矩阵手册

[http://matrixcookbook.com]

Kaare Brandt Petersen
Michael Syskind Pedersen

版本：2012年11月15日

介绍

这是什么？ 这些页面是关于矩阵及其相关事项的事实（身份、近似、不等式、关系等）的集合。它以这种形式收集起来，以方便任何想要快速查阅的人。

免责声明：这里介绍的身份、近似和关系显然不是发明的，而是从大量来源中收集、借用和复制的。 这些来源包括在互联网上找到的类似但较短的笔记和书籍附录-请参阅参考文献以获取完整列表。

错误：很可能存在错误、拼写错误和错误，我们对此表示歉意，并将非常感谢在cookbook@2302.dk接收更正。

持续进行中：保持一个涉及矩阵的大型关系存储库的项目自然是持续进行的，版本将在标题中的日期中显示。

建议：欢迎您提供有关额外内容或某些主题的详细说明 acookbook@2302.dk。

关键词：矩阵代数，矩阵关系，矩阵恒等式，行列式的导数，逆矩阵的导数，矩阵的微分。

致谢：我们要感谢以下人员对我们的贡献和建议：Bill Baxter，Brian Templeton，Christian Rishøj，Christian Schr¨oppel，Dan Boley，Douglas L. Theobald，Esben Hoegh-Rasmussen，Evripidis Karseras，Georg Martius，Glynne Casteel，Jan Larsen，Jun Bin Gao，J¨urgen Struckmeier，Kamil Dedecius，Karim T. Abou-Moustafa，Korbinian Strimmer，Lars Christiansen，Lars Kai Hansen，Leland Wilkinson，Liguo He，Loic Thibaut，Markus Froeb，Michael Hubatka，Miguel Bar˜ao，Ole Winther，Pavel Sakov，Stephan Hattinger，Troels Pedersen，Vasile Sima，Vincent Rabaud，Zhaoshui He。我们还要感谢奥蒂康基金会资助我们的博士研究。

目录

1 基础知识 **6**
1.1 迹 . . . 6
1.2 行列式 . . . 6
1.3 特殊情况 2x2 . . . 7

2 导数 **8**
2.1 行列式的导数 . . . 8
2.2 逆矩阵的导数 . . . 9
2.3 特征值的导数 . . . 10
2.4 矩阵、向量和标量形式的导数 . . . 10
2.5 迹的导数 . . . 12
2.6 向量范数的导数 . . . 14
2.7 矩阵范数的导数 . . . 14
2.8 结构化矩阵的导数 . . . 14

3 逆矩阵 **17**
3.1 基本 . . . 17
3.2 精确关系 . . . 18
3.3 对逆矩阵的影响 . . . 20
3.4 近似 . . . 20
3.5 广义逆矩阵 . . . 21
3.6 伪逆矩阵 . . . 21

4 复数矩阵 **24**
4.1 复数导数 . . . 24
4.2 高阶和非线性导数 . . . 26
4.3 复数和的逆 . . . 27

5 解和分解 **28**
5.1 线性方程的解 . . . 28
5.2 特征值和特征向量 . . . 30
5.3 奇异值分解 . . . 31
5.4 三角分解 . . . 32
5.5 LU分解 . . . 32
5.6 LDM分解 . . . 33
5.7 LDL分解 . . . 33

6 统计与概率 **34**
6.1 矩的定义 . . . 34
6.2 线性组合的期望 . . . 35
6.3 加权标量变量 . . . 36

7 多元分布 **37**
7.1 柯西分布 . . . 37
7.2 狄利克雷分布 . . . 37
7.3 正态分布 . . . 37
7.4 正态逆伽马分布 . . . 37
7.5 高斯分布 . . . 37
7.6 多项分布 . . . 37

7.7 学生t分布 . . . 37
7.8 Wishart . . . 38
7.9 逆Wishart分布 . . . 39

8 高斯分布 **40**
8.1 基础知识 . . . 40
8.2 矩的定义 . . . 42
8.3 其他 . . . 44
8.4 高斯混合模型 . . . 44

9个特殊矩阵 **46**
9.1 块矩阵 . . . 46
9.2 离散傅里叶变换矩阵 . . . 47
9.3 Hermite矩阵和skew-Hermite矩阵 . . . 48
9.4 幂等矩阵 . . . 49
9.5 正交矩阵 . . . 49
9.6 正定和半正定矩阵 . . . 50
9.7 单入口矩阵 . . . 52
9.8 对称矩阵，反对称矩阵 . . . 54
9.9 Toeplitz矩阵 . . . 54
9.10 过渡矩阵 . . . 55
9.11 单位矩阵，排列和移位 . . . 56
9.12 Vandermonde矩阵 . . . 57

10 函数和运算符 **58**
10.1 函数和级数 . . . 58
10.2 Kronecker和Vec运算符 . . . 59
10.3 向量范数 . . . 61
10.4 矩阵范数 . . . 61
10.5 秩 . . . 62
10.6 包含Dirac Delta函数的积分 . . . 62
10.7 其他 . . . 63

一维结果 **64**
A.1 高斯分布 . . . 64
A.2 一维高斯混合 . . . 65

B 证明和细节 **66**
B.1 其他证明 . . . 66

符号和术语

$\mathbf{A}$	矩阵
$\mathbf{A}_{ij}$	为某种目的而索引的矩阵
$\mathbf{A}_i$	为某种目的而索引的矩阵
$\mathbf{A}^{ij}$	为某种目的而索引的矩阵
$\mathbf{A}^n$	为某种目的而索引的矩阵或 一个方阵的n次幂
$\mathbf{A}^{-1}$	矩阵 $\mathbf{A}$的逆矩阵
$\mathbf{A}^+$	矩阵 $\mathbf{A}$的伪逆矩阵(见第3.6节)
$\mathbf{A}^{1/2}$	矩阵的平方根（如果存在），不是逐元素的
$(\mathbf{A})_{ij}$	矩阵 $\mathbf{A}$的（i,j）元素
A_{ij}	矩阵 $\mathbf{A}$的（i,j）元素
$[\mathbf{A}]_{ij}$	矩阵 $\mathbf{A}$的 ij子矩阵，即删除第i行和第j列
$\mathbf{a}$	向量（列向量）
$\mathbf{a}_i$	为某种目的而索引的向量
a_i	向量 $\mathbf{a}$的第i个元素
a	标量
$\Re z$	标量的实部
$\Re \mathbf{z}$	向量的实部
$\Re \mathbf{Z}$	矩阵的实部
$\Im z$	标量的虚部
$\Im \mathbf{z}$	向量的虚部
$\Im \mathbf{Z}$	矩阵的虚部
$\det(\mathbf{A})$	矩阵 $\mathbf{A}$的行列式
$\mathrm{Tr}(\mathbf{A})$	矩阵 $\mathbf{A}$的迹
$\mathrm{diag}(\mathbf{A})$	矩阵 $\mathbf{A}$的对角矩阵，即 $(\mathrm{diag}(\mathbf{A}))_{ij} = \delta_{ij} A_{ij}$
$\mathrm{eig}(\mathbf{A})$	矩阵 $\mathbf{A}$的特征值
$\mathrm{vec}(\mathbf{A})$	矩阵 $\mathbf{A}$的向量形式（见第10.2.2节）
$\sup$	集合的上确界
$\lvert\lvert\mathbf{A}\rvert\rvert$	矩阵范数（下标表示范数类型）
$\mathbf{A}^T$	转置矩阵
$\mathbf{A}^{-T}$	转置的逆矩阵和逆的转置矩阵， $\mathbf{A}^{-T} = (\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}$.
$\mathbf{A}^*$	共轭矩阵
$\mathbf{A}^H$	转置共轭矩阵（厄米矩阵）
$\mathbf{A} \circ \mathbf{B}$	哈达玛积（逐元素乘积）
$\mathbf{A} \otimes \mathbf{B}$	克罗内克积
$\mathbf{0}$	零矩阵。所有元素都为零。
$\mathbf{I}$	单位矩阵
$\mathbf{J}^{ij}$	单元素矩阵，在$(i,\ j)$处为1，其他地方为零
$\mathbf{\Sigma}$	正定矩阵
$\mathbf{\Lambda}$	对角矩阵

1 基础知识

$$
\begin{aligned}
(\mathbf{AB})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1} && (1)\\
(\mathbf{ABC}...)^{-1} &= ...\mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1} && (2)\\
(\mathbf{A}^T)^{-1} &= (\mathbf{A}^{-1})^T && (3)\\
(\mathbf{A}+\mathbf{B})^T &= \mathbf{A}^T+\mathbf{B}^T && (4)\\
(\mathbf{AB})^T &= \mathbf{B}^T\mathbf{A}^T && (5)\\
(\mathbf{ABC}...)^T &= ...\mathbf{C}^T\mathbf{B}^T\mathbf{A}^T && (6)\\
(\mathbf{A}^H)^{-1} &= (\mathbf{A}^{-1})^H && (7)\\
(\mathbf{A}+\mathbf{B})^H &= \mathbf{A}^H+\mathbf{B}^H && (8)\\
(\mathbf{AB})^H &= \mathbf{B}^H\mathbf{A}^H && (9)\\
(\mathbf{ABC}...)^H &= ...\mathbf{C}^H\mathbf{B}^H\mathbf{A}^H && (10)
\end{aligned}
$$

1.1 迹

$$
\begin{aligned}
\mathrm{Tr}(\mathbf{A}) &= \textstyle\sum_i A_{ii} && (11)\\
\mathrm{Tr}(\mathbf{A}) &= \textstyle\sum_i \lambda_i, \qquad \lambda_i = \mathrm{eig}(\mathbf{A}) && (12)\\
\mathrm{Tr}(\mathbf{A}) &= \mathrm{Tr}(\mathbf{A}^T) && (13)\\
\mathrm{Tr}(\mathbf{AB}) &= \mathrm{Tr}(\mathbf{BA}) && (14)\\
\mathrm{Tr}(\mathbf{A}+\mathbf{B}) &= \mathrm{Tr}(\mathbf{A})+\mathrm{Tr}(\mathbf{B}) && (15)\\
\mathrm{Tr}(\mathbf{ABC}) &= \mathrm{Tr}(\mathbf{BCA}) = \mathrm{Tr}(\mathbf{CAB}) && (16)\\
\mathbf{a}^T\mathbf{a} &= \mathrm{Tr}(\mathbf{a}\mathbf{a}^T) && (17)
\end{aligned}
$$

1.2 行列式

设 $\mathbf{A}$为一个 $n \times n$矩阵。

$$
\begin{aligned}
\det(\mathbf{A}) &= \textstyle\prod_i \lambda_i \qquad \lambda_i = \mathrm{eig}(\mathbf{A}) && (18)\\
\det(c\mathbf{A}) &= c^n \det(\mathbf{A}), \qquad \text{如果 } \mathbf{A} \in \mathbb{R}^{n\times n} && (19)\\
\det(\mathbf{A}^T) &= \det(\mathbf{A}) && (20)\\
\det(\mathbf{AB}) &= \det(\mathbf{A})\det(\mathbf{B}) && (21)\\
\det(\mathbf{A}^{-1}) &= 1/\det(\mathbf{A}) && (22)\\
\det(\mathbf{A}^n) &= \det(\mathbf{A})^n && (23)\\
\det(\mathbf{I}+\mathbf{u}\mathbf{v}^T) &= 1+\mathbf{u}^T\mathbf{v} && (24)
\end{aligned}
$$

对于 n= 2:

$$\det(\mathbf{I}+\mathbf{A}) = 1+\det(\mathbf{A})+\mathrm{Tr}(\mathbf{A}) \tag{25}$$

对于 n= 3:

$$\det(\mathbf{I}+\mathbf{A}) = 1+\det(\mathbf{A})+\mathrm{Tr}(\mathbf{A})+\frac{1}{2}\mathrm{Tr}(\mathbf{A})^2-\frac{1}{2}\mathrm{Tr}(\mathbf{A}^2) \tag{26}$$

对于 $n = 4$:

$$\begin{aligned}\det(\mathbf{I}+\mathbf{A}) &= 1+\det(\mathbf{A})+\mathrm{Tr}(\mathbf{A})+\frac{1}{2} \\ &+\mathrm{Tr}(\mathbf{A})^2-\frac{1}{2}\mathrm{Tr}(\mathbf{A}^2) \\ &+\frac{1}{6}\mathrm{Tr}(\mathbf{A})^3-\frac{1}{2}\mathrm{Tr}(\mathbf{A})\mathrm{Tr}(\mathbf{A}^2)+\frac{1}{3}\mathrm{Tr}(\mathbf{A}^3)\end{aligned} \tag{27}$$

对于小 ε，以下近似成立

$$\det(\mathbf{I}+\varepsilon\mathbf{A}) \cong 1+\det(\mathbf{A})+\varepsilon\mathrm{Tr}(\mathbf{A})+\frac{1}{2}\varepsilon^2\mathrm{Tr}(\mathbf{A})^2-\frac{1}{2}\varepsilon^2\mathrm{Tr}(\mathbf{A}^2) \tag{28}$$

1.3 特殊情况 2x2

考虑矩阵 $\mathbf{A}$

$$\mathbf{A}=\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22}\end{bmatrix}$$

行列式和迹

$$\det(\mathbf{A}) = A_{11}A_{22}-A_{12}A_{21} \tag{29}$$

$$\mathrm{Tr}(\mathbf{A}) = A_{11}+A_{22} \tag{30}$$

特征值

$$\lambda^2-\lambda\cdot\mathrm{Tr}(\mathbf{A})+\det(\mathbf{A})=0$$

$$\lambda_1=\frac{\mathrm{Tr}(\mathbf{A})+\sqrt{\mathrm{Tr}(\mathbf{A})^2-4\det(\mathbf{A})}}{2} \qquad \lambda_2=\frac{\mathrm{Tr}(\mathbf{A})-\sqrt{\mathrm{Tr}(\mathbf{A})^2-4\det(\mathbf{A})}}{2}$$

$$\lambda_1+\lambda_2=\mathrm{Tr}(\mathbf{A}) \qquad \lambda_1\lambda_2=\det(\mathbf{A})$$

特征向量

$$\mathbf{v}_1 \propto \begin{bmatrix} A_{12} \\ \lambda_1-A_{11}\end{bmatrix} \qquad \mathbf{v}_2 \propto \begin{bmatrix} A_{12} \\ \lambda_2-A_{11}\end{bmatrix}$$

逆

$$\mathbf{A}^{-1}=\frac{1}{\det(\mathbf{A})}\begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11}\end{bmatrix} \tag{31}$$

2 导数

本节涵盖了关于矩阵 $\mathbf{X}$的多个表达式的微分。 请注意，我们始终假设 $\mathbf{X}$没有特殊结构，即 $\mathbf{X}$的元素是独立的（例如不对称、Toeplitz、正定）。 有关结构化矩阵的微分，请参见第2.8节。 基本假设可以用以下公式表示

$$\frac{\partial X_{kl}}{\partial X_{ij}} = \delta_{ik}\delta_{lj} \tag{32}$$

例如，向量形式如下,

$$\left[\frac{\partial \mathbf{x}}{\partial y}\right]_i = \frac{\partial x_i}{\partial y} \qquad \left[\frac{\partial x}{\partial \mathbf{y}}\right]_i = \frac{\partial x}{\partial y_i} \qquad \left[\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right]_{ij} = \frac{\partial x_i}{\partial y_j}$$

以下规则是普遍适用的，在推导表达式的微分时非常有用（[19]）:

$$\begin{aligned}
\partial \mathbf{A} &= 0 \qquad (\mathbf{A}\text{ 是常数}) && (33)\\
\partial(\alpha\mathbf{X}) &= \alpha\partial\mathbf{X} && (34)\\
\partial(\mathbf{X}+\mathbf{Y}) &= \partial\mathbf{X}+\partial\mathbf{Y} && (35)\\
\partial(\mathrm{Tr}(\mathbf{X})) &= \mathrm{Tr}(\partial\mathbf{X}) && (36)\\
\partial(\mathbf{X}\mathbf{Y}) &= (\partial\mathbf{X})\mathbf{Y}+\mathbf{X}(\partial\mathbf{Y}) && (37)\\
\partial(\mathbf{X}\circ\mathbf{Y}) &= (\partial\mathbf{X})\circ\mathbf{Y}+\mathbf{X}\circ(\partial\mathbf{Y}) && (38)\\
\partial(\mathbf{X}\otimes\mathbf{Y}) &= (\partial\mathbf{X})\otimes\mathbf{Y}+\mathbf{X}\otimes(\partial\mathbf{Y}) && (39)\\
\partial(\mathbf{X}^{-1}) &= -\mathbf{X}^{-1}(\partial\mathbf{X})\mathbf{X}^{-1} && (40)\\
\partial(\det(\mathbf{X})) &= \mathrm{Tr}(\mathrm{adj}(\mathbf{X})\partial\mathbf{X}) && (41)\\
\partial(\det(\mathbf{X})) &= \det(\mathbf{X})\mathrm{Tr}(\mathbf{X}^{-1}\partial\mathbf{X}) && (42)\\
\partial(\ln(\det(\mathbf{X}))) &= \mathrm{Tr}(\mathbf{X}^{-1}\partial\mathbf{X}) && (43)\\
\partial\mathbf{X}^T &= (\partial\mathbf{X})^T && (44)\\
\partial\mathbf{X}^H &= (\partial\mathbf{X})^H && (45)
\end{aligned}$$

2.1 行列式的导数

2.1.1 一般形式

$$\begin{aligned}
\frac{\partial \det(\mathbf{Y})}{\partial x} &= \det(\mathbf{Y})\mathrm{Tr}\left[\mathbf{Y}^{-1}\frac{\partial\mathbf{Y}}{\partial x}\right] && (46)\\
\sum_k \frac{\partial \det(\mathbf{X})}{\partial X_{ik}} X_{jk} &= \delta_{ij}\det(\mathbf{X}) && (47)\\
\frac{\partial^2 \det(\mathbf{Y})}{\partial x^2} &= \det(\mathbf{Y})\left[\text{迹}\left[\mathbf{Y}^{-1}\frac{\partial\frac{\partial\mathbf{Y}}{\partial x}}{\partial x}\right]\right.\\
&\quad +\text{迹}\left[\mathbf{Y}^{-1}\frac{\partial\mathbf{Y}}{\partial x}\right]\text{迹}\left[\mathbf{Y}^{-1}\frac{\partial\mathbf{Y}}{\partial x}\right]\\
&\quad \left.-\text{迹}\left[\left(\mathbf{Y}^{-1}\frac{\partial\mathbf{Y}}{\partial x}\right)\left(\mathbf{Y}^{-1}\frac{\partial\mathbf{Y}}{\partial x}\right)\right]\right] && (48)
\end{aligned}$$

2.1.2 线性形式

$$\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} = \det(\mathbf{X})(\mathbf{X}^{-1})^T \tag{49}$$

$$\sum_k \frac{\partial \det(\mathbf{X})}{\partial X_{ik}} X_{jk} = \delta_{ij} \det(\mathbf{X}) \tag{50}$$

$$\frac{\partial \det(\mathbf{AXB})}{\partial \mathbf{X}} = \det(\mathbf{AXB})(\mathbf{X}^{-1})^T = \det(\mathbf{AXB})(\mathbf{X}^T)^{-1} \tag{51}$$

2.1.3 方阵形式

如果 $\mathbf{X}$是方阵且可逆，则

$$\frac{\partial \det(\mathbf{X}^T\mathbf{AX})}{\partial \mathbf{X}} = 2\det(\mathbf{X}^T\mathbf{AX})\mathbf{X}^{-T} \tag{52}$$

如果 $\mathbf{X}$不是方阵但 $\mathbf{A}$是对称的，则

$$\frac{\partial \det(\mathbf{X}^T\mathbf{AX})}{\partial \mathbf{X}} = 2\det(\mathbf{X}^T\mathbf{AX})\mathbf{AX}(\mathbf{X}^T\mathbf{AX})^{-1} \tag{53}$$

如果 $\mathbf{X}$不是方阵且 $\mathbf{A}$不对称，则

$$\frac{\partial \det(\mathbf{X}^T\mathbf{AX})}{\partial \mathbf{X}} = \det(\mathbf{X}^T\mathbf{AX})(\mathbf{AX}(\mathbf{X}^T\mathbf{AX})^{-1} + \mathbf{A}^T\mathbf{X}(\mathbf{X}^T\mathbf{A}^T\mathbf{X})^{-1}) \tag{54}$$

2.1.4 其他非线性形式

一些特殊情况是（见[9, 7]）

$$\frac{\partial \ln \det(\mathbf{X}^T\mathbf{X})|}{\partial \mathbf{X}} = 2(\mathbf{X}^+)^T \tag{55}$$

$$\frac{\partial \ln \det(\mathbf{X}^T\mathbf{X})}{\partial \mathbf{X}^+} = -2\mathbf{X}^T \tag{56}$$

$$\frac{\partial \ln|\det(\mathbf{X})|}{\partial \mathbf{X}} = (\mathbf{X}^{-1})^T = (\mathbf{X}^T)^{-1} \tag{57}$$

$$\frac{\partial \det(\mathbf{X}^k)}{\partial \mathbf{X}} = k\det(\mathbf{X}^k)\mathbf{X}^{-T} \tag{58}$$

2.2 逆矩阵的导数

从[27]我们有基本等式

$$\frac{\partial \mathbf{Y}^{-1}}{\partial x} = -\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\mathbf{Y}^{-1} \tag{59}$$

由此可见

$$\frac{\partial(\mathbf{X}^{-1})_{kl}}{\partial X_{ij}} = -(\mathbf{X}^{-1})_{ki}(\mathbf{X}^{-1})_{jl} \tag{60}$$

$$\frac{\partial\mathbf{a}^T\mathbf{X}^{-1}\mathbf{b}}{\partial\mathbf{X}} = -\mathbf{X}^{-T}\mathbf{a}\mathbf{b}^T\mathbf{X}^{-T} \tag{61}$$

$$\frac{\partial\det(\mathbf{X}^{-1})}{\partial\mathbf{X}} = -\det(\mathbf{X}^{-1})(\mathbf{X}^{-1})^T \tag{62}$$

$$\frac{\partial\mathrm{Tr}(\mathbf{A}\mathbf{X}^{-1}\mathbf{B})}{\partial\mathbf{X}} = -(\mathbf{X}^{-1}\mathbf{B}\mathbf{A}\mathbf{X}^{-1})^T \tag{63}$$

$$\frac{\partial\mathrm{Tr}((\mathbf{X}+\mathbf{A})^{-1})}{\partial\mathbf{X}} = -((\mathbf{X}+\mathbf{A})^{-1}(\mathbf{X}+\mathbf{A})^{-1})^T \tag{64}$$

从[32]我们有以下结果：设 $\mathbf{A}$是一个$n\times n$可逆方阵， $\mathbf{W}$是 $\mathbf{A}$的逆矩阵， $J(\mathbf{A})$是一个关于 $\mathbf{A}$的n元可微函数，那么 J关于 $\mathbf{A}$和 $\mathbf{W}$的偏导数满足

$$\frac{\partial J}{\partial\mathbf{A}} = -\mathbf{A}^{-T}\frac{\partial J}{\partial\mathbf{W}}\mathbf{A}^{-T}$$

2.3 特征值的导数

$$\frac{\partial}{\partial\mathbf{X}}\sum\mathrm{eig}(\mathbf{X}) = \frac{\partial}{\partial\mathbf{X}}\mathrm{Tr}(\mathbf{X}) = \mathbf{I} \tag{65}$$

$$\frac{\partial}{\partial\mathbf{X}}\prod\mathrm{eig}(\mathbf{X}) = \frac{\partial}{\partial\mathbf{X}}\det(\mathbf{X}) = \det(\mathbf{X})\mathbf{X}^{-T} \tag{66}$$

如果 $\mathbf{A}$是实对称矩阵， λ_i和 $\mathbf{v}_i$是 $\mathbf{A}$的不同特征值和特征向量（见(276)），则[33]$\partial\lambda_i = \mathbf{v}_i{}^T\partial(\mathbf{A})\,\mathbf{v}_i$

$$\tag{67}$$

$$\partial\mathbf{v}_i = (\lambda_i\mathbf{I}-\mathbf{A})^+\partial(\mathbf{A})\mathbf{v}_i \tag{68}$$

2.4 矩阵、向量和标量形式的导数

2.4.1 一阶

$$\frac{\partial\mathbf{x}^T\mathbf{a}}{\partial\mathbf{x}} = \frac{\partial\mathbf{a}^T\mathbf{x}}{\partial\mathbf{x}} = \mathbf{a} \tag{69}$$

$$\frac{\partial\mathbf{a}^T\mathbf{X}\mathbf{b}}{\partial\mathbf{X}} = \mathbf{a}\mathbf{b}^T \tag{70}$$

$$\frac{\partial\mathbf{a}^T\mathbf{X}^T\mathbf{b}}{\partial\mathbf{X}} = \mathbf{b}\mathbf{a}^T \tag{71}$$

$$\frac{\partial\mathbf{a}^T\mathbf{X}\mathbf{a}}{\partial\mathbf{X}} = \frac{\partial\mathbf{a}^T\mathbf{X}^T\mathbf{a}}{\partial\mathbf{X}} = \mathbf{a}\mathbf{a}^T \tag{72}$$

$$\frac{\partial\mathbf{X}}{\partial X_{ij}} = \mathbf{J}^{ij} \tag{73}$$

$$\frac{\partial(\mathbf{X}\mathbf{A})_{ij}}{\partial X_{mn}} = \delta_{im}(\mathbf{A})_{nj} = (\mathbf{J}^{mn}\mathbf{A})_{ij} \tag{74}$$

$$\frac{\partial(\mathbf{X}^T\mathbf{A})_{ij}}{\partial X_{mn}} = \delta_{in}(\mathbf{A})_{mj} = (\mathbf{J}^{nm}\mathbf{A})_{ij} \tag{75}$$

2.4.2 二阶

$$\frac{\partial}{\partial X_{ij}} \sum_{klmn} X_{kl} X_{mn} = 2 \sum_{kl} X_{kl} \quad (76)$$

$$\frac{\partial \mathbf{b}^T \mathbf{X}^T \mathbf{X} \mathbf{c}}{\partial \mathbf{X}} = \mathbf{X}(\mathbf{b}\mathbf{c}^T + \mathbf{c}\mathbf{b}^T) \quad (77)$$

$$\frac{\partial (\mathbf{B}\mathbf{x}+\mathbf{b})^T \mathbf{C} (\mathbf{D}\mathbf{x}+\mathbf{d})}{\partial \mathbf{x}} = \mathbf{B}^T \mathbf{C}(\mathbf{D}\mathbf{x}+\mathbf{d}) + \mathbf{D}^T \mathbf{C}^T (\mathbf{B}\mathbf{x}+\mathbf{b}) \quad (78)$$

$$\frac{\partial (\mathbf{X}^T \mathbf{B} \mathbf{X})_{kl}}{\partial X_{ij}} = \delta_{lj} (\mathbf{X}^T \mathbf{B})_{ki} + \delta_{kj} (\mathbf{B}\mathbf{X})_{il} \quad (79)$$

$$\frac{\partial (\mathbf{X}^T \mathbf{B} \mathbf{X})}{\partial X_{ij}} = \mathbf{X}^T \mathbf{B} \mathbf{J}^{ij} + \mathbf{J}^{ji} \mathbf{B} \mathbf{X} \qquad (\mathbf{J}^{ij})_{kl} = \delta_{ik}\delta_{jl} \quad (80)$$

有关单入矩阵 $\mathbf{J}^{ij}$的有用性质，请参见第9.7节

$$\frac{\partial \mathbf{x}^T \mathbf{B} \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{B} + \mathbf{B}^T)\mathbf{x} \quad (81)$$

$$\frac{\partial \mathbf{b}^T \mathbf{X}^T \mathbf{D} \mathbf{X} \mathbf{c}}{\partial \mathbf{X}} = \mathbf{D}^T \mathbf{X} \mathbf{b} \mathbf{c}^T + \mathbf{D} \mathbf{X} \mathbf{c} \mathbf{b}^T \quad (82)$$

$$\frac{\partial}{\partial \mathbf{X}} (\mathbf{X}\mathbf{b}+\mathbf{c})^T \mathbf{D} (\mathbf{X}\mathbf{b}+\mathbf{c}) = (\mathbf{D}+\mathbf{D}^T)(\mathbf{X}\mathbf{b}+\mathbf{c})\mathbf{b}^T \quad (83)$$

假设 $\mathbf{W}$ 是对称的，那么

$$\frac{\partial}{\partial \mathbf{s}} (\mathbf{x}-\mathbf{A}\mathbf{s})^T \mathbf{W} (\mathbf{x}-\mathbf{A}\mathbf{s}) = -2\mathbf{A}^T \mathbf{W} (\mathbf{x}-\mathbf{A}\mathbf{s}) \quad (84)$$

$$\frac{\partial}{\partial \mathbf{x}} (\mathbf{x}-\mathbf{s})^T \mathbf{W} (\mathbf{x}-\mathbf{s}) = 2\mathbf{W}(\mathbf{x}-\mathbf{s}) \quad (85)$$

$$\frac{\partial}{\partial \mathbf{s}} (\mathbf{x}-\mathbf{s})^T \mathbf{W} (\mathbf{x}-\mathbf{s}) = -2\mathbf{W}(\mathbf{x}-\mathbf{s}) \quad (86)$$

$$\frac{\partial}{\partial \mathbf{x}} (\mathbf{x}-\mathbf{A}\mathbf{s})^T \mathbf{W} (\mathbf{x}-\mathbf{A}\mathbf{s}) = 2\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) \quad (87)$$

$$\frac{\partial}{\partial \mathbf{A}} (\mathbf{x}-\mathbf{A}\mathbf{s})^T \mathbf{W} (\mathbf{x}-\mathbf{A}\mathbf{s}) = -2\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s})\mathbf{s}^T \quad (88)$$

作为具有复数值的情况，以下成立

$$\frac{\partial (a - \mathbf{x}^H \mathbf{b})^2}{\partial \mathbf{x}} = -2\mathbf{b}(a - \mathbf{x}^H \mathbf{b})^* \quad (89)$$

这个公式也被称为LMS算法[14]

2.4.3 高阶和非线性

$$\frac{\partial (\mathbf{X}^n)_{kl}}{\partial X_{ij}} = \sum_{r=0}^{n-1} (\mathbf{X}^r \mathbf{J}^{ij} \mathbf{X}^{n-1-r})_{kl} \quad (90)$$

有关以上内容的证明，请参见B.1.3。

$$\frac{\partial}{\partial \mathbf{X}} \mathbf{a}^T \mathbf{X}^n \mathbf{b} = \sum_{r=0}^{n-1} (\mathbf{X}^r)^T \mathbf{a} \mathbf{b}^T (\mathbf{X}^{n-1-r})^T \quad (91)$$

$$\begin{aligned}\frac{\partial}{\partial \mathbf{X}}\mathbf{a}^T(\mathbf{X}^n)^T\mathbf{X}^n\mathbf{b} &= \sum_{r=0}^{n-1}\Big[\mathbf{X}^{n-1-r}\mathbf{a}\mathbf{b}^T(\mathbf{X}^n)^T\mathbf{X}^r \\ &\quad +(\mathbf{X}^r)^T\mathbf{X}^n\mathbf{a}\mathbf{b}^T(\mathbf{X}^{n-1-r})^T\Big] \qquad (92)\end{aligned}$$

有关证明，请参见B.1.3。
假设 $\mathbf{s}$和 $\mathbf{r}$是 $\mathbf{x}$的函数，即 $\mathbf{s}=\mathbf{s}(\mathbf{x}),\mathbf{r}=\mathbf{r}(\mathbf{x})$，且 $\mathbf{A}$是一个常数，那么

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{s}^T\mathbf{A}\mathbf{r} = \left[\frac{\partial \mathbf{s}}{\partial \mathbf{x}}\right]^T\mathbf{A}\mathbf{r}+\left[\frac{\partial \mathbf{r}}{\partial \mathbf{x}}\right]^T\mathbf{A}^T\mathbf{s} \qquad (93)$$

$$\begin{aligned}\frac{\partial}{\partial \mathbf{x}}\frac{(\mathbf{A}\mathbf{x})^T(\mathbf{A}\mathbf{x})}{(\mathbf{B}\mathbf{x})^T(\mathbf{B}\mathbf{x})} &= \frac{\partial}{\partial \mathbf{x}}\frac{\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}^T\mathbf{B}\mathbf{x}} &(94)\\ &= 2\frac{\mathbf{A}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{B}\mathbf{x}}-2\frac{\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}\mathbf{B}^T\mathbf{B}\mathbf{x}}{(\mathbf{x}^T\mathbf{B}^T\mathbf{B}\mathbf{x})^2} &(95)\end{aligned}$$

2.4.4 梯度和海森矩阵

利用上述结果，我们有梯度和海森矩阵

$$\begin{aligned} f &= \mathbf{x}^T\mathbf{A}\mathbf{x}+\mathbf{b}^T\mathbf{x} &(96)\\ \nabla_{\mathbf{x}}f=\frac{\partial f}{\partial \mathbf{x}} &= (\mathbf{A}+\mathbf{A}^T)\mathbf{x}+\mathbf{b} &(97)\\ \frac{\partial^2 f}{\partial \mathbf{x}\partial \mathbf{x}^T} &= \mathbf{A}+\mathbf{A}^T &(98)\end{aligned}$$

2.5 迹的导数

假设 $F(\mathbf{X})$ 是 X的每个元素的可微函数。那么有

$$\frac{\partial \mathrm{Tr}(F(\mathbf{X}))}{\partial \mathbf{X}} = f(\mathbf{X})^T$$

其中 $f(\cdot)$ 是 $F(\cdot)$ 的标量导数。

2.5.1 一阶

$$\begin{aligned}\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}) &= \mathbf{I} &(99)\\ \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}\mathbf{A}) &= \mathbf{A}^T &(100)\\ \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\mathbf{X}\mathbf{B}) &= \mathbf{A}^T\mathbf{B}^T &(101)\\ \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\mathbf{X}^T\mathbf{B}) &= \mathbf{B}\mathbf{A} &(102)\\ \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{A}) &= \mathbf{A} &(103)\\ \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\mathbf{X}^T) &= \mathbf{A} &(104)\\ \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\otimes\mathbf{X}) &= \mathrm{Tr}(\mathbf{A})\mathbf{I} &(105)\end{aligned}$$

2.5.2 二阶

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^2) &= 2\mathbf{X}^T && (106)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^2\mathbf{B}) &= (\mathbf{XB}+\mathbf{BX})^T && (107)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{BX}) &= \mathbf{BX}+\mathbf{B}^T\mathbf{X} && (108)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{BXX}^T) &= \mathbf{BX}+\mathbf{B}^T\mathbf{X} && (109)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{XX}^T\mathbf{B}) &= \mathbf{BX}+\mathbf{B}^T\mathbf{X} && (110)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{XBX}^T) &= \mathbf{XB}^T+\mathbf{XB} && (111)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{BX}^T\mathbf{X}) &= \mathbf{XB}^T+\mathbf{XB} && (112)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{XB}) &= \mathbf{XB}^T+\mathbf{XB} && (113)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AXBX}) &= \mathbf{A}^T\mathbf{X}^T\mathbf{B}^T+\mathbf{B}^T\mathbf{X}^T\mathbf{A}^T && (114)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{X}) &= \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{XX}^T) \quad = \quad 2\mathbf{X} && (115)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{B}^T\mathbf{X}^T\mathbf{CXB}) &= \mathbf{C}^T\mathbf{XBB}^T+\mathbf{CXBB}^T && (116)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\left[\mathbf{X}^T\mathbf{BXC}\right] &= \mathbf{BXC}+\mathbf{B}^T\mathbf{XC}^T && (117)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AXBX}^T\mathbf{C}) &= \mathbf{A}^T\mathbf{C}^T\mathbf{XB}^T+\mathbf{CAXB} && (118)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\Big[(\mathbf{AXB}+\mathbf{C})(\mathbf{AXB}+\mathbf{C})^T\Big] &= 2\mathbf{A}^T(\mathbf{AXB}+\mathbf{C})\mathbf{B}^T && (119)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}\otimes\mathbf{X}) &= \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X})\mathrm{Tr}(\mathbf{X}) = 2\mathrm{Tr}(\mathbf{X})\mathbf{I} && (120)
\end{aligned}$$

见[7]。

2.5.3 高阶

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^k) &= k(\mathbf{X}^{k-1})^T && (121)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AX}^k) &= \sum_{r=0}^{k-1}(\mathbf{X}^r\mathbf{AX}^{k-r-1})^T && (122)\\
\tfrac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\left[\mathbf{B}^T\mathbf{X}^T\mathbf{CXX}^T\mathbf{CXB}\right] &= \mathbf{CXX}^T\mathbf{CXBB}^T\\
&\quad +\mathbf{C}^T\mathbf{XBB}^T\mathbf{X}^T\mathbf{C}^T\mathbf{X}\\
&\quad +\mathbf{CXBB}^T\mathbf{X}^T\mathbf{CX}\\
&\quad +\mathbf{C}^T\mathbf{XX}^T\mathbf{C}^T\mathbf{XBB}^T && (123)
\end{aligned}$$

2.5.4 其他

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{A}\mathbf{X}^{-1}\mathbf{B}) = -(\mathbf{X}^{-1}\mathbf{B}\mathbf{A}\mathbf{X}^{-1})^T = -\mathbf{X}^{-T}\mathbf{A}^T\mathbf{B}^T\mathbf{X}^{-T} \tag{124}$$

假设 $\mathbf{B}$ 和 $\mathbf{C}$ 是对称的，那么

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}\Big[(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}\mathbf{A}\Big] = -(\mathbf{C}\mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1})(\mathbf{A}+\mathbf{A}^T)(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} \tag{125}$$

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}\Big[(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{B}\mathbf{X})\Big] = -2\mathbf{C}\mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{B}\mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} + 2\mathbf{B}\mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} \tag{126}$$

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}\Big[(\mathbf{A}+\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{B}\mathbf{X})\Big] = -2\mathbf{C}\mathbf{X}(\mathbf{A}+\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{B}\mathbf{X}(\mathbf{A}+\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} + 2\mathbf{B}\mathbf{X}(\mathbf{A}+\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} \tag{127}$$

见[7]。

$$\frac{\partial \mathrm{Tr}(\sin(\mathbf{X}))}{\partial \mathbf{X}} = \cos(\mathbf{X})^T \tag{128}$$

2.6 向量范数的导数

2.6.1 二范数

$$\frac{\partial}{\partial \mathbf{x}} ||\mathbf{x}-\mathbf{a}||_2 = \frac{\mathbf{x}-\mathbf{a}}{||\mathbf{x}-\mathbf{a}||_2} \tag{129}$$

$$\frac{\partial}{\partial \mathbf{x}} \frac{\mathbf{x}-\mathbf{a}}{\|\mathbf{x}-\mathbf{a}\|_2} = \frac{\mathbf{I}}{\|\mathbf{x}-\mathbf{a}\|_2} - \frac{(\mathbf{x}-\mathbf{a})(\mathbf{x}-\mathbf{a})^T}{\|\mathbf{x}-\mathbf{a}\|_2^3} \tag{130}$$

$$\frac{\partial ||\mathbf{x}||_2^2}{\partial \mathbf{x}} = \frac{\partial ||\mathbf{x}^T\mathbf{x}||_2}{\partial \mathbf{x}} = 2\mathbf{x} \tag{131}$$

2.7 矩阵范数的导数

有关矩阵范数的更多信息，请参阅第10.4节。

2.7.1 弗罗贝尼乌斯范数

$$\frac{\partial}{\partial \mathbf{X}} ||\mathbf{X}||_{\mathrm{F}}^2 = \frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{X}\mathbf{X}^H) = 2\mathbf{X} \tag{132}$$

参见（248）。 请注意，这也是方程119中结果的特殊情况。

2.8 结构化矩阵的导数

假设矩阵 $\mathbf{A}$具有某种结构，即对称的，Toeplitz的等等。
在这种情况下，前一节的导数通常不适用。
相反，考虑以下关于标量函数
$f(\mathbf{A})$

$$\frac{\text{df的一般规}}{dA_{ij}} = \sum_{kl} \frac{\partial f}{\partial A_{kl}} \frac{\partial A_{kl}}{\partial A_{ij}} = \mathrm{Tr}\left[\left[\frac{\partial f}{\partial \mathbf{A}}\right]^T \frac{\partial \mathbf{A}}{\partial A_{ij}}\right] \tag{133}$$

本文档中，关于自身的矩阵的矩阵差异被称为 $\mathbf{A}$的结构矩阵，并且简单地定义为

$$\frac{\partial \mathbf{A}}{\partial A_{ij}} = \mathbf{S}^{ij} \tag{134}$$

如果 $\mathbf{A}$没有特殊结构，我们只需简单地将 $\mathbf{S}^{ij} = \mathbf{J}^{ij}$，也就是说，结构矩阵就是单个元素的矩阵。 许多结构都可以用单个元素矩阵表示，更多结构矩阵的示例请参见第9.7.6节。

2.8.1 链式法则

有时候目标是找到另一个矩阵的函数的导数。 设 $\mathbf{U} = f(\mathbf{X})$，目标是找到函数g($\mathbf{U}$)相对于 $\mathbf{X}$的导数：$\partial g(\mathbf{U})$

$$\frac{}{\partial \mathbf{X}} = \frac{\partial g(f(\mathbf{X}))}{\partial \mathbf{X}} \tag{135}$$

然后链式法则可以写成以下形式：

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\partial g(\mathbf{U})}{\partial x_{ij}} = \sum_{k=1}^{M}\sum_{l=1}^{N} \frac{\partial g(\mathbf{U})}{\partial u_{kl}} \frac{\partial u_{kl}}{\partial x_{ij}} \tag{136}$$

使用矩阵符号，可以写成：

$$\frac{\partial g(\mathbf{U})}{\partial X_{ij}} = \mathrm{Tr}\Big[(\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}})^T \frac{\partial \mathbf{U}}{\partial X_{ij}}\Big]. \tag{137}$$

2.8.2 对称

如果 $\mathbf{A}$是对称的，那么 $\mathbf{S}^{ij} = \mathbf{J}^{ij} + \mathbf{J}^{ji} - \mathbf{J}^{ij}\mathbf{J}^{ij}$，因此

$$\frac{df}{d\mathbf{A}} = \left[\frac{\partial f}{\partial \mathbf{A}}\right] + \left[\frac{\partial f}{\partial \mathbf{A}}\right]^T - \mathrm{diag}\left[\frac{\partial f}{\partial \mathbf{A}}\right] \tag{138}$$

也就是说，例如（[5]）：

$$\frac{\partial \mathrm{Tr}(\mathbf{AX})}{\partial \mathbf{X}} = \mathbf{A} + \mathbf{A}^T - (\mathbf{A} \circ \mathbf{I})\text{，参见（14} \tag{139}$$

$$\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} = \det(\mathbf{X})(2\mathbf{X}^{-1} - (\mathbf{X}^{-1} \circ \mathbf{I})) \tag{140}$$

$$\frac{\partial \ln \det(\mathbf{X})}{\partial \mathbf{X}} = 2\mathbf{X}^{-1} - (\mathbf{X}^{-1} \circ \mathbf{I}) \tag{141}$$

2.8.3 对角

如果 $\mathbf{X}$是对角的，那么（[19]）：

$$\frac{\partial \mathrm{Tr}(\mathbf{AX})}{\partial \mathbf{X}} = \mathbf{A} \circ \mathbf{I} \tag{142}$$

2.8.4 Toeplitz

就像对称矩阵和对角矩阵一样，Toeplitz矩阵也具有特殊的结构，在对具有Toeplitz结构的矩阵进行导数时需要考虑这一点。

$$
\begin{aligned}
&\frac{\partial \mathrm{Tr}(\mathbf{AT})}{\partial \mathbf{T}} \\
=\quad &\frac{\partial \mathrm{Tr}(\mathbf{TA})}{\partial \mathbf{T}} \\
=\quad &\begin{bmatrix}
\mathrm{Tr}(\mathbf{A}) & \mathrm{Tr}([\mathbf{A}^T]_{n1}) & \mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{n-1,2}) & \cdots & A_{n1} \\
\mathrm{Tr}([\mathbf{A}^T]_{1n})) & \mathrm{Tr}(\mathbf{A}) & \ddots & \ddots & \vdots \\
\mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{2,n-1}) & \ddots & \ddots & \ddots & \mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{n-1,2}) \\
\vdots & \ddots & \ddots & \ddots & \mathrm{Tr}([\mathbf{A}^T]_{n1}) \\
A_{1n} & \cdots & \mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{2,n-1}) & \mathrm{Tr}([\mathbf{A}^T]_{1n})) & \mathrm{Tr}(\mathbf{A})
\end{bmatrix} \\
\equiv\quad &\boldsymbol{\alpha}(\mathbf{A})
\end{aligned} \tag{143}
$$

正如可以看到的，导数 $\boldsymbol{\alpha}(\mathbf{A})$也具有Toeplitz结构。 对角线上的每个值是 $\mathbf{A}$中对角线上的所有值的和，主对角线旁边的对角线上的值等于 $\mathbf{A}^T$中主对角线旁边的对角线上的值的和。 这个结果仅对无约束的Toeplitz矩阵有效。 如果Toeplitz矩阵也是对称的，同样的导数得到

$$
\frac{\partial \mathrm{Tr}(\mathbf{AT})}{\partial \mathbf{T}} = \frac{\partial \mathrm{Tr}(\mathbf{TA})}{\partial \mathbf{T}} = \boldsymbol{\alpha}(\mathbf{A}) + \boldsymbol{\alpha}(\mathbf{A})^T - \boldsymbol{\alpha}(\mathbf{A}) \circ \mathbf{I} \tag{144}
$$

3 逆矩阵

3.1 基本

3.1.1 定义

矩阵 $\mathbf{A}$的逆矩阵 $\mathbf{A}^{-1}$是指满足以下条件的矩阵A $\in \mathbb{C}^{n\times n}$：

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}, \tag{145}$$

其中 $\mathbf{I}$是 $n \times n$单位矩阵。 如果 $\mathbf{A}^{-1}$存在，则称 $\mathbf{A}$为非奇异矩阵。
否则， $\mathbf{A}$被称为奇异矩阵（参见例如[12]）。

3.1.2 余子式和伴随矩阵

矩阵 $\mathbf{A}$的子矩阵，记为 $[\mathbf{A}]_{ij}$，是通过删除 $\mathbf{A}$的第 i行和第 j列得到的一个 $(n-1) \times (n-1)$ 矩阵。 矩阵的 (i,j)余子式定义为

$$\mathrm{cof}(\mathbf{A}, i, j) = (-1)^{i+j} \det([\mathbf{A}]_{ij}), \tag{146}$$

余子式矩阵可以由余子式创建

$$\mathrm{cof}(\mathbf{A}) = \begin{bmatrix} \mathrm{cof}(\mathbf{A},1,1) & \cdots & \mathrm{cof}(\mathbf{A},1,n) \\ \vdots & \mathrm{cof}(\mathbf{A},i,j) & \vdots \\ \mathrm{cof}(\mathbf{A},n,1) & \cdots & \mathrm{cof}(\mathbf{A},n,n) \end{bmatrix} \tag{147}$$

伴随矩阵是余子式矩阵的转置

$$\mathrm{adj}(\mathbf{A}) = (\mathrm{cof}(\mathbf{A}))^T, \tag{148}$$

3.1.3 行列式

矩阵 $\mathbf{A} \in \mathbb{C}^{n\times n}$的行列式定义如下（见[12]）

$$\det(\mathbf{A}) = \sum_{j=1}^{n} (-1)^{j+1} A_{1j} \text{行列式} ([\mathbf{A}]_{1j}) \tag{149}$$

$$= \sum_{j=1}^{n} A_{1j} \text{余子式} (\mathbf{A}, 1, j). \tag{150}$$

3.1.4 构造

逆矩阵可以通过伴随矩阵构造

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \cdot \mathrm{adj}(\mathbf{A}) \tag{151}$$

对于2 ×2矩阵的情况，请参见第1.3节。

3.1.5 条件数

矩阵 $c(\mathbf{A})$的条件数是矩阵的最大奇异值与最小奇异值之比（参见第5.3节的奇异值）

$$c(\mathbf{A}) = \frac{d_+}{d_-} \tag{152}$$

条件数可以用来衡量矩阵的奇异性。如果条件数很大，表示矩阵接近奇异。 条件数也可以从矩阵范数估计。在这里

$$c(\mathbf{A}) = \|\mathbf{A}\| \cdot \|\mathbf{A}^{-1}\|, \tag{153}$$

其中$\|\cdot\|$是一种范数，例如1-范数，2-范数， ∞-范数或者Frobenius范数（详见第10.4节关于矩阵范数的更多内容）。

矩阵 $\mathbf{A}$的2-范数等于 $\sqrt{(\max(\text{eig}(\mathbf{A}^H\mathbf{A})))}$ [1_2, p.57]。对于对称矩阵，这可以简化为 $||\mathbf{A}||_2 = \max(|\text{eig}(\mathbf{A})|)$ [1_2, p.394]。如果矩阵是对称且正定的， $||\mathbf{A}||_2 = \max(\text{eig}(\mathbf{A}))$。 基于$_2$-范数的条件数因此简化为

$$\|\mathbf{A}\|_2\|\mathbf{A}^{-1}\|_2 = \max(\text{eig}(\mathbf{A}))\,\max(\text{eig}(\mathbf{A}^{-1})) = \frac{\max(\text{eig}(\mathbf{A}))}{\min(\text{eig}(\mathbf{A}))}. \tag{154}$$

3.2 精确关系

3.2.1 基础

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{155}$$

3.2.2 伍德伯里恒等式

伍德伯里恒等式有很多变种。 其中的一种可以在[12]中找到

$$(\mathbf{A} + \mathbf{CBC}^T)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}(\mathbf{B}^{-1} + \mathbf{C}^T\mathbf{A}^{-1}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{A}^{-1} \tag{156}$$

$$(\mathbf{A} + \mathbf{UBV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{B}^{-1} + \mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1} \tag{157}$$

如果 $\mathbf{P}, \mathbf{R}$是正定的，那么（见[30]）

$$(\mathbf{P}^{-1} + \mathbf{B}^T\mathbf{R}^{-1}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{R}^{-1} = \mathbf{PB}^T(\mathbf{BPB}^T + \mathbf{R})^{-1} \tag{158}$$

3.2.3 凯拉斯变体

$$(\mathbf{A} + \mathbf{BC})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1} \tag{159}$$

参见[4, 第153页]。

3.2.4 谢尔曼-莫里森

$$(\mathbf{A} + \mathbf{bc}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{bc}^T\mathbf{A}^{-1}}{1 + \mathbf{c}^T\mathbf{A}^{-1}\mathbf{b}} \tag{160}$$

3.2.5 西尔尔恩恒等式集合

以下一组恒等式可在[25, 第151页]中找到,

$$\begin{aligned}
(\mathbf{I}+\mathbf{A}^{-1})^{-1} &= \mathbf{A}(\mathbf{A}+\mathbf{I})^{-1} && (161)\\
(\mathbf{A}+\mathbf{B}\mathbf{B}^T)^{-1}\mathbf{B} &= \mathbf{A}^{-1}\mathbf{B}(\mathbf{I}+\mathbf{B}^T\mathbf{A}^{-1}\mathbf{B})^{-1} && (162)\\
(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1} &= \mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B} = \mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A} && (163)\\
\mathbf{A}-\mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A} &= \mathbf{B}-\mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B} && (164)\\
\mathbf{A}^{-1}+\mathbf{B}^{-1} &= \mathbf{A}^{-1}(\mathbf{A}+\mathbf{B})\mathbf{B}^{-1} && (165)\\
(\mathbf{I}+\mathbf{A}\mathbf{B})^{-1} &= \mathbf{I}-\mathbf{A}(\mathbf{I}+\mathbf{B}\mathbf{A})^{-1}\mathbf{B} && (166)\\
(\mathbf{I}+\mathbf{A}\mathbf{B})^{-1}\mathbf{A} &= \mathbf{A}(\mathbf{I}+\mathbf{B}\mathbf{A})^{-1} && (167)
\end{aligned}$$

3.2.6 内积逆的秩-1更新

记 $\mathbf{A}=(\mathbf{X}^T\mathbf{X})^{-1}$并且 $\mathbf{X}$被扩展以包括一个新的列向量
在末尾$\tilde{\mathbf{X}} = [\mathbf{X} \quad \mathbf{v}]$. 然后 [34]

$$(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1} = \begin{bmatrix} \mathbf{A}+\frac{\mathbf{A}\mathbf{X}^T\mathbf{v}\mathbf{v}^T\mathbf{X}\mathbf{A}^T}{\mathbf{v}^T\mathbf{v}-\mathbf{v}^T\mathbf{X}\mathbf{A}\mathbf{X}^T\mathbf{v}} & \frac{-\mathbf{A}\mathbf{X}^T\mathbf{v}}{\mathbf{v}^T\mathbf{v}-\mathbf{v}^T\mathbf{X}\mathbf{A}\mathbf{X}^T\mathbf{v}} \\ \frac{-\mathbf{v}^T\mathbf{X}\mathbf{A}^T}{\mathbf{v}^T\mathbf{v}-\mathbf{v}^T\mathbf{X}\mathbf{A}\mathbf{X}^T\mathbf{v}} & \frac{1}{\mathbf{v}^T\mathbf{v}-\mathbf{v}^T\mathbf{X}\mathbf{A}\mathbf{X}^T\mathbf{v}} \end{bmatrix}$$

3.2.7 Moore-Penrose逆的秩-1更新

下面是实值矩阵的Moore-Penrose伪逆的秩-1更新，证明可以在[18]中找到。 矩阵 $\mathbf{G}$定义如下:

$$(\mathbf{A}+\mathbf{c}\mathbf{d}^T)^+ = \mathbf{A}^+ + \mathbf{G} \tag{168}$$

使用以下符号

$$\begin{aligned}
\beta &= 1+\mathbf{d}^T\mathbf{A}^+\mathbf{c} && (169)\\
\mathbf{v} &= \mathbf{A}^+\mathbf{c} && (170)\\
\mathbf{n} &= (\mathbf{A}^+)^T\mathbf{d} && (171)\\
\mathbf{w} &= (\mathbf{I}-\mathbf{A}\mathbf{A}^+)\mathbf{c} && (172)\\
\mathbf{m} &= (\mathbf{I}-\mathbf{A}^+\mathbf{A})^T\mathbf{d} && (173)
\end{aligned}$$

解决方案分为六种不同情况，取决于实体 $||\mathbf{w}||$,
$||\mathbf{m}||$, 和 β. 请注意，对于任何（列）向量 $\mathbf{v}$，有 $\mathbf{v}^+ =$
$\mathbf{v}^T(\mathbf{v}^T\mathbf{v})^{-1} = \frac{\mathbf{v}^T}{||\mathbf{v}||^2}$. 解决方案是:

第1种情况：如果 $||\mathbf{w}|| \neq 0$ 和 $||\mathbf{m}|| \neq 0$。那么

$$\begin{aligned}
\mathbf{G} &= -\mathbf{v}\mathbf{w}^+ - (\mathbf{m}^+)^T\mathbf{n}^T + \beta(\mathbf{m}^+)^T\mathbf{w}^+ && (174)\\
&= -\frac{1}{||\mathbf{w}||^2}\mathbf{v}\mathbf{w}^T - \frac{1}{||\mathbf{m}||^2}\mathbf{m}\mathbf{n}^T + \frac{\beta}{||\mathbf{m}||^2||\mathbf{w}||^2}\mathbf{m}\mathbf{w}^T && (175)
\end{aligned}$$

第2个案例中的第6个情况：如果 $||\mathbf{w}|| = 0$且 $||\mathbf{m}|| \neq 0$且 $\beta = 0$。那么

$$\begin{aligned}
\mathbf{G} &= -\mathbf{v}\mathbf{v}^+\mathbf{A}^+ - (\mathbf{m}^+)^T\mathbf{n}^T && (176)\\
&= -\frac{1}{||\mathbf{v}||^2}\mathbf{v}\mathbf{v}^T\mathbf{A}^+ - \frac{1}{||\mathbf{m}||^2}\mathbf{m}\mathbf{n}^T && (177)
\end{aligned}$$

第3个案例中的第6个情况：如果 $||\mathbf{w}||= 0$且 $\beta= 0$。那么

$$\mathbf{G} = \frac{1}{\beta}\mathbf{m}\mathbf{v}^T\mathbf{A}^+ - \frac{\beta}{||\mathbf{v}||^2||\mathbf{m}||^2+|\beta|^2}\left(\frac{||\mathbf{v}||^2}{\beta}\mathbf{m}+\mathbf{v}\right)\left(\frac{||\mathbf{m}||^2}{\beta}(\mathbf{A}^+)^T\mathbf{v}+\mathbf{n}\right)^T \tag{178}$$

第4个案例中的第6个情况：如果 $||\mathbf{w}|| \neq 0$且 $||\mathbf{m}||= 0$且 $\beta= 0$。那么

$$\begin{aligned}\mathbf{G} &= -\mathbf{A}^+\mathbf{n}\mathbf{n}^+ - \mathbf{v}\mathbf{w}^+ & (179)\\ &= -\frac{1}{||\mathbf{n}||^2}\mathbf{A}^+\mathbf{n}\mathbf{n}^T - \frac{1}{||\mathbf{w}||^2}\mathbf{v}\mathbf{w}^T & (180)\end{aligned}$$

第5个案例：如果 $||\mathbf{m}||= 0$且 $\beta= 0$。那么

$$\mathbf{G} = \frac{1}{\beta}\mathbf{A}^+\mathbf{n}\mathbf{w}^T - \frac{\beta}{||\mathbf{n}||^2||\mathbf{w}||^2+|\beta|^2}\left(\frac{||\mathbf{w}||^2}{\beta}\mathbf{A}^+\mathbf{n}+\mathbf{v}\right)\left(\frac{||\mathbf{n}||^2}{\beta}\mathbf{w}+\mathbf{n}\right)^T \tag{181}$$

第6个案例：如果 $||\mathbf{w}||= 0$且 $||\mathbf{m}||= 0$且 $\beta= 0$。那么

$$\begin{aligned}\mathbf{G} &= -\mathbf{v}\mathbf{v}^+\mathbf{A}^+ - \mathbf{A}^+\mathbf{n}\mathbf{n}^+ + \mathbf{v}^+\mathbf{A}^+\mathbf{n}\mathbf{v}\mathbf{n}^+ & (182)\\ &= -\frac{1}{||\mathbf{v}||^2}\mathbf{v}\mathbf{v}^T\mathbf{A}^+ - \frac{1}{||\mathbf{n}||^2}\mathbf{A}^+\mathbf{n}\mathbf{n}^T + \frac{\mathbf{v}^T\mathbf{A}^+\mathbf{n}}{||\mathbf{v}||^2||\mathbf{n}||^2}\mathbf{v}\mathbf{n}^T & (183)\end{aligned}$$

3.3 对逆矩阵的影响

$$\text{If} \quad (\mathbf{A}+\mathbf{B})^{-1} = \mathbf{A}^{-1}+\mathbf{B}^{-1} \quad \text{然后} \quad \mathbf{A}\mathbf{B}^{-1}\mathbf{A} = \mathbf{B}\mathbf{A}^{-1}\mathbf{B} \tag{184}$$

参见[25]。

3.3.1 正定恒等式

假设 $\mathbf{P}$，$\mathbf{R}$为正定且可逆的，那么

$$(\mathbf{P}^{-1}+\mathbf{B}^T\mathbf{R}^{-1}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{R}^{-1} = \mathbf{P}\mathbf{B}^T(\mathbf{B}\mathbf{P}\mathbf{B}^T+\mathbf{R})^{-1} \tag{185}$$

参见[30]。

3.4 近似

下面的恒等式被称为矩阵的$Neuman$级数，当 $|\lambda_i| <1$对于所有特征值 λ_i时成立

$$(\mathbf{I}-\mathbf{A})^{-1} = \sum_{n=0}^{\infty}\mathbf{A}^n \tag{186}$$

等价于

$$(\mathbf{I}+\mathbf{A})^{-1} = \sum_{n=0}^{\infty}(-1)^n\mathbf{A}^n \tag{187}$$

当所有特征值λ_i的绝对值都小于1时，对于 $n\to\infty$，有 $\mathbf{A}\to 0$，并且以下近似成立

$$\begin{aligned}(\mathbf{I}-\mathbf{A})^{-1} &\cong \mathbf{I}+\mathbf{A}+\mathbf{A}^2 & (188)\\ (\mathbf{I}+\mathbf{A})^{-1} &\cong \mathbf{I}-\mathbf{A}+\mathbf{A}^2 & (189)\end{aligned}$$

以下近似来自于[22]，在 $\mathbf{A}$大且对称时成立

$$\mathbf{A} - \mathbf{A}(\mathbf{I}+\mathbf{A})^{-1}\mathbf{A} \cong \mathbf{I} - \mathbf{A}^{-1} \tag{190}$$

如果 σ^2相对于 $\mathbf{Q}$和 $\mathbf{M}$很小

$$(\mathbf{Q}+\sigma^2\mathbf{M})^{-1} \cong \mathbf{Q}^{-1} - \sigma^2\mathbf{Q}^{-1}\mathbf{M}\mathbf{Q}^{-1} \tag{191}$$

证明:

$$\begin{aligned}
(\mathbf{Q}+\sigma^2\mathbf{M})^{-1} &= && (192)\\
(\mathbf{Q}\mathbf{Q}^{-1}\mathbf{Q}+\sigma^2\mathbf{M}\mathbf{Q}^{-1}\mathbf{Q})^{-1} &= && (193)\\
((\mathbf{I}+\sigma^2\mathbf{M}\mathbf{Q}^{-1})\mathbf{Q})^{-1} &= && (194)\\
\mathbf{Q}^{-1}(\mathbf{I}+\sigma^2\mathbf{M}\mathbf{Q}^{-1})^{-1} & && (195)
\end{aligned}$$

这可以用泰勒展开式重写:

$$\begin{aligned}
\mathbf{Q}^{-1}(\mathbf{I}+\sigma^2\mathbf{M}\mathbf{Q}^{-1})^{-1} &= && (196)\\
\mathbf{Q}^{-1}(\mathbf{I}-\sigma^2\mathbf{M}\mathbf{Q}^{-1}+(\sigma^2\mathbf{M}\mathbf{Q}^{-1})^2-...) &\cong \mathbf{Q}^{-1}-\sigma^2\mathbf{Q}^{-1}\mathbf{M}\mathbf{Q}^{-1} && (197)
\end{aligned}$$

3.5 广义逆矩阵

3.5.1 定义

矩阵 $\mathbf{A}$的广义逆矩阵是任何满足 (见
[26]) 的矩阵A−

$$\mathbf{A}\mathbf{A}^{-}\mathbf{A} = \mathbf{A} \tag{198}$$

矩阵 $\mathbf{A}^{-}$不唯一。

3.6 伪逆矩阵

3.6.1 定义

矩阵 $\mathbf{A}$的伪逆矩阵（或摩尔-彭罗斯逆矩阵）是满足的矩阵 $\mathbf{A}^{+}$

I	$\mathbf{A}\mathbf{A}^{+}\mathbf{A} = \mathbf{A}$
II	$\mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+} = \mathbf{A}^{+}$
III	$\mathbf{A}\mathbf{A}^{+}$ 对称
IV	$\mathbf{A}^{+}\mathbf{A}$ 对称

矩阵 $\mathbf{A}^{+}$是唯一的且总是存在的。 请注意，在复数矩阵的情况下，对称条件被一个具有埃尔米特性质的条件所替代。

3.6.2 属性

假设 $\mathbf{A}^+$是 $\mathbf{A}$的伪逆矩阵，则 (见 [3] 中的一些性质)

$$
\begin{aligned}
(\mathbf{A}^+)^+ &= \mathbf{A} && (199)\\
(\mathbf{A}^T)^+ &= (\mathbf{A}^+)^T && (200)\\
(\mathbf{A}^H)^+ &= (\mathbf{A}^+)^H && (201)\\
(\mathbf{A}^*)^+ &= (A^+)^* && (202)\\
(\mathbf{A}^+\mathbf{A})\mathbf{A}^H &= \mathbf{A}^H && (203)\\
(\mathbf{A}^+\mathbf{A})\mathbf{A}^T &= \mathbf{A}^T && (204)\\
(c\mathbf{A})^+ &= (1/c)\mathbf{A}^+ && (205)\\
\mathbf{A}^+ &= (\mathbf{A}^T\mathbf{A})^+\mathbf{A}^T && (206)\\
\mathbf{A}^+ &= \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^+ && (207)\\
(\mathbf{A}^T\mathbf{A})^+ &= \mathbf{A}^+(\mathbf{A}^T)^+ && (208)\\
(\mathbf{A}\mathbf{A}^T)^+ &= (\mathbf{A}^T)^+\mathbf{A}^+ && (209)\\
\mathbf{A}^+ &= (\mathbf{A}^H\mathbf{A})^+\mathbf{A}^H && (210)\\
\mathbf{A}^+ &= \mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^+ && (211)\\
(\mathbf{A}^H\mathbf{A})^+ &= \mathbf{A}^+(\mathbf{A}^H)^+ && (212)\\
(\mathbf{A}\mathbf{A}^H)^+ &= (\mathbf{A}^H)^+\mathbf{A}^+ && (213)\\
(\mathbf{A}\mathbf{B})^+ &= (\mathbf{A}^+\mathbf{A}\mathbf{B})^+(\mathbf{A}\mathbf{B}\mathbf{B}^+)^+ && (214)\\
f(\mathbf{A}^H\mathbf{A}) - f(0)\mathbf{I} &= \mathbf{A}^+[f(\mathbf{A}\mathbf{A}^H) - f(0)\mathbf{I}]\mathbf{A} && (215)\\
f(\mathbf{A}\mathbf{A}^H) - f(0)\mathbf{I} &= \mathbf{A}[f(\mathbf{A}^H\mathbf{A}) - f(0)\mathbf{I}]\mathbf{A}^+ && (216)
\end{aligned}
$$

其中 $\mathbf{A} \in \mathbb{C}^{n\times m}$.

假设 $\mathbf{A}$具有满秩，则

$$
\begin{aligned}
(\mathbf{A}\mathbf{A}^+)(\mathbf{A}\mathbf{A}^+) &= \mathbf{A}\mathbf{A}^+ && (217)\\
(\mathbf{A}^+\mathbf{A})(\mathbf{A}^+\mathbf{A}) &= \mathbf{A}^+\mathbf{A} && (218)\\
\mathrm{Tr}(\mathbf{A}\mathbf{A}^+) &= \mathrm{rank}(\mathbf{A}\mathbf{A}^+) \quad (\text{见 [26]}) && (219)\\
\mathrm{Tr}(\mathbf{A}^+\mathbf{A}) &= \mathrm{rank}(\mathbf{A}^+\mathbf{A}) \quad (\text{见 [26]}) && (220)
\end{aligned}
$$

对于两个矩阵，有

$$
\begin{aligned}
(\mathbf{A}\mathbf{B})^+ &= (\mathbf{A}^+\mathbf{A}\mathbf{B})^+(\mathbf{A}\mathbf{B}\mathbf{B}^+)^+ && (221)\\
(\mathbf{A}\otimes\mathbf{B})^+ &= \mathbf{A}^+\otimes\mathbf{B}^+ && (222)
\end{aligned}
$$

3.6.3 构造

假设 $\mathbf{A}$ 具有满秩，则

$$
\begin{array}{llll}
\mathbf{A}\ n\times n & \text{方阵} & \mathrm{rank}(\mathbf{A}) = n & \Rightarrow \quad \mathbf{A}^+ = \mathbf{A}^{-1}\\
\mathbf{A}\ n\times m & \text{宽} & \mathrm{rank}(\mathbf{A}) = n & \Rightarrow \quad \mathbf{A}^+ = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\\
\mathbf{A}\ n\times m & \text{高} & \mathrm{rank}(\mathbf{A}) = m & \Rightarrow \quad \mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T
\end{array}
$$

所谓的“宽版本”也被称为右逆，而“高版本”被称为左逆。

假设 $\mathbf{A}$没有满秩，即 $\mathbf{A}$是 $n \times m$的矩阵，且$\text{rank}(\mathbf{A}) = r < \min(n, m)$。伪逆矩阵 $\mathbf{A}^+$可以通过奇异值分解 $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$构造，即

$$\mathbf{A}^+ = \mathbf{V}_r \mathbf{D}_r^{-1} \mathbf{U}_r^T \tag{223}$$

其中 $\mathbf{U}_r$、 $\mathbf{D}_r$和 $\mathbf{V}_r$是删除了退化行和列的矩阵。另一种方式是：总是存在两个满秩矩阵 $\mathbf{C}$ $n \times r$和 $\mathbf{D}$ $r \times m$，使得 $\mathbf{A} = \mathbf{C}\mathbf{D}$。利用这些矩阵，有以下关系成立

$$\mathbf{A}^+ = \mathbf{D}^T(\mathbf{D}\mathbf{D}^T)^{-1}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T \tag{224}$$

参见[3]。

4复数矩阵

复数标量积 $r = pq$可以写成

$$\begin{bmatrix} \Re r \\ \Im r \end{bmatrix} = \begin{bmatrix} \Re p & -\Im p \\ \Im p & \Re p \end{bmatrix} \begin{bmatrix} \Re q \\ \Im q \end{bmatrix} \tag{225}$$

4.1 复数导数

为了对复数z进行表达式 $f(\ z)$ 的微分，需要满足柯西-黎曼方程组（[7]）：$df($
$z)$

$$\overline{\quad dz \quad} = \frac{\partial \Re(f(z))}{\partial \Re z} + i\frac{\partial \Im(f(z))}{\partial \Re z} \tag{226}$$

和

$$\frac{df(z)}{dz} = -i\frac{\partial \Re(f(z))}{\partial \Im z} + \frac{\partial \Im(f(z))}{\partial \Im z} \tag{227}$$

或者更简洁地表示为：

$$\frac{\partial f(z)}{\partial \Im z} = i\frac{\partial f(z)}{\partial \Re z}. \tag{228}$$

满足区域R中点的柯西-黎曼方程的复函数被称为在该区域R中是解析的。一般来说，涉及复共轭或共轭转置的表达式不满足柯西-黎曼方程。 为了避免这个问题，使用了更广义的复导数定义（[24]，[6]）：

- 广义复导数：

$$\frac{df(z)}{dz} = \frac{1}{2}\Big(\frac{\partial f(z)}{\partial \Re z} - i\frac{\partial f(z)}{\partial \Im z}\Big). \tag{229}$$

- 共轭复导数：

$$\frac{df(z)}{dz^*} = \frac{1}{2}\Big(\frac{\partial f(z)}{\partial \Re z} + i\frac{\partial f(z)}{\partial \Im z}\Big). \tag{230}$$

当 f是解析函数时，广义复导数等于普通导数。 对于非解析函数，如 $f(z) = z^*$ ，导数等于零。 当 f 是解析函数时，共轭复导数等于零。 在推导复梯度时，例如[21]使用了共轭复导数。

注意：

$$\frac{df(z)}{dz} = \frac{\partial f(z)}{\partial \Re z} + i\frac{\partial f(z)}{\partial \Im z}. \tag{231}$$

- 复梯度向量：如果 f是一个关于复向量 $\mathbf{z}$的实函数，
  那么复梯度向量由以下公式给出（[14, p. 798]）

$$\begin{aligned} \nabla f(\mathbf{z}) &= 2\frac{df(\mathbf{z})}{d\mathbf{z}^*} \\ &= \frac{\partial f(\mathbf{z})}{\partial \Re \mathbf{z}} + i\frac{\partial f(\mathbf{z})}{\partial \Im \mathbf{z}}. \end{aligned} \tag{232}$$

- 复梯度矩阵：如果 f是一个关于复矩阵 $\mathbf{Z}$的实函数，
  那么复梯度矩阵由以下公式给出（[2]）

$$
\begin{aligned}
\nabla f(\mathbf{Z}) &= 2\frac{df(\mathbf{Z})}{d\mathbf{Z}^*} \\
&= \frac{\partial f(\mathbf{Z})}{\partial \Re \mathbf{Z}} + i\frac{\partial f(\mathbf{Z})}{\partial \Im \mathbf{Z}}.
\end{aligned}
\tag{233}
$$

  这些表达式可用于梯度下降算法。

4.1.1 复数的链式法则

当复数函数$u = f(x)$是非解析的时候，链式法则会更加复杂。 对于非解析函数，可以应用以下链式法则（[7]）$\partial g(u)$

$$
\begin{aligned}
\frac{}{\partial x} &= \frac{\partial g}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial g}{\partial u^*}\frac{\partial u^*}{\partial x} \\
&= \frac{\partial g}{\partial u}\frac{\partial u}{\partial x} + \left(\frac{\partial g^*}{\partial u}\right)^*\frac{\partial u^*}{\partial x}
\end{aligned}
\tag{234}
$$

注意，如果函数是解析的，第二项将减少为零，函数将简化为常规的众所周知的链式法则。 对于标量函数 $g(\mathbf{U})$的矩阵导数，链式法则可以写成以下形式：

$$
\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\mathrm{Tr}((\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}})^T \partial \mathbf{U})}{\partial \mathbf{X}} + \frac{\mathrm{Tr}((\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}^*})^T \partial \mathbf{U}^*)}{\partial \mathbf{X}}.
\tag{235}
$$

4.1.2 复数的迹导数

如果导数涉及复数，通常会涉及共轭转置。 显示复数导数最有用的方法是分别显示对实部和虚部的导数。一个简单的例子是：

$$
\frac{\partial \mathrm{Tr}(\mathbf{X}^*)}{\partial \Re \mathbf{X}} = \frac{\partial \mathrm{Tr}(\mathbf{X}^H)}{\partial \Re \mathbf{X}} = \mathbf{I} \tag{236}
$$

$$
i\frac{\partial \mathrm{Tr}(\mathbf{X}^*)}{\partial \Im \mathbf{X}} = i\frac{\partial \mathrm{Tr}(\mathbf{X}^H)}{\partial \Im \mathbf{X}} = \mathbf{I} \tag{237}
$$

由于这两个结果具有相同的符号，应使用共轭复数导数（230）。

$$
\frac{\partial \mathrm{Tr}(\mathbf{X})}{\partial \Re \mathbf{X}} = \frac{\partial \mathrm{Tr}(\mathbf{X}^T)}{\partial \Re \mathbf{X}} = \mathbf{I} \tag{238}
$$

$$
i\frac{\partial \mathrm{Tr}(\mathbf{X})}{\partial \Im \mathbf{X}} = i\frac{\partial \mathrm{Tr}(\mathbf{X}^T)}{\partial \Im \mathbf{X}} = -\mathbf{I} \tag{239}
$$

在这里，这两个结果具有不同的符号，应使用广义复数导数（229）。 因此，即使 $\mathbf{X}$是一个复数，也可以看出（100）成立。

$$
\frac{\partial \mathrm{Tr}(\mathbf{A}\mathbf{X}^H)}{\partial \Re \mathbf{X}} = \mathbf{A} \tag{240}
$$

$$
i\frac{\partial \mathrm{Tr}(\mathbf{A}\mathbf{X}^H)}{\partial \Im \mathbf{X}} = \mathbf{A} \tag{241}
$$

$$\frac{\partial \mathrm{Tr}(\mathbf{A}\mathbf{X}^*)}{\partial \Re \mathbf{X}} = \mathbf{A}^T \tag{242}$$
$$i\frac{\partial \mathrm{Tr}(\mathbf{A}\mathbf{X}^*)}{\partial \Im \mathbf{X}} = \mathbf{A}^T \tag{243}$$

$$\frac{\partial \mathrm{Tr}(\mathbf{X}\mathbf{X}^H)}{\partial \Re \mathbf{X}} = \frac{\partial \mathrm{Tr}(\mathbf{X}^H\mathbf{X})}{\partial \Re \mathbf{X}} = 2\Re \mathbf{X} \tag{244}$$
$$i\frac{\partial \mathrm{Tr}(\mathbf{X}\mathbf{X}^H)}{\partial \Im \mathbf{X}} = i\frac{\partial \mathrm{Tr}(\mathbf{X}^H\mathbf{X})}{\partial \Im \mathbf{X}} = i2\Im \mathbf{X} \tag{245}$$

通过将(244)和(245)插入(229)和(230)中，可以看到

$$\frac{\partial \mathrm{Tr}(\mathbf{X}\mathbf{X}^H)}{\partial \mathbf{X}} = \mathbf{X}^* \tag{246}$$
$$\frac{\partial \mathrm{Tr}(\mathbf{X}\mathbf{X}^H)}{\partial \mathbf{X}^*} = \mathbf{X} \tag{247}$$

由于函数$\mathrm{Tr}(\mathbf{X}\mathbf{X}^H)$是复矩阵 $\mathbf{X}$的实函数，复梯度矩阵(233)由以下给出

$$\nabla \mathrm{Tr}(\mathbf{X}\mathbf{X}^H) = 2\frac{\partial \mathrm{Tr}(\mathbf{X}\mathbf{X}^H)}{\partial \mathbf{X}^*} = 2\mathbf{X} \tag{248}$$

4.1.3涉及行列式的复导数

这里提供一个计算示例。 目标是找到关于$\mathbf{X} \in \mathbb{C}^{m\times n}$的$\det(\mathbf{X}^H\mathbf{A}\mathbf{X})$的导数。通过对 $\mathbf{X}$的实部和虚部进行求导，利用(42)和(37)，可以计算出$\det(\mathbf{X}^H\mathbf{A}\mathbf{X})$ (详见附录B.1.4)

$$\begin{aligned}\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} &= \frac{1}{2}\Big(\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Re \mathbf{X}} - i\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Im \mathbf{X}}\Big) \\ &= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\left((\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^H\mathbf{A}\right)^T \end{aligned} \tag{249}$$

和复共轭导数产生

$$\begin{aligned}\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \mathbf{X}^*} &= \frac{1}{2}\Big(\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Re \mathbf{X}} + i\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Im \mathbf{X}}\Big) \\ &= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\mathbf{A}\mathbf{X}(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1} \end{aligned} \tag{250}$$

4.2 高阶和非线性导数

$$\frac{\partial}{\partial \mathbf{x}}\frac{(\mathbf{A}\mathbf{x})^H(\mathbf{A}\mathbf{x})}{(\mathbf{B}\mathbf{x})^H(\mathbf{B}\mathbf{x})} = \frac{\partial}{\partial \mathbf{x}}\frac{\mathbf{x}^H\mathbf{A}^H\mathbf{A}\mathbf{x}}{\mathbf{x}^H\mathbf{B}^H\mathbf{B}\mathbf{x}} \tag{251}$$
$$= 2\frac{\mathbf{A}^H\mathbf{A}\mathbf{x}}{\mathbf{x}^H\mathbf{B}\mathbf{B}\mathbf{x}} - 2\frac{\mathbf{x}^H\mathbf{A}^H\mathbf{A}\mathbf{x}\mathbf{B}^H\mathbf{B}\mathbf{x}}{(\mathbf{x}^H\mathbf{B}^H\mathbf{B}\mathbf{x})^2} \tag{252}$$

4.3 复数和的逆

给定实矩阵 $\mathbf{A}, \mathbf{B}$找到复数和 $\mathbf{A} + i\mathbf{B}$的逆。 形成辅助矩阵

$$\begin{aligned} \mathbf{E} &= \mathbf{A} + t\mathbf{B} && (253) \\ \mathbf{F} &= \mathbf{B} - t\mathbf{A}, && (254) \end{aligned}$$

并找到一个值为 t 的 $\mathbf{E}^{-1}$ 存在。然后

$$\begin{aligned} (\mathbf{A} + i\mathbf{B})^{-1} &= (1 - it)(\mathbf{E} + i\mathbf{F})^{-1} && (255) \\ &= (1 - it)((\mathbf{E} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F})^{-1} - i(\mathbf{E} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F})^{-1}\mathbf{F}\mathbf{E}^{-1}) && (256) \\ &= (1 - it)(\mathbf{E} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F})^{-1}(\mathbf{I} - i\mathbf{F}\mathbf{E}^{-1}) && (257) \\ &= (\mathbf{E} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F})^{-1}((\mathbf{I} - t\mathbf{F}\mathbf{E}^{-1}) - i(t\mathbf{I} + \mathbf{F}\mathbf{E}^{-1})) && (258) \\ &= (\mathbf{E} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F})^{-1}(\mathbf{I} - t\mathbf{F}\mathbf{E}^{-1}) \\ &\quad -i(\mathbf{E} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F})^{-1}(t\mathbf{I} + \mathbf{F}\mathbf{E}^{-1}) && (259) \end{aligned}$$

5 解和分解

5.1 线性方程的解

5.1.1 简单线性回归

假设我们有数据 (x_n, y_n) for $n = 1, ..., N$ and are seeking the parameters $a, b \in \mathbb{R}$ such that $y_i \cong ax_i + b$. 使用最小二乘误差函数，可以用以下符号表示最优的 a, b值

$$\mathbf{x} = (x_1, ..., x_N)^T \qquad \mathbf{y} = (y_1, ..., y_N)^T \qquad \mathbf{1} = (1, ..., 1)^T \in \mathbb{R}^{N\times 1}$$

和

$$\begin{array}{lll} R_{xx} = \mathbf{x}^T\mathbf{x} & R_{x1} = \mathbf{x}^T\mathbf{1} & R_{11} = \mathbf{1}^T\mathbf{1} \\ R_{yx} = \mathbf{y}^T\mathbf{x} & R_{y1} = \mathbf{y}^T\mathbf{1} & \end{array}$$

如

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} R_{xx} & R_{x1} \\ R_{x1} & R_{11} \end{bmatrix}^{-1} \begin{bmatrix} R_{x,y} \\ R_{y1} \end{bmatrix} \tag{260}$$

5.1.2 线性系统中的存在性

假设 $\mathbf{A}$是 $n \times m$的，并考虑线性系统

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{261}$$

构造增广矩阵 $\mathbf{B} = [\mathbf{A}\ \mathbf{b}]$ 然后

条件	解
$\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{B}) = m$	唯一解 $\mathbf{x}$
$\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{B}) < m$	多个解 $\mathbf{x}$
$\mathrm{rank}(\mathbf{A}) < \mathrm{rank}(\mathbf{B})$	无解 $\mathbf{x}$

5.1.3 标准方阵

假设 $\mathbf{A}$是方阵且可逆，则

$$\mathbf{A}\mathbf{x} = \mathbf{b} \qquad \Rightarrow \qquad \mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \tag{262}$$

5.1.4 退化的方阵

假设 $\mathbf{A}$是一个 $n\times n$的矩阵，但是它的秩为$r < n$。 在这种情况下，方程组 $\mathbf{A}\mathbf{x} = \mathbf{b}$的解为

$$\mathbf{x} = \mathbf{A}^{+}\mathbf{b}$$

其中 $\mathbf{A}^{+}$ 是秩缺失矩阵的伪逆，根据第3.6.3节的描述构造。

5.1.5 克莱姆法则

方程

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{263}$$

其中 $\mathbf{A}$是方阵，如果 $\mathbf{x}$的第 i个元素可以表示为

$$x_i = \frac{\det \mathbf{B}}{\det \mathbf{A}}, \tag{264}$$

其中 $\mathbf{B}$等于 $\mathbf{A}$，但是 $\mathbf{A}$的第 i列被 $\mathbf{b}$替换。

5.1.6 过定的矩形方程组

假设 $\mathbf{A}$是一个 $n \times m$的矩阵，$n > m$(高矩阵)，且$\text{rank}(\mathbf{A}) = m$，那

$$\text{么方程组}\mathbf{A}\mathbf{x} = \mathbf{b}\text{的} \quad \Rightarrow \quad \mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} = \mathbf{A}^+\mathbf{b} \tag{265}$$

也就是说如果存在解 $\mathbf{x}$的话！如果没有解，下面的内容可能会有用：$\mathbf{A}\mathbf{x} = \mathbf{b}$

$$\Rightarrow \quad \mathbf{x}_{min} = \mathbf{A}^+\mathbf{b} \tag{266}$$

现在，$\mathbf{x}_{min}$是向量 $\mathbf{x}$，它最小化了 $||\mathbf{A}\mathbf{x} - \mathbf{b}||^2$，也就是说，它是最“接近正确”的向量。 矩阵 $\mathbf{A}^+$是 $\mathbf{A}$的伪逆矩阵。参见[3]。

5.1.7 欠定矩形

假设 $\mathbf{A}$是一个 $n \times m$的矩阵，且$n < m$（“宽”），且$\text{rank}(\mathbf{A}) = n$。

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad \Rightarrow \quad \mathbf{x}_{min} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b} \tag{267}$$

该方程有很多解 $\mathbf{x}$。 但是，$\mathbf{x}_{min}$是最小化了$||\mathbf{A}\mathbf{x} - \mathbf{b}||^2$的解，也是具有最小范数 $||\ \mathbf{x}||^2$的解。 对于矩阵版本也是一样的：假设 $\mathbf{A}$是一个 $n \times m$的矩阵， $\mathbf{X}$是一个 $m \times n$的矩阵， $\mathbf{B}$是一个 $n \times n$的矩阵，那么$\mathbf{A}\mathbf{X} = \mathbf{B}$

$$\Rightarrow \quad \mathbf{X}_{min} = \mathbf{A}^+\mathbf{B} \tag{268}$$

方程有很多解 $\mathbf{X}$。 但是 $\mathbf{X}_{min}$是最小化$||\mathbf{A}\mathbf{X} - \mathbf{B}||^2$的解，也是具有最小范数 $||\mathbf{X}||^2$的解。参见[3]。

类似但不同：假设 $\mathbf{A}$是方阵 $n \times n$，矩阵 $\mathbf{B}_0$， $\mathbf{B}_1$是 $n \times N$的矩阵，其中$N > n$，那么如果 $\mathbf{B}_0$具有最大秩

$$\mathbf{A}\mathbf{B}_0 = \mathbf{B}_1 \quad \Rightarrow \quad \mathbf{A}_{min} = \mathbf{B}_1\mathbf{B}_0^T(\mathbf{B}_0\mathbf{B}_0^T)^{-1} \tag{269}$$

其中 $\mathbf{A}_{min}$表示在最小二乘意义下最优的矩阵。 一种解释是 $\mathbf{A}$是将 $\mathbf{B}_0$的列向量映射到 $\mathbf{B}_1$的列向量的线性近似。

5.1.8 线性形式和零解

$$\mathbf{A}\mathbf{x} = \mathbf{0}\text{，对于所有的}\mathbf{x} \quad \Rightarrow \quad \mathbf{A} = \mathbf{0} \tag{270}$$

5.1.9 方阵形式和零

如果 $\mathbf{A}$是对称的，则

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{0}, \quad \forall\mathbf{x} \quad \Rightarrow \quad \mathbf{A} = \mathbf{0} \tag{271}$$

5.1.10 李亚普诺夫方程

$$\mathbf{AX} + \mathbf{XB} = \mathbf{C} \tag{272}$$
$$\text{vec}(\mathbf{X}) = (\mathbf{I} \otimes \mathbf{A} + \mathbf{B}^T \otimes \mathbf{I})^{-1}\text{vec}(\mathbf{C}) \tag{273}$$

有关克罗内克积和vec运算符的详细信息，请参见第10.2.1节和10.2.2节。

5.1.11 封装求和

$$\sum_n \mathbf{A}_n \mathbf{X} \mathbf{B}_n = \mathbf{C} \tag{274}$$
$$\text{vec}(\mathbf{X}) = \left(\sum_n \mathbf{B}_n^T \otimes \mathbf{A}_n\right)^{-1} \text{vec}(\mathbf{C}) \tag{275}$$

有关克罗内克积和vec运算符的详细信息，请参见第10.2.1节和10.2.2节。

5.2 特征值和特征向量

5.2.1 定义

特征向量 $\mathbf{v}_i$和特征值 λ_i满足以下条件

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i \tag{276}$$

5.2.2 分解

对于具有与维度相同数量的不同特征值的矩阵 $\mathbf{A}$，以下等式成立，其中 $\mathbf{V}$的列是特征向量，$(\mathbf{D})_{ij} = \delta_{ij}\lambda_i$，$\mathbf{AV} = \mathbf{VD}$

(277)对

于有缺陷的矩阵 $\mathbf{A}$，即具有较少不同特征值的矩阵，以下分解称为Jordan规范形式，成立

$$\mathbf{AV} = \mathbf{VJ} \tag{278}$$

其中 $\mathbf{J}$是一个块对角矩阵，块 $\mathbf{J}_i = \lambda_i \mathbf{I} + \mathbf{N}$。 矩阵
$\mathbf{J}_i$的维度与相同特征值 λ_i的数量相等， $\mathbf{N}$
是一个大小相同的方阵，其超对角线上为1，其他位置为0。

对于所有矩阵 $\mathbf{A}$，存在矩阵 $\mathbf{V}$和 $\mathbf{R}$使得

$$\mathbf{AV} = \mathbf{VR}\text{也成立} \tag{279}$$

其中 $\mathbf{R}$是上三角矩阵，其对角线上是特征值 λ_i

5.2.3 一般性质假设 $\mathbf{A} \in \mathbb{R}^{n\times m}$和 $\mathbf{B} \in \mathbb{R}$

$^{m\times n}$，则 $\mathbf{ei}\text{g}(\mathbf{AB}) = \text{eig}(\mathbf{BA})$

(280)

$$\text{rank}(\mathbf{A}) = r \quad \Rightarrow \text{最多有 } r\text{个非零特征值 } \lambda_i \tag{281}$$

5.2.4 对称性

假设 $\mathbf{A}$ 是对称矩阵，则

$$\begin{aligned} \mathbf{V}\mathbf{V}^T &= \mathbf{I} \quad (\text{即 } \mathbf{V} \text{ 是正交矩阵}) & (282)\\ \lambda_i &\in \mathbb{R} \quad (\text{即特征值 } \lambda_i \text{是实数}) & (283)\\ \mathrm{Tr}(\mathbf{A}^p) &= \textstyle\sum_i \lambda_i^p & (284)\\ \mathrm{eig}(\mathbf{I}+c\mathbf{A}) &= 1+c\lambda_i & (285)\\ \mathrm{eig}(\mathbf{A}-c\mathbf{I}) &= \lambda_i - c & (286)\\ \mathrm{eig}(\mathbf{A}^{-1}) &= \lambda_i^{-1} & (287) \end{aligned}$$

对于一个对称的正矩阵 $\mathbf{A}$,

$$\mathrm{eig}(\mathbf{A}^T\mathbf{A}) = \mathrm{eig}(\mathbf{A}\mathbf{A}^T) = \mathrm{eig}(\mathbf{A}) \circ \mathrm{eig}(\mathbf{A}) \tag{288}$$

5.2.5 特征多项式

矩阵 $\mathbf{A}$的特征多项式为

$$\begin{aligned} 0 &= \det(\mathbf{A}-\lambda\mathbf{I}) & (289)\\ &= \lambda^n - g_1\lambda^{n-1} + g_2\lambda^{n-2} - ... + (-1)^n g_n & (290) \end{aligned}$$

注意，系数 g_j对于 $j=1,...,n$在旋转下是不变的矩阵 $\mathbf{A}$. 因此， g_j是 $\mathbf{A}$的所有子矩阵的行和列的行列式之和取j行和列的情况下。 也就是说， g_1是 $\mathbf{A}$的迹， g_2是从 $\mathbf{A}$中删除所有但两行和两列后可以形成的所有($n(n-1)/2$)个子矩阵的行列式之和，以此类推 – 参见[17]。

5.3 奇异值分解

任意 $n\times m$矩阵 $\mathbf{A}$可以表示为

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T, \tag{291}$$

其中

$$\begin{aligned} \mathbf{U} &= \text{矩阵 } \mathbf{A}\mathbf{A}^T \text{ 的特征向量，维度为 } n\times n\\ \mathbf{D} &= \sqrt{\mathrm{diag}(\mathrm{eig}(\mathbf{A}\mathbf{A}^T))} \quad n\times m & (292)\\ \mathbf{V} &= \text{矩阵 } \mathbf{A}^T\mathbf{A} \text{ 的特征向量，维度为 } m\times m \end{aligned}$$

5.3.1 对称方阵分解为平方

假设 $\mathbf{A}$ 为 $n\times n$ 的对称方阵。则

$$\begin{bmatrix} \mathbf{A} \end{bmatrix} = \begin{bmatrix} \mathbf{V} \end{bmatrix}\begin{bmatrix} \mathbf{D} \end{bmatrix}\begin{bmatrix} \mathbf{V}^T \end{bmatrix}, \tag{293}$$

其中 $\mathbf{D}$是 $\mathbf{A}$的特征值构成的对角矩阵， $\mathbf{V}$是正交矩阵，是 $\mathbf{A}$的特征向量。

5.3.2 方阵分解为平方

假设 $\mathbf{A}\in\mathbb{R}^{n\times n}$。则

$$\begin{bmatrix} \mathbf{A} \end{bmatrix} = \begin{bmatrix} \mathbf{V} \end{bmatrix}\begin{bmatrix} \mathbf{D} \end{bmatrix}\begin{bmatrix} \mathbf{U}^T \end{bmatrix}, \tag{294}$$

其中 $\mathbf{D}$是对角线，其元素为 $\mathbf{A}\mathbf{A}^T$的特征值的平方根， $\mathbf{V}$是 $\mathbf{A}\mathbf{A}^T$的特征向量， $\mathbf{U}^T$是 $\mathbf{A}^T\mathbf{A}$的特征向量。

5.3.3 方阵分解为矩形

假设 $\mathbf{V}_*\mathbf{D}_*\mathbf{U}_*^T = \mathbf{0}$，那么我们可以将 $\mathbf{A}$的奇异值分解展开为

$$\left[\begin{array}{c}\mathbf{A}\end{array}\right] = \left[\begin{array}{c|c}\mathbf{V} & \mathbf{V}_*\end{array}\right]\left[\begin{array}{c|c}\mathbf{D} & \mathbf{0}\\ \hline \mathbf{0} & \mathbf{D}_*\end{array}\right]\left[\begin{array}{c}\mathbf{U}^T\\ \hline \mathbf{U}_*^T\end{array}\right], \tag{295}$$

其中 $\mathbf{A}$的奇异值分解为 $\mathbf{A} = \mathbf{VDU}^T$。

5.3.4 矩形分解 I

假设 $\mathbf{A}$是 $n \times m$的矩阵， $\mathbf{V}$是 $n \times n$的矩阵， $\mathbf{D}$是 $n \times n$的矩阵， $\mathbf{U}^T$是 $n \times m$

$$\left[\begin{array}{ccc} & \mathbf{A} & \end{array}\right] = \left[\begin{array}{c}\mathbf{V}\end{array}\right]\left[\begin{array}{c}\mathbf{D}\end{array}\right]\left[\begin{array}{ccc} & \mathbf{U}^T & \end{array}\right], \tag{296}$$

其中 $\mathbf{D}$是对角线上的特征值的平方根 $\mathbf{AA}^T$, $\mathbf{V}$是 $\mathbf{AA}^T$的特征向量 $\mathbf{U}^T$是 $\mathbf{A}^T\mathbf{A}$的特征向量。

5.3.5 矩形分解 II

假设 $\mathbf{A}$是 $n \times m$的矩阵 $\mathbf{V}$是 $n \times m$的矩阵 $\mathbf{D}$是 $m \times m$的矩阵 $\mathbf{U}^T$是 $m \times m$

$$\left[\begin{array}{ccc} & \mathbf{A} & \end{array}\right] = \left[\begin{array}{ccc} & \mathbf{V} & \end{array}\right]\left[\begin{array}{ccc} & & \\ & \mathbf{D} & \\ & & \end{array}\right]\left[\begin{array}{ccc} & & \\ & \mathbf{U}^T & \\ & & \end{array}\right] \tag{297}$$

5.3.6 矩形分解 III

假设 $\mathbf{A}$是 $n \times m$的矩阵 $\mathbf{V}$是 $n \times n$的矩阵 $\mathbf{D}$是 $n \times m$的矩阵 $\mathbf{U}^T$是 $m \times m$

$$\left[\begin{array}{ccc} & \mathbf{A} & \end{array}\right] = \left[\begin{array}{c}\mathbf{V}\end{array}\right]\left[\begin{array}{ccc} & \mathbf{D} & \end{array}\right]\left[\begin{array}{ccc} & & \\ & \mathbf{U}^T & \\ & & \end{array}\right], \tag{298}$$

其中 $\mathbf{D}$是对角线上的特征值的平方根 $\mathbf{AA}^T$, $\mathbf{V}$是 $\mathbf{AA}^T$的特征向量 $\mathbf{U}^T$是 $\mathbf{A}^T\mathbf{A}$的特征向量。

5.4 三角分解

5.5 LU分解

假设 $\mathbf{A}$是一个具有非零主子式的方阵，则

$$\mathbf{A} = \mathbf{LU} \tag{299}$$

其中 $\mathbf{L}$是一个唯一的单位下三角矩阵， $\mathbf{U}$是一个唯一的上三角矩阵。

5.5.1 乔列斯基分解

假设 $\mathbf{A}$是一个对称正定的方阵，那么

$$\mathbf{A} = \mathbf{U}^T\mathbf{U} = \mathbf{LL}^T, \tag{300}$$

其中 $\mathbf{U}$是一个唯一的上三角矩阵， $\mathbf{L}$是一个下三角矩阵。

5.6 LDM分解

假设 $\mathbf{A}$是一个具有非零主子式[1]的方阵，那么

$$\mathbf{A} = \mathbf{LDM}^T \tag{301}$$

其中 $\mathbf{L}$，$\mathbf{M}$是唯一的单位下三角矩阵，$\mathbf{D}$是一个唯一的对角矩阵。

5.7 LDL分解

LDL分解是LDM分解的特殊情况。假设
$\mathbf{A}$是一个非奇异的对称正定方阵，那么

$$\mathbf{A} = \mathbf{LDL}^T = \mathbf{L}^T\mathbf{DL} \tag{302}$$

其中 $\mathbf{L}$是一个单位下三角矩阵，$\mathbf{D}$是一个对角矩阵。如果 $\mathbf{A}$也是正定的，那么 $\mathbf{D}$的对角线元素严格为正。

[1]如果与一个主子式相对应的矩阵是较大矩阵的二次上左部分（即由行和列从1到k的矩阵元素组成），则该主子式称为前导主子式。对于一个n乘n的方阵，有n个前导主子式。[31]

6 统计与概率

6.1 矩的定义

假设 $\mathbf{x} \in \mathbb{R}^{n\times 1}$ 是一个随机变量

6.1.1 均值

均值向量 $\mathbf{m}$的定义如下

$$(\mathbf{m})_i = \langle x_i \rangle \tag{303}$$

6.1.2 协方差

协方差矩阵 $\mathbf{M}$的定义如下

$$(\mathbf{M})_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \tag{304}$$

或者可以写成

$$\mathbf{M} = \langle (\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^T \rangle \tag{305}$$

6.1.3 第三阶矩

第三阶中心矩阵 - 在某些情况下称为共偏斜度 - 使用以下符号定义

$$m^{(3)}_{ijk} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle)(x_k - \langle x_k \rangle) \rangle \tag{306}$$

如

$$\mathbf{M}_3 = \left[m^{(3)}_{::1} m^{(3)}_{::2} ... m^{(3)}_{::n} \right] \tag{307}$$

其中':'表示给定索引内的所有元素。 $\mathbf{M}_3$也可以表示为

$$\mathbf{M}_3 = \langle (\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^T \otimes (\mathbf{x} - \mathbf{m})^T \rangle \tag{308}$$

6.1.4 四阶矩

第四阶中心矩阵 - 在某些情况下也称为峰度 - 使用以下符号定义

$$m^{(4)}_{ijkl} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle)(x_k - \langle x_k \rangle)(x_l - \langle x_l \rangle) \rangle \tag{309}$$

在

$$\mathbf{M}_4 = \left[m^{(4)}_{::11} m^{(4)}_{::21} ... m^{(4)}_{::n1} | m^{(4)}_{::12} m^{(4)}_{::22} ... m^{(4)}_{::n2} | ... | m^{(4)}_{::1n} m^{(4)}_{::2n} ... m^{(4)}_{::nn} \right] \tag{310}$$

或者可以表示为

$$\mathbf{M}_4 = \langle (\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^T \otimes (\mathbf{x} - \mathbf{m})^T \otimes (\mathbf{x} - \mathbf{m})^T \rangle \tag{311}$$

6.2 线性组合的期望

6.2.1 线性形式

假设 $\mathbf{X}$和 $\mathbf{x}$是一个矩阵和一个随机变量向量。那么（见
参见[26]）

$$E[\mathbf{AXB}+\mathbf{C}] = \mathbf{A}E[\mathbf{X}]\mathbf{B}+\mathbf{C} \quad (312)$$
$$\mathrm{Var}[\mathbf{Ax}] = \mathbf{A}\mathrm{Var}[\mathbf{x}]\mathbf{A}^T \quad (313)$$
$$\mathrm{Cov}[\mathbf{Ax},\mathbf{By}] = \mathbf{A}\mathrm{Cov}[\mathbf{x},\mathbf{y}]\mathbf{B}^T \quad (314)$$

假设 $\mathbf{x}$是一个具有均值 $\mathbf{m}$的随机向量，那么（见[7]）

$$E[\mathbf{Ax}+\mathbf{b}] = \mathbf{Am}+\mathbf{b} \quad (315)$$
$$E[\mathbf{Ax}] = \mathbf{Am} \quad (316)$$
$$E[\mathbf{x}+\mathbf{b}] = \mathbf{m}+\mathbf{b} \quad (317)$$

6.2.2 二次型

假设 $\mathbf{A}$是对称的， $\mathbf{c} = E[\mathbf{x}]$ 和 $\mathbf{\Sigma} = \mathrm{Var}[\mathbf{x}]$. 还假设所有坐标 x_i是独立的，具有相同的中心矩 $\mu_1, \mu_2, \mu_3, \mu_4$并且记 $\mathbf{a}= \mathrm{diag}(\mathbf{A})$. 那么（参见[26]）

$$E[\mathbf{x}^T\mathbf{Ax}] = \mathrm{Tr}(\mathbf{A\Sigma})+\mathbf{c}^T\mathbf{Ac} \quad (318)$$
$$\mathrm{Var}[\mathbf{x}^T\mathbf{Ax}] = 2\mu_2^2\mathrm{Tr}(\mathbf{A}^2)+4\mu_2\mathbf{c}^T\mathbf{A}^2\mathbf{c}+4\mu_3\mathbf{c}^T\mathbf{Aa}+(\mu_4-3\mu_2^2)\mathbf{a}^T\mathbf{a} \quad (319)$$

另外，假设 $\mathbf{x}$是一个均值为 $\mathbf{m}$，协方差为 $\mathbf{M}$的随机向量（见[7]）

$$E[(\mathbf{Ax}+\mathbf{a})(\mathbf{Bx}+\mathbf{b})^T] = \mathbf{AMB}^T+(\mathbf{Am}+\mathbf{a})(\mathbf{Bm}+\mathbf{b})^T \quad (320)$$
$$E[\mathbf{xx}^T] = \mathbf{M}+\mathbf{mm}^T \quad (321)$$
$$E[\mathbf{xa}^T\mathbf{x}] = (\mathbf{M}+\mathbf{mm}^T)\mathbf{a} \quad (322)$$
$$E[\mathbf{x}^T\mathbf{ax}^T] = \mathbf{a}^T(\mathbf{M}+\mathbf{mm}^T) \quad (323)$$
$$E[(\mathbf{Ax})(\mathbf{Ax})^T] = \mathbf{A}(\mathbf{M}+\mathbf{mm}^T)\mathbf{A}^T \quad (324)$$
$$E[(\mathbf{x}+\mathbf{a})(\mathbf{x}+\mathbf{a})^T] = \mathbf{M}+(\mathbf{m}+\mathbf{a})(\mathbf{m}+\mathbf{a})^T \quad (325)$$

$$E[(\mathbf{Ax}+\mathbf{a})^T(\mathbf{Bx}+\mathbf{b})] = \mathrm{Tr}(\mathbf{AMB}^T)+(\mathbf{Am}+\mathbf{a})^T(\mathbf{Bm}+\mathbf{b}) \quad (326)$$
$$E[\mathbf{x}^T\mathbf{x}] = \mathrm{Tr}(\mathbf{M})+\mathbf{m}^T\mathbf{m} \quad (327)$$
$$E[\mathbf{x}^T\mathbf{Ax}] = \mathrm{Tr}(\mathbf{AM})+\mathbf{m}^T\mathbf{Am} \quad (328)$$
$$E[(\mathbf{Ax})^T(\mathbf{Ax})] = \mathrm{Tr}(\mathbf{AMA}^T)+(\mathbf{Am})^T(\mathbf{Am}) \quad (329)$$
$$E[(\mathbf{x}+\mathbf{a})^T(\mathbf{x}+\mathbf{a})] = \mathrm{Tr}(\mathbf{M})+(\mathbf{m}+\mathbf{a})^T(\mathbf{m}+\mathbf{a}) \quad (330)$$

见[7]。

6.2.3 立方形式

假设 $\mathbf{x}$是一个具有独立坐标、均值 $\mathbf{m}$、协方差 $\mathbf{M}$和中心矩 $\mathbf{v}_3 = E[(\mathbf{x}-\mathbf{m})^3]$的随机向量。然后（参见[7]）

$$
\begin{aligned}
E[(\mathbf{Ax}+\mathbf{a})(\mathbf{Bx}+\mathbf{b})^T(\mathbf{Cx}+\mathbf{c})] &= \mathbf{A}\mathrm{diag}(\mathbf{B}^T\mathbf{C})\mathbf{v}_3 \\
&\quad +\mathrm{Tr}(\mathbf{BMC}^T)(\mathbf{Am}+\mathbf{a}) \\
&\quad +\mathbf{AMC}^T(\mathbf{Bm}+\mathbf{b}) \\
&\quad +(\mathbf{AMB}^T+(\mathbf{Am}+\mathbf{a})(\mathbf{Bm}+\mathbf{b})^T)(\mathbf{Cm}+\mathbf{c}) \\
E[\mathbf{xx}^T\mathbf{x}] &= \mathbf{v}_3 + 2\mathbf{Mm} + (\mathrm{Tr}(\mathbf{M})+\mathbf{m}^T\mathbf{m})\mathbf{m} \\
E[(\mathbf{Ax}+\mathbf{a})(\mathbf{Ax}+\mathbf{a})^T(\mathbf{Ax}+\mathbf{a})] &= \mathbf{A}\mathrm{diag}(\mathbf{A}^T\mathbf{A})\mathbf{v}_3 \\
&\quad +[2\mathbf{AMA}^T+(\mathbf{Ax}+\mathbf{a})(\mathbf{Ax}+\mathbf{a})^T](\mathbf{Am}+\mathbf{a}) \\
&\quad +\mathrm{Tr}(\mathbf{AMA}^T)(\mathbf{Am}+\mathbf{a}) \\
E[(\mathbf{Ax}+\mathbf{a})\mathbf{b}^T(\mathbf{Cx}+\mathbf{c})(\mathbf{Dx}+\mathbf{d})^T] &= (\mathbf{Ax}+\mathbf{a})\mathbf{b}^T(\mathbf{CMD}^T+(\mathbf{Cm}+\mathbf{c})(\mathbf{Dm}+\mathbf{d})^T) \\
&\quad +(\mathbf{AMC}^T+(\mathbf{Am}+\mathbf{a})(\mathbf{Cm}+\mathbf{c})^T)\mathbf{b}(\mathbf{Dm}+\mathbf{d})^T \\
&\quad +\mathbf{b}^T(\mathbf{Cm}+\mathbf{c})(\mathbf{AMD}^T-(\mathbf{Am}+\mathbf{a})(\mathbf{Dm}+\mathbf{d})^T)
\end{aligned}
$$

6.3 加权标量变量

假设 $\mathbf{x}\in\mathbb{R}^{n\times 1}$是一个随机变量， $\mathbf{w}\in\mathbb{R}^{n\times 1}$是一个常数向量，$y$是线性组合 $y = \mathbf{w}^T\mathbf{x}$。进一步假设 $\mathbf{m}, \mathbf{M}_2, \mathbf{M}_3, \mathbf{M}_4$表示变量 $\mathbf{x}$的均值、协方差和中心三阶和四阶矩阵。那么有

$$
\begin{aligned}
\langle y\rangle &= \mathbf{w}^T\mathbf{m} && (331)\\
\langle (y-\langle y\rangle)^2\rangle &= \mathbf{w}^T\mathbf{M}_2\mathbf{w} && (332)\\
\langle (y-\langle y\rangle)^3\rangle &= \mathbf{w}^T\mathbf{M}_3\mathbf{w}\otimes\mathbf{w} && (333)\\
\langle (y-\langle y\rangle)^4\rangle &= \mathbf{w}^T\mathbf{M}_4\mathbf{w}\otimes\mathbf{w}\otimes\mathbf{w} && (334)
\end{aligned}
$$

7 多元分布

7.1 柯西分布

柯西分布向量的密度函数 $\mathbf{t} \in \mathbb{R}^{P\times 1}$，给定为

$$p(\mathbf{t}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \pi^{-P/2}\frac{\Gamma(\frac{1+P}{2})}{\Gamma(1/2)}\frac{\det(\boldsymbol{\Sigma})^{-1/2}}{\left[1 + (\mathbf{t}-\boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\mathbf{t}-\boldsymbol{\mu})\right]^{(1+P)/2}} \tag{335}$$

其中 $\boldsymbol{\mu}$是位置， $\boldsymbol{\Sigma}$是正定的，Γ表示伽玛函数。 柯西分布是学生-t分布的特例
。

7.2 狄利克雷分布

狄利克雷分布是与多项分布相比的一种“逆”分布，适用于有界连续变量 $\mathbf{x}= [x_1, \dots, x_P]$[16, p. 44]

$$p(\mathbf{x}|\boldsymbol{\alpha}) = \frac{\Gamma\left(\sum_p^P \alpha_p\right)}{\prod_p^P \Gamma(\alpha_p)}\prod_p^P x_p^{\alpha_p - 1}$$

7.3 正态分布

正态分布也被称为高斯分布。见第8节。

7.4 正态逆伽马分布

7.5 高斯分布

见第8节。

7.6 多项分布

如果向量 $\mathbf{n}$包含计数，即 $(\mathbf{n})_i \in 0, 1, 2, ...$，则离散多项式分布对于 $\mathbf{n}$给出

$$P(\mathbf{n}|\mathbf{a}, n) = \frac{n!}{n_1! \dots n_d!}\prod_i^d a_i^{n_i}, \qquad \sum_i^d n_i = n \tag{336}$$

其中 a_i是概率，即 $0 \le a_i \le 1$，且 $\sum_i a_i = 1.$

7.7 学生t分布

学生-t分布向量 $\mathbf{t} \in \mathbb{R}^{P\times 1}$的密度由以下公式给出

$$p(\mathbf{t}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = (\pi\nu)^{-P/2}\frac{\Gamma(\frac{\nu+P}{2})}{\Gamma(\nu/2)}\frac{\det(\boldsymbol{\Sigma})^{-1/2}}{\left[1 + \nu^{-1}(\mathbf{t}-\boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\mathbf{t}-\boldsymbol{\mu})\right]^{(\nu+P)/2}} \tag{337}$$

其中 $\boldsymbol{\mu}$是位置，比例矩阵 $\boldsymbol{\Sigma}$是对称的、正定的， ν是自由度，Γ表示伽玛函数
。 对于 $\nu= 1$，学生-t分布变成柯西分布（见第7.1节）。

7.7.1 平均值

$$\mathrm{E}(\mathbf{t}) = \boldsymbol{\mu}, \qquad \nu > 1 \tag{338}$$

7.7.2 方差

$$\mathrm{cov}(\mathbf{t}) = \frac{\nu}{\nu-2}\boldsymbol{\Sigma}, \qquad \nu > 2 \tag{339}$$

7.7.3 众数

概念 $mode$意味着最可能值的位置

$$\mathrm{mode}(\mathbf{t}) = \boldsymbol{\mu} \tag{340}$$

7.7.4 完整矩阵版本

如果不是一个向量 $\mathbf{t} \in \mathbb{R}^{P\times 1}$，而是一个矩阵 $\mathbf{T} \in \mathbb{R}^{P\times N}$，则对于T的学生-t
分布为

$$\begin{aligned} p(\mathbf{T}|\mathbf{M},\boldsymbol{\Omega},\boldsymbol{\Sigma},\nu) &= \pi^{-NP/2}\prod_{p=1}^{P}\frac{\Gamma\left[(\nu+P-p+1)/2\right]}{\Gamma\left[(\nu-p+1)/2\right]}\times \\ &\quad \nu\det(\boldsymbol{\Omega})^{-\nu/2}\det(\boldsymbol{\Sigma})^{-N/2}\times \\ &\quad \det\left[\boldsymbol{\Omega}^{-1}+(\mathbf{T}-\mathbf{M})\boldsymbol{\Sigma}^{-1}(\mathbf{T}-\mathbf{M})^{\mathsf{T}}\right]^{-(\nu+P)/2} \end{aligned} \tag{341}$$

其中 $\mathbf{M}$是位置， $\boldsymbol{\Omega}$是缩放矩阵， $\boldsymbol{\Sigma}$是正定的， ν是
自由度，Γ表示伽玛函数。

7.8 Wishart

对于 $\mathbf{M} \in \mathbb{R}^{P\times P}$的中心Wishart分布， $\mathbf{M}$是正定的，其中
m可以看作是自由度参数[16，方程3.8.1] [8，
第2.5节]，[11]

$$\begin{aligned} p(\mathbf{M}|\boldsymbol{\Sigma},m) &= \frac{1}{2^{mP/2}\pi^{P(P-1)/4}\prod_p^P\Gamma[\frac{1}{2}(m+1-p)]}\times \\ &\quad \det(\boldsymbol{\Sigma})^{-m/2}\det(\mathbf{M})^{(m-P-1)/2}\times \\ &\quad \exp\left[-\frac{1}{2}\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\mathbf{M})\right] \end{aligned} \tag{342}$$

7.8.1 平均值

$$E(\mathbf{M}) = \mathbf{m}\boldsymbol{\Sigma} \tag{343}$$

7.9 逆Wishart分布

对于 $\mathbf{M} \in \mathbb{R}^{P\times P}$的（正常的）逆Wishart分布， $\mathbf{M}$是正定的，其中 m可以被视为自由度参数[11]

$$\begin{aligned} p(\mathbf{M}|\mathbf{\Sigma}, m) \quad = \quad & \frac{1}{2^{mP/2}\pi^{P(P-1)/4}\prod_p^P \Gamma[\frac{1}{2}(m+1-p)]} \times \\ & \det(\mathbf{\Sigma})^{m/2}\det(\mathbf{M})^{-(m-P-1)/2} \times \\ & \exp\left[-\frac{1}{2}\mathrm{Tr}(\mathbf{\Sigma}\mathbf{M}^{-1})\right] \end{aligned} \tag{344}$$

7.9.1 平均值

$$E(\mathbf{M}) = \mathbf{\Sigma}\frac{1}{m-P-1} \tag{345}$$

8 高斯分布

8.1 基础知识

8.1.1 密度和归一化

$\mathbf{x}\sim\mathcal{N}(\mathbf{m},\mathbf{\Sigma})$的密度为

$$p(\mathbf{x})=\frac{1}{\sqrt{\det(2\pi\mathbf{\Sigma})}}\exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})\right] \tag{346}$$

请注意，如果 $\mathbf{x}$是d维的，则$\det(2\pi\mathbf{\Sigma})=(2\pi)^d\det(\mathbf{\Sigma})$。
积分和归一化

$$\begin{aligned}
\int\exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})\right]d\mathbf{x} &= \sqrt{\det(2\pi\mathbf{\Sigma})} \\
\int\exp\left[-\frac{1}{2}\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}+\mathbf{m}^T\mathbf{\Sigma}^{-1}\mathbf{x}\right]d\mathbf{x} &= \sqrt{\det(2\pi\mathbf{\Sigma})}\exp\left[\frac{1}{2}\mathbf{m}^T\mathbf{\Sigma}^{-1}\mathbf{m}\right] \\
\int\exp\left[-\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}+\mathbf{c}^T\mathbf{x}\right]d\mathbf{x} &= \sqrt{\det(2\pi\mathbf{A}^{-1})}\exp\left[\frac{1}{2}\mathbf{c}^T\mathbf{A}^{-T}\mathbf{c}\right]
\end{aligned}$$

如果 $\mathbf{X}=[\mathbf{x}_1\mathbf{x}_2...\mathbf{x}_n]$ 和 $\mathbf{C}=[\mathbf{c}_1\mathbf{c}_2...\mathbf{c}_n]$，那么

$$\int\exp\left[-\frac{1}{2}\mathrm{Tr}(\mathbf{X}^T\mathbf{A}\mathbf{X})+\mathrm{Tr}(\mathbf{C}^T\mathbf{X})\right]d\mathbf{X}=\sqrt{\det(2\pi\mathbf{A}^{-1})}^n\exp\left[\frac{1}{2}\mathrm{Tr}(\mathbf{C}^T\mathbf{A}^{-1}\mathbf{C})\right]$$

密度的导数为

$$\frac{\partial p(\mathbf{x})}{\partial\mathbf{x}} = -p(\mathbf{x})\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{m}) \tag{347}$$

$$\frac{\partial^2 p}{\partial\mathbf{x}\partial\mathbf{x}^T} = p(\mathbf{x})\left(\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})^T\mathbf{\Sigma}^{-1}-\mathbf{\Sigma}^{-1}\right) \tag{348}$$

8.1.2 边缘分布

假设 $\mathbf{x}\sim\mathcal{N}_{\mathbf{x}}(\mu,\mathbf{\Sigma})$ 其中

$$\mathbf{x}=\begin{bmatrix}\mathbf{x}_a\\ \mathbf{x}_b\end{bmatrix}\qquad \boldsymbol{\mu}=\begin{bmatrix}\boldsymbol{\mu}_a\\ \boldsymbol{\mu}_b\end{bmatrix}\qquad \mathbf{\Sigma}=\begin{bmatrix}\mathbf{\Sigma}_a & \mathbf{\Sigma}_c\\ \mathbf{\Sigma}_c^T & \mathbf{\Sigma}_b\end{bmatrix} \tag{349}$$

然后

$$p(\mathbf{x}_a) = \mathcal{N}_{\mathbf{x}_a}(\boldsymbol{\mu}_a,\mathbf{\Sigma}_a) \tag{350}$$

$$p(\mathbf{x}_b) = \mathcal{N}_{\mathbf{x}_b}(\boldsymbol{\mu}_b,\mathbf{\Sigma}_b) \tag{351}$$

8.1.3 条件分布

假设 $\mathbf{x}\sim\mathcal{N}_{\mathbf{x}}(\mu,\mathbf{\Sigma})$ 其中

$$\mathbf{x}=\begin{bmatrix}\mathbf{x}_a\\ \mathbf{x}_b\end{bmatrix}\qquad \boldsymbol{\mu}=\begin{bmatrix}\boldsymbol{\mu}_a\\ \boldsymbol{\mu}_b\end{bmatrix}\qquad \mathbf{\Sigma}=\begin{bmatrix}\mathbf{\Sigma}_a & \mathbf{\Sigma}_c\\ \mathbf{\Sigma}_c^T & \mathbf{\Sigma}_b\end{bmatrix} \tag{352}$$

然后

$$p(\mathbf{x}_a|\mathbf{x}_b) = \mathcal{N}_{\mathbf{x}_a}(\hat{\mu}_a, \hat{\mathbf{\Sigma}}_a) \qquad \begin{cases} \hat{\boldsymbol{\mu}}_a &= \boldsymbol{\mu}_a + \mathbf{\Sigma}_c \mathbf{\Sigma}_b^{-1}(\mathbf{x}_b - \boldsymbol{\mu}_b) \\ \hat{\mathbf{\Sigma}}_a &= \mathbf{\Sigma}_a - \mathbf{\Sigma}_c \mathbf{\Sigma}_b^{-1} \mathbf{\Sigma}_c^T \end{cases} \tag{353}$$

$$p(\mathbf{x}_b|\mathbf{x}_a) = \mathcal{N}_{\mathbf{x}_b}(\hat{\mu}_b, \hat{\mathbf{\Sigma}}_b) \qquad \begin{cases} \hat{\boldsymbol{\mu}}_b &= \boldsymbol{\mu}_b + \mathbf{\Sigma}_c^T \mathbf{\Sigma}_a^{-1}(\mathbf{x}_a - \boldsymbol{\mu}_a) \\ \hat{\mathbf{\Sigma}}_b &= \mathbf{\Sigma}_b - \mathbf{\Sigma}_c^T \mathbf{\Sigma}_a^{-1} \mathbf{\Sigma}_c \end{cases} \tag{354}$$

注意，协方差矩阵是块矩阵的舒尔补，详见9.1.5。

8.1.4 线性组合

假设 $\mathbf{x} \sim \mathcal{N}(\mathbf{m}_x, \mathbf{\Sigma}_x)$ 和 $\mathbf{y} \sim \mathcal{N}(\mathbf{m}_y, \mathbf{\Sigma}_y)$，则

$$\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} + \mathbf{c} \sim \mathcal{N}(\mathbf{A}\mathbf{m}_x + \mathbf{B}\mathbf{m}_y + \mathbf{c}, \mathbf{A}\mathbf{\Sigma}_x\mathbf{A}^T + \mathbf{B}\mathbf{\Sigma}_y\mathbf{B}^T) \tag{355}$$

8.1.5 重新排列均值

$$\mathcal{N}_{\mathbf{A}\mathbf{x}}[\mathbf{m}, \mathbf{\Sigma}] = \sqrt{\frac{\det(2\pi(\mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{A})^{-1})}{\sqrt{\det(2\pi\mathbf{\Sigma})}}} \mathcal{N}_{\mathbf{x}}[\mathbf{A}^{-1}\mathbf{m}, (\mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{A})^{-1}] \tag{356}$$

如果 $\mathbf{A}$是方阵且可逆，简化为

$$\mathcal{N}_{\mathbf{A}\mathbf{x}}[\mathbf{m}, \mathbf{\Sigma}] = \frac{1}{|\det(\mathbf{A})|} \mathcal{N}_{\mathbf{x}}[\mathbf{A}^{-1}\mathbf{m}, (\mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{A})^{-1}] \tag{357}$$

8.1.6 重排成方阵形式

如果 $\mathbf{A}$是对称矩阵，则

$$\begin{aligned} -\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} &= -\frac{1}{2}(\mathbf{x} - \mathbf{A}^{-1}\mathbf{b})^T\mathbf{A}(\mathbf{x} - \mathbf{A}^{-1}\mathbf{b}) + \frac{1}{2}\mathbf{b}^T\mathbf{A}^{-1}\mathbf{b} \\ -\frac{1}{2}\mathrm{Tr}(\mathbf{X}^T\mathbf{A}\mathbf{X}) + \mathrm{Tr}(\mathbf{B}^T\mathbf{X}) &= -\frac{1}{2}\mathrm{Tr}[(\mathbf{X} - \mathbf{A}^{-1}\mathbf{B})^T\mathbf{A}(\mathbf{X} - \mathbf{A}^{-1}\mathbf{B})] + \frac{1}{2}\mathrm{Tr}(\mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}) \end{aligned}$$

8.1.7 两个平方形式的和

在向量形式中（假设 $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2$ 是对称的）

$$-\frac{1}{2}(\mathbf{x} - \mathbf{m}_1)^T\mathbf{\Sigma}_1^{-1}(\mathbf{x} - \mathbf{m}_1) \tag{358}$$

$$-\frac{1}{2}(\mathbf{x} - \mathbf{m}_2)^T\mathbf{\Sigma}_2^{-1}(\mathbf{x} - \mathbf{m}_2) \tag{359}$$

$$= -\frac{1}{2}(\mathbf{x} - \mathbf{m}_c)^T\mathbf{\Sigma}_c^{-1}(\mathbf{x} - \mathbf{m}_c) + C \tag{360}$$

$$\mathbf{\Sigma}_c^{-1} = \mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1} \tag{361}$$

$$\mathbf{m}_c = (\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1})^{-1}(\mathbf{\Sigma}_1^{-1}\mathbf{m}_1 + \mathbf{\Sigma}_2^{-1}\mathbf{m}_2) \tag{362}$$

$$C = \frac{1}{2}(\mathbf{m}_1^T\mathbf{\Sigma}_1^{-1} + \mathbf{m}_2^T\mathbf{\Sigma}_2^{-1})(\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1})^{-1}(\mathbf{\Sigma}_1^{-1}\text{米}_1 + \mathbf{\Sigma}_2^{-1}\text{米}_2) \tag{363}$$

$$-\frac{1}{2}\left(\text{米}_1^T\mathbf{\Sigma}_1^{-1}\text{米}_1 + \text{米}_2^T\mathbf{\Sigma}_2^{-1}\text{米}_2\right) \tag{364}$$

在迹的表述中（假设 $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2$是对称的）

$$
\begin{aligned}
& -\frac{1}{2}\mathrm{Tr}((\mathbf{X}-\mathbf{M}_1)^T\mathbf{\Sigma}_1^{-1}(\mathbf{X}-\mathbf{M}_1)) && (365)\\
& -\frac{1}{2}\mathrm{Tr}((\mathbf{X}-\mathbf{M}_2)^T\mathbf{\Sigma}_2^{-1}(\mathbf{X}-\mathbf{M}_2)) && (366)\\
= & -\frac{1}{2}\mathrm{Tr}[(\mathbf{X}-\mathbf{M}_c)^T\mathbf{\Sigma}_c^{-1}(\mathbf{X}-\mathbf{M}_c)] + C && (367)
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{\Sigma}_c^{-1} &= \mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1} && (368)\\
\mathbf{M}_c &= (\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1})^{-1}(\mathbf{\Sigma}_1^{-1}\mathbf{M}_1 + \mathbf{\Sigma}_2^{-1}\mathbf{M}_2) && (369)\\
C &= \frac{1}{2}\mathrm{Tr}\Big[(\mathbf{\Sigma}_1^{-1}\mathbf{M}_1 + \mathbf{\Sigma}_2^{-1}\mathbf{M}_2)^T(\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1})^{-1}(\mathbf{\Sigma}_1^{-1}\mathbf{M}_1 + \mathbf{\Sigma}_2^{-1}\mathbf{M}_2)\Big]\\
&\quad -\frac{1}{2}\mathrm{Tr}(\mathbf{M}_1^T\mathbf{\Sigma}_1^{-1}\mathbf{M}_1 + \mathbf{M}_2^T\mathbf{\Sigma}_2^{-1}\mathbf{M}_2) && (370)
\end{aligned}
$$

8.1.8 高斯密度的乘积

设 $\mathcal{N}_{\mathbf{x}}(\mathbf{m}, \mathbf{\Sigma})$ 表示 $\mathbf{x}$的密度，则

$$
\mathcal{N}_{\mathbf{x}}(\mathbf{m}_1, \mathbf{\Sigma}_1) \cdot \mathcal{N}_{\mathbf{x}}(\mathbf{m}_2, \mathbf{\Sigma}_2) = c_c \mathcal{N}_{\mathbf{x}}(\mathbf{m}_c, \mathbf{\Sigma}_c) \tag{371}
$$

$$
\begin{aligned}
c_c &= \mathcal{N}_{\mathbf{m}_1}(\mathbf{m}_2, (\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2))\\
&= \frac{1}{\sqrt{\det(2\pi(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2))}} \exp\left[-\frac{1}{2}(\mathbf{m}_1 - \mathbf{m}_2)^T(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2)^{-1}(\mathbf{m}_1 - \mathbf{m}_2)\right]\\
\mathbf{m}_c &= (\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1})^{-1}(\mathbf{\Sigma}_1^{-1}\mathbf{m}_1 + \mathbf{\Sigma}_2^{-1}\mathbf{m}_2)\\
\mathbf{\Sigma}_c &= (\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1})^{-1}
\end{aligned}
$$

但请注意，该乘积未归一化为密度 $\mathbf{x}$。

8.2 矩的定义

8.2.1 线性形式的均值和协方差

一阶和二阶矩。假设 $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$

$$
E(\mathbf{x}) = \mathbf{m} \tag{372}
$$

$$
\mathrm{Cov}(\mathbf{x}, \mathbf{x}) = \mathrm{Var}(\mathbf{x}) = \mathbf{\Sigma} = E(\mathbf{x}\mathbf{x}^T) - E(\mathbf{x})E(\mathbf{x}^T) = E(\mathbf{x}\mathbf{x}^T) - \mathbf{m}\mathbf{m}^T \tag{373}
$$

对于任何其他分布，高斯分布也成立

$$
\begin{aligned}
E[\mathbf{A}\mathbf{x}] &= \mathbf{A}E[\mathbf{x}] && (374)\\
\mathrm{Var}[\mathbf{A}\mathbf{x}] &= \mathbf{A}\mathrm{Var}[\mathbf{x}]\mathbf{A}^T && (375)\\
\mathrm{Cov}[\mathbf{A}\mathbf{x}, \mathbf{B}\mathbf{y}] &= \mathbf{A}\mathrm{Cov}[\mathbf{x}, \mathbf{y}]\mathbf{B}^T && (376)
\end{aligned}
$$

8.2.2 平方形式的均值和方差

平方形式的均值和方差：假设 $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$

$$\begin{aligned} E(\mathbf{x}\mathbf{x}^T) &= \mathbf{\Sigma} + \mathbf{m}\mathbf{m}^T && (377) \\ E[\mathbf{x}^T\mathbf{A}\mathbf{x}] &= \mathrm{Tr}(\mathbf{A}\mathbf{\Sigma}) + \mathbf{m}^T\mathbf{A}\mathbf{m} && (378) \\ \mathrm{Var}(\mathbf{x}^T\mathbf{A}\mathbf{x}) &= \mathrm{Tr}[\mathbf{A}\mathbf{\Sigma}(\mathbf{A}+\mathbf{A}^T)\mathbf{\Sigma}] + ... \\ & \quad + \mathbf{m}^T(\mathbf{A}+\mathbf{A}^T)\mathbf{\Sigma}(\mathbf{A}+\mathbf{A}^T)\mathbf{m} && (379) \\ E[(\mathbf{x}-\mathbf{m}')^T\mathbf{A}(\mathbf{x}-\mathbf{m}')] &= (\mathbf{m}-\mathbf{m}')^T\mathbf{A}(\mathbf{m}-\mathbf{m}') + \mathrm{Tr}(\mathbf{A}\mathbf{\Sigma}) && (380) \end{aligned}$$

如果 $\mathbf{\Sigma} = \sigma^2\mathbf{I}$且 $\mathbf{A}$是对称的，则

$$\mathrm{Var}(\mathbf{x}^T\mathbf{A}\mathbf{x}) = 2\sigma^4\mathrm{Tr}(\mathbf{A}^2) + 4\sigma^2\mathbf{m}^T\mathbf{A}^2\mathbf{m} \tag{381}$$

假设 $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ 且 $\mathbf{A}$ 和 $\mathbf{B}$ 是对称矩阵，那么

$$\mathrm{Cov}(\mathbf{x}^T\mathbf{A}\mathbf{x}, \mathbf{x}^T\mathbf{B}\mathbf{x}) = 2\sigma^4\mathrm{Tr}(\mathbf{A}\mathbf{B}) \tag{382}$$

8.2.3 三次形式

假设 $\mathbf{x}$是一个具有独立坐标、均值 $\mathbf{m}$和协方差 $\mathbf{M}$的随机向量

$$\begin{aligned} E[\mathbf{x}\mathbf{b}^T\mathbf{x}\mathbf{x}^T] &= \mathbf{m}\mathbf{b}^T(\mathbf{M}+\mathbf{m}\mathbf{m}^T) + (\mathbf{M}+\mathbf{m}\mathbf{m}^T)\mathbf{b}\mathbf{m}^T \\ & \quad + \mathbf{b}^T\mathbf{m}(\mathbf{M}-\mathbf{m}\mathbf{m}^T) && (383) \end{aligned}$$

8.2.4 四次形式的均值

$$\begin{aligned} E[\mathbf{x}\mathbf{x}^T\mathbf{x}\mathbf{x}^T] &= 2(\mathbf{\Sigma}+\mathbf{m}\mathbf{m}^T)^2 + \mathbf{m}^T\mathbf{m}(\mathbf{\Sigma}-\mathbf{m}\mathbf{m}^T) \\ & \quad + \mathrm{Tr}(\mathbf{\Sigma})(\mathbf{\Sigma}+\mathbf{m}\mathbf{m}^T) \\ E[\mathbf{x}\mathbf{x}^T\mathbf{A}\mathbf{x}\mathbf{x}^T] &= (\mathbf{\Sigma}+\mathbf{m}\mathbf{m}^T)(\mathbf{A}+\mathbf{A}^T)(\mathbf{\Sigma}+\mathbf{m}\mathbf{m}^T) \\ & \quad + \mathbf{m}^T\mathbf{A}\mathbf{m}(\mathbf{\Sigma}-\mathbf{m}\mathbf{m}^T) + \mathrm{Tr}[\mathbf{A}\mathbf{\Sigma}](\mathbf{\Sigma}+\mathbf{m}\mathbf{m}^T) \\ E[\mathbf{x}^T\mathbf{x}\mathbf{x}^T\mathbf{x}] &= 2\mathrm{Tr}(\mathbf{\Sigma}^2) + 4\mathbf{m}^T\mathbf{\Sigma}\mathbf{m} + (\mathrm{Tr}(\mathbf{\Sigma}) + \mathbf{m}^T\mathbf{m})^2 \\ E[\mathbf{x}^T\mathbf{A}\mathbf{x}\mathbf{x}^T\mathbf{B}\mathbf{x}] &= \mathrm{Tr}[\mathbf{A}\mathbf{\Sigma}(\mathbf{B}+\mathbf{B}^T)\mathbf{\Sigma}] + \mathbf{m}^T(\mathbf{A}+\mathbf{A}^T)\mathbf{\Sigma}(\mathbf{B}+\mathbf{B}^T)\mathbf{m} \\ & \quad + (\mathrm{Tr}(\mathbf{A}\mathbf{\Sigma}) + \mathbf{m}^T\mathbf{A}\mathbf{m})(\mathrm{Tr}(\mathbf{B}\mathbf{\Sigma}) + \mathbf{m}^T\mathbf{B}\mathbf{m}) \end{aligned}$$

$$\begin{aligned} & E[\mathbf{a}^T\mathbf{x}\mathbf{b}^T\mathbf{x}\mathbf{c}^T\mathbf{x}\mathbf{d}^T\mathbf{x}] \\ = \; & (\mathbf{a}^T(\mathbf{\Sigma}+\mathbf{m}\mathbf{m}^T)\mathbf{b})(\mathbf{c}^T(\mathbf{\Sigma}+\mathbf{m}\mathbf{m}^T)\mathbf{d}) \\ & + (\mathbf{a}^T(\mathbf{\Sigma}+\mathbf{m}\mathbf{m}^T)\mathbf{c})(\mathbf{b}^T(\mathbf{\Sigma}+\mathbf{m}\mathbf{m}^T)\mathbf{d}) \\ & + (\mathbf{a}^T(\mathbf{\Sigma}+\mathbf{m}\mathbf{m}^T)\mathbf{d})(\mathbf{b}^T(\mathbf{\Sigma}+\mathbf{m}\mathbf{m}^T)\mathbf{c}) - 2\mathbf{a}^T\mathbf{m}\mathbf{b}^T\mathbf{m}\mathbf{c}^T\mathbf{m}\mathbf{d}^T\mathbf{m} \end{aligned}$$

$$\begin{aligned} & E[(\mathbf{A}\mathbf{x}+\mathbf{a})(\mathbf{B}\mathbf{x}+\mathbf{b})^T(\mathbf{C}\mathbf{x}+\mathbf{c})(\mathbf{D}\mathbf{x}+\mathbf{d})^T] \\ = \; & [\mathbf{A}\mathbf{\Sigma}\mathbf{B}^T + (\mathbf{A}\mathbf{m}+\mathbf{a})(\mathbf{B}\mathbf{m}+\mathbf{b})^T][\mathbf{C}\mathbf{\Sigma}\mathbf{D}^T + (\mathbf{C}\mathbf{m}+\mathbf{c})(\mathbf{D}\mathbf{m}+\mathbf{d})^T] \\ & + [\mathbf{A}\mathbf{\Sigma}\mathbf{C}^T + (\mathbf{A}\mathbf{m}+\mathbf{a})(\mathbf{C}\mathbf{m}+\mathbf{c})^T][\mathbf{B}\mathbf{\Sigma}\mathbf{D}^T + (\mathbf{B}\mathbf{m}+\mathbf{b})(\mathbf{D}\mathbf{m}+\mathbf{d})^T] \\ & + (\mathbf{B}\mathbf{m}+\mathbf{b})^T(\mathbf{C}\mathbf{m}+\mathbf{c})[\mathbf{A}\mathbf{\Sigma}\mathbf{D}^T - (\mathbf{A}\mathbf{m}+\mathbf{a})(\mathbf{D}\mathbf{m}+\mathbf{d})^T] \\ & + \mathrm{Tr}(\mathbf{B}\mathbf{\Sigma}\mathbf{C}^T)[\mathbf{A}\mathbf{\Sigma}\mathbf{D}^T + (\mathbf{A}\mathbf{m}+\mathbf{a})(\mathbf{D}\mathbf{m}+\mathbf{d})^T] \end{aligned}$$

$$\begin{aligned}
& E[(\mathbf{Ax}+\mathbf{a})^T(\mathbf{Bx}+\mathbf{b})(\mathbf{Cx}+\mathbf{c})^T(\mathbf{Dx}+\mathbf{d})] \\
= & \operatorname{Tr}[\mathbf{A\Sigma}(\mathbf{C}^T\mathbf{D}+\mathbf{D}^T\mathbf{C})\mathbf{\Sigma B}^T] \\
& +[(\mathbf{Am}+\mathbf{a})^T\mathbf{B}+(\mathbf{Bm}+\mathbf{b})^T\mathbf{A}]\mathbf{\Sigma}[\mathbf{C}^T(\mathbf{Dm}+\mathbf{d})+\mathbf{D}^T(\mathbf{Cm}+\mathbf{c})] \\
& +[\operatorname{Tr}(\mathbf{A\Sigma B}^T)+(\mathbf{Am}+\mathbf{a})^T(\mathbf{Bm}+\mathbf{b})][\operatorname{Tr}(\mathbf{C\Sigma D}^T)+(\mathbf{Cm}+\mathbf{c})^T(\mathbf{Dm}+\mathbf{d})]
\end{aligned}$$

参见 [7].

8.2.5 矩

$$E[\mathbf{x}] = \sum_k \rho_k \mathbf{m}_k \tag{384}$$

$$\operatorname{Cov}(\mathbf{x}) = \sum_k \sum_{k'} \rho_k \rho_{k'}(\mathbf{\Sigma}_k + \mathbf{m}_k\mathbf{m}_k^T - \mathbf{m}_k\mathbf{m}_{k'}^T) \tag{385}$$

8.3 其他

8.3.1 白化

假设 $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$ 那么

$$\mathbf{z} = \mathbf{\Sigma}^{-1/2}(\mathbf{x}-\mathbf{m}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{386}$$

相反地，如果 $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 那么可以通过设置

$$\mathbf{x} = \mathbf{\Sigma}^{1/2}\mathbf{z} + \mathbf{m} \sim \mathcal{N}(\mathbf{m}, \mathbf{\Sigma}) \text{ 来生成数据} \tag{387}$$

注意 $\mathbf{\Sigma}^{1/2}$表示满足 $\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{1/2} = \mathbf{\Sigma}$的矩阵，它存在且唯一，因为 $\mathbf{\Sigma}$是正定的。

8.3.2 卡方连接

假设 $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$ 并且 $\mathbf{x}$是 n维的，那么

$$z = (\mathbf{x}-\mathbf{m})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{m}) \sim \chi_n^2 \tag{388}$$

其中 χ_n^2和表示自由度为 n的卡方分布。

8.3.3 熵

D维高斯分布的熵

$$H(\mathbf{x}) = -\int \mathcal{N}(\mathbf{m}, \mathbf{\Sigma}) \ln \mathcal{N}(\mathbf{m}, \mathbf{\Sigma}) d\mathbf{x} = \ln \sqrt{\det(2\pi\mathbf{\Sigma})} + \frac{D}{2} \tag{389}$$

8.4 高斯混合模型

8.4.1 密度

如果变量 $\mathbf{x}$服从高斯混合分布，则其密度函数为

$$p(\mathbf{x}) = \sum_{k=1}^K \rho_k \frac{1}{\sqrt{\det(2\pi\mathbf{\Sigma}_k)}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m}_k)^T\mathbf{\Sigma}_k^{-1}(\mathbf{x}-\mathbf{m}_k)\right] \tag{390}$$

其中 ρ_k之和为1，且$\mathbf{\Sigma}_k$都是正定的。

8.4.2 导数

定义 $p(\mathbf{s}) = \sum_k \rho_k \mathcal{N}_\mathbf{s}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ 得到

$$\frac{\partial \ln p(\mathbf{s})}{\partial \rho_j} = \frac{\rho_j \mathcal{N}_\mathbf{s}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_k \rho_k \mathcal{N}_\mathbf{s}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \frac{\partial}{\partial \rho_j} \ln[\rho_j \mathcal{N}_\mathbf{s}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)] \tag{391}$$

$$= \frac{\rho_j \mathcal{N}_\mathbf{s}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_k \rho_k \mathcal{N}_\mathbf{s}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \frac{1}{\rho_j} \tag{392}$$

$$\frac{\partial \ln p(\mathbf{s})}{\partial \boldsymbol{\mu}_j} = \frac{\rho_j \mathcal{N}_\mathbf{s}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_k \rho_k \mathcal{N}_\mathbf{s}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \frac{\partial}{\partial \boldsymbol{\mu}_j} \ln[\rho_j \mathcal{N}_\mathbf{s}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)] \tag{393}$$

$$= \frac{\rho_j \mathcal{N}_\mathbf{s}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_k \rho_k \mathcal{N}_\mathbf{s}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \left[\boldsymbol{\Sigma}_j^{-1}(\mathbf{s} - \boldsymbol{\mu}_j)\right] \tag{394}$$

$$\frac{\partial \ln p(\mathbf{s})}{\partial \boldsymbol{\Sigma}_j} = \frac{\rho_j \mathcal{N}_\mathbf{s}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_k \rho_k \mathcal{N}_\mathbf{s}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \frac{\partial}{\partial \boldsymbol{\Sigma}_j} \ln[\rho_j \mathcal{N}_\mathbf{s}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)] \tag{395}$$

$$= \frac{\rho_j \mathcal{N}_\mathbf{s}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_k \rho_k \mathcal{N}_\mathbf{s}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \frac{1}{2} \left[-\boldsymbol{\Sigma}_j^{-1} + \boldsymbol{\Sigma}_j^{-1}(\mathbf{s} - \boldsymbol{\mu}_j)(\mathbf{s} - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}\right] \tag{396}$$

但是 ρ_k和 $\boldsymbol{\Sigma}_k$需要受到限制。

9个特殊矩阵

9.1 块矩阵

设 $\mathbf{A}_{ij}$表示 $\mathbf{A}$的第 ij个块。

9.1.1 乘法

假设块的维度匹配，我们有

$$\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right] \left[\begin{array}{c|c} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \hline \mathbf{B}_{21} & \mathbf{B}_{22} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \hline \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{array}\right]$$

9.1.2 行列式

行列式可以通过以下方式表示

$$\mathbf{C}_1 = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} \quad (397)$$
$$\mathbf{C}_2 = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \quad (398)$$

作为

$$\text{行列式}\left(\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right]\right) = \det(\mathbf{A}_{22}) \cdot \det(\mathbf{C}_1) = \det(\mathbf{A}_{11}) \cdot \det(\mathbf{C}_2)$$

9.1.3 逆矩阵

逆矩阵可以通过以下方式表示

$$\mathbf{C}_1 = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} \quad (399)$$
$$\mathbf{C}_2 = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \quad (400)$$

as

$$\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right]^{-1} = \left[\begin{array}{c|c} \mathbf{C}_1^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{C}_2^{-1} \\ \hline -\mathbf{C}_2^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{C}_2^{-1} \end{array}\right]$$
$$= \left[\begin{array}{c|c} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{C}_2^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{C}_1^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \hline -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{C}_1^{-1} & \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{C}_1^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{array}\right]$$

9.1.4 块对角

对于块对角矩阵，我们有

$$\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{A}_{22} \end{array}\right]^{-1} = \left[\begin{array}{c|c} (\mathbf{A}_{11})^{-1} & \mathbf{0} \\ \hline \mathbf{0} & (\mathbf{A}_{22})^{-1} \end{array}\right] \quad (401)$$
$$\text{行列式}\left(\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{A}_{22} \end{array}\right]\right) = \text{行列式}\ (\mathbf{A}_{11}) \cdot \text{行列式}\ (\mathbf{A}_{22}) \quad (402)$$

9.1.5 Schur补

将矩阵视为

$$\left[\begin{array}{c|c}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \hline \mathbf{A}_{21} & \mathbf{A}_{22}\end{array}\right]$$

上述矩阵的块 $\mathbf{A}_{11}$的舒尔补是矩阵（在上文中表示为 $\mathbf{C}_2$）

$$\mathbf{A}_{22}-\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$$

上述矩阵的块 $\mathbf{A}_{22}$的舒尔补是矩阵（在上文中表示为 $\mathbf{C}_1$）

$$\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$$

使用舒尔补，可以重写块矩阵的逆

$$\begin{aligned}&\left[\begin{array}{c|c}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \hline \mathbf{A}_{21} & \mathbf{A}_{22}\end{array}\right]^{-1}\\ =\;&\left[\begin{array}{c|c}\mathbf{I} & \mathbf{0}\\ \hline -\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{I}\end{array}\right]\left[\begin{array}{c|c}(\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} & \mathbf{0}\\ \hline \mathbf{0} & \mathbf{A}_{22}^{-1}\end{array}\right]\left[\begin{array}{c|c}\mathbf{I} & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\\ \hline \mathbf{0} & \mathbf{I}\end{array}\right]\end{aligned}$$

舒尔补在解决以下形式的线性系统时很有用

$$\left[\begin{array}{c|c}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \hline \mathbf{A}_{21} & \mathbf{A}_{22}\end{array}\right]\left[\begin{array}{c}\mathbf{x}_1\\ \hline \mathbf{x}_2\end{array}\right]=\left[\begin{array}{c}\mathbf{b}_1\\ \hline \mathbf{b}_2\end{array}\right]$$

对于 $\mathbf{x}_1$，它具有以下方程

$$(\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})\mathbf{x}_1=\mathbf{b}_1-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{b}_2$$

当适当的逆存在时，可以解出 $\mathbf{x}_1$，然后将其插入方程中以解出 $\mathbf{x}_2$。

9.2 离散傅里叶变换矩阵

DFT矩阵是一个 $N\times N$对称矩阵 $\mathbf{W}_N$，其中k,n元素为

$$W_N^{kn}=e^{\frac{-j2\pi kn}{N}} \tag{403}$$

因此，离散傅里叶变换（DFT）可以表示为

$$X(k)=\sum_{n=0}^{N-1}x(n)W_{Nkn}\text{。} \tag{404}$$

同样，逆离散傅里叶变换（IDFT）可以表示为

$$x(n)=\frac{1}{N}\sum_{k=0}^{N-1}X(k)W_N^{-kn}\text{。} \tag{405}$$

向量 $\mathbf{x}=[x(0),x(1),\cdots,x(N-1)]^T$的DFT可以以矩阵形式表示

$$\mathbf{X}=\mathbf{W}_N\mathbf{x}, \tag{406}$$

其中 $\mathbf{X}= [X(0), X(1),\cdots, x(N-1)]^T$. 逆离散傅里叶变换（IDFT）同样给出为

$$\mathbf{x} = \mathbf{W}_N^{-1}\mathbf{X}. \tag{407}$$

一些 $\mathbf{W}_N$的性质存在：

$$\mathbf{W}_N^{-1} = \frac{1}{N}\mathbf{W}_N^* \tag{408}$$
$$\mathbf{W}_N\mathbf{W}_N^* = N\mathbf{I} \tag{409}$$
$$\mathbf{W}_N^* = \mathbf{W}_N^H \tag{410}$$

如果 $W_N = e^{\frac{-j2\pi}{N}}$，那么 [23]

$$W_N^{m+N/2} = -W_N^m \tag{411}$$

注意，离散傅里叶变换（DFT）矩阵是一个范德蒙德矩阵。

循环矩阵和离散傅里叶变换（DFT）之间存在以下重要关系

$$\mathbf{T}_C = \mathbf{W}_N^{-1}(\mathbf{I}\circ(\mathbf{W}_N\mathbf{t}))\mathbf{W}_N, \tag{412}$$

其中 $\mathbf{t}= [t_0, t_1,\cdots, t_{n-1}]^T$是 $\mathbf{T}_C$的第一行.

9.3 Hermite矩阵和skew-Hermite矩阵

一个矩阵 $\mathbf{A}\in\mathbb{C}^{m\times n}$被称为 *Hermitian*如果

$$\mathbf{A}^H = \mathbf{A}$$

对于实值矩阵，Hermitian和symmetric矩阵是等价的.

$$\mathbf{A}\text{是Hermitian} \quad\Leftrightarrow\quad \mathbf{x}^H\mathbf{A}\mathbf{x}\in\mathbb{R}, \quad \forall\mathbf{x}\in\mathbb{C}^{n\times 1} \tag{413}$$
$$\mathbf{A}\text{是Hermitian} \quad\Leftrightarrow\quad \text{eig}(\mathbf{A})\in\mathbb{R} \tag{414}$$

注意

$$\mathbf{A} = \mathbf{B} + i\mathbf{C}$$

其中 $\mathbf{B},\mathbf{C}$ 是Hermitian，那么

$$\mathbf{B} = \frac{\mathbf{A}+\mathbf{A}^H}{2}, \qquad \mathbf{C} = \frac{\mathbf{A}-\mathbf{A}^H}{2i}$$

9.3.1 Skew-Hermitian

如果一个矩阵 $\mathbf{A}$被称为*skew-hermitian*如果

$$\mathbf{A} = -\mathbf{A}^H$$

对于实值矩阵，反厄米矩阵和反对称矩阵是等价的。

$$\mathbf{A}\text{厄米矩阵} \quad\Leftrightarrow\quad i\mathbf{A}\text{是反厄米矩阵} \tag{415}$$
$$\mathbf{A}\ \text{反厄米矩阵} \quad\Leftrightarrow\quad \mathbf{x}^H\mathbf{A}\mathbf{y} = -\mathbf{x}^H\mathbf{A}^H\mathbf{y}, \quad \forall\mathbf{x},\mathbf{y} \tag{416}$$
$$\mathbf{A}\ \text{反厄米矩阵} \quad\Rightarrow\quad \text{eig}(\mathbf{A}) = i\lambda, \quad \lambda\in\mathbb{R} \tag{417}$$

9.4 幂等矩阵

矩阵 $\mathbf{A}$是幂等矩阵，如果

$$\mathbf{AA} = \mathbf{A}$$

幂等矩阵 $\mathbf{A}$和 $\mathbf{B}$，具有以下性质

$$\mathbf{A}^n = \mathbf{A}, \quad \text{对于} n = 1, 2, 3, \ldots \tag{418}$$

$$\mathbf{I} - \mathbf{A} \quad \text{是幂等的} \tag{419}$$

$$\mathbf{A}^H \quad \text{是幂等的} \tag{420}$$

$$\mathbf{I} - \mathbf{A}^H \quad \text{是幂等的} \tag{421}$$

$$\text{如果 } \mathbf{AB} = \mathbf{BA} \quad \Rightarrow \quad \mathbf{AB} \text{ 是幂等的} \tag{422}$$

$$\text{rank}(\mathbf{A}) = \text{Tr}(\mathbf{A}) \tag{423}$$

$$\mathbf{A}(\mathbf{I} - \mathbf{A}) = \mathbf{0} \tag{424}$$

$$(\mathbf{I} - \mathbf{A})\mathbf{A} = \mathbf{0} \tag{425}$$

$$\mathbf{A}^+ = \mathbf{A} \tag{426}$$

$$f(s\mathbf{I} + t\mathbf{A}) = (\mathbf{I} - \mathbf{A})f(s) + \mathbf{A}f(s + t) \tag{427}$$

注意 $\mathbf{A} - \mathbf{I}$不一定是幂等的。

9.4.1 零幂

如果矩阵 $\mathbf{A}$是零幂的，则

$$\mathbf{A}^2 = \mathbf{0}$$

零幂矩阵具有以下性质：

$$f(s\mathbf{I} + t\mathbf{A}) = \mathbf{I}f(s) + t\mathbf{A}f'(s) \tag{428}$$

9.4.2 单位势

如果矩阵 $\mathbf{A}$是单位势的，则

$$\mathbf{AA} = \mathbf{I}$$

单位势矩阵具有以下性质：

$$f(s\mathbf{I} + t\mathbf{A}) = [(\mathbf{I} + \mathbf{A})f(s + t) + (\mathbf{I} - \mathbf{A})f(s - t)]/2 \tag{429}$$

9.5 正交矩阵

如果一个方阵 $\mathbf{Q}$是正交的，当且仅当，

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}$$

然后 $\mathbf{Q}$具有以下特性

- 它的特征值位于单位圆上。
- 它的特征向量是酉的，即长度为一。
- 正交矩阵的逆矩阵也是正交的。

正交矩阵 $\mathbf{Q}$的基本性质

$$\begin{aligned}\mathbf{Q}^{-1} &= \mathbf{Q}^T \\ \mathbf{Q}^{-T} &= \mathbf{Q} \\ \mathbf{Q}\mathbf{Q}^T &= \mathbf{I} \\ \mathbf{Q}^T\mathbf{Q} &= \mathbf{I} \\ \det(\mathbf{Q}) &= \pm 1\end{aligned}$$

9.5.1 正交对称

一个同时正交和对称的矩阵被称为正交对称矩阵[20]。因此，$\mathbf{Q}^T{}_+\mathbf{Q}_+ = \mathbf{I}$

$$\tag{430}$$

$$\mathbf{Q}_+ = \mathbf{Q}_+^T \tag{431}$$

正交对称矩阵的幂由以下规则给出

$$\mathbf{Q}_+^k = \frac{1+(-1)^k}{2}\mathbf{I} + \frac{1+(-1)^{k+1}}{2}\mathbf{Q}_+ \tag{432}$$

$$= \frac{1+\cos(k\pi)}{2}\mathbf{I} + \frac{1-\cos(k\pi)}{2}\mathbf{Q}_+ \tag{433}$$

9.5.2 正交斜对称

一个同时正交和反对称的矩阵被称为正交斜对称矩阵[20]。因此，$\mathbf{Q}_-^H\mathbf{Q}_- = \mathbf{I}$

$$\tag{434}$$

$$\mathbf{Q}_- = -\mathbf{Q}_-^H \tag{435}$$

正交斜对称矩阵的幂由以下规则给出

$$\mathbf{Q}_-^k = \frac{i^k+(-i)^k}{2}\mathbf{I} - i\frac{i^k-(-i)^k}{2}\mathbf{Q}_- \tag{436}$$

$$= \cos(k\frac{\pi}{2})\mathbf{I} + \sin(k\frac{\pi}{2})\mathbf{Q}_- \tag{437}$$

9.5.3 分解

一个方阵 $\mathbf{A}$总是可以写成一个对称矩阵 $\mathbf{A}_+$和一个反对称矩阵 $\mathbf{A}_-$的和

$$\mathbf{A} = \mathbf{A}_+ + \mathbf{A}_- \tag{438}$$

9.6 正定和半正定矩阵

9.6.1 定义

矩阵 $\mathbf{A}$是正定的当且仅当

$$\mathbf{x}^T\mathbf{A}\mathbf{x} > \mathbf{0}, \quad \forall \mathbf{x} = \mathbf{0} \tag{439}$$

矩阵 $\mathbf{A}$是半正定的当且仅当

$$\mathbf{x}^T\mathbf{A}\mathbf{x} \geq \mathbf{0}, \quad \forall \mathbf{x} \tag{440}$$

注意，如果 $\mathbf{A}$是正定的，则 $\mathbf{A}$也是半正定的。

9.6.2 特征值

以下关于特征值成立：

$$\begin{aligned} \mathbf{A}\text{正定} \quad &\Leftrightarrow \text{特征值}\left(\frac{\mathbf{A}+\mathbf{A}^H}{2}\right) > 0 \\ \mathbf{A}\text{半正定} \quad &\Leftrightarrow \text{特征值}\left(\frac{\mathbf{A}+\mathbf{A}^H}{2}\right) \geq 0 \end{aligned} \tag{441}$$

9.6.3 迹

以下关于迹成立：

$$\begin{aligned} \mathbf{A}\text{正定} \quad &\Rightarrow \quad \mathrm{Tr}(\mathbf{A}) > 0 \\ \mathbf{A}\ \text{半正定} \quad &\Rightarrow \quad \mathrm{Tr}(\mathbf{A}) \geq 0 \end{aligned} \tag{442}$$

9.6.4 逆

如果 $\mathbf{A}$是正定的，则 $\mathbf{A}$可逆且 $\mathbf{A}^{-1}$也是正定的。

9.6.5 对角线

如果 $\mathbf{A}$是正定的，则 $A_{ii} > 0$，$\forall i$

9.6.6 分解 I

矩阵 $\mathbf{A}$是秩为 r的半正定矩阵 $\Leftrightarrow$存在一个秩为 r的矩阵 $\mathbf{B}$使得 $\mathbf{A} = \mathbf{B}\mathbf{B}^T$

矩阵 $\mathbf{A}$是正定的 $\Leftrightarrow$存在可逆矩阵 $\mathbf{B}$使得 $\mathbf{A} = \mathbf{B}\mathbf{B}^T$

9.6.7 分解 II

假设 $\mathbf{A}$是一个 $n \times n$ 半正定矩阵，那么存在一个 $n \times r$ 秩为 r 的矩阵$\mathbf{B}$使得 $\mathbf{B}^T\mathbf{A}\mathbf{B} = \mathbf{I}$.

9.6.8 零方程

假设 $\mathbf{A}$是半正定的，那么 $\mathbf{X}^T\mathbf{A}\mathbf{X} = \mathbf{0} \quad \Rightarrow \quad \mathbf{A}\mathbf{X} = \mathbf{0}$

9.6.9 乘积的秩

假设 $\mathbf{A}$是正定的，那么 $\mathrm{rank}(\mathbf{B}\mathbf{A}\mathbf{B}^T) = \mathrm{rank}(\mathbf{B})$

9.6.10 正定性质

如果 $\mathbf{A}$是一个 $n \times n$ 正定矩阵， $\mathbf{B}$是一个 $r \times n$ 秩为 r 的矩阵，那么 $\mathbf{B}\mathbf{A}\mathbf{B}^T$
是正定的.

9.6.11 外积

如果 $\mathbf{X}$是一个 $n \times r$的矩阵，其中 $n \leq r$且 $\mathrm{rank}(\mathbf{X}) = n$，则 $\mathbf{X}\mathbf{X}^T$是正定的。

9.6.12 小扰动

如果 $\mathbf{A}$是正定的且 $\mathbf{B}$是对称的，则对于足够小的t， $\mathbf{A}-t\mathbf{B}$是正定的。

9.6.13 Hadamard不等式

如果 $\mathbf{A}$是一个正定或半正定矩阵，则

$$\det(\mathbf{A}) \leq \prod_i A_{ii}$$

参见[15, pp.477]

9.6.14 Hadamard乘积关系

假设 $\mathbf{P}=\mathbf{A}\mathbf{A}^T$和 $\mathbf{Q}=\mathbf{B}\mathbf{B}^T$是半正定矩阵，则有

$$\mathbf{P}\circ\mathbf{Q}=\mathbf{R}\mathbf{R}^T$$

其中 $\mathbf{R}$的列按以下方式构造： $\mathbf{r}_{i+(j-1)N}$ ^{A} =一个$_i$ ○一个$_j$，对于 i= 1， 2， ..., N_A和 j= 1， 2， ..., N_B。结果尚未发表，但由Pavel Sakov和Craig Bishop报道。

9.7 单入口矩阵

9.7.1 定义

单元素矩阵 $\mathbf{J}$i^j $\in$ $\mathbb{R}^{n\times n}$被定义为除了在(i,j)处为1之外，其他地方都为零的矩阵。 在一个4×4的例子中，可能会有

$$\mathbf{J}^{23}=\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{443}$$

当处理涉及矩阵的表达式的导数时，单元素矩阵非常有用。

9.7.2 交换和零

假设 $\mathbf{A}$ 为 $n\times m$， $\mathbf{J}^{ij}$ 为 $m\times p$

$$\mathbf{A}\mathbf{J}^{ij}=\begin{bmatrix} \mathbf{0} & \mathbf{0} & \ldots & \mathbf{A}_i & \ldots & \mathbf{0} \end{bmatrix} \tag{444}$$

即一个零矩阵，其中第i列用 $\mathbf{A}$的第j列代替。 假设 $\mathbf{A}$是一个 $n\times m$的矩阵， $\mathbf{J}$ ^{ij}是一个 $p\times n$的矩阵

$$\mathbf{J}^{ij}\mathbf{A}=\begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{A}_j \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \tag{445}$$

即一个 $p \times m$的零矩阵，其中第j行用 $\mathbf{A}$的第i行代替。

9.7.3 重写元素的乘积

$$A_{ki}B_{jl} = (\mathbf{A}\text{的第}_i\text{行乘以 }\mathbf{B}\text{的第}_j\text{列})\text{的第}_{kl}\text{个元素} = (\mathbf{A}\mathbf{J}^{ij}\mathbf{B})\text{的第}_{kl}\text{个元素} \tag{446}$$

$$A_{ik}B_{lj} = (\mathbf{A}\text{的第}_i\text{行乘以 }\mathbf{B}\text{的第}_j\text{列的转置})\text{的第}_{kl}\text{个元素} = (\mathbf{A}\text{的转置}\mathbf{J}^{ij}\mathbf{B}\text{的} \tag{447}$$

$$A_{ik}B_{jl} = (\mathbf{A}\text{的第}_i\text{行乘以 }\mathbf{B}\text{的第}_j\text{列})\text{的第}_{kl}\text{个元素} = (\mathbf{A}\text{的转置}\mathbf{J}^{ij}\mathbf{B} \tag{448}$$

$$A_{ki}B_{lj} = (\mathbf{A}\mathbf{e}_i\mathbf{e}_j^T\mathbf{B}^T)_{kl} \quad = \quad (\mathbf{A}\mathbf{J}^{ij}\mathbf{B}^T)_{kl} \tag{449}$$

9.7.4 单元素矩阵的性质

If $i = j$

$$\mathbf{J}^{ij}\mathbf{J}^{ij} = \mathbf{J}^{ij} \qquad (\mathbf{J}^{ij})^T(\mathbf{J}^{ij})^T = \mathbf{J}^{ij}$$
$$\mathbf{J}^{ij}(\mathbf{J}^{ij})^T = \mathbf{J}^{ij} \qquad (\mathbf{J}^{ij})^T\mathbf{J}^{ij} = \mathbf{J}^{ij}$$

If $i = j$

$$\mathbf{J}^{ij}\mathbf{J}^{ij} = \mathbf{0} \qquad (\mathbf{J}^{ij})^T(\mathbf{J}^{ij})^T = \mathbf{0}$$
$$\mathbf{J}^{ij}(\mathbf{J}^{ij})^T = \mathbf{J}^{ii} \qquad (\mathbf{J}^{ij})^T\mathbf{J}^{ij} = \mathbf{J}^{jj}$$

9.7.5 标量表达式中的单元素矩阵

假设 $\mathbf{A}$是 $n \times m$的矩阵， $\mathbf{J}$是 $m \times n$的矩阵，则

$$\mathrm{Tr}(\mathbf{A}\mathbf{J}^{ij}) = \mathrm{Tr}(\mathbf{J}^{ij}\mathbf{A}) = (\mathbf{A}^T)_{ij} \tag{450}$$

假设 $\mathbf{A}$是 $n \times n$的矩阵, $\mathbf{J}$是 $n \times m$的矩阵, $\mathbf{B}$是 $m \times n$的矩阵, 那么

$$\mathrm{Tr}(\mathbf{A}\mathbf{J}^{ij}\mathbf{B}) = (\mathbf{A}^T\mathbf{B}^T)_{ij} \tag{451}$$
$$\mathrm{Tr}(\mathbf{A}\mathbf{J}^{ji}\mathbf{B}) = (\mathbf{B}\mathbf{A})_{ij} \tag{452}$$
$$\mathrm{Tr}(\mathbf{A}\mathbf{J}^{ij}\mathbf{J}^{ij}\mathbf{B}) = \mathrm{diag}(\mathbf{A}^T\mathbf{B}^T)_{ij} \tag{453}$$

假设 $\mathbf{A}$是 $n \times n$的矩阵, $\mathbf{J}^{ij}$是 $n \times m$的矩阵, $\mathbf{B}$是 $m \times n$的矩阵, 那么

$$\mathbf{x}^T\mathbf{A}\mathbf{J}^{ij}\mathbf{B}\mathbf{x} = (\mathbf{A}^T\mathbf{x}\mathbf{x}^T\mathbf{B}^T)_{ij} \tag{454}$$
$$\mathbf{x}^T\mathbf{A}\mathbf{J}^{ij}\mathbf{J}^{ij}\mathbf{B}\mathbf{x} = \mathrm{diag}(\mathbf{A}^T\mathbf{x}\mathbf{x}^T\mathbf{B}^T)_{ij} \tag{455}$$

9.7.6 结构矩阵

结构矩阵的定义如下

$$\frac{\partial \mathbf{A}}{\partial A_{ij}} = \mathbf{S}^{ij} \tag{456}$$

如果 $\mathbf{A}$没有特殊结构，则

$$\mathbf{S}^{ij} = \mathbf{J}^{ij} \tag{457}$$

如果 $\mathbf{A}$是对称的

$$\mathbf{S}^{ij} = \mathbf{J}^{ij} + \mathbf{J}^{ji} - \mathbf{J}^{ij}\mathbf{J}^{ij} \tag{458}$$

9.8 对称矩阵，反对称矩阵

9.8.1 对称矩阵

如果矩阵 $\mathbf{A}$是对称的

$$\mathbf{A} = \mathbf{A}^T \tag{459}$$

对称矩阵具有许多重要的性质，例如它们的特征值是实数，特征向量正交。

9.8.2 反对称矩阵/斜对称矩阵

反对称矩阵也被称为斜对称矩阵。 它具有以下定义的属性

$$\mathbf{A} = -\mathbf{A}^T \tag{460}$$

由此可见，反对称矩阵的对角线始终为零。$n \times n$的反对称矩阵还具有以下属性
。

$$\begin{aligned} \det(\mathbf{A}^T) &= \det(-\mathbf{A}) = (-1)^n \det(\mathbf{A}) && (461) \\ -\det(\mathbf{A}) &= \det(-\mathbf{A}) = 0, \quad \text{if } n \text{ is odd} && (462) \end{aligned}$$

反对称矩阵的特征值位于虚轴上，而特征向量是酉的。

9.8.3 分解

一个方阵 $\mathbf{A}$总是可以写成一个对称矩阵 $\mathbf{A}_+$和一个反对称矩阵 $\mathbf{A}_-$的和

$$\mathbf{A} = \mathbf{A}_+ + \mathbf{A}_- \tag{463}$$

这样的分解可以是

$$\mathbf{A} = \frac{\mathbf{A} + \mathbf{A}^T}{2} + \frac{\mathbf{A} - \mathbf{A}^T}{2} = \mathbf{A}_+ + \mathbf{A}_- \tag{464}$$

9.9 Toeplitz矩阵

Toeplitz矩阵 $\mathbf{T}$是一种每个对角线上的元素都相同的矩阵。 在方阵的情况下，它具有以下结构：

$$\mathbf{T} = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_{12} \\ t_{n1} & \cdots & t_{21} & t_{11} \end{bmatrix} = \begin{bmatrix} t_0 & t_1 & \cdots & t_{n-1} \\ t_{-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_1 \\ t_{-(n-1)} & \cdots & t_{-1} & t_0 \end{bmatrix} \tag{465}$$

一个Toeplitz矩阵是对称的。 如果一个矩阵是对称的（或正交对称的），这意味着该矩阵关于其东北-西南对角线（反对角线）对称[12]。 对称矩阵是更大的矩阵类别，因为对称矩阵不一定具有Toeplitz结构。这

是一些特殊情况的Toeplitz矩阵。 对称Toeplitz矩阵的表示为：

$$\mathbf{T} = \begin{bmatrix} t_0 & t_1 & \cdots & t_{n-1} \\ t_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_1 \\ t_{n-1} & \cdots & t_1 & t_0 \end{bmatrix} \tag{466}$$

循环Toeplitz矩阵：

$$\mathbf{T}_C = \begin{bmatrix} t_0 & t_1 & \cdots & t_{n-1} \\ t_{n-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_1 \\ t_1 & \cdots & t_{n-1} & t_0 \end{bmatrix} \tag{467}$$

上三角Toeplitz矩阵：

$$\mathbf{T}_U = \begin{bmatrix} t_0 & t_1 & \cdots & t_{n-1} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_1 \\ 0 & \cdots & 0 & t_0 \end{bmatrix}, \tag{468}$$

和下三角Toeplitz矩阵：

$$\mathbf{T}_L = \begin{bmatrix} t_0 & 0 & \cdots & 0 \\ t_{-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ t_{-(n-1)} & \cdots & t_{-1} & t_0 \end{bmatrix} \tag{469}$$

9.9.1 Toeplitz矩阵的性质

Toeplitz矩阵具有一些计算优势。 两个Toeplitz矩阵的加法可以在 $\mathcal{O}(n)$ 次浮点运算中完成，两个Toeplitz矩阵的乘法可以在 $\mathcal{O}(n \ln n)$ 次浮点运算中完成。 T oeplitz方程组可以在 $\mathcal{O}(n^2)$ 次浮点运算中求解。 正定Toeplitz矩阵的逆矩阵也可以在 $\mathcal{O}(n^2)$ 次浮点运算中找到。Toeplitz矩阵的逆矩阵是对称的。 两个下三角Toeplitz矩阵的乘积是一个Toeplitz矩阵。 有关Toeplitz矩阵和循环矩阵的更多信息可以在[13, 7]中找到。

9.10 过渡矩阵

一个方阵 $\mathbf{P}$是一个过渡矩阵，也被称为随机矩阵或
概率矩阵，如果

$$0 \leq (\mathbf{P})_{ij} \leq 1, \qquad \sum_j (\mathbf{P})_{ij} = 1$$

过渡矩阵通常描述了从状态 i到状态 j在一步中移动的概率，并且与马尔可夫过程密切相关。过渡矩阵

具有以下属性

$$\begin{aligned}
\mathrm{Prob}[i \to j \text{ 在1步中}] &= (\mathbf{P})_{ij} && (470)\\
\mathrm{Prob}[i \to j \text{ 在2步中}] &= (\mathbf{P}^2)_{ij} && (471)\\
\mathrm{Prob}[i \to j \text{在 } k \text{步中}] &= (\mathbf{P}^k)_{ij} && (472)\\
\text{如果所有行都相同} &\Rightarrow \mathbf{P}^n = \mathbf{P} && (473)\\
\boldsymbol{\alpha}\mathbf{P} &= \boldsymbol{\alpha}, \quad \boldsymbol{\alpha}\text{被称为不变} && (474)
\end{aligned}$$

其中 $\boldsymbol{\alpha}$是所谓的稳态概率向量，即 $0 \leq \alpha_i \leq 1$ 和 $\sum_i \alpha_i = 1$.

9.11 单位矩阵，排列和移位

9.11.1 单位向量

设 $\mathbf{e}_i \in \mathbb{R}^{n\times 1}$为第 i个单位向量，即在所有条目中都为零，除了第 i个条目为1。

9.11.2 行和列

$$\begin{aligned}
\mathbf{A}\text{的第}i\text{行} &= \mathbf{e}_i^T \mathbf{A} && (475)\\
\mathbf{A}\text{的第}_j\text{列} &= \mathbf{A}\mathbf{e}_j && (476)
\end{aligned}$$

9.11.3 排列

设 $\mathbf{P}$为某个排列矩阵，例如

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{e}_2 & \mathbf{e}_1 & \mathbf{e}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{e}_2^T \\ \mathbf{e}_1^T \\ \mathbf{e}_3^T \end{bmatrix} \quad (477)$$

对于排列矩阵，有以下性质

$$\mathbf{P}\mathbf{P}^T = \mathbf{I} \quad (478)$$

以及

$$\mathbf{A}\mathbf{P} = \begin{bmatrix} \mathbf{A}\mathbf{e}_2 & \mathbf{A}\mathbf{e}_1 & \mathbf{A}\mathbf{e}_3 \end{bmatrix} \qquad \mathbf{P}\mathbf{A} = \begin{bmatrix} \mathbf{e}_2^T\mathbf{A} \\ \mathbf{e}_1^T\mathbf{A} \\ \mathbf{e}_3^T\mathbf{A} \end{bmatrix} \quad (479)$$

也就是说，第一个矩阵的列是 $\mathbf{A}$的列按照排列顺序排列，而第二个矩阵的行是 $\mathbf{A}$的行按照排列顺序排列。

9.11.4 平移、移动或滞后算子

设 $\mathbf{L}$表示4×4

示例中定义的滞后（或“平移”或“移动”）算子

$$\mathbf{L} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (480)$$

即一个零矩阵，其次对角线上为1，$(\mathbf{L})_{ij} = \delta_{i,j+1}$。对于一些信号 x_t，其中 $t = 1, ..., N$，滞后算子的n次幂会改变索引，即

$$(\mathbf{L}^n\mathbf{x})_t = \begin{cases} 0 & \text{对于 } t = 1, .., n \\ \text{对于}_{t} =_{n} + 1, ..., N\text{，滞后算子的}n\text{次幂会将索引向后移动}n\text{个位置} \end{cases} \tag{481}$$

一个相关但略有不同的矩阵是在4x4示例中定义的“循环平移”算子

$$\hat{\mathbf{L}} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{482}$$

即一个矩阵，其定义为 $(\hat{\mathbf{L}})_{ij} = \delta_{i,j+1} + \delta_{i,1}\delta_{j,dim(\mathbf{L})}$。对于一个信号 $\mathbf{x}$，它的作用是

$$(\hat{\mathbf{L}}^n\mathbf{x})_t = x_{t'}, \quad t' = [(t-n) \mod N] + 1 \tag{483}$$

也就是说，$\hat{\mathbf{L}}$就像移位运算符 $\mathbf{L}$一样，只是它将信号“包裹”起来，就像周期性地移动（用信号的末尾替换零值）。

请注意，$\hat{\mathbf{L}}$是可逆的和正交的，即

$$\hat{\mathbf{L}}^{-1} = \hat{\mathbf{L}}^T \tag{484}$$

9.12 Vandermonde矩阵

Vandermonde矩阵的形式为[15]

$$\mathbf{V} = \begin{bmatrix} 1 & v_1 & v_1^2 & \cdots & v_1^{n-1} \\ 1 & v_2 & v_2^2 & \cdots & v_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & v_n & v_n^2 & \cdots & v_n^{n-1} \end{bmatrix}. \tag{485}$$

$\mathbf{V}$的转置也被称为范德蒙德矩阵。行列式由[29]给出

$$\det \mathbf{V} = \prod_{i>j}(v_i - v_j) \tag{486}$$

10 函数和运算符

10.1 函数和级数

10.1.1 有限级数

$$(\mathbf{X}^n - \mathbf{I})(\mathbf{X} - \mathbf{I})^{-1} = \mathbf{I} + \mathbf{X} + \mathbf{X}^2 + ... + \mathbf{X}^{n-1} \tag{487}$$

10.1.2 标量函数的泰勒展开

考虑一些以向量$\mathbf{x}$为参数的标量函数 $f(\ \mathbf{x})$。
我们可以围绕 $\mathbf{x}_0$进行泰勒展开

$$f(\mathbf{x}) \cong f(\mathbf{x}_0) + \mathbf{g}(\mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \mathbf{H}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \tag{488}$$

哪里

$$\mathbf{g}(\mathbf{x}_0) = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}\Big|_{\mathbf{x}_0} \qquad \mathbf{H}(\mathbf{x}_0) = \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^T}\Big|_{\mathbf{x}_0}$$

10.1.3 无穷级数矩阵函数

对于一维解析函数，可以通过无穷级数为方阵 $\mathbf{X}$定义一个矩阵函数

$$\mathbf{f}(\mathbf{X}) = \sum_{n=0}^{\infty} c_n \mathbf{X}^n \tag{489}$$

假设极限存在且有限。 如果系数 c_n满足 $\sum_n c_n x^n < \infty$，则可以证明上述级数存在且有限，参见[1]。 因此，对于任何解析函数 $f(x)$，都存在一个相应的矩阵函数 $\mathbf{f}(\mathbf{x})$通过泰勒展开构造。 利用这一点，可以证明以下结果：

1）矩阵 $\mathbf{A}$是其特征多项式的零点[1]：

$$p(\lambda) = \det(\mathbf{I}\lambda - \mathbf{A}) = \sum_n c_n \lambda^n \qquad \Rightarrow \qquad p(\mathbf{A}) = \mathbf{0} \tag{490}$$

2) 如果 $\mathbf{A}$是方阵，则成立 [1]

$$\mathbf{A} = \mathbf{U}\mathbf{B}\mathbf{U}^{-1} \quad \Rightarrow \quad \mathbf{f}(\mathbf{A}) = \mathbf{U}\mathbf{f}(\mathbf{B})\mathbf{U}^{-1} \tag{491}$$

3) 在使用幂级数时，一个有用的事实是

$$\mathbf{A}^n \to \mathbf{0} \text{当 } n \to \infty \qquad \text{if} \qquad |\mathbf{A}| < 1 \tag{492}$$

10.1.4 身份和交换

对于一个解析矩阵函数 $\mathbf{f}(\mathbf{X})$，有以下关系成立

$$\mathbf{f}(\mathbf{AB})\mathbf{A} = \mathbf{A}\mathbf{f}(\mathbf{BA}) \tag{493}$$

参见 B.1.2 以获取证明。

10.1.5 指数矩阵函数

类似于普通标量指数函数，可以定义指数和对数矩阵函数：

$$e^{\mathbf{A}} \equiv \sum_{n=0}^{\infty} \frac{1}{n!}\mathbf{A}^n = \mathbf{I} + \mathbf{A} + \frac{1}{2}\mathbf{A}^2 + ... \tag{494}$$

$$e^{-\mathbf{A}} \equiv \sum_{n=0}^{\infty} \frac{1}{n!}(-1)^n\mathbf{A}^n = \mathbf{I} - \mathbf{A} + \frac{1}{2}\mathbf{A}^2 - ... \tag{495}$$

$$e^{t\mathbf{A}} \equiv \sum_{n=0}^{\infty} \frac{1}{n!}(t\mathbf{A})^n = \mathbf{I} + t\mathbf{A} + \frac{1}{2}t^2\mathbf{A}^2 + ... \tag{496}$$

$$\ln(\mathbf{I}+\mathbf{A}) \equiv \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}\mathbf{A}^n = \mathbf{A} - \frac{1}{2}\mathbf{A}^2 + \frac{1}{3}\mathbf{A}^3 - ... \tag{497}$$

指数函数的一些性质[1]

$$e^{\mathbf{A}}e^{\mathbf{B}} = e^{\mathbf{A}+\mathbf{B}} \quad \text{if} \quad \mathbf{AB} = \mathbf{BA} \tag{498}$$

$$(e^{\mathbf{A}})^{-1} = e^{-\mathbf{A}} \tag{499}$$

$$\frac{d}{dt}e^{t\mathbf{A}} = \mathbf{A}e^{t\mathbf{A}} = e^{t\mathbf{A}}\mathbf{A}, \quad t \in \mathbb{R} \tag{500}$$

$$\frac{d}{dt}\mathrm{Tr}(e^{t\mathbf{A}}) = \mathrm{Tr}(\mathbf{A}e^{t\mathbf{A}}) \tag{501}$$

$$\det(e^{\mathbf{A}}) = e^{\mathrm{Tr}(\mathbf{A})} \tag{502}$$

10.1.6 三角函数

$$\sin(\mathbf{A}) \equiv \sum_{n=0}^{\infty} \frac{(-1)^n\mathbf{A}^{2n+1}}{(2n+1)!} = \mathbf{A} - \frac{1}{3!}\mathbf{A}^3 + \frac{1}{5!}\mathbf{A}^5 - ... \tag{503}$$

$$\cos(\mathbf{A}) \equiv \sum_{n=0}^{\infty} \frac{(-1)^n\mathbf{A}^{2n}}{(2n)!} = \mathbf{I} - \frac{1}{2!}\mathbf{A}^2 + \frac{1}{4!}\mathbf{A}^4 - ... \tag{504}$$

10.2 Kronecker和Vec运算符

10.2.1 克罗内克积

一个 $m \times n$矩阵 $\mathbf{A}$和一个 $r \times q$矩阵 $\mathbf{B}$的克罗内克积是一个
$mr \times nq$ 矩阵， $\mathbf{A} \otimes \mathbf{B}$ 定义为

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} A_{11}\mathbf{B} & A_{12}\mathbf{B} & ... & A_{1n}\mathbf{B} \\ A_{21}\mathbf{B} & A_{22}\mathbf{B} & ... & A_{2n}\mathbf{B} \\ \vdots & & & \vdots \\ A_{m1}\mathbf{B} & A_{m2}\mathbf{B} & ... & A_{mn}\mathbf{B} \end{bmatrix} \tag{505}$$

克罗内克积具有以下性质（见[19]）

$$\begin{aligned}
\mathbf{A}\otimes(\mathbf{B}+\mathbf{C}) &= \mathbf{A}\otimes\mathbf{B}+\mathbf{A}\otimes\mathbf{C} && (506)\\
\mathbf{A}\otimes\mathbf{B} &= \mathbf{B}\otimes\mathbf{A} \quad \text{一般来说} && (507)\\
\mathbf{A}\otimes(\mathbf{B}\otimes\mathbf{C}) &= (\mathbf{A}\otimes\mathbf{B})\otimes\mathbf{C} && (508)\\
(\alpha_A\mathbf{A}\otimes\alpha_B\mathbf{B}) &= \alpha_A\alpha_B(\mathbf{A}\otimes\mathbf{B}) && (509)\\
(\mathbf{A}\otimes\mathbf{B})^T &= \mathbf{A}^T\otimes\mathbf{B}^T && (510)\\
(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) &= \mathbf{AC}\otimes\mathbf{BD} && (511)\\
(\mathbf{A}\otimes\mathbf{B})^{-1} &= \mathbf{A}^{-1}\otimes\mathbf{B}^{-1} && (512)\\
(\mathbf{A}\otimes\mathbf{B})^{+} &= \mathbf{A}^{+}\otimes\mathbf{B}^{+} && (513)\\
\mathrm{rank}(\mathbf{A}\otimes\mathbf{B}) &= \mathrm{rank}(\mathbf{A})\mathrm{rank}(\mathbf{B}) && (514)\\
\mathrm{Tr}(\mathbf{A}\otimes\mathbf{B}) &= \mathrm{Tr}(\mathbf{A})\mathrm{Tr}(\mathbf{B}) = \mathrm{Tr}(\mathbf{\Lambda}_A\otimes\mathbf{\Lambda}_B) && (515)\\
\det(\mathbf{A}\otimes\mathbf{B}) &= \det(\mathbf{A})^{\mathrm{rank}(\mathbf{B})}\det(\mathbf{B})^{\mathrm{rank}(\mathbf{A})} && (516)\\
\{\mathrm{eig}(\mathbf{A}\otimes\mathbf{B})\} &= \{\mathrm{eig}(\mathbf{B}\otimes\mathbf{A})\} \quad \text{如果 } \mathbf{A},\mathbf{B}\text{是方阵} && (517)\\
\{\mathrm{eig}(\mathbf{A}\otimes\mathbf{B})\} &= \{\mathrm{eig}(\mathbf{A})\mathrm{eig}(\mathbf{B})^T\} && (518)\\
& \quad \text{如果 } \mathbf{A},\mathbf{B}\text{是对称的方阵} &&\\
\mathrm{eig}(\mathbf{A}\otimes\mathbf{B}) &= \mathrm{eig}(\mathbf{A})\otimes\mathrm{eig}(\mathbf{B}) && (519)
\end{aligned}$$

其中 $\{\lambda_i\}$表示值 λ_i的集合，即无特定顺序或结构的值， $\mathbf{\Lambda}_A$表示具有特征值的对角矩阵
$\mathbf{A}$。

10.2.2 向量运算符

向量运算符应用于矩阵 $\mathbf{A}$将列堆叠成向量，即对于一个2×2的矩阵

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \qquad \mathrm{vec}(\mathbf{A}) = \begin{bmatrix} A_{11} \\ A_{21} \\ A_{12} \\ A_{22} \end{bmatrix}$$

向量运算符的性质包括（参见[19]）

$$\begin{aligned}
\mathrm{vec}(\mathbf{AXB}) &= (\mathbf{B}^T\otimes\mathbf{A})\mathrm{vec}(\mathbf{X}) && (520)\\
\mathrm{Tr}(\mathbf{A}^T\mathbf{B}) &= \mathrm{vec}(\mathbf{A})^T\mathrm{vec}(\mathbf{B}) && (521)\\
\mathrm{vec}(\mathbf{A}+\mathbf{B}) &= \mathrm{vec}(\mathbf{A}) + \mathrm{vec}(\mathbf{B}) && (522)\\
\mathrm{vec}(\alpha\mathbf{A}) &= \alpha\cdot\mathrm{vec}(\mathbf{A}) && (523)\\
\mathbf{a}^T\mathbf{XBX}^T\mathbf{c} &= \mathrm{vec}(\mathbf{X})^T(\mathbf{B}\otimes\mathbf{ca}^T)\mathrm{vec}(\mathbf{X}) && (524)
\end{aligned}$$

参见 B.1.1 以证明等式 524。

10.3 向量范数

10.3.1 例子

$$||\mathbf{x}||_1 = \sum_i |x_i| \quad (525)$$

$$||\mathbf{x}||_2^2 = \mathbf{x}^H\mathbf{x} \quad (526)$$

$$||\mathbf{x}||_p = \left[\sum_i |x_i|^p\right]^{1/p} \quad (527)$$

$$||\mathbf{x}||_\infty = \max_i |x_i| \quad (528)$$

进一步阅读，例如 [12, p. 52]

10.4 矩阵范数

10.4.1 定义

矩阵范数是一个满足

$$||\mathbf{A}|| \geq 0\text{的映射} \quad (529)$$

$$||\mathbf{A}|| = 0 \Leftrightarrow \mathbf{A} = \mathbf{0} \quad (530)$$

$$||c\mathbf{A}|| = |c|||\mathbf{A}||, \qquad c \in \mathbb{R} \quad (531)$$

$$||\mathbf{A}+\mathbf{B}|| \leq ||\mathbf{A}|| + ||\mathbf{B}|| \quad (532)$$

10.4.2 引导范数或算子范数

引导范数是由向量范数引导的矩阵范数，由以下公式定义

$$||\mathbf{A}|| = \sup\{||\mathbf{A}\mathbf{x}|| \quad | \quad ||\mathbf{x}|| = 1\} \quad (533)$$

左边的$||\cdot||$表示引导矩阵范数，而右边的$||\cdot||$表示向量范数。 对于引导范数，有$||\mathbf{I}|| \ = 1$

$$(534)$$

$$||\mathbf{A}\mathbf{x}|| \leq ||\mathbf{A}|| \cdot ||\mathbf{x}||, \qquad \text{对于所有 } \mathbf{A}, \ \mathbf{x} \quad (535)$$

$$||\mathbf{A}\mathbf{B}|| \leq ||\mathbf{A}|| \cdot ||\mathbf{B}||, \qquad \text{对于所有 } \mathbf{A}, \mathbf{B} \quad (536)$$

10.4.3 例子

$$||\mathbf{A}||_1 = \max_j \sum_i |A_{ij}| \quad (537)$$

$$||\mathbf{A}||_2 = \sqrt{\max \text{eig}(\mathbf{A}^H\mathbf{A})} \quad (538)$$

$$||\mathbf{A}||_p = (\max_{||\mathbf{x}||_p=1} ||\mathbf{A}\mathbf{x}||_p)^{1/p} \quad (539)$$

$$||\mathbf{A}||_\infty = \max_i \sum_j |A_{ij}| \quad (540)$$

$$||\mathbf{A}||_\text{F} = \sqrt{\sum_{ij} |A_{ij}|^2} = \sqrt{\text{Tr}(\mathbf{A}\mathbf{A}^H)} \qquad \text{(Frobenius)} \quad (541)$$

$$||\mathbf{A}||_{\max} = \max_{ij} |A_{ij}| \tag{542}$$

$$||\mathbf{A}||_{\mathrm{KF}} = ||\mathrm{sing}(\mathbf{A})||_1 \qquad \text{(Ky Fan)} \tag{543}$$

其中 $\mathrm{sing}(\mathbf{A})$ 是矩阵 $\mathbf{A}$的奇异值向量。

10.4.4 不等式

E. H. Ras$m$ussen 在尚未发表的资料中推导和收集了以下不等式。 它们被整理在下表中，假设 $\mathbf{A}$是一个 $m \times n$ 的矩阵， $d=\mathrm{rank}(\mathbf{A})$

	$\|\|\mathbf{A}\|\|$最大值	$\|\|\mathbf{A}\|\|_1$ 范数	$\|\|\mathbf{A}\|\|$无穷范数	$\|\|\mathbf{A}\|\|_2$ 范数	$\|\|\mathbf{A}\|\|_\mathrm{F}$ 范数	$\|\|\mathbf{A}\|\|_\mathrm{KF}$ 范数
$\|\|\mathbf{A}\|\|$最大值		1	1	1	1	1
$\|\|\mathbf{A}\|\|_1$	m		m	$\sqrt{米}$	$\sqrt{米}$	$\sqrt{米}$
$\|\|\mathbf{A}\|\|$无穷范数	n	n		$\sqrt{恩}$	$\sqrt{恩}$	$\sqrt{恩}$
$\|\|\mathbf{A}\|\|_2$ 范数	$\sqrt{米恩}$	$\sqrt{恩}$	$\sqrt{来}$		1	1
$\|\|\mathbf{A}\|\|_\mathrm{F}$	$\sqrt{米恩}$	$\sqrt{恩}$	$\sqrt{来}$	$\sqrt{d}$		1
$\|\|\mathbf{A}\|\|_\mathrm{KF}$ 范数	$\sqrt{米恩德}$	$\sqrt{恩德}$	$\sqrt{米德}$	d	$\sqrt{d}$	

这些应该被理解为，例如

$$||\mathbf{A}||_2 \leq \sqrt{来} \cdot ||\mathbf{A}||无穷范数 \tag{544}$$

10.4.5 条件数

矩阵 $\mathbf{A}$的2范数等于 $\sqrt{(\max(\mathrm{eig}(\mathbf{A}^T\mathbf{A})))}$ [12, p.57]。对于对称的、正定的矩阵，这可以简化为$\max(\mathrm{eig}(\mathbf{A}))$ 基于2范数的条件数因此简化为

$$\|\mathbf{A}\|_2 \|\mathbf{A}^{-1}\|_2 = \max(\mathrm{eig}(\mathbf{A}))\, \max(\mathrm{eig}(\mathbf{A}^{-1})) = \frac{\max(\mathrm{eig}(\mathbf{A}))}{\min(\mathrm{eig}(\mathbf{A}))}. \tag{545}$$

10.5 秩

10.5.1 斯尔维斯特不等式

如果 $\mathbf{A}$是 $m \times n$， $\mathbf{B}$是 $n \times r$，那么

$$\mathrm{rank}(\mathbf{A}) + \mathrm{rank}(\mathbf{B}) - n \leq \mathrm{rank}(\mathbf{AB}) \leq \min\{\mathrm{rank}(\mathbf{A}), \mathrm{rank}(\mathbf{B})\} \tag{546}$$

10.6 包含Dirac Delta函数的积分

假设 $\mathbf{A}$ 是方阵，那么

$$\int p(\mathbf{s})\delta(\mathbf{x} - \mathbf{As})d\mathbf{s} = \frac{1}{\det(\mathbf{A})} p(\mathbf{A}^{-1}\mathbf{x}) \tag{547}$$

假设 $\mathbf{A}$是“欠定”的，即“高瘦”的，那么

$$\int p(\mathbf{s})\delta(\mathbf{x} - \mathbf{As})d\mathbf{s} = \left\{ \begin{array}{ll} \frac{1}{\sqrt{\det(\mathbf{A}^T\mathbf{A})}} p(\mathbf{A}^+\mathbf{x}) & \text{如果 } \mathbf{x} = \mathbf{AA}^+\mathbf{x} \\ 0 & \text{其他地方} \end{array} \right\} \tag{548}$$

见[9]

10.7 其他

对于任意的 $\mathbf{A}$都成立

$$\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{A}^T) = \mathrm{rank}(\mathbf{A}\mathbf{A}^T) = \mathrm{rank}(\mathbf{A}^T\mathbf{A}) \tag{549}$$

它成立

$$\mathbf{A}\text{是正定的} \Leftrightarrow \text{存在}\mathbf{B}\text{可逆，使得 } \mathbf{A} = \mathbf{B}\mathbf{B}^T \tag{550}$$

一维结果

A.1 高斯分布

A.1.1 密度

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{551}$$

A.1.2 归一化

$$\int e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds = \sqrt{2\pi\sigma^2} \tag{552}$$

$$\int e^{-(ax^2+bx+c)} dx = \sqrt{\frac{\pi}{a}} \exp\left[\frac{b^2-4ac}{4a}\right] \tag{553}$$

$$\int e^{c_2x^2+c_1x+c_0} dx = \sqrt{\frac{\pi}{-克_2}} \exp\left[\frac{克_1^2 - 4克_2克_0}{-4克_2}\right] \tag{554}$$

A.1.3 导数

$$\frac{\partial p(克)}{\partial \mu} = p(x)\frac{(克-\mu)}{\sigma^2} \tag{555}$$

$$\frac{\partial \ln p(克)}{\partial \mu} = \frac{(克-\mu)}{\sigma^2} \tag{556}$$

$$\frac{\partial p(克)}{\partial \sigma} = p(x)\frac{1}{\sigma}\left[\frac{(克-\mu)^2}{\sigma^2} - 1\right] \tag{557}$$

$$\frac{\partial \ln p(克)}{\partial \sigma} = \frac{1}{\sigma}\left[\frac{(克-\mu)^2}{\sigma^2} - 1\right] \tag{558}$$

A.1.4 完成平方

$$克_2克^2 + 克_1克 + 克_0 = -克(克-克)^2 + 克$$

$$-克 = 克_2 \qquad 克 = \frac{1}{2}\frac{克_1}{克_2} \qquad 克 = \frac{1}{4}\frac{克_1^2}{克_2} + 克_0$$

或

$$克_2克^2 + 克_1克 + 克_0 = \frac{1}{2\sigma^2}(克-\mu)^2 + 克$$

$$\mu = \frac{-克_1}{2c_2} \qquad \sigma^2 = \frac{-1}{2c_2} \qquad d = c_0 - \frac{c_1^2}{4c_2}$$

A.1.5 矩阵的矩

如果密度由以下方式表示

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(s-\mu)^2}{2\sigma^2}\right] \quad \text{或者} \quad p(x) = C\exp(c_2x^2 + c_1x) \tag{559}$$

那么前几个基本矩为

$$
\begin{aligned}
\langle x\rangle &= \mu &&= \tfrac{-c_1}{2\text{克}_2}\\
\langle x^2\rangle &= \sigma^2+\mu^2 &&= \tfrac{-1}{2\text{克}_2}+\left(\tfrac{-c_1}{2c_2}\right)^2\\
\langle x^3\rangle &= 3\sigma^2\mu+\mu^3 &&= \tfrac{\text{克}_1}{(2c_2)^2}\left[3-\tfrac{c_1^2}{2c_2}\right]\\
\langle x^4\rangle &= \mu^4+6\mu^2\sigma^2+3\sigma^4 &&= \left(\tfrac{\text{克}_1}{2c_2}\right)^4+6\left(\tfrac{\text{克}1}{2\text{克}2}\right)^2\left(\tfrac{-1}{2c_2}\right)+3\left(\tfrac{1}{2c_2}\right)^2
\end{aligned}
$$

以及中心矩是

$$
\begin{aligned}
\langle(x-\mu)\rangle &= 0 &&= 0\\
\langle(x-\mu)^2\rangle &= \sigma^2 &&= \left[\tfrac{-1}{2c_2}\right]\\
\langle(x-\mu)^3\rangle &= 0 &&= 0\\
\langle(x-\mu)^4\rangle &= 3\sigma^4 &&= 3\left[\tfrac{1}{2\text{克}2}\right]^2
\end{aligned}
$$

一种伪矩（非归一化积分）可以很容易地推导出来

$$
\int \exp(c_2x^2+c_1x)x^n dx = Z\langle x^n\rangle = \sqrt{\frac{\pi}{-c_2}}\exp\left[\frac{c_1^2}{-4\text{克}_2}\right]\langle x^n\rangle \qquad (560)
$$

从非中心矩可以推导出其他实体，如

$$
\begin{aligned}
\langle x^2\rangle-\langle x\rangle^2 &= \sigma^2 &&= \tfrac{-1}{2\text{克}2}\\
\langle x^3\rangle-\langle x^2\rangle\langle x\rangle &= 2\sigma^2\mu &&= \tfrac{2c_1}{(2c_2)^2}\\
\langle x^4\rangle-\langle x^2\rangle^2 &= 2\sigma^4+4\mu^2\sigma^2 &&= \tfrac{2}{(2c_2)^2}\left[1-4\tfrac{c_1^2}{2c_2}\right]
\end{aligned}
$$

A.2 一维高斯混合

A.2.1 密度和归一化

$$
p(s)=\sum_k^K \frac{\rho_k}{\sqrt{2\pi\sigma_k^2}}\exp\left[-\frac{1}{2}\frac{(s-\mu_k)^2}{\sigma_k^2}\right] \qquad (561)
$$

A.2.2 矩

MoG的一个有用的事实是

$$
\langle x^n\rangle=\sum_k \rho_k\langle x^n\rangle_k \qquad (562)
$$

其中 $\langle\cdot\rangle_k$表示对第 k个分量的平均。 我们可以从密度函数计算前四个矩

$$
\begin{aligned}
p(x) &= \sum_k \rho_k\frac{1}{\sqrt{2\pi\sigma_k^2}}\exp\left[-\frac{1}{2}\frac{(x-\mu_k)^2}{\sigma_k^2}\right] && (563)\\
p(x) &= \sum_k \rho_k C_k\exp\left[c_{k2}x^2+c_{k1}x\right] && (564)
\end{aligned}
$$

如

$$
\begin{array}{rclcl}
\langle x\rangle & = & \sum_k \rho_k\mu_k & = & \sum_k \rho_k \left[\frac{-c_{k1}}{2c_{k2}}\right] \\
\langle x^2\rangle & = & \sum_k \rho_k(\sigma_k^2+\mu_k^2) & = & \sum_k \rho_k \left[\frac{-1}{2c_{k2}}+\left(\frac{-c_{k1}}{2c_{k2}}\right)^2\right] \\
\langle x^3\rangle & = & \sum_k \rho_k(3\sigma_k^2\mu_k+\mu_k^3) & = & \sum_k \rho_k \left[\frac{c_{k1}}{(2c_{k2})^2}\left[3-\frac{c_{k1}^2}{2c_{k2}}\right]\right] \\
\langle x^4\rangle & = & \sum_k \rho_k(\mu_k^4+6\mu_k^2\sigma_k^2+3\sigma_k^4) & = & \sum_k \rho_k \left[\left(\frac{1}{2c_{k2}}\right)^2\left[\left(\frac{c_{k1}}{2c_{k2}}\right)^2-6\frac{c_{k1}^2}{2c_{k2}}+3\right]\right]
\end{array}
$$

如果所有的高斯函数都是以中心为基准的，即 $\mu_k=0$ 对于所有的 k, 那么

$$
\begin{array}{rclcl}
\langle x\rangle & = & 0 & = & 0 \\
\langle x^2\rangle & = & \sum_k \rho_k\sigma_k^2 & = & \sum_k \rho_k \left[\frac{-1}{2c_{k2}}\right] \\
\langle x^3\rangle & = & 0 & = & 0 \\
\langle x^4\rangle & = & \sum_k \rho_k 3\sigma_k^4 & = & \sum_k \rho_k 3\left[\frac{-1}{2c_{k2}}\right]^2
\end{array}
$$

从非中心矩可以推导出其他实体，如

$$
\begin{array}{rcl}
\langle x^2\rangle-\langle x\rangle^2 & = & \sum_{k,k'} \rho_k\rho_{k'}\left[\mu_k^2+\sigma_k^2-\mu_k\mu_{k'}\right] \\
\langle x^3\rangle-\langle x^2\rangle\langle x\rangle & = & \sum_{k,k'} \rho_k\rho_{k'}\left[3\sigma_k^2\mu_k+\mu_k^3-(\sigma_k^2+\mu_k^2)\mu_{k'}\right] \\
\langle x^4\rangle-\langle x^2\rangle^2 & = & \sum_{k,k'} \rho_k\rho_{k'}\left[\mu_k^4+6\mu_k^2\sigma_k^2+3\sigma_k^4-(\sigma_k^2+\mu_k^2)(\sigma_{k'}^2+\mu_{k'}^2)\right]
\end{array}
$$

A.2.3 导数

定义 $p(s)=\sum$ 对于第j个分量的参数 $\theta_j$of，我们得到

$$
\frac{\partial \ln p(s)}{\partial \theta_j}=\frac{\rho_j\mathcal{N}_s(\mu_j,\sigma_j^2)}{\sum_k \rho_k\mathcal{N}_s(\mu_k,\sigma_k^2)}\frac{\partial \ln(\rho_j\mathcal{N}_s(\mu_j,\sigma_j^2))}{\partial \theta_j} \tag{565}
$$

也就是说,

$$
\frac{\partial \ln p(s)}{\partial \rho_j} = \frac{\rho_j\mathcal{N}_s(\mu_j,\sigma_j^2)}{\sum_k \rho_k\mathcal{N}_s(\mu_k,\sigma_k^2)}\frac{1}{\rho_j} \tag{566}
$$

$$
\frac{\partial \ln p(s)}{\partial \mu_j} = \frac{\rho_j\mathcal{N}_s(\mu_j,\sigma_j^2)}{\sum_k \rho_k\mathcal{N}_s(\mu_k,\sigma_k^2)}\frac{(s-\mu_j)}{\sigma_j^2} \tag{567}
$$

$$
\frac{\partial \ln p(s)}{\partial \sigma_j} = \frac{\rho_j\mathcal{N}_s(\mu_j,\sigma_j^2)}{\sum_k \rho_k\mathcal{N}_s(\mu_k,\sigma_k^2)}\frac{1}{\sigma_j}\left[\frac{(s-\mu_j)^2}{\sigma_j^2}-1\right] \tag{568}
$$

注意 ρ_k必须受到约束，成为适当的比率。通过定义比率为 $\rho_j=e^{r_j}/\sum_k e^{r_k}$，我们得到

$$
\frac{\partial \ln p(s)}{\partial r_j}=\sum_l \frac{\partial \ln p(s)}{\partial \rho_l}\frac{\partial \rho_l}{\partial r_j} \quad \text{哪里} \quad \frac{\partial \rho_l}{\partial r_j}=\rho_l(\delta_{lj}-\rho_j) \tag{569}
$$

B 证明和细节

B.1 其他证明

B.1.1 方程式 524 的证明

以下证明是 Florian Roemer 的工作。请注意下面的向量和矩阵可以是复数，符号 $\mathbf{X}^H$用于转置和共轭，而 $\mathbf{X}^T$仅用于复数矩阵的转置。

定义行向量 $\mathbf{y} = \mathbf{a}^H\mathbf{X}\mathbf{B}$和列向量 $\mathbf{z} = \mathbf{X}^H\mathbf{c}$。 然后，$\mathbf{a}^T\mathbf{X}\mathbf{B}\mathbf{X}^T\mathbf{c} = \mathbf{y}\mathbf{z} = \mathbf{z}$

$^T\mathbf{y}^T$请注意， $\mathbf{y}$可以重写为 $\mathrm{vec}(\mathbf{y})^T$，这与vec(

$\mathrm{conj}(\mathbf{y}))^H = \mathrm{vec}(\mathbf{a}^T\mathrm{conj}(\mathbf{X})\mathrm{conj}(\mathbf{B}))^H$ 相同

其中"conj"表示复共轭。 应用线性形式的 vec 规则
方程式 520，我们得到

$$\mathbf{y} = (\mathbf{B}^H \otimes \mathbf{a}^T\mathrm{vec}(\mathrm{conj}(\mathbf{X}))^H = \mathrm{vec}(\mathbf{X})^T(\mathbf{B} \otimes \mathrm{conj}(\mathbf{a}))$$其中我们还使

用了 Kronecker 乘积的转置规则。对于 $\mathbf{y}^T$，这得到了 $(\mathbf{B}^T \otimes \mathbf{a}^H)\mathrm{vec}(\mathbf{X})$。类似地，我们可以重写 $\mathbf{z}$，这与$\mathrm{vec}(\mathbf{z}^T) = \mathrm{vec}(\mathbf{c}^T\mathrm{conj}(\mathbf{X}))$ 相同。再次应用方程式 520，我们得到$\mathbf{z} = (\mathbf{I} \otimes \mathbf{c}^T)\mathrm{vec}(\mathrm{conj}(\mathbf{X}))$

其中 $\mathbf{I}$是单位矩阵。 对于 $\mathbf{z}^T$，我们得到$\mathrm{vec}(\mathbf{X})(\mathbf{I} \otimes \mathbf{c})$。 最后，原始表达式为 $\mathbf{z}$ $^T\mathbf{y}^T$，现在采用以下形式$\mathrm{vec}(\mathbf{X})^H(\mathbf{I} \otimes \mathbf{c})(\mathbf{B}^T \otimes \mathbf{a}^H)\mathrm{vec}(\mathbf{X})$

最后一步是应用Kronecker积的乘积规则，并通过此方式组合Kronecker积。 这给出了$\mathrm{vec}(\mathbf{X})^H(\mathbf{B}^T \otimes \mathbf{c}\mathbf{a}^H)\mathrm{vec}(\mathbf{X})$这是所需的结果。

B.1.2 方程式 493 的证明

对于任何矩阵参数$\mathbf{X}$的解析函数 $\mathbf{f}(\mathbf{X})$，有

$$\begin{aligned}
\mathbf{f}(\mathbf{A}\mathbf{B})\mathbf{A} &= \left(\sum_{n=0}^{\infty} c_n(\mathbf{A}\mathbf{B})^n\right)\mathbf{A} \\
&= \sum_{n=0}^{\infty} c_n(\mathbf{A}\mathbf{B})^n\mathbf{A} \\
&= \sum_{n=0}^{\infty} c_n\mathbf{A}(\mathbf{B}\mathbf{A})^n \\
&= \mathbf{A}\sum_{n=0}^{\infty} c_n(\mathbf{B}\mathbf{A})^n \\
&= \mathbf{A}\mathbf{f}(\mathbf{B}\mathbf{A})
\end{aligned}$$

B.1.3 方程 91 的证明

基本上我们需要计算

$$
\begin{aligned}
\frac{\partial(\mathbf{X}^n)_{kl}}{\partial X_{ij}} &= \frac{\partial}{\partial X_{ij}} \sum_{u_1,\ldots,u_{n-1}} X_{k,u_1} X_{u_1,u_2} \ldots X_{u_{n-1},l} \\
&= \delta_{k,i}\delta_{u_1,j} X_{u_1,u_2} \ldots X_{u_{n-1},l} \\
&\quad + X_{k,u_1}\delta_{u_1,i}\delta_{u_2,j} \ldots X_{u_{n-1},l} \\
&\quad \vdots \\
&\quad + X_{k,u_1} X_{u_1,u_2} \ldots \delta_{u_{n-1},i}\delta_{l,j} \\
&= \sum_{r=0}^{n-1} (\mathbf{X}^r)_{ki} (\mathbf{X}^{n-1-r})_{jl} \\
&= \sum_{r=0}^{n-1} (\mathbf{X}^r \mathbf{J}^{ij} \mathbf{X}^{n-1-r})_{kl}
\end{aligned}
$$

利用第 9.7.4 节中找到的单个元素矩阵的性质，结果很容易得出。

B.1.4 方程式 571 的详细信息

$$
\begin{aligned}
\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X}) &= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\mathrm{Tr}[(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\partial(\mathbf{X}^H\mathbf{A}\mathbf{X})] \\
&= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\mathrm{Tr}[(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}(\partial(\mathbf{X}^H)\mathbf{A}\mathbf{X} + \mathbf{X}^H\partial(\mathbf{A}\mathbf{X}))] \\
&= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\Big(\mathrm{Tr}[(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\partial(\mathbf{X}^H)\mathbf{A}\mathbf{X}] \\
&\quad + \mathrm{Tr}[(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^H\partial(\mathbf{A}\mathbf{X})]\Big) \\
&= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\Big(\mathrm{Tr}[\mathbf{A}\mathbf{X}(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\partial(\mathbf{X}^H)] \\
&\quad + \mathrm{Tr}[(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^H\mathbf{A}\partial(\mathbf{X})]\Big)
\end{aligned}
$$

首先，找到对 $\mathbf{X}$的实部的导数

$$
\begin{aligned}
\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Re\mathbf{X}} &= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\Big(\frac{\mathrm{Tr}[\mathbf{A}\mathbf{X}(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\partial(\mathbf{X}^H)]}{\partial \Re\mathbf{X}} \\
&\quad + \frac{\mathrm{Tr}[(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^H\mathbf{A}\partial(\mathbf{X})]}{\partial \Re\mathbf{X}}\Big) \\
&= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\Big(\mathbf{A}\mathbf{X}(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1} + ((\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^H\mathbf{A})^T\Big)
\end{aligned}
$$

通过计算，使用了(100)和(240)。此外，通过使用(241)，找到了对 $\mathbf{X}$的虚部的导数即 $\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})$

$$
\begin{aligned}
\frac{}{\partial \Im\mathbf{X}} &= i\det(\mathbf{X}^H\mathbf{A}\mathbf{X})\Big(\frac{\mathrm{Tr}[\mathbf{A}\mathbf{X}(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\partial(\mathbf{X}^H)]}{\partial \Im\mathbf{X}} \\
&\quad + \frac{\mathrm{Tr}[(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^H\mathbf{A}\partial(\mathbf{X})]}{\partial \Im\mathbf{X}}\Big) \\
&= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\Big(\mathbf{A}\mathbf{X}(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1} - ((\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^H\mathbf{A})^T\Big)
\end{aligned}
$$

因此，导数为

$$
\begin{aligned}
\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} &= \frac{1}{2}\Big(\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Re\mathbf{X}} - i\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Im\mathbf{X}}\Big) \\
&= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\Big((\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^H\mathbf{A}\Big)^T
\end{aligned}
$$

而复共轭导数为

$$\begin{aligned}\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \mathbf{X}^*} &= \frac{1}{2}\Big(\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Re \mathbf{X}} + i\,\frac{\partial \det(\mathbf{X}^H\mathbf{A}\mathbf{X})}{\partial \Im \mathbf{X}}\Big)\\ &= \det(\mathbf{X}^H\mathbf{A}\mathbf{X})\mathbf{A}\mathbf{X}(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1}\end{aligned}$$

注意，对于实数 $\mathbf{X}$, $\mathbf{A}$，(249)和(250)的和减少到(54)。
类似的计算产生

$$\begin{aligned}\frac{\partial \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)}{\partial \mathbf{X}} &= \frac{1}{2}\Big(\frac{\partial \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)}{\partial \Re \mathbf{X}} - i\,\frac{\partial \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)}{\partial \Im \mathbf{X}}\Big)\\ &= \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)\left(\mathbf{A}\mathbf{X}^H(\mathbf{X}\mathbf{A}\mathbf{X}^H)^{-1}\right)^T \qquad (570)\end{aligned}$$

和

$$\begin{aligned}\frac{\partial \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)}{\partial \mathbf{X}^*} &= \frac{1}{2}\Big(\frac{\partial \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)}{\partial \Re \mathbf{X}} + i\,\frac{\partial \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)}{\partial \Im \mathbf{X}}\Big)\\ &= \det(\mathbf{X}\mathbf{A}\mathbf{X}^H)(\mathbf{X}\mathbf{A}\mathbf{X}^H)^{-1}\mathbf{X}\mathbf{A} \qquad (571)\end{aligned}$$

参考文献

[1] Karl Gustav Andersson and Lars-Christer Boiers. *Ordinaera differentialekvationer*. Studenterlitteratur, 1992.

[2] Jörn Anemüller, Terrence J. Sejnowski, and Scott Makeig. Complex independent component analysis of frequency-domain electroencephalographic data. *Neural Networks*, 16(9):1311–1323, November 2003.

[3] S. Barnet. 矩阵. 方法和应用. 牛津应用数学与计算科学系列. 克拉伦登出版社, 1990.

[4] Christopher Bishop. 神经网络模式识别. 牛津大学出版社, 1995.

[5] Robert J. Boik. 讲义: 统计学550. 在线, 2002年4月22日. 讲义.

[6] D. H. Brandwood. 复杂梯度算子及其在自适应阵列理论中的应用.*IEE Proceedings*, 130(1):11–16, 1983年2月. *P*TS. F和H.

[7] M. Brookes. 矩阵参考手册, 2004年. 网站于2004年5月20日.

[8] Contradsen K.,统计学导论, IMM讲义, 1984年.

[9] Mads Dyrholm. 一些矩阵结果, 2004年. 网站于2004年8月23日.

[10] Nielsen F. A.,公式, 神经研究单位和丹麦技术大学, 2002年.

[11] Gelman A. B., J. S. Carlin, H. S. Stern, D. B. Rubin, 贝叶斯数据分析, Chapman and Hall / CRC, 1995年.

[12] Gene H. Golub and Charles F. van Loan.矩阵计算. 巴尔的摩约翰斯霍普金斯大学出版社, 第三版, 1996年.

[13] Robert M. Gray. Toeplitz和循环矩阵: 一份综述. 技术报告, 信息系统实验室, 电气工程系, 斯坦福大学, 加利福尼亚州斯坦福, 94305, 2002年8月.

[14] Simon Haykin. 自适应滤波器理论 . 普林斯顿大学出版社, 上塞德尔河, 新泽西州, 第四版, 2002年.

[15] 罗杰・霍恩和查尔斯・约翰逊。矩阵分析。 剑桥大学出版社，1985年。

[16] Mardia K. V.，J.T. Kent和J.M. Bibby，多元分析，学术出版社有限公司，1979年。

[17] Mathpages关于“特征值问题和矩阵不变量”的内容，

http://www.mathpages.com/home/kmath128.htm

[18] Carl D. Meyer. 修改矩阵的广义逆。*SIAM*应用数学杂志，24(3):315–323，1973年5月。

[19] Thomas P. Minka. 对统计学有用的旧和新的矩阵代数，2000年12月。笔记。

[20] Daniele Mortari正交-斜对称和正交-对称矩阵三角学，*John Lee Junkins*天体动力学研讨会，*AAS* 03–265，2003年5月。 德克萨斯州A&M大学，College Station，TX

[21] L. Parra 和 C. Spence. 非平稳源的卷积盲分离。 在*IEEE Transactions Speech and Audio Processing*, 页码 320–327, 2000年5月。

[22] Kaare Brandt Petersen, Jiucang Hao 和 Te-Won Lee. 用于过完备表示的生成和过滤方法。 神经信息处理 - 信件和评论, 卷 8(1), 2005年。

[23] John G. Proakis 和 Dimitris G. Manolakis. 数字信号处理。 Prentice-Hall, 1996年。

[24] Laurent Schwartz. *Cours d'Analyse*, 第二卷。 Hermann, Paris, 1967年。如[14]中引用。

[25] Shayle R. Searle. 矩阵代数在统计学中的应用。 John Wiley and Sons, 1982年。

[26] G. Seber 和 A. Lee. 线性回归分析. John Wiley and Sons, 2002.

[27] S. M. Selby.标准数学表. CRC Press, 1974.

[28] Inna Stainvas. 微分计算中的矩阵代数. 神经计算研究小组, 信息工程, 阿斯顿大学, 英国, 八月 2002. 笔记.

[29] P. P. Vaidyanathan. 多速率系统和滤波器组. Prentice Hall, 1993.

[30] Max Welling. 卡尔曼滤波器. 讲义.

[31] 维基百科关于子式: "子式 (线性代数)",

`http://en.wikipedia.org/wiki/Minor_(linear_algebra)`

[32] Zhaoshui He, Shengli Xie, 等, "基于稀疏表示的频域卷积盲源分离", IEEE Transactions onAudio, Speech and Language Processing, vol.15(5):1551-1563, 2007年7月.

[33] Karim T. Abou-Moustafa关于广义特征值问题的特征值和特征向量的导数. 麦吉尔技术报告，2010年10月.

[34] Mohammad Emtiyaz Khan当添加/删除一列时，更新矩阵的逆. E*mt C*S, *U* B*C* 2008年2月27日

索引

反对称，54

分块矩阵，46

链式法则，15
乔列斯基分解，32
共偏度，34
共偏斜度，34
条件数，62
克莱默法则，29

复矩阵的导数，24
行列式的导数，8
迹的导数，12
逆的导数，9
对称矩阵的导数，15
Toeplitz矩阵的导数，16
狄利克雷分布，37

特征值，30
特征向量，30
指数矩阵函数，59

高斯，条件，40
高斯，熵，44
高斯，线性组合，41
高斯，边际，40
高斯，密度的乘积，42
广义逆，21

哈达玛不等式，52
共轭，48

幂等，49

克罗内克积，59

LDL分解，33
LDM分解，33
线性回归，28
LU分解，32
李雅普诺夫方程，30

摩尔-彭罗斯逆，21
多项式分布，37

幂零，49
矩阵的范数，61
向量的范数，61

正态逆伽玛分布，37
正态逆威夏特分布，39

正交，49

矩阵的幂级数, 58
概率矩阵, 55
伪逆, 21

舒尔补, 41, 47
单元素矩阵, 52
奇异值分解 (SVD), 31
斜厄米矩阵, 48
斜对称矩阵, 54
随机矩阵, 55
学生-t分布, 37
Sylvester不等式, 62
对称矩阵, 54

泰勒展开, 58
Toeplitz矩阵, 54
过渡矩阵, 55
三角函数, 59

幂等矩阵, 49

范德蒙矩阵, 57
向量算子, 59, 60

Wishart分布, 38
Woodbury恒等式, 18