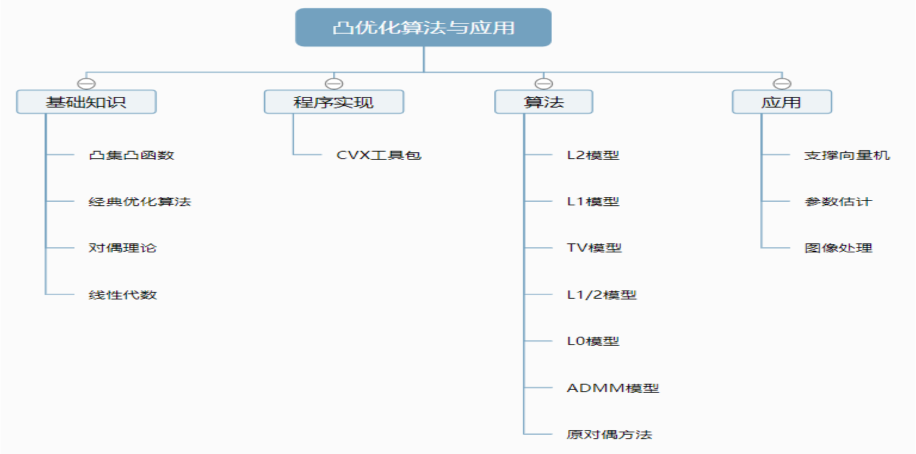
凸优化算法与应用

王俊刚

西北工业大学数学与统计学院
Email: wangjungang@nwpu.edu.cn

2021 年 9 月

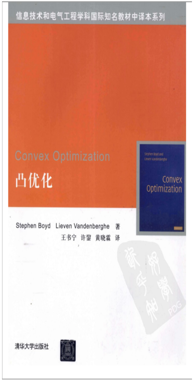
教学内容

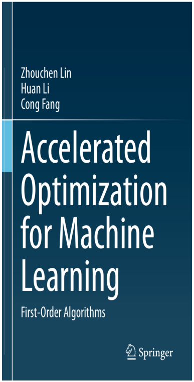


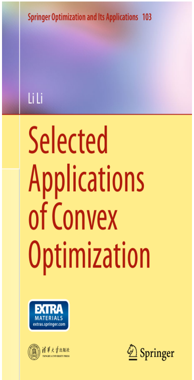
凸优化算法与应用
基础知识
程序实现
算法
应用
凸集凸函数
经典优化算法
对偶理论
线性代数
CVX工具包
L2模型
L1模型
TV模型
L1/2模型
L0模型
ADMM模型
原对偶方法
支撑向量机
参数估计
图像处理

参考书目


现代数学译丛 14
非线性最优化基础
[日] Masao Fukushima 著
林贵华 译
科学出版社


信息技术和电气工程学科国际知名教材中译本系列
Convex Optimization
凸优化
Convex Optimization
Stephen Boyd Lieven Vandenberghe 著
王书宁 许鋆 黄晓霖 译
清华大学出版社


Zhouchen Lin
Huan Li
Cong Fang
Accelerated Optimization for Machine Learning
First-Order Algorithms
Springer


Springer Optimization and Its Applications 103
Li Li
Selected Applications of Convex Optimization
EXTRA MATERIALS
清华大学出版社
Springer

成绩相关

- 30%：考勤+期中汇报ppt(凸优化结合自己的研究方向)
- 70%：期末大论文+编程实现

第一部分

凸优化基础

第一节 凸集与凸函数

定义 1.1

集合Ω 称为凸集，如果对所有的 $x_1, x_2 \in \Omega$ 和$\lambda \in [0,1]$, 有 $\lambda x_1 + (1-\lambda)x_2 \in \Omega$.

常见的保凸运算(参见参考书二)：

- 交集运算(包括有限交和无限交)、集合的和、笛卡尔乘积
- 仿射函数(形如 $f(x) = Ax + b$ 的映射)、分式线性函数 $f(x) = (Ax+b)/(c^T x + d)$ 等

如判断多面体是否凸时，经常把多面体表示成半空间的交.

凸函数

定义 1.2

设$\Omega \subset R^n$ 是一个凸集，称函数$f(x) : \Omega \mapsto R$是一个凸函数，如果对所有的$x, y \in \Omega$和$\lambda \in [0, 1]$有

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y). \tag{1.1}$$

上述不等式严格成立时，称 $f(x)$ 为严格凸函数.

凸函数

凸函数的等价定义有很多，等价的定义都可以用来作为判别方法.

- 当函数一阶可导时，刻画凸的方法是

$$f(y) \geq f(x) + \nabla f(x)^T (y - x), x, y \in \Omega. \tag{1.2}$$

- 当函数二阶可导时，判断凸的方法是

$$\nabla^2 f(x) \geq 0, x \in \Omega. \tag{1.3}$$

凹函数

定义 1.3

设 $\Omega \subset R^n$ 是一个凸集，称函数 $f(x) : \Omega \mapsto R$ 是一个凹函数，如果 $-f(x)$ 是一个凸函数，即对所有的 $x, y \in \Omega$ 和 $\lambda \in [0, 1]$ 有

$$f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y). \tag{1.4}$$

定理 1.1 (Jensen不等式)

设$\Omega \subset R^n$是一个凸集，$f(x): \Omega \mapsto R$是一个凸函数，则对所有的$x_1, \ldots, x_m \in \Omega, \lambda_i \geq 0, \sum_{i=1}^m \lambda_i = 1, m \in N$, 有下式成立

$$f(\lambda_1 x_1 + \cdots + \lambda_m x_m) \leq \lambda_1 f(x_1) + \cdots + \lambda_m f(x_m). \tag{1.5}$$

证明.

关于变量个数用数学归纳法即可. □

第二节 共轭函数和次微分

共轭函数

定义 2.1 (共轭函数或Fenchel变换)

称函数

$$f^*(\xi) = \sup_{x \in V} \langle \xi, x \rangle - f(x) \tag{2.1}$$

为 f 的共轭函数.

共轭函数由来

对于一个凸函数 f, 采用类似的策略,我们将其用仿射函数 $\langle \xi, x\rangle - \alpha$ 来表示.对于一个给定的斜率 $\xi \in V^*$, 我们只要去选取一个最好的 α即可.

$$\begin{aligned} & f(x) \ge \langle \xi, x\rangle - \alpha, \quad \forall x \in V \\ \Leftrightarrow & \alpha \ge \langle \xi, x\rangle - f(x), \quad \forall x \in V \\ \Leftrightarrow & \alpha \ge \sup_{x\in V} \{\langle \xi, x\rangle - f(x)\}. \end{aligned} \tag{2.2}$$

α 的最佳选择为

$$\alpha = f^*(\xi) = \sup_{x\in V} \langle \xi, x\rangle - f(x). \tag{2.3}$$

如果这个上确界是有限的, 则 $\langle \xi, x\rangle - f^*(\xi)$ 是 f 的斜率为 ξ的最佳仿射超平面. α 是截距的相反数.

共轭函数举例

例 1 (绝对值函数的共轭函数)

若$f(x)=|x|$,则

$$f^*(\xi)=\begin{cases}0, & |\xi|\le 1,\\ +\infty, & |\xi|>1.\end{cases}\tag{2.4}$$

证明.

$f^*(\xi)=\sup_x\{\langle\xi,x\rangle-|x|\}$.
当$|\xi|>1$时，取$x=t\xi$,并令$t\to+\infty$, 则$f^*(\xi)=+\infty$.
当$|\xi|\le 1$时, 因$\langle\xi,x\rangle-|x|\le 0$, 再取$x=0$即得结论. □

共轭函数举例

例 2 (欧氏范数的共轭函数)

若$f(x) = \|x\|$,则

$$f^*(\xi) = \begin{cases} 0, & \|\xi\|_* \le 1, \\ +\infty, & \|\xi\|_* > 1, \end{cases} \tag{2.5}$$

其中$\|\cdot\|_*$是原范数的共轭范数。

证明.

$f^*(\xi) = \sup_x\{\langle \xi, x\rangle - \|x\|\}$. 当$\|\xi\|_* > 1$时，存在 z 满足 $\|z\| \le 1, \langle z, \xi\rangle > 1$.
取$x = tz$, 并令$t \to +\infty$, 则$f^*(\xi) = +\infty$. 当
$\|\xi\|_* \le 1$时,因$\langle \xi, x\rangle - \|x\| \le \|\xi\|_*\|x\| - \|x\| \le 0$, 再取 $x = 0$ 即得结论. □

Legendre变换

定义 2.2 (Legendre变换)

设$f \in C^2(R^n, R)$,其导函数$\xi = \nabla f(x)$有逆函数为ψ, 记作$x = \psi(\xi)$. 称

$$f^*(\xi) = \langle \xi, x \rangle - f(x) = \langle \xi, \psi(\xi) \rangle - f \circ \psi(\xi) \tag{2.6}$$

为f的*Legendre*变换.

注 2.1

在共轭函数定义 $f^*(\xi) = \sup_{x \in V} \langle \xi, x \rangle - f(x)$ 中，如果 f 可导且右边的上确界取到,则 $\nabla f(\bar{x}) = \xi$,代入得到

$$f^*(\xi) = \langle \xi, \psi(\xi) \rangle - f \circ \psi(\xi). \tag{2.7}$$

因此前者是后者在凸函数情形的推广.

$$f^*(\xi) = \langle \xi, x \rangle - f(x) = \langle \xi, \psi(\xi) \rangle - f \circ \psi(\xi)$$

例 3

$f(x) = \frac{x^p}{p}, p > 1$. 因$f'(x) = x^{p-1} = \xi \Rightarrow x = \xi^{\frac{1}{p-1}}$, 从而

$$f^*(\xi) = \xi x - f(x) = \xi \xi^{\frac{1}{p-1}} - \frac{1}{p} \xi^{\frac{p}{p-1}} = \frac{1}{q} \xi^q,$$

其中$\frac{1}{p} + \frac{1}{q} = 1$.

Legendre变换举例

$$f^*(\xi) = \langle \xi, x \rangle - f(x) = \langle \xi, \psi(\xi) \rangle - f \circ \psi(\xi)$$

例 4 (二次型函数的共轭函数)

$f(x) = \frac{1}{2}x^TAx$. 因$f'(x) = Ax = \xi \Rightarrow x = A^{-1}\xi$, 从而

$$\begin{aligned} f^*(\xi) &= \langle \xi, x \rangle - f(x) \\ &= \langle \xi, A^{-1}\xi \rangle - \frac{1}{2}(A^{-1}\xi)^T A(A^{-1}\xi) \\ &= \frac{1}{2}\xi^T A^{-T}\xi. \end{aligned}$$

次微分

Legendre变换的问题在于要求函数梯度的逆处处存在. 凸函数不一定满足这个条件，因此需要将梯度做推广.

定义 2.3 (次微分)

满足

$$f(y) - f(x) \geq \langle \xi, y - x \rangle, y \in R^n \tag{2.8}$$

的向量$\xi \in R^n$称为凸函数f在点x处的次梯度. 次梯度的全体构成的集合记为$\partial f(x)$，称之为 f 在 x 处的次微分.

次微分的逆

注意到$x \mapsto \partial f(x)$是一个集值映射.

$$
\begin{aligned}
\xi \in \partial f(x) &\Leftrightarrow f(y) \geq f(x) + \langle \xi, y - x \rangle, \forall y \in D(f). \\
&\Leftrightarrow f(x) - \langle \xi, x \rangle = \min_{y \in D(f)} \{ f(y) - \langle \xi, y \rangle \}.
\end{aligned} \tag{2.9}
$$

其逆映射也是一个集值映射. 以$\psi(\xi)$表示$\partial f(x)$的逆映射，则

$$
\psi(\xi) = \{ x \in D(f) | \xi \in \partial f(x) \}. \tag{2.10}
$$

即，$x \in \psi(\xi) \Leftrightarrow x$使函数$f(y) - \langle \xi, y \rangle$达到极小值.

共轭函数的性质

下面介绍共轭函数的几个性质.

定理 2.1

1) *(Young不等式).* $\langle \xi, x\rangle \le f(x) + f^*(\xi)$.
2) $f(x) + f^*(\xi) = \langle x, \xi\rangle \Longleftrightarrow \xi \in \partial f(x)$.

证明.

1). 由共轭函数定义立得. 2) 由次微分的定义有

$$\begin{aligned}\xi \in \partial f(x) &\Longleftrightarrow \langle \xi, y - x\rangle \le f(y) - f(x), \forall y \in R^n \\ &\Longleftrightarrow \langle \xi, y\rangle - f(y) \le \langle \xi, x\rangle - f(x), \forall y \in R^n \\ &\Longleftrightarrow f^*(\xi) \le \langle \xi, x\rangle - f(x).\end{aligned} \tag{2.11}$$

□

定义 2.4

称$\mathrm{dom} f = \{x \in R^n | f(x) < +\infty\}$为$f$的有效域.当凸函数$f$满足 $\mathrm{dom} f \neq \emptyset$ 且对 $\forall x, f(x) > -\infty$ 时, 称之为正常凸函数 *(proper convex function)*.

定义 2.5

若对任意收敛于x的点列$\{x^k\} \subset R^n$均有

$$f(x) \leq \liminf_{k \to \infty} f(x^k), \tag{2.12}$$

则称 f 在 x 处下半连续*(lower semicontinuous)*.

共轭函数的性质

定理 2.2 (Fenchel-Moreau定理)

若f是一个正常的，下半连续的凸函数，则$f^{**} = f$.

该定理的条件用函数的图可以很直观地描述,即等价于函数的图像

$$\operatorname{epi} f = \{(x, t) \in \mathbf{R}^n \times \mathbf{R} | f(x) \leq t\}$$

是非空闭凸集.
由Fenchel-Moreau定理易得原函数的次微分和共轭函数的次微分之间的联系.

定理 2.3

对于一个正常的，下半连续的凸函数，我们有

$$\xi \in \partial f(x) \Longleftrightarrow x \in \partial f^*(\xi). \tag{2.13}$$

这个性质会用于后面原问题和对偶问题的相互转化.

示性函数与支撑函数

定义 2.6 (示性函数)

给定集合 $S \subseteq \mathbf{R}^n$, 由

$$\delta_S(\boldsymbol{x}) = \begin{cases} 0, & \boldsymbol{x} \in S \\ +\infty, & \boldsymbol{x} \notin S \end{cases}$$

定义的函数 $\delta_S : \mathbf{R}^n \to (-\infty, +\infty]$ 称为集合 S 的示性函数 *(indicator function)*.

定义 2.7 (支撑函数)

对于非空凸集 $S \subseteq \mathbf{R}^n$ 的示性函数 δ_S, 其共轭函数 $\delta_S^* : \mathbf{R}^n \to (-\infty, +\infty]$ 称为集合 S 的支撑函数 *(support function)*.

定理 2.4

示性函数与支撑函数具有如下性质.

1. 非空凸集 $S \subseteq \mathbf{R}^n$ 的示性函数 δ_S 为正常凸函数. 进一步, 若 S 为闭集, 则 δ_S 为闭正常凸函数.
2. 非空凸集 $S \subseteq \mathbf{R}^n$ 的支撑函数 δ_S^* 为正齐次闭正常凸函数，即 $\delta_S^*(\lambda \boldsymbol{y}) = \lambda \delta_S^*(\boldsymbol{y}), \lambda > 0$. 反之, 若函数 $f : \mathbf{R}^n \to (-\infty, +\infty]$ 为正齐次闭正常凸函数, 则 f 必为某非空闭凸集 $S \subseteq \mathbf{R}^n$ 的支撲函数.
3. 恒取有限值的正齐次凸函数 $f : \mathrm{R}^n \to \mathrm{R}$ 必为某非空紧凸集 $S \subseteq \mathbf{R}^n$ 的支撑函数.

支撑函数举例

1. 当 $S = \{x \in \mathbf{R} \mid -2 \leqslant x \leqslant 1\}$ 时, $\delta_S^*(y) = \begin{cases} y, & y \geqslant 0 \\ -2y, & y < 0 \end{cases}$.
2. 当 $S = \{x \in \mathbf{R} \mid 1 \leqslant x \leqslant 3\}$ 时, $\delta_S^*(y) = \begin{cases} 3y, & y \geqslant 0 \\ y, & y < 0 \end{cases}$.
3. 当 $S = \{x \in \mathbf{R} \mid x \geqslant 2\}$ 时, $\delta_S^*(y) = \begin{cases} +\infty, & y > 0 \\ 2y, & y \leqslant 0 \end{cases}$.
4. 当 $S = \{\boldsymbol{x} \in \mathbf{R}^n \mid \|\boldsymbol{x}\| \leqslant 1\}$ 时, $\delta_S^*(\boldsymbol{y}) = \|\boldsymbol{y}\|_*$.

共轭函数应用举例

下面的例子给出了上述概念和性质的一些应用.
当 F 是凸集 K上的示性函数时,我们有

$$
\begin{aligned}
& y = (I + \partial F)^{-1} x \\
& \Rightarrow x = y + \partial F(y) \\
& \Rightarrow 0 \in y - x + \partial F(y) \\
& \Rightarrow y = argmin_w \frac{\|w - x\|^2}{2} + F(w) = argmin_{w \in K} \frac{\|w - x\|^2}{2} = \Pi_K(x),
\end{aligned}
$$

其中 $\Pi_K(x)$ 表示 x 在 K 上的投影.

共轭函数应用举例

当 F^* 是凸集 K上的示性函数时,我们有

$$
\begin{aligned}
& y = (I + \partial F)^{-1} x \\
& \Rightarrow x \in y + \partial F(y) \\
& \Rightarrow x - y \in \partial F(y) \\
& \Rightarrow y \in \partial F^*(x - y) \\
& \Rightarrow 0 \in x - y - x + \partial F^*(x - y) \\
& \Rightarrow x - y = argmin_w \frac{\|w - x\|^2}{2} + F^*(w) = argmin_{w \in K} \frac{\|w - x\|^2}{2} = \Pi_K(x) \\
& \Rightarrow y = x - \Pi_K(x).
\end{aligned}
$$

共轭函数应用举例

在一般情形, 我们有如下的结论

$$
\begin{aligned}
& y = (I + \partial F)^{-1}(x) \\
& \Rightarrow x \in y + \partial F(y) \\
& \Rightarrow x - y \in \partial F(y) \\
& \Rightarrow y \in \partial F^*(x - y) \\
& \Rightarrow x \in x - y + \partial F^*(x - y) \\
& \Rightarrow x - y = (I + \partial F^*)^{-1}(x) \\
& \Rightarrow x = (I + \partial F)^{-1}(x) + (I + \partial F^*)^{-1}(x).
\end{aligned}
$$

上述公式告诉我们,只要有一个逆已知,另一个逆就知道了. 我们通常去求其中更容易的一个.

第三节 KKT 条件

考虑如下约束极小化问题

$$\begin{aligned} &\min_x f(x) \\ &\text{ s.t. } g_i(x) \leq 0, i = 1, \ldots, m. \end{aligned} \tag{3.1}$$

优化问题(3.1)的Lagrangian函数定义为

$$L(\boldsymbol{x}, \alpha) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \alpha_i g_i(\boldsymbol{x}), \tag{3.2}$$

其中$\alpha = [\alpha_1, \ldots, \alpha_m]^T \in \mathbb{R}^{m+}$称为对偶变量或Lagrange乘子.
Lagrange对偶函数定义为Lagrangian函数关于$\boldsymbol{x}$的极小值

$$q(\alpha) = \inf_{\boldsymbol{x} \in \Omega} L(\boldsymbol{x}, \alpha). \tag{3.3}$$

记原问题(3.1)的最优值为 $\bar{f}$.容易看出 $q(\alpha) \leq \bar{f}$. 我们称之为弱对偶性.

为了得到f^*的下界,我们考虑如下的Lagrange对偶问题

$$\begin{aligned} &\max_{\alpha} q(\alpha) \\ &\text{s.t.} \quad \alpha_i \geq 0, i = 1, \ldots, m. \end{aligned} \tag{3.4}$$

当原问题的最优值和对偶问题的最优值相等时,称之为强对偶. Karush-Kuhn-Tucker(KKT) 条件可用来刻画强对偶成立的准则.

定理 3.1 (KKT鞍点条件)

如果存在一对$(\bar{\boldsymbol{x}}, \bar{\alpha})$使下式

$$L(\bar{\boldsymbol{x}}, \alpha) \leq L(\bar{\boldsymbol{x}}, \bar{\alpha}) \leq L(\boldsymbol{x}, \bar{\alpha}) \tag{3.5}$$

对所有的$\boldsymbol{x} \in \Omega$和$\alpha \in \mathbb{R}^{n+}$成立,则$\bar{\boldsymbol{x}}$是原问题(3.1)的解.

证明.

设 $\bar{q}$ 为(3.4)的最优值,由 $q(\alpha) \leq \bar{f}$ 知 $\bar{q} \leq \bar{f}$. 又因为

$$f(\bar{\boldsymbol{x}}) = \sup_{\alpha \geq 0} L(\bar{\boldsymbol{x}}, \alpha) = L(\bar{\boldsymbol{x}}, \bar{\alpha}) = \min_{x} L(\boldsymbol{x}, \bar{\alpha}) = q(\bar{\alpha}).$$

从而有

$$\bar{f} = \min_{x} f(\boldsymbol{x}) \leq f(\bar{\boldsymbol{x}}) = q(\bar{\alpha}) \leq \bar{q}.$$

综合以上得到

$$f(\bar{\boldsymbol{x}}) = \bar{f} = \bar{q} = L(\bar{\boldsymbol{x}}, \bar{\alpha}).$$

□

定理 3.1 (KKT鞍点条件)

如果存在一对$(\bar{\boldsymbol{x}}, \bar{\alpha})$使下式

$$L(\bar{\boldsymbol{x}}, \alpha) \leq L(\bar{\boldsymbol{x}}, \bar{\alpha}) \leq L(\boldsymbol{x}, \bar{\alpha}) \tag{3.5}$$

对所有的$\boldsymbol{x} \in \Omega$和$\alpha \in \mathbb{R}^{n+}$成立,则$\bar{\boldsymbol{x}}$是原问题(3.1)的解.

从(3.5)还能看出

$$\bar{\alpha}_i g_i(\bar{\boldsymbol{x}}) = 0, i = 1, \ldots, m. \tag{3.6}$$

该条件称为互补性松弛(complementary slackness)条件.

下面考虑带有等式和不等式约束的优化问题

$$\begin{aligned} &\min_{x} f(\boldsymbol{x}) \\ &\text{s.t. } g_i(\boldsymbol{x}) \leq 0, i = 1, \ldots, m, \\ &h_j(\boldsymbol{x}) = 0, j = 1, \ldots, l. \end{aligned} \tag{3.7}$$

其Lagrangian函数定义为

$$L(\boldsymbol{x}, \alpha, \boldsymbol{\beta}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \alpha_i g_i(\boldsymbol{x}) + \sum_{j=1}^{l} \beta_j h_j(\boldsymbol{x}), \alpha_i \geq 0, \beta_j \in \mathbb{R}. \tag{3.8}$$

相应的KKT鞍点条件为

定理 3.2

如果存在$\left(\bar{\boldsymbol{x}}, \bar{\alpha}, \bar{\boldsymbol{\beta}}\right)$使下式

$$L\left(\bar{\boldsymbol{x}}, \alpha, \boldsymbol{\beta}\right) \leq L\left(\bar{\boldsymbol{x}}, \bar{\alpha}, \bar{\boldsymbol{\beta}}\right) \leq L\left(\boldsymbol{x}, \bar{\alpha}, \bar{\boldsymbol{\beta}}\right) \tag{3.9}$$

对所有的$\boldsymbol{x} \in \Omega, \boldsymbol{\alpha} \in \mathbb{R}^{n+}$, *and* $\boldsymbol{\beta} \in \mathbb{R}^{l}$成立,则$\bar{\boldsymbol{x}}$是问题(3.7)的解.

KKT条件是所有优化问题的充分条件,同时也是所有凸优化问题的必要条件.需要指出的是,对于等式约束,我们可以将其分解为两个不等式约束,从而当成不等式约束问题去处理.

当优化问题的目标函数和约束函数均可微时,我们可以直接利用KKT条件得到解.

定理 3.3 (可微凸优化问题的KKT条件)

如果问题(3.1)的目标函数和约束函数均可微,若存在$\bar{\boldsymbol{x}} \in \Omega, \bar{\alpha} \in \mathbb{R}^{m+}$使下式成立

$$\begin{aligned}
&\frac{\partial L(\bar{\boldsymbol{x}}, \bar{\alpha})}{\partial \boldsymbol{x}} = \frac{\partial f(\bar{\boldsymbol{x}})}{\partial \boldsymbol{x}} + \sum_{i=1}^{m} \bar{\alpha}_i \frac{\partial g_i(\bar{\boldsymbol{x}})}{\partial \boldsymbol{x}} = 0 \ (\textit{Saddle point at } \bar{\boldsymbol{x}}\ ), \\
&\frac{\partial L(\bar{\boldsymbol{x}}, \bar{\alpha})}{\partial \alpha_i} = g_i(\bar{\boldsymbol{x}}) \leq 0, i = 1, \ldots, m\,(\ \textit{Saddle point at } \bar{\alpha}), \\
&\bar{\alpha}_i g_i(\bar{\boldsymbol{x}}) = 0, i = 1, \ldots, m \ (\textit{Zero KKT-gap})\,.
\end{aligned} \tag{3.10}$$

则$\bar{\boldsymbol{x}}$ 是原问题的最优解.

设Y和Z分别是R^n, R^m的非空子集,给定$X\times Y$上的函数$L(x,y): X\times Y\to[-\infty,+\infty]$, 下面研究$\inf_x\sup_y L(x,y)$和$\sup_y\inf_x L(x,y)$的关系. 定义两个函数如下：

$$\begin{aligned} f(\boldsymbol{x}) &= \sup_{\boldsymbol{y}\in Y} L(\boldsymbol{x},\boldsymbol{y}), \\ g(y) &= \inf_{x\in X} L(x,y). \end{aligned} \tag{3.11}$$

下面研究问题这两个函数对应的优化问题

$$\min_{\boldsymbol{x}\in X} f(\boldsymbol{x}) \tag{3.12}$$

与

$$\max_{y\in Y} g(y) \tag{3.13}$$

之间的关系.

引理 3.1

对任意$x \in X, y \in Y$均有$g(y) \leq f(\boldsymbol{x})$成立. 进一步,还有

$$\sup_{y \in Y} g(y) \leqslant \inf_{x \in X} f(x). \tag{3.14}$$

证明.

由不等式

$$g(\boldsymbol{y}) \leqslant L(\boldsymbol{x}, \boldsymbol{y}) \leqslant f(\boldsymbol{x}) \tag{3.15}$$

直接得到结论. □

定义 3.1

若点$(\bar{x},\bar{y})\in X\times Y$满足

$$L(\bar{x},y)\leqslant L(\bar{x},\bar{y})\leqslant L(x,\bar{y}),\quad x\in X, y\in Y, \tag{3.16}$$

则称$(\bar{x},\bar{y})$为函数Φ的鞍点.

下面说明L的鞍点和问题(3.12),(3.13)最优解的等价性.

定理 3.4

点$(\bar{x},\bar{y})\in X\times Y$为函数$L(x,y): X\times Y\to[-\infty,+\infty]$的鞍点的充要条件是$\bar{x}\in X$与$\bar{y}\in Y$满足

$$f(\bar{x})=\inf_{x\in X} f(x)=\sup_{y\in Y} g(y)=g(\bar{y}). \tag{3.17}$$

证明.

先证必要性.设$(\bar{x},\bar{y})$为L的鞍点,则

$$f(\bar{x})=\sup_{y\in Y}L(\bar{x},y)=L(\bar{x},\bar{y})=\inf_{x\in X}L(x,\bar{y})=g(\bar{y}) \tag{3.18}$$

成立,因此

$$\inf_{\boldsymbol{x}\in X}f(\boldsymbol{x})\leqslant f(\overline{\boldsymbol{x}})=g(\overline{\boldsymbol{y}})\leqslant\sup_{y\in Y}g(\boldsymbol{y}). \tag{3.19}$$

而由引理3.1知

$$\inf_{\boldsymbol{x}\in X}f(\boldsymbol{x})\geqslant\sup_{y\in Y}g(\boldsymbol{y}). \tag{3.20}$$

从而结论成立. 下证充分性.由于

$$f(\bar{x})=\sup_{y\in Y}L(\bar{x},y)\geqslant L(\bar{x},\bar{y})\geqslant\inf_{x\in X}L(x,\bar{y})=g(\bar{y}). \tag{3.21}$$

从而由条件知

$$\sup_{y\in Y}L(\bar{x},y)=L(\bar{x},\bar{y})=\inf_{x\in X}L(x,\bar{y}). \tag{3.22}$$

这说明$(\bar{x},\bar{y})\in X\times Y$为函数$L(x,y):X\times Y\to[-\infty,+\infty]$的鞍点. □

第四节 对偶理论

对偶定理在线性规划中非常重要,在非线性规划中,对偶性理论在构造求解方法等方面也极为有用. 给定一个优化问题后,我们要得到其对偶问题, 需要先对原问题做一个扰动, 得到一个二元目标函数. 不同扰动形式会导出不同的对偶问题. 我们在写出问题不同的等价形式时实际上也给出了问题的不同扰动形式.
假设原始问题为

$$\min_x \Phi(x,0), \tag{4.1}$$

其中 $\Phi : X \times X \to R$ 是凸函数. 通过合适的选取 Φ, 标准的凸问题都可以写成该形式. 扰动问题为

$$\min_x \Phi(x,y), \tag{4.2}$$

其中 $y \in Y, y \neq 0$.
设

$$h(y) = \inf_x \Phi(x,y),$$

我们的优化问题实际是求 $h(0)$.
假设 h 满足 Fenchel-Moreau 定理的条件，因此，$h(0) = h^{**}(0)$. 对偶问题就是求 $h^{**}(0)$. 换句话说,对偶问题就是求

$$\max_{y^* \in Y^*} -h^*(y^*). \tag{4.3}$$

称 $-h^*(y^*)$ 为对偶函数.

将对偶问题用 Φ来表示. 根据 $\Phi^*(x^*, y^*) = \sup_{x,y}(x, x^*) + (y, y^*) - \Phi(x, y)$ 及

$$
\begin{aligned}
h^*(y^*) &= \sup_y (y, y^*) - h(y) \\
&= \sup_y \left((y, y^*) - \inf_x \Phi(x, y)\right) \\
&= \sup_{x,y} (y, y^*) - \Phi(x, y) \\
&= \Phi^*(0, y^*).
\end{aligned}
\tag{4.4}
$$

我们可将对偶问题转化为

$$
\max_{y^* \in Y^*} -h^*(y^*) = \max_{y^* \in Y^*} -\Phi^*(0, y^*). \tag{4.5}
$$

定理 4.1

对$\bar{x} \in X$和$\bar{y}^* \in Y^*$,下列结论等价.

1. $\bar{x}$ 是问题*(P)*的解, $\bar{y}^*$ 是问题 (P^*)的解,且 $\min(P) = \max(P^*)$.
2. $\Phi(\bar{x}, 0) + \Phi^*(0, \bar{y}^*) = 0$.
3. $(0, \bar{y}^*) \in \partial\Phi(\bar{x}, 0)$.

所以原问题为

$$\min_{x\in X} \Phi(x,0), \tag{4.6}$$

而对偶问题为

$$\min_{y^*\in Y^*} \Phi^*(0,y^*). \tag{4.7}$$

两个问题的相似性在数学上看是美的.我们可以看到如果求对偶问题的对偶问题,会回到原问题 (假定$\Phi = \Phi^{**}$).当然前提是 V and V^* 自然同构.

注 4.1

后面均利用 $\Phi^*(0,y^*)$ 求对偶问题，而不再借助h *(*过河拆桥术*)*.

下面用一个具体例子说明上述过程.考虑优化问题

$$\min_x f(x) \tag{4.8}$$

$$\text{subject to } g(x) \le 0, \tag{4.9}$$

该问题可以扰动为

$$\min_x f(x) \tag{4.10}$$

$$\text{subject to } g(x) + y \le 0. \tag{4.11}$$

写成二元函数的形式为

$$\Phi(x, y) = \begin{cases} f(x), & \text{if } g(x) + y \le 0 \\ +\infty, & \text{otherwise.} \end{cases} \tag{4.12}$$

为了得到对偶问题,我们需要计算 $-\Phi^*(0, y^*)$.

$$
\begin{aligned}
-\Phi^*(0, y^*) &= -\sup_{g(x)+y\le 0}\{\langle y^*, y\rangle - f(x)\} \\
&= -\sup_{q\ge 0, x}\{\langle y^*, -g(x)-q\rangle - f(x)\} \\
&= \inf_{q\ge 0, x}\langle y^*, g(x)+q\rangle + f(x).
\end{aligned}
\tag{4.13}
$$

我们可以先关于 q 取极小, 除非 $y^* \ge 0$, 否则我们会得到 $-\infty$. 因此对偶函数为

$$
-\Phi^*(0, y^*) = \begin{cases} \inf_x f(x) + \langle y^*, g(x)\rangle, & \text{if } y^* > 0, \\ -\infty, & \text{else.} \end{cases}
\tag{4.14}
$$

这正是我们熟知的结果.

当$g(x) = Ax - b$时,对偶函数为

$$
\begin{aligned}
-\Phi^*(0, y^*) &= \inf_x f(x) + (y^*, Ax) - (y^*, b) \\
&= -\left(\sup_x -f(x) - (y^*, Ax)\right) - (y^*, b) \\
&= -\left(\sup_x (-A^* y^*, x) - f(x)\right) - (y^*, b) \\
&= -f^*(-A^* y^*) - (y^*, b).
\end{aligned} \tag{4.15}
$$

下面我们来推导求解这个问题的算法.记

$$
L(x, y^*) = f(x) + (y^*, Ax - b), \tag{4.16}
$$

$$
L_c(x, y^*) = L(x, y^*) + \frac{c}{2}\|Ax - b\|^2. \tag{4.17}
$$

则对偶问题是求如下问题的极大值

$$
-h^*(y^*) = \min_x L(x, y^*). \tag{4.18}
$$

即求$h^*(y^*)$的最小值.

显格式推导：若对y采用显格式 $y^{k+1}=y^k-c\nabla h^*(y^k)$,并对变量采用交替格式,则有如下显格式算法：

$$\begin{cases} x^{k+1}=argmin_x L(x,y^k), \\ y^{k+1}=y^k+c(Ax^{k+1}-b), \end{cases} \tag{4.19}$$

其中第二式是是对$h(y)=L(x^{k+1},y)$求导得到的.

注 4.2

这个算法以及后面的算法本质都是在求导数。优化h^*用的是最速下降法，可以换成其他方法.

隐格式推导： 对y采用隐格式即 $y^{k+1} = y^k - c\nabla h^*(y^{k+1}) \Leftrightarrow y^{k+1} = Prox_{ch^*} y^k$.

引理 4.1

设 $x^* = argmin_x L(x, y)$, 则 $b - Ax^* \in \partial h^*(y)$.

证明：

$$h^*(y) = -L(x^*, y) \Rightarrow \nabla h^*(y) = b - Ax^*.$$

根据引理，我们得到算法

$$\begin{cases} x^{k+1} = argmin_x L(x, y^{k+1}), \\ y^{k+1} = y^k + c(Ax^{k+1} - b). \end{cases} \tag{4.20}$$

这个算法直观上不可行的!

稍加分析: 当$x^{k+1} = argmin_x L(x, y^{k+1})$时, 有$b - Ax^{k+1} \in \partial h(y^{k+1})$. 从而有$y^{k+1} = y^k + c(Ax^{k+1} - b)$. 进一步考察$x^{k+1}$满足的条件

$$0 \in \partial_x L(x^{k+1}, y^{k+1}) = \partial f(x^{k+1}) + A^T y^{k+1} = \partial f(x^{k+1}) + A^T(y^k + c(Ax^{k+1} - b)), \tag{4.21}$$

即

$$x^{k+1} = argmin_x L_c(x, y^k). \tag{4.22}$$

从而得到隐格式如下：

$$\begin{cases} x^{k+1} = argmin_x L_c(x, y^k), \\ y^{k+1} = y^k + c(Ax^{k+1} - b). \end{cases} \tag{4.23}$$

变形后的方法可行，称之为增广Lagrange方法.

注 4.3

如果约束条件本来就是 $AX = b$, 则增广*Lagrange*方法中增加的项就是零,这时增广项看起来是很自然的.这种思想也会用到后面的*ADMM*方法中.除用隐格式的方法解释步长和增广项系数的关系外,用最优性条件也能解释.最优性条件要求

$$A\bar{x} - b = 0, \quad \nabla f(\bar{x}) + A^T\bar{y} = 0. \tag{4.24}$$

由于x^{k+1}极小化$L_\rho(x, y^k)$,故

$$\begin{aligned} 0 &= \nabla_x L_\rho\left(x^{k+1}, y^k\right) \\ &= \nabla_x f\left(x^{k+1}\right) + A^T\left(y^k + \rho\left(Ax^{k+1} - b\right)\right) \\ &= \nabla_x f\left(x^{k+1}\right) + A^T y^{k+1}. \end{aligned} \tag{4.25}$$

因此步长和系数一致才能保证最优性条件每一步都成立.而第一个最优性条件需要随着迭代收敛后保证.

增广Lagrange方法的缺点是即使 $f(x)$是可分离变量的,增广后也不能并行计算了.为克服这个缺点,提出后来的ADMM方法.

乘子交替法(Alternating Direction Method of Multipliers)是处理如下问题的

$$\begin{aligned} \text{minimize} \quad & f(x) + g(z), \\ \text{subject to} \quad & Ax + Bz = c, \end{aligned} \tag{4.26}$$

其中$x \in \mathbf{R}^n, z \in \mathbf{R}^m, A \in \mathbf{R}^{p\times n}, B \in \mathbf{R}^{p\times m}, c \in \mathbf{R}^p$. 该问题的增广Lagrange函数为

$$L_\rho(x,z,y) = f(x) + g(z) + y^T(Ax + Bz - c) + (\rho/2)\|Ax + Bz - c\|_2^2. \tag{4.27}$$

ADMM算法如下：

$$\begin{cases} x^{k+1} := \underset{x}{\operatorname{argmin}} L_\rho\left(x, z^k, y^k\right), \\ z^{k+1} := \underset{z}{\operatorname{argmin}} L_\rho\left(x^{k+1}, z, y^k\right), \\ y^{k+1} := y^k + \rho\left(Ax^{k+1} + Bz^{k+1} - c\right). \end{cases} \tag{4.28}$$

其中 $\rho > 0$.

而乘子法对应的形式为

$$\begin{cases} \left(x^{k+1}, z^{k+1}\right) := \underset{x,z}{\operatorname{argmin}} L_\rho\left(x, z, y^k\right), \\ \qquad\qquad y^{k+1} := y^k + \rho\left(Ax^{k+1} + Bz^{k+1} - c\right). \end{cases} \tag{4.29}$$

ADMM与该方法的区别是对两个变量采用先后(交替)顺序进行更新, 类似于乘子方法的Gauss-Seidel变形.若采用Jacobi变形就可以并行了.

ADMM可以通过合并一次项和二次项及缩放对偶变量得到一个更方便的形式.定义残差$r = Ax + Bz - c$,则有

$$y^T r + \frac{\rho}{2}\|r\|_2^2 = \frac{\rho}{2}\|r + \frac{y}{\rho}\|_2^2 - \frac{1}{2\rho}\|y\|_2^2 = \frac{\rho}{2}\|r + u\|_2^2 - \frac{\rho}{2}\|u\|_2^2. \tag{4.30}$$

其中$u = \frac{y}{\rho}$ 是伸缩后的对偶变量.由此我们得到ADMM的新形式如下：

$$\begin{cases} x^{k+1} := \underset{x}{\operatorname{argmin}}\left(f(x) + \frac{\rho}{2}\left\|Ax + Bz^k - c + u^k\right\|_2^2\right), \\ z^{k+1} := \underset{z}{\operatorname{argmin}}\left(g(z) + \frac{\rho}{2}\left\|Ax^{k+1} + Bz - c + u^k\right\|_2^2\right), \\ u^{k+1} := u^k + Ax^{k+1} + Bz^{k+1} - c. \end{cases} \tag{4.31}$$

该形式与原型是等价,但是形式上更简洁,因此普遍被采用.

下面用两种方法推导一个常见泛函的最优性条件. 目标函数为

$$F(x) = f(x) + \varphi(\Lambda x). \tag{4.32}$$

扰动函数为:

$$\Phi(x, y) = f(x) + \varphi(\Lambda x + y). \tag{4.33}$$

方法一：基于Fenchel-Moreau定理推导. 目标函数的共轭函数为：

$$\begin{aligned}\Phi^*(x^*, y^*) &= \sup_{x\in X}\sup_{y\in Y}\{\langle x^*, x\rangle + \langle y^*, y\rangle - \Phi(x, y)\} \\ &= \sup_{x\in X}\left\{\langle x^*, x\rangle - f(x) + \sup_{y\in Y}[\langle y^*, y\rangle - \varphi(\Lambda x + y)]\right\},\end{aligned} \tag{4.34}$$

其中

$$\begin{aligned}&\sup_{y\in Y}[\langle y^*, y\rangle - \varphi(\Lambda x + y)] = \sup_{y\in Y}[\langle y^*, \Lambda x + y\rangle - \varphi(\Lambda x + y) - \langle y^*, \Lambda x\rangle] \\ &= \sup_{z\in Y}[\langle y^*, z\rangle - \varphi(z)] - \langle y^*, \Lambda x\rangle = \varphi^*(y^*) - \langle y^*, \Lambda x\rangle.\end{aligned} \tag{4.35}$$

从而有

$$\begin{aligned}\Phi^*(x^*, y^*) &= \sup_{x\in X}(\langle x^*, x\rangle - \langle y^*, \Lambda x\rangle - f(x) + \varphi^*(y^*)\} \\ &= \sup_{x\in X}\{\langle x^* - \Lambda^* y^*, x\rangle - f(x) + \varphi^*(y^*)\} \\ &= f^*(x^* - \Lambda^* y^*) + \varphi^*(y^*).\end{aligned} \tag{4.36}$$

最优性条件为：

$$
\begin{aligned}
0 &= \Phi(\bar{x}, 0) + \Phi^*(0, \bar{y}^*) = f(\bar{x}) + f^*(-\Lambda^* \bar{y}^*) + \varphi(\Lambda \bar{x}) + \varphi^*(\bar{y}^*) \\
&= [f(\bar{x}) + f^*(-\Lambda^* \bar{y}^*) - \langle -\Lambda^* \bar{y}^*, \bar{x} \rangle] + [\varphi(\Lambda \bar{x}) + \varphi^*(\bar{y}^*) - \langle \bar{y}^*, \Lambda \bar{x} \rangle].
\end{aligned}
\tag{4.37}
$$

由于方括号里面的项均非负,从而有

$$
\begin{aligned}
f(\bar{x}) + f^*(-\Lambda^* \bar{y}^*) - \langle -\Lambda^* \bar{y}^*, \bar{x} \rangle = 0, \\
\varphi(\Lambda \bar{x}) + \varphi^*(\bar{y}^*) - \langle \bar{y}^*, \Lambda \bar{x} \rangle = 0.
\end{aligned}
\tag{4.38}
$$

由Fenchel-Moreau定理得到

$$
\begin{aligned}
-\Lambda^* \bar{y}^* \in \partial f(\bar{x}), \\
\bar{y}^* \in \partial \varphi(\Lambda \bar{x}).
\end{aligned}
\tag{4.39}
$$

上述公式本质上给出了导数的计算公式.

方法小结：

1. 对目标函数进行扰动，得到扰动函数;
2. 计算扰动函数的共轭函数;
3. 计算最优解满足的方程;
4. 利用Fenchel-Moreau定理得到结果.

方法二：基于Lagrange函数推导.
Lagrange 函数 $L: X \times Y^* \to (-\infty, \infty]$ 定义如下：

$$-L(x, y^*) = \sup_{y \in Y} \{\langle y^*, y\rangle - \Phi(x, y)\} = \Phi_x^*(y^*). \tag{4.40}$$

注意到：

$$\begin{aligned} \Phi^*(x^*, y^*) &= \sup_{x \in X, y \in Y} \{\langle x^*, x\rangle + \langle y^*, y\rangle - \Phi(x, y)\} \\ &= \sup_{x \in X} \left\{\langle x^*, x\rangle + \sup_{y \in Y} \{\langle y^*, y\rangle - \Phi(x, y)\}\right\} \\ &= \sup_{x \in X} \{\langle x^*, x\rangle - L(x, y^*)\}. \end{aligned} \tag{4.41}$$

因此

$$-\Phi^*(0, y^*) = \inf_{x \in X} L(x, y^*). \tag{4.42}$$

因此共轭问题等价于

$$\sup_{y^* \in Y^*} \inf_{x \in X} L(x, y^*). \tag{4.43}$$

假设Φ是正常下半连续凸的,则

$$\begin{aligned}\Phi(x,y)=\Phi_x^{**}(x,y)&=\sup_{y^*\in Y^*}\{\langle y^*,y\rangle-\Phi_x^*(y^*)\}\\&=\sup_{y^*\in Y^*}\{\langle y^*,y\rangle+L(x,y^*)\}.\end{aligned}\tag{4.44}$$

因此

$$\Phi(x,0)=\sup_{y^*\in Y^*}L(x,y^*).\tag{4.45}$$

因此原问题等价于

$$\inf_{x\in X}\sup_{y^*\in Y}L(x,y^*).\tag{4.46}$$

Lagrange 函数 L 与函数 Φ之间有如下关系成立:

$$L(\boldsymbol{x},y^*)=\inf_{\boldsymbol{y}}\{\Phi(\boldsymbol{x},\boldsymbol{y})-\langle y^*,\boldsymbol{y}\rangle\mid\boldsymbol{y}\in\mathbf{R}^m\}=-\Phi_x^*(y^*),\tag{4.47}$$

$$\Phi(\boldsymbol{x},\boldsymbol{y})=\sup_{y^*}\{\langle y^*,\boldsymbol{y}\rangle+L(\boldsymbol{x},y^*)\mid\lambda\in\mathbf{R}^m\}=\Phi_x^{**}(y).\tag{4.48}$$

以上分析表明，本节定义的Lagrange函数能用来刻画原问题和对偶问题.

下面利用定义的Lagrange函数推导最优性条件. 将目标函数的扰动函数代入Lagrange函数并利用共轭函数定义简单推导得

$$L(x, y^*) = f(x) + \langle y^*, \Lambda x\rangle - \varphi^*(y^*). \tag{4.49}$$

根据鞍点定理有 $0 \in \frac{\partial L}{\partial x}, 0 \in \frac{\partial L}{y^*}$, 即

$$\begin{aligned} -\Lambda^* \bar{y}^* &\in \partial f(\bar{x}), \\ \Lambda \bar{x} &\in \partial \varphi^*(\bar{y}^*). \end{aligned} \tag{4.50}$$

此结论和前面推导的结果相同.

方法小结:

1. 对目标函数进行扰动, 得到扰动函数;
2. 利用扰动函数计算Lagrange函数;
3. 对Lagrange函数求导, 并令导数为零即可.

第五节 矩阵微分[1,2]

[1] 张贤达. 矩阵分析与应用[M]. 清华大学出版社, 2013.

[2] Magnus J R, Neudecker H. Matrix differential calculus with applications in statistics and econometrics[M]. John Wiley and Sons, 2019.

引理 5.1

$$\mathrm{vec}(AXB) = B^T \otimes A\mathrm{vec}(X).$$

证明.

$$(AXB)_j = AXB_j = \sum_{k=1}^{n} B_{k,j} AX_k.$$

按列写出系数即可得到结论. □

引理 5.2

$$d(X^{-1}) = -X^{-1}dX\ X^{-1}.$$

证明.

$$0 = dI = d(XX^{-1}) = dXX^{-1} + XdX^{-1}.$$

□

引理 5.3

$$\frac{dX}{dx_{ij}} = e_i e_j^T,$$

$$\frac{dX^T}{dx_{ij}} = e_j e_i^T,$$

$$\frac{dX^{-1}}{dx_{ij}} = -X^{-1} e_i e_j^T X^{-1} = -X^{-1} e_i (X^{-T} e_j)^T.$$

证明.

显然. □

引理 5.4

$$\frac{d}{dX}\mathrm{tr}(AXB) = A^TB^T, \frac{d}{dX}\mathrm{tr}(AX^TB) = BA,$$

$$\frac{d}{dX}\mathrm{tr}(AX^{-1}B) = -X^{-T}A^TB^TX^{-T}.$$

证明.

$$\frac{d}{dx_{ij}}\mathrm{tr}(AXB) = \mathrm{tr}(Ae_ie_j^TB) = \mathrm{tr}(e_j^TBAe_i) = e_j^TBAe_i = (BA)_{ji} = (A^TB^T)_{ij}.$$

$$\frac{d}{dx_{ij}}\mathrm{tr}(AX^TB) = \mathrm{tr}(Ae_je_i^TB) = \mathrm{tr}(e_i^TBAe_j) = e_i^TBAe_j = (BA)_{ij}.$$

$$\begin{aligned}\frac{d}{dx_{ij}}\mathrm{tr}(AX^{-1}B) &= -\mathrm{tr}(AX^{-1}e_ie_j^TX^{-1}B) = -\mathrm{tr}(e_j^TX^{-1}BAX^{-1}e_i)\\ &= -e_j^TX^{-1}BAX^{-1}e_i = -(X^{-1}BAX^{-1})_{ji}\\ &= -(X^{-T}A^TB^TX^{-T})_{ij}.\end{aligned}$$

□

引理 5.5

$$\frac{d}{dX}(a^TX^{-1}b) = -X^{-T}ab^TX^{-T}.$$

证明.

$$\begin{aligned}\frac{d}{dx_{ij}}(a^TX^{-1}b) &= -a^TX^{-1}e_ie_j^TX^{-1}b = -a^TX^{-1}e_i * e_j^TX^{-1}b \\ &= -e_i^TX^{-T}a * b^TX^{-T}e_j = -(X^{-T}ab^TX^{-T})_{ij}.\end{aligned}$$

□

引理 5.6

$$d(\det(X)) = (X^*)^T : dX = \det(X)X^{-T} : dX.$$

证明.

由 $\det(X)I = XX^*$知, 对每个 i 有, $\det(X) = \sum_j x_{ij}X^*_{ji}$. 从而

$$\frac{d}{dx_{ij}}\det(X) = X^*_{ji} = (X^*)^T_{ij}.$$

□

引理 5.7

$$d\left(\det(A^T XB)\right) = d(\det(B^T X^T A)) = A(A^T XB)^{*T} B^T : dX = \det(A^T XB) X^{-T} : dX.$$

证明.

$$\begin{aligned}
d(\det(A^T XB)) &= (A^T XB)^{*T} : d(A^T XB) \\
&= \mathrm{vec}(A^T XB)^{*T} : (B \otimes A)^T \mathrm{vec}(dX) \\
&= (B \otimes A\mathrm{vec}(A^T XB)^{*T}) : \mathrm{vec}(dX) \\
&= A(A^T XB)^{*T} B^T : dX \\
&= \det(A^T XB)(A(A^T XB)^{-T} B^T) : dX \\
&= \det(A^T XB)(AA^{-1} X^{-T} B^{-T} B^T) : dX \\
&= \det(A^T XB) X^{-T} : dX.
\end{aligned}$$

□

引理 5.8

$$d(ln(\det(A^TXB))) = X^{-T} : dX.$$

证明.

$$\begin{aligned} d\left(\ln\left(\det(A^TXB)\right)\right) &= \det(A^TXB)^{-1}d(\det(A^TXB)) \\ &= \det(A^TXB)^{-1}\det(A^TXB)X^{-T} : dX \\ &= X^{-T} : dX. \end{aligned}$$

□

引理 5.9

$$d\left(\det(X)^k\right) = k\,\det(X)^k\ X^{-T} : dX.$$

证明.

$$\begin{aligned} d\left(\det(X)^k\right) &= k\ det(X)^{k-1} d(\det(X)) \\ &= k\,\det(X)^{k-1}\det(X)X^{-T} : dX \\ &= k\,\det(X)^k X^{-T} : dX. \end{aligned}$$

□

CVX软件包安装

下载网址:http://cvxr.com/cvx/
下载点击cvx_setup.m文件安装即可.