矩阵代数与应用

上海大学计算机工程与科学学院

2024年1月

第六章 逆矩阵与矩阵方程求解

基本问题

$B_{p\times q}, C_{q\times n}$是任意矩阵.

$$\begin{aligned} vec(BC) &= (C^T \otimes I_p)vec(B) \\ &= (C^T \otimes B)vec(I_q) \\ &= (I_n \otimes B)vec(C) \end{aligned}$$

- ◆ 线性方程组

$$A_{m\times n} X_{n\times 1} = b_{m\times 1}$$

- ◆ 矩阵方程

$$A_{m\times n} x_{n\times p} = b_{m\times p} \quad \Rightarrow (I_p \otimes A_{m\times n})vec(x_{n\times p}) = vec(b_{m\times p})$$

若A为正方矩阵（m=n）且非奇异，则

$$x = A^{-1} b$$

线性方程组可以看成矩阵方程的特例(p=1)

α=独立方程个数(rank([A,b])，β=独立未知量(参数)个数(rank(A))

(1)适定(well-determined)方程:当(α=β=n)

(2)欠定(under-determined)方程:当(α=β<n)

(3)超定(over-determined)方程:当(α>β)

一致方程：存在精确解
(consistent equation)
非一致方程：无精确解
(inconsistent equation)

注:自由未知参量(可以任意取值)个数为n-β,n为未知数的个数. **超定方程**:方程数大于未知数个数(不精确说法,在上下文含义清楚的情况下,亦可如此说)

$$\begin{cases} x_1 + x_2 = 2 \\ x_1 - x_2 = 4 \end{cases}$$

$$\begin{cases} x_1 + x_2 + x_3 = 2 \\ x_1 - x_2 + 6x_3 = 4 \end{cases}$$

$$\begin{cases} x_1 + x_2 + x_3 = 2 \\ x_1 - x_2 + 6x_3 = 4 \\ 0 = 7 \end{cases}$$

$$\begin{cases} x_1 + x_2 + x_3 = 2 \\ x_1 - x_2 + 6x_3 = 4 \\ x_1 + 3x_2 + x_3 = 2 \\ x_1 + x_2 + 3x_3 = 2 \end{cases}$$

一致方程：存在精确解 (consistent equation)
非一致方程：无精确解 (inconsistent equation)

α=独立方程个数(rank([A,b])，β=独立未知量(参数)个数(rank(A))

(1)适定(well-determined)方程:当(α=β=n)

(2)欠定(under-determined)方程:当(α=β<n)

(3)超定(over-determined)方程:当(α>β)

矩阵的秩

（1）秩是一个正整数.

（2）秩等于或小于矩阵的行数或列数.

（3）若 $\operatorname{rank}\left(A_{n\times n}\right)=n$，则称$A$是**满秩(full rank)矩阵**，此时$A$非奇异，且其$n$个列向量$a_1,\cdots,a_n$ 线性无关.

（4）若 $\operatorname{rank}\left(A_{m\times n}\right)<\min\{m,n\}$，则称$A$是**秩亏缺(rank deficient)**的. 一个秩亏缺的正方矩阵为奇异矩阵.

（5）若$\operatorname{rank}\left(A_{m\times n}\right)=m(<n)$，则称矩阵$A$是**满行秩(full row rank)**矩阵.

（6）若 $\operatorname{rank}\left(A_{m\times n}\right)=n(<m)$，则称矩阵$A$是**满列秩(full column rank)**矩阵.

（7）任何矩阵A左乘满列秩矩阵或者右乘满行秩矩阵后，矩阵A的秩保持不变.

（8） $\operatorname{rank}(AB)\le\min\{\operatorname{rank}(A),\operatorname{rank}(B)\}$.

线性空间

定义 设V是一个非空集合,P为一个数域．如果对于任意两个元素$\vec{\alpha}, \vec{\beta} \in V$，总有唯一的一个元素$\vec{\gamma} \in V$与之对应，称之为$\vec{\alpha}$与$\vec{\beta}$的和，记作

$$\vec{\gamma} = \vec{\alpha} + \vec{\beta}$$

若对于任一数$\lambda \in P$与任一元素$\vec{\alpha} \in V$，总有唯一的一个元素$\vec{\delta} \in V$与之对应，称为λ与$\vec{\alpha}$的积，记作

$$\vec{\delta} = \lambda\vec{\alpha}$$

如果上述的两种运算满足以下八条运算规律:

(1) $\vec{\alpha}+\vec{\beta}=\vec{\beta}+\vec{\alpha}$; (2) $\left(\vec{\alpha}+\vec{\beta}\right)+\vec{\gamma}=\vec{\alpha}+\left(\vec{\beta}+\vec{\gamma}\right)$;

(3) $\boldsymbol{V}$中存在零元素$\vec{0}$,对$\forall\vec{\alpha}\in\boldsymbol{V}$,都有$\vec{\alpha}+\vec{0}=\vec{\alpha}$;

(**4**)对任何$\vec{\alpha}\in\boldsymbol{V}$,都有$\vec{\alpha}$的负元素$\vec{\beta}\in\boldsymbol{V}$,使$\vec{\alpha}+\vec{\beta}=\vec{\mathbf{0}}$;

(5) $\lambda\left(\mu\vec{\alpha}\right)=\left(\lambda\mu\right)\vec{\alpha}$; $\lambda,\mu\in P$

(6) $\left(\lambda+\mu\right)\vec{\alpha}=\lambda\vec{\alpha}+\mu\vec{\alpha}$;

(7) $\lambda\left(\vec{\alpha}+\vec{\beta}\right)=\lambda\vec{\alpha}+\lambda\vec{\beta}$.   (8) $1\vec{\alpha}=\vec{\alpha}$;

那么V就称为数域 P 上的线性空间.

注：线性空间的元素统称为“向量”，但它可以是通常的向量，也可以是矩阵、多项式、函数等.

线性空间的简单性质:

1.零元素是唯一的； $\vec{0}_1=\vec{0}_1+\vec{0}_2=\vec{0}_2$

2.负元素是唯一的；

$$-\vec{\alpha}_1=(-\vec{\alpha}_1)+\vec{0}=(-\vec{\alpha}_1)+(\vec{\alpha}+(-\vec{\alpha}_2))$$

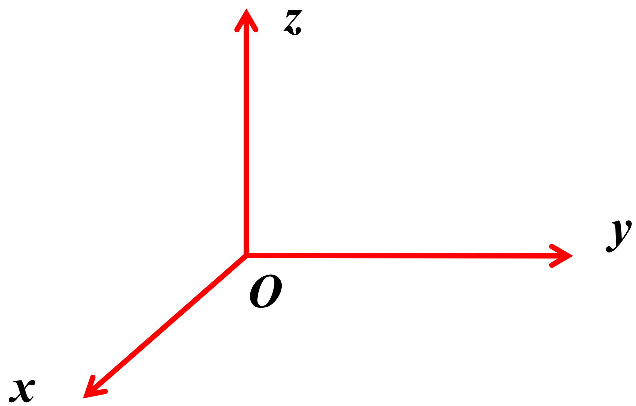
$$=((-\vec{\alpha}_1)+\vec{\alpha})+(-\vec{\alpha}_2)=\vec{0}+(-\vec{\alpha}_2)=-\vec{\alpha}_2$$

3. $0\vec{\alpha}=\vec{0},\ k\vec{0}=\vec{0},\ (-1)\vec{\alpha}=-\vec{\alpha};$

4. 如果 $k\vec{\alpha}=\vec{0}$，那么 $k=0$ 或 $\vec{\alpha}=\vec{0}.$

线性子空间

对三维几何空间：



任何过原点的平面是R^3的子集.

在该平面上的所有向量对于向量的加法和数乘运算构成一个二维的线性空间.

R^3的线性子空间

定义： 设W是数域F上线性空间V的非空子集合. 如果W中的向量对V中所定义的向量加法和数乘运算也构成F上的线性空间，则称W为V的线性子空间,简称子空间.

定理: W是V的非空子集合，则W是V的子空间的充要条件是

$$\forall \vec{\alpha}, \vec{\beta} \in W, \forall k \in F, \text{有} \quad k\vec{\alpha} + \vec{\beta} \in W.$$

注　V的子空间 $\begin{cases} V\text{和零子空间是}V\text{的平凡子空间;} \\ \text{其它子空间称为}V\text{的真子空间.} \end{cases}$

生成子空间

设 $\vec{\alpha}_1, \vec{\alpha}_2, \cdots, \vec{\alpha}_s \in V$，则

$$L(\vec{\alpha}_1, \vec{\alpha}_2, \cdots, \vec{\alpha}_s) = \{k_1\vec{\alpha}_1 + k_2\vec{\alpha}_2 + \cdots + k_s\vec{\alpha}_s \mid k_1, k_2, \cdots, k_s \in F\}$$

是 V 的子集. 上述集合记为 L_α.

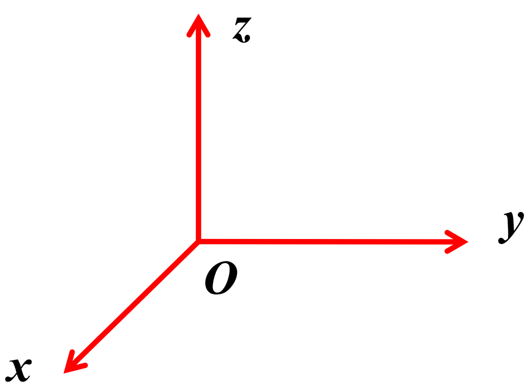
易证 L_α 是 V 的子空间.

我们称 L_α 是由 $\vec{\alpha}_1, \vec{\alpha}_2, \cdots, \vec{\alpha}_s$ 生成的子空间. $\vec{\alpha}_1, \vec{\alpha}_2, \cdots, \vec{\alpha}_s$ 是它的**生成向量组**.

基、维数和坐标

线性空间中含有无穷多个向量. 如何找出有限个向量刻划空间中的所有向量?

如三维几何空间:



可以用$\vec{i},\vec{j},\vec{k}$刻划R^3中的任意向量.

$$\vec{r}=(x,y,z)=x\vec{i}+y\vec{j}+z\vec{k}$$

$$\vec{i}=(1,0,0);\ \vec{j}=(0,1,0);\ \vec{k}=(0,0,1).$$

3个最基本的向量$\vec{i},\vec{j},\vec{k}$构成坐标系

$\vec{r}=x\vec{i}+y\vec{j}+z\vec{k}$中$x,y,z$称为在该坐标系下的坐标.

设$\vec{\alpha}_{i_1},\vec{\alpha}_{i_2},\ldots,\vec{\alpha}_{i_r}$是线性空间$V$中的一个向量组，若满足

(1)$\vec{\alpha}_{i_1},\vec{\alpha}_{i_2},\ldots,\vec{\alpha}_{i_r}$线性无关;

(2)V中任意一个向量$\vec{\alpha}$都可由$\vec{\alpha}_{i_1},\vec{\alpha}_{i_2},\ldots,\vec{\alpha}_{i_r}$线性表示：

$$\vec{\alpha}=x_1\vec{\alpha}_{i_1}+x_2\vec{\alpha}_{i_2}+x_r\vec{\alpha}_{i_r}$$

则称：$\vec{\alpha}_{i_1},\vec{\alpha}_{i_2},\ldots,\vec{\alpha}_{i_r}$是$V$的一组基.

基中所含向量的个数，即r为V的维数，记为$\dim(V)=r$.

$x_1,x_2,\cdots,x_r$称为$\vec{\alpha}$在该基下的坐标.

注: (1)规定$V=\{\vec{0}\}$为零维空间.

(2)有限维线性空间V的基不唯一.

问题：宇宙是几维的？ （超弦理论）

A={(x,y):x,y∈ R}

B={(x,y,0):x,y∈ R}

C={(x,y,1):x,y∈ R}

D={(x,y,z):x,y, $z \in R$}

E={(x,y,z,1):x,y, $z \in R$}

F={(x,y,z,0,0):x,y, $z \in R$}

G={(x,y,z,1,2):x,y, $z \in R$}

空间？ 维数？

Span(A):A中向量的所有线性组合的全体。

问题：

Span(C),Span(E),Span(G)

的维数?

3， 4， 4?

向量组的秩

定义 生成子空间$L(\vec{\alpha}_1,\vec{\alpha}_2,\cdots,\vec{\alpha}_s)$的维数称为**向量组$\vec{\alpha}_1,\vec{\alpha}_2,\cdots,\vec{\alpha}_s$的秩**，记为$r(\vec{\alpha}_1,\vec{\alpha}_2,\cdots,\vec{\alpha}_s)$.

设子空间$L_\alpha = L(\vec{\alpha}_1,\vec{\alpha}_2,\cdots,\vec{\alpha}_s)$的维数为$r$,如何确定$L_\alpha$的一组基（或向量组的秩）?

矩阵的行秩与列秩

给定矩阵A，

称矩阵A的行向量组生成的子空间$\mathrm{R}(A)$,
对应空间的维数为矩阵的行秩；

称矩阵A的列向量组生成的子空间$\mathrm{C}(A)$，
对应空间的维数为矩阵的列秩.

定理 矩阵的行秩、列秩和矩阵的秩相等.

无交连、正交与正交补

交空间的定义

设V_1、V_2为线性空间V的子空间，则集合

$$V_1 \cap V_2 = \{a \mid a \in V_1 \text{且} a \in V_2\}$$

也为V的子空间，称之为V_1与V_2的**交空间**.

事实上，$\because 0 \in V_1, 0 \in V_2$，$\therefore 0 \in V_1 \cap V_2 \neq \varnothing$

任取 $\alpha, \beta \in V_1 \cap V_2$，即 $\alpha, \beta \in V_1$，且$\alpha, \beta \in V_2$，

则 $k\alpha + l\beta \in V_1 \cap V_2$

故 $V_1 \cap V_2$为V的子空间.

$\forall \vec{\alpha}, \vec{\beta} \in W, \forall k \in F$，

有 $k\vec{\alpha} + \vec{\beta} \in W$.

多个子空间的交

$$V_1 \cap V_2 \cap \cdots \cap V_k = \bigcap_{i=1}^{k} V_i = \{\alpha \mid \alpha \in V_i, i = 1,2,3,\cdots,k\}$$

$V_1, V_2, \cdots, V_k$ 为线性空间V的子空间，则上述集合

也为V的子空间，称为 $V_1, V_2, \cdots, V_k$ 的交空间.

若这些子空间共同的唯一向量为零向量，则称子空间 $V_1, V_2, \cdots, V_k$ 无交连。

无交连的子空间的并集张成的子空间，称为这些子空间的直和，记作

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_k \quad or \quad V = \bigoplus_{i=1}^{k} V_i$$

若一向量与子空间S的所有向量都正交，则称该向量正交于子空间S. 若$S_1, S_2, \cdots, S_n$为正交子空间，

记作$S_i \perp S_j, i \neq j$，若$a_i \perp a_j$对所有的

$a_i \in S_i, a_j \in S_j (i \neq j)$恒成立。

正交补空间：

W是与子空间**U**正交的所有向量的集合组成一个向量子空间，称这样的W为U的一个**正交补空间，**可记作$W = U^{\perp}$.

U和W是V的一个**直和分解，且** $\dim(U) + \dim(W) = \dim(V)$

- **无交连**是比正交更弱的条件. 两个子空间无交连, 只是表明这两个子空间没有任何一对非零的共同向量. 而两个子空间正交时, 在两个子空间分别任取一个向量都是正交的. **无交连**的两个子空间不一定正交, 但是正交的两个子空间必定是无交连的.
- 正交补空间是一个比正交子空间更严格的概念. 两个子空间正交但是不一定是正交补. 可能存在多个子空间和同一个子空间正交.

列空间、行空间与零空间

- 令 $A = [\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_p]_{m\times p}$ 为列向量张成的子空间,
- **行空间**为复共轭行向量张成的子空间,

$$Col\left(A\right) = Span\left\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_p\right\}$$

- 复矩阵A的行空间与复共轭转置矩阵的列空间等价,

$$Row(A) = Span\{\vec{r}_1^{\,*}, \cdots, \vec{r}_m^{\,*}\}$$

- **矩阵的零空间也称为矩阵的核,**

$$Row(A) = Col(A^{\mathrm{H}}) = Span\{\vec{r}_1^{\,*}, \cdots, \vec{r}_m^{\,*}\}$$

$$Null(A) = N(A) = Ker(A) = \{\vec{x} \in C^n | A\vec{x} = 0\}$$

- **A的值域，**

$$Range(A)=R(A)=\{\vec{y}\in C^m | A\vec{x}=\vec{y}, \vec{x}\in C^n\}$$

$$Range(A)=Col(A), Range(A^H)=Col(A^H)$$

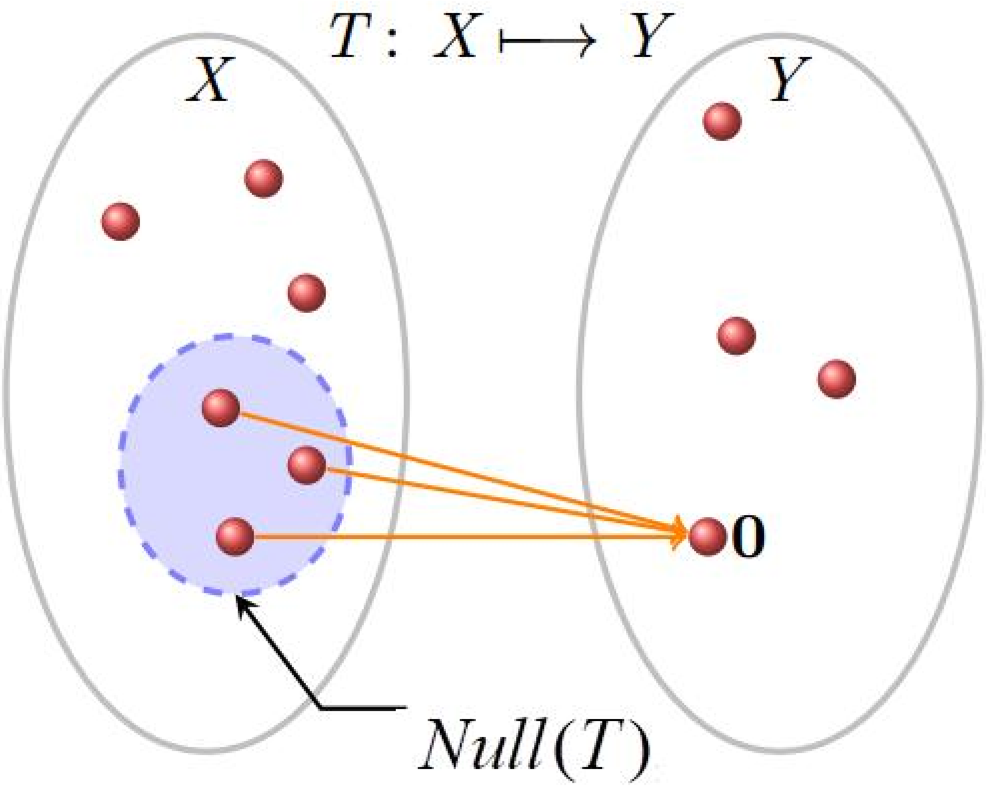
- **零空间的维数称为A的零化维（nullity），是自由变量的个数，**

$$\text{nullity}(A)=\dim(Null(A))$$

$$\begin{cases} x_1+x_2+x_3=2 \\ x_1-x_2+6x_3=4 \end{cases}$$

特解和通解

列空间、行空间与零空间



列空间、行空间与零空间

$$(Row(A))^{\perp}=Null(A),(Col(A))^{\perp}=Null(A^{H})$$

零空间 Null $(\boldsymbol{A})$	列空间 Col $(\boldsymbol{A})$
Null $(\boldsymbol{A})$ 是 $\mathbb{C}^{m}$ 的子空间	Col $(\boldsymbol{A})$ 是 $\mathbb{C}^{n}$ 的子空间
Null $(\boldsymbol{A})$ 为隐含定义，与 $\boldsymbol{A}$ 的列向量无直接关系	Col $(\boldsymbol{A})$ 为显式定义，直接由 $\boldsymbol{A}$ 的所有列向量张成
Null $(\boldsymbol{A})$ 的基应满足 $\boldsymbol{Ax}=\mathbf{0}$	Col $(\boldsymbol{A})$ 的基是 $\boldsymbol{A}$ 的主元列
Null $(\boldsymbol{A})$ 与矩阵 $\boldsymbol{A}$ 的元素无任何明显关系	矩阵 $\boldsymbol{A}$ 的每一列都在 Col $(\boldsymbol{A})$ 内
Null $(\boldsymbol{A})$ 的典型向量 $\boldsymbol{v}$ 满足 $\boldsymbol{Av}=\mathbf{0}$	Col $(\boldsymbol{A})$ 的典型向量满足 $\boldsymbol{Ax}=\boldsymbol{v}$ 为一致方程
$\boldsymbol{v}\in$ Null $(\boldsymbol{A})$ 的条件：$\boldsymbol{Av}=\mathbf{0}$	$\boldsymbol{v}\in$ Col $(\boldsymbol{A})$ 的条件：$[\boldsymbol{A},\boldsymbol{v}]$ 与 $\boldsymbol{A}$ 具有相同的秩

- **定义**：一个子空间 $S \subseteq C^n$ 称为（相对于）A不变的，若 $x \in S \Rightarrow Ax \in S$.
- 由A的特征向量张成的子空间S是相对于A**不变的子空间.**
- 矩阵A的任意一个特征值λ，零空间 $\mathrm{Null}(A-\lambda I)$是相对于A不变的子空间。

$$u \in \mathrm{Null}(A-\lambda I) \Rightarrow (A-\lambda I)u = 0,$$
$$(A-\lambda I)Au = \lambda(A-\lambda I)u = 0$$

- 零空间 $\mathrm{Null}(A-\lambda I)$ 称为矩阵A与特征值 λ 对应的**特征空间(**eigenspace).

6.0 逆矩阵

定义：称 $n\times n$ 矩阵 A 可逆，若存在一个 $n\times n$ 矩阵 A^{-1} 满足

$AA^{-1}=A^{-1}A=I$. 并称 A^{-1} 是实矩阵 A 的逆矩阵.

关于矩阵的非奇异性/可逆性，下列叙述等价：

(1) A 非奇异.

(2) A^{-1} 存在.

(3) $\text{rank}(A)=n$.

(4) A 的行(列)线性无关.

(5) A 的值域维数为 n, 零空间维数为0. （等价不等价？）

(6) $\det(A)\neq 0$.

(7) $Ax=b$ 对每一个 b 有唯一的解.

(8) $Ax=0$ 只有解 $x=0$.

$\text{Range}(A)=\{Y:Y=AX, X\in C^n\}$

$\text{Null}(A)=\{X:AX=0\}$

逆矩阵 A^{-1} 的性质：

（1）$AA^{-1} = A^{-1}A = I$.

（2）A^{-1} 是唯一的.

（3）$|A^{-1}| = \dfrac{1}{|A|}$.

（4）逆矩阵是非奇异的.

（5）$(A^{-1})^{-1} = A$.

（6）$(A^{H})^{-1} = (A^{-1})^{H}$.

（7）若 $A^{H}=A$，则 $(A^{-1})^{H}=A^{-1}$.

（8）$(A^{*})^{-1}=(A^{-1})^{*}$.

（9）若A和B均可逆，则

$$(AB)^{-1}=B^{-1}A^{-1},$$

更一般地

$$(ABC)^{-1}=C^{-1}B^{-1}A^{-1}.$$

（10）若 $A=diag(a_1,a_2,\cdots,a_m)$，则

$$A^{-1}=diag(a_1^{-1},a_2^{-1},\cdots,a_m^{-1}).$$

（11）若A 非奇异，则

A为正交矩阵 $\Leftrightarrow A^{-1}=A^{T}$

A为酉矩阵 $\Leftrightarrow A^{-1}=A^{H}$

矩阵求逆的几个公式

◊ 矩阵求逆引理(Sherman-Morrison公式)

$$(A+xy^H)^{-1}=A^{-1}-\frac{A^{-1}xy^HA^{-1}}{1+y^HA^{-1}x}$$

◊ Woodbury 公式

$$\begin{aligned}(A+UBV)^{-1}&=A^{-1}-A^{-1}UB(B+BVA^{-1}UB)^{-1}BVA^{-1}\\&=A^{-1}-A^{-1}U(I+BVA^{-1}U)^{-1}BVA^{-1}\\&=A^{-1}-A^{-1}U(B^{-1}+VA^{-1}U)^{-1}VA^{-1}\end{aligned}$$

◊ Duncan-Guttman公式

$$(A-UD^{-1}V)^{-1}=A^{-1}-A^{-1}U(D+VA^{-1}U)^{-1}VA^{-1}$$

矩阵求逆引理的典型应用:相关矩阵求逆的迭代递推.

补充内容，仅供参考. **Sherman-Morrison**公式和**Duncan-Guttman**公式都是**Woodbury** 公式的特例.

Sherman-Morrison公式的证明：

前提： $A, A+xy^H$ 可逆

$$(A+xy^H)(A^{-1}-\frac{A^{-1}xy^HA^{-1}}{1+y^HA^{-1}x})$$

$$=AA^{-1}-A\frac{A^{-1}xy^HA^{-1}}{1+y^HA^{-1}x}+xy^HA^{-1}-xy^H\frac{A^{-1}xy^HA^{-1}}{1+y^HA^{-1}x}$$

$$=I-\frac{xy^HA^{-1}}{1+y^HA^{-1}x}+xy^HA^{-1}-\frac{xy^HA^{-1}xy^HA^{-1}}{1+y^HA^{-1}x}$$

$$=I+xy^HA^{-1}-\frac{xy^HA^{-1}+xy^HA^{-1}xy^HA^{-1}}{1+y^HA^{-1}x}$$

$$=I+xy^HA^{-1}-\frac{xy^HA^{-1}+x(y^HA^{-1}x)y^HA^{-1}}{1+y^HA^{-1}x}$$

$$=I+xy^HA^{-1}-\frac{xy^HA^{-1}+xy^HA^{-1}(y^HA^{-1}x)}{1+y^HA^{-1}x}$$

$$=I+xy^HA^{-1}-\frac{xy^HA^{-1}(1+y^HA^{-1}x)}{1+y^HA^{-1}x}=I$$

分块矩阵的求逆公式

（1）矩阵A可逆时

$$\begin{bmatrix} A & U \\ V & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}U(D - VA^{-1}U)^{-1}VA^{-1} & -A^{-1}U(D - VA^{-1}U)^{-1} \\ -(D - VA^{-1}U)^{-1}VA^{-1} & (D - VA^{-1}U)^{-1} \end{bmatrix}$$

（2）矩阵A和D可逆时

$$\begin{bmatrix} A & U \\ V & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - UD^{-1}V)^{-1} & -A^{-1}U(D - VA^{-1}U)^{-1} \\ -D^{-1}V(A - UD^{-1}V)^{-1} & (D - VA^{-1}U)^{-1} \end{bmatrix}$$

或者

$$\begin{bmatrix} A & U \\ V & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - UD^{-1}V)^{-1} & -(A - UD^{-1}V)^{-1}UD^{-1} \\ -(D - VA^{-1}U)^{-1}VA^{-1} & (D - VA^{-1}U)^{-1} \end{bmatrix}$$

或者

$$\begin{bmatrix} A & U \\ V & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - UD^{-1}V)^{-1} & -(V - DU^{-1}A)^{-1} \\ (U - AV^{-1}D)^{-1} & (D - VA^{-1}U)^{-1} \end{bmatrix}$$

$$\begin{bmatrix} A & U \\ V & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}U(D-VA^{-1}U)^{-1}VA^{-1} & -A^{-1}U(D-VA^{-1}U)^{-1} \\ -(D-VA^{-1}U)^{-1}VA^{-1} & (D-VA^{-1}U)^{-1} \end{bmatrix}$$

的推导：

$$\begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} A^{-1} & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} A & U \\ V & D \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} I & A^{-1}U \\ V & D \end{bmatrix},$$

$$\begin{bmatrix} I & 0 \\ V & I \end{bmatrix}\begin{bmatrix} I & 0 \\ -V & I \end{bmatrix}\begin{bmatrix} I & A^{-1}U \\ V & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ V & I \end{bmatrix}\begin{bmatrix} I & A^{-1}U \\ 0 & D-VA^{-1}U \end{bmatrix},$$

$$\begin{bmatrix} I & 0 \\ 0 & D-VA^{-1}U \end{bmatrix}\begin{bmatrix} I & 0 \\ 0 & (D-VA^{-1}U)^{-1} \end{bmatrix}\begin{bmatrix} I & A^{-1}U \\ 0 & D-VA^{-1}U \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & D-VA^{-1}U \end{bmatrix}\begin{bmatrix} I & A^{-1}U \\ 0 & I \end{bmatrix},$$

$$\begin{bmatrix} I & A^{-1}U \\ 0 & I \end{bmatrix}\begin{bmatrix} I & -A^{-1}U \\ 0 & I \end{bmatrix}\begin{bmatrix} I & A^{-1}U \\ 0 & I \end{bmatrix} = \begin{bmatrix} I & A^{-1}U \\ 0 & I \end{bmatrix}\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix};$$

$$\begin{bmatrix} A & U \\ V & D \end{bmatrix}^{-1} = \begin{bmatrix} I & -A^{-1}U \\ 0 & I \end{bmatrix}\begin{bmatrix} I & 0 \\ 0 & (D-VA^{-1}U)^{-1} \end{bmatrix}\begin{bmatrix} I & 0 \\ -V & I \end{bmatrix}\begin{bmatrix} A^{-1} & 0 \\ 0 & I \end{bmatrix}$$

Hermitian 矩阵的求逆引理

令Hermitian 矩阵的分块形式为：$R_{m+1} = \begin{bmatrix} R_m & r_m \\ r_m^H & \rho_m \end{bmatrix}$

则可用R_m^{-1} 递推R_{m+1}^{-1} ：

$$R_{m+1}^{-1} = \begin{bmatrix} R_m^{-1} & 0_m \\ 0_m^H & 0 \end{bmatrix} + \frac{1}{\beta_m} \begin{bmatrix} b_m b_m^H & b_m \\ b_m^H & 1 \end{bmatrix}$$

其中

$$b_m = -R_m^{-1} r_m$$

$$\beta_m = \rho_m - r_m^H R_m^{-1} r_m = \rho_m + r_m^H b_m$$

注：可以直接验证，也可以用上页分块矩阵求逆那样的思路推导.

6.1 广义逆矩阵

$\mathrm{rank}(AB) \le \min\{\mathrm{rank}(A), \mathrm{rank}(B)\}$.

左逆矩阵：满足 $LA = I_{n\times n}$ 的矩阵L称为A的左逆矩阵.

定理：仅当 $m \ge n$ 时，矩阵 $A \in C^{m\times n}$ 可能有左逆矩阵.

右逆矩阵：满足 $AR = I_{m\times m}$ 的矩阵R称为A的右逆矩阵.

定理：仅当 $m \le n$ 时，矩阵 $A \in C^{m\times n}$ 可能有右逆矩阵.

- 若 $m>n$ ，并且A是满列秩的，则

$$L = (A^H A)^{-1} A^H$$

为A的一个左逆矩阵，称其为A的左伪逆矩阵.

- 若 $m<n$ ，并且A是满行秩的，则

$$R = A^H (AA^H)^{-1}$$

为A的一个右逆矩阵，称其为A的右伪逆矩阵.

注：左(右)逆矩阵不一定唯一；左(右)伪逆矩阵是特指.

请注意如下推导中的"**不一定**"

$$L_{n\times m}A_{m\times n}=I_{n\times n}$$

$$\Rightarrow L_{n\times m}A_{m\times n}A_{m\times n}^{H}=I_{n\times n}A_{m\times n}^{H}$$

$$\Rightarrow L_{n\times m}(A_{m\times n}A_{m\times n}^{H})(A_{m\times n}A_{m\times n}^{H})^{+}=I_{n\times n}A_{m\times n}^{H}(A_{m\times n}A_{m\times n}^{H})^{+}$$

$$\overset{\text{不一定}}{\Rightarrow} L_{n\times m}=I_{n\times n}A_{m\times n}^{H}(A_{m\times n}A_{m\times n}^{H})^{+}$$

$$\because (A_{m\times n}A_{m\times n}^{H})(A_{m\times n}A_{m\times n}^{H})^{+}\overset{\text{不一定}}{=}I_{m\times m}$$

+：表示某种逆

基本广义逆矩阵

定理： 设 $A \in C^{m\times n}$, 则一致方程 $Ax=b$ 对任意 $b\neq 0$ 有解 $x_0 = Gb$, 当且仅当

$$AGA = A.$$

定义： $m\times n$ 矩阵A的**基本广义逆矩阵** A^- 是一个满足如下约束的 $n\times m$ 矩阵

$$AA^-A = A.$$

基本广义逆矩阵的等价定义

（1） A^-A为幂等矩阵，且$\mathrm{rank}(A^-A)=\mathrm{rank}(A)$

(2) AA^-为幂等矩阵，且$\mathrm{rank}(AA^-)=\mathrm{rank}(A)$

$$B \sim \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}, B^2 = B$$

定理： 矩阵 $A \in C^{m\times n}$有唯一基本广义逆 A^- 的充要条件为非奇异, 且此时

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

基本广义逆矩阵定理的证明

<u>充分性</u>:

因为$AGA=A$, 所以对于任意的 X, 都有$AGAX=AX$,

令$AX=b$(此方程有解, 使其成立的 X是存在的), 因而

$AGb=b$成立, $X=Gb$是方程的一个解.

<u>必要性</u>:

因为$AGb=b$对任意的 b成立.任取一个 X, 令$AX=b$,

所以$AGAX=AX$对任意的 X也成立.

因而, $AGA=A$.

A^- 的性质

（1） $(A^-)^H = (A^H)^-$.

（2） $(\lambda A)^- = \lambda^\dagger A^-$，其中$\lambda^\dagger = \begin{cases} \lambda^{-1}, \lambda \neq 0 \\ 0, \lambda = 0 \end{cases}$

$\mathrm{Range}(A) = \{Y : Y = AX, X \in \mathrm{C}^n\}$
$\mathrm{Null}(A) = \{X : AX = 0\}$

（3） 若S和T非奇异，则 $T^{-1}A^-S^{-1} = (SAT)^-$.

（4） $\mathrm{rank}(A^-) \geq \mathrm{rank}(A)$.

（5） AA^- 和 A^-A 均为幂等矩阵且与A同秩.

（6） $\mathrm{Range}(AA^-) = \mathrm{Range}(A), \mathrm{Null}(A^-A) = \mathrm{Null}(A),$

$\mathrm{Range}((AA^-)^H) = \mathrm{Range}(A^H).$

（7） $A^-A = I_n$ 的充要条件是$\mathrm{rank}(A) = n,$

$AA^- = I_m$ 的充要条件是$\mathrm{rank}(A) = m.$

（8） 若$A \in C^{m\times n}$，$B \in C^{n\times q}$，则

$AB(AB)^-A = A$的充要条件是 $\mathrm{rank}(AB) = \mathrm{rank}(A),$

$B(AB)^-AB = B$的充要条件是 $\mathrm{rank}(AB) = \mathrm{rank}(B).$

Moore-Penrose逆矩阵

定义： 令A为一任意$m\times n$ 矩阵.若 $n\times m$矩阵G满足以下四个条件:

（1）$AGA=A$;

（2）$GAG=G$;

（3）$(AG)^H=AG$, 即AG为Hermitian矩阵;

（4）$(GA)^H=GA$, 即GA为Hermitian矩阵.

则称矩阵G为A的Moore-Penrose广义逆，并记为$A^\dagger$.

定理： 对任意 $A\in C^{m\times n}$， $A^\dagger$存在并且唯一.

科学史：Moore于1935年从投影的角度，证明了矩阵A的广义逆矩阵A†必须满足两个条件，但这两个条件不方便使用. Penrose于1955年提出了定义广义逆矩阵的以上四个条件. 1956年Rado证明了Penrose四条件与Moore两条件等价. 人们后来将广义逆矩阵需要满足的四个条件习惯上称为Moore-Penrose条件. 并将这种广义逆矩阵称为Moore-Penrose逆矩阵.（参见:张贤达,矩阵分析与应用,2013,P62）

广义逆矩阵分类

①只满足条件(1)和(2)的矩阵$G=A^\dagger$称为A的自反广义逆矩阵 (reflexive generalized inverse);

②满足条件(1),(2)和(3)的矩阵$A^\dagger$称为A的正规化广义逆矩阵 (normalized generalized inverse);

③满足条件(1),(2)和(4)的矩阵$A^\dagger$称为A的弱广义逆矩阵(weak generalized inverse);

④满足四个条件的矩阵$A^\dagger$称为A的Moore-Penrose逆矩阵.

◊ 一般地，对于只满足部分条件的广义逆，可以用满足的条件编号来表示和命名.

如：$A^{(1,2,4)}$. $A\{1,2,3\}$ $A\{1,2,3,4\}$

```
a=rand(3,4)
r=a'*inv(a*a')
a*r*a
pinv(a)
```

```
a =
    0.6948    0.0344    0.7655    0.4898
    0.3171    0.4387    0.7952    0.4456
    0.9502    0.3816    0.1869    0.6463
r =
    0.7065   -0.9063    0.6464
   -1.8947    1.6039    0.7169
    0.8579    0.6181   -0.8958
   -0.1681    0.2069    0.4327
ans =
    0.6948    0.0344    0.7655    0.4898
    0.3171    0.4387    0.7952    0.4456
    0.9502    0.3816    0.1869    0.6463
ans =
    0.7065   -0.9063    0.6464
   -1.8947    1.6039    0.7169
    0.8579    0.6181   -0.8958
   -0.1681    0.2069    0.4327
```

Moore-Penrose逆矩阵与其他广义逆矩阵的关系：

（1）非奇异矩阵$A_{n\times n}$的逆矩阵 A^{-1}满足Moore-Penrose逆矩阵的所有四个条件，即此时$A^{\dagger} = A^{-1}$.

（2）满列秩矩阵$A_{m\times n}(m>n)$的左伪逆矩阵$(A^H A)^{-1}A^H$ 满足Moore-Penrose逆矩阵的全部四个条件，即此时$A^{\dagger}=(A^H A)^{-1}A^H$.

（3）满行秩矩阵 $A_{m\times n}(m<n)$ 的右伪逆矩阵$A^H(AA^H)^{-1}$ 也满足Moore-Penrose逆矩阵的所有四个条件，即此时$A^{\dagger}= A^H(AA^H)^{-1}$

（4）左逆矩阵 $L_{n\times m}$ 是弱广义逆矩阵 $A^{(1,2,4)}$.

（5）右逆矩阵是正规化广义逆矩阵 $A^{(1,2,3)}$.

（6） A^{-} 只满足Moore-Penrose逆矩阵条件(1).

(7) $A^{\dagger}$是唯一的.

(8) $(AB)^{\dagger}=B^{\dagger}A^{\dagger}$不一定成立.

(9) $A^{\dagger} = A^{(1,4)}AA^{(1,3)}$

(10) $A^{\dagger}=(A^{H}A)^{\dagger}A^{H}$ 或$A^{\dagger}=A^{H}(AA^{H})^{\dagger}$

注： 除$A^{\dagger}$唯一确定外，只满足部分Moore-Penrose条件的其余各种广义逆都不是唯一确定的. 此时，每一种广义逆矩阵都包含着一类矩阵，即为A{1,2,4}等(即 $A^{(1,2,4)} \in A\{1,2,4\}$).

若A是$m\times n$矩阵，则 A^+是$n\times m$矩阵，但一般 $AA^+\neq I_{m\times m}, A^+A\neq I_{n\times n}$.

```
B =
     1     0     0     0
     0     2     0     0
     0     0     3     0
>> pinv(B)
ans =
    1.0000         0         0
         0    0.5000         0
         0         0    0.3333
         0         0         0
>> B*pinv(B)
ans =
     1     0     0
     0     1     0
     0     0     1
>> pinv(B)*B
ans =
     1     0     0     0
     0     1     0     0
     0     0     1     0
     0     0     0     0
```

```
A =
    0.8147    0.9134    0.2785    0.9649
    0.9058    0.6324    0.5469    0.1576
    0.1270    0.0975    0.9575    0.9706
>> pinv(A)
ans =
   -0.0050    0.7369   -0.2626
    0.5386    0.0485   -0.3691
   -0.8371    0.8121    0.7604
    0.7724   -0.9025    0.3516
>> A*pinv(A)
ans =
    1.0000    0.0000   -0.0000
   -0.0000    1.0000    0.0000
   -0.0000   -0.0000    1.0000
>> pinv(A)*A
ans =
    0.6301    0.4358    0.1502   -0.1435
    0.4358    0.4866   -0.1769    0.1691
    0.1502   -0.1769    0.9391    0.0583
   -0.1435    0.1691    0.0583    0.9443
```

$(AB)^{\dagger}=B^{\dagger}A^{\dagger}$不一定成立

```
B =
     1     0     0     0
     0     1     0     0
     0     0     1     0
     0     0     0     1

A =
     1     0     0     0
     0     1     0     0
     0     0     1     0

>> pinv(A*B)-pinv(B)*pinv(A)

ans =
     0     0     0
     0     0     0
     0     0     0
     0     0     0
```

Yes,成立

```
>> A=rand(3,4)
A =
    0.5688    0.3371    0.3112    0.6020
    0.4694    0.1622    0.5285    0.2630
    0.0119    0.7943    0.1656    0.6541

>> B=rand(4,4)
B =
    0.6892    0.2290    0.5383    0.1067
    0.7482    0.9133    0.9961    0.9619
    0.4505    0.1524    0.0782    0.0046
    0.0838    0.8258    0.4427    0.7749

>> pinv(A*B)-pinv(B)*pinv(A)

ans =
    1.2782   -0.6291   -0.6509
  -17.2657    8.4973    8.7917
    7.9809   -3.9278   -4.0639
    9.9134   -4.8789   -5.0479
```

No,不成立

$A^\dagger$的性质

（1） $(A^\dagger)^\dagger=A$.

（2） $(A^H)^\dagger=(A^\dagger)^H$, $(A^T)^\dagger=(A^\dagger)^T$.

（3） $(A^HA)^\dagger=A^\dagger(A^H)^\dagger$， $(AA^H)^\dagger=(A^H)^\dagger A^\dagger$, 但一般情况下$(AB)^\dagger\neq B^\dagger A^\dagger$.

（4） 若 $\Sigma=\begin{bmatrix}\Sigma_1 & O\\ O & O\end{bmatrix}_{m\times n}$， 其中 $\Sigma_1=diag(\sigma_1,\sigma_2,\ldots,\sigma_r),(\sigma_i\neq 0,i=1,\ldots,r)$则

$$\Sigma^\dagger=\begin{bmatrix}\Sigma_1^\dagger & 0\\ 0 & 0\end{bmatrix}_{n\times m}, \text{其中 } \Sigma_1^\dagger=diag(\sigma_1^{-1},\sigma_2^{-1},\ldots,\sigma_r^{-1}).$$

（5） $rank(A^\dagger)=rank(A)$

（6） $Range(A^\dagger)=Range(A^H),Null(A^\dagger)=Null(A^H)$

（7） 若矩阵A的满秩分解为$A=FG$， 则

$$A^\dagger=(FG)^\dagger=G^H(GG^H)^{-1}(F^HF)^{-1}F^H$$

对于$m\times n$的矩阵A， 假设其秩为r， 若存在秩同样为r两个矩阵： $F_{m\times r}$(列满秩)和$G_{r\times n}$(行满秩)， 使得$A=FG$， 则称其为矩阵A的满秩分解.

Moore-Penrose逆矩阵的计算方法

此页†与+表示相同的意思

1.方程求解法

Step 1：分别求解矩阵方程

$$AA^{H}X^{H}=A\rightarrow X^{H}$$

$$A^{H}AY=A^{H}\rightarrow Y$$

Step 2：计算广义逆矩阵 $A^{\dagger}=XAY$

2.列递推算法

初始值

令 $k=2,3,\ldots,n$，进行如下递推计算：

$$\boldsymbol{d}_k=\boldsymbol{A}_{k-1}^{\dagger}\boldsymbol{a}_k$$

$$\boldsymbol{b}_k=\begin{cases}(1+\boldsymbol{d}_k^{\mathrm{H}}\boldsymbol{d}_k)^{-1}\boldsymbol{d}_k^{\mathrm{H}}\boldsymbol{A}_{k-1}^{\dagger}, & \boldsymbol{d}_k^{\mathrm{H}}\boldsymbol{d}_k\neq-1\\ (\boldsymbol{a}_k-\boldsymbol{A}_{k-1}\boldsymbol{d}_k)^{\dagger}, & \boldsymbol{d}_k^{\mathrm{H}}\boldsymbol{d}_k=-1\end{cases}$$

$$\boldsymbol{A}_k^{\dagger}=\begin{bmatrix}\boldsymbol{A}_{k-1}^{\dagger}-\boldsymbol{d}_k\boldsymbol{b}_k\\ \boldsymbol{b}_k\end{bmatrix}$$

3.迹方法

令 $A_{m\times n}$ 的秩为r.

此页†与+表示相同的意思

Step 1: 计算 $B = A^T A.$

Step 2: 令 $C_1 = I.$

Step 3: 计算

$$C_{i+1} = \frac{1}{i} tr(C_i B) I - C_i B, i = 1,2,\ldots, r-1$$

Step 4: 计算

$$A^+ = \frac{r}{tr(C_r B)} C_r A^T.$$

注意，$C_{i+1} B = O, tr(C_i B) \neq 0.$

4.奇异值分解(SVD)法

若$A=U\Sigma V^H$为 $A_{m\times n}$ 的奇异值分解，其中$\Sigma_{m\times n}$为对角矩阵，其元素包含奇异值σ_i，则

$$A^+ = V\Sigma^+ U^H,$$

此页†与+表示相同的意思

其中$\Sigma^\dagger$为一$n\times m$对角阵，其对角元素为

$$\begin{cases} 1/\sigma_i, \sigma_i \neq 0 \\ 0, \sigma_i = 0 \end{cases}$$

5.满秩分解法

若$A=FG$ 是矩阵$A_{m\times n}$的满秩分解，则

$$A^+ = G^H(F^H A G^H)^{-1} F^H.$$

注:满秩分解可以由奇异值分解得到. 求矩阵的Moore-Penrose逆，奇异值分解方法比较简单易用.

$$\begin{pmatrix} A_{m\times n} & I_{m\times m} \\ I_{n\times n} & 0_{n\times m} \end{pmatrix} \to \begin{pmatrix} I_{m\times n} & B_{m\times m} \\ C_{n\times n} & 0_{n\times m} \end{pmatrix} \to \begin{pmatrix} \begin{pmatrix} I_{r\times r} & 0 \\ 0 & 0 \end{pmatrix} & B_{m\times m} \\ C_{n\times n} & 0_{n\times m} \end{pmatrix}$$

$$\Rightarrow A_{m\times n} = B_{m\times m}^{-1} \begin{pmatrix} I_{r\times r} & 0 \\ 0 & 0 \end{pmatrix} C_{n\times n}^{-1} = B_{m\times m}^{-1} \begin{pmatrix} I_{r\times r} \\ 0 \end{pmatrix} \begin{pmatrix} I_{r\times r} & 0 \end{pmatrix} C_{n\times n}^{-1}$$

$$F = B_{m\times m}^{-1} \begin{pmatrix} I_{r\times r} \\ 0 \end{pmatrix}, G = \begin{pmatrix} I_{r\times r} & 0 \end{pmatrix} C_{n\times n}^{-1}$$

$$A = FG$$

$$\boxed{\mathrm{A}^{+} = G^{H}(F^{H}AG^{H})^{-1}F^{H}.}$$

Matlab求逆运算左除(\)和右除(/),inv,pinv的用法及区别

inv

若A为非奇异方阵,则存在逆矩阵,可利用inv求逆:inv(A).

左除(\)和右除(/)

对于非奇异方阵A，若需如下运算:inv(A)*B，则可以用矩阵左除(\)替代，形式如下:A\B，即A\B = inv(A)*B.

pinv

若需要求逆的矩阵A为奇异矩阵或者非方阵，则并不存在逆矩阵，此时可以使用pinv(A)求其伪逆(广义逆):

X=pinv(A); %返回矩阵 A 的 Moore-Penrose 伪逆.

X=pinv(A,tol); %其中, tol为误差.

```
A=rand(3,4);
[U,S,V]=svd(A)
SS=size(S);
SSS=min(SS(1),SS(2));
S1=S;
for II=1:SSS
    if S(II,II)>eps
        S1(II,II)=1/S(II,II);
    end
end
PA=(U*S1*V')'
PA1=pinv(A)
```

```
U =
   -0.7350   -0.0625    0.6751
   -0.4743   -0.6643   -0.5778
   -0.4846    0.7449   -0.4586
S =
    1.9188         0         0         0
         0    0.7543         0         0
         0         0    0.2598         0
V =
   -0.6132    0.4201    0.4111   -0.5277
   -0.5262    0.1633    0.1102    0.8273
   -0.1764    0.4704   -0.8599   -0.0905
   -0.5621   -0.7587   -0.2818   -0.1702
PA =
    1.2685   -1.1327   -0.1561
    0.4745   -0.2589    0.0995
   -2.2061    1.5419    2.0271
   -0.4541    1.4337   -0.1099
PA1 =
    1.2685   -1.1327   -0.1561
    0.4745   -0.2589    0.0995
   -2.2061    1.5419    2.0271
   -0.4541    1.4337   -0.1099
```

各种逆之间的关系

1) $A\{1\} \supset A\{1,2\} \supset A\{1,2,3\} \supset A\{1,2,3,4\}$

2)左(右)逆矩阵不一定唯一，也不一定存在；
左(右)伪逆矩阵是特指.

3)彭罗斯逆矩阵必定存在且唯一.

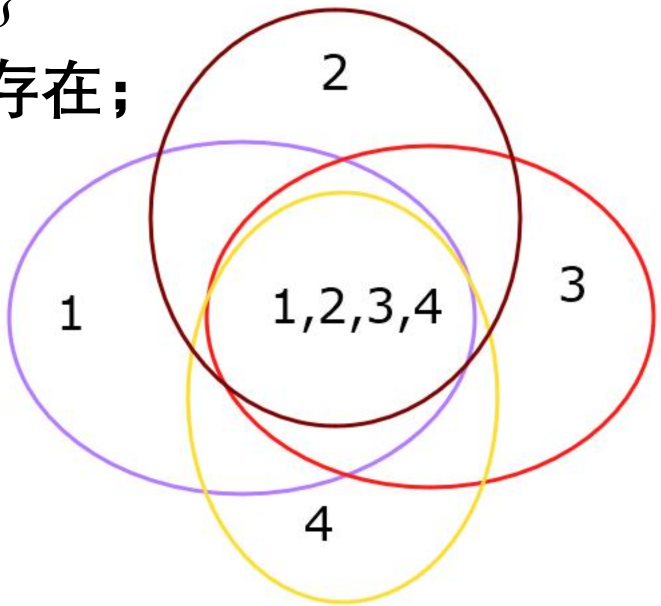
4)基本广义逆矩阵的全体：A{1}.

5)左逆矩阵若存在，必定属于A{1,2,4}.

6)右逆矩阵若存在，必定属于A{1,2,3}.

7)各种逆的讨论与矩阵方程求解问题相关.

8)基于彭罗斯逆，可以构造其它几种逆的通用表达式.



应用:线性方程组的求解

考虑关于$x \in C^n$ 的非齐次线性方程组

$$A_{m\times n}x_{n\times 1} = b_{m\times 1}$$

一致方程

矛盾方程

一致方程判断定理: 线性方程程 $Ax=b$ 是一致的，当且仅当

$$\operatorname{rank}([A,b]) = \operatorname{rank}(A)$$

定理1：设 $A\in C^{m\times n},B\in C^{p\times q},D\in C^{m\times q}$，则关于 $X\in C^{n\times p}$ 的矩阵方程

$$AXB=D$$

注:相容=有解

相容的充要条件是 $AA^{(1)}DB^{(1)}B=D$

且其通解为 $X=A^{(1)}DB^{(1)}+Y-A^{(1)}AYBB^{(1)}$

其中 $Y\in C^{n\times p}$是任意矩阵.

推论1：线性方程组 $Ax=b$相容的充要条件为

$$AA^{(1)}b=b$$

相容

且其通解为

$$x=A^{(1)}b+(I-A^{(1)}A)z,\forall z\in C^{n}$$

推论2：设 $A\in C^{m\times n}$， $A^{(1)}\in A\{1\}$，则

$$A\{1\}=\{A^{(1)}+Z-A^{(1)}AZAA^{(1)})\mid Z\in C^{n\times m}\}$$

定理1的证明.

必要性:

$$AXB = D \Rightarrow AA^{-}AXBB^{-}B = AA^{-}DB^{-}B \Rightarrow AXB = AA^{-}DB^{-}B$$

$$\Rightarrow D = AA^{-}DB^{-}B$$

充分性:

令 $X = A^{(1)}DB^{(1)}$ ， 则

$$AXB = AA^{(1)}DB^{(1)}B = D,$$

所以方程有解.

通解:

$$A(X - Y + Y)B = D \Rightarrow A(X - Y)B + AYB = D \Rightarrow A(X - Y)B = D - AYB$$

$\Rightarrow X - Y = A^{(1)}(D - AYB)B^{(1)}$是该方程的一个特解(对任意$Y$都成立).

$\Rightarrow X = A^{(1)}DB^{(1)} - A^{(1)}AYBB^{(1)} + Y$是其通解(由$Y$的任意性)

引理1：若线性方程组 $Ax=b$ 相容，则其极小范数解 x^o 唯一,且

$$x^o \in Range\,(A^H)$$

引理2：集合$A\{1,4\}$由矩阵方程

$$XA = A^{(1,4)}A$$

的所有解X组成，其中

$$A^{(1,4)} \in A\{1,4\}$$

验证满足条件1和条件4即可

$$(1)\,AXA = AA^{(1,4)}A = A$$

$$(4)(XA)^H = (A^{(1,4)}A)^H = A^{(1,4)}A = XA$$

推论3：设$A \in C^{m\times n}$，$A^{(1,4)} \in A\{1,4\}$，则

$$A\{1,4\} = \{A^{(1,4)} + Z(I - AA^{(1,4)}) \mid Z \in C^{n\times m}\}$$

定理2：设方程组 $Ax=b$ 相容，则其极小范数解为

$$x = A^{(1,4)}b$$

反之，设 $\overline{A} \in C^{n\times m}$.若对所有 $b \in Range(A), x = \overline{A}b$ 是方程的最小二乘解，则 $\overline{A} \in A\{1,4\}$.

引理1的证明思路:

$\forall z \in C^n$

$\| x \|_2^2 = \| A^{(1)}b + (I - A^{(1)}A)z \|_2^2$

$= \{A^{(1)}b + (I - A^{(1)}A)z\}^H \{A^{(1)}b + (I - A^{(1)}A)z\}$

$\Rightarrow$ 极值点z，代入通解公式即可得极小范数解

(需要用到矩阵微分知识)

引理3： 设$A\in C^{m\times n}$,则集合<u>A{1,3}</u>由矩阵方程

$$AX=AA^{(1,3)}$$

的所有解X组成，其中 $A^{(1,3)}\in A\{1,3\}$.

验证满足条件1和条件3即可

推论4： 设 $A\in C^{m\times n}$, $A^{(1,3)}\in A\{1,3\}$ 则

$$A\{1,3\}=\{A^{(1,3)}+(I-A^{(1,3)}A)Z\mid Z\in C^{n\times m}\}.$$

定理3： 设 $A\in C^{m\times n}, b\in C^{m}, A^{(1,3)}\in A\{1,3\}$，则矛盾线性方程组 $Ax=b$ 的最小二乘解为 $x=A^{(1,3)}b$. 反之，若对所有 $\bar{A}\in C^{n\times m}$，$b\in C^{m}, x=\bar{A}b$是方程的最小二乘解，则

$$\bar{A}\in A\{1,3\}.$$

推论5：x是方程组$Ax=b$的最小二乘解的充要条件是，x为

$$A^{H}Ax=A^{H}b$$

的解.

定理4：设 $A\in C^{m\times n}$, $b\in C^{m}$，则方程组$Ax=b$的极小范数最小二乘解为

此页+与†表示相同的意思

$$x=A^{+}b.$$

反之，设 $\overline{A}\in C^{m\times n}$. 若对所有$b\in C^{m}$, $x=\overline{A}b$ 是方程的极小范数最小二乘解，则 $\overline{A}=A^{+}$.

备注：（1）若 $Ax=b$ 相容，则极小范数最小二乘解$x=A^{\dagger}b$与极小范数解一致.

（2）当A满列秩时, $A^{+}=(A^{H}A)^{-1}A^{H}$，此时，超定方程或适定方程$Ax=b$中唯一的极小范数最小二乘解 $x=A^{+}B=(A^{H}A)^{-1}A^{H}b$ 也即为最小二乘解.

（3）当A满行秩时, $A^{+}=A^{H}(AA^{H})^{-1}$，此时，超定方程或适定方程 $Ax=b$ 中唯一的极小范数最小二乘解 $x=A^{+}b=A^{H}(AA^{H})^{-1}b$ 也即为极小范数解.

以上所说最小二乘解，都没有考虑系数矩阵和数据向量的误差或噪声问题!

定理5：若矩阵方程不相容，则其极小范数最小二乘解为

$$X^0 = \arg \min_{\min \|AXB-D\|_F} \|X\|_F$$

$$= A^\dagger D B^\dagger$$

不相容

符号和记号：

$$\min_{x \in D} f(x) = \min\{f(x) : x \in D\}$$

$$\arg\min_{x \in D} f(x) = \{x_0 : f(x_0) = \min\{f(x) : x \in D\}\}$$

$$x = \arg\min_{x \in D} f(x) \Leftrightarrow x \in \{x_0 : f(x_0) = \min\{f(x) : x \in D\}\}$$

$$X^0 = \arg \min_{\min \|AXB-D\|_F} \|X\|_F = \arg \min_{X = \arg\min \|AXB-D\|_F} \|X\|_F$$

例 求如下方程的通解、极小范数解

$$\begin{cases} x_1 + x_2 + x_3 = 2 \\ x_1 - x_2 + 6x_3 = 4 \end{cases}$$

解：令 $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 6 \end{pmatrix}, b = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$，其通解为 $X = A^+ b + (I - A^+ A)z, z \in C^3$

其极小范数解为函数 $y = \| A^+ b + (I - A^+ A)z \|$ 的极小值点 $, z \in C^3$

例 求 $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 6 \end{pmatrix}$ 的 $Moore-Penrose$ 逆.

解:
$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & -1 & 6 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \xrightarrow{\text{第一行加到第二行}} \begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 7 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \xrightarrow{\text{第一行减去第二行}} \begin{pmatrix} 0 & 1 & -6 & 0 & -1 \\ 1 & 0 & 7 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 7 & 1 & 1 \\ 0 & 1 & -6 & 0 & -1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

$$\xrightarrow{\text{第一列乘以}-7\text{，加到第三列}} \begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & -6 & 0 & -1 \\ 1 & 0 & -7 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \xrightarrow{\text{第二列乘以}6\text{，加到第三列}} \begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & -7 & 0 & 0 \\ 0 & 1 & 6 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

$A = FG.$ 其中，$F = \left(\begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}^{-1}; \begin{pmatrix} 0 & 0 \end{pmatrix} \right) = \begin{pmatrix} 1 & 1 \\ 0 & -1 \\ 0 & 0 \end{pmatrix}, G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & -7 \\ 0 & 1 & 6 \\ 0 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & 7 \\ 0 & 1 & -6 \end{pmatrix}$

$$F=\begin{pmatrix}1&0\\0&-1\end{pmatrix},G=\begin{pmatrix}1&0&7\\0&1&-6\end{pmatrix}$$

$$A^{+}=G^{H}(F^{H}AG^{H})^{-1}F^{H}=\begin{pmatrix}0.4568&-0.0617\\0.5309&-0.0988\\0.0123&0.1605\end{pmatrix}$$

$$pinv(A)=\begin{pmatrix}0.4103&-0.0385\\0.5641&-0.1154\\0.0256&0.1538\end{pmatrix}$$

例 令$A=\begin{pmatrix}1&1&1\\1&-1&6\end{pmatrix},b=\begin{pmatrix}2\\4\end{pmatrix}$，$X=A^{+}b+(I-A^{+}A)z,z\in C^{3}$，求$\|A^{+}b+(I-A^{+}A)z\|$的极值，及极值点

解：极值为0，极值点为$X_0=A^{+}b.$

例已知$A=\begin{pmatrix}1&1&1\\1&-1&6\end{pmatrix}$,求$A$的非零奇异值和奇异值 分解.

解：$AA^{T}=\begin{pmatrix}3&6\\6&38\end{pmatrix}=\begin{pmatrix}-0.9864&0.1644\\0.1644&0.9864\end{pmatrix}\begin{pmatrix}2&0\\0&39\end{pmatrix}\begin{pmatrix}-0.9864&0.1644\\0.1644&0.9864\end{pmatrix}^{-1}$

$$A^{T}A=\begin{pmatrix}2&0&7\\0&2&-5\\7&-5&37\end{pmatrix}=\begin{pmatrix}-0.7926&0.5812&0.1843\\0.5661&0.8137&-0.1316\\0.2265&0&0.9740\end{pmatrix}\begin{pmatrix}0&0&0\\0&2&0\\0&0&39\end{pmatrix}\begin{pmatrix}-0.7926&0.5812&0.1843\\0.5661&0.8137&-0.1316\\0.2265&0&0.9740\end{pmatrix}^{-1}$$

$$A=\begin{pmatrix}0.1644&-0.9864\\0.9864&0.1644\end{pmatrix}\begin{pmatrix}\sqrt{39}&0&0\\0&\sqrt{2}&0\end{pmatrix}\begin{pmatrix}0.1843&0.5812&-0.9726\\-0.1316&0.8137&0.5661\\0.9740&0&0.2265\end{pmatrix}$$

6.3 最小二乘法

$\tilde{X}$为解向量(其随机性由误差向量决定)

当方程

$$A_{m\times n}X_{n\times 1}=b_{m\times 1},\ A_{m\times n}X_{n\times 1}=b_{m\times 1}+e_{m\times 1}$$

是超定方程或矛盾方程时，解如何确定？(**矛盾是因为模型不准确，由误差引起**)，求如下最优化问题：

$$\min_X \| e_{m\times 1} \|_2^2 = \| A_{m\times n}X_{n\times 1}-b_{m\times 1} \|_2^2 .$$

若A满列秩，得

$$X_{LS}=(A^TA)^{-1}A^Tb.$$

左伪逆矩阵.

此方法称为普通最小二乘法（ordinary least squares,OLS)

定理(Gauss-Markov):线性方程组

$$A_{m\times n}X_{n\times 1}=b_{m\times 1}+e_{m\times 1}$$

的误差向量的均值和协方差矩阵分别为 $\mathbf{E}\{e\}=0, \mathbf{E}\{ee^H\}=\sigma^2 I$
则该方程存在最优无偏解的充分必要条件是rank(A)=n. 此时

$$X_{LS}=(A^TA)^{-1}A^Tb \qquad Var(X_{LS})\le Var(\tilde{X})$$

无偏估计：估计量的数学期望等于被估计参数的真实值.

$$(AX-b)^H(AX-b)$$

$$=(X^HA^H-b^H)(AX-b)$$

$$=X^HA^HAX-X^HA^Hb-b^HAX+b^Hb$$

$$\Rightarrow A^HAX=A^Hb(\text{求导数,极值条件})$$

$$\Rightarrow X=(A^HA)^{-1}A^Hb$$

误差由数据向量引起!

- **普通最小二乘问题(ordinary least squares,OLS)**

求X,使得数据向量引起的误差最小:

$$\min_X \| e_{m\times 1} \|_2^2 = \| A_{m\times n} X_{n\times 1} - b_{m\times 1} \|_2^2 .$$

普通最小二乘解也可以写成:

$$X_{LS} = \arg\min_X \| AX - b \|_2^2 .$$

若rank(A)=n,则

$$X_{LS} = (A^T A)^{-1} A^T b.$$

若rank(A)<n,则

$$X_{LS} = (A^T A)^{+} A^T b.$$

此页†与+表示相同的意思

如果误差分量是独立同分布的,则最小二乘解是无偏最优解(Gauss-Markov定理). 若不是独立同分布的,则不是无偏最优解.

```
clear all
MU=1.5;
SIGMA=1;
A= normrnd(MU,SIGMA,3,4)
b= normrnd(MU,SIGMA,3,1)
XLS1=(A'*A)\(A'*b)
XLS2=pinv(A'*A)*(A'*b)
```

```
A =
    1.7157    1.6049    0.8331   -0.4330
    0.3342    2.2223    1.6873    1.0610
    0.3520    4.0855    1.4175   -0.2947

b =
    2.3404
    0.6120
    1.6001

XLS1 =
    0.7875
    0.0107
    0.7260
   -0.8482

XLS2 =
    0.9836
    0.2391
    0.1103
   -0.4092
```

```
clear all
MU=1.5;
SIGMA=1;
A= normrnd(MU,SIGMA,4,3)
b= normrnd(MU,SIGMA,4,1)
XLS1=(A'*A)\(A'*b)
XLS2=pinv(A'*A)*(A'*b)
```

```
A =
    0.9555    2.2394    0.6604
    1.8035    3.2119    2.8546
    0.8997    1.3059    0.4278
    1.9900   -0.6384    2.4610

b =
    1.6240
    2.9367
   -0.4609
    1.3023

XLS1 =
   -0.7449
    0.3593
    1.1630

XLS2 =
   -0.7449
    0.3593
    1.1630
```

误差由数据矩阵引起!

● 数据最小二乘问题(data least squares,DLS)

对于超定方程$AX=b$, 假定数据向量b无观测误差或噪声， 只有数据矩阵A有观测误差或噪声， 令$A=A_0+E$, 并且E的**每个元素是零均值、等方差的独立同分布的高斯随机变量**.

用校正矩阵ΔA补偿数据矩阵中存在的误差矩阵， 使得

$$\begin{cases}(A+\Delta A)X=b\\ \mathbf{E}(A_0+\Delta A+E)=A_0+\Delta A\end{cases} \Rightarrow (A_0+\Delta A)X=b.$$

注： 仅知道A的值， 真实值A_0是不可观测的. 只能基于$AX=b$估算X.

此时， 求如下最优解问题的方法称为数据最小二乘(Data Least Squares, DLS)法.

$$X_{DLS}=\arg\min_X \|\Delta A\|_2^2 \quad \text{subject to} \quad b\in \text{Range}(A+\Delta A).$$

注：

$$b\in \text{Range}(A+\Delta A) \Leftrightarrow \exists X, s.t. (A+\Delta A)X=b.$$

迹函数的性质:

$$\| \Delta A \|_2^2 = trace\,(\Delta A(\Delta A)^H), trace\,(BC) = trace\,(CB),$$

$$< A, B >= vec(A)^H vec(B) = trace\,(A^H B).$$

推导1:

$$(A+\Delta A)X = b \Rightarrow AX - b = -\Delta AX$$

$$\Rightarrow < AX-b, AX-b >=< -\Delta AX, -\Delta AX >$$

$$\Rightarrow (AX-b)^H (AX-b) = X^H (\Delta A)^H \Delta AX$$

$$\Rightarrow \frac{(AX-b)^H (AX-b)}{X^H X} = \frac{X^H (\Delta A)^H \Delta AX}{X^H X}, X \neq 0$$

$$(\Delta A)^H \Delta A = \sigma^2 I,$$

$$\frac{X^H (\Delta A)^H \Delta AX}{X^H X} = \sigma^2$$

$$\because \| \Delta A \|_F^2 = trace((\Delta A)^H \Delta A) = trace(diag(\sigma^2, \cdots, \sigma^2)) = n\sigma^2$$

$$\Rightarrow X_{DLS} = \arg\min_{X\neq 0} \| \Delta A \|_2^2 = \arg\min_{X \neq 0} \frac{(AX-b)^H (AX-b)}{X^H X}.$$

推导2：该问题等价于求X, 使得

$J(X)=\|\Delta A\|_F^2=trace(\Delta A(\Delta A)^H)$达到最小,同时$(A+\Delta A)X=b$.

令$L(X)=trace(\Delta A(\Delta A)^H)+2\lambda^H(AX+\Delta AX-b)$,则

$$\partial L/\partial(\Delta A)^{\mathrm{H}}=2(\Delta A)^{\mathrm{H}}+2X\lambda^{\mathrm{H}} \quad (\text{参见Page223})$$

令$\partial L/\partial(\Delta A)^{\mathrm{H}}=0$,得

$$\Delta A=-\lambda X^H$$

代入$AX+\Delta AX-b=0$,得$\lambda=\dfrac{AX-b}{X^HX}$,从而$\Delta A=-\dfrac{(AX-b)X^H}{X^HX}$,

因而,目标函数 (objective/ cost function)

$$J(X)=trace(\Delta A(\Delta A)^H)=trace(\frac{(AX-b)X^H}{X^HX}\cdot\frac{X(AX-b)^H}{X^HX})$$

$$=trace(\frac{(AX-b)(AX-b)^H}{X^HX})=trace(\frac{(AX-b)^H(AX-b)}{X^HX})$$

$$=\frac{(AX-b)^H(AX-b)}{X^HX}.$$

由此得 $$X_{DLS}=\arg\min_{X\neq 0}\frac{(AX-b)^H(AX-b)}{X^HX}.$$

- **Tikhonov(吉洪诺夫)正则化方法**

解 $A_{m\times n}X_{n\times 1}=b_{m\times 1}(m>n)$ 的两个基本假设：(1)数据矩阵非奇异或者列满秩，(2)数据向量和数据矩阵存在加性噪声或误差(最好是独立同分布)． **矩阵秩亏缺时，问题如何解决?**

1963年Tikhonov提出了用正则化最小二乘代价函数改进普通最小二乘法的目标函数

$$J(X)=\frac{1}{2}(\| AX-b\|_2^2+\lambda\| X\|_2^2)$$

式中$\lambda\geqq 0$称为正则化参数(regularization parameters). 注意到

$$\frac{\partial J(X)}{\partial X^H}=A^H AX-A^H b+\lambda X$$

令$\partial J(X)/\partial X^H=0$,得解

$$X_{Tik}=(A^H A+\lambda I)^{-1}A^H b.$$

$$X_{Tik}=\arg\min_X(\| AX-b\|_2^2+\lambda\| X\|_2^2).$$

```
clear all
MU=1.5;
SIGMA=1;
A= normrnd(MU,SIGMA,4,3)
b= normrnd(MU,SIGMA,4,1)
l=1
T=(A'*A+l.*eye(3,3))\A'*b
l=2
T=(A'*A+l.*eye(3,3))\A'*b
l=10
T=(A'*A+l.*eye(3,3))\A'*b
l=100
T=(A'*A+l.*eye(3,3))\A'*b
```

```
A =
    0.6345   -0.8299    1.9517
    1.3235    0.0509    1.3697
    2.2914    1.8335    1.6837
    0.1680    1.8914    1.0238
b =
    2.3620
    0.1383
    1.9550
    0.6513

l =  1
T =
    0.1236
   -0.0306
    0.7559

l =  2
T =
    0.1839
   -0.0142
    0.6488

l = 10
T =
    0.2003
    0.0439
    0.3673

l = 100
T =
    0.0525
    0.0222
    0.0758
```

Tikhonov正则化方法：以$(A^{\mathrm{H}}A+\lambda I)$代替**协方差矩阵**$(A^{\mathrm{H}}A)^{-1}$的方法，称为松弛法(relaxation method), λ称为松弛因子.

反正则化方法(deregularizd method)：若数据矩阵A满列秩，则可以采用与正则化相反的方法，给被噪声污染的协方差矩阵$A^{\mathrm{H}}A$加一个很小的负扰动矩阵，使$A^{\mathrm{H}}A$消去部分干扰. 此时，解为

$$X = (A^H A - \lambda I)^{-1} A^H b.$$

Matlab里用：
norm(A)
norm(A,'fro')

关于条件数的说明：

$$\mathrm{cond}(A) = \| A \|_2 \cdot \| A^{-1} \|_2$$

$$\| A \|_2 = \max_{X \neq 0} \frac{\| AX \|_2}{\| X \|_2} \text{(诱导}L_2\text{范数)}, \quad \| A \|_F = \sqrt{trace(A^H A)}$$

因而

$$\mathrm{cond}(A^H A + \lambda I) = \mathrm{cond}(diag(\sigma_1 + \lambda, \sigma_2 + \lambda, \cdots, \sigma_n + \lambda)) = \frac{\sigma_1 + \lambda}{\sigma_n + \lambda}.$$

$$\mathrm{cond}(A^H A + \lambda I) = \frac{\sigma_1 + \lambda}{\sigma_n + \lambda} \leq (\lambda + \| A \|_F^2) / \lambda.$$

此处，奇异值为协方差矩阵$A^{\mathrm{H}}A$的

$$\|A\|_F = \left[\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right]^{1/2} = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2}$$

```
A=rand(3,4)
[U,S,V]=svd(A)
cond(A'*A+0.1*eye(4,4))
cond(A'*A+1*eye(4,4))
cond(A'*A+10*eye(4,4))
cond(A'*A+100*eye(4,4))
```

```
A =
    0.0357    0.6787    0.3922    0.7060
    0.8491    0.7577    0.6555    0.0318
    0.9340    0.7431    0.1712    0.2769
U =
   -0.4150    0.9089   -0.0404
   -0.6625   -0.3323   -0.6713
   -0.6236   -0.2519    0.7401
S =
    1.9059         0         0         0
         0    0.7689         0         0
         0         0    0.3833         0
V =
   -0.6085   -0.6307    0.3124    0.3665
   -0.6543    0.2314    0.0362   -0.7190
   -0.3693    0.1243   -0.8588    0.3328
   -0.2554    0.7302    0.4045    0.4878
```

Matlab函数cond( )用的是诱导L2范数.

```
A =
    0.9572    0.1419    0.7922
    0.4854    0.4218    0.9595
    0.8003    0.9157    0.6557

>> cond(A)
ans =
    5.7137

>> norm(A)*norm(inv(A))
ans =
    5.7137

>> norm(A,'fro')*norm(inv(A),'fro')
ans =
    7.2384
```

ans = 37.3264（大）
ans = 4.6326
ans = 1.3633
ans = 1.0363（小）

请注意两者的区别:

$$\|A\|_p = \max_{X\neq 0}\frac{\|AX\|_p}{\|X\|_p}\text{(诱导}L_p\text{范数)}$$

$$\|A\|_p = \left(\sum_{i,j}|a_{ij}|^p\right)^{\frac{1}{p}}\text{(普通}L_p\text{范数)}$$

```
a =
     1     0     0
     0     2     0
     0     0     3
>> norm(a)
ans =
     3
>> norm(a,'fro')
ans =
    3.7417
```

```
>> A=rand(3,5)
A =
    0.9572    0.1419    0.7922    0.0357    0.6787
    0.4854    0.4218    0.9595    0.8491    0.7577
    0.8003    0.9157    0.6557    0.9340    0.7431
>> norm(A)
ans =
    2.6755
>> norm(A,'fro')
ans =
    2.8264
```

令$A=U\Sigma V^H$是A的SVD分解，则 $A^H A=V\Sigma^2 V^H$,因而

$$X_{DS}=(A^H A)^{-1}A^H b=V\Sigma^{-1}U^H b$$

$$X_{Tik}=(A^H A+\sigma_{\min}^2 I)^{-1}A^H b=V(\Sigma^2+\sigma_{\min}^2)^{-1}\Sigma U^H b$$

$$(\Sigma^2+\sigma_{\min}^2)^{-1}\Sigma=diag\left(\frac{\sigma_1}{\sigma_1^2+\sigma_{\min}^2},\cdots,\frac{\sigma_n}{\sigma_n^2+\sigma_{\min}^2}\right)$$

Tikhonov正则化解有如下性质:

此页†与+表示相同的意思

$$X=(A^H A+\lambda I)^{-1}A^H b.$$
$$X=(A^H A-\lambda I)^{-1}A^H b.$$

(1)X_{Tik}关于数据向量是线性的 .(线性性)

(2)$\lim\limits_{\lambda\to 0}X_{Tik}=X_{LS}=A^{+}b,\ \lim\limits_{\lambda\to\infty}X_{Tik}=0$.(极限)

(3)$X_{Tik}=\arg\min\limits_{A^H(AX-b)=-\lambda X}\|X\|_2$.(在可行解中,有最小$L_2$范数)

(4)在$[0,\infty)$上，X_{Tik}是关于λ的光滑向量函数 .(正则化路径)

$\frac{\partial J(X)}{\partial X^H}=A^H AX-A^H b+\lambda X$	$J(X)=\frac{1}{2}(\|AX-b\|_2^2+\lambda\|X\|_2^2)$

$$\boxed{\frac{\partial J(X)}{\partial X^H} = A^H AX - A^H b + \lambda X}$$

$$\boxed{X^H A^H AX - X^H A^H b + \lambda X^H X = 0}$$

$$\begin{aligned}
&(AX-b)^H (AX-b) + \lambda \| X \|^2 \\
&= (X^H A^H - b^H)(AX-b) + \lambda \| X \|^2 \\
&= X^H A^H AX - 2X^H A^H b + b^H b + \lambda \| X \|^2 \\
&= b^H b - X^H A^H b \\
&= b^H b - \lambda \| X \|^2 - X^H A^H AX
\end{aligned}$$

- **总体最小二乘问题(total least squares,TLS)**

方程$AX=b$的A和b都存在误差或噪声. $A=A_0+E, b=b_0+e.$
A_0和b_0分别代表不可观测的无误差数据矩阵和数据向量.

总体最小二乘问题可以用约束优化问题叙述为:

$$\text{TLS：} \min_{\Delta A, \Delta b, X} \| [\Delta A, \Delta b] \|_2^2$$
$$\text{subject to } (A+\Delta A)X = b+\Delta b.$$

注意到

$$(A+\Delta A)X = b+\Delta b \Rightarrow AX-b = -(\Delta A X - \Delta b) = -[\Delta A, \Delta b][X^H, -1]^H$$

可用数据最小二乘解的“推导2”的方式，得

$$X_{TLS} = \arg\min_{X\neq 0} \frac{(AX-b)^H (AX-b)}{X^H X+1}.$$

```
clear all
MU=1.5;
SIGMA=1;
A= normrnd(MU,SIGMA,4,3)
b= normrnd(MU,SIGMA,4,1)
[U,S,V]=svd([A,b])
l=S(4,4)
XTLS=(A'*A-l*l*eye(3,3))\A'*b
XLS=(A'*A)\A'*b
```

```
A =
    1.2219    0.2872    2.5826
    1.9227    1.5662    2.5061
   -0.1702    2.1524    0.8491
    1.9716    1.8271    1.7571

b =
    0.5556
    0.1782
    2.4248
    1.5000
```

```
U =
   -0.4307   -0.3975    0.7767    0.2308
   -0.5571   -0.3833   -0.3060   -0.6702
   -0.4064    0.8297    0.2776   -0.2634
   -0.5823    0.0816   -0.4755    0.6544
S =
    5.9691         0         0         0
         0    2.7639         0         0
         0         0    1.0908         0
         0         0         0    0.6182
V =
   -0.4483   -0.4353   -0.5721    0.5313
   -0.4917    0.4416   -0.4836   -0.5740
   -0.6494   -0.4122    0.5861   -0.2547
   -0.3681    0.6676    0.3088    0.5688

l =
    0.6182
XTLS =                          XLS =
   -0.9340                         -0.6157
    1.0091                          0.9467
    0.4477                          0.2591
```

总体最小二乘解的闭式解

解析解(analytical solution)：用容易计算的公式给出的自变量和因变量之间的对应关系的函数, 也称为**闭式解(closed-form solution).**

数值解(numerical solution)：采用某种计算方法,如有限元的方法, 数值逼近,插值的方法, 得到的解.

若增广矩阵$B=[A,b]$的奇异值为$\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_{n+1}$,则总体最小二乘解可表示为

$$X_{TLS}=(A^H A-\sigma_{n+1}^2 I)^{-1} A^H b.$$

$$X_{TLS}=\arg\min_{X\neq 0} \frac{(AX-b)^H (AX-b)}{X^H X+1}.$$

若增广矩阵具有比较好的性质，则总体最小二乘解等价于特殊参数的吉洪诺夫反正则化方法(deregularizd method)解

解向量的比较

$$AX = b$$

$$X_{LS} = \arg\min_{X} \| AX - b \|_2^2 .$$

$$X_{LS} = (A^T A)^+ A^T b.$$

$$X_{DLS} = \arg\min_{X \neq 0} \frac{(AX-b)^H (AX-b)}{X^H X}.$$

$$X_{Tik} = \arg\min_{X} (\| AX - b \|_2^2 + \lambda \| X \|_2^2).$$

$$X_{Tik} = (A^H A + \lambda I)^{-1} A^H b.$$

$$X_{TLS} = \arg\min_{X \neq 0} \frac{(AX-b)^H (AX-b)}{X^H X + 1}.$$

$$X_{TLS} = (A^H A - \sigma_{n+1}^2 I)^{-1} A^H b.$$

注：此闭式解解对增广矩阵的奇异值有要求

也就是说，当考虑系数矩阵和数据向量在数据采集过程中有误差和噪声时，总可以将方程求解问题转化为最优化问题. 针对不同特点，最优化函数形式不一样.

再论最小二乘拟合（参见第二章最小二乘拟合）

设 $(x_i, y_i), i=1,2,\ldots,m$ 为一组观测数据.试寻找一个n次非零多项式

$$f(x)=a_0+a_1x+\cdots+a_nx^n=\sum_{i=0}^{n}a_ix^i\,(n<m)$$

以最小化

$$J=\sum_{j=1}^{m}\left|y_j-f(x_j)\right|^2$$

$$=\sum_{j=1}^{m}\left|y_j-\sum_{i=0}^{n}a_i(x_j)^i\right|^2$$

为此，令

$$\frac{\partial J}{\partial a_k}=-2\sum_{j=1}^{m}(x_j)^k\left[y_j-\sum_{i=0}^{n}a_i(x_j)^i\right]=0, k=0,1,\ldots,n$$

即

$$\frac{\partial J}{\partial a_k}=-2\sum_{j=1}^{m}(x_j)^k y_j+2\sum_{i=0}^{n}a_i\sum_{j=1}^{m}(x_j)^{i+k}=0,\quad k=0,1,\ldots,n$$

若记 $\alpha_{k+i}=\sum_{j=1}^{m}(x_j)^{k+i}, \beta_k=\sum_{j=1}^{m}(x_j)^k y_j, \quad k=0,1,\ldots,n$ ，则上式可得

$$\sum_{j=1}^{m}(x_j)^k y_j=\sum_{i=0}^{n} a_i \sum_{j=1}^{m}(x_j)^{k+i} \Rightarrow \beta_k=\sum_{i=0}^{n}\alpha_{k+i}a_i$$

即 $H_{n+1}a=b$ 其中

$$H_{n+1}=\begin{array}{l} \left[\begin{array}{ccccc} \alpha_0 & \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_{n+1} \\ \alpha_2 & \alpha_3 & \alpha_4 & \cdots & \alpha_{n+2} \\ \vdots & \vdots & \vdots & & \vdots \\ \alpha_n & \alpha_{n+1} & \alpha_{n+2} & \cdots & \alpha_{2n} \end{array}\right] \end{array} \begin{array}{l} \leftarrow k=0 \\ \leftarrow k=1 \\ \leftarrow k=2 \\ \quad\vdots \\ \leftarrow k=n \end{array}$$

为一Hankel矩阵，

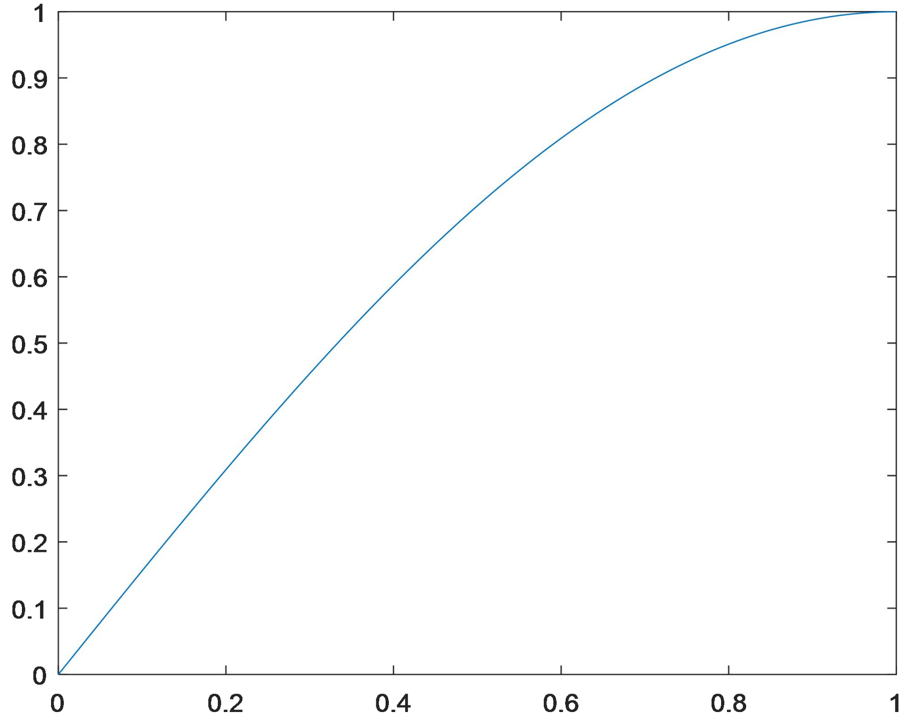
$$a=(a_0,a_1,\cdots,a_n)^T, \qquad b=(\beta_0,\beta_1,\cdots,\beta_n)^T$$

即最小二乘数据拟合问题转化为求解一以Hankel矩阵为系数矩阵的线性方程.

以上方法可用于一般函数的多项式拟合

```
N=7;%阶数, N+1个系数
M=20;%数据采样点个数
x=rand(M,1);
y=sin(pi/2*x);
a=zeros(1,2*N+1);
b=zeros(N+1,1);
for k=1:2*N+1
    for ii=1:M
        a(k)=a(k)+x(ii)^(k-1);
    end
end
for k=1:N+1
    for ii=1:M
        b(k)=b(k)+y(ii)*x(ii)^(k-1);
    end
end
```

```
H=zeros(N+1,N+1);
for ii=1:N+1
    for jj=1:N+1
        H(ii,jj)=a(ii+jj-1);
    end
end
disp(H);
A=H\b;
disp(A');
tt=0:0.01:1;
yy=zeros(1,length(tt));
for ii=1:length(tt)
    for jj=1:N+1
        yy(ii)=yy(ii)+A(jj)*tt(ii)^(jj-1);
    end
end
plot(tt,yy);
```




```
>> ttt
  20.0000    9.8706    6.4347    4.6450    3.5666    2.8572    2.3589    1.9910
   9.8706    6.4347    4.6450    3.5666    2.8572    2.3589    1.9910    1.7085
   6.4347    4.6450    3.5666    2.8572    2.3589    1.9910    1.7085    1.4849
   4.6450    3.5666    2.8572    2.3589    1.9910    1.7085    1.4849    1.3036
   3.5666    2.8572    2.3589    1.9910    1.7085    1.4849    1.3036    1.1537
   2.8572    2.3589    1.9910    1.7085    1.4849    1.3036    1.1537    1.0279
   2.3589    1.9910    1.7085    1.4849    1.3036    1.1537    1.0279    0.9209
   1.9910    1.7085    1.4849    1.3036    1.1537    1.0279    0.9209    0.8289

  -0.0000    1.5708   -0.0000   -0.6457   -0.0011    0.0820   -0.0026   -0.0034
```

案例7：最小二乘拟合

设$(x_i, y_i), i=1,2,\ldots,m$ 为一组观测数据.试寻找一个n次非零多项式

$$f(x)=a_0+a_1x+\cdots+a_nx^n=\sum_{i=0}^{n}a_ix^i$$

令$a=(a_0,a_1,\cdots,a_n)^T, v(x)=(1,x,x^2,\cdots,x^n)$, 则$f(x)=v(x)a$.令

$$b=(y_1,y_2,\cdots,y_m)^T, A=(v(x_1);v(x_2);\cdots;v(x_m)),$$

则以上标量方程可记为

$$b=Aa$$

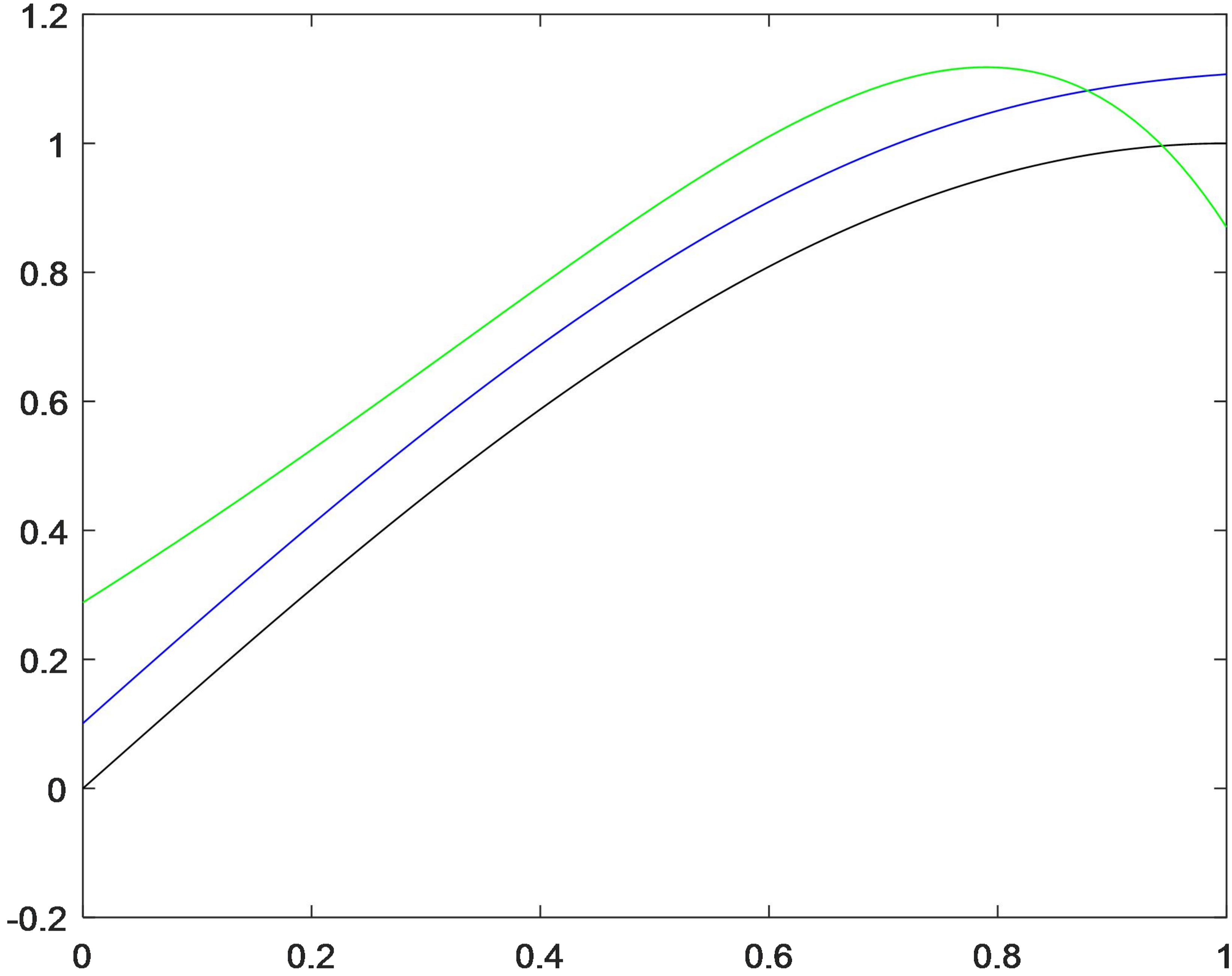
可分三种情况讨论以上最小二乘拟合问题((n+1)>m,n+1=m,(n+1)<m).

$$b = Aa$$

```
clear all; rng default
N=7;%阶数，N+1个系数
M=20;%数据采样点个数
x=rand(M,1);
b=sin(pi/2*x);%函数取值无噪声！！！
a=zeros(N+1,1);
A=zeros(M,N+1);
for ii=1:M
    for jj=1:N+1
        A(ii,jj)=x(ii)^(jj-1);
    end
end
BltN=A(1:N-1,:);
BeqN=A(1:N+1,:);
BgtN=A;
%(1)方程个数等于未知数个数
B=BeqN;
bb=b(1:N+1);
a=B\bb;
%a=(B'*B+.00012*eye(1+N,1+N))\B'*bb;
tt=0:0.01:1;
yy=zeros(1,length(tt));
for ii=1:length(tt)
    for jj=1:N+1
        yy(ii)=yy(ii)+a(jj)*tt(ii)^(jj-1);
    end
end
plot(tt,yy,'Color','k');
hold on
%(2)方程个数大于未知数个数
B=BgtN;
bb=b;
a=(B'*B+.00012*eye(1+N,1+N))\B'*bb;
yy=zeros(1,length(tt));
for ii=1:length(tt)
    for jj=1:N+1
        yy(ii)=yy(ii)+a(jj)*tt(ii)^(jj-1);
    end
end
plot(tt,yy+0.1,'Color','b');
hold on
%(3)方程个数小于未知数个数
B=BltN;
bb=b(1:N-1);
a=(B'*B+.0012*eye(1+N,1+N))\B'*bb;
yy=zeros(1,length(tt));
for ii=1:length(tt)
    for jj=1:N+1
        yy(ii)=yy(ii)+a(jj)*tt(ii)^(jj-1);
    end
end
plot(tt,yy+0.2,'Color','g');
```

数据实验：请试总体最小二乘法

函数取值无噪声的请况下，不同最小二乘法的效果比较！



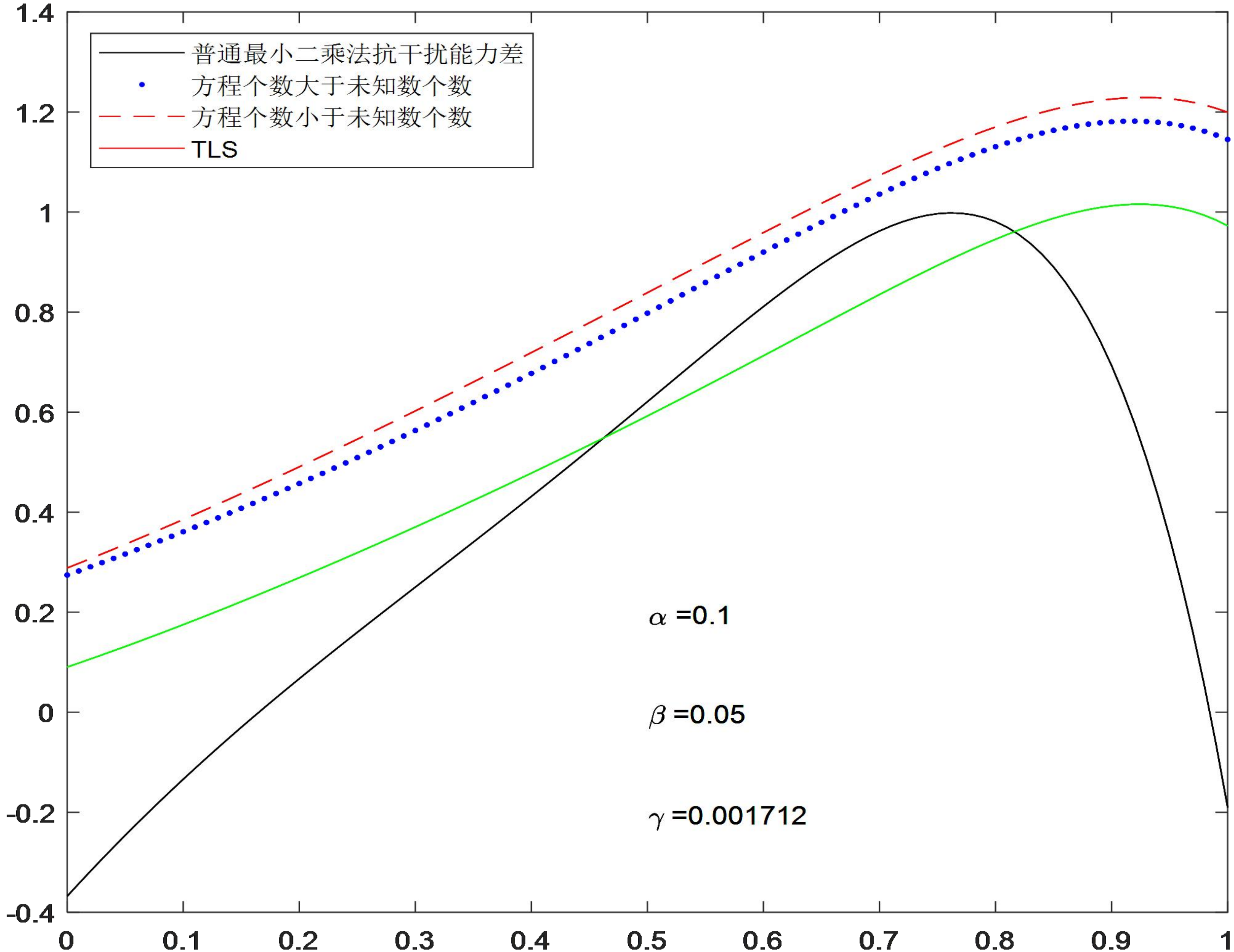
$$b = Aa$$

```
clear all; rng default
N=7;%阶数，N+1个系数
M=20;%数据采样点个数
x=rand(M,1);alpha=0.1;
b=sin(pi/2*x)+randn*alpha;%取值有噪声！
a=zeros(N+1,1);
A=zeros(M,N+1);beta=0.05;
for ii=1:M
    for jj=1:N+1
        A(ii,jj)=x(ii)^(jj-1)+randn*beta;%有噪声！
    end
end
BltN=A(1:N-1,:);
BeqN=A(1:N+1,:);
BgtN=A;
%(1):方程个数等于未知数个数
B=BeqN;
bb=b(1:N+1); a=B\bb;
%a=(B'*B+.00012*eye(1+N,1+N))\B'*bb;
tt=0:0.01:1;
yy=zeros(1,length(tt));
for ii=1:length(tt)
    for jj=1:N+1
        yy(ii)=yy(ii)+a(jj)*tt(ii)^(jj-1);
    end
end
plot(tt,yy,'Color','k');
hold on
%(2)方程个数大于未知数个数
B=BgtN;
bb=b;
a=(B'*B+.00012*eye(1+N,1+N))\B'*bb;
yy=zeros(1,length(tt));
for ii=1:length(tt)
    for jj=1:N+1
        yy(ii)=yy(ii)+a(jj)*tt(ii)^(jj-1);
    end
end
plot(tt,yy+0.1,'Color','b');
hold on
%(3)方程个数小于未知数个数
B=BltN;
bb=b(1:N-1);
a=(B'*B+.0012*eye(1+N,1+N))\B'*bb;
yy=zeros(1,length(tt));
for ii=1:length(tt)
    for jj=1:N+1
        yy(ii)=yy(ii)+a(jj)*tt(ii)^(jj-1);
    end
end
plot(tt,yy+0.2,'Color','g');
```

数据实验：请试总体最小二乘法

```
%(4)方程个数小于未知数个数
B=BltN;
bb=b(1:N-1);
%这是总体最小二乘法的简单处理方案，调整参数gama
gama=.001712;
a=(B'*B-gama*eye(1+N,1+N))\B'*bb;
yy=zeros(1,length(tt));
for ii=1:length(tt)
    for jj=1:N+1
        yy(ii)=yy(ii)+a(jj)*tt(ii)^(jj-1);
    end
end
plot(tt,yy+0.,'Color','r');
legend('普通最小二乘法抗干扰能力差','方程个数大于未知数个数','方程个数小于未知数个数','TLS','Location','NorthWest');
text(0.5,0.2,['\alpha =' ,num2str(alpha)]);
text(0.5,0,['\beta =' ,num2str(beta)]);
text(0.5,-0.2,['\gamma =' ,num2str(gamma)]);
```

函数取值有噪声的请况下，不同最小二乘法的效果比较！



6.6 稀疏矩阵方程求解

对于方程

$$A_{m\times n}X_{n\times 1}=b_{m\times 1},\quad A_{m\times n}x_{n\times p}=b_{m\times p},$$

若它的标量方程个数远远小于未知参量个数，则称该方程为稀疏矩阵方程(稀疏线性方程组).

再论范数

令$V^{m\times n}$ 是复矩阵空间. 函数$\|X\|:V^{m\times n}\mapsto R$称为矩阵$X$的范数，若对所有$X,Y\in V^{m\times n}$，满足

(1) $\|X\|\geq 0$ **,当且仅当 X=0时等号成立;(非负性)**

(2) $\|cX\|=|c|\|X\|,\forall c\in C$;**(正比例性)**

(3) $\|X+Y\|\leq\|X\|+\|Y\|$; **(三角不等式)**

(4) $\|XZ\|\leq\|X\|\cdot\|Z\|,\forall Z\in V^{n\times k}$.

◆几种常用的向量范数

$Supp(x) = |\{x_i \neq 0 : x = (x_1, x_2, \cdots, x_i, \cdots, x_n)\}|$

l_0 (准/拟)范数: $\|x\|_0 = Supp(x)$

l_1 范数: $\|x\|_1 = \sum_{i=1}^{n} |x_i|$

l_2 范数(Euclidean范数或Frobenius范数):

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

l_∞ 范数: $\|x\|_\infty = \max(|x_1|, |x_2|, \cdots, |x_n|)$

l_p 范数(Holder范数):

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

加权范数(椭圆范数): $\|x\|_A = \sqrt{x^H A x}$ （A为Hermitian正定）

◆ 几种常用的矩阵范数 $\left(A \in C^{m\times n}\right)$

$$\|A\|_1 = \sum_i^m \sum_j^n | a_{ij} |,$$

$$\|A\|_\infty = \max_{i,j}\{| a_{ij} |\},$$

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n \left|a_{ij}\right|^2}$$ (Frobenius范数)

$$\|A\|_\Omega = \sqrt{tr\left(A^H \Omega A\right)}, (\Omega > 0)$$ (Mahalanobis范数)

从属范数(诱导范数): $\|A\|_P = \max_{\|x\|_P=1} \|Ax\|_P$

◆ 列和范数: $\|A\|_1 = \max_j \sum_{i=1}^m \left|a_{i,j}\right|$

◆ 谱范数: $\|A\|_2 = \sqrt{\lambda_{\max}\left(A^H A\right)} = \| A \|_{spec}$

◆ 行和范数: $\|A\|_\infty = \max_i \sum_{j=1}^n \left|a_{i,j}\right|$

考虑稀疏矩阵方程 $Y=\Phi X$,其中$\Phi\in R^{m\times n}, X\in R^{n}, Y\in R^{m}$,且$m\prec\prec n$.
则稀疏矩阵方程的求解问题是L_0范数最小化问题

$$(\mathrm{P}_0)\qquad \min_{\mathrm{X}} \|X\|_0 \ \text{ subject to } Y=\Phi X$$

(P_0) 问题非常难解，是**NP难问题**(?), 可将该问题近似并转化为L_1范数下的最优化问题.

$$(\mathrm{P}_1)\qquad \min_{\mathrm{X}} \|X\|_1 \ \text{ subject to } Y=\Phi X$$

$$(\mathrm{P}_{1,0})\qquad \min_{\mathrm{X}} \|X\|_1 \ \text{ subject to } \|Y-\Phi X\|_2\le\varepsilon$$

与L_0范数最优化问题相似，L_1范数最优化问题也有两种变种:

$$(\mathrm{P}_{1,1})\qquad \min_{\mathrm{X}} \|Y-\Phi X\|_2 \ \text{ subject to } \|X\|_1\le \mathrm{K}$$

$$(\mathrm{P}_{1,2})\qquad \min_{\mathrm{X}} \|Y-\Phi X\|_2^2+\lambda\|X\|_1$$ **(Lasso的优化目标)**

https://blog.csdn.net/u014295667/article/details/47090639

K-稀疏：一个向量X是K稀疏的，若

$$X \in \mathrm{C}^N, \| X \|_0 \leq \mathrm{K}, \text{其中} \mathrm{K} \in \{1,2,\cdots,N\}$$

稀疏度：

$$\text{sparseness}(X) = \frac{\sqrt{n} - \| X \|_1 / \| X \|_2}{\sqrt{n} - 1}$$

求稀疏矩阵方程的稀疏度为s的整体最优解的朴素方法：**令r=[ns]，选取r个未知数，令其他未知数都等于零，求取对应超定方程的最小二乘解. 从全部最小二乘解中确定最优解.**

以上朴素方法不可行.可行的算法一般有：正交匹配追踪算法，LASSO算法与LARS算法，同伦算法，Bregman迭代算法等.

LARS algorithm

https://www.mathworks.com/matlabcentral/fileexchange/23186-lars-algorithm

https://cosx.org/2011/04/an-introduction-to-lars/

https://cosx.org/2011/04/modified-lars-and-lasso/

https://zhuanlan.zhihu.com/p/46999826

LASSO(The Least Absolute Shrinkage and Selectionator operator)问题是一类有约束的最优化问题，是一种经典的松弛算法模型, 1996年由Tibshirani提出. 基本思想：基于L1范数诱导最优化问题(**P12**)的稀疏解. LASSO三种求解方法：闭式解、LARS、CD.

LARS(Least Angle Regression)算法, 由Efron于2004年提出, 是一种解决Lasso问题的算法.

LASSO

https://ww2.mathworks.cn/help/stats/lasso-regularization.html?requestedDomain=cn

lasso

Lasso or elastic net regularization for linear models

Syntax

```
B = lasso(X,y)
B = lasso(X,y,Name,Value)
[B,FitInfo] = lasso( ___ )
```

Description

`B = lasso(X,y)` returns fitted least-squares regression coefficients for linear models of the predictor data `X` and the response `y`. Each column of `B` corresponds to a particular regularization coefficient in `Lambda`. By default, `lasso` performs lasso regularization using a geometric sequence of `Lambda` values.

`B = lasso(X,y,Name,Value)` fits regularized regressions with additional options specified by one or more name-value pair arguments. For example, `'Alpha',0.5` sets elastic net as the regularization method, with the parameter `Alpha` equal to 0.5.

`[B,FitInfo] = lasso( ___ )` also returns the structure `FitInfo`, which contains information about the fit of the models, using any of the input arguments in the previous syntaxes.

```
rng default % For reproducibility
X = randn(100,5);
weights = [0;2;0;-3;0];
% Only two nonzero coefficients
y = X*weights + randn(100,1)*0.1;
 % Small added noise
 B = lasso(X,y);
B(:,25)

ans =
        0
   1.6093
        0
  -2.5865
        0
```

```
X=normrnd(0,1,3,5)
Y=rand(3,1)
B=lasso(X,Y)
whos
B(:,1)

 Name      Size            Bytes  Class
Attributes

  B         5x39            1560  double
  X         3x5              120  double
  Y         3x1               24  double
  ans       5x1               40  double

ans =
   -0.5196
    0.0210
         0
         0
         0
```

案例8：ARMA模型

时间序列：具有某种统计规律的数值按其发生的时间先后顺序排列而成的数列. 时间序列分析的主要目的是根据已有的历史数据对未来进行预测.

ARMA 模型(Auto-Regressive and Moving Average Model)：是研究时间序列的重要方法，由自回归模型(简称AR模型)与滑动平均模型(简称MA模型)混合构成.

定义：若时间序列$\{y_t\}(t=1,2,\cdots)$可以表示为$\{y_{t-1},y_{t-2},\cdots,y_{t-p}\}$和残差项$U_t$的线性组合，即

$$y_t=\alpha_1 y_{t-1}+\alpha_2 y_{t-2}+\cdots+\alpha_p y_{t-p}+U_t$$

则称上式为y_t的p阶自回归模型，记为$AR(p)$，$\{y_t\}$也称为自回归序列.

定义：若时间序列$\{y_t\}(t=1,2,\cdots)$可以表示为残差项$\{U_t,U_{t-1},\cdots,U_{t-q}\}$的线性组合，即

$$y_t=U_t-\beta_1 U_{t-1}-\beta_2 U_{t-2}-\cdots-\beta_q U_{t-q}$$

则称上式为y_t的q阶滑动平均模型，记为$MA(q)$，$\{y_t\}$也称为滑动平均序列.

定义：对于时间序列$\{y_t\}(t=1,2,\cdots)$，称如下模型

$$y_t=\alpha_1 y_{t-1}+\alpha_2 y_{t-2}+\cdots+\alpha_p y_{t-p}+U_t-\beta_1 U_{t-1}-\cdots-\beta_q U_{t-q}$$

为y_t的自回归滑动平均模型,记为$ARMA(p,q)$.其中，

$\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\cdots,\alpha_p),\boldsymbol{\beta}=(\beta_1,\beta_2,\cdots,\beta_q)$分别称为自回归与滑动平均参数.

问题：谐波频率估计

$$x(n)=\sum_{i=1}^{p}A_i\sin(2\pi f_i n+\theta_i)+w(n)$$

(A_i,f_i,θ_i)为第i个谐波信号的(幅度,频率,相位),$w(n)$是加性白噪声.

参考三角函数的和、差角公式

$$\sin(\alpha+\beta+\gamma)=2\sin(\alpha+\gamma)\cos\beta-\sin(\alpha-\beta+\gamma)$$

可知该过程服从如下ARMA模型(推导过程略.)

$$x(n)+\sum_{i=1}^{2p}a_i x(n-i)=w(n)+\sum_{i=1}^{2p}a_i w(n-i)$$

令

$$R_x(k)=\frac{1}{M+2p-k+1}\sum_{i=1}^{M+2p-k+1}x(i)x(i+k)$$

则，如下差分方程(修正Yule-Walker方程)成立

$$\sum_{i=1}^{2p} a_i R_x(k-i) = R_x(k), \forall k > 2p$$

令

$$R = \begin{pmatrix} R_x(2p) & R_x(2p-1) & \cdots & R_x(1) \\ R_x(2p+1) & R_x(2p) & \cdots & R_x(2) \\ \vdots & \vdots & \cdots & \vdots \\ R_x(2p+M) & R_x(2p+M-1) & \cdots & R_x(M+1) \end{pmatrix}_{(M+1)\times 2p}$$

$$a = (a_1, a_2, \cdots, a_{2p})^T_{1\times 2p}$$

$$r = (R_x(2p+1), R_x(2p+2), \cdots, R_x(2p+M+1))^T_{1\times(M+1)}$$

则修正Yule-Walker方程的矩阵形式为

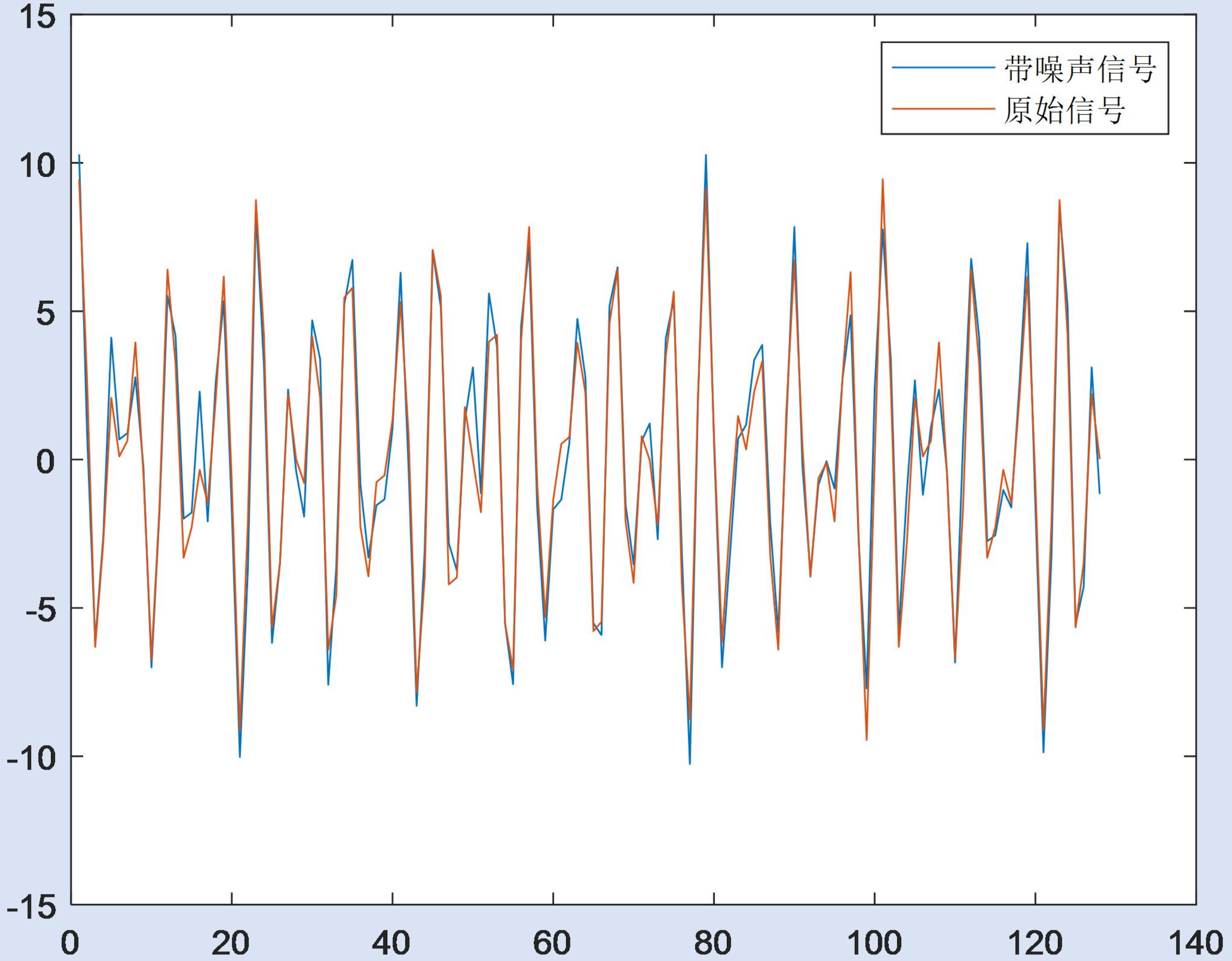
$$Ra = -r$$

适当选择M>>p，至少做M+2p+1次数据实验，求以上方程的最小二乘解，可得向量a. 根据a，可求(A_i, f_i, θ_i) .

f_i为$P(z) = 1 + \sum_{i=1}^{2p} a_i z^{-i}$的根的倒数（取其中的非负值).

```
clear all
p=2;
n=128;
A1=sqrt(18);
A2=sqrt(18);
A3=sqrt(2);
f1=0.18;
f2=0.27;
f3=0.23;
for ii=1:n
    x(ii)=A1*sin(2*pi*f1*ii)+...
A2*sin(2*pi*f2*ii)+A3*sin(2*pi*f3*ii);
    xx(ii)=x(ii)+normrnd(0,1,1,1);
end
nn=1:n;
plot(nn,[xx;x]);
 legend('带噪声信号','原始信号');
 M=n-2*p-1;
 Rx=zeros(1,n);
for ii=1:(2*p+M)
    for jj=1:n-ii
        Rx(ii)=Rx(ii)+xx(jj)*xx(jj+ii);
    end
    Rx(ii)=Rx(ii)/(n-ii);
end
for ii=1:(1+M)
    for jj=1:2*p
        R(ii,jj)=Rx(2*p+ii-1-jj+1);
    end
end
r=zeros(2*p,1);
for ii=1:(1+M)
    r(ii)=-Rx(2*p+ii);
end
a=inv(R'*R)*R'*r
P=zeros(2*p+1,1);
for ii=1:2*p
    P(ii)=a(2*p-ii+1);
end
P(2*p+1)=1;
f=roots( P)
z=zeros(p,1);
ff=zeros(p,1);
for ii=1:p
    z(ii)=1/f(2*ii-1);
end
for ii=1:p
    ff(ii)=1/(2*pi)*atan(imag(z(ii))/real(z(ii)));
end
ff
```

数据实验：请补齐总体最小二乘解的代码



a =

-0.5877
1.7752
-0.5855
0.9892

f =

0.4236 + 0.9094i
0.4236 - 0.9094i
-0.1277 + 0.9941i
-0.1277 - 0.9941i

ff =

-0.1806
0.1806
0.2297
-0.2297

数据实验

实验1：1月15日

1.编程验证Kronecker积与Khatri-Rao积的乘积的关系

$$(A\otimes B)(C*D)=AC*BD$$

2.编写程序，求矩阵方程 $AX+BXC+XD=E$

3.编程实现vec()的逆运算

4.编写程序，求一组数据的多项式**最小二乘拟合，并估算其精度**

实验2：1月15日

5.随机生成10000组数据，对其进行主成分分析

6.修改案例3中的程序，对图像进行次成分分析

实验3：1月22日

7. 请编程解决习题6.19所述问题(总体最小二乘法方法不限)

8.请补齐本PPT90页程序中的总体最小二乘法方法部分

作业:

6.4
6.5
6.7
6.12
6.13