Mathematic Notes

邵华

2016 年 1 月 28 日

目　录

第一章　矩阵计算

第 1.1 节　文中符号

- 用大写字母 **A** 表示矩阵，小写字母 **a** 表示列向量，a 表示实数或者复数标量
- $+, -, *$ 分别表示矩阵加，减乘
- $\mathbf{A^T}$ 表示 **A** 的转置
- $\mathbf{A^H}$ 表示 **A** 的共轭转置，如果 **A** 是实矩阵，$\mathbf{A^T} = \mathbf{A^H}$
- $\mathbf{A^{-1}}, \mathbf{A}^{\#}, \mathbf{A}^{+}$ 分别表示逆，一般逆和伪逆
- $.*, ./$ 表示逐元素的乘法和除法
- $|\mathbf{A}|, ||\mathbf{A}||_{\mathbf{F}}$ 代表行列式和 Frobenius 范数
- $||\mathbf{a}||$ 代表欧几里得范数
- $|a|$ 代表 a 的绝对值

第 1.2 节　矩阵推导结论

$$\xi_i = \frac{||\mathrm{HT}||_{\mathrm{i,i}}^2}{[\mathrm{HTT^H H^H}]_{\mathrm{i,i}}} \tag{1.1}$$

将 **HT** 进行 SVD 分解

$$\mathbf{HT} = \mathbf{U}\Lambda\mathbf{V^H} \tag{1.2}$$

则有：

$$\xi_i = \frac{||\mathbf{HT}||_{i,i}^2}{[\mathbf{HTT^H H^H}]_{\mathbf{i,i}}} = \frac{|u_i^H \Lambda v_i|^2}{|u_i^H \Lambda^2 v_i|} \tag{1.3}$$

其中，u_i, v_i 分别是 $\mathbf{U^H}, \mathbf{V^H}$ 的第 i 列

Cauchy-Schwarz 不等式：

$$|\langle x, y\rangle|^2 \leq \langle x, x\rangle \cdot \langle y, y\rangle \tag{1.4}$$

Cauchy-Schwarz 积分不等式：

$$(\int_a^b f(x)g(x)dx)^2 \leq \int_a^b f^2(x)dx \int_a^b g^2(x)dx \tag{1.5}$$

$Hölder$ 不等式，当 $p, q \geq 1$，且 $\frac{1}{p}+\frac{1}{q}=1$

$$|\sum_{i=1}^n a_i b_i| \leq (\sum_{i=1}^n |a_i|^p)^{\frac{1}{p}} (\sum_{i=1}^n |b_i|^q)^{\frac{1}{q}} \tag{1.6}$$

Minkowski 积分不等式：

$$(\int_a^b |f(x)g(x)|^p dx)^{\frac{1}{p}} \leq (\int_a^b |f(x)|^p dx)^{\frac{1}{p}} (\int_a^b |g(x)|^p dx)^{\frac{1}{p}} \tag{1.7}$$

当 *p=q=2* 时候，*Minkowski* 积分不等式和 $Hölder$ 不等式退化成 *Cauchy* 不等式。

Jesen 不等式，给出了积分的凸函数值和凸函数积分值之间的关系：若 $f(x)$ 是 区间 $[a, b]$ 的凸函数，对任意的 $x_i, x_2, x_3, \cdots, x_n \in (a, b)$，有

$$f(\frac{x_n + x_2 + \cdots + x_n}{n}) \leq \frac{1}{n}[f(x_1) + f(x_2) + \cdots + f(x_n)] \tag{1.8}$$

$$\mathbf{Tr}[\mathbf{A}\mathbf{A}^{\mathbf{H}}] = ||\mathbf{vec}(\mathbf{A})||^2 = ||\mathbf{vec}(\mathbf{A}^{\mathbf{H}})||^2 \tag{1.9}$$

$$\sum_i \lambda_i ||\mathbf{A}e_i||^2 = \mathbf{Tr}\{\mathbf{A}\Lambda\mathbf{A}^{\mathbf{H}}\} \tag{1.10}$$

第 1.3 节 矩阵微积分

1.3.1 符号表示

- $d/dx(\mathbf{y})$ 是一个向量，第 i 个元素为 $dy(i)/dx$
- $d/d\mathbf{x}(y)$ 是一个向量，第 i 个元素为 $dy/dx(i)$
- $d/d\mathbf{x}(\mathbf{y}^{\mathbf{T}})$ 是一个矩阵，第 (i, j) 个元素为 $dy(j)/dx(i)$
- $d/dx(\mathbf{Y})$ 是一个矩阵，第 (i, j) 个元素为 $dy(i, j)/dx$

- $d/d\mathbf{X}(y)$ 是一个矩阵，第 (i,j) 个元素是 $dy/dx(i,j)$

1.3.2 线性乘积的导数

- $d/dx(\mathbf{AYB}) = \mathbf{A} * d/dx(\mathbf{Y}) * \mathbf{B}$
- $d/dx(\mathbf{Ay}) = \mathbf{A} * d/dx(\mathbf{y})$
- $d/d\mathbf{x}(\mathbf{x}^{\mathbf{T}}\mathbf{A}) = \mathbf{A}$
- $d/d\mathbf{x}(\mathbf{x}^{\mathbf{T}}) = \mathbf{I}$
- $d/d\mathbf{x}(\mathbf{x}^{\mathbf{T}}\mathbf{a}) = d/d\mathbf{x}(\mathbf{a}^{\mathbf{T}}\mathbf{x}) = \mathbf{a}$
- $d/d\mathbf{X}(\mathbf{a}^{\mathbf{T}}\mathbf{X}\mathbf{b}) = \mathbf{a}\mathbf{b}^{\mathbf{T}}$
- $d/d\mathbf{X}(\mathbf{a}^{\mathbf{T}}\mathbf{X}\mathbf{a}) = d/d\mathbf{X}(\mathbf{a}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{a}) = \mathbf{a}\mathbf{a}^{\mathbf{T}}$
- $d/d\mathbf{X}(\mathbf{a}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{b}) = \mathbf{b}\mathbf{a}^{\mathbf{T}}$
- $d/dx(\mathbf{YZ}) = \mathbf{Y} * d/dx(\mathbf{Z}) + d/dx(\mathbf{Y}) * \mathbf{Z}$

1.3.3 二次乘积的导数

- $d/d\mathbf{x}(\mathbf{Ax}+\mathbf{b})^{\mathbf{T}}\mathbf{C}(\mathbf{Dx}+\mathbf{e}) = \mathbf{A}^{\mathbf{T}}\mathbf{C}(\mathbf{Dx}+\mathbf{e}) + \mathbf{D}^{\mathbf{T}}\mathbf{C}^{\mathbf{T}}(\mathbf{Ax}+\mathbf{b})$
- $d/d\mathbf{x}(\mathbf{x}^{\mathbf{T}}\mathbf{Cx}) = (\mathbf{C}+\mathbf{C}^{\mathbf{T}})\mathbf{x}$
- $d/d\mathbf{x}(\mathbf{x}^{\mathbf{T}}\mathbf{Cx}) = 2Cx$
- [C:symmetric]$d/d\mathbf{x}\mathbf{x}^{\mathbf{T}}\mathbf{x} = 2x$
- $d/d\mathbf{x}(\mathbf{Ax}+\mathbf{b})^{\mathbf{T}}(\mathbf{Dx}+\mathbf{e}) = \mathbf{A}^{\mathbf{T}}(\mathbf{Dx}+\mathbf{e}) + \mathbf{D}^{\mathbf{T}}(\mathbf{Ax}+\mathbf{b})$
- $d/d\mathbf{x}(\mathbf{Ax}+\mathbf{b})^{\mathbf{T}}(\mathbf{Ax}+\mathbf{b}) = \mathbf{2A}^{\mathbf{T}}(\mathbf{Ax}+\mathbf{b})$
- [C:symmetric]$d/d\mathbf{x}(\mathbf{Ax}+\mathbf{b})^{\mathbf{T}}\mathbf{C}(\mathbf{Ax}+\mathbf{b}) = \mathbf{2A}^{\mathbf{T}}\mathbf{C}(\mathbf{Ax}+\mathbf{b})$
- $d/d\mathbf{X}(\mathbf{a}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{b}) = \mathbf{X}(\mathbf{a}\mathbf{b}^{\mathbf{T}} + \mathbf{b}\mathbf{a}^{\mathbf{T}})$
- $d/d\mathbf{X}(\mathbf{a}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{a}) = \mathbf{2Xba}^{\mathbf{T}}$
- $d/d\mathbf{X}(\mathbf{a}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{CXb}) = \mathbf{C}^{\mathbf{T}}\mathbf{Xab}^{\mathbf{T}} + \mathbf{CXba}^{\mathbf{T}}$
- $d/d\mathbf{X}(\mathbf{a}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{CXa}) = (\mathbf{C}+\mathbf{C}^{\mathbf{T}})\mathbf{Xaa}^{\mathbf{T}}$
- [C:Symmetric] $d/d\mathbf{X}(\mathbf{a}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{CXa}) = \mathbf{2CXaa}^{\mathbf{T}}$
- $d/d\mathbf{X}((\mathbf{Xa}+\mathbf{b})^{\mathbf{T}})\mathbf{C}(\mathbf{Xa}+\mathbf{b}) = (\mathbf{C}+\mathbf{C}^{\mathbf{T}})(\mathbf{Xa}+\mathbf{b})\mathbf{a}^{\mathbf{T}}$

1.3.4 三次乘积的导数

$$d/d\mathbf{x}(\mathbf{x^T A x x^T}) = (\mathbf{A} + \mathbf{A^T})\mathbf{x x^T} + \mathbf{x^T A x I} \tag{1.11}$$

1.3.5 逆矩阵的导数

$$d/d\mathbf{x}(\mathbf{Y^{-1}}) = -\mathbf{Y^{-1}} d/dx(\mathbf{Y})\mathbf{Y^{-1}} \tag{1.12}$$

1.3.6 迹的导数

- $d/d\mathbf{X}(\mathbf{tr}(\mathbf{X})) = \mathbf{I}$
- $d/d\mathbf{X}(\mathbf{tr}(\mathbf{X^k}) = \mathbf{k}(\mathbf{X^{k-1}})^\mathbf{T}$
- $d/d\mathbf{X}(\mathbf{tr}(\mathbf{AX^k})) = \sum_{\mathbf{r=0:k-1}} (\mathbf{X^r A X^{k-r-1}})^\mathbf{T}$
- $d/d\mathbf{X}(\mathbf{tr}(\mathbf{AX^{-1}B})) = -(\mathbf{X^{-1}BAX^{-1}})^\mathbf{T}$
- $d/d\mathbf{X}(\mathbf{tr}(\mathbf{AX^{-1}})) = d/d\mathbf{X}(\mathbf{tr}(\mathbf{X^{-1}A})) = -\mathbf{X^{-T}A^T x^{-T}}$
- $d/d\mathbf{X}(\mathbf{tr}(\mathbf{A^T X B^T})) = d/d\mathbf{X}(\mathbf{tr}(\mathbf{B X^T A})) = \mathbf{AB}$
- $d/d\mathbf{X}(\mathbf{tr}(\mathbf{XA^T})) = d/d\mathbf{X}(\mathbf{tr}(\mathbf{X^T A})) = d/d\mathbf{X}(\mathbf{tr}(\mathbf{A^T X})) = d/d\mathbf{X}(\mathbf{tr}(\mathbf{AX^T})) = \mathbf{A}$
- $d/d\mathbf{X}(\mathbf{tr}(\mathbf{AXBX^T})) = \mathbf{A^T X B^T} + \mathbf{AXB}$
- $d/d\mathbf{X}(\mathbf{tr}(\mathbf{XAX^T})) = \mathbf{X}(\mathbf{A} + \mathbf{A^T})$
- $d/d\mathbf{X}(\mathbf{tr}(\mathbf{X^T A X})) = \mathbf{X^T}(\mathbf{A} + \mathbf{A^T})$
- $d/d\mathbf{X}(\mathbf{tr}(\mathbf{AX^T X})) = (\mathbf{A} + \mathbf{A^T})\mathbf{X}$
- $d/d\mathbf{X}(\mathbf{tr}(\mathbf{AXBX})) = \mathbf{A^T X^T B^T} + \mathbf{B^T X^T A^T}$
- [C:symmetric]$d/d\mathbf{X}(\mathbf{tr}((\mathbf{X^T C X})^{-1}\mathbf{A}) = d/d\mathbf{X}(\mathbf{tr}(\mathbf{A}(\mathbf{X^T C X})^{-1})) = -(\mathbf{CX}(\mathbf{X^T C X})^{-1})(\mathbf{A}+\mathbf{A^T})(\mathbf{X^T C X})^{-1}$
- [B,C:Symmetric]
- $d/d\mathbf{X}(\mathbf{tr}((\mathbf{X^T C X})^{-1}(\mathbf{X^T B X})) = d/d\mathbf{X}(\mathbf{tr}((\mathbf{X^T B X})(\mathbf{X^T C X})^{-1} = -\mathbf{2}(\mathbf{CX}(\mathbf{X^T C X})^{-1}\mathbf{X^T B X}(\mathbf{X^T C X})^{-1} + \mathbf{2BX}(\mathbf{X^T C X})^{-1}$

1.3.7 行列式的导数

- $d/d\mathbf{X}(\mathbf{det}(\mathbf{X})) = d/d\mathbf{X}(\mathbf{det}(\mathbf{X^T})) = \mathbf{det}(\mathbf{X}) * \mathbf{X^{-T}}$
- $d/d\mathbf{X}(\mathbf{det}(\mathbf{AXB})) = \mathbf{det}(\mathbf{AXB}) * \mathbf{X^{-T}}$
- $d/d\mathbf{X}(\mathbf{ln}(\mathbf{det}(\mathbf{AXB}))) = \mathbf{X^{-T}}$

- $d/d\mathbf{X}(\mathbf{det}(\mathbf{X^k})) = \mathbf{k} * \mathbf{det}(\mathbf{X^k}) * \mathbf{X^{-T}}$
- $d/d\mathbf{X}(\mathbf{ln}(\mathbf{det}(\mathbf{X^k}))) = \mathbf{kX^{-T}}$
- [Real] $d/d\mathbf{X}(\mathbf{det}(\mathbf{X^TCX})) = \mathbf{det}(\mathbf{X^TCX}) * (\mathbf{C} + \mathbf{C^T})\mathbf{X}(\mathbf{X^TCX})^{-1}$
- [C:Real,Symmetric]$d/d\mathbf{X}(\mathbf{det}(\mathbf{X^TCX})) = \mathbf{2det}(\mathbf{X^TCX}) * \mathbf{CX}(\mathbf{X^TCX})^{-1}$
- [C:Real,Symmetric]$d/d\mathbf{X}(\mathbf{ln}(\mathbf{det}(\mathbf{X^TCX}))) = \mathbf{2CX}(\mathbf{X^TCX})^{-1}$

1.3.8 雅可比

如果 **y** 是 **x** 的函数，则有 $d\mathbf{y^T}/d\mathbf{x}$ 是 **y** 相对于 **x** 的雅可比矩阵。$|d\mathbf{y^T}/d\mathbf{x}|$ 称为 **y** 相对于 **x** 的雅可比行列式，它主要用在积分中变换积分变量：

$$\int (f(\mathbf{y})d\mathbf{y}) = \int (f(\mathbf{y}(\mathbf{x}))|d\mathbf{y^T}/d\mathbf{x}|dx) \tag{1.13}$$

1.3.9 Hessian 矩阵

如果 f 是 x 的函数，则对称矩阵 $d^2f/dx^2 = d/dx^T(df/dx)$ 被称为函 数 f(x) 的 Hessian 矩阵。对于 $df/d\mathbf{x} = 0$ 的点，若 Hessian 矩阵是正定，负 定和不定的情况，分别为最小值，最大值，和鞍点。

- $d^2/d\mathbf{x^2}(\mathbf{AX} + \mathbf{B})^\mathbf{T}\mathbf{C}(\mathbf{Dx} + \mathbf{e}) = \mathbf{A^TCD} + \mathbf{D^TC^TA}$
- $d^2/d\mathbf{x^2}(\mathbf{a^Tx}) = \mathbf{0}$
- $d^2/d\mathbf{x^2}(\mathbf{x^Tx}) = \mathbf{2I}$
- $d^2/d\mathbf{x^2}(\mathbf{x^TCx}) = \mathbf{C} + \mathbf{C^T}$
- $d^2/d\mathbf{x^2}(\mathbf{Ax} + \mathbf{b})^\mathbf{T}(\mathbf{Dx} + \mathbf{e}) = \mathbf{A^TD} + \mathbf{D^TA}$
- $d^2/d\mathbf{x^2}(\mathbf{Ax} + \mathbf{b})^\mathbf{T}(\mathbf{Ax} + \mathbf{b}) = \mathbf{2A^TA}$
- [C:Symmetric] $d^2/d\mathbf{x^2}(\mathbf{Ax} + \mathbf{b})^\mathbf{T}\mathbf{C}(\mathbf{Ax} + \mathbf{b}) = \mathbf{2A^TCA}$

第 1.4 节 二阶锥规划

1.4.1 基本形式 [1]

$$\begin{aligned} &\textbf{minimize } f^T(x) \\ &\textbf{subject to } ||A_ix + b_i|| \le c_i^Tx + d_i,\ i = 1, 2, \cdots, N \end{aligned} \tag{1.14}$$

其中，$x \in \mathbb{R}^n$ 是优化变量，优化问题的参数为 $f \in \mathbb{R}^n, A_i \in \mathbb{R}^{(n_i-1)\times n}, b_i \in \mathbb{R}^{n_i-1}, c_i \in \mathbb{R}^n, d_i \in \mathbb{R}$。范 数表示欧几里得范数，即:$||u|| = (u^T u)^{1/2}$。约束

$$||A_i x + b_i|| \le c_i^T + d_i \tag{1.15}$$

为维度 n_i 的二阶锥约束。因为标准锥，或者 k 维单位二阶凸锥被定义为:

$$\mathscr{C}_k = \left\{ \begin{bmatrix} u \\ t \end{bmatrix} | u \in \mathbb{R}^{k-1}, t \in \mathbb{R}, ||u|| \le t \right\} \tag{1.16}$$

又被称做二次冰锥，活着洛伦兹锥。由于二阶锥约束的点集是单位二阶锥在仿射 映射下的原象：

$$||A_i x + b_i|| \le c_i^T x + d_i \Longleftrightarrow \begin{bmatrix} A_i \\ c_i^T \end{bmatrix} x + \begin{bmatrix} b_i \\ d_i \end{bmatrix} \in \mathscr{C}_{n_i} \tag{1.17}$$

因此是一个凸约束。故 SOCP 是一个凸优化问题。

1.4.2　半正定问题

二阶锥规划可以转化成如下形式的半正定问题：

$$\begin{aligned} &\textbf{minimize } f^T x \\ &\textbf{subject to } \begin{bmatrix} (c_i^T x + d_i)I & A_i x + b_i \\ (A_i x + b_i)^T & c_i^T x + d_i \end{bmatrix} \succeq 0,\ i = 1, \cdots, N \end{aligned} \tag{1.18}$$

然而，通过求解半正定规划来求解二次锥优化并不是一个好的办法，因为内点法 对于 SOCP 问题，比 SDP 问题有更好的收敛速度。

1.4.3　能够变换为 SOCP 问题的形式

二次约束的二次规划

范数的和与最大值问题

双曲约束问题

矩阵泛函问题

可表示为二阶锥的函数和集合

第二章　最优化问题与 KKT 条件

第 2.1 节　优化问题与对偶函数、对偶问题

考虑如下的最优化问题：

$$
\begin{aligned}
&\textbf{minimize}\ f_0(x)\\
&\textbf{subject to}\ f_i(x) \le 0,\ i=1,2,\cdots,m\\
&\qquad h_i(x)=0,\ i=1,2,\cdots,p
\end{aligned}
\tag{2.1}
$$

其自变量为 $x \in \mathbf{R}^n$，设问题的定义域为 $\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom} f_i \cap \bigcap_{i=1}^{p} \mathbf{dom} h_i$，优化问题的最优值为 p^*。注意，这里并没有假设(2.1)是一个凸优化问题。

问题(2.1)的 Lagrange 函数 $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R} \to \mathbf{R}$ 为：

$$
L(x,\lambda,v) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} v_i h_i(x) \tag{2.2}
$$

Lagrange 对偶函数g 为 $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$，关于 **Lagrange 函数**关于 x 的最小值：

$$
g(\lambda,v) = \mathbf{inf}_{x\in\mathcal{D}} L(x,\lambda,v) \tag{2.3}
$$

由于 **Lagrange 对偶函数**是关于 (λ, v) 的仿射函数的逐点下确界，所 以，即使原问题不是凸的，对偶函数也是**凹函数**。

于是，有 $\lambda \ge 0, v$，下面式子成立：

$$
g(\lambda,v) \le p^* \tag{2.4}
$$

即 $g(\lambda,v)$ 是愿问题最优解的一个下界。

Lagrange 对偶问题：

$$
\begin{aligned}
&\textbf{maximize}\ g(\lambda,v)\\
&\textbf{subject to}\ \lambda \ge 0
\end{aligned}
\tag{2.5}
$$

设 Lagrange 对偶问题具有可行解，即满足 $\lambda \le 0$ 和 $g(\lambda,v) > -\infty$ 的一组 (λ,v)。称 (λ^*,v^*) 为 对偶最优解，或者最优 Lagrange 乘子，对应的最优值为 d^*。根据定义，如 下额式子成立:

$$
d^* \le p^* \tag{2.6}
$$

定义差值 $p^* - d^*$ 为原问题的最有对偶间隙。

第 2.2 节　弱对偶性与强对偶性

(2.6)称为弱对偶性。若 $p^*=d^*$，则称强对偶性成立。

对于一般的情况，强对偶性不成立。但是如果原问题是凸问题，则强对偶性通常（但不总是）成立。强对偶性成立的条件：1. 原问题是凸问题，2.Slater 条件成 立：存在一点 $x\in \textbf{relint}\,\mathcal{D}$，使得下面式子成立了：

$$f_i(x)<0,\ i=1,\cdots,m,\ Ax=b \tag{2.7}$$

满足上面式子的点又被称为是严格可行点，这是因为不等式约束严格成立。

若化的 Slater 条件：若不等式约束函数 f_i 中有写函数是仿射函数，则仿射 不等式不需要严格成立:

$$f_i(x)\le 0, i=1,2\cdots,k,\ f_i(x)<0,\ i=k+1,\cdots,m,\ Ax=b \tag{2.8}$$

其中，$f_1,f_2,\cdots,f_k$ 是仿射的。

注意到当所有的约束条件都是线性等式或者不等式，并且 $\textbf{dom}f_0$ 是开集时，条件(2.8)就是可行性条件。

第 2.3 节　互补松性和 KKT 条件

令 x^* 和 (λ^*,v^*) 分别是原问题和对偶问题的最优解，则有：

$$\begin{aligned}
f_0(x^*)&=g(\lambda^*,v^*)\\
&=\textbf{inf}_x L(x^*,\lambda^*,v^*)\\
&\le f_0(x^*)+\sum_{i=1}^m\lambda_i^* f_i(x^*)+\sum_{i=1}^p v_i^* h_i(x^*)\\
&\le f_0(x^*)
\end{aligned} \tag{2.9}$$

第一个等式说明最优对偶间隙为零，第二个是对偶函数的定义，第三个不等式根 据 Lagrange 函数关于 x 求下确界小于等于其在 $x=x^*$ 处的值求得来。由此可 得一个重要的结论：

$$\sum_{i=1}^m\lambda_i^* f_i(x^*)=0 \tag{2.10}$$

由于每一项都非正，因此有:

$$\lambda_i^* f_i(x^*)=0,\ i=1,\cdots,m. \tag{2.11}$$

上面的条件称为**互补松弛性.**

Karush-Kuhn-Tucker(KKT) 条件:

$$
\begin{aligned}
f_x(x^*) \le 0,\ i = 1, \cdots, m \\
h_i(x^*) = 0,\ i = 1, \cdots, m \\
\lambda_i^* \ge 0,\ i = 1, \cdots, m \\
\lambda_i^* f_i(x^*) = 0,\ i = 1, \cdots, m \\
\Delta f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \Delta f_i(x^*) + \sum_{i=1}^{*} v_i^* \Delta h_i(x^*) = 0
\end{aligned}
\tag{2.12}
$$

若某个凸优化问题具有可微的目标函数和约束函数，且其满足 Slater 条件，那么 KKT 条件是最优性的充要条件：Slater 条件意味着最优对偶间隙为零，且对偶最 优解可以达到，因此 x 是原问题最优解，当且仅当存在 (λ, v)，二者满 足 KKT 条件。

第三章　概率和分布相关

第 3.1 节　正态分布、Gamma 分布、卡方分布、Nakagami-m 分布

3.1.1　Basic: Exponential Distributon

PDF:

$$f(x;\lambda) = \lambda e^{-\lambda x} \quad x \geq 0 \tag{3.1}$$

λ is the **Rate** parameter, while $\beta = 1/\lambda$ is the **Scale** parameter.
CDF:

$$F(x;\lambda) = 1 - e^{-\lambda x} \quad x \geq 0 \tag{3.2}$$

$\mathbb{E}[X] = \frac{1}{\lambda} = \beta$, **Var**$[X] = \frac{1}{\lambda^2}$
Properties:

- $X_i \sim \mathrm{Exp}(\lambda)$, then $\min(X_1, X_2, \cdots, X_n) \sim \mathrm{Exp(n\lambda)}$
- $X_i \sim \mathrm{Exp}(\lambda)$, then $X_1 + X_2 + \cdots + X_k \sim \mathrm{Erlang(k,\lambda)} = \mathrm{Gamma(k,\lambda^{-1})} \quad \mathrm{(k,\theta) - form}$

3.1.2　Normal Distribution

Notation:

$$\mathcal{N}(\mu, \sigma^2) \tag{3.3}$$

PDF:

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.4}$$

CDF:

$$F(x;\mu,\sigma) = \frac{1}{2}[1 + \mathrm{erf}(\frac{(x-\mu)^2}{\sigma\sqrt{2}})] \tag{3.5}$$

运算：

- $X_1 \sim \mathcal{N}(\mu_1, \sigma_1)$,$X_2 \sim \mathcal{N}(\mu_2, \sigma_2)$, $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$
- $X \sim \mathcal{N}(\mu, \sigma)$,then $X^2/\sigma^2 \sim \chi_1(\mu^2/\sigma^2)$,$\chi_1$ 为自由度为 1 的卡方分布

- $Z = X_1X_2$，其中 X_2, X_2 为标准正态分布，则 Z 为“乘积”正态 分布，pdf 为：

$$f_z(z) = \pi^{-1}K_0(|z|) \tag{3.6}$$

其中 K_0 为修正第二类 Bessel 函数

微分方程：

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + (x^2 - \alpha^2)y = 0 \tag{3.7}$$

对于任意的复数 α，方程的标准解被称为 Bessel 函数。对于正的 α 或者整数，如果方程的解在 $x = 0$ 处有届，则解被称为第一类 Bessel 函数，记为 $J_\alpha(x)$; 如果在 $x = 0$ 处为奇异并且多值，则该解被称为第二类 Bessel 函数，记为 $Y_\alpha(x)$. $H^1_\alpha(x) = J_\alpha(x) + iY_\alpha(x), H^2_\alpha(x) = J_\alpha(x) - iY_\alpha(x)$ 为另外两个线性无关的解，被称为 Hankel 函数。
微分方程：

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} - (x^2 + \alpha^2)y = 0 \tag{3.8}$$

对应的解，被称为修正的 Bessel 函数，第一类和第二类分别记录为 I_α, K_α

- $Z = \sqrt{X_1^2 + X_2^2}$, 其中 X_1, X_2 为标准正态分布，则 Z 为 Rayleigh 分布
- $Z = X_1 \div X_2$, 其中 X_1, X_2 为标准正态分布, 则 Z 为 Cauchy(0,1) 分 布
- $Z = X_1^2 + X_2^2 + \cdots + X_n^2$, 其中 $X_i \sim \mathcal{N}(0, 1)$, 则 Z 为 χ_n 自由度为 n 的卡方分布
- $X_1, X_2 \cdots X_n \sim \mathcal{N}(0, 1)$,$Y_1, Y_2, \cdots Y_m \sim \mathcal{N}(0, 1)$, 则 它们的归一化平方和之比:

$$F = \frac{(X_1^2 + X_2^2 + \cdots + X_n^2)/n}{(Y_1^2 + Y_2^2 + \cdots + Y_m^2)/m} \sim F_{n,m} \tag{3.9}$$

其中 $F_{n,m}$ 为自由度为 (n, m) 的 F 分布。

3.1.3 Gamma Distribution

Notation: Gamma 分布具有两种表示，形状一尺度 (shape-scale)

$$X \sim \Gamma(k, \theta) = Gamma(k, \theta) \tag{3.10}$$

或者形状一率 (shape-rate)

$$X \sim \Gamma(\alpha, \beta) = Gamma(\alpha, \beta) \tag{3.11}$$

其中 $\alpha = k, \beta = \frac{1}{\theta}$

PDF:

$$f(x; k, \theta) = \frac{x^{k-1}e^{-\frac{x}{\theta}}}{\theta^k\Gamma(k)} \tag{3.12}$$

for $x > 0, k, \theta > 0$

$$g(x;\alpha,\beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-x\beta}}{\Gamma(\alpha)} \tag{3.13}$$

for $x > 0, \alpha, \beta > 0$.

CDF:

$$F(x;k,\theta) = \int_0^x f(u;k,\theta)du = \frac{\gamma(k, \frac{x}{\theta})}{\Gamma(k)} \tag{3.14}$$

其中 $\gamma(.,.)$ 为不完全 gamma 函 数:$\gamma(s,x) = \int_0^x t^{s-1} e^{-t} dt$. 若 k 为正整数，则 gamma 分布为 Erlang 分布.

Summation:

$$\sum_{i=1}^N X_i \sim Gamma(\sum_{i-1}^N k_i, \theta) \tag{3.15}$$

其中，所有的 X_i 相互独立，且 $X_i \sim Gamma(k_i, \theta)$, 具有相 同的尺度参数。

Scaling:

如果 $X \sim Gamma(k,\theta)$，对于 $c > 0$，则有 $cX \sim Gamma(k, c\theta) = Gamma(\alpha, \beta/c)$.

事实上，如果 X 为一个指数随机变量，“率”rate 参数为 λ，则 cX 为一 个率参数 rate 参数为 λ/c 的随机变量。由于

$$X \sim \mathrm{Exp}(\lambda) = X \sim Gamma(1, \lambda^{-1}) \quad (k,\theta)\mathrm{-form} \tag{3.16}$$

故有此结论。

3.1.4 Chi-Square Distribution

如果 $Z_1, Z_2, Z_3 \cdots Z_k$ 是独立标准正态随机变量，则他们的平房和 $Q = \sum_{i=1}^k Z_i^2$ 服从卡方分布，表示为:

$$Q \sim \chi^2(k) \quad \text{or} \quad Q \sim \chi_k^2 \tag{3.17}$$

PDF:

$$f(x;k) = \frac{x^{k/2-1} e^{-x/2}}{2^{k/2}\Gamma(\frac{k}{2})} \quad x > 0; \tag{3.18}$$

$\Gamma(\frac{1}{2}) = \sqrt{\pi}$

当 $k = 1$ 时，卡方分布为正态分布的平方；$k = 2$ 时，为参数为 1/2 的指数分 布,Rayleigh 分布的平方；$k = 3$ 时，为参数为 1 的 Maxwell 分布的平方 $X \sim$ Maxwell(1) then $X^2 \sim \chi^2(2)$。

可加性：独立卡方分布随机变量的和仍旧为卡方分布，其自由度为所有随机变量 自由度之和。

如果 $X_1, X_2 \cdots, X_n$ 是独立的卡方分布随机变量，自由 度分别为 k_i，则 $Y = X_1 + X_2 + \cdots + X_n$ 为自由度为 $\sum_{i=0}^n k_i$ 的卡方分布。

3.1.5 Nakagami-m Distribution

PDF:

$$f(x;m,\Omega)=\frac{2m^m}{\Gamma(m)\Omega^m}x^{2m-1}\text{exp}(-\frac{m}{\Omega}x^2) \tag{3.19}$$

其中，m,Ω 可以表示为:

$$m=\frac{\mathbb{E}^2[X^2]}{\text{Var}[X^2]},\quad \Omega=\mathbb{E}[X^2] \tag{3.20}$$

Generation:

- 从 Gamma 分布产生，给定一个 $Y\sim\Gamma(k,\theta)$，通过设置 $k=m,\theta=\Omega/m$, 求取 Y 的根号 $X=\sqrt{Y}$, 则有 $X\sim$ Nakagami(m,Ω)
- 从 Chi-square 产生, 将卡方分布的自由度 k 变为 $2m$, 同时通过随机变量 的尺度变换得到。$Y\sim\chi(2m)$,$X=\sqrt{\Omega/2m}Y$ 服从参数为 (m,Ω) 的 Nakagami 分布.

第 3.2 节　Moment-Generating Function

对于一个随机变量 x 和对应的概率密度函数 $f(x)$, 如果存在一个 h 使得在 $|t|<h$ 时

$$M(t)=\mathbb{E}(e^{tx}) \tag{3.21}$$

$M(t)$ 被称为 MGF 函数。

如果随机变量 X 和 Y 相互独立，则 MGF 函数满足:

$$\begin{aligned}M_{x+y}(t)&=\mathbb{E}(e^{t(x+y)})\\&=\mathbb{E}(e^{tx}e^{ty})\\&=\mathbb{E}(e^{tx})\mathbb{E}(e^{ty})\\&=M_x(t)M_y(t)\end{aligned} \tag{3.22}$$

如果 $M(t)$ 在 0 点可微，则在 0 点的 n 阶矩可以用 MGF 在 0 点的微分表示:

$$\begin{aligned}M(t)&=\mathbb{E}(e^{tx})\quad M(0)=1\\M^{'}(t)&=\mathbb{E}(xe^{tx})\quad M^{'}(0)=\mathbb{E}(x)\\M^{''}(t)&=\mathbb{E}(x^2e^{tx})\quad M^{''}(0)=\mathbb{E}(x^2)\\M^{(n)}(t)&=\mathbb{E}(x^ne^{tx})\quad M^{(n)}(0)=\mathbb{E}(x^n)\end{aligned} \tag{3.23}$$

对于实分量的随机变量向量 X,MGF 函数可以表示为:

$$M_X(t)=\mathbb{E}(e^{<t,X>}) \tag{3.24}$$

其中 t 是一个向量，$<.,.>$ 表示向量点积.

常见分布的 MGF 函数:

Distribution	Moment-generating function
Bernoulli $P(X=1)=p$	$1-p+pe^t$
Geometric $(1-p)^{k-1}p$	$\frac{pe^t}{1-(1-p)e^t}, \forall t < -ln(1-p)$
Binomial $B(n,p)$	$(1-p+pe^t)^n$
Poisson Pois(λ)	$e^{\lambda(e^t-1)}$
Uniform continuous $U(a,b)$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
Uniform discrete $U(a,b)$	$\frac{e^{at}-e^{(b+1)t}}{(b-a+1)(1-e^t)}$
Normal $N(\mu,\sigma^2)$	$e^{t\mu+\frac{1}{2}\sigma^2t^2}$
Multivariate normal $N(\mu,\Sigma)$	$e^{t^T\mu+\frac{1}{2}t^T\Sigma t}$
Chi-square χ_k^2	$(1-2k)^{-k/2}$
Gamma $\Gamma(k,\theta)$	$(1-t\theta)^{-k}$
Exponential Exp(λ)	$(1-t\lambda^{-1})^{-1}, \quad t<\lambda$
Cauchy (μ,θ)	does not exist
Rayleigh(σ)	$1+\sigma te^{\sigma^2t^2/2}\sqrt{\frac{\pi}{2}}(erf(\frac{\sigma t}{\sqrt{2}})+1)$

表 3.1:

参考文献

[1] Miguel Sousa Lobo, Lieven Vandenberghe, Stephen Boyd, and Hervé Lebret. Applications of second-order cone programming. *Linear Algebra and its Applications*, 284(1–3):193 – 228, 1998. International Linear Algebra Society (ILAS) Symposium on Fast Algorithms for Control, Signals and Image Processing.