对偶理论:一般的约束优化问题

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x), \\ \text{s.t.} \quad & c_i(x) \le 0,\ i \in \mathcal{I}, \\ & c_i(x) = 0,\ i \in \mathcal{E}, \end{aligned}$$

其中c_i 为定义在$\mathbb{R}^n$ 或其子集上的实值函数, $\mathcal{I}$ 和$\mathcal{E}$ 分别表示不等式约束和等式约束对应的下标集合且各下标互不相同.

- 这个问题的可行域定义为

$$\mathcal{X} = \{x \in \mathbb{R}^n \mid c_i(x) \le 0,\ i \in \mathcal{I} \text{ 且 } c_i(x) = 0,\ i \in \mathcal{E}\}.$$

- 通过将$\mathcal{X}$ 的示性函数加到目标函数中可以得到无约束优化问题.但是转化后问题的目标函数是不连续的、不可微的以及不是有限的,这导致我们难以分析其理论性质以及设计有效的算法.

拉格朗日函数

一般的约束优化问题:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & c_i(x) \le 0,\ i \in \mathcal{I},\ |\mathcal{I}| = m \\ & c_i(x) = 0,\ i \in \mathcal{E},\ |\mathcal{E}| = p \end{aligned}$$

变量$x \in \mathbb{R}^n$, 最优值为p^*, 定义域为

$$\mathcal{X} = \{x \in \mathbb{R}^n \mid c_i(x) \le 0,\ i \in \mathcal{I} \text{ 且 } c_i(x) = 0,\ i \in \mathcal{E}\}$$

拉格朗日函数$L : \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^p \ \to\ \mathbb{R}$

$$L(x, \lambda, \nu) = f(x) + \sum_{i \in \mathcal{I}} \lambda_i c_i(x) + \sum_{i \in \mathcal{E}} \nu_i c_i(x)$$

- λ_i 为第i 个不等式约束对应的拉格朗日乘子
- ν_i 为第i 个等式约束对应的拉格朗日乘子

拉格朗日对偶函数

拉格朗日对偶函数$g : \mathbb{R}^m_+ \times \mathbb{R}^p \to [-\infty, +\infty)$

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathbb{R}^n} L(x, \lambda, \nu) \\ &= \inf_{x \in \mathbb{R}^n} \left(f(x) + \sum_{i \in \mathcal{I}} \lambda_i c_i(x) + \sum_{i \in \mathcal{E}} \nu_i c_i(x) \right) \end{aligned}$$

定理 (弱对偶原理)

若$\lambda \geq 0$, 则$g(\lambda, \nu) \leq p^*$.

证明:若$\tilde{x} \in \mathcal{X}$, 则

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) \leq L(\tilde{x}, \lambda, \nu) \leq f(\tilde{x}),$$

对$\tilde{x}$取下界得

$$g(\lambda, \nu) \leq \inf_{\tilde{x} \in \mathcal{X}} f(\tilde{x}) = p^*.$$

拉格朗日对偶问题

拉格朗日对偶问题:

$$\max_{\lambda\geq 0,\nu} g(\lambda,\nu) = \max_{\lambda\geq 0,\nu} \inf_{x\in\mathbb{R}^n} L(x,\lambda,\nu)$$

- 称λ 和ν 为对偶变量,设最优值为q^*
- q^* 为p^* 的最优下界,称$p^* - q^*$ 为对偶间隙
- 拉格朗日对偶问题是一个凸优化问题
- $\mathrm{dom} g = \{(\lambda,\nu) \mid \lambda \geq 0, g(\lambda,\nu) > -\infty\}$,称其元素为对偶可行解

例: 标准形式线性规划及其对偶

$$\begin{array}{lll} \min & c^Tx \\ \text{s.t.} & Ax = b \\ & x \geq 0 \end{array} \qquad\qquad \begin{array}{ll} \max & -b^T\nu \\ \text{s.t.} & A^T\nu + c \geq 0 \end{array}$$

实例:线性方程组具有最小模的解

$$\begin{array}{ll} \min & x^T x \\ \text{s.t.} & Ax = b \end{array}$$

对偶函数

- 拉格朗日函数为$L(x,\nu) = x^T x + \nu^T(Ax - b)$
- 求L 关于x 的最小值, 由一阶条件:

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

- 将上式代入L 得到对偶函数g:

$$g(\nu) = L((-1/2)A^T\nu, \nu) = -\frac{1}{4}\nu^T A A^T \nu - b^T\nu$$

它是关于ν 的凹函数

弱对偶性: $p^* \geq -(1/4)\nu^T A A^T \nu - b^T\nu, \ \forall \nu$

实例:标准形式线性规划

$$
\begin{aligned}
\min \quad & c^T x \\
\text{s.t.} \quad & Ax = b, \quad x \geq 0
\end{aligned}
$$

对偶函数

- 拉格朗日函数为

$$
\begin{aligned}
L(x, \lambda, \nu) &= c^T x + \nu^T (Ax - b) - \lambda^T x \\
&= -b^T \nu + (c + A^T \nu - \lambda)^T x
\end{aligned}
$$

- L 关于x 线性, 因此

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

g 在仿射集$\{(\lambda, \nu) \mid A^T \nu - \lambda + c = 0\}$ 上线性, 因此是凹函数

弱对偶性: $p^* \geq -b^T \nu$ 若$A^T \nu + c \geq 0$

线性规划问题的对偶

$$\begin{aligned}\min_{x} \quad & c^{\mathrm{T}}x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0\end{aligned}$$

- 拉格朗日函数:

$$L(x, s, \nu) = c^{\mathrm{T}}x + \nu^{\mathrm{T}}(Ax - b) - s^{\mathrm{T}}x = -b^{\mathrm{T}}\nu + (A^{\mathrm{T}}\nu - s + c)^{\mathrm{T}}x$$

- 对偶函数:

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

- 对偶问题:

$$\begin{aligned}\max_{s,\nu} \quad & -b^{\mathrm{T}}\nu, \\ \text{s.t.} \quad & A^{\mathrm{T}}\nu - s + c = 0, \\ & s \geq 0.\end{aligned} \quad \overset{y=-\nu}{\Longleftrightarrow} \quad \begin{aligned}\max_{s,y} \quad & b^{\mathrm{T}}y, \\ \text{s.t.} \quad & A^{\mathrm{T}}y + s = c, \\ & s \geq 0.\end{aligned}$$

- 若保留约束$x \geq 0$, 则拉格朗日函数为

$$L(x,y) = c^{\mathrm{T}}x - y^{\mathrm{T}}(Ax-b) = b^{\mathrm{T}}y + (c - A^{\mathrm{T}}y)^{\mathrm{T}}x.$$

- 对偶问题需要将$x \geq 0$ 添加到约束里：

$$\max_{y} \left\{ \inf_{x} \; b^{\mathrm{T}}y + (c - A^{\mathrm{T}}y)^{\mathrm{T}}x, \quad \text{s.t.} \;\; x \geq 0 \right\} \Rightarrow \begin{array}{ll} \max\limits_{y} & b^{\mathrm{T}}y, \\ \text{s.t.} & A^{\mathrm{T}}y \leq c. \end{array}$$

  此对偶问题可以通过将上页最后一个问题中的变量s 消去得到.

- 视$\max \; b^{\mathrm{T}}y$ 为$\min \; -b^{\mathrm{T}}y$, 对偶问题的拉格朗日函数为

$$L(y,x) = -b^{\mathrm{T}}y + x^{\mathrm{T}}(A^{\mathrm{T}}y - c) = -c^{\mathrm{T}}x + (Ax-b)^{\mathrm{T}}y.$$

线性规划问题的对偶

- 因此得到对偶函数

$$g(x)=\inf_{y} L(y,x)=\begin{cases}-c^{\mathrm{T}}x, & Ax=b,\\ -\infty, & \text{其他}.\end{cases}$$

- 相应的对偶问题是

$$\begin{aligned}\max_{x}\quad & -c^{\mathrm{T}}x,\\ \text{s.t.}\quad & Ax=b,\\ & x\geq 0.\end{aligned}$$

该问题与原始问题完全等价，这表明线性规划问题与其对偶问题互为对偶.

$$\begin{array}{ll} \min & \|x\| \\ \text{s.t.} & Ax = b \end{array}$$

对偶函数

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

其中 $\|v\|_* = \sup_{\|u\| \leq 1} u^T v$ 是 $\|\cdot\|$ 的对偶范数

证明: 利用 $\inf_x(\|x\| - y^T x)$ 在 $\|y\|_* \leq 1$ 时等于0 否则等于 $-\infty$

- 若 $\|y\|_* \leq 1$, 则 $x - y^T x \geq 0$ 对任意 x 都成立, 当 $x = 0$ 时取等
- 若 $\|y\|_* > 1$, 取 $x = tu$, 其中 $\|u\| \leq 1$, $u^T y = \|y\|_* > 1$:

$$\|x\| - y^T x = t(\|u\| - \|y\|_*) \to -\infty \quad \text{当} t \to \infty$$

弱对偶性: $p^* \geq b^T \nu$ 若 $\|A^T \nu\|_* \leq 1$

拉格朗日对偶与共轭函数

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & Ax \le b, \quad Cx = d \end{array}$$

对偶函数

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathrm{dom}\, f_0} \left(f_0(x) + (A^T\lambda + C^T\nu)^T x - b^T\lambda - d^T\nu \right) \\ &= -f_0^*(-A^T\lambda - C^T\nu) - b^T\lambda - d^T\nu \end{aligned}$$

- 回顾共轭函数的定义$f^*(y) = \sup_{x \in \mathrm{dom}\, f}(y^T x - f(x))$
- 在f_0 的共轭函数已知时可以简化对偶函数的推导

例：最大化熵

$$f_0(x) = \sum_{i=1}^n x_i \log x_i, \qquad f_0^*(y) = \sum_{i=1}^n e^{y_i - 1}$$

弱对偶性与强对偶性

弱对偶性: $d^* \le p^*$

- 对凸问题与非凸问题都成立
- 可导出复杂问题的非平凡下界, 例如, SDP问题

$$\begin{array}{ll} \max & -\mathbf{1}^T\nu \\ \text{s.t.} & W + \mathrm{diag}(\nu) \succeq 0 \end{array}$$

给出了二路划分问题的一个下界:

$$\min\ x^TWx, \quad \text{s.t. } x_i^2 = 1, \quad i = 1, \dots, n$$

强对偶性: $d^* = p^*$

- 对一般问题而言通常不成立
- (通常) 对凸问题成立
- 称保证凸问题强对偶性成立的条件为**约束品性**

问题形式与对偶性

- 一个问题不同等价形式的对偶可能差异巨大
- 当对偶问题难以推导或没有价值时，可以尝试改写原问题的形式

常用的改写技巧

- 引入新变量与等式约束
- 将显式约束隐式化或将隐式约束显式化
- 改变目标函数或者约束函数的形式
  例如，用 $\phi(f_0(x))$ 取代 $f_0(x)$，其中 ϕ 是凸的增函数

引入新变量与等式约束

$$\min \quad f_0(Ax+b)$$

- 对偶函数为常数
- 强对偶性成立, 但对偶问题无意义

改写原问题及其对偶

$$\begin{array}{ll} \min & f_0(y) \\ \text{s.t.} & Ax+b-y=0 \end{array} \qquad \begin{array}{ll} \max & b^T\nu - f_0^*(\nu) \\ \text{s.t.} & A^T\nu = 0 \end{array}$$

对偶函数为

$$\begin{aligned} g(\nu) &= \inf_{x,y}(f_0(y) - \nu^T y + \nu^T Ax + b^T\nu) \\ &= \begin{cases} -f_0^*(\nu) + b^T\nu & A^T\nu = 0 \\ -\infty & \text{其他} \end{cases} \end{aligned}$$

范数逼近问题: $\min \|Ax-b\|$

$$\begin{aligned} \min \quad & \|y\| \\ \text{s.t.} \quad & y = Ax - b \end{aligned}$$

由$\|\cdot\|$ 的共轭函数知其对偶函数为:

$$\begin{aligned} g(\nu) &= \inf_{x,y}(\|y\| + \nu^T y - \nu^T Ax + b^T\nu) \\ &= \begin{cases} b^T\nu + \inf_y(\|y\| + \nu^T y) & A^T\nu = 0 \\ -\infty & \text{其他} \end{cases} \\ &= \begin{cases} b^T\nu & A^T\nu = 0, \quad \|\nu\|_* \le 1 \\ -\infty & \text{其他} \end{cases} \end{aligned}$$

范数逼近问题的对偶

$$\begin{aligned} \max \quad & b^T\nu \\ \text{s.t.} \quad & A^T\nu = 0, \quad \|\nu\|_* \le 1 \end{aligned}$$

带边界约束的线性规划: 原问题与对偶问题

$$
\begin{array}{ll}
\min & c^T x \\
\text{s.t.} & Ax = b \\
& -\mathbf{1} \le x \le \mathbf{1}
\end{array}
\qquad
\begin{array}{ll}
\max & -b^T \nu - \mathbf{1}^T \lambda_1 - \mathbf{1}^T \lambda_2 \\
\text{s.t.} & c + A^T \nu + \lambda_1 - \lambda_2 = 0 \\
& \lambda_1 \ge 0, \quad \lambda_2 \ge 0
\end{array}
$$

通过隐式化边界约束改写原问题

$$
\begin{array}{ll}
\min & f_0(x) = \begin{cases} c^T x & -\mathbf{1} \le x \le \mathbf{1} \\ -\infty & \text{其他} \end{cases} \\
\text{s.t.} & Ax = b
\end{array}
$$

对偶函数为

$$
\begin{aligned}
g(\nu) &= \inf_{-\mathbf{1} \le x \le \mathbf{1}} (c^T x + \nu^T (Ax - b)) \\
&= -b^T \nu - \|A^T \nu + c\|_1
\end{aligned}
$$

对偶问题: $\max \ -b^T \nu - \|A^T \nu + c\|_1$

带广义不等式约束优化问题

- 问题

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x), \\ \text{s.t.} \quad & c_i(x) \le 0, \ i \in \mathcal{I}, \\ & c_i(x) = 0, \ i \in \mathcal{E}, \end{aligned}$$

  中的不等式约束$c_i(x)$, $i \in \mathcal{I}$ 都是实值函数的形式.

- 在许多实际应用中, 我们还会遇到大量带广义不等式约束的优化问题, 例如自变量x 可能取值于半正定矩阵空间中.

- 对于这类约束我们不易将其化为$c_i(x) \le 0$ 的形式, 此时又该如何构造拉格朗日对偶函数呢?

适当锥与广义不等式

定义 (适当锥)

称满足如下条件的锥K 为适当锥(proper cone)：

1. K 是凸锥；
2. K 是闭集；
3. K 是实心的(solid), 即$\mathbf{int}\, K \neq \emptyset$ ；
4. K 是尖的(pointed), 即对任意非零向量x, 若$x \in K$, 则$-x \notin K$, 也即K 中无法容纳直线.

- 适当锥K 可以诱导出广义不等式, 它定义了全空间上的偏序关系：
$$x \preceq_K y \iff y - x \in K.$$
- 类似地, 可以定义严格广义不等式：
$$x \prec_K y \iff y - x \in \mathbf{int}\, K.$$
- 当$K = \mathbb{R}^n_+$ 时, $x \preceq_K y$ 是我们之前经常使用的记号$x \leq y$
- 当$K = \mathcal{S}^n_+$ 时, $X \preceq_K Y$ 表示$Y - X \succeq 0$, 即$Y - X$ 是半正定矩阵

对偶锥与拉格朗日乘子

- 对广义不等式，该如何构造广义不等式约束所对应的乘子？

定义 (对偶锥)

令K 为全空间Ω 的子集，称集合

$$K^* = \{y \in \Omega \mid \langle x, y\rangle \geq 0,\ \forall x \in K\}$$

为其对偶锥.

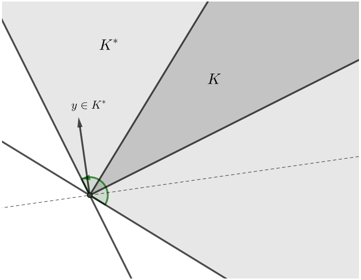

Figure: $\mathbb{R}^2$ 平面上的锥K 及其对偶锥K^*

对偶锥与拉格朗日乘子:注记

- 如果$K=\mathbb{R}^n_+, \Omega=\mathbb{R}^n$ 并且定义$\langle x,y\rangle = x^{\mathrm{T}}y$, 那么易知$K^*=\mathbb{R}^n_+$.
- 假设$K=\mathcal{S}^n_+, \Omega=\mathcal{S}^n$ 并且定义
$$\langle X,Y\rangle = \mathrm{Tr}(XY^{\mathrm{T}}),$$
可以证明
$$\langle X,Y\rangle \geq 0,\ \forall X\in\mathcal{S}^n_+ \quad \Longleftrightarrow \quad Y\in\mathcal{S}^n_+,$$
即半正定锥的对偶锥仍为半正定锥.
- 称满足$K=K^*$ 的锥K 为**自对偶锥**, 因此非负锥和半正定锥都是自对偶锥.
- 直观来说, 对偶锥K^* 中向量和原锥K 中向量的内积恒非负, 这一性质可以被用来构造拉格朗日对偶函数.

广义不等式约束优化问题拉格朗日函数的构造

- 广义不等式约束优化问题:

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & c_i(x) \preceq_{K_i} 0,\ i\in\mathcal{I} \\ & c_i(x) = 0,\ i\in\mathcal{E} \end{aligned}$$

  其中$c_i:\mathbb{R}^n\to\mathbb{R}^{k_i},\ k_i\in\mathbb{N}_+, i\in\mathcal{I}$ 为向量值函数, f 与$c_i, i\in\mathcal{E}$ 为实值函数, $K_i, i\in\mathcal{I}$ 为适当锥.

- 拉格朗日函数L :

$$L(x,\lambda,\nu) = f(x) + \sum_{i\in\mathcal{I}}\langle c_i(x),\lambda_i\rangle + \sum_{i\in\mathcal{E}}\nu_i c_i(x), \quad \lambda_i\in K_i^*, \nu_i\in\mathbb{R}.$$

- 容易验证$L(x,\lambda,\nu)\le f(x),\ \forall\, x\in\mathcal{X},\ \lambda_i\in K_i^*, \nu_i\in\mathbb{R}$.
- 对偶函数$g(\lambda,\nu) = \inf_{x\in\mathbb{R}^n} L(x,\lambda,\nu)$, 对偶问题为

$$\max_{\lambda_i\in K_i^*,\ \nu_i\in\mathbb{R}} \quad g(\lambda,\nu).$$

半定规划问题的对偶问题

$$\begin{aligned} \min_{X \in \mathcal{S}^n} \quad & \langle C, X \rangle \\ \text{s.t.} \quad & \langle A_i, X \rangle = b_i, \ i = 1, 2, \cdots, m \\ & X \succeq 0 \end{aligned}$$

其中$A_i \in \mathcal{S}^n$, $i = 1, 2, \cdots, m$, $C \in \mathcal{S}^n, b \in \mathbb{R}^m$

- 拉格朗日函数:

$$L(X, y, S) = \langle C, X \rangle - \sum_{i=1}^{m} y_i(\langle A_i, X \rangle - b_i) - \langle S, X \rangle, \quad S \succeq 0$$

- 对偶函数:

$$g(y, S) = \inf_X \ L(X, y, S) = \begin{cases} b^{\mathrm{T}} y, & \sum_{i=1}^m y_i A_i - C + S = 0 \\ -\infty, & \text{其他} \end{cases}$$

- 对偶问题:

$$\min_{y \in \mathbb{R}^m} \quad -b^{\mathrm{T}} y, \quad \text{s.t.} \quad \sum_{i=1}^{m} y_i A_i - C + S = 0, \ S \succeq 0$$

半定规划对偶问题的对偶问题

$$\min_{y\in\mathbb{R}^m} \quad -b^{\mathrm{T}}y, \quad \text{s.t.} \quad \sum_{i=1}^m y_iA_i \preceq C$$

- 拉格朗日函数:

$$\begin{aligned} L(y,X) &= -b^{\mathrm{T}}y + \langle X, \sum_{i=1}^m y_iA_i - C\rangle \\ &= \sum_{i=1}^m y_i(-b_i + \langle A_i, X\rangle) - \langle C, X\rangle \end{aligned}$$

- 对偶函数:

$$g(X) = \inf_y L(y,X) = \begin{cases} -\langle C,X\rangle, & \langle A_i, X\rangle = b_i,\ i=1,2,\cdots,m \\ -\infty, & \text{其他} \end{cases}$$

- 对偶问题:

$$\min_{X\in\mathcal{S}^n} \quad \langle C,X\rangle, \quad \text{s.t.} \quad \langle A_i, X\rangle = b_i,\ i=1,2,\cdots,m,\ X\succeq 0$$

ℓ_1 正则化问题的对偶

$$\min_{x\in\mathbb{R}^n} \quad \frac{1}{2}\|Ax-b\|^2+\mu\|x\|_1$$

令$r=Ax-b$, 问题等价于$\min_{x\in\mathbb{R}^n}\frac{1}{2}\|r\|^2+\mu\|x\|_1$, s.t. $r=Ax-b$

- 拉格朗日函数:

$$\begin{aligned}L(x,r,\lambda)&=\frac{1}{2}\|r\|^2+\mu\|x\|_1-\langle\lambda,Ax-b-r\rangle\\&=\frac{1}{2}\|r\|^2+\lambda^{\mathrm{T}}r+\mu\|x\|_1-(A^{\mathrm{T}}\lambda)^{\mathrm{T}}x+b^{\mathrm{T}}\lambda\end{aligned}$$

- 对偶函数:

$$g(\lambda)=\inf_{x,r} L(x,r,\lambda)=\begin{cases}b^{\mathrm{T}}\lambda-\frac{1}{2}\|\lambda\|^2, & \|A^{\mathrm{T}}\lambda\|_\infty\le\mu\\ -\infty, & \text{其他}\end{cases}$$

- 对偶问题:

$$\max \quad b^{\mathrm{T}}\lambda-\frac{1}{2}\|\lambda\|^2, \quad \text{s.t.} \quad \|A^{\mathrm{T}}\lambda\|_\infty\le\mu$$