凸优化 & KKT 条件

Yimeng Ren

1 凸集和凸函数

1.1 凸集

$C \in \mathbb{R}^p$ is convex if:

$$\begin{aligned}&\text{all } \beta, \beta' \in C \text{ ajd all scalar } s \in [0,1]\\&\text{we have: } s\beta + (1-s)\beta' \in C\end{aligned}$$

1.2 凸函数

$f : \mathbb{R}^P \to \mathbb{R}$ is convex function if:

$$\beta, \beta' \in dom(f), s \in [0,1]$$

we have

$$f(\beta(s)) = f\left(s\beta + (1-s)\beta'\right) \leq sf(\beta) + (1-s)f\left(\beta'\right)$$

2 Lagrange 对偶

optimization problem:

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

where $x \in \mathbf{R}^n$ and $\mathcal{D} = \bigcap_{i=0}^{m} \operatorname{dom} f_i \cap \bigcap_{i=1}^{p} \operatorname{dom} h_i \neq \Phi$

denote the solution is $x^\star$ and corresponding value of function is $p^\star$.

the **Lagrange Function** $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ is:

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

the **Lagrange Dual Function**$g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ 为 $L(x, \lambda, \nu)$ 对 $x \in \mathcal{D}$ 取最小值:

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

对偶函数构成了原问题最优值 $p^\star$ 的下界，即对于任意 $\lambda \geq 0$ and ν，有:

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

即我们可以得到和参数 λ, ν 相关的一个下界。那么从 Lagrange 函数可以得到的**最好**下界是什么?

Lagrange 对偶问题:

$$\begin{aligned} &\max\nolimits_{\lambda,\nu} g(\lambda, \nu) \\ &\text{s.t. } \lambda \succeq \mathbf{0} \end{aligned}$$

记对偶问题的最优值为 $d^\star$，对应的最优解为 $(\lambda^\star, \nu^\star)$。

2.1 弱对偶性

对偶问题的最优值为 $d^\star$，原问题的最优值为 $p^\star$，则 $d^\star$ 是对偶函数 $g(\lambda, \nu)$ 能给出的 $p^\star$ 的最好的下界，且有:

$$d^\star \leq p^\star$$

称为弱对偶性。即使原问题非凸，弱对偶性也成立。定义**最优对偶间隙**为 $p^\star - d^\star$

2.2 强对偶性

如果等式

$$d^\star = p^\star$$

成立，则称强对偶性成立。强对偶性的成立需要满足一定的条件。

3 最优性条件

假设强对偶性成立（原对偶问题有相同的最优值），$x^\star$ 为原问题的最优解，$(\lambda^\star, \nu^\star)$ 为对偶问题的最优解，这表明：

$$\begin{aligned} f_0\left(\mathbf{x}^\star\right) &= g\left(\lambda^\star, \nu^\star\right) \\ &= \min_{\mathbf{x}\in\mathcal{D}}\left(f_0(\mathbf{x})+\sum_{i=1}^m \lambda_i^\star f_i(\mathbf{x})+\sum_{i=1}^p \nu_i^\star h_i(\mathbf{x})\right) \\ &\leq f_0\left(\mathbf{x}^\star\right)+\sum_{i=1}^m \lambda_i^\star f_i\left(\mathbf{x}^\star\right)+\sum_{i=1}^p \nu_i^\star h_i\left(\mathbf{x}^\star\right) \\ &\leq f_0\left(\mathbf{x}^\star\right) \end{aligned}$$

最后一个不等式成立是因为 $\lambda^\star \geq 0, f_i(x^\star) \leq 0\ (i=1,...,m), h_i(x^\star)=0\ (i=1,...p)$。但是由于 $f_0\left(\mathbf{x}^\star\right)=f_0\left(\mathbf{x}^\star\right)$，两个不等号都应该取等号，即

$$f_0\left(\mathbf{x}^\star\right)+\sum_{i=1}^m \lambda_i^\star f_i\left(\mathbf{x}^\star\right)+\sum_{i=1}^p \nu_i^\star h_i\left(\mathbf{x}^\star\right)=f_0\left(\mathbf{x}^\star\right)$$

1. 由于 $h_i\left(\mathbf{x}^\star\right)=0$，所以有

$$\sum_{i=1}^m \lambda_i^\star f_i\left(\mathbf{x}^\star\right)=0$$

又因为 $\lambda_i^\star \geq 0, \quad f_i\left(\mathbf{x}^\star\right) \leq 0$，所以 $\lambda_i^\star f_i\left(\mathbf{x}^\star\right) \leq 0$，因此每一项都应当为 0，即

$$\lambda_i^\star f_i\left(\mathbf{x}^\star\right)=0, i=1,\ldots,m$$

2. 又要求

$$\min_{\mathbf{x}}\left(f_0(\mathbf{x})+\sum_{i=1}^m \lambda_i^\star f_i(\mathbf{x})+\sum_{i=1}^p \nu_i^\star h_i(\mathbf{x})\right)=f_0\left(\mathbf{x}^\star\right)+\sum_{i=1}^m \lambda_i^\star f_i\left(\mathbf{x}^\star\right)+\sum_{i=1}^p \nu_i^\star h_i\left(\mathbf{x}^\star\right)$$

即 $x^\star$ 为 $L\left(\mathbf{x}, \lambda^\star, \nu^\star\right)$ 的最小值点，因此，$L\left(\mathbf{x}, \lambda^\star, \nu^\star\right)$ 在此处梯度必须为 0，即

$$\nabla f_0\left(\mathbf{x}^\star\right)+\sum_{i=1}^m \lambda_i^\star \nabla f_i\left(\mathbf{x}^\star\right)+\sum_{i=1}^p \nu_i^\star \nabla h_i\left(\mathbf{x}^\star\right)=\mathbf{0}$$

此时，得到 **KKT 条件 (Karush–Kuhn–Tucker conditions, KKT conditions)** 条件为:

$$
\begin{aligned}
f_i\left(\mathbf{x}^{\star}\right) &\leq 0, i=1, \ldots, m \\
h_i\left(\mathbf{x}^{\star}\right) &=0, i=1, \ldots, p \\
\lambda_i^{\star} &\geq 0, i=1, \ldots, m \\
\lambda_i^{\star} f_i\left(\mathbf{x}^{\star}\right) &=0, i=1, \ldots, m \\
\nabla f_0\left(\mathbf{x}^{\star}\right)+\sum_{i=1}^m \lambda_i^{\star} \nabla f_i\left(\mathbf{x}^{\star}\right)+\sum_{i=1}^p \nu_i^{\star} \nabla h_i\left(\mathbf{x}^{\star}\right) &=\mathbf{0}
\end{aligned}
$$

对目标函数和约束函数可微，且强对偶性成立的优化问题（无论原问题是否是凸优化问题），KKT 条件是最优解的必要条件，即原对偶问题的任意一对最优解都必须满足 KKT 条件。