

矩阵论进阶知识

本文依据 方保镕、周继东、李医民《矩阵论》（第2版，清华大学出版社，2013）的体系，梳理矩阵论的进阶理论与应用。教材分 上篇（基础篇）与 下篇（应用篇）共 8 章；本书将其中超出入门范围的内容加以归纳，并与姊妹篇 FundamentalsOfMatrixTheory.md 衔接。

前置阅读：FundamentalsOfMatrixTheory.md 已覆盖矩阵运算、行列式、秩、逆、特征值、基本分解与范数入门。本文在此基础上深入几何观点、标准形、广义逆、矩阵微积分及特殊矩阵等主题。

导读：与教材章节对照

教材章节	本文章节	核心进阶内容	基础篇对应
第1章 矩阵的几何理论	§ 1	线性算子、内积空间、等积变换、埃尔米特理论	§ 1–§ 2、§ 6–§ 7
第2章 λ 矩阵与 Jordan 标准形	§ 2	λ 矩阵、Smith 型、不变因子、Jordan 块	§ 8（特征值）
第3章 矩阵的分解	§ 3	LDU、极分解、谱分解、正规矩阵	§ 9
第4章 赋范线性空间与矩阵范数	§ 4	范数公理、诱导范数、谱半径	§ 10
第5章 矩阵微积分及其应用	§ 5	矩阵函数、矩阵求导、矩阵指数	—
第6章 广义逆矩阵及其应用	§ 6	Moore–Penrose 伪逆、最小二乘	§ 4、§ 9 满秩分解
第7章 几类特殊矩阵与特殊积	§ 7	非负矩阵、M 矩阵、Kronecker 积等	§ 6
第8章 矩阵在数学内外的应用	§ 8	综合应用导引	§ 11

上篇 基础篇

§ 1 矩阵的几何理论

教材强调：矩阵论是 **多维高等数学**——高等数学研究一维变量之间的函数关系，矩阵论研究多维向量（原像与像）之间的 **线性算子** 及其矩阵表示。

1.1 线性空间与线性算子

设 V, W 为域 $\mathbb{F}$ ($\mathbb{R}$ 或 $\mathbb{C}$) 上的有限维线性空间，**线性算子** $T: V \rightarrow W$ 满足

$$T(\alpha \mathbf{u} + \beta \mathbf{v}) = \alpha T(\mathbf{u}) + \beta T(\mathbf{v}), \quad \forall \alpha, \beta \in \mathbb{F}, \mathbf{u}, \mathbf{v} \in V.$$

取 $V = \mathbb{F}^n$, $W = \mathbb{F}^m$, 标准基 $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, 则 T 由 $m \times n$ 矩阵 A 完全确定:

$$T(\mathbf{x}) = A\mathbf{x}, \quad A = (T(\mathbf{e}_1) \mid \dots \mid T(\mathbf{e}_n)).$$

换基: 若 P 为 V 中新基到旧基的过渡矩阵, Q 为 W 中类似过渡矩阵, 则同一算子在新基下的矩阵为

$$B = Q^{-1}AP.$$

核与像 (与 FundamentalsOfMatrixTheory.md §4 一致):

概念	定义	维数关系
$\ker(T) = \ker(A)$	$T(\mathbf{x}) = \mathbf{0}$ 的解空间	$\dim \ker(A) + \text{rank}(A) = n$
$\text{im}(T) = \text{col}(A)$	T 的值域, 即 A 的列空间	$\dim \text{im}(A) = \text{rank}(A)$

1.2 内积空间与等积变换

内积空间 $(V, \langle \cdot, \cdot \rangle)$ 在向量空间上增加内积, 诱导 **范数** $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ 与 **角度**。

在 $\mathbb{R}^n$ 上常用 **欧氏内积** $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$; 在 $\mathbb{C}^n$ 上用 **标准内积** $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^* \mathbf{y}$ 。

等积变换 (保距变换): 线性算子 T 满足 $\|T(\mathbf{x})\| = \|\mathbf{x}\|$ 对一切 $\mathbf{x}$ 成立。

域	等积变换的矩阵	条件
$\mathbb{R}$	正交矩阵 Q	$Q^T Q = I$

域	等积变换的矩阵	条件
$\mathbb{C}$	酉矩阵 U	$U^*U = I$

几何意义：旋转、反射等不改变长度与夹角。实情形见 FundamentalsOfMatrixTheory.md § 6。

1.3 埃尔米特变换与矩阵不等式

埃尔米特变换 ($\mathbb{C}$) / **对称变换** ($\mathbb{R}$)：满足 $\langle T\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, T\mathbf{y} \rangle$ ，矩阵形式为 $A^* = A$ (或 $A^T = A$)。

谱定理 (埃尔米特情形)： n 阶埃尔米特矩阵 A 可酉对角化：

$$A = U\Lambda U^*, \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n) \in \mathbb{R}.$$

正定与半正定 (深化 FundamentalsOfMatrixTheory.md § 6–§ 7)：

类型	定义	等价条件 (埃尔米特时)
正定	$\mathbf{x}^* A \mathbf{x} > 0, \forall \mathbf{x} \neq \mathbf{0}$	特征值全 > 0 ；存在可逆 P 使 $A = P^* P$
半正定	$\mathbf{x}^* A \mathbf{x} \geq 0$	特征值全 ≥ 0
负定	$\mathbf{x}^* A \mathbf{x} < 0, \forall \mathbf{x} \neq \mathbf{0}$	特征值全 < 0

矩阵不等式 (Loewner 序)：对埃尔米特矩阵 A, B ，记 $A \succ B$ (或 $A > B$) 当 $A - B$ 正定； $A \succeq B$ 当 $A - B$ 半正定。

常用结论：

- 若 $A \succeq O, B \succeq O$ ，则 $A + B \succeq O$ 。
- 若 A 埃尔米特，则 $\lambda_{\min}(A) \leq \mathbf{x}^* A \mathbf{x} / \mathbf{x}^* \mathbf{x} \leq \lambda_{\max}(A)$ 。
- Rayleigh 商** $R(\mathbf{x}) = \frac{\mathbf{x}^* A \mathbf{x}}{\mathbf{x}^* \mathbf{x}}$ 的极值为最大、最小特征值。

§ 2 λ 矩阵与 Jordan 标准形

当矩阵不可对角化时，特征值分解需推广为 **Jordan 标准形**。教材通过 **λ 矩阵** 理论建立这一标准形。

2.1 λ 矩阵

λ 矩阵 $A(\lambda)$ ：元素为多项式 $a_{ij}(\lambda) \in \mathbb{F}[\lambda]$ 的矩阵。特别地，**特征矩阵** $\lambda I - A$ 是 λ 矩阵。

两种等价关系：

关系	变换	不变量
相抵	$A \sim B$: 存在可逆 λ 矩阵 P, Q 使 $PAQ = B$	秩、不变因子
相似	$A \sim B$: 存在可逆 P 使 $P^{-1}AP = B$	特征多项式、初等因子、Jordan 形

2.2 Smith 标准形（相抵下）

对域 $\mathbb{F}$ 上秩为 r 的 $m \times n$ λ 矩阵 $A(\lambda)$ ，存在可逆 λ 矩阵 P, Q 使

$$PA(\lambda)Q = \text{diag}(d_1(\lambda), \dots, d_r(\lambda), 0, \dots, 0),$$

其中 $d_i(\lambda)$ 为 **不变因子**，满足 $d_i \mid d_{i+1}$ 且为首一多项式。

在 $\mathbb{Z}$ 上（整数矩阵情形，见 SimulationTest1_Answers.md）：用 **初等行、列变换** 化为对角形，对角元 $d_i \in \mathbb{Z}$ 满足 $d_i \mid d_{i+1}$ 。

初等因子：将不变因子在 $\mathbb{F}$ 上分解为不可约因子的幂，每个因子幂 $p(\lambda)^k$ 对应一个 Jordan 块。

2.3 Jordan 标准形（相似下）

对 n 阶方阵 A ，存在可逆 P 使

$$P^{-1}AP = J = \text{diag}(J_{n_1}(\lambda_1), \dots, J_{n_k}(\lambda_k)),$$

其中 **Jordan 块**

$$J_m(\lambda) = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}_{m \times m}$$

求法概要：

- 1. 求特征多项式 $\det(\lambda I - A)$ ，得各特征值 λ_i 及其 **代数重数** n_i 。
- 2. 对每个 λ_i ，求 **几何重数** $g_i = n - \text{rank}(A - \lambda_i I)$ 。**块的个数** $= g_i$ ，**各块阶数之和** $= n_i$ ，二者联立确定块结构； $1 < g_i < n_i$ 时须进一步求 $\dim \ker(A - \lambda_i I)^k$ 的维数链。
- 3. 或：对特征矩阵 $\lambda I - A$ 求 Smith 标准形，读初等因子 $(\lambda - \lambda_i)^{n_j}$ ，每个对应 $J_{n_j}(\lambda_i)$ 。

例 1 ($g_i = 1$)： $A = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$ ，特征值 $\lambda = 2$ ， $n = 2$ ， $g = 1$ ，故恰有一个 2 阶块：

$$A \sim J_2(2) = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$$

例 2 ($1 < g < n$): $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $\lambda = 0$, $n = 3$, $g = 2$ 。记 $\nu_k = \dim \ker(A^k)$, 则 $\nu_1 = 2$, $\nu_2 = 3$ 。阶数 $\geq k$ 的块数 $= \nu_k - \nu_{k-1}$:

- $k = 2$: $\nu_2 - \nu_1 = 1$ 个块 ≥ 2 阶 $\rightarrow$ 一个 $J_2(0)$, 余下一个 $J_1(0)$ 。

$$A \sim \text{diag}(J_2(0), J_1(0)) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

可对角化的判据: A 有 n 个线性无关特征向量 $\Leftrightarrow$ 每个特征值的几何重数等于代数重数 $\Leftrightarrow$ Jordan 形为对角阵。

§ 3 矩阵的分解（深化）

FundamentalsOfMatrixTheory.md § 9 已给出 LU、Cholesky、QR、特征值分解、SVD、满秩分解的分步算例。本节按教材补充 **LDU**、**极分解** 与 **谱分解**。

3.1 三角分解与 LDU 分解

LU 分解 (Doolittle): $A = LU$, L 单位下三角, U 上三角。

若 A 为方阵且各顺序主子式均非零, 还可写成 **LDU 分解**:

$$A = LDU,$$

L 单位下三角, D 对角, U 单位上三角。 D 的对角元为消元主元。

Crout 分解: U 为单位上三角, 主元记入 L 的对角。

3.2 QR 分解与 Householder 法

Gram-Schmidt 正交化可手算 QR, 但数值不稳定。工程实现常用 **Householder 反射**:

对列向量 $\mathbf{a}$, 构造 $H = I - 2\mathbf{u}\mathbf{u}^T / \|\mathbf{u}\|^2$, 使 $H\mathbf{a} = \|\mathbf{a}\|\mathbf{e}_1$, 逐列消元得上三角 R , $Q = H_1H_2 \cdots H_k$ 。

3.3 满秩分解与 SVD

满秩分解 $A = FG$ (F 列满秩 $m \times r$, G 行满秩 $r \times n$): 行最简形主元列作 F , 主元行作 G 。详见 FundamentalsOfMatrixTheory.md § 9。

奇异值分解 $A = U\Sigma V^T$ 与 **极分解**:

- 对任意 $m \times n$ 实矩阵 A ，极分解为 $A = QP$ ， Q 列正交 ($m \times n$)， P 半正定 ($n \times n$)。
- 由 SVD: $A = U\Sigma V^T = (UV^T)(V\Sigma V^T)$ ，其中 UV^T 在方阵情形为正交因子。

3.4 谱分解

正规矩阵： $AA^* = A^*A$ 。实情形即 $AA^T = A^T A$ （对称矩阵正规；正交矩阵亦正规）。

谱定理：正规矩阵可酉（实正交）对角化 $A = U\Lambda U^*$ 。

谱分解（特征投影形式）：设 $\lambda_1, \dots, \lambda_k$ 为互异特征值， P_i 为到特征子空间 $\ker(A - \lambda_i I)$ 的投影，则

$$A = \sum_{i=1}^k \lambda_i P_i, \quad P_i P_j = \delta_{ij} P_i, \quad \sum_i P_i = I.$$

单纯矩阵（可对角化矩阵）的谱分解即 $A = PDP^{-1}$ ， P 的列为特征向量。

§ 4 赋范线性空间与矩阵范数

4.1 向量范数

范数 $\|\cdot\| : V \rightarrow \mathbb{R}$ 满足：

- 正定性： $\|\mathbf{x}\| \geq 0$ ，且 $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$
- 齐次性： $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$
- 三角不等式： $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

常用向量范数 ($\mathbf{x} = (x_1, \dots, x_n)^T$)：

名称	定义	记号
1-范数	$\ \mathbf{x}\ _1 = \sum_i$	$\ \mathbf{x}\ _1$
2-范数	$\ \mathbf{x}\ _2 = \sqrt{\sum_i x_i^2}$	欧氏长度
∞ -范数	$\ \mathbf{x}\ _\infty = \max_i$	$\ \mathbf{x}\ _\infty$

范数等价：有限维空间上任意两种范数等价，即存在 $c_1, c_2 > 0$ 使 $c_1 \|\mathbf{x}\|_\alpha \leq \|\mathbf{x}\|_\beta \leq c_2 \|\mathbf{x}\|_\alpha$ 。

4.2 矩阵范数

矩阵范数 除满足正定性、齐次性、三角不等式外，还需 **次可乘性**（相容性）：

$$\|AB\| \leq \|A\| \|B\|.$$

范数	定义	说明
Frobenius	$\ A\ _F = \sqrt{\sum_{i,j} a_{ij}^2}$	a_{ij}
谱范数 (2-范数)	$\ A\ _2 = \sigma_{\max}(A)$	最大奇异值
1-范数	$\ A\ _1 = \max_j \sum_i a_{ij} $	a_{ij}
∞ -范数	$\ A\ _{\infty} = \max_i \sum_j a_{ij} $	a_{ij}

诱导范数： $\|A\|_{(\alpha,\beta)} = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_{\beta}}{\|\mathbf{x}\|_{\alpha}}$ 。谱范数即由 2-范数诱导。

谱半径： $\rho(A) = \max_i |\lambda_i|$ 。一般 $\rho(A) \leq \|A\|$ ；当 A 正规时 $\|A\|_2 = \rho(A)$ 。

条件数（见 FundamentalsOfMatrixTheory.md § 10）：

$$\kappa(A) = \|A\| \|A^{-1}\|, \quad \kappa(A) \geq 1.$$

κ 越大，解 $A\mathbf{x} = \mathbf{b}$ 对扰动越敏感。

下篇 应用篇

§ 5 矩阵微积分及其应用

5.1 矩阵函数

对标量函数 f ，方阵 A 的 **矩阵函数** $f(A)$ 通过 Jordan 形或谱分解定义：

- 若 $A = PJP^{-1}$ ，则 $f(A) = Pf(J)P^{-1}$ ， $f(J)$ 对 Jordan 块逐块计算。
- 对 Jordan 块 $J_m(\lambda)$ ， $f(J_m(\lambda))$ 为上三角 Toeplitz 阵，对角元 $f(\lambda)$ ，上超对角为 $f'(\lambda), f''(\lambda)/2!, \dots$

常用矩阵函数：

函数	定义	应用
e^A	矩阵指数	线性微分方程 $\dot{\mathbf{x}} = A\mathbf{x}$ 的解 $\mathbf{x}(t) = e^{At}\mathbf{x}(0)$

函数	定义	应用
$\ln A$	矩阵对数（谱值 > 0 时）	协方差分解、流形优化
$A^{1/2}$	矩阵平方根（半正定 A ）	Cholesky 的推广

矩阵指数：若 $A = U\Lambda U^{-1}$ 可对角化，则 $e^A = U\text{diag}(e^{\lambda_i})U^{-1}$ 。

5.2 矩阵求导

设 $f: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ 为标量函数。

常见结论 ($A \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^n$):

$$\begin{aligned}\frac{\partial}{\partial A} \text{tr}(AB) &= B^T, & \frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^T A \mathbf{x}) &= (A + A^T) \mathbf{x}, \\ \frac{\partial}{\partial A} \|A\|_F^2 &= 2A, & \frac{\partial}{\partial \mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|_2^2 &= 2A^T(A\mathbf{x} - \mathbf{b}).\end{aligned}$$

最后一式即最小二乘梯度，用于优化与机器学习。

§ 6 广义逆矩阵及其应用

当 A 不可逆甚至非方阵时，**Moore–Penrose 伪逆** A^+ 推广了逆矩阵。

6.1 定义（Penrose 方程）

A^+ 是满足以下四式的唯一矩阵：

$$AA^+A = A, \quad A^+AA^+ = A^+, \quad (AA^+)^* = AA^+, \quad (A^+A)^* = A^+A.$$

性质：

情形	结果
A 可逆	$A^+ = A^{-1}$
A 列满秩	$A^+ = (A^*A)^{-1}A^*$
A 行满秩	$A^+ = A^*(AA^*)^{-1}$
一般情形	由 SVD 或满秩分解计算

6.2 由满秩分解求 A^+

设 $A = FG$ （满秩分解， $\text{rank}(A) = r$ ），则

$$A^+ = G^T(GG^T)^{-1}(F^TF)^{-1}F^T.$$

分步算例见 SimulationTest1_Answers.md 第1题。

6.3 由 SVD 求 A^+

设 $A = U\Sigma V^T$ ， $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r, 0, \dots)$ ， $\sigma_i > 0$ 。则

$$A^+ = V\Sigma^+U^T, \quad \Sigma^+ = \text{diag}\left(\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_r}, 0, \dots\right).$$

6.4 线性方程组的应用

对 $A\mathbf{x} = \mathbf{b}$ ：

问题	解
相容（有解）	极小范数解 $\mathbf{x} = A^+\mathbf{b}$
不相容（最小二乘）	极小范数最小二乘解 $\mathbf{x} = A^+\mathbf{b}$

此时 $A\mathbf{x} = AA^+\mathbf{b}$ 为 $\mathbf{b}$ 在 $\text{col}(A)$ 上的正交投影。

§ 7 几类特殊矩阵与特殊积

7.1 非负矩阵与正矩阵

非负矩阵 $A \geq O$ ： $a_{ij} \geq 0$ 。 正矩阵 $A > O$ ： $a_{ij} > 0$ 。

Perron–Frobenius 定理（正矩阵 A ）：

- 存在唯一最大正特征值 $\rho(A)$ （谱半径），其余特征值模严格小于 $\rho(A)$ 。
- $\rho(A)$ 对应正特征向量。
- $\rho(A)$ 在连通正矩阵情形下为单特征值。

应用于 Markov 链、人口模型、PageRank 等。

7.2 M 矩阵与 H 矩阵

M 矩阵：可表示为 $A = sI - B$ ，其中 $B \geq O$ ， $\rho(B) < s$ 。等价地，非对角元 ≤ 0 ，且 A 的所有顺序主子式 > 0 （或等价条件之一成立）。

性质：M 矩阵非奇异，且 $A^{-1} \geq O$ 。用于偏微分方程离散、稳定性分析。

H 矩阵：其比较矩阵 B ($b_{ii} = a_{ii}$, $b_{ij} = -|a_{ij}|$, $i \neq j$) 为 M 矩阵。

7.3 随机矩阵与双随机矩阵

随机矩阵（行随机）：各行和为 1，元素非负。描述概率转移。

双随机矩阵：行和、列和均为 1，元素非负。Birkhoff 定理：双随机矩阵是置换矩阵的凸组合。

7.4 循环矩阵与 Toeplitz 矩阵

Toeplitz 矩阵：每条对角线元素相同；**循环矩阵** 由第一行生成，可用 FFT 快速乘法，在信号处理中重要。

Hankel 矩阵：反对角线元素相同。

7.5 矩阵特殊积

积	定义	性质
Kronecker 积 $A \otimes B$	分块矩阵 $(a_{ij}B)$	$(A \otimes B)(C \otimes D) = AC \otimes BD$
Hadamard 积 $A \circ B$	逐元素相乘 $(a_{ij}b_{ij})$	保持对称性；Schur 积定理：半正定阵 Hadamard 积仍半正定
反积 (Khatri-Rao)	列-wise Kronecker	张量分解

vec 技巧： $\text{vec}(ABC) = (C^T \otimes A)\text{vec}(B)$ ，将矩阵方程化为向量形式。

§ 8 矩阵在数学内外的应用

教材第 8 章贯穿各领域，此处归纳与本文各章的对应关系。

应用领域	所用矩阵工具	本文章节
数值线性代数	LU/QR/SVD、条件数、迭代法	§ 3、§ 4
常微分方程	e^{At} 、Jordan 形	§ 2、§ 5
最小二乘与回归	A^+ 、QR、正规方程	§ 3、§ 6
控制理论	可控性矩阵、特征值、Lyapunov 方程	§ 1、§ 2、§ 7
图论与网络	邻接矩阵、Laplacian、随机矩阵	§ 7、FundamentalsOfMatrixTheory.md § 11
信号与图像	Toeplitz、SVD 压缩、FFT	§ 3、§ 7
量子力学	埃尔米特算符、酉演化	§ 1
机器学习	梯度、伪逆、正定核矩阵	§ 5、§ 6、§ 7

附录 A：学习建议

- 先巩固基础篇：**FundamentalsOfMatrixTheory.md 中的运算、秩、特征值、基本分解应熟练。
- 几何观点贯穿：**每遇矩阵运算，追问其作为线性算子的几何意义（核、像、保距、谱）。
- 标准形是核心：**Jordan 形解决"不可对角化"；Smith 形解决" λ 矩阵相抵"。
- 分解即算法：**LU/QR/SVD 不仅是理论，更是数值计算的基础（LAPACK）。
- 习题配套：**教材附录含 15 套模拟自测；可参考 SimulationTest1_Answers.md 的解答风格。

附录 B：推荐阅读

资料	说明
方保镕、周继东、李医民《矩阵论》（第2版）	本文主依据
方保镕《矩阵论千题习题详解》	配套习题解答
FundamentalsOfMatrixTheory.md	本仓库入门篇

资料	说明
Golub & Van Loan, <i>Matrix Computations</i>	数值线性代数经典

小结：方保镕等《矩阵论》以 **几何理论** 为起点，经 **λ 矩阵与 Jordan 形** 完成标准形理论，以 **分解与范数** 连接数值计算，再以 **矩阵微积分、广义逆、特殊矩阵** 支撑工程应用。掌握这一体系，即可从"会算矩阵"进阶到"理解矩阵作为多维线性工具"的本质。