矩阵分析及其应用

中国人民大学 管华

2017-08-31 更新

目录

第一章 矩阵基础知识 7

1.1 线性空间与线性映射 7

1.1.1 线性空间与线性子空间 7

1.1.2 Euclid 空间、酉空间 8

1.1.3 线性映射及其矩阵表示 10

1.1.4 几个重要的线性子空间及其性质 10

1.2 矩阵的数值特征 10

1.2.1 秩 10

1.2.2 行列式 10

1.2.3 迹 11

1.2.4 特征值、特征向量和特征多项式 11

1.3 矩阵的标准形 11

1.3.1 等价变换下的标准形 11

1.3.2 相似变换下的 Jordan 标准形 12

1.3.3 相合变换下的标准形 13

1.4 半正定和正定矩阵 13

1.5 矩阵求逆公式 13

1.5.1 Leverrier-Faddeev 算法 13

1.5.2 分块求逆公式 13

1.5.3 Sherman-Motrison-Woodbury 公式 13

1.6 Hadamard 与 Kronecker 积 13

1.6.1 Hadamard 积及其性质 13

1.6.2 Kronecker 积及其性质 13

第二章 向量范数和矩阵范数 **15**

2.1 向量范数 15

2.1.1 向量范数的定义 15

2.1.2 常用向量范数 15

2.1.3 向量范数的分析性质 15

2.1.4 向量范数的代数性质 16

2.2 矩阵范数 16

2.2.1 矩阵范数的定义及分析性质 16

2.2.2 常用的矩阵范数 16

2.2.3 由向量范数诱导的矩阵范数 16

2.3 一些应用 17

2.3.1 谱半径与矩阵范数 17

2.3.2 矩阵逆与线性方程组解的扰动问题 17

2.3.3 条件数 17

第三章 矩阵函数和矩阵微积分 **19**

3.1 矩阵序列和矩阵级数 19

3.1.1 矩阵序列 19

3.1.2 矩阵级数 19

3.1.3 矩阵幂级数 20

3.2 矩阵函数 20

3.2.1 矩阵函数的定义与性质 20

3.2.2 矩阵函数值的计算 21

3.3 矩阵的微分和积分 22

3.3.1 以一元函数为元素的矩阵的微积分 22

3.3.2 函数对向量的微分 22

3.3.3 函数对矩阵的微分 24

3.3.4 矩阵对矩阵的微分 25

3.4 一些应用 25

3.4.1 特征多项式系数的表示 25

3.4.2 线性常系数微分方程组的求解 25

3.4.3 矩阵最优低秩逼近 25

第四章 矩阵分解 **25**
4.1 满秩分解 25
4.2 三角分解 25
4.2.1 LU 分解 25
4.2.2 LDU 分解 26
4.2.3 LU 分解的算法 26
4.2.4 Cholesky 分解 26
4.3 QR 分解 26
4.3.1 QR 分解 26
4.3.2 Gram-Schmidt 算法及其修正 26
4.3.3 Householder 变换法 26
4.3.4 Givens 旋转法 27
4.4 奇异值分解 27
4.4.1 定义及性质 27
4.4.2 极分解 27
4.5 矩阵的同时对角化 27
4.5.1 Hermite 矩阵和正规矩阵同时对角化 27
4.5.2 广义奇异值分解 27
4.6 一些应用 27
4.6.1 随机向量的模拟 27
4.6.2 基于 QR 分解的最小二乘算法 27
4.6.3 矩阵的最优逼近 27

第五章 特征值分析 **29**
5.1 特征值的连续性 29
5.2 特征值的估计 29
5.2.1 特征值的界 29
5.2.2 特征值所在的区域 29
5.3 Hermite 矩阵的特征值及其极性 29
5.3.1 Rayleigh 商 29
5.3.2 广义 Rayleigh 商 29
5.3.3 特征值的分隔 30
5.3.4 Hermite 扰动下的特征值 30
5.4 一些应用 30

5.4.1 与对角矩阵相似的矩阵特征值的扰动 30
5.4.2 主成分分析 30
5.4.3 概率分布的 Wasserstein 距离 30

第六章 广义逆矩阵 31

6.1 投影矩阵 31
6.2 广义逆矩阵及其性质 31
6.2.1 广义逆的定义 31
6.2.2 广义逆的性质 32
6.2.3 广义逆的等价形式 32
6.2.4 广义逆的反序法则 32
6.2.5 广义逆矩阵的连续性问题 32
6.3 广义逆的计算方法 32
6.3.1 单个矩阵的广义逆 32
6.3.2 更新矩阵的广义逆 32
6.3.3 分块算法 32
6.4 一些应用 32
6.4.1 矩阵方程、线性方程组的解与广义逆 32
6.4.2 精确初始化的最小二乘递推算法 32

第三章 矩阵函数和矩阵微积分

3.1 矩阵序列和矩阵级数

3.1.1 矩阵序列

引理 3.1.1

$\boldsymbol{A}$is$n \times n$, 若存在一种矩阵范数 $\|\cdot\|$ 使得 $\|\boldsymbol{A}\| < 1$, 则 $\boldsymbol{A}^k \to \boldsymbol{O}$。

定理 3.1.1

设 $\boldsymbol{A}$is$n \times n$，则 $\boldsymbol{A}^k \to \boldsymbol{O} \Longleftrightarrow \rho(\boldsymbol{A}) < 1$

定义 3.1.1 (收敛矩阵)

$\boldsymbol{A}$is$n \times n$

$$\boldsymbol{A}^k \to \boldsymbol{O}$$

定义 3.1.2 (界)

$$\left|a_{ij}^{(k)}\right| < C$$

并称 C 为 $\{\boldsymbol{A}^{(k)}\}$ 的界。

推论 3.1.1

$$\left\|\boldsymbol{A}^k\right\|^{\frac{1}{k}} \to \rho\boldsymbol{A}$$

3.1.2 矩阵级数

定义 3.1.3 (矩阵级数)

$$\boldsymbol{A}^{(0)} + \boldsymbol{A}^{(1)} + \cdots \boldsymbol{A}^{(k)} + \cdots = \sum \boldsymbol{A}^{(k)}$$

定义 3.1.4

矩阵级数是发散的

绝对收敛

3.1.3 矩阵幂级数

定理 3.1.2 (幂级数)

$$\sum \boldsymbol{A}^k = \boldsymbol{I} + \boldsymbol{A} + \boldsymbol{A}^2 + \cdots + \boldsymbol{A}^k + \cdots = (\boldsymbol{I} - \boldsymbol{A})^{-1}$$

收敛的充要条件是 $\boldsymbol{A}$ 为收敛矩阵，且在收敛是，其和为 $(\boldsymbol{I} - \boldsymbol{A})^{-1}$.

定理 3.1.3

$f(z) = \sum c_k z^k$ 的收敛半径为 r. 若 $\boldsymbol{A} \text{is} n \times n$ 满足 $\rho(\boldsymbol{A}) < r$, 则矩阵幂级数 $\sum c_k \boldsymbol{A}^k$ **绝对收敛**；若 $rho(\boldsymbol{A}) > r$, 则矩阵幂级数发散。

3.2 矩阵函数

3.2.1 矩阵函数的定义与性质

定义 3.2.1

当矩阵 $\boldsymbol{A} \text{is} n \times n$ 的谱半径 $\rho(\boldsymbol{A}) < r$ 时，矩阵幂级数 $\sum c_k \boldsymbol{A}^k$ 收敛，称其和为 **矩阵函数**，记为

$$f(\boldsymbol{A}) = \sum_{0}^{\infty} c_k \boldsymbol{A}^k$$

例 3.2.1

$$\begin{cases} f(z) = \dfrac{1}{1-z} = \sum z^k, & |z| < 1 \\ f(\boldsymbol{A}) = \sum \boldsymbol{A}^k = (\boldsymbol{I} - \boldsymbol{A})^{-1}, & \rho(\boldsymbol{A}) < 1 \end{cases}$$

■

定理 3.2.1

若 $\boldsymbol{A}, \boldsymbol{B}$ 可交换，即 $\boldsymbol{AB} = \boldsymbol{BA}$, 则

$$e^{\boldsymbol{A}} e^{\boldsymbol{B}} = e^{\boldsymbol{B}} e^{\boldsymbol{A}} = e^{\boldsymbol{A}+\boldsymbol{B}}$$

同时，

$$\begin{cases} e^{\boldsymbol{A}}e^{-\boldsymbol{A}} = e^{-\boldsymbol{A}}e^{\boldsymbol{A}} = \boldsymbol{I} \\ \left(e^{\boldsymbol{A}}\right)^{-1} = e^{-\boldsymbol{A}} \\ \left(e^{\boldsymbol{A}}\right)^{k} = e^{k\boldsymbol{A}} \end{cases}$$

3.2.2　矩阵函数值的计算

待定系数法

数项级数求和法

对角形法

若 $\boldsymbol{A}$is$n \times n$ 相似于对角矩阵 Λ, 即存在非奇异矩阵 $\boldsymbol{P}$, 使得

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$$

则有

$$f(\boldsymbol{A}) = \boldsymbol{P} \cdot \mathrm{diag}\left(f(\lambda_1), \cdots, f(\lambda_n)\right) \cdot \boldsymbol{P}^{-1}$$

当 $\boldsymbol{A}$ 相似于对角矩阵是，矩阵幂级数的求和问题可以转化为求变换矩阵的问题。

Jordan 标准形法

定义 3.2.2
设 $\boldsymbol{A}$ 的 Jordan 标准形为 $\boldsymbol{J}$, 则存在可逆矩阵 $\boldsymbol{P}$ 使得

$$\begin{cases} \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J} = \mathrm{diag}(\boldsymbol{J}_1, \cdots, \boldsymbol{J}_s) \\ f(\boldsymbol{A}) = \boldsymbol{P} \cdot \mathrm{diag}(f(\boldsymbol{J}_1), \cdots, f(\boldsymbol{J}_s)) \cdot \boldsymbol{P}^{-1} \end{cases}$$

去掉了收敛矩阵的限制。

定理 3.2.2
对于 $f(\boldsymbol{A})$ 与矩阵的 Jordan 标准形 $\boldsymbol{J}$ 中 Jordan 块的排列顺序无关，与变换矩阵 $\boldsymbol{P}$ 的选取无关。函数可相加，可相乘。

$$\begin{cases} f(z) = f_1(z) + f_2(z) \Longrightarrow f(\boldsymbol{A}) = f_1(\boldsymbol{A}) + f_2(\boldsymbol{A}) \\ f(z) = f_1(z)f_2(z) \Longrightarrow f(\boldsymbol{A}) = f_1(\boldsymbol{A})f_2(\boldsymbol{A}) \end{cases}$$

3.3 矩阵的微分和积分

3.3.1 以一元函数为元素的矩阵的微积分

$$\boldsymbol{A}(t)=\begin{bmatrix} a_{11}(t) & \cdots & a_{1n}(t) \\ \vdots & & \vdots \\ a_{m1}(t) & \cdots & a_{mn}(t) \end{bmatrix}$$

3.3.2 函数对向量的微分

定义 3.3.1

设 $f(\boldsymbol{x})$ 为纯量函数，其中 $\boldsymbol{x}=[x_1,\cdots,x_n]^T\in\mathbb{C}^n$，则

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}=\begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}_1} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}_n} \end{bmatrix}$$

例 3.3.1 $\boldsymbol{A}\text{is}n\times n,\ \boldsymbol{x}=\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

$$f(\boldsymbol{x})=\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}=\sum\sum a_{ij}x_iy_j$$

对 $\forall k=1,\cdots,n$, 有

$$\frac{\partial f(\boldsymbol{x})}{\partial x_k}=\frac{\partial}{\partial x_k}\left(\sum\sum a_{ij}x_iy_j\right)=\sum a_{ik}x_i+\sum a_{kj}x_j$$

所以

$$\frac{\partial \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{A}^T\boldsymbol{x}$$

若 $\boldsymbol{A}$ 为对称矩阵，则

$$\frac{\partial \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}}=2\boldsymbol{A}\boldsymbol{x}$$

■

定义 3.3.2

$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_{[n\times 1]}, f(\boldsymbol{x}) = \big[f_1(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x})\big]_{1\times m}$

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\boldsymbol{x})}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial f_1(\boldsymbol{x})}{\partial x_n} & \cdots & \frac{\partial f_m(\boldsymbol{x})}{\partial x_n} \end{bmatrix}_{n\times m}$$

定义 3.3.3 (Jacobi 矩阵)

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_{[n\times 1]}, f(\boldsymbol{x}) = \begin{bmatrix} f_1(\boldsymbol{x}) \\ \vdots \\ f_m(\boldsymbol{x}) \end{bmatrix}$$

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^T} = \begin{bmatrix} \frac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\boldsymbol{x})}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\boldsymbol{x})}{\partial x_n} \end{bmatrix}_{m\times n}$$

定理 3.3.1 (链式法则)

$$\frac{\partial f}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{y}^T}{\partial \boldsymbol{x}} \frac{\partial f}{\partial \boldsymbol{y}} \Longleftrightarrow \frac{\partial f(\boldsymbol{y}(\boldsymbol{x}))}{\partial \boldsymbol{x}} = \frac{\partial \left(\boldsymbol{y}(\boldsymbol{x})\right)^T}{\partial \boldsymbol{x}} \frac{\partial f(\boldsymbol{y})}{\partial \boldsymbol{y}}$$

定义 3.3.4

$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ 纯量函数关于向量的二阶微分是由 n^2 个二阶偏导组成的 $n \times n$ 阶矩阵，称为 Hessian 矩阵。

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x} \partial \boldsymbol{x}^T} = \frac{\partial}{\partial \boldsymbol{x}^T}(\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}})$$

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x} \partial \boldsymbol{x}^T} = \begin{bmatrix} \frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x}_1 \partial \boldsymbol{x}_1} & cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x}_1 \partial \boldsymbol{x}_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x}_n \partial \boldsymbol{x}_1} & cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x}_n \partial \boldsymbol{x}_n} \end{bmatrix}$$

特别的，若 $f(\boldsymbol{x})=\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$，则

$$\frac{\partial^2(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x})}{\partial\boldsymbol{x}\partial\boldsymbol{x}^T}=\boldsymbol{A}+\ MA^T$$

定义 3.3.5 (向量值函数 $\boldsymbol{f}(\boldsymbol{x})$ 对向量 $\boldsymbol{x}$ 的微分)
$\boldsymbol{x}=[x_1,\cdots,x_n]^T$，$f(\boldsymbol{x})=\left[f_1(\boldsymbol{x}),\cdots,f_m(\boldsymbol{x})\right]^T$

$$\frac{\partial f(\boldsymbol{x})}{\partial\boldsymbol{x}}=\begin{bmatrix}\frac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\boldsymbol{x})}{\partial x_1}\\ \vdots & & \vdots\\ \frac{\partial f_1(\boldsymbol{x})}{\partial x_n} & \cdots & \frac{\partial f_m(\boldsymbol{x})}{\partial x_n}\end{bmatrix}$$

定义 3.3.6 (Jacobi 矩阵)
$\boldsymbol{x}=[x_1,\cdots,x_n]^T$，$f(\boldsymbol{x})=\left[f_1(\boldsymbol{x}),\cdots,f_m(\boldsymbol{x})\right]^T$

$$\frac{\partial f(\boldsymbol{x})}{\partial\boldsymbol{x}^T}=\begin{bmatrix}\frac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\boldsymbol{x})}{\partial x_n}\\ \vdots & & \vdots\\ \frac{\partial f_m(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\boldsymbol{x})}{\partial x_n}\end{bmatrix}$$

定义 3.3.7 (Hessian 矩阵)

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial\boldsymbol{x}\partial\boldsymbol{x}^T}=\frac{\partial}{\partial\boldsymbol{x}^T}\left(\frac{\partial f(\boldsymbol{x})}{\partial\boldsymbol{x}}\right)=\begin{bmatrix}\frac{\partial^2 f(\boldsymbol{x})}{\partial x_1\partial x_1} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1\partial x_n}\\ \vdots & & \vdots\\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n\partial x_1} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n\partial x_n}\end{bmatrix}$$

若 $f(\boldsymbol{x})=\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$，则

$$\frac{\partial^2(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x})}{\partial\boldsymbol{x}\partial\boldsymbol{x}^T}=\boldsymbol{A}+\boldsymbol{A}^T$$

3.3.3 函数对矩阵的微分

定义与性质

定义 3.3.8 (纯量函数 $\boldsymbol{f}(A)$ 对矩阵 A 的微分定义)

$$\frac{\partial f(\boldsymbol{A})}{\partial\boldsymbol{A}}=\begin{bmatrix}\frac{\partial f(\boldsymbol{A})}{\partial a_{11}} & \cdots & \frac{\partial f(\boldsymbol{A})}{\partial a_{1n}}\\ \vdots & \ddots & \vdots\\ \frac{\partial f(\boldsymbol{A})}{\partial a_{m1}} & \cdots & \frac{\partial f(\boldsymbol{A})}{\partial a_{mn}}\end{bmatrix}$$

迹的梯度矩阵

行列式的梯度矩阵

3.3.4　矩阵对矩阵的微分

定义 3.3.9 (矩阵 $\boldsymbol{F}(\boldsymbol{X})$ 对 $\boldsymbol{X}$ 的微分)
$\boldsymbol{X} \in \mathbb{C}^{m\times n}$，$f_{ij}(\boldsymbol{X})$ 为 mn 元纯量函数 $(i\colon 1 \to p,\ j\colon 1 \to q)$，记矩阵函数 $\boldsymbol{F}(\boldsymbol{X}) = \big(f_{ij}(\boldsymbol{X})\big)$

$$\frac{\partial \boldsymbol{F}(\boldsymbol{X})}{\partial \boldsymbol{X}} = \left[\operatorname{vec}\left(\frac{\partial f_{11}}{\partial \boldsymbol{X}}\right), \operatorname{vec}\left(\frac{\partial f_{12}}{\partial \boldsymbol{X}}\right), \ldots, \operatorname{vec}\left(\frac{\partial f_{pq}}{\partial \boldsymbol{X}}\right)\right]$$

3.4　一些应用

3.4.1　特征多项式系数的表示

3.4.2　线性常系数微分方程组的求解

3.4.3　矩阵最优低秩逼近