理解拉格朗日对偶

本笔记是对凸优化理论中拉格朗日对偶性的总结与理解，主要参考了以下资料：

[Convex Optimization, Stephen Boyd and Lieven Vandenberghe, Cambridge University](#)

[如何通俗地讲解对偶问题？尤其是拉格朗日对偶lagrangian duality？](#)

本笔记的目的在于帮助读者更好地理解拉格朗日对偶性的概念，因此不会包含严谨的数学证明过程。

1. 拉格朗日对偶问题

我们先给出拉格朗日对偶性的定义形式。

考虑具有等式约束和不等式约束的原始优化问题：

$$\begin{array}{ll} \min_x & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, \quad i=1,\ldots,m \\ & h_i(x)=0, \quad i=1,\ldots,p \end{array}$$

这个问题的定义域是目标函数、等式约束函数和不等式约束函数定义域的交集，即 $\mathcal{D}=\bigcap_{i=0}^m \operatorname{dom} \mathrm{f}_i \cap \bigcap_{i=1}^p \operatorname{dom} \mathrm{h}_i$，将原始问题的最优值记为$p^*$。

构造拉格朗日函数：

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

我们将拉格朗日函数$L(x,\lambda,\nu)$在定义域上针对x求下界的结果称为拉格朗日对偶函数(Lagrange dual function)，它是拉格朗日乘子λ和ν 的函数，即

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

上式最右端括号中的函数可以看作是关于一组(λ,ν)的仿射函数，因此函数g是关于这个仿射函数的逐点下确界。由于凹函数（仿射函数既凸且凹）的逐点下确界仍是凹函数，因此g**总是凹函数，不管原问题是否为凸。**

函数g的另一个重要性质是它给出了原问题最优值p^*的一个下界，即对任意的$\lambda \succeq 0$和ν下式成立：

$$g(\lambda,\nu)\leqslant p^* \tag{1.1}$$

可以很容易地验证这个重要的性质。设$\tilde{x}$是原问题的一个可行点，即$f_i(\tilde{x})\leqslant 0$且$h_i(\tilde{x})=0$，假设$\lambda\succeq 0$，我们有

$$\sum_{i=1}^m \lambda_i f_i(\tilde{x})+\sum_{i=1}^p \nu_i h_i(\tilde{x})\leqslant 0$$

这是因为左边的第一项非正而第二项为零。根据上述不等式，有

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

因此

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

对于每一个可行点$\tilde{x}$都满足$g(\lambda,\nu)\leqslant f_0(\tilde{x})$，因此不等式(1.1)成立。

很自然地，我们希望找到一个最大的下界来逼近优化值p^*，因此有如下优化问题：

$$\begin{array}{ll}\max_{\lambda,\nu} & g(\lambda,\nu)\\ \text{s.t.} & \lambda\succeq 0\end{array}$$

我们称这个问题为原问题的**对偶问题**。通过上面的分析我们知道，只有当$\lambda\succeq 0$时，$g(\lambda,\nu)$才能成为p^*的一个下界，我们称这种情况为**对偶可行**(Dual feasible)。

拉格朗日对偶问题最重要的性质是不管原问题是否是凸的，它的**对偶问题总是一个凸优化问题**（因为对偶问题是在求一个凹函数的最大值）。对偶问题给出了原始问题最优值的一个下界，当原始问题不便于求解时，我们可以通过求解它的对偶问题来找到原始最优值的下界。另外，当满足一定条件时（强对偶性），原始问题和对偶问题的解是完全等价的。

2. 拉格朗日对偶的由来

理解拉格朗日对偶有很多种方式，在我的另一篇关于KKT条件的笔记中，我们还会对这个问题进行进一步的探讨。这里给出其中一种，参考了知乎答主**@又红又正**的回答https://www.zhihu.com/question/58584814, 介绍如何一步步引入拉格朗日对偶的过程。

为方便起见，我们考虑仅含不等式约束的优化问题：

$$\begin{array}{ll}\min_x & f_0(x)\\ \text{s.t.} & f_i(x)\leq 0,\quad i=1,\ldots,m\end{array}\tag{2.1}$$

我们的目标是要找到问题(2.1)最优值的一个最好的下界。首先，我们考虑方程组

$$\begin{array}{l}f_0(\mathbf{x})<v\\ f_i(\mathbf{x})\leq 0, i=1,2,\ldots,m\end{array}\tag{2.2}$$

若方程组(2.2)无解，则v是问题(2.1)的一个下界，因为此时没有比v更小的$f_0(x)$。同时我们注意到，当方程组(2.2)有解时，对于任意的$\lambda\geq 0$，以下不等式成立：

$$f_0(x)+\sum_{i=1}^m\lambda_i f_i(x)<f_0(x)<v\tag{2.3}$$

因此，根据逆否命题，我们可以推出，方程组(2.2)无解的充要条件是：存在$\lambda\geq 0$，使得方程(2.3)不成立，即有

$$\tilde{v}=\min_x f_0(x)+\sum_{i=1}^m\lambda_i f_i(x)\geq v>f_0(x)\tag{2.4}$$

即$\tilde{v}$也是$f_0(x)$的下界（注意此时$\tilde{v}>f_0(x)$不成立，即$\tilde{v}\leq f_0(x)$）。由于我们要找的是最好的下界，因此我们要最大化$\tilde{v}$的值，于是有

$$\tilde{v}=\max_{\lambda\geq 0}\min_x f_0(x)+\sum_{i=1}^m\lambda_i f_i(x)\tag{2.5}$$

注意到，式(2.5)实际上就是原问题(2.1)的对偶问题。整个推理逻辑是先根据式(2.5)取$\tilde{v}$和λ，可得式(2.4)成立，从而导出式(2.3)无解，再导出方程组(2.2)无解，从而得到$\tilde{v}$是问题(2.1)的下界。