@TOC

对偶

对偶理论 $primal\ dual$

有约束问题的无约束形式

对于有约束优化问题

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \leqslant 0, & i=1,\cdots,m \\ & h_i(x)=0, & i=1,\cdots,p \end{array}$$

定义域

$$\mathcal{D}=\bigcap_{i=0}^{m} \operatorname{dom} f_i \cap \bigcap_{i=1}^{p} \operatorname{dom} h_i$$

可转化为无约束优化问题,即

$$\text{minimize} \quad f_0(x)+\sum_{i=1}^{m} I_{-}(f_i(x))+\sum_{i=1}^{p} I_0(h_i(x))$$

其中

$$I_{-}(u)=\begin{cases} 0 & u \leqslant 0 \\ \infty & u \geq 0 \end{cases}$$

$$I_0(u)=\begin{cases} 1 & u \in \{0\} \\ 0 & u \notin \{0\} \end{cases}$$

$I_0(u)$ 可看成是集合 $\{0\}$ 的示性函数, $I_{-}(u)$ 可看成是对原问题违反约束 $f_i(x)$ 的一种惩罚,一旦原问题违反约束, 则 $f_i(x)$ 的值会大于0, 则 $I_{-}(u)=\infty$, 整个无约束函数的最小值为 ∞.

Lagrange 函数,自变量为 x, λ, v

Lagrange 函数的定义为:

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

可以看出,有约束问题的无约束形式与 Lagrange 函数有些相似,我们可以认为, Lagrange 函数用线性的惩罚函数(或者惩罚力度比较弱的惩罚函数)代替了函数 $I_0(u)$, $I_{-}(u)$.

对于不等式约束,如果 $f_i(x)\geq 0$, 那么惩罚力度开始增加, 如果$f_i(x)<0$,不但没有惩罚,反而会有奖励,即 minimize $L(x,\lambda,v)$ 会取到更小的值.

对于等式约束,待补充.

用线性函数 $\lambda_i u_i$ 去逼近 $I_-(u)$ 是远远不够的,但是线性函数至少可以看成是示性函数的一个下估计, 我们可以得到, 对偶函数是原问题最优函数值的一个下界.

Lagrange 对偶函数,自变量为 λ, v

Lagrange 对偶函数的定义为:

$$g(\lambda, v)=\inf_{x\in\mathcal{D}} L(x,\lambda,v)=\inf_{x\in\mathcal{D}}\left(f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)+\sum_{i=1}^{p}v_i h_i(x)\right),$$
$$\lambda\in R_+^m, v\in R^p, \mathcal{D}\text{是原问题的定义域}$$

Lagrange **对偶函数相比于** Lagrange **函数,多出了对 $x(x\in\mathcal{D})$ 求最大下确界的过程,变量 x 被对偶函数所隐藏.**

注意,这里 Lagrange **对偶函数和原问题的定义域是相同的,都为 $\mathcal{D}$, 这很关键.**

证明 Lagrange **对偶函数构成了原问题最优值 p^* 的下界**

对任意的 $\lambda\succeq 0$和 v, 下式成立: $\preceq$

$$g(\lambda, v)\leqslant p^*$$

证明:

设 $\tilde{x}$ 是原问题在定义域 $\mathcal{D}$ 中的一个可行点(不一定是最优点), 正是由于 $\tilde{x}\in\mathcal{D}$,所以有 $f_i(\tilde{x})\leqslant 0, h_i(\tilde{x})=0$, 根据假设 $\lambda\succeq 0$, 我们有

$$\begin{aligned}
&\sum_{i=1}^{m}\lambda_i f_i(\tilde{x})+\sum_{i=1}^{p}v_i h_i(\tilde{x})\\
&=\sum_{i=1}^{m}\lambda_i f_i(\tilde{x})+\sum_{i=1}^{p}v_i * 0\\
&=\sum_{i=1}^{m}\lambda_i f_i(\tilde{x})\leqslant 0
\end{aligned}$$

则

$$\begin{aligned}
L(\tilde{x},\lambda,v)&=f_0(\tilde{x})+\sum_{i=1}^{m}\lambda_i f_i(\tilde{x})+\sum_{i=1}^{p}v_i h_i(\tilde{x})\\
&=f_0(\tilde{x})+\sum_{i=1}^{m}\lambda_i f_i(\tilde{x})\leqslant f_0(\tilde{x})
\end{aligned}$$

式 $\inf\limits_{x\in\mathcal{D}} L(x,\lambda,v)$ 表明函数 $L(x,\lambda,v)$ 在集合 $\mathcal{D}$ 中的下界, 因此有

$$g(\lambda,v)=\inf_{x\in\mathcal{D}} L(x,\lambda,v)\leqslant L(\tilde{x},\lambda,v)\leqslant f_0(\tilde{x})$$

因此 Lagrange 对偶函数构成了原问题最优值 p^* 的下界.

证明 Lagrange 对偶函数为凹函数

无论原问题是否为凸问题, 无论$f_0(x), f_i(x), h_i(x)$ 是什么类型的函数. 但 Lagrange 对偶函数是一个凹函数, 证明思路如下: 集合 $\mathcal{D}$ 无论是否连续,都可以看成是变量 x 的有限集合,当对偶函数 $g(\lambda,v)$ 中的 λ,v 为定值时,则对于集合\mathcal{D}中的每一个 x_i 都有 $f_i(x_i)$, $h_i(x_i)$ 与之对应,

令

$$\begin{aligned}
\lambda &= [\,\lambda_1,\lambda_2,\cdots,\lambda_m\,] \\
v &= [\,v_1,v_2,\cdots,v_p\,] \\
F(x) &= [\,f_1(x),f_2(x),\cdots,f_m(x)\,] \\
H(x) &= [\,h_1(x),h_2(x),\cdots,h_p(x)\,] \\
\mathcal{D} &= \{x_0,x_1,\cdots,x_n\}, \quad \text{当集合}\mathcal{D}\text{为连续值集合时}, n\to\infty
\end{aligned}$$

且

$$\begin{aligned}
g(\lambda,v) &= \inf_{x\in\mathcal{D}} L(x,\lambda,v) \\
&= \inf_{x\in\mathcal{D}} \left(f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)+\sum_{i=1}^{p}v_i h_i(x)\right) \\
&= \inf_{x\in\mathcal{D}} \left(f_0(x)+\lambda F(x)^T+vH(x)^T\right)
\end{aligned}$$

则有

$$\begin{aligned}
g(\lambda,v) \quad &= \inf_{x\in\mathcal{D}} \quad L(x,\lambda,v) \\
&= \inf \quad \big[\big(f_0(x_0)+\lambda F(x_0)^T+vH(x_0)^T\big), \\
&\qquad\quad \big(f_0(x_1)+\lambda F(x_1)^T+vH(x_1)^T\big), \\
&\qquad\qquad\qquad \cdots, \\
&\qquad\quad \big(f_0(x_n)+\lambda F(x_n)^T+vH(x_n)^T\big)\big]
\end{aligned}$$

为了方便书写和观察，令

$$\begin{aligned}
A_i &= f_0(x_i) \\
B_i &= F(x_i) \\
C_i &= H(x_i)
\end{aligned}$$

则

$$
\begin{aligned}
g(\lambda, v) &= \inf_{x \in \mathcal{D}} \quad L(x, \lambda, v) \\
&= \inf \quad \big[\big(f_0(x_0) + \lambda F(x_0)^T + vH(x_0)^T\big), \\
&\qquad\quad \big(f_0(x_1) + \lambda F(x_1)^T + vH(x_1)^T\big), \\
&\qquad\quad \cdots, \\
&\qquad\quad \big(f_0(x_n) + \lambda F(x_n)^T + vH(x_n)^T\big)\big] \\
&= \inf \quad \big[\big(A_0 + \lambda B_0^T + vC_0^T\big), \\
&\qquad\quad \big(A_1 + \lambda B_1^T + vC_1^T\big), \\
&\qquad\quad \cdots, \\
&\qquad\quad \big(A_n + \lambda B_n^T + vC_n^T\big)\big]
\end{aligned}
$$

易知 $A_i + \lambda B_i^T + vC_i^T$ 是关于变量 λ, v 的仿射函数 因为 **凹函数的逐点最大下确界仍是凹函数(仿射函数即是凹函数,也是凸函数)**, 因此 Lagrange 对偶函数仍是凹函数.

Lagrange **对偶问题**

Lagrange 对偶问题的形式如下所示:

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

Lagrange **对偶问题** 可以用来寻找 Lagrange **对偶函数** 的最大值.由前面的内容可知, Lagrange **对偶函数** 可以在 λ, v 确定的情况下,对 x $(x \in \mathcal{D})$ 求 Lagrange **函数** 的最大下确界(最小值),我们将 Lagrange 对偶问题写成如下形式:

$$
\max_{\lambda \succeq 0} \inf_{x \in \mathcal{D}} L(x, \lambda, v)
$$

或者如下形式(当 max 可以取到时)

$$
\sup_{\lambda \succeq 0} \inf_{x \in \mathcal{D}} L(x, \lambda, v)
$$

对 Lagrange **对偶的鞍点解释, 矮个子里挑高个子 和 高个子里挑矮个子**

为了简化讨论,假设原问题没有等式约束

$$
\begin{aligned}
&\text{minimize} \quad f_0(x) \\
&\text{subject to} \quad f_i(x) \leqslant 0, \qquad i = 1, \cdots, m
\end{aligned}
$$

原问题的最优值 p^* 与下面这个问题的最优值等价:

构造原问题的 Lagrange 函数,即:

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

求 Lagrange 函数在x确定的情况下, 对 $\lambda \succeq 0$ 求最小上确界:

$$\begin{aligned}\sup_{\lambda\succeq 0} L(x,\lambda) &= \sup_{\lambda\succeq 0}\left(f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)\right)\\ &= \begin{cases} f_0(x) & f_i(x)\leqslant 0, i=1,\cdots,m\\ \infty & \text{otherwise}\end{cases}\end{aligned}$$

当$f_i(x)>0$时,只需要让相应的 $\lambda_i=+\infty$ 其余 $\lambda_j=0\ (i\neq j)$,则 Lagrange 函数对 $\lambda\succeq 0$ 求最小上确界的结果为 $+\infty$, 此时无论$f_0(x)$ 的值是多少都无意义, 在这里不考虑 $f_0(x_i)=\pm\infty$. 这种情况下, 我们将无法取得一个有效值, 因此讨论这种情况的最小上确界无意义.

求得 $\sup\limits_{\lambda\succeq 0} L(x,\lambda)$ 的结果之后, 再求在 $x\in\mathcal{D}$ 的情况下对 $\sup\limits_{\lambda\succeq 0} L(x,\lambda)$ 求最大下确界, 即 $\inf\limits_{x\in\mathcal{D}}\ \sup\limits_{\lambda\succeq 0} L(x,\lambda)$, 这等价于求解原问题. 令原问题的最优值为 p^* :

$$\begin{aligned}p^* &= \inf_{x\in\mathcal{D}}\ \sup_{\lambda\succeq 0} L(x,\lambda)\\ &= \inf_{x\in\mathcal{D}}\ \sup_{\lambda\succeq 0}\left(f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)\right)\\ &= \inf_{x\in\mathcal{D}}\ \begin{cases} f_0(x) & f_i(x)\leqslant 0, i=1,\cdots,m\\ \infty & \text{otherwise}\end{cases}\\ &= \inf_{x\in\mathcal{D}}\ f_0(x) \quad \text{subject to} \quad f_i(x)\leqslant 0, i=1,\cdots,m\\ &= \min_{x\in\mathcal{D}}\ f_0(x) \quad \text{subject to} \quad f_i(x)\leqslant 0, i=1,\cdots,m\\ &\quad (\text{如果最小值可以取到, 则 sup 与min等价})\end{aligned}$$

设对偶问题的最优值为 d^*, 根据对偶问题的定义:

$$\begin{aligned}d^* &= \sup_{\lambda\succeq 0}\ g(\lambda,v)\\ &= \sup_{\lambda\succeq 0}\ \inf_{x\in\mathcal{D}}\ L(x,\lambda)\\ &= \sup_{\lambda\succeq 0}\ \inf_{x\in\mathcal{D}}\ \left(f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)\right)\end{aligned}$$

因此**弱对偶性**可以表示为下述不等式

$$\begin{aligned}\text{对偶问题的最优值 } d^* &\leqslant \text{原问题的最优值 } p^*\\ d^* &\leqslant p^*\\ \sup_{\lambda\succeq 0}\ \inf_{x\in\mathcal{D}}\ L(x,\lambda) &\leqslant \inf_{x\in\mathcal{D}}\ \sup_{\lambda\succeq 0}\ L(x,\lambda)\end{aligned}$$

而**强对偶性**可以表示为下述等式

$$\begin{aligned}\text{对偶问题的最优值 } d^* &= \text{原问题的最优值 } p^*\\ d^* &= p^*\\ \sup_{\lambda\succeq 0}\ \inf_{x\in\mathcal{D}}\ L(x,\lambda) &= \inf_{x\in\mathcal{D}}\ \sup_{\lambda\succeq 0}\ L(x,\lambda)\end{aligned}$$

当 $p^*=d^*$ **时,强对偶成立**, 强对偶性意味着对 x 求极小和对 $\lambda\succeq 0$ 求极大可以互换而不影响结果.

鞍点定义

我们称一对 $\tilde{w} \in W$ 是函数 f (以及 W 和 Z) 的鞍点，如果对任意 $w \in W$ 和 $z \in Z$ 下式成立

$$f(\tilde{w}, z) \leqslant f(\tilde{w}, \tilde{z}) \leqslant f(w, \tilde{z}).$$

换言之，$g_1(w) = f(w, \tilde{z})$ 在 $\tilde{w}$ 处取得最小值(关于变量 $w \in W$)，$g_2(z) = f(\tilde{w}, z)$ 在 $\tilde{z}$ 处取得最大值（关于变量 $z \in W$):

$$f(\tilde{w}, \tilde{z}) = \inf_{w \in W} f(w, \tilde{z}), \quad f(\tilde{w}, \tilde{z}) = \sup_{z \in Z} f(\tilde{w}, z)$$

上式意味着强极大极小性质成立，且共同值为 $f(\tilde{w}, \tilde{z})$. 回到我们关于 Lagrange 对偶的讨论,如果 x^* 和 λ^* 分别是原问题和对偶问题的最优点,且强对偶性成立,则它们是 Lagrange 函数的一个鞍点.

反过来同样成立: 如果 (x, λ) 是 Lagrange 函数的一个鞍点，那么 x 是原问题的最优解，λ 是对偶问题的最优解,且最优对偶间隙为零.

对偶的操作

在本文中, 我们可以看到对偶操作消去了一些变量, 用集合中显性的值来取代某一变量, 另外, 构造了许多的仿射函数, 最后,进行取最大下界或最小上界操作. 我们应该学会这种变换思想

参考资料

- [1] Convex Optimization – Boyd and Vandenberghe
- [2] 中科大-凸优化