凸优化笔记: 对偶理论

Apple Zhang

2023 年 3 月 30 日

1 基本概念

定义一个优化问题 (不一定是凸问题) 为:

$$\begin{aligned} p^* = \min_{\mathbf{x}} \ & f_0(\mathbf{x}) \\ \text{s.t.} \ & f_i(\mathbf{x}) \le 0, \quad i = 1, 2, \cdots, m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, 2, \cdots, n. \end{aligned} \tag{1}$$

其中，p^* 表示该问题的最优值. 定义该问题的**定义域**为

$$\mathcal{D} = \left(\bigcap_{i=0}^{m} \mathbf{dom}\, f_i \right) \cap \left(\bigcap_{i=1}^{n} \mathbf{dom}\, h_i \right). \tag{2}$$

可行域 为满足约束条件的 $\mathbf{x}$ 的集合.

$$\mathcal{F} = \{ \mathbf{x} \in \mathcal{D} \mid f_i(\mathbf{x}) \le 0,\ i = 1, 2, \cdots, m,\ h_i(\mathbf{x}) = 0,\ i = 1, 2, \cdots, n \}. \tag{3}$$

特别地，当 $f_0, f_1, \cdots, f_m$ 都是凸函数 (convex function)、h_i 均为仿射函数时 (affine function)，则 (1) 是一个凸问题.

定义 1. *优化问题 (1) 的拉格朗日函数定义为:*

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{n} \nu_i h_i(\mathbf{x}). \tag{4}$$

定义 2. *定义拉格朗日函数 $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ 的对偶函数为:*

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \min_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}). \tag{5}$$

对于对偶函数，其具有以下性质.

性质 1. *对于任何优化问题，其对偶问题 $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ 都是凹函数.*

证明. 对任意 $\mathbf{x} \in \mathcal{D}$， $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ 关于 $\boldsymbol{\lambda}$ 和 $\boldsymbol{\nu}$ 是仿射函数，所以对偶函数 $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ 关于 $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ 的逐点下确界是凹函数. $\square$

性质 2. 对任意 $\lambda_i \geq 0\ (i=1,2,\cdots,m)$ 和 $\boldsymbol{\nu} \in \mathbb{R}^n$，对偶函数是原问题最优值的下界，即

$$g(\boldsymbol{\lambda},\boldsymbol{\nu}) \leq p^*. \tag{6}$$

证明. 对原问题 (1) 的最优解 $\mathbf{x}^*$，有 $f_i(\mathbf{x}^*) \leq 0$ 和 $h_i(\mathbf{x}^*) = 0$. 因此

$$\sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}^*) + \sum_{i=1}^{n} h_i(\mathbf{x}^*) \leq 0. \tag{7}$$

从而拉格朗日函数满足:

$$L(\mathbf{x}^*,\boldsymbol{\lambda},\boldsymbol{\nu}) = f_0(\mathbf{x}^*) + \sum_{i=1}^{n} \lambda_i f_i(\mathbf{x}^*) + \sum_{i=1}^{n} h_i(\mathbf{x}^*) \leq f_0(\mathbf{x}^*). \tag{8}$$

因此

$$g(\boldsymbol{\lambda},\boldsymbol{\nu}) = \min_{\mathbf{x}\in\mathcal{D}} L(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\nu}) \leq L(\mathbf{x}^*,\boldsymbol{\lambda},\boldsymbol{\nu}) \leq f_0(\mathbf{x}^*) = p^*. \tag{9}$$

□

根据对偶函数，定义原问题的**Lagrange 对偶问题**为

定义 3. 问题 (1) 的 *Lagrange* 对偶问题（或简称对偶问题）定义为

$$d^* = \max_{\boldsymbol{\lambda},\boldsymbol{\nu}} g(\boldsymbol{\lambda},\boldsymbol{\nu}), \quad \text{s.t. } \boldsymbol{\lambda} \geq \mathbf{0}. \tag{10}$$

其中，d^* 表示对偶问题的最优值.

根据性质 2，可知 $d^* \leq p^*$，该性质称为**弱对偶性**.

2 强对偶性

如果等式

$$d^* = p^*, \tag{11}$$

成立，则称强对偶性成立. 一般情况下，强对偶性不一定成立. Slater 条件给出了强对偶性的一个**充分不必要条件**.

定理 1. 如果原问题是**凸问题**，若存在一点 $\mathbf{x} \in \mathbf{relint}\,\mathcal{D}$ 使得

$$f_i(\mathbf{x}) < 0\ (i=1,2,\cdots,m), \quad h_i(\mathbf{x}) = 0\ (i=1,2,\cdots,n). \tag{12}$$

以上条件称为 *Slater* 条件，若原问题是凸问题且满足 *Slater* 条件，则强对偶性成立. 若 $f_i(\mathbf{x})$ 都是仿射函数，则 *Slater* 条件可以改为：存在 $\mathbf{x} \in \mathbf{relint}\,\mathcal{D}$ 使得

$$f_i(\mathbf{x}) \leq 0\ (i=1,2,\cdots,m), \quad h_i(\mathbf{x}) = 0\ (i=1,2,\cdots,n). \tag{13}$$

并非所有凸问题都满足 Slater 条件，但满足该条件的凸问题占绝大多数. 因此一般认为我们能见到的凸问题都满足强对偶性. 另外 Slater 条件是强对偶性的充分不必要条件，因此不满足该条件的优化问题也可能有强对偶性.

3 鞍点解释

3.1 鞍点

对于函数 $f(w,z)$，其中 $w \in W$、$z \in Z$，其总是满足极大极小不等式，即:

$$\max_{z\in Z}\min_{w\in W} f(w,z) \le \min_{w\in W}\max_{z\in Z} f(w,z). \tag{14}$$

事实上，

$$\min_{w\in W} f(w,z) \le f(w,z) \le \max_{z\in Z} f(w,z) \tag{15}$$

则对任意 $w \in W$ 和 $z \in Z$ 有

$$\min_{w\in W} f(w,z) \le \max_{z\in Z} f(w,z) \tag{16}$$

因此无论 w, z 如何取值，左侧的最大值总是小于右侧的最小值，即

$$\max_{z\in Z}\min_{w\in W} f(w,z) \le \min_{w\in W}\max_{z\in Z} f(w,z).$$

如果 (14) 中的等号可以取到，即

$$\max_{z\in Z}\min_{w\in W} f(w,z) = \min_{w\in W}\max_{z\in Z} f(w,z) \tag{17}$$

则称 $f(w,z)$ 具有鞍点性质，(w,z) 是函数 f 的鞍点.

定义 4. *对于函数 $f(w,z)$，若存在 $\tilde{w} \in W$ 以及 $\tilde{z} \in Z$ 满足*

$$f(\tilde{w},z) \le f(\tilde{w},\tilde{z}) \le f(w,\tilde{z}), \tag{18}$$

则称 $(\tilde{w},\tilde{z})$ 为鞍点.

不等式 (18) 表明 $f(w,\tilde{z})$ 在 $\tilde{w}$ 取得最小值，$f(\tilde{w},z)$ 在 $\tilde{z}$ 取得最大值:

$$\min_{w\in W} f(w,\tilde{z}) = f(\tilde{w},\tilde{z}) = \max_{z\in Z} f(\tilde{w},z). \tag{19}$$

3.2 强对偶性的鞍点解释

为了方便讨论，本节假设原问题中不存在等式约束，即

$$\begin{aligned} p^* = \min_{\mathbf{x}}\ & f_0(\mathbf{x}) \\ \text{s.t.}\ & f_i(\mathbf{x}) \le 0, \quad i = 1,2,\cdots,m \end{aligned} \tag{20}$$

定理 2. *原问题 (20) 的最优值 p^* 等于:*

$$\min_{\mathbf{x}\in\mathcal{D}}\max_{\boldsymbol{\lambda}\ge\mathbf{0}} L(\mathbf{x},\boldsymbol{\lambda}). \tag{21}$$

证明. 已知对偶问题为

$$L(\mathbf{x},\boldsymbol{\lambda}) = f_0(\mathbf{x}) + \sum_{i=1}^{n}\lambda_i f_i(\mathbf{x}). \tag{22}$$

根据 $f_i(\mathbf{x})$ 的取值不同可作以下讨论:

- 若 $\forall i=1,2,\cdots,m$ 均有 $f_i(\mathbf{x})\le 0$，则 $L(\mathbf{x},\boldsymbol{\lambda})$ 关于每一个 λ_i 都为单调递减. 由此则说明 $\max_{\boldsymbol{\lambda}\ge\mathbf{0}} L(\mathbf{x},\boldsymbol{\lambda})=L(\mathbf{x},\mathbf{0})=f_0(\mathbf{x})$.
- 若 $\exists i_0$ 使得 $f_{i_0}(\mathbf{x})>0$，则 $L(\mathbf{x},\boldsymbol{\lambda})$ 关于 λ_{i_0} 单调递增，因此取 $\lambda_{i_0}=+\infty$ 可以取最大值. 即 $\max_{\boldsymbol{\lambda}\ge\mathbf{0}} L(\mathbf{x},\boldsymbol{\lambda})=+\infty$.

因此有

$$\theta(\mathbf{x})=\max_{\boldsymbol{\lambda}\ge\mathbf{0}} L(\mathbf{x},\boldsymbol{\lambda})=\begin{cases} f_0(\mathbf{x}), & f_i(\mathbf{x})\le 0, i=1,2,\cdots,m \\ +\infty, & \text{otherwise.}\end{cases} \tag{23}$$

对于 $x\in\mathcal{D}$ 最小化 $\theta(\mathbf{x})$，由于 $f_0(\mathbf{x})<+\infty$，则有

$$\min_{\mathbf{x}\in\mathcal{D}}\max_{\boldsymbol{\lambda}\ge\mathbf{0}} L(\mathbf{x},\boldsymbol{\lambda})=\min_{\mathbf{x}\in\mathcal{D}}\theta(\mathbf{x})=\min_{\mathbf{x}} f_0(\mathbf{x}),\quad f_i(\mathbf{x})\le 0,\ i=1,2,\cdots,m. \tag{24}$$

这就说明了 (21) 的最优值等于 p^*. □

根据以上的结论，可以总结以下两个式子:

$$p^*=\min_{\mathbf{x}\in\mathcal{D}}\max_{\boldsymbol{\lambda}\ge\mathbf{0}} L(\mathbf{x},\boldsymbol{\lambda}), \tag{25}$$

$$d^*=\max_{\boldsymbol{\lambda}\ge\mathbf{0}}\min_{\mathbf{x}\in\mathcal{D}} L(\mathbf{x},\boldsymbol{\lambda}), \tag{26}$$

讨论拉格朗日函数的鞍点性质，有以下的鞍点定理.

定理 3 (鞍点定理)**.** *$L(\mathbf{x},\boldsymbol{\lambda})$ 有鞍点 $(\tilde{\mathbf{x}},\tilde{\boldsymbol{\lambda}})$ 是强对偶性存在，且 $\tilde{\mathbf{x}}$ 为原问题的最优解、$\tilde{\boldsymbol{\lambda}}$ 为对偶问题最优解的充要条件.*

证明. $(\Rightarrow)$ 若 $L(\mathbf{x},\boldsymbol{\lambda})$ 有鞍点 $(\tilde{\mathbf{x}},\tilde{\boldsymbol{\lambda}})$，则显然 $p^*=d^*$，即强对偶性存在. 由于

$$\tilde{\mathbf{x}}=\arg\min_{\mathbf{x}\in\mathcal{D}}\max_{\boldsymbol{\lambda}\ge\mathbf{0}} L(\mathbf{x},\boldsymbol{\lambda})=\arg\min_{\mathbf{x}} f_0(\mathbf{x}),\quad f_i(\mathbf{x})\le 0,\ i=1,2,\cdots,m. \tag{27}$$

$$\tilde{\boldsymbol{\lambda}}=\arg\max_{\boldsymbol{\lambda}\ge\mathbf{0}}\min_{\mathbf{x}\in\mathcal{D}} L(\mathbf{x},\boldsymbol{\lambda})=\arg\max_{\boldsymbol{\lambda}\ge\mathbf{0}} g(\boldsymbol{\lambda}). \tag{28}$$

因此 $\tilde{\mathbf{x}}$ 为原问题的最优解、$\tilde{\boldsymbol{\lambda}}$ 为对偶问题最优解.

$(\Leftarrow)$ 若已知 $\tilde{\mathbf{x}}$ 是原问题最优解，$\tilde{\boldsymbol{\lambda}}$ 为对偶问题最优解，且强对偶性存在，则有

$$\begin{aligned} f_0(\tilde{\mathbf{x}}) &= g(\tilde{\boldsymbol{\lambda}}) && (29)\\ &= \min_{\mathbf{x}\in\mathcal{D}}\left\{f_0(\mathbf{x})+\sum_{i=1}^n\tilde{\lambda}_i f_i(\mathbf{x})\right\}, \quad \text{definition of } g(\tilde{\boldsymbol{\lambda}}) && (30)\\ &\le f_0(\tilde{\mathbf{x}})+\sum_{i=1}^n\tilde{\lambda}_i f_i(\tilde{\mathbf{x}}), \quad \min_{\mathbf{x}\in\mathcal{D}} L(\mathbf{x},\boldsymbol{\lambda})\le L(\tilde{\mathbf{x}},\boldsymbol{\lambda}) && (31)\\ &\le f_0(\tilde{\mathbf{x}}), \quad \lambda_i f_i(\tilde{\mathbf{x}})\le 0. && (32)\end{aligned}$$

由于等式两端 $f_0(\mathbf{x})$ 推知中间的 $\le$ 全部都要取等号. 因此，根据第三行可知 $f_0(\tilde{\mathbf{x}})=L(\tilde{\mathbf{x}},\tilde{\boldsymbol{\lambda}})$，根据第二行和第三行可知

$$L(\tilde{\mathbf{x}},\tilde{\boldsymbol{\lambda}})\le L(\mathbf{x},\tilde{\boldsymbol{\lambda}}). \tag{33}$$

根据第三行和第四行可知 $\tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) = 0$.

$$L(\tilde{\mathbf{x}}, \boldsymbol{\lambda}) = f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{n} \lambda_i f_i(\tilde{\mathbf{x}}) \le f_0(\tilde{\mathbf{x}}) = L(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}). \tag{34}$$

综合 (33) 和 (34) 可得

$$L(\tilde{\mathbf{x}}, \boldsymbol{\lambda}) \le L(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}) \le L(\mathbf{x}, \tilde{\boldsymbol{\lambda}}) \tag{35}$$

因此 $(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}})$ 为 $L(\mathbf{x}, \boldsymbol{\lambda})$ 的鞍点. □

鞍点定理说明如果强对偶性满足，则最优解是拉格朗日函数 $L(\mathbf{x}, \boldsymbol{\lambda})$ 的鞍点. 因此如果已经求得对偶问题的最优解 $\tilde{\boldsymbol{\lambda}}$，则原问题的最优解为

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} L(\mathbf{x}, \tilde{\boldsymbol{\lambda}}). \tag{36}$$

4 KKT 条件

对于任何优化问题

$$\begin{aligned} &\min_{\mathbf{x}} f_0(\mathbf{x}) \\ &\text{s.t. } f_i(\mathbf{x}) \le 0, \quad i = 1, 2, \cdots, m \\ &\quad\quad h_i(\mathbf{x}) = 0, \quad i = 1, 2, \cdots, n. \end{aligned} \tag{37}$$

假设 $f_0, f_1, \cdots, f_m$ 以及 $h_1, h_2, \cdots, h_n$ 均可微（简称这样的问题为可微凸问题），并且对偶间隙为零（即强对偶性成立）.

定义 5 (KKT条件)**.** *设 $\tilde{\mathbf{x}}$、$\tilde{\boldsymbol{\lambda}}$ 和 $\tilde{\boldsymbol{\nu}}$ 为原问题和对偶问题的某一组最优解. 则该问题的* ***Karush-Kuhn-Tucker (KKT)*** *条件为*

$$f_i(\tilde{\mathbf{x}}) \le 0, \quad i = 1, 2, \cdots, m \tag{38}$$

$$h_i(\tilde{\mathbf{x}}) = 0, \quad i = 1, 2, \cdots, n \tag{39}$$

$$\tilde{\lambda}_i \ge 0, \quad i = 1, 2, \cdots, m \tag{40}$$

$$\tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) = 0, \quad i = 1, 2, \cdots, m \tag{41}$$

$$\nabla_{\mathbf{x}} f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla_{\mathbf{x}} f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{n} \tilde{\nu}_i \nabla_{\mathbf{x}} h_i(\mathbf{x}) = 0. \tag{42}$$

KKT 条件可以分为三个部分：

- 满足可行域，即(38),(39);
- 互补松弛，即(40),(41);
- 导数为零，即(42).

KKT 条件是重要的优化手段，如果强对偶性成立，则 KKT 条件给出了原问题最优解的必要条件.

定理 4. *对可微的任意优化问题，若强对偶性成立，则 KKT 条件是原问题取得最优解的必要条件.*

证明. 已知强对偶性成立，设 $\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}$ 为原问题、对偶问题的最优解，则有

$$
\begin{aligned}
&f_i(\tilde{\mathbf{x}}) \leq 0, \quad i = 1, 2, \cdots, m, &(43)\\
&h_i(\tilde{\mathbf{x}}) = 0, \quad i = 1, 2, \cdots, n. &(44)\\
&\tilde{\lambda}_i \geq 0. &(45)
\end{aligned}
$$

以及

$$
\begin{aligned}
d^* = g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) &= \min_{\mathbf{x}} L(\mathbf{x}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) \\
&\leq L(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) \\
&= f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{n} \tilde{\nu}_i h_i(\tilde{\mathbf{x}}) \\
&\leq f_0(\tilde{\mathbf{x}}) = p^*.
\end{aligned} \tag{46}
$$

注意到强对偶性存在，即 $d^* = p^*$，则两个不等号都应该取等号，因此就可以推知

$$L(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) = \min_{\mathbf{x}} L(\mathbf{x}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}). \tag{47}$$

而又由于 $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ 对 $\mathbf{x}$ 可微，因此其一阶最优性条件需满足:

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) = \nabla_{\mathbf{x}} f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla_{\mathbf{x}} f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{n} \tilde{\nu}_i \nabla_{\mathbf{x}} h_i(\mathbf{x}) = 0. \tag{48}$$

而同时，根据 (46) 的最后两行，得知

$$\sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) = 0. \tag{49}$$

由于 $\tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) \leq 0$，但对每项求和却为零，说明

$$\lambda_i f_i(\tilde{\mathbf{x}}) = 0. \tag{50}$$

以上证明了 KKT 条件的必要性. □

更进一步，对于可微凸问题，有以下的结论.

定理 5. *如果 (37) 是凸问题，并且 $f_0, f_1, \cdots, f_m$ 以及 $h_1, h_2, \cdots, h_n$ 均可微. 则满足 KKT 条件的是原问题和对偶问题的最优解.*

证明. 若已知 $\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*$ 满足 KKT 条件，则有

$$
\begin{aligned}
g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) &= \min_{\mathbf{x}} f_0(\mathbf{x}) + \sum_{i=1}^{n} \lambda_i^* f_i(\mathbf{x}) + \sum_{i=1}^{n} \nu_i^* h_i(\mathbf{x}) \\
&= f_0(\mathbf{x}^*) + \sum_{i=1}^{n} \lambda_i^* f_i(\mathbf{x}^*) + \sum_{i=1}^{n} \nu_i^* h_i(\mathbf{x}^*) \\
&= f_0(\mathbf{x}^*)
\end{aligned} \tag{51}
$$

其中，第一行到第二行的原因是 $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ 关于 $\mathbf{x}$ 为凸函数，且有 (42). 第二行到第三行的原因是 $\lambda_i^* f_i(\mathbf{x}^*) = 0$ 以及 $h_i(\mathbf{x}^*) = 0$. 等式 (51) 说明了问题 (37) 的对偶间隙为零（强对偶性成立），因此说明 $\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*$ 为原问题和对偶问题的最优解. $\square$

根据以上的结论容易得出:

定理 6. *对于满足 Slater 条件的可微凸问题，KKT 条件是最优性的充要条件.*

因此，KKT 条件对于凸优化是十分重要的. 大多凸优化算法都需要查找满足 KKT 条件的解，从而就能找到原问题的最优解.

参考资料

[1] Boyd S, Boyd S P, Vandenberghe L. Convex optimization[M]. Cambridge university press, 2004.

[2] `https://www.bilibili.com/video/BV1Jt411p7jE`.