凸优化进阶之对偶理论

一：一般优化问题

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

问题的定义域$\mathcal{D} = (\bigcap_{i=0}^{m} \operatorname{dom} f_i) \bigcap \left(\bigcap_{i=0}^{p} \operatorname{dom} h_i\right)$.需要注意的是：定义域与可行域是不同的。

二：拉格朗日函数

2.1 函数基本介绍

- 拉格朗日函数将目标函数和约束条件整合到了一起。

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

- 主变量：$\mathbf{x}$
- 对偶变量：$\lambda \geq 0$,即$\lambda = [\lambda_1, \lambda_2, \cdots \lambda_m]^T \geq 0$;而$\mathbf{v}$可以大于等于0，也可以小于等于0.
- 意义解释：这其实是一种添加惩罚的方式。如果约束条件$f_i(\mathbf{x}) \geq 0$,那么就相当于加了一个正数($\lambda \geq 0$)，使得拉格朗日函数变大，而我们的目标在于最小化该函数，因此就会强迫约束条件小于等于0。

2.2 函数的主问题分析

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

- 主问题:

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

- 主问题分析：我们的目标其实是$\min_{\mathbf{x}} f_0(\mathbf{x})$,即我们要求函数$f_0(\mathbf{x})$的最小值，而不是$maxL$的最小值。那么原函数和$maxL$之间有什么关系呢?

  我们单独看max的部分：

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

在这个函数中，λ和v是变量。当x在可行域范围内时，$f_i(\mathbf{x})\leq 0,\lambda\geq 0$,所以括号中第一项小于等于0；第二项中由于$h_i(\mathbf{x})=0$,所以第二项也为0。这样也就是说$\max_{\lambda,v}\left(\sum_{i=1}^{m}\lambda_i f_i(\mathrm{x})+\sum_{i=1}^{p}v_i h_i(\mathrm{x})\right)$的值为0。所以最小化$maxL$和最小化目标函数是等效的。

2.3 函数的对偶问题分析

- 拉格朗日对偶函数

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

  - 需要注意：这个对偶函数是定义在函数定义域上的，不是可行域。
  - 回忆：逐点最大：$f_1,\cdots,f_m$凸，则$f(\mathrm{x})=\max\{f_1(\mathrm{x}),\cdots,f_m(\mathrm{x})\}$凸。$f(\mathbf{x},\mathbf{y})$对于每个$\mathrm{y}\in\mathcal{A}$凸，则$\max_{y\in\mathcal{A}} f(x,y)$凸。
  - $g(\boldsymbol{\lambda},\boldsymbol{v})$其实是关于$\lambda,v$的仿射函数，所以是既凸且凹的函数。所以该函数的逐点下确界总是凹的。即$g(\boldsymbol{\lambda},\boldsymbol{v})$是一个凹函数。
  - 如果$\tilde{\mathbf{x}}$是一个可行域中的点，则：

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

    - 因为$g(\boldsymbol{\lambda},\boldsymbol{v})$是拉格朗日函数的最小值，所以它小于等于**可行域**中的任何一个函数值。
    - 又因为

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

      - 因为对于可行域中的点，$f_i(\widetilde{\mathbf{x}})\leq 0$,$h_i(\widetilde{\mathbf{x}})=0$.所以$L(\widetilde{\mathbf{x}},\boldsymbol{\lambda},\boldsymbol{v})\leq f_0(\widetilde{\mathbf{x}})$.这也就是说$g(\boldsymbol{\lambda},\boldsymbol{v})\leq f_0(\widetilde{\mathbf{x}})$,即$g(\boldsymbol{\lambda},\boldsymbol{v})$小于等于$f_0(\mathbf{x})$中任意一点的函数值，当然也就小于其最小值了，即$g(\boldsymbol{\lambda},\boldsymbol{v})\leq f_0(\mathbf{x}^*)=p^*$。
      - 从上边的分析中可以看出，对偶问题小于等于原问题最优解的下界。如果我们能够求出对偶问题的上界，那么就可以确定一个原问题的下界。接下来我们就想办法求对偶问题的上界，即最大值。
- 拉格朗日对偶问题：

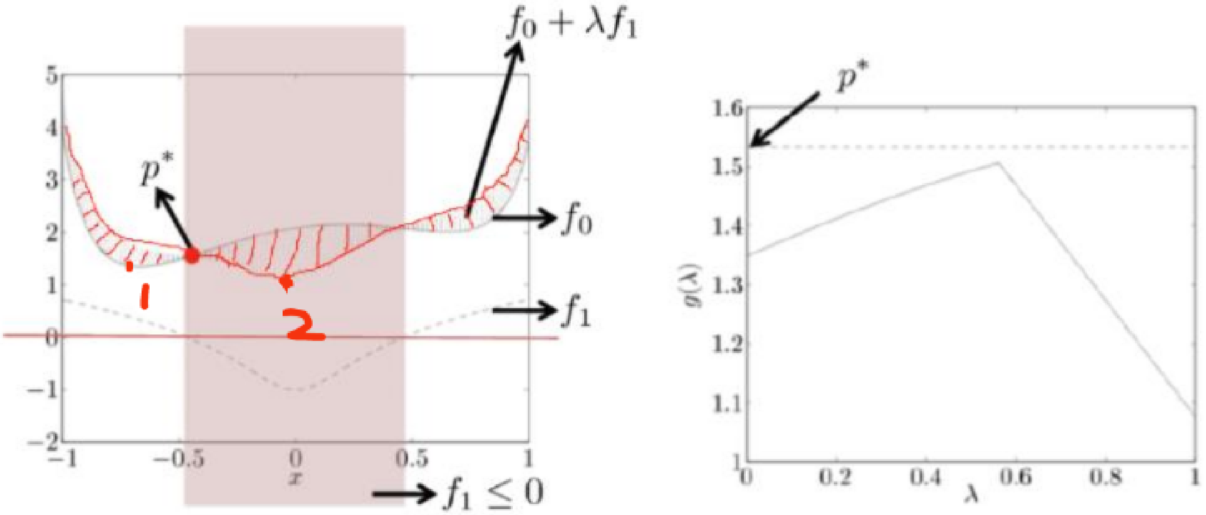
$$\begin{array}{ll}\text{maximize} & g(\lambda,v)\\ \text{subject to} & \lambda\geq 0\end{array}$$

- 目标函数：$\max_{\lambda\geq 0,\mathrm{v}}\min_{\mathrm{x}\in\mathcal{D}} L(\mathrm{x},\lambda,v)$

- 这其实是一个凹函数在凸集上的最大化问题。是一个凸优化问题。设其最优值为d^*，对应的极值点为λ^*, v^*。
- 从以上分析中我们就可以得出一个结论：不管原问题是不是一个凸优化问题，他的对偶问题一定是一个凸优化问题。$g(\boldsymbol{\lambda}^*, \mathbf{v}^*) = d^* \leq p^*$,即对偶问题的最大值小于等于原问题的最小值。

2.4 对偶问题的几何解释

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



- 可行域为图中阴影部分；定义域为[-1,1]
- 原问题的最小值（即可行域中的最小值）在p^*x处。
- 当$\lambda = 0$时，$L(\mathbf{x}, \lambda) = f_0(\mathbf{x})$.函数在**定义域**上的最小值在1处，约为1.3。当$\lambda = 1$时，$L(\mathbf{x}, \lambda) = f_0(\mathbf{x}) + f_1(\mathbf{x})$,此时函数的最小值在2处，约为0.8.当λ在[0,1]之间变化时，对偶问题的最大值始终小于原函数的最小值（参考右边图）。

2.5 强弱对偶问题解释

- 弱对偶：$d^* \leq p^*$，无论原问题是不是凸优化问题，总成立
- 强对偶：$d^* = p^*$，
  - 该条件通常不成立
  - 但是对于凸优化问题通常成立
  - 凸优化问题可以改写为：

$$\begin{gathered} minimize\ f_0(\mathbf{x}) \\ subject\ to\ f_i(\mathbf{x}) \leq 0\ for\ i=1,2,\cdots m \\ \mathbf{A}\mathbf{x}=\mathbf{b} \end{gathered}$$

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

- 说明：如果没有不等式约束，只有$\mathbf{A}\mathbf{x}=\mathbf{b}$，那么一定是强对偶；如果一个存在不等式约束，那么满足slater条件时，是强对偶；但是需要注意的是：不满足slater条件，不代表一定不是强对偶问题。

2.6 从对偶问题解主问题

- 假定强对偶问题成立，$(\mathrm{x}^*,\lambda^*,v^*)$是主问题和对偶问题的最优解，那么

$$\begin{aligned} p^* &= f_0\left(\mathrm{x}^*\right)=d^*=g\left(\lambda^*,v^*\right) \\ &= \min_{\mathbf{x}}\left(f_0(\mathrm{x})+\sum_{i=1}^m \lambda_i^* f_i(\mathrm{x})+\sum_{i=1}^p v_i^* h_i(\mathrm{x})\right) \\ &\text{最小值肯定小于等于任意一个x对应的函数值，因此} \\ &\leq f_0\left(\mathrm{x}^*\right)+\sum_{i=1}^m \lambda_i^* f_i\left(\mathrm{x}^*\right)+\sum_{i=1}^p v_i^* h_i\left(\mathrm{x}^*\right) \\ &\text{因为x星肯定在函数的可行域内，因此后边两项均为非正值，因此} \\ &\leq f_0\left(\mathrm{x}^*\right) \end{aligned}$$

- 观察上式首尾两项，可得$f_0(\mathrm{x}^*)\leq f_0(\mathrm{x}^*)$，因此小于等于号可变为等号。因此也就可以得到：
  - 结论1：$\lambda_i^* f_i\left(\mathrm{x}^*\right)=0$
  - 结论2：$L\left(\mathrm{x},\lambda^*,v^*\right)$关于$\mathbf{x}^*$处取极小值，有$\nabla_{\mathrm{x}} L\left(\mathrm{x}^*,\lambda^*,v^*\right)=0$
    - $g\left(\lambda^*,v^*\right)=L\left(\mathrm{x}^*,\lambda^*,v^*\right)$；对偶函数的最优解就在拉格朗日函数取极值的时候
    - $g\left(\lambda^*,v^*\right)=\min_x L\left(\mathrm{x},\lambda^*,v^*\right)$

2.7 KKT条件

- 凸优化问题：

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

- 拉格朗日函数

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

- 凸优化问题强对偶成立的充要条件：

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

- KKT问题的几何解释：https://www.zhihu.com/question/58584814/answer/159863739. 或查看参考资料中的文件。

2.8 主对问题思考

- 主问题：

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

- 对偶问题

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

- 主对关系（强对偶成立时，max和min可以互换，即可以取等号）

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

三：具体计算案例

3.1最小二范数问题

- 求解下列问题

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

分析一下问题：首先目标函数是一个二范数，可以看作$\mathbf{x}^{\mathbf{T}} \mathbf{I} \mathbf{x}$，其中$\mathbf{I}$为正定矩阵。所以是一个凸函数。另外，约束条件为等式约束，是超平面的集合，是凸集。因此问题是一个凸优化问题。

解法：

1. 写出拉格朗日函数：

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

2. 写出对偶函数：

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

则:

$$
\begin{aligned}
&\nabla_{\mathbf{x}} L(\mathbf{x}, \mathbf{v})=\mathbf{0} \Rightarrow 2 \mathbf{x}+\mathbf{A}^{T} \mathbf{v}=\mathbf{0} \Rightarrow \mathbf{x}^{*}(\mathbf{v})=-\tfrac{1}{2} \mathbf{A}^{T} \mathbf{v} \\
&g(\mathbf{v})=L\left(\mathbf{x}^{*}(\mathbf{v}), \mathbf{v}\right)=-\tfrac{1}{4} \mathbf{v}^{T}\left(\mathbf{A} \mathbf{A}^{T}\right) \mathbf{v}-\mathbf{v}^{T} \mathbf{b}
\end{aligned}
$$

g(v)是一个凹函数

3. 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:

$$
\begin{aligned}
\mathbf{v} * &=-2\left(\mathbf{A} \mathbf{A}^{T}\right)^{-1} \mathbf{b} \\
d^{*} &=\mathbf{b}^{T}\left(\mathbf{A} \mathbf{A}^{T}\right)^{-1} \mathbf{b}
\end{aligned}
$$

由于$p^{*}=d^{*}$,则

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

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

参考资料

1. https://www.zhihu.com/question/58584814/answer/159863739