矩阵理论复习笔记

前言

本复习笔记是我个人在学习矩阵理论的过程中整理、总结而成，包含课本的内容、上课 PPT 涉及到的内容以及不懂地方的补充知识，希望能够对你有所帮助。

全书排版是利用 LaTeX 完成的，这也是对我使用 LaTeX 的一次较大工程的练手，希望我能在撰写完之后对于 LaTeX 使用有更深层次的理解。

因本人水平有限，故本总结笔记如有不当之处，敬请指出，本人不胜感激！

尹彦江

2024 年 7 月 10 日

目录

第零章　线性代数基本知识

本章主要针对矩阵理论中所涉及到的线性代数知识进行复习与回顾，随着课程的进行，当认为有必要需要补充相关的线性代数基本知识时，会补充在本节。

本章会涉及以下内容：

- 行列式
- 矩阵
- 向量组、线性方程组
- 特征值与特征向量，矩阵相似理论

注意：本章只作为一个对于基本知识的回顾与复习，并不会涉及到更详细的解释，如果需要查看更详细的解释，请翻阅对应的线性代数教材或上网搜集资料。[1]

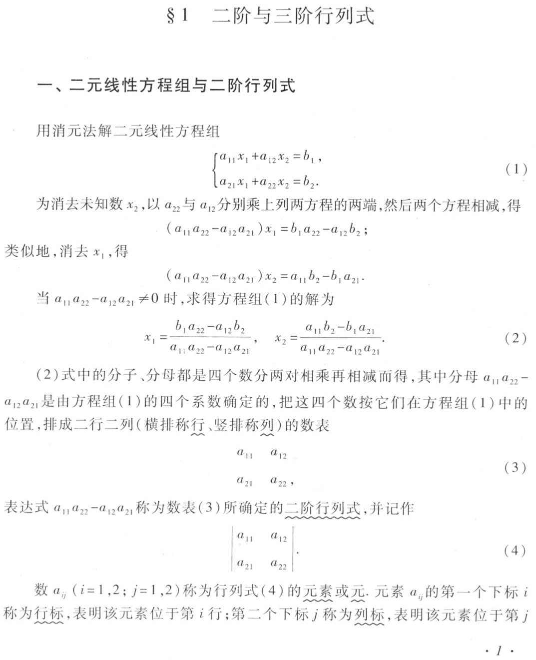
[1]本章的大多数内容是在《2025 张宇考研数学基础 30 讲——线性代数分册》的基础上整理总 (抄) 结 (写) 而成

0.1 行列式

行列式作为后续矩阵理论中涉及特征值与特征向量计算的基础，因此有必要复习巩固行列式的定义、行列式的相关性质以及具体行列式的计算。

0.1.1 行列式的定义与计算方法

何为行列式？如果查阅线性代数的教材（这里以同济六版《线性代数》为例），会发现教材是从特例到一般进行推广，首先给出二阶行列式和三阶行列式应如何计算，后面给出全排列的逆序数的概念，最后推广至 n 阶行列式得出 n 阶行列式的定义的，而在介绍二阶行列式的时候，是先从解一个二元线性方程组得出的：



§1 二阶与三阶行列式

一、二元线性方程组与二阶行列式

用消元法解二元线性方程组

$$\begin{cases} a_{11}x_1+a_{12}x_2=b_1, \\ a_{21}x_1+a_{22}x_2=b_2. \end{cases} \tag{1}$$

为消去未知数 x_2，以 a_{22} 与 a_{12} 分别乘上列两方程的两端，然后两个方程相减，得

$$(a_{11}a_{22}-a_{12}a_{21})x_1=b_1a_{22}-a_{12}b_2;$$

类似地，消去 x_1，得

$$(a_{11}a_{22}-a_{12}a_{21})x_2=a_{11}b_2-b_1a_{21}.$$

当 $a_{11}a_{22}-a_{12}a_{21}\neq 0$ 时，求得方程组(1)的解为

$$x_1=\frac{b_1a_{22}-a_{12}b_2}{a_{11}a_{22}-a_{12}a_{21}},\quad x_2=\frac{a_{11}b_2-b_1a_{21}}{a_{11}a_{22}-a_{12}a_{21}}. \tag{2}$$

(2)式中的分子、分母都是四个数分两对相乘再相减而得，其中分母 $a_{11}a_{22}-a_{12}a_{21}$ 是由方程组(1)的四个系数确定的，把这四个数按它们在方程组(1)中的位置，排成二行二列（横排称行、竖排称列）的数表

$$\begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22}, \end{matrix} \tag{3}$$

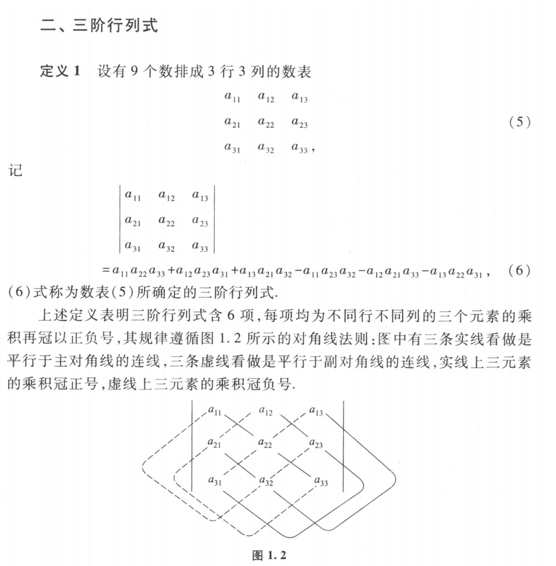
表达式 $a_{11}a_{22}-a_{12}a_{21}$ 称为数表(3)所确定的二阶行列式，并记作

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}. \tag{4}$$

数 a_{ij} $(i=1,2;\ j=1,2)$ 称为行列式(4)的元素或元. 元素 a_{ij} 的第一个下标 i 称为行标，表明该元素位于第 i 行；第二个下标 j 称为列标，表明该元素位于第 j

· 1 ·

(a) 教材中对于二阶行列式的介绍



二、三阶行列式

定义 1 设有 9 个数排成 3 行 3 列的数表

$$\begin{matrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33}, \end{matrix} \tag{5}$$

记

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}-a_{11}a_{23}a_{32}-a_{12}a_{21}a_{33}-a_{13}a_{22}a_{31}, \tag{6}$$

(6)式称为数表(5)所确定的三阶行列式.

上述定义表明三阶行列式含 6 项，每项均为不同行不同列的三个元素的乘积再冠以正负号，其规律遵循图 1.2 所示的对角线法则：图中有三条实线看做是平行于主对角线的连线，三条虚线看做是平行于副对角线的连线，实线上三元素的乘积冠正号，虚线上三元素的乘积冠负号.

图 1.2

(b) 教材中对于二三行列式的介绍

图 1: 《线性代数》教材中对于二阶、三阶行列式的介绍

这样的定义是从代数的角度出发的，可能对于行列式的理解并不是很清楚，接下来会给出除课本之外的另一种理解行列式的方法。

行列式的第一种定义

假设我们有如下的一个二阶行列式：

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

我们将该行列式的第一行看作一个向量，即 $\boldsymbol{a_1} = [a_{11}, a_{12}]$，将第二行看作另外一个向量，即 $\boldsymbol{a_2} = [a_{21}, a_{22}]$，将这两个向量放在平面直角坐标系中，并以这两条边构建成一个平行四边形，如下图所示：

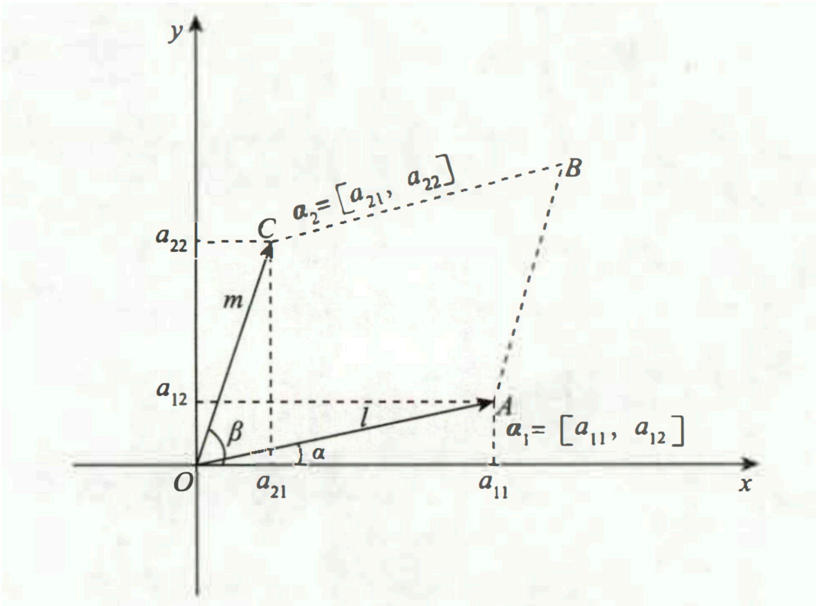


图 2: 将两个向量放在平面直角坐标系中表示的效果图

那么，这个平行四边形的面积是多少呢？我们设 $\boldsymbol{a_1}$ 与 x 轴的夹角为 α, $\boldsymbol{a_2}$ 与 x 轴的夹角为 β, 设 $||\boldsymbol{a_1}|| = l, ||\boldsymbol{a_2}|| = m$。

则，平行四边形的面积为：

$$\begin{aligned} S &= l \cdot m \cdot \sin(\beta - \alpha) \\ &= l \cdot m(\sin\beta\cos\alpha - \cos\beta\sin\alpha) \\ &= l\cos\alpha \cdot m\sin\beta - l\sin\alpha \cdot \cos\beta \\ &= a_{11}a_{22} - a_{12}a_{21} \end{aligned}$$

这是不是与二阶行列式的结果一样呢？所以我们可以得出一个结论：二阶行列式可以表示为一个平行四边形的面积。

接下来我们讨论三阶行列式，与二阶行列式相似，假设我们有如下三阶行列式:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

其实,三阶行列式的几何意义即为是一个由三个三维向量 $\boldsymbol{a_1} = [a_{11}, a_{12}, a_{13}]$, $\boldsymbol{a_2} = [a_{21}, a_{22}, a_{23}]$, $\boldsymbol{a_3} = [a_{31}, a_{32}, a_{33}]$ 组成的一个平行六面体的体积，如下图:

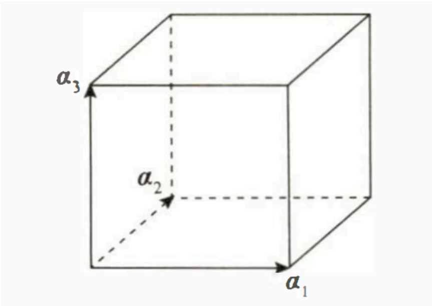


图 3: 三阶行列式的几何意义示意

再进行推广，我们可以得到 n 阶行列式的定义:

定义 0.1.1

一个 n 阶行列式

$$D_n = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{1n} & \cdots & a_{nn} \end{vmatrix}$$

其是由 n 个 n 维向量:

$$\boldsymbol{a_1} = [a_{11}, a_{12}, \cdots, a_{1n}]$$

$$\boldsymbol{a_2} = [a_{21}, a_{22}, \cdots, a_{2n}]$$

$$\cdots$$

$$\boldsymbol{a_n} = [a_{n1}, a_{n2}, \cdots, a_{nn}]$$

组成的，其结果是以这 n 个向量为边的 n 维图形的“体积”。

这里可以结合后面线性相关和线性无关的性质：如果一个方阵的行列式为 0，就说这个向量组是线性相关的，如果用行列式的几何意义理解将会变得十分简单，因为如果一个向量组线性相关，就说明这个向量组至少有一个向量是与其他的向量共线的，如果共线，就说明这个 n 维图形将会少一条边，少一条边“体积”自然为 0。

行列式的第二种定义

后面行列式的第二种与第三种定义与教材中对于行列式的定义相同，第二种行列式引入了“逆序数”这一概念来辅助定义行列式，即一个 n 阶行列式:

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{1n} & \cdots & a_{nn} \end{vmatrix} = \sum_{j_1j_2\cdots j_n} (-1)^{r(j_1j_2\cdots j_n)} a_{1j_1}a_{2j_2}\cdots a_{nj_n}$$

这里 $r(j_1j_2\cdots j_n)$ 表示的是数列 $j_1, j_2, \cdots, j_n$ 的逆序数[2]。

注. 如何理解上面的公式：计算一个 n 阶行列式只需要找出其所有的行和所有的列的展开乘以对应的系数（-1 的正负），最后将所有的和加起来即为行列式的值。

行列式的第三种定义

这里介绍行列式的第三种定义——代数余子式与余子式形式，这一种定义也是计算行列式的常用方法。

首先，回顾一下什么是余子式，什么是代数余子式:

- 余子式：将一个 n 阶行列式，去掉元素 a_{ij} 所在的行和列（即第 i 行和第 j 列）之后得到的新的行列式，记为 $\boldsymbol{M_{ij}}$。
- 代数余子式：余子式乘以 $(-1)^{i+j}$ 记为代数余子式，记为 $\boldsymbol{A_{ij}}$。

根据上面的介绍，不难得出下列关系:

$$A_{ij} = (-1)^{i+j}M_{ij}$$

由此，可以得出行列式的第三种定义:

定义 0.1.2

可以按照某一行某一列来展开行列式，即:

$$|\boldsymbol{A}| = \begin{cases} a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in} = \sum\limits_{j=1}^{n} a_{ij}A_{ij} & (i = 1, 2, \cdots, n) \to (\text{按行展开}) \\ a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj} = \sum\limits_{i=1}^{n} a_{ij}A_{ij} & (j = 1, 2, \cdots, n) \to (\text{按列展开}) \end{cases}$$

[2]关于逆序数的详细解释，请参阅: https://zhuanlan.zhihu.com/p/398845457

注. 行列式的某一行或者某一列乘以其他非对应行或非对应列的代数余子式后再求和，其结果为 0。

例 0.1.1. 试计算行列式

$$D=\begin{vmatrix} 2 & -1 & 0 & 0 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 2 & -1 \\ -1 & 0 & 0 & 2 \end{vmatrix}$$

解. 计算行列式的值，按照定义三来解题的话，尽可能找“0”多的行或者列进行展开，这样可以简单计算。

将行列式按第一列展开，得

$$\begin{aligned} D &= 2\begin{vmatrix} 2 & -1 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & 2 \end{vmatrix} - 1\times(-1)^{(4+1)}\begin{vmatrix} -1 & 0 & 0 \\ 2 & -1 & 0 \\ 0 & 2 & -1 \end{vmatrix} \\ &= 16-1=15 \end{aligned}$$

□

0.1.2 行列式的性质

只知道了行列式的计算还不够，运用好行列式的性质可以在处理复杂行列式的时候及时发现其规律对其进行处理，从而达到简便计算的效果。

行列式有下列性质：

- 行列互换，其值不变，即 $|\boldsymbol{A}|=|\boldsymbol{A}^T|$
- 若行列式中的某一行或者某一列的元素全为 0，则行列式为 0
- 若行列式中某一行或者某一列有公因子 $k(k\neq 0)$，则可以将该公因子提到行列式外

面，即：

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ ka_{11} & ka_{12} & \cdots & ka_{1n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{1nn} \end{vmatrix} = \begin{vmatrix} ka_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ ka_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ ka_{n1} & a_{n2} & \cdots & a_{1nn} \end{vmatrix} = k \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{1nn} \end{vmatrix}$$

注.

1. 这叫做行列式的“倍乘”性质
2. 请注意在行列式中可以提出公因子和矩阵里可以提出公因子的区别，在矩阵中需要要求所有所有元素都含有公因子才可以提出来。

- 行列式中某一行或者某一列的元素均是某两个数之和，则可以将行列式拆成两个行列式之和，即

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{i1}+b_{i1} & a_{i2}+b_{i2} & \cdots & a_{in}+b_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ b_{i1} & b_{i2} & \cdots & b_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

注. 请仔细留意该性质成立的条件

- 行列式中两行或两列互换，行列式变号
- 行列式中两行或两列的元素对应成比例，则行列式结果为 0
- 行列式中的某一行或某一列的 k 倍加到另外一行或另外一列，行列式不变

注. 这也叫行列式的倍加性质

注. 以上的性质可以结合行列式的几何意义辅助理解

例 0.1.2. 设关于 λ 的方程

$$\begin{vmatrix} \lambda-1 & -2 & 3 \\ 1 & \lambda-4 & 3 \\ -1 & a & \lambda-5 \end{vmatrix} = 0$$

有二重根，求 a 的值

解. 将行列式的第一行与第二行相减，得

$$\begin{vmatrix} \lambda-2 & 2-\lambda & 0 \\ 1 & \lambda-4 & 3 \\ -1 & a & \lambda-5 \end{vmatrix}$$

将行列式的第一列加到第二列，得

$$\begin{vmatrix} \lambda-2 & 0 & 0 \\ 1 & \lambda-3 & 3 \\ -1 & a-1 & \lambda-5 \end{vmatrix}$$

按第一行展开，有

$$\begin{aligned}(\lambda-2)\begin{vmatrix} \lambda-3 & 3 \\ a-1 & \lambda-5 \end{vmatrix} &= (\lambda-2)[(\lambda-3)(\lambda-5)-3(a-1)] \\ &= (\lambda-2)(\lambda^2-8\lambda+18-3a)=0\end{aligned}$$

若想让函数有一个二重根，则需要满足 $\lambda^2-8\lambda+18-3a$ 有二重根。

由求根公式得，$\triangle=-\frac{b^2-4ac}{2a}=(-8)^2-4(18-3a)=0$

解得 $a=\frac{2}{3}$ □

0.2 矩阵

在介绍了行列式之后，本小节主要介绍矩阵的运算性质、矩阵的逆、矩阵的秩、一些特殊的矩阵，以及矩阵的初等变换等内容。

0.2.1 矩阵的运算性质

矩阵的运算性质如下:

- 矩阵的加法: 对应项元素相加即可
- 矩阵的乘法：左边矩阵的每一行乘以右边矩阵的每一列，将求得的和写在对应位置，请注意矩阵乘法需要满足左侧矩阵的列与右侧矩阵的行相等才能进行

- 数乘矩阵：$k\boldsymbol{A}=\boldsymbol{A}k=k\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\vdots&\vdots&&\vdots\\a_{m1}&a_{m2}&\cdots&a_{mn}\end{bmatrix}=\begin{bmatrix}ka_{11}&ka_{12}&\cdots&ka_{1n}\\ka_{21}&ka_{22}&\cdots&ka_{2n}\\\vdots&\vdots&&\vdots\\ka_{m1}&ka_{m2}&\cdots&ka_{mn}\end{bmatrix}$

  注. 注意这里与前面行列式倍乘性质的区别

- 矩阵的加法满足交换律、结合律，矩阵的数乘满足分配律、交换律和结合律，**但矩阵之间的乘法只满足结合律和分配律，一般不满足交换律**
- 转置矩阵满足下面的运算律
  1. $(\boldsymbol{A}^T)^T=A$
  2. $(k\boldsymbol{A})^T=k\boldsymbol{A}^T$
  3. $(\boldsymbol{A}+\boldsymbol{B})^T=\boldsymbol{A}^T+\boldsymbol{B}^T$
  4. $(\boldsymbol{A}\boldsymbol{B})^T=\boldsymbol{B}^T\boldsymbol{A}^T\rightarrow$“穿脱原则”

0.2.2 矩阵的逆

定义 0.2.1: 逆矩阵定义

设 $\boldsymbol{A},\boldsymbol{B}$ 是 n 阶方阵，E 是 n 阶单位矩阵，若 $\boldsymbol{A}\boldsymbol{B}=\boldsymbol{B}\boldsymbol{A}=\boldsymbol{E}$，则称 $\boldsymbol{A}$ 是可逆矩阵，$\boldsymbol{B}$ 是 $\boldsymbol{A}$ 的逆矩阵，记为 $\boldsymbol{A}^{-1}$。

注. $\boldsymbol{A}$ 矩阵可逆的充要条件是 $|\boldsymbol{A}|\neq 0$。

逆矩阵的性质与重要公式

设 $\boldsymbol{A},\boldsymbol{B}$ 为同阶可逆矩阵，则

1. $(\boldsymbol{A}^{-1})^{-1}=\boldsymbol{A}$
2. 若 $k\neq 0$，则 $(k\boldsymbol{A})^{-1}=\frac{1}{k}\boldsymbol{A}^{-1}$
3. $\boldsymbol{A}\boldsymbol{B}$ 也可逆，则 $(\boldsymbol{A}\boldsymbol{B})^{-1}=\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$
4. $\boldsymbol{A}^T$ 也可逆，且 $(\boldsymbol{A}^T)^{-1}=(\boldsymbol{A}^{-1})^T$
5. $|\boldsymbol{A}^{-1}|=|\boldsymbol{A}|^{-1}$

注. $\boldsymbol{A}+\boldsymbol{B}$ 不一定可逆，且 $(\boldsymbol{A}+\boldsymbol{B})^{-1}\neq\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1}$

0.2.3 矩阵的伴随

伴随矩阵的定义

定义 0.2.2

$$A^* = \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{bmatrix}$$

其中 A_{ij} 代表代数余子式。

伴随矩阵的性质与重要公式

伴随矩阵具有下列的性质:

- 对任意 n 阶方阵 $\boldsymbol{A}$，都有伴随矩阵 $\boldsymbol{A}^*$, 且有公式

$$\boldsymbol{A}\boldsymbol{A}^* = \boldsymbol{A}^*\boldsymbol{A} = |\boldsymbol{A}|\boldsymbol{E}$$

$$|\boldsymbol{A}|^* = |\boldsymbol{A}|^{n-1}$$

当 $|\boldsymbol{A}| \neq 0$ 的时候，有

- $\boldsymbol{A}^* = |\boldsymbol{A}|\boldsymbol{A}^{-1}$
- $\boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|}\boldsymbol{A}^*$
- $\boldsymbol{A} = |\boldsymbol{A}|(\boldsymbol{A}^*)^{-1}$
- $(k\boldsymbol{A})(k\boldsymbol{A})^* = |k\boldsymbol{A}|\boldsymbol{E}$
- $\boldsymbol{A}^T(\boldsymbol{A}^T)^* = |\boldsymbol{A}^T|\boldsymbol{E}$
- $\boldsymbol{A}^{-1}(\boldsymbol{A}^{-1})^* = |\boldsymbol{A}^{-1}|\boldsymbol{E}$
- $\boldsymbol{A}^*(\boldsymbol{A}^*)^* = |\boldsymbol{A}^*|\boldsymbol{E}$
- $(\boldsymbol{A}^T)^* = (\boldsymbol{A}^*)^T$
- $(\boldsymbol{A}^{-1})^* = (\boldsymbol{A}^*)^{-1}$
- $(\boldsymbol{A}\boldsymbol{B})^* = \boldsymbol{B}^*\boldsymbol{A}^*$
- $(\boldsymbol{A}^*)^* = |\boldsymbol{A}|^{n-2}\boldsymbol{A}$

注. $(\boldsymbol{A} + \boldsymbol{B}^*) \neq \boldsymbol{A}^* + \boldsymbol{B}^*$

0.2.4 初等矩阵与初等变换

初等变换

有如下三种初等变换:

1. 一个非零常数乘矩阵的某一行或某一列 $\rightarrow$ 倍乘性质

2. 矩阵中的某两行或两列互换位置 $\rightarrow$ 互换性质

3. 将某一行或某一列的 k 倍加到另外一行或另外一列中 $\rightarrow$ 倍加性质

以上三种变换叫做**初等行变换/初等列变换**

初等矩阵

由单位矩阵经过一次初等变换得到的矩阵叫做初等矩阵以三阶矩阵为例，有如下三种初等矩阵：

1. $E_2(k) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $\boldsymbol{E}$ 的第 2 行或第 2 列乘 k 倍，称为倍乘初等矩阵

2. $E_{12} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $\boldsymbol{E}$ 的第 1,2 行或第 1,2 列互换，称为互换初等矩阵

3. $E_{31}(k) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ k & 0 & 1 \end{bmatrix}$, $\boldsymbol{E}$ 的第 1 行的 k 倍加到第 3 行，或第 3 列的 k 倍加到第 1 列，称为倍加初等矩阵

行阶梯型矩阵和行最简阶梯型矩阵

具有如下特性的矩阵称为**行阶梯形矩阵**:

- 若有零行，必位于非零行的下方
- 各个非零行左起第一个非零元素的列标从上到下是严格增大的

当每一非零行的第一个元素为 1，并且这些非零元素所在列的其他元素均为 0, 则这个矩阵就变成了**最简行阶梯型矩阵**.

用初等变换求逆矩阵

$$\left[\boldsymbol{A} \vdots \boldsymbol{E}\right] \xrightarrow{\text{初等行变换}} \left[\boldsymbol{E} \vdots \boldsymbol{A}^{-1}\right]$$

$$\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{E} \end{bmatrix} \xrightarrow{\text{初等列变换}} \begin{bmatrix} \boldsymbol{E} \\ \boldsymbol{A}^{-1} \end{bmatrix}$$

0.2.5 矩阵的秩

矩阵的秩作为后续解线性方程组、向量组的关键性质，需要认真把握

定义 0.2.3: 矩阵的秩的定义

设 $\boldsymbol{A}$ 是 $m \times n$ 阶矩阵，若存在 k 阶子式（即，任取 k 行与 k 列构成的行列式）不为零，而任意 $k+1$ 阶子式（如果有的话）全为 0，则 $r(\boldsymbol{A}) = k$, 并且若 $\boldsymbol{A}$ 为 $n \times n$ 阶矩阵，则

$$r(\boldsymbol{A})_{n\times n} = n \Leftrightarrow |\boldsymbol{A}| \neq 0 \Leftrightarrow \boldsymbol{A}\text{可逆}$$

如果用定义求矩阵的秩较为麻烦，可以将行列式化成行阶梯型矩阵，其非零行的行数即为 $\boldsymbol{A}$ 的秩.

矩阵的秩的几个重要式子

设 $\boldsymbol{A}$ 是 $m \times n$ 阶矩阵，$\boldsymbol{B}$ 是满足有关矩阵运算要求的矩阵，则:

- $0 \leq r(\boldsymbol{A}) \leq \min\{m, n\}$
- $r(k\boldsymbol{A}) = r(\boldsymbol{A})(k \neq 0)$
- $r(\boldsymbol{AB}) \leq \min\{r(\boldsymbol{A}, r(\boldsymbol{B}))\}$
- $r(\boldsymbol{A} + \boldsymbol{B}) \leq r(\boldsymbol{A}) + r(\boldsymbol{B})$
- $r(\boldsymbol{A}^*) = \begin{cases} n, & r(\boldsymbol{A} = n) \\ 1, & r(\boldsymbol{A} = n-1) \\ 0, & r(\boldsymbol{A} < n-1) \end{cases}$ 其中，$\boldsymbol{A}$ 为 $n(n \geq 2)$ 阶方阵

- 设 $\boldsymbol{A}$ 为 $m \times n$ 阶矩阵，$\boldsymbol{P,Q}$ 分别是 m,n 阶可逆矩阵，则

$$r(\boldsymbol{A}) = r(\boldsymbol{PA}) = r(\boldsymbol{AQ}) = r(\boldsymbol{PAQ})$$

- 若 $\boldsymbol{A}_{m\times n}\boldsymbol{B}_{n\times s} = \boldsymbol{O}$, 则 $r(\boldsymbol{A} + r(\boldsymbol{B}) \leq n)$
- $r(\boldsymbol{A}) = r(\boldsymbol{A}^T) = r(\boldsymbol{A}\boldsymbol{A^T}) = r(\boldsymbol{A^T}\boldsymbol{A})$

关于矩阵的秩后续更详细介绍，请看向量组、线性方程组部分。

0.3 矩阵的秩、向量组、线性方程组

向量组与线性方程组的联系十分紧密，可以说这两个概念分别一个是从向量角度讨论，一个是在方程组的角度讨论，其本质是相同的，而在其中起重要作用的因子就是上一节最后介绍的矩阵的秩。

0.3.1 向量以及向量组

在本小节中，重点回顾向量组的线性相关、线性无关的概念及其相关性质。

线性相关和线性无关

定义 0.3.1: 线性相关

对于 m 个 n 维向量 $a_1, a_2, \cdots, a_m$，若**存在**一组**不全为零**的数 $k_1, k_2, \cdots, k_m$，使得

$$k_1a_1 + k_2a_2 + \cdots + k_ma_m = 0$$

则称向量组 $a_1, a_2, \cdots, a_m$ **线性相关**.

定义 0.3.2: 线性无关

若**不存在不全为零**的数 $k_1, k_2, \cdots, k_m$，使得

$$k_1a_1 + k_2a_2 + \cdots + k_ma_m = 0$$

则称向量组 $a_1, a_2, \cdots, a_m$ **线性无关**，即，当且仅当 $k_1 = k_2 = \cdots = k_n = 0$ 的时候，才能使等式成立.

判别线性相关性的几大定理

1. 向量组 $a_1, a_2, \cdots, a_n (n \geq 2)$ 线性相关的充要条件是向量组中至少有一个向量可以有其他的 $n-1$ 个向量表示

   注. 其逆否命题为：向量组 $a_1, a_2, \cdots, a_n (n \geq 2)$ 线性无关的充要条件是向量组中的任何一个向量都不能由其余的 $n-1$ 个向量表示

2. 若向量组 $a_1, a_2, \cdots, a_n$ 线性无关，而 $\beta, a_1, a_2, \cdots, a_m$ 线性相关，则 β 可以由 $a_1, a_2, \cdots, a_n$ 线性表示，且表示法唯一

3. 如果向量组 $\beta_1, \beta_2, \cdots, \beta_t$ 可以由向量组 $a_1, a_2, \cdots, a_s$ 线性表示，且 $t > s$, 则 $\beta_1, \beta_2, \cdots, \beta_t$ 线性相关（以少表多，多的相关）

4. 如果向量组 $a_1, a_2, \cdots, a_m$ 有一部分向量线性相关，则整个向量组也线性相关

   注. 其逆否命题为：如果向量组 $a_1, a_2, \cdots, a_m$ 线性无关，则其任意一部分向量组都线性无关

5. 如果一组 n 维向量 $a_1, a_2, \cdots, a_s$ 线性无关，那么把这些向量各任意添加 m 个分量所得到的新向量 $(n+m$ 维) 组 $a_1^*, a_2^*, \cdots, a_s^*$ 也是线性无关的

   注. 其逆否命题为：如果向量组 $a_1, a_2, \cdots, a_s$ 线性相关，那么他们各去掉相同的若干分量所得到的新向量组也是线性相关的

极大线性无关组

定义 0.3.3: 极大线性无关组

在向量组 $a_1, a_2, \cdots, a_s$ 中，若存在部分组 $a_{i_1}, a_{i_2}, \cdots, a_{i_r}$ 满足：

1. $a_{i_1}, a_{i_2}, \cdots, a_{i_r}$ 线性无关
2. 向量组中的任一向量 $a_i (i = 1, 2, \cdots, s)$ 均可由 $a_{i_1}, a_{i_2}, \cdots, a_{i_r}$ 线性表示

则称向量组 $a_{i_1}, a_{i_2}, \cdots, a_{i_r}$ 是元向量组的一个**极大线性无关组**

需要注意的是，向量组的极大线性无关组一般不唯一，一个线性无关向量组的极大线性无关组就是该向量组本身

介绍完极大线性无关组后，结合前面对于矩阵的秩的相关定义，又可以建立起下面的关系：

1. 向量组 $a_1, a_2, \cdots, a_s$ 的极大线性无关组 $a_{i_1}, a_{i_2}, \cdots, a_{i_r}$ 中所含向量的个数 r 即为向量组的秩

2. $r(\boldsymbol{A}) = \boldsymbol{A}$的行秩($\boldsymbol{A}$的行向量组的秩) $= \boldsymbol{A}$的列秩 ($\boldsymbol{A}$的列向量组的秩)

介绍完了线性相关和线性无关以及极大线性无关组的相关内容之后，接下来我们引入方程组的内容。

注. 本小节没有涉及到等价向量组的相关内容，暂时认为对后续学习不重要，若后续该概念在课程学习中多次提及再来补充

0.3.2 方程组

本小节主要介绍齐次线性方程组和非齐次线性方程组的结构及其解的性质。

齐次线性方程组的结构及其解的性质

齐次方程组的结构如下：

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0, \\ \qquad\qquad\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0, \end{cases}$$

这是一个 m 个方程，n 个未知数的齐次线性方程组

如果用向量表示，则上述方程组可以表示为：

$$x_1\boldsymbol{a_1} + x_2\boldsymbol{a_2} + \cdots + x_n\boldsymbol{a_n} = 0$$

其中

$$\boldsymbol{a_j} = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}, j = 1, 2, \cdots, n$$

用矩阵表示为

$$\boldsymbol{A}_{m\times n}\boldsymbol{x} = \boldsymbol{0}$$

其中

$$\boldsymbol{A}_{m\times n}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \boldsymbol{x}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

方程组有解的条件为:

- 当 $r(\boldsymbol{A})=\boldsymbol{n}(a_1,a_2,\cdots,a_n$ 线性无关时)，方程组有**唯一零解**
- 当 $r(\boldsymbol{A})<\boldsymbol{n}(a_1,a_2,\cdots,a_n$ 线性相关时)，方程组有**非零解** (无穷多解)，且有 $n-r$ 个线性无关解

非齐次线性方程组的结构及其解的性质

非齐次线性方程组的结构如下

$$\begin{cases} a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1, \\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2, \\ \cdots \\ a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n=b_m, \end{cases}$$

这是一个 m 个方程，n 个未知数的非齐次线性方程组, 其向量形式为 $x_1\boldsymbol{a_1}+x_2\boldsymbol{a_2}+\cdots+x_n\boldsymbol{a_n}=\boldsymbol{b}$，其中

$$\boldsymbol{a}_j=\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}, j=1,2,\cdots,n, \boldsymbol{b}=\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

用矩阵表示为

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$$

其中

$$\boldsymbol{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \boldsymbol{x}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

矩阵 $\left[\begin{array}{cccc|c} a_11 & a_12 & \cdots & a_1n & b_1 \\ a_21 & a_22 & \cdots & a_2n & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_m1 & a_m2 & \cdots & a_mn & b_m \end{array}\right]$ 称为矩阵 $\boldsymbol{A}$ 的**增广矩阵**，简记为 $[\boldsymbol{A} \vdots \boldsymbol{b}]$

方程组有解的条件为:

- 若 $r(\boldsymbol{A}) \neq r([\boldsymbol{A}, \boldsymbol{b}])$ ($\boldsymbol{b}$不能由$a_1, a_2, \cdots, a_n$ 线性表示)，方程组**无解**
- 当 $r(\boldsymbol{A}) = r([\boldsymbol{A}, \boldsymbol{b}]) = \boldsymbol{n}$ ($a_1, a_2, \cdots, a_n$ 线性无关，$a_1, a_2, \cdots, a_n, \boldsymbol{b}$ 线性相关)，方程组有**唯一解** (无穷多解)
- 若 $r(\boldsymbol{A}) = r([\boldsymbol{A}, \boldsymbol{b}]) = r < n$，则方程组有**无穷多解**

非齐次线性方程组与齐次线性方程组的解的关系

非齐次线性方程组的通解 = 非齐次线性方程组的一个特解 + 齐次线性方程组的通解, 即

1. 求出方程组 $\boldsymbol{Ax} = \boldsymbol{0}$ 的通解 $k_1\xi_1 + k_2\xi_2 + \cdots + k_{n-r}\xi_{n-r}$
2. 求出 $\boldsymbol{Ax} = \boldsymbol{b}$ 的一个特解 $\boldsymbol{\eta}$
3. 则 $\boldsymbol{Ax} = \boldsymbol{b}$ 的通解为 $k_1\xi_1 + k_2\xi_2 + \cdots + k_{n-r}\xi_{n-r} + \boldsymbol{\eta}$, 其中 $k_1, k_2, \cdots, k_{n-r}$ 为任意常数

注. 本小节没有涉及基础解系、两个方程组的公共解以及同解方程组的相关内容，其中基础解系概念较为简单故不概述，后面方程组的公共解以及同解方程组因为课程中暂未涉及，故也没有介绍，若后续课程中需要用到这两部分的内容再在这里补充。

0.4 特征值与特征向量、相似理论

前面三节的内容都是为了本节的内容做铺垫的，因此在回顾本节内容之前请先确保前三节的大题框架没有问题。

0.4.1 特征值与特征向量的定义

定义 0.4.1: 特征值与特征向量

设 $\boldsymbol{A}$ 是 n 阶矩阵，λ 是一个数，若存在 n 维非零列向量 $\boldsymbol{\xi}$，使得

$$\boldsymbol{A}\boldsymbol{\xi} = \lambda\boldsymbol{\xi} \tag{1}$$

则称 λ 是 $\boldsymbol{A}$ 的**特征值**，$\boldsymbol{\xi}$ 是 $\boldsymbol{A}$ 对应于特征值 λ 的**特征向量**

0.4.2 求解特征值与特征向量的步骤

由定义中给出的式子，可以得到

$$(\lambda\boldsymbol{E} - \boldsymbol{A})\boldsymbol{\xi} = \boldsymbol{0}$$

由于 $\boldsymbol{\xi} \neq \boldsymbol{0}$，故其次方程组 $(\lambda\boldsymbol{E} - \boldsymbol{A})\boldsymbol{\xi} = \boldsymbol{0}$ 有非零解，对应行列式 $|\lambda\boldsymbol{E} - \boldsymbol{A}| = 0$

解该行列式形成的方程，可以得出特征值，之后根据接出的特征值回带入其次线性方程组中，即可求得该特征值对应的特征向量。

0.4.3 特征值、特征向量的性质与重要结论

特征值的性质与重要结论

1. λ_0 是 $\boldsymbol{A}$ 的特征值 $\Leftrightarrow |\lambda_0\boldsymbol{E} - \boldsymbol{A}| = 0$
2. λ_0 不是 $\boldsymbol{A}$ 的特征值 $\Leftrightarrow |\lambda_0\boldsymbol{E} - \boldsymbol{A}| \neq 0$
3. 若 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 是 $\boldsymbol{A}$ 的 n 个特征值，则

$$\begin{cases} |\boldsymbol{A}| = \lambda_1\lambda_2\cdots\lambda_n \\ \operatorname{tr}(\boldsymbol{A}) = \lambda_1 + \lambda_2 + \cdots + \lambda_n, \quad \operatorname{tr}(\boldsymbol{A})\text{叫做矩阵 } \boldsymbol{A} \text{ 的}\textbf{迹} \end{cases}$$

特征向量的性质与重要结论

1. $\boldsymbol{\xi}(\neq \boldsymbol{0})$ 是 $\boldsymbol{A}$ 的属于 λ_0 的特征向量 $\Leftrightarrow \boldsymbol{\xi}$ 是 $(\lambda_0\boldsymbol{E} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ 的非零解
2. k 重特征值 λ 至多只有 k 个线性无关的特征向量
3. 若 $\boldsymbol{\xi_1}, \boldsymbol{\xi_2}$ 是 $\boldsymbol{A}$ 的属于不同特征值 λ_1, λ_2 的特征向量，则 $\boldsymbol{\xi_1}, \boldsymbol{\xi_2}$ 线性无关

4. 若 $\boldsymbol{\xi_1}, \boldsymbol{\xi_2}$ 是 $\boldsymbol{A}$ 的属于同一特征值 λ 的特征向量，则非零向量 $k_1\boldsymbol{\xi_1} + k_2\boldsymbol{\xi_2}$ 仍是 $\boldsymbol{A}$ 的属于特征值 λ 的特征向量

5. 若 $\boldsymbol{\xi_1}, \boldsymbol{\xi_2}$ 是 $\boldsymbol{A}$ 的属于不同特征值 λ_1, λ_2 的特征向量，则当 $k_1 \neq 0, k_2 \neq 0$ 时，$k_1\boldsymbol{\xi_1} + k_2\boldsymbol{\xi_2}$ 不是 $\boldsymbol{A}$ 的任何特征值的特征向量

6. 设 λ_1, λ_2 是 $\boldsymbol{A}$ 的两个不同的特征值，$\boldsymbol{\xi}$ 是对应于 λ_1 的特征向量，则 $\boldsymbol{\xi}$ 不是对应于 λ_2 的特征向量（即一个特征向量不能属于两个不同的特征值）

0.4.4　矩阵的相似理论

矩阵的相似

定义 0.4.2: 矩阵的相似

设 $\boldsymbol{A}, \boldsymbol{B}$ 是两个 n 阶方阵，若存在 n 阶可逆矩阵 $\boldsymbol{P}$，使得 $\boldsymbol{P^{-1}AP} = \boldsymbol{B}$，则称 $\boldsymbol{A}$ 相似于 $\boldsymbol{B}$，记为 $\boldsymbol{A} \sim \boldsymbol{B}$

矩阵相似有如下特点：

1. **反身性**: $\boldsymbol{A} \sim \boldsymbol{A}$
2. **对称性**: 若 $\boldsymbol{A} \sim \boldsymbol{B}$, 则 $\boldsymbol{B} \sim \boldsymbol{A}$
3. **传递性**: 若 $\boldsymbol{A} \sim \boldsymbol{B}, \boldsymbol{B} \sim \boldsymbol{C}$，则 $\boldsymbol{A} \sim \boldsymbol{C}$

相似矩阵的性质

若 $\boldsymbol{A} \sim \boldsymbol{B}$，则

- $|\boldsymbol{A}| = |\boldsymbol{B}|$
- $r(\boldsymbol{A}) = r(\boldsymbol{B})$
- $\mathrm{tr}(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{B})$
- $\lambda_{\boldsymbol{A}} = \lambda_{\boldsymbol{B}}$，或 $|\lambda\boldsymbol{E} - \boldsymbol{A}| = |\lambda\boldsymbol{E} - \boldsymbol{B}|$
- $r(\lambda\boldsymbol{E} - \boldsymbol{A}) = r(\lambda\boldsymbol{E} - \boldsymbol{B})$
- $\boldsymbol{A}, \boldsymbol{B}$ 的各阶主子式之和分别相等

矩阵的相似对角化

定义 0.4.3

设 $\boldsymbol{A}$ 是 n 阶方阵，若存在 n 阶可逆矩阵 $\boldsymbol{P}$，使得 $\boldsymbol{P^{-1}AP} = \boldsymbol{\Lambda}$，其中 $\boldsymbol{\Lambda}$ 是对角矩阵，则称 $\boldsymbol{A}$ 可以相似对角化，记 $\boldsymbol{A} \sim \boldsymbol{\Lambda}$，称 $\boldsymbol{\Lambda}$ 是 $\boldsymbol{A}$ 的相似标准型

判断矩阵能否相似对角化的条件

充要条件：

- n 阶矩阵 $\boldsymbol{A}$ 可以相似对角化 $\Leftrightarrow$ $\boldsymbol{A}$ 有 n 个线性无关的特征向量
- n 阶矩阵 $\boldsymbol{A}$ 可以相似对角化 $\Leftrightarrow$ $\boldsymbol{A}$ 对应于每个 k_i 重特征值都有 k_i 个线性无关的特征向量

充分条件：

- n 阶矩阵 $\boldsymbol{A}$ 有 n 个不同的特征值 $\Rightarrow$ n 阶矩阵 $\boldsymbol{A}$ 可以相似对角化
- n 阶矩阵 $\boldsymbol{A}$ 为实对称矩阵 $\Rightarrow$ n 阶矩阵 $\boldsymbol{A}$ 可以相似对角化

0.5 结语

本章主要对于课上出现的内容涉及到线性代数的一部分知识进行了复习，其余没有出现的部分代表暂时在这门课程中没有用到，随着课程的进行，如果后续会有相关知识涉及到了这一章没有出现的内容，会在这里进行补充。

第一章　线性代数基础

本章将会正式进入矩阵理论的知识内容中，首先是第一章，这一章的名字叫做“线性代数基础”，按照课本的话来讲，这一部分不是单纯的对于线性代数知识的简单回顾与复习，而是在已经掌握线性代数的知识的基础上进行深化，同时，这一章也是后面内容的基础。

本章将会涉及到以下内容：

- 线性空间与子空间
- 空间分解与维数定理
- 特征值与特征向量
- 线性变换
- …

注意：之后的内容会随着课程的进行进行及时更新

1.1 线性空间与子空间

提到"空间"一词，很多人对这个概念应该并不陌生，我们从出生便降临在这个世界中，这个世界便是一个三维空间，我们使用计算机浏览互联网，这也可以称作一个网络空间...... 类似的例子还有很多很多。

1.1.1 线性空间

在这里我们要讨论的概念叫"线性空间"，是数学上的空间，上面提到的我们生活在的三维空间，抽象出来也属于线性空间。

线性空间的例子，除了上面提到的三维空间，在平面直角坐标系 x, y 组成的一个平面也是一个空间（二维空间），那究竟何为线性空间呢？有什么标准来判断其是否是一个空间？

判断能否组成一个空间的定义如下：

定义 1.1.1: 判断空间的定义

设 V 是一个非空集合，P 是一个数域，在集合 V 的元素之间定义加法 $v = \alpha + \beta$，定义数量乘法 $\delta = k\alpha$，如果加法与数量乘法满足下列规则：

- $\alpha + \beta = \beta + \alpha$
- $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$
- $\exists \mathbf{0} \in V, \forall \alpha \in V$, 有$\alpha + \mathbf{0} = \alpha$（存在零元素）
- $\forall \alpha \in V, \exists \beta \in V$, 使得$\alpha + \beta = \mathbf{0}$（存在负元素）
- $1\alpha = \alpha$
- $k(l\alpha) = (kl)\alpha$
- $(k + l)\alpha = k\alpha + l\alpha$
- $k(\alpha + \beta) = k\alpha + k\beta$

则 V 称为数域 P 上的**线性空间**

在这里我们需要明确下面一个概念：何为数域?

定义 1.1.2: 数域的概念

如果一个由数字构成的集合（叫做数集）P，这个数集对于加法、减法、乘法、除法（除数不为 0）封闭，则就把这个数集 P 叫做数域

这里又引出了一个新的词——封闭，何为封闭？封闭的概念较为简单：如果一个集合中的某两个数做某一运算之后的结果仍然在该集合中，那么就称该集合对于该运算是封闭的。

如果还是对于这一概念不理解，希望下面这个例子能够帮你理解：

例 1.1.1. 全体整数组成的集合 $\mathbb{Z}$ 是否是一个数域？

解. 整数集包含两大部分：**正整数和负整数**，0 是整数，但 0 既不是正数也不是负数

接下来我们来判断整数集是否对于加法封闭：

我们从小学数学的知识就可以得知，两个整数相加依旧是整数，所以整数集对于加法是封闭的

依次类推，两个整数相减，相乘，结果依旧是一个整数，所以很显然，整数集对于四则运算中的加法、减法和乘法都是封闭的

最后，整数集对于除法是否是封闭的呢？

很显然不是，举一个最简单的例子，$a=1, b=2$，a 除以 b 的结果是 $\frac{1}{2}$，它并不是一个整数，而是一个分数，或者说小数，又可以说是一个有理数，所以整数集对于除法并不是封闭的。

综上，可以断定，整数集并不是一个数域。 □

注. *定义1.1.1中的八条性质说明了什么？*

左侧四条定义了空间对于加法需要满足以下特性：加法的交换律、加法的结合律、存在零元素、存在父元素

右侧的四条定义了空间的数量乘法需要满足以下特性：数量乘法的结合律、数量乘法的分配律（分配律分为两个标量相加的分配律，以及两个向量相加的分配律）。

通常，定义1.1.1给出的八个条件即为判断一个集合是否能构成空间的依据，请看下面的例题。

例 1.1.2. 设多项式集合

$$P_n[x] = \{a_{n-1}x^{n-1} + \cdots + a_1x + a_0 | a_i \in P, i = 0, 1, \cdots, n-1\}$$

这里 P_n 代表**次数不超过 n 的多项式**, 请问 $P_n[x]$ 是否能构成一个线性空间？

解. 按照定义1.1.1的八条规律，依次来判断

首先，多项式的加法一定满足交换律和结合律（由小学数学学过的加法交换律和加法结合律就能得知），同样的，我们可以在这个多项式集合中找到一个元素 0，使得该多项式与 0

相加的结果依旧是该多项式（很明显，这个元素 0 就是数字 0，即 $a_{n-1}, a_{n-2}, \cdots, a_1, a_0$ 均为 0 的时候），此外，我们可以构造出下面的一个多项式集合，令其与 $P_n[x]$ 相加的结果为 0：

$$Q_n[x] = -P_n[x]$$

综上，该集合满足空间定义中对于加法的规律，接下来判断乘法

很显然，存在一个元素 1，使得该集合与 1 相乘的结果就是其本身（这个元素 1 就是数字 1，即 $a_{n-1} = a_{n-2} = \cdots = a_1 = 0, a_0 = 1$ 的时候）。同时，由小学数学和初中数学的知识可以知道，多项式的乘法满足数乘的结合律与分配律

综上，该多项式集合是一个线性空间。 □

注. 上面的例子大多数情况下运用了一些"显然，由……的知识可以得知"，没有具体写明如何得出的结果，有以下两点原因，第一点原因是在写这段文字的时候确实懒得打这么长的公式了，二是认为大家应该能明白上面判断的过程，所以就没有写明公式，当然写出公式也是可以的，如果后面确实不理解上面是如何"显然"得来的，会重新更新这部分的例子，用公式表明。

1.1.2 线性空间的维数

在了解了何为线性空间之后，接下来我们再来了解一下如何形容一个线性空间，维数便是形容线性空间的一个度量，我们前面一直所说的"三维空间"中的"三维"，就表明该空间的维数为 3

定义 1.1.3: 线性空间维数的定义

在线性空间 V 中，如过有 n 个向量 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 线性无关，而 V 中任意 $n+1$ 个向量线性相关，则称 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 为 V 的一组**基底**，由于线性空间的所有基底总含有相同数目的向量，则 n 称为线性空间 V 的**维数**，常记为 $\dim V = n$

上面的定义说明了，一个空间中线性无关向量的个数其实就是该空间的维数，同时，这一系列线性无关的向量便可以构成该空间的一组基，回顾第零章线性表示相关的内容，我们同样可以得知：该空间的任意向量都可以由这一组基来线性表示。

注. 请注意，向量组/线性空间的维数与向量的维数是两个不同的概念，一定要注意区分。

向量的维数：向量有几行，一般就说是向量的维数为几，如 $\boldsymbol{\alpha}=[1,2,3]^T$，那么这就是一个三维向量

向量组/线性空间的维数：向量组中线性无关向量的个数，如 $A=\begin{bmatrix}1&1&1\\2&2&2\\3&3&3\end{bmatrix}$，虽然 A 是一个 3×3 的矩阵，但是经过分析可以看出该矩阵其实只有一个线性无关的向量，故该向量组/空间是 *1* 维的，但是其中按列分块出的的三个列向量却是三维的向量。

1.2 空间分解与维数定理

在中学阶段物理课程中，我们接触过力的分解，了解过一个矢量是可以进行分解的，同样也可以进行合成，这就是平行四边形法则或者三角形法则。

在线性代数中，我们也了解过基向量的概念，也了解到了在空间中的一个向量其实就是一系列基向量合成而成，相对应的，一个空间中的向量也可以分解到各个构成他的基向量所在的空间中。

为什么需要将向量，还有后面要涉及到的空间进行分解呢？答案是因为方便研究问题。

本小节我们将目光再放远一些，了解一下空间的分解。

1.2.1 空间的和与维数定理

定义 1.2.1

设 $\boldsymbol{V}_1,\boldsymbol{V}_2$ 是线性空间 $\boldsymbol{V}$ 的子空间，则 $\boldsymbol{V}_1$ 与 $\boldsymbol{V}_2$ 的和为

$$\boldsymbol{V}_1+V_2=\{\alpha_1+\alpha_2|\forall\alpha_1\in\boldsymbol{V}_1,\forall\alpha_2\in\boldsymbol{V}_2\}$$

这条定义其实就是说两个空间的和就相当于在这两个空间里面各自任取一个向量求和，最后的结果就是空间的和。

两个空间求和之后生成的新空间，新空间的维数应该如何计算呢？这就是下面的维数定理。

定理 1.2.1: 维数定理

设 $\boldsymbol{V}_1$ 和 $\boldsymbol{V}_2$ 是线性空间 $\boldsymbol{V}$ 的子空间，则

$$\dim(\boldsymbol{V}_1) + \dim(\boldsymbol{V}_2) = \dim(\boldsymbol{V}_1 + \boldsymbol{V}_2) + \dim(\boldsymbol{V}_1 \cap \boldsymbol{V}_2)$$

维数定理表明，两个空间求和之后，其维数应该等于各自的维数减掉公共部分的维数，这很好理解，相当于公共部分的维数被多加了一次，这类似于两个集合相加。

说完了维数定理，接下来再来说一个新的概念——直和：

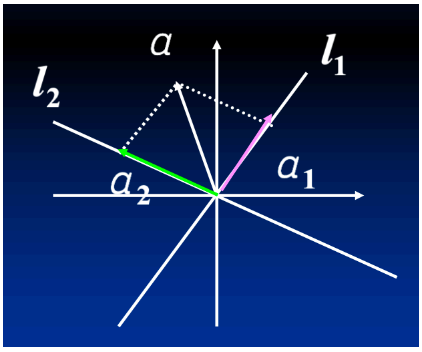
定义 1.2.2

设 $\boldsymbol{V}_1$ 和 $\boldsymbol{V}_2$ 是线性空间 $\boldsymbol{V}$ 的子空间，弱队 $\forall\alpha \in V_1 + V_2$，有

$$\alpha = \alpha_1 + \alpha_2\ (\alpha_1 \in \boldsymbol{V}_1, \alpha_2 \in \boldsymbol{V}_2)$$

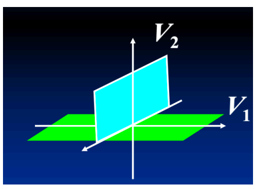
且是唯一的，这个和 $\boldsymbol{V}_1 + \boldsymbol{V}_2$ 就被称为直和，记为 $\boldsymbol{V}_1 \oplus \boldsymbol{V}_2$

如何理解直和呢？我们来看一个二维平面上的例子：



在这个例子中，向量 α 沿着两条直线 l_1, l_2 分解成了两个部分，这个时候，假设 l_1 是固定的，那么 l_2 就一定固定且唯一（可以画一个三角形，两条边都确定了那第三边一定也是确定的），那么这个时候就可以说 $l_1 + l_2$ 构成的二维空间是直和，二维空间的分解具有唯一性。

那事情到了三维空间，还是一样的吗？



在三维空间中，假设有一个向量 α，在 $\boldsymbol{V}_1$ 中的分量，那他在 $\boldsymbol{V}_2$ 中的分量会有无数种，

这就是说，三维空间的分解不具有唯一性。

如果只利用直和的定义来判断直和是否成立有些太过麻烦，下面的定理能够方便我们判断直和：

定理 1.2.2

设 $\boldsymbol{V}_1, \boldsymbol{V}_2$ 是线性空间 $\boldsymbol{V}$ 的子空间，则下列命题等价：

1. $\boldsymbol{V}_1 + \boldsymbol{V}_2$ 是直和
2. 零向量的表示方法唯一
3. $\boldsymbol{V}_1 \cap \boldsymbol{V}_2 = \{\mathbf{0}\}$

也就是说，如果两个空间没有相同的部分，那它们的和一定是直和，分解出的基向量也一定没有相同的，维数也就等于两个空间各自维数之和。这就类似于两个互不相交的集合，他们的和可以类比为空间的"直和"。

前面所讲述的是两个空间内的直和，接下来给出一般情况下的直和及其性质：

定义 1.2.3

设 $\boldsymbol{V}_1, \boldsymbol{V}_2, \cdots, \boldsymbol{V}_s$ 是线性空间 $\boldsymbol{V}$ 的子空间，如果和 $\boldsymbol{V}_1 + \boldsymbol{V}_2 + \cdots + \boldsymbol{V}_s$ 中的每个向量 α 的分解式

$$\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_s, \alpha_i \in \boldsymbol{V}_i \quad (i = 1, 2, \cdots, s)$$

是唯一的，这个和 $\boldsymbol{V}_1 + \boldsymbol{V}_2 + \cdots + \boldsymbol{V}_s$ 就称为直和，记为

$$\boldsymbol{V}_1 \oplus \boldsymbol{V}_2 \oplus \cdots \oplus \boldsymbol{V}_s$$

定理 1.2.3

设 $\boldsymbol{V}_1, \boldsymbol{V}_2, \cdots, \boldsymbol{V}_s$ 是线性空间 $\boldsymbol{V}$ 的子空间，则下列命题相互等价：

1. $\boldsymbol{W} = \boldsymbol{V}_1 + \boldsymbol{V}_2 + \cdots + \boldsymbol{V}_s$ 是直和
2. 零向量的表示方法唯一
3. $\boldsymbol{V}_i \cap (\sum_{j \neq i} \boldsymbol{V}_j) = \{\mathbf{0}\}$
4. $\dim(\boldsymbol{W}) = \sum \dim(\boldsymbol{V}_i)$

1.3 特征值与特征向量

本节课的内容虽然叫做“特征值与特征向量”，但内容并不只是简单的线性代数中介绍的特征值与特征向量，我们要更加深入理解特征值与特征向量这两个概念。

1.3.1 内容回顾

特征值与特征向量

先来回顾一下线性代数中学到的特征值与特征向量：

定义 1.3.1

设 $\boldsymbol{A} \in \mathbb{P}^{n\times n}, \alpha \in \mathbb{P}^n, \lambda \in \mathbb{P}$，若

$$\boldsymbol{A}\alpha = \lambda\alpha \ (\alpha \neq 0)$$

则称 λ 为 $\boldsymbol{A}$ 的一个**特征值**，α 为 $\boldsymbol{A}$ 对应于 λ 的一个**特征向量**。

特征向量具有下面的性质：

1. 设 $\boldsymbol{A}\alpha = \lambda\alpha \ (\alpha \neq 0)$，则

$$\boldsymbol{A}(k\alpha) = k(\boldsymbol{A}\alpha) = k(\lambda\alpha) = \lambda(k\alpha)$$

2. 设 $\boldsymbol{A}\alpha_i = \lambda\alpha_i \ (i = 1, 2, \cdots, s)$，则

$$\boldsymbol{A}(k_1\alpha_1 + k_2\alpha_2 + \cdots + k_s\alpha_s) = \lambda(k_1\alpha_1 + k_2\alpha_2 + \cdots + k_s\alpha_s)$$

特征子空间

对于所有的特征向量 α，设 $\boldsymbol{V}_\lambda = \{\alpha | \boldsymbol{A}\alpha = \lambda\alpha, \alpha \in \mathbb{R}^n\}$，由特征值和特征向量的性质可知：

$$\forall \alpha, \beta \in \boldsymbol{V}_\lambda, \alpha + \beta \in \boldsymbol{V}_\lambda \ \text{(加法封闭)}$$

$$\forall \alpha \in \boldsymbol{V}_\lambda, k \in \mathbb{R}, k\alpha \in \boldsymbol{V}_\lambda \ \text{(数乘封闭)}$$

因此，可以把 $\boldsymbol{V}_\lambda$ 称作一个空间，这个空间叫做 $\boldsymbol{A}$ 的**特征子空间**。

1.3.2 谱、几何重数和代数重数

回顾了已经学过的知识之后，接下来就是新内容了，首先介绍几个基本概念

定义 1.3.2

$\boldsymbol{A}$ 的所有特征值的全体称为 $\boldsymbol{A}$ 的**谱**，记为 $\lambda(\boldsymbol{A})$。

定义 1.3.3

我们把

$$|\lambda\boldsymbol{E}-\boldsymbol{A}|=(\lambda-\lambda_1)^{n_1}\cdots(\lambda-\lambda_r)^{n_r}$$

叫做 $\boldsymbol{A}$ 的特征多项式，$\sum\limits_{i=1}^{r}n_i=n$，这里 n_i 叫做特征值 λ_i 的**代数重数**。
如果

$$\operatorname{rank}(\lambda_i\boldsymbol{E}-\boldsymbol{A})=n-m_i$$

m_i 叫做 λ_i 的**几何重数**。

几何重数与代数重数与后面 Jordan 标准型的内容紧密相关，代数重数就是特征方程中，该特征值的次数，几何重数与该特征值对应特征向量的特征矩阵的秩紧密相关。

1.3.3 对角化与 Jordan 标准型

在线性代数中，我们学过对一个矩阵进行相似对角化，了解了其步骤，并且了解能对一个矩阵进行相似对角化的条件。可现实中，能够对角化的矩阵毕竟是少数，将对角化的概念一般化，就是马上就要介绍的 Jordan 标准型。

Jordan 块与 Jordan 标准型

定理 1.3.1

设 $\boldsymbol{A} \subseteq \mathbb{C}^{n\times n}$ 有 r 个不同的特征值：$\lambda_1,\lambda_2,\cdots,\lambda_r$，其代数重数分别为 $n_1,n_2,\cdots,n_r$，则必存在可逆矩阵 $\boldsymbol{P}\in\mathbb{C}^{n\times n}$，使得

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{J}=diag(\boldsymbol{J}_1(\lambda_1),\cdots,\boldsymbol{J}_r(\lambda_r))$$

其中，矩阵 $\boldsymbol{J}$ 叫做 $\boldsymbol{A}$ 的 Jordan 标准型，$\boldsymbol{J}_1(\lambda_1),\cdots,\boldsymbol{J}_r(\lambda_r)$，叫做一个 Jordan 块，其结构为

$$\boldsymbol{J}_i(\lambda_i)=\begin{bmatrix}\lambda_i & 1 & & \\ & \lambda_i & 1 & \\ & & \ddots & 1 \\ & & & \lambda_i\end{bmatrix}$$

Jordan 标准型有下面的几个结论（**一定要记住**）：

1. Jordan 块的个数 k 是线性无关特征向量的个数
2. 当且仅当 $k=n$ 的时候矩阵可以对角化
3. 一个已知特征值的 Jordan 块个数是该特征值的几何重数，同时也是相应特征子空间的维数，一个已知特征值的 Jordan 块的阶数之和是该特征值的代数重数。
4. 特征值的几何重数小于等于其代数重数
5. 矩阵不同特征值对应的特征向量线性无关

那 Jordan 标准型和矩阵对角化有什么样的关系呢？看下面的定理：

定理 1.3.2

设 $\boldsymbol{A}\subseteq\mathbb{C}^{n\times n}$，则下列命题等价：

- $\boldsymbol{A}$ 是可对角化矩阵
- $\mathbb{C}^n$ 存在由 $\boldsymbol{A}$ 的特征向量构成的一组基
- $\boldsymbol{A}$ 的 Jordan 标准型中的 Jordan 块都是一阶的
- $m_i=n_i$ $(i=1,2,\cdots,r)$（代数重数 = 几何重数）

1.3.4 特征值与特征向量的几何性质

从线性变换开始说起

我们将线性空间到自身的映射称之为**变换**，如果在空间 $\boldsymbol{V}$ 中存在变换 T，并且满足

1. $\forall \alpha, \beta \in \boldsymbol{V}, T(\alpha+\beta)=T(\alpha)+T(\beta)$

2. $\forall k \in \mathbb{P}, T(k\alpha)=kT(\alpha)$

那么就成 T 是 $\boldsymbol{V}$ 的**线性变换**。

线性变换的特征值

从纯数值上理解了特征值之后，接下来换一个角度，从线性变换的角度理解特征值与特征向量：

定义 1.3.4

设 T 是线性空间 $V_n(C)$ 的一个线性变换，如果存在 $\lambda \in \mathbb{C}$ 和非零向量 $\xi \in \boldsymbol{V}_n(C)$，使得 $T\xi=\lambda\xi$，则 λ 叫做 T 的特征值，ξ 叫做 T 的属于特征值 λ 的特征向量。

那这个线性变换的 T 与最开始从数值角度分析的特征值与特征向量的矩阵 $\boldsymbol{A}$ 有何关系呢？下面推导一下。

线性变换与矩阵

我们假设 $\boldsymbol{V}$ 是 n 维空间，$\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 为该空间下的基，T 为 $\boldsymbol{V}$ 上的线性变换，则有

$$T\varepsilon_1=\sum_{i=1}^{n} a_{i1}\varepsilon_i,\ T\varepsilon_2=\sum_{i=1}^{n} a_{i2}\varepsilon_2,\ \cdots,\ T\varepsilon_n=\sum_{i=1}^{n} a_{in}\varepsilon_n$$

则有

$$\begin{aligned} T(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n) &= (T\varepsilon_1, T\varepsilon_2, \cdots, T\varepsilon_n) \\ &= (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \\ &= (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)\boldsymbol{A} \end{aligned}$$

这里的矩阵 $\boldsymbol{A}$ 称为线性变换 T 在基 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 下的**矩阵**。

由此，我们从向量方程 $T\xi = \lambda\xi$ 推出了矩阵方程 $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$，反之，给定一个矩阵 $\boldsymbol{A}$，也可以得出，在基底 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 下存在唯一的线性变换 T 与之对应。

1.4 初等矩阵、酉矩阵和酉变换

“初等矩阵”这个概念，我们在线性代数中学到过，可在这里，“初等矩阵”说的是：

定义 1.4.1

设 $u, v \in \mathbb{C}^n, \sigma \in \mathbb{C}$，称

$$\boldsymbol{E}(u, v, \sigma) = \boldsymbol{E} - \sigma\boldsymbol{u}\boldsymbol{v}^H$$

为初等矩阵

在这里，我们把线性代数中的“初等矩阵”更名为“初等变换矩阵”。

- $\boldsymbol{E}_{ij} = \boldsymbol{E} - (e_i - e_j)(e_i - e_j)^T = \boldsymbol{E}(e_i - e_j, e_i - e_j, 1)$
- $\boldsymbol{E}_{ij}(k) = \boldsymbol{E} + ke_je_i^T = \boldsymbol{E}(e_j, e_i, -k)$
- $\boldsymbol{E}_i(k) = \boldsymbol{E} - (1-k)e_ie_i^T = \boldsymbol{E}(e_i, e_i, 1-k)$

1.4.1 初等矩阵的相关性质

接下来介绍初等矩阵的相关性质，具体的证明过程都省略掉了（不想写证明过程了），感兴趣的可以翻看课本。

初等矩阵的特征向量

- 当 $u \perp v$ 时，设 $u_1, u_2, \cdots, u_{n-1}$ 是与 u 垂直的空间（正交补子空间）的一组基，他们也是 $\boldsymbol{E}(u, v, \sigma)$ 的 $n-1$ 个线性无关的特征向量。
- 若 u 不与 v 垂直时，设 $u_1, u_2, \cdots, u_{n-1}$ 是与 u 垂直的空间（正交补子空间）的一组基，则 $u, u_!, \cdots, u_{n-1}$ 是 $\boldsymbol{E}$ 的 n 个无关特征向量。

初等矩阵的特征值

$$\lambda(\boldsymbol{E}(u, v, \sigma)) = \{1, 1, \cdots, 1, 1 - \sigma v^H u\}$$

初等矩阵的行列式

$$\det(\boldsymbol{E}(u,v,\sigma)) = 1-\sigma v^H u$$

初等矩阵的逆

$$\boldsymbol{E}(u,v,\sigma)^{-1} = \boldsymbol{E}(u,v,\frac{\sigma}{\sigma v^H u-1}),\ (1-\sigma v^H u \neq 0)$$

其他性质

非零向量 $a,b,\in\mathbb{C}^n$，存在 u,v,σ，使得

$$\boldsymbol{E}(u,v,\sigma)a = b,\ (\sigma u = \frac{a-b}{v^H a})$$

1.4.2 初等下三角阵

定义 1.4.2

$$L_i(l_i) = \boldsymbol{E}(l_i,e_i,1) = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & -l_{i+1,i} & \ddots & & \\ & & \vdots & & 1 & \\ & & -l_{n,i} & & & 1 \end{bmatrix}$$

其中

$$\begin{cases} l_i = (0,\cdots,0,l_{i+1},\cdots,l_{n,i})^T \\ e_i = (0,\cdots,0,1,0,\cdots,0)^T \end{cases}$$

$$l_{k,i} = \frac{a_{k,i}}{a_{i,i}}(k=i+1,\cdots,n), \boldsymbol{E}(l_i,e_i,1)\boldsymbol{A} = \tilde{\boldsymbol{A}} \Rightarrow \tilde{a}_{k,i} = 0\ (k=i+1,\cdots,n)$$

1.4.3 初等酉阵 (Householder 变换)

定义 1.4.3

$$\boldsymbol{H}(u)=\boldsymbol{E}(u,u;2)=\boldsymbol{E}-2uu^H,\ (u^Hu=1)$$

1. $\boldsymbol{H}(u)^H=\boldsymbol{H}(u)-\boldsymbol{H}(u)^{-1}$
2. $\boldsymbol{H}(u)(a+ru)=a-ru,\ \forall a\in u^{\perp},r\in\mathbb{C}$（镜像变换）

1.5 欧式空间上的度量

本节只介绍相关的定理和性质，不涉及具体证明，本节内容无非就是将之前所学过的熟知的一些性质放到向量空间中罢了。

定义 1.5.1

在线性空间 $\boldsymbol{V}_n(\boldsymbol{P})$ 上，若映射

$$\boldsymbol{V}_n(\boldsymbol{P})\times\boldsymbol{V}_n(\boldsymbol{P})\to\boldsymbol{P}$$

1. $<\boldsymbol{x},\boldsymbol{x}>\geq 0;<\boldsymbol{x},\boldsymbol{x}>=0$ 当且仅当 $\boldsymbol{x}=0$
2. $<\boldsymbol{x},\boldsymbol{y}>=\overline{<\boldsymbol{y},\boldsymbol{x}>},\forall\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{V}_n(\mathbb{P})$
3. $<\lambda\boldsymbol{x},\boldsymbol{y}>=\bar{\lambda}<\boldsymbol{x},\boldsymbol{y}>,\forall\lambda\in\mathbb{P},\forall\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{V}_n(\mathbb{P})$
4. $<\boldsymbol{x}+\boldsymbol{y},\boldsymbol{z}>=<\boldsymbol{x},\boldsymbol{z}>+<\boldsymbol{y},\boldsymbol{z}>,\ \forall\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\in\boldsymbol{V}_n(\mathbb{P})$

则称 $<\boldsymbol{x},\boldsymbol{y}>$ 是 $\boldsymbol{V}_n(\mathbb{P})$ 上的内积

定义 1.5.2: 向量长度

$$\|\boldsymbol{x}\|=\sqrt{<\boldsymbol{x},\boldsymbol{x}>}$$

定义 1.5.3: 向量之间的距离

$$\|\boldsymbol{x}-\boldsymbol{y}\|$$

定义 1.5.4: 向量正交

$$< \boldsymbol{x}, \boldsymbol{y} >= 0$$

定义 1.5.5: 勾股定理

$$\boldsymbol{x} \perp \boldsymbol{y} \rightarrow \|\boldsymbol{x}+\boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2$$

定理 1.5.1: 垂线段最短定理

欧式空间 $\boldsymbol{V}_n(R)$ 中的一个固定向量和一个子空间中各向量的距离"垂线最短"

定义 1.5.6: Gram 行列式

$$\boldsymbol{G}(\alpha_1, \cdots, \alpha_k) = \begin{vmatrix} < \alpha_1, \alpha_1 > & < \alpha_1, \alpha_2 > & \cdots & < \alpha_1, \alpha_k > \\ < \alpha_2, \alpha_1 > & < \alpha_2, \alpha_2 > & \cdots & < \alpha_2, \alpha_k > \\ \vdots & \vdots & & \vdots \\ < \alpha_k, \alpha_1 > & < \alpha_k, \alpha_2 > & \cdots & < \alpha_k, \alpha_k > \end{vmatrix}$$

Gram 行列式的性质

如果 n 维欧式空间 $\boldsymbol{V}$ 中向量组 $\alpha_1, \alpha_2, \cdots, \alpha_k$ 线性无关，则将向量组 $\alpha_1, \alpha_2, \cdots, \alpha_k$ 正交化后，它的 Gram 行列式不变，即

$$\boldsymbol{G}(\alpha_1, \alpha_2, \cdots, \alpha_k) = \boldsymbol{G}(\beta_1, \beta_2, \cdots, \beta_k)$$

定理 1.5.2

设 $\alpha_1, \alpha_2, \cdots, \alpha_k$ 是线性子空间 $\boldsymbol{V}_1$ 的一组基，则向量 α 给出的点到线性流形 $\boldsymbol{P} = \alpha_0 + \boldsymbol{V}_1$ 的距离 d 为

$$d^2 = \frac{\boldsymbol{G}(\alpha_1, \cdots, \alpha_k, \alpha - \alpha_0)}{\boldsymbol{G}(\alpha_1, \cdots, \alpha_k)}$$

1.6 酉空间的分解与投影

1.6.1 何为酉空间?

引用网上的一句话[1]，可以十分简洁的回答出何为酉空间:

欧几里得空间的定义：定义了内积的实线性空间。

欧几里得空间定义推广到复线性空间就是酉空间。

1.6.2 何为投影?

投影这个概念并不陌生，在中学阶段就接触过投影，但那个时候接触到的投影大多数都是将其分解到 x 轴或者 y 轴这样，这其实是后面要介绍的一种比较特殊的投影——正交投影，其实，就如同分解一样，投影也可以不必正交。

何为投影呢？用下面这个公式可以简洁说明：

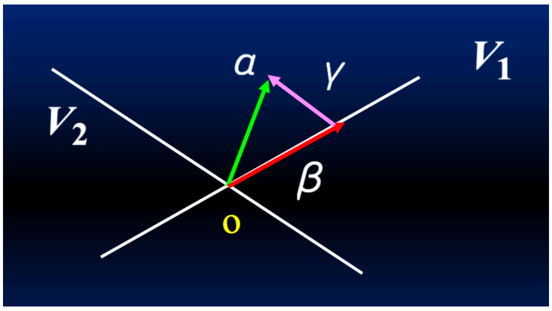
假设有变换 T，又有

$$T(\alpha) = T(\beta + \gamma) = \beta$$

那么

$$T \circ T(\alpha) = T^2(\alpha) = T[T(\beta + \gamma)] = T(\beta) = \beta = T(\alpha)$$

如果只写公式，这并不好理解，所以请看下面这幅图：



图中，向量 α 可以表示为 β 与 γ 的和，结合定义，变换 T 成功实现了将平面上的点沿着 $\boldsymbol{V}_2$ 投影到了 $\boldsymbol{V}_1$ 中，并把 $\boldsymbol{V}_1$ 上的点变为 $\boldsymbol{V}_1$ 上的同一点，因此，T 的作用就可以把空间中的点限制到 $\boldsymbol{V}_1$ 坐标轴中，这是一种恒等变换，具有 $T^2 = T$ 的性质，这就是投影。

定义 1.6.1

若变换满足 $T^2 = T \Leftrightarrow T$ 是 $\boldsymbol{V}_n(\mathbb{C})$ 上的投影

[1]引用链接：https://zhuanlan.zhihu.com/p/29312940(虽然这篇文章里也是引的，但是实在懒得找源头了)

投影满足下面的性质，这里只给出性质不给出证明，感兴趣的可以翻书看如何证明的

1. 设 T 是 $\boldsymbol{V}_n(\mathbb{C})$ 上的投影，则

$$\boldsymbol{V}_n(\mathbb{C}) = \boldsymbol{R}(T) \oplus \boldsymbol{N}(T)$$

   这里 $\boldsymbol{R}(T)$ 叫做值域，$\boldsymbol{N}(T)$ 叫做核。

2. 设 $\boldsymbol{V}_n(\mathbb{C}) = \boldsymbol{V}_1 \oplus \boldsymbol{V}_2$，则存在投影 T, 使得

$$R(T) = \boldsymbol{V}_1, N(T) = \boldsymbol{V}_2$$

显然，上面两条性质互为逆。

1.6.3 正交投影

最后要介绍的就是正交投影了，这里就简单说一下

定义 1.6.2

设 T 是投影，若

$$R^{\perp}(T) = N(T)$$

则 T 是正交投影

正交投影具有下面的性质:

定理 1.6.1

设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}, \boldsymbol{A}^2 = \boldsymbol{A}$，则

$$\boldsymbol{A}\text{是正交投影} \Leftrightarrow \boldsymbol{A}^H = \boldsymbol{A}$$

第二章　向量与矩阵的范数

本章将会介绍范数这一概念，主要内容包括介绍向量的范数、矩阵的范数和算子范数，同时在本章的最后会介绍范数的相关应用——借助范数来判断一个矩阵的逆矩阵或求解线性方程组因摄动带来的误差是否处于一个可接受的范围内。

为什么要引入范数？在本章中应时刻牢记范数是一个数而非其他之物，将矩阵、向量类的抽象概念转换成一个“大小”往往能够简化问题。

本章将会涉及到以下内容：

- 向量的范数
- 矩阵的范数
- 算子范数
- 范数的应用

注意：本章内的内容具有强连续性，需要按照顺序从头开始了解，同时，本章内容会涉及多个定理、性质的证明，需要静下心来仔细复盘证明方法，但本章内容与第一章的联系不大，因此如果觉得第一章理解起来有难度可以先跳过第一章的内容继续学习第二章。

2.1 向量的范数

本节我们首先介绍向量的范数，在开始详细介绍向量的范数前我们可以先简单“感受”一下何为范数。

2.1.1 范数的直观感受

在第一章，我们了解过向量的长度：

$$\|\boldsymbol{x}\| = \sqrt{<\boldsymbol{x}, \boldsymbol{x}>} = \boldsymbol{x}^H\boldsymbol{x}$$

有了向量的长度，向量便在实际空间内有了意义，也给后续定义距离等概念打下了基础。

其实，向量的距离就是向量的 2-范数，范数的种类有很多，每一种范数的计算方法都不一样但大同小异，这里利用向量 2-范数作说明是因为向量的 2-范数便于理解（距离），同时也引出下面的对范数的意义的概括：这便是范数的具体意义——对“距离”概念的延伸，是一种广义的“长度”。

为什么说是广义的“长度”呢？继续阅读就会发现，向量范数并不止向量的 2-范数，还有其他类型的向量范数。

定义 2.1.1: 向量范数的定义

设映射 $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$ 满足：

1. 正定条件：$\|\boldsymbol{x}\| \geq 0$, 当且仅当 $\boldsymbol{x} = \boldsymbol{0}$ 时 $|\boldsymbol{x}| = 0$
2. 齐次条件：$\|\lambda\boldsymbol{x}\| = |\lambda|\|\boldsymbol{x}\|, \lambda \in \mathbb{C}, \boldsymbol{x} \in \mathbb{C}^n$
3. 三角不等式：$\|\boldsymbol{x} + \boldsymbol{y}\| \leq +\|\boldsymbol{y}\|, \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$

则称映射 $\|\cdot\|$ 为 $\mathbb{C}^n$ 上向量 $\boldsymbol{x}$ 的范数，定义了范数的 $\mathbb{C}^n$ 又叫做一个线性赋范空间。

定义2.1.1给出了能够成为向量范数的条件：是一个映射，并且能够满足三个条件，就可以将这种映射叫做向量的范数，这也是今后判断给出一种映射是否为向量范数的方法。

2.1.2 向量范数的性质

定义2.1.1中给出的能够成为向量范数的三个条件也是向量范数的三个性质，除此之外，根据上面的定义，我们还可以得出向量范数的下列性质：

1. 零向量的范数是 0

   这一条性质不难理解，直接使用向量范数的正定条件即可得出。

2. 当 $\boldsymbol{x} \neq \mathbf{0}$ 时，$\|\frac{1}{\|\boldsymbol{x}\|}\boldsymbol{x}\| = 1$

   这一条性质通过向量范数的齐次条件可以得出：向量范数是一个数，故 $\|\frac{1}{\|\boldsymbol{x}\|}\boldsymbol{x}\| = \frac{1}{\|\boldsymbol{x}\|}\|\boldsymbol{x}\| = 1$

3. 对任意 $\boldsymbol{x} \in \mathbb{C}^n$，有 $\|-\boldsymbol{x}\| = \|\boldsymbol{x}\|$

   同样地，利用齐次性依旧可以得到这一条性质：$\|-\boldsymbol{x}\| = |-1| \cdot \|\boldsymbol{x}\| = \|\boldsymbol{x}\|$

4. 对任意 $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$，有 $|\|\boldsymbol{x}\| - \|\boldsymbol{y}\|| \leq \|\boldsymbol{x} - \boldsymbol{y}\|$

   上述不等式的证明过程如下：

   证明.

   $$\|\boldsymbol{x}\| = \|\boldsymbol{x} - \boldsymbol{y} + \boldsymbol{y}\| \leq \|\boldsymbol{x} - \boldsymbol{y}\| + \|\boldsymbol{y}\| (\text{范数的三角不等式})$$

   从而

   $$\|\boldsymbol{x} - \boldsymbol{y}\| = \|\boldsymbol{y} - \boldsymbol{x}\| \geq \|\boldsymbol{y}\| - \|\boldsymbol{x}\| (\text{依旧是范数的三角不等式})$$

   那么

   $$\|\boldsymbol{x}\| - \|\boldsymbol{y}\| \geq -\|\boldsymbol{x} - \boldsymbol{y}\| (\text{不等式左右两边同时乘-1，不等式方向改变})$$

   于是就有了

   $$|\|\boldsymbol{x}\| - \|\boldsymbol{y}\|| \leq \|\boldsymbol{x} - \boldsymbol{y}\|$$

   证毕 □

2.1.3 对向量范数三大性质的理解

在2.1.2节中列出的向量范数的性质，那如何理解向量的范数三大性质呢？

对于正定性的理解：可以从 2-范数来看，一般来说，我们认为 2-范数表示的是向量的距离，即

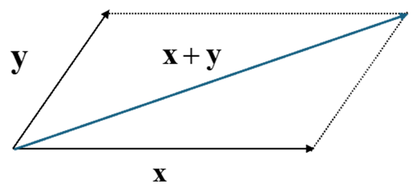
$$||\boldsymbol{x}||_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

(这里 $x_1, x_2, \cdots, x_n$ 代表向量 $\boldsymbol{x}$ 的各个分量), 在后面介绍到 p-范数的时候也可以从其式子来看出非负性。

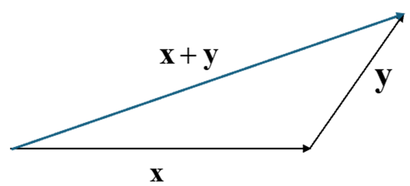
对于齐次性的理解：仍以二维空间上的 2-范数为例，对于向量 $\boldsymbol{x}, \lambda\boldsymbol{x}$ 我们可以将其理解为对该向量进行伸缩变换，当 λ 为正时，新的伸缩变换之后的向量与原向量方向相同，当 λ

为负时，新的伸缩变换之后的向量与原向量方向相反，由于 2-范数描述的为距离，所以 $|\lambda|||x||$ 就代表将向量 $\boldsymbol{x}$ 的距离乘以缩放因子的绝对值（因为范数是个数，所以忽略掉正负带来的影响，因此加绝对值）。

对于三角不等式的理解：依旧以二维空间上的 2-范数为例，依据向量相加的平行四边形法则：我们可以画出 $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{x}+\boldsymbol{y}$ 的关系：



如果我们把向量 $\boldsymbol{y}$ 平移一下，就会得到下面这幅图：



这就构成了一个三角形，依据三角形两边之和大于第三边的性质，从而可以轻易理解范数性质 3——三角不等式的含义。

2.1.4 常见的向量范数

本节介绍整个向量范数的重点——常见的向量范数经常使用的向量范数有下面三个：

$$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

$$\|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}$$

$$\|\boldsymbol{x}\|_\infty = \max_{1\le i\le n} |x_i|$$

他们分别为向量 1-范数，向量 2-范数和向量无穷范数，接下来将会一一给出这三种范数是向量范数的证明。

向量 1-范数的证明与 Minkovski 不等式

正定性证明:

若向量 $\boldsymbol{x} \neq 0$，则表明 $\boldsymbol{x}$ 至少有一个分量 $x_k \neq 0$，从而 $|x_k| > 0$, 故 $\|\boldsymbol{x}\|_1 > 0$

当 $\boldsymbol{x} = \boldsymbol{0}$ 时，显然向量 $\boldsymbol{x}$ 的所有分量必为 0，故 1-范数一定为 0，正定性证毕。

齐次性证明:

$$\begin{aligned}\|\lambda \boldsymbol{x}\|_1 &= |\lambda x_1| + |\lambda x_2| + \cdots + |\lambda x_n| \\ &= |\lambda||x_1| + |\lambda||x_2| + \cdots + |\lambda||x_n| \\ &= |\lambda|(|x_1| + |x_2| + \cdots + |x_n|) \\ &= |\lambda|\|\boldsymbol{x}\|_1\end{aligned}$$

齐次性证毕。

三角不等式证明:

$$\|\boldsymbol{x} + \boldsymbol{y}\|_1 = \sum_{i=1}^{n} |x_i + y_i| \tag{1}$$

由 $|a + b| \leq |a| + |b|$

式 1 可以变成

$$\sum_{i=1}^{n} |x_i + y_i| \leq \sum_{i=1}^{n} |x_i| + \sum_{i=1}^{n} |y_i| = \|\boldsymbol{x}\|_1 + \|\boldsymbol{y}\|_1$$

即

$$\|\boldsymbol{x} + \boldsymbol{y}\|_1 \leq \|\boldsymbol{x}\|_1 + \|\boldsymbol{y}\|_1$$

这个不等式也叫 Minkovski 不等式。

三角不等式证毕，因此，1-范数是向量范数。

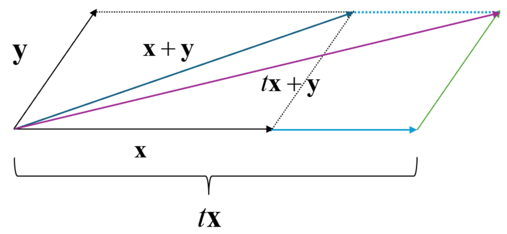
向量 2-范数的证明与 Cauchy 不等式 (柯西不等式)

在正式介绍向量 2-范数之前，需要了解一下关于 Cauchy 不等式的有关内容。

Cauchy 不等式的内容为:

$$| < x, y > | \leq \|x\|_2 \cdot \|y\|_2$$

即，向量 $\boldsymbol{x}$ 与向量 $\boldsymbol{y}$ 的内积 ($< \cdot, \cdot >$ 表示两个向量的内积) 的绝对值小于等于两个向量的 2-范数。



下面给出在二维欧式空间里 Cauchy 不等式的证明：

如图所示，假设 $t>0$，向量 $\boldsymbol{x},\boldsymbol{y},\boldsymbol{x}+\boldsymbol{y}$ 和向量 $t\boldsymbol{x},y,t\boldsymbol{x}+\boldsymbol{y}$ 的情况如图，接下来针对向量 $\boldsymbol{x}$ 与向量 $\boldsymbol{y}$ 线性相关与线性无关时分类讨论：

当 $\boldsymbol{x}$ 与向量 $\boldsymbol{y}$ 线性无关时，两个向量的关系可以由上图表示，此时两向量不共线，可由反证法证明得 $t\boldsymbol{x}+\boldsymbol{y}\neq\mathbf{0}$。

所以

$$
\begin{aligned}
|t\boldsymbol{x}+\boldsymbol{y}|^2 &=< t\boldsymbol{x}+\boldsymbol{y}, t\boldsymbol{x}+\boldsymbol{y} > \\
&=<\boldsymbol{x},\boldsymbol{x}>t^2+2<\boldsymbol{x},\boldsymbol{y}>t+<\boldsymbol{y},\boldsymbol{y}>
\end{aligned}
$$

这是一个可以看做自变量为 t 的二次函数，令 $f(t)=<\boldsymbol{x},\boldsymbol{x}>t^2+2<\boldsymbol{x},\boldsymbol{y}>t+<\boldsymbol{y},\boldsymbol{y}>$。

因为 $t\boldsymbol{x}$ 与 $\boldsymbol{y}$ 线性无关，所以 $t\boldsymbol{x}+\boldsymbol{y}\neq\mathbf{0}$，故 $|t\boldsymbol{x}+\boldsymbol{y}|^2$ 一定大于 0，所以 $f(t)>0$。

由于 $f(t)$ 是一个二次函数，二次函数值恒大于零，代表二次函数与 x 轴没有交点，意味着 $\triangle=4<\boldsymbol{x},\boldsymbol{y}>-4<\boldsymbol{x},\boldsymbol{x}><\boldsymbol{y},\boldsymbol{y}><0$

即：

$$
<\boldsymbol{x},\boldsymbol{y}>^2\leq\|x\|_2^2\|y\|_2^2
$$

当 $\boldsymbol{x}$ 与向量 $\boldsymbol{y}$ 线性相关时，显而易见，两向量共线，即有关系

$$
\boldsymbol{y}=t\boldsymbol{x}
$$

此时

$$
\begin{aligned}
<\boldsymbol{x},\boldsymbol{y}>^2 &=<\boldsymbol{x},\boldsymbol{y}>\cdot<\boldsymbol{x},\boldsymbol{y}> \\
&=<\boldsymbol{x},\boldsymbol{kx}>\cdot<\boldsymbol{x},\boldsymbol{kx}> \quad =k\|\boldsymbol{x}\|_2^2\cdot k\|\boldsymbol{x}\|_2^2=k^2\|\boldsymbol{x}\|_2^2\cdot\|\boldsymbol{x}\|_2^2 \\
&=\|\boldsymbol{y}\|_2^2\|\boldsymbol{x}\|_2^2
\end{aligned}
$$

故在欧式空间里，Cauchy 不等式成立。

需要特别注意的是，上面对于 Cauchy 不等式的证明是在欧式空间内的，如果放到酉空间中，因为酉空间内 $<\boldsymbol{x},\boldsymbol{y}>$ 的结果不一定与 $<\boldsymbol{y},\boldsymbol{x}>$ 的结果相同，故不能进行上面这样

的处理，如何证明 Cauchy 不等式在酉空间内依旧成立呢？请看后续对于 Hölder 不等式的证明。

接下来开始证明向量 2-范数是向量范数:

正定性证明:

若向量 $\boldsymbol{x} \neq 0$，则表明 $\boldsymbol{x}$ 至少有一个分量 $x_k \neq 0$，从而 $|x_k| > 0$, 故 $\|\boldsymbol{x}\|_2 > 0$

当 $\boldsymbol{x} = \boldsymbol{0}$ 时，显然向量 $\boldsymbol{x}$ 的所有分量必为 0，故 2-范数一定为 0，正定性证毕。

齐次性证明:

$$\begin{aligned}
\|\lambda \boldsymbol{x}\|_2 &= \sqrt{|\lambda x_1|^2 + |\lambda x_2|^2 + \cdots + |\lambda x_n|^2} \\
&= \sqrt{|\lambda|^2|x_1|^2 + |\lambda|^2|x_2|^2 + \cdots + |\lambda|^2|x_n|^2} \\
&= |\lambda|\sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2} \\
&= |\lambda|\|\boldsymbol{x}\|_2
\end{aligned}$$

齐次性证毕。

三角不等式证明:

$$\begin{aligned}
\|\boldsymbol{x}+\boldsymbol{y}\|_2^2 &=< \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{x}+\boldsymbol{y} > \\
&= (\boldsymbol{x}+\boldsymbol{y})^H(\boldsymbol{x}+\boldsymbol{y}) \\
&= (\boldsymbol{x}^h + \boldsymbol{y}^h)(\boldsymbol{x}+\boldsymbol{y}) \\
&= \boldsymbol{x}^H\boldsymbol{x} + \boldsymbol{x}^H\boldsymbol{y} + \boldsymbol{y}^H\boldsymbol{x} + \boldsymbol{y}^H\boldsymbol{y} \\
&= |\boldsymbol{x}^H\boldsymbol{x} + \boldsymbol{x}^H\boldsymbol{y} + \boldsymbol{y}^H\boldsymbol{x} + \boldsymbol{y}^H\boldsymbol{y}|
\end{aligned} \tag{2}$$

(2-范数一般可以理解为长度，长度一定非负，故加绝对值前后大小不变)

由 $|a+b| \leq |a| + |b|$ $|ab| = |a| \cdot |b|$

式 2 可以变为

$$\begin{aligned}
|\boldsymbol{x}^H\boldsymbol{x} + \boldsymbol{x}^H\boldsymbol{y} + \boldsymbol{y}^H\boldsymbol{x} + \boldsymbol{y}^H\boldsymbol{y}| &\leq |\boldsymbol{x}^H\boldsymbol{x}| + |\boldsymbol{x}^H\boldsymbol{y}| + |\boldsymbol{y}^H\boldsymbol{x}| + |\boldsymbol{y}^H\boldsymbol{y}| \\
&= \|\boldsymbol{x}\|_2^2 + 2\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2 + \|\boldsymbol{y}\|_2^2 \\
&= (\|\boldsymbol{x}\|_2 + \|\boldsymbol{y}\|_2)^2
\end{aligned}$$

即

$$\|\boldsymbol{x}+\boldsymbol{y}\|_2^2 \leq (\|\boldsymbol{x}\|_2 + \|\boldsymbol{y}\|_2)^2$$

三角不等式证毕，因此，2-范数是向量范数。

向量无穷范数的证明

正定性证明：

若向量 $\boldsymbol{x} \neq 0$，则表明 $\boldsymbol{x}$ 至少有一个分量 $x_k \neq 0$，从而 $|x_k| > 0$, 故 $\|\boldsymbol{x}\|_\infty > 0$

当 $\boldsymbol{x} = \mathbf{0}$ 时，显然向量 $\boldsymbol{x}$ 的所有分量必为 0，故无穷范数一定为 0，正定性证毕。

齐次性证明：

$$\begin{aligned}\|\lambda\boldsymbol{x}\|_\infty &= \max_{1<i\leq n}|\lambda x_i| \\ &= |\lambda|\max_{1<i\leq n}|x_i| \\ &= |\lambda|\|\boldsymbol{x}\|_\infty\end{aligned}$$

齐次性证毕。

三角不等式证明：

$$\|\boldsymbol{x}+\boldsymbol{y}\|_\infty = \max_{1<i\leq n}|x_i+y_i| \leq \max_{1<i\leq n}|x_i| + \max_{1<i\leq n}|y_i| = \|\boldsymbol{x}\|_\infty + \|\boldsymbol{y}\|_\infty$$

三角不等式证毕，因此，无穷范数是向量范数。

2.1.5 向量 P-范数

在2.1.4小节中，着重介绍了常见的三种向量范数并给出了其能够成为向量范数的证明，不难发现这三种范数都有一定的相似的结构，能否有一种公式能够总括这些常见的范数呢？

实际上，2.1.4介绍的三种常见范数都是本小节要介绍的 P-范数的特殊情况，向量 P-范数又叫做 Hölder 范数，其形式如下

定义 2.1.2: 向量 P-范数的定义

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}, 1 \leq p < \infty$$

通过上面的定义，不难发现，当 $p=1$ 的时候，其为向量 1-范数，当 $p=2$ 的时候，其为向量 2-范数，当 $p\to\infty$ 的时候，其为向量无穷范数。

那如何证明向量 P 范数依旧是向量范数呢？接下来给出向量 P 范数是向量范数的证明，但在此之前，还需要介绍一些背景知识。

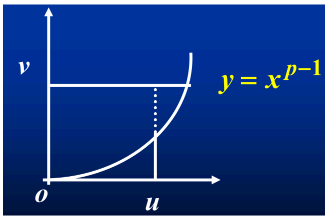
Young 不等式 (杨氏不等式)

接下来介绍 Young 不等式，Young 不等式的内容如下：若 u 和 v 是非负实数，p, q 都是正实数，且满足条件 $p > 1$ 和 $\frac{1}{p} + \frac{1}{q} = 1$，则恒有不等式

$$uv \leq \frac{1}{p}u^p + \frac{1}{q}v^q$$

下面给出 Young 不等式的证明：

如下图



我们可以根据该图得出如下三部分：

- uv: 矩形的面积
- $\int_0^u x^{p-1}\mathrm{d}x$: 以 (u, u^{p-1}) 与 x 轴包裹的小曲面三角形的面积
- $\int_0^v y^{\frac{1}{p-1}}\mathrm{d}y$: 以 $(v^{\frac{1}{p-1}}, v)$ 与 y 轴包裹的小曲面三角形的面积

看图，当 u 与 v 并不相交于函数同一点，即 $v \neq u^{p-1}$ 的时候，很明显能知道矩形的面积是小于两部分曲面三角形的面积的和的，而当 $v = u^{p-1}$ 时，矩形的面积恰好等于两部分曲面三角形的面积的和。

有了上面的认识，接下来将通过严格的数学推导理性的证明 Young 不等式：

首先，注意到

$$\frac{1}{q} + \frac{1}{p} = 1$$

对其进行变形，有

$$pq = p + q$$

同时，

$$\begin{aligned}\int_0^u x^{p-1}\mathrm{d}x &= \frac{1}{p}x^p|_{x=0}^{x=u} \\ &= \frac{1}{p}u^p\end{aligned}$$

下面处理 $\int_0^v y^{\frac{1}{p-1}}\mathrm{d}y$：

由 Young 不等式的前提，得到了 $pq = p + q$，从而有

$$p = (p-1)q$$

同时有

$$\frac{q}{p} = \frac{1}{p-1}$$

所以

$$\begin{aligned}\int_0^v y^{\frac{1}{p-1}}\mathrm{d}y &= \int_0^v y^{\frac{q}{p}}\mathrm{d}y \\ &= \frac{1}{\frac{q}{p}+1}y^{\frac{q}{p}+1}|_{y=0}^{y=v} \quad = \frac{1}{\frac{p+q}{p}}v^{\frac{p+q}{p}} = \frac{1}{\frac{pq}{p}}v^{\frac{pq}{p}} \quad = \frac{1}{q}v^q\end{aligned}$$

故

$$uv \le \frac{1}{p}u^p + \frac{1}{q}v^q$$

证毕。

Hölder 不等式

在2.1.4节介绍 Cauchy 不等式的最后提到，用原本的在欧式空间中证明 Cauchy 不等式在酉空间中依旧成立是不适用的，那应该如何证明在酉空间中，Cauchy 不等式依旧成立呢？这就是本节要讨论的 Hölder 不等式，其内容如下：

若 $p, q > 1$，且 $\frac{1}{p} + \frac{1}{q} = 1$，则对 $\boldsymbol{C}^n$ 中任意向量 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T, \boldsymbol{y} = (y_1, y_2, \cdots, y_n)^T$ 都有

$$\sum_{i=1}^n |x_i||y_i| \le (\sum_{i=1}^n |x_i|^p)^{1/p}(\sum_{i=1}^n |y_i|^q)^{1/q}$$

这里看等号右边的这两个乘积项，不难发现这与范数的定义很相似，就是 $\|\boldsymbol{x}\|_p$ 与 $\|\boldsymbol{y}\|_q$

下面给出 Hölder 不等式的证明：

令

$$u = \frac{|x_i|}{\|\boldsymbol{x}\|_p}, v = \frac{|y_i|}{\|\boldsymbol{y}\|_q}$$

由 Young 不等式，可以得到

$$uv = \frac{|x_i||y_i|}{\|\boldsymbol{x}\|_p\|\boldsymbol{y}\|_q} \le \frac{1}{p}\frac{|x_i|^p}{\|\boldsymbol{x}\|_p^p} + \frac{1}{q}\frac{|y_i|^q}{\|\boldsymbol{y}\|_q^q} \qquad (i = 1, 2, \cdots, n)$$

等式左右两侧分别进行累加，从 1 加至 n，从而有

$$\sum_{i=1}^n \frac{|x_i||y_i|}{\|\boldsymbol{x}\|_p\|\boldsymbol{y}\|_q} \le \frac{1}{p\|\boldsymbol{x}\|_p^p}\sum_{i=1}^n |x_i|^p + \frac{1}{q\|\boldsymbol{y}\|_q^q}\sum_{i=1}^n |y_i|^q \tag{1}$$

由 p-范数的式子

$$\|\boldsymbol{x}\|_p=\left(\sum_{i=1}^n|x_i|^p\right)^{1/p}$$

可以得出

$$\|\boldsymbol{x}\|_p^p=\sum_{i=1}^n|x_i|^p$$

从而有

$$\frac{\sum\limits_{i=1}^n|x_i|^p}{\|\boldsymbol{x}\|_p^p}=1 \tag{2}$$

故，不等式 1 的右侧不等号

$$\frac{1}{p\|\boldsymbol{x}\|_p^p}\sum_{i=1}^n|x_i|^p+\frac{1}{q\|\boldsymbol{y}\|_q^q}\sum_{i=1}^n|y_i|^q=\frac{1}{p}\sum_{i=1}^n\frac{|x_i|^p}{\|\boldsymbol{x}\|_p^p}+\frac{1}{q}\sum_{i=1}^n\frac{|y_i|^q}{\|\boldsymbol{y}\|_q^q} \tag{3}$$

由式 2 的结果，式 3 的结果最后为

$$\frac{1}{p}+\frac{1}{q}=1(\text{根据 Young 不等式的前提})$$

故式 1 化简为

$$\sum_{i=1}^n\frac{|x_i||y_i|}{\|\boldsymbol{x}\|_p\|\boldsymbol{y}\|_q}\le 1$$

等式左右两侧同时乘 $\|\boldsymbol{x}\|_p\|\boldsymbol{y}\|_q$，有

$$\sum_{i=1}^n|x_i|\cdot|y_i|\le\|\boldsymbol{x}\|_p\|\boldsymbol{y}\|_q$$

即

$$\sum_{i=1}^n|x_i||y_i|\le(\sum_{i=1}^n|x_i|^p)^{1/p}(\sum_{i=1}^n|y_i|^q)^{1/q}$$

证毕。

上面的是 Hölder 不等式的全部证明过程，那 Hölder 不等式与 Cauchy 不等式的关系是什么呢?

仔细观察 Hölder 不等式的右侧，可以发现，当 $p=q=2$ 时，这就是 Cauchy 不等式的右侧，那左侧的 $|<\boldsymbol{x},\boldsymbol{y}>|$ 要怎么得出来呢?

根据内积的定义，有

$$<\boldsymbol{x},\boldsymbol{y}>=\boldsymbol{x}^H\boldsymbol{y}=\bar{x}_1y_1+\bar{x}_2y_2+\cdots+\bar{x}_ny_n$$

故等式两侧都加绝对值，有

$$|<\boldsymbol{x},\boldsymbol{y}>|=|\boldsymbol{x}^H\boldsymbol{y}|=|\bar{x}_1y_1+\bar{x}_2y_2+\cdots+\bar{x}_ny_n| \tag{4}$$

根据绝对值三角不等式

$$|a+b|\leq|a|+|b|$$

式 4 的结果为

$$|<\boldsymbol{x},\boldsymbol{y}>|\leq|\bar{x}_1y_1|+|\bar{x}_2y_2|+\cdots+|\bar{x}_ny_n| \tag{5}$$

又根据性质

$$|ab|=|a|\cdot|b|$$

式 5 的结果为

$$|<\boldsymbol{x},\boldsymbol{y}>|\leq|\bar{x}_1y_1|+|\bar{x}_2y_2|+\cdots+|\bar{x}_ny_n|=|x_1||y_1|+|x_2||y_2|+\cdots+|x_n||y_n|=\sum_{i=1}^{n}|x_i||y_i| \tag{5}$$

所以我们就得出了如下的不等关系：

$$|<\boldsymbol{x},\boldsymbol{y}>|\leq\sum_{i=1}^{n}|x_i||y_i|\leq(\sum_{i=1}^{n}|x_i|^p)^{1/p}(\sum_{i=1}^{n}|y_i|^q)^{1/q}$$

这样，Cauchy 不等式在酉空间下依旧成立，证毕。

P-范数是向量范数的证明

有了前面这些定理，接下来开始证明 P-范数是向量范数：

正定性证明：

若向量 $\boldsymbol{x}\neq0$，则表明 $\boldsymbol{x}$ 至少有一个分量 $x_k\neq0$，从而 $|x_k|>0$, 故 $\|\boldsymbol{x}\|_p>0$

当 $\boldsymbol{x}=\mathbf{0}$ 时，显然向量 $\boldsymbol{x}$ 的所有分量必为 0，故 p-范数一定为 0，正定性证毕。

齐次性证明：

$$\begin{aligned}\|\lambda\boldsymbol{x}\|_p&=\sqrt[p]{|\lambda x_1|^p+|\lambda x_2|^p+\cdots+|\lambda x_n|^p}\\&=\sqrt[p]{|\lambda|^p|x_1|^p+|\lambda|^p|x_2|^p+\cdots+|\lambda|^p|x_n|^p}\\&=|\lambda|\sqrt[p]{|x_1|^p+|x_2|^p+\cdots+|x_n|^p}\\&=|\lambda|\|\boldsymbol{x}\|_p\end{aligned}$$

齐次性证毕。

三角不等式证明:

当 $p=1$ 的时候，很明显，这是 1-范数，1-范数的三角不等式证明过程请参见2.1.4节。

当 $p>1$ 时

$$\begin{aligned}\sum_{i=1}^{n}(|x_i|+|y_i|)^p &= \sum_{i=1}^{n}(|x_i|+|y_i|)\cdot(|x_i|+|y_i|)^{p-1} \\ &= \sum_{i=1}^{n}|x_i|(|x_i|+|y_i|)^{p-1}+\sum_{i=1}^{n}|y_i|(|x_i|+|y_i|)^{p-1}\end{aligned} \tag{8}$$

由 Hölder 不等式，式 8 可以变为

$$\begin{aligned}\sum_{i=1}^{n}|x_i|(|x_i|+|y_i|)^{p-1}+\sum_{i=1}^{n}|y_i|(|x_i|+|y_i|)^{p-1} \le \sum_{i=1}^{n}\left[(|x_i|+|y_i|)^{(p-1)q}\right]^{\frac{1}{q}}\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}} \\ +\sum_{i=1}^{n}\left[(|x_i|+|y_i|)^{(p-1)q}\right]^{\frac{1}{q}}\left(\sum_{i=1}^{n}|y_i|^p\right)^{\frac{1}{p}}\end{aligned}$$

注意到不等式右侧都有项 $\sum\limits_{i=1}^{n}\left[(|x_i|+|y_i|)^{(p-1)q}\right]^{\frac{1}{q}}$，提取公因式，有

$$\left(\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}+\left(\sum_{i=1}^{n}|y_i|^p\right)^{\frac{1}{p}}\right)\left[\sum_{i=1}^{n}(|x_i|+|y_i|)^{(p-1)q}\right]^{\frac{1}{q}}$$

所以经过整理，我们得到了下面这个式子:

$$\sum_{i=1}^{n}(|x_i|+|y_i|)^p \le \left(\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}+\left(\sum_{i=1}^{n}|y_i|^p\right)^{\frac{1}{p}}\right)\left[\sum_{i=1}^{n}(|x_i|+|y_i|)^{(p-1)q}\right]^{\frac{1}{q}} \tag{9}$$

由 Young 不等式的前提条件（如果忘记了请看 1.4 节）$p=(p-1)q$, 式 9 可以变为

$$\sum_{i=1}^{n}(|x_i|+|y_i|)^p \le \left(\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}+\left(\sum_{i=1}^{n}|y_i|^p\right)^{\frac{1}{p}}\right)\left[\sum_{i=1}^{n}(|x_i|+|y_i|)^{p}\right]^{\frac{1}{q}} \tag{10}$$

注意到不等号左右两侧都有项 $\sum\limits_{i=1}^{n}(|x_i|+|y_i|)^p$，故提公因式，有

$$\left[\sum_{i=1}^{n}(|x_i|+|y_i|)^p\right]^{1-\frac{1}{q}} \le \left(\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}+\left(\sum_{i=1}^{n}|y_i|^p\right)^{\frac{1}{p}}\right) \tag{11}$$

又由 Young 不等式的前提条件: $\frac{1}{p}+\frac{1}{q}=1$，故式 11 可以化简为:

$$\left[\sum_{i=1}^{n}(|x_i|+|y_i|)^p\right]^{\frac{1}{p}} \le \left(\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}+\left(\sum_{i=1}^{n}|y_i|^p\right)^{\frac{1}{p}}\right) \tag{12}$$

接下来，只需要证明

$$\|\boldsymbol{x}+\boldsymbol{y}\|_p \le \left[\sum_{i=1}^{n}(|x_i|+|y_i|)^p\right]^{\frac{1}{p}}$$

即可

由 p-范数的定义和绝对值三角不等式，有

$$\|\boldsymbol{x}+\boldsymbol{y}\|_p = \left(\sum_{i=1}^{n}(|x_i+y_i|)^p\right)^{\frac{1}{p}} \le \left(\sum_{i=1}^{n}(|x_i|+|y_i|)^p\right)^{\frac{1}{p}} \le \|\boldsymbol{x}\|_p + \|\boldsymbol{y}\|_p$$

故三角不等式成立，故 p 范数是向量范数，证毕。

证明 p-范数的三角不等式比较复杂，主要涉及下面五步：

1. 将起点 $\sum\limits_{i=1}^{n}(|x_i|+|y_i|)^p$ 变换为 1 与 $p-1$ 两个幂相乘的形式，并应用分配律展开。
2. 在第一步的基础上，应用 Hölder 不等式进行放大。
3. 提公因式 $\sum\limits_{i=1}^{n}\left[(|x_i|+|y_i|)^{(p-1)q}\right]^{\frac{1}{q}}$，并应用 Young 不等式的前提条件进行化简
4. 不等式左右两侧同时除以 $\sum\limits_{i=1}^{n}(|x_i|+|y_i|)^p$，并继续应用 Young 不等式的前提条件进行化简。
5. 应用 $|x_i+y_i| \le |x_i|+|y_i|$，得到最后的结论。

由此，通过 P-范数的定义，我们知道了一个范数有无数种种类描述，并非只有向量 1-范数、向量 2-范数或向量无穷范数，同时经过2.1.4和2.1.5两个小节的介绍，对于向量的范数有了更直观的感受——范数的确是一种广义的“距离”。

2.1.6 向量范数的应用

在本节的最后一小节，介绍一下向量范数的三大应用——利用高维空间的范数定义低维空间的向量范数、向量范数的等价和利用向量范数研究向量序列的收敛性

利用高维空间的向量范数定义低维空间的向量范数

定理 2.1.1

设 $\|\cdot\|$ 是 $\mathbb{C}^n$ 上的范数，$A \in \mathbb{C}_n^{m\times n}$(表示 A 矩阵的列数等于矩阵的秩，即 A 矩阵列满秩)，则 $\|A\cdot\|$ 是 $\mathbb{C}^n$ 上的范数。

定理 2.1.1 的作用是什么呢？在很多情况下，如果我们知道高维空间内的向量范数，因为各种原因求得低维空间内的向量范数较为困难或者根本无法求得，那么在这种情况下，我们

可以构造一个列满秩的矩阵，与低维空间内的该向量 (即定理中的“·”) 相乘，便可求出低维空间内的向量的范数。

为什么可以这样做呢？还记得在上一章中讲过的线性变换的概念吗？

假设有一个 m 维的向量 $\boldsymbol{y}$，还有一个 n 维的向量 $\boldsymbol{x}(m>n)$，假设有一个矩阵 A，便可以做如下线性变换：

$$\boldsymbol{y}=A\boldsymbol{x}$$

这样，通过一个矩阵 A，便成功的将低维的向量 $\boldsymbol{x}$ 升维至高维。

如果上面这样说还是觉得不理解，接下来通过一个具体的例子来展示一下何为“升维”

例 2.1.1. 假设有矩阵 $A=\begin{bmatrix}1&1\\0&1\\1&1\end{bmatrix}$(A 是一个三维矩阵)，向量 $\boldsymbol{x}=[x_1,x_2]^T$，($\boldsymbol{x}$ 是一个二维矩阵)

那么 $A\boldsymbol{x}=\begin{bmatrix}1&1\\0&1\\1&1\end{bmatrix}\cdot\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}x_1+x_2\\x_2\\x_1+x_2\end{bmatrix}$，很显然，$A\boldsymbol{x}$ 变为了一个三维矩阵，这就是通过一个矩阵实现将低维向量升维至高维空间。

那么在范数层面上，也可以实现一样的功能，公式如下：

$$\|\boldsymbol{x}\|_n\overset{def}{=}\|\boldsymbol{y}\|_m=\|A\boldsymbol{x}\|_m$$

下面给出使用高维空间向量范数来定义低维空间向量范数合理性的证明，依旧是从范数的三个性质出发

证明.

1. 正定性

   由于 $x\neq 0$，并且 A 是列满秩矩阵，由线性代数的知识可以得出齐次线性方程组 $A\boldsymbol{x}=\boldsymbol{0}$ 只有零解，故范数 $\|A\boldsymbol{x}\|>0$

   因此，由 $A\boldsymbol{x}$ 定义的 $\boldsymbol{x}$ 的范数 $\|\boldsymbol{x}\|$ 也大于零，正定性证毕。

2. 齐次性

   $\|\lambda\boldsymbol{x}\|_n\overset{def}{=}\|A\lambda x\|_m=\|\lambda A\boldsymbol{x}\|_m$

   由于 λ 是一个数，同时 $\|A\boldsymbol{x}\|_m$ 已经是一个范数，满足范数的齐次性，因此上式的结果为 $|\lambda|\|A\boldsymbol{x}\|$，齐次性证毕。

3. 三角不等式

$\|\boldsymbol{x_1}+\boldsymbol{x_2}\|_n=\|A(\boldsymbol{x_1}+\boldsymbol{x_2})\|_m=\|A\boldsymbol{x_1}+A\boldsymbol{x_2}\|_m$

显然，$\|A\boldsymbol{x_1}\|$ 与范数 $\|A\boldsymbol{x_2}\|$ 已经是范数，由范数的三角不等式，有

$\|A\boldsymbol{x_1}+A\boldsymbol{x_2}\|_m\leq\|A\boldsymbol{x_1}\|_m+\|A\boldsymbol{x_2}\|_m=\|\boldsymbol{x_1}\|_n+\|\boldsymbol{x_2}\|_n$，三角不等式证毕。

□

下面用这个定理来看一道例题

例 2.1.2. 设 $\|\boldsymbol{x}\|$ 是 $\mathbb{P}^n$ 中的向量范数，$A\in\mathbb{P}^{n\times n}$，则 $\|A\boldsymbol{x}\|$ 也是 $\mathbb{P}^n$ 中的向量范数的充要条件是 A 可逆。

解. 先证充分性（右推左）：

显然，由定理2.1.1可知，在 A 可逆的条件下，其一定为列满秩矩阵，因此 $\|A\boldsymbol{x}\|$ 一定为向量范数。

再证必要性（左推右）：

使用反证法，假设 A 不可逆，故齐次线性方程 $A\boldsymbol{x}=\boldsymbol{0}$ 存在非零解 $\boldsymbol{x_0}$

因此，$\|\boldsymbol{x_0}\|\overset{def}{=}\|A\boldsymbol{x_0}\|=\boldsymbol{0}$，这与范数的正定性相矛盾。

证毕。 □

向量范数的等价

定义 2.1.3: 向量范数等价的定义

设在 $V_n(P)$ 上定义了 $\|\boldsymbol{x}\|_a,\|\boldsymbol{x}\|_b$ 两种向量范数，若存在常数 $C_1>0,C_2>0$，使得 $\forall x\in V_n(P)$，都有

$$\|\boldsymbol{x}\|_a\leq C_1\|\boldsymbol{x}\|_b,$$

$$\|\boldsymbol{x}\|_b\leq C_2\|\boldsymbol{x}\|_b$$

则称 $\|\boldsymbol{x}\|_a$ 与 $\|\boldsymbol{x}\|_b$ 等价

向量范数等价的概念是为后面利用向量范数判断向量序列的收敛性做服务的，由定义2.1.3里的不等关系，我们可以构造出下面的两组不等式：

$$\frac{1}{C_1}\|\boldsymbol{x}\|_a\leq\|\boldsymbol{x}\|_b\leq C_2\|\boldsymbol{x}\|_b$$

和

$$\frac{1}{C_2}\|\boldsymbol{x}\|_b\leq\|\boldsymbol{x}\|_a\leq C_1\|\boldsymbol{x}\|_b$$

由于范数是一个数，故上式的不等式可以由夹逼定理确定向量范数不受范数的种类影响而能够唯一收敛到一个值。

利用向量范数判断向量序列的收敛性

何为向量序列？在高等数学的课程中，我们接触过数列这一概念，如，等差数列，等比数列这些都是数列，下面将数列这一概念延伸一下便能够得到向量序列的知识。

假设有如下的向量

$$\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(k)}, \cdots$$

其中

$$\boldsymbol{x}^{(1)} = \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ \vdots \\ x_n^{(1)} \end{bmatrix}, \boldsymbol{x}^{(2)} = \begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \\ \vdots \\ x_n^{(2)} \end{bmatrix}, \cdots, \boldsymbol{x}^{(k)} = \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \\ \vdots \\ x_n^{(k)} \end{bmatrix}, \cdots$$

这个时候，序列 $x_1^{(1)}, x_1^{(2)}, \cdots, x_1^{(k)}, \cdots$，序列 $x_2^{(1)}, x_2^{(2)}, \cdots, x_2^{(k)}, \cdots$ 等（把这些向量的每一行组合在一起）就是一个**数列**，而把这些数列组合在一起，就是这些向量，这些向量组成了一个**向量序列**。

在高等数学的课程中，我们了解过数列极限 (下面这段定义摘录自《张宇考研数学基础 30 讲——高等数学分册》)：

设 x_n 为一数列，若存在常数 a，对于任意的 ε(不论它多么小), 总存在正整数 N，使得当 $n > N$ 时，$|x_n - a| < \varepsilon$ 恒成立，则称常数 a 是数列 x_n 的极限，或者称数列 x_n 收敛于 a，记为

$$\lim_{n\to\infty} x_n = a\text{或}x_n \to a \quad (n \to \infty)$$

将数列极限这一概念延伸到向量序列的极限中，上面介绍的每一个数列如果都能收敛，那么就可以说整个向量序列也是收敛的，趋近于某一个向量。

如果向量序列是收敛的，会有如下等式：

$$\lim_{k\to\infty} \boldsymbol{x}^{(\boldsymbol{k})} = \boldsymbol{a}$$

上式还有下面这种利用范数来判断向量序列是否收敛的定义：

定理 2.1.2

设 $\|\cdot\|$ 是 $\mathbb{C}^n$ 上的任一向量范数，则

$$\lim_{k\to\infty} \boldsymbol{x}^{(k)} = \boldsymbol{a} \Longleftrightarrow \lim_{k\to\infty} \|\boldsymbol{x}^{(k)} - \boldsymbol{a} = 0\|$$

上面这两个式子的等价证明如下：

证明.

因为

$$\lim_{k\to\infty} \boldsymbol{x}^{(k)} = \boldsymbol{a}$$

所以

$$\lim_{k\to\infty} \boldsymbol{x}_i^{(k)} = \boldsymbol{a}_i \ (i = 1, 2, \cdots, n)$$

所以

$$\lim_{k\to\infty} |\boldsymbol{x}_i^{(k)} - \boldsymbol{a}_i| = 0$$

这里注意到 $\boldsymbol{x}_i^{(k)}$ 和 $\boldsymbol{a}_i$ 的任意性, $\lim\limits_{k\to\infty} |\boldsymbol{x}_i^{(k)} - \boldsymbol{a}_i| = 0$ 这个等式表明的就是向量序列的每一个分量所构成的数列都是收敛的，那自然存在下面的关系：

$$\lim_{k\to\infty} \max_{1\le i\le n} \{|\boldsymbol{x}_i^{(k)} - \boldsymbol{a}_i|\} = 0$$

既然每一个分量构成的数列是收敛的，那自然最大的那一行分量构成的数列依旧是收敛的，仔细看这个式子，可以与向量的无穷范数建立起了联系，即

$$\lim_{k\to\infty} \|\boldsymbol{x}_i^{(k)} - \boldsymbol{a}_i\|_\infty = 0$$

证毕。 □

2.2 矩阵的范数

上一节介绍了向量范数的有关知识，包括向量范数的性质、判断条件、常见的向量范数等，本节将眼光放到矩阵范数中。

2.2.1 从向量范数到矩阵范数

上一节我们知道了任意的向量范数 (即向量 p-范数) 的形式，那矩阵的范数又应该如何理解呢？

矩阵范数和向量范数的形式及其类似，甚至可以通过向量范数的形式引出矩阵范数，将 $\mathbb{P}^n$ 上的范数推广到 $\mathbb{P}^{m\times n}$ 上。

为什么可以这样推广呢？这里需要做这样一步认定：我们可以将一个 $m\times n$ 的矩阵 A 看作是数域 $\mathbb{P}^{mn}$ 上的一个向量，如此就可以将一个矩阵当做一个向量来处理，自然就可以用前一节的向量范数的相关概念来定义矩阵范数。

既然如此，那矩阵范数的定义就不难理解了：

定义 2.2.1: 矩阵范数的定义

设 $A\in\mathbb{P}^{m\times n}$，若映射 $\|\cdot\|$: $\mathbb{P}^{m\times n}\to\mathbb{R}$ 满足：

- 正定性：$\|A\|\geq 0$，当且仅当 $A=0$ 时，$\|A\|=0$
- 齐次性：$\|\lambda A\|=|\lambda|\|A\|,\forall\lambda\in\mathbb{R},\forall A\in\mathbb{P}^{m\times n}$
- 三角不等式：$\|A+B\|\leq\|A\|+\|B\|,\forall A,B\in P^{m\times n}$

则称映射 $\|\cdot\|$ 为 $\mathbb{P}^{m\times n}$ 上的**矩阵范数**。

2.2.2 常见的矩阵范数

与常见的向量范数相似，常见的矩阵范数有矩阵 1-范数，矩阵 2-范数和矩阵无穷范数：

$$\|A\|_{m_1}=\sum_{j=-1}^{n}\sum_{i=1}^{m}|a_{ij}|$$

$$\|A\|_{m_2}=\left(\sum_{j=1}^{n}\sum_{i=1}^{m}|a_{ij}|^2\right)^{\frac{1}{2}}$$

$$\|A\|_{m_\infty}=\max_{i,j}\{|a_{ij}|\},1\leq i\leq m,1\leq j\leq n$$

由于矩阵范数和向量范数的相似性，故上面的三种常见矩阵范数的证明过程就不详细说了，矩阵范数更值得关注的是矩阵范数的相容性质。

2.2.3 矩阵范数的相容

矩阵范数相容的定义如下：

定义 2.2.2: 范数相容的定义

设 $\|\cdot\|_a$: $\mathbb{P}^{m\times l}\to\mathbb{R}, \|\cdot\|_b:\mathbb{P}^{l\times n}\to\mathbb{R}, \|\cdot\|_c:\mathbb{P}^{m\times n}\to\mathbb{R}$ 是矩阵范数，如果

$$\|AB\|_c\le\|A\|_a\cdot\|B\|_b$$

则称矩阵范数 $\|\cdot\|_a, \|\cdot\|_b$ 和 $\|\cdot\|_c$ 相容。

如果

$$\|AB\|\le\|A\|\cdot\|B\|$$

则称 $\|\cdot\|$ 是子相容的矩阵范数

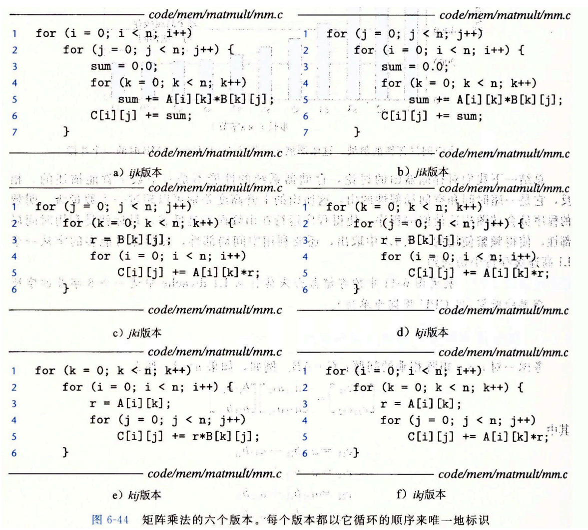
注. 今后若无特殊说明，所有的“相容范数”均为自相容范数。

有矩阵范数相容的作用是什么呢? 我们知道，矩阵乘法的法则是这样的 (下面这段话摘自同济大学《工程数学——线性代数》):

设 $A=(a_{ij})$ 是一个 $m\times s$ 矩阵，$B=(b_{ij})$ 是一个 $x\times n$ 矩阵，那么规定矩阵 A 与矩阵 B 的乘积是一个 $m\times n$ 矩阵 $C=(c_{ij})$，其中

$$c_{ij}=a_{i1}b_{1j}+a_{i2}b_{2j}+\cdots+a_{is}b_{sj}=\sum_{k=1}^{s}a_{ik}b_{kj},\quad(i=1,2,\cdots,m;j=1,2,\cdots,n)$$

通俗点说，两个矩阵相乘，得到的新矩阵的对应元素是第一个矩阵的行乘以第二个元素的列的结果求和得出的，那么这个过程用计算机来表述的话，代码是这样的 (以 C 语言为例):


code/mem/matmult/mm.c

```
for (i = 0; i < n; i++)
    for (j = 0; j < n; j++) {
        sum = 0.0;
        for (k = 0; k < n; k++)
            sum += A[i][k]*B[k][j];
        C[i][j] += sum;
    }
```

a) ijk版本

```
for (j = 0; j < n; j++)
    for (i = 0; i < n; i++) {
        sum = 0.0;
        for (k = 0; k < n; k++)
            sum += A[i][k]*B[k][j];
        C[i][j] += sum;
    }
```

b) jik版本

```
for (j = 0; j < n; j++)
    for (k = 0; k < n; k++) {
        r = B[k][j];
        for (i = 0; i < n; i++)
            C[i][j] += A[i][k]*r;
    }
```

c) jki版本

```
for (k = 0; k < n; k++)
    for (j = 0; j < n; j++) {
        r = B[k][j];
        for (i = 0; i < n; i++)
            C[i][j] += A[i][k]*r;
    }
```

d) kji版本

```
for (k = 0; k < n; k++)
    for (i = 0; i < n; i++) {
        r = A[i][k];
        for (j = 0; j < n; j++)
            C[i][j] += r*B[k][j];
    }
```

e) kij版本

```
for (i = 0; i < n; i++)
    for (k = 0; k < n; k++) {
        r = A[i][k];
        for (j = 0; j < n; j++)
            C[i][j] += A[i][k]*r;
    }
```

f) ikj版本

图 6-44 矩阵乘法的六个版本。每个版本都以它循环的顺序来唯一地标识

图 2.1: 矩阵乘法代码示意

图2.1的代码节选自《深入理解计算机系统》，在书中对于矩阵乘法给出了六个版本的算

法，但算法中 i,j,k 变量的不同顺序仅仅是对访问内存的次数有影响[1]，对于整体的时间复杂度，依旧为 $O(n^3)$，这个算法的时间复杂度在面对两个极为庞大的矩阵进行乘法运算是要消耗非常大的时间的，如果在这种情况下计算矩阵范数是一件十分耗时的事情。

但幸运的是，在许多工程中，我们其实没有必要知道两个矩阵相乘的范数的精确数值（在很多情况下都是这样，我们只需要一个近似的数值，或者说了解到一个上界或者下界就可以，比如上文中提到的时间复杂度大 O 记法，它仅仅描述了一个算法所耗费时间的最坏情况，即一个上界），在这种情况下，矩阵的相容性就会起到作用，计算两个单独的矩阵的范数再相乘一定比计算两个矩阵先相乘再求范数省时。[2]

2.2.4 常见矩阵范数的相容性

在了解了何为矩阵范数相容的概念后，接下来就要讨论常见矩阵范数（矩阵 1-范数，矩阵 2-范数和矩阵无穷范数）的相容性了。

矩阵 1-范数的相容性

前面已经介绍过矩阵 1-范数的结构

$$\|A\|_{m_1} = \sum_{j=1}^{n}\sum_{i=1}^{m}|a_{ij}|$$

下面来证明矩阵 1-范数是自相容范数：

证明. 若证明矩阵 1-范数是自相容范数，只需证

$$\|\boldsymbol{AB}\|_{m_1} \le \|\boldsymbol{A}\|_{m_1}\|\boldsymbol{B}\|_{m_1}$$

设 $\boldsymbol{A} \in \mathbb{P}^{m\times l}, \boldsymbol{B} \in \mathbb{P}^{l\times n}$, 根据矩阵 1-范数的定义，有

$$\begin{aligned}
\|\boldsymbol{AB}\|_{m_1} &= \sum_{i=1}^{m}\sum_{j=1}^{n}|a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{il}b_{lj}| \quad (\text{矩阵乘法}) \\
&\le \sum_{i=1}^{m}\sum_{j=1}^{n}(|a_{i1}||b_{1j}| + |a_{i2}||b_{2j}| + \cdots + |a_{il}||b_{lj}|) \quad (\text{绝对值三角不等式}) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{l}|a_{ik}||b_{kj}| \quad (\text{用一个连加号概括上式})
\end{aligned}$$

[1]至于为什么交换变量顺序会对访问内存有影响，这里涉及到了程序的局部性原理，已超出了本门课程的范畴，具体请翻阅计算机组成原理相关教材。

[2]关于算法的时间复杂度，如何衡量一个计算机算法的好坏，具体请翻阅计算机算法方面的教材。

交换一下连加号的顺序，原式变为

$$\sum_{k=1}^{l}\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ik}||b_{kj}|$$

由于连加号的计算是由内向外计算的，因此会首先计算 $\sum_{j=1}^{n}|a_{ik}||b_{kj}|$，这个时候会发现 a_{ik} 并没有出现下标 j，也就是说 $|a_{jk}|$ 并不会受 j 的影响，因此，在 $\sum_{j=1}^{n}|a_{ik}||b_{kj}|$ 中，$|a_{ik}|$ 可以被认为是一个公因子提到公式外面去，则原式可以变为

$$\sum_{k=1}^{l}\sum_{i=1}^{m}|a_{ik}|\sum_{j=1}^{n}|b_{kj}|$$

此时观察 $\sum_{i=1}^{m}|a_{ik}|\sum_{j=1}^{n}|b_{kj}|$，会发现 $\sum_{j=1}^{n}|b_{kj}|$ 并不受 i 的影响，因此可以把这部分整体当做一个公因式提到外面，于是式子就会变成

$$\sum_{k=1}^{l}\left[\left(\sum_{i=1}^{m}|a_{ik}|\right)\left(\sum_{j=1}^{n}|b_{kj}|\right)\right]$$

这里，$(\sum_{i=1}^{m}|a_{ik}|)$ 叫做 A 的**绝对列和**，$\left(\sum_{j=1}^{n}|b_{kj}|\right)$ 叫做 B 的**绝对行和**，这两个概念在下一节算子范数中还会有应用。

如果想继续提公因式，需要对式子进行放缩，这里引入矩阵 A 的矩阵 1-范数，上式变为

$$\sum_{k=1}^{l}\left(\sum_{j=1}^{n}|b_{kj}|\right)\left(\sum_{k=1}^{l}\sum_{i=1}^{m}|a_{ik}|\right)$$

这相当于是在原本 A 矩阵的绝对列和的基础上把所有的 A 元素相加，显然结果更大，因此有如下不等关系

$$\sum_{k=1}^{l}\left[\left(\sum_{i=1}^{m}|a_{ik}|\right)\left(\sum_{j=1}^{n}|b_{kj}|\right)\right]\leq\sum_{k=1}^{l}\left(\sum_{j=1}^{n}|b_{kj}|\right)\left(\sum_{k=1}^{l}\sum_{i=1}^{m}|a_{ik}|\right)$$

这种情况下可以发现，$\left(\sum_{k=1}^{l}\sum_{i=1}^{m}|a_{ik}|\right)$ 并不受 $\sum_{k=1}^{l}\left(\sum_{j=1}^{n}|b_{kj}|\right)\left(\sum_{k=1}^{l}\sum_{i=1}^{m}|a_{ik}|\right)$ 中最外层的连加号 $\sum_{k=1}^{l}$ 的影响，因此 $\left(\sum_{k=1}^{l}\sum_{i=1}^{m}|a_{ik}|\right)$ 又是一个公因式，将其提出，最后式子变为

$$\sum_{k=1}^{l}\left[\left(\sum_{i=1}^{m}|a_{ik}|\right)\left(\sum_{j=1}^{n}|b_{kj}|\right)\right]\leq\left(\sum_{k=1}^{l}\sum_{i=1}^{m}|a_{ik}|\right)\left(\sum_{k=1}^{l}\sum_{j=1}^{n}|b_{kj}|\right)$$

注意到不等式右侧就是 A 矩阵的矩阵 1-范数和 B 矩阵的矩阵 1-范数，因此得到了

$$\|\boldsymbol{AB}\|_{m_1}\leq\|\boldsymbol{A}\|_{m_1}\|\boldsymbol{B}\|_{m_1}$$

也就说明矩阵 1-范数是自相容范数，证毕。 □

矩阵 2-范数的相容性

前面已经介绍过矩阵 2-范数的结构

$$\|\boldsymbol{A}\|_{m_2}=\left(\sum_{j=1}^{n}\sum_{i=1}^{m}|a_{ij}|^2\right)^{\frac{1}{2}}$$

下面来证明矩阵 2-范数是自相容范数：

证明. 若想证矩阵 2-范数是自相容范数，只需证

$$\|\boldsymbol{AB}\|_{m_2}^2\leq\|\boldsymbol{A}\|_{m_2}^2\cdot\|\boldsymbol{B}\|_{m_2}^2$$

根据矩阵 2-范数的定义，有

$$\begin{aligned}\|\boldsymbol{AB}\|_{m_2}^2&=\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{i1}b_{1j}+a_{i2}b_{2j}+\cdots+a_{il}b_{lj}|^2\\&\leq\sum_{i=1}^{m}\sum_{j=1}^{n}[|a_{i1}||b_{1j}|+\cdots+|a_{il}||b_{lj}|]^2\\&=\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\sum_{k=1}^{l}|a_{ik}||b_{kj}|\right)^2\end{aligned}$$

前面三步的方法跟证明矩阵 1-范数是自相容范数的流程是一样的，接下来是不一样的

根据 Cauchy 不等式, 有

$$\|\boldsymbol{AB}\|_{m_2}^2\leq\sum_{i=1}^{m}\left[\sum_{j=1}^{n}\left(\sum_{k=1}^{l}|a_{ik}|^2\right)\left(\sum_{k=1}^{l}|b_{kj}|^2\right)\right]$$

请注意不等式右侧的 $\sum_{j=1}^{n}\left(\sum_{k=1}^{l}|a_{ik}|^2\right)\left(\sum_{k=1}^{l}|b_{kj}|^2\right)$ 这一部分，依旧是由于计算连加是从内向外计算的，因此项 $\sum_{k=1}^{l}|a_{ik}|^2$ 并不受 $\sum_{j=1}^{n}$ 中 j 的影响，故可以当作公因式提出，如此式子就变为了

$$\|\boldsymbol{AB}\|_{m_2}^2\leq\sum_{i=1}^{m}\left(\sum_{k=1}^{l}|a_{ik}|^2\right)\left(\sum_{j=1}^{n}\sum_{k=1}^{l}|b_{kj}|^2\right)$$

此时注意到在不等式的右侧，项 $\sum_{j=1}^{n}\sum_{k=1}^{l}|b_{kj}|^2$ 依旧不受 $\sum_{i=1}^{m}$ 中 i 的影响，故整体可以当作公因式提出，如此式子就变成了

$$\|\boldsymbol{AB}\|_{m_2}^2\leq\left(\sum_{i=1}^{m}\sum_{k=1}^{l}|a_{ik}|^2\right)\left(\sum_{j=1}^{n}\sum_{k=1}^{l}|b_{kj}|^2\right)$$

即

$$\|\boldsymbol{AB}\|_{m_2}^2 \leq \|\boldsymbol{A}\|_{m_2}^2 \|\boldsymbol{B}\|_{m_2}^2$$

故矩阵 2-范数是自相容范数，证毕。 □

矩阵 2-范数也叫 Frobenius 范数，其也是用途最多的一种范数，矩阵 2-范数还有下面的性质

定理 2.2.1: 矩阵 2-范数 (Frobenius 范数) 的性质

设 $A \in \mathbb{P}^{n\times n}$,

1. 若 $\boldsymbol{A} = (a_1, a_2, \cdots, a_n)$, 则

$$\|\boldsymbol{A}\|_F^2 \overset{①}{=} \|\boldsymbol{A}\|_{m_2}^2 \overset{②}{=} \sum_{i=1}^{n} \|a_i\|_2^2$$

   其中，$\|a_i\|_2^2 = a_i^H a_i$ 是 $\mathbb{P}^n$ 中的向量范数

2. $\|\boldsymbol{A}\|_{m_2}^2 \overset{③}{=} \operatorname{tr}(\boldsymbol{A}^H\boldsymbol{A}) \overset{④}{=} \sum_{i=1}^{n} \lambda_i(\boldsymbol{A}^H\boldsymbol{A})$

3. 对任意的酉矩阵 $\boldsymbol{U}$ $\boldsymbol{V} \in \mathbb{P}^{n\times n}$，有

$$\|\boldsymbol{A}\|_{m_2} = \|\boldsymbol{U}^H\boldsymbol{A}\boldsymbol{V}\|_{m_2} = \|\boldsymbol{U}\boldsymbol{A}\boldsymbol{V}^H\|_{m_2}$$

这里只给出定理2.2.1中前两条的证明，第三条的证明在后面的酉不变范数中。

证明.

① 的证明省略，因为这就是定义——矩阵 2-范数也叫 Frobenius 范数，这无需证明。

④ 的证明省略，因为根据矩阵特征值与矩阵的迹的关系，矩阵特征值的和就是矩阵的迹，这也无需证明。

②, ③, ④ 的证明如下：设矩阵

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

则

$$\boldsymbol{A}^H=\begin{bmatrix}\bar{a}_{11} & \bar{a}_{21} & \cdots & \bar{a}_{n1}\\ \bar{a}_{21} & \bar{a}_{22} & \cdots & \bar{a}_{n2}\\ \vdots & \vdots & & \vdots\\ \bar{a}_{1n} & \bar{a}_{2n} & \cdots & \bar{a}_{nn}\end{bmatrix}$$

则

$$\boldsymbol{A}^H\cdot\boldsymbol{A}=\begin{bmatrix}\sum\limits_{i=1}^{n}|a_{i1}|^2 & * & * & *\\ * & \sum\limits_{i=1}^{n}|a_{i2}|^2 & * & *\\ \vdots & \vdots & & \vdots\\ * & * & * & \sum\limits_{i=1}^{n}|a_{in}|^2\end{bmatrix}$$

注. 注意，这里的 $*$ 代表的是计算出来的结果，但这部分的结果对于证明过程没有意义，故不进行计算用 $*$ 代替。

仔细观察便能发现，对角线上的元素分别为：第一列元素的和的平方、第二列元素的和的平方，一直到第 n 列元素和的平方，这恰恰等于矩阵的矩阵 2-范数，而根据矩阵的迹的计算方法，矩阵的迹等于矩阵对角线元素的和，因此，就有了如下等式

$$\operatorname{tr}(A^HA)=\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2=\|A\|_{m2}$$

至此，性质 1 与 2 证毕。 □

矩阵无穷范数的相容性

矩阵的无穷范数是不相容的，证明不相容，只需要给出一个反例即可，即只需给出两个矩阵 $\boldsymbol{A},\boldsymbol{B}$，使得

$$\|\boldsymbol{AB}\|_{m\infty}\nleq\|\boldsymbol{A}\|_{m\infty}\|\boldsymbol{B}\|_{m\infty}$$

即可

事实上，若

$$\boldsymbol{A}=\boldsymbol{B}=\begin{bmatrix}0 & 1\\ 1 & 1\end{bmatrix}$$

则

$$\boldsymbol{AB}=\begin{bmatrix}0 & 1\\ 1 & 1\end{bmatrix}^2=\begin{bmatrix}1 & 1\\ 1 & 2\end{bmatrix}$$

故

$$\|\boldsymbol{AB}\|_{m_\infty}=2,\quad \|\boldsymbol{A}\|_{m_\infty}\|\boldsymbol{B}\|_{m_\infty}=1,\quad \|\boldsymbol{AB}\|_{m_\infty}>\|\boldsymbol{A}\|_{m_\infty}\|\boldsymbol{B}\|_{m_\infty}$$

因此，矩阵的无穷范数不是自相容范数。

2.2.5 酉不变范数

酉不变范数就是定理2.2.1中的性质 3，下面给出其证明：

证明.

$$\begin{aligned}\|\boldsymbol{UA}\|_{m_2}^2&=\operatorname{tr}\left[(\boldsymbol{UA})^H\boldsymbol{UA}\right]\\&=\operatorname{tr}\left[\boldsymbol{A}^H\boldsymbol{U}^H\boldsymbol{UA}\right](\text{注意到}\boldsymbol{U}^H\boldsymbol{U}=\boldsymbol{E})\\&=\operatorname{tr}\left[\boldsymbol{A}^H\boldsymbol{A}\right]\end{aligned}$$

$$\begin{aligned}\|\boldsymbol{AV}\|_{m_2}^2&=\operatorname{tr}\left[(\boldsymbol{AV})^H\boldsymbol{AV}\right]\\&=\operatorname{tr}\left[\boldsymbol{V}^H\boldsymbol{A}^H\boldsymbol{AV}\right](\text{注意到}\boldsymbol{V}^H=\boldsymbol{V}^{-1})\end{aligned}$$

根据矩阵相似的定义，注意到

$$\boldsymbol{V^{-1}A^HAV}\sim\boldsymbol{A}^H\boldsymbol{A}$$

故

$$\text{原式}=\operatorname{tr}\left[\boldsymbol{A}^H\boldsymbol{A}\right](\text{因为矩阵相似，故特征值相等，因此矩阵的迹也相等})$$

证毕。 □

在上面的过程中，分别利用酉矩阵与其共轭转置相乘为单位阵以及酉矩阵与其共轭转置互为逆矩阵从而导出矩阵相似的性质进行推导，将上面的两种情况合成在一起就是定理2.2.1中的第三条的所有。

2.3 算子范数

本章的最后一个大知识点为将要讨论的算子范数，算子范数成功的将向量范数与矩阵范数联系在一起，如此便可通过向量范数来定义矩阵的算子范数。

2.3.1 向量范数与矩阵范数的相容关系

在前面的两个小节中，我们讨论了向量范数和矩阵范数，我们在引入矩阵范数的时候，是通过与向量范数相似的方式来定义矩阵范数，可如果将二者放在一起，他们的相容性关系是

如何的呢?

定义 2.3.1: 矩阵范数与向量范数相容

设 $\|\cdot\|_a$ 是 $\mathbb{P}^n$ 上的向量范数，$\|*\|_m$ 是 $\mathbb{P}^{n\times n}$ 上的矩阵范数，且

$$\|\boldsymbol{Ax}\|_a \leq \|\boldsymbol{A}\|_m\|\boldsymbol{x}\|_a$$

则称 $\|*\|_m$ 为与向量范数 $\|\cdot\|_a$ 相容的矩阵范数

定义2.3.1描述了矩阵范数与向量范数相容的条件，接下来我们来看两个例子:

例 2.3.1. 设 $\boldsymbol{x}\in\mathbb{P}^n,\boldsymbol{A}\in\mathbb{P}^{n\times n}$，则

$$\|\boldsymbol{A}\|_{m_1}=\sum_{j=1}^{n}\sum_{i=1}^{n}|a_{ij}|$$

是与向量范数 $\|\cdot\|_1$ 相容的矩阵范数。

解. 证明方法与前文证明矩阵 1-范数是自相容范数思路相似，只需要将 $\boldsymbol{B}$ 矩阵换为向量 $\boldsymbol{x}$ 即可，故不赘述。 □

例 2.3.2. 设 $x\in\mathbb{P}^n,A\in\mathbb{P}^{n\times n}$，则 $\|\boldsymbol{A}\|_{m_2}$ 是与 $\|\boldsymbol{x}\|_2$ 相容的矩阵范数

解. 证明方法与前文证明矩阵 2-范数是自相容范数思路相似，只需要将 $\boldsymbol{B}$ 矩阵换为向量 $\boldsymbol{x}$ 即可，故不赘述。 □

可是我们知道，向量范数有很多种（参考介绍向量范数的 $p-$ 范数），那对于任意一种向量范数，是否也存在与该向量范数相容的矩阵范数呢？这就是下面要介绍的算子范数:

2.3.2 算子范数

下面给出算子范数的定义:

定义 2.3.2: 算子范数的定义

设 $\|\boldsymbol{x}\|_a$ 是 $\mathbb{P}^n$ 上的向量范数，$\boldsymbol{A}\in\mathbb{P}^{n\times n}$，则

$$\|\boldsymbol{A}\|_a=\max_{\boldsymbol{x}\neq\boldsymbol{0}}\frac{\|\boldsymbol{Ax}\|_a}{\|\boldsymbol{x}\|_a}$$

是与向量范数 $\|\boldsymbol{x}\|_a$ 相容的矩阵范数，并且称该矩阵范数是从属于向量范数 $\|\boldsymbol{x}\|$ 的算子范数。

定义2.3.2说了两件事情：算子范数是与定义它的向量范数是相容的；同时，说了这个公式代表的事算子范数，为什么算子范数是与定义它的向量范数是相容的呢？证明如下：

证明. 由算子范数的定义，有

$$\|\boldsymbol{A}\|_a = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_a}{\|\boldsymbol{x}\|_a}$$

那就会有如下关系：

$$\|\boldsymbol{A}\|_a = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_a}{\|\boldsymbol{x}\|_a} \geq \frac{\|\boldsymbol{A}\boldsymbol{x}\|_a}{\|\boldsymbol{x}\|_a}$$

化简一下，就会有

$$\|\boldsymbol{A}\boldsymbol{x}\|_a \leq \|\boldsymbol{A}\|_a \cdot \|\boldsymbol{x}\|_a$$

也就成功说明了，算子范数是与定义它的向量范数是相容的，证毕。

□

请注意，定义2.3.2还有另外一种形式，即

$$\|\boldsymbol{A}\|_a = \max_{\|\boldsymbol{u}\|_a=1} \|\boldsymbol{A}\boldsymbol{u}\|_a$$

两种形式的定义是怎么来的呢？推导如下：

$$\|\boldsymbol{A}\boldsymbol{x}\|_a = \|\boldsymbol{A} \cdot \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\|_a$$

令 $u = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_a}$，所以，原式变为

$$\max_{\|\boldsymbol{u}\|_a=1} \|\boldsymbol{A}\boldsymbol{u}\|_a$$

接下来，我们给出算子范数是范数的证明:

1. 正定性证明

若 $\boldsymbol{A}$ 不为 $\boldsymbol{0}$，则存在 $\boldsymbol{x_0} \neq \boldsymbol{0}$，使得 $\boldsymbol{A}\boldsymbol{x_0} \neq \boldsymbol{0}$

因此，$\|\boldsymbol{A}\boldsymbol{x_0}\|_a > 0$, 正定性证毕。

2. 齐次性证明

$$\|\lambda\boldsymbol{A}\|_a = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|\lambda|\boldsymbol{A}\boldsymbol{x}|\|_a}{\|\boldsymbol{x}\|_a}$$

$$\begin{aligned} &= \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{|\lambda|\|\boldsymbol{A}\boldsymbol{x}\|_a}{\|\boldsymbol{x}\|_a} \\ &= |\lambda| \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_a}{\|\boldsymbol{x}\|_a} \\ &= |\lambda|\|\boldsymbol{A}\|_a \end{aligned}$$

齐次性证毕。

3. 三角不等式证明

$$\begin{aligned}\|\boldsymbol{A}+\boldsymbol{B}\|_a &= \max_{\boldsymbol{x}\neq\boldsymbol{0}}\frac{\|(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x}\|}{\|\boldsymbol{x}\|_a} = \max_{\boldsymbol{x}\neq\boldsymbol{0}}\frac{\|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{x}\|_a}{\|\boldsymbol{x}\|_a}\\ &\leq \max_{\boldsymbol{x}\neq\boldsymbol{0}}\frac{\|\boldsymbol{A}\boldsymbol{x}\|_a+\|\boldsymbol{B}\boldsymbol{x}\|_a}{\|\boldsymbol{x}\|_a}\end{aligned}$$

令 $f(\boldsymbol{x})=\max\limits_{\boldsymbol{x}\neq\boldsymbol{0}}\frac{\|\boldsymbol{A}\boldsymbol{x}\|_a}{\|\boldsymbol{x}\|_a}, g(\boldsymbol{x})=\max\limits_{\boldsymbol{x}\neq\boldsymbol{0}}\frac{\|\boldsymbol{B}\boldsymbol{x}\|_a}{\|\boldsymbol{x}\|_a}$, 故

$$\max_{\boldsymbol{x}\neq\boldsymbol{0}}\frac{\|\boldsymbol{A}\boldsymbol{x}\|_a+\|\boldsymbol{B}\boldsymbol{x}\|_a}{\|\boldsymbol{x}\|_a}\leq f(\boldsymbol{x})+g(\boldsymbol{x})=\|\boldsymbol{A}\|_a+\|\boldsymbol{B}\|_b$$

至此，三角不等式证毕。

算子范数的相关定理

定理 2.3.1: 算子范数一定是自相容范数

设 $\|\boldsymbol{x}\|_a$ 是 $\mathbb{P}^n$ 上的向量范数，$\boldsymbol{A},\boldsymbol{B}\in\mathbb{P}^{n\times n}$, $\|\boldsymbol{A}\|_a$ 是从属于 $\|\boldsymbol{x}\|_a$ 的算子范数，则它是相容的矩阵范数，即

$$\|\boldsymbol{A}\boldsymbol{B}\|_a\leq\|\boldsymbol{A}\|_a\cdot\|\boldsymbol{B}\|_a$$

定理2.3.1表明，任何一个算子范数都是自相容范数，证明如下：

证明. 由算子范数定义，有：

$$\begin{aligned}\|\boldsymbol{A}\boldsymbol{B}\|_a &= \max_{\boldsymbol{x}\neq\boldsymbol{0}}\frac{\|\boldsymbol{A}\boldsymbol{B}\boldsymbol{x}\|_a}{\|\boldsymbol{x}\|_a}\\ &\leq \max_{\boldsymbol{x}\neq\boldsymbol{0}}\frac{\|\boldsymbol{A}\|_a\cdot\|\boldsymbol{B}\boldsymbol{x}\|_a}{\|\boldsymbol{x}\|_a}\\ &= \|\boldsymbol{A}\|_a\cdot\max_{\boldsymbol{x}\neq\boldsymbol{0}}\frac{\|\boldsymbol{B}\boldsymbol{x}\|_a}{\|\boldsymbol{x}\|_a}=\|\boldsymbol{A}\|_a\cdot\|\boldsymbol{B}\|_a\end{aligned}$$

证毕。 □

上面的定理与讨论当给定向量范数和矩阵时，就能确定与给定向量范数相容的矩阵范数，那如果给定矩阵范数时，能否找到与其相容的向量范数呢？

定理 2.3.2

设 $\|\cdot\|_m$ 是相容的矩阵范数，则存在向量范数 $\|\boldsymbol{x}\|$，使得

$$\|\boldsymbol{A}\boldsymbol{x}\|\leq\|\boldsymbol{A}\|_m\cdot\|\boldsymbol{x}\|$$

定理2.3.2的证明如下：

证明. 定义下列向量范数：

$$\|\boldsymbol{x}\| = \|\boldsymbol{x}\boldsymbol{a}^H\|_m, \quad (\boldsymbol{a} \neq \boldsymbol{0})$$

正定性：由于 $\boldsymbol{x} \neq \boldsymbol{0}$，故 $\boldsymbol{x}\boldsymbol{a}^H \neq \boldsymbol{0}$，因此 $\|\boldsymbol{x}\| = \|\boldsymbol{x}\boldsymbol{a}^H\|_m > 0$，正定性证毕。

齐次性：

$$\|\lambda\boldsymbol{x}\| = \|\lambda\boldsymbol{x}\boldsymbol{a}^H\|_m$$

右侧的 $\|\lambda\boldsymbol{x}\boldsymbol{a}^H\|_m$ 是一个矩阵范数，由矩阵范数的齐次性，有

$$\|\lambda\boldsymbol{x}\boldsymbol{a}^H\|_m = |\lambda|\|\boldsymbol{x}\boldsymbol{a}^H\|_m$$

齐次性证毕。

三角不等式：

$$\begin{aligned}\|\boldsymbol{x_1} + \boldsymbol{x_2}\| &= \|(\boldsymbol{x_1} + \boldsymbol{x_2})\boldsymbol{a}^H\|_m \\ &= \|\boldsymbol{x_1}\boldsymbol{a}^H + \boldsymbol{x_2}\boldsymbol{a}^H\|_m\end{aligned}$$

由矩阵范数的三角不等式，有

$$\|\boldsymbol{x_1}\boldsymbol{a}^H + \boldsymbol{x_2}\boldsymbol{a}^H\|_m \leq \|\boldsymbol{x_1}\boldsymbol{a}^H\|_m + \|\boldsymbol{x_2}\boldsymbol{a}^H\|_m = \|\boldsymbol{x_1}\| + \|\boldsymbol{x_2}\|$$

三角不等式证毕，因此，我们可以通过给定的矩阵范数找到其相容的向量范数 □

矩阵范数与特征值之间的关系

定理 2.3.3

如果 $\|\cdot\|_m : \mathbb{C}^{n\times n} \to \mathbb{R}$ 是一相容的矩阵范数，则对任一 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，有

$$|\lambda_i| \leq \|\boldsymbol{A}\|_m$$

其中，λ_i 是 $\boldsymbol{A}$ 的特征值。

定理2.3.3指出了矩阵范数和矩阵特征值之间的关系，由此我们可以矩阵特征值来判断矩阵范数，同理也可以根据矩阵范数判断矩阵特征值。

定理2.3.3的证明如下：

证明.

根据特征值的定义，对于矩阵 $\boldsymbol{A}$ 的任一特征值 λ_i，有

$$\boldsymbol{A}\boldsymbol{x} = \lambda_i \boldsymbol{x}$$

因此，有

$$\|\boldsymbol{A}\boldsymbol{x}\| = \|\lambda_i \boldsymbol{x}\| = |\lambda_i| \|\boldsymbol{x}\|$$

由定理2.3.2, 若矩阵范数相容，则必定存在一个向量的向量范数与之相容，故，有

$$\|\boldsymbol{A}\boldsymbol{x}\| \leq \|\boldsymbol{A}\|_m \|\boldsymbol{x}\|$$

因此，就有

$$|\lambda_i| \|\boldsymbol{x}\| \leq \|\boldsymbol{A}\|_m \|\boldsymbol{x}\|$$

由于 $\|\boldsymbol{x}\| \neq \boldsymbol{0}$，因此，两侧可以约掉 $\boldsymbol{x}$，于是式子就变成了

$$|\lambda_i| \leq \|\boldsymbol{A}\|_m$$

证毕。 □

2.3.3 算子范数的计算

本小节着重介绍算子范数的计算，由于特殊的定义关系，有时算子范数并不一定能够计算出来，本小节着重介绍算子 1-范数，算子 2-范数（也叫谱范数）和算子无穷范数的计算。

算子 1-范数的计算

从属于向量范数 $\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|$ 的算子范数为

$$\|\boldsymbol{A}\|_1 = \max_j \sum_{i=1}^{n} |a_{ij}|$$

又称为**极大列和范数**。

这个公式的意思是，首先先分别计算出每一列的绝对列和出来，然后再找出来最大的那个绝对列和，这就是 $\max\limits_j$ 的含义，而不是找到数字最大的那一列。

如何证明这个等式是正确的呢？往常证明等式成立往往是通过从等式的左侧或右侧出发，经过一系列的变形之后推导到了等式的右侧/左侧，证明了等式的成立，但在本节证明等式成立中，我们并不会采用这种方法，而是首先假设等式的左侧小于等于右侧，证明其成立，而后假设等式的左侧大于等于右侧，证明其成立，两者一结合便只有相等这一种选项。

证明.

假设 $\|\boldsymbol{A}\|_1 \leq \max\limits_j \sum_{i=1}^n |a_{ij}|$，首先证明该不等式的成立。

根据矩阵与向量的乘法，有

$$\|\boldsymbol{A}\boldsymbol{x}\|_1 = \sum_{i=1}^n |a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n|$$

根据绝对值三角不等式，有如下关系

$$\sum_{i=1}^n |a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n| \leq \sum_{i=1}^n \sum_{j=1}^n |a_{ij}||x_j|$$

由于多重累加可以交换顺序，因此上面的式子可以变成

$$\sum_{j=1}^n \sum_{i=1}^n |a_{ij}||x_j|$$

此时，注意到 $\sum_{i=1}^n |a_{ij}||x_j|$ 中的 $|x_j|$ 其实并不会受 j 的变化而变化，因此可以当成公因式提出，于是式子就成了

$$\sum_{j=1}^n (|x_j| \sum_{i=1}^n |a_{ij}|)$$

仔细观察 $\sum_{i=1}^n |a_{ij}|$，不难发现，这其实就是矩阵 $\boldsymbol{A}$ 的绝对列和，接下来，对式子做类似证明矩阵范数相容性那样进行放缩，有

$$\begin{aligned}\sum_{j=1}^n (|x_j| \sum_{i=1}^n |a_{ij}|) &\leq \sum_{j=1}^n (|x_j| \cdot \max_j \sum_{i=1}^n |a_{ij}|) \\ &= \max_j \sum_{i=1}^n |a_{ij}| \cdot \sum_{j=1}^n |x_j|\end{aligned}$$

注意到 $\sum_{j=1}^n |x_j|$ 其实就是向量 $\boldsymbol{x}$ 的向量 1-范数，因此有

$$\max_j \sum_{i=1}^n |a_{ij}| \cdot \sum_{j=1}^n |x_j| = \max_j \sum_{i=1}^n |a_{ij}| \cdot \|\boldsymbol{x}\|_1$$

整体的式子是这样的：

$$\|\boldsymbol{A}\boldsymbol{x}\|_1 \leq \max_j \sum_{i=1}^n |a_{ij}| \cdot \|\boldsymbol{x}\|_1$$

移项，于是就有了

$$\frac{\|\boldsymbol{A}\boldsymbol{x}\|_1}{\|\boldsymbol{x}\|_1} \leq \max_j \sum_{i=1}^n |a_{ij}|$$

或许会有人会发现，左侧 $\frac{\|\boldsymbol{A}\boldsymbol{x}\|_1}{\|\boldsymbol{x}\|_1}$ 缺少了 $\max\limits_{\boldsymbol{x}\neq\boldsymbol{0}}$，这是为什么呢？

因为在证明的过程中，我们假设的是对任意的 $\boldsymbol{x} \neq \boldsymbol{0}$ 都成立，因此不必加上 $\max\limits_{\boldsymbol{x}\neq\boldsymbol{0}}$

很好，我们顺利证明出了 $\|\boldsymbol{A}\|_1 \le \max\limits_j \sum_{i=1}^n |a_{ij}|$, 接下来继续证明 $\|\boldsymbol{A}\|_1 \ge \max\limits_j \sum_{i=1}^n |a_{ij}|$

由于涉及到某一列的绝对列和最大，所以不妨假设矩阵 $\boldsymbol{A}$ 的第 s 列的绝对列和最大，即矩阵 $\boldsymbol{A}$ 的形式为

$$\boldsymbol{A} = [\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \cdots, \boldsymbol{\alpha}_s, \cdots, \boldsymbol{\alpha_n}]$$

此时，取特殊向量 $\boldsymbol{x_0} = \boldsymbol{\varepsilon}_s$，这里的 $\boldsymbol{\varepsilon_s}$ 代表的是单位矩阵 $\boldsymbol{E}$ 的第 s 列的列向量

由向量范数以及单位矩阵的性质，不难得出下列等式关系

$$\|\boldsymbol{x_0}\|_1 = \|\boldsymbol{\varepsilon_s}\|_1 = 1$$

左乘一个矩阵 $\boldsymbol{A}$，有

$$\|\boldsymbol{A}\boldsymbol{x_0}\|_1 = \|\boldsymbol{A}\boldsymbol{\varepsilon}_s\|_1 = \|\boldsymbol{\alpha}_s\|_1 = \max_j \sum_{i=1}^n |a_{ij}|$$

由于 $\|\boldsymbol{x_0}\|_1 = 1$，因此对于上面的等式同时除以 $\|\boldsymbol{x_0}\|_1$，对式子结果不影响，但我们得到了

$$\frac{\|\boldsymbol{A}\boldsymbol{x_0}\|_1}{\|\boldsymbol{x_0}\|_1} = \max_j \sum_{i=1}^n |a_{ij}|$$

此时请注意，由算子范数最开始的定义，算子范数其实是需要找出

$$\max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_1}{\|\boldsymbol{x}\|_1}$$

这个值的范围是限定在整个矩阵中的，而非仅仅某一列或某一行，这是一种全局最大，但是看上面的证明，我们仅仅是找到了矩阵的某一行最大，是一种局部的最大值，全局最大值一定是大于等于局部最大值的，因此，我们有

$$\|\boldsymbol{A}\|_1 = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_1}{\|\boldsymbol{x}\|_1} \ge \frac{\|\boldsymbol{A}\boldsymbol{x_0}\|_1}{\|\boldsymbol{x_0}\|_1} = \max_j \sum_{i=1}^n |a_{ij}|$$

由此，我们证明了 $\|\boldsymbol{A}\|_1 \ge \max\limits_j \sum_{i=1}^n |a_{ij}|$，结合上面已经证明出的 $\|\boldsymbol{A}\|_1 \le \max\limits_j \sum_{i=1}^n |a_{ij}|$，我们就可以说，$\|\boldsymbol{A}\|_1 = \max\limits_j \sum_{i=1}^n |a_{ij}|$，证明完毕。

□

算子无穷范数的计算

从属于 $\|\boldsymbol{x}\|_\infty$ 的算子范数为

$$\|\boldsymbol{A}\|_\infty = \max_i (\sum_{j=1}^n |a_{ij}|)$$

也被称作**极大行和范数**.

与证明算子 1-范数的方式类似，我们依旧需要构建出两个不等式：$\|\boldsymbol{A}\|_\infty = \max\limits_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|\boldsymbol{Ax}\|_\infty}{\|\boldsymbol{x}\|_\infty} \geq \max\limits_i(\sum_{j=1}^n |a_{ij}|)$ 和 $\|\boldsymbol{A}\|_\infty = \max\limits_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|\boldsymbol{Ax}\|_\infty}{\|\boldsymbol{x}\|_\infty} \leq \max\limits_i(\sum_{j=1}^n |a_{ij}|)$，并依次证明成立，最后推出不等关系，证明过程如下：

证明. 首先证明

$$\|\boldsymbol{A}\|_\infty = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|\boldsymbol{Ax}\|_\infty}{\|\boldsymbol{x}\|_\infty} \leq \max_i(\sum_{j=1}^n |a_{ij}|)$$

与证明算子 1-范数的过程类似，有

$$\|\boldsymbol{Ax}\|_\infty = \max_i |a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n|$$

根据绝对值三角不等式，有如下关系

$$\max_i |a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n| \leq \max_i \left(\sum_{j=1}^n |a_{ij}||x_j|\right)$$

仔细观察 $\sum_{j=1}^n |a_{ij}|$，不难发现，这其实就是矩阵 $\boldsymbol{A}$ 的绝对行和，接下来，对式子做类似证明矩阵范数相容性那样进行放缩，有

$$\begin{aligned}
\max_i \left(\sum_{j=1}^n |a_{ij}||x_j|\right) &\leq \max_i \left(\sum_{j=1}^n |a_{ij}| \cdot \max_j |x_j|\right) \\
&= \max_i \sum_{j=1}^n |a_{ij}| \cdot \max_j |x_j| \\
&= \left(\max_i \sum_{j=1}^n |a_{ij}|\right) \cdot \|\boldsymbol{x}\|_\infty
\end{aligned}$$

整体的式子是这样的：

$$\|\boldsymbol{Ax}\|_\infty \leq \max_i \sum_{j=1}^n |a_{ij}| \cdot \|\boldsymbol{x}\|_\infty$$

移项，于是就有了

$$\frac{\|\boldsymbol{Ax}\|_\infty}{\|\boldsymbol{x}\|_\infty} \leq \max_i \sum_{j=1}^n |a_{ij}|$$

故，只要有 $\boldsymbol{x} \neq \boldsymbol{0}$，就会有上式成立

接下来，证明

$$\max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|\boldsymbol{Ax}\|_\infty}{\|\boldsymbol{x}\|_\infty} \geq \max_i(\sum_{j=1}^n |a_{ij}|)$$

但与假设最大绝对列和不同的是，我们其实并没有办法通过简单的一个矩阵乘一个行的形式来构造绝对行和，为此，我们采取下面的方式：

令 $\mu = \sum_{j=1}^{n}|a_{sj}| = \max\limits_{i}\sum_{j=1}^{n}|a_{ij}|$，其中 $a_{sj} = |a_{sj}| \cdot e^{i\theta j} \quad (j = 1, 2, \cdots, n)$ [1]

由于复数表示的性质，可以知道 $e^{i\theta j}$ 的值是永远小于等于 1 的，构造一个特殊的向量 $\boldsymbol{z}$，其形式为

$$\boldsymbol{z} = (e^{-i\theta_1}, e^{-i\theta_2}, \cdots, e^{-i\theta_n})$$

由欧拉公式可知，$\|e^{i\theta}\| = 1$，故 $\|\boldsymbol{z}\|_\infty = 1$

故 $\|\boldsymbol{Az}\|_\infty = \max\limits_{i}|a_{i1}e^{-i\theta_1} + \cdots + a_{in}e^{-i\theta_n}|$

假设矩阵 $\boldsymbol{A}$ 的第 s 行的绝对行和最大，即

$$\sum_{j=1}^{n}|a_{sj}| = \max_{i}\sum_{j=1}^{n}|a_{ij}|$$

于是我们有

$$\|\boldsymbol{Az}\|_\infty = \max_{i}|a_{i1}e^{-i\theta_1} + \cdots + a_{in}e^{-i\theta_n}| \geq |a_{s1}e^{-i\theta_1} + \cdots + a_{sn}e^{-i\theta_n}|$$

为什么这个不等关系会成立呢，因为 $\max\limits_{i}|a_{i1}e^{-i\theta_1} + \cdots + a_{in}e^{-i\theta_n}|$ 是所有分量的模的最大值，而 $|a_{s1}e^{-i\theta_1} + \cdots + a_{sn}e^{-i\theta_n}|$ 仅仅只是其中一个分量。

注意到

$$a_{si}e^{-i\theta_i} = |a_{si}|e^{i\theta_1} \cdot e^{-i\theta_1} = |a_{s1}|$$

所以

$$|a_{s1}e^{-i\theta_1} + \cdots + a_{sn}e^{-i\theta_n}| = |a_{s1}| + |a_{s2}| + \cdots + |a_{sn}| = \max_{i}\sum_{j=1}^{n}|a_{ij}|$$

故

$$\max_{\boldsymbol{x}\neq\boldsymbol{0}}\frac{\|\boldsymbol{Ax}\|_\infty}{\|\boldsymbol{x}\|_\infty} \geq \frac{\|\boldsymbol{Az}\|}{\|\boldsymbol{z}\|} = \max_{i}\sum_{j=1}^{n}|a_{ij}|$$

证毕。 □

算子 2-范数（谱范数）的计算

从属于 $\|\boldsymbol{x}\|_2$ 的算子范数（又称为谱范数）为

$$\|\boldsymbol{A}\|_2 = \sqrt{r(\boldsymbol{A}^H\boldsymbol{A})}$$

[1]这里涉及到了复数的表示形式和欧拉公式的内容，欧拉公式的表示为 $e^{i\theta} = \cos\theta + \sin\theta$，根据该公式，复数还可以使用 $z = \rho(\cos\theta + i\sin\theta)$ 的形式表示。

这里出现了一个新的东西：$r(\boldsymbol{A}^H\boldsymbol{A})$，这是什么东西呢，这个东西表示 $\boldsymbol{A}^H\boldsymbol{A}$ 的**谱半径**，谱半径的计算方法为：

$$r(\boldsymbol{A}) = \max_i |\lambda_i|$$

即，一个矩阵的谱半径的值就是其特征向量绝对值的最大值。

利用算子范数的定义，即算子 2-范数的计算公式为

$$\|\boldsymbol{A}\|_2 = \max_{\|\boldsymbol{u}\|_2=1} \|\boldsymbol{A}\boldsymbol{u}\|_2 = \sqrt{r(\boldsymbol{A}^H\boldsymbol{A})}$$

这里，依旧只需要证明

$$\max_{\|\boldsymbol{u}\|_2=1} \|\boldsymbol{A}\boldsymbol{u}\|_2 \geq \sqrt{r(\boldsymbol{A}^H\boldsymbol{A})}$$

和

$$\max_{\|\boldsymbol{u}\|_2=1} \|\boldsymbol{A}\boldsymbol{u}\|_2 \leq \sqrt{r(\boldsymbol{A}^H\boldsymbol{A})}$$

即可

这里要简单介绍一下 Hermite 矩阵的概念

定义 2.3.3: Hermite 矩阵

设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，若 $\boldsymbol{A}^H = \boldsymbol{A}$，则称 $\boldsymbol{A}$ 是 Hermite 矩阵；若 $\boldsymbol{A}^H = -\boldsymbol{A}$，则称 $\boldsymbol{A}$ 是反 Hermite 矩阵。

Hermite 矩阵还拥有如下性质 (这里只写后面需要用到的，具体更多的性质请看后面的相关内容)：

- $\boldsymbol{A}$ 是正定矩阵
- $\boldsymbol{A}$ 的特征值全为正实数

因此，对于这里的谱范数，我们不妨设矩阵 $\boldsymbol{A}^H\boldsymbol{A}$ 的特征值有如下关系：

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

这里 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 分别为单位特征向量 $x_1, x_2, \cdots, x_n$ 对应的特征值，又由于矩阵 $\boldsymbol{A}^H\boldsymbol{A}$ 是 Hermite 矩阵，故其所有的特征向量均正交。

故

$$r(\boldsymbol{A}^H\boldsymbol{A}) = \lambda_1$$

任取 $\boldsymbol{u} = a_1x_1 + a_2x_2 + \cdots + a_nx_n$，并且令 $\|\boldsymbol{u}\|_2^2 = 1$

因此，有

$$\|\boldsymbol{u}\|_2^2 = 1$$

证明. 先证

$$\max_{\|\boldsymbol{u}\|_2=1} \|\boldsymbol{A}\boldsymbol{u}\|_2 \leq \sqrt{r(\boldsymbol{A}^H\boldsymbol{A})}$$

由定义，有

$$\begin{aligned}
\|\boldsymbol{A}\boldsymbol{u}\|_2^2 &= \boldsymbol{u}^H\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{u} \\
&= \boldsymbol{u}^H\boldsymbol{A}^H\boldsymbol{A}(a_1x_1 + a_2x_2 + \cdots + a_nx_n) \\
&= \boldsymbol{u}^H(a_1\boldsymbol{A}^H\boldsymbol{A}x_1 + a_2\boldsymbol{A}^H\boldsymbol{A}x_2 + \cdots + a_n\boldsymbol{A}^H\boldsymbol{A}x_n) \\
&= \boldsymbol{u}^H(a_1\lambda_1x_1 + a_2\lambda_2x_2 + \cdots + a_n\lambda_nx_n) \\
&= (\bar{a}_1x_1^H + \cdots + \bar{a}_nx_n^H)\cdot(a_1\lambda_1x_1 + a_2\lambda_2x_2 + \cdots + a_n\lambda_nx_n) \\
&= \lambda_1|a_1|^2 + \lambda_2|a_2|^2 + \cdots + \lambda_n|a_n|^2
\end{aligned}$$

由于最开始的特征值之间的大小关系，我们有:

$$\begin{aligned}
\lambda_1|a_1|^2 + \lambda_2|a_2|^2 + \cdots + \lambda_n|a_n|^2 &\leq \lambda_1(|a_1|^2|a_2|^2 + \cdots + |a_n|^2) \\
&= \lambda_1\|\boldsymbol{u}\|_2^2 = \lambda_1
\end{aligned}$$

即

$$\|\boldsymbol{A}\boldsymbol{u}\|_2^2 \leq \lambda_1 = r(\boldsymbol{A}^H\boldsymbol{A})$$

接下来证明

$$\max_{\|\boldsymbol{u}\|_2=1} \|\boldsymbol{A}\boldsymbol{u}\|_2 \geq \sqrt{r(\boldsymbol{A}^H\boldsymbol{A})}$$

由特征值和特征向量的定义，有

$$\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{x_1} = \lambda_1\boldsymbol{x_1}$$

等式左右两侧同时乘 $\boldsymbol{x}_1^H$，有

$$\boldsymbol{x}_1^H\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{x}_1 = \lambda_1\boldsymbol{x}_1^H\boldsymbol{x_1}^H\boldsymbol{x_1} = \lambda_1$$

即

$$\max_{\|\boldsymbol{u}\|_2=1} \|\boldsymbol{A}\boldsymbol{u}\|_2^2 \geq \|\boldsymbol{A}\boldsymbol{x_1}\|_2^2 = \boldsymbol{x_1}^H\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{x_1} = \boldsymbol{x_1}^H\lambda_1\boldsymbol{x} = \lambda_1$$

证毕。

（数学真好玩，证明真有趣，我要天天学数学数学 ） □

谱范数的其他性质

首先介绍谱范数的酉不变性质

定理 2.3.4: 酉不变性质

设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，则

1. $\|\boldsymbol{A}\|_2 = \|\boldsymbol{A}^H\|_2 = \|\boldsymbol{A}^T\|_2 = \|\bar{\boldsymbol{A}}\|_2$
2. $\|\boldsymbol{A}^H\boldsymbol{A}\|_2 = \|\boldsymbol{A}\boldsymbol{A}^H\|_2 = \|\boldsymbol{A}\|_2^2$
3. 对任何 n 阶酉矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{V}$，都有

$$\|\boldsymbol{U}\boldsymbol{A}\|_2 = \|\boldsymbol{A}\boldsymbol{V}\|_2 = \|\boldsymbol{U}\boldsymbol{A}\boldsymbol{V}\|_2 = \|\boldsymbol{A}\|_2$$

上面三个性质的证明如下：

证明.

性质 1：

$\|\boldsymbol{A}\|_2 = \sqrt{r(\boldsymbol{A}^H\boldsymbol{A})}$ 因此，有 $\|\boldsymbol{A^H}\|_2 = \sqrt{r[(\boldsymbol{A}^H)^H\boldsymbol{A}^H]} = \sqrt{r(\boldsymbol{A}\boldsymbol{A}^H)} = \sqrt{r(\boldsymbol{A}^H\boldsymbol{A})} = \|\boldsymbol{A}\|_2$

$\|\boldsymbol{A^T}\|_2 = \sqrt{r[(\boldsymbol{A}^T)^H\boldsymbol{A}^T]} = \sqrt{r(\boldsymbol{A}\boldsymbol{A}^H)^T} = \sqrt{r(\boldsymbol{A}\boldsymbol{A}^H)} = \|\boldsymbol{A}^H\|_2$

由于 $\bar{\boldsymbol{A}} = (\boldsymbol{A^T})^H$，矩阵的转置与求共轭均不会影响矩阵的范数，因此显然相等，性质 1 证明完毕。

性质 2：

$\|\boldsymbol{A}^H\boldsymbol{A}\|_2 = \sqrt{r[(\boldsymbol{A}^H\boldsymbol{A})^H\boldsymbol{A}^H\boldsymbol{A}]} = \sqrt{r(\boldsymbol{A}^H\boldsymbol{A})^2} = r(\boldsymbol{A}^H\boldsymbol{A}) = \|\boldsymbol{A}\|_2^2$

性质 3：

性质 3 的证明方式与矩阵范数中的酉不变范数的证明方式类似，在这里不再赘述

□

接下来介绍谱范数的下一条性质：

定理 2.3.5

设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，则

- $\|\boldsymbol{A}\|_2 = \max\limits_{\|\boldsymbol{x}\|_2=\|\boldsymbol{y}\|_2=1} |\boldsymbol{y}^H\boldsymbol{A}\boldsymbol{x}|$
- $\|\boldsymbol{A}\|_2^2 \le \|\boldsymbol{A}\|_1\|\boldsymbol{A}\|_\infty$

定理2.3.5的证明如下:

对于第一条性质，采用的证明思路是这样的:

1. 首先证明右侧的值是一个上界
2. 其次证明这个上界是可以达到的

首先证明第一项，由 Cauchy 不等式，有

$$|\boldsymbol{y}^H(\boldsymbol{Ax})| = | < \boldsymbol{y}, \boldsymbol{Ax} > | \le \|\boldsymbol{y}\|_2 \cdot \|\boldsymbol{Ax}\|_2$$

由范数相容性可知

$$\|\boldsymbol{Ax}\|_2 \le \|\boldsymbol{A}\|_2 \|\boldsymbol{x}\|_2$$

因此，不等式左右两侧同乘 $\|\boldsymbol{y}\|_2$，有

$$\|\boldsymbol{y}\|_2 \cdot \|\boldsymbol{Ax}\|_2 \le \|\boldsymbol{y}\|_2 \cdot \|\boldsymbol{A}\|_2 \|\boldsymbol{x}\|_2$$

由于 $\|\boldsymbol{x}\|_2 = \|\boldsymbol{y}\|_2 = 1$ 因此

$$| < \boldsymbol{y}, \boldsymbol{Ax} > | \le \|\boldsymbol{A}\|_2$$

证明了右侧的确是一个上界，接下来证明该上界可以达到

由算子范数的定义

$$\|\boldsymbol{A}\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \|\boldsymbol{Ax}\|_2$$

因此，$\exists x_0$，有 $\|\boldsymbol{A}\|_2 = \|\boldsymbol{Ax_0}\|_2$

对 $\boldsymbol{Ax_0}$ 进行单位化, 有 $\boldsymbol{y}_0 = \frac{\boldsymbol{Ax_0}}{\|\boldsymbol{Ax_0}\|_2}$

因此，

$$|\boldsymbol{y}_0^H \cdot \boldsymbol{Ax_0}| = |\frac{(\boldsymbol{Ax_0})^H \boldsymbol{Ax_0}}{\|\boldsymbol{Ax_0}\|_2}|$$

分子恰好是 $\boldsymbol{Ax_0}$ 的内积，内积同时也是向量 2-范数，故

$$|\frac{(\boldsymbol{Ax_0})^H \boldsymbol{Ax_0}}{\|\boldsymbol{Ax_0}\|_2}| = \frac{\|\boldsymbol{Ax_0}\|_2^2}{\|\boldsymbol{Ax_0}\|_2} = \|\boldsymbol{Ax_0}\|_2 = \|\boldsymbol{A}\|_2$$

如此，证明可以达到这个上界

因此

$$\max_{\|\boldsymbol{x}\|_2=\|\boldsymbol{y}\|_2=1} |\boldsymbol{y}^H \boldsymbol{Ax}| = \|\boldsymbol{A}\|_2$$

第二条性质的证明:

$\|\boldsymbol{A}\|_2^2 = r(\boldsymbol{A}^H\boldsymbol{A})$

由于 $r(\boldsymbol{A}^H\boldsymbol{A}) = \max_i |\lambda_i|$，故根据定理2.3.3，有

$$r(\boldsymbol{A}\boldsymbol{A}^H) \leq \|\boldsymbol{A}^H\|_\infty \|\boldsymbol{A}\|_\infty$$

这里请回忆一下计算算子无穷范数的时候给出的计算方式，不难发现这里与标准算子无穷范数相比，多了一个转置共轭，算子无穷范数又叫行和范数，转置一下就变成了列和范数，这恰好就是算子 1-范数的形式，因此，上式可以变为

$$\|\boldsymbol{A}\|_1 \|\boldsymbol{A}\|_\infty$$

即

$$\|\boldsymbol{A}\|_2^2 = r(\boldsymbol{A}^H\boldsymbol{A}) \leq \|\boldsymbol{A}\|_1 \|\boldsymbol{A}\|_\infty$$

证毕。

2.4 酉不变范数

考试不考，上课没讲，略。

2.5 矩阵的测度

考试不考，上课没讲，略。

2.6 范数的应用

这一章虽然老师讲了，但是考试不考，主要就是希望各位能够培养一个严谨的精神，一个小小的误差可能会导致求出的结果与理想结果差距甚远，学习方法只需要看老师这一节 PPT 的几个例子即可。

第三章　矩阵的分解

矩阵分解是将一个矩阵分解为比较简单的或者具有某种特性的若干矩阵的和或者乘积，往往分解出的矩阵我们可以方便的研究其矩阵的秩、特征值、奇异值等信息，这为对于原始矩阵的研究或者处理带来极大的便利性。

本节着重会介绍下面的几种矩阵分解方法:

- 矩阵的三角分解
- 矩阵的谱分解
- 矩阵的满秩分解（最大秩分解）
- 矩阵的奇异值分解

注意：本节内容与上一章知识完全没有关系，但需要了解线性代数中对于矩阵的特征值、特征向量、实对称矩阵的性质等内容，如果遗忘了，请翻阅第零章的相关知识。

3.1　矩阵的三角分解

3.1.1　常见的三角矩阵及其性质

在开始本节内容之前，先来认识一下常见的三角矩阵

- 正线上三角阵

$$\boldsymbol{R}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

- 单位上三角阵

$$\boldsymbol{R}=\begin{bmatrix}1 & a_{12} & \cdots & a_{1n}\\ 0 & 1 & \cdots & a_{2n}\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & 1\end{bmatrix}$$

- 正线下三角阵

$$\boldsymbol{R}=\begin{bmatrix}a_{11} & 0 & \cdots & 0\\ a_{21} & a_{22} & \cdots & 0\\ \vdots & \vdots & & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{bmatrix}$$

- 单位下三角阵

$$\boldsymbol{R}=\begin{bmatrix}1 & 0 & \cdots & 0\\ a_{21} & a1 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ a_{n1} & a_{n2} & \cdots & 1\end{bmatrix}$$

常见的三角矩阵具有下面的性质:

定理 3.1.1: 三角阵的性质

1. 上 (下) 三角矩阵 $\boldsymbol{R}$ 的逆也为上 (下) 三角矩阵，对角元是原来元素的倒数。
2. 两个上 (下) 三角矩阵 $\boldsymbol{R_1}, \boldsymbol{R_2}$ 的乘积 $\boldsymbol{R_1 R_2}$ 也是上 (下) 三角矩阵。
3. 酉矩阵 $\boldsymbol{U}$ 的逆 $\boldsymbol{U}^{-1}$ 也是酉矩阵。
4. 两个酉矩阵之积 $\boldsymbol{U}_1\boldsymbol{U}_2$ 也是酉矩阵。

牢记上面的性质，这些性质会在后面反复用到。

3.1.2 n 阶矩阵的三角分解

定理 3.1.2

设 $\boldsymbol{A} \in \mathbb{C}_n^{n\times n}$，则 $\boldsymbol{A}$ 可以唯一的分解为

$$\boldsymbol{A} = \boldsymbol{U}_1\boldsymbol{R}$$

其中 $\boldsymbol{U}_1$ 是酉矩阵，$\boldsymbol{R}$ 是正线上三角矩阵。

或 $\boldsymbol{A}$ 可以唯一地分解为

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}_2$$

其中 $\boldsymbol{L}$ 是正线下三角矩阵，$\boldsymbol{U}_2$ 是酉矩阵

这就是 n 阶矩阵的三角分解。证明过程//TODO,（后面如果没时间了就看看课本吧，这个证明可以看看）

除此之外，还有下面的推论：

推论 3.1.3

设 $\boldsymbol{A} \in \mathbb{R}_n^{n\times n}$，则 $\boldsymbol{A}$ 可以唯一地分解为

$$\boldsymbol{A} = \boldsymbol{Q}_1\boldsymbol{R}$$

其中，$\boldsymbol{Q}_1$ 是正交矩阵，$\boldsymbol{R}$ 是正线上三角矩阵。

或，$\boldsymbol{A}$ 可以唯一分解为

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{Q}_2$$

其中，$\boldsymbol{L}$ 是正线下三角矩阵，$\boldsymbol{Q}_2$ 是正交矩阵.

推论 3.1.4

设 $\boldsymbol{A}$ 是实对称正定矩阵，则存在唯一的正线上三角矩阵，使得

$$\boldsymbol{A} = \boldsymbol{R}^T\boldsymbol{R}$$

推论 3.1.5

设 $\boldsymbol{A}$ 是正定 Hermite 矩阵，则存在唯一的正线上三角矩阵，使得

$$\boldsymbol{A} = \boldsymbol{R}^H\boldsymbol{R}$$

证明省略，后面的不考，本小节结束。

3.1.3 任意矩阵的三角分解

定理 3.1.6

设 $\boldsymbol{A}$ 为行满秩矩阵或者列满秩矩阵，则

- 设 $\boldsymbol{A} \in \mathbb{C}_n^{m\times n}$，则存在 m 阶酉矩阵 $\boldsymbol{U}$ 即 n 阶正线上三角矩阵 $\boldsymbol{R}$，使得

$$\boldsymbol{A} = \boldsymbol{U}\begin{bmatrix}\boldsymbol{R}\\ 0\end{bmatrix}$$

- 设 $\boldsymbol{A} \in \mathbb{C}_m^{m\times n}$，则存在 n 阶酉矩阵 $\boldsymbol{U}$ 即 m 阶正线下三角矩阵 $\boldsymbol{L}$，使得

$$\boldsymbol{A} = (\boldsymbol{L} \quad 0)\boldsymbol{U}$$

最后还有一个较为重要的在后面奇异值分解中会用到的定理：

定理 3.1.7

设 $\boldsymbol{A} \in \mathbb{C}_r^{m\times n}$，则存在酉矩阵 $\boldsymbol{U} \in \mathbb{U}m \times m$ 和 $\boldsymbol{V} \in \mathbb{U}^{n\times n}$ 及 r 阶正线下三角矩阵 $\boldsymbol{L}$，使得

$$\boldsymbol{A} = \boldsymbol{U}\begin{bmatrix}\boldsymbol{L} & 0\\ 0 & 0\end{bmatrix}\boldsymbol{V}$$

3.2 矩阵的谱分解

接下来了解一下矩阵的谱分解，矩阵的谱分解与矩阵的特征值与特征向量关系密切。

3.2.1 单纯矩阵的谱分解

在介绍单纯矩阵的谱分解前，需要明白何为单纯矩阵，而想要明白何为单纯矩阵，又需要了解下面的两个概念：代数重复度和几何重复度。

定义 3.2.1: 代数重复度

设 $\lambda_1, \lambda_2, \cdots, \lambda_k$ 是 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ 的相异特征值，其重数分别为 $r_1, r_2, \cdots, r_k$，则称 r_i 为矩阵 $\boldsymbol{A}$ 的特征值 λ_i 的**代数重复度**。

代数重复度与第一章提到的代数重数概念十分相似，其实他们就是一回事，如果能明白这个关系，下面的几何重复度理解起来也就没什么难度了。

定义 3.2.2: 几何重复度

齐次方程组 $\boldsymbol{A}\boldsymbol{x} = \lambda_i \boldsymbol{x}\ (i = 1, 2, \cdots, k)$ 的解空间 $\boldsymbol{V}_{\lambda_i}$ 称为 $\boldsymbol{A}$ 的对应于特征值 λ_i 的特征空间，则 $\boldsymbol{V}_{\lambda_i}$ 的维数称为 $\boldsymbol{A}$ 的特征值 λ_i 的**几何重复度**。

了解了这两个概念，接下来就要介绍何为单纯矩阵了：

定义 3.2.3

若矩阵 $\boldsymbol{A}$ 的每个特征值的代数重复度与几何重复度相等，则称矩阵 $\boldsymbol{A}$ 为**单纯矩阵**。

在了解了何为单纯矩阵之后，接下来介绍何为单纯矩阵的谱分解：

定理 3.2.1: 单纯矩阵谱分解

设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ 是单纯矩阵，则 $\boldsymbol{A}$ 可分解为一系列幂等矩阵 $\boldsymbol{A}_i\ (i = 1, 2, \cdots, n)$ 的加权和，即

$$\boldsymbol{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{A}_i$$

其中，$\lambda_i\ (i = 1, 2, \cdots, n)$ 是 $\boldsymbol{A}$ 的特征值，并且 $\boldsymbol{A}_i$ 还有如下的性质：

1. 幂等性：$\boldsymbol{A}_i^2 = \boldsymbol{A}_i$
2. 分离性：$\boldsymbol{A}_i\boldsymbol{A}_j = 0\ (i \neq j)$
3. 可加性：$\sum\limits_{i=1}^{n} \boldsymbol{A}_i = \boldsymbol{E}_n$

这些性质的证明就略过了，感兴趣可以翻翻书。

上面介绍的是如何把单纯矩阵进行谱分解，下面的定理说的就是谱分解中的矩阵 $\boldsymbol{A}_i$ 与单纯矩阵之间的关系：

定理 3.2.2

设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，它有 k 个相异特征值 $\lambda_i\ (i=1,2,\cdots,k)$，则 $\boldsymbol{A}$ 是单纯矩阵的充要条件是存在 k 个矩阵 $\boldsymbol{A}_i\ (i=1,2,\cdots,k)$ 满足

1. $\boldsymbol{A}_i\boldsymbol{A}_j = \begin{cases} \boldsymbol{A}_i, & i=j \\ 0, & i\neq j \end{cases}$
2. $\sum_{i=1}^{k}\boldsymbol{A}_i = \boldsymbol{E}_n$
3. $\boldsymbol{A} = \sum_{i=1}^{k}\lambda_i\boldsymbol{A}_i$

3.2.2 正规矩阵及其分解

了解了单纯矩阵的分解之后，下面是正规矩阵的分解

定义 3.2.4: 正规矩阵定义

若 n 阶复矩阵 $\boldsymbol{A}$ 满足

$$\boldsymbol{A}\boldsymbol{A}^H = \boldsymbol{A}^H\boldsymbol{A}$$

则称 $\boldsymbol{A}$ 为**正规矩阵**。

正规矩阵有下面的性质：

1. 若 $\boldsymbol{A}$ 为正规矩阵且 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 酉相似，则 $\boldsymbol{B}$ 也为正规矩阵
2. (Schur 定理) 设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，则存在酉矩阵 $\boldsymbol{U}$，使得

   $$\boldsymbol{A} = \boldsymbol{U}\mathbb{R}\boldsymbol{U}^H$$

   其中，$\mathbb{R}$ 是一个上三角矩阵且主对角线上的元素为 $\boldsymbol{A}$ 的特征值。
3. 设 $\boldsymbol{A}$ 是正规矩阵且是三角矩阵，则 $\boldsymbol{A}$ 是对角矩阵。

什么样的矩阵能够成为正规矩阵呢？

定理 3.2.3

n 阶复矩阵 $\boldsymbol{A}$ 是正规矩阵的充要条件是 $\boldsymbol{A}$ 与对角阵酉相似，即存在 n 阶酉矩阵 $\boldsymbol{U}$，使得

$$\boldsymbol{A}=\boldsymbol{U}diag(\lambda_1,\lambda_2,\cdots,\lambda_n)\boldsymbol{U}^H$$

其中，$\lambda_1,\lambda_2,\cdots,\lambda_n$ 是 $\boldsymbol{A}$ 的 n 个特征值。

或者可以用下面这个定理来判断：

定理 3.2.4

设 $\boldsymbol{A}\in\mathbb{C}^{n\times n}$，它有 k 个相异特征值 $\lambda_i\ (i=1,2,\cdots,k)$，则 $\boldsymbol{A}$ 是单纯矩阵的充要条件是存在 k 个矩阵 $\boldsymbol{A}_i\ (i=1,2,\cdots,k)$ 满足

1. $\boldsymbol{A}_i\boldsymbol{A}_j=\begin{cases}\boldsymbol{A}_i, & i=j\\ 0, & i\neq j\end{cases}$
2. $\sum_{i=1}^{k}\boldsymbol{A}_i=\boldsymbol{E}_n$
3. $\boldsymbol{A}=\sum_{i=1}^{k}\lambda_i\boldsymbol{A}_i$
4. $\boldsymbol{A}_i^H=\boldsymbol{A}_i\ (i=1,2,\cdots,k)$

3.3 矩阵的满秩分解

3.3.1 满秩分解的定义

在了解了矩阵的谱分解之后，再来看一种较为简单的分解——矩阵的满秩分解，也叫最大秩分解。

定理 3.3.1

设 $\boldsymbol{A}\in\mathbb{C}_r^{m\times n}$，则存在矩阵 $\boldsymbol{B}\in\mathbb{C}_r^{m\times r},\boldsymbol{D}\in\mathbb{C}_r^{r\times n}$，使得

$$\boldsymbol{A}=\boldsymbol{B}\boldsymbol{D}$$

满秩分解很好理解，即，把一个矩阵分解成一个行满秩矩阵和一个列满秩矩阵相乘的形式即可。

3.3.2 满秩分解的步骤

对一个矩阵进行满秩分解的步骤如下:

1. 进行初等行变换,化为**行最简阶梯形矩阵** $\widetilde{\boldsymbol{A}}$
2. 找到 $\boldsymbol{A}$ 的一个极大线性无关组,这就是列满秩矩阵 $\boldsymbol{B}$
3. 在 $\widetilde{\boldsymbol{A}}$ 中的所有非零行构成矩阵 $\boldsymbol{D}$

3.4 矩阵的奇异值分解

在本节的最后了解一下矩阵的奇异值分解,矩阵的奇异值分解常用来进行图像处理,对图像进行压缩,因此具有较为重要的应用场景。

3.4.1 为什么需要进行奇异值分解?

前面对于矩阵的分解,大多数都有要求,要求矩阵需要是方阵,这就又对矩阵的种类进行了限制,如果想要对于任意的矩阵进行分解,由第一节的最后一个定理可知,存在酉矩阵 $\boldsymbol{U},\boldsymbol{V}$ 和 r 阶上三角矩阵或者下三角矩阵 $\boldsymbol{R}(L)$,使得

$$\boldsymbol{A}=\boldsymbol{U}\begin{bmatrix}\boldsymbol{R} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}\boldsymbol{V}$$

但这样子依旧要求是方阵,奇异值分解就是在这个基础上而来。

如何把一个矩阵变成方阵呢?这里采用的方法是研究 $\boldsymbol{A}^H\boldsymbol{A}$,这的确变成了一个方阵,可为什么要这样处理呢?这就要给出下面的定理

定理 3.4.1

设 $\boldsymbol{A}\in\mathbb{C}_r^{m\times n}$,则有

1. $\operatorname{rank}(\boldsymbol{A})=\operatorname{rank}(\boldsymbol{A}^H\boldsymbol{A})=\operatorname{rank}(\boldsymbol{A}\boldsymbol{A}^H)$
2. $\boldsymbol{A}^H\boldsymbol{A},\boldsymbol{A}\boldsymbol{A}^H$ 的特征值均为非负实数
3. $\boldsymbol{A}^H\boldsymbol{A},\boldsymbol{A}\boldsymbol{A}^H$ 的非零特征值相同

上面的定理是为下面矩阵的奇异值做铺垫,上面的这些定理仿佛在说:"我们可以用 $\boldsymbol{A}^H\boldsymbol{A}$ 或者 $\boldsymbol{A}\boldsymbol{A}^H$ 来研究 $\boldsymbol{A}$"。

3.4.2 奇异值

定义 3.4.1

设 $\boldsymbol{A} \in \mathbb{C}_r^{m\times n}, \boldsymbol{A}^H\boldsymbol{A}$ 的特征值为

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{r+1} = \cdots = \lambda_n = 0$$

则称 $\sigma_i = \sqrt{\lambda_i}\ (i = 1, 2, \cdots, r)$ 为 $\boldsymbol{A}$ 的**正奇异值**。

有了奇异值的概念之后，就有了下面奇异值分解的定理：

定理 3.4.2: 奇异值分解

设 $\boldsymbol{A} \in \mathbb{C}_r^{m\times n}, \sigma_1, \sigma_2, \cdots, \sigma_r$ 是 $\boldsymbol{A}$ 的 r 个正奇异值，则存在酉矩阵 $\boldsymbol{U} \in \mathbb{C}^{m\times m}$ 和 $\boldsymbol{V} \in \mathbb{C}^{n\times n}$，使得

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{D} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix} \boldsymbol{V}$$

其中,$\boldsymbol{D} = diag(\sigma_1, \sigma_2, \cdots, \sigma_r), |\sigma_i| = \sigma_i$

这个定理与第一节最后的定理很类似，不同之处就是将左上角的元素换成了对角矩阵。最后简单介绍一下酉等价的概念，以及两个酉等价矩阵的奇异值关系。

定义 3.4.2

设 $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{m\times n}$，如果存在酉矩阵 $\boldsymbol{U} \in \mathbb{C}^{m\times m}$ 和 $\boldsymbol{V} \in \mathbb{C}^{n\times n}$，使得

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{B}\boldsymbol{V}$$

则称 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 酉等价，若两个矩阵酉等价，则他们有相同的正奇异值。

3.4.3 求矩阵奇异值分解的方法

求矩阵 $\boldsymbol{A}$ 的奇异值分解的方法如下：

1. 求 $\boldsymbol{A}^H\boldsymbol{A}$ 的特征值以及特征向量
2. 构造酉矩阵 $\boldsymbol{V}$
3. 构造酉矩阵 $\boldsymbol{U}$

4. 给出矩阵 $\boldsymbol{A}$ 的奇异值分解

第四章　特征值的估计与摄动

在学习线性代数的时候，我们了解过如何求一个矩阵的特征值，并得到与该特征值对应的特征向量，如，对一个 n 阶方阵 $\boldsymbol{A}$，求特征值的方法是构建特征方程

$$|\lambda \boldsymbol{E} - \boldsymbol{A}| = 0$$

所得到的 λ_i 便是求解出的特征值。

这种构建特征方程求解特征值的方法求出来的解尽管精确，但有很大的局限性，如果在面对阶数较大的矩阵中，该方法需要庞大的计算量才可以计算出矩阵的特征值，同时，面对高阶行列式，对其进行处理，判断其解的情况也是一件需要耗费功夫的事情。

幸运的是，在工程中，很多情况下我们往往并不需要了解矩阵特征值精确解，而只需要了解该特征值的一个大概范围即可，这就是本章要讨论的问题——特征值的估计。

由于考试，本章只介绍 Gerschgorin(盖尔) 圆盘定理的知识，本章的其他内容如果有机会会在后面补上。

4.1 Gerschgorin(盖尔) 圆盘定理

本节所涉及的圆盘定理可以帮助确定一个矩阵的特征值的大致范围。

4.1.1 第一圆盘定理

行盖尔圆盘和列盖尔圆盘

在介绍第一圆盘定理前，需要了解何为行盖尔圆盘和列盖尔圆盘，这两个概念会在本节内容中反复提到，因此需要掌握。

定义 4.1.1: 行盖尔圆盘和列盖尔圆盘

设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，则

$$S_i = \{z \in \mathbb{C} : |z - a_{ii}| \leq R_i\}$$

为矩阵 $\boldsymbol{A}$ 在复平面上的第 i 个行盖尔圆，其中

$$R_i = R_i(\boldsymbol{A}) = \sum_{j\neq i} |a_{ij}| \quad (i = 1, 2, \cdots, n)$$

叫做 S_i 的**半径**。

类似的

$$G_i = \{z \in \mathbb{C} : |z - a_{ii}| \leq C_i\}, C_i = \sum_{i\neq j} |a_{ji}|$$

叫做列盖尔圆，半径为 C_i

可能一下子看到这个定义不是很理解，下面给出通俗的说法

行盖尔圆：以矩阵对角线元素为圆心，当前行其余元素的绝对值的和为半径的一个圆。

列盖尔圆：以矩阵对角线元素为圆心，当前列其余元素的绝对值的和为半径的一个圆。

有了这两个概念，才有下面的第一圆盘定理：

定理 4.1.1: 第一圆盘定理

设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，则 $\boldsymbol{A}$ 的任一特征值

$$\lambda_i \in S = \bigcup_{j=1}^{n} S_j \quad (i = 1, 2, \cdots, n)$$

定理4.1.1表明：矩阵 $\boldsymbol{A}$ 的任意特征值都在行盖尔圆中，证明过程就不写在这里了，感兴趣的可以翻阅教材。

下面通过一个例题具体感受一下第一圆盘定理的应用：

例 4.1.1. 估计矩阵

$$\boldsymbol{A}=\begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{3}{2} & i & 0 \\ 0 & -\frac{i}{2} & 5 & -\frac{i}{2} \\ -1 & 0 & 0 & 5i \end{bmatrix}$$

的特征值的分布范围 (i 为虚数单位)

解. 根据第一圆盘定理，可以画出如下的 4 个行盖尔圆：

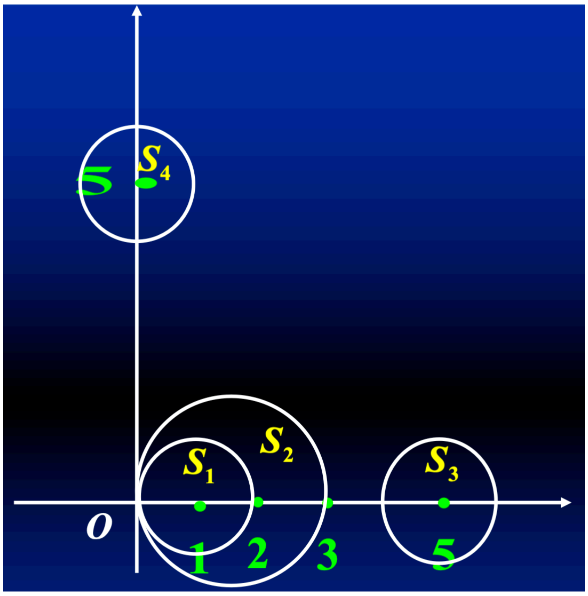
$$S_1:|z-1|\leq 1$$

$$S_2:|z-\frac{3}{2}|\leq\frac{3}{2}$$

$$S_3:|z-5|\leq 1$$

$$S_4:|z-5i|\leq 1$$

将这四个盖尔圆画出来就如下图所示：



这就是如何画出盖尔圆，这个矩阵的特征值就会分布在盖尔圆内。 □

同样的，我们还有如下推论：

推论 4.1.2

设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，则 $\boldsymbol{A}$ 的任一特征值

$$\lambda_i \in \bigcup_{j=1}^{n} G_i$$

也就是说，矩阵的特征值同样分布在列盖尔圆内。

4.1.2 第二圆盘定理

第一圆盘定理只说明了矩阵 $\boldsymbol{A}$ 的特征值都在盖尔圆中，但并没有说哪个连通部分有哪些特征值，要让特征值的分布更明确，就需要第二圆盘定理。

定理 4.1.3: 第二圆盘定理

设 n 阶方阵 $\boldsymbol{A}$ 的 n 个盖尔圆盘中有 k 个圆盘的并形成一个连通区域 $\boldsymbol{G}$，且它与余下的 $n-k$ 个圆盘都不相交，则在该区域 $\boldsymbol{G}$ 中恰好有 $\boldsymbol{A}$ 的 k 个特征值。

这里仅仅解释一下何为"连通"，关于该定理的证明就不赘述了，详细可以看 PPT 或者课本。

以例4.1.1画出的盖尔圆为例，S_1, S_2 重叠在一起，它们的并集是一个连通区域，交结在一起的盖尔圆所构成的连通区域为并集 S 的一个连通部分，孤立的盖尔圆就是一个连通部分，所以，S_1 与 S_2 的并集，即 S_2、S_3、S_4 各是一个连通部分。

4.1.3 圆盘定理的其他推论

接下来介绍的圆盘定理的其他推论更有助于借助圆盘来判断矩阵特征值的情况：

推论 4.1.4

设 n 阶方阵 $\boldsymbol{A}$ 的 n 个盖尔圆盘两两互不相交，则 $\boldsymbol{A}$ 相似于对角阵。

如果 n 个盖尔圆盘两两互不相交，则说明每一个盖尔圆盘就包含一个矩阵的特征值，这也就是说 n 个盖尔圆盘就会有 n 个互不相同的特征值，由线性代数的知识可知，矩阵 $\boldsymbol{A}$ 可以相似对角化。

推论 4.1.5

设 n 阶实矩阵 $\boldsymbol{A}$ 的 n 个盖尔圆盘两两互不相交，则 $\boldsymbol{A}$ 的特征值全为实数。

如果 $\boldsymbol{A}$ 不是实数而是复数的话，圆盘里面会有两个特征值，这与第二圆盘定理相矛盾。

4.1.4 缩小特征值的范围

圆盘定理虽然可以帮助确定特征值的范围，但这个范围我们还可以通过一些处理进行缩小，这就要利用到下面在线性代数中学到的性质：对于一个矩阵 $\boldsymbol{A}$，乘上一个可逆矩阵，不改变矩阵特征值。

对于一个矩阵 $\boldsymbol{A}$，我们做出如下处理：左乘一个可逆矩阵 $\boldsymbol{S}^{-1}$，右乘一个 $\boldsymbol{S}$，于是，对下面的一个新的矩阵画盖尔圆盘:

$$\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$$

如果对于 $\boldsymbol{S}$ 的选择较为恰当，是能够得到更精确的特征值范围的，通常，一个比较特别的选择是选取

$$\boldsymbol{S}=\boldsymbol{D}=diag(p_1,p_2,\cdots,p_n)\quad p_i>0\ (i=1,2,\cdots,n)$$

这样的一个对角矩阵，于是，新的矩阵就成了

$$\begin{aligned}
\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}&=\begin{bmatrix} p_1^{-1} & & & \\ & p_2^{-1} & & \\ & & \ddots & \\ & & & p_4^{-1}\end{bmatrix}\cdot\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{bmatrix}\cdot\begin{bmatrix} p_1 & & & \\ & p_2 & & \\ & & \ddots & \\ & & & p_4\end{bmatrix}\\
&=\begin{bmatrix} \frac{a_{11}}{p_1} & \frac{a_{12}}{p_1} & \cdots & \frac{a_{1n}}{p_1} \\ \frac{a_{21}}{p_2} & \frac{a_{22}}{p_2} & \cdots & \frac{a_{2n}}{p_2} \\ \vdots & \vdots & & \vdots \\ \frac{a_{n1}}{p_n} & \frac{a_{n2}}{p_n} & \cdots & \frac{a_{nn}}{p_n}\end{bmatrix}\cdot\begin{bmatrix} p_1 & & & \\ & p_2 & & \\ & & \ddots & \\ & & & p_4\end{bmatrix}\\
&=\begin{bmatrix} a_{11} & \frac{p_2}{p_1}a_{12} & \cdots & \frac{p_n}{p_1}a_{1n} \\ \frac{p_1}{p_2}a_{21} & a_{22} & \cdots & \frac{p_n}{p_2}a_{2n} \\ \vdots & \vdots & & \vdots \\ \frac{p_1}{p_n}a_{n1} & \frac{p_2}{p_n}a_{n2} & \cdots & a_{nn}\end{bmatrix}
\end{aligned}$$

这个时候下的盖尔圆的半径可以受 $\boldsymbol{S}$ 矩阵的对角元素来影响，可以看到，以第一行为例，可以增大 p_1 的值，减小 $p_2, \cdots, p_n$ 的值，这个时候盖尔圆 S_1 就会变小，可相对应的会发现，$S_2, \cdots, S_n$ 的半径就会变大。

采用这种方式和正常求盖尔圆的对于特征值影响到底有多大呢？请看下面的例子：

例 4.1.2. 估计矩阵

$$\boldsymbol{A} = \begin{bmatrix} 0.9 & 0.01 & 0.12 \\ 0.01 & 0.8 & 0.13 \\ 0.01 & 0.02 & 0.4 \end{bmatrix}$$

的特征值的范围。

解.

如果采用原始方法，求解的盖尔圆的信息如下：

$$S_1 : |z - 0.9| \leq 0.13$$

$$S_2 : |z - 0.8| \leq 0.14$$

$$S_3 : |z - 0.4| \leq 0.03$$

若取 $p_1 = p_2 = 1, p_3 = 0.1$，则

$$\boldsymbol{D} = diag(1, 1, 0.1)$$

新的盖尔圆盘信息就成了：

$$S_1' : |z - 0.9| \leq 0.022$$

$$S_2' : |z - 0.8| \leq 0.023$$

$$S_3' : |z - 0.4| \leq 0.3$$

对比一下就会发现，前面两个盖尔圆盘 (S_1, S_2) 的半径的确变小了，代表特征值的范围更加精确，但相对应的，S_3 的半径变大，特征值的范围更加模糊。 □

4.1.5 盖尔圆盘的其他性质

在本节的最后，介绍一下其他盖尔圆盘的性质

行对角占优和列对角占优

定义 4.1.2: 行对角占优和列对角占优

设 $\boldsymbol{a} \in \mathbb{C}^{n\times n}$，则

1. 若 $|a_{ii}| \geq R_i$，则称该矩阵为**行对角占优**。
2. 若 $|a_{ii}| \geq C_i$，则称该矩阵为**列对角占优**。

如果去掉等于号，上面的两个就成了**行严格对角占优**和**列严格对角占优**。

这两个概念是为了下面的定理服务的，这里不给出证明，感兴趣的可以自己翻课本

定理 4.1.6

设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ 行（或列）严格对角占优，则

1. $\boldsymbol{A}$ 可逆，且 $\lambda_i \in \bigcup_{i=1}^{n} S_i \quad (S_i = \{z \in \mathbb{C} : |z - a_{ii}| \leq |a_{ii}|\})$
2. 若 $\boldsymbol{A}$ 的所有主对角元都为正数，则 $\boldsymbol{A}$ 的特征值都有正实部
3. 若 $\boldsymbol{A}$ 为 Hermite 矩阵，且所有主对角元都为正数，则 $\boldsymbol{A}$ 的特征值都为正数。

第五章　矩阵分析

在学习微积分的课程中，我们对于函数研究过其极限、求导、求积分等操作，能否将这些数学工具应用到矩阵中呢？这就是本章矩阵分析要干的事情。

本章内容包括：

- 矩阵序列与矩阵级数
- 矩阵函数

5.1　矩阵序列与矩阵级数

5.1.1　矩阵序列

从向量序列到矩阵序列

在第二章介绍向量范数的应用的时候，我们曾经接触过向量序列的敛散性问题：

$$\lim_{k\to\infty} x^k = \lim_{k\to\infty} \|x^k - a\| = 0$$

也在相对应的地方介绍了何为向量序列，向量序列也是通过数列的情况延伸而来，再对向量序列进行延伸，便有了矩阵序列的概念

定义 5.1.1: 矩阵序列

设 $m \times n$ 型矩阵序列为 $\boldsymbol{A}^{(k)}$，其中

$$\boldsymbol{A}^{(k)} = \begin{bmatrix} a_{11}^{(k)} & a_{12}^{(k)} & \cdots & a_{1n}^{(k)} \\ a_{21}^{(k)} & a_{22}^{(k)} & \cdots & a_{2n}^{(k)} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}^{(k)} & a_{m1}^{(k)} & \cdots & a_{mn}^{(k)} \end{bmatrix}, \quad k = 1, 2, \cdots$$

矩阵序列的极限

与向量序列的极限类似，矩阵序列的极限定义如下：

定义 5.1.2

若

$$\lim_{k\to\infty} a_{ij}^{(k)} = a_{ij}$$

则

$$\lim_{k\to\infty} \boldsymbol{A}^{(k)} = \boldsymbol{A}$$

矩阵序列极限的性质

定理 5.1.1

设 $\lim\limits_{k\to\infty} \boldsymbol{A}^{(k)=\boldsymbol{A}}$, $\lim\limits_{k\to\infty} \boldsymbol{B}^{(k)=\boldsymbol{B},\alpha,\beta\in\mathbb{C}}$ 则

1. $\lim\limits_{k\to\infty}(\alpha\boldsymbol{A}^{(k)} + \beta\boldsymbol{B}^{(k)}) = \alpha\boldsymbol{A} + \beta\boldsymbol{B}$
2. $\lim\limits_{k\to\infty} \boldsymbol{A}^{(k)}\boldsymbol{B}^{(k)} = \boldsymbol{A}\boldsymbol{B}$
3. 当 $\boldsymbol{A}^(k)$ 与 $\boldsymbol{A}$ 都可逆时，$\lim\limits_{k\to\infty}(\boldsymbol{A}^{(k)})^{-1} = \boldsymbol{A}^{-1}$

下面的定理给出矩阵序列收敛条件：

定理 5.1.2

设 $\|\cdot\|$ 是 $\mathbb{C}^{m\times n}$ 上任一矩阵范数，$\mathbb{C}^{m\times n}$ 中矩阵序列收敛于 $\boldsymbol{A}$ 的充要条件是

$$\lim_{k\to\infty} \|\boldsymbol{A}^{(k)} - \boldsymbol{A}\| = 0$$

与向量序列的收敛判断类似，定理5.1.2指出了矩阵序列收敛条件，下面的定义将会指出何为收敛矩阵。

定义 5.1.3

设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，若 $\lim\limits_{k\to\infty} \boldsymbol{A}^k = 0$($k$为正整数)，则称 $\boldsymbol{A}$ 为**收敛矩阵**.

下面举一个例子，来表示何为收敛矩阵

例 5.1.1.

$$A^{(k)} = \begin{bmatrix} \frac{1}{2^k} & 0 \\ 0 & \frac{1}{2^k} \end{bmatrix}$$

我们知道

$$\lim_{k\to\infty} \frac{1}{2^k} = 0$$

自然就可以推出

$$\lim_{k\to\infty} \boldsymbol{A}^k = 0$$

就代表矩阵 $\boldsymbol{A}$ 是一个收敛矩阵。

下面介绍一下矩阵能够成为收敛矩阵的条件

定理 5.1.3

设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，则 $\boldsymbol{A}$ 为收敛矩阵的充要条件是谱半径 $r(\boldsymbol{A}) < 1$

证明如下：

证明.

充分性证明：

由 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，再由 Jordan 标准型的知识可知，存在矩阵 $\boldsymbol{P}$，使得

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J} = diag(J_{r_1}(\lambda_1), J_{r_2}(\lambda_2), \cdots, J_{r_s}(\lambda_s))$$

由矩阵乘法的知识，不难看出

$$\boldsymbol{A}^k = \boldsymbol{P}\boldsymbol{J}^k\boldsymbol{P}^{-1}$$

而

$$\boldsymbol{J}_{r_i}^k(\lambda_i) = \begin{bmatrix} f_k(\lambda_i) & f_k'(\lambda_i) & \cdots & \frac{f_k^{(r_i-1)}(\lambda_i)}{(r_i-1)!} \\ & f_k(\lambda_i) & \cdots & \frac{f_k^{(r_i-2)}(\lambda_i)}{(r_i-2)!} \\ & & \ddots & \vdots \\ & & & f_k(\lambda_i) \end{bmatrix}$$

其中，$f_k(\lambda_i) = \lambda_i^k$

由于 $r(\boldsymbol{A}) < 1$，则可以知道矩阵 $\boldsymbol{A}$ 的所有特征值均小于 1，那不难看出，当 $k \to \infty$ 时，$f_k(\lambda_i) = \lambda_i^k \to 0$, 从而 $f_k^{(l)}(\lambda_i) \to 0, \boldsymbol{J}_{r_i}^k(\lambda_i) \to 0$

因此，$\boldsymbol{J} \to 0$, 故 $\boldsymbol{A} \to 0$，从而知道 $\boldsymbol{A}$ 是一个收敛矩阵，充分性证毕。

必要性证明：

由于 $\boldsymbol{A}^k \to 0$，故 $\boldsymbol{J}_{r_i}^k(\lambda_i) \to 0$, 故 $\lambda_i^k \to 0$，从而可以知道 $\boldsymbol{A}$ 的所有特征值绝对值必定小于 1，因此 $r(\boldsymbol{A}) < 1$，必要性证毕。 □

5.1.2 矩阵级数

介绍完矩阵序列，接下来介绍矩阵级数。

定义 5.1.4

设 $\boldsymbol{A}^{(k)}$ 是 $\mathbb{C}^{m\times n}$ 的矩阵序列，称

$$\sum_{k=1}^{\infty} \boldsymbol{A}^{(k)} = \boldsymbol{A}^{(1)} + \boldsymbol{A}^{(2)} + \cdots + \boldsymbol{A}^{(k)} + \cdots$$

为**矩阵级数**，称 $\boldsymbol{S}^{(N)} = \sum\limits_{k=1}^{N} \boldsymbol{A}^{(k)}$ 为矩阵级数的**部分和**，如果 $\lim\limits_{N\to\infty} \boldsymbol{S}^{(N)} = \boldsymbol{S}$，则称 $\sum\limits_{k=1}^{\infty} \boldsymbol{A}^{(k)}$ **收敛**

上面的定义是从矩阵序列的角度定义的矩阵级数，接下来，与研究数项级数类似，如何判断矩阵级数绝对收敛?

定义 5.1.5

如果 mn 个数项级数

$$\sum_{k=1}^{\infty} a_{ij}^{(k)}, i = 1, 2, \cdots, m; j = 1, 2, \cdots, n$$

都绝对收敛，则称矩阵级数 $\sum_{k=1}^{infty} \boldsymbol{A}^{(k)}$ 绝对收敛。

上面的定义是通过数项级数绝对收敛推出矩阵级数的绝对收敛，还可以通过利用范数来确定矩阵级数是否绝对收敛：

定理 5.1.4

在 $\mathbb{C}^{n\times n}$ 中，$\sum\limits_{k=1}^{\infty} \boldsymbol{A}^{(k)}$ 绝对收敛的充要条件是正项级数 $\sum\limits_{k=1}^{\infty} \|\boldsymbol{A}^{(k)}\|$ 收敛。

证明如下：

证明.

若 $\sum\limits_{k=1}^{\infty} \boldsymbol{A}^{(k)}$ 绝对收敛，则 $\sum\limits_{k=1}^{\infty} |a_{ij}^{(k)} \leq M$

因此，根据矩阵 1-范数，有

$$\sum_{k=1}^{N} \|\boldsymbol{A}^{(k)}\|_{m_1} = \sum_{k=1}^{N} \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}^{(k)}| \right) \leq mnM$$

因此，若矩阵 1-范数收敛，则任意矩阵范数一定收敛（小收敛则大收敛）若任意矩阵范数收敛，那必定矩阵 1-范数收敛，由于一定存在

$$|a_{ij}^{(k)}| \leq \|\boldsymbol{A}^{(k)}\|_{m_1}$$

则可以得出 $\sum_{k=1}^{\infty} a_{ij}^{(k)}$ 绝对收敛，证毕。 □

5.1.3 矩阵幂级数

定理 5.1.5

方阵 $\boldsymbol{A}$ 的 Neumann 级数

$$\sum_{k=0}^{\infty} \boldsymbol{A}^k = \boldsymbol{I} + \boldsymbol{A} + \boldsymbol{A}^2 + \cdots + \boldsymbol{A}^k + \cdots$$

收敛的充要条件是 $r(\boldsymbol{A}) < 1$，且收敛时，和为 $(\boldsymbol{I} - \boldsymbol{A})^{-1}$

证明过程略，感兴趣的可以看书。

在学习微积分的无穷级数内容中，我们知道，幂级数存在收敛半径，那矩阵幂级数的收敛半径要如何求呢？下面的定理给出了答案：

定理 5.1.6

设幂级数

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

的收敛半径为 r，如果方阵 $\boldsymbol{A}$ 满足 $r(\boldsymbol{A}) < r$，则矩阵幂级数绝对收敛，如果 $r(\boldsymbol{A}) > r$，则矩阵幂级数发散。

5.2 矩阵函数

上一节是为本节内容作为铺垫的，本节将会利用矩阵级数来研究矩阵函数。

5.2.1 函数与幂级数

说到函数，可能大多数人都会想到函数都是类似于下面的形式：

$$f(x) = x, f(x) = x^2, \cdots$$

可在最一开始说，我们将会用矩阵级数的内容来研究矩阵函数，这两者存在何种关系呢？

如果还记得当时学无穷级数——函数展开成幂级数和幂级数的和函数的相关知识的话，就会知道，函数和无穷级数是离不开的，比如，你可能还记得下面这个式子：

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 + \cdots$$

这就建立起了一个函数与幂级数的关系，事实上，常见初等函数都可以展开成幂级数的形式，这里就不一一介绍了，具体请翻阅高等数学教材。

这种幂级数与函数之间的关系放到矩阵中依旧成立，这就是为什么说用矩阵幂级数的内容来研究矩阵函数的原因，下面给出矩阵函数的定义：

定义 5.2.1

设幂级数 $\sum\limits_{k=0}^{\infty} c_k \boldsymbol{z}^k$ 的收敛半径为 r，且当 $\boldsymbol{z} < r$ 时，幂级数收敛域 $f(\boldsymbol{z})$，即

$$f(\boldsymbol{z}) = \sum_{l=0}^{\infty} c_k \boldsymbol{z}^k, \quad |\boldsymbol{z}| < r$$

如果 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ 满足 $r(\boldsymbol{A}) < r$，则称收敛的矩阵幂级数 $\sum\limits_{k=0}^{\infty} c_k \boldsymbol{z}^k$ 的和记为矩阵函数，记为 $f(\boldsymbol{A})$，即

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

将 $f(\boldsymbol{A})$ 的方阵 $\boldsymbol{A}$ 换成 $\boldsymbol{At}, \boldsymbol{t}$ 为参数，就会得到

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

定义太长可以不看，简单来说就是定义前面的介绍，与普通的数项函数与幂级数的概念类似。

5.2.2 常见的矩阵函数

下面介绍了一些常见的矩阵函数，**这些函数的展开式应熟稔于心** (每天起床头件事，先背一遍展开式 (手动狗头))：

1. $$e^{\boldsymbol{A}} = \boldsymbol{E} + \boldsymbol{A} + \frac{1}{2!}\boldsymbol{A}^2 + \cdots + \frac{1}{k!}\boldsymbol{A}^k + \cdots = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!}\boldsymbol{A}^{2k+1}$$

2. $$\sin \boldsymbol{A} = \boldsymbol{A} - \frac{1}{3!}\boldsymbol{A}^3 + \frac{1}{5!}\boldsymbol{A}^5 - \frac{1}{7!}\boldsymbol{A}^7 + \cdots = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!}\boldsymbol{A}^{(2k+1)}$$

3. $$\cos \boldsymbol{A} = \boldsymbol{E} - \frac{1}{2!}\boldsymbol{A}^2 + \frac{1}{4!}\boldsymbol{A}^4 - \frac{1}{6!}\boldsymbol{A}^6 + \cdots = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!}\boldsymbol{A}^{(2k)}$$

4. $$(\boldsymbol{E} - \boldsymbol{A})^{-1} = \boldsymbol{E} + \boldsymbol{A} + \boldsymbol{A}^2 + \cdots + \boldsymbol{A}^k + \cdots = \sum_{k=0}^{\infty} \boldsymbol{A}^k$$

5. $$\ln(\boldsymbol{E} + \boldsymbol{A}) = \boldsymbol{A} - \frac{1}{2}\boldsymbol{A}^2 + \frac{1}{3}\boldsymbol{A}^3 - \frac{1}{4}\boldsymbol{A}^4 + \cdots = \sum_{k=0}^{\infty} \frac{(-1)^k}{k+1}\boldsymbol{A}^{k+1}$$

5.2.3 矩阵函数值的计算

只有函数的形式还没有用，我们最终要像计算普通函数那样把矩阵函数的函数值给计算出来，有三种方法可以计算矩阵函数值。

相似对角化计算矩阵函数值

若矩阵 $\boldsymbol{A}$ 可以相似对角化，则存在下面的等式

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, \cdots, \lambda_n) = \boldsymbol{D}$$

故

$$\begin{aligned} f(\boldsymbol{A}) &= \sum_{k=0}^{\infty} c_k \boldsymbol{A}^k = \sum_{k=0}^{\infty} c_k (\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1})^k = \boldsymbol{P}\left(\sum_{k=0}^{\infty} c_k \boldsymbol{D}^k\right)\boldsymbol{P}^{-1} \\ &= \boldsymbol{P}\begin{bmatrix} \sum\limits_{k=0}^{\infty} c_k\lambda_1^k & & \\ & \ddots & \\ & & \sum\limits_{k=0}^{\infty} c_k\lambda_n^k \end{bmatrix}\boldsymbol{P}^{-1} = \boldsymbol{P}\begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{bmatrix}\boldsymbol{P}^{-1} \end{aligned}$$

其中

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

因此，使用相似对角化的方法求解矩阵函数值的步骤为：

1. 将矩阵进行相似对角化
2. 直接代公式算答案

注意：使用该方法前，请务必确认该矩阵是否可以相似对角化，有关相似对角化的内容，可以参考第零章相似理论的相关知识，或查阅第一章 Jordan 标准型的相关知识。

使用 Jordan 标准型的方式计算矩阵函数值

设

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{J}=diag(\boldsymbol{J}_1,\boldsymbol{J}_2,\cdots,\boldsymbol{J}_s),\boldsymbol{J}_i=\begin{bmatrix}\lambda_1 & 1 & & & \\ & \ddots & \ddots & & \\ & & & \lambda_i & 1 \\ & & & & \lambda_i\end{bmatrix}_{m_i\times m_i}$$

则

$$f(\boldsymbol{J}_i)=\sum_{k=0}^{\infty}a_k\boldsymbol{J}_i^k=\begin{bmatrix}f(\lambda_i) & \frac{1}{1!}f'(\lambda_i) & \cdots & \frac{1}{(m_i-1)!}f^{(m_i-1)}(\lambda_i) \\ & f(\lambda_i) & \cdots & \vdots \\ & & \ddots & \frac{1}{1!}f'(\lambda_i) \\ & & & f(\lambda_i)\end{bmatrix}$$

故

$$\begin{aligned}f(\boldsymbol{A})&=\sum_{k=0}^{\infty}a_k\boldsymbol{P}\boldsymbol{J}^k\boldsymbol{P}^{-1}=(\sum_{k=0}^{\infty}a_k\boldsymbol{J}^k)\boldsymbol{P}^{-1}\\&=\boldsymbol{P}\begin{bmatrix}\sum\limits_{k=0}^{\infty}a_k\boldsymbol{J}_1^k & & \\ & \ddots & \\ & & \sum\limits_{k=0}^{\infty}a_k\boldsymbol{J}_s^k\end{bmatrix}\boldsymbol{P}^{-1}=\boldsymbol{P}\begin{bmatrix}f(\boldsymbol{J}_1) & & \\ & \ddots & \\ & & f(\boldsymbol{J}_s)\end{bmatrix}\boldsymbol{P}^{-1}\end{aligned}$$

因此，使用 Jordan 标准型的方法求解矩阵函数值的步骤为：

1. 求 Jordan 标准型
2. 代公式计算

使用数项级数求和方式计算矩阵函数值

第一种方法有局限性，第二种方法需要较大计算量，下面将要介绍的这种方法计算量不大，但是需要记住常见函数的矩阵函数展开式。

在开始说计算方法之前，首先介绍哈密尔顿-凯莱定理：

定理 5.2.1: 哈密尔顿-凯莱定理

设 $\boldsymbol{A}$ 是数域 $\mathbb{P}$ 上的一个 $n\times n$ 矩阵，$f(\lambda)=|\lambda\boldsymbol{E}-\boldsymbol{A}|$ 是 $\boldsymbol{A}$ 的特征多项式，则

$$f(\boldsymbol{A})=\boldsymbol{A}^n-b_{n-1}\boldsymbol{A}^{n-1}-\cdots-b_1\boldsymbol{A}-b_0\boldsymbol{E}=0$$

这个定理有什么用呢？我们继续看。

由定理，我们可以得到下面的等式关系：

$$\boldsymbol{A}^n=b_{n-1}\boldsymbol{A}^{n-1}+\cdots+b_1\boldsymbol{A}+b_0\boldsymbol{E}$$

如果我们想要计算 $\boldsymbol{A}^{n+1}$，那只需要计算

$$\boldsymbol{A}^n\cdot\boldsymbol{A}$$

的结果即可，即

$$\boldsymbol{A}^{n+1}=\boldsymbol{A}^n\boldsymbol{A}=b_0\boldsymbol{A}+b_1\boldsymbol{A}^2+\cdots+b_{n-1}\boldsymbol{A}^n$$

而

$$\boldsymbol{A}^n=b_{n-1}\boldsymbol{A}^{n-1}+\cdots+b_1\boldsymbol{A}+b_0\boldsymbol{E}$$

因此，我们重新将式子整理一下合并同类项就可以得到一个新的式子

由此，我们会发现，再计算更高次的矩阵乘法时，我们完全可以用低阶矩阵乘积的结果来表示出来，从而简化了计算，这就是该定理的作用。

可这对计算矩阵函数值有什么用呢？由矩阵幂级数的知识，一个矩阵函数可以写成

$$f(\boldsymbol{A})=c_0\boldsymbol{E}+c_1\boldsymbol{A}+\cdots+c_k\boldsymbol{A}^k+\cdots$$

的形式，由哈密尔顿-凯莱定理，最高次项可以用低阶来表示，因此可以重新合并同类项，最终就会变成

$$(c_0+c_nb_0+c_{n+1}b_0^{(1)}+\cdots)\boldsymbol{E}+(c_1+c_nb_1+c_{n+1}b_1^{(1)}+\cdots)\boldsymbol{A}+\cdots$$

的形式，而

$$(c_0 + c_n b_0 + c_{n+1} b_0^{(1)} + \cdots)\text{和}(c_1 + c_n b_1 + c_{n+1} b_1^{(1)} + \cdots)$$

都是数项级数，因此可以使用数项级数的方式来求解矩阵函数函数值。

因此，使用求解数项级数的方法求解矩阵函数值的步骤为:

1. 求特征方程 $|\lambda \boldsymbol{E} - \boldsymbol{A}|$ 的值，利用哈密尔顿-凯莱定理进行简化

2. 代公式计算

（如果只看上面这些抽象方法肯定看不懂，因此每一个方法应该都要加一个例子的，但是时间有限并且这些例子书上都有，所以可以看书或者 PPT 的例子就不在这里写了，等有时间再写）

5.2.4 矩阵函数的其他性质

本节的最后，再介绍一些矩阵函数的其他性质:

1. 如果 $\boldsymbol{AB} = \boldsymbol{BA}$，则 $e^{\boldsymbol{A}} e^{\boldsymbol{B}} = e^{\boldsymbol{B}} e^{\boldsymbol{A}} = e^{\boldsymbol{A}+\boldsymbol{B}}$

2. 如果 $\boldsymbol{AB} = \boldsymbol{BA}$，则

   (a) $\cos(\boldsymbol{A} + \boldsymbol{B}) = \cos \boldsymbol{A} \cos \boldsymbol{B} - \sin \boldsymbol{A} \sin \boldsymbol{B}$

   (b) $\sin(\boldsymbol{A} + \boldsymbol{B}) = \sin \boldsymbol{A} \cos \boldsymbol{B} + \cos \boldsymbol{A} \sin \boldsymbol{B}$

这些性质与代数上的性质都类似，只不过需要注意一下这些性质的成立是基于 $\boldsymbol{AB} = \boldsymbol{BA}$ 的情况下才可以，其他的就没有什么需要注意的了。

第六章　广义逆矩阵

在学习线性代数的时候，我们了解过矩阵逆的概念，同时我们也知道一个矩阵如果有逆矩阵，必须要满足两个条件：

- 必须是方阵
- 这个方阵的行列式不为 0

我们说，如果一个矩阵满足上面的两个条件，就说该矩阵是可逆的，也就是说，存在一个矩阵 $\boldsymbol{B}$，使得

$$\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{E}$$

成立，这个时候就说矩阵 $\boldsymbol{B}$ 是矩阵 $\boldsymbol{A}$ 的逆矩阵，记作 $\boldsymbol{A}^{-1}$。

引入矩阵的逆的作用是什么呢？我们在解线性方程组 $\boldsymbol{Ax} = \boldsymbol{b}$ 的时候，如果存在矩阵 $\boldsymbol{A}$ 的逆矩阵，那么这个线性方程组的解就唯一，即

$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$$

为方程组的解，这便是矩阵的逆的最大的作用——解线性方程组。

可是，世界上存在的矩阵多种多样，能够满足存在逆矩阵的条件的矩阵只占很小的一部分，那这样的话求解线性方程组就没办法了吗?

本章介绍广义逆矩阵，广义逆矩阵的引入就是为了解决不满足矩阵的逆的两个条件的矩阵定义逆矩阵的问题，在学习了本章之后就会发现，其实大多数的矩阵虽然没有严格的逆矩阵，但大多会存在广义逆矩阵。

1995 年，R.Penrose 证明了，对于任何矩阵 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$，若存在唯一的矩阵 $\boldsymbol{G} \in \mathbb{C}^{n\times m}$，同时满足下面 4 个方程：

$$\boldsymbol{AGA} = \boldsymbol{A}$$

$$\boldsymbol{GAG} = \boldsymbol{G}$$

$$(\boldsymbol{GA})^{\boldsymbol{H}} = \boldsymbol{GA}$$

$$(\boldsymbol{AG})^{\boldsymbol{H}} = \boldsymbol{AG}$$

中的一个或几个，则将矩阵 $\boldsymbol{G}$ 称为矩阵 $\boldsymbol{A}$ 的**广义逆矩阵**，又叫 Moore-Penrose 逆 (M-P 逆)。

本章涉及到的内容包括：

- 矩阵的单边逆
- 广义逆矩阵 $\boldsymbol{A}^{-}$
- 自反广义逆矩阵 $\boldsymbol{A}_r^{-}$
- M-P 广义逆矩阵 $\boldsymbol{A}^{+}$
- 广义逆矩阵的应用

6.1 矩阵的单边逆

6.1.1 依旧从解方程组说起

我们接着就本章序言的问题讨论，假设有一个任意矩阵 $\boldsymbol{A}$，假设线性方程组

$$\boldsymbol{Ax} = \boldsymbol{b}$$

现在由于 $\boldsymbol{A}$ 并不一定是一个可逆矩阵，因此没有办法直接采取消去的办法，采取左右两侧都乘 $\boldsymbol{A}^{-1}$ 的方式来解方程，但事情真的到此就结束了吗？

根据矩阵的乘法，或许可能存在这样的一个矩阵 $\boldsymbol{B}$，使得

$$\boldsymbol{BA} = \boldsymbol{E}$$

成立。

这个时候，只需要在方程组左右两侧同时左乘 $\boldsymbol{B}$，那这个等式就会变成

$$\boldsymbol{BAx} = \boldsymbol{Bb}$$

由于 $\boldsymbol{BA} = \boldsymbol{E}$，因此式子就成了

$$\boldsymbol{x} = \boldsymbol{Bb}$$

如此，我们便在不用借助逆矩阵的情况下，依旧求出来了方程组的解，此时这个 $\boldsymbol{B}$ 也并不是 $\boldsymbol{A}$ 的逆，其实，并不需要一定要找到 $\boldsymbol{A}$ 矩阵的逆，只需要找到一个矩阵 $\boldsymbol{B}$ 使得 $\boldsymbol{BA}=\boldsymbol{E}$ 就可以了，这就是**矩阵的单边逆**。

定义 6.1.1: 左逆和右逆

设 $\boldsymbol{A}\in\mathbb{C}^{m\times n}$，如果有 $G\in\mathbb{C}^{n\times m}$，使得

$$\boldsymbol{GA}=\boldsymbol{E}_n$$

则称 $\boldsymbol{G}$ 为 $\boldsymbol{A}$ 的**左逆矩阵**，记为 $\boldsymbol{G}=\boldsymbol{A}_L^{-1}$.

如果

$$\boldsymbol{AG}=\boldsymbol{E}_m$$

则称 $\boldsymbol{G}$ 为 $\boldsymbol{A}$ 的**右逆矩阵**，记为 $\boldsymbol{G}=\boldsymbol{A}_R^{-1}$.

左逆和右逆的出现使得我们对于矩阵的逆的条件得以放宽，只要能够满足上式成立，依旧能够使得消去因子成立，那是所有的矩阵都存在左逆和右逆吗？请看下面的定理6.1.1

定理 6.1.1: 左逆右逆存在的条件

设 $\boldsymbol{A}\in\mathbb{C}^{m\times n}$，则

1. $\boldsymbol{A}$ 左可逆的充要条件是 $\boldsymbol{A}$ 为列满秩矩阵。
2. $\boldsymbol{A}$ 右可逆的充要条件是 $\boldsymbol{A}$ 为行满秩矩阵。

定理6.1.1的证明如下（以左可逆矩阵为例）

充分性证明：

由于 $\boldsymbol{A}$ 是列满秩矩阵，因此 $\boldsymbol{A}^H\boldsymbol{A}$ 一定是满秩 n 阶方阵，同时该矩阵可逆

故存在下列等式

$$(\boldsymbol{A}^H\boldsymbol{A})^{-1}(\boldsymbol{A}^H\boldsymbol{A})=\boldsymbol{E}_n$$

去掉括号，有

$$(\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H\boldsymbol{A}=\boldsymbol{E}_n$$

此时，$(\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H$ 为左可逆矩阵

必要性证明：

由线性代数矩阵的秩的知识，有

$$\operatorname{rank}(\boldsymbol{AB}) \leq \min\{\operatorname{rank}(\boldsymbol{A}), \operatorname{rank}(\boldsymbol{B})\}$$

由于 $\boldsymbol{A}$ 存在左可逆矩阵，故 $\boldsymbol{A}_L^{-1}\boldsymbol{A} = \boldsymbol{E}_n$

因此，有

$$\operatorname{rank}(\boldsymbol{A}) \geq \operatorname{rank}(\boldsymbol{A}_L^{-1}\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{E}_n) = n$$

又由于 $\boldsymbol{A}$ 的列数为 n，因此由线性代数的知识可知，$\operatorname{rank}\boldsymbol{A} \leq n$

故 $\operatorname{rank}(\boldsymbol{A}) = n$, 证毕。

在了解了任意一个矩阵左逆右逆存在的条件后，还有下面一个推论

推论 6.1.2

设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$，则

1. $\boldsymbol{A}$ 左可逆的充要条件为 $N(\boldsymbol{A}) = \{0\}$
2. $\boldsymbol{A}$ 右可逆的充要条件为 $R(\boldsymbol{A}) = \mathbb{C}^m$

这里需要解释一下 $N(\boldsymbol{A})$ 和 $R(\boldsymbol{A})$ 是什么:

$$N(\boldsymbol{A}) = \{\boldsymbol{x}|\boldsymbol{Ax} = \boldsymbol{0}, \boldsymbol{x} \in \mathbb{C}^n\}$$

$$R(\boldsymbol{A}) = \{\boldsymbol{y}|\boldsymbol{y} = \boldsymbol{Ax}, \forall \boldsymbol{x} \in \mathbb{C}^n\}$$

$N(\boldsymbol{A})$ 叫做**核**，表示的是方程组 $\boldsymbol{Ax} = \boldsymbol{0}$ 的解空间，$R(\boldsymbol{A})$ 表示**值域**，同时有下面的不等关系（证明略）:

$$\dim[R(\boldsymbol{A})] = \operatorname{rank}(\boldsymbol{A})$$

推论6.1.2的证明很简单:

对于 (1)，由于 $N(\boldsymbol{A}) = \{0\}$，可以得知方程组 $\boldsymbol{Ax} = \boldsymbol{0}$ 只有零解，由此可知 $\boldsymbol{A}$ 列满秩。

对于 (2)，由于 $R(\boldsymbol{A}) = \mathbb{C}^m$，因此可知 $\dim[R(\boldsymbol{A})] = \operatorname{rank}(\boldsymbol{A}) = m$，因此 $\boldsymbol{A}$ 行满秩。

6.1.2 求矩阵左逆/右逆的方法

求矩阵左逆的方法

在线性代数中是如何求矩阵的逆的呢？我们通过将矩阵 $\boldsymbol{A}$ 与单位矩阵拼凑在一起，经过初等行变换，将左侧 $\boldsymbol{A}$ 化为 $\boldsymbol{E}$，那右侧的 $\boldsymbol{E}$ 就是 $\boldsymbol{A}^{-1}$ 了，即

$$(\boldsymbol{A} \quad \boldsymbol{E}) \xrightarrow{\text{行}} (\boldsymbol{E} \quad \boldsymbol{A}^{-1})$$

上面的过程其实可以转换为

$$\boldsymbol{E}_n \cdots \boldsymbol{E}_2 \boldsymbol{E}_1 \boldsymbol{A} = \boldsymbol{E}$$

仔细看，这是不是与 $\boldsymbol{A}_L^{-1}$ 的形式很相似，这个时候就可以把 $\boldsymbol{E}_n \cdots \boldsymbol{E}_2 \boldsymbol{E}_1$ 当做 $\boldsymbol{A}_L^{-1}$。

那么，上面的过程用一个严谨的公式是要如何表示呢？

$$[\boldsymbol{A} \quad \boldsymbol{E}] \xrightarrow{\text{行}} \begin{bmatrix} \boldsymbol{E}_n & \boldsymbol{G} \\ \boldsymbol{0} & * \end{bmatrix}$$

这里的 $\boldsymbol{G}$ 就是矩阵的左逆。

求矩阵右逆的方法

用类似求矩阵左逆的方法，只不过现在变成右乘一系列的初等矩阵 (即对矩阵做初等列变换) 即可求得矩阵右逆，公式如下：

$$\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{E} \end{bmatrix} \xrightarrow{\text{列变换}} \begin{bmatrix} \boldsymbol{E} & \boldsymbol{0} \\ \boldsymbol{G} & * \end{bmatrix}$$

6.1.3 矩阵单边逆的其他性质

本小节的最后一部分介绍一下矩阵单边逆的其他性质，这些性质了解即可，证明过程无需掌握，因此就不写在这里了，感兴趣的可以翻阅教材查看。

矩阵的单边逆不唯一

定理 6.1.3

设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ 是左可逆矩阵，则

$$\boldsymbol{G} = (\boldsymbol{A}_1^{-1} - \boldsymbol{B}\boldsymbol{A_2}\boldsymbol{A_1}^{-1} \quad \boldsymbol{B})\boldsymbol{P}$$

是 $\boldsymbol{A}$ 的左逆矩阵，其中 $\boldsymbol{B} \in \mathbb{C}^{n\times(m-n)}$ 为任意矩阵，行初等变换对应的矩阵 $\boldsymbol{P}$ 满足 $\boldsymbol{PA} = \begin{bmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A_2} \end{bmatrix}$, $\boldsymbol{A}_1$ 是 n 阶可逆方阵。

定理 6.1.4

设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ 是右可逆矩阵，则

$$\boldsymbol{G} = \boldsymbol{Q}\begin{bmatrix} \boldsymbol{A}_1^{-1} - \boldsymbol{A}_1^{-1}\boldsymbol{A}_2\boldsymbol{D} \\ \boldsymbol{D} \end{bmatrix}$$

是 $\boldsymbol{A}$ 的右逆矩阵，其中 $\boldsymbol{D} \in \mathbb{C}^{(n-m)\times m}$ 为任意矩阵，列初等变换对应的矩阵 $\boldsymbol{Q}$ 满足 $\boldsymbol{AQ} = (\boldsymbol{A}_1 \quad \boldsymbol{A}_2)$, $\boldsymbol{A}_1$ 是 m 阶可逆方阵。

6.1.4 矩阵单边逆与线性方程组的关系

定理 6.1.5

设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ 是左可逆矩阵，$\boldsymbol{A}_L^{-1}$ 是 $\boldsymbol{A}$ 的左逆矩阵，则方程组 $\boldsymbol{Ax} = \boldsymbol{b}$ 有解的充要条件是

$$(\boldsymbol{E}_m - \boldsymbol{A}\boldsymbol{A}_L^{-1})\boldsymbol{b} = \boldsymbol{0}$$

若等式成立，则方程组 $\boldsymbol{Ax} = \boldsymbol{b}$ 有唯一解

$$\boldsymbol{x} = (\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H\boldsymbol{b}$$

定理 6.1.6

设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ 是右可逆矩阵，则 $\boldsymbol{Ax} = \boldsymbol{b}$ 对任何 $b \in \mathbb{C}^m$ 都有解，若 $\boldsymbol{b} \neq \boldsymbol{0}$，则方程组的解可表示为

$$\boldsymbol{x} = \boldsymbol{A}_R^{-1}\boldsymbol{b}$$

其中，$\boldsymbol{x} = \boldsymbol{A}_R^{-1}$ 是 $\boldsymbol{A}$ 的一个右逆矩阵

6.2 广义逆矩阵 $\boldsymbol{A}^-$

在上一节，我们认识到了矩阵的单边逆，在本节我们继续将矩阵逆的范围进行扩大，来了解一下矩阵的广义逆 $\boldsymbol{A}^-$。

6.2.1 还是从解线性方程组说起...

上一节的最后的两个定理指出了线性方程组有解的条件以及在有解的情况下的解是什么，可是解存在的最大的前提是矩阵的左逆或者右逆需要存在，可我们还知道，如果矩阵的左逆或者右逆存在，那必然该矩阵是行满秩或列满秩矩阵，事实上，这类的矩阵依旧并不能代表所有的矩阵，更广泛的矩阵并不一定行满秩或列满秩。

由线性代数的知识，我们知道，方程组 $\boldsymbol{Ax} = \boldsymbol{b}$ 有解的充要条件是

$$\mathrm{rank}(\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{A} \quad \boldsymbol{b})$$

如果方程组有解，那必定存在矩阵 $\boldsymbol{G} \in \mathbb{C}^{\boldsymbol{n}\times\boldsymbol{m}}$，把解表示为

$$\boldsymbol{x} = \boldsymbol{Gb}$$

此时整个方程也就变为

$$\boldsymbol{AGb} = \boldsymbol{b}$$

这个时候，矩阵 $\boldsymbol{G}$ 就是 $\boldsymbol{A}$ 的**广义逆矩阵**，记为 $\boldsymbol{G} = \boldsymbol{A}^-$。

定义 6.2.1: 广义逆矩阵的定义

设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$，如果存在矩阵 $\boldsymbol{G} \in \mathbb{C}^{m\times n}$，使得

$$\boldsymbol{AGb} = \boldsymbol{b} \quad (\forall \boldsymbol{b} \in R(\boldsymbol{A}))$$

则称 $\boldsymbol{G}$ 为 $\boldsymbol{A}$ 的**广义逆矩阵**，记为 $\boldsymbol{G} = \boldsymbol{A}^-$

6.2.2 广义逆矩阵的性质

接下来介绍一些广义逆矩阵的性质

定理 6.2.1

设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$，则 $\boldsymbol{A}$ 存在广义逆矩阵的充要条件是存在 $\boldsymbol{G} \in \mathbb{C}^{n\times m}$，使其满足

$$\boldsymbol{AGA} = \boldsymbol{A}$$

证明.

充分性证明：

由于 $\forall \boldsymbol{b} \in R(\boldsymbol{A})$，因此可以知道方程 $\boldsymbol{Ax} = \boldsymbol{b}$ 有解，不妨设解为 $\boldsymbol{x}_0$

因此有 $\boldsymbol{Ax}_0 = \boldsymbol{b}$，同时有

$$\boldsymbol{AGb} = \boldsymbol{AGAx}_0 = \boldsymbol{Ax}_0 = \boldsymbol{b}$$

由定义6.2.1，可知 $\boldsymbol{G}$ 为 $\boldsymbol{A}$ 的广义逆矩阵，证毕。

必要性证明：

任取 $\boldsymbol{u} \in \mathbb{C}^n$，则必定存在矩阵 $\boldsymbol{A}$，使得

$$\boldsymbol{Au} = \boldsymbol{b} \in R(\boldsymbol{A})$$

因此，也一定有

$$\boldsymbol{AGAu} = \boldsymbol{Au}$$

请注意到 $\boldsymbol{u}$ 的任意性，不管 $\boldsymbol{u}$ 取何值，一定存在

$$\boldsymbol{AGA} = \boldsymbol{A}$$

证毕。 □

广义逆矩阵的秩

推论 6.2.2

设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$，且 $\boldsymbol{A}^- \in \mathbb{C}^{n\times m}$ 是 $\boldsymbol{A}$ 的一个广义逆矩阵，则

$$\operatorname{rank}(\boldsymbol{A}^-) \geq \operatorname{rank}(\boldsymbol{A})$$

证明.

$$\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A}) \leq \operatorname{rank}(\boldsymbol{A}\boldsymbol{A}^-) \leq \operatorname{rank}(\boldsymbol{A}^-)$$

□

广义逆矩阵的其他性质

定理 6.2.3

设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}, \lambda \in \mathbb{C}$，则

1. $(\boldsymbol{A}^T)^- = (\boldsymbol{A}^-)^T, (\boldsymbol{A}^H)^- = (\boldsymbol{A}^-)^H$
2. $\boldsymbol{A}\boldsymbol{A}^-$ 与 $\boldsymbol{A}^-\boldsymbol{A}$ 都是幂等矩阵，且
$$\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}\boldsymbol{A}^-) = \operatorname{rank}(\boldsymbol{A}^-\boldsymbol{A})$$
3. $\lambda^-\boldsymbol{A}^-$ 为 $\lambda\boldsymbol{A}$ 的广义逆矩阵，其中 $\lambda^- = \begin{cases} 0, & \lambda = 0 \\ \lambda^{-1}, & \lambda \neq \end{cases}$
4. 设 $\boldsymbol{S}$ 是 m 阶可逆矩阵，$\boldsymbol{T}$ 是 n 阶可逆矩阵，且 $\boldsymbol{B} = \boldsymbol{S}\boldsymbol{A}\boldsymbol{T}$，则 $\boldsymbol{T}^{-1}\boldsymbol{A}^-\boldsymbol{S}^{-1}$ 是 $\boldsymbol{B}$ 的广义逆矩阵
5. $R(\boldsymbol{A}\boldsymbol{A}^-) = R(\boldsymbol{A}), N(\boldsymbol{A}^-\boldsymbol{A}) = N(\boldsymbol{A})$

证明.

1 证明：

由于 $\boldsymbol{A}^-$ 是 $\boldsymbol{A}$ 的广义逆，因此有

$$\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{A}$$

两侧同时转置，有

$$\boldsymbol{A}^T(\boldsymbol{A}^-)^T\boldsymbol{A}^T = \boldsymbol{A}^T$$

由定理6.2.1，故

$$(\boldsymbol{A}^-)^T = (\boldsymbol{A}^T)^-$$

2 证明：

$$(\boldsymbol{A}\boldsymbol{A}^-)^2 = (\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A})\boldsymbol{A}^-$$

注意到

$$\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A} = \boldsymbol{A}$$

因此原式就变成了

$$(\boldsymbol{A}\boldsymbol{A}^-)^2 = \boldsymbol{A}\boldsymbol{A}^-$$

这就是 $\boldsymbol{A}\boldsymbol{A}^-$ 是幂等矩阵的原因，同理 $\boldsymbol{A}^-\boldsymbol{A}$ 也可以用类似的方法证明。

由线性代数的知识，可以得出如下不等关系:

$$\text{rank}(\boldsymbol{A}) \geq \text{rank}(\boldsymbol{A}\boldsymbol{A}^-) \rightarrow \boldsymbol{A}\text{是}\boldsymbol{A}\boldsymbol{A}^-\text{里的因子}$$

同时还有

$$\text{rank}(\boldsymbol{A}\boldsymbol{A}^-) \geq \text{rank}(\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A}) \rightarrow \boldsymbol{A}\boldsymbol{A}^-\text{是}\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A}\text{里的因子}$$

又由于 $\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A} = \boldsymbol{A}$, 因此整个的不等关系为

$$\text{rank}(\boldsymbol{A}) \geq \text{rank}(\boldsymbol{A}\boldsymbol{A}^-) \geq \text{rank}(\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A}) = \text{rank}(\boldsymbol{A})$$

因此得到

$$\text{rank}(\boldsymbol{A}) = \text{rank}(\boldsymbol{A}\boldsymbol{A}^-)$$

3 证明：若 $\lambda = neq0$，则有

$$(\lambda\boldsymbol{A})(\lambda^-\boldsymbol{A}^-)(\lambda\boldsymbol{A}) = (\lambda\lambda^-\lambda)\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A}$$

由于 $\lambda^- = \lambda^{-1}, \boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A} = \boldsymbol{A}$，故原式变成了

$$(\lambda\boldsymbol{A})(\lambda^-\boldsymbol{A}^-)(\lambda\boldsymbol{A})\lambda\boldsymbol{A}$$

若 $\lambda = 0$, 则有

$$(\lambda\boldsymbol{A})(\lambda^-\boldsymbol{A}^-)(\lambda\boldsymbol{A}) = 0 = \lambda\boldsymbol{A}$$

证毕。

4 证明：只需验证 $\boldsymbol{AGA}=\boldsymbol{A}$ 即可，过程就不写了

5 证明：

在这里遇到了需要证明两个集合相等的问题，或者说要证明两个空间相等的问题，解决这种问题的思路分成以下两步：

- 首先证明左侧或右侧的集合包含在对侧的集合中，如果集合相等那也一定存在集合包含关系，只不过是互相包含罢了。
- 之后证明空间的维数相等即可。

对于 $R(\boldsymbol{AA}^-)$ 与 $R(\boldsymbol{A})$，显然有

$$R(\boldsymbol{AA}^-)\subseteq R(\boldsymbol{A})$$

同时我们还有

$$\operatorname{rank}(\boldsymbol{AA}^-)=\operatorname{rank}(\boldsymbol{A})$$

因此可以得出 $R(\boldsymbol{AA}^-)=R(\boldsymbol{A})$

对于 $N(\boldsymbol{AA}^-)$ 与 $N(\boldsymbol{A})$，有

$$N(\boldsymbol{A})\subseteq N(\boldsymbol{AA}^-)$$

依旧根据

$$\operatorname{rank}(\boldsymbol{AA}^-)=\operatorname{rank}(\boldsymbol{A})$$

得到了两个集合相等，证毕。 □

6.3 自反广义逆矩阵 $\boldsymbol{A}_r^{-1}$

何为自反广义逆矩阵呢？在学习线性代数逆矩阵的时候，有这样一条已经熟知的性质：

$$(\boldsymbol{A}^{-1})^{-1}=\boldsymbol{A}$$

这叫做逆矩阵的自反性质，可这样的性质在广义逆矩阵中还成立吗?

假设有矩阵 $\boldsymbol{A}^-=\boldsymbol{0},\boldsymbol{0}^-=\boldsymbol{E}$，那 $(\boldsymbol{A}^-)^-=(\boldsymbol{0})^-=\boldsymbol{0}^-=\boldsymbol{E}\neq\boldsymbol{A}$，通过这个小例子就知道了广义逆矩阵并不一定具有自反性。

6.3.1 自反广义逆矩阵的定义

定义 6.3.1

设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$, 如果有 $\boldsymbol{G} \in \mathbb{C}^{n\times m}$，使得

$$\boldsymbol{AGA} = \boldsymbol{A}, \boldsymbol{GAG} = \boldsymbol{G}$$

同时成立，则称 $\boldsymbol{G}$ 为 $\boldsymbol{A}$ 的**自反广义逆矩阵**。

根据定义可以看到，自反广义逆矩阵在广义逆矩阵的前提下又增加了 $\boldsymbol{GAG} = \boldsymbol{G}$ 的条件，是所有的矩阵都拥有自反广义逆矩阵吗？在下面的定理中给出了答案（证明略）

定理 6.3.1

任何矩阵都有自反广义逆矩阵。

定理 6.3.2

设 $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{C}^{n\times m}$ 均为 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ 的广义逆矩阵，则

$$\boldsymbol{Z} = \boldsymbol{XAY}$$

是 $\boldsymbol{A}$ 的自反广义逆矩阵。

除此之外，如何判断一个广义逆矩阵是自反广义逆矩阵呢？请看下面的定理

定理 6.3.3

$\boldsymbol{A} \in \mathbb{C}^{m\times n}, \boldsymbol{A}^-$ 是 $\boldsymbol{A}$ 的广义逆矩阵，则 $\boldsymbol{A}^-$ 是 $\boldsymbol{A}$ 的自反广义逆矩阵的充要条件是

$$\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}^-)$$

定理6.3.3从矩阵的秩的角度表明了一个广义逆矩阵如何成为自反广义逆矩阵的条件，证明如下：

证明.

必要性证明：

因为 $\boldsymbol{A}^-$ 是自反广义逆矩阵，根据自反广义逆矩阵，有

$$\boldsymbol{AA}^-\boldsymbol{A} = \boldsymbol{A}, \boldsymbol{A}^-\boldsymbol{AA}^- = \boldsymbol{A}^-$$

由前半段，可知 $\mathrm{rank}(\boldsymbol{A}^{-}) \geq \mathrm{rank}(\boldsymbol{A})$

由后半段，可知 $\mathrm{rank}(\boldsymbol{A}^{-}) \leq \mathrm{rank}(\boldsymbol{A})$

所以 $\mathrm{rank}(\boldsymbol{A}^{-}) = \mathrm{rank}(\boldsymbol{A})$

充分性证明：

//TODO(懒得写了！) □

6.4 M-P 广义逆矩阵 $\boldsymbol{A}^{+}$

6.4.1 $\boldsymbol{A}^{+}$ 及其性质

本节课继续介绍另外一种广义逆矩阵：M-P 广义逆矩阵 $\boldsymbol{A}^{+}$。

定义 6.4.1

设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$, 如果有 $\boldsymbol{G} \in \mathbb{C}^{n\times m}$，使得

$$\boldsymbol{AGA} = \boldsymbol{A}, \boldsymbol{GAG} = \boldsymbol{G}, (\boldsymbol{AG})^{H} = \boldsymbol{AG}, (\boldsymbol{GA})^{H} = \boldsymbol{GA}$$

则称 $\boldsymbol{G}$ 是 $\boldsymbol{A}$ 的 **M-P 广义逆矩阵**，记为 $\boldsymbol{G} = A^{+}$.

定义6.4.1只是说明 A^{+} 的性质，并没有告诉计算方式，如何计算 A^{+}? 请看下面的定理：

定理 6.4.1

设 $\boldsymbol{A} \in \mathbb{C}_{r}^{m\times n}, \boldsymbol{A} = \boldsymbol{BD}$ 是 $\boldsymbol{A}$ 的最大秩分解，则

$$\boldsymbol{A}^{+} = \boldsymbol{G} = \boldsymbol{D}^{H}(\boldsymbol{DD}^{H})^{-1}(\boldsymbol{B}^{H}\boldsymbol{B})^{-1}\boldsymbol{B}^{H}$$

如何证明该定理的合理性呢？只需要验证定义6.4.1的四个等式即可。

如果留意前面的单边逆、广义逆和自反广义逆矩阵，会发现大多数矩阵的单边逆、广义逆和自反广义逆并不一定是唯一的，下面的定理给出了 $\boldsymbol{A}^{+}$ 的唯一性。

定理 6.4.2

设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$，则 $\boldsymbol{A}^{+}$ 是唯一的。

证明过程可以不用理解，感兴趣可以查阅教材。

$\boldsymbol{A}^+$ 的性质

定理 6.4.3

设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$，则有

1. $(\boldsymbol{A}^+)^+ = \boldsymbol{A}$
2. $(\boldsymbol{A}^T)^+ = (\boldsymbol{A}^+)^T, (\boldsymbol{A}^H)^+ = (\boldsymbol{A}^+)^H$
3. $\boldsymbol{A}^+ = (\boldsymbol{A}^H\boldsymbol{A})^+\boldsymbol{A}^H = \boldsymbol{A}^H(\boldsymbol{A}\boldsymbol{A}^H)^+$
4. $R(\boldsymbol{A}^+) = r(\boldsymbol{A}^H)$
5. $\boldsymbol{A}\boldsymbol{A}^+ = P_{R(\boldsymbol{A})}, \boldsymbol{A}^+\boldsymbol{A} = P_{R(\boldsymbol{A}^H)}$
6. $R(\boldsymbol{A}) = R(\boldsymbol{A}^H) \Leftrightarrow \boldsymbol{A}\boldsymbol{A}^+ = \boldsymbol{A}^+\boldsymbol{A}$

如果对证明过程感兴趣，可以翻阅教材查看。

除此之外，还有最后一个关于 $\boldsymbol{A}^+$ 的性质:

定理 6.4.4

设 $\boldsymbol{A} \in \mathbb{C}^{m\times l}, \boldsymbol{B} \in \mathbb{C}^{l\times n}$，则

$$(\boldsymbol{A}\boldsymbol{B})^+ = \boldsymbol{B}^+\boldsymbol{A}^+ \Leftrightarrow \begin{cases} R(\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{B}) \subseteq R(\boldsymbol{B}) \\ R(\boldsymbol{B}\boldsymbol{B}^H\boldsymbol{A}^H) \subseteq R(\boldsymbol{A}^H) \end{cases}$$

6.4.2 计算 $\boldsymbol{A}^+$ 的方法

最大秩分解法

用最大秩分解的方式计算 $\boldsymbol{A}^+$ 就是前面的定理6.4.1，这里不再赘述，但下面这一条计算 $\boldsymbol{A}^+$ 的方式需要了解一下:

引理 6.4.5

设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$，则

1. 如果 $\boldsymbol{A}$ 是行满秩矩阵，则 $\boldsymbol{A}^+ = \boldsymbol{A}^H(\boldsymbol{A}\boldsymbol{A}^H)^{-1}$
2. 如果 $\boldsymbol{A}$ 是列满秩矩阵，则 $\boldsymbol{A}^+ = (\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H$

奇异值分解法

定理 6.4.6

设 $\boldsymbol{A} \in \mathbb{C}_r^{m\times n}$ 的奇异值分解为

$$\boldsymbol{A} = \boldsymbol{U}\begin{bmatrix} \boldsymbol{D}_r & 0 \\ 0 & 0 \end{bmatrix}\boldsymbol{V}$$

则有

1. $\boldsymbol{A}^+ = \boldsymbol{V}^H\boldsymbol{D}^+\boldsymbol{U}^H$
2. $\|\boldsymbol{A}^+\|_F^2 = \sum\limits_{i=1}^{r}\frac{1}{\sigma_i^2}$
3. $\|\boldsymbol{A}^+\|^2 = \frac{1}{\min\limits_{i\le i\le r}\sigma_i}$

证明. 看看书上或者 PPT 的证明吧，再也不想证明了。。。 □

6.5 广义逆矩阵的应用

最后，介绍一下广义逆矩阵在解方程方面的应用。

6.5.1 广义逆矩阵在解矩阵方程上的应用

定理 6.5.1

设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}, \boldsymbol{B} \in \mathbb{C}^{p\times q}, \boldsymbol{D} \in \mathbb{C}^{m\times q}$，则矩阵方程

$$\boldsymbol{AXB} = \boldsymbol{D}$$

有解的充要条件是存在 $\boldsymbol{A}^-$和$\boldsymbol{B}^-$，使得

$$\boldsymbol{AA}^-\boldsymbol{DB}^-\boldsymbol{B} = \boldsymbol{D}$$

成立，在有解的条件下，矩阵方程 $\boldsymbol{AXB} = \boldsymbol{D}$ 的通解为

$$\boldsymbol{X} = \boldsymbol{A}^-\boldsymbol{DB}^- + \boldsymbol{Y} - \boldsymbol{A}^-\boldsymbol{AYBB}^-, \quad \forall \boldsymbol{Y} \in \mathbb{C}^{n\times p}$$

在此基础上堆式子进行变形，就产生了下面的两个推论

推论 6.5.2

设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}, \boldsymbol{D} \in \mathbb{C}^{m\times p}$，则矩阵方程

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{D}$$

有解的充要条件是存在 $\boldsymbol{A}^{-}$，使得

$$\boldsymbol{A}\boldsymbol{A}^{-}\boldsymbol{D} = \boldsymbol{D}$$

成立，在有解的条件下，矩阵方程 $\boldsymbol{A}\boldsymbol{X} = \boldsymbol{D}$ 的通解为

$$\boldsymbol{X} = \boldsymbol{A}^{-}\boldsymbol{D} + \boldsymbol{Y} - \boldsymbol{A}\boldsymbol{A}^{-}\boldsymbol{Y}, \quad \forall \boldsymbol{Y} \in \mathbb{C}^{n\times p}$$

推论 6.5.3

设 $\boldsymbol{B} \in \mathbb{C}^{m\times n}, \boldsymbol{D} \in \mathbb{C}^{p\times n}$，则矩阵方程

$$\boldsymbol{X}\boldsymbol{B} = \boldsymbol{D}$$

有解的充要条件是存在 $\boldsymbol{B}^{-}$，使得

$$\boldsymbol{D}\boldsymbol{B}^{-}\boldsymbol{B} = \boldsymbol{D}$$

成立，在有解的条件下，矩阵方程 $\boldsymbol{X}\boldsymbol{B} = \boldsymbol{D}$ 的通解为

$$\boldsymbol{X} = \boldsymbol{D}\boldsymbol{B}^{-} + \boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{B}\boldsymbol{B}^{-}, \quad \forall \boldsymbol{Y} \in \mathbb{C}^{p\times m}$$

6.5.2 广义逆矩阵在解线性方程组上的应用

推论 6.5.4

设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}, \boldsymbol{b} \in \mathbb{C}^{m}$，则方程组 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 有解的充要条件是存在 $\boldsymbol{A}^{-}$，使得

$$\boldsymbol{A}\boldsymbol{A}^{-}\boldsymbol{b} = \boldsymbol{b}$$

成立，此时 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 的通解为

$$\boldsymbol{x} = \boldsymbol{A}^{-}\boldsymbol{b} + (\boldsymbol{E}_n - \boldsymbol{A}^{-}\boldsymbol{A})\boldsymbol{u}, \quad \forall \boldsymbol{u} \in \mathbb{C}^{n}$$

相容方程的最小范数解

如果方程组 $\boldsymbol{Ax} = \boldsymbol{b}$ 有解，那么就叫该方程组为相容的方程组，将所有解中最小的解称为**最小范数解**, 并且最小范数解是唯一的，证明过程可以利用勾股定理进行证明。

不相容方程组的解

如果方程组 $\boldsymbol{Ax} = \boldsymbol{b}$ 无解，就叫该方程组为不相容方程组，此时，虽然没有解，但可以求出离理想结果最近的解，用数学语言描述就是：令 $f(\boldsymbol{x}) = \|\boldsymbol{Ax} - \boldsymbol{b}\|_2^2$，存在 $\boldsymbol{x}_0$ 使得 $f(\boldsymbol{x}_0)$ 最小，这个时候称 $\boldsymbol{x}_0$ 叫做方程组的最小二乘解

不相容方程组 $\boldsymbol{Ax} = \boldsymbol{b}$ 的最小二乘解的通解为

$$\boldsymbol{x} = \boldsymbol{Gb} + (\boldsymbol{E} - \boldsymbol{A}^{-}\boldsymbol{A})\boldsymbol{u}, \quad \forall \boldsymbol{u} \in \mathbb{C}^n$$

这个时候 $\boldsymbol{Gb} = \boldsymbol{A}^{+}\boldsymbol{b}$ 记为方程组的最佳逼近解。