

浅谈矩阵范数

王尔卓

2022 年 12 月 4 日

摘要

本文主要参考于品的数学分析和丘维生的高等代数上册, 从矩阵范数的定义出发, 讨论了矩阵指数函数的可微性, 并给出了 Cayley-Hamilton 定理的拓扑解释。

目录

1	基本性质	2
2	矩阵级数	3
3	Cayley-Hamilton 定理的拓扑解释	6
4	矩阵指数函数的计算	7

1 基本性质

Definition 1.1. V 是 \mathbb{R} 上的线性空间, $\|\cdot\|: V \rightarrow \mathbb{R}_{\geq 0}$ 是一个函数, 满足:

1. $\|x\| = 0$ 当且仅当 $x = 0$
2. $x \in V, \lambda \in \mathbb{R}$ 有 $\|\lambda x\| = |\lambda| \|x\|$
3. $\|x + y\| \leq \|x\| + \|y\|$

则称这个线性空间为**赋范线性空间**, 如果以度量 $d(x, y) = \|x - y\|$ 诱导出的拓扑空间完备, 则称这个这个线性空间为**Banach 空间**。

我们考虑所有 \mathbb{R} 上 n 级矩阵构成的集合 $M_n(\mathbb{R})$, 其关于通常的加法和数乘法显然构成一个线性空间, 现在对其上任意元素 $A = (a_{ij})$ 赋予范数:

$$\|A\| = \sqrt{\sum_{i=1}^n \sum_{j=1}^n a_{ij}^2}$$

根据欧式空间本身的完备性, 该范数使 $M_n(\mathbb{R})$ 构成一个 Banach 空间, 我们现在证明该范数的一个简单性质。

Proposition 1.2. $\|AB\| \leq C \|A\| \|B\|$

Proof: 设 $A = (a_{ij}), B = (b_{ij}), C = (c_{ij})$, 令 $C = n^2$, 则有:

$$\|AB\| = \|C\| \leq n \max_{1 \leq i, j \leq n} |c_{ij}| \leq n^2 \max_{1 \leq i, j \leq n} |a_{ij}| \max_{1 \leq i, j \leq n} |b_{ij}| \leq C \|A\| \|B\|$$

我们定义赋范线性空间中的收敛为以度量 $d(x, y) = \|x - y\|$ 诱导出的拓扑空间中的收敛, 并定义一个赋范线性空间空间中的级数 $\sum_{n=1}^{\infty} u_n$ 绝对收敛为级数 $\sum_{n=1}^{\infty} \|u_n\|$ 收敛, 不难由柯西列的定义得到, 对于一个 Banach 空间, 其上绝对收敛的级数一定收敛。

接下来, 我们不加证明的将数学分析中绝对收敛级数性质搬运到 $M_n(\mathbb{R})$, 其证明的精神与数学分析相同, 但在证明过程中需要抛弃掉一切有关序关系的工具, 因为我们并未对 $M_n(\mathbb{R})$ 赋予序结构。

Proposition 1.3. $M_n(\mathbb{R})$ 中绝对收敛的级数收敛。

Proposition 1.4. $M_n(\mathbb{R})$ 中绝对收敛的级数在对其进行重排后依然绝对收敛, 且取值相同。

Proposition 1.5. $M_n(\mathbb{R})$ 中绝对收敛的级数任意加括号后也绝对收敛, 且取值相同。

Proposition 1.6. $M_n(\mathbb{R})$ 绝对收敛的两级数的柯西乘积也绝对收敛, 且其值等于两级数取得乘积。

有了上述性质, 我们可以研究矩阵级数了。

2 矩阵级数

Definition 2.1. 我们定义指数矩阵函数为

$$\exp : M_n(\mathbb{R}) \rightarrow M_n(\mathbb{R})$$

$$A \rightarrow e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!}$$

要证明这个定义是良定义，由完备性，只需证 $\forall \varepsilon > 0$ ，存在 $N > 0$ ，任意 $m > n > N$ ，有：

$$\left\| \sum_{k=n}^m \frac{A^k}{k!} \right\| < \varepsilon$$

事实上注意到：

$$\left\| \sum_{k=n}^m \frac{A^k}{k!} \right\| \leq \sum_{k=n}^m \left\| \frac{A^k}{k!} \right\| \leq \sum_{k=n}^m C^{k-1} \frac{\|A\|^k}{k!}$$

则由 e^x 幂级数展开的收敛域与柯西收敛准则可判断出指数矩阵函数的收敛性。接下来我们阐明几条矩阵级数的性质。

Proposition 2.2. $A, B \in M_n(\mathbb{R})$, A 与 B 可交换，则：

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

Proof:

$$\begin{aligned} e^A e^B &= \left(\sum_{n=0}^{\infty} \frac{A^n}{n!} \right) \left(\sum_{n=0}^{\infty} \frac{B^n}{n!} \right) = \sum_{n=0}^{\infty} \sum_{j+i=n} \frac{B^j}{j!} \frac{A^i}{i!} \\ &= \sum_{n=0}^{\infty} \sum_{j+i=n} \frac{B^j}{j!} \frac{A^i}{i!} \frac{(i+j)!}{n!} \\ &= \sum_{n=0}^{\infty} \frac{(A+B)^n}{n!} \end{aligned}$$

Corollary 2.3. $(e^A)^{-1} = e^{-A}$

Corollary 2.4. A 是 n 级反对称矩阵，则 e^A 是正交矩阵。

Proof: 可以验证 A, A^T 可交换，则：

$$I = e^O = e^{A+A^T} = e^A e^{A^T} = e^A (e^A)^T$$

最后一步是因为我们可以把矩阵收敛等价于其 n^2 个位置上的级数分别收敛。

Proposition 2.5. 如果 T 是可逆矩阵，则

$$e^{T^{-1}AT} = T^{-1}e^AT$$

Proof: 从 Proposition 1.2 可以看出, 取极限可以和左乘或右乘一个矩阵交换, 因此我们有:

$$e^{T^{-1}AT} = \sum_{k=0}^{\infty} \frac{(T^{-1}AT)^k}{k!} = T^{-1} \left(\sum_{k=0}^{\infty} \frac{A^k}{k!} \right) T^{-1} = T^{-1} e^A T$$

Proposition 2.6.

$$A = \begin{bmatrix} 0 & x \\ -x & 0 \end{bmatrix}$$

则

$$e^A = \begin{bmatrix} \cos x & \sin x \\ \sin x & \cos x \end{bmatrix}$$

Proof: 这事实上是 $e^{ix} = \cos x + i \sin x$ 矩阵形式的表达。记 $J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ 注意到 J 的幂具有模 4 的周期性, 即:

$$J^{4k+1} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, J^{4k+2} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, J^{4k+3} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, J^{4k} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

又因为:

$$\frac{A^m}{m!} = x^m \frac{J^m}{m!}$$

从而

$$e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!} = \sum_{n=0}^{\infty} x^n \frac{J^n}{n!}$$

由三角函数的定义:

$$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}, \quad \cos x = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}$$

得到 $e^A = \begin{bmatrix} \cos x & \sin x \\ \sin x & \cos x \end{bmatrix}$

下面一个定理刻画了指数矩阵函数的可微性, 从而也得到了连续性。

Theorem 2.7. 指数矩阵函数

$$\exp : M_n(\mathbb{R}) \rightarrow M_n(\mathbb{R})$$

$$A \rightarrow e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!}$$

是 $M_n(\mathbb{R}) \rightarrow M_n(\mathbb{R})$ 的可微映射, 其在 A 处的微分为线性映射:

$$d\exp(A)(V) = \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{k=0}^{n-1} A^k V A^{n-1-k}$$

Proof: 我们先假设 $A \neq O$, , 我们要取计算 A 处的微分, 由于:

$$e^{A+V} - e^V = \sum_{n=0}^{\infty} \frac{(A+V)^n - A^n}{n!}$$

对于 $n \geq 2$, 我们单独考虑里面的项:

$$(A+V)^n - A^n = \sum_{k=0}^{n-1} A^k V A^{n-1-k} + Q_n(V)$$

其中 $Q_n(V)$ 为包含所有包含超过两个 V 的项, 我们不妨设 $\|V\| \leq \|A\|$ (因为可微性只需要体现在局部), 从而对 $Q_n(V)$ 的范数大小有估计:

$$\|Q_n(V)\| \leq 2^n C^{n-1} \|V\|^2 \|A\|^{n-2}$$

从而

$$\sum_{n=2}^{\infty} \left\| \frac{Q_n(V)}{n!} \right\| \leq \sum_{n=2}^{\infty} \frac{2^n C^{n-1} \|A\|^{n-2}}{n!} \|V\|^2 \leq \frac{e^{2C\|A\|}}{\|A\|^2} \|V\|^2 = \mathcal{O}(\|V\|^2)$$

从而有:

$$\|e^{A+V} - e^V - \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{k=0}^{n-1} A^k V A^{n-1-k}\| = \mathcal{O}(\|V\|^2) \quad (1)$$

进而得到 $A \neq O$ 时结论成立, 而 $A = O$ 时, 估计式 (1) 代入检验仍然成立, 因此我们证明了证明了该结论。

Corollary 2.8.

$$\exp : M_n(\mathbb{R}) \rightarrow M_n(\mathbb{R})$$

$$A \rightarrow e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!}$$

当 A, V 可交换时,

$$d\exp(A)(V) = e^A V$$

Banach 空间本身可以定义可微, 自然地我们会去研究形如 $t \rightarrow \exp(At)$ 这样函数的可微性.

Definition 2.9. 对于一个 **Banach 空间** V , 考虑一个函数:

$$f : \mathbb{R} \rightarrow V$$

$$t \rightarrow f(t)$$

f 在 t 处可微是指存在 $v \in V$ 使得

$$\lim_{h \rightarrow 0} \frac{\|f(t+h) - f(t) - hv\|}{h} = 0$$

此时 v 称为 f 在 t 处的导数 (微分)

Theorem 2.10. 考虑函数:

$$\varphi : t \rightarrow \exp(At)$$

其在 Definition 2.9 的意义下可微, 且微分为 $A \exp(At)$.

Proof: 注意到:

$$\begin{aligned} \left\| \frac{\exp(A(t+h)) - \exp(At) - A \exp(At)h}{h} \right\| &= \frac{1}{h} \left\| \exp(At) \sum_{k=2}^{\infty} \frac{(Ah)^k}{k!} \right\| \\ &= h \left\| \exp(At) \sum_{k=2}^{\infty} \frac{(Ah)^{k-2}}{k!} \right\| = \mathcal{O}(h) \rightarrow 0 \end{aligned}$$

证毕!

Corollary 2.11. 齐次线性微分方程组

$$x' = Ax$$

其中一个基解矩阵为 $\exp(At)$

Proof: 在 Definition 2.9 的意义下, 对 $M_n(\mathbb{R})$ 上元素作微分等价于对矩阵中的每个元素作微分, 因此 $\exp(At)$ 显然是该微分方程组的解矩阵, 要证明是基解矩阵只需证明 $\exp(At)$ 始终满秩, 这由 Theorem 4.2 是平凡的.

3 Cayley-Hamilton 定理的拓扑解释

我们在代数学中接触过 Cayley-Hamilton 定理, 这是指对于一个域 F 上矩阵 A , 其特征多项式 $p(x) \in F[x]$ 满足 $P(A) = O$, 而现在我们用矩阵范数重新证明复数域上的 Cayley-Hamilton 定理。

考虑 \mathbb{C} 上 n 级矩阵构成的集合 $M_n(\mathbb{C})$ 上矩阵, 现在对其上任意元素 $A = (a_{ij})$ 赋予范数:

$$\|A\| = \sqrt{\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2}$$

这则 $d(A, B) = \|A - B\|$ 显然是 $M_n(\mathbb{C})$ 上的一个度量, 从而诱导出一个拓扑空间。我们先来刻画这个拓扑空间的性质。

Proposition 3.1. 对于一个多项式 $f(x) \in \mathbb{C}[x]$:

$$\varphi : A \rightarrow f(A)$$

是 $M_n(\mathbb{C})$ 到 $M_n(\mathbb{C})$ 的连续映射。

Proof: 注意到 $M_n(\mathbb{C})$ 到 $M_n(\mathbb{C})$ 的连续映射关于矩阵加法具有可加性, 关于数乘也封闭, 因此只需证明 $A \rightarrow A^k$ 是连续函数, 取 $\|B\|$ 充分小, 从而注意到估计式:

$$\|(A+B)^k - A^k\| \ll \|B\|$$

因此 $A \rightarrow A^k$ 也连续, 从而证明了该结论。

其次我们陈述一个高等代数中的简单命题作为引理:

Lemma 3.2. 如果矩阵 $A \in M_n(\mathbb{C})$ 可对角化, 则对于其特征多项式 $p(x)$ 有 $p(A) = O$

Proof: 只需注意到对角化后对角线上所有元素为特征多项式的根。

下面一个引理是证明的核心

Lemma 3.3. 对于 $A \in M_n(\mathbb{C}), \forall \varepsilon > 0$, 存在 $\|B\| < \varepsilon$, 使得 $A+B$ 可对角化。

Proof: 将 A 相似成若当标准型 $P^{-1}AP$, 对 A 的对角线元素用取满足 $\|B\| < \varepsilon$ 的 B 进行微小的扰动, 使得 $P^{-1}AP - B$ 的对角线上元素各不相同, 从而 $P^{-1}AP - B$ 可对角化, 因此 $A - PBP^{-1}$ 也可对角化, 而 $\|PBP^{-1}\| \leq C\|P\|\|P^{-1}\|\varepsilon$, 由于 P 是取定的, 所以我们已经证明了该结论。

Theorem 3.4. $A \in M_n(\mathbb{C})$, $p(x)$ 为其特征多项式, 则 $p(A) = O$

Proof: 由 Lemma 3.2 与 Lemma 3.3, 得到 A 的任意邻域内一定有元素 Y 使得 $p(Y) = 0$, 再用 Proposition 3.1 得到的连续性, 有

$$p(A) = O$$

证明完毕。

从而我们可以对复数域上矩阵的 Cayley-Hamilton 定理通过矩阵范数给出一个拓扑学解释, 即:

可对角化方阵在所有方阵中稠密

4 矩阵指数函数的计算

借助若当标准型, 我们可以给出如何具体计算矩阵指数函数, 注意到在 Section 2 中, 我们定义的 $M_n(\mathbb{C})$ 上的拓扑与我们最开始定义的 $M_n(\mathbb{R})$ 中的拓扑诸性质类似, 包括完备性, 数乘保极限, 左乘保极限等等, 因此我们讲它们不加证明的迁移。

Definition 4.1. 我们定义(复)指数矩阵函数为

$$\exp : M_n(\mathbb{C}) \rightarrow M_n(\mathbb{C})$$

$$A \rightarrow e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!}$$

首先这个定义是矩阵指数函数的推广，其次，由 Proposition 2.5，我们只需对任意一个矩阵的若当标准型求出其矩阵指数函数即可，再由若当块呈分块对角形状，我们只需对某个若当块求出其矩阵指数函数。若当块特征值为 0 的情况是平凡的，因为此时该矩阵幂零，矩阵指数函数只是一个有限和，当若当块特征值非零时，即若当块形如：

$$J = \begin{bmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & & 0 \\ 0 & 0 & \lambda & & 0 \\ \vdots & & & \ddots & 1 \\ 0 & \dots & & 0 & \lambda \end{bmatrix}_{n \times n} = \lambda I + H_{n-1} \quad \lambda \in \mathbb{C}$$

由 Proposition 2.2：

$$e^J = e^{\lambda I + H_{n-1}} = e^{\lambda I} e^{H_{n-1}} = e^\lambda e^{H_{n-1}}$$

其中 $e^{H_{n-1}}$ 是一个 n 级上三角矩阵，满足 $a_{ij} = \begin{cases} 0 & i < j \\ \frac{1}{(j-i)!} & j \geq i \end{cases}$ ，为了后面叙述方便起见，将 $e^{H_{n-1}}$ 记为 E_n ，并将矩阵指数函数的具体计算展示为如下定理：

Theorem 4.2. 对于一个复数域上的矩阵 A ，其若当标准型为 $J = \text{diag}(J_1, J_2, \dots, J_n)$ ，第 k 个若当块阶数为 a_k ，特征值为 λ_k ，且有可逆矩阵 P 使得 $PAP^{-1} = J$ ，则：

$$\exp(A) = P^{-1} \exp JP = \text{diag}(e^{\lambda_1} E_{a_1}, e^{\lambda_2} E_{a_2}, \dots, e^{\lambda_n} E_{a_n})$$

Corollary 4.3. 若 $\lambda_1, \lambda_2, \dots, \lambda_n$ 为 A 的特征值，若 $\mu_1, \mu_2, \dots, \mu_n$ 为 B 的特征值 (考虑重数)，则

$$\begin{aligned} \text{tr}(\exp(A)) &= \sum_{i=1}^n e^{\lambda_i} \\ \text{tr}(\exp A \otimes B) &= \sum_{i=1}^n \sum_{j=1}^m e^{\lambda_i \mu_j} \end{aligned}$$

Proof: 前者平凡，后者只需注意到 $(A \otimes B)(C \otimes D) = AC \otimes BD$ ，因此对于可逆矩阵 P, Q 而言，

$$(P \otimes Q)^{-1} = P^{-1} \otimes Q^{-1}$$

如果通过可逆矩阵 P, Q 将 A, B 转化为若当标准型 D_1, D_2 ，即

$$P^{-1}AP = D_1, Q^{-1}AQ = D_2$$

则有：

$$(P \otimes Q)^{-1}(A \otimes B)(P \otimes Q) = P^{-1} \otimes Q^{-1}(A \otimes B)(P \otimes Q) = D_1 \otimes D_2$$

从而看出 $A \otimes B$ 的特征值 (考虑重数) 为 $\lambda_i \mu_j, 1 \leq i, j \leq n$