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

1.1　矩阵序列和矩阵级数

1.1.1　矩阵序列

引理 1.1.1

$\boldsymbol{A}$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 1.1.1

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

定义 1.1.1 (收敛矩阵)

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

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

定义 1.1.2 (界)

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

并称 C 为 $\{\boldsymbol{A}^{(k)}\}$ 的界。

推论 1.1.1

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

1.1.2　矩阵级数

定义 1.1.3 (矩阵级数)

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

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)

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

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 1.1.3

$f(z) = \sum c_k z^k$ 的收敛半径为 r. 若 $\boldsymbol{A}$ 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。

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 $\boldsymbol{A}$ 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 1.2.1

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

■

定理 1.2.1

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,

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

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 $\boldsymbol{A}$ 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 1.2.2
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 1.2.2
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。

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

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

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

1.3.2 函数对向量的微分

定义 1.3.1

设 $f(\boldsymbol{x})$ 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 1.3.1 $\boldsymbol{A}$ 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 1.3.2

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

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

定义 1.3.3 (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 1.3.1 (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 1.3.4 (Hessian 矩阵)

$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ 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

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

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)

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

定理 1.3.2

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

记 $\boldsymbol{E}_{ij}=\boldsymbol{e}_i\boldsymbol{e}_j^T$, 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

$$\frac{\partial \mathrm{Tr}\boldsymbol{X}}{\partial \boldsymbol{X}}=\boldsymbol{I}$$

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

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

定理 1.3.4

$$\frac{\partial \mathrm{Tr}(\boldsymbol{A}\boldsymbol{X})}{\partial \boldsymbol{X}} = \boldsymbol{A}^T$$
$$\frac{\partial \mathrm{Tr}(\boldsymbol{A}\boldsymbol{X})}{\partial \boldsymbol{X}} = \boldsymbol{A} + \boldsymbol{A}^T - \mathrm{diag}\boldsymbol{A}$$
$$\frac{\partial \mathrm{Tr}(\boldsymbol{A}\boldsymbol{X}^{-1})}{\partial \boldsymbol{X}} = -\left(\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X}^{-1}\right)^T$$

定理 1.3.5

$$\frac{\partial \mathrm{Tr}(\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X})}{\partial \boldsymbol{X}} = (\boldsymbol{A} + \boldsymbol{A}^T)\boldsymbol{X}$$
$$\frac{\partial \boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^T}{\partial \boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{A} + \boldsymbol{A}^T)$$
$$\frac{\partial \mathrm{Tr}(\boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B})}{\partial \boldsymbol{X}} = -\left(\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{A}\boldsymbol{X}^{-1}\right)^T$$

行列式的梯度矩阵

定理 1.3.6
若 $\boldsymbol{X}$ 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 $\boldsymbol{X}$ 为对称矩阵，则

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

1.3.4 矩阵对矩阵的微分

定义 1.3.6 (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 $\boldsymbol{F}(\boldsymbol{X}) = \left(f_{ij}(\boldsymbol{X})\right)$

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

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