第一讲　线性空间

一、 线性空间的定义及性质

[知识预备]

★集合：笼统的说是指一些事物（或者对象）组成的整体。

集合的表示：枚举、表达式

集合的运算：并（$\bigcup$），交（$\bigcap$）

另外，集合的“和”（＋）：并不是严格意义上集合的运算，因为它限定了集合中元素须有可加性。

★数域：一种数集，对四则运算封闭（除数不为零）。比如有理数域、实数域（R）和复数域（C）。实数域和复数域是工程上较常用的两个数域。

线性空间是线性代数最基本的概念之一，也是学习现代矩阵论的重要基础。

1．线性空间的定义：

设$\boldsymbol{V}$是一个非空集合，其元素用$\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$等表示；$\boldsymbol{K}$是一个数域，其元素用$\boldsymbol{k},\boldsymbol{l},\boldsymbol{m}$等表示。如果$\boldsymbol{V}$满足[如下 8 条性质，分两类]：

（I）在$\boldsymbol{V}$中定义一个“加法”运算，即当$\boldsymbol{x}$，$\boldsymbol{y}\in\boldsymbol{V}$时，有唯一的和$\boldsymbol{x}+\boldsymbol{y}\in\boldsymbol{V}$（封闭性），且加法运算满足下列性质:

（1）结合律　$\boldsymbol{x}+(\boldsymbol{y}+\boldsymbol{z})=(\boldsymbol{x}+\boldsymbol{y})+\boldsymbol{z}$；

（2）交换律　$\boldsymbol{x}+\boldsymbol{y}=\boldsymbol{y}+\boldsymbol{x}$；

（3）零元律　存在零元素$\boldsymbol{O}$，使$\boldsymbol{x}+\boldsymbol{O}=\boldsymbol{x}$；

（4）负元律　对于任一元素$\boldsymbol{x}\in\boldsymbol{V}$，存在一元素$\boldsymbol{y}\in\boldsymbol{V}$，使$\boldsymbol{x}+\boldsymbol{y}=\boldsymbol{O}$，且称$\boldsymbol{y}$为$\boldsymbol{x}$的负元素，记为$(-\boldsymbol{x})$。则有$\boldsymbol{x}+(-\boldsymbol{x})=\boldsymbol{O}$。

（II）在$\boldsymbol{V}$中定义一个“数乘”运算，即当$\boldsymbol{x}\in\boldsymbol{V},\boldsymbol{k}\in\boldsymbol{K}$时，有唯一的$\boldsymbol{kx}\in\boldsymbol{V}$（封闭性），且数乘运算满足下列性质:

（5）数因子分配律　$\boldsymbol{k}(\boldsymbol{x}+\boldsymbol{y})=\boldsymbol{kx}+\boldsymbol{ky}$；

（6）分配律　$(\boldsymbol{k}+\boldsymbol{l})\boldsymbol{x}=\boldsymbol{kx}+\boldsymbol{lx}$；

（7）结合律　$\boldsymbol{k}(\boldsymbol{lx})=(\boldsymbol{kl})\boldsymbol{x}$；

（8）恒等律　$\boldsymbol{1x}=\boldsymbol{x}$；　[数域中一定有$\boldsymbol{1}$]

则称$\boldsymbol{V}$为数域$\boldsymbol{K}$上的线性空间。

注意以下几点：

1）线性空间是基于一定数域来的。同一个集合，对于不同数域，就可能构成不同的线性空间，甚至对有的数域能构成线性空间，而对其他数域不能构成线性空间。

2）两种运算、八条性质。数域$\boldsymbol{K}$中的运算是具体的四则运算，而$\boldsymbol{V}$中所定义的加法运算和数乘运算则是抽象的、形式的。

3）除了两种运算和八条性质外，还应注意唯一性、封闭性是否满足。

当数域 $\boldsymbol{K}$ 为实数域时， $\boldsymbol{V}$ 就称为实线性空间； $\boldsymbol{K}$ 为复数域， $\boldsymbol{V}$ 就称为复线性空间。

例1．　设 $\boldsymbol{R}^+ =${全体正实数}，其“加法”及“数乘”运算定义为

$$x \boxplus y = xy \quad , \qquad k \circ x = x^k$$

证明：$\boldsymbol{R}^+$ 是实数域 $\boldsymbol{R}$ 上的线性空间。

[证明] 首先需要证明两种运算的唯一性和封闭性

①唯一性和封闭性

唯一性显然

若 $x > 0, y > 0$，$k \in R$，则有

$$x \boxplus y = xy \in R^+ ，\ k \circ x = x^k \in R^+ \quad \text{封闭性得证。}$$

②八条性质

（1）　$x \boxplus (y \boxplus z) = x(yz) = (xy)z = (x \boxplus y) \boxplus z$

（2）　$x \boxplus y = xy = yx = y \boxplus x$

（3）　$1$ 是零元素　　$x \boxplus 1 = x \cdot 1 = x$　[ $x \boxplus O = x \to xO = x \to O = 1$ ]

（4）　$1/x$ 是 $x$ 的负元素　　$x \boxplus 1/x = x \cdot 1/x = 1$　　[ $x + y = O$ ]

（5）　$k \circ (x \boxplus y) = (xy)^k = x^k y^k = (k \circ x) \boxplus (k \circ y)$　　[数因子分配律]

（6）　$(k+l) \circ x = x^{k+l} = x^k x^l = (k \circ x) \boxplus (l \circ x)$　[分配律]

（7）　$k \circ (l \circ x) = (x^l)^k = x^{kl} = (kl) \circ x$　　　[结合律]

（8）　$1 \circ x = x^1 = x$　　　　　　[恒等律]

由此可证， $\boldsymbol{R}^+$ 是实数域 $\boldsymbol{R}$ 上的线性空间。

2．定理：线性空间具有如下性质

（1）　零元素是唯一的，任一元素的负元素也是唯一的。

（2）　如下恒等式成立： $0x = O$，$(-1)x = (-x)$。

[证明]（1）采用反证法：

①零元素是唯一的。 设存在两个零元素 $O_1$ 和 $O_2$，则由于 $O_1$ 和 $O_2$ 均为零元素，按零元律有

[交换律]

$$O_1 + O_2 = O_1 = O_2 + O_1 = O_2$$

所以 $\boldsymbol{O}_1=\boldsymbol{O}_2$

即$\boldsymbol{O}_1$和$\boldsymbol{O}_2$相同，与假设相矛盾，故只有一个零元素。

②任一元素的负元素也是唯一的。假设$\forall \boldsymbol{x}\in \boldsymbol{V}$，存在两个负元素$\boldsymbol{y}$和$\boldsymbol{z}$，则根据负元律有

$$\boldsymbol{x}+\boldsymbol{y}=\boldsymbol{O}=\boldsymbol{x}+\boldsymbol{z}$$

$$\boldsymbol{y}=\boldsymbol{y}+\boldsymbol{O}=\boldsymbol{y}+(\boldsymbol{x}+\boldsymbol{z})=(\boldsymbol{y}+\boldsymbol{x})+\boldsymbol{z}=\boldsymbol{O}+\boldsymbol{z}=\boldsymbol{z}$$

[零元律] [结合律] [零元律]

即$\boldsymbol{y}$和$\boldsymbol{z}$相同，故负元素唯一。

（2） ①：设$\boldsymbol{w}=0\boldsymbol{x}$，则$\boldsymbol{x}+\boldsymbol{w}=1\boldsymbol{x}+0\boldsymbol{x}=(1+0)\boldsymbol{x}=\boldsymbol{x}$，故$\boldsymbol{w}=\boldsymbol{O}$。

[恒等律]

②：设$\boldsymbol{w}=(-1)\boldsymbol{x}$，则$\boldsymbol{x}+\boldsymbol{w}=1\boldsymbol{x}+(-1)\boldsymbol{x}=0\boldsymbol{x}=\boldsymbol{O}$，故$\boldsymbol{w}=-\boldsymbol{x}$。

3．线性相关性

线性空间中相关性概念与线性代数中向量组线性相关性概念类似。

•线性组合：$\forall \boldsymbol{x}_1,\boldsymbol{x}_2\cdots,\boldsymbol{x}_m\in \boldsymbol{V},\boldsymbol{c}_1,\boldsymbol{c}_2\cdots,\boldsymbol{c}_m\in \boldsymbol{K}$

$$\boldsymbol{c}_1\boldsymbol{x}_1+\boldsymbol{c}_2\boldsymbol{x}_2+\cdots+\boldsymbol{c}_m\boldsymbol{x}_m=\sum_{i=1}^{m}\boldsymbol{c}_i\boldsymbol{x}_i$$

称为元素组$\boldsymbol{x}_1,\boldsymbol{x}_2\cdots,\boldsymbol{x}_m$的一个线性组合。

•线性表示：$\boldsymbol{V}$中某个元素$\boldsymbol{x}$可表示为其中某个元素组的线性组合，则称$\boldsymbol{x}$可由该元素组线性表示。

•线性相关性：如果存在一组不全为零的数$\boldsymbol{c}_1,\boldsymbol{c}_2\cdots,\boldsymbol{c}_m\in \boldsymbol{K}$，使得对于元素$\boldsymbol{x}_1,\boldsymbol{x}_2\cdots,\boldsymbol{x}_m\in \boldsymbol{V}$有

$$\sum_{i=1}^{m}\boldsymbol{c}_i\boldsymbol{x}_i=\boldsymbol{0}$$

则称元素组$\boldsymbol{x}_1,\boldsymbol{x}_2\cdots,\boldsymbol{x}_m$线性相关，否则称其线性无关。线性相关性概念是个非常重要的概念，有了线性相关性才有下面的线性空间的维数、基和坐标。

4．线性空间的维数

定义：线性空间$\boldsymbol{V}$中最大线性无关元素组所含元素个数称为$\boldsymbol{V}$的维数，记为$\dim \boldsymbol{V}$。

本课程只考虑有限维情况，对于无限维情况不涉及 。

例 2. 全体 m×n 阶实矩阵的集合构成一个实线性空间（对于矩阵加法和数对矩阵的数乘运算），求其维数。

[解] 一个直接的方法就是找一个最大线性无关组，其元素尽可能简单。

令 $\boldsymbol{E_{ij}}$ 为这样的一个 m×n 阶矩阵，其 $\boldsymbol{(i,j)}$ 元素为 1，其余元素为零。

显然，这样的矩阵共有 m×n 个，构成一个具有 m×n 个元素的线性无关元素组 $\left\{\boldsymbol{E_{11}, E_{12}, \cdots, E_{1n}; E_{21}, E_{22,} \cdots, E_{2n}; \cdots; E_{m1}, E_{m2}, \cdots, E_{mn}}\right\}$。另一方面，还需说明元素个数最大。对于任意的 $\boldsymbol{A = (a_{ij})_{m\times n}}$，都可由以上元素组线性表示，

$$\boldsymbol{A = \sum_{i,j} a_{ij} E_{ij} \to \sum_{i,j} a_{ij} E_{ij} - A = 0}$$

即 $\left\{\boldsymbol{E_{ij}} \middle| \boldsymbol{i = 1,2\cdots, m; j = 1,2\cdots, n}\right\}$构成了最大线性无关元素组，所以该空间的维数为 m×n。

二、 线性空间的基与坐标

1.  基的定义：设 $\boldsymbol{V}$ 是数域 $\boldsymbol{K}$ 上的线性空间， $\boldsymbol{x_1, x_2 \cdots, x_r\ (r \geq 1)}$ 是属于 $\boldsymbol{V}$ 的 $\boldsymbol{r}$ 个任意元素，如果它满足

（1） $\boldsymbol{x_1, x_2 \cdots, x_r}$ 线性无关；

（2）$\boldsymbol{V}$ 中任一向量 $x$ 均可由 $\boldsymbol{x_1, x_2 \cdots, x_r}$ 线性表示。

则称 $\boldsymbol{x_1, x_2 \cdots, x_r}$ 为 $\boldsymbol{V}$ 的一个基，并称 $\boldsymbol{x_1, x_2 \cdots, x_r}$ 为该基的基元素。

•基正是 $\boldsymbol{V}$ 中最大线性无关元素组； $\boldsymbol{V}$ 的维数正是基中所含元素的个数。
•基是不唯一的，但不同的基所含元素个数相等。

例3　考虑全体复数所形成的集合 $\boldsymbol{C}$ 。如果 $\boldsymbol{K = C}$ （复数域），则该集合对复数加法和复数的乘法构成线性空间，其基可取为 1，空间维数为 1；如果取 $\boldsymbol{K = R}$ （实数域），则该集合对复数加法及实数对复数的数乘构成线性空间，其基可取为 $\boldsymbol{\{1, i\}}$，空间维数为 2。

| 数域 $K$ | 两种运算 | 基 | 一般元素 | 空间类型 | 维数 |
|---|---|---|---|---|---|
| 复数域 $\boldsymbol{C}$ | （1）复数加法；<br>（2）复数对复数的数乘 | $\boldsymbol{\{1\}}$ | $\boldsymbol{c = c \cdot 1}$ | 复线性空间 | 1 |
| 实数域 $R$ | （1）复数加法；<br>（2）实数对复数的数乘 | $\boldsymbol{\{1, i\}}$ | $\boldsymbol{c = a \cdot 1 + b \cdot i}$ | 实线性空间 | 2 |

2. 坐标的定义：称线性空间$V^n$的一个基 $x_1, x_2 \cdots, x_n$ 为$V^n$的一个坐标系，

$\forall x \in V^n$，它在该基下的线性表示为：

$$\sum_{i=1}^{n} \xi_i x_i \quad (\xi_i \in K, x_i \in V^n, i = 1,2\cdots,n)$$

则称$\xi_1, \xi_2 \cdots, \xi_n$为$x$在该坐标系中的坐标或分量，记为$(\xi_1, \xi_2 \cdots, \xi_n)^T$

讨论：（1）一般来说，线性空间及其元素是抽象的对象，不同空间的元素完全可以具有千差万别的类别及性质。但坐标表示却把它们统一了起来，坐标表示把这种差别留给了基和基元素，由坐标所组成的新向量仅由数域中的数表示出来。

（2）更进一步，原本抽象的“加法”及 “数乘”经过坐标表示就演化为向量加法及数对向量的数乘。

$1^\circ$
$$\begin{aligned} x + y &= (\xi_1 x_1 + \xi_2 x_2 + \cdots + \xi_n x_n) + (\eta_1 x_1 + \eta_2 x_2 + \cdots + \eta_n x_n) \\ &= (\xi_1 + \eta_1) x_1 + (\xi_2 + \eta_2) x_2 + \cdots + (\xi_n + \eta_n) x_n \end{aligned}$$

正对应

$$\begin{cases} x = (\xi_1, \xi_2 \cdots, \xi_n) \\ y = (\eta_1, \eta_2 \cdots, \eta_n) \end{cases} \to x + y = (\xi_1 + \eta_1, \xi_2 + \eta_2 \cdots, \xi_n + \eta_n)$$

$2^\circ$ $\quad kx = k(\xi x + \xi x + \cdots + \xi x) = (k\xi_1) x_1 + (k\xi_2) x_2 + \cdots + (k\xi_n) x_n$

$$\to (k\xi_1, k\xi_2 \cdots, k\xi_n)$$

正对应 $\quad x = (\xi_1, \xi_2 \cdots, \xi_n) \to kx = (k\xi_1, k\xi_2 \cdots, k\xi_3)$

（3）显然，同一元素在不同坐标系中的坐标是不同的。后面我们还要研究这一变换关系。

## 三、 基变换与坐标变换

基是不唯一的，因此，需要研究基改变时坐标变换的规律。

设$x_1, x_2 \cdots, x_n$是$V^n$的旧基，$y_1, y_2 \cdots, y_n$是$V^n$的新基，由于两者都是基，所以可以相互线性表示

$$y_j = \sum_{i=1}^{n} c_{ij} x_i \quad (i = 1,2\cdots,n)$$

即
$$[y_1, y_2 \cdots, y_n] = [x_1, x_2 \cdots, x_n] \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix} = [x_1, x_2 \cdots, x_n] C$$

其中$C$称为过渡矩阵，上式就给出了基变换关系，可以证明，$C$是可逆的。

设$x \in V^n$，它在旧基下的线性表示为

$$x = \sum_{i=1}^{n} \xi_i x_i = [x_1, x_2 \cdots, x_n] \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix}$$

它在新基下的线性表示为

$$x = \sum_{i=1}^{n} \xi_i' y_i = [y_1, y_2 \cdots, y_n] \begin{bmatrix} \xi_1' \\ \xi_2' \\ \vdots \\ \xi_n' \end{bmatrix}$$

则

$$[y_1, y_2 \cdots, y_n] \begin{bmatrix} \xi_1' \\ \xi_2' \\ \vdots \\ \xi_n' \end{bmatrix} = [x_1, x_2 \cdots, x_n] \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix}$$

由于基元素的线性无关性，得到坐标变换关系

$$C \begin{bmatrix} \xi_1' \\ \xi_2' \\ \vdots \\ \xi_n' \end{bmatrix} = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix} \rightarrow \begin{bmatrix} \xi_1' \\ \xi_2' \\ \vdots \\ \xi_n' \end{bmatrix} = C^{-1} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix}$$

作业：P25－26　 3，5，7，9

补充：证明对于线性空间的零元素 $O$， $\forall k \in K$， 均有 $kO = O$。

# 第二讲　线性子空间

## 一、线性子空间的定义及其性质

1. 定义：设$V_1$是数域$K$上的线性空间$V$的一个非空子集合，且对$V$已有的线性运算满足以下条件

（1）　如果$x, y \in V_1$，则$x + y \in V_1$；

（2）　如果$x \in V_1$，$k \in K$，则$kx \in V_1$，

则称$V_1$是$V$的一个线性子空间或子空间。

2. 性质：（1）线性子空间$V_1$与线性空间$V$享有共同的零元素；

（2）$V_1$中元素的负元素仍在$V_1$中。

[证明]（1）$0x = O$

$x \in V_1 \subset V$

$\therefore$ $V$中的零元素也在$V_1$中，$V_1$与$V$享有共同的零元素。

（2）$\forall x \in V_1$

$(-1)x = (-x) \in V_1$　封闭性

$\therefore$ $V_1$中元素的负元素仍在$V_1$中

3. 分类：子空间可分为平凡子空间和非平凡子空间

平凡子空间：{0}和$V$本身

非平凡子空间：除以上两类子空间

4. 生成子空间：设$x_1, x_2 \cdots, x_m$为$V$中的元素，它们的所有线性组合的集合

$$\left\{ \sum_{i=1}^{m} k_i x_i \middle| k_i \in K, i = 1,2\cdots,m \right\}$$

也是$V$的线性子空间，称为由$x_1, x_2 \cdots, x_m$生（张）成的子空间，记为$L(x_1, x_2 \cdots, x_m)$或者$Span(x_1, x_2 \cdots, x_m)$。

若$x_1, x_2 \cdots, x_m$线性无关，则

$$\dim\{L(x_1, x_2 \cdots, x_m)\} = m$$

5. 基扩定理：设$V_1$是数域$K$上的线性空间$V^n$的一个$m$维子空间，$x_1, x_2 \cdots, x_m$是$V_1$的一个基，则这$m$个基向量必可扩充为$V^n$的一个基；换言之，在$V^n$中必可找到$n-m$个元素

$x_{m+1}, x_{m+2} \cdots, x_n$，使得 $x_1, x_2 \cdots, x_n$ 成为 $V^n$ 的一个基。这 $n-m$ 个元素必不在 $V_1$ 中。

二、子空间的交与和

1. 定义：设 $V_1$、$V_2$ 是线性空间 $V$ 的两个子空间，则

$$V_1 \cap V_2 = \{x | x \in V_1, x \in V_2\}$$

$$V_1 + V_2 = \{x + y | x \in V_1, y \in V_2\}$$

分别称为 $V_1$ 和 $V_2$ 的交与和。

2. 定理：若 $V_1$ 和 $V_2$ 是线性空间 $V$ 的两个子空间，则 $V_1 \cap V_2$，$V_1 + V_2$ 均为 $V$ 的子空间

［证明］（1）$\forall x, y \in V_1 \cap V_2$

$$x + y \in V_1 \qquad x + y \in V_2$$

$$\therefore \ x + y \in V_1 \cap V_2$$

$$\forall x \in V_1 \cap V_2 \quad k \in K$$

$$kx \in V_1 \quad kx \in V_2 \quad \therefore \ kx \in V_1 \cap V_2$$

$\therefore$ $V_1 \cap V_2$ 是 $V$ 的一个线性子空间。

（2）$\forall x_1, x_2 \in V_1 \quad \forall y_1, y_2 \in V_2$

$$(x_1 + y_1) \in V_1 + V_2 \qquad (x_2 + y_2) \in V_1 + V_2 \qquad (x_1 + x_2) \in V_1 \qquad (y_1 + y_2) \in V_2$$

$$(x_1 + y_1) + (x_2 + y_2) = (x_1 + x_2) + (y_1 + y_2) \in V_1 + V_2$$

$$\forall k \in K \qquad kx_1 \in V_1 \qquad ky_1 \in V_2$$

$$k(x_1 + y_1) = kx_1 + ky_1 \in V_1 + V_2$$

$\therefore$ $V_1 + V_2$ 是 $V$ 的子空间。

3. 维数公式：若 $V_1$、$V_2$ 是线性空间 $V$ 的子空间，则有

$$\dim(V_1 + V_2) + \dim(V_1 \cap V_2) = \dim V_1 + \dim V_2$$

［证明］设 $\dim V_1 = n_1, \dim V_2 = n_2, \dim(V_1 \cap V_2) = m$

需要证明 $\dim(V_1 + V_2) = n_1 + n_2 - m$

设 $x_1, x_2 \cdots, x_m$ 是 $V_1 \cap V_2$ 的一个基，根据基扩定理

存在 1） $y_1, y_2 \cdots, y_{n1-m} \in V_1$，使 $x_1, x_2 \cdots, x_m, y_1, y_2 \cdots, y_{n1-m}$ 成为 $V_1$ 的一个基；

2） $z_1, z_2 \cdots, z_{n2-m} \in V_2$，使 $x_1, x_2 \cdots, x_m, z_1, z_2 \cdots, z_{n2-m}$

成为 $V_2$ 的一个基；

考察 $x_1, x_2 \cdots, x_m, y_1, y_2 \cdots, y_{n1-m}, z_1, z_2 \cdots, z_{n2-m}$，

若能证明它为 $V_1 + V_2$ 的一个基，则有 $\dim(V_1 + V_2) = n_1 + n_2 - m$。

成为基的两个条件：

1） 它可以线性表示 $V_1 + V_2$ 中的任意元素

2） 线性无关

显然条件 1）是满足的，现在证明条件 2)，采用反证法。

假定上述元素组线性相关，则存在一组不全为 0 的数 $k_1, k_2 \cdots, k_m, p_1, p_2 \cdots, p_{n1-m}, q_1, q_2 \cdots, q_{n2-m}$ 使

$$\sum k_i x_i + \sum p_i y_i + \sum q_i z_i = 0$$

令 $z = \sum q_i z_i \in V_2$，则

$$\sum k_i x_i + \sum p_i y_i = -z \in V_1$$

$$\therefore \quad -z \in V_1 \cap V_2$$

则 $-z$ 可用 $x_1, x_2 \cdots, x_m$ 线性表示，

$$\therefore \quad p_i = 0$$

则，$\sum k_i x_i + \sum q_i z_i = 0$

因为，$x_1, x_2 \cdots, x_m, z_1, z_2 \cdots, z_{n2-m}$ 线性无关

$$\therefore \quad q_i = 0 \quad k_i = 0$$

这与假设矛盾，所以上述元素线性无关，可作为 $V_1 + V_2$ 的一个基。

$$\therefore \quad \dim(V_1 + V_2) = n_1 + n_2 - m$$

三、子空间的直和

1. 定义：设 $V_1$、$V_2$ 是线性空间 $V$ 的子空间，若其和空间 $V_1 + V_2$ 中的任一元素只能唯一的表示为 $V_1$ 的一

个元素与$V_2$的一个元素之和，即$\forall x \in V_1 + V_2$，存在唯一的$y \in V_1, z \in V_2$，使$x = y + z$，则称$V_1 + V_2$为$V_1$与$V_2$的直和，记为$V_1 \oplus V_2$

子空间的直和并不是一种特殊的和，仍然是

$$V_1 + V_2 = \{x + y | x \in V_1, y \in V_2\},$$

反映的是两个子空间的关系特殊。

2. 定理：如下四种表述等价

（1）$V_1 + V_2$成为直和$V_1 \oplus V_2$

（2）$V_1 \cap V_2 = \{0\}$

（3）$\dim(V_1 + V_2) = \dim V_1 + \dim V_2$

（4）$x_1, x_2 \cdots, x_s$为$V_1$的基，$y_1, y_2 \cdots, y_t$为$V_2$的基，则$x_1, x_2 \cdots, x_s, y_1, y_2 \cdots, y_t$为$V_1 + V_2$的基

[证明]（2）和（3）的等价性显然

采用循环证法：（1）$\rightarrow$（2）$\rightarrow$（4）$\rightarrow$（1）

（1）$\rightarrow$（2）：已知$V_1 + V_2 = V_1 \oplus V_2$

假定$x \neq 0$且$x \in V_1 \cap V_2$，则

$$0 = 0 + 0 = x + (-x)$$

$0 \in V_1 + V_2$， $V_1$， $0 \in V_2$， $x \in V_1$， $-x \in V_2$

说明对 0 元素存在两种分解，这与直和的定义矛盾，所以假定不成立，在$V_1 \cap V_2$中只能存在0 元素，即$V_1 \cap V_2 = \{0\}$

（2）$\rightarrow$（4）：已知$V_1 \cap V_2 = \{0\}$

成为基的两个条件：

1） 可以线性表示$V_1 + V_2$中的任意元素

2）线性无关

$\forall x \in V_1$， $y \in V_2$，存在如下坐标表示式

$$x = \sum_{i=1}^{s} \xi_i x_i \qquad y = \sum_{i=1}^{t} \eta_i y_i$$

$x + y$可表示$V_1 + V_2$中的任一元素，

$\therefore\ x_1, x_2\cdots, x_s, y_1, y_2\cdots, y_t$ 可表示$V_1+V_2$中的任意元素。

假设$x_1, x_2\cdots, x_s, y_1, y_2\cdots, y_t$线性相关，即存在不全为 0 的$\xi_1,\xi_2\cdots,\xi_s,\eta_1,\eta_2\cdots,\eta_t$ 使

$$\sum_{i=1}^{s}\xi_i x_i+\sum_{i=1}^{t}\eta_i y_i=0$$

而$x=\sum_{i=1}^{s}\xi_i x_i\in V_1\quad y=\sum_{i=1}^{t}\eta_i y_i\in V_2$

$\therefore\quad \sum_{i=1}^{s}\xi_i x_i=-y\in V_2$

$\therefore\quad \sum_{i=1}^{s}\xi_i x_i\in V_1\bigcap V_2$

$\therefore\quad \sum_{i=1}^{s}\xi_i x_i=0$

$\therefore\quad \xi_1=\xi_2=\cdots=\xi_s=0$

同理$\eta_1=\eta_2=\cdots=\eta_s=0$

这与其线性相关性矛盾，$x_1, x_2\cdots, x_s, y_1, y_2\cdots, y_t$线性无关

$\therefore\quad x_1, x_2\cdots, x_s, y_1, y_2\cdots, y_t$ 可作为$V_1+V_2$的基

（4）$\to$（1）：已知（4）成立

在$x_1, x_2\cdots, x_s, y_1, y_2\cdots, y_t$这组基下

$\forall x\in V_1+V_2$存在唯一的坐标$\xi_1,\xi_2\cdots,\xi_s,\eta_1,\eta_2\cdots,\eta_t$使

$$x=\sum_{i=1}^{s}\xi_i x_i+\sum_{i=1}^{t}\eta_i y_i$$

$$\sum_{i=1}^{s}\xi_i x_i\in V_1\qquad \sum_{i=1}^{t}\eta_i y_i\in V_2$$

$\therefore\quad V_1+V_2$成为直和

作业：P25－26，11、12、13

第三讲 线性变换及其矩阵

一、线性变换及其运算

定义：设$\boldsymbol{V}$是数域$\boldsymbol{K}$上的线性空间，$\boldsymbol{T}$是$\boldsymbol{V}$到自身的一个映射，使得对于$\boldsymbol{V}$中的任意元素$\boldsymbol{x}$均存在唯一的$\boldsymbol{y} \in \boldsymbol{V}$与之对应，则称$\boldsymbol{T}$为$\boldsymbol{V}$的一个变换或算子，记为

$$\boldsymbol{Tx} = \boldsymbol{y}$$

称$\boldsymbol{y}$为$\boldsymbol{x}$在变换$\boldsymbol{T}$下的象，$\boldsymbol{x}$为$\boldsymbol{y}$的原象。

若变换$\boldsymbol{T}$还满足

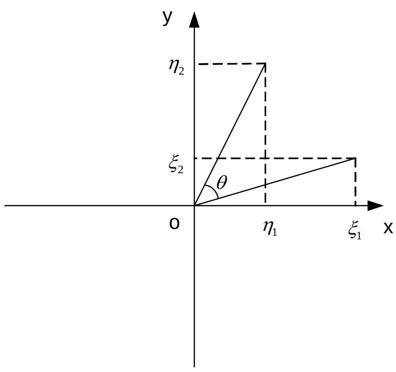
$$\boldsymbol{T}(\boldsymbol{kx}+\boldsymbol{ly}) = \boldsymbol{k}(\boldsymbol{Tx}) + \boldsymbol{l}(\boldsymbol{Ty}) \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{V}, \boldsymbol{k}, \boldsymbol{l} \in \boldsymbol{K}$$

称 $T$ 为线性变换。

［例 1］二维实向量空间$\mathbf{R}^2 = \left\{ \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \middle| \ \xi_{\mathrm{i}} \in \mathbf{R} \right\}$，将其绕原点旋转$\boldsymbol{\theta}$角的操作就是一个线性变换。

［证明］ $\boldsymbol{x} = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \qquad \boldsymbol{y} = \boldsymbol{Tx} = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} \qquad \begin{cases} \eta_1 = \xi_1 \cos\theta - \xi_2 \sin\theta \\ \eta_2 = \xi_1 \sin\theta + \xi_2 \cos\theta \end{cases}$

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \in \mathbf{R}^2$$




可见该操作为变换，下面证明其为线性变换

$$\forall \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \boldsymbol{z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \in \mathbf{R}^2，\ \boldsymbol{k}, \boldsymbol{l} \in \mathbf{R}$$

$$\boldsymbol{kx} + \boldsymbol{lz} = \begin{bmatrix} kx_1 \\ kx_2 \end{bmatrix} + \begin{bmatrix} lz_1 \\ lz_2 \end{bmatrix} = \begin{bmatrix} kx_1 + lz_1 \\ kx_2 + lz_2 \end{bmatrix}$$

$$T(kx+lz)=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}kx_1+lz_1\\ kx_2+lz_2\end{bmatrix}$$
$$=k\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}+l\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}z_1\\ z_2\end{bmatrix}$$
$$=k(Tx)+l(Tz)$$

$\therefore$ $T$ 是线性变换。

［例 2］次数不超过 $n$ 的全体实多项式 $\mathbf{P_n}$ 构成实数域上的一个 $n+1$ 维的线性空间，其基可选为 $\{1,x,x^2,\cdots,x^n\}$，微分算子 $D=\dfrac{d}{dx}$ 是 $\mathbf{P_n}$ 上的一个线性变换。

［证明］显然 $D$ 对 $\mathbf{P_n}$ 而言是变换，

要证明 $D$ 满足线性变换的条件

$$\forall f,g\in\mathbf{P_n},\ \ k,l\in\mathbf{R}$$

$$D(kf+lg)=k(Df)+l(Dg)$$

$\therefore$ $D$ 是 $\mathbf{P_n}$ 上的线性变换。

2. 性质

（1） 线性变换把零元素仍变为零元素

（2） 负元素的象为原来元素的象的负元素

（3） 线性变换把线性相关的元素组仍变为线性相关的元素组

［证明］线性变换 $T(kx+ly)=k(Tx)+l(Ty)$

（1）$T(O)=T(0x)=0(Tx)=O$

（2）$T(-x)=(-1)(Tx)=-(Tx)$

（3）元素组 $x_1,x_2,\cdots,x_m$ 线性相关，即存在一组不全为零的数 $k_1,k_2,\cdots,k_m$ 使

$$\sum_{i=1}^{m}k_ix_i=0$$

则 $$T(\sum_{i=1}^{m}k_ix_i)=\sum_{i=1}^{m}k_i(Tx_i)=T(0)=0$$

$\therefore$ $\{Tx_i\}$ 线性相关。

［得证］

应该注意，线性无关的元素组经过线性变换不一定再是线性无关的，变换后的情况

与元素组和线性变换有关。若线性变换$\mathbf{T}$将所有线性无关的元素组仍变换为线性无关的元素组，则称之为满秩的线性变换。

3. 线性变换的运算

（1） 恒等变换$\boldsymbol{T_e}$：$\forall \boldsymbol{x} \in \boldsymbol{V}, \boldsymbol{T_e x} = \boldsymbol{x}$

（2） 零变换$\boldsymbol{T_0}$：$\forall \boldsymbol{x} \in \boldsymbol{V}, \boldsymbol{T_0 x} = \boldsymbol{0}$

（3） 变换的相等：$\boldsymbol{T_1}$、$\boldsymbol{T_2}$是$\boldsymbol{V}$ 的两个线性变换，$\forall \boldsymbol{x} \in \boldsymbol{V}$，均有$\boldsymbol{T_1 x} = \boldsymbol{T_2 x}$，则称

$\boldsymbol{T_1} = \boldsymbol{T_2}$

（4） 线性变换的和$\boldsymbol{T_1} + \boldsymbol{T_2}$：$\forall \boldsymbol{x} \in \boldsymbol{V}$，$(\boldsymbol{T_1} + \boldsymbol{T_2})\boldsymbol{x} = \boldsymbol{T_1 x} + \boldsymbol{T x_2}$

（5） 线性变换的数乘$\boldsymbol{kT}$：$\forall \boldsymbol{x} \in \boldsymbol{V}$，$(\boldsymbol{kT})\boldsymbol{x} = \boldsymbol{k}(\boldsymbol{Tx})$

负变换：$(-\boldsymbol{T})\boldsymbol{x} = -(\boldsymbol{Tx})$

（6） 线性变换的乘积$\boldsymbol{T_1 T_2}$：$\forall \boldsymbol{x} \in \boldsymbol{V}$，$(\boldsymbol{T_1 T_2})\boldsymbol{x} = \boldsymbol{T_1}(\boldsymbol{T_2 x})$

（7） 逆变换$\boldsymbol{T}^{-1}$：$\forall \boldsymbol{x} \in \boldsymbol{V}$，若存在线性变换$\boldsymbol{S}$使得$(\boldsymbol{ST})\boldsymbol{x} \equiv \boldsymbol{x}$，则称$\boldsymbol{S}$为$\boldsymbol{T}$的逆变换$\boldsymbol{S} = \boldsymbol{T}^{-1}$

（8） 线性变换的多项式：

$$\boldsymbol{T}^n = \underbrace{\boldsymbol{TT\cdots T}}_{n\text{个}}\text{，并规定 } \boldsymbol{T}^0 = \boldsymbol{T_e}$$

$$f(\boldsymbol{T}) = \sum_{n=0}^{N} a_n \boldsymbol{T}^n \quad \rightarrow \quad f(\boldsymbol{T})\boldsymbol{x} = \sum_{n=0}^{N} a_n \boldsymbol{T}^n \boldsymbol{x}$$

需要说明的是：

1）$\boldsymbol{T_e}$也称为单位变换；

2）和矩阵的乘积一样，线性变换的乘积不满足交换律；

3）不是所有的变换都具有逆变换，只有满秩变换才有逆变换，$\boldsymbol{ST} = \boldsymbol{T_e}$；

4）恒等变换、零变换、线性变换的和、乘积、多项式及逆变换（若存在）均为线性变换。

二、线性变换的矩阵表示

线性变换用矩阵表示，将抽象的线性变换转化为具体的矩阵形式。

设$\boldsymbol{T}$ 是线性空间 $\mathbf{V^n}$ 的一个线性变换，且 $\{\boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_n}\}$是 $\mathbf{V^n}$ 的一个基，$\forall \boldsymbol{x} \in \mathbf{V^n}$，存在唯一的坐标表示

$$x=\left[x_1,x_2,\cdots,x_n\right]\begin{bmatrix}\xi_1\\\xi_2\\\vdots\\\xi_n\end{bmatrix}=\xi_1x_1+\xi_2x_2+\cdots+\xi_nx_n$$

$$Tx=T(\xi_1x_1+\xi_2x_2+\cdots+\xi_nx_n)$$

$$=\left[Tx_1,Tx_2,\cdots,Tx_n\right]\begin{bmatrix}\xi_1\\\xi_2\\\vdots\\\xi_n\end{bmatrix}$$

$$=\left[T(x_1,x_2,\cdots,x_n)\right]\begin{bmatrix}\xi_1\\\xi_2\\\vdots\\\xi_n\end{bmatrix}$$

因此，要确定线性变换$T$，只需确定基元素在该变换下的象就可以了。

$$Tx_i=\left[x_1,x_2,\cdots,x_n\right]\begin{bmatrix}a_{i1}\\a_{i2}\\\vdots\\a_{in}\end{bmatrix}$$

$$T\left[x_1,x_2,\cdots,x_n\right]=\left[x_1,x_2,\cdots,x_n\right]\begin{bmatrix}a_{11}&a_{21}&\cdots&a_{n1}\\a_{12}&a_{22}&\cdots&a_{n2}\\\vdots&\vdots&\ddots&\vdots\\a_{1n}&a_{2n}&\cdots&a_{nn}\end{bmatrix}=\left[x_1,x_2,\cdots,x_n\right]A$$

对于任意元素$x$，在该基下，变换后$Tx$的坐标表示为

$$Tx=\left[x_1,x_2,\cdots,x_n\right]\begin{bmatrix}\eta_1\\\eta_2\\\vdots\\\eta_n\end{bmatrix}$$

同时

$$Tx=\left[T(x_1,x_2,\cdots,x_n)\right]\begin{bmatrix}\xi_1\\\xi_2\\\vdots\\\xi_n\end{bmatrix}=\left[x_1,x_2,\cdots,x_n\right]A\begin{bmatrix}\xi_1\\\xi_2\\\vdots\\\xi_n\end{bmatrix}$$

对比可知：

$$\begin{bmatrix}\eta_1\\ \eta_2\\ \vdots\\ \eta_n\end{bmatrix}=A\begin{bmatrix}\xi_1\\ \xi_2\\ \vdots\\ \xi_n\end{bmatrix}$$

即：
$$x\leftrightarrow\begin{bmatrix}\xi_1\\ \xi_2\\ \vdots\\ \xi_n\end{bmatrix}$$

$$Tx\leftrightarrow A\begin{bmatrix}\xi_1\\ \xi_2\\ \vdots\\ \xi_n\end{bmatrix}$$

1. 定义：把$A$称为$T$在基$\{x_1,x_2,\cdots,x_n\}$下的矩阵。

2. 定理：设$\{x_1,x_2,\cdots,x_n\}$是$\mathbf{V}^n$的一个基，$T_1$、$T_2$在该基下的矩阵分别为$A$、$B$。则有

（1）$(T_1+T_2)[x_1,x_2,\cdots,x_n]=[x_1,x_2,\cdots,x_n](A+B)$

（2）$kT_1[x_1,x_2,\cdots,x_n]=[x_1,x_2,\cdots,x_n](kA)$

（3）$(T_1T_2)[x_1,x_2,\cdots,x_n]=[x_1,x_2,\cdots,x_n](AB)$

（4）$T^{-1}[x_1,x_2,\cdots,x_n]=[x_1,x_2,\cdots,x_n]A^{-1}$

推论 1. 设$f(t)=\sum\limits_{i=0}^{m}a_it^i$为纯量 $t$ 的 $m$ 次多项式，$T$为线性空间$\mathbf{V}^n$的一个线性变换，且在$\mathbf{V}^n$的基$\{x_1,x_2,\cdots,x_n\}$下的矩阵为$\mathbf{A}$，则

$$f(T)[x_1,x_2,\cdots,x_n]=[x_1,x_2,\cdots,x_n]f(A)$$

其中 $f(T)=a_0T_e+a_1T+a_2T^2+\cdots+a_nT^n$

$$f(A)=a_0I+a_1A+a_2A^2+\cdots+a_nA^n$$

推论 2. 设线性变换$T$在$\mathbf{V}^n$的基$\{x_1,x_2,\cdots,x_n\}$下的矩阵为$A$，元素$x$在该基下的坐标为$(\xi_1,\xi_2,\cdots,\xi_n)$，则$Tx$在该基下的坐标$(\eta_1,\eta_2,\cdots,\eta_n)$满足

$$\begin{bmatrix}\eta_1\\ \eta_2\\ \vdots\\ \eta_n\end{bmatrix}=A\begin{bmatrix}\xi_1\\ \xi_2\\ \vdots\\ \xi_n\end{bmatrix}$$

3. 相似矩阵

设 $\boldsymbol{T}$ 在 $\mathbf{V^n}$ 的两个基 $\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}$ 及 $\{\boldsymbol{x}_1^{'},\boldsymbol{x}_2^{'},\cdots,\boldsymbol{x}_n^{'}\}$ 的矩阵分别为 $\boldsymbol{A}$ 和 $\boldsymbol{B}$，且 $[\boldsymbol{x}_1^{'},\boldsymbol{x}_2^{'},\cdots,\boldsymbol{x}_n^{'}]=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]\boldsymbol{C}$，则

$$\boldsymbol{B}=\boldsymbol{C}^{-1}\boldsymbol{A}\boldsymbol{C}$$

即 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为相似矩阵。

［证明］

$$\boldsymbol{T}[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]\boldsymbol{A}$$

$$\boldsymbol{T}[\boldsymbol{x}_1^{'},\boldsymbol{x}_2^{'},\cdots,\boldsymbol{x}_n^{'}]=[\boldsymbol{x}_1^{'},\boldsymbol{x}_2^{'},\cdots,\boldsymbol{x}_n^{'}]\boldsymbol{B}$$

$$\boldsymbol{T}[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]\boldsymbol{C}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]\boldsymbol{C}\boldsymbol{B}$$

$$[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]\boldsymbol{A}\boldsymbol{C}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]\boldsymbol{C}\boldsymbol{B}$$

$$\Rightarrow \boldsymbol{A}\boldsymbol{C}=\boldsymbol{C}\boldsymbol{B}\quad \text{即}\ \boldsymbol{B}=\boldsymbol{C}^{-1}\boldsymbol{A}\boldsymbol{C}$$

定理：$\boldsymbol{n}$ 阶方阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 相似的充要条件是 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为同一线性变换在不同基下的矩阵。

［证明］必要性：已知 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 相似，即存在可逆矩阵 $\boldsymbol{P}$ 使 $\boldsymbol{B}=\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$

选取一个基 $\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}$，定义

$$\boldsymbol{T}[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]\boldsymbol{A}$$

考虑 $[\boldsymbol{x}_1^{'},\boldsymbol{x}_2^{'},\cdots,\boldsymbol{x}_n^{'}]=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]\boldsymbol{P}$ 可作为基，且

$$\begin{aligned}\boldsymbol{T}[\boldsymbol{x}_1^{'},\boldsymbol{x}_2^{'},\cdots,\boldsymbol{x}_n^{'}]&=\boldsymbol{T}[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]\boldsymbol{P}\\&=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]\boldsymbol{A}\boldsymbol{P}\\&=[\boldsymbol{x}_1^{'},\boldsymbol{x}_2^{'},\cdots,\boldsymbol{x}_n^{'}]\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}\\&=[\boldsymbol{x}_1^{'},\boldsymbol{x}_2^{'},\cdots,\boldsymbol{x}_n^{'}]\boldsymbol{B}\end{aligned}$$

$\therefore$ $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为同一线性变换在不同基下的矩阵。

充分性的证明由相似矩阵定义的证明给出。

三、线性变换及矩阵的值域和核

1. 定义：设$\boldsymbol{T}$是线性空间$\mathbf{V^n}$的线性变换，称

$$\boldsymbol{R(T)}=\left\{\boldsymbol{Tx}\mid \boldsymbol{x}\in\mathbf{V^n}\right\}$$ 为$\boldsymbol{T}$的值域；

$$\boldsymbol{N(T)}=\left\{\boldsymbol{x}\mid \boldsymbol{x}\in\mathbf{V^n},\boldsymbol{Tx}=\mathbf{0}\right\}$$ 称为$\boldsymbol{T}$的核。

$\boldsymbol{R(T)}$和$\boldsymbol{N(T)}$均为$\mathbf{V^n}$的子空间。

设$\boldsymbol{A}$为$\boldsymbol{m\times n}$阶矩阵，称

$$\boldsymbol{R(A)}=\left\{\boldsymbol{Ax}\mid \boldsymbol{x}\in\mathbf{R^n}\ \mathbf{or}\quad \boldsymbol{x}\in\mathbf{C^n}\right\}$$ 为矩阵$\boldsymbol{A}$的值域；

$$\boldsymbol{N(A)}=\left\{\boldsymbol{x}\mid \boldsymbol{x}\in\mathbf{R^n}\ \mathbf{or}\quad \boldsymbol{x}\in\mathbf{C^n},\boldsymbol{Ax}=\mathbf{0}\right\}$$ 为$\boldsymbol{A}$的核。

$\mathbf{dim}\,\boldsymbol{R(T)}$、$\mathbf{dim}\,\boldsymbol{N(T)}$称为$\boldsymbol{T}$的秩和零度；

$\mathbf{dim}\,\boldsymbol{R(A)}$、$\mathbf{dim}\,\boldsymbol{N(A)}$称为$\boldsymbol{A}$的秩和零度。

2. 定理：（1）$\mathbf{dim}\,\boldsymbol{R(T)}+\mathbf{dim}\,\boldsymbol{N(T)}=\mathbf{dim}\,\mathbf{V^n}$

（2）$\mathbf{dim}\,\boldsymbol{R(A)}=\mathbf{rank}(\boldsymbol{A})$

（3）$\mathbf{dim}\,\boldsymbol{R(A)}+\mathbf{dim}\,\boldsymbol{N(A)}=\boldsymbol{n}$，$\boldsymbol{n}$为$\boldsymbol{A}$的列数。

若$\boldsymbol{A}$是线性变换$\boldsymbol{T}$的矩阵，则

$$\mathbf{dim}\,\boldsymbol{R(T)}=\mathbf{dim}\,\boldsymbol{R(A)},\quad \mathbf{dim}\,\boldsymbol{N(T)}=\mathbf{dim}\,\boldsymbol{N(A)}$$

作业：P77－78，1、26、7

第四讲 矩阵的对角化

基

| | 元素 | 坐标向量 |
|---|---|---|
| 加法 | 元素加法 | 坐标向量的加法 |
| 数乘 | 数与元素“乘” | 数与坐标向量相乘 |
| 线性变换及其作用 | 对应关系 | 矩阵与坐标列向量的乘积 |

对任何线性空间，给定基后，我们对元素进行线性变换或线性运算时，只需用元素的坐标向量以及线性变换的矩阵即可，因此，在后面的内容中着重研究矩阵和向量。

---

对角矩阵的形式比较简单，处理起来较方便，比如求解矩阵方程 $\boldsymbol{Ax}=\boldsymbol{b}$ 时，将矩阵 $\boldsymbol{A}$ 对角化后很容易得到方程的解。对角化的过程实际上是一个去耦的过程。以前我们学习过相似变化对角化。那么，一个方阵是否总可以通过相似变化将其对角化呢？或者对角化需要什么样的条件呢？如果不能对角化，我们还可以做哪些处理使问题变得简单呢？

一、 特征值与特征向量

1. 定义：对 $\boldsymbol{m}$ 阶方阵 $\boldsymbol{A}$，若存在数 $\boldsymbol{\lambda}$，及非零向量（列向量）$\boldsymbol{x}$，使得 $\boldsymbol{Ax}=\boldsymbol{\lambda x}$，则称 $\boldsymbol{\lambda}$ 为 $\boldsymbol{A}$ 的特征值，$\boldsymbol{x}$ 为 $\boldsymbol{A}$ 的属于特征值 $\boldsymbol{\lambda}$ 的特征向量。
   - 特征向量不唯一
   - 特征向量非零
   - $(\lambda I-A)x=0$ 有非零解，则 $\det(\lambda I-A)=0$，称 $\det(\lambda I-A)$ 为 $A$ 的特征多项式。

[例 1] $\boldsymbol{A}=\begin{bmatrix}1&2&2\\2&1&2\\2&2&1\end{bmatrix}$，求其特征值和特征向量。

[解] $$\det(\lambda I-A)=\begin{vmatrix}\lambda-1&-2&-2\\-2&\lambda-1&-2\\-2&-2&\lambda-1\end{vmatrix}=0$$

$$(\lambda+1)^2(\lambda-5)=0$$

$$\lambda_1=\lambda_2=-1 \qquad \lambda_3=5$$

属于特征值 $\lambda=-1$ 的特征向量 $\qquad (-I-A)x=0$

$$\begin{bmatrix}2&2&2\\2&2&2\\2&2&2\end{bmatrix}\begin{bmatrix}\xi_1\\\xi_2\\\xi_3\end{bmatrix}=0 \qquad \xi_1+\xi_2+\xi_3=0 \qquad \begin{cases}\xi_1=\xi_1\\\xi_2=\xi_2\\\xi_3=-\xi_1-\xi_2\end{cases}$$

可取基础解系为 $x_1=\begin{bmatrix}1\\0\\-1\end{bmatrix}$ $x_2=\begin{bmatrix}0\\1\\-1\end{bmatrix}$

属于 $\lambda=5$ 的特征向量 $(5I-A)x=0$

$$\begin{bmatrix}4 & -2 & -2\\-2 & 4 & -2\\-2 & -2 & 4\end{bmatrix}\begin{bmatrix}\xi_1\\\xi_2\\\xi_3\end{bmatrix}=0 \qquad \xi_1=\xi_2=\xi_3$$

可取基础解系为 $x_3=\begin{bmatrix}1\\1\\1\end{bmatrix}$

2. 矩阵的迹与行列式

$\mathrm{tr}A=\sum_{i=1}^{n}a_{ii}$ 所有对角元素之和

$$\det A=\prod_{i=1}^{n}\lambda_i \qquad \mathrm{tr}A=\sum_{i=1}^{n}\lambda_i$$

3. 两个定理

（1） 设 $A$、$B$ 分别为 $m\times n$ 和 $n\times m$ 阶矩阵，则

$$\mathrm{tr}(AB)=\mathrm{tr}(BA)$$

（2）sylvster 定理：设 $A$、$B$ 分别为 $m\times n$ 和 $n\times m$ 阶矩阵，则

$$\det(\lambda I_m-AB)=\lambda^{m-n}\det(\lambda I_n-BA)$$

即：$AB$ 与 $BA$ 的特征值只差零特征值的个数，非零特征值相同。

二、 矩阵对角化的充要条件

定理：$n$ 阶方阵 $A$ 可通过相似变换对角化的充要条件是它具有 $n$ 个线性无关的特征向量。

［证明］充分性：已知 $A$ 具有 $n$ 个线性无关的特征向量 $x_1,x_2,\cdots,x_n$，则

$$Ax_i=\lambda_i x_i \qquad i=1,2,\cdots,n$$

$$A\begin{bmatrix}x_1 & x_2 & \cdots & x_n\end{bmatrix}=\begin{bmatrix}\lambda_1x_1 & \lambda_2x_2 & \cdots & \lambda_nx_n\end{bmatrix}$$

$$=\begin{bmatrix}x_1 & x_2 & \cdots & x_n\end{bmatrix}\begin{bmatrix}\lambda_1 & & & 0\\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n\end{bmatrix}$$

$x_1, x_2, \cdots, x_n$ 线性无关，故 $P = [x_1 \ \ x_2 \ \ \cdots \ \ x_n]$ 为满秩矩阵，

令 $\Lambda = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$，则有

$$AP = P\Lambda$$

$$P^{-1}AP = \Lambda$$

必要性：已知存在可逆方阵 $\mathbf{P}$，使 $P^{-1}AP = \Lambda = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$

将 $P$ 写成列向量 $P = [P_1 \ \ P_2 \ \ \cdots \ \ P_n]$，$P_n$ 为 $n$ 维列向量

$$[AP_1 \ \ AP_2 \ \ \cdots \ \ AP_n] = [\lambda_1 P_1 \ \ \lambda_2 P_2 \ \ \cdots \ \ \lambda_n P_n]$$

可见，$\lambda_i$ 为 $A$ 的特征值，$P_i$ 为 $A$ 的特征向量，

$\therefore$ $A$ 具有 $n$ 个线性无关的特征向量。

推论：$n$ 阶方阵有 $n$ 个互异的特征值，则必可对角化。（充分条件）

三、 内积空间

1. Euclid 空间

设 $\mathbf{V}$ 是实线性空间（$k \in \mathbf{R}$），对于 $\mathbf{V}$ 中任何两个元素 $x$、$y$ 均按某一规则存在一个实数与之对应，记为 $(x, y)$，若它满足

（1）交换律 $(x, y) = (y, x)$

（2）分配律 $(x, y+z) = (x, y) + (x, z)$

（3）齐次律 $(kx, y) = k(x, y)$

（4）非负性 $(x, x) \geq 0$，当且仅当 $x = 0$ 时，$(x, x) = 0$

则称 $(x, y)$ 为 $x$ 与 $y$ 的内积，定义了内积的实线性空间称为 Euclid 空间。

对于一个给定的线性空间，可以定义多种内积，较典型的如三维向量空间的数量积就满足以上四条性质，构成内积。以 $n$ 维向量空间为例：

$$x = [\xi_1 \ \ \xi_2 \ \ \cdots \ \ \xi_n]^T, \quad y = [\eta_1 \ \ \eta_2 \ \ \cdots \ \ \eta_n]^T$$

可定义内积$\left(x,y\right)=\sum_{i=1}^{n}w_i\xi_i\eta_i$，它满足内积的四条性质：

（1）$\left(x,y\right)=\sum_{i=1}^{n}w_i\xi_i\eta_i=\sum_{i=1}^{n}w_i\eta_i\xi_i=\left(y,x\right)$

（2）$\left(x,y+z\right)=\sum_{i=1}^{n}w_i\xi_i(\eta_i+\varsigma_i)=\sum_{i=1}^{n}w_i\xi_i\eta_i+\sum_{i=1}^{n}w_i\xi_i\varsigma_i=\left(x,y\right)+\left(x,z\right)$

（3）$\left(kx,y\right)=\sum_{i=1}^{n}w_i(k\xi_i)\eta_i=k\sum_{i=1}^{n}w_i\xi_i\eta_i=k\left(x,y\right)$

（4）$\left(x,x\right)=\sum_{i=1}^{n}w_i\xi_i^2\geq 0$　当且仅当$x_i=0$时，$\left(x,x\right)=0$

该内积可写为：$\left(x,y\right)=x^{\mathrm{T}}Wy$，其中$W=\begin{bmatrix} w_1 & & & 0 \\ & w_2 & & \\ & & \ddots & \\ 0 & & & w_n \end{bmatrix}$

更一般的，对正定实对称矩阵$A$，$\left(x,y\right)=x^{\mathrm{T}}Ay$也满足内积的定义。

正定：（1）特征值全为正（2）各阶顺序主子式大于 0

2. 酉空间：

设$\mathbf{V}$是复线性空间（$k\in\mathbf{C}$），对于$\mathbf{V}$中任何两个元素$x$、$y$均按某一规则存在一个复数与之对应，记为$\left(x,y\right)$，若它满足

（1）交换律 $\left(x,y\right)=\overline{\left(y,x\right)}$

（2）分配律 $\left(x,y+z\right)=\left(x,y\right)+\left(x,z\right)$

（3）齐次律 $\left(kx,y\right)=k\left(x,y\right)$　　or　$\left(x,ky\right)=\bar{k}\left(x,y\right)$

（4）非负性 $\left(x,x\right)\geq 0$，当且仅当$x=0$时，$\left(x,x\right)=0$

则称$\left(x,y\right)$为$x$与$y$的内积，定义了内积的复线性空间称为酉空间。

以$n$维向量空间为例，$A$为厄米（$A^{\mathrm{H}}=A$）正定（$x^{\mathrm{H}}Ax>0$）矩阵，

$$\left(x,y\right)=x^{\mathrm{T}}A\bar{y}=\sum_{i=1}^{n}\sum_{j=1}^{n}\xi_i a_{ij}\bar{\eta}_j$$

较常见的比如$A=\mathbf{diag}[w_1\quad w_2\quad \mathrm{L}\quad w_n]$，$w_i>0$

最简单：实 $\left(x,y\right)=x^{\mathrm{T}}y$

复 $\left(x,y\right)=x^{\mathrm{T}}\overline{y}$

3. 正交性：若$\left(x,y\right)=0$，则称 $x$ 与 $y$ 正交。

$$x\text{与}y\text{的夹角：}\cos\alpha=\frac{(x,y)}{|x||y|},\ \alpha\text{称为}x\text{与}y\text{的夹角。}$$

4. Gram-Schmidt 正交化手续

设 $x_1,x_2,\cdots,x_n$ 为一组线性无关的元素或向量，可以进行如下正交归一化操作（正交规范化或正交单位化）：

1° $y_1=\dfrac{x_1}{|x_1|}$

2° $x_2'=x_2+k_{21}y_1$ 选择合适的 $k_{21}$ 使 $x_2'$ 与 $\mathbf{y}_1$ 正交，

$$(x_2',y_1)=(x_2,y_1)+k_{21}(y_1,y_1)=0$$

$$k_{21}=-(x_2,y_1)$$

$$y_2=\frac{x_2'}{|x_2'|}$$

3° $x_3'=x_3+k_{31}y_1+k_{32}y_2$ 选择 $k_{31}$、$k_{32}$ 使 $x_3'$ 与 $y_1$ 和 $y_2$ 均正交

$$(x_3',y_1)=(x_3',y_2)=0$$

$$(x_3',y_1)=(x_3,y_1)+k_{31}=0\ \rightarrow k_{31}=-\left(x_3,y_1\right)$$

$$(x_3',y_2)=(x_3,y_2)+k_{32}=0\ \rightarrow k_{32}=-\left(x_3,y_2\right)$$

$$y_3=\frac{x_3'}{|x_3'|}$$

一般的，$x_i'=x_i+\sum_{j=1}^{i-1}k_{ij}y_j\quad i=1,2,\cdots,n$

$$k_{ij}=-\left(x_i,y_j\right)$$

$$y_i=\frac{x_i'}{|x_i'|}$$

$y_1, y_2, \cdots, y_n$成为一组正交归一化向量：$(y_i, y_j) = \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$

若 $x_1, x_2, \cdots, x_n$ 为一组基元素，则 $y_1, y_2, \cdots, y_n$ 成为标准正交基。

作业：P106－107　　1(1)(2),2,4,5,10,11

第五讲 对角化与 Jordan 标准形

一、正规矩阵

1. 实对称矩阵与厄米矩阵

实对称矩阵：实矩阵 $\boldsymbol{A}$ $\quad \boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}$

厄米矩阵：复矩阵 $\boldsymbol{A}$ $\quad \boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$

实反对称矩阵：实矩阵 $\boldsymbol{A}$ $\quad \boldsymbol{A}^{\mathrm{T}}=-\boldsymbol{A}$

反厄米矩阵：复矩阵 $\boldsymbol{A}$ $\quad \boldsymbol{A}^{\mathrm{H}}=-\boldsymbol{A}$

2. 正交矩阵和酉矩阵

正交矩阵：实矩阵 $\boldsymbol{A}$ $\quad \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{I}$ （$\boldsymbol{A}^{-1}=\boldsymbol{A}^{\mathrm{T}}$）

酉矩阵：复矩阵 $\boldsymbol{A}$ $\quad \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{I}$ （$\boldsymbol{A}^{-1}=\boldsymbol{A}^{\mathrm{H}}$）

3. 正交相似变换和酉相似变换

$\boldsymbol{P}$ 为正交矩阵， $\boldsymbol{A}$ 为实矩阵， $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ 为对 $\boldsymbol{A}$ 的正交相似变换；

$\boldsymbol{P}$ 为酉矩阵， $\boldsymbol{A}$ 为复矩阵， $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ 为对 $\boldsymbol{A}$ 的酉相似变换。

4. 正规矩阵

实矩阵 $\boldsymbol{A}$，若满足 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$，则 $\boldsymbol{A}$ 为实正规矩阵；

复矩阵 $\boldsymbol{A}$，若满足 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$，则 $\boldsymbol{A}$ 为复正规矩阵。

显然，实对称矩阵、实反对称矩阵、正交矩阵均为实正规矩阵；
厄米矩阵、反厄米矩阵、酉矩阵均为复正规矩阵。

5. 相似矩阵具有相同的特征多项式 $\longrightarrow$ 相同的特征值、迹、行列式。

$$\begin{aligned}\det(\lambda \boldsymbol{I}-\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}) &= \det[\boldsymbol{P}^{-1}(\lambda \boldsymbol{I}-A)\boldsymbol{P}]\\ &= \det(\boldsymbol{P}^{-1})\det(\lambda \boldsymbol{I}-A)\det(\boldsymbol{P})\\ &= \det(\boldsymbol{P}^{-1})\det(\boldsymbol{P})\det(\lambda \boldsymbol{I}-A)\\ &= \det(\lambda \boldsymbol{I}-A)\end{aligned}$$

（$\det(\boldsymbol{A}\boldsymbol{B})=\det(A)\det(\boldsymbol{B})$）

二、酉对角化

1. Schur 引理：设数 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 是 $\boldsymbol{n}$ 阶方阵 $\boldsymbol{A}$ 的特征值，则存在酉矩阵 $\boldsymbol{U}$，使

$$U^{-1}AU=\begin{bmatrix}\lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n\end{bmatrix}$$

［证明］设 $x_1$ 是 $A$ 的属于特征值 $\lambda_1$ 的特征向量，即 $Ax_1=\lambda_1 x_1$，

$u_1=\dfrac{x_1}{|x_1|}$，并由其扩充为一组标准正交向量 $u_1,u_2,\cdots,u_n$

$$u_i^{\mathrm{H}}u_j=\begin{cases}0 & i\neq j\\ 1 & i=j\end{cases}$$

令 $U_0=[u_1\ \ u_2\ \ \cdots\ \ u_n]$，$U_0$ 为酉矩阵

$$U_0^HU_0=\begin{bmatrix}u_1^H\\ u_2^H\\ \vdots\\ u_n^H\end{bmatrix}[u_1\ \ u_2\ \ \cdots\ \ u_n]=\begin{bmatrix}u_1^Hu_1 & u_1^Hu_2 & \cdots & u_1^Hu_n\\ u_2^Hu_1 & u_2^Hu_2 & \cdots & u_2^Hu_n\\ \vdots & \vdots & \ddots & \vdots\\ u_n^Hu_1 & u_n^Hu_2 & \cdots & u_n^Hu_n\end{bmatrix}=I_n$$

对 $A$ 进行酉相似变换：

$$U_0^HAU_0=\begin{bmatrix}u_1^H\\ u_2^H\\ \vdots\\ u_n^H\end{bmatrix}A[u_1\ \ u_2\ \ \cdots\ \ u_n]=(u_i^HAu_j)_{n\times n}$$

第一列：$u_i^{\mathrm{H}}Au_1=u_i^{\mathrm{H}}\lambda_1u_1=\lambda_1u_i^{\mathrm{H}}u_1=\begin{cases}0 & i\neq 1\\ \lambda_1 & i=1\end{cases}$

$$U_0^HAU_0=\left[\begin{array}{c|c}\lambda & *\\ \hline 0 & \\ \vdots & (A_1)_{(n-1)\times(n-1)}\\ 0 & \end{array}\right]$$

$$(A_1)_{(n-1)(n-1)}=\begin{bmatrix}u_2^H\\ u_3^H\\ \vdots\\ u_n^H\end{bmatrix}A[u_2\ \ u_3\ \ \cdots\ \ u_n]=\begin{bmatrix}u_2^HAu & \cdots & u_2^HAu_n\\ \vdots & \ddots & \vdots\\ u_n^HAu_2 & \cdots & u_n^HAu_n\end{bmatrix}$$

相似矩阵具有相同的特征值，因此，对于 $A_1$，其特征值为 $\lambda_2,\cdots,\lambda_n$，与上相同，

可得一个酉矩阵$\boldsymbol{U}_1$，使得

$$\boldsymbol{U}_1^H \boldsymbol{A}_1 \boldsymbol{U}_1 = \left[\begin{array}{c|c} \lambda_2 & * \\ \hline \mathbf{0} & \\ \vdots & (\boldsymbol{A}_2)_{(n-2)\times(n-2)} \\ \mathbf{0} & \end{array}\right]$$

依次类推，分别可找到酉矩阵$\boldsymbol{U}_2, \boldsymbol{U}_3, \cdots, \boldsymbol{U}_{n-2}$使

$$\boldsymbol{U}_2^H \boldsymbol{A}_2 \boldsymbol{U}_2 = \left[\begin{array}{c|c} \lambda_3 & * \\ \hline \mathbf{0} & \\ \vdots & (\boldsymbol{A}_3)_{(n-3)\times(n-3)} \\ \mathbf{0} & \end{array}\right]$$

$$\vdots$$

$$\vdots$$

$$\boldsymbol{U}_{n-2}^{\mathrm{H}} \boldsymbol{A}_{n-2} \boldsymbol{U}_{n-2} = \begin{bmatrix} \lambda_{n-1} & * \\ \mathbf{0} & \lambda_n \end{bmatrix}$$

令$\boldsymbol{U} = \boldsymbol{U}_0 \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{U}_1 \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{U}_2 \end{bmatrix} \cdots \begin{bmatrix} \boldsymbol{I}_{n-2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{U}_{n-2} \end{bmatrix}$

$\boldsymbol{U}$ 是酉矩阵，$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U} = \boldsymbol{I}$

$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U} = ?$

$$\boldsymbol{U}^H \boldsymbol{A} \boldsymbol{U} = \begin{bmatrix} \boldsymbol{I}_{n-2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{U}_{n-2}^H \end{bmatrix} \cdots \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{U}_1^H \end{bmatrix} \boldsymbol{U}_0^H \boldsymbol{A} \boldsymbol{U}_0 \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{U}_1 \end{bmatrix} \cdots \begin{bmatrix} \boldsymbol{I}_{n-2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{U}_{n-2} \end{bmatrix}$$

$$\boldsymbol{U}_0^{\mathrm{H}} \boldsymbol{A} \boldsymbol{U}_0 = \begin{bmatrix} \lambda_1 & * \\ \mathbf{0} & \boldsymbol{A}_1 \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{U}_1^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} \lambda_1 & * \\ \mathbf{0} & \boldsymbol{A}_1 \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{U}_1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & * \\ \mathbf{0} & \boldsymbol{U}_1^{\mathrm{H}} \boldsymbol{A}_1 \boldsymbol{U}_1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & * & * \\ \mathbf{0} & \lambda_2 & * \\ \mathbf{0} & \mathbf{0} & \boldsymbol{A}_2 \end{bmatrix}$$

$$\vdots$$

$$\vdots$$

$$U^H AU = \begin{bmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$$

［得证］

什么样的矩阵能够通过酉相似变换成为对角阵呢？

2. 定理：$n$阶方阵$A$，酉相似于对角阵的充要条件是：$A$为正规阵（实或复）。

［证明］由 Schur 引理：存在酉矩阵$U$使得

$$\Lambda = U^H AU = \begin{bmatrix} \lambda_1 & & & t_{ij} \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix} \quad 1 \le i \le j \le n$$

$\lambda_1, \lambda_2, \cdots, \lambda_n$是$A$的特征值。

$$\Lambda^H = (U^H AU)^H = \begin{bmatrix} \overline{\lambda_1} & & & 0 \\ & \overline{\lambda_2} & & \\ & & \ddots & \\ t_{ij} & & & \overline{\lambda_n} \end{bmatrix}$$

充分性:已知$A$为正规阵，即$A^{\mathrm{H}}A = AA^{\mathrm{H}}$，要证$t_{ij} = 0$

$$\begin{cases} \Lambda\Lambda^{\mathrm{H}} = U^{\mathrm{H}}AA^{\mathrm{H}}U \\ \Lambda^{\mathrm{H}}\Lambda = U^{\mathrm{H}}A^{\mathrm{H}}AU \end{cases}$$

$$\Lambda\Lambda^{\mathrm{H}} = \Lambda^{\mathrm{H}}\Lambda$$

$$\Lambda^H\Lambda = \begin{bmatrix} |\lambda_1|^2 & & \\ & |\lambda_2|^2 & \\ & & \ddots \end{bmatrix}$$

$$\Lambda\Lambda^H = \begin{bmatrix} |\lambda_1|^2 + \sum |t_{1j}|^2 & & \\ & |\lambda_2|^2 + \sum |t_{2j}|^2 & \\ & & \ddots \end{bmatrix}$$

由对角元素相等可得$t_{1j} = 0$，$t_{2j} = 0$，$\cdots$，$t_{nj} = 0$

$\therefore \quad t_{ij} = 0$

$$\therefore \quad \boldsymbol{U}^{H}\boldsymbol{A}\boldsymbol{U}=\begin{bmatrix}\lambda_1 & & & \mathbf{0}\\ & \lambda_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \lambda_n\end{bmatrix}$$

必要性：已知存在酉矩阵$\boldsymbol{U}$ 使$\boldsymbol{U}^{H}\boldsymbol{A}\boldsymbol{U}=\begin{bmatrix}\lambda_1 & & & \mathbf{0}\\ & \lambda_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \lambda_n\end{bmatrix}=\Lambda$，要证$\boldsymbol{A}$为正规矩阵。

$$\begin{cases}\Lambda\Lambda^{\mathrm{H}}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}\\ \Lambda^{\mathrm{H}}\Lambda=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}\end{cases}$$

$$\because \quad \Lambda\Lambda^{\mathrm{H}}=\Lambda^{\mathrm{H}}\Lambda$$

$$\therefore \quad \boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}$$

$\because \boldsymbol{U}$ 可逆

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

［得证］

说明：（1）不能酉对角化的矩阵仍有可能采用其它可逆变换将其对角化，例如

$$\boldsymbol{A}=\begin{bmatrix}\mathbf{1} & \mathbf{2}\\ \mathbf{0} & \mathbf{3}\end{bmatrix} \qquad \boldsymbol{A}^{\mathrm{T}}=\begin{bmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{2} & \mathbf{3}\end{bmatrix}$$

$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\neq\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$　　$\boldsymbol{A}$不是正规矩阵

但$\boldsymbol{\lambda}(\boldsymbol{A})=\mathbf{1},\mathbf{3}$，两个特征值互异，可以相似变换对角化。可见，$\boldsymbol{A}$可以对角化，但不能酉对角化。

（2）实正规矩阵一般不能通过正交相似变换对角化。（若特征值全为实数，则可正交相似对角化）

如$\boldsymbol{A}=\begin{bmatrix}\mathbf{1} & \mathbf{2}\\ \mathbf{-2} & \mathbf{1}\end{bmatrix}$，特征值为$1\pm 2j$，$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\begin{bmatrix}\mathbf{5} & \mathbf{0}\\ \mathbf{0} & \mathbf{5}\end{bmatrix}$ 正规阵，但不可能正交对角化。

不能对角化的矩阵一定具有多重特征值，对于不能对角化的矩阵也希望找到某种标准形式，使之尽量接近对角化的形式——Jordan 标准形。

三、Jordan 标准形

1. Jordan 标准形的存在定理

任何方阵$\boldsymbol{A}$均可通过某一相似变换化为如下 Jordan 标准形：

$$J=\begin{bmatrix} J_1(\lambda_1) & & & 0 \\ & J_2(\lambda_2) & & \\ & & \ddots & \\ 0 & & & J_s(\lambda_s) \end{bmatrix}$$

其中 $J_i(\lambda_i)=\begin{bmatrix} \lambda_i & 1 & & 0 \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{bmatrix}$ 称为 Jordan 块矩阵。$\lambda_1,\lambda_2,\cdots,\lambda_s$ 为 $A$ 的特征值，可以是多重的。

说明：（1）$J_i(\lambda_i)$ 中的特征值全为 $\lambda_i$，但是对于不同的 $i$、$j$，有可能 $\lambda_i=\lambda_j$，即多重特征值可能对应多个 Jordan 块矩阵。

（2）Jordan 标准形是唯一的，这种唯一性是指：各 Jordan 块矩阵的阶数和对应的特征值是唯一的，但是各 Jordan 块矩阵的位置可以变化。

2. 多项式矩阵（又称为 $\lambda$ 阵）

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

称为 $\lambda$ 的多项式矩阵，其中矩阵元素 $a_{ij}(\lambda)$ 为 $\lambda$ 的多项式。

- 多项式矩阵的初等变换

初等变换的目的是为了在保持矩阵原有属性的前提下形式上变得简单。

（1） 互换两行（列）

（2） 以非零常数乘以某行（列） ［这里不能乘以 $\lambda$ 的多项式或零，这样有可能改变原来矩阵的秩和属性］

（3） 将某行（列）乘以 $\lambda$ 的多项式加到另一行（列）

- 多项式矩阵的标准形式：采用初等变换可将多项式矩阵化为如下形式：

$$A(\lambda)\to\begin{bmatrix} d_1(\lambda) & & & & & & 0 \\ & d_2(\lambda) & & & & & \\ & & \ddots & & & & \\ & & & d_r(\lambda) & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ 0 & & & & & & 0 \end{bmatrix}$$

其中，多项式 $d_i(\lambda)$ 是首一多项式（首项系数为 1，即最高幂次项的系数为 1），且 $d_1(\lambda)|d_2(\lambda)$、$d_2(\lambda)|d_3(\lambda)$、$\cdots$、$d_{r-1}(\lambda)|d_r(\lambda)$，即 $d_i(\lambda)$ 是 $d_{i+1}(\lambda)$ 的因式。

（1） 多项式矩阵的标准形式不随所采用的初等变换而变，故称 $d_i(\lambda)$ 为不变因子。

（2） 不变因子又可采用如下方法求得：设 $D_i(\lambda)$ 为 $A(\lambda)$ 的所有 $i$ 阶子行列式的最大公因式，则 $d_i(\lambda)=\dfrac{D_i(\lambda)}{D_{i-1}(\lambda)}$，$D_0(\lambda)=1$。$D_i(\lambda)$ 称为 i 阶行列式因子。

（3） 将每个不变因子化为不可约因式，这些不可约因式称为 $A(\lambda)$ 的初等因子，全体初等因子称为初等因子组。例如：

$$d_1(\lambda)=(\lambda-2)^2(\lambda-3)\to(\lambda-2)^2\text{和}(\lambda-3)$$

$$d_2(\lambda)=(\lambda-2)^2(\lambda-3)^5\to(\lambda-2)^2\text{和}(\lambda-3)^5$$

初等因子组中应包括两个 $(\lambda-2)^2$。

3. Jordan 标准形的求法

（1） 求出特征矩阵 $(\lambda I-A)$ 的初等因子组，设为 $(\lambda-\lambda_1)^{m_1}$、$(\lambda-\lambda_2)^{m_2}$、⋯、$(\lambda-\lambda_s)^{m_s}$。

（2） 写出各 Jordan 块矩阵（一个初等因子对应一个 Jordan 块矩阵）

$$(\lambda-\lambda_i)^{m_i}\to J_i(\lambda_i)=\begin{bmatrix}\lambda_i & 1 & & 0\\ & \lambda_i & \ddots & \\ & & \ddots & 1\\ 0 & & & \lambda_i\end{bmatrix}_{m_i\times m_i}$$

（3） 合成 Jordan 矩阵： $J=\begin{bmatrix}J_1 & & & 0\\ & J_2 & & \\ & & \ddots & \\ 0 & & & J_s\end{bmatrix}$

例：求矩阵 $A=\begin{bmatrix}2 & 1 & 0 & -1 & -1 & 0\\ 1 & 2 & 0 & 0 & -1 & 1\\ -1 & -1 & 4 & 1 & 0 & -1\\ 0 & -1 & 0 & 3 & 1 & 0\\ 0 & 0 & 0 & 0 & 4 & 0\\ 1 & 0 & 0 & 0 & -1 & 3\end{bmatrix}$ 的 Jordan 标准形。

［解］写出特征矩阵

$$(\lambda I-A)=\begin{bmatrix}\lambda-2 & -1 & 0 & 1 & 1 & 0\\ -1 & \lambda-2 & 0 & 0 & 1 & -1\\ 1 & 1 & \lambda-4 & -1 & 0 & 1\\ 0 & 1 & 0 & \lambda-3 & -1 & 0\\ 0 & 0 & 0 & 0 & \lambda-4 & 0\\ -1 & 0 & 0 & 0 & 1 & \lambda-3\end{bmatrix}$$

第 1～4 行与第 1、2、4、5 列交叉的元素形成的四阶子式为

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

第 1、2、3、5 行与 1、3、4、5 列交叉的元素形成的四阶子式为

$$\begin{vmatrix}\lambda-2 & 0 & 1 & 1\\ -1 & 0 & 0 & 1\\ 1 & \lambda-4 & -1 & 0\\ 0 & 0 & 0 & \lambda-4\end{vmatrix}=-(\lambda-4)^2$$

这两个子式的公因式为 1，故 $D_4(\lambda)=1\Rightarrow D_1(\lambda)=D_2(\lambda)=D_3(\lambda)=1$

第 1～5 行与第 1、2、3、5、6 列交叉的元素形成的五阶子式为

$$\begin{vmatrix}\lambda-2 & -1 & 0 & 1 & 0\\ -1 & \lambda-2 & 0 & 1 & -1\\ 1 & 1 & \lambda-4 & 0 & 1\\ 0 & 1 & 0 & -1 & 0\\ 0 & 0 & 0 & \lambda-4 & 0\end{vmatrix}=-(\lambda-2)(\lambda-4)^2$$

第 1、2、3、4、6 行与第 1、2、4、5、6 列交叉的元素形成的五阶子式为

$$\begin{vmatrix}\lambda-2 & -1 & 1 & 1 & 0\\ -1 & \lambda-2 & 0 & 1 & -1\\ 1 & 1 & -1 & 0 & 1\\ 0 & 1 & \lambda-3 & -1 & 0\\ -1 & 0 & 0 & 1 & \lambda-3\end{vmatrix}=4(\lambda-2)^3$$

其它五阶子式均含 $(\lambda-2)$ 因式，故 $D_5(\lambda)=(\lambda-2)$

特征值行列式为 $D_6(\lambda)=(\lambda-2)^3(\lambda-4)^3$，从而有

$$d_1(\lambda)=d_2(\lambda)=d_3(\lambda)=d_4(\lambda)=1, \quad d_5(\lambda)=(\lambda-2),$$

$$d_6(\lambda)=(\lambda-2)^2(\lambda-4)^3$$

- 初等因子组为

$$(\lambda-2)\ , \quad (\lambda-2)^2\ , \quad (\lambda-4)^3$$

- 相应的 Jordan 块为

$$[2]\ , \quad \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}, \quad \begin{bmatrix} 4 & 1 & 0 \\ 0 & 4 & 1 \\ 0 & 0 & 4 \end{bmatrix}$$

- Jordan 标准形为

$$\begin{bmatrix} 2 & & & & & 0 \\ & 2 & 1 & & & \\ & & 2 & & & \\ & & & 4 & 1 & \\ & & & & 4 & 1 \\ 0 & & & & & 4 \end{bmatrix}$$

作业：P106　1(1)(2),　2,　4,　5,　10
　　　P79　19（1）（3）

第六讲 Jordon 标准形的变换与应用

一、 Jordon 标准形变换矩阵的求法

$$P^{-1}AP = J \quad \to \quad AP = PJ$$

$1^{o}$ 将 $P$ 按 $J$ 的结构写成列块的形式

$$\begin{array}{ccccc} P = [ & P_1 & P_2 & \cdots & P_r ] \\ & \uparrow & \uparrow & & \uparrow \\ & m_1\text{列} & m_2\text{列} & & m_r\text{列} \end{array}$$

$$\to \quad A\begin{bmatrix} P_1 & P_2 & \cdots & P_r \end{bmatrix} = \begin{bmatrix} P_1 & P_2 & \cdots & P_r \end{bmatrix}\begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_r \end{bmatrix}$$

$$\to AP_i = P_i J_i \quad (i = 1,2\cdots,r)$$

$2^{o}$ 求解 $r$ 个矩阵方程 $AP_i = P_i J_i \quad (i = 1,2\cdots,r)$

$3^{o}$ 将 $r$ 个 $P_i$ 合成变换矩阵 $P = \begin{bmatrix} P_1 & P_2 & \cdots & P_r \end{bmatrix}$

★ 关于方程 $AP_i = P_i J_i$ 的求解

$$P_i = \begin{bmatrix} P_{i1} & P_{i2} & \cdots & P_{im} \end{bmatrix}$$

$$A\begin{bmatrix} P_{i1} & P_{i2} & \cdots & P_{im} \end{bmatrix} = \begin{bmatrix} P_{i1} & P_{i2} & \cdots & P_{im} \end{bmatrix}\begin{bmatrix} \lambda_i & 1 & & 0 \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{bmatrix}$$

$$AP_{i1} = \lambda_i P_{i1} \quad \to \quad (A - \lambda_i I)P_{i1} = 0$$

$$AP_{i2} = P_{i1} + \lambda_i P_{i2} \quad \to \quad (A - \lambda_i I)P_{i2} = P_{i1} \quad \to (A - \lambda_i I)^2 P_{i2} = 0$$

$$\vdots$$

$$AP_{im_i} = P_{im_i - 1} + \lambda_i P_{im_i} \to (A - \lambda_i I)P_{im_i} = P_{im_i - 1} \to (A - \lambda_i I)^{m_i} P_{im_i} = 0$$

两种具体做法:（ⅰ）按照 $P_{i1} \to P_{i2} \to \cdots \to P_{im}$ 的顺序求解，即先求出特征向量 $P_{i1}$，然后由后续方程求出 $P_{i2}$、$P_{i3}$、$\cdots$；（ⅱ）先求 $(A - \lambda_i I)^{m_i} P_{im_i} = 0$ 的解向量 $P_{im_i}$，然后直接得到 $P_{im_i - 1} \to P_{im_i - 2} \to \cdots \to P_{i1}$。

前一做法由于$(A-\lambda_i I)$为奇异矩阵，每一步均存在多解及无解问题，故各步之间不能完全独立，前一步尚需依赖后一步、再后一步、…，直至最后一步才能完全确定一些待定系数；而后一做法仅出现一次求解方程，其余为直接赋值，无上述问题。但又可能导致低阶$P_{im_i}$出现零向量的问题。

由于 $P_{i1}=(A-\lambda_i I)^{m_i-1}P_{im_i}$

$P_{i2}=(A-\lambda_i I)^{m_i-2}P_{im_i}$

$\vdots$

$P_{im_{i-1}}=(A-\lambda_i I)P_{im_i}$

故$P_{im_i}$应满足：$(A-\lambda_i I)^{m_i}P_{im_i}=0$但$(A-\lambda_i I)^{m_i-1}P_{im_i}\neq 0$

同一特征值可能出现在不同的 Jordan 块中，对于这种情况，按各 Jordan 块阶数高低一次进行处理，高阶先处理，低阶后处理，同阶同时处理。

（1） 最高阶（没有属于同一特征值的 Jordan 块同阶）可按下述方法求出$P_{im_i}$，即使$(A-\lambda_i I)^{m_i}x=0$

但$(A-\lambda_i I)^{m_i-1}x\neq 0$的$x$作为$P_{im_i}$。

然后由方程$P_{i(j-1)}=(A-\lambda_i I)P_j$依次求出$P_{im_{i-1}},P_{im_{i-2}},\cdots,$直至$P_{ij}$且$j$等于下一个属于同一特征值的 Jordan 块的阶数。

（2） 对于上述新 Jordan 块，它的$P_{im_i}$不仅要考虑到满足

$$(A-\lambda_i I)^{m_i}x=0\text{但}(A-\lambda_i I)^{m_i-1}x\neq 0，$$

而且还应与前述$P_{ij}$线性无关。

（3） 其它属于同一特征值的 Jordan 块处理时，按照（2）的原则处理即可。

（4） 出现多个属于同一特征值的 Jordan 块同阶时，还应考虑线性无关问题。

例：求$A=\begin{bmatrix}2&1&0&-1&-1&0\\1&2&0&0&-1&1\\-1&-1&4&1&0&-1\\0&-1&0&3&1&0\\0&0&0&0&4&0\\1&0&0&0&-1&3\end{bmatrix}$的 Jordan 标准形及其变换矩阵。

[解]：上一讲已求出其 Jordan 标准形，也可按如下方法求得。

（$\boldsymbol{\lambda I - A}$）可采用初等变换化为$\begin{bmatrix} 1 & & & & & 0 \\ & 1 & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & (\lambda-2) & \\ 0 & & & & & (\lambda-2)^2(\lambda-4)^3 \end{bmatrix}$

按此得出 Jordon 标准形

$$\begin{bmatrix} 2 & & & & & 0 \\ & 2 & 1 & & & \\ & & 2 & & & \\ & & & 4 & 1 & \\ & & & & 4 & 1 \\ 0 & & & & & 4 \end{bmatrix}$$

同时可见 $\det(\lambda I - A) = (\lambda-2)^3(\lambda-4)^3$，即 $\lambda = 2$ 与 $\lambda = 4$ 均为三重特征值.

下面求变换矩阵 $P$

（1）$\lambda_3 = 4$ 的 Jordon 矩阵仅有一块，$m_3 = 3$

$$P_3 = \begin{bmatrix} P_{31} & P_{32} & P_{33} \end{bmatrix}$$

先求 $P_{33}$，$P_{33}$ 应满足 $(A-4I)^3 P_{33} = 0 \quad (A-4I)^2 P_{33} \neq 0$

$$(A-4I) = \begin{bmatrix} -2 & 1 & 0 & -1 & -1 & 0 \\ 1 & -2 & 0 & 0 & -1 & 1 \\ -1 & -1 & 0 & 1 & 0 & -1 \\ 0 & -1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & -1 \end{bmatrix} \qquad (A-4I)^2 = \begin{bmatrix} 5 & -3 & 0 & 3 & 0 & 1 \\ -3 & 5 & 0 & -1 & 0 & -3 \\ 0 & 0 & 0 & 0 & 4 & 0 \\ -1 & 3 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -3 & 1 & 0 & -1 & 0 & 1 \end{bmatrix}$$

$$(A-4I)^3=\begin{bmatrix} -12 & 8 & 0 & -8 & 0 & -4 \\ 8 & -12 & 0 & 4 & 0 & 8 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & -8 & 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 8 & -4 & 0 & 4 & 0 & 0 \end{bmatrix}$$

求 $(A-4I)^3x=0$ $(A-4I)^2x=0$ 设 $x=\begin{bmatrix}\xi_1 & \xi_2 & \xi_3 & \xi_4 & \xi_5 & \xi_6\end{bmatrix}^{\mathrm{T}}$

$(A-4I)^3x=0$ 其通解为 $\xi_1\begin{bmatrix}1\\1\\0\\-1\\0\\1\end{bmatrix}+\xi_3\begin{bmatrix}0\\0\\1\\0\\0\\0\end{bmatrix}+\xi_5\begin{bmatrix}0\\0\\0\\0\\1\\0\end{bmatrix}$

$(A-4I)^2x=0$ 其通解为 $\xi_1\begin{bmatrix}1\\1\\0\\-1\\0\\1\end{bmatrix}+\xi_3\begin{bmatrix}0\\0\\1\\0\\0\\0\end{bmatrix}$

可取 $P_{33}=\begin{bmatrix}0 & 0 & 0 & 0 & 1 & 0\end{bmatrix}^{\mathrm{T}}$

$$P_{32}=(A-4I)P_{33}=\begin{bmatrix}-1 & -1 & 0 & 1 & 0 & -1\end{bmatrix}^{\mathrm{T}}$$

$$P_{31}=(A-4I)P_{32}=\begin{bmatrix}0 & 0 & 4 & 0 & 0 & 0\end{bmatrix}^{\mathrm{T}}$$

（2）对 $\lambda_{1,2}=2$ 存在两个 Jordan 块，$J_1=\begin{bmatrix}2\end{bmatrix}$ ，$J_2=\begin{bmatrix}2 & 1\\0 & 2\end{bmatrix}$,

分别对应 $P_1$，$P_2=\begin{bmatrix}P_{21} & P_{22}\end{bmatrix}$

从 $P_{22}$ 入手：$(A-2I)^2P_{22}=0$，$(A-2I)P_{22}\neq 0$

$$(A-2I)=\begin{bmatrix}0&1&0&-1&-1&0\\1&0&0&0&-1&1\\-1&-1&2&1&0&-1\\0&-1&0&1&1&0\\0&0&0&0&2&0\\1&0&0&0&-1&1\end{bmatrix}\quad (A-2I)^2=\begin{bmatrix}1&1&0&-1&-4&1\\1&1&0&-1&-4&1\\-4&-4&4&4&4&-4\\-1&-1&0&1&4&-1\\0&0&0&0&4&0\\1&1&0&-1&-4&1\end{bmatrix}$$

$$(A-2I)^2x=0$$

$$\to\quad x=\begin{bmatrix}\xi_1&\xi_2&\xi_3&\xi_4&\xi_5&\xi_6\end{bmatrix}^{\mathrm{T}}=\begin{bmatrix}\xi_1&\xi_2&0&\xi_1+\xi_2+\xi_6&0&\xi_6\end{bmatrix}^{\mathrm{T}}$$

$$(A-2I)x=0$$

$$\to x=\begin{bmatrix}\xi_1&\xi_2&\xi_3&\xi_4&\xi_5&\xi_6\end{bmatrix}^{\mathrm{T}}=\begin{bmatrix}\xi_1&\xi_2&0&\xi_2&0&-\xi_1\end{bmatrix}^{\mathrm{T}}$$

取 $P_{22}=\begin{bmatrix}-1&1&0&0&0&0\end{bmatrix}^{\mathrm{T}}$

$$P_{21}=(A-2I)P_{22}=\begin{bmatrix}-1&1&0&-1&0&-1\end{bmatrix}^{\mathrm{T}}$$

$P_1$：$(A-2I)P_1=0$ $\quad P_1$ 与 $P_{21}$ 应线性无关，可取 $P_1=\begin{bmatrix}1&0&0&0&0&-1\end{bmatrix}^{\mathrm{T}}$

（3）合成变换矩阵

$$P=\begin{bmatrix}1&1&-1&0&-1&0\\0&-1&1&0&-1&0\\0&0&0&4&0&0\\0&-1&0&0&1&0\\0&0&0&0&0&1\\-1&-1&0&0&-1&0\end{bmatrix}\qquad P^{-1}\text{存在}$$

可以验证：$P^{-1}AP=J$

二、 Jordan 标准形的幂及多项式

$$J^k=\begin{bmatrix}J_1^k&&&\\&J_2^k&&\\&&\ddots&\\&&&J_s^k\end{bmatrix},\ \text{即}\ J_i=\begin{bmatrix}\lambda_i&1&&0\\&\lambda_i&\ddots&\\&&\ddots&1\\0&&&\lambda_i\end{bmatrix},$$

$$J_i^2=\begin{bmatrix}\lambda_i^2 & 2\lambda_i & 1 & & & 0\\ & \lambda_i^2 & 2\lambda_i & 1 & & \\ & & \lambda_i^2 & 2\lambda_i & \ddots & \\ & & & \ddots & \ddots & 1\\ & & & & \lambda_i^2 & 2\lambda_i\\ 0 & & & & & \lambda_i^2\end{bmatrix}$$

$J_i^k$ 亦为类似的上三角形条带矩阵, 在与主对角线平行的斜线上各元素相等. 其中 $J_i^k$ 第一行的元素依次为

$$\begin{cases}\lambda_i^k \quad k\lambda_i^{k-1} \quad \dfrac{k!}{2!(k-2)!}\lambda_i^{k-2} \quad \ldots \quad C_k^t\lambda_i^t \quad \ldots \quad 1 \quad 0 \quad \ldots \quad 0 \qquad (m_i-k-1\text{个}0),(m_i>k)\\ \lambda_i^k \quad k\lambda_i^{k-1} \quad \dfrac{k!}{2!(k-2)!}\lambda_i^{k-2} \quad \ldots \quad C_k^t\lambda_i^t \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad C_k^{k-m_i+1}\lambda_i^{k-m_i+1},(m_i\le k)\end{cases}$$

设有多项式 $f(x)=\sum_{k=0}^{m}a_kx^k=a_0+a_1x+\ldots+a_mx^m$. 则

$$f(J)=\sum_{k=0}^{m}a_kJ^k=a_0I+a_1J+\cdots+a_mJ^m=\begin{bmatrix}f(J_1) & & & \\ & f(J_2) & & \\ & & \ddots & \\ & & & f(J_s)\end{bmatrix}$$

又 $$f(\lambda_i)=a_0+a_1\lambda_i+\text{L}+a_m\lambda_i^m=\sum_{k=0}^{m}a_k\lambda_i^k$$

$$f^{'}(\lambda_i)=a_1+2a_2\lambda_i+\cdots+ma_m\lambda_i^{m-1}=\sum_{k=1}^{m}ka_k\lambda_i^{k-1}$$

……

$$f^{(m)}(\lambda_i)=m!a_m$$

$$f^{(m+l)}(\lambda_i)=0,(l=1,2,\ldots)$$

故

$$f(J_i)=a_0\begin{bmatrix}1 & & & & & 0\\ & 1 & & & & \\ & & 1 & & & \\ & & & \ddots & & \\ & & & & 1 & \\ 0 & & & & & 1\end{bmatrix}+a_1\begin{bmatrix}\lambda & 1 & & & & 0\\ & \lambda & 1 & & & \\ & & \lambda & & & \\ & & & & 1 & \\ & & & & \lambda & 1\\ 0 & & & & & \lambda\end{bmatrix}+a_2\begin{bmatrix}\lambda_i^2 & 2\lambda_i & 1 & & & 0\\ & \lambda_i^2 & 2\lambda_i & 1 & & \\ & & \lambda_i^2 & \ddots & \ddots & \\ & & & \ddots & 2\lambda_i & 1\\ & & & & \lambda_i^2 & 2\lambda_i\\ 0 & & & & & \lambda_i^2\end{bmatrix}+\cdots$$

$$
=\begin{bmatrix}
f(\lambda_i) & f'(\lambda_i) & \frac{1}{2!}f^{(2)}(\lambda_i) & \cdots & \frac{1}{m!}f^{(m)}(\lambda_i) & \cdots & \frac{1}{(m_i-1)!}f^{(m_i-1)}(\lambda_i) \\
 & f(\lambda_i) & f'(\lambda_i) & \frac{1}{2!}f^{(2)}(\lambda_i) & & & \vdots \\
 & & f(\lambda_i) & f'(\lambda_i) & \ddots & & \vdots \\
 & & & f(\lambda_i) & \ddots & \ddots & \vdots \\
 & & & & \ddots & \ddots & \frac{1}{2!}f^{(2)}(\lambda_i) \\
 & & & & & \ddots & f'(\lambda_i) \\
 & & & & & & f(\lambda_i)
\end{bmatrix}
$$

这就是说，$f(J_i)$ 仍为上三角矩阵，在与主对角线平行的斜线上各元素均相等，而其第一行元素依次为

$$
f(\lambda_i) \quad f'(\lambda_i) \quad \frac{1}{2!}f^{(2)}(\lambda_i) \quad \cdots \quad \frac{1}{(m_i-1)!}f^{(m_i-1)}(\lambda_i)
$$
（无论 $m_i$ 是否大于 $m$）

若 $A=P^{-1}JP$，则 $f(A)=P^{-1}f(J)P$ 计算十分方便,无需再采用矩阵乘积.

作业:　　P107　　11

第七讲 矩阵级数与矩阵函数

一、 矩阵序列

1. 定义：设有矩阵序列$\left\{A^{(k)}\right\}$，其中$A^{(k)}=\left(a_{ij}^{(k)}\right)$，且当$k\to\infty$时$a_{ij}^{(k)}\to a_{ij}$，则称$\left\{A^{(k)}\right\}$收敛，并把$A=\left(a_{ij}\right)$叫做$\left\{A^{(k)}\right\}$的极限，或称$\left\{A^{(k)}\right\}$收敛于$A$，记为

$$\lim_{k\to\infty}A^{(k)}=A \quad 或 \quad A^{(k)}\underset{k\to\infty}{\to}A$$

不收敛的矩阵序列则称为发散的,其中又分为有界和无界的情况.

2. 收敛矩阵序列的性质：

设$\left\{A^{(k)}\right\}$,$\left\{B^{(k)}\right\}$分别收敛于$A$,$B$则

(1) $\alpha A^{(k)}+\beta B^{(k)}\underset{k\to\infty}{\to}\alpha A+\beta B$

(2) $A^{(k)}B^{(k)}\underset{k\to\infty}{\to}AB$

(3) $(A^{(k)})^{-1}\underset{k\to\infty}{\to}A^{-1}$，若$(A^{(k)})^{-1},A^{-1}$存在

(4) $PA^{(k)}Q\underset{k\to\infty}{\to}PAQ$

3 收敛矩阵：设$A$为方阵,且当$k\to\infty$时$A^k\to 0$，则称$A$为收敛矩阵.

[定理] 方阵$A$为收敛矩阵的充要条件是$A$的所有特征值的模值均小于1.

证明：对任何方阵$A$,均存在可逆矩阵$P$，使得

$$A=PJP^{-1}$$

其中$J$ 为$A$的 Jordan 标准形

$$J=\begin{bmatrix}J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_s\end{bmatrix},\quad J_i=\begin{bmatrix}\lambda_i & 1 & & 0\\ & \lambda_i & \ddots & \\ & & \ddots & 1\\ 0 & & & \lambda_i\end{bmatrix}$$

$$A^k=PJ^kP^{-1}=P\begin{bmatrix}J_1^k & & & \\ & J_2^k & & \\ & & \ddots & \\ & & & J_s^k\end{bmatrix}P^{-1}$$

$$J_i^k=\begin{bmatrix}\lambda_i^k & k\lambda_i^{k-1} & \cdots & \dfrac{k!}{(m_i-1)!(k-m_i+1)!}\lambda_i^{k-m_i+1}\\ & \ddots & \ddots & \ddots\\ & & \ddots & \ddots\\ & & & \end{bmatrix},\ \text{当}\,k>m_i$$

$A^k\to 0$ 就等价于 $J_i^k\to 0(i=1,2,...,s)$，等价于 $\lambda_i^k\to 0(i=1,2,...,s)$，而这只有 $|\lambda_i|<1$ 才可能也必能.

［得证］

二、 矩阵级数

1. 定义：矩阵序列 $\{A^{(k)}\}$ 的无穷和 $A^{(1)}+A^{(2)}+\cdots+A^{(k)}+\cdots$ 叫做矩阵级数，而 $S^{(N)}=\sum_{k=1}^{N}A^{(k)}$ 称为其部分和，若矩阵序列 $\{S^{(N)}\}$ 收敛，且有极限 $S$，则称该级数收敛，且有和 $S$．记为

$$\sum_{k=1}^{\infty}A^{(k)}=S$$

不收敛的级数必为发散的.

若矩阵级数 $\sum_{k=1}^{\infty}A^{(k)}$ 的所有元素 $\sum_{k=1}^{\infty}a_{ij}^{(k)}$ 均绝对收敛，则称该级数为绝对收敛.

2. 绝对收敛矩阵级数的性质

(1) 绝对收敛级数一定收敛，且任意调换它的项所得的级数仍收敛，并具有相同的和.

(2) $\sum_{k=1}^{\infty}A^{(k)}$ 绝对收敛，则 $\sum_{k=1}^{\infty}PA^{(k)}Q$ 也绝对收敛且等于 $P\sum_{k=1}^{\infty}A^{(k)}Q$

(3) $\sum_{k=1}^{\infty}A^{(k)}$，$\sum_{k=1}^{\infty}B^{(k)}$ 均绝对收敛，且和分别为 $S_1,S_2$ 则

$$\sum_{k=1}^{\infty}(\sum_{i=1}^{k}A^{(i)}B^{(k+1-i)})=S_1S_2$$

三、 方阵的幂级数

$A$ 为方阵，$\sum_{k=0}^{\infty}c_kA^k,(A^0=I)$ 称为 $A$ 的幂级数. $\sum_{k=0}^{\infty}A^k$ 称为 $A$ 的 Neumann 级数.

1. Neumann 级数收敛的充要条件

［定理］Neumann 级数收敛的充要条件是 $A$ 为收敛矩阵，且在收敛时其和为 $(I-A)^{-1}$.

证明：［必要性］

级数$\sum_{k=0}^{\infty} A^k$收敛，其元素为

$$\delta_{ij} + (A)_{ij} + (A^2)_{ij} + (A^3)_{ij} + \cdots$$

显然也是收敛的. 作为数项级数，其通项趋于零是级数收敛的必要条件. 故

$$(A^k)_{ij} \underset{k\to\infty}{\longrightarrow} 0，即 A^k \underset{k\to\infty}{\longrightarrow} 0$$

也就是说$A$为收敛矩阵.

[充分性]:

$A$为收敛矩阵，则其特征值的模值均小于 1. 设$A$的特征值为$\lambda$，$(I-A)$的特征值为$\mu$. 则由

$$\det(\mu I-(I-A)) = \det((\mu-1)I+A) = (-1)^n \det((1-\mu)I-A)$$

可见$1-\mu=\lambda \to \mu = 1-\lambda$

故$0<|\mu|<2 \to \mu \neq 0$，$(I-A)$的行列式不为零, $(I-A)^{-1}$存在.

而$(I+A+A^2+...+A^k)(I-A) = I-A^{k+1}$

右乘$(I-A)^{-1}$得

$$I+A+A^2+...+A^k = (I-A^{k+1})(I-A)^{-1}$$

当$k\to\infty$时，$A^{k+1}\to 0$，故$A^{k+1}(I-A)^{-1}\to 0$. 所以

$$\sum_{i=0}^{\infty} A^i = \lim_{k\to\infty}\sum_{i=0}^{k} A^i = (I-A)^{-1}$$

即 Neumann 级数收敛于$(I-A)^{-1}$.

2. 收敛圆

[定理] 若矩阵$A$的特征值全部落在幂级数$\varphi(z)=\sum_{k=0}^{\infty} c_k z^k$的收敛圆内，则矩阵幂级数$\varphi(A)=\sum_{k=0}^{\infty} c_k A^k, (A^0=I)$是绝对收敛的. 反之，若$A$存在落在$\varphi(z)$的收敛圆外的特征值，则$\varphi(A)$是发散的.

证明略.

[推论] 若幂级数在整个复平面上收敛，则对任何的方阵$A$，$\varphi(A)$均收敛.

四、 矩阵函数

如：$e^A$，$\sin A$，$\cos A$

以矩阵为自变量的“函数”（实际上是“函矩阵”）

我们知道，$e^z=1+z+\frac{1}{2!}z^2+\cdots=\sum_{n=0}^{\infty}\frac{1}{n!}z^n$

$$\sin(z)=\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)!}z^{2n+1}$$

$$\cos(z)=\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n)!}z^{2n}$$

均为整个复平面上收敛的级数，故对任何的方阵 $A$

$$e^A=\sum_{n=0}^{\infty}\frac{1}{n!}A^n$$

$$\sin(A)=\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)!}A^{2n+1}$$

$$\cos(A)=\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n)!}A^{2n}$$

均绝对收敛．三者分别称为矩阵指数函数、矩阵正弦函数、矩阵余弦函数。

［性质］

$$e^{jA}=\cos A+j\sin A$$

$$\cos A=\frac{1}{2}(e^{jA}+e^{-jA})$$

$$\sin A=\frac{1}{2j}(e^{jA}-e^{-jA})$$

$$\cos(-A)=\cos A$$

$$\sin(-A)=-\sin A$$

$$\left.\begin{aligned}\cos(A\pm B)=\cos A\cos B\mp\sin A\sin B\\ \sin(A\pm B)=\sin A\cos B\pm\cos A\sin B\end{aligned}\right\}\leftarrow AB=BA$$

但是一般来说 $e^Ae^B$，$e^Be^A$，$e^{A+B}$ 三者互不相等．例如

$$A=\begin{bmatrix}1&1\\0&0\end{bmatrix},\quad B=\begin{bmatrix}1&-1\\0&0\end{bmatrix}，则$$

$$A^2=\begin{bmatrix}1&1\\0&0\end{bmatrix}=A^3=A^4=\cdots$$

$$B^2=\begin{bmatrix}1&-1\\0&0\end{bmatrix}=B^3=B^4=\cdots$$

$$e^A=I+(\sum_{n=1}^{\infty}\frac{1}{n!})A=I+(e-1)A=\begin{bmatrix}e&e-1\\0&1\end{bmatrix}$$

$$e^B=I+(\sum_{n=1}^{\infty}\frac{1}{n!})B=I+(e-1)B=\begin{bmatrix}e&1-e\\0&1\end{bmatrix}$$

可见$e^Ae^B\neq e^Be^A$

$$A+B=\begin{bmatrix}2&0\\0&0\end{bmatrix},\ (A+B)^2=2\begin{bmatrix}2&0\\0&0\end{bmatrix}=2(A+B),\ (A+B)^3=2^2(A+B),\cdots$$

$$e^{A+B}=I+(\sum_{n=1}^{\infty}\frac{1}{n!}2^{n-1})(A+B)=I+\frac{1}{2}(e^2-1)(A+B)=\begin{bmatrix}e^2&0\\0&1\end{bmatrix}$$

所以，$e^{A+B}\neq e^Ae^B$ ，$e^{A+B}\neq e^Be^A$

[定理] 若$AB=BA$，则$e^{A+B}=e^Ae^B=e^Be^A$

证明:

$$\begin{aligned}e^Ae^B&=(I+A+\frac{1}{2!}A^2+...)(I+B+\frac{1}{2!}B^2+...)\\&=I+(A+B)+\frac{1}{2!}(A^2+2AB+B^2)+\frac{1}{3!}(A^3+3A^2B+3AB^2+B^3)+...\\&=I+(A+B)+\frac{1}{2!}(A+B)^2+\frac{1}{3!}(A+B)^3+...=e^{A+B}\end{aligned}$$

$$(A+B)^2=(A+B)(A+B)=A^2+AB+BA+B^2=A^2+2AB+B^2$$

$$(A+B)^3=\cdots=A^3+3A^2B+3AB^2+B^3$$

同理，有$e^Be^A=e^{A+B}$

[推论] $e^Ae^{-A}=e^{-A}e^A=e^0=I$，$(e^A)^{-1}=e^{-A},(e^A)^m=e^{mA}$，$e^A$总存在逆阵。

五、 矩阵函数的初步计算

1. Hamilton-Cayley 定理

$n$阶矩阵$A$是其特征多项式的零点，即令

$$\varphi(\lambda)=\det(\lambda I-A)=\lambda^n+c_1\lambda^{n-1}+\cdots+c_{n-1}\lambda+c_n$$

则$\varphi(A)=A^n+c_1A^{n-1}+\cdots+c_{n-1}A+c_nI=0$

[证明]：设$A$的特征值为$\lambda_1,\lambda_2,\cdots,\lambda_n$，则$\varphi(\lambda)$又可写成

$$\varphi(\lambda)=(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n)$$

由 Schur 引理知，存在酉矩阵$U$，使得

$$U^{-1}AU=\begin{bmatrix}\lambda_1 & & & *\\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n\end{bmatrix}$$

而$U^{-1}\varphi(A)U=\varphi(U^{-1}AU)=(U^{-1}AU-\lambda_1 I)(U^{-1}AU-\lambda_2 I)\cdots(U^{-1}AU-\lambda_n I)$

$$=\begin{bmatrix}0 & & & & *\\ & \lambda_2-\lambda_1 & & & \\ & & \lambda_3-\lambda_1 & & \\ & & & \ddots & \\ 0 & & & & \lambda_n-\lambda_1\end{bmatrix}\begin{bmatrix}\lambda_1-\lambda_2 & & & & *\\ & 0 & & & \\ & & \lambda_3-\lambda_2 & & \\ & & & \ddots & \\ 0 & & & & \lambda_n-\lambda_2\end{bmatrix}\cdots\begin{bmatrix}\lambda_1-\lambda_n & & & *\\ & \lambda_2-\lambda_n & & \\ & & \ddots & \\ & & & \lambda_{n-1}-\lambda_n & \\ 0 & & & & 0\end{bmatrix}$$

$$=\begin{bmatrix}0 & 0 & & \\ \vdots & \vdots & & \\ & & * & \\ \vdots & \vdots & & \\ 0 & 0 & & \end{bmatrix}\begin{bmatrix}\lambda_1-\lambda_3 & & & & \\ & \lambda_2-\lambda_3 & & * & \\ & & 0 & & \\ & & & \lambda_4-\lambda_3 & \\ 0 & & & & \ddots\end{bmatrix}\cdots\begin{bmatrix}\lambda_1-\lambda_n & & & & \\ & \lambda_2-\lambda_n & & * & \\ & & \ddots & & \\ & & & \lambda_{n-1}-\lambda_n & \\ 0 & & & & 0\end{bmatrix}$$

$$=\begin{bmatrix}0 & 0 & 0 & \\ 0 & 0 & 0 & \\ \vdots & \vdots & \vdots & * \\ & & & \\ 0 & 0 & 0 & \end{bmatrix}\begin{bmatrix}\lambda_1-\lambda_4 & & & & *\\ & \lambda_2-\lambda_4 & & & \\ & & \lambda_3-\lambda_4 & & \\ & & & 0 & \\ 0 & & & & \ddots\end{bmatrix}\cdots\begin{bmatrix}\lambda_1-\lambda_n & & & & *\\ & \lambda_2-\lambda_n & & & \\ & & \ddots & & \\ & & & \lambda_{n-1}-\lambda_n & \\ 0 & & & & 0\end{bmatrix}$$

$$=\begin{bmatrix}0 & 0 & \cdots & \cdots & 0\\ 0 & 0 & \cdots & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ & & & & \\ 0 & 0 & \cdots & \cdots & 0\end{bmatrix}=0$$

即$\varphi(A)=0$

2. 零化多项式

多项式 $f(z)$，若 $f(A)=0$，则称其为 $A$ 的零化多项式。

由以上定理可知，方阵 $A$ 的特征多项式为 $A$ 的零化多项式。

3. 矩阵指数函数、正弦函数、余弦函数的计算

例：已知四阶矩阵的特征值是 $\pi$ 、$-\pi$ 、 0、 0，求 $\sin A$ 、 $\cos A$ 、 $e^A$

解：$\varphi(\lambda)=(\lambda-\pi)(\lambda+\pi)(\lambda-0)(\lambda-0)=\lambda^4-\pi^2\lambda^2$

故 $\varphi(A)=A^4-\pi^2A^2=0 \to A^4=\pi^2A^2, A^5=\pi^2A^3, A^6=\pi^2A^4=\pi^4A^2,\cdots$

$$\begin{aligned}\sin(A)&=A+\sum_{n=1}^{\infty}\frac{(-1)^n}{(2n+1)!}A^{2n+1}=A+\sum_{n=1}^{\infty}\frac{(-1)^n}{(2n+1)!}\pi^{2(n-1)}A^3\\&=A+\frac{1}{\pi^3}\left(\sum_{n=1}^{\infty}\frac{(-1)^n}{(2n+1)!}\pi^{2n+1}\right)A^3\\&=A+\frac{1}{\pi^3}(\sin\pi-\pi)A^3=A-\pi^{-2}A^3\end{aligned}$$

$$\begin{aligned}\cos(A)&=I+\sum_{n=1}^{\infty}\frac{(-1)^n}{(2n)!}A^{2n}=I+\sum_{n=1}^{\infty}\frac{(-1)^n}{(2n)!}\pi^{2(n-1)}A^2\\&=I+\frac{1}{\pi^2}(\cos\pi-1)A^2=I-2\pi^{-2}A^2\end{aligned}$$

$$\begin{aligned}e^A&=\sum_{n=0}^{\infty}\frac{1}{n!}A^n=I+A+\sum_{n=1}^{\infty}\frac{1}{(2n)!}A^{2n}+\sum_{n=1}^{\infty}\frac{1}{(2n+1)!}A^{2n+1}\\&=I+A+\sum_{n=1}^{\infty}\frac{1}{(2n)!}\pi^{2(n-1)}A^2+\sum_{n=1}^{\infty}\frac{1}{(2n+1)!}\pi^{2(n-1)}A^3\\&=I+A+\frac{\cosh\pi-1}{\pi^2}A^2+\frac{\sinh\pi-\pi}{\pi^3}A^3\end{aligned}$$

作业　P163　 3，4，5

第八讲 矩阵函数的求法

一、利用 Jordan 标准形求矩阵函数。

对于矩阵的多项式，我们曾导出 $f(A)=Pf(J)P^{-1}$， $f$ ：多项式

$$f(J)=\begin{bmatrix} f(J_1) & & & & \\ & f(J_2) & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & f(J_s) \end{bmatrix}$$

$$f(J_i)=\begin{bmatrix} f(\lambda_i) & f'(\lambda_i) & \frac{1}{2!}f''(\lambda_i) & \cdots & \frac{1}{(m_i-1)!}f^{(m_i-1)}(\lambda_i) \\ & \ddots & \ddots & \ddots & \end{bmatrix}$$

实际上，以上结果不仅对矩阵的多项式成立，对矩阵的幂级数也成立。由此引出矩阵函数的另一种定义及计算方法。

1. 定义：设 $n$ 阶矩阵 $A$ 的 Jordan 标准形为 $J$

$$J=\begin{bmatrix} J_1 & & & & \\ & J_2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & J_s \end{bmatrix}，\quad J(\lambda_i)=\begin{bmatrix} \lambda_i & 1 & & & 0 \\ & \lambda_i & 1 & & \\ & & \lambda_i & \ddots & \\ & & & \ddots & 1 \\ 0 & & & & \lambda_i \end{bmatrix}$$

且有非奇异矩阵 $P$ 使得： $P^{-1}AP=J$

对于函数 $f(z)$，若下列函数

$$f(\lambda_i), f'(\lambda_i), \cdots, f^{(m_i-1)}(\lambda_i) \qquad (\lambda=1,2,\cdots,s)$$

均有意义，则称矩阵函数 $f(A)$ 有意义，且

$$f(A)=Pf(J)P^{-1}=P\begin{bmatrix} f(J_1) & & & & \\ & f(J_2) & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & f(J_s) \end{bmatrix}P^{-1}$$

$$f(J_i)=\begin{bmatrix} f(\lambda_i) & f'(\lambda_i) & \frac{1}{2!}f''(\lambda_i) & \cdots & \frac{1}{(m_i-1)!}f^{(m_i-1)}(\lambda_i) \\ & \ddots & \ddots & \ddots & \end{bmatrix}_{m_i\times m_i}$$

2. 矩阵函数的求法（步骤）：

$1^\circ$ 求出 $A$ 的 Jordan 标准形及变换矩阵 $P$， $P^{-1}AP=J$

$2^\circ$ 对于 $J$ 的各 Jordan 块 $J_i$ 求出 $f(J_i)$，即计算出

$$f(\lambda_i), f'(\lambda_i), \ldots\ldots, f^{(m_i-1)}(\lambda_i)$$

并按照顺序构成 $f(J_i)$，

$$f(J_i)=\begin{bmatrix} f(\lambda_i) & f'(\lambda_i) & \frac{1}{2!}f''(\lambda_i) & \cdots & \frac{1}{(m_i-1)!}f^{(m_i-1)}(\lambda_i) \\ & \ddots & \ddots & \ddots & \end{bmatrix}_{m_i\times m_i}$$

$3^\circ$ 合成 $f(J)=\begin{bmatrix} f(J_1) & & & & \\ & f(J_2) & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & f(J_s) \end{bmatrix}$

$4^\circ$ 矩阵乘积给出 $f(A)=Pf(J)P^{-1}$

需要说明的是，计算结果与 Jordan 标准形中 Jordan 块的顺序无关。

例 1 （教材 P70 例 1.27）． $A=\begin{bmatrix} 1 & 2 & 3 & 4 \\ & 1 & 2 & 3 \\ & & 1 & 2 \\ & & & 1 \end{bmatrix}$，求 $\sqrt{A}$

［解］ $1^{o}$ 求出 $J$ 及 $P$

$$J=\begin{bmatrix} 1 & 1 & 0 & 0 \\ & 1 & 1 & 0 \\ & & 1 & 1 \\ & & & 1 \end{bmatrix}, P=\begin{bmatrix} 8 & 4 & 0 & 0 \\ & 4 & -1 & 1 \\ & & 2 & -2 \\ & & & 1 \end{bmatrix}, P^{-1}=\frac{1}{16}\begin{bmatrix} 2 & -2 & -1 & 0 \\ & 4 & 2 & 0 \\ & & 8 & 16 \\ & & & 16 \end{bmatrix}$$

$2^{\circ}$ 求出 $f(\lambda_i), f'(\lambda_i), \ldots\ldots, f^{(m_i-1)}(\lambda_i)$ 并构成 $f(J_i)$：

$$\lambda_1 = 1, m_1 = 4, f(z) = \sqrt{z}$$

$$f(1) = 1,$$

$$f'(1) = \frac{1}{2} z^{-\frac{1}{2}} |_{z=1} = \frac{1}{2}, f''(1) = -\frac{1}{4} z^{-\frac{3}{2}} |_{z=1} = -\frac{1}{4}, f'''(1) = \frac{3}{8} z^{-\frac{5}{2}} |_{z=1} = \frac{3}{8}$$

$$f(J_1) = \begin{bmatrix} 16 & 8 & -2 & 1 \\ & 16 & 8 & -2 \\ & & 16 & 8 \\ & & & 16 \end{bmatrix} \frac{1}{16}$$

$3^{\circ}$ 合成 $f(J) = f(J_1)$

$4^{\circ}$ 求 $f(A) = Pf(J)P^{-1}$， $f(A) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ & 1 & 1 & 1 \\ & & 1 & 1 \\ & & & 1 \end{bmatrix}$

说明：

（1） $f(z) = \sqrt{z}$，在 $z = 0$ 不存在泰勒展开，如按原先的幂级数定义，则根本无从谈 $f(A)$ 的计算，可见新的定义延拓了原来的定义；

（2）$[f(A)]^2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ & 1 & 1 & 1 \\ & & 1 & 1 \\ & & & 1 \end{bmatrix}^2 = \begin{bmatrix} 1 & 2 & 3 & 4 \\ & 1 & 2 & 3 \\ & & 1 & 2 \\ & & & 1 \end{bmatrix} = A$，

可见这样的 $\sqrt{A}$ 确与 $A^2$ 构成反函数；

（3）矩阵函数的种类不仅是我们介绍的这种，如辛矩阵。以 $A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$ 为例，以我们这里的定义，$\sqrt{A} = \begin{bmatrix} \pm i & 0 \\ 0 & \pm i \end{bmatrix}$，但 $B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ 亦满足 $B^2 = A$，即 B 也可以看作某种 $\sqrt{A}$

二、利用零化多项式求解矩阵函数.

利用 Jordan 标准形求解矩阵函数的方法比较复杂，它需要求 $J$

和 $\boldsymbol{P}$ 。下面我们介绍根据零化多项式求解矩阵函数的一种方法。

定理：$\boldsymbol{n}$阶方阵 $\boldsymbol{A}$ 的最小多项式等于它的特征矩阵的第 $\boldsymbol{n}$ 个（也就是最后一个）不变因子 $\boldsymbol{d_n(\lambda)}$ 。（可参见张远达《线性代数原理》P215）

设 $\boldsymbol{n}$ 阶方阵 $\boldsymbol{A}$ 的不变因子反向依次为 $\boldsymbol{d_n(\lambda), d_{n-1}(\lambda), \cdots, d_1(\lambda)}$，由它们给出的初等因子分别为

$$(\lambda-\lambda_1)^{m_1},(\lambda-\lambda_2)^{m_2},\cdots,(\lambda-\lambda_r)^{m_r}\text{；}(\lambda-\lambda_{r+1})^{m_{r+1}},\cdots,(\lambda-\lambda_s)^{m_s}$$

$$\sum_{i=1}^{s} m_i = n$$

由于 $\boldsymbol{d_1(\lambda)\,|\,d_2(\lambda), d_2(\lambda)\,|\,d_3(\lambda), \cdots, d_{n-1}(\lambda)\,|\,d_n(\lambda)}$，故

$1^{\mathbf{o}}$ $\boldsymbol{\lambda_{r+1} \sim \lambda_s}$ 必定出现在 $\boldsymbol{\lambda_1 \sim \lambda_r}$ 中；

$2^{\mathbf{o}}$ 若 $\boldsymbol{\lambda_i(i>r)=\lambda_j(j\le r)}$ 则 $\boldsymbol{m_i \le m_j}$

根据上述定理，$\boldsymbol{A}$ 的最小多项式

$$\varphi_0(\lambda)=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_r)^{m_r}$$

即 $$(\lambda_1 I - A)^{m_1}(\lambda_2 I - A)^{m_2}\cdots(\lambda_r I - A)^{m_r} = O$$

令 $\boldsymbol{m=\sum_{i=1}^{r} m_i}$ ，则可见 $\boldsymbol{A^m}$ 可以由 $\boldsymbol{A^0=I, A, A^2, \cdots, A^{m-1}}$ 线性表示，从而 $\boldsymbol{A^{m+i}(i>0)}$ 亦可由 $\boldsymbol{A^0=I, A, A^2, \cdots, A^{m-1}}$ 线性表示。所以，矩阵函数 $\boldsymbol{f(A)}$ 若存在，也必定可由 $\boldsymbol{A^0 \sim A^{m-1}}$ 线性表示。

因此，我们定义一个系数待定的($m$-1)次多项式 $\boldsymbol{g(\lambda)=\sum_{i=0}^{m-1} c_i \lambda^i}$，根据以上论述，适当选择系数 $\boldsymbol{c_0 \sim c_{m-1}}$，就可以使 $f(A)=g(A)$

又，假设 $\boldsymbol{J}$ 、$\boldsymbol{P}$ 分别为 $\boldsymbol{A}$ 的 Jordan 标准形及相应变换矩阵：$\boldsymbol{A=PJP^{-1}}$

则 $\boldsymbol{f(A)=Pf(J)P^{-1}}$，$\boldsymbol{g(A)=Pg(J)P^{-1} \rightarrow f(J)=g(J) \rightarrow f(J_i)=g(J_i)}$

$$\Rightarrow f(\lambda_i)=g(\lambda_i), f'(\lambda_i)=g'(\lambda_i), \cdots, f^{(m_i-1)}(\lambda_i)=g^{(m_i-1)}(\lambda_i) \quad (i=1,2,\cdots,r)$$

由于$g(A)$为待定系数的多项式，上面就成为关于$c_0 \sim c_{m-1}$的线性方程组。且方程的个数为$m=\sum_{i=1}^{r} m_i$等于未知数个数，正好可以确定$c_0 \sim c_{m-1}$

由此给出根据最小多项式求矩阵函数的一般方法。

$1^o$ 求出最小多项式

$$\varphi_0(\lambda)=d_n(\lambda)=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_r)^{m_r},\sum_{i=1}^{r} m_i=m;$$

（或者特征多项式$\varphi(\lambda)=(\lambda-\lambda_1)^{n_1}(\lambda-\lambda_2)^{n_2}\cdots(\lambda-\lambda_r)^{n_r},\sum_{i=1}^{r} n_i=n$）

$2^o$ 形式上写出待定多项式

$$g(\lambda)=\sum_{i=0}^{m-1} c_i\lambda^i=c_0+c_1\lambda+c_2\lambda^2+\cdots+c_{m-1}\lambda^{m-1}$$

（或者$g(\lambda)=\sum_{i=0}^{n-1} c_i\lambda^i=c_0+c_1\lambda+c_2\lambda^2+\cdots+c_{n-1}\lambda^{n-1}$）

$3^o$ 求解关于$c_0 \sim c_{m-1}$的线性方程组

$$g^{(k)}(\lambda_i)=f^{(k)}(\lambda_i)\quad (k=1,2,\cdots,m_i;i=1,2,\cdots,r)$$

（或者$k=1,2,\cdots,n_i;i=1,2,\cdots,r$）

$4^o$ 求出$g(A)$，即可得$f(A)=g(A)$.

从推导的过程看，似乎不仅最小多项式可用于矩阵函数的计算，一般零化多项式也可以，其中以特征多项式最为方便。（但$k=1,2,\cdots,n_i$的根据仍需充分作证）

例 2、采用新方法计算$A=\begin{bmatrix}1&2&3&4\\ &1&2&3\\ & &1&2\\ & & &1\end{bmatrix}$的函数 $\sqrt{A}$。（$f(\lambda)=\sqrt{\lambda}$）

[解] $1^o$ $\varphi(\lambda)=\varphi_0(\lambda)=(\lambda-1)^4.\quad m_1=4=m=n,\lambda_1=1$；

$2^o$ $g(\lambda)=c_0+c_1\lambda+c_2\lambda^2+c_3\lambda^3$

$3^o$ 方程组为

$$g(1)=f(1)=1=c_0+c_1+c_2+c_3 \qquad g'(1)=f'(1)=\frac{1}{2}=c_1+2c_2+3c_3$$

$$g''(1)=f''(1)=-\frac{1}{4}=2c_2+6c_3 \qquad g'''(1)=f'''(1)=\frac{3}{8}=6c_3$$

$$\rightarrow \qquad c_3=\frac{1}{16}, c_2=-\frac{5}{16}, c_1=\frac{15}{16}, c_0=\frac{5}{16}$$

$4^{o}$ $g(A)=\frac{1}{16}(5I+15A-5A^2+A^3)$

$$A^2=\begin{bmatrix}1&4&10&20\\&1&4&10\\&&1&4\\&&&1\end{bmatrix},\quad A^3=\begin{bmatrix}1&6&21&56\\&1&6&21\\&&1&6\\&&&1\end{bmatrix}$$

$\therefore$

$$f(A)=\frac{1}{16}\left\{\begin{bmatrix}5&0&0&0\\&5&0&0\\&&5&0\\&&&5\end{bmatrix}+\begin{bmatrix}15&30&45&60\\&15&30&45\\&&15&30\\&&&15\end{bmatrix}-\begin{bmatrix}5&20&50&100\\&5&20&50\\&&5&20\\&&&5\end{bmatrix}+\begin{bmatrix}1&6&21&56\\&1&6&21\\&&1&6\\&&&1\end{bmatrix}\right\}$$

$$=\begin{bmatrix}1&1&1&1\\&1&1&1\\&&1&1\\&&&1\end{bmatrix}$$

与 Jordan 标准形方法完全一致。

作业： P163 6

第九讲 矩阵微分方程

一、矩阵的微分和积分

1. 矩阵导数定义：若矩阵 $A(t)=(a_{ij}(t))_{m\times n}$ 的每一个元素 $a_{ij}(t)$ 是变量 $t$ 的可微函数，则称 $A(t)$ 可微，其导数定义为

$$\frac{dA}{dt}=A'(t)=\left(\frac{da_{ij}}{dt}\right)_{m\times n}$$

由此出发，函数可以定义高阶导数，类似地，又可以定义偏导数。

2. 矩阵导数性质：若 $A(t)$, $B(t)$ 是两个可进行相应运算的可微矩阵，则

（1）$\frac{d}{dt}[A(t)\pm B(t)]=\frac{dA}{dt}\pm\frac{dB}{dt}$

（2）$\frac{d}{dt}[A(t)B(t)]=\frac{dA}{dt}B+A\frac{dB}{dt}$

（3）$\frac{d}{dt}[a(t)A(t)]=\frac{da}{dt}A+a\frac{dA}{dt}$

（4）$\frac{d}{dt}\left(e^{tA}\right)=Ae^{tA}=e^{tA}A \qquad \frac{d}{dt}\left(\cos\left(tA\right)\right)=-A\sin\left(tA\right)$

$\frac{d}{dt}\left(\sin\left(tA\right)\right)=A\cos\left(tA\right)$

（$A$ 与 $t$ 无关）

此处仅对 $\frac{d}{dt}(e^{tA})=Ae^{tA}=e^{tA}A$ 加以证明

证明：

$$\frac{d}{dt}(e^{tA})=\frac{d}{dt}(I+tA+\frac{1}{2!}t^2A^2+\frac{1}{3!}t^3A^3+\cdots)=A+tA^2+\frac{1}{2!}t^2A^3+\cdots$$

$$=A(I+tA+\frac{1}{2!}t^2A^2+\cdots)=Ae^{tA}$$

又 $=(I+tA+\frac{1}{2!}t^2A^2+\cdots)A=e^{tA}A$

3. 矩阵积分定义：若矩阵 $A(t)=(a_{ij}(t))_{m\times n}$ 的每个元素 $a_{ij}(t)$ 都是区间 $[t_0,t_1]$ 上的可积函数，则称 $A(t)$ 在区间 $[t_0,t_1]$ 上可积，并定义 $A(t)$ 在 $[t_0,t_1]$ 上的积分为

$$\int_{t_0}^{t_1} A(t)dt = \left(\int_{t_0}^{t_1} a_{ij}(t)dt\right)_{m\times n}$$

4. 矩阵积分性质

（1）$\int_{t_0}^{t_1}[A(t)\pm B(t)]dt = \int_{t_0}^{t_1} A(t)dt \pm \int_{t_0}^{t_1} B(t)dt$

（2）$\int_{t_0}^{t_1}[A(t)B]dt = \left(\int_{t_0}^{t_1} A(t)dt\right)B,\quad \int_{t_0}^{t_1}[AB(t)]dt = A\left(\int_{t_0}^{t_1} B(t)dt\right)$

（3）$\frac{d}{dt}\int_a^t A(t')dt' = A(t),\quad \int_a^b A'(t)dt = A(b)-A(a)$

二、 一阶线性齐次常系数常微分方程组

设有一阶线性齐次常系数常微分方程组

$$\begin{cases} \frac{dx_1}{dt} = a_{11}x_1(t) + a_{12}x_2(t) + \cdots + a_{1n}x_n(t) \\ \frac{dx_2}{dt} = a_{21}x_1(t) + a_{22}x_2(t) + \cdots + a_{2n}x_n(t) \\ \vdots \\ \frac{dx_n}{dt} = a_{n1}x_1(t) + a_{n2}x_2(t) + \cdots + a_{nn}x_n(t) \end{cases}$$

式中 $t$ 是自变量，$x_i = x_i(t)$ 是 $t$ 的一元函数 $(i = 1,2,\cdots,n)$，$a_{ij}(i,j = 1,2,\cdots,n)$ 是常系数。

令

$$x(t) = [x_1(t), x_2(t),\cdots, x_n(t)]^{\mathrm{T}},\quad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

则原方程组变成如下矩阵方程

$$\frac{dx}{dt} = Ax(t)$$

其解为

$$x(t) = e^{tA}x(0) = e^{tA}c \xrightarrow{\text{更一般的}} x(t) = e^{(t-t_0)A}x(t_0)$$

对该解求导，可以验证

$$\frac{dx(t)}{dt} = Ae^{tA}c = Ax(t) \text{ 且 } t = 0 \text{时，} x(t) = e^{0A}c = Ic = c = x(0)$$

表明 $x(t)$ 确为方程的解，积分常数亦正确。

例：求解微分方程组 $\begin{cases} \dfrac{dx_1}{dt} = x_2 \\ \dfrac{dx_2}{dt} = -x_1 \end{cases}$，　初始条件为 $\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}$

解：$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$，$f(A) = e^{tA} \quad \to \quad f(\lambda) = e^{t\lambda}$

$1^o$ 求出 $A$ 的特征多项式，$\varphi(\lambda) = \begin{vmatrix} \lambda & -1 \\ 1 & \lambda \end{vmatrix} = (\lambda^2 + 1) = (\lambda - j)(\lambda + j)$，　$j = \sqrt{-1}$

$$\lambda_1 = j, m_1 = 1; \lambda_2 = -j, m_2 = 1$$

$2^o$ 定义待定系数的多项式 $g(\lambda) = c_0 + c_1\lambda$

$3^o$ 解方程
$$g(\lambda_1) = f(\lambda_1) = e^{jt} = \cos t + j\sin t = c_0 + jc_1$$
$$g(\lambda_2) = f(\lambda_2) = e^{-jt} = \cos t - j\sin t = c_0 - jc_1$$

$$\begin{cases} c_0 = \cos t \\ c_1 = \sin t \end{cases}$$

$4^o$

$$g(A) = c_0 I + c_1 A = \begin{bmatrix} \cos t & 0 \\ 0 & \cos t \end{bmatrix} + \begin{bmatrix} 0 & \sin t \\ -\sin t & 0 \end{bmatrix} = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix}$$
$$= f(A) = e^{tA}$$

$$x(t) = e^{tA}x(0) = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix}\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} r_1\cos t + r_2\sin t \\ r_2\cos t - r_1\sin t \end{bmatrix} = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

三、　一阶线性非齐次常系数常微分方程组

$$
\begin{cases}
\dfrac{dx_1}{dt}=a_{11}x_1(t)+a_{12}x_2(t)+\cdots+a_{1n}x_n(t)+b_1(t)\\
\dfrac{dx_2}{dt}=a_{21}x_1(t)+a_{22}x_2(t)+\cdots+a_{2n}x_n(t)+b_2(t)\\
\vdots\\
\dfrac{dx_n}{dt}=a_{n1}x_1(t)+a_{n2}x_2(t)+\cdots+a_{nn}x_n(t)+b_n(t)
\end{cases}
$$

令

$$x(t)=[x_1(t),x_2(t),\cdots,x_n(t)]^{\mathrm{T}}$$

$$b(t)=[b_1(t),b_2(t),\cdots,b_n(t)]^{\mathrm{T}}$$

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

方程组化为矩阵方程

$$\frac{dx}{dt}=Ax+b$$

采用常数变易法求解之；齐次方程组的解为 $e^{tA}c$ ，可设非齐次方程组的解为 $e^{tA}c(t)$ ，代入方程，得：

$$\frac{dx}{dt}=\frac{d}{dt}(e^{tA})c(t)+e^{tA}\frac{dc}{dt}=A\cancel{x(t)}+e^{tA}\frac{dc}{dt}=A\cancel{x(t)}+b(t)\to\frac{\mathrm{dc}}{\mathrm{dt}}=\mathrm{e}^{-tA}\mathrm{b(t)}$$

$\therefore\ c(t)=\int_0^t e^{-sA}b(s)ds\ \leftarrow$ 由积分性质(3)可验证 $c(t)$ 是解。

加上初始条件，有

$$x(t)=e^{tA}\left[c(0)+\int_0^t e^{-sA}b(s)ds\right]$$

说明:高阶常微分方程常可以化为一阶常微分方程组来处理，

如： $$a\frac{d^2y}{dt^2}+b\frac{dy}{dt}+cy=f$$

令 $x_1=y,x_2=\dfrac{dy}{dt}$ ，则可得

$$
\begin{cases}
\dfrac{dx_1}{dt} = x_2 \\
\dfrac{dx_2}{dt} = \dfrac{1}{a}(f - cx_1 - bx_2) = -\dfrac{c}{a}x_1 - \dfrac{b}{a}x_2 + \dfrac{f}{a}
\end{cases}
$$

一般地，$n$ 阶常微分方程可以化为 $n$ 个一阶常微分方程组成的方程组。

作业：p170-171　　5、9

　　　p177　　3、4

第十讲 矩阵的三角分解

一、 Gauss 消元法的矩阵形式

$n$元线性方程组

$$\begin{cases} a_{11}\xi_1 + a_{12}\xi_2 + \cdots + a_{1n}\xi_n = b_1 \\ a_{21}\xi_1 + a_{22}\xi_2 + \cdots + a_{2n}\xi_n = b_2 \\ \vdots \\ a_{n1}\xi_1 + a_{n2}\xi_2 + \cdots + a_{nn}\xi_n = b_n \end{cases} \rightarrow Ax = b$$

$$\left( \begin{aligned} & A = (a_{ij}) \\ & x = [\xi_1, \xi_2, \cdots \xi_n]^{\mathrm{T}} \\ & b = [b_1, b_2 \cdots b_n]^{\mathrm{T}} \end{aligned} \right)$$

设$A^{(0)} = A = \left(a_{ij}\right)_{n\times n}$，设 $A$ 的 $k$ 阶顺序主子式为$\Delta_k$，若$\Delta_1 = a_{11}^{(0)} \neq 0$，

可以令$c_{i1} = \dfrac{a_{i1}^{(0)}}{a_{11}^{(0)}}$

并构造 Frobenius 矩阵

$$L_1 = \begin{bmatrix} 1 & & & 0 \\ c_{21} & 1 & & \\ \vdots & & \ddots & \\ c_{n1} & 0 & & 1 \end{bmatrix}_{n\times n} \rightarrow L_1^{-1} = \begin{bmatrix} 1 & & & 0 \\ -c_{21} & 1 & & \\ \vdots & & \ddots & \\ -c_{n1} & 0 & & 1 \end{bmatrix}$$

计算可得

$$A^{(1)} = L_1^{-1}A^{(0)} = \begin{bmatrix} a_{11}^{(0)} & a_{12}^{(0)} & \cdots & a_{1n}^{(0)} \\ & a_{22}^{(1)} & \cdots & a_{2n}^{(1)} \\ 0 & \vdots & \ddots & \vdots \\ & a_{n2}^{(1)} & \cdots & a_{nn}^{(1)} \end{bmatrix} \rightarrow A^{(0)} = L_1 A^{(1)}$$

第三种初等变换不改变行列式，故$\Delta_2 = \mathrm{a}_{11}^{(0)}\mathrm{a}_{22}^{(1)}$，若$\Delta_2 \neq 0$，则$\mathrm{a}_{22}^{(1)} \neq 0$，又可定义

$c_{i2} = \dfrac{a_{i2}^{(1)}}{a_{22}^{(1)}}(i = 3, 4, \cdots n)$，并构造 Frobenius 矩阵

$$L_2=\begin{bmatrix}1&&&&\\&1&&&\\&c_{32}&\ddots&&\\&\vdots&&&\\&c_{n2}&&&1\end{bmatrix}\quad\rightarrow\quad L_2^{-1}=\begin{bmatrix}1&&&&\\&1&&&\\&-c_{32}&\ddots&&\\&\vdots&&&\\&-c_{n2}&&&1\end{bmatrix}$$

$$A^{(2)}=L_2^{-1}A^{(1)}=\begin{bmatrix}a_{11}^{(0)}&a_{12}^{(0)}&a_{13}^{(0)}&\cdots&a_{1n}^{(0)}\\&a_{22}^{(1)}&a_{23}^{(1)}&\cdots&a_{2n}^{(1)}\\&&a_{33}^{(2)}&\cdots&a_{3n}^{(2)}\\&&\vdots&&\vdots\\&&a_{n3}^{(2)}&\cdots&a_{nn}^{(2)}\end{bmatrix}\quad\rightarrow\quad A^{(1)}=L_2A^{(2)}$$

依此类推，进行到第$(r-1)$步，则可得到

$$A^{(r-1)}=\begin{bmatrix}a_{11}^{(0)}&\cdots&a_{1r-1}^{(0)}&a_{1r}^{(0)}&\cdots&a_{1n}^{(0)}\\&\ddots&\vdots&\vdots&\cdots&\\&&a_{r-1r-1}^{(r-2)}&a_{r-1r}^{(r-2)}&\cdots&a_{r-1n}^{(r-2)}\\&&&a_{rr}^{(r-1)}&\cdots&a_{rn}^{(r-1)}\\&&&\vdots&&\vdots\\&&&a_{nr}^{(r-1)}&\cdots&a_{nn}^{(r-1)}\end{bmatrix}\quad(r=2,3,\cdots,n)$$

则 $A$ 的 $r$ 阶顺序主子式 $\Delta_r=a_{11}^{(0)}a_{22}^{(1)}\cdots a_{r-1r-1}^{(r-2)}a_{rr}^{r-1}$，若 $\Delta_r\neq 0$，则 $a_{rr}^{r-1}\neq 0$ 可定义

$c_{ir}=\dfrac{a_{ir}^{r-1}}{a_{rr}^{r-1}}$，并构造 Frobenius 矩阵

$$L_r=\begin{bmatrix}1&&&&\\&\ddots&&&\\&&1&&\\&&c_{r+1r}&1&\\&&\vdots&&\ddots&\\&&c_{nr}&&&1\end{bmatrix}\quad\rightarrow\quad L_r^{-1}=\begin{bmatrix}1&&&&\\&\ddots&&&\\&&1&&\\&&-c_{r+1r}&1&\\&&\vdots&&\ddots&\\&&-c_{nr}&&&1\end{bmatrix}$$

$$A^{(r)}=L_r^{-1}A^{r-1}=\begin{bmatrix} a_{11}^{(0)} & \cdots & a_{1r}^{(0)} & a_{1r+1}^{(0)} & \cdots & a_{1n}^{(0)} \\ & \ddots & \vdots & \vdots & \cdots & \\ & & a_{rr}^{(r-1)} & a_{rr+1}^{(r-1)} & \cdots & a_{rn}^{(r-1)} \\ & & & a_{r+1r+1}^{(r)} & \cdots & a_{r+1n}^{(r)} \\ & & & \vdots & & \vdots \\ & & & a_{nr+1}^{(r)} & \cdots & a_{nn}^{(r)} \end{bmatrix} \to A^{(r-1)}=L_rA^{(r)}$$

$$(r=2,3,\cdots,n-1)$$

直到第（$n$—1）步，得到

$$A^{(n-1)}=\begin{bmatrix} a_{11}^{(0)} & a_{12}^{(0)} & \cdots & a_{1n}^{(0)} \\ & a_{22}^{(1)} & \cdots & a_{2n}^{(1)} \\ & & \ddots & \vdots \\ & & & a_{nn}^{(n-1)} \end{bmatrix}$$ 则完成了消元的过程

而消元法能进行下去的条件是$\Delta_r \neq 0 \quad (r=1,2,\cdots,n-1)$

二、 $LU$分解与$LDU$分解

$$A=A^{(0)}=L_1A^{(1)}=L_1L_2A^{(2)}=\cdots=L_1L_2L_3\cdots L_{n-1}A^{(n-1)}$$

容易求出

$$L=L_1L_2\cdots L_{n-1}=\begin{bmatrix} 1 & & & & \\ c_{21} & 1 & & & \\ \vdots & & \ddots & & \\ c_{n-11} & c_{n-12} & & 1 & \\ c_{n1} & c_{n2} & & c_{nn-1} & 1 \end{bmatrix}$$ 为下三角矩阵

令$U=A^{(n-1)}$为上三角矩阵，则

$A=LU$ （L: lower U: upper L: left R: right）

以上将 $A$ 分解成一个下三角矩阵与上三角矩阵的乘积，就称为 $LU$ 分解或 $LR$ 分解。

$LU$ 分解不唯一，显然，令 $D$ 为对角元素不为零的 $n$ 阶对角阵，则

$$A=LU=LDD^{-1}U=\hat{L}\hat{U}$$

可以采用如下的方法将分解完全确定，即要求

（1） $L$ 为单位下三角矩阵 或

（2） $U$为单位上三角矩阵 或

（3） 将 $A$ 分解为 $LDU$，其中 $L$，$U$ 分别为单位下三角，单位上三角矩阵，$D$ 为对角阵

$$D=\mathrm{diag}[d_1,d_2,\cdots,d_n]，而 d_k=\Delta_k/\Delta_{k-1}$$

$(k=1,2,\cdots,n)$ $\Delta_0=1$。

$n$ 阶非奇异矩阵 $A$ 有三角分解 $LU$ 或 $LDU$ 的充要条件是 $A$ 的顺序主子式 $\Delta_r \neq 0$ $(r=1,2,\cdots,n)$

$n$ 个顺序主子式全不为零的条件实际上是比较严格的，特别是在数值计算中，$a_{kk}^{(k-1)}$ 很小时可能会带来大的计算误差。因此，有必要采取选主元的消元方法，这可以是列主元（在 $a_{kk}^{(k-1)},a_{k+1\ k}^{(k-1)},\cdots,a_{nk}^{(k-1)}$ 中选取模最大者作为新的 $a_{kk}^{(k-1)}$）、行主元（在 $a_{kk}^{(k-1)},a_{kk+1}^{(k-1)},\cdots,a_{kn}^{(k-1)}$ 中选取模最大者作为新的 $a_{kk}^{(k-1)}$）全主元（在所有 $a_{ij}^{(k-1)}$（$k \le i,j \le n$）中选模最大者作为新的 $a_{kk}^{(k-1)}$）。之所以这样做，其理论基础在于对于任何可逆矩阵 $A$，存在置换矩阵 $P$ 使得 $PA$ 的所有顺序主子式全不为零。

列主元素法：在矩阵的某列中选取模值最大者作为新的对角元素，选取范围为对角线元素以下的各元素。比如第一步：找第一个未知数前的系数 $|a_{i1}|$ 最大的一个，将其所在的方程作为第一个方程，即交换矩阵的两行，自由项也相应变换；第二步变换时，找 $|a_{i2}|(i \ge 2)$ 中最大的一个，然后按照第一步的方法继续。

行主元素法：在矩阵的某行中选取模值最大者作为新的对角元素，选取范围为对角线元素以后的各元素，需要记住未知数变换的顺序，最后再还原回去。因此需要更多的存储空间，不如列主元素法方便。

全主元素法：若某列元素均较小或某行元素均较小时，可在各行各列中选取模值最大者作为对角元素。与以上两种方法相比，其计算稳定性更好，精度更高，计算量增大。

$$Ax=b$$

$$A=LU$$

$$\begin{cases} Ly=b \\ Ux=y \end{cases} \leftarrow \text{两个三角形方程回代即可}$$

三、其他三角分解

1. 定义 设 $A$ 具有唯一的 $LDU$ 分解

（1） 若将 $D$，$U$ 结合起来得 $A=L\hat{U}$（$\hat{U}=DU$），则称为 $A$ 的 Doolittle 分解

（2） 若将 $L$，$D$ 结合起来得 $A=\hat{L}U$（$\hat{L}=LD$），则称为 $A$ 的 Crout 分解

2. 算法

Crout 分解，设

$$\hat{L}=\begin{bmatrix} l_{11} & & & \\ l_{21} & l_{22} & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix}，\ U=\begin{bmatrix} 1 & u_{12} & \cdots & u_{1n} \\ & 1 & \cdots & u_{2n} \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}$$

由 $A=\hat{L}U$ 乘出得

（1）$l_{i1}=a_{i1}$ (第1列)　（$i=1,2,3,\cdots n$）　($A,\hat{L}$第1列)

（2）$u_{1j}=a_{1j}/l_{11}$ (第1行)　（$j=2,3,\cdots n$）　($A,U$第1行)

（3）$l_{i2}=a_{i2}-l_{i1}u_{12}$ (第2列)　($i=2,3,\cdots n$)　($A,\hat{L}$第2列)

（4）$u_{2j}=\dfrac{1}{l_{22}}\left(a_{2j}-l_{21}u_{1j}\right)$　（$j=3,4,\cdots n$）　($A,U$第2行)

（5）一般地，对 $A$, $\hat{L}$ 的第 $k$ 列运算，有

$$l_{ik}=a_{ik}-\sum_{m=1}^{k-1}l_{im}u_{mk} \qquad (k=1,2,\cdots,n;i=k,k+1,\cdots,n)$$

（6） 对 $A$，$U$的第 $k$ 行运算，有

$$u_{kj}=\frac{1}{l_{kk}}(a_{kj}-\sum_{m=1}^{k-1}l_{km}u_{mj}) \qquad (k=1,2,\cdots n-1;j=k+1,k+2,\cdots n)$$

直至最后，得到的 $l_{ij},u_{ij}$ 恰可排成

$$\begin{bmatrix} l_{11} & u_{12} & u_{13} & \cdots & u_{1n} \\ l_{21} & l_{22} & u_{23} & \cdots & u_{2n} \\ l_{31} & l_{32} & l_{33} & \cdots & u_{3n} \\ \vdots & \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & l_{n3} & & l_{nn} \end{bmatrix}$$ 先算列后算行

3. 厄米正定矩阵的 Cholesky 分解

$$A=GG^{\mathrm{H}}$$

$$g_{ij}=\begin{cases}\left(a_{ii}-\sum_{k=1}^{i-1}\left|g_{ik}\right|^2\right)^{1/2} & i=j\\ \frac{1}{g_{jj}}(a_{ij}-\sum_{k=1}^{j-1}g_{ik}\overline{g_{ik}}) & i>j\\ 0 & i<j\end{cases}$$

理论上，Cholesky 具有中间量 $g_{ij}$ 可以控制（$\left|g_{ij}\right|\le\sqrt{\left|a_{ii}\right|}$）的好处，应较稳健，但实际计算中发现，对希尔伯特矩阵问题，不如全主元方法。

作业：p195　2、3

第十一讲 矩阵的QR分解

一．Givens矩阵与Givens变换

1． 定义：设实数c与s满足$c^2+s^2=1$，称

$$
T_{ij}=\begin{bmatrix}
1 & & & & & & & & & & \\
 & \ddots & & & & & & & & & \\
 & & 1 & & & & & & & & \\
 & & & c & & & & s & & & \\
 & & & & 1 & & & & & & \\
 & & & \vdots & & \ddots & & & & & \\
 & & & & & & 1 & & & & \\
 & & & -s & & & & c & & & \\
 & & & & & & & & 1 & & \\
 & & & \vdots & & & & & & \ddots & \\
 & & & & & & & & & & 1
\end{bmatrix}
\begin{array}{l}
\\ \\ \\ \leftarrow(i) \\ \\ \\ \\ \leftarrow(j) \\ \\ \\ \\
\end{array}
$$

（$i<j$）

为 Givens 矩阵（初等旋转矩阵），也记作$T_{ij}=T_{ij}(c,s)$。由 Givens 矩阵所确定的线性变换称为Givens变换（初等旋转变换）。

说明：（1）实数$c^2+s^2=1$，故存在$\theta$，使$c=\cos(\theta),s=\sin(\theta)$。

（2）$y=T_{ij}x$中$T_{ij}$确定了将向量 x 变成 y 的一种变换，正是 Givens 变换。二阶情况下，$y=\begin{bmatrix}\cos(\theta) & \sin(\theta)\\ -\sin(\theta) & \cos(\theta)\end{bmatrix}x$ 确定的正是平面直角坐标系中绕原点的一个旋转变换（旋转$\theta$度）。

（3）以上实Givens矩阵也可推广称为复初等旋转矩阵。

$$
\mathbf{U}_{ik}=\begin{array}{c}\left[\begin{array}{ccccccccccc}
1 &&&&&&&&&& \\
& \ddots &&&&&&&&& \\
&& 1 &&&&&&&& \\
&&& ce^{j\theta_1} &&&& se^{j\theta_2} &&& \\
&&&& 1 &&&&&& \\
&&& \vdots && \ddots &&&&& \\
&&&&&& 1 &&&& \\
&&& -se^{j\theta_3} &&&& ce^{j\theta_4} &&& \\
&&&&&&&& 1 && \\
&&& \vdots &&&&&& \ddots & \\
&&&&&&&&&& 1
\end{array}\right]\end{array}
\begin{array}{l} \\ \\ \\ \leftarrow(\mathbf{i}) \\ \\ \\ \\ \leftarrow(\mathbf{k}) \\ \\ \\ \\ \end{array}
$$

其中 c 与 s 仍为满足$\mathbf{c^2+s^2=1}$的实数，$\theta_1,\theta_2,\theta_3,\theta_4$为实角度。

显然，$\det(\mathbf{U}_{ik})=\mathbf{c}^2\mathbf{e}^{j(\theta_1+\theta_4)}+\mathbf{s}^2\mathbf{e}^{j(\theta_2+\theta_3)}$

当$\theta_1+\theta_4=\theta_2+\theta_3$时，$\det(\mathbf{U}_{ik})=\mathbf{e}^{j(\theta_1+\theta_4)}$

当$\theta_1+\theta_4=\theta_2+\theta_3=2n\pi$时，$\det(\mathbf{U}_{ik})=1$

2. 性质

（1）$\left[\mathbf{T}_{ij}(\mathbf{c},\mathbf{s})\right]^{-1}=\left[\mathbf{T}_{ij}(\mathbf{c},\mathbf{s})\right]^{T}=\mathbf{T}_{ij}(\mathbf{c},-\mathbf{s})$，$-\mathbf{s}=-\sin(\theta)=\sin(-\theta)$,旋转$\theta$度再反向旋转度$\theta$

$$\det\left[\mathbf{T}_{ij}\left(\mathbf{c},\mathbf{s}\right)\right]=1$$

（2）设$\mathbf{x}=\left[\xi_1\quad\xi_2\quad\cdots\quad\xi_n\right]^T$，$\mathbf{y}=\mathbf{T}_{ij}\mathbf{x}=\left[\eta_1\quad\eta_2\quad\cdots\quad\eta_n\right]^T$，则有

$$\begin{cases}\eta_i=c\xi_i+s\xi_j\\ \eta_j=-s\xi_i+c\xi_j\\ \eta_k=\xi_k\qquad(k\neq i,j)\end{cases}$$

当$\xi_i^2+\xi_j^2\neq 0$时，总可以选$\mathbf{c}=\dfrac{\xi_i}{\sqrt{\xi_i^2+\xi_j^2}}$，$\mathbf{s}=\dfrac{\xi_j}{\sqrt{\xi_i^2+\xi_j^2}}$使

$$\begin{cases}\eta_i=\sqrt{\xi_i^2+\xi_j^2}\\ \eta_j=0\end{cases}\rightarrow\mathbf{T}_{ij}\mathbf{x}=\left[\xi_1\quad\xi_2\quad\cdots\quad\sqrt{\xi_i^2+\xi_j^2}\quad\cdots\quad 0\quad\cdots\quad\xi_n\right]^T$$ 定理1.

设 $\mathbf{x}=\begin{bmatrix}\xi_1 & \xi_2 & \cdots & \xi_n\end{bmatrix}^T \neq \mathbf{0}$，则存在有限个 Givens 矩阵的乘积 T，使得 $\mathbf{Tx}=|\mathbf{x}|\mathbf{e}_1$

说明：（1）$|\mathbf{x}|=\sqrt{\|\mathbf{x}\|_2^2}=\sqrt{\mathbf{x}^T\mathbf{x}}$（x 为实数时），$|\mathbf{x}|=\sqrt{\mathbf{x}^H\mathbf{x}}$（x 为复数时）。

（2）$\mathbf{e}_1=\begin{bmatrix}1 & 0 & 0 & \cdots & 0\end{bmatrix}^T$

[证明]：$\xi_1 \neq 0$ 的情形

（1） 构造 $\mathbf{T}_{12}(c,s): c=\dfrac{\xi_1}{\sqrt{\xi_1^2+\xi_2^2}}, s=\dfrac{\xi_2}{\sqrt{\xi_1^2+\xi_2^2}}$

$$\mathbf{T}_{12}\mathbf{x}=\begin{bmatrix}\sqrt{\xi_1^2+\xi_2^2} & 0 & \xi_3 & \xi_4 & \cdots & \xi_n\end{bmatrix}^T$$

（2） 对 $\mathbf{T}_{12}\mathbf{x}$ 再考虑 $\mathbf{T}_{13}(c,s): c=\dfrac{\sqrt{\xi_1^2+\xi_2^2}}{\sqrt{\xi_1^2+\xi_2^2+\xi_3^2}}, s=\dfrac{\xi_3}{\sqrt{\xi_1^2+\xi_2^2+\xi_3^2}}$

$$\mathbf{T}_{13}\,\mathbf{T}_{12}\mathbf{x}=\begin{bmatrix}\sqrt{\xi_1^2+\xi_2^2+\xi_3^2} & 0 & 0 & \xi_4 & \cdots & \xi_n\end{bmatrix}^T$$

（3） 依此类推，构造

$$\mathbf{T}_{1k}(c,s): c=\frac{\sqrt{\xi_1^2+\cdots+\xi_k^2}}{\sqrt{\xi_1^2+\xi_2^2+\cdots+\xi_k^2}}, s=\frac{\xi_k}{\sqrt{\xi_1^2+\xi_2^2+\cdots+\xi_k^2}}$$

(k=2, 3, ⋯. . n)

$$\mathbf{T}_{1k}\left\langle \mathbf{T}_{1k-1}\left\{\mathbf{T}_{1k-2}\left[\cdots\mathbf{T}_{13}\left(\mathbf{T}_{12}\mathbf{x}\right)\right]\right\}\right\rangle$$
$$=\begin{bmatrix}\sqrt{\xi_1^2+\xi_2^2+\cdots+\xi_k^2} & 0 & 0 & \cdots & 0 & \xi_{k+1} & \cdots & \xi_n\end{bmatrix}^T$$

直至 k=n。令 $\mathbf{T}=\mathbf{T}_{1n}\mathbf{T}_{1n-1}\mathbf{T}\cdots\mathbf{T}_{12}$，则有

$$\mathbf{Tx}=\begin{bmatrix}\sqrt{\xi_1^2+\xi_2^2+\cdots\xi_n^2} & 0 & 0 & \cdots & 0\end{bmatrix}^T=|\mathbf{x}|\mathbf{e}_1$$

$\xi_1=0$ 的情形，从第一个不为零的 $\xi_i$ 开始运用上述方法即可

推论：对于任何非零列向量 $\mathbf{x}\in\mathbf{R}^n$ 及任何单位列向量 $\mathbf{z}(|\mathbf{z}|=1)$，均存在着有限个 Givens 矩阵的乘积 T，使 $\mathbf{Tx}=|\mathbf{x}|\mathbf{z}$。

[证明]：由上述定理，对 x 存在有限个 Givens 矩阵 $\mathbf{T}_{12}^{(1)},\mathbf{T}_{13}^{(1)},\cdots,\mathbf{T}_{1n}^{(1)}$ 的乘积

$$\mathbf{T}^{(1)}=\mathbf{T}_{1n}^{(1)}\mathbf{T}_{1n-1}^{(1)}\cdots\mathbf{T}_{13}^{(1)}\mathbf{T}_{12}^{(1)}\text{，使 }\mathbf{T}^{(1)}\mathbf{x}=|\mathbf{x}|\mathbf{e}_1$$

对 z 同理存在有限个 Givens 矩阵 $\mathbf{T}_{12}^{(2)},\mathbf{T}_{13}^{(2)},\cdots,\mathbf{T}_{1n}^{(2)}$ 的乘积

$$\mathbf{T}^{(2)}=\mathbf{T}_{1n}^{(2)}\mathbf{T}_{1n-1}^{(2)}\cdots\mathbf{T}_{13}^{(2)}\mathbf{T}_{12}^{(2)}\text{，使 }\mathbf{T}^{(2)}\mathbf{z}=|\mathbf{z}|\mathbf{e}_1=\mathbf{e}_1$$

即，

$$\mathbf{T}^{(1)}\mathbf{x}=|\mathbf{x}|\mathbf{T}^{(2)}\mathbf{z}=\mathbf{T}^{(2)}(|\mathbf{x}|\mathbf{z})\rightarrow\left(\mathbf{T}^{(2)}\right)^{-1}\mathbf{T}^{(1)}\mathbf{x}=|\mathbf{x}|\mathbf{z}$$

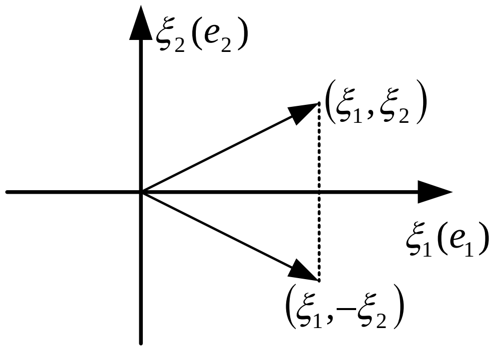
$$\Rightarrow\left(\mathbf{T}_{1n}^{(2)}\mathbf{T}_{1n-1}^{(2)}\cdots\mathbf{T}_{12}^{(2)}\right)^{-1}\left(\mathbf{T}_{1n}^{(1)}\mathbf{T}_{1n-1}^{(1)}\cdots\mathbf{T}_{12}^{(1)}\right)\mathbf{x}=|\mathbf{x}|\mathbf{z}$$

其中

$$\begin{aligned}&\left(\mathbf{T}_{1n}^{(2)}\mathbf{T}_{1n-1}^{(2)}\cdots\mathbf{T}_{12}^{(2)}\right)^{-1}\left(\mathbf{T}_{1n}^{(1)}\mathbf{T}_{1n-1}^{(1)}\cdots\mathbf{T}_{12}^{(1)}\right)\\&=\left(\mathbf{T}_{12}^{(2)}\right)^{-1}\left(\mathbf{T}_{13}^{(2)}\right)^{-1}\cdots\left(\mathbf{T}_{1n}^{(2)}\right)^{-1}\mathbf{T}_{1n}^{(1)}\mathbf{T}_{1n-1}^{(1)}\cdots\mathbf{T}_{12}^{(1)}\\&=\left(\mathbf{T}_{12}^{(2)}\right)^{\mathrm{T}}\left(\mathbf{T}_{13}^{(2)}\right)^{\mathrm{T}}\cdots\left(\mathbf{T}_{1n}^{(2)}\right)^{\mathrm{T}}\mathbf{T}_{1n}^{(1)}\mathbf{T}_{1n-1}^{(1)}\cdots\mathbf{T}_{12}^{(1)}\end{aligned}$$

为有限个 Givens 矩阵的乘积。

二 、Householder 矩阵与 Householder 变换




平面直角坐标系中，将向量 $\mathbf{x}=\begin{bmatrix}\xi_1 & \xi_2\end{bmatrix}^{\mathrm{T}}$ 关于 $\mathbf{e}_1$ 轴作镜像变换，则得到

$$\mathbf{y}=\begin{bmatrix}\xi_1\\-\xi_2\end{bmatrix}=\begin{bmatrix}1&0\\0&-1\end{bmatrix}\begin{bmatrix}\xi_1\\\xi_2\end{bmatrix}=\left(\mathbf{I}-2\mathbf{e}_2\mathbf{e}_2^{\mathrm{T}}\right)\mathbf{x}=\mathbf{H}\mathbf{x}$$

一般地，可将其推广

1. 定义：设单位列向量 $\mathbf{u}\in\mathbf{R}^n$，称 $\mathbf{H}=\mathbf{I}-2\mathbf{u}\mathbf{u}^{\mathrm{T}}$ 为 Householder 矩阵（初等反射矩阵），由 Householder 矩阵所确定的线性变换（ $\mathbf{y}=\mathbf{H}\mathbf{x}$ ）称为 Householder 变换

2 . 性质

（1）$\mathbf{H}^{\mathrm{T}}=\mathbf{H}$（实对称），$\mathbf{H}^{-1}=\mathbf{H}^{\mathrm{T}}$（正交），$\mathbf{H}^2=\mathbf{I}$（对合），$\mathbf{H}^{-1}=\mathbf{H}$ （自逆），$\det\mathbf{H}=-1$

为证明第 5 条，可利用如下引理。

引理：设 $\mathbf{A}\in\mathbf{R}^{m\times n},\mathbf{B}\in\mathbf{R}^{n\times m}$，则 $\det\left(\mathbf{I}_m+\mathbf{A}\mathbf{B}\right)=\det\left(\mathbf{I}_n+\mathbf{B}\mathbf{A}\right)$

[证明]：参考如下的分块矩阵$\begin{bmatrix} \mathbf{I_n} & \mathbf{B} \\ -\mathbf{A} & \mathbf{I_m} \end{bmatrix}$的行列式，有

$$\det\begin{bmatrix} \mathbf{I_n} & \mathbf{B} \\ -\mathbf{A} & \mathbf{I_m} \end{bmatrix} = \det\begin{bmatrix} \mathbf{I_n} & \mathbf{B} \\ \mathbf{0} & \mathbf{I_m}+\mathbf{AB} \end{bmatrix} = \det(\mathbf{I_m}+\mathbf{AB}) = \det\begin{bmatrix} \mathbf{I_n}+\mathbf{BA} & \mathbf{B} \\ \mathbf{0} & \mathbf{I_m} \end{bmatrix}$$ 故，

$$= \det(\mathbf{I_n}+\mathbf{BA})$$

$$\det(\mathbf{H}) = \det\left[\mathbf{I}+\left(-2\mathbf{uu^T}\right)\right] = \mathbf{1}+\mathbf{u^T}(-2\mathbf{u}) = \mathbf{1}-\mathbf{2} = -\mathbf{1}$$

$$\bullet \mathbf{H^T} = \left(\mathbf{I}-2\mathbf{uu^T}\right)^{\mathbf{T}} = \mathbf{I}-2\left(\mathbf{u^T}\right)^{\mathbf{T}}\mathbf{u^T} = \mathbf{H}$$

$$\bullet \mathbf{H}^2 = \left(\mathbf{I}-2\mathbf{uu^T}\right)\left(\mathbf{I}-2\mathbf{uu^T}\right) = \mathbf{I}-2\mathbf{uu^T}-2\mathbf{uu^T}+4\mathbf{uu^Tuu^T} = \mathbf{I}$$

定理 2. 对于任何非零列向量$\mathbf{x}\in\mathbf{R^n}$及任何单位列向量$\mathbf{z}\in\mathbf{R^n}$，存在 Householder 矩阵 H，使得

$$\mathbf{Hx} = |\mathbf{x}|\mathbf{z}\text{。}$$

[证明] 当$\mathbf{x}=|\mathbf{x}|\mathbf{z}$时，选 u 满足$\mathbf{u^Tx}=\mathbf{0}$，则$\mathbf{Hx}=\left(\mathbf{I}-2\mathbf{uu^T}\right)\mathbf{x}=\mathbf{x}=|\mathbf{x}|\mathbf{z}$

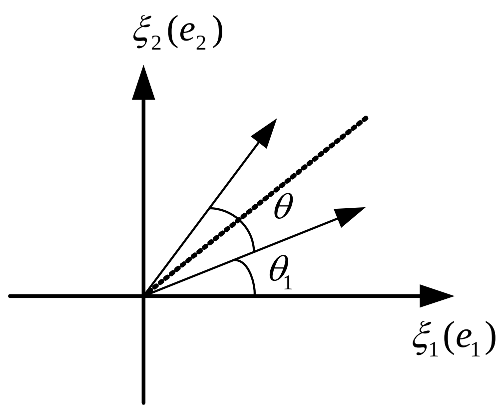
当$\mathbf{x}\neq|\mathbf{x}|\mathbf{z}$时，选$\mathbf{u}=\dfrac{\mathbf{x}-|\mathbf{x}|\mathbf{z}}{\left|\mathbf{x}-|\mathbf{x}|\mathbf{z}\right|}$，有

$$\left|\mathbf{x}-|\mathbf{x}|\mathbf{z}\right|^2 = \left(\mathbf{x}-|\mathbf{x}|\mathbf{z}\right)^{\mathbf{T}}\left(\mathbf{x}-|\mathbf{x}|\mathbf{z}\right) = \mathbf{xx^T}+|\mathbf{x}|^2\mathbf{z^Tz}-|\mathbf{x}|\left(\mathbf{z^Tx}+\mathbf{x^Tz}\right)$$
$$= 2\left(\mathbf{x^Tx}-|\mathbf{x}|\mathbf{z^Tx}\right) = 2\left(\mathbf{x}-|\mathbf{x}|\mathbf{z}\right)^{\mathbf{T}}\mathbf{x}$$

$$\mathbf{Hx} = \left(\mathbf{I}-2\frac{\mathbf{x}-|\mathbf{x}|\mathbf{z}}{\left|\mathbf{x}-|\mathbf{x}|\mathbf{z}\right|}\frac{\mathbf{x^T}-|\mathbf{x}|\mathbf{z^T}}{\left|\mathbf{x}-|\mathbf{x}|\mathbf{z}\right|}\right)\mathbf{x} = \mathbf{x}-\left(\mathbf{x}-|\mathbf{x}|\mathbf{z}\right) = |\mathbf{x}|\mathbf{z}$$

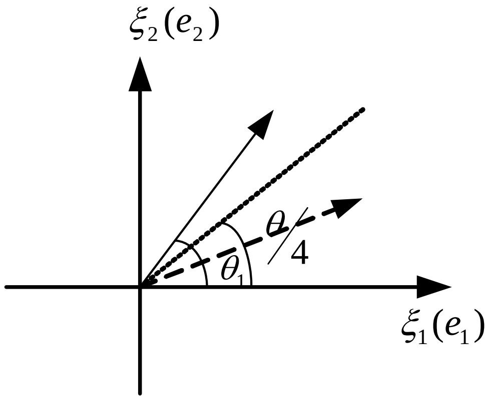
定理 3. 初等旋转矩阵（Givens 矩阵）是两个初等反射矩阵的乘积。

证明参见$\mathbf{p_{202}}$，较容易。我们这里主要是给出一种几何解释。



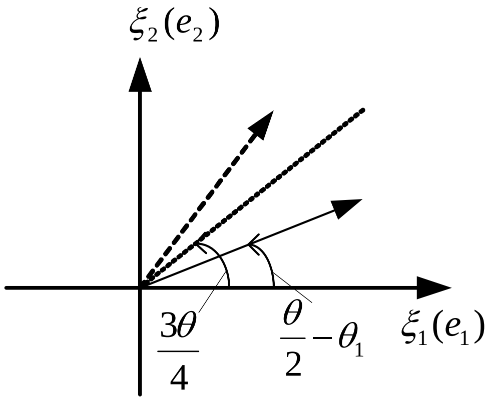

从表明上看，似乎一种反射变换即可代替旋转变换。实际上是不对的，因为这样的反射变换对应的

对称轴沿$\left(\theta_1+\theta_2/2\right)$方向，与$\theta_1$有关




实际上，旋转变换可由这样两次反射变换的作用来代替。

首先，关于沿$\left(\theta/4\right)$对称轴作反射变换，则原向量沿$\theta_1$方向转至$\left(-\theta_1+\theta/2\right)$。




其次，关于沿$\left(3\theta/4\right)$对称轴作反射变换，则向量反射至$\left(\frac{\theta}{2}-\theta_1\right)+2\left(\frac{\theta}{4}+\theta_1\right)=\theta_1+\theta$。

正是原向量沿$\theta_1$方向转$\theta$的结果。

- 旋转变换可用两个反射变换的连续作用来代替，即$\mathbf{T_{ij}}=\mathbf{H_v H_u}$。但是反射变换却不可能用多个旋转变换的连续作用来代替。这是因为$\mathbf{det}\left(\mathbf{T_{ij}}\right)\equiv\mathbf{1},\mathbf{det}\left(\mathbf{H}\right)\equiv-\mathbf{1}$。由两个－1 的乘积可得 1，但多个 1 的乘积只能是 1，不是－1。

三、 QR 分解

1. 定义：如果实（复）矩阵 A 可化为正交（酉）矩阵 Q 与实（复）上三角矩阵 R 的乘积，即$\mathbf{A}=\mathbf{QR}$，则称上式为 A 的 QR 分解。

2. 定理：设 A 是 n 阶的非奇异矩阵，则存在正交（酉）矩阵 Q 与实（复）上三角矩阵 R 使得$\mathbf{A}=\mathbf{QR}$，且除去相差一个对角元素的绝对值（模）全为 1 的对角因子外，上述分解唯一。

[证明]：设 A 记为$\mathbf{A}=\left[\mathbf{a_1}\quad\mathbf{a_2}\quad\cdots\quad\mathbf{a_n}\right]$，A 非奇异$\rightarrow\mathbf{a_1},\mathbf{a_2},\cdots,\mathbf{a_n}$线性无关

采用 Gram-schmidt 正交化方法将它们正交化，可得

$$\begin{cases} \mathbf{b}_1=\mathbf{a}_1 & \mathbf{q}_1=\mathbf{b}_1/|\mathbf{b}_1| \\ \mathbf{b}_2=\mathbf{a}_2-\mathbf{k}_{21}\mathbf{b}_1 & \mathbf{q}_2=\mathbf{b}_2/|\mathbf{b}_2| \\ \mathbf{b}_3=\mathbf{a}_3-\mathbf{k}_{31}\mathbf{b}_1-\mathbf{k}_{32}\mathbf{b}_2 & \vdots\rightarrow(\mathbf{q}_i,\mathbf{q}_j)=\delta_{ij} \\ \vdots & \\ \mathbf{b}_n=\mathbf{a}_n-\mathbf{k}_{n1}\mathbf{b}_1-\mathbf{k}_{n2}\mathbf{b}_2-\cdots-\mathbf{k}_{nn-1}\mathbf{b}_{n-1} & \mathbf{q}_n=\mathbf{b}_n/|\mathbf{b}_n| \end{cases}$$

$$\mathbf{k}_{ij}=\frac{(\mathbf{a}_i,\mathbf{a}_j)}{(\mathbf{b}_j,\mathbf{b}_j)}\leftarrow(\mathbf{a}_i,\mathbf{b}_j)=0$$

$$\begin{aligned}[\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n]&=[\mathbf{b}_1 \quad \mathbf{b}_2 \quad \cdots \quad \mathbf{b}_n]\begin{bmatrix}1 & \mathbf{k}_{21} & \cdots & \mathbf{k}_{n1}\\ & 1 & \cdots & \mathbf{k}_{n2}\\ & & \ddots & \vdots\\ & & & 1\end{bmatrix}\\ &=[\mathbf{b}_1 \quad \mathbf{b}_2 \quad \cdots \quad \mathbf{b}_n]\mathbf{C}\\ &=[\mathbf{q}_1 \quad \mathbf{q}_2 \quad \cdots \quad \mathbf{q}_n]\begin{bmatrix}|\mathbf{b}_1| & & & \\ & |\mathbf{b}_2| & & \\ & & \ddots & \\ & & & |\mathbf{b}_n|\end{bmatrix}\mathbf{C}\\ &=\mathbf{QR}\end{aligned}$$

Q 是正交（酉）矩阵

R 是实（复）上三角矩阵

唯一性：采用反证法。设存在两个 QR 分解，$\mathbf{A}=\mathbf{QR}=\mathbf{Q}_1\mathbf{R}_1$，则

$$\mathbf{Q}=\mathbf{Q}_1\mathbf{R}_1\mathbf{R}^{-1}=\mathbf{Q}_1\mathbf{D} \qquad \rightarrow \mathbf{D}\text{为上三角矩阵}$$

而 $\mathbf{I}=\mathbf{Q}^H\mathbf{Q}=(\mathbf{Q}_1\mathbf{D})^H(\mathbf{Q}_1\mathbf{D})=\mathbf{D}^H\mathbf{D}\rightarrow\mathbf{D}$为酉（正交）矩阵

故，D 只能为对角阵

$$\mathbf{D}=\begin{bmatrix}\mathbf{d}_{11} & \mathbf{d}_{12} & \cdots & \mathbf{d}_{1n}\\ & \mathbf{d}_{22} & \cdots & \mathbf{d}_{2n}\\ & & \ddots & \vdots\\ & & & \mathbf{d}_{nn}\end{bmatrix} \qquad \rightarrow \mathbf{d}_{12}=0;\mathbf{d}_{13}=\mathbf{d}_{23}=0;\cdots$$

。D 是对角元素绝对值

（模）全为 1 的对角阵。
这一证明方法可推广为：

定理 5： 设 A 是 $\mathbf{m\times n}$ 的实（复）矩阵，且其 n 个列线性无关，则 A 具有分解 $\mathbf{A=QR}$。其中 Q 是 $\mathbf{m\times n}$ 阶实（复）矩阵，且满足 $\mathbf{Q^TQ=I(Q^HQ=I)}$，R 是 n 阶实（复）非奇异上三角矩阵。除了相差一个对角元素的绝对值（模）全为 1 的对角阵因子外，上述分解唯一。

3. 求 QR 分解的方法

[方法一]采用 Givens 方法

将 n 阶非奇异矩阵 A 写为

$\mathbf{A}=\begin{bmatrix}\mathbf{b}^{(1)^T}\\ *\end{bmatrix}^{\mathbf{T}}$ 则存在有限个 Givens 矩阵的乘积 $\mathbf{T_1}$，使得

$$\mathbf{T_1b^{(1)}=\left|b^{(1)}\right|e_1\rightarrow T_1A=\begin{bmatrix}a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)}\\ 0 & & & \\ \vdots & & A^{(1)} & \\ 0 & & & \end{bmatrix}}$$

$\mathbf{A^{(1)}}$ 写成 $\mathbf{A}^{(1)}=\begin{bmatrix}\mathbf{b}^{(2)^T}\\ *\end{bmatrix}^{\mathbf{T}}\rightarrow$ 存在 $\mathbf{T_2}$，使得

$$\mathbf{T_2A^{(1)}=\begin{bmatrix}a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2n}^{(2)}\\ 0 & & & \\ \vdots & & A^{(2)} & \\ 0 & & & \end{bmatrix}}$$

$\vdots$

$\mathbf{A^{(n-2)}}$ 写成 $\mathbf{A}^{(n-2)}=\begin{bmatrix}\mathbf{b}^{(n-1)^T}\\ *\end{bmatrix}^{\mathbf{T}}\rightarrow$ 存在 $\mathbf{T_{n-1}}$，使得

$$\mathbf{T_{n-1}A^{(n-2)}=\begin{bmatrix}a_{n-1,n-1}^{(n-1)} & a_{n-1,n}^{(n-1)}\\ 0 & a_{nn}^{(n-1)}\end{bmatrix}}$$

令 $\mathbf{T}=\begin{bmatrix}\mathbf{I_{n-2}} & \mathbf{0}\\ \mathbf{0} & \mathbf{T_{n-1}}\end{bmatrix}\begin{bmatrix}\mathbf{I_{n-3}} & \mathbf{0}\\ \mathbf{0} & \mathbf{T_{n-2}}\end{bmatrix}\cdots\begin{bmatrix}\mathbf{I_2} & \mathbf{0}\\ \mathbf{0} & \mathbf{T_3}\end{bmatrix}\begin{bmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & \mathbf{T_2}\end{bmatrix}\mathbf{T_1}$，则有

$$TA=\begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ & & \ddots & \vdots \\ & & & a_{nn}^{n-1} \end{bmatrix}=R \to A=T^{-1}R=QR$$

其中，R 为上三角矩阵，Q=$T^{-1}$ 为正交矩阵

[方法二]采用 Householde 方法

$A=\begin{bmatrix} b^{(1)} & * \end{bmatrix}$ 存在 $H_1$，使得

$$H_1 b^{(1)}=\left|b^{(1)}\right| e_1 \to H_1 A=\begin{bmatrix} \left|b^{(1)}\right| e_1^n & A^{(1)} \end{bmatrix}$$

$A^{(1)}=\begin{bmatrix} b^{(2)} & * \end{bmatrix}$ 存在 $H_2$，使得 $H_2 A^{(1)}=\begin{bmatrix} \left|b^{(2)}\right| e_1^{n-1} & A^{(2)} \end{bmatrix}\cdots$

$A^{(n-2)}=\begin{bmatrix} b^{(n-1)} & b^{(n)} \end{bmatrix}$ 存在 $H_{n-1}$，使得 $H_{n-1}A^{(n-2)}=\begin{bmatrix} a_{n-1,n-1}^{(n-1)} & a_{n-1,n}^{(n-1)} \\ 0 & a_{nn}^{(n-1)} \end{bmatrix}$

令 $S=\begin{bmatrix} I_{n-2} & 0 \\ 0 & H_{n-1} \end{bmatrix}\begin{bmatrix} I_{n-3} & 0 \\ 0 & H_{n-2} \end{bmatrix}\cdots\begin{bmatrix} I_2 & 0 \\ 0 & H_3 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & H_2 \end{bmatrix}H_1$

$$H_{l+1}=I_{n-l}-2uu^T(u\in R^{n-l},u^Tu=1)$$

则

$$\begin{bmatrix} I_l & 0 \\ 0 & H_{l+1} \end{bmatrix}=\begin{bmatrix} I_l & 0 \\ 0 & I_{n-l} \end{bmatrix}-2\begin{bmatrix} 0 & 0 \\ 0 & uu^T \end{bmatrix}=I_n-2\begin{bmatrix} 0 \\ u \end{bmatrix}\begin{bmatrix} 0^T & u^T \end{bmatrix}=I_n-2uu^T$$

$$SA=\begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ & & \ddots & \vdots \\ & & & a_{nn}^{n-1} \end{bmatrix}=R\text{，}S^{-1}=Q \text{ 为正交矩阵}$$

以上两种方法中的前一种方法可推广到复矩阵的情况。

3. Gram-schmidt 正交归一化方法

$A=\begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$，各列向量线性无关可进行正交化

$$q_1=a_1\Big/\left|a_1\right| \qquad y_1=a_1$$

$$q_2=\frac{a_2+k_{21}q_1}{\left|a_2+k_{21}q_1\right|},k_{21}=-\langle a_2,q_1\rangle \qquad y_2=a_2+k_{21}q_1$$

$$\vdots$$

$$\mathbf{q}_l=\frac{\mathbf{a}_l+\sum_{j=1}^{l-1}\mathbf{k}_{lj}\mathbf{q}_j}{\left|\mathbf{a}_l+\sum_{j=1}^{l-1}\mathbf{k}_{lj}\mathbf{q}_j\right|},\mathbf{k}_{lj}=-\left\langle\mathbf{a}_l,\mathbf{q}_j\right\rangle \qquad \mathbf{y}_l=\mathbf{a}_l+\sum_{j=1}^{l-1}\mathbf{k}_{lj}\mathbf{q}_j$$

$$\mathbf{q}_i^H\mathbf{q}_j=\delta_{ij}=\begin{cases}0 & i\neq j\\ 1 & i=j\end{cases}\rightarrow \mathbf{Q}=\begin{bmatrix}\mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n\end{bmatrix}$$，满足 $\mathbf{Q}^T\mathbf{Q}=\mathbf{I}$

改写：

$$\mathbf{a}_1=\mathbf{q}_1|\mathbf{y}_1|$$

$$\mathbf{a}_2=\mathbf{q}_2|\mathbf{y}_2|-\mathbf{k}_{21}\mathbf{q}_1$$

$$\vdots$$

$$\mathbf{a}_l=\mathbf{q}_l|\mathbf{y}_l|-\sum_{j=1}^{l-1}\mathbf{k}_{lj}\mathbf{q}_j$$

$$\begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n\end{bmatrix}=\begin{bmatrix}\mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n\end{bmatrix}\begin{bmatrix}|\mathbf{y}_1| & -\mathbf{k}_{21} & -\mathbf{k}_{31} & \cdots & -\mathbf{k}_{n1}\\ & |\mathbf{y}_2| & -\mathbf{k}_{32} & \cdots & -\mathbf{k}_{n2}\\ & & |\mathbf{y}_3| & \cdots & \\ & & & \ddots & \vdots\\ & & & & |\mathbf{y}_n|\end{bmatrix}$$

$$=\mathbf{QR}$$

作业：p219-220，1、7、8

# 第十二讲　满秩分解与奇异值分解

## 一、矩阵的满秩分解

1. 定义：设 $\mathbf{A}\in \mathbf{C}_r^{m\times n}(\mathbf{r}>0)$，若存在矩阵 $\mathbf{F}\in \mathbf{C}_r^{m\times r}$ 及 $\mathbf{G}\in \mathbf{C}_r^{r\times n}$，使得

$\mathbf{A}=\mathbf{FG}$，则称其为 $\mathbf{A}$ 的一个满秩分解。

说明：（1）$\mathbf{F}$ 为列满秩矩阵，即列数等于秩；$\mathbf{G}$ 为行满秩矩阵，即行数等于秩。

（2）满秩分解不唯一。$\forall \mathbf{D}\in \mathbf{C}_r^{r\times r}$（$\mathbf{r}$ 阶可逆方阵），则

$$\mathbf{A}=\mathbf{FG}=\mathbf{F}(\mathbf{D}\mathbf{D}^{-1})\mathbf{G}=(\mathbf{FD})(\mathbf{D}^{-1}\mathbf{G})=\mathbf{F}_1\mathbf{G}_1，且\ \mathbf{F}_1\in \mathbf{C}_r^{m\times r},\mathbf{G}_1\in \mathbf{C}_r^{r\times n}$$

2. 存在性定理：任何非零矩阵均存在满秩分解

证明：采用构造性证明方法。设 $\mathbf{A}\in \mathbf{C}_r^{m\times n}$，则存在初等变换矩阵 $\mathbf{E}\in \mathbf{C}_m^{m\times m}$，

使 $$\mathbf{EA}=\mathbf{B}=\begin{array}{c}\left[\begin{matrix}\mathbf{G}\\ \cdots\cdots\\ \mathbf{O}\end{matrix}\right]\end{array}\begin{array}{l}\mathbf{r}行\\ \\ (\mathbf{m}-\mathbf{r})行\end{array}$$，其中 $\mathbf{G}\in \mathbf{C}_r^{r\times n}$

将 $\mathbf{A}$ 写成 $\mathbf{A}=\mathbf{E}^{-1}\mathbf{B}$，并把 $\mathbf{E}^{-1}$ 分块成 $\mathbf{E}^{-1}=[\underset{\mathbf{r}列}{\mathbf{F}}\ |\ \underset{(\mathbf{m}-\mathbf{r})列}{\mathbf{S}}]$，其中 $\mathbf{F}\in \mathbf{C}_r^{m\times r}$

$$\therefore \mathbf{A}=\left[\begin{matrix}\mathbf{F} & \vdots & \mathbf{S}\end{matrix}\right]\left[\begin{matrix}\mathbf{G}\\ \cdots\\ \mathbf{O}\end{matrix}\right]=\mathbf{FG}$$　　是满秩分解。

3. Hermite 标准形（行阶梯标准形）

设 $\mathbf{B}\in \mathbf{C}_r^{m\times n}(\mathbf{r}>0)$，且满足

（1）　$\mathbf{B}$ 的前 $\mathbf{r}$ 行中每一行至少含一个非零元素（称为非零行），且第一个非零元素为 1，而后 $(\mathbf{m}-\mathbf{r})$ 行的元素全为零（称为零行）；

（2）　若 $\mathbf{B}$ 中第 $\mathbf{i}$ 行的第一个非零元素（即 1）在第 $\mathbf{j}_i$ 列 $(\mathbf{i}=1,2,\ldots,\mathbf{r})$，则

$$\mathbf{j}_1<\mathbf{j}_2<\ldots<\mathbf{j}_r；$$

（3）　矩阵 $\mathbf{B}$ 的第 $\mathbf{j}_1$ 列，第 $\mathbf{j}_2$ 列，…，第 $\mathbf{j}_r$ 列合起来恰为 $\mathbf{m}$ 阶单位方阵 $\mathbf{I}_m$ 的前 $\mathbf{r}$ 列（即 $\mathbf{j}_1,\mathbf{j}_2,\ldots,\mathbf{j}_r$ 列上除了前述的 1 外全为 0）则称 $\mathbf{B}$ 为 Hermite 标准形。

例 1 $$\mathbf{B}_1=\begin{bmatrix}1&2&0&0&-1&3\\0&0&1&0&2&2\\0&0&0&1&-1&1\\0&0&0&0&0&0\\0&0&0&0&0&0\end{bmatrix}_{5\times6}\in\mathbf{C}_3^{5\times6}$$ 为 Hermite 标准形

$$\mathbf{B}_2=\begin{bmatrix}0&0&1&0&2\\0&0&0&1&3\\0&0&0&0&0\\0&0&0&0&0\end{bmatrix}_{4\times5}\in\mathbf{C}_2^{4\times5}$$ 也是 Hermite 标准形

4. 满秩分解的一种求法

设 $\mathbf{A}\in\mathbf{C}_r^{m\times n}$，

（1） 采用行初等变换将 $\mathbf{A}$ 化成 Hermite 标准形，其矩阵形式为 $\mathbf{EA}=\mathbf{B}$，其中 $\mathbf{B}$ 为 Hermite 标准形定义中给出的形状；

（2） 选取置换矩阵

$1^\circ$ $\mathbf{P}$ 的第 $\mathbf{i}$ 列为 $\mathbf{e}_{j_i}$，即该列向量除第 $\mathbf{j_i}$ 个元素为 1 外，其余元素全为零 $(\mathbf{i}=1,2,\ldots,\mathbf{r})$，其中 $\mathbf{j_i}$ 为 Hermite 标准形中每行第一个非零元素（即 1）所在的列数；

$2^\circ$ 其它 $(\mathbf{n}-\mathbf{r})$ 列只需确保 $\mathbf{P}$ 为置换矩阵即可（$\mathbf{P}$ 的每一行，每一列均只有一个非零元素，且为 1）；

$3^\circ$ 用 $\mathbf{P}$ 右乘任何矩阵（可乘性得到满足时），即可将该矩阵的第 $\mathbf{j_i}$ 列置换到新矩阵（即乘积矩阵）的第 $\mathbf{i}$ 列

$4^\circ$ 令 $\mathbf{P}=[\underset{\mathbf{r}列}{\mathbf{P}_1}\quad|\quad\underset{(\mathbf{n}-\mathbf{r})列}{*}]$，即 $\mathbf{P}_1=\begin{bmatrix}\mathbf{e}_{j_1} & \mathbf{e}_{j_2}\ldots\mathbf{e}_{j_r}\end{bmatrix}_{\mathbf{n}\times\mathbf{r}}\in\mathbf{C}_r^{n\times r}$

（3）令 $\mathbf{G}=\mathbf{B}$ 的前 $\mathbf{r}$ 行 $\in\mathbf{C}_n^{r\times n}$，$\mathbf{F}=\mathbf{AP}_1\in\mathbf{C}_r^{m\times r}$ 则 $\mathbf{A}=\mathbf{FG}$

证明：$\mathbf{EA}=\mathbf{B}=\begin{bmatrix}\mathbf{G}\\\mathbf{O}\end{bmatrix}$，$\mathbf{A}=\mathbf{E}^{-1}\mathbf{B}=[\mathbf{F}\quad|\quad\mathbf{S}]\begin{bmatrix}\mathbf{G}\\\mathbf{O}\end{bmatrix}=\mathbf{FG}$ 则 $\mathbf{F}\in\mathbf{C}_r^{m\times r}$，$\mathbf{G}\in\mathbf{C}_r^{r\times n}$，$\mathbf{G}$ 已知，但 $\mathbf{F}=?$，当然可以通过求出 $\mathbf{E},\mathbf{E}^{-1}$ 再将 $\mathbf{E}^{-1}$ 分块得到，但这样 $\mathbf{G}$ 就没必要采用 Hermite 标准形形式，注意到 $\mathbf{BP}_1=\begin{bmatrix}\mathbf{I}_r\\\mathbf{O}\end{bmatrix}$，则 $\mathbf{AP}_1=\mathbf{E}^{-1}\mathbf{BP}_1=[\mathbf{F}\quad|\quad\mathbf{S}]\begin{bmatrix}\mathbf{I}_r\\\mathbf{O}\end{bmatrix}=\mathbf{F}$ 证毕

例 2　$\mathbf{A}=\begin{bmatrix}1&2&3&0\\0&2&1&-1\\1&0&2&1\end{bmatrix}$求其满秩分解

解：(1)首先求出$\mathbf{A}$的秩。显然，前两行互相独立，而第三行可由第一行减去第二行得到，故$\mathbf{r}=2$。

(2)进行初等变换将$\mathbf{A}$化为 Hermite 标准型。

$$[\mathbf{A}\quad|\quad\mathbf{I}_3]=\begin{bmatrix}1&2&3&0&.&1&0&0\\0&2&1&-1&.&0&1&0\\1&0&2&1&.&0&0&1\end{bmatrix}\underset{\rightarrow}{(3)-(1)+(2)}$$

$$\begin{bmatrix}1&2&3&0&.&1&0&0\\0&2&1&-1&.&0&1&0\\0&0&0&0&.&-1&1&1\end{bmatrix}\underset{\rightarrow}{(1)-(2)}$$

$$\begin{bmatrix}1&0&2&1&.&1&-1&0\\0&2&1&-1&.&0&1&0\\0&0&0&0&.&-1&-1&1\end{bmatrix}\underset{\rightarrow}{(2)/2}$$

$$\begin{bmatrix}1&0&2&1&.&1&-1&0\\0&1&1/2&-1/2&.&0&1/2&0\\0&0&0&0&.&-1&1&1\end{bmatrix},$$

即

$$\mathbf{E}=\begin{bmatrix}1&-1&0\\0&1/2&0\\-1&1&1\end{bmatrix},\quad \mathbf{B}=\begin{bmatrix}1&0&2&1\\0&1&1/2&-1/2\\0&0&0&0\end{bmatrix},\quad \mathbf{G}=\begin{bmatrix}1&0&2&1\\0&1&1/2&-1/2\end{bmatrix}$$

(3)求出$\mathbf{P}_1$及$\mathbf{AP}_1$

由$\mathbf{B}$可见，$\mathbf{j}_1=\mathbf{1},\mathbf{j}_2=\mathbf{2}$故$\mathbf{P}_1=\begin{bmatrix}1&0\\0&1\\0&0\\0&0\end{bmatrix}$，$\mathbf{F}=\mathbf{AP}_1=\begin{bmatrix}1&2\\0&2\\1&0\end{bmatrix}$

验证：$\mathbf{FG}=\begin{bmatrix}1&2\\0&2\\1&0\end{bmatrix}\begin{bmatrix}1&0&2&1\\0&1&1/2&-1/2\end{bmatrix}=\begin{bmatrix}1&2&3&0\\0&2&1&-1\\1&0&2&1\end{bmatrix}$

而 $\mathbf{E}^{-1}=\begin{bmatrix}1&2&0\\0&2&0\\1&0&1\end{bmatrix}$

二、酉对角分解与奇异值分解

1. 厄米矩阵的谱分解

$\mathbf{A}$ 为厄米矩阵，则存在酉矩阵 $\mathbf{U}$，使

$$\mathbf{U}^{\mathbf{H}}\mathbf{A}\mathbf{U}=\begin{bmatrix}\lambda_1 & & & \mathbf{O}\\ & \lambda_2 & & \\ & & \ddots & \\ \mathbf{O} & & & \lambda_n\end{bmatrix}=\Lambda$$

将 $\mathbf{U}$ 写成列向量形式，即 $\mathbf{U}=\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_n\end{bmatrix}$，则

$$\mathbf{A}=\mathbf{U}\Lambda\mathbf{U}^{\mathbf{H}}=\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_n\end{bmatrix}\begin{bmatrix}\lambda_1 & & & & \mathbf{O}\\ & \lambda_2 & & & \\ & & . & & \\ & & & . & \\ \mathbf{O} & & & & \lambda_n\end{bmatrix}\begin{bmatrix}\mathbf{u}_1^{\mathbf{H}}\\ \mathbf{u}_2^{\mathbf{H}}\\ .\\ .\\ \mathbf{u}_n^{\mathbf{H}}\end{bmatrix}=\sum_{i=1}^{n}\lambda_i\mathbf{u}_i\mathbf{u}_i^{\mathbf{H}}$$

2. 非奇异矩阵的酉对角分解

定理：设 $\mathbf{A}$ 为 $\mathbf{n}$ 阶非奇异矩阵，则存在 $\mathbf{n}$ 阶酉矩阵 $\mathbf{U}$ 及 $\mathbf{V}$，使得

$$\mathbf{U}^{\mathbf{H}}\mathbf{A}\mathbf{V}=\begin{bmatrix}\sigma_1 & & & & \mathbf{O}\\ & \sigma_2 & & & \\ & & . & & \\ & & & . & \\ \mathbf{O} & & & & \sigma_n\end{bmatrix},\quad \sigma_i>0(i=1,2,\dots,n)$$

（若将 $\mathbf{U},\mathbf{V}$ 写成 $\mathbf{U}=\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_n\end{bmatrix},\mathbf{V}=\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n\end{bmatrix}$，则

$\mathbf{A}=\sum_{i=1}^{n}\sigma_i\mathbf{u}_i\mathbf{v}_i^{n}$）

证：$\mathbf{A}^{\mathbf{H}}\mathbf{A}$ 也为 $\mathbf{n}$ 阶非奇异矩阵，而且是厄米、正定矩阵，故存在 $\mathbf{n}$ 阶酉矩阵 $\mathbf{V}$，使

$$V^H(A^HA)V=\begin{bmatrix}\sigma_1^2 & & & & O\\ & \sigma_2^2 & & & \\ & & . & & \\ & & & . & \\ O & & & & \sigma_n^2\end{bmatrix}$$ $\sigma_i^2$ 为 $A^HA$ 的特征值。

令 $\sum=\begin{bmatrix}\sigma_1 & & & & O\\ & \sigma_2 & & & \\ & & . & & \\ & & & . & \\ O & & & & \sigma_n\end{bmatrix}$，则 $V^HA^HAV=\sum^2$

令 $U^H=\sum^{-1}V^HA^H,\rightarrow U=AV\sum^{-1}$，则

$$U^HU=\sum^{-1}(V^HA^HAV)\sum^{-1}=I_n$$

即 $U$ 也是酉矩阵，而且 $U^HAV=\sum^{-1}V^HA^HAV=\sum$ 证毕

酉对角分解的求法正如证明中所给：先对 $A^HA$ 对角化（酉对角化），求出变换矩阵 $V$，再令 $U=AV\sum^{-1}$ 即可。

3. 一般矩阵的奇异值分解

定理：设 $A\in C_r^{m\times n}$，则存在 $m$ 阶酉矩阵 $U$ 及 $n$ 阶酉矩阵 $V$，使

$$U^HAV=\begin{array}{c}\left[\begin{array}{cccc|c}\sigma_1 & & & 0 & \\ & \sigma_2 & & & O\\ & & \ddots & & \\ 0 & & & \sigma_r & \\ \hline & O & & & O\end{array}\right]\begin{array}{l}\\ \\ r\text{行}\\ \\ (m-r)\text{行}\end{array}\\ \quad r\text{列}\qquad (n-r)\text{列}\end{array}$$ 即

$$A=U\begin{bmatrix}\sigma_1 & & & O\\ & \sigma_2 & & \\ & & . & \\ & & & \sigma_r\\ O & & & O\end{bmatrix}V^H$$

证：首先考虑 $A^HA$。因为 $rank(A^HA)=rank(AA^H)=rankA$，故 $A^HA\in C_r^{n\times n}$，而且是厄米、半正定的，存在 $n$ 阶酉矩阵 $V$，使

$$\mathbf{V}^H(\mathbf{A}^H\mathbf{A})\mathbf{V}=\begin{bmatrix}\sigma_1^2 & & & & \mathbf{O}\\ & \sigma_2^2 & & & \\ & & \cdot & & \\ & & & \sigma_r^2 & \\ \mathbf{O} & & & & \mathbf{O}\end{bmatrix}_{n\times n}$$

令 $\sum=\begin{bmatrix}\sigma_1 & & & & \mathbf{O}\\ & \sigma_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ \mathbf{O} & & & & \sigma_r\end{bmatrix}$，$\mathbf{V}=\begin{matrix}[\mathbf{V}_1 & | & \mathbf{V}_2]\\ \mathbf{r}\text{列} & & (\mathbf{n}-\mathbf{r})\text{列}\end{matrix}$ 则

$$\mathbf{V}^H(\mathbf{A}^H\mathbf{A})\mathbf{V}=\begin{bmatrix}\mathbf{V}_1^H(\mathbf{A}^H\mathbf{A})\mathbf{V}_1 & \mathbf{V}_1^H(\mathbf{A}^H\mathbf{A})\mathbf{V}_2\\ \mathbf{V}_2^H(\mathbf{A}^H\mathbf{A})\mathbf{V}_1 & \mathbf{V}_2^H(\mathbf{A}^H\mathbf{A})\mathbf{V}_2\end{bmatrix}=\begin{bmatrix}\sum^2 & \mathbf{O}\\ \mathbf{O} & \mathbf{O}\end{bmatrix}$$

$$\mathbf{V}_1^H(\mathbf{A}^H\mathbf{A})\mathbf{V}_1=\sum^2 \qquad \mathbf{V}_1^H(\mathbf{A}^H\mathbf{A})\mathbf{V}_2=\mathbf{O}_{r\times(n-r)}$$

$$\mathbf{V}_2^H(\mathbf{A}^H\mathbf{A})\mathbf{V}_2=\mathbf{O}_{(n-r)\times(n-r)}$$

令 $\mathbf{U}_1=\mathbf{A}\mathbf{V}_1\sum^{-1}$ 则 $\mathbf{U}_1^H\mathbf{A}\mathbf{V}_1=\sum$，又 $(\mathbf{A}\mathbf{V}_2)^H(\mathbf{A}\mathbf{V}_2)=\mathbf{0}\to\mathbf{A}\mathbf{V}_2=\mathbf{0}$

在 $\mathbf{U}_1$ 的基础上构造酉矩阵 $\mathbf{U}=[\mathbf{U}_1 \;|\; \mathbf{U}_2]$，即 $\mathbf{U}^H\mathbf{U}=\mathbf{I}$

这由前面基扩充定理可知是可行的，

$$\mathbf{U}_1^H\mathbf{U}_1=\mathbf{I}_r,\mathbf{U}_1^H\mathbf{U}_2=\mathbf{O}_{r\times(n-r)},\mathbf{U}_2^H\mathbf{U}_2=\mathbf{I}_{n-r}$$

故

$$\mathbf{U}^H\mathbf{A}\mathbf{V}=\begin{bmatrix}\mathbf{U}_1^H\\ \mathbf{U}_2^H\end{bmatrix}\mathbf{A}[\mathbf{V}_1 \quad \mathbf{V}_2]=\begin{bmatrix}\mathbf{U}_1^H\mathbf{A}\mathbf{V}_1 & \mathbf{U}_1^H\mathbf{A}\mathbf{V}_2\\ \mathbf{U}_2^H\mathbf{A}\mathbf{V}_1 & \mathbf{U}_2^H\mathbf{A}\mathbf{V}_2\end{bmatrix}$$

其中已知

$$\mathbf{U}_1^H\mathbf{A}\mathbf{V}_1=\sum \quad \text{而} \quad \mathbf{U}_1^H\mathbf{A}\mathbf{V}_2=\mathbf{0},\mathbf{U}_2^H\mathbf{A}\mathbf{V}_2=\mathbf{0} \quad (\because \mathbf{A}\mathbf{V}_2=\mathbf{0})$$

$$\mathbf{U}_2^H\mathbf{A}\mathbf{V}_1=\mathbf{U}_2^H(\mathbf{U}_1\sum)=(\mathbf{U}_2^H\mathbf{U}_1)\sum=\mathbf{0}$$

故定理得证。

奇异值分解的求法可按证明步骤求之。

作业： P225　1(2)，　2，　5

　　　 P233　1

# 第十三讲 Penrose 广义逆矩阵(I)

## 一、Penrose 广义逆矩阵的定义及存在性

所谓广义，即推广了原有概念或结果。我们知道，逆矩阵概念是针对非奇异的（或称为满秩的）方阵。故这一概念可推广到：（1）奇异方阵；（2）非方矩阵。事实上， Penrose 广义逆矩阵涵盖了两种情况。

对于满秩方阵 A, $A^{-1}$存在，且 $AA^{-1}=A^{-1}A=I$ 　故，当然有

$$\begin{cases} AA^{-1}A = A \\ A^{-1}AA^{-1} = A \\ (AA^{-1})^H = AA^{-1} \\ (A^{-1}A)^H = A^{-1}A \end{cases}$$

这四个对满秩方阵显然成立的等式构成了 Penrose 广义逆的启示。

1. Penrose 定义：设 $A\in C^{m\times n}$，若 $Z\in C^{n\times m}$ 且使如下四个等式成立，

$$AZA = A,\quad ZAZ = Z,\quad (AZ)^H = AZ,\quad (ZA)^H = ZA$$

则称 Z 为 A 的 Moore-Penrose(广义)逆，记为，$A^{\dagger}$。

而上述四个等式又依次称为 Penrose 方程（i）,(ii) ,(iii) ,(iv)。

2. Moore-Penrose 逆的存在性和唯一性

定理：任给 $A\in C^{m\times n}$，$A^{\dagger}$ 均存在且唯一。

证明：存在性. 　$\forall$ $A\in C_r^{m\times n}$，均存在酉矩阵 $U\in C_m^{m\times m}$，$V\in C_n^{n\times n}$ 使

$$U^H AV = D = \begin{bmatrix} \sigma_1 & & & & \vdots & \\ & \sigma_2 & & & \vdots & \\ & & \ddots & & \vdots & 0 \\ & & & \sigma_r & \vdots & \\ \cdots & \cdots & \cdots & \cdots & \vdots & \cdots \\ & 0 & & & \vdots & 0 \end{bmatrix}_{m\times n} \quad 即\ A = UDV^H$$

其中，$\sigma_1^2,\sigma_2^2,\cdots,\sigma_r^2$ 是 $A^H A$ 的全部非零特征值。

此时，令$Z=V\tilde{D}U^H \in C_r^{n\times m}$ 　　则

$$\tilde{D}=\begin{bmatrix} \sigma_1^{-1} & & & & \vdots & \\ & \sigma_2^{-1} & & & \vdots & \\ & & \ddots & & \vdots & 0 \\ & & & \sigma_r^{-1} & \vdots & \\ \cdots & \cdots & \cdots & \cdots & \vdots & \cdots \\ & & 0 & & \vdots & 0 \end{bmatrix}_{n\times m}$$

（i） $AZA=(UDV^H)(V\tilde{D}U^H)(UDV^H)=UD\tilde{D}DV^H=UDV^H=A$

（ii） $ZAZ=(V\tilde{D}U^H)(UDV^H)(V\tilde{D}U^H)=V\tilde{D}D\tilde{D}U^H=V\tilde{D}U^H=Z$

（iii） $(AZ)^H=[(UDV^H)(V\tilde{D}U^H)]^H=(UD\tilde{D}U^H)^H=UD\tilde{D}U^H=AZ$

（iv） $(ZA)^H=(V\tilde{D}DV^H)^H=V\tilde{D}DV^H=ZA$

即，$Z=A^{\dagger}$

其中 $D\tilde{D}=\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{m\times m}$ $\tilde{D}D=\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n\times n}$ $AZ=UD\tilde{D}U^H$，$ZA=V\tilde{D}DV^H$

唯一性：设 Z，Y 均满足四个 Penrose 方程，则

$$\begin{aligned} Z&=ZAZ=Z(AZ)^H=ZZ^HA^H=ZZ^H(AYA)^H=Z(AZ)^H(AY)^H=Z(AZ)(AY)\\ &=ZAY=(ZA)^HY=A^HZ^HY=A^H\ Z^H(YAY)=A^HZ^H(YA)^HY=A^HZ^HA^HY^HY\\ &=(AZA)^HY^HY=A^HY^HY=(YA)^HY=YAY=Y \end{aligned}$$

即，满足四个 Penrose 方程的 Z 是唯一的.

该证明实际上给出了 Moore-Penrose 逆的一种构造方法。由 $A^{\dagger}$ 的唯一性可知：(1)当 *A* 为满秩方阵时，$A^{\dagger}=A^{-1}$；(2) $A^{\dagger}$ 实际上还是一个限制相当严格，狭窄的量，可考虑更加放宽。

3. ｛$i,j,\cdots,l$｝-逆的定义：$\forall A\in C^{m\times n}$，若 $Z\in C^{n\times m}$ 且满足 Penrose 方程中的第 $(i),(j),\cdots,(l)$ 个方程，则称 Z 为 A 的 $\{i,j,\cdots,l\}$-逆，记为 $A^{(i,j,\cdots,l)}$，其全体记为 $A\{i,j,\cdots,l\}$。$\{i,j,\cdots,l\}$-逆共有 $C_4^1+C_4^2+C_4^3+C_4^4=15$ 类，但实际上常用的为如下 5 类：A{1}，A{1,2}，A{1,3}，A{1,4}，A{1,2,3,4}$=A^{\dagger}$

二、{1}-逆的性质

引理： $\mathrm{rank}(AB)\le\min\ (\mathrm{rank}A,\mathrm{rank}\ B\ )$

证明：矩阵的秩＝行秩＝列秩.　将$\mathbf{A}$、$\mathbf{B}$写成 $(\mathbf{A}\in\mathbf{C}^{m\times n},\ \mathbf{B}\in\mathbf{C}^{n\times p})$

$$\mathbf{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}=[a_1 \quad a_2 \quad \cdots \quad a_n]$$

$$\mathbf{B}=\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix}=\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

（1）设$\mathrm{rank}(A)=r$，则必存在 $a_{l_1},a_{l_2},\cdots,a_{l_r}$　($l_1,l_2,\cdots,l_r$两两不同）成为线性无关的向量组。所以，其它列向量$a_i$可表示为:

$$a_i=\sum_{k=1}^{r}p_{ik}a_{l_k} \qquad (i=1,2,\cdots,n)$$

$$AB=[a_1 \quad a_2 \quad \cdots \quad a_n]\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix}=[\ \sum_{i=1}^{n}b_{i1}a_i \quad \sum_{i=1}^{n}b_{i2}a_i \quad \cdots \quad \sum_{i=1}^{n}b_{ip}a_i]$$

可见 AB 的各列向量均为$a_{l_1},a_{l_2},\cdots,a_{l_r}$的线性组合。亦即

$$\mathrm{rank}(AB)\le r=\mathrm{rank}(A)$$

（2）同理。设$\mathrm{rank}(B)=s$，则必存在 $b_{m_1},b_{m_2},\cdots,b_{m_s}$　成为线性无关的向量组。所以，其它列向量$b_i$可表示为:

$$b_i=\sum_{k=1}^{s}q_{ik}b_{m_k} \qquad (i=1,2,\cdots,n)$$

$$\mathbf{AB}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}=\begin{bmatrix} \sum_{i=1}^{n} a_{1i}b_i \\ \sum_{i=1}^{n} a_{2i}b_i \\ \vdots \\ \sum_{i=1}^{n} a_{mi}b_i \end{bmatrix}$$

可见，AB 的各行向量均为 $\mathbf{b}_{m_1},\mathbf{b}_{m_2},\cdots,\mathbf{b}_{m_s}$ 的线性组合，故

$$\operatorname{rank}(AB)\le \operatorname{rank} B = s$$

合起来即 $\operatorname{rank}(AB)\le \min(\operatorname{rank}A,\operatorname{rank} B)$

定理：设 $A\in C^{m\times n}, B\in C^{n\times P}$，$\lambda\in C$，$\lambda^{\dagger}=\begin{cases}\lambda^{-1} & \lambda\ne 0\\ 0 & \lambda=0\end{cases}$ 则

（1） $(A^{(1)})^H\in A^H\{1\}$

（2） $\lambda^{\dagger}A^{(1)}\in(\lambda A)\{1\}$

（3）S、T 为可逆方阵且与 A 可乘，则

$$T^{-1}A^{(1)}S^{-1}\in(SAT)\{1\}\ ,\ (S\in C_m^{m\times m},T\in C_n^{n\times n})$$

（4） $\operatorname{rank}（A^{(1)})\ge \operatorname{rank}A$

（5） $AA^{(1)}$和$A^{(1)}A$均为幂等矩阵且与$A$同秩 （$P^2=P$）

（6） $R(AA^{(1)})=R(A),\quad N(A^{(1)}A)=N(A),\quad R((A^{(1)}A)^H)=R(A^H)$

（7） $A^{(1)}A=I_n \iff \operatorname{rank}(A)=n$

$$AA^{(1)}=I_m \iff \operatorname{rank}(A)=m$$

（8） $AB(AB)^{(1)}A=A \iff \operatorname{rank}(AB)=\operatorname{rank}(A)$

$B(AB)^{(1)}AB=B \iff \operatorname{rank}(AB)=\operatorname{rank}(B)$

证明：（1） $A^H(A^{(1)})^HA^H=(AA^{(1)}A)^H=A^H \quad \to (A^{(1)})^H\in A^H\{1\}$

（2） $\lambda=0$ 时，$\lambda A=0_{m\times n}$，$\lambda^{\dagger}A^{(1)}=0_{n\times m}$. 显然成立.

$\lambda\ne 0$ 时，$(\lambda A)(\lambda^{\dagger}A^{(1)})(\lambda A)=(\lambda\lambda^{-1}\lambda)(AA^{(1)}A)=\lambda A$

(3) $(SAT)(T^{-1}A^{(1)}S^{-1})(SAT)=S(AA^{(1)}A)T=SAT$

(4) $rank(A)=rank(AA^{(1)}A)\le rank(A^{(1)})$

(5) $AA^{(1)}A=A \quad \to \begin{cases} AA^{(1)}A\cdot A^{(1)}=A\cdot A^{(1)} & \to (AA^{(1)})^2=AA^{(1)} \\ A^{(1)}\cdot AA^{(1)}A=A^{(1)}\cdot A & \to (A^{(1)}A)^2=A^{(1)}A \end{cases}$

又 $$\begin{aligned} & rank(A)=rank(AA^{(1)}A)\le rank(AA^{(1)})\le rank(A) \\ & \to rank(AA^{(1)})=rank(A) \end{aligned}$$

同理，$rank(A^{(1)}A)=rank(A)$

(6) $\bullet\ R(A)=\{Ax \mid x\in C^n\}\subset C^m$，$R(AA^{(1)})=\{AA^{(1)}y \mid y\in C^m\}\subset C^m$

$\Rightarrow R(A)\supseteq R(AA^{(1)})\supseteq R(AA^{(1)}A) \quad \to R(A)=R(AA^{(1)})$

又法：将　写成 $A=[a_1a_2\cdots a_n],\ \ AA^{(1)}=[b_1b_2\cdots b_m].$

$a_i,b_j$ 均为 m 维列向量，则

$$R(A)=span\{a_1,a_2,\cdots,a_n\}=\{\sum_{i=1}^{n}\xi_i a_i \mid \xi_i\in C\} \quad \to dim\ R(A)=rank(A)$$

$$R(AA^{(1)})=span\{b_1,b_2,\cdots,b_m\}=\{\sum_{i=1}^{m}\eta_i b_i \mid \eta_i\in C\}$$

$$\to dim\ R(AA^{(1)})=rank(AA^{(1)})$$

即 $dim\ R(A)=dim\ R(AA^{(1)})$ 　且 $R(A)\supseteq R(AA^{(1)})$

故 $R(A)=R(AA^{(1)})$

$\bullet N(A)=\{x \mid Ax=0, x\in C^n\}\subseteq C^m$，$N(A^{(1)}A)=\{x \mid A^{(1)}Ax=0\ , x\in C^n\}\subseteq C^n$

$\Rightarrow \ N(A)\subseteq N(A^{(1)}A)\subseteq N(AA^{(1)}A)=N(A) \quad \to N(A)=N(A^{(1)}A)$

又法：$dim\ N(A)=n-rank(A)=n-rank(A^{(1)}A)=dim\ N(A^{(1)}A)$

又 $N(A)\subseteq N(A^{(1)}A)$ 　故 $N(A)=N(A^{(1)}A)$

在 $R(AA^{(1)})=R(A)$ 中，将 $A$ 换为 $A^H$，$A^{(1)}$ 换为 $(A^{(1)})^H$，则有

$$R(A^H)=R(A^H(A^{(1)})^H)=R((A^{(1)}A)^H)$$

(7) 以 $AA^{(1)}=I_m \quad \Leftrightarrow \quad rankA=m$ 为例.

$$\Rightarrow: \quad \mathrm{rank}A = \mathrm{rank}(AA^{(1)}) = \mathrm{rank}(I_m) = m.$$

$$\Leftarrow: \quad \mathrm{rank}(AA^{(1)}) = \mathrm{rank}A = m$$

即 $AA^{(1)}$ 为 m 阶满秩可逆方阵, $(AA^{(1)})^{-1}$ 存在。

又 $AA^{(1)}$ 幂等：（$AA^{(1)})^2 = AA^{(1)}$，乘以（$AA^{(1)})^{-1}$，得 $AA^{(1)} = I_m$

(8) $R(A) = \{Ax \mid x \in C^n\} \subseteq C^m$

$$R(AB) = \{ABy \mid y \in C^p\} \subseteq C^m \quad \to R(A) \supseteq R(AB)$$

•对 $AB(AB)^{(1)}A = A \quad \Leftrightarrow \quad \mathrm{rank}(AB) = \mathrm{rank}(A)$

$$\Rightarrow: \quad \mathrm{rank}A = \mathrm{rank}(AB(AB)^{(1)}A) \le \mathrm{rank}(AB) \le \mathrm{rank}(A)$$
$$\to \quad \mathrm{rank}(AB) = \mathrm{rank}A$$

$$\Leftarrow: \quad \mathrm{rank}A = \dim R(A), \quad \mathrm{rank}(AB) = \dim R(AB)$$
故 $R(A) = R(AB)$

即，$\forall x \in C^n, \exists y \in C^p$，使 $Ax = ABy$. 故

$$Ax = ABy = AB(AB)^{(1)}ABy = AB(AB)^{(1)}Ax$$
$$(\text{注意} \forall x \in C^n) \to AB(AB)^{(1)}A = A$$

• 对 $B(AB)^{(1)}AB = B \quad \Leftrightarrow \quad \mathrm{rank}(AB) = \mathrm{rank}(B)$

$$\Rightarrow: \quad \mathrm{rank}B = \mathrm{rank}(B(AB)^{(1)}AB) \le \mathrm{rank}(AB) \le \mathrm{rank}(B)$$
$$\to \quad \mathrm{rank}(AB) = \mathrm{rank}B$$

$$\Leftarrow: \quad R(B(AB)^{(1)}AB) = \{B(AB)^{(1)}ABy \mid y \in C^p\} \subseteq R(B) = \{Bx \mid x \in C^p\}$$

又，$\mathrm{rank}B = \mathrm{rank}(AB) = \mathrm{rank}(AB(AB)^{(1)}AB) \le \mathrm{rank}(B(AB)^{(1)}AB) \le \mathrm{rank}B$

$$\to \mathrm{rank}B = \mathrm{rank}(B(AB)^{(1)}AB) \quad \Rightarrow \quad R(B) = R(B(AB)^{(1)}AB)$$

即，$\forall x \in C^p$， $\exists y \in C^p$，使 $B(AB)^{(1)}ABy = Bx$. 故

$$Bx = B(AB)^{(1)}ABy = B(AB)^{(1)}AB(AB)^{(1)}ABy = B(AB)^{(1)}ABx$$

定理：矩阵 A 当且仅当 A 为满秩方阵时具有唯一的{1}逆，此时

$$A^{(1)} = A^{-1}$$

作业：P306　　3，4，5

# 第十四讲　Penrose 广义逆（$\Pi$）

## 一、$\{1\}$-逆与$\{1,2\}$-逆

定理 1：设 Y，Z $\in A\{1\}$，则 $YAZ\in A\{1,2\}$.

证：已知 AYA = AZA = A　故

(i)　A(YAZ)A = AZA = A；

(ii)　(YAZ)A(YAZ)=YAYAZ=YAZ .

定理 2：给定矩阵 A 及 $Z\in A\{1\}$，则 $Z\in A\{1,2\}$的充要条件是

$$\mathbf{rankA = rankZ}$$

证：必要性.　$Z\in A\{1,2\}$　则　(i) AZA=A;　(ii) ZAZ=Z

$\rightarrow A\in Z\{1,2\}$

而由 $\mathrm{rank}A^{(1)}\geq \mathrm{rank}A$ 可知 $\mathrm{rank}Z\geq \mathrm{rank}A$，$\mathrm{rank}A\geq \mathrm{rank}Z$

$\Rightarrow$ rankZ=rankA

充分性.　因为 $R(ZA)\subseteq R(Z)$，　而 rankZ=rankA，$Z\in A\{1\}$

故　rank(ZA)=rank(A)=rank(Z)　$\rightarrow$ R(ZA)=R(Z)

$\forall\, \mathbf{e}\in C^{\mathbf{m}}$，$\exists\, \mathbf{u}\in C^{\mathbf{n}}$，使 $ZA\mathbf{u}=Z\mathbf{e}$

$\rightarrow$　$ZA[\mathbf{u_1}\,\mathbf{u_2}\cdots\mathbf{u_m}]=Z[\mathbf{e_1}\mathbf{e_2}\cdots\mathbf{e_m}]$

$[\mathbf{e_1}\mathbf{e_2}\cdots\mathbf{e_m}]=I_{\mathbf{m}}$，　$[\mathbf{u_1}\,\mathbf{u_2}\cdots\mathbf{u_m}]=U$（$\mathbf{u_i}$=n 维，$\mathbf{e_j}$＝m 维）

$\Rightarrow\exists$ U 使 Z=ZAU

故　ZAZ=ZA(ZAU)=ZAU=Z　$\rightarrow$ Z 满足 Penrose 方程(ii)

可见，$Z\in A\{1,2\}$.

## 二、$\{1\}$-逆与$\{1,2,3\}$-逆、$\{1,2,4\}$-逆

引理：对任意矩阵 A 均有　$\mathrm{rank}(A^{\mathbf{H}}A)$ = rankA = $\mathrm{rank}(AA^{\mathbf{H}})$

证：　$\forall\, \mathbf{x}\in N(A)$，即 $A\mathbf{x}=0$，则 $A^{\mathbf{H}}A\mathbf{x}=0$　$\rightarrow N(A)\subseteq N(A^{\mathbf{H}}A)$

另一方面 $\forall\, \mathbf{x}\in$ $\mathbf{N(A^H A)}$，则

$$\mathbf{x^H A^H A x}=0=\left(\mathbf{Ax}\right)^{\mathbf{H}}\left(\mathbf{Ax}\right) \Rightarrow \mathbf{Ax=0} \to \mathbf{N(A^H A)} \subseteq \mathbf{N}\left(\mathbf{A}\right)$$

$\therefore \mathbf{N(A^H A)=N}\left(\mathbf{A}\right)$，又 $\mathbf{A^H A}$ 与 A 的列数均为 n，

dimN(A)=n－rankA， dim $\mathbf{N(A^H A)}$=n－rank$\mathbf{(A^H A)}$

$\Rightarrow$ rank$\mathbf{(A^H A)}$=rankA .

A$\leftrightarrow \mathbf{A^H}$，则 rank$\mathbf{(AA^H)}$=rank$\mathbf{A^H}$=rankA.

定理 3： 给定矩阵 A， 则 Y=$\mathbf{(A^H A)^{(1)} A^H} \in$A$\left\{1,2,3\right\}$

Z=$\mathbf{A^H (AA^H)^{(1)}} \in$A$\left\{1,2,4\right\}$

证：显然 R$\mathbf{(A^H A)} \subseteq$R（$\mathbf{A^H}$），又由引理可知 R$\mathbf{(A^H A)}$=R（$\mathbf{A^H}$），

即存在 U 使 $\mathbf{A^H}=\mathbf{A^H A}$U $\to$ A=$\mathbf{U^H A^H A}$

AYA=（$\mathbf{U^H A^H A}$）[$\mathbf{(A^H A)^{(1)} A^H}$]A$\overset{\mathbf{(i)}}{=}\mathbf{U^H A^H A}$=A 满足(i) $\to$Y$\in$A$\left\{1\right\}$

可见 rankY$\geq$rankA

但 rankY=rank$\left(\left(\mathbf{A^H A}\right)^{(1)}\mathbf{A^H}\right)\leq$ rank$\mathbf{A^H}$=rankA.

即 rankY=rankA. $\to$Y$\in$ A$\left\{1,2\right\}$

AY=（$\mathbf{U^H A^H A}$）$\mathbf{(A^H A)^{(1)} A^H}$=$\mathbf{U^H A^H A\ (A^H A)^{(1)} A^H}$AU

=$\mathbf{U^H\ (A^H A)U}$=$\mathbf{(AY)^{(H)}}$

$\Rightarrow$ Y$\in$ A$\left\{3\right\}$ 综合之，即 Y$\in$A$\left\{1,2,3\right\}$

同理可证另式。

三、关于 $\boldsymbol{A}^{+}$

定理 4：给定矩阵 A， $\boldsymbol{A}^{+}$=$\mathbf{A^{(1,4)}\ AA^{(1,3)}}$

证：(1)由定理 1 知， $\mathbf{A^{(1,4)}\ AA^{(1,3)}} \overset{\Delta}{=}$ X$\in$A$\left\{1,2\right\}$

(2) AX=$\mathbf{A\ A^{(1,4)}\ AA^{(1,3)}} \overset{\mathbf{i}}{=} \mathbf{AA^{(1,3)}} \overset{\mathbf{iii}}{=}$ （$\mathbf{AA^{(1,3)}}$）$^{\mathbf{H}}$=$\left(\mathbf{AX}\right)^{\mathbf{H}}$

(3) XA=$\mathbf{A^{(1,4)}\ AA^{(1,3)}\ A} \overset{\mathbf{i}}{=} \mathbf{A^{(1,4)}\ A} \overset{\mathbf{iv}}{=}$ （$\mathbf{A^{(1,4)}A)^H}=\left(\mathbf{XA}\right)^{\mathbf{H}}$

$\Rightarrow$ X$\in$A$\{1,2,3,4\}$=$A^+$

定理 5： 给定矩阵 A，则

(1) rank $A^+$=rankA

(2) $(A^+)^+$=A

(3) $(A^H)^+=(A^+)^H$ ， $(A^T)^+=(A^+)^T$

(4) $(A^H A)^+=A^+(A^H)^+$ ， $(A\ A^H)^+=(A^H)^+A^+$

(5) $A^+=(A^H A)^+A^H=A^H(AA^H)^+$

(6) $R(A^+)=R(A^H),N(A^+)=N(A^H)$

证：(1) $A^+\in$A$\{1,2\}$ $\rightarrow$ rank $A^+$=rankA

(2) Penrose 方程中 (i)$\leftrightarrow$(ii), (iii)$\leftrightarrow$(iv) 互为对称

故 $(A^+)^+$=A.

(3) 直接采用四个方程验证即可。

(4) 同上。

(5) 证 X=$(A^H A)^+A^H$ ， 由定理 3 知 X$\in$A$\{1,2,3\}$ ,且

$$XA=(A^H A)^+A^H A=\underset{\substack{\uparrow \\ (A^H A)^+\text{当然是}(A^H A)^{(4)}}}{((A^H A)^+A^H A)^H}=(XA)^H$$

$\Rightarrow$X$\in$A$\{1,2,3,4\}$

另式同理可证。

(6) R($A^+$)$\overset{(5)}{=}$ R($A^H(AA^H)^+$)$\subseteq$R($A^H$)

N($A^+$)=$N((A^H A)^+A^H)\supseteq$ N($A^H$) ,而 rank $A^+$=rankA＝rank $A^H$

$\therefore$ R($A^+$)=R($A^H$) ,N($A^+$)=N($A^H$)

推论 1：若 $A\in C_n^{m\times n}$(列满秩矩阵) ， 则 $A^+=(A^H A)^{-1}A^H$

$\mathbf{A} \in \mathbf{C}_{\mathbf{m}}^{\mathbf{m\times n}}$（行满秩矩阵） ，则 $\mathbf{A}^{+} = \mathbf{A}^{\mathrm{H}}(\mathbf{A}\mathbf{A}^{\mathrm{H}})^{-1}$

推论 2： 对非零列向量 $\alpha$，$\alpha^{+} = \left(\alpha^{\mathbf{H}}\alpha\right)^{-1}\alpha^{\mathbf{H}}$；

对非零行向量 $\beta$， $\beta^{+} = \beta^{\mathbf{H}}\left(\beta\beta^{\mathbf{H}}\right)^{-1}$；$\alpha^{\mathbf{H}}\alpha, \beta\beta^{\mathbf{H}}$ 均为数。

A,B 可逆，则 $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$，但一般 $(\boldsymbol{AB})^{+} \neq \boldsymbol{B}^{+}\boldsymbol{A}^{+}$

如 $\mathbf{A} = [1 \quad 0]$，$\mathbf{B} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$， $\mathbf{AB} = [1]$， BA=$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$

$(\boldsymbol{AB})^{+}$=[1] ， $\boldsymbol{A}^{+} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$， $\boldsymbol{B}^{+} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}$， $\boldsymbol{B}^{+}\boldsymbol{A}^{+} = \begin{bmatrix} \frac{1}{2} \end{bmatrix}$

作业 ：P 306-307 6、8、11、12

第十五讲 投影矩阵与Moore-Penrose逆

一、投影算子与投影矩阵

设L，M为$\mathbf{C}^n$的子空间并构成直和$\mathbf{L}+\mathbf{M}=\mathbf{L}\oplus\mathbf{M}=\mathbf{C}^n$.即

$\forall \mathbf{x}\in\mathbf{C}^n$，$\exists$唯一的$\mathbf{y}\in\mathbf{L},\mathbf{z}\in\mathbf{M}$使x=y+z

称y为x沿着M到L的投影。

1. 定义：将任意$\mathbf{x}\in\mathbf{C}^n$变为其沿着M到L的投影的变换称为沿着M到L的投影算子，记为$\mathbf{P}_{\mathbf{L},\mathbf{M}}$即$\mathbf{P}_{\mathbf{L},\mathbf{M}}\ \mathbf{x}=\mathbf{y}\in\mathbf{L}$,投影算子是线性变换，其矩阵称为投影矩阵,仍记为$\mathbf{P}_{\mathbf{L},\mathbf{M}}$。

2. 充要条件

引理：设n阶方阵E为幂等矩阵，则$\mathbf{N(E)}=\mathbf{R(I-E)}$

证明：

$$\begin{aligned}&\because \mathbf{E}^2=\mathbf{E}\to\mathbf{E(I-E)}=\mathbf{O}\to\forall\mathbf{x}\in\mathbf{C}^n,\mathbf{E[(I-E)x]}=\mathbf{0}\\&\quad\to\mathbf{E[R(I-E)]}=\mathbf{0}\Rightarrow\mathbf{R(I-E)}\subseteq\mathbf{N(E)}\end{aligned}$$

另一方面，$\forall\mathbf{x}\in\mathbf{N(E)},$即$\mathbf{Ex}=\mathbf{0}$，则

$$\begin{aligned}&\mathbf{x}=\mathbf{Ix}-O=\mathbf{Ix}-\mathbf{Ex}=\mathbf{(I-E)x}\in\mathbf{R}(\mathbf{I-E})\\&\Rightarrow\mathbf{N}(\mathbf{E})\subseteq\mathbf{R}(\mathbf{I-E})\\&\therefore\mathbf{N}(\mathbf{E})=\mathbf{R}(\mathbf{I-E})\end{aligned}$$

定理：n阶方阵P成为投影矩阵的充要条件是P为幂等矩阵。

证明：充分性

$$\mathbf{P}^2=\mathbf{P},\forall\mathbf{x}\in\mathbf{C}^n,\text{令}\mathbf{y}=\mathbf{Px}\in\mathbf{R(P)},\mathbf{z}=\mathbf{(I-P)x}\in\mathbf{R}(\mathbf{I-P})=\mathbf{N}(\mathbf{P})\text{。}$$

若 $\mathbf{R}(\mathbf{P})\bigcap\mathbf{N}(\mathbf{P})=\{\mathbf{0}\},\text{则}\mathbf{P}=\mathbf{P}_{\mathbf{R}(\mathbf{P}),\mathbf{N}(\mathbf{P})}$确为投影矩阵，下面证之

$$\forall\mathbf{x}\in\mathbf{R(P)}\bigcap\mathbf{N(P)},$$

一方面，因$\mathbf{x}\in\mathbf{R(P)},\text{存在}\mathbf{u}\in\mathbf{C}^n\text{使}\mathbf{x}=\mathbf{Pu}$

$$\begin{aligned}&\text{另一方面}\mathbf{x}\in\mathbf{N(P)},\text{即}\mathbf{Px}=\mathbf{0}.\text{但}\mathbf{Px}=\mathbf{P}^2\mathbf{u}=\mathbf{Pu}=\mathbf{x}\to\mathbf{x}=\mathbf{0}\\&\qquad\Rightarrow\mathbf{R(P)}\bigcap\mathbf{N(P)}=\{\mathbf{0}\}\text{。}\end{aligned}$$

必要性 $\mathbf{P}=\mathbf{P}_{\mathbf{L},\mathbf{M}}$ 故 $\forall\mathbf{x}\in\mathbf{C}^n$，$\exists$唯一分解 $\mathbf{y}\in\mathbf{L},\mathbf{z}\in\mathbf{M}$ 使

$$\mathbf{x}=\mathbf{y}+\mathbf{z}\text{且}\ \mathbf{Px}=\mathbf{y}$$

$$\rightarrow P^2x = Py = y = Px \xrightarrow{x任意} P^2 = P$$

$$\underset{y = y + 0}{\uparrow}$$

3. 投影矩阵的构造

设已知 $C^n$ 的子空间 L、M 构成直和 $L \oplus M = C^n$，下面构造 $P_{L,M}$。

取 L 的一个基 $\left\{x_1, x_2 \cdots x_r\right\}$（设 L 为 $r$ 维子空间），M 的一个基 $\left\{y_1, y_2 \cdots y_{n-r}\right\}$（则 M 的维数为 n-r）。由直和关系知 $\left\{x_1, x_2 \cdots x_r;\ y_1, y_2 \cdots y_{n-r}\right\}$ 即构成 $C^n$ 的一个基。故，如令

$$X = \left[x_1, x_2 \cdots x_r\right], \quad Y = \left[y_1, y_2 \cdots y_{n-r}\right]$$

则[X　Y]为可逆方阵。另一方面

$$x_i \in L \rightarrow P_{L,M}x_i = x_i; y_i \in M \rightarrow P_{L,M}y_i = 0$$

即 $$P_{L,M}\left[X\ Y\right] = \left[X\ O\right] \rightarrow P_{L,M} = \left[X\ O\right]\left[X\ Y\right]^{-1}$$

可见，$P_{L,M}$ 的秩为 r（$\mathrm{rank}(P_{L,M}) = \dim R(P_{L,M}) = \dim L$）

二、正交投影算子与正交投影矩阵

L 为 $C^n$ 的子空间，其正交补空间 $L^{\perp} = \left\{x \middle| (x, y) = 0, x \in C^n, y \in L\right\}$（无特别声明取 $x^H y$）

1. 定义：设 L 是 $C^n$ 的子空间，则称沿着 $L^{\perp}$ 到 L 的投影算子 $P_{L,L^{\perp}}$ 为正交投影算子，简记为 $P_L$。正交投影算子的矩阵称为正交投影矩阵，仍记为 $P_L$。

2. 充要条件：n 阶方阵 P 为正交投影矩阵的充要条件是 P 为幂等的厄米矩阵。

证明：首先证明两个引理:

（1）对 n 阶方阵 A，$\forall x \in C^n$ 均有 $x^H Ax = 0$ 则 A=0，

（2）N（$P^H$）=$R^{\perp}(P)$

(1)证明：设 A=$(a_{ij})_{n\times n}$ 取 $x = \left[0 \cdots 0 \cdots 1_{(第i个)} \cdots 0\right]^T$ 则 $x^H Ax = a_{ii} = 0$

再取 $x=\left[0\cdots0,\ \xi_i,0\cdots0,\ \xi_j,0\cdots0\right]^T(i\neq j)$，注意到$a_{ii}=0$,

则$x^HAx=\overline{\xi_i}a_{ij}\xi_j+\overline{\xi_j}a_{ji}\xi_i$

$$\rightarrow\begin{cases}\xi_i=\xi_j=1,则a_{ij}+a_{ji}=0\\ \xi_i=1,\xi_j=\sqrt{-1},则a_{ij}-a_{ji}=0\end{cases}\rightarrow a_{ij}=a_{ji}=0统一考虑即A=0$$

（2）证明：$\forall x\in N(P^H),即P^Hx=0\rightarrow x^HP=0\rightarrow$

$\forall y\in C^n,\ x^H(Py)=0$

由y的任意性，知$x\perp R(P)\Rightarrow N(P^H)\subseteq R^{\perp}(P)$

另一方面，设$x\in R^{\perp}(P)即\forall y\in C^n,均有x^H(Py)=0\rightarrow x^HP=0$

$$\rightarrow P^Hx=0\rightarrow x\in N(P^H)\Rightarrow N(P^H)\supseteq R^{\perp}(P)$$

所以 $N(P^H)=R^{\perp}(P)$

现在证明该充要条件。

充分性：

$$P^2=P,\ P^H=P\rightarrow P=P_{R(p),N(p)}=P_{R(p),N(p^H)}=P_{R(p),R^{\perp}(p)}=P_{R(p)}$$

必要性：

$$P=P_L$$

$$\forall x\in C^n,可唯一地分解成y=Px\in L,z=(I-P)x\in L^{\perp}使x=y+z$$

又 $y\in L,z\in L^{\perp}\rightarrow y^Hz=0\rightarrow x^HP^H(I-P)x=0\xrightarrow{x任意}P^H(I-P)=0$

$P^H=P^HP=(P^HP)^H=(P^H)^H=P$， $P$ 为厄米矩阵。

幂等已由上一定理得知。

3. 正交投影矩阵的构造

设 L 的一个基为$\{x_1,x_2\cdots x_r\}$，$L^{\perp}$ 的一个基为$\{y_1,y_2\cdots y_{n-r}\}$。则$x_i^Hy_j=0$

令$X=[x_1,x_2\cdots x_r]$，$Y=[y_1,y_2\cdots y_{n-r}]$ 则$X^HY=0$

$$(A^{-1}=(A^HA)^{-1}A^H)$$

$$P_L=[X\ O][X\ Y]^{-1}=[X\ O]\{[X\ Y]^H[X\ Y]\}^{-1}[X\ Y]^H$$

$$=[X\ O]\begin{bmatrix} X^HX & X^HY \\ Y^HX & Y^HY \end{bmatrix}^{-1}[X\ Y]^H=[X\ O]\begin{bmatrix} X^HX & 0 \\ 0 & Y^HY \end{bmatrix}^{-1}\begin{bmatrix} X^H \\ Y^H \end{bmatrix}$$

$$=\begin{bmatrix} X(X^HX)^{-1} & O \end{bmatrix}\begin{bmatrix} X^H \\ Y^H \end{bmatrix}=X(X^HX)^{-1}X^H$$

三、投影矩阵与广义逆矩阵

(i) $AXA=A\rightarrow AX=P_{R(A),N(AX)}$

（ii）$XAX=X\rightarrow XA=P_{R(XA),N(A)}$

$$\begin{cases} AXA=A \\ (AX)^H=AX \end{cases}\rightarrow AX=P_{R(A)}$$

$$\begin{cases} XAX=X \\ (XA)^H=XA \end{cases}\rightarrow XA=P_{R(XA)}$$

Moore 定义：设 $A\in C^{m\times n},X\in C^{n\times m}$ 且 $AX=P_{R(A)}$， $XA=P_{R(A)}$

则 X 为 A 的 Moore 广义逆矩阵。事实上，Moore 广义逆矩阵正是 $A^\dagger$

作业 P295 1、4

第十六讲 广义逆的计算及应用

一、 由Hermite标准形求{1}-逆

任何矩阵都可由初等行变换化为Hermite标准形。设$\mathbf{A}\in\mathbf{C}_r^{m\times n}$，存在满秩矩阵$\mathbf{E}\in\mathbf{C}_m^{m\times m}$，使$\mathbf{EA}=\mathbf{B}$（Hermite标准形），采用置换矩阵$\mathbf{P}$：

$$\mathbf{P}=\begin{bmatrix}\mathbf{e}_{l_1} & \mathbf{e}_{l_2} & \cdots & | & \text{其它}\mathbf{e}_i\end{bmatrix}_{n\times n}$$

$$\mathbf{EAP}=\begin{bmatrix}\mathbf{I}_r & \mathbf{K}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}$$

$$\mathbf{A}=\mathbf{E}^{-1}\begin{bmatrix}\mathbf{I}_r & \mathbf{K}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{P}^{-1}$$

1. 求{1}-逆的方法

$$\mathbf{A}\{1\}=\left\{\mathbf{P}\begin{bmatrix}\mathbf{I}_r & \mathbf{M}\\ \mathbf{N} & \mathbf{L}\end{bmatrix}_{n\times m}\mathbf{E}\,\middle|\,\mathbf{KN}=\mathbf{0}\right\}$$ （取阶数合适的M、L）

[证明]令$\mathbf{X}=\mathbf{P}\begin{bmatrix}\mathbf{I}_r & \mathbf{M}\\ \mathbf{N} & \mathbf{L}\end{bmatrix}\mathbf{E}$，则

$$\begin{aligned}\mathbf{AXA}&=\mathbf{E}^{-1}\begin{bmatrix}\mathbf{I}_r & \mathbf{K}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{P}^{-1}\mathbf{P}\begin{bmatrix}\mathbf{I}_r & \mathbf{M}\\ \mathbf{N} & \mathbf{L}\end{bmatrix}\mathbf{E}\mathbf{E}^{-1}\begin{bmatrix}\mathbf{I}_r & \mathbf{K}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{P}^{-1}\\
&=\mathbf{E}^{-1}\begin{bmatrix}\mathbf{I}_r+\mathbf{KN} & \mathbf{M}+\mathbf{KL}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{I}_r & \mathbf{K}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{P}^{-1}\\
&=\mathbf{E}^{-1}\begin{bmatrix}\mathbf{I}_r+\mathbf{KN} & (\mathbf{I}_r+\mathbf{KN})\mathbf{K}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{P}^{-1}\\
&=\mathbf{E}^{-1}\begin{bmatrix}\mathbf{I}_r & \mathbf{K}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{P}^{-1}\\
&=\mathbf{A}\end{aligned}$$

2. {1,2}-逆

当$\mathbf{X}\in\mathbf{A}\{1\}$时，由定理可知：$\mathbf{rankX}=\mathbf{rankA}$是$\mathbf{X}\in\mathbf{A}\{1,2\}$的充要条件。

$\mathbf{X}=\mathbf{P}\begin{bmatrix}\mathbf{I}_r & \mathbf{M}\\ \mathbf{N} & \mathbf{L}\end{bmatrix}\mathbf{E}$，$\mathbf{P}$、$\mathbf{E}$为满秩方阵

$$\therefore\quad \mathbf{rankX}=\mathbf{rank}\begin{bmatrix}\mathbf{I}_r & \mathbf{M}\\ \mathbf{N} & \mathbf{L}\end{bmatrix}=\mathbf{rankA}=\mathbf{r}$$

$$\begin{bmatrix}\mathbf{I}_r & \mathbf{M}\\ \mathbf{N} & \mathbf{L}\end{bmatrix}\sim\begin{bmatrix}\mathbf{I}_r & \mathbf{M}\\ \mathbf{0} & \mathbf{L}-\mathbf{NM}\end{bmatrix}\to\mathbf{L}-\mathbf{NM}=\mathbf{0}$$

$$\therefore\ A\{1,2\}=\left\{P\begin{bmatrix} I_r & M \\ N & L \end{bmatrix}_{n\times m}E \,\middle|\, KN=0, L=NM\right\}$$

二、　由满秩分解求广义逆

对 A 进行满秩分解：$A=FG$，$A\in C_r^{m\times n}$，$F\in C_r^{m\times r}$，$G\in C_r^{r\times n}$

[定理] 设 $A\in C_r^{m\times n}$，其满秩分解为 $A=FG$，则

（1）$G^{(i)}F^{(1)}\in A\{i\} \quad i=1,2,4$

（2）$G^{(1)}F^{(i)}\in A\{i\} \quad i=1,2,3$

（3）$G^{(1)}F^{+}\in A\{1,2,3\}$，$G^{+}F^{(1)}\in A\{1,2,4\}$

（4）$A^{+}=G^{+}F^{(1,3)}=G^{(1,4)}F^{+}$

（5）$A^{+}=G^{+}F^{+}=G^{H}(GG^{H})^{-1}(F^{H}F)^{-1}F^{H}=G^{H}(F^{H}AG^{H})^{-1}F^{H}$

证明思路：（1）（2）代入相应的 Penrose 方程即可证之，

由（1）（2）$\Rightarrow$（3）$\Rightarrow$（4）$\Rightarrow$（5）

三、　矩阵方程 $AXB=D$ 的相容性条件及通解

定理 1. 矩阵方程 $AXB=D$ 相容（有解）的充要条件：$AA^{(1)}DB^{(1)}B=D$

在相容情况下矩阵方程的通解为：

$$\left\{A^{(1)}DB^{(1)}+Y-A^{(1)}AYBB^{(1)} \mid Y\text{为阶数合适的任意矩阵}\right\}$$

[证明] 相容性条件的充分性：

已知 $AA^{(1)}DB^{(1)}B=D$，显然有解 $X=A^{(1)}DB^{(1)}$

相容性条件的必要性：已知 $AXB=D$ 有解，设某个解为 $X$，即

$$D=AXB=AA^{(1)}AXBB^{(1)}B=AA^{(1)}DB^{(1)}B$$

现在证明通解："通解"有两个含义：（1）解集合中的任何元素为方程的解；（2）方程的任何解均可由集合中的元素表现出来。

（1）　令 $X=A^{(1)}DB^{(1)}+Y-A^{(1)}AYBB^{(1)}$

$$AXB=D+AYB-AYB=D$$

$\therefore$ 集合中的元素为方程的解

（2）　设 $X$ 为方程的解，即 $AXB=D$

$$X=A^{(1)}DB^{(1)}+X-A^{(1)}DB^{(1)}=A^{(1)}DB^{(1)}+X-A^{(1)}AXBB^{(1)}$$

对应于集合中 $Y=X$ 的情况。

[得证]

由上述证明可见：（1）通解中两个$\mathbf{A}^{(1)}$及两个$\mathbf{B}^{(1)}$完全可以不同。

（2）通解集合中，不同的$\mathbf{Y}$完全可能对应同一个解。

推论 1. 线性方程组$\mathbf{Ax}=\mathbf{b}$有解的充要条件为：$\mathbf{AA}^{(1)}\mathbf{b}=\mathbf{b}$

且通解为$\left\{\mathbf{A}^{(1)}\mathbf{b}+(\mathbf{I}_n-\mathbf{A}^{(1)}\mathbf{A})\mathbf{y}\,|\,\mathbf{y}\text{为列向量}\right\}$

推论 2. $\mathbf{A}\{1\}$ $(\mathbf{AXA}=\mathbf{A}\text{的解})$为如下集合：

$$\left\{\mathbf{A}^{(1)}\mathbf{AA}^{(1)}+\mathbf{Y}-\mathbf{A}^{(1)}\mathbf{AYAA}^{(1)}\right\}\qquad \text{（四个}\mathbf{A}^{(1)}\text{可互不相同）}$$

四、 极小范数解

在方程有解时，完全可能是具有无穷多个解，实际中常常希望研究其中具有特定性质的解，例如范数最小的解，即极小范数解。

引理 1. 方程$\mathbf{Ax}=\mathbf{b}$若有解，则必存在唯一的极小范数解（对 2-范数），且该解在$\mathbf{R}(\mathbf{A}^H)$中。

[证明] 设$\mathbf{x}$是方程$\mathbf{Ax}=\mathbf{b}$的解，可将其分解为$\mathbf{x}=\mathbf{x}_0+\mathbf{y}$，其中

$$\mathbf{x}_0\in\mathbf{R}(\mathbf{A}^H)=\mathbf{N}^{\perp}(\mathbf{A})\to\mathbf{x}_0\perp\mathbf{N}(\mathbf{A}),\ \ \mathbf{y}\in\mathbf{N}(\mathbf{A})$$

$$\|\mathbf{x}\|_2^2=\|\mathbf{x}_0+\mathbf{y}\|_2^2=(\mathbf{x}_0+\mathbf{y})^H(\mathbf{x}_0+\mathbf{y})=\mathbf{x}_0^H\mathbf{x}_0+\mathbf{y}^H\mathbf{y}\ =\|\mathbf{x}_0\|_2^2+\|\mathbf{y}\|_2^2\ge\|\mathbf{x}_0\|_2^2$$

而$\mathbf{Ax}=\mathbf{Ax}_0+\mathbf{Ay}=\mathbf{Ax}_0+\mathbf{0}=\mathbf{Ax}_0=\mathbf{b}$

即：$\mathbf{x}_0$也是方程的解，也就是$\mathbf{R}(\mathbf{A}^H)$中存在$\mathbf{Ax}=\mathbf{b}$的解。

假设$\mathbf{R}(\mathbf{A}^H)$中存在方程$\mathbf{Ax}=\mathbf{b}$的两个解$\mathbf{x}_1$和$\mathbf{x}_2$，即$\mathbf{Ax}_1=\mathbf{Ax}_2=\mathbf{b}$

$\to\mathbf{Ax}_1-\mathbf{Ax}_2=\mathbf{0}\to(\mathbf{x}_1-\mathbf{x}_2)\in\mathbf{N}(\mathbf{A})$ 同时$(\mathbf{x}_1-\mathbf{x}_2)\in\mathbf{N}^{\perp}(\mathbf{A})$

$\therefore\qquad(\mathbf{x}_1-\mathbf{x}_2)\in\mathbf{N}(\mathbf{A})\bigcap\mathbf{N}^{\perp}(\mathbf{A})=\{\mathbf{0}\}$

$\therefore\qquad\mathbf{x}_1=\mathbf{x}_2$

也就是说在$\mathbf{R}(\mathbf{A}^H)$中方程$\mathbf{Ax}=\mathbf{b}$只有唯一的解（若方程有解）

$\therefore$ 方程的任何其它解的 2-范数均大于$\mathbf{x}_0$的 2-范数

$\therefore$ $\mathbf{x}_0$是极小范数解

[得证]

由证明可知，方程$\mathbf{Ax}=\mathbf{b}$在$\mathbf{R}(\mathbf{A}^H)$的解必定是极小范数解。

引理 2. $\mathbf{A}\{1,4\}$由如下方程的通解构成$\mathbf{XA}=\mathbf{A}^{(1,4)}\mathbf{A}$，其中$\mathbf{A}^{(1,4)}$是 A 的某一个{1,4}-逆。

[证明]一方面：上述方程的解一定是 A 的某一个{1,4}-逆，设 X 为其解

（ⅰ） $\mathbf{AXA} = \mathbf{AA}^{(1,4)}\mathbf{A} = \mathbf{A}$

(iv) $\mathbf{XA} = \mathbf{A}^{(1,4)}\mathbf{A}$ 是厄米矩阵

另一方面：A 的任何{1,4}-逆均满足上述方程，设 X 是 A 的{1,4}-逆，$\mathbf{A}^{(1,4)}$ 是某个给定的{1,4}-逆，X 满足(ⅰ)(iv)Penrose 方程

$$\mathbf{A}^{(1,4)}\mathbf{A} \overset{(i)}{=} \mathbf{A}^{(1,4)}\mathbf{AXA} \overset{(iv)}{=} \left(\mathbf{A}^{(1,4)}\mathbf{A}\right)^{\mathbf{H}}\left(\mathbf{XA}\right)^{\mathbf{H}} = \left(\mathbf{XAA}^{(1,4)}\mathbf{A}\right)^{\mathbf{H}} = \left(\mathbf{XA}\right)^{\mathbf{H}} = \mathbf{XA}$$

[得证]

以上引理说明，对于 $\mathbf{X} \in \mathbf{A}\{1,4\}$，$\mathbf{XA}$ 是个不变量。

定理 2. 设方程 $\mathbf{Ax} = \mathbf{b}$ 相容，则 $\mathbf{x} = \mathbf{A}^{(1,4)}\mathbf{b}$ 是方程的极小范数解；反之，若对任意 $\mathbf{b} \in \mathbf{R(A)}$，存在 $\mathbf{X}$ 使得 $\mathbf{Xb}$ 成为该方程的极小范数解，则 $\mathbf{X} \in \mathbf{A}\{1,4\}$。

[证明] 先证前半部分。推论 1 $\rightarrow$ $\mathbf{A}^{(1)}\mathbf{b}$ 是 $\mathbf{Ax} = \mathbf{b}$ 的解

$$\begin{cases} \mathbf{A}^{(1,4)} \in \mathbf{A}\{1\} \rightarrow \mathbf{x} = \mathbf{A}^{(1,4)}\mathbf{b}\text{是方程的解} \\ \mathbf{A}^{(1,4)} \in \mathbf{A}\{4\} \rightarrow \mathbf{x} = \mathbf{A}^{(1,4)}\mathbf{b} = \mathbf{A}^{(1,4)}\mathbf{AA}^{(1)}\mathbf{b} = \left(\mathbf{A}^{(1,4)}\mathbf{A}\right)^{\mathbf{H}}\mathbf{A}^{(1)}\mathbf{b} \end{cases}$$

$$= \mathbf{A}^{\mathbf{H}}(\mathbf{A}^{(1,4)})^{\mathbf{H}}\mathbf{A}^{(1)}\mathbf{b} \in \mathbf{R}(\mathbf{A}^{\mathbf{H}})$$

由引理 1 知，$\mathbf{A}^{(1,4)}\mathbf{b}$ 是极小范数解。

后半部分:，存在对于任意 $\mathbf{b} \in \mathbf{R(A)}$，均有 $\mathbf{Xb}$ 为 $\mathbf{Ax} = \mathbf{b}$ 的极小范数解，即 $\mathbf{Xb} = \mathbf{A}^{(1,4)}\mathbf{b}$ 为极小范数解。

因为 $\forall\mathbf{b} \in \mathbf{R(A)}$，上式都成立，将 $\mathbf{b}$ 依次取为 $\mathbf{A}$ 的各列，合起来得

$$\mathbf{XA} = \mathbf{A}^{(1,4)}\mathbf{A}$$

由引理 2 知 $\mathbf{X} \in \mathbf{A}\{1,4\}$

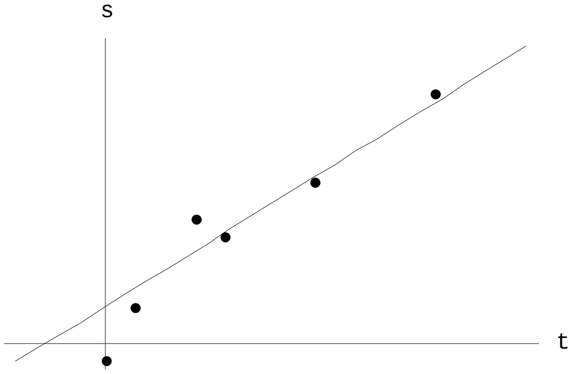
定理 3. 设 $\mathbf{A} \in \mathbf{C}^{\mathbf{m}\times\mathbf{n}}$，则 $\mathbf{A}\{1,4\} = \left\{\mathbf{A}^{(1,4)} + \mathbf{Z}(\mathbf{I} - \mathbf{AA}^{(1,4)}) \mid \mathbf{Z} \in \mathbf{C}^{\mathbf{m}\times\mathbf{n}}\right\}$

该定理的证明可由引理 2 结合定理 1 给出。

作业：P332　2　3(1)(2)

第十七讲 矛盾方程（组）的解---最小二乘法

## 一、从实验数据处理谈起




设有一组实验数据$(t_1, s_1)$，$(t_2, s_2)$，……，$(t_n, s_n)$，希望由实验数据拟合给定规律，从而测出待测量的有关参数。

假定规律为：$s=c_1\mathbf{t}+\mathbf{c_2}$，由于存在误差 $s_i \neq c_1\mathbf{t_i}+\mathbf{c_2} \qquad (\mathbf{i=1,2,\cdots,n})$，令

$$\mathbf{A}=\begin{Bmatrix}\mathbf{t_1} & \mathbf{1}\\ \mathbf{t_2} & \mathbf{1}\\ \vdots & \vdots\\ \mathbf{t_n} & \mathbf{1}\end{Bmatrix},\mathbf{x}=\begin{Bmatrix}\mathbf{c_1}\\ \mathbf{c_2}\end{Bmatrix},\mathbf{b}=\begin{Bmatrix}\mathbf{s_1}\\ \mathbf{s_2}\\ \vdots\\ \mathbf{s_n}\end{Bmatrix},$$

则:Ax=b实际无解,或者说矩阵方程Ax=b成为矛盾方程（不自洽、非相容），虽说无解，但在物理上看，我们需要而且也理当有“解”。怎么办？

一般处理是，定义一种目标函数，例如：

$$\mathbf{E(c_1,c_2)=\sum_{i=1}^{n}w_i(s_i-c_1t_i-c_2)^2} \qquad \mathbf{w_i>0}\text{为加权系数}$$

使误差$\mathbf{E(c_1,c_2)}$最小化。$w_i=1$(i=1～n)时$\mathbf{E(c_1,c_2)}=\|\mathbf{Ax-b}\|_2^2$

## 二、　最小二乘法（解）

对于矛盾方程Ax=b，最小二乘法是求其“解”的一种方法。即求使$\|\mathbf{Ax-b}\|_2=\mathbf{min}$的解。

引理：设$\mathbf{A\in C^{m\times n}}$,A{1,3}由如下方程的通解构成：

$$\mathbf{AX=AA^{(1,3)}\to A\{1,3\}=\{A^{(1,3)}+(I-A^{(1,3)}A)Z\,\big|\,Z\in C^{n\times m}\}}$$

其中，$A^{(1,3)}$为A{1,3}中的某个矩阵。

证：$1^{\circ}$ 方程既然相容，设X是其某个解，则

$$\mathbf{(i)\quad AXA=AA^{(1,3)}A=A\to X\in A\{1\}}$$

$$\mathbf{(iii)\quad (AX)^H=(AA^{(1,3)})^H=AA^{(1,3)}=AX\to X\in A\{3\}}$$

即方程的解必在 A{1,3}中。

2° 设X为A的一个{1,3}-逆矩阵，则

$$\begin{aligned}\mathbf{AX}&=\mathbf{AA}^{(1,3)}\mathbf{AX}\overset{\text{iii}}{=}\left(\mathbf{AA}^{(1,3)}\right)^{\mathbf{H}}\left(\mathbf{AX}\right)^{\mathbf{H}}\\&=\left(\mathbf{A}^{(1,3)}\right)^{\mathbf{H}}\mathbf{A}^{\mathbf{H}}\mathbf{X}^{\mathbf{H}}\mathbf{A}^{\mathbf{H}}\\&=\left(\mathbf{A}^{(1,3)}\right)^{\mathbf{H}}(\mathbf{AXA})^{\mathbf{H}}\\&=\left(\mathbf{AA}^{(1,3)}\right)^{\mathbf{H}}=\mathbf{AA}^{(1,3)}\end{aligned}$$

即，A的{1,3}-逆矩阵必满足方程AX=AA$^{(1,3)}$

$$\begin{aligned}\therefore\quad \mathbf{A}\{1,3\}&=\left\{方程\mathbf{AX}=\mathbf{AA}^{(1,3)}的所有解\right\}\\&=\left\{\mathbf{A}^{(1,3)}+(\mathbf{I}-\mathbf{A}^{(1,3)}\mathbf{A})\mathbf{Z}\,\middle|\,\mathbf{Z}\in\mathbf{C}^{n\times m}\right\}\end{aligned}$$

令 $\mathbf{X}=\mathbf{A}^{(1,3)}+(\mathbf{I}-\mathbf{A}^{(1,3)}\mathbf{A})\mathbf{Z}$，则

$$\text{(i)}\ \mathbf{AXA}=\mathbf{AA}^{(1,3)}\mathbf{A}+\mathbf{AZA}-\mathbf{AA}^{(1,3)}\mathbf{AZA}=\mathbf{A}\qquad \mathbf{X}\in\mathbf{A}\{1\}$$

$$\text{(iii)}\ \mathbf{AX}=\mathbf{AA}^{(1,3)}+(\mathbf{A}-\mathbf{AA}^{(1,3)}\mathbf{A})\mathbf{Z}=\mathbf{AA}^{(1,3)}=(\mathbf{AX})^{\mathbf{H}}\qquad \mathbf{X}\in\mathbf{A}\{3\}$$

定理：矩阵方程Ax=b的最小二乘解为 $\mathbf{x}=\mathbf{A}^{(1,3)}\mathbf{b}$，其中A$^{(1,3)}$为A的任何一个{1,3}-逆矩阵，

反之，存在X，对于任何 $\mathbf{b}\in\mathbf{C}^{m}$ 均有Xb成为Ax=b的最小二乘解，则 $\mathbf{X}\in\mathbf{A}\{1,3\}$。

证明：

$$\mathbf{Ax}-\mathbf{b}=(\mathbf{Ax}-\mathbf{P}_{\mathbf{R(A)}}\mathbf{b})+(\mathbf{P}_{\mathbf{R(A)}}\mathbf{b}-\mathbf{b})$$

$$(\mathbf{Ax}-\mathbf{P}_{\mathbf{R(A)}}\mathbf{b})\in\mathbf{R(A)},(\mathbf{P}_{\mathbf{R(A)}}\mathbf{b}-\mathbf{b})=-(\mathbf{I}-\mathbf{P}_{\mathbf{R(A)}})\mathbf{b}=-\mathbf{P}_{\mathbf{R}^{\perp}\mathbf{(A)}}\mathbf{b}\in\mathbf{R}^{\perp}(\mathbf{A})$$

所以，$\left\|\mathbf{Ax}-\mathbf{b}\right\|_2^2=\left\|\mathbf{Ax}-\mathbf{P}_{\mathbf{R(A)}}\mathbf{b}\right\|_2^2+\left\|\mathbf{P}_{\mathbf{R(A)}}\mathbf{b}-\mathbf{b}\right\|_2^2\geq\left\|\mathbf{b}-\mathbf{P}_{\mathbf{R(A)}}\mathbf{b}\right\|_2^2$，

故 $\left\|\mathbf{Ax}-\mathbf{b}\right\|_2^2$ 取得极小值的条件是x为方程 $\mathbf{Ax}=\mathbf{P}_{\mathbf{R(A)}}\mathbf{b}$ 的解。任取一个 $\mathbf{A}^{(1,3)}\in\mathbf{A}\{1,3\}$，我们知道 $\mathbf{AA}^{(1,3)}=\mathbf{P}_{\mathbf{R(A)}}$。而对于 $\mathbf{x}=\mathbf{A}^{(1,3)}\mathbf{b}$，有 $\mathbf{Ax}=\mathbf{AA}^{(1,3)}\mathbf{b}=\mathbf{P}_{\mathbf{R(A)}}\mathbf{b}$（但最小二乘解是否一定具有A$^{(1,3)}$b的形式呢？）

方程 $\mathbf{Ax}=\mathbf{AA}^{(1,3)}\mathbf{b}$ 的通解为

$$\begin{aligned}\mathbf{x}&=\left\{\mathbf{A}^{(1,3)}\mathbf{AA}^{(1,3)}\mathbf{b}+\mathbf{y}-\mathbf{A}^{(1,3)}\mathbf{Ay}\,\middle|\,\mathbf{y}\in\mathbf{C}^{n}\right\}\qquad \mathbf{y}=\mathbf{A}^{(1,3)}\mathbf{b}+\mathbf{z}\\&=\left\{\mathbf{A}^{(1,3)}\mathbf{b}+(\mathbf{I}-\mathbf{A}^{(1,3)}\mathbf{A})\mathbf{z}\,\middle|\,\mathbf{z}\in\mathbf{C}^{n}\right\}\end{aligned}$$

显然最小二乘解并不一定都具有A$^{(1,3)}$b的形式。

反之，若对于 $\forall \mathbf{b}\in \mathbf{C}^{m}, \mathbf{x}=\mathbf{X}\mathbf{b}$ 均使 $\mathbf{A}x=P_{R(A)}\mathbf{b}=\mathbf{A}\mathbf{A}^{(1,3)}\mathbf{b}$，即 $\forall \mathbf{b}$, 有 $\mathbf{A}\mathbf{X}\mathbf{b}=\mathbf{A}\mathbf{A}^{(1,3)}\mathbf{b}\to \mathbf{A}\mathbf{X}=\mathbf{A}\mathbf{A}^{(1,3)}\to \mathbf{X}\in \mathbf{A}\{1,3\}$

推论：x 是方程 Ax＝b 的最小二乘解的充要条件是，x 为方程 $\mathbf{A}^{H}\mathbf{A}\mathbf{x}=\mathbf{A}^{H}\mathbf{b}$ 的解。

证：$\mathbf{x}$ 为最小二乘解 $\Leftrightarrow \mathbf{A}\mathbf{x}=\mathbf{P}_{R(A)}\mathbf{b}$，而 $\mathbf{b}=\mathbf{P}_{R(A)}\mathbf{b}+\mathbf{P}_{N(A^{H})}\mathbf{b}$，故

$$\mathbf{x}\text{为最小二乘解}\Leftrightarrow \mathbf{A}\mathbf{x}-\mathbf{b}=-\mathbf{P}_{N(A^{H})}\mathbf{b}\in N(\mathbf{A}^{H})\to \mathbf{A}^{H}(\mathbf{A}\mathbf{x}-\mathbf{b})=\mathbf{0}$$

最小二乘解一般不唯一。

三、　极小范数最小二乘解

定理 2 ：设 $\mathbf{A}\in \mathbf{C}^{m\times n}, \mathbf{b}\in \mathbf{C}^{m}$，则 x＝$\boldsymbol{A}^{+}$b 是方程 Ax＝b 的极小范数最小二乘解。反之，若存在 $\mathbf{X}\in \mathbf{C}^{n\times m}$，若对于所有 $\mathbf{b}\in \mathbf{C}^{m}$，x＝Xb 均成为方程 Ax＝b 的极小范数最小二乘解，则 X＝$\boldsymbol{A}^{+}$。

证：最小二乘解满足 Ax ＝ $AA^{(1,3)}b$，其极小范数解唯一，且为 $x=A^{(1,4)}(AA^{(1,3)}b)=A^{+}b$，反之，$\forall \mathbf{b}\in \mathbf{C}^{m}, \mathbf{X}\mathbf{b}$ 均成为唯一的极小范数最小二乘解 $\boldsymbol{A}^{+}\boldsymbol{b}$，所以：X＝$\boldsymbol{A}^{+}$。

定理 3：矩阵方程 AXB＝D 的极小范数最小二乘解唯一，且为 $\boldsymbol{X}=\boldsymbol{A}^{+}\boldsymbol{D}\boldsymbol{B}^{+}$

证明略（教材 P86）

作业：P343－344，1，2，5

第十八讲　全面最小二乘法

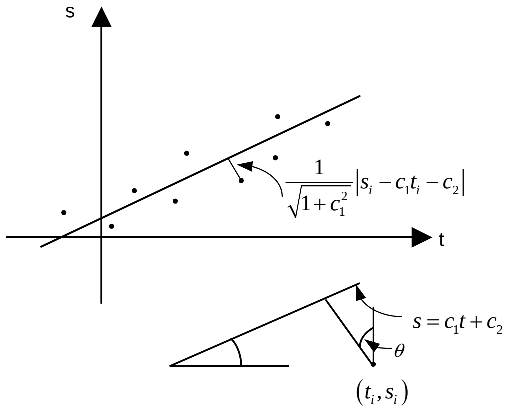
一、　法向回归

一组测量数据$\left(t_i,s_i\right)$,欲拟和直线

$$s=c_1t+c_2$$

最小二乘法采取目标函数：$E\left(c_1,c_2\right)=\sum_{i=1}^{n}\left|s_i-c_1t_i-c_2\right|^2=\min$

它隐含了在测量中，$t_i$是精确测量的，只有$s_i$才测得不准确，而在实际测量中，$t_i$，$s_i$都是无法准确测量的，因此，采用法向回归更有可能。




点$\left(t_i,s_i\right)$到直线$s=c_1t+c_2$的距离为

$$\frac{1}{\sqrt{1+c_1^2}}\left|s_i-c_1t_i-c_2\right|$$

故法向回归的目标函数为

$$E\left(c_1,c_2\right)=\left(\frac{1}{\sqrt{1+c_1^2}}\right)^2\sum_{i=1}^{n}\left|s_i-c_1t_i-c_2\right|^2=\min$$

$$\frac{\partial E}{\partial c_2}=\frac{1}{1+c_1^2}\sum_{i=1}^{n}(-2)\left(s_i-c_1t_i-c_2\right)=0\rightarrow c_2=\frac{1}{n}\sum_{i=1}^{n}s_i-c_1t_i$$

$$\frac{\partial E}{\partial c_1} = -\frac{2c_1}{\left(1+c_1^2\right)^2}\sum_{i=1}^{n}\left(s_i - c_1 t_i - c_2\right)^2 + \frac{2}{1+c_1^2}\sum_{i=1}^{n}\left(-t_i\right)\left(s_i - c_1 t_i - c_2\right)$$

$$= \frac{2}{1+c_1^2}\sum_{i=1}^{n}\left(c_1 c_2 - c_1 s_i - t_i\right)\left(s_i - c_1 t_i - c_2\right)$$

$$= \frac{2}{1+c_1^2}\left\{c_1 c\sum_{i=1}^{n}\ \left(s_i - c_1 t_i - c_2\right) - c_1\sum_{i=1}^{n}s_i\left(s_i - c_1 t_i - c_2\right) - \sum_{i=1}^{n}t_i\left(s_i - c_1 t_i - c_2\right)\right\}$$

$$= \frac{-2}{1+c_1^2}\left\{c_1\sum_{i=1}^{n}s_i\left(s_i - c_1 t_i - c_2\right) + \sum_{i=1}^{n}t_i\left(s_i - c_1 t_i - c_2\right)\right\} = 0$$

将 $c_2$ 代入之，可得

$$\begin{cases} c_1 = \dfrac{\left(l_{ss} - l_{tt}\right) + \sqrt{\left(l_{ss} - l_{tt}\right)^2 + 4l_{st}^2}}{2l_{st}} \\ c_2 = \bar{s} - c_1 \bar{t} \end{cases}$$

其中

$$\begin{cases} \bar{s} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}s_i \\ \bar{t} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}t_i \\ l_{ss} = \displaystyle\sum_{i=1}^{n}\left(s_i - \bar{s}\right)^2 = \sum_{i=1}^{n}s_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n}s_i\right)^2, \\ l_{st} = \displaystyle\sum_{i=1}^{n}\left(s_i - \bar{s}\right)\left(t_i - \bar{t}\right) = \sum_{i=1}^{n}s_i t_i - \frac{1}{n}\left(\sum_{i=1}^{n}s_i\right)\left(\sum_{i=1}^{n}t_i\right) \\ l_{tt} = \displaystyle\sum_{i=1}^{n}\left(t_i - \bar{t}\right)^2 = \sum_{i=1}^{n}t_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n}t_i\right)^2 \end{cases}$$

另一种推导方法：

$$E\left(c_1, c_2\right) = \frac{1}{1+c_1^2}\sum_{i=1}^{n}\left(s_i - c_1 t_i - c_2\right)^2$$

$$\frac{\partial E}{\partial c_2} = 0 \rightarrow c_2 = \frac{1}{n}\sum_{i=1}^{n}\left(s_i - c_1 t_i\right) = \bar{s} - c_1\bar{t} \Rightarrow E\left(c_1, c_2\right) = \frac{\sum_{i=1}^{n}\left[\left(s_i - \bar{s}\right) - c_1\left(t_i - \bar{t}\right)\right]^2}{1+c_1^2}$$

$$E(c_1,c_2)=\frac{l_{ss}-2c_1l_{st}+l_{tt}c_1^2}{1+c_1^2}$$

$$\frac{\partial E}{\partial c_1}=0\to c_1=\frac{(l_{ss}-l_{tt})\pm\sqrt{(l_{ss}-l_{tt})^2+4l_{st}^2}}{2l_{st}}$$

“±”中，“－”对应的 E 的最大值

作为比较，最小二乘法$\sum_{i=1}^{n}\left|s_i-c_1t_i-c_2\right|^2=\min$给出

$$\begin{cases} c_1=l_{st}/l_{tt} \\ c_2=\bar{s}-c_1\bar{t} \end{cases}$$

例1.　　7 点测量

$(t_i,s_i)=(0,3.1),(0.5,3.9),(1,5.2),(1.5,6.0),(2,6.9),(2.5,8.0),(3.0,9.1)$ 拟合直线$c_1t+c_2=s$

解：计算结果$\bar{t}=1.5,\bar{s}\approx 6.02857,l_{tt}=7,l_{ss}=27.8743,l_{st}=13.95$

最小二乘法给出$c_1=1.99286,c_2=3.03929$

全面最小二乘法（法向回归）给出$c_1=1.99709,c_2=3.03293$

测量数据误差小，分布合理时，两种方法效果非常接近。

二 、全面最小二乘法（Totally Least Square Method）

当方程$Ax=b$成为矛盾方程时，采用最小二乘法求解的观点实际上认为 b 存在误差，而 A 不存在误差，故应有$\varepsilon$，使得

$$Ax=b+\varepsilon$$

$\varepsilon$应尽量小以使得不至于严重得破坏方程$\to\left\|\varepsilon\right\|_2=\min$

全面最小二乘法采取如下观点解决矛盾方程的问题，不仅 b 存在误差，A 也存在误差，故，存在 E 和$\varepsilon$，使

$$(A+E)x=b+\varepsilon$$

$E$、$\varepsilon$也应该尽量小，以使得不至于严重偏离原方程$\to\left\|[E\,|\,\varepsilon]\right\|_F=\min$

$$(A+E)x=b+\varepsilon\Leftrightarrow\left([A\,|\,b]+[E\,|\,\varepsilon]\right)\begin{bmatrix}x\\-1\end{bmatrix}=0$$

记$\mathbf{C}=\left[\mathbf{A}\,|\,\mathbf{b}\right],\Delta=\left[\mathbf{E}\,|\,\boldsymbol{\varepsilon}\right],\mathbf{v}=\begin{bmatrix}\mathbf{x}\\-\mathbf{1}\end{bmatrix}$，则全面最小二乘解即求如下方程

$$\left(\mathbf{C}+\Delta\right)\mathbf{v}=\mathbf{0}$$

的非零解 v，且 v 的最后分量不能为零，而其中$\Delta$应满足$\left\|\Delta\right\|_F=\min$

引理：设$\mathbf{X}\in\mathbf{C}_r^{m\times n}$，且存在奇异值分解，$\mathbf{X}=\mathbf{U}\begin{bmatrix}\sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \mathbf{0}\end{bmatrix}_{m\times n}\mathbf{V}^H$，其中

$\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_r>\mathbf{0}$。又设

$$\mathbf{Y}=\mathbf{U}\begin{bmatrix}\sigma_1 & & & \\ & \ddots & & \\ & & \sigma_s & \\ & & & \mathbf{0}\end{bmatrix}_{m\times n}\mathbf{V}^n\qquad(s<r)$$ 则

$$\left\|\mathbf{X}-\mathbf{Y}\right\|_F=\min_{\substack{z\in C^{m\times n}\\ \mathrm{rank}z=s}}\left\|\mathbf{X}-\mathbf{Z}\right\|_F$$

首先来考虑 F-范数。设$\mathbf{P}_{m\times n}=\mathbf{UQV}^H,\mathbf{U}$、$\mathbf{V}$分别为 m 阶、n 阶酉矩阵。Q 为$m\times n$阶矩阵（上式不一定是奇异值分解）。则

$$\begin{aligned}\left\|\mathbf{P}\right\|_F^2&=\sum_{i,j}\left|p_{ij}\right|^2=\sum_{i=1}^{m}\left(\sum_{j=1}^{n}p_{ij}\overline{p_{ij}}\right)=\sum_{i=1}^{m}\left(\mathbf{PP}^H\right)_{ii}=\mathrm{tr}\left(\mathbf{PP}^H\right)=\mathrm{tr}\left(\mathbf{P}^H\mathbf{P}\right)\\&=\mathrm{tr}\left(\mathbf{UQV}^H\mathbf{VQ}^H\mathbf{U}^H\right)=\mathrm{tr}\left(\mathbf{UQQ}^H\mathbf{U}^H\right)=\mathrm{tr}\left(\mathbf{Q}^H\mathbf{U}^H\mathbf{UQ}\right)\\&=\mathrm{tr}(\mathbf{Q}^H\mathbf{Q})=\mathrm{tr}\left(\mathbf{QQ}^H\right)=\left\|\mathbf{Q}\right\|_F^2\end{aligned}$$

（按照教材上的说法，正交相抵或酉相抵的矩阵 F 范数相同）

$\left\|\mathbf{X}-\mathbf{Y}\right\|_F^2=\sum_{i=s+1}^{r}\sigma_i^2$，又令$\mathbf{Z}=\mathbf{UTV}^H\leftarrow\mathbf{T}=\mathbf{U}^H\mathbf{ZV}$，则

$$\left\|\mathbf{X}-\mathbf{Z}\right\|_F^2=\sum_{i=1}^{r}\left|t_{ii}-\sigma_i\right|^2+\sum_{\substack{i=1\\i\neq j}}^{r}\sum_{j=1}^{n}\left|t_{ij}\right|^2+\sum_{i=r+1}^{m}\sum_{j=1}^{n}\left|t_{ij}\right|^2$$

对任意 Z 矩阵而言，各$t_{ij}$之间完全独立，则$\left\|\mathbf{X}-\mathbf{Z}\right\|_F$是可能等于零的。但是

$\mathbf{rank(Z)=s<r}$ 。故 $\left\|\mathbf{X-Z}\right\|_F$ 不可能为零。详细论证可知 $\mathbf{t_{ij}=0(i\neq j), t_{ii}=0(i>s), t_{ii}=\sigma_i(i=1,2,\cdots,s)}$ 时，$\left\|\mathbf{X-Z}\right\|_F$ 最小

下面仅考虑在实际应用中非常常见的一种情况：$\mathbf{A\in C_n^{m\times n}}$，$\left[\mathbf{A}\middle|\mathbf{b}\right]\in \mathbf{C_{n+1}^{m\times n}}$，即 A 是列满秩的，$\left[\mathbf{A}\middle|\mathbf{b}\right]$ 也是列满秩的。这样，系数矩阵与增广矩阵的秩不相等，方程 $\mathbf{Ax=b}$ 不相容。

定理 1： 设 $\mathbf{A\in C_n^{m\times n}}, \left[\mathbf{A}\middle|\mathbf{b}\right]\in \mathbf{C_{n+1}^{m\times(n+1)}}$ 具有如下的奇异值分解

$$\mathbf{C=U}\begin{bmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_{n+1} \\ & \mathbf{0} & \end{bmatrix}\mathbf{V^H}, (\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_{n+1})$$

则使方程 $\left(\mathbf{C}+\Delta\right)\mathbf{v=0}$ 具有非零解，且 F 范数最小的 $\Delta$ 存在，并且 $\left\|\Delta\right\|_F=\sigma_{n+1}$

证明：方程 $\left(\mathbf{C}+\Delta\right)\mathbf{v=0}$ 要有非零解，必须 $\mathbf{rank(C+\Delta)<n+1}$，故由引理知

$$\begin{aligned}\min\left\|\Delta\right\|_F &= \min_{\mathrm{rank}\left(C+\Delta\right)<n+1}\left\|C-\left(C+\Delta\right)\right\|_F \\ &= \min_{\mathrm{rank}\left(C+\Delta\right)=n}\left\|C-\left(C+\Delta\right)\right\| = \sigma_{n+1}\end{aligned}$$

显然满足

$$\Delta=\mathbf{U}\begin{bmatrix}\mathbf{0} & & & \\ & \ddots & & \\ & & \mathbf{0} & \\ & & & \sigma_{n+1} \\ & & \mathbf{O} & \end{bmatrix}\mathbf{V^H}$$

定理 2： 设 $\sigma_{n+1}$ 为 C 的 n-k+1 重奇异值，且 $\mathbf{v_{k+1}, v_{k+2},\cdots v_{n+1}}$ 相应的为 $\mathbf{C^HC}$ 的属于 $\left(\mathbf{n-k+1}\right)$ 重特征值 $\sigma_{n+1}^2$ 的正交归一特征向量，则使方程 $\left(\mathbf{C}+\Delta\right)\mathbf{v=0}$ 具有非零的解且 F 范数最小的 $\Delta$ 为

$$\Delta=-\mathbf{Cv_sv_s^H}\big/\mathbf{v_s^Hv_s}$$

而方程的解则为 $\mathbf{v=v_s}$，其中 $\mathbf{v_s\in S_c}\overset{\Delta}{=}\mathbf{span}\left\{\mathbf{v_{k+1}, v_{k+2},\cdots v_{n+1}}\right\}$

证：（1） 显然 $\mathbf{CC^Hv_s}=\sigma_{n+1}^2\mathbf{v_s}$

$$\|\Delta\|_F^2=\left\|\mathbf{C}\mathbf{v}_s\mathbf{v}_s^H\right\|_F^2\Big/\left(\mathbf{v}_s^H\mathbf{v}_s\right)^2=\mathbf{tr}\left(\mathbf{v}_s\mathbf{v}_s^H\mathbf{C}^H\mathbf{C}\mathbf{v}_s\mathbf{v}_s^H\right)\Big/\left(\mathbf{v}_s^H\mathbf{v}_s\right)^2$$

$$=\frac{\sigma_{n+1}^2}{\left(\mathbf{v}_s^H\mathbf{v}_s\right)^2}\mathbf{tr}\left(\mathbf{v}_s\mathbf{v}_s^H\mathbf{v}_s\mathbf{v}_s^H\right)=\frac{\sigma_{n+1}^2}{\mathbf{v}_s^H\mathbf{v}_s}\mathbf{tr}\left(\mathbf{v}_s\mathbf{v}_s^H\right)=\frac{\sigma_{n+1}^2}{\mathbf{v}_s^H\mathbf{v}_s}\mathbf{tr}\left(\mathbf{v}_s^H\mathbf{v}_s\right)$$

$$=\sigma_{n+1}^2$$

（2） $$\left(\mathbf{C}+\Delta\right)\mathbf{v}_s=\mathbf{C}\mathbf{v}_s-\frac{\mathbf{C}\mathbf{v}_s\mathbf{v}_s^H\mathbf{v}_s}{\mathbf{v}_s^H\mathbf{v}_s}=\mathbf{0}$$

（3） $\forall \mathbf{v}\in\mathbf{C}^{n+1}$,有

$$\mathbf{v}=\left(\sum_{i=1}^{n-k+1}\mathbf{v}_{k+i}\mathbf{v}_{k+i}^H\right)\mathbf{v}+\left(\mathbf{I}_{n+1}-\sum_{i=1}^{n-k+1}\mathbf{v}_{k+i}\mathbf{v}_{k+i}^H\right)\mathbf{v}\overset{\Delta}{=}\mathbf{v}_s+\mathbf{v}_T$$

虽然，$\left(\mathbf{C}-\mathbf{C}\dfrac{\mathbf{v}\mathbf{v}^H}{\mathbf{v}^H\mathbf{v}}\right)\mathbf{v}=\mathbf{0}$,但 $\left\|-\mathbf{C}\dfrac{\mathbf{v}\mathbf{v}^H}{\mathbf{v}^H\mathbf{v}}\right\|_F>\sigma_{n+1}$

$$\because\quad \mathbf{C}^H\mathbf{C}\mathbf{v}=\mathbf{C}^H\mathbf{C}\left(\mathbf{v}_s+\mathbf{v}_T\right)=\sigma_{n+1}^2\mathbf{v}_s+\mathbf{C}^H\mathbf{C}\mathbf{v}_T>\sigma_{n+1}^2\mathbf{v}_s+\sigma_{n+1}^2\mathbf{v}_T$$

$$=\sigma_{n+1}^2\left(\mathbf{v}_s+\mathbf{v}_T\right)=\sigma_{n+1}^2\mathbf{v}$$

$$\therefore\quad \left\|\mathbf{C}\frac{\mathbf{v}\mathbf{v}^H}{\mathbf{v}^H\mathbf{v}}\right\|_F^2=\frac{1}{\left(\mathbf{v}^H\mathbf{v}\right)^2}\mathbf{tr}\left(\mathbf{v}\mathbf{v}^H\mathbf{C}^H\mathbf{C}\mathbf{v}\mathbf{v}^H\right)>\frac{\sigma_{n+1}^2}{\left(\mathbf{v}^H\mathbf{v}\right)^2}\mathbf{tr}\left(\mathbf{v}\mathbf{v}^H\mathbf{v}\mathbf{v}^H\right)=\sigma_{n+1}^2$$

定理 3：在定理 2 的条件下，全面最小二乘解存在的充要条件为：向量 $\mathbf{e}_{n+1}=\begin{bmatrix}\underbrace{\mathbf{0}\ \cdots\ \mathbf{0}}_{n个} & 1\end{bmatrix}^T$ 不正交于 $\mathbf{S}_c$。此时，$\forall\mathbf{v}\in\left\{\mathbf{q}=\begin{bmatrix}\mathbf{y}\\ \alpha\end{bmatrix}\middle|\mathbf{q}\in\mathbf{S}_c,\alpha\neq\mathbf{0}\right\}$，则最小二乘解为

$$\mathbf{x}=\frac{1}{\alpha}\mathbf{y}$$

说明：（1）最小二乘解一定存在，但全面最小二乘解不一定

（2）存在全面最小二乘解时，若 $\sigma_{n+1}$ 为 C 的单重奇异值，全面最小二乘解唯一，否则，解不唯一

例 2. 采用全面最小二乘法重新研究（上例）法向回归的问题

$$\mathbf{Ax}=\mathbf{b} \qquad \mathbf{A}=\begin{bmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \\ t_n & 1 \end{bmatrix} \qquad \mathbf{x}=\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \qquad \mathbf{b}=\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}$$

$$\mathbf{C}=\begin{bmatrix} t_1 & 1 & s_1 \\ t_2 & 1 & s_2 \\ \vdots & \vdots & \vdots \\ t_n & 1 & s_n \end{bmatrix}$$

$$\mathbf{C}^T\mathbf{C}=\begin{bmatrix} t_1 & t_2 & \cdots & t_n \\ 1 & 1 & \cdots & 1 \\ s_1 & s_2 & \cdots & s_n \end{bmatrix}\begin{bmatrix} t_1 & 1 & s_1 \\ t_2 & 1 & s_2 \\ \vdots & \vdots & \vdots \\ t_n & 1 & s_n \end{bmatrix}=\begin{bmatrix} \sum t_i^2 & \sum t_i & \sum s_i t_i \\ \sum t_i & n & \sum s_i \\ \sum s_i t_i & \sum s_i & \sum s_i^2 \end{bmatrix}$$

$$\mathbf{C}^T\mathbf{C}=\begin{bmatrix} 22.75 & 10.5 & 77.25 \\ 10.5 & 7 & 42.2 \\ 77.25 & 42.2 & 282.28 \end{bmatrix},\ \lambda = 309.7754, 2.249389, 0.0051987257$$

$$\mathbf{v}_3=\begin{bmatrix} 1.990944 & 3.044416 & -1 \end{bmatrix}^T$$

$$\begin{matrix} \downarrow & \downarrow \\ c_1 & c_2 \end{matrix}$$

（对应$\lambda_3$）

与法向回归结果并不相同，但亦十分接近。值得注意的是$\bar{s} \neq c_1\bar{t}+c_2$（全面最小二乘解）

第十九讲　范数理论及其应用

## 一、向量范数

范数可以看作长度概念的推广，主要用于逼近的程度。

1. 向量范数定义：设 V 为数域 K 上的向量空间，若对于 V 的任一向量 x，对应一个实值函数$\|\mathbf{x}\|$，并满足以下三个条件：

（1）非负性 $\|\mathbf{x}\|\geq \mathbf{0}$，等号当且仅当 x=0 时成立；

（2）齐次性 $\|\alpha\mathbf{x}\|=|\alpha|\|\mathbf{x}\|,\alpha\in\mathbf{k},\mathbf{x}\in\mathbf{V};$

（3）三角不等式$\|\mathbf{x}+\mathbf{y}\|\leq\|\mathbf{x}\|+\|\mathbf{y}\|,\mathbf{x},\mathbf{y}\in\mathbf{V}$。

则称$\|\mathbf{x}\|$为 V 中向量 x 的范数，简称为向量范数。

例 1. $\mathbf{x}\in\mathbf{C}^{\mathbf{n}}$，它可表示成$\mathbf{x}=\begin{bmatrix}\xi_1 & \xi_2 & \cdots & \xi_n\end{bmatrix}^{\mathrm{T}}$，$\xi_i\in\mathbf{C}$，

$$\|\mathbf{x}\|_2\overset{\Delta}{=}\left(\sum_{i=1}^{n}|\xi_i|^2\right)^{1/2}$$就是一种范数

证明：（i）非负性 $\|\mathbf{x}\|_2=\left(\sum_{i=1}^{n}|\xi_i|^2\right)^{1/2}\geq\mathbf{0}$，

当且仅当$\xi_i=\mathbf{0}(\mathbf{i}=\mathbf{1},\mathbf{2},\cdots,\mathbf{n})$时，即 x＝0 时，$\|\mathbf{x}\|_2$＝0

（ii）齐次性 $$\|\alpha\mathbf{x}\|_2=\left(\sum_{i=1}^{n}|\alpha\xi_i|^2\right)^{1/2}=|\alpha|\left(\sum_{i=1}^{n}|\xi_i|^2\right)^{1/2}=|\alpha|\|\mathbf{x}\|_2$$

（iii）$\mathbf{y}=\begin{bmatrix}\eta_1 & \eta_2 & \cdots & \eta_n\end{bmatrix}^{\mathrm{T}}$，$\eta_i\in\mathbf{C}$

$$\mathbf{x}+\mathbf{y}=\begin{bmatrix}\xi_1+\eta_1 & \xi_2+\eta_2 & \cdots & \xi_n+\eta_n\end{bmatrix}^{\mathrm{T}}$$

$$\|\mathbf{x}+\mathbf{y}\|_2^2=\sum_{i=1}^{n}|\xi_i+\eta_i|^2$$

$$|\xi_i+\eta_i|^2=|\xi_i|^2+|\eta_i|^2+2\mathrm{Re}\left(\overline{\xi_i}\eta_i\right)\leq|\xi_i|^2+|\eta_i|^2+2|\xi_i||\eta_i|$$

$$\|\mathbf{x}+\mathbf{y}\|_2^2\leq\|\mathbf{x}\|_2^2+\|\mathbf{y}\|_2^2+2\sum_{i=1}^{n}|\xi_i||\eta_i|$$

$$\left(\|\mathbf{x}\|_2+\|\mathbf{y}\|_2\right)^2=\|\mathbf{x}\|_2^2+\|\mathbf{y}\|_2^2+2\|\mathbf{x}\|_2\|\mathbf{y}\|_2$$

根据 Hölder 不等式：

$$\sum_{i=1}^{n} a_i b_i \le \left(\sum_{i=1}^{n} a_i^p\right)^{1/p} \left(\sum_{i=1}^{n} b_i^q\right)^{1/q}, \quad p,q>1, \frac{1}{p}+\frac{1}{q}=1, a_i, b_i>0$$

$$\|x\|_2 \|y\|_2 = \left(\sum_{i=1}^{n} |\xi_i|^2\right)^{1/2} \left(\sum_{i=1}^{n} |\eta_i|^2\right)^{1/2} \ge \sum_{i=1}^{n} |\xi_i||\eta_i|$$

$$\therefore \quad \|x+y\|_2 \le \|x\|_2 + \|y\|_2$$

2. 两类向量范数

（1）$\|x\|_2 = \left(x^H x\right)^{1/2}$

推广到 $\|x\|_A = \left(x^H A x\right)^{1/2}$，A 为厄米正定矩阵（椭圆范数）

当$A = W = \mathrm{diag}\begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix}$，$w_i > 0$

$$\|x\|_w = \left(\sum_{i=1}^{n} w_i |\xi_i|^2\right)^{1/2} \quad \text{加权范数}$$

（2）$\|x\|_p = \left(\sum_{i=1}^{n} |\xi_i|^p\right)^{1/p}$ （p≥1）,称为向量的 p-范数或$l_p$范数。

证明：$\|x\|_p$显然满足非负性和齐次性

$$y = \begin{bmatrix} \eta_1 & \eta_2 & \cdots & \eta_n \end{bmatrix}^T$$

$$\|x\|_p = \left(\sum_{i=1}^{n} |\xi_i|^p\right)^{1/p}, \quad \|y\|_p = \left(\sum_{i=1}^{n} |\eta_i|^p\right)^{1/p}, \quad \|x+y\|_p = \left(\sum_{i=1}^{n} |\xi_i+\eta_i|^p\right)^{1/p}$$

$$\left(\|x+y\|_p\right)^p = \sum_{i=1}^{n} |\xi_i+\eta_i|^p = \sum_{i=1}^{n} |\xi_i+\eta_i|^{p-1} |\xi_i+\eta_i|$$

$$\le \sum_{i=1}^{n} |\xi_i+\eta_i|^{p-1} |\xi_i| + \sum_{i=1}^{n} |\xi_i+\eta_i|^{p-1} |\eta_i|$$

应用 Hölder 不等式

$$\sum_{i=1}^{n} |\xi_i+\eta_i|^{p-1} |\xi_i| \le \left[\sum_{i=1}^{n} |\xi_i+\eta_i|^{(p-1)q}\right]^{1/q} \left[\sum_{i=1}^{n} |\xi_i|^p\right]^{1/p}$$

$$\sum_{i=1}^{n} |\xi_i+\eta_i|^{p-1} |\eta_i| \le \left[\sum_{i=1}^{n} |\xi_i+\eta_i|^{(p-1)q}\right]^{1/q} \left[\sum_{i=1}^{n} |\eta_i|^p\right]^{1/p}$$

$$\frac{1}{p}+\frac{1}{q}=1 \Rightarrow (p-1)q = p$$

$$\therefore \quad \sum_{i=1}^{n}\left|\xi_i+\eta_i\right|^p \le \left(\sum_{i=1}^{n}\left|\xi_i+\eta_i\right|^p\right)^{1/q}\left[\left(\sum_{i=1}^{n}\left|\xi_i\right|^p\right)^{1/p}+\left(\sum_{i=1}^{n}\left|\eta_i\right|^p\right)^{1/p}\right]$$

$$\left(\sum_{i=1}^{n}\left|\xi_i+\eta_i\right|^p\right)^{1/p} \le \left(\sum_{i=1}^{n}\left|\xi_i\right|^p\right)^{1/p}+\left(\sum_{i=1}^{n}\left|\eta_i\right|^p\right)^{1/p}$$

即 $\|x+y\|_p \le \|x\|_p+\|y\|_p$

3. 向量范数的等价性

定理 1. 设$\|\ \|_\alpha$、$\|\ \|_\beta$为$C^n$的两种向量范数，则必定存在正数 m、M，使得 $m\|x\|_\alpha \le \|x\|_\beta \le M\|x\|_\alpha$，（m、M 与 x 无关），它就称为向量范数的等价性。

同时有$\dfrac{1}{M}\|x\|_\beta \le \|x\|_\alpha \le \dfrac{1}{m}\|x\|_\beta$

二、矩阵范数

1. 矩阵范数定义：设$k^{m\times n}(k=c或R)$表示数域 k 上全体$m\times n$阶矩阵的集合。若对于$k^{m\times n}$中任一矩阵 A，均对应一个实值函数，并满足以下四个条件：

（1）非负性：$\|A\|\ge 0$ ，等号当且仅当 A=0 时成立；

（2）齐次性：$\|\alpha A\|=|\alpha|\|A\|,\alpha\in k;$

（3）三角不等式：$\|A+B\|\le\|A\|+\|B\|, A,B\in k^{m\times n}$

则称$\|A\|$为广义矩阵范数；

（4）相容性： $\|AB\|\le\|A\|\|B\|$

则称$\|A\|$为矩阵范数。

2. 常用的矩阵范数

（1）p-范数：$\|A\|_p=\max\dfrac{\|Ax\|_p}{\|x\|_p}$，

$A=\left(a_{ij}\right)_{m\times n}$，x 为所有可能的向量，$x=\left[\xi_1\quad \xi_2\quad \cdots\quad \xi_n\right]^T$，

$\|\alpha x\|_p=|\alpha|\|x\|_p$，$\|Ax\|_p=\dfrac{1}{|\alpha|}\|A(\alpha x)\|_p\ \ (\alpha\ne 0)$

$$\therefore \quad \|A\|_p = \max_{\|x\|_p=1} \|Ax\|_p$$

$$\|A\|_1 = \max_{\|x\|_1=1} \|Ax\|_1$$

$$\|x\|_1 = \sum_{i=1}^{n} |\xi_i| = 1 \text{，} \quad \|Ax\|_1 = \sum_{i=1}^{m} \left| \sum_{j=1}^{n} a_{ij}\xi_j \right|$$

可以证明：$\|A\|_1 = \max\limits_{1\le j\le n} \sum\limits_{i=1}^{n} |a_{ij}|$ 列（和）范数

$\|A\|_2 = \max\limits_{1\le i\le n} \sqrt{\lambda_i (A^H A)}$ 谱范数

$\|A\|_\infty = \max\limits_{1\le i\le m} \sum\limits_{j=1}^{n} |a_{ij}|$ 行（和）范数

$$\|x\|_\infty = \left( \sum_{i=1}^{n} |\xi_i|^p \right)^{1/p} \Bigg|_{p\to\infty} = \max_{1\le i\le n} |\xi_i|$$

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

（2）Frobenius 范数（F-范数）和导出性范数

F-范数： $\|A\|_F = \left( \sum\limits_{i,j=1}^{n} |a_{ij}|^2 \right)^{1/2}$

导出性范数：设$\|x\|$为数域 k 上 n 维向量空间$k^n$（k=R 或 C）的一种向量范数。可定义矩阵范数为：

$$\|A\| = \max_{\|x\|=1} \|Ax\| = \max_{x\ne 0} \left( \frac{\|Ax\|}{\|x\|} \right)$$

三、应用

逼近和误差估计是矩阵范数应用的主要领域。

矩阵条件数：$\mathrm{cond}(A) = \|A\|\|A^{-1}\|$

由相容性可知：$\|A\|\|A^{-1}\| \ge \|AA^{-1}\| = \|I\|$

$$\|x\| = \|Ix\| \le \|I\|\|x\| \Rightarrow \|I\| \ge 1$$

对于导出性范数 $\|I\| = 1$

$$\therefore \quad \mathrm{cond}(\mathbf{A}) \ge 1$$

条件数反映了误差放大的程度，条件数越大，矩阵越病态。

对于方程 $\mathbf{Ax}=\mathbf{b} \quad \to \mathbf{x}=\mathbf{A}^{-1}\mathbf{b}$

考虑两种情况：(1) b 存在误差； (2) A 存在误差

（1） b 存在误差 $\Delta\mathbf{b}$，求出的 x 存在误差 $\Delta\mathbf{x}$， $\Delta\mathbf{x}=\mathbf{A}^{-1}\Delta\mathbf{b}$

$$\|\Delta\mathbf{x}\| \le \|\mathbf{A}^{-1}\|\|\Delta\mathbf{b}\|$$

考察相对误差，求 $\dfrac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|}$

$$\|\mathbf{b}\| \le \|\mathbf{A}\|\|\mathbf{x}\| \to \|\mathbf{x}\| \ge \frac{\|\mathbf{b}\|}{\|\mathbf{A}\|}$$

$$\therefore \quad \frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \le \|\mathbf{A}\|\|\mathbf{A}^{-1}\|\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|}$$

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \Big/ \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} \le \|\mathbf{A}\|\|\mathbf{A}^{-1}\| = \mathrm{cond}(\mathbf{A})$$

（2） A 存在误差 $\Delta\mathbf{A}$，求出的解 x 存在误差 $\Delta\mathbf{x}$

$$\mathbf{Ax}=\mathbf{b} \qquad (\mathbf{A}+\Delta\mathbf{A})(\mathbf{x}+\Delta\mathbf{x})=\mathbf{b}$$

$$\to \mathbf{A}\Delta\mathbf{x} = -\Delta\mathbf{A}\mathbf{x} - \Delta\mathbf{A}\Delta\mathbf{x}$$

忽略高阶小量得： $\|\Delta\mathbf{x}\| \le \|\mathbf{A}^{-1}\|\|\Delta\mathbf{A}\|\|\mathbf{x}\|$

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \le \|\mathbf{A}^{-1}\|\|\mathbf{A}\|\frac{\|\Delta\mathbf{A}\|}{\|\mathbf{A}\|}$$

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \Big/ \frac{\|\Delta\mathbf{A}\|}{\|\mathbf{A}\|} \le \|\mathbf{A}^{-1}\|\|\mathbf{A}\| = \mathrm{cond}(\mathbf{A})$$

常用条件数用 $\|\mathbf{A}\|_2$ 来考虑：

$$\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^H\mathbf{A})}$$

$$\|\mathbf{A}^{-1}\|_2 = \sqrt{\lambda_{\min}^{-1}(\mathbf{A}^H\mathbf{A})}$$

$$\mathrm{cond}(\mathbf{A}) = \sqrt{\frac{\lambda_{\max}(\mathbf{A}^H\mathbf{A})}{\lambda_{\min}(\mathbf{A}^H\mathbf{A})}}$$

作业：P275　1、2

第二十讲 矩阵特征值估计

特征值计算较困难，希望找到简便的特征值界限或分布范围的估计方法。

一、 特征值界的估计

定理 1. 设$\mathbf{A}\in\mathbf{R}^{n\times n}$，$\lambda$为 A 的任意特征值，则有

$$\left|\mathrm{Im}(\lambda)\right|\le M\sqrt{\frac{n(n-1)}{2}}$$

其中，$M=\max\limits_{1\le i,j\le n}\left|\frac{a_{ij}-a_{ji}}{2}\right|$

证明：设 x 为 A 的属于特征值$\lambda$的单位特征向量，即$Ax=\lambda x$，$x^H x=1$，则

$$\lambda=x^H Ax \ \rightarrow\ \bar{\lambda}=\left(x^H Ax\right)^H=x^H A^H x$$

$$\lambda-\bar{\lambda}=2j\mathrm{Im}(\lambda)=x^H\left(A-A^H\right)x=x^H\left(A-A^T\right)x$$

将 x 写成$x=\left[\xi_1,\xi_2,\cdots,\xi_n\right]^T$

$$x^H\left(A-A^T\right)x=\sum_{i=1}^{n}\sum_{j=1}^{n}\bar{\xi}_i\left(a_{ij}-a_{ji}\right)\xi_j$$

$$2\left|\mathrm{Im}(\lambda)\right|=\left|\sum_{i=1}^{n}\sum_{j=1}^{n}\bar{\xi}_i\left(a_{ij}-a_{ji}\right)\xi_j\right|$$

$$\le\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\bar{\xi}_i\left(a_{ij}-a_{ji}\right)\xi_j\right|$$

$$=\sum_{i,j=1}^{n}{}'\left|\xi_i\xi_j\right|\left|a_{ij}-a_{ji}\right| \qquad \text{（}\textstyle\sum{}'\text{ 表示不含 i＝j）}$$

$$\le 2M\sum_{i,j=1}^{n}{}'\left|\xi_i\xi_j\right|$$

$$\left|\mathrm{Im}(\lambda)\right|^2\le M^2\left(\sum_{i,j=1}^{n}{}'\left|\xi_i\xi_j\right|\right)^2$$

$$\le M^2 n(n-1)\sum_{i,j=1}^{n}{}'\left|\xi_i\xi_j\right|^2$$

$$=M^2 n(n-1)\sum_{i,j=1}^{n}{}'\left|\xi_i\right|^2\left|\xi_j\right|^2$$

$$\sum_{i,j=1}^{n}{}' \left|\xi_i\right|^2 \left|\xi_j\right|^2 = \sum_{i,j=1}^{n} \left|\xi_i\right|^2 \left|\xi_j\right|^2 - \sum_{i=1}^{n} \left|\xi_i\right|^4 \le \sum_{i=1}^{n} \left|\xi_i\right|^2 - \sum_{i=1}^{n} \left|\xi_i\right|^4$$

$$= \sum_{i=1}^{n} \left|\xi_i\right|^2 \left(1 - \left|\xi_i\right|^2\right)$$

不妨写为： $= \left|\xi_1\right|^2 \left(1 - \left|\xi_1\right|^2\right) + \left|\xi_2\right|^2 \left(1 - \left|\xi_2\right|^2\right) + \sum_{i=3}^{n} \left|\xi_i\right|^2 \left(1 - \left|\xi_i\right|^2\right)$

$$\le \left( \frac{\left|\xi_1\right|^2 + \left(1 - \left|\xi_1\right|^2\right)}{2} \right)^2 + \left( \frac{\left|\xi_2\right|^2 + \left(1 - \left|\xi_2\right|^2\right)}{2} \right)^2 + \sum_{i=3}^{n} \left|\xi_i\right|^2 \left(1 - \left|\xi_i\right|^2\right)$$

$$\le \frac{1}{2}$$

取等号的条件为$\left|\xi_1\right|^2 = \left|\xi_2\right|^2 = \dfrac{1}{2}$，但$\left\|x\right\|^2 = 1$，所以其它$\left|\xi_i\right|^2 = 0$

$$\therefore \quad \left|\mathrm{Im}(\lambda)\right| \le M\sqrt{\frac{n(n-1)}{2}}$$

定理 2. 设$A \in R^{n\times n}$，$\lambda$为 A 的任意特征值，则有

$$\left|\lambda\right| \le n\rho \qquad \left|\mathrm{Re}(\lambda)\right| \le \frac{1}{2} n\tau \qquad \left|\mathrm{Im}(\lambda)\right| \le \frac{1}{2} ns$$

其中，$\rho = \max\limits_{1\le i,j\le n} \left|a_{ij}\right|$，$\tau = \max\limits_{1\le i,j\le n} \left|a_{ij} + \overline{a_{ji}}\right|$，$s = \max\limits_{1\le i,j\le n} \left|a_{ij} - \overline{a_{ji}}\right|$

二、 盖尔圆法

定义：设$A = \left(a_{ij}\right)_{n\times n} \in C^{n\times n}$，由方程$\left|z - a_{ii}\right| \le R_i = \sum\limits_{\substack{j=1 \\ i\ne j}}^{n} \left|a_{ij}\right|$所确定的圆称为 A 的第 i 个盖尔圆，$R_i$称为盖尔圆的半径。

定理 3：矩阵 A 的所有特征值均落在它的所有盖尔圆的并集之中。

证明：设$A = \left(a_{ij}\right)_{n\times n} \in C^{n\times n}$，$\lambda$为 A 的某一个特征值，x 为相应的特征向量，将 x 写成$x = \left[\xi_1, \xi_2, \cdots, \xi_n\right]^T$，设$\left|\xi_{i_0}\right| = \max\left|\xi_i\right|$

由$Ax = \lambda x$，考虑$i_0$行

$$\sum_{j=1}^{n} a_{i_0 j}\xi_j = \lambda \xi_{i_0}$$

$$\left(\lambda - a_{i_0 i_0}\right)\xi_{i_0} = \sum_{j=1}^{n}{}' a_{i_0 j}\xi_j \quad \left(j \neq i_0\right)$$

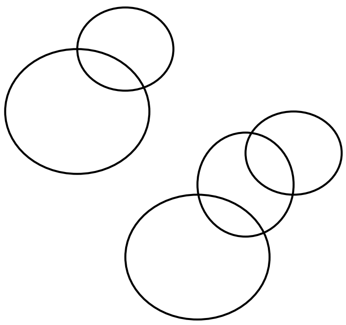
$$\left|\lambda - a_{i_0 i_0}\right| = \left|\sum_{j=1}^{n}{}' a_{i_0 j}\frac{\xi_j}{\xi_{i_0}}\right| \le \sum_{j=1}^{n}{}' \left|a_{i_0 j}\right|\left|\frac{\xi_j}{\xi_{i_0}}\right| \le R_{i_0}$$

对于 A 的特征值$\lambda$，一定存在$i_0\left(1 \le i_0 \le n\right)$，使$\lambda$落在 A 的第$i_0$个盖尔圆中，对于每个特征值都有相同的结论。

定理 4. 将矩阵 A 的全体盖尔圆的并集按连通部分分成若干个子集，（一个子集由完全连通的盖尔圆组成，不同子集没有相连通的部分），对每个子集，若它恰好由 K 个盖尔圆组成，则该子集中恰好包含 A 的 K 个特征值。

说明：盖尔圆相互重叠时重复计算，特征值相重时也重复计算

证明：设$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \in C^{n\times n}$，



令$B\left(u\right) = \begin{bmatrix} a_{11} & ua_{12} & ua_{13} & \cdots & ua_{1n} \\ ua_{21} & a_{22} & ua_{23} & \cdots & ua_{2n} \\ ua_{31} & ua_{32} & a_{33} & \cdots & ua_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ ua_{n1} & ua_{n2} & ua_{n3} & \cdots & a_{nn} \end{bmatrix}$

$0 \le u \le 1$，$B\left(0\right) = \mathrm{diag}\left[a_{11} \quad a_{22} \quad \cdots \quad a_{nn}\right]$，$B\left(1\right) = A$

$B\left(u\right)$的特征多项式是 u 的多项式，其特征值是 u 的连续函数，观察 u（$0 \le u \le 1$）变化的过程中$B\left(u\right)$特征值的变化，特征值只能在盖尔圆连通的子集内变动，而不能跨出连通子集。

由此可见，由 K 个盖尔圆组成的连通子集恰好包含 K 个特征值。

应该注意到：连通的这些盖尔圆中，有些盖尔圆可能包含两个或多个特征值，而其它盖尔圆中可能无特征值。

推论 1. 孤立盖尔圆中恰好包含一个特征值。

推论 2. 实矩阵的孤立盖尔圆恰好包含一个实特征值。

推论 3. 盖尔圆方法中盖尔圆半径可以按列求和。（因为方阵转置后特征值不变）

推论 4. 盖尔圆半径变为$\mathbf{R_i} = \sum_{j=1}^{n}{}' \frac{\alpha_i}{\alpha_j}\left|\mathbf{a_{ij}}\right|$，两个盖尔圆定理仍然成立。

说明如下：相似矩阵$\mathbf{P^{-1}AP}$与 A 具有相同的特征值，取

$$\mathbf{P} = \mathbf{diag}\begin{bmatrix}\alpha_1 & \alpha_2 & \cdots & \alpha_n\end{bmatrix} \ (\alpha_i > 0)$$

$$\mathbf{B} = \mathbf{P^{-1}AP} = \begin{bmatrix} \frac{1}{\alpha_1} & & & \mathbf{0} \\ & \frac{1}{\alpha_2} & & \\ & & \ddots & \\ \mathbf{0} & & & \frac{1}{\alpha_n} \end{bmatrix} \begin{bmatrix}\mathbf{a_{ij}}\end{bmatrix} \begin{bmatrix} \alpha_1 & & & \mathbf{0} \\ & \alpha_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \alpha_n \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\alpha_i}{\alpha_j}\mathbf{a_{ij}} \end{bmatrix}$$

根据推论 4，选取适当的$\alpha_i$使盖尔圆变大或变小，可以对特征值进行隔离。但有时这种隔离特征值的方法会失效，如对于那些对角线上由相同元素组成的矩阵，此时盖尔圆的圆心相同。

作业：P261　2，3，4

第二十一讲 广义特征值与极小极大原理

一、 广义特征值问题

1、定义：设 A、B 为 n 阶方阵，若存在数$\lambda$，使得方程$\mathbf{Ax}=\lambda\mathbf{Bx}$存在非零解，则称$\lambda$为 A 相对于 B 的广义特征值，x 为 A 相对于 B 的属于广义特征值$\lambda$的特征向量。

- 是标准特征值问题的推广，当 B＝I（单位矩阵）时，广义特征值问题退化为标准特征值问题。
- 特征向量是非零的
- 广义特征值的求解

$$(\mathbf{A}-\lambda\mathbf{B})\mathbf{x}=\mathbf{0} \quad 或者 \quad (\lambda\mathbf{B}-\mathbf{A})\mathbf{x}=\mathbf{0}$$

$$\rightarrow 特征方程 \ \det(\mathbf{A}-\lambda\mathbf{B})=\mathbf{0}$$

求得$\lambda$后代回原方程$\mathbf{Ax}=\lambda\mathbf{Bx}$可求出 x

本课程进一步考虑 A、B 厄米且为正定矩阵的情况。

2、等价表述

（1） B 正定，$\mathbf{B}^{-1}$存在 $\rightarrow \mathbf{B}^{-1}\mathbf{Ax}=\lambda\mathbf{x}$，广义特征值问题化为了标准特征值问题，但一般来说，$\mathbf{B}^{-1}\mathbf{A}$一般不再是厄米矩阵。

（2） B 厄米，存在 Cholesky 分解，$\mathbf{B}=\mathbf{GG}^{\mathbf{H}}$，G 满秩

$$\mathbf{Ax}=\lambda\mathbf{GG}^{\mathbf{H}}\mathbf{x} \qquad 令\ \mathbf{G}^{\mathbf{H}}\mathbf{x}=\mathbf{y}$$

则 $\mathbf{G}^{-1}\mathbf{A}\left(\mathbf{G}^{\mathbf{H}}\right)^{-1}\mathbf{y}=\lambda\mathbf{y}$ 也成为标准特征值问题。

$\mathbf{G}^{-1}\mathbf{A}\left(\mathbf{G}^{\mathbf{H}}\right)^{-1}$为厄米矩阵，广义特征值是实数，可以按大小顺序排列$\lambda_1\le\lambda_2\le\cdots\le\lambda_n$，一定存在一组正交归一的特征向量，即存在$\mathbf{y}_1,\mathbf{y}_2\cdots,\mathbf{y}_n$满足

$$\mathbf{G}^{-1}\mathbf{A}\left(\mathbf{G}^{\mathbf{H}}\right)^{-1}\mathbf{y}_i=\lambda\mathbf{y}_i$$

$$\mathbf{y}_i^{\mathbf{H}}\mathbf{y}_j=\delta_{ij}=\begin{cases}1 & i=j\\ 0 & i\neq j\end{cases}$$

还原为$\mathbf{x}_i=\left(\mathbf{G}^{\mathbf{H}}\right)^{-1}\mathbf{y}_i$ (i=1,2,⋯,n)，则

$$\mathbf{y}_i^{\mathbf{H}}\mathbf{y}_j=\left(\mathbf{x}_i^{\mathbf{H}}\mathbf{G}\right)\left(\mathbf{G}^{\mathbf{H}}\mathbf{x}_j\right)=\mathbf{x}_i^{\mathbf{H}}\mathbf{Bx}_j=\delta_{ij}=\begin{cases}1 & i=j\\ 0 & i\neq j\end{cases} \qquad （带权正交）$$

二、 瑞利商

A、B 为 n 阶厄米矩阵，且 B 正定，称$\mathbf{R}(\mathbf{x})=\dfrac{\mathbf{x}^{\mathbf{H}}\mathbf{Ax}}{\mathbf{x}^{\mathbf{H}}\mathbf{Bx}}(\mathbf{x}\neq\mathbf{0})$为 A 相对于 B 的瑞利

商。

$\mathbf{x_1, x_2 \cdots, x_n}$ 线性无关，所以，$\forall \mathbf{x} \in \mathbf{C}^n$，存在 $\mathbf{a_1, a_2 \cdots, a_n} \in \mathbf{C}$，使得

$$\mathbf{x} = \sum_{i=1}^{n} \mathbf{a_i x_i}$$

$$\mathbf{x^H B x} = \left(\sum_{i=1}^{n} \mathbf{a_i x_i}\right)^H \mathbf{B}\left(\sum_{j=1}^{n} \mathbf{a_j x_j}\right) = \sum_{i,j=1}^{n} \overline{\mathbf{a_i}} \mathbf{a_j x_i^H B x_j} = \sum_{i=1}^{n} |\mathbf{a_i}|^2$$

$$\mathbf{x^H A x} = \sum_{i,j=1}^{n} \overline{\mathbf{a_i}} \mathbf{a_j x_i^H A x_j} = \sum_{i,j=1}^{n} \overline{\mathbf{a_i}} \mathbf{a_j x_i^H} \lambda_i \mathbf{B x_j} = \sum_{i=1}^{n} \lambda_i |\mathbf{a_i}|^2$$

$$\therefore \quad \mathbf{R(x)} = \frac{\sum_{i=1}^{n} \lambda_i |\mathbf{a_i}|^2}{\sum_{i=1}^{n} |\mathbf{a_i}|^2}$$

● $\min\limits_{x \neq 0} \mathbf{R(x)} = \lambda_1 \qquad \max\limits_{x \neq 0} \mathbf{R(x)} = \lambda_n$

证明：$\mathbf{R(x)} = \dfrac{\mathbf{x^H A x}}{\mathbf{x^H B x}} = \dfrac{(\mathbf{kx})^H \mathbf{A}(\mathbf{kx})}{(\mathbf{kx})^H \mathbf{B}(\mathbf{kx})}$ k 为非零常数

可取 $\mathbf{k} = \dfrac{1}{\|\mathbf{x}\|}$， $\|\mathbf{kx}\| = 1$

$$\therefore \quad \mathbf{R(x)} = \left.\frac{\mathbf{x^H A x}}{\mathbf{x^H B x}}\right|_{\|x\|=1}$$ （闭区域）

当 $\mathbf{x} = \mathbf{x_1}$ 或 $\mathbf{a_i} = 0 (\mathbf{i} = 2, 3, \cdots, \mathbf{n})$ 时，$\mathbf{R(x)} = \lambda_1$

$$\lambda_i \geq \lambda_1 \qquad \mathbf{R(x)} \geq \lambda_1 \frac{\sum_{i=1}^{n} |\mathbf{a_i}|^2}{\sum_{i=1}^{n} |\mathbf{a_i}|^2} = \lambda_1$$

$$\therefore \min_{x \neq 0} \mathbf{R(x)} = \lambda_1$$

另一方面，$\lambda_i \leq \lambda_n \qquad \mathbf{R(x)} \leq \lambda_n \dfrac{\sum_{i=1}^{n} |\mathbf{a_i}|^2}{\sum_{i=1}^{n} |\mathbf{a_i}|^2} = \lambda_n$

$$\therefore \quad \max_{x\neq 0} R(x) = \lambda_n$$

［证毕］

当 B＝I 时，标准特征值问题 $Ax = \lambda x$ （$A^H = A$）

$$\begin{cases} \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n \\ x_i^H x_j = \delta_{ij} \end{cases}$$

则 $\min\limits_{(x\neq 0)} \dfrac{x^H A x}{x^H x} = \lambda_1 \qquad \max\limits_{(x\neq 0)} \dfrac{x^H A x}{x^H x} = \lambda_n$

进一步分析可得

$$\min_{x\neq 0} R(x)\Big|_{a_1=0} = \lambda_2 \qquad\qquad \max_{x\neq 0} R(x)\Big|_{a_n=0} = \lambda_{n-1}$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$\min_{x\neq 0} R(x)\Big|_{a_1=a_2=\cdots=a_k=0} = \lambda_{k+1} \qquad \max_{x\neq 0} R(x)\Big|_{a_n=a_{n-1}=\cdots=a_{n-k}=0} = \lambda_{n-k-1}$$

定理 1. 设 $L = \mathrm{span}\{x_r, x_{r+1}, \cdots, x_s\}$ $\left(\lambda_r \le \lambda_{r+1} \le \cdots \le \lambda_s\right)$ , 则

$$\min_{\substack{x\neq 0\\ x\in L}} R(x) = \lambda_r \qquad \max_{\substack{x\neq 0\\ x\in L}} R(x) = \lambda_s$$

这一结果不便于应用，希望对上述结果进行改造，改造成不依赖于 $x_i$ 的一种表达方式。

$a_1 = 0$ 和 $a_n = 0$ 的情况均对应于 x 在（n-1）维的子空间内变动，x 在 L 中变动是在一个(s-r+1)维子空间中变化。

一般的，x 在 $C^n$ 的(n-1)维子空间 $V_{n-1}$ 中变动时，

$$\min_{\substack{x\neq 0\\ x\in V_{n-1}}} R(x) \le \lambda_2 \qquad \max_{\substack{x\neq 0\\ x\in V_{n-1}}} R(x) \ge \lambda_{n-1}$$

即，对于不同的 $V_{n-1}$，$R(x)$ 的最小值及最大值有可能不同，其中各个最小值中最大者为 $\lambda_2$，各个最大值中的最小者为 $\lambda_{n-1}$

$$\max_{V_{n-1}\in C^n}\left[\min_{\substack{x\neq 0\\ x\in V_{n-1}}} R(x)\right] = \lambda_2 \qquad \min_{V_{n-1}\in C^n}\left[\max_{\substack{x\neq 0\\ x\in V_{n-1}}} R(x)\right] = \lambda_{n-1}$$

定理 2. 设 $V_k$ 是 $C^n$ 的一个 k 维子空间，则

$$\max_{V_k\in C^n}\left[\min_{\substack{x\neq 0\\ x\in V_k}} R(x)\right]=\lambda_{n-k+1}\qquad \min_{V_k\in C^n}\left[\max_{\substack{x\neq 0\\ x\in V_k}} R(x)\right]=\lambda_k$$

以上两式称为广义特征值的极小极大原理。

- B＝I 时，标准特征值问题同样存在上述关系。
- 矩阵奇异值问题：$\left[\sigma(A)\right]^2=\lambda\left(A^H A\right)$　（非零）

$$R(x)=\frac{x^H\left(A^H A\right)x}{x^H x}=\frac{\|Ax\|_2^2}{\|x\|_2^2}$$

$$\sigma_{n-k+1}=\max_{V_k\in C^n}\left[\min_{\substack{x\neq 0\\ x\in V_k}}\frac{\|Ax\|_2}{\|x\|_2}\right]$$

$$\sigma_k=\min_{V_k\in C^n}\left[\max_{\substack{x\neq 0\\ x\in V_k}}\frac{\|Ax\|_2}{\|x\|_2}\right]$$

1．线性空间的定义：

设$\boldsymbol{V}$是一个非空集合，其元素用$\boldsymbol{x,y,z}$等表示；$\boldsymbol{K}$是一个数域，其元素用$\boldsymbol{k,l,m}$等表示。如果$\boldsymbol{V}$满足[如下 8 条性质，分两类]：

（I）在$\boldsymbol{V}$中定义一个“加法”运算，即当$\boldsymbol{x}$，$\boldsymbol{y} \in \boldsymbol{V}$时，有唯一的和$\boldsymbol{x}+\boldsymbol{y} \in \boldsymbol{V}$（封闭性），且加法运算满足下列性质：

（1）结合律　$\boldsymbol{x}+(\boldsymbol{y}+\boldsymbol{z})=(\boldsymbol{x}+\boldsymbol{y})+\boldsymbol{z}$；

（2）交换律　$\boldsymbol{x}+\boldsymbol{y}=\boldsymbol{y}+\boldsymbol{x}$；

（3）零元律　存在零元素$\boldsymbol{O}$，使$\boldsymbol{x}+\boldsymbol{O}=\boldsymbol{x}$；

（4）负元律　对于任一元素$\boldsymbol{x} \in \boldsymbol{V}$，存在一元素$\boldsymbol{y} \in \boldsymbol{V}$，使$\boldsymbol{x}+\boldsymbol{y}=\boldsymbol{O}$，且称$\boldsymbol{y}$为$\boldsymbol{x}$的负元素，记为$(-\boldsymbol{x})$。则有$\boldsymbol{x}+(-\boldsymbol{x})=\boldsymbol{O}$。

（II）在$\boldsymbol{V}$中定义一个“数乘”运算，即当$\boldsymbol{x} \in \boldsymbol{V}, \boldsymbol{k} \in \boldsymbol{K}$时，有唯一的$\boldsymbol{kx} \in \boldsymbol{V}$（封闭性），且数乘运算满足下列性质：

（5）数因子分配律　$\boldsymbol{k}(\boldsymbol{x}+\boldsymbol{y})=\boldsymbol{kx}+\boldsymbol{ky}$；

（6）分配律　$(\boldsymbol{k}+\boldsymbol{l})\boldsymbol{x}=\boldsymbol{kx}+\boldsymbol{lx}$；

（7）结合律　$\boldsymbol{k}(\boldsymbol{lx})=(\boldsymbol{kl})\boldsymbol{x}$；

（8）恒等律　$\boldsymbol{1x}=\boldsymbol{x}$；　［数域中一定有$\boldsymbol{1}$］

则称$\boldsymbol{V}$为数域$\boldsymbol{K}$上的线性空间。

注意以下几点：

1）线性空间是基于一定数域来的。同一个集合，对于不同数域，就可能构成不同的线性空间，甚至对有的数域能构成线性空间，而对其他数域不能构成线性空间。

2）两种运算、八条性质。数域$\boldsymbol{K}$中的运算是具体的四则运算，而$\boldsymbol{V}$中所定义的加法运算和数乘运算则是抽象的、形式的。

3）除了两种运算和八条性质外，还应注意唯一性、封闭性是否满足。

当数域$\boldsymbol{K}$为实数域时，$\boldsymbol{V}$就称为实线性空间；$\boldsymbol{K}$为复数域，$\boldsymbol{V}$就称为复线性空间。

2．线性相关性

线性空间中相关性概念与线性代数中向量组线性相关性概念类似。

•线性组合：$\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \cdots, \boldsymbol{x}_m \in \boldsymbol{V}, \boldsymbol{c}_1, \boldsymbol{c}_2 \cdots, \boldsymbol{c}_m \in \boldsymbol{K}$

$$\boldsymbol{c}_1\boldsymbol{x}_1+\boldsymbol{c}_2\boldsymbol{x}_2+\cdots+\boldsymbol{c}_m\boldsymbol{x}_m=\sum_{i=1}^{m}\boldsymbol{c}_i\boldsymbol{x}_i$$

称为元素组$\boldsymbol{x}_1, \boldsymbol{x}_2 \cdots, \boldsymbol{x}_m$的一个线性组合。

•线性表示：$\boldsymbol{V}$中某个元素$\boldsymbol{x}$可表示为其中某个元素组的线性组合，则称$\boldsymbol{x}$可由该元素

组线性表示。

•线性相关性：如果存在一组不全为零的数 $\boldsymbol{c_1, c_2 \cdots, c_m \in K}$，使得对于元素 $\boldsymbol{x_1, x_2 \cdots, x_m \in V}$ 有

$$\sum_{i=1}^{m} c_i x_i = 0$$

则称元素组 $\boldsymbol{x_1, x_2 \cdots, x_m}$ 线性相关，否则称其线性无关。线性相关性概念是个非常重要的概念，有了线性相关性才有下面的线性空间的维数、基和坐标。

3．线性空间的维数

定义：线性空间 $\boldsymbol{V}$ 中最大线性无关元素组所含元素个数称为 $\boldsymbol{V}$ 的维数，记为 $\mathbf{dim}\,\boldsymbol{V}$ 。

4、基的定义：设 $\boldsymbol{V}$ 是数域 $\boldsymbol{K}$ 上的线性空间，$\boldsymbol{x_1, x_2 \cdots, x_r (r \geq 1)}$ 是属于 $\boldsymbol{V}$ 的 $\boldsymbol{r}$ 个任意元素，如果它满足

（1）$\boldsymbol{x_1, x_2 \cdots, x_r}$ 线性无关；

（2）$\boldsymbol{V}$ 中任一向量 $x$ 均可由 $\boldsymbol{x_1, x_2 \cdots, x_r}$ 线性表示。

则称 $\boldsymbol{x_1, x_2 \cdots, x_r}$ 为 $\boldsymbol{V}$ 的一个基，并称 $\boldsymbol{x_1, x_2 \cdots, x_r}$ 为该基的基元素。

•基正是 $\boldsymbol{V}$ 中最大线性无关元素组；$\boldsymbol{V}$ 的维数正是基中所含元素的个数。

•基是不唯一的，但不同的基所含元素个数相等。

5、坐标的定义：称线性空间 $\boldsymbol{V^n}$ 的一个基 $\boldsymbol{x_1, x_2 \cdots, x_n}$ 为 $\boldsymbol{V^n}$ 的一个坐标系，$\boldsymbol{\forall x \in V^n}$，它在该基下的线性表示为：

$$\sum_{i=1}^{n} \xi_i x_i \quad (\xi_i \in K, x_i \in V^n, i = 1,2 \cdots, n)$$

则称 $\boldsymbol{\xi_1, \xi_2 \cdots, \xi_n}$ 为 $\boldsymbol{x}$ 在该坐标系中的坐标或分量，记为 $\boldsymbol{(\xi_1, \xi_2 \cdots, \xi_n)^T}$

讨论：（1）一般来说，线性空间及其元素是抽象的对象，不同空间的元素完全可以具有千差万别的类别及性质。但坐标表示却把它们统一了起来，坐标表示把这种差别留给了基和基元素，由坐标所组成的新向量仅由数域中的数表示出来。

（2）更进一步，原本抽象的“加法”及 “数乘”经过坐标表示就演化为向量加法及数对向量的数乘。

1°
$$\begin{aligned} x + y &= (\xi_1 x_1 + \xi_2 x_2 + \cdots + \xi_n x_n) + (\eta_1 x_1 + \eta_2 x_2 + \cdots + \eta_n x_n) \\ &= (\xi_1 + \eta_1) x_1 + (\xi_2 + \eta_2) x_2 + \cdots + (\xi_n + \eta_n) x_n \end{aligned}$$

正对应

$$\begin{cases} x = (\xi_1, \xi_2 \cdots, \xi_n) \\ y = (\eta_1, \eta_2 \cdots, \eta_n) \end{cases} \rightarrow x + y = (\xi_1 + \eta_1, \xi_2 + \eta_2 \cdots, \xi_n + \eta_n)$$

$2^{\circ}$ $kx = k(\xi x + \xi x + \cdots + \xi x) = (k\xi_1)x_1 + (k\xi_2)x_2 + \cdots + (k\xi_n)x_n$

$\rightarrow (k\xi_1, k\xi_2 \cdots, k\xi_n)$

正对应 $x = (\xi_1, \xi_2 \cdots, \xi_n) \rightarrow kx = (k\xi_1, k\xi_2 \cdots, k\xi_3)$

（3）显然，同一元素在不同坐标系中的坐标是不同的。后面我们还要研究这一变换关系。

6、定义：设$V_1$是数域$K$上的线性空间$V$的一个非空子集合，且对$V$已有的线性运算满足以下条件

（1） 如果$x, y \in V_1$，则$x + y \in V_1$；

（2） 如果$x \in V_1$，$k \in K$，则$kx \in V_1$，

则称$V_1$是$V$的一个线性子空间或子空间。

7.定义：设$V_1$、$V_2$是线性空间$V$的两个子空间，则

$$V_1 \cap V_2 = \{x | x \in V_1, x \in V_2\}$$

$$V_1 + V_2 = \{x + y | x \in V_1, y \in V_2\}$$

分别称为$V_1$和$V_2$的交与和。

定理：若$V_1$和$V_2$是线性空间$V$的两个子空间，则$V_1 \cap V_2$，$V_1 + V_2$均为$V$的子空间

8. 向量范数定义：设 V 为数域 K 上的向量空间，若对于 V 的任一向量 x，对应一个实值函数$\|\mathbf{x}\|$，并满足以下三个条件：

（1）非负性 $\|\mathbf{x}\| \geq 0$，等号当且仅当 x=0 时成立；

（2）齐次性 $\|\alpha\mathbf{x}\| = |\alpha| \|\mathbf{x}\|, \alpha \in \mathbf{k}, \mathbf{x} \in \mathbf{V};$

（3）三角不等式$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|, \mathbf{x}, \mathbf{y} \in \mathbf{V}$。

则称$\|\mathbf{x}\|$为 V 中向量 x 的范数，简称为向量范数。

9. 矩阵范数定义：设$\mathbf{k}^{\mathbf{m} \times \mathbf{n}} (\mathbf{k} = \mathbf{c}或\mathbf{R})$表示数域 k 上全体$\mathbf{m} \times \mathbf{n}$阶矩阵的集合。若对于$\mathbf{k}^{\mathbf{m} \times \mathbf{n}}$中任一矩阵 A，均对应一个实值函数，并满足以下四个条件：

（1）非负性：$\|\mathbf{A}\| \geq 0$ ，等号当且仅当 A=0 时成立；

（2）齐次性：$\|\alpha \mathbf{A}\| = |\alpha| \|\mathbf{A}\|, \alpha \in \mathbf{k};$

（3）三角不等式：$\|\mathbf{A}+\mathbf{B}\| \le \|\mathbf{A}\| + \|\mathbf{B}\|, \mathbf{A},\mathbf{B} \in \mathbf{k}^{m\times n}$

则称$\|\mathbf{A}\|$为广义矩阵范数；

（4）相容性： $\|\mathbf{AB}\| \le \|\mathbf{A}\| \|\mathbf{B}\|$

则称$\|\mathbf{A}\|$为矩阵范数。

范数计算：

矩阵和向量的相容性：

10 特征值、特征向量．定义：对$\boldsymbol{m}$阶方阵$\boldsymbol{A}$，若存在数$\boldsymbol{\lambda}$，及非零向量（列向量）$\boldsymbol{x}$，使得$\boldsymbol{Ax}=\boldsymbol{\lambda x}$,则称$\boldsymbol{\lambda}$为$\boldsymbol{A}$的特征值，$\boldsymbol{x}$为$\boldsymbol{A}$的属于特征值$\boldsymbol{\lambda}$的特征向量。

- 特征向量不唯一
- 特征向量非零
- $(\lambda I - A)x = 0$有非零解，则$\det(\lambda I - A) = 0$，称$\det(\lambda I - A)$为$\boldsymbol{A}$的特征多项式。

11．一般矩阵的奇异值分解

定理：设$\mathbf{A} \in \mathbf{C}_r^{m\times n}$，则存在$\mathbf{m}$阶酉矩阵$\mathbf{U}$及$\mathbf{n}$阶酉矩阵$\mathbf{V}$，使

$$\mathbf{U}^H\mathbf{A}\mathbf{V} = \begin{array}{c} \left[\begin{array}{cccc|c} \sigma_1 & & & \mathbf{0} & \\ & \sigma_2 & & & \mathbf{O} \\ & & \ddots & & \\ \mathbf{0} & & & \sigma_r & \\ \hline & \mathbf{O} & & & \mathbf{O} \end{array}\right] \begin{array}{l} \\ \\ \mathbf{r}\text{行} \\ \\ (\mathbf{m}-\mathbf{r})\text{行} \end{array} \\ \mathbf{r}\text{列} \qquad (\mathbf{n}-\mathbf{r})\text{列} \end{array}$$

即

$$\mathbf{A} = \mathbf{U}\begin{bmatrix} \sigma_1 & & & & \mathbf{O} \\ & \sigma_2 & & & \\ & & . & & \\ & & & \sigma_r & \\ \mathbf{O} & & & & \mathbf{O} \end{bmatrix}\mathbf{V}^H$$

12、Penrose 定义：设 A∈C$^{\mathbf{m\times n}}$，若 Z∈$C^{\mathbf{n\times m}}$且使如下四个等式成立，

AZA = A,  ZAZ = Z,  $(AZ)^{\mathbf{H}}$ = AZ ,  $(ZA)^{\mathbf{H}}$ = ZA

则称 Z 为 A 的 Moore-Penrose(广义)逆，记为，$A^{\dagger}$。

而上述四个等式又依次称为 Penrose 方程（i）,(ii) ,(iii) ,(iv)。

13、正弦函数、余弦函数的计算

利用 Jordan 标准形求矩阵函数

利用零化多项式求解矩阵函数

14、投影变换算子

设 L，M 为 $\mathbf{C}^n$ 的子空间并构成直和 $\mathbf{L}+\mathbf{M}=\mathbf{L}\oplus\mathbf{M}=\mathbf{C}^n$. 即

$\forall \mathbf{x}\in\mathbf{C}^n$，$\exists$ 唯一的 $\mathbf{y}\in\mathbf{L},\mathbf{z}\in\mathbf{M}$ 使 x=y+z

称 y 为 x 沿着 M 到 L 的投影。

定义:将任意 $\mathbf{x}\in\mathbf{C}^n$ 变为其沿着 M 到 L 的投影的变换称为沿着 M 到 L 的投影算子,记为 $\mathbf{P}_{\mathbf{L},\mathbf{M}}$

即 $\mathbf{P}_{\mathbf{L},\mathbf{M}}\ \mathbf{x}=\mathbf{y}\in\mathbf{L}$,投影算子是线性变换，其矩阵称为投影矩阵,仍记为 $\mathbf{P}_{\mathbf{L},\mathbf{M}}$。

充要条件：n 阶方阵 P 为正交投影矩阵的充要条件是 P 为幂等的厄米矩阵。

定理：n 阶方阵 P 成为投影矩阵的充要条件是 P 为幂等矩阵。

投影矩阵的构造

15、盖尔圆法

定义：设 $\mathbf{A}=\left(\mathbf{a}_{\mathbf{ij}}\right)_{\mathbf{n}\times\mathbf{n}}\in\mathbf{C}^{\mathbf{n}\times\mathbf{n}}$,由方程 $\left|\mathbf{z}-\mathbf{a}_{\mathbf{ii}}\right|\le\mathbf{R}_{\mathbf{i}}=\sum_{\substack{\mathbf{j}=1\\ \mathbf{i}\neq\mathbf{j}}}^{\mathbf{n}}\left|\mathbf{a}_{\mathbf{ij}}\right|$ 所确定的圆称为 A 的第 i 个盖尔圆，$\mathbf{R}_{\mathbf{i}}$ 称为盖尔圆的半径。

定理 3：矩阵 A 的所有特征值均落在它的所有盖尔圆的并集之中。

推论 1. 孤立盖尔圆中恰好包含一个特征值。

推论 2. 实矩阵的孤立盖尔圆恰好包含一个实特征值。

推论 3. 盖尔圆方法中盖尔圆半径可以按列求和。(因为方阵转置后特征值不变)

推论 4. 盖尔圆半径变为 $\mathbf{R}_{\mathbf{i}}=\sum_{\mathbf{j}=1}^{\mathbf{n}}{}'\,\dfrac{\alpha_{\mathbf{i}}}{\alpha_{\mathbf{j}}}\left|\mathbf{a}_{\mathbf{ij}}\right|$，两个盖尔圆定理仍然成立。

# 复习课

## 线性空间

一、 线性空间的定义及性质和基本概念

一个集合定义了一种加法运算，且与一个数域相联系定义了数乘运算，满足一定的条件，构成线性空间。

基本概念：线性相关、线性无关、基、坐标、线性空间的维数、线性子空间、子空间和（交、直和）、生成子空间

线性空间的基与坐标

线性变换定义及其运算

线性变换及其矩阵

矩阵变换、投影变换

线性变换的矩阵表示

范数定义，矩阵范数的相容性条件

二、 对角化与 Jordan 标准形

三、 矩阵级数与矩阵函数计算

四、 矩阵分解、约当标准型，计算矩阵函数（会算）

五、 广义逆、满秩分解、最小二乘、正交投影（基本概念）

六、 矩阵特征值估计

盖尔圆法（基本概念）

1、 线性变换

2、 向量数目、解空间维数

3、 线性空间向量之间关系

4、 矩阵范数相容性

5、 特征值分解，公式 Ax=nx

6、 SVD 去掉一个奇异值、向量，范数、矩阵范数

7、 广义逆

8、 矩阵函数计算

9、 投影变换算子

10、 盖尔圆