# 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	矩阵函数和矩阵微积分	5
1.1	矩阵序列和矩阵级数	5
	1.1.1 矩阵序列	5
	1.1.2 矩阵级数	5
	1.1.3 矩阵幂级数	6
1.2	矩阵函数	6
	1.2.1 矩阵函数的定义与性质	6
	1.2.2 矩阵函数值的计算	7
1.3	矩阵的微分和积分	8
	1.3.1 以一元函数为元素的矩阵的微积分	8
	1.3.2 函数对向量的微分	8
	1.3.3 函数对矩阵的微分	10
	1.3.4 矩阵对矩阵的微分	11
1.4	一些应用 1	12
	1.4.1 特征多项式系数的表示	12
	1.4.2 线性常系数微分方程组的求解	12
	1.4.3 矩阵最优低秩逼近	12

4 目录

# 第一章 矩阵函数和矩阵微积分

# 1.1 矩阵序列和矩阵级数

### 1.1.1 矩阵序列

引理 1.1.1

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

定理 1.1.1

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

定义 1.1.1 (收敛矩阵)

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

 $m{A}^k o m{O}$ 

定义 1.1.2 (界)

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

并称 C 为  $\{A^{(k)}\}$  的界。

推论 1.1.1

$$\left\|oldsymbol{A}^k
ight\|^{rac{1}{k}}
ightarrow
hooldsymbol{A}$$

1.1.2 矩阵级数

定义 1.1.3 (矩阵级数)

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

#### 定义 1.1.4

矩阵级数是发散的

绝对收敛

#### 1.1.3 矩阵幂级数

定理 1.1.2 (幂级数)

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

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

#### 定理 1.1.3

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

### 1.2 矩阵函数

### 1.2.1 矩阵函数的定义与性质

#### 定义 1.2.1

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

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

#### 例 1.2.1

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

#### 定理 1.2.1

若 A, B 可交换, 即 AB = BA,则

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

1.2 矩阵函数 7

同时,

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

#### 1.2.2 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 A is  $n \times n$  相似于对角矩阵 Λ, 即存在非奇异矩阵 P, 使得

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

则有

$$f(\mathbf{A}) = \mathbf{P} \cdot \operatorname{diag} (f(\lambda_1), \cdots, f(\lambda_n)) \cdot \mathbf{P}^{-1}$$

当 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 1.2.2

设 A 的 Jordan 标准形为 J, 则存在可逆矩阵 P 使得

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

去掉了收敛矩阵的限制。

#### 定理 1.2.2

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

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

## 1.3 矩阵的微分和积分

#### 1.3.1 以一元函数为元素的矩阵的微积分

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

#### 1.3.2 函数对向量的微分

#### 定义 1.3.1

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

$$rac{\partial f(oldsymbol{x})}{\partial oldsymbol{x}} = egin{bmatrix} rac{\partial f(oldsymbol{x})}{\partial oldsymbol{x}_1} \ dots \ rac{\partial f(oldsymbol{x})}{\partial oldsymbol{x}_n} \end{bmatrix}$$

例 1.3.1 
$$\boldsymbol{A}$$
 is  $n \times n$ ,  $\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ 

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \sum \sum a_{ij} x_i y_j$$

 $対 \forall k = 1, \cdots, n, 有$ 

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

所以

$$rac{\partial oldsymbol{x}^T oldsymbol{A} oldsymbol{x}}{\partial oldsymbol{x}} = oldsymbol{A} oldsymbol{x} + oldsymbol{A}^T oldsymbol{x}$$

若 A 为对称矩阵,则

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

#### 1.3 矩阵的微分和积分

定义 
$$\begin{bmatrix} \mathbf{x} \cdot \mathbf{x} \\ \mathbf{x} \end{bmatrix} \mathbf{2}$$
  $\mathbf{x} = \begin{bmatrix} f_1(\mathbf{x}), \dots, f_m(\mathbf{x}) \end{bmatrix}_{1 \times m}$ 

$$rac{\partial f(m{x})}{\partial m{x}} = egin{bmatrix} rac{\partial f_1(m{x})}{\partial x_1} & \cdots & rac{\partial f_m(m{x})}{\partial x_1} \ dots & dots \ rac{\partial f_1(m{x})}{\partial x_n} & \cdots & rac{\partial f_m(m{x})}{\partial x_n} \end{bmatrix}_{n imes m}$$

9

#### 定义 1.3.3 (Jacobi 矩阵)

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_{[n \times 1]}, f(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix}$$
$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^T} = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m(\mathbf{x})}{\partial x_n} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_n} \end{bmatrix}$$

#### 定理 1.3.1 (链式法则)

$$rac{\partial f}{\partial oldsymbol{x}} = rac{\partial oldsymbol{y}^T}{\partial oldsymbol{x}} rac{\partial f}{\partial oldsymbol{y}} \Longleftrightarrow rac{\partial f(oldsymbol{y}(oldsymbol{x}))}{\partial oldsymbol{x}} = rac{\partial \left(oldsymbol{y}(oldsymbol{x})
ight)^T}{\partial oldsymbol{x}} rac{\partial f(oldsymbol{y})}{\partial oldsymbol{y}}$$

定义 [1x3] 4 (Hessian 矩阵)  $x = \begin{bmatrix} \vdots \\ x \end{bmatrix}$  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。

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

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

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

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

#### 1.3.3 函数对矩阵的微分

定义与性质

定义 1.3.5 (纯量函数 f(A) 对矩阵 A 的微分定义)

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

#### 定理 1.3.2

$$\frac{\partial g}{\partial \mathbf{A}} = \frac{\partial g}{\partial y} \frac{\partial y}{\partial \mathbf{A}} \Longleftrightarrow \frac{\partial g(f(\mathbf{A}))}{\partial \mathbf{A}} = \frac{\partial g(y)}{\partial y} \frac{\partial f(\mathbf{A})}{\partial \mathbf{A}}$$

记  $E_{ij} = e_i e_i^T$ , 即第 i 行第 j 列元素为 1, 其余元素为零的矩阵。

$$\frac{\partial \left(\boldsymbol{a}^{T}\boldsymbol{X}^{-1}\boldsymbol{b}\right)}{\partial \boldsymbol{X}} = -\left(\boldsymbol{A}^{-1}\right)^{T}\boldsymbol{a}\boldsymbol{b}^{T}\left(\boldsymbol{X}^{-1}\right)^{T}$$

迹的梯度矩阵

定理 1.3.3

$$\frac{\partial \text{Tr} \boldsymbol{X}}{\partial \boldsymbol{X}} = \boldsymbol{I}$$
$$\frac{\partial \text{Tr} (\boldsymbol{X}^{-1})}{\partial \boldsymbol{X}} = -(\boldsymbol{X}^{-2})^T$$

$$\operatorname{Tr}(\boldsymbol{A}\boldsymbol{X}) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_{ij}$$

11

#### 定理 1.3.4

$$\begin{split} \frac{\partial \text{Tr}(\boldsymbol{A}\boldsymbol{X})}{\partial \boldsymbol{X}} &= \boldsymbol{A}^T \\ \frac{\partial \text{Tr}(\boldsymbol{A}\boldsymbol{X})}{\partial \boldsymbol{X}} &= \boldsymbol{A} + \boldsymbol{A}^T - \text{diag}\boldsymbol{A} \\ \frac{\partial \text{Tr}(\boldsymbol{A}\boldsymbol{X}^{-1})}{\partial \boldsymbol{X}} &= -\left(\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X}^{-1}\right)^T \end{split}$$

#### 定理 1.3.5

$$\begin{split} \frac{\partial \text{Tr}(\boldsymbol{X}^{T}\boldsymbol{A}\boldsymbol{X})}{\partial \boldsymbol{X}} &= (\boldsymbol{A} + \boldsymbol{A}^{T})\boldsymbol{X} \\ \frac{\partial \boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{T}}{\partial \boldsymbol{X}} &= \boldsymbol{X}(\boldsymbol{A} + \boldsymbol{A}^{T}) \\ \frac{\partial \text{Tr}(\boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B})}{\partial \boldsymbol{X}} &= -\left(\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{A}\boldsymbol{X}^{-1}\right)^{T} \end{split}$$

#### 行列式的梯度矩阵

定理 1.3.6

若 X 中的元素相互独立,则

$$\frac{\partial \det \boldsymbol{X}}{\partial \boldsymbol{X}} = \det \boldsymbol{X} \cdot (\boldsymbol{X}^{-1})^T$$

若 X 为对称矩阵,则

$$\frac{\partial \det \boldsymbol{X}}{\partial \boldsymbol{X}} = \det \boldsymbol{X} \cdot (2\boldsymbol{X}^{-1} - \operatorname{diag}(\boldsymbol{X}^{-1}))$$

#### 1.3.4 矩阵对矩阵的微分

定义 1.3.6 (矩阵 F(X) 对 X 的微分)

 $X \in \mathbb{C}^{m \times n}$ ,  $f_{ij}(X)$  为 mn 元纯量函数  $(i: 1 \rightarrow p, j: 1 \rightarrow q)$ , 记矩阵函数  $F(X) = (f_{ij}(X))$ 

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

# 1.4 一些应用

- 1.4.1 特征多项式系数的表示
- 1.4.2 线性常系数微分方程组的求解
- 1.4.3 矩阵最优低秩逼近