对偶理论

Lagrange对偶函数

- 考虑标准形式的优化问题：

$$\text{minimize } f_0(x)$$
$$\text{subject to } f_i(x) \le 0,\ \ i = 1, . . , m$$
$$h_i(x) = 0,\ \ i = 1, \dots, p$$

变量$x \in R^n$，定义域为$\mathcal{D}$,其最优值为p^*

- Lagrange 对偶的基本思想是在优化目标函数中考虑约束条件，即添加约束条件的加权和，定义上述优化问题的**Lagrange函数**$L: R^n \times R^m \times R^p \to R$为

$$L(x, \lambda, \nu) = \mathrm{f}_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

定义域为$\mathrm{dom}L = \mathcal{D} \times R^m \times R^p$

Lagrange对偶函数

$$L(x,\lambda,\nu) = \mathrm{f}_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

定义域为$\mathrm{dom}L = \mathcal{D} \times R^m \times R^p$

- 目标函数和约束函数的加权和
- λ_i为第i个不等式约束$f_i(x) \le 0$对应的**Lagrange乘子**
- ν_i为第i个等式约束$h_i(x) = 0$对应的**Lagrange乘子**

Lagrange对偶函数

- Lagrange对偶函数： $g: R^m \times R^p \rightarrow R$

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

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

- 对偶函数的性质：
  - ➢对偶函数为凹函数（一族关于(λ,v)的仿射函数的逐点下确界）
  - ➢$\forall\lambda \succcurlyeq 0, \forall v, g(\lambda,v) \le p^*$，即对偶函数构成了原问题最优值的下界

证明：如果$\tilde{x}$是可行的并且$\lambda \succcurlyeq 0$,则$f_0(\tilde{x}) \ge L(\tilde{x},\lambda,\nu) \ge \inf_{x\in\mathcal{D}} L(x,\lambda,\nu) = g(\lambda,\nu)$

在所有可行的$\tilde{x}$中取最小值可得$p^* \ge g(\lambda,\nu)$

举例--线性方程组的最小二乘解

考虑如下问题：

$$\text{minimize } x^T x$$
$$\text{subject to } Ax = b$$

该问题的Lagrange函数是$L(x,\nu) = x^T x + \nu^T(Ax - b)$，对偶函数是$g(\nu) = \inf_x L(x,\nu)$

因为$L(x,\nu)$是x的二次凸函数，可以通过求解下式来得到函数的最小值

$$\nabla_x L(x,\nu) = 2x + A^T\nu = 0 \Rightarrow x =- (1/2)A^T\nu$$

此时对偶函数为$g(\nu) = L\big((-1/2)A^T\nu, \nu\big) =- \frac{1}{4}\nu^T AA^T\nu - b^T\nu$

这是一个关于ν的二次凹函数，对于任意的$\nu \in R^p$，有

$$P^* \geq - (1/4)\nu^T AA^T\nu - b^T\nu$$

举例--标准形式的线性规划

考虑如下问题：

$$\begin{aligned} &\text{minimize } c^T x \\ &\text{subject to } Ax = b \\ &x \succcurlyeq 0 \end{aligned}$$

该问题的Lagrange函数是$L(x,\nu,\lambda) = c^T x - \sum_{i=1}^{n} \lambda_i x_i + \nu^T(Ax - b) = -b^T\nu + (c + A^T\nu - \lambda)^T x$

对偶函数是$g(\lambda,\nu) = \inf_x L(x,\lambda,\nu) = -b^T\nu + \inf_x (c + A^T\nu - \lambda)^T x$

线性函数只有恒为零时才有下界，因此只有$c + A^T\nu - \lambda = 0$时$g(\lambda,\nu) = -b^T\nu$其余情况为$-\infty$:

$$g(\lambda,\nu) = \inf_x L(x,\lambda,\nu) = \begin{cases} -b^T\nu & c + A^T\nu - \lambda = 0 \\ -\infty & \text{其他} \end{cases}$$

举例—双向划分问题

考虑如下问题：

$$\text{minimize } x^T W x$$
$$\text{subject to } x_i^2 = 1,\ \ i = 1, .., n$$

该问题是一个非凸问题，可行集包含了2^n个点（$x_i = \pm 1$）

看成n个元素的集合上的双向划分问题：$\{1, \ldots, n\} = \{i | x_i = -1\} \cup \{i | x_i = 1\}$

矩阵系数W_{ij}看成分量i和j在同一个分区内的成本，-W_{ij}看成分量i和j不在同一个分区内的成本，问题转变为考虑分量间所有配对的成本

$$g(\nu) = \inf_x \left(x^T W x + \sum_i \nu_i (x^2 - 1)\right) = \inf_x x^T \big(W + diag(\nu)\big) x - 1^T \nu$$

$$= \begin{cases} -1^T \nu & W + diag(\nu) \succcurlyeq 0 \\ -\infty & \text{其他} \end{cases}$$

对偶问题

从Lagrange函数能够得到的最好下界是什么?

$$\text{maximize } g(\lambda, \nu)$$
$$\text{subject to } \lambda \succcurlyeq 0$$

上述问题被称为**Lagrange对偶问题**，该问题是一个凸优化问题，最优值记为d^*，满足$\lambda \succcurlyeq 0$和$g(\lambda,\nu) > -\infty$的(λ,ν)称为对偶可行解。

举例：标准形式的线性规划及其对偶

$$\begin{array}{c} \text{mimimize } c^T x \\ \text{subject to } Ax = b \\ x \succcurlyeq 0 \end{array}$$

$$\begin{array}{ll} \text{maximize} & g(\lambda,\nu) = \begin{cases} -b^T\nu & A^T\nu - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases} \\ \text{subject to} & \lambda \succeq 0. \end{array}$$

$$\begin{array}{ll} \text{maximize} & -b^T\nu \\ \text{subject to} & A^T\nu - \lambda + c = 0 \\ & \lambda \succeq 0. \end{array}$$

$$\longrightarrow$$

$$\begin{array}{c} \text{maximize } -b^T\nu \\ \text{subject to } A^T\nu + c \succcurlyeq 0 \end{array}$$

强弱对偶

- **弱对偶性：** $d^* \leq p^*$

对于所有的凸问题和非凸问题来说始终成立，对于p、d无限时也成立

最优对偶间隙 $p^* - d^*$

当原问题很难求解时，弱对偶问题可以给出原版问题最优值的一个下界，如双向划分问题，其对偶问题是一个半正定规划：

$$\text{maximize } -1^T\nu$$
$$\text{subject to } W + diag(\nu) \succcurlyeq 0$$

强对偶性： $d^* = p^*$

不是普遍成立；对于凸问题，通常（但不总是）成立

在凸问题中，强对偶性成立的条件，称为**约束准则**

Slater约束准则

- 对于一个凸问题：

$$\text{minimize } f_0(x)$$
$$\text{subject to } f_i(x) \le 0,\ \ i = 1, \dots, m$$
$$Ax = b$$

存在一点$x \in relint\,\mathcal{D}$，使得下式成立：

$$f_i(x) < 0, i = 1, \dots, m, \quad Ax = b$$

- 满足上述条件的点称为**严格可行**，当Slater条件成立时，强对偶性成立
- 可以进一步改进，如果前面的k个约束函数$f_1, \dots, f_k$是仿射的，则条件可以弱化：

$$f_i(x) \le 0, i = 1, \dots, k,\ \ f_i(x) < 0, i = k+1, \dots, m,\ \ Ax = b$$

即仿射不等式不需要严格成立

举例—不等式形式的线性规划

- 线性规划问题：

$$\text{minimize } c^T x$$
$$\text{subject to } Ax \preccurlyeq b$$

对偶函数：

$$g(\lambda) = \inf_x ((c + A^T\lambda)^T x - b^T\lambda) = \begin{cases} -b^T\lambda, & A^T\lambda + c = 0 \\ -\infty, & \text{其他} \end{cases}$$

对偶问题

$$\text{maximize } -b^T\lambda$$
$$\text{subject to } A^T\lambda + c = 0,\ \lambda \succcurlyeq 0$$

- 从Slater条件可知：如果存在$\tilde{x}$使得$A\tilde{x} \prec b$，则 $p^* = d^*$
- 实际上，只要原问题和对偶问题可行， $p^* = d^*$

举例—二次规划

- 二次规划问题:

$$\text{minimize } x^T P x$$
$$\text{subject to } Ax \preccurlyeq b$$

对偶函数:

$$g(\lambda) = \inf_x (x^T P x + \lambda^T (Ax - b)) = -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

对偶问题

$$\text{maximize } -(1/4)\lambda^T A P^{-1} A^T \lambda - b^T \lambda$$
$$\text{subject to } \lambda \succcurlyeq 0$$

- 从Slater条件可知：如果存在$\tilde{x}$使得$A\tilde{x} \prec b$，则 $p^* = d^*$
- 实际上，$p^* = d^*$一直成立

互补松弛性

假设原问题强对偶性成立，令x^*是原问题的最优解，(λ^*, ν^*)是对偶问题的最优解

$$
\begin{aligned}
f_0(x^*) = g(\lambda^*, \nu^*) &= \inf_x \left(f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right) \\
&\le f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*) \\
&\le f_0(x^*)
\end{aligned}
$$

因此，上述两个不等式均可以取等号，从中可以看出：

- x^*是Lagrange函数$L(x, \lambda^*, \nu^*)$的一个最小点
- 对于$i = 1, \ldots, m$存在$\lambda_i^* f_i(x^*) = 0$，进一步可以推导出：

$$\lambda_i^* > 0 \Rightarrow f_i(x^*) = 0,\ \ f_i(x^*) < 0 \Rightarrow \lambda_i^* = 0$$

KKT最优性条件

以下四个条件称为KKT条件

1. 原问题可行性：$f_i(x^*) \le 0, i = 1, \dots, m, h_i(x^*) = 0, i = 1, \dots, p$
2. 对偶问题可行性：$\lambda \ge 0$
3. 互补松弛性：$\lambda_i f_i(x^*) = 0, i = 1, \dots, m$
4. 极值条件：$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i \nabla h_i(x^*) = 0$

对于目标函数和约束函数可微的任意优化问题，如果强对偶性存在，则任何一对原问题最优解和对偶问题最优解必须满足KKT条件

凸问题的KKT条件

当原问题是凸问题时，满足KKT条件的点也是原、对偶的最优解：

- 可行性条件说明可行解
- 从互补松弛性：$f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
- 从第四条条件和凸性：$g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

$$\begin{aligned} g(\tilde{\lambda}, \tilde{\nu}) &= L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \\ &= f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\tilde{x}) \\ &= f_0(\tilde{x}), \end{aligned}$$

因此，$f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$，即最优对偶间隙为0

- 若某个凸优化问题具有可微的目标函数和约束函数，且满足Slater条件，则KKT条件是其最优性的充要条件

举例：注水（1）

考虑如下优化问题：

$$\text{minimize } -\sum_{i=1}^{n} \log(x_i + \alpha_i)$$
$$\text{subject to } x \succcurlyeq 0,\ \ 1^T x = 1$$

- 在信息论中的意义：将值为1的总功率分配给不同的信道，使得总的通信功率最大

对不等式变量$x^* \succcurlyeq 0$引入Lagrange乘子$\lambda^* \in R^n$，对等式约束$1^T x = 1$引入乘子$\nu^* \in R$，通过KKT条件得到：

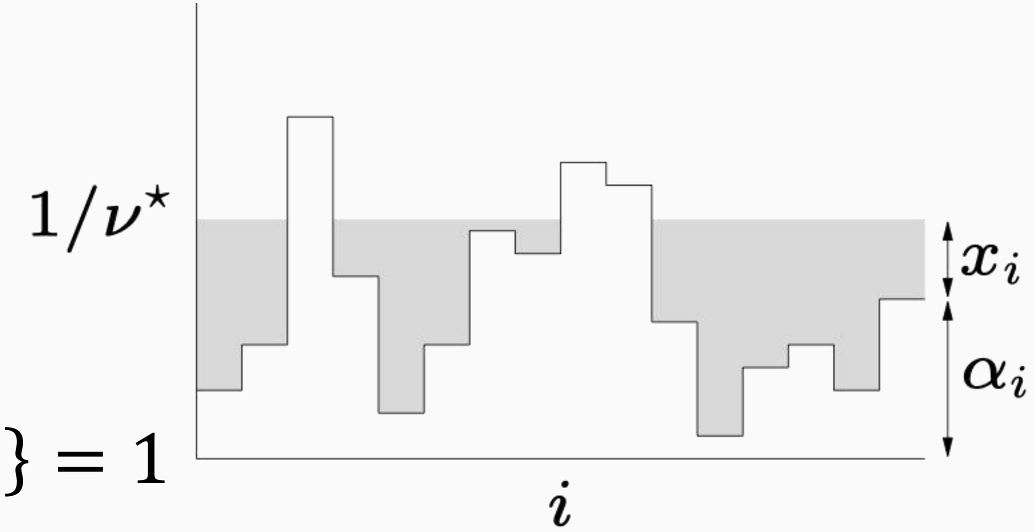
$$x^* \succcurlyeq 0,\ \ 1^T x^* = 1,\ \ \lambda^* \succcurlyeq 0,\ \ \lambda_i^* x_i^* = 0,\ \ i = 1, \dots, n$$
$$\frac{1}{x_i^* + \alpha_i} + \lambda_i^* = \nu^*$$

举例：注水（2）

$$\text{minimize} \ -\sum_{i=1}^{n} \log(x_i + \alpha_i)$$
$$\text{subject to } x \succcurlyeq 0,\ \mathbf{1}^T x = 1$$

求解得:

- 当$\nu^* < 1/\alpha_i$时：有$x_i = \frac{1}{\nu} - \alpha_i$
- 当$\nu^* \geq 1/\alpha_i$时：有$x_i = 0$
- 确定ν的方法： $\mathbf{1}^T x = \sum_{i=1}^{n} \max\{0, 1/\nu - \alpha_i\} = 1$



解释:

我们可以将α_i看作第i片区域的水平线，然后对于整个区域进行注水，使其具有深度$1/\nu$，所需要的总水量为$\sum_{i=1}^{n} \max\{0, 1/\nu - \alpha_i\}$，不断注水，直到总水量为1，第$i$个区域的水位深度即为最优$x_i^*$.

对偶与问题重构

- 一个问题的等式约束不同可以导致对偶问题的截然不同
- 当一个问题的对偶问题难以解决时重构原问题可能会有奇效

常见的重构

- 引入新的变量和等式约束
- 将隐式约束变成显式约束，或者反之亦然，例如可以将显式约束并入目标函数的定义域
- 转换目标函数或者约束函数，如用原目标$f_0(x)$的增函数$\emptyset(f_0(x))$取代原目标函数

引入新的变量以及相应的等式约束(1)

考虑如下无约束问题：

$$\text{minimize } f_0(Ax+b)$$

其Lagrange 对偶函数是常数p^*，所以虽然强对偶性存在，但是其Lagrange对偶问题没什么意义

转换问题为下列优化问题：

$$\begin{aligned} &\text{minimize } f_0(y) \\ &\text{subject to } Ax+b=y \end{aligned}$$

则变换之后的Lagrange函数为：

$$L(x,y,\nu)=f_0(y)+\nu^T(Ax+b-y)$$

minimize $f_0(Ax+b)$

minimize $f_0(y)$
subject to $Ax+b=y$

引入新的变量以及相应的等式约束(2)

从中可以看出，除非$A^T\nu=0$，否则其对偶函数$g(\nu)=-\infty$，若$A^T\nu=0$，则有

$$g(\nu)=b^T\nu+\inf_y\left(f_0(y)-\nu^Ty\right)=b^T\nu-f_0^*(\nu)$$

其中f_0^*是f_0的共轭函数，则原问题的对偶问题可以转换为：

$$\begin{aligned}&\text{maximize } b^T\nu-f_0^*(\nu)\\&\text{subject to } A^T\nu=0\end{aligned}$$

可见，经过变换之后问题的对偶问题显然比原问题的对偶问题具有意义的多。

范数逼近问题(1)

考虑如下无约束范数逼近问题：

$$\text{minimize } \|Ax-b\|$$

其中$\|\cdot\|$代表任意范数，对于上述问题，其Lagrange对偶函数也是常数，对于此类问题进行变化：

$$\begin{aligned}&\text{minimize } \|y\| \\ &\text{subject to } Ax-b=y\end{aligned}$$

其对偶函数为：

$$g(\nu)=\inf_{x,y}\left(\|y\|+\nu^T y-\nu^T Ax+b^T\nu\right)$$

$$=\begin{cases} b^T\nu+\inf\limits_{y}\left(\|y\|+\nu^T y\right) & A^T\nu=0 \\ -\infty & \text{其他}\end{cases}$$

范数逼近问题(2)

$$\text{minimize } \|Ax - b\|$$

$$\updownarrow$$

$$\begin{aligned} &\text{minimize } \|y\| \\ &\text{subject to } Ax - b = y \end{aligned}$$

$$g(\nu) = \begin{cases} b^T\nu & A^T\nu = 0,\ \|\nu\|_* \leq 1 \\ -\infty & \text{其他} \end{cases}$$

因此，其变换后的问题的Lagrange对偶问题为

$$\begin{aligned} &\text{maximize } b^T\nu \\ &\text{subject to } \ \|\nu\|_* \leq 1,\ \ A^T\nu = 0 \end{aligned}$$

隐式约束(1)

考虑一下具有框约束的线性规划问题：

$$\begin{aligned} &\text{minimize } c^T x \\ &\text{subject to} \quad Ax = b,\ \ l \preccurlyeq x \preccurlyeq u \end{aligned}$$

易得该问题的对偶问题为：

$$\begin{aligned} &\text{maximize } \ -b^T\nu - \lambda_1^T u + \lambda_2^T l \\ &\text{subject to } A^T\nu + \lambda_1 - \lambda_2 + c = 0 \\ &\lambda_1 \succcurlyeq 0,\ \ \lambda_2 \succcurlyeq 0 \end{aligned}$$

将原问题进行重构：

$$\begin{aligned} &\text{minimize } f_0(x) = \begin{cases} c^T x, & l \preccurlyeq x \preccurlyeq u \\ \infty, & \text{其他} \end{cases} \\ &\text{subject to} \quad Ax = b \end{aligned}$$

隐式约束(2)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b,\ l \preccurlyeq x \preccurlyeq u \end{array} \longleftrightarrow \begin{array}{ll} \text{minimize} & f_0(x) = \begin{cases} c^T x, & l \preccurlyeq x \preccurlyeq u \\ \infty, & \text{其他} \end{cases} \\ \text{subject to} & Ax = b \end{array}$$

经过转换之后，该问题的对偶函数为：

$$g(\nu) = \inf_{l \preccurlyeq x \preccurlyeq u} (c^T x + \nu^T (Ax - b))$$

$$= -b^T \nu - u^T (A^T \nu + c)^- + l^T (A^T \nu + c)^+$$

其中，$y_i^+ = \max\{y_i, 0\}$，$y_i^- = \max\{-y_i, 0\}$,可以得到g的解析表达式，它是一个凹的分片线性函数。

新问题的对偶问题是一个无约束问题：

$$\text{maximize}\ -b^T \nu - u^T (A^T \nu + c)^- + l^T (A^T \nu + c)^+$$