

LP

- 可以发现

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

- 对偶问题

$$d^* = \max_{\boldsymbol{\lambda}} -\mathbf{b}^T\boldsymbol{\lambda}$$
$$s.t.\ \mathbf{A}^T\boldsymbol{\lambda} + \mathbf{c} = \mathbf{0}$$
$$\boldsymbol{\lambda} \geq \mathbf{0}$$

- 对偶的对偶

本章总结

本章参考资料

- Boyd, Stephen, and Lieven Vandenberghe. Convex optimization. Cambridge university press, 2004.
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.