机器学习与模式识别

第0讲 数学基础提要

2020~2021学年

相关数学基础

M1　齐次线性方程组的求解(基础解系)

M2　向量的内积、方阵的特征分解

M3　向量范数与矩阵范数

M4　多元函数对矩阵变量的求导

M5　矩阵的QR分解与奇异值分解

M6　投影矩阵、广义逆矩阵、正定矩阵

(概率论与数理统计：自行复习)

M1-线性方程组求解（一）齐次线性方程组解的性质

1．解向量的概念

设有齐次线性方程组

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

若记

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

则上述方程组（1）可写成向量方程

$$Ax=0.$$

若 $x_1=\xi_{11}, x_2=\xi_{21}, \cdots, x_n=\xi_{n1}$ 为方程 $Ax=0$ 的

解，则

$$
\boldsymbol{x} = \xi_1 = \begin{pmatrix} \xi_{11} \\ \xi_{21} \\ \vdots \\ \xi_{n1} \end{pmatrix}
$$

称为方程组(1) 的**解向量**，它也就是向量方程(2)的解.

2．齐次线性方程组解的性质

（1）若 $x=\xi_1, x=\xi_2$ 为 $Ax=0$ 的解，则

$$x=\xi_1+\xi_2$$

也是 $Ax=0$ 的解.

（2）若 $x=\xi_1$ 为 $Ax=0$的解， k 为实数，则 $x=k\xi_1$也是$Ax=0$的解.

由以上两个性质可知，方程组的全体解向量所组成的集合，对于加法和数乘运算是封闭的，因此构成一个向量空间，称此向量空间为齐次线性方程组 $Ax=0$ 的解空间.

1. 基础解系的定义

$\eta_1,\eta_2,\cdots,\eta_t$称为齐次线性方程组 $Ax=0$的基础解系,如果

(1)$\eta_1,\eta_2,\cdots,\eta_t$是$Ax=0$的一组线性无关的解;

(2)$Ax=0$的任一解都可由$\eta_1,\eta_2,\cdots,\eta_t$线性表出.

如果$\eta_1,\eta_2,\cdots,\eta_t$为齐次线性方程组$Ax=0$的一组基础解系,那么,$Ax=0$的通解可表示为

$$x = k_1\eta_1 + k_2\eta_2 + \cdots + k_t\eta_t$$

其中$k_1,k_2,\cdots,k_{n-r}$是任意常数.

2．齐次线性方程组基础解系的求法

设齐次线性方程组的系数矩阵为 A，并不妨设A的前 r 个列向量线性无关.于是 A 可化为

$$A \sim \begin{pmatrix} 1 & \cdots & 0 & b_{11} & \cdots & b_{1,n-r} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 1 & b_{r1} & \cdots & b_{r,n-r} \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \end{pmatrix}$$

$$Ax = 0 \Leftrightarrow \begin{pmatrix} 1 & \cdots & 0 & b_{11} & \cdots & b_{1,n-r} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 1 & b_{r1} & \cdots & b_{r,n-r} \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ \vdots \\ x_n \end{pmatrix} = 0$$

$$\Leftrightarrow \begin{cases} x_1 = -b_{11} x_{r+1} - \cdots - b_{1,n-r} x_n \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_r = -b_{r1} x_{r+1} - \cdots - b_{r,n-r} x_n \end{cases}$$

现对$x_{r+1},\cdots,x_n$取下列 $n-r$ 组数：

$$\begin{pmatrix} x_{r+1} \\ x_{r+2} \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \cdots, \quad \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

分别代入$\begin{cases} x_1 = -b_{11}x_{r+1} - \cdots - b_{1,n-r}x_n \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_r = -b_{r1}x_{r+1} - \cdots - b_{r,n-r}x_n \end{cases}$

依次得 $\begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix} = \begin{pmatrix} -b_{11} \\ \vdots \\ -b_{r1} \end{pmatrix}, \begin{pmatrix} -b_{12} \\ \vdots \\ -b_{r2} \end{pmatrix}, \cdots, \begin{pmatrix} -b_{1,n-r} \\ \vdots \\ -b_{r,n-r} \end{pmatrix}.$

从而求得原方程组的 $n-r$ 个解:

$$\xi_1 = \begin{pmatrix} -b_{11} \\ \vdots \\ -b_{r1} \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \xi_2 = \begin{pmatrix} -b_{12} \\ \vdots \\ -b_{r2} \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \cdots, \xi_{n-r} = \begin{pmatrix} -b_{1,n-r} \\ \vdots \\ -b_{r,n-r} \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

1 向量内积的定义

定义 设有n维向量

$$x=\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},\quad y=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

令$[x,y]=x_1y_1+x_2y_2+\cdots+x_ny_n$，$[x,y]$称为向量$x$与$y$的内积.

内积的矩阵表示

$$[x, y] = x^T y,$$

其中x, y都是列向量.

内积满足下列运算规律(其中x, y, z为n维向量,λ为实数):

(1)$[x, y] = [y, x]$;

(2)$[\lambda x, y] = \lambda[x, y]$;

(3)$[x + y, z] = [x, z] + [y, z]$.

2　向量的长度

定义　令

$$\|x\| = \sqrt{[x, x]} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2},$$

$\|x\|$称为n维向量x的长度(或范数).

向量的长度具有下列性质:

(1)非负性　当$x \neq 0$时,$\|x\| > 0$;当$x = 0$时,$\|x\| = 0$;

(2)齐次性　$\|\lambda x\| = |\lambda| \|x\|$;

(3)三角不等式　$\|x + y\| \leq \|x\| + \|y\|$.

当$\|x\| = 1$时，称x为单位向量.

向量的内积满足施瓦茨不等式

$$[x,y]^2 \le [x,x][y,y],$$

从而有 $$\left|\frac{[x,y]}{\|x\|\|y\|}\right| \le 1, \quad (\text{当}\|x\|\|y\| \ne 0\text{时}).$$

3　向量的夹角

定义　当$\|x\| \neq 0, \|y\| \neq 0$时,

$$\theta = \arccos\frac{[x, y]}{\|x\|\|y\|}$$

称为n维向量x与y的夹角.

当$[x, y] = 0$时,称向量x与y正交.

若$x = 0$,则x与任何向量都正交.

4　正交向量组的性质

所谓正交向量组，是指一组两两正交的非零向量．向量空间的基若是正交向量组，就称为正交基.

定理　若n维向量$a_1,a_2,\cdots,a_r$是一组两两正交的非零向量,则$a_1,a_2,\cdots,a_r$线性无关.

定义　设n维向量$e_1,e_2,\cdots,e_r$是向量空间$V(V\subset R^n)$的一个基,如果$e_1,e_2,\cdots,e_r$两两正交,则称$e_1,e_2,\cdots,e_r$是V的一个规范正交基.

若$e_1, e_2, \cdots, e_r$是V的一个规范正交基，那么V中任一向量a都可表为

$$a = \lambda_1 e_1 + \lambda_2 e_2 + \cdots + \lambda_r e_r,$$

其中 $\lambda_i = e_i^T a = [a, e_i], (i = 1, 2, \cdots, r).$

施密特正交化方法

设$a_1, a_2, \cdots, a_r$是向量空间V的一个基，要求V的一个规范正交基，只需把$a_1, a_2, \cdots, a_r$这个基规范正交化.

第一步　正交化

取　　$b_1 = a_1;$

$$b_2 = a_2 - \frac{[b_1, a_2]}{[b_1, b_1]} b_1;$$

$$\cdots\cdots\cdots$$

$$b_r = a_r - \frac{[b_1, a_r]}{[b_1, b_1]} b_1 - \frac{[b_2, a_r]}{[b_2, b_2]} b_2 - \cdots - \frac{[b_{r-1}, a_r]}{[b_{r-1}, b_{r-1}]} b_{r-1}.$$

则$b_1, b_2, \cdots, b_r$两两正交,且与$a_1, a_2, \cdots, a_r$等价.

第二步　单位化

取　$\boldsymbol{e}_1=\dfrac{1}{\|\boldsymbol{b}_1\|}\boldsymbol{b}_1,\boldsymbol{e}_2=\dfrac{1}{\|\boldsymbol{b}_2\|}\boldsymbol{b}_2,\cdots,\boldsymbol{e}_r=\dfrac{1}{\|\boldsymbol{b}_r\|}\boldsymbol{b}_r$,就得$V$的一个规范正交基.

5　正交矩阵与正交变换

定义　如果n阶矩阵A满足

$$A^T A = E \qquad (\text{即} A^{-1} = A^T),$$

那么称A为正交矩阵.

方阵 A 为正交矩阵的充分必要条件是 A的行（列）向量都是单位向量，且两两正交.

正交矩阵A的n个列(行)向量构成向量空间 R^n 的一个规范正交基.

定义　若 P 为正交矩阵，则线性变换 $y = Px$ 称为正交变换.

正交变换的特性在于保持线段的长度不变.

设 $y = Px$ 为正交变换,则有

$$\|y\| = \sqrt{y^T y} = \sqrt{x^T P^T P x} = \sqrt{x^T x} = \|x\|.$$

定义 设A是n阶矩阵,如果数λ和n维非零列向量x使关系式

$$Ax = \lambda x$$

成立,那么,这样的数λ称为方阵A的特征值,非零向量x称为A的对应于特征值λ的特征向量.

$|A-\lambda E| = 0$称为方阵A的特征方程.

$f(\lambda) = |A-\lambda E|$称为方阵$A$的特征多项式.

n阶方阵A有n个特征值.若$A=(a_{ij})$的特征值为$\lambda_1,\lambda_2,\cdots,\lambda_n$,则有

(1) $\lambda_1+\lambda_2+\cdots+\lambda_n=a_{11}+a_{22}+\cdots+a_{nn}$;

(2) $\lambda_1\lambda_2\cdots\lambda_n=|A|$.

1、有关特征值的一些结论

设λ是$A=(a_{ij})_{n\times n}$的特征值,则

(1)λ也是A^T的特征值;

(2)λ^k是A^k的特征值(k为任意自然数);$\varphi(\lambda)$是$\varphi(A)$的特征值.其中$\varphi(\lambda)=a_0+a_1\lambda+\cdots+a_m\lambda^m$,

$\varphi(A)=a_0E+a_1A+\cdots+a_mA^m$.

(3)当A可逆时,$\frac{1}{\lambda}$是A^{-1}的特征值;$\frac{1}{\lambda}\cdot|A|$是$A^*$的特征值.

2、有关特征向量的一些结论

定理 设 $\lambda_1, \lambda_2, \cdots, \lambda_m$ 是方阵A的m个特征值, $p_1, p_2, \cdots, p_m$ 依次是与之对应的特征向量,如果 $\lambda_1, \lambda_2, \cdots, \lambda_m$ 各不相等,则 $p_1, p_2, \cdots, p_m$ 线性无关.

即属于不同特征值的特征向量是线性无关的.

定理 属于同一个特征值的特征向量的非零线性组合仍是属于这个特征值的特征向量.

3、特征值与特征向量的求法

第一步　计算A 的特征多项式；

第二步　求出特征方程的全部根，即得A 的全部特征值；

第三步　将每一个特征值代入相应的线性方程组，求出基础解系，即得该特征值的特征向量.

例 计算3阶实矩阵$A=\begin{pmatrix}3&2&4\\2&0&2\\4&2&3\end{pmatrix}$的全部特征值和特征向量.

解 第一步 计算A的特征多项式

$$f(\lambda)=\left|\lambda E-A\right|=\begin{vmatrix}\lambda-3&-2&-4\\-2&\lambda&-2\\-4&-2&\lambda-3\end{vmatrix}$$

$$=(\lambda-8)(\lambda+1)^2.$$

第二步　求出特征多项式$f(\lambda)$的全部根,即A的全部特征值.

令$f(\lambda)=0$,解之得$\lambda_1=8,\lambda_2=\lambda_3=-1$,为$A$的全部特征值.

第三步　求出A的全部特征向量

对$\lambda_1=8$,求相应线性方程组$(\lambda_1 E-A)x=0$的一个基础解系.

$$\begin{cases} 5x_1 - 2x_2 - 4x_3 = 0, \\ -2x_1 + 8x_2 - 2x_3 = 0, \\ -4x_1 - 2x_2 + 5x_3 = 0, \end{cases}$$

化简求得此方程组的一个基础解系

$$\alpha_1 = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}.$$

属于$\lambda_1 = 8$的全部特征向量为$k_1\alpha_1$($k_1 \neq 0$为实数).

同理对$\lambda_2=\lambda_3=-1$,求相应线性方程组$(\lambda_2 E-A)x=0$的一个基础解系:

$$\begin{cases}-4x_1-2x_2-4x_3=0,\\ -2x_1-x_2-2x_3=0,\\ -4x_1-2x_2-4x_3=0,\end{cases}$$

求解得此方程组的一个基础解系:

$$\alpha_2=\begin{pmatrix}1\\0\\-1\end{pmatrix},\qquad \alpha_2=\begin{pmatrix}1\\-2\\0\end{pmatrix}.$$

于是A的属于$\lambda_2=\lambda_3=-1$的全部特征向量为

$$k_2\alpha_2+k_3\alpha_3,$$

k_2,k_3是不全为零的实数.

从而A的全部特征向量为$k_1\alpha_1;k_2\alpha_2+k_3\alpha_3$,这里$k_1\neq 0$为实数,$k_2,k_3$是不全为零的实数.

定义 设A,B都是n阶矩阵,若有可逆矩阵P,使

$$P^{-1}AP=B,$$

则称B是A的相似矩阵,或说矩阵A与B相似.

对A进行运算 $P^{-1}AP$称为对A进行相似变换,可逆矩阵P称为把A变成B的相似变换矩阵.

矩阵之间的相似具有(1)自反性；(2)对称性；(3)传递性.

1、有关相似矩阵的性质

(1)若 A 与 B 相似，则 A 与 B 的特征多项式相同，从而 A 与 B 的特征值亦相同.

(2)若 A 与对角矩阵

$$\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

相似,则 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 是 A 的 n 个特征值.

(3)若$A=PBP^{-1}$,则$A^k=PB^kP^{-1}$,

$\varphi(A)=P\varphi(B)P^{-1}$.

特别地,若有可逆阵P,使$P^{-1}AP=\Lambda$为对角阵,

则有$A^k=P\Lambda^kP^{-1}$,$\varphi(A)=P\varphi(\Lambda)P^{-1}$.

(4) A 能对角化的充分必要条件是 A有 n个线性无关的特征向量.

(5) A有n个互异的特征值，则A与对角阵相似.

2、实对称矩阵的相似矩阵

(1)实对称矩阵的特征值为实数.

(2)实对称矩阵的属于不同特征值的特征向量必正交.

(3)若λ是实对称矩阵A的r重特征值,则对应λ的必有r个线性无关的特征向量.

(4)实对称矩阵必可对角化.即若A为n阶实对称阵,则必有正交阵P,使得$P^{-1}AP=\Lambda$,其中Λ是以A的n个特征值为对角元素的对角阵.

定义 含有n个变量$x_1, x_2, \cdots, x_n$的二次齐次函数

$$f(x_1, x_2, \cdots, x_n) = a_{11}x_1^2 + a_{22}x_2^2 + \cdots + a_{nn}x_n^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + \cdots + 2a_{n-1,n}x_{n-1}x_n$$

称为二次型.

二次型可记作$f=x^TAx$,其中$A^T=A$. A称为二次型f的矩阵, f称为对称阵A的二次型,对称阵A的秩称为二次型f的秩.

二次型与它的矩阵是一一对应的.

当a_{ij}是复数时, f称为复二次型;当a_{ij}是实数时, f称为实二次型.

正定二次型

定义 设有实二次型$f(x)=x^T Ax$,如果对任何$x\neq 0$,都有$f(x)>0$(显然$f(0)=0$),则称f为正定二次型,并称对称矩阵A是正定的;如果对任何$x\neq 0$,都有$f(x)<0$,则称f为负定二次型,并称对称矩阵A是负定的.

M3-向量范数与矩阵范数（一）向量范数

定义：如果V是数域K上的线性空间，且对于V的任一向量x，对应一个实数值$\|x\|$，满足以下三个条件

1) 非负性： $\|x\| \geq 0$， 且 $\|x\| = 0 \Leftrightarrow x = 0$

2) 齐次性： $\|kx\| = |k| \cdot \|x\|, \quad \forall k \in K$

3) 三角不等式：$\|x + y\| \leq \|x\| + \|y\|$

则称$\|x\|$为V上向量x的范数，简称为向量范数。

注意：2)中 $|k|$ 当K为实数时为绝对值，

当K为复数域时为复数的模。

向量的范数具有下列简单性质：

(1) 当 $\|x\| \neq 0$ 时，$\left\| \frac{1}{\|x\|} x \right\| = 1$ $\quad \because \left\| \frac{1}{\|x\|} x \right\| = \frac{1}{\|x\|} \|x\| = 1$

(2) $\forall x \in V$ ，$\|-x\| = \|x\|$ $\quad \because \|-x\| = |-1| \|x\| = \|x\|$

(3) $\forall x, y \in V$ ，$\|x\| - \|y\| \leq \|x - y\|$

$\because \|x\| = \|(x - y) + y\| \leq \|x - y\| + \|y\| \Rightarrow \|x\| - \|y\| \leq \|x - y\|$

(4) $\forall x, y \in V$ ，$\big| \|x\| - \|y\| \big| \leq \|x - y\|$

$\|\boldsymbol{x}\| = \|(\boldsymbol{x} - \boldsymbol{y}) + \boldsymbol{y}\| \leq \|\boldsymbol{x} - \boldsymbol{y}\| + \|\boldsymbol{y}\| \quad \Rightarrow \|\boldsymbol{x}\| - \|\boldsymbol{y}\| \leq \|\boldsymbol{x} - \boldsymbol{y}\|$

同样 $\quad \|\boldsymbol{y}\| - \|\boldsymbol{x}\| \leq \|\boldsymbol{y} - \boldsymbol{x}\| = \|\boldsymbol{x} - \boldsymbol{y}\|$

例1：线性空间 C^n，设 $x=(x_1,x_2,\cdots,x_n)^T \in C^n$

1： $\|x\|_1 = \sum|\xi_i|$ 是一种向量范数，记为**1-范数**

2： $\|x\| = \sqrt{(x,x)}$ 是一种向量范数，记**2-范数**

3: $\|x\| = \max\limits_i |x_i|$ 是一种向量范数,记为 ∞ **-范数**

4： $\|x\|_p = \left(\sum\limits_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} \quad (1 \le p < \infty)$

是一种向量范数，记为p-范数或 l_p 范数

例2：线性空间V^n中，任取它的一组基 $x_1,\cdots,x_n$

则对于任意向量x,它可以表示为

$$x=\xi_1 x_1+\cdots+\xi_n x_n$$

与 $\alpha=(\xi_1,\cdots,\xi_n)^T\in C^n$ 是同构的

所以 $\|x\|_p=\|\alpha\|_p$ 是V^n中元素x的p—范数

例3: $C[a,b]$为闭区间 $[a,b]$上的所有实连续函数所成线性空间，可以验证以下定义式均满足范数条件

$$\|f(x)\|_1=\int_a^b f(x)dt \qquad \|f(x)\|_\infty=\max_{t\in[a,b]}|f(x)|$$

$$\|f(x)\|_p=\left(\int_a^b |f(x)|^p dt\right)^{\frac{1}{p}},1<p<\infty$$

例4：设A为n阶实对称正定矩阵，对$x \in Rn$,

定义$\|x\|_A = \left(x^T A x\right)^{1/2}$称为加权范数或椭圆范数

由正定矩阵定义可知$\|x\|_A = 0 \Leftrightarrow x = 0$; $\|x\|_A \neq 0 \Leftrightarrow x \neq 0$

对任意数 $\alpha \in R$ ，有

$$\|\alpha x\|_A = \sqrt{(\alpha x)^T A \alpha x} = \sqrt{\alpha^2 x^T A x} = |\alpha| \sqrt{x^T A x} = |\alpha| \|x\|_A$$

由A正定且实对称 $\Rightarrow \exists$ 正交矩阵Q，使得

$$Q^T A Q = diag(\lambda_1, \cdots, \lambda_n), \quad \lambda_i > 0, i = 1, \cdots, n$$

定义 $B = diag(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_n})Q^T$ 可得 $A = B^T B$

$$\because \|x\|_A = \left(x^T B^T B x\right)^{1/2} = \left[\left(Bx\right)^T \left(Bx\right)\right]^{1/2} = \|Bx\|_2$$

$$\therefore \|x + y\|_A = \|B(x + y)\|_2 \le \|Bx\|_2 + \|By\|_2 = \|x\|_A + \|y\|_A$$

M3-向量范数与矩阵范数（二）向量范数的等价性

定义：有限维线性空间 V^n 中任意两个向量范数

$\|x\|_\alpha$ 和 $\|x\|_\beta$，如果存在着正常数 c_1 和 c_2，

使得 $c_1\|x\|_\beta \le \|x\|_\alpha \le c_2\|x\|_\beta \qquad (\forall x \in V^n)$

则称范数 $\|x\|_\alpha$ 与 $\|x\|_\beta$ 等价

（1）自反性：$1\cdot\|x\|_\alpha \le \|x\|_\alpha \le 1\cdot\|x\|_\alpha, \forall x \in V^n$

（2）对称性：$\frac{1}{c_2}\|x\|_\alpha \le \|x\|_\beta \le \frac{1}{c_1}\|x\|_\alpha, \forall x \in V^n$

（3）传递性：$\left.\begin{array}{l} c_1\|x\|_\beta \le \|x\|_\alpha \le c_2\|x\|_\beta \\ c_3\|x\|_\gamma \le \|x\|_\beta \le c_4\|x\|_\gamma \end{array}\right\} \quad \forall x \in V^n$

$$\Rightarrow c_5\|x\|_\gamma \le \|x\|_\alpha \le c_6\|x\|_\gamma$$

定理：有限维线性空间中任意两个向量范数都等价。

证明思路　1)范数等价为等价关系,满足传递性;

2)任意范数为坐标函数的连续函数；

3)在单位超球面上有大于零的极大极小值,与2-范数等价。

M3-向量范数与矩阵范数（三）矩阵范数

定义：矩阵空间 $C^{m\times n}$ 中，$\forall A\in C^{m\times n}$，

定义实数值 $\|A\|$，且满足以下条件

1)正定条件：$\|A\|\geq 0$，且 $\|A\|=0 \Leftrightarrow A=0_{m\times n}$

2)齐次条件：$\|kA\|=|k|\bullet\|A\|$，$\forall k\in K$

3)三角不等式：$\|A+B\|\leq\|A\|+\|B\|$，$B\in C^{m\times n}$

则称$\|A\|$为A的广义范数。

若对于 $C^{m\times n}, C^{n\times l}$ 及 $C^{m\times n}$ 上的同类广义矩阵范数有

4)相容条件：$\|AB\|\leq\|A\|\|B\|$，$B\in C^{n\times l}$

则称$\|A\|$为A的范数。

定义：设 $C^{m\times n}$ 的矩阵函数 $\|A\|_M$，C^m 与 C^n 中的同类范数 $\|x\|_V$ ，若 $\|Ax\|_V \le \|A\|_M \|x\|_V$

则称矩阵范数 $\|A\|_M$ 与向量范数 $\|x\|_V$ 相容

M3-向量范数与矩阵范数（四）矩阵的从属范数

定理：对 C^m与 C^n上的同类向量范数 $\|x\|_V$，定义

$$\|A\| = \max_{\|x\|_V=1} \|Ax\|_V \qquad \left(\forall A_{m\times n}, x \in C^n\right)$$

则 $\|A\|$ 是 $C^{m\times n}$ 中矩阵A 的范数，且 $\|A\|$ 与 $\|x\|_V$ 相容

$\|A\|$称为由 $\|x\|_V$导出的矩阵范数（或称为从属范数）

等价定义：

$$\max_{\|x\|_V=1} \|Ax\|_V = \max_{x\neq\theta} \frac{\|Ax\|_V}{\|x\|_V}$$

注：

（1）一般的矩阵范数：$\because I = I \cdot I$

$$\|I\| \le \|I\| \cdot \|I\| \qquad \therefore \|I\| \ge 1$$

例如：$\|I\|_{m1} = n, \quad \|I\|_F = \sqrt{n}$

（2）矩阵的从属范数：$\|I\| = \max\limits_{\|x\|_V = 1} \|Ix\|_V = 1$

（3）常用的从属范数：

$$\|x\|_V \qquad \|x\|_1 \qquad \|x\|_2 \qquad \|x\|_\infty$$

$$\|A\|_M \qquad \|A\|_1 \qquad \|A\|_2 \qquad \|A\|_\infty$$

定理： 设 $A=\left(a_{ij}\right)_{m\times n}$ ，则

（1）列和范数： $\|A\|_1=\max\limits_{j}\left\{\sum\limits_{i=1}^{m}\left|a_{ij}\right|\right\}$

（2）谱范数： $\|A\|_2=\sqrt{\lambda_1},\lambda_1=\max\left\{\lambda\left(A^H A\right)\right\}$

（3）行和范数： $\|A\|_\infty=\max\limits_{i}\left\{\sum\limits_{j=1}^{m}\left|a_{ij}\right|\right\}$

M4-多元函数对矩阵变量的求导

（一）数量函数关于向量（矩阵）的微分

定义1　设$y=f(X)=f(x_1,x_2,\cdots,x_n)$

对$x_1,x_2,\cdots,x_n$有偏导数，定义$y=f(X)$

对向量$X=(x_1,x_2,\cdots,x_n)^T$的导数为

$$\frac{df}{dX}=\left(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\cdots,\frac{\partial f}{\partial x_n}\right)^T=gradf,$$

而数量函数$y=f(X)$对向量$X^T=(x_1,x_2,\cdots,x_n)$

的导数定义为

$$\frac{df}{dX^T}=\left(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\cdots,\frac{\partial f}{\partial x_n}\right)。\quad 显然\left(\frac{df}{dX^T}\right)^T=\frac{df}{dX}。$$

一般地，若 $y=f\left(X\right)=$
$f\left(x_{11},x_{12},\cdots,x_{1n},\ x_{21},x_{22},\cdots,x_{2n},\cdots,\ x_{m1},x_{m2},\cdots,x_{mn}\right)$
对每个x_{ij}有偏导数，则定义$y=f\left(X\right)$
对矩阵$X=\left(x_{ij}\right)_{m\times n}$的导数为

$$\frac{df}{dX}=\begin{pmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \frac{\partial f}{\partial x_{21}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{2n}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \frac{\partial f}{\partial x_{m2}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{pmatrix}。$$

性质1　设$f(X)$，$g(X)$为矩阵X的数量函数，如果$\frac{df}{dX}$及$\frac{dg}{dX}$存在，则

$$(1)\ \frac{d}{dX}\left[f(X)+g(X)\right]=\frac{df(X)}{dX}+\frac{dg(X)}{dX};$$

$$(2)\ \frac{d}{dX}\left[f(X)g(X)\right]=f(X)\frac{dg(X)}{dX}+g(X)\frac{df(X)}{dX}$$

例1　设自变量矩阵及其数量函数分别为

$$X=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix},$$

$$y=f(X)=x_{11}{}^2+x_{12}{}^2+\cdots+x_{1n}{}^2+x_{21}{}^2+$$

$$x_{22}{}^2+\cdots+x_{2n}{}^2+\cdots+x_{m1}{}^2+x_{m2}{}^2+\cdots+x_{mn}{}^2$$

$$=\sum_{i=1}^{m}\sum_{j=1}^{n}x_{ij}{}^2,$$ 求 $\dfrac{dy}{dX}$。

解　由于$\dfrac{\partial f}{\partial x_{ij}} = 2x_{ij}$

$(i = 1, 2, \cdots, m;\ j = 1, 2, \cdots, n)$

$$\frac{dy}{dX} = \left(\frac{\partial f}{\partial x_{ij}} \right)_{m\times n} = \left(2x_{ij} \right)_{m\times n} = 2X$$

性质2　设 $X=\left(x_{ij}\right)_{m\times n}$，$B=\left(b_{ij}\right)_{n\times m}$，$A=\left(a_{ij}\right)_{m\times m}$，则

(1) $\dfrac{d}{dX}tr\left(XX^{T}\right)=2X$;

(2) $\dfrac{d}{dX}tr\left(BX\right)=\dfrac{d}{dX}tr\left(X^{T}B^{T}\right)=B^{T}$;

(3) $\dfrac{d}{dX}tr\left(XAX^{T}\right)=\left(A+A^{T}\right)X$。

例2　设$y=f(X)=k^{\mathrm{T}}X=\sum_{j=1}^{n}k_j x_j$，其中，

$X=(x_1,x_2,\cdots,x_n)^T$， $k=(k_1,k_2,\cdots,k_n)^T$，

求$\dfrac{dy}{dX}$。

解： $$\frac{dy}{dx_j}=\frac{d(\sum_{j=1}^{n}k_j x_j)}{dx_j}=k_j$$

$$\frac{dy}{dX}=\left(\frac{\partial y}{\partial x_1},\frac{\partial y}{\partial x_2},\cdots,\frac{\partial y}{\partial x_n}\right)^T=k$$

例3　设二次型 $y=f(X)=X^{T}AX=\sum_{i,j=1}^{n}a_{ij}x_{i}x_{j}$,其中,

$X=(x_1,x_2,\cdots,x_n)^T$, $A=(a_{ij})_{n\times n}$,

且$A^T=A$,求$\dfrac{dy}{dX}$。

解　因为 $\dfrac{dy}{dX}=\left(\dfrac{\partial y}{\partial x_1},\dfrac{\partial y}{\partial x_2},\cdots,\dfrac{\partial y}{\partial x_n}\right)^T$

故只需求出$\dfrac{\partial y}{\partial x_k},k=1,2,\cdots,n$即可。

由于$y=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}x_ix_j=\sum_{i\neq k}\sum_{j\neq k}a_{ij}x_ix_j+$

$2\sum_{i\neq k}a_{ik}x_ix_k+a_{kk}x_k^2$,故$\dfrac{\partial y}{\partial x_k}=2\sum_{i\neq k}a_{ik}x_i+2a_{kk}x_k$

$=2\sum_{i=1}^{n}a_{ik}x_i=2\sum_{j=1}^{n}a_{kj}x_j$,从而$\dfrac{dy}{dX}=2AX$。

例14：$A\in R^{m\times n}, b\in R^{m}$，若 $x\in R^{n}$ 使得 $\|Ax-b\|_2=\min$，

则 $A^T Ax=A^T b$

解：$$f(x)=\|Ax-b\|_2^2=(Ax-b)^T(Ax-b)$$

$$=x^T A^T Ax-2b^T Ax+b^T b$$

$$g(x)=b^T Ax=b_1\sum_{j=1}^{n}a_{1j}\xi_j+\cdots+b_m\sum_{j=1}^{n}a_{mj}\xi_j$$

$$\frac{dg}{dx}=\begin{bmatrix}\dfrac{\partial g}{\partial \xi_1}\\ \vdots\\ \dfrac{\partial g}{\partial \xi_n}\end{bmatrix}=\begin{bmatrix}b_1a_{11}+\cdots+b_m a_{m1}\\ \vdots\\ b_1a_{1n}+\cdots+b_m a_{mn}\end{bmatrix}=A^T b$$

$$\frac{df}{dx}=2A^T Ax-2A^T b=0\quad\Rightarrow A^T Ax=A^T b$$

【注】$r(A^T A)=r(A)\Rightarrow r(A^T A\,|\,A^T b)=r(A^T A)\Rightarrow A^T Ax=A^T b$ 有解

M4-多元函数对矩阵变量的求导

（二）向量函数对向量（矩阵）的导数

定义2　设$a_1(X), a_2(X), \cdots, a_m(X)$对$x_i\ (i=1,2,\cdots,n)$的偏导数都存在，定义向量函数$a^T(X)$对$X$的导数为一个$n\times m$阶矩阵，它的第$i$列向量就是$a_i(X)$对$X$的导数，即

$$\frac{da^T(X)}{dX}=\begin{pmatrix} \frac{\partial a_1(X)}{\partial x_1} & \frac{\partial a_2(X)}{\partial x_1} & \cdots & \frac{\partial a_m(X)}{\partial x_1} \\ \frac{\partial a_1(X)}{\partial x_2} & \frac{\partial a_2(X)}{\partial x_2} & \cdots & \frac{\partial a_m(X)}{\partial x_2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial a_1(X)}{\partial x_n} & \frac{\partial a_2(X)}{\partial x_n} & \cdots & \frac{\partial a_m(X)}{\partial x_n} \end{pmatrix}_{n\times m},$$

同理，$a(X)$对X^T的导数定义为

$$\frac{da(X)}{dX^T}=\begin{pmatrix} \frac{\partial a_1(X)}{\partial x_1} & \frac{\partial a_1(X)}{\partial x_2} & \cdots & \frac{\partial a_1(X)}{\partial x_n} \\ \frac{\partial a_2(X)}{\partial x_1} & \frac{\partial a_2(X)}{\partial x_2} & \cdots & \frac{\partial a_2(X)}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial a_m(X)}{\partial x_1} & \frac{\partial a_m(X)}{\partial x_2} & \cdots & \frac{\partial a_m(X)}{\partial x_n} \end{pmatrix}_{m\times n}$$

显然有:(1) $\frac{da^T(X)}{dX}=\left[\frac{da(X)}{dX^T}\right]^T$, (2) $\frac{dX}{dX^T}=I_n=\frac{dX^T}{dX}$

性质2　设A为$s\times m$阶常数矩阵，$f(X)$为向量X的数量函数，$a(X)$，$b(X)$为X的m维列向量函数，则(1) $\frac{d}{dX}\left[a^T(X)+b^T(X)\right]=\frac{da^T(X)}{dX}+\frac{db^T(X)}{dX}$;

(2) $\frac{d}{dX}\left[f(X)a^T(X)\right]=\frac{df(X)}{dX}a^T(X)+f(X)\frac{da^T(X)}{dX}$;

(3) $\frac{d}{dX^T}\left[Aa(X)\right]=A\frac{da(X)}{dX^T}$;

(4) $\frac{d}{dX}\left[a^T(X)b(X)\right]=\frac{da^T(X)}{dX}b(X)+\frac{db^T(X)}{dX}a(X)$;

(5) $\frac{d}{dX^T}\left[a^T(X)b(X)\right]=b^T(X)\frac{da(X)}{dX^T}+a^T(X)\frac{db(X)}{dX^T}$。

例15: $x=\begin{bmatrix}\xi_1\\ \vdots\\ \xi_n\end{bmatrix}$, $F(x)=\left[f_1(x), f_2(x), \cdots, f_l(x)\right]$

$$\frac{dF}{dx}=\begin{bmatrix}\frac{\partial f_1}{\partial \xi_1} & \cdots & \frac{\partial f_l}{\partial \xi_1}\\ \vdots & & \vdots\\ \frac{\partial f_1}{\partial \xi_n} & \cdots & \frac{\partial f_l}{\partial \xi_n}\end{bmatrix}$$

例16: $A=\left(a_{ij}\right)_{n\times n}$, $x=\begin{bmatrix}\xi_1\\ \vdots\\ \xi_n\end{bmatrix}$ $Ax=\begin{bmatrix}\sum_{j=1}^{n} a_{1j}\xi_j\\ \vdots\\ \sum_{j=1}^{n} a_{nj}\xi_j\end{bmatrix}$

$$\frac{d(Ax)}{dx^T}=\begin{bmatrix}a_{11} & \cdots & a_{1n}\\ \vdots & & \vdots\\ a_{n1} & \cdots & a_{nn}\end{bmatrix}=A$$

M5-矩阵的QR分解与奇异值分解

（一）方阵的正交三角分解（QR分解）

定理6： $A_{n\times n}$ 可逆 $\Rightarrow \exists$ 正交矩阵$\boldsymbol{Q}$，可逆上三角矩阵$\boldsymbol{R}$，使得$\boldsymbol{A}=\boldsymbol{QR}$。

证明：$\boldsymbol{A}=(\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n)$ 可逆 $\Rightarrow \boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n$ 线性无关，

正交化后可得：

$$\begin{cases} \boldsymbol{b}_1=\boldsymbol{a}_1 \\ \boldsymbol{b}_2=\boldsymbol{a}_2-k_{21}\boldsymbol{b}_1 \\ \quad\cdots\cdots \\ \boldsymbol{b}_n=\boldsymbol{a}_n-k_{n,n-1}\boldsymbol{b}_{n-1}-\cdots-k_{n1}\boldsymbol{b}_1 \end{cases} \qquad \begin{cases} \boldsymbol{a}_1=\boldsymbol{b}_1 \\ \boldsymbol{a}_2=k_{21}\boldsymbol{b}_1+\boldsymbol{b}_2 \\ \quad\cdots\cdots \\ \boldsymbol{a}_n=k_{n1}\boldsymbol{b}_1+\cdots+k_{n,n-1}\boldsymbol{b}_{n-1}+\boldsymbol{b}_n \end{cases}$$

$$(a_1,a_2,\cdots,a_n)=(b_1,b_2,\cdots,b_n)K$$

$$=(q_1,q_2,\cdots,q_n)\begin{bmatrix}|b_1| & & & \\ & |b_2| & & \\ & & \ddots & \\ & & & |b_n|\end{bmatrix}\begin{bmatrix}1 & k_{21} & \cdots & k_{n1} \\ & 1 & \cdots & k_{n2} \\ & & \ddots & \vdots \\ & & & 1\end{bmatrix}$$

令 $Q=(q_1,q_2,\cdots,q_n),R=\begin{bmatrix}|b_1| & & & \\ & |b_2| & & \\ & & \ddots & \\ & & & |b_n|\end{bmatrix}\begin{bmatrix}1 & k_{21} & \cdots & k_{n1} \\ & 1 & \cdots & k_{n2} \\ & & \ddots & \vdots \\ & & & 1\end{bmatrix}$

则$A=QR$，其中 $q_i=\dfrac{b_i}{|b_i|}\quad(i=1,2,\cdots,n)$

定理7：$A_{m\times n}$ 列满秩 $\Rightarrow \exists$ 矩阵 $Q_{m\times n}$ 满足 $Q^H Q = I$，可逆上三角矩阵 $R_{n\times n}$ ，使得$A=QR$。

M5-矩阵的QR分解与奇异值分解
（二）奇异值分解（SVD）

一、预备知识

(1) $\forall A_{m\times n}$, $\left(A^{\mathrm{H}}A\right)_{n\times n}$ 是 **Hermite**（半）正定矩阵.

$$\forall x \neq 0,\ x^{\mathrm{H}}A^{\mathrm{H}}Ax = \left(Ax\right)^{\mathrm{H}}\left(Ax\right) = \left|Ax\right|^2 \geq 0$$

(2) 齐次方程组 $Ax = 0$ 与 $A^{\mathrm{H}}Ax = 0$ 同解

若 $Ax = 0$，则 $A^{\mathrm{H}}Ax = 0$；

反之，$A^{\mathrm{H}}Ax = 0$ $\Rightarrow \left|Ax\right|^2 = \left(Ax\right)^{\mathrm{H}}\left(Ax\right) = x^{\mathrm{H}}\left(A^{\mathrm{H}}Ax\right) = 0$

$\Rightarrow Ax = 0$

(3) $\mathbf{rank}\,A = \mathbf{rank}(A^{\mathrm{H}}A)$

$$S_1 = \{x \mid Ax = 0\}, \quad S_2 = \{x \mid A^{\mathrm{H}}Ax = 0\}$$

$$S_1 = S_2 \Rightarrow \dim S_1 = \dim S_2 \Rightarrow n - r_A = n - r_{A^{\mathrm{H}}A}$$

$$\Rightarrow r_A = r_{A^{\mathrm{H}}A}$$

(4) $A = O_{m\times n} \Leftrightarrow A^{\mathrm{H}}A = O_{n\times n}$

必要性．左乘即得；

充分性 $r_A = r_{A^{\mathrm{H}}A} = 0 \Rightarrow A = O$

二、正交对角分解

定理15： $A_{n\times n}$ 可逆 $\Rightarrow \exists$ 酉矩阵 $U_{n\times n}, V_{n\times n}$ ，使得

$$U^{H}AV=\begin{bmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_n\end{bmatrix}\overset{\Delta}{=}D\quad(\sigma_i>0)$$

三、奇异值分解

$A_{m\times n}\in C_r^{m\times n}(r\geq 1)\Rightarrow A^H A\in C_r^{n\times n}$ 半正定

$A^H A$ 的特征值： $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_r\geq\lambda_{r+1}=\cdots=\lambda_n=0$

A的奇异值： $\sigma_i=\sqrt{\lambda_i},\quad i=1,2,\cdots n$

特点：（1）A的奇异值个数等于A的列数

（2）A的非零奇异值个数等于 $\mathrm{rank}\,A$

定理16: $A_{m\times n}\in C_r^{m\times n}(r\geq 1), \Sigma_r=\begin{bmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r\end{bmatrix}\Rightarrow$

存在酉矩阵 $U_{m\times m}$ 及 $V_{n\times n}$ ，使得 $U^{H}AV=\begin{bmatrix}\Sigma_r & O\\ O & O\end{bmatrix}_{m\times n}\overset{\Delta}{=}D$

[注]: 称 $A=UDV^H$ 为A的奇异值分解

(1) U与V不唯一；

(2) U的列为AA^H的特征向量，V的列为A^HA的特征向量

(3) 称U的列为A的左奇异向量，称V的列为A的右奇异向量.

定理17：$A_{m\times n}\in C_r^{m\times n}(r\geq 0)$ 的奇异值分解 $A=U\begin{bmatrix}\Sigma & O\\ O & O\end{bmatrix}V^H$

中，划分 $U=(u_1,u_2,\cdots,u_m),V=(v_1,v_2,\cdots,v_n)$，则有

(1) $N(A)=\text{span}\{v_{r+1},v_{r+2},\cdots,v_n\}$；

(2) $R(A)=\text{span}\{u_1,u_2,\cdots,u_r\}$；

(3) $A=\sigma_1u_1v_1^{\mathrm{H}}+\sigma_2u_2v_2^{\mathrm{H}}+\cdots+\sigma_ru_rv_r^{\mathrm{H}}$

M6-投影矩阵、广义逆矩阵（一）投影矩阵

1、投影变换

定义：向量空间C^n中，子空间L与M满足$C^n = L \oplus M$，

对 $\forall x \in C^n$，分解式 $x = y + z, y \in L, z \in M$ 唯一。

称变换 $T_{L,M}(x) = y$ 为沿着M到L的投影

性质(1)：$T_{L,M}$ 是线性变换

性质(2)：$R(T_{L,M}) = L, N(T_{L,M}) = M$

性质(3)：$\forall x \in L \Rightarrow T_{L,M}(x) = x \quad \forall x \in M \Rightarrow T_{L,M}(x) = \theta$

[注] $T_{L,M}$ 是L中的单位变换

$T_{L,M}$ 是M中的零变换

2、投影矩阵

定义：取线性空间C^n的基为$e_1,e_2,\dots,e_n$时，元素x与它的坐标"形式一致"。称 $T_{L,M}$ 在该基下的矩阵记为投影矩阵$P_{L,M}$

性质(4)： $T_{L,M}(x)=y \Leftrightarrow P_{L,M}x=y$

$$x \in L \Rightarrow T_{L,M}(x)=x \Rightarrow P_{L,M}x=x$$

$$x \in M \Rightarrow T_{L,M}(x)=\theta \Rightarrow P_{L,M}x=\theta$$

4、正交投影变换

欧氏空间 C^n 中，子空间L给定，取 $M = L^{\perp}$,

则 $\mathbf{C}^n = L \oplus M$

正交投影变换$T_L = T_{L,M}$; 正交投影矩阵 $P_L = P_{L,M}$

定理2：方阵 $P = P_L \Leftrightarrow P^2 = P, P^H = P$

4、正交投影矩阵的确定办法

$$\left.\begin{array}{l} L\text{的基为}\quad x_1,\cdots,x_r:\quad X=(x_1,\cdots,x_r) \\ L^{\perp}\ \text{的基为}\ y_1,\cdots,y_{n-r}:Y=(y_1,\cdots,y_{n-r}) \end{array}\right\} \Rightarrow \begin{cases} X^H Y=O \\ Y^H X=O \end{cases}$$

已求得 $P_L = P_{L,L^{\perp}} = (X|O)\cdot(X|Y)^{-1}$

因为 $(X|Y)^{H}\cdot(X\mid Y)=\begin{pmatrix} X^{H} \\ Y^{H} \end{pmatrix}\cdot(X\mid Y)=\begin{bmatrix} X^{H}X & O \\ O & Y^{H}Y \end{bmatrix}$

所以 $(X\mid Y)^{-1}=\begin{bmatrix} (X^{H}X)^{-1} & O \\ O & (Y^{H}Y)^{-1} \end{bmatrix}\cdot(X\mid Y)^{H}=\begin{bmatrix} (X^{H}X)^{-1}X^{H} \\ (Y^{H}Y)^{-1}Y^{H} \end{bmatrix}$

于是 $P_L=(X|O)\cdot\begin{bmatrix} (X^{H}X)^{-1}X^{H} \\ (Y^{H}Y)^{-1}Y^{H} \end{bmatrix}=X\cdot(X^{H}X)^{-1}\cdot X^{H}$

M6-投影矩阵、广义逆矩阵（二）广义逆矩阵

定义：对 $A_{m\times n}$ 若有 $X_{n\times m}$ 满足Penrose方程

（1） $AXA=A$ （2） $XAX=X$

（3） $(AX)^H=AX$ （4） $(XA)^H=XA$

称X为A的M-P逆，记作A^+.(Moore 1920，Penrose1955)

例如 $A_{m\times n}$ 可逆，$X=A^{-1}$ 满足P-方程：$A^+=A^{-1}$

$A=O_{m\times n}, X=O_{n\times m}$ 满足P-方程： $O^+_{m\times n}=O_{n\times m}$

$A=\begin{bmatrix}1\\1\end{bmatrix}, X=\begin{bmatrix}\frac{1}{2} & \frac{1}{2}\end{bmatrix}$ 满足P-方程：$A^+=\begin{bmatrix}\frac{1}{2} & \frac{1}{2}\end{bmatrix}$

例5： 设 $A \in C_r^{m\times n}$ 的奇异值分解为 $A = U\begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix}_{m\times n} V^H$

则 $A^+ = V\begin{bmatrix} \Sigma_r^{-1} & O \\ O & O \end{bmatrix}_{n\times m} U^H$

直接验证即可。

进一步的有

$$A = (U_s, U_n)\begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix}_{m\times n} (V_s, V_n)^H$$

$$= U_s \Sigma_r V_s^H$$

式子中，下标s改为r，n改为n-r。

$$A^+ = V_s \Sigma_r^{-1} U_s^H$$

定理10： (1) $r_{A^+} = r_A$ (2) $\left(A^+\right)^+ = A$

(3) $\left(A^H\right)^+ = \left(A^+\right)^H$ $\left(A^T\right)^+ = \left(A^+\right)^T$

(4) $\left(A^H A\right)^+ = A^+\left(A^H\right)^+$ $\left(AA^H\right)^+ = \left(A^H\right)^+ A^+$

(5) $A^+ = \left(A^H A\right)^+ A^H = A^H\left(AA^H\right)^+$

(6) $R\left(A^+\right) = R\left(A^H\right), N\left(A^+\right) = N\left(A^H\right)$

M6-投影矩阵、广义逆矩阵（三）正定矩阵

正定矩阵

定义：令A是n阶Hermite矩阵,若对任意复n维列向量

$x \neq 0$ ，都有 $x^H A x \geq 0$

则称A为**Hermite非负定矩阵**，简称为**非负定矩阵**。对任意复n维列向量 $x \neq 0$ ，都有

$$x^H A x > 0$$

称为**Hermite正定矩阵**，简称为**正定矩阵**

性质(1)：若A正定，且k为正常数，则kA正定

性质(2)：若A、B正定，则$A+B$正定

性质(3)：若A正定，则 A^{-1} 正定

正定矩阵的判断准则

Hermite矩阵A是正定矩阵 $\Longleftrightarrow$ A的特征值都为正数

因为如果 $Ax=\lambda x$ ，则有

$$0<x^H Ax = x^H \lambda x = \lambda x^H x = \lambda \|x\|^2$$

一定有 $\lambda>0$。反之，必然存在酉矩阵U，使得

$$A=U\Lambda U^H,\quad \Lambda=diag(\lambda_1,\cdots,\lambda_n)$$

因为$\lambda_i>0$，且对任意 $x\neq 0$ 有 $Ux=y\neq 0$

从而$x^H Ax=\left(U^H y\right)^H A\left(U^H y\right) = y^H UAU^H y = y^H\Lambda y>0$

所以A是正定的

Hermite矩阵A是非负定矩阵 $\Longleftrightarrow$ A的特征值都为非负数

Hermite矩阵A是正定矩阵（非负定）

$\Leftrightarrow$ A的特征值都为正数（非负数）

$\Leftrightarrow$ $\mathrm{tr}\,A > \lambda_i(A), \quad (or\ \mathrm{tr}\,A \geq \lambda_i(A)) \quad i = 1, \cdots, n$

$\Leftrightarrow$ $A = P^H P$ P是n阶非奇异矩阵(矩阵)

$\Leftrightarrow$ $A = B^2$ B是n阶正定矩阵(非负定矩阵)

$\Leftrightarrow$ A的n个顺序主子式为正数(非负数)

- A，C是正定矩阵，且$AC=CA$，则AC为正定矩阵
- A是正定矩阵，C是非负定矩阵，且$AC=CA$，
- A，B是n阶Hermite矩阵，若A-B为正定矩阵(非负定矩阵)，则称A大于B，记为$A>B$.(A不小于B记 $A \geq B$)也可记为$B<A$.($B \leq A$)

$A \geq B$ 的充要条件是，对任意复n维列向量x都有

$$x^H Ax \geq x^H Bx$$

对A，B是实对角阵时，$A = diag(a_1, a_2, \cdots, a_n)$

$B = diag(b_1, b_2, \cdots, b_n)$，则有 $a_i \geq b_i$

课后作业

- 见另文。
- 下次上课前提交。
- 最好使用电子档。

End of This Part