系统辨识三要素：
数据、模型类和准则

最小二乘法辨识

传递函数 参数模型

↑

响应曲线 非参数模型

古典辨识： 脉冲响应法，阶跃响应法，频域响应法，

近代辨识： 最小二乘法、极大似然法

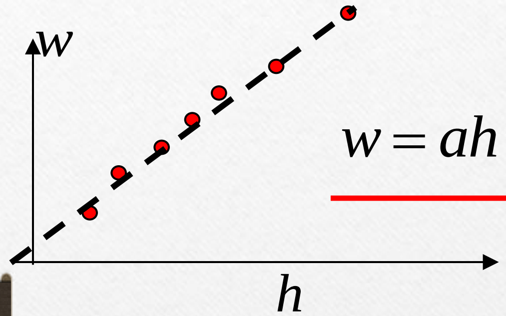
辨识对象： 以单输入单输出系统差分方程为模型

微分方程的离散化

辨识内容： 系统模型参数

例: 体重和身高 (假设: 线性)

最小二乘问题(Least-squares problems)



$$w = ah + b$$

$$\boxed{\begin{aligned} ah_1 + b &= w_1 \\ ah_2 + b &= w_2 \\ &\vdots \\ ah_k + b &= w_k \end{aligned}}$$

Matrix form

$$\begin{bmatrix} h_1 & 1 \\ h_2 & 1 \\ \vdots & \vdots \\ h_k & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{bmatrix}$$

$$Ax = b$$

参数 $x = \begin{bmatrix} a \\ b \end{bmatrix}$

LS: $\text{minimize} \quad \|Ax - b\|_2^2$

- 解析形式解: $x^\star = (A^T A)^{-1} A^T b$

基本的最小二乘估计

限定模型的种类

解决问题： 在模型阶次n已知的情况下，根据系统的输入输出数据，估计出系统差分方程的各项系数。

1.基于输入/输出数据的系统模型描述

$$\begin{Bmatrix} u(0), y(0), u(1), y(1), \ldots\ldots u(n), y(n), \\ u(n+1), y(n+1) \ldots u(n+N), y(n+N) \ldots \end{Bmatrix}$$

n+N个点

SISO系统的差分方程为

模型. 给定输入和输出序列, 待辨识的参数 a_i, b_i

$$\begin{cases} x(k)+a_1x(k-1)+\cdots+a_nx(k-n)=b_0u(k)+\cdots+b_mu(k-m) \\ y(k)=x(k)+v(k) \end{cases}$$

式中，$x(k)$为理论输出值，$y(k)$为实际观测值，$v(k)$为观测噪声。则有：

系统内部状态, 无法观测　　输出数据

$$x(k)=y(k)-v(k)$$

将$x(k)$代入上式，可得输入输出数据方程为： 消去了中间变量x(k)

$$y(k)+a_1y(k-1)+\cdots+a_ny(k-n)=b_0u(k)+b_1u(k-1)+\cdots+b_nu(k-n)+\xi(k)$$

$$\xi(k)=v(k)+\sum_{i=1}^{n}a_iv(k-i)$$

则当前输出为：

$$y(k)=-a_1y(k-1)-\cdots-a_ny(k-n)+b_0u(k)+b_1u(k-1)+\cdots+b_nu(k-n)+\xi(k)$$

设观测数据有(n+$\overset{\downarrow\text{任意一个大数}}{N}$)个，令k分别等于n+1,⋯,n+N,则有：

↑系统阶次　　为了让差分方程启动　　共N个方程

$$\begin{aligned}
&k=n+1 \quad y(n+1)=-a_1y(n)-\cdots-a_ny(1)+b_0u(n+1)+\cdots+b_nu(1)+\xi(n+1)\\
&k=n+2 \quad y(n+2)=-a_1y(n+1)-\cdots-a_ny(2)+b_0u(n+2)+\cdots+b_nu(2)+\xi(n+2)\\
&\qquad\vdots\\
&\qquad y(n+N)=-a_1y(n+N-1)-\cdots-a_ny(N)+b_0u(n+N)+\cdots+b_nu(N)+\xi(n+N)
\end{aligned}$$

上式写成向量形式为：

$$\begin{bmatrix} y(n+1) \\ y(n+2) \\ \vdots \\ y(n+N) \end{bmatrix} = \begin{bmatrix} -y(n) & \cdots & -y(1) & u(n+1) & \cdots & u(1) \\ -y(n+1) & \cdots & -y(2) & u(n+2) & \cdots & u(2) \\ \vdots & & \vdots & \vdots & & \vdots \\ -y(n+N-1) & \cdots & -y(N) & u(n+N) & \cdots & u(N) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \\ b_o \\ \vdots \\ b_n \end{bmatrix} + \begin{bmatrix} \xi(n+1) \\ \xi(n+2) \\ \vdots \\ \xi(n+N) \end{bmatrix}$$

记为：

$$\mathbf{Y}_{N\times 1} = \mathbf{\Phi}_{N\times(2n+1)}\,\mathbf{\theta}_{(2n+1)\times 1} + \boldsymbol{\xi}_{N\times 1} \qquad \mathrm{N}$$

$\mathbf{Y}_{N\times 1}$ → **输出向量**

$\mathbf{\Phi}_{N\times(2n+1)}$ → **测量矩阵**

$\mathbf{\theta}_{(2n+1)\times 1}$ → **参数矩阵**

$\boldsymbol{\xi}_{N\times 1}$ → **噪声矩阵**

N → **数据长度**

即

$$\mathbf{Y} = \mathbf{\Phi}\,\boldsymbol{\theta} + \xi \qquad (3.1)$$

噪声矩阵

若N=(2n+1)且ξ=0，则上式中的$\mathbf{\Phi}$阵为(2n+1)×(2n+1)的方阵。由此，可解得θ的唯一解为：

$$\boldsymbol{\theta} = \mathbf{\Phi}^{-1}Y$$

而在实际工程中，ξ肯定不等于0，且N>>(2n+1)，即方程个数远大于未知数，故而上述θ的解不成立。

当前任务： 在存在噪声ξ和数据长度N>>(2n+1)的情况下，如何进行参数θ的估计。

2.基本的最小二乘法(LS)

辨识准则：残差平方和最小。

(1)残差e

$\mathbf{e}=\mathbf{Y}-\hat{\mathbf{Y}}$，$\hat{\mathbf{Y}}$为模型的计算值，即 $\hat{\mathbf{Y}}=\Phi\hat{\boldsymbol{\theta}}$

(2)指标函数J　$\mathbf{e}=\mathbf{Y}-\Phi\hat{\boldsymbol{\theta}}$

$$J=\sum_{k=n+1}^{n+N}e^2(k)=\mathbf{e}\mathbf{e}^T=(\mathbf{Y}-\Phi\hat{\boldsymbol{\theta}})^T(\mathbf{Y}-\Phi\hat{\boldsymbol{\theta}})$$

故而，最小二乘法辨识就是使J最小的参数估计方法。

即有：

$$\hat{\theta}=\min_{\theta}\ J$$

下面我们推导θ估计值的计算方法。

J取得最小值，也即J为极值，则有：

$$\frac{\partial J}{\partial \hat{\boldsymbol{\theta}}} = 0$$

$$\Rightarrow \frac{\partial[(\mathbf{Y}-\boldsymbol{\Phi}\hat{\boldsymbol{\theta}})^T(\mathbf{Y}-\boldsymbol{\Phi}\hat{\boldsymbol{\theta}})]}{\partial \hat{\boldsymbol{\theta}}} = 0$$

$$\Rightarrow -2\boldsymbol{\Phi}^{\mathrm{T}}(\mathbf{Y}-\boldsymbol{\Phi}\hat{\boldsymbol{\theta}}) = 0$$

$$\Rightarrow \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\hat{\boldsymbol{\theta}} = \boldsymbol{\Phi}^{\mathrm{T}}\mathbf{Y}$$

其中，$(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})$ 为(2n+1)×(2n+1)的方阵。 $\boldsymbol{\Phi}_{N\times(2n+1)}$ $\boldsymbol{\theta}_{(2n+1)\times 1}$

若其逆阵存在，则：

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{Y} \qquad (3.2)$$

上式即为最小二乘法的参数估计结果。

$$\mathbf{Y} = \boldsymbol{\Phi}\,\boldsymbol{\theta} + \boldsymbol{\xi}$$

作业 1：已知某一单输入单输出线性系统的差分方程形式为

$$y(k) = -a_1 y(k-1) + b_0 u(k) + b_1 u(k-1) + \xi(k)$$

但其参数 a_1 , b_0 ， b_1为未知数，且$\xi(k)$为不相关的随机序列。经过辨识试验，测得5组输入输出数据为

$$u(1) = 2.1 \qquad u(2) = -2.7 \qquad u(3) = 0.8$$

$$u(4) = 1.5 \qquad u(5) = -2.1$$

$$y(1) = 0.3 \qquad y(2) = 0.5 \qquad y(3) = -0.2$$

$$y(4) = 0.6 \qquad y(5) = 0.83$$

试求出其最优参数估计。

解：由系统方程 $y(k) = -a_1 y(k-1) + b_0 u(k) + b_1 u(k-1) + \xi(k)$，

设：

$$Y = \begin{pmatrix} y(2) \\ y(3) \\ y(4) \\ y(5) \end{pmatrix} \qquad \phi = \begin{pmatrix} -y(1) & u(2) & u(1) \\ -y(2) & u(3) & u(2) \\ -y(3) & u(4) & u(3) \\ -y(4) & u(5) & u(4) \end{pmatrix} \qquad \theta = \begin{pmatrix} a_1 \\ b_0 \\ b_1 \end{pmatrix}$$

满足：

$$Y = \phi\theta + \xi$$

根据最小二乘法，其参数估计的结果为：$\hat{\theta} = (\phi^T \phi)^{-1} \phi^T Y$

代码：

```
yk = [0.3 0.5 -0.2 0.6 0.83];
u1 = [2.1 -2.7 0.8 1.5 -2.1];

Y = [yk(2); yk(3); yk(4); yk(5)];
X = [-yk(1), u1(2), u1(1);
     -yk(2), u1(3), u1(2);
     -yk(3), u1(4), u1(3);
     -yk(4), u1(5), u1(4)];
% 计算参数估计结果 theta_hat = inv(X' * X) * X' * Y
theta_hat = (X' * X) \ (X' * Y);
disp(theta_hat);
```

结果：a_1 = -1.4862, b_0 = 0.3456, b_1 = 0.4569

理论上，偏导为0只能说明J取得极值

使J为极小值的条件为：$\dfrac{\partial^2 J}{\partial \hat{\boldsymbol{\theta}}^2} > 0$

$$\frac{\partial^2 J}{\partial \hat{\boldsymbol{\theta}}^2} = 2\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} \Rightarrow \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} > 0$$

3. 最小二乘法对输入信号的要求

主要讨论 $\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} > 0$对输入信号u(k)的要求。

$$\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} = \left[\begin{array}{cccc} -y(n) & -y(n+1) & \cdots & -y(n+N-1) \\ \vdots & \vdots & & \vdots \\ -y(1) & -y(2) & \cdots & -y(N) \\ \hdashline u(n+1) & u(n+2) & \cdots & u(n+N) \\ \vdots & \vdots & & \vdots \\ u(1) & u(2) & \cdots & u(N) \end{array}\right] \times \left[\begin{array}{ccc:ccc} -y(n) & \cdots & -y(1) & u(n+1) & \cdots & u(1) \\ -y(n+1) & \cdots & -y(2) & u(n+2) & \cdots & u(2) \\ \vdots & & \vdots & \vdots & & \vdots \\ -y(n+N-1) & \cdots & -y(N) & u(n+N) & \cdots & u(N) \end{array}\right]$$

$$\begin{bmatrix} Y \\ U \end{bmatrix}_{2\times1} \times \begin{bmatrix} Y & U \end{bmatrix}_{1\times2} = \begin{bmatrix} & \end{bmatrix}_{2\times2}$$

$$= \begin{bmatrix} \boldsymbol{\Phi}_{yy} & \boldsymbol{\Phi}_{yu} \\ \boldsymbol{\Phi}_{uy} & \boldsymbol{\Phi}_{uu} \end{bmatrix}$$

其中，
$$\boldsymbol{\Phi}_{yy} = \begin{bmatrix} \sum\limits_{k=n}^{n+N-1} y^2(k) & \sum\limits_{k=n}^{n+N-1} y(k)y(k-1) & \cdots & \sum\limits_{k=n}^{n+N-1} y(k)y(k-n+1) \\ \sum\limits_{k=n}^{n+N-1} y(k-1)y(k) & \sum\limits_{k=n}^{n+N-1} y^2(k-1) & \cdots & \sum\limits_{k=n}^{n+N-1} y(k-1)y(k-n+1) \\ \vdots & \vdots & \cdots & \vdots \\ \sum\limits_{k=n}^{n+N-1} y(k-n+1)y(k) & \sum\limits_{k=n}^{n+N-1} y(k-n+1)y(k-1) & \cdots & \sum\limits_{k=n}^{n+N-1} y^2(k-n+1) \end{bmatrix}$$

$$\mathbf{\Phi}_{yu}=\begin{bmatrix} -\sum_{k=n}^{n+N-1} y(k)u(k+1) & -\sum_{k=n}^{n+N-1} y(k)u(k) & \cdots & -\sum_{k=n}^{n+N-1} y(k)u(k-n+1) \\ -\sum_{k=n}^{n+N-1} y(k-1)u(k+1) & -\sum_{k=n}^{n+N-1} y(k-1)u(k) & \cdots & -\sum_{k=n}^{n+N-1} y(k-1)u(k-n+1) \\ \vdots & \vdots & & \vdots \\ -\sum_{k=n}^{n+N-1} y(k-n+1)u(k+1) & -\sum_{k=n}^{n+N-1} y(k-n+1)u(k) & \cdots & -\sum_{k=n}^{n+N-1} y(k-n+1)u(k-n+1) \end{bmatrix}$$

$$\mathbf{\Phi}_{uy}=\begin{bmatrix} -\sum_{k=n}^{n+N-1} y(k)u(k+1) & -\sum_{k=n}^{n+N-1} y(k-1)u(k+1) & \cdots & -\sum_{k=n}^{n+N-1} y(k-n+1)u(k+1) \\ -\sum_{k=n}^{n+N-1} y(k)u(k) & -y(k-1)u(k) & \cdots & -\sum_{k=n}^{n+N-1} y(k-n+1)u(k) \\ \vdots & \vdots & & \vdots \\ -\sum_{k=n}^{n+N-1} y(k)u(k-n+1) & -\sum_{k=n}^{n+N-1} y(k-1)u(k-n+1) & \cdots & -\sum_{k=n}^{n+N-1} y(k-n+1)u(k-n+1) \end{bmatrix}$$

$$\mathbf{\Phi}_{uu}=\begin{bmatrix} \sum_{k=n}^{n+N-1} u^2(k+1) & \sum_{k=n}^{n+N-1} u(k+1)u(k) & \cdots & \sum_{k=n}^{n+N-1} u(k+1)u(k-n+1) \\ \sum_{k=n}^{n+N-1} u(k)u(k+1) & \sum_{k=n}^{n+N-1} u^2(k) & \cdots & \sum_{k=n}^{n+N-1} u(k)u(k-n+1) \\ \vdots & \vdots & & \vdots \\ \sum_{k=n}^{n+N-1} u(k-n+1)u(k+1) & \sum_{k=n}^{n+N-1} u(k-n+1)u(k) & \cdots & \sum_{k=n}^{n+N-1} u^2(k-n+1) \end{bmatrix}$$

则当N→∞时，有：

自相关函数

$$\lim_{N\to\infty}\frac{1}{N}\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}=\begin{bmatrix}\mathbf{R}_y & \mathbf{R}_{yu}\\ \mathbf{R}_{uy} & \mathbf{R}_u\end{bmatrix}=\mathbf{R}$$

互相关函数

$$\mathbf{R}_y=\begin{bmatrix}R_y(0) & R_y(1) & \cdots & R_y(n-1)\\ R_y(1) & R_y(0) & \cdots & R_y(n-2)\\ \vdots & \vdots & & \vdots\\ R_y(n-1) & R_y(n-2) & \cdots & R_y(0)\end{bmatrix}$$

$$\mathbf{R}_{uy}=\begin{bmatrix}-R_{uy}(-1) & -R_{uy}(0) & \cdots & R_{uy}(n-1)\\ -R_{uy}(-2) & -R_{uy}(-1) & \cdots & R_{uy}(n-2)\\ \vdots & \vdots & & \vdots\\ -R_{uy}(-n) & -R_{uy}(-n+1) & \cdots & R_{uy}(0)\end{bmatrix}=\mathbf{R}_{yu}^{T}$$

$$\mathbf{R}_u=\begin{bmatrix}R_u(0) & R_u(1) & \cdots & R_u(n)\\ R_u(1) & R_u(0) & \cdots & R_u(n-1)\\ \vdots & \vdots & & \vdots\\ R_u(n) & R_u(n-1) & \cdots & R_u(0)\end{bmatrix}$$

$$\mathbf{R}_u = \begin{bmatrix} R_u(0) & R_u(1) & \cdots & R_u(n) \\ R_u(1) & R_u(0) & \cdots & R_u(n-1) \\ \vdots & \vdots & & \vdots \\ R_u(n) & R_u(n-1) & \cdots & R_u(0) \end{bmatrix}$$

于是有：

J取得极小值→ $\Phi^T\Phi$正定→R正定→R_u正定。

因此：

J取得极小值的必要条件为R_u为正定阵。

这就是最小二乘法对输入信号的要求。

定义：

如果序列{u(k)}的(n+1)阶方阵Ru是正定的，则称序列{u(k)}为(n+1)阶持续激励信号。

因此，最小二乘法对输入信号的要求为：

{u(k)}为(n+1)阶持续激励信号

考查

$$\mathbf{R}_u=\begin{bmatrix} R_u(0) & R_u(1) & \cdots & R_u(n) \\ R_u(1) & R_u(0) & \cdots & R_u(n-1) \\ \vdots & \vdots & & \vdots \\ R_u(n) & R_u(n-1) & \cdots & R_u(0) \end{bmatrix}$$

若Ru为强对角线占优矩阵，则Ru正定。

哪些输入信号{u(k)}的Ru是强对角线占优矩阵？以下输入信号均能满足Ru正定的要求：

总结：白噪声 —— 平稳随机过程 $w(t)$

- 均值为零，自相关函数为脉冲函数，功率谱密度为非零常数

$\mu_w=E\{w(t)\}=0$ $R_w(\tau)=\sigma^2\delta(\tau)$ $S_w(w)=\sigma^2$ $(-\infty<w<+\infty)$

⇒对角线上无穷大，其它元素均为0

(1)白噪声序列；

(2)伪随机二位式噪声序列； 即M序列

(3)有色噪声随机信号序列。 扫频信号

工程上常用“伪随机二位式噪声序列”、“有色噪声随机信号序列”作为输入信号。

4. 最小二乘估计的概率性质

$$\begin{cases} \mathbf{Y} = \mathbf{\Phi}\,\boldsymbol{\theta} + \boldsymbol{\xi} \\ \hat{\boldsymbol{\theta}} = (\mathbf{\Phi}^{T}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{T}\mathbf{Y} \end{cases}$$

最小二乘估计的概率性质主要有以下四方面：

(1)估计的无偏性；

LS无偏估计的充分条件为：

{ξ(k)}为零均值不相关随机序列，且与{u(k)}无关

(2)估计的一致性；

一致性估计的充分条件为：

{ξ(k)}为零均值不相关随机序列，且与{u(k)}无关

eg 白噪声序列、M序列

(3)估计的有效性；

(4)估计的渐进正态性。

我们主要讨论前两项：无偏性和一致性。

即使在v(k)为白噪声的条件下，{ξ(k)}为相关随机序列。故而**基本最小二乘估计是有偏估计**，必须对基本最小二乘法进行改进。

最小二乘法的修正算法：广义最小二乘和辅助变量方法

(1)估计的无偏性

无偏性估计的定义：

若 $E\{\hat{\boldsymbol{\theta}}\} = E\{\boldsymbol{\theta}\} = \boldsymbol{\theta}$，则称 $\hat{\boldsymbol{\theta}}$是参数θ的无偏估计。

下面讨论无偏估计的条件。

$$\begin{cases} \mathbf{Y} = \boldsymbol{\Phi}\,\boldsymbol{\theta} + \boldsymbol{\xi} \\ \hat{\boldsymbol{\theta}} = (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{Y} \end{cases} \tag{3.3}$$

$$E[\hat{\boldsymbol{\theta}}] = E[(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{Y}] = E[(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{\Phi}\,\boldsymbol{\theta} + \boldsymbol{\xi})]$$

$$\Rightarrow E[\hat{\boldsymbol{\theta}}] = E[\boldsymbol{\theta} + (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\,\boldsymbol{\xi}] = \boldsymbol{\theta} + E[(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\,\boldsymbol{\xi}]$$

LS无偏估计的充要条件为：

$$E[(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\,\boldsymbol{\xi}] = 0$$

下面讨论无偏估计的充分条件。

$$y(k)=-a_1y(k-1)-\cdots-a_ny(k-n)+b_0u(k)+b_1u(k-1)+\cdots+b_nu(k-n)+\xi(k)$$

由上式可知：

y(k)只与ξ(k),ξ(k-1),ξ(k-2)···相关，而与ξ(k+1),ξ(k+2),ξ(k+3) ···不相关。

考查充要条件 $E[(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathrm{T}}\xi]=0$

$$\mathbf{\Phi}^{T}\xi=\begin{bmatrix} -y(n) & -y(n+1) & \cdots & -y(n+N-1) \\ \vdots & \vdots & & \vdots \\ -y(1) & -y(2) & \cdots & -y(N) \\ u(n+1) & u(n+2) & \cdots & u(n+N) \\ \vdots & \vdots & & \vdots \\ u(1) & u(2) & \cdots & u(N) \end{bmatrix}\begin{bmatrix} \xi(n+1) \\ \xi(n+2) \\ \vdots \\ \xi(n+N) \end{bmatrix}$$

在每个时刻的数值是独立的　相关. eg. 阶跃信号、正弦信号

若{ξ(k)}为零均值不相关随机序列，且与{u(k)}无关。

则由上式可知，Φ^{T}与ξ不相关。

则有：

$E(XY) \xrightarrow{X,Y不相关} E(X)E(Y)$

$$E[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta} + E[(\boldsymbol{\Phi}^{\mathbf{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathbf{T}}\ \xi]$$

$$= \boldsymbol{\theta} + E[(\boldsymbol{\Phi}^{\mathbf{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathbf{T}}]E[\ \xi]$$

$$= \boldsymbol{\theta} + 0$$

$$= \boldsymbol{\theta}$$

可见，在上述条件下我们得到了参数θ的无偏估计。

LS无偏估计的充分条件为：

{ξ(k)}为零均值不相关随机序列，且与{u(k)}无关。

(2)一致性估计

一致性估计的定义：

若参数估计值以概率1收敛于真值θ，则称估计值具有一致性。

或采用下述定义：

若 $\lim\limits_{N\to\infty} Var\tilde{\boldsymbol{\theta}} = 0$，则称 $\hat{\boldsymbol{\theta}}$ 是参数θ的一致性估计。

式中，$Var[\tilde{\boldsymbol{\theta}}]$ 为估计误差 $\tilde{\boldsymbol{\theta}}$ 的方差。

下面讨论一致性估计的充分条件。

$$\begin{aligned}
\tilde{\boldsymbol{\theta}} &= \boldsymbol{\theta} - \hat{\boldsymbol{\theta}} = \boldsymbol{\theta} - (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\ \mathbf{Y} \\
&= \boldsymbol{\theta} - (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{\Phi}\ \boldsymbol{\theta} + \boldsymbol{\xi}) \\
&= -(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\xi}
\end{aligned}$$

估计误差的方差为：

$$Var\tilde{\boldsymbol{\theta}} = E[\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^T] = E[(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T(\xi\xi^T)\boldsymbol{\Phi}(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}]$$

同样，假设{ξ(k)}为零均值不相关随机序列，且与{u(k)}无关。则Φ^{T}与ξ不相关,且有：

$$E\left\{\xi\xi^T\right\} = \sigma^2\mathbf{I}_N$$

$$\begin{aligned}\Rightarrow Var\tilde{\boldsymbol{\theta}} &= E[(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\sigma^2\mathbf{I}_N\boldsymbol{\Phi}(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}] \\ &= E[\sigma^2(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\boldsymbol{\Phi}(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}] \\ &= \sigma^2 E[(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}]\end{aligned}$$

$$\Rightarrow \lim_{N\to\infty} Var\tilde{\boldsymbol{\theta}} = \lim_{N\to\infty}\sigma^2 E[(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}] = \lim_{N\to\infty}\frac{\sigma^2}{N}E[(\frac{1}{N}\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}] = \lim_{N\to\infty}\frac{\sigma^2}{N}\underset{\uparrow\text{自相关矩阵}}{\mathbf{R}^{-1}} = 0$$

一致性估计的充分条件为：

{ξ(k)}为零均值不相关随机序列，且与{u(k)}无关。

(3) 估计值的有效性

有效性的定义：若参数估计误差的方差达到最小值，则称该估计值是有效估计值。

LS有效估计值的充分条件：$\{\xi(k)\}$是零均值且服从正态分布的白噪声序列。

(4) 估计值的渐近正态性

渐近正态性的定义：若参数估计值服从正态分布，则称该估计值是渐近正态的。

LS渐近正态性的充分条件：$\{\xi(k)\}$是零均值且服从正态分布的白噪声序列。

(5) 基本最小二乘估计存在问题

$$\begin{cases} \mathbf{Y} = \mathbf{\Phi}\,\boldsymbol{\theta} + \boldsymbol{\xi} \\ \hat{\boldsymbol{\theta}} = (\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{Y} \end{cases}$$

无偏性和一致性估计的充分条件均为：{ξ(k)}为零均值不相关随机序列，且与{u(k)}无关。

考查 $\begin{cases} \xi(k) = v(k) + \sum_{i=1}^{n} a_i v(k-i) \\ \xi(k+1) = v(k+1) + \sum_{i=1}^{n} a_i v(k+1-i) \end{cases} \Rightarrow \xi(k),\ \xi(k+1)$ 相关

也就是说即使在v(k)为白噪声的条件下，{ξ(k)}为相关随机序列。故而**基本最小二乘估计是有偏估计**，必须对基本最小二乘法进行改进。

最小二乘法的修正算法：广义最小二乘和辅助变量方法

递推最小二乘法 在线辨识方法 要会推导过程和公式

解决问题： (n+N)组观测数据时的参数估计值已知，现在又得到了一组新的观测值(u(n+N+1),y(n+N+1))，如何采用最小二乘法进行在线估计新的估计值问题。

1.递推算法推导

假设已获取了数据长度为N的I/O数据，则由LS估计有：

$$
\begin{cases}
\mathbf{Y}_N = \mathbf{\Phi}_N \boldsymbol{\theta} + \boldsymbol{\xi}_N \\
\hat{\boldsymbol{\theta}}_N = (\mathbf{\Phi}_N{}^{\mathrm{T}} \mathbf{\Phi}_N)^{-1} \mathbf{\Phi}_N{}^{\mathrm{T}} \mathbf{Y}_N \\
\tilde{\boldsymbol{\theta}}_N = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_N = -(\mathbf{\Phi}_N^T \mathbf{\Phi}_N)^{-1} \mathbf{\Phi}_N^T \boldsymbol{\xi}_N \\
\mathrm{Var}\tilde{\boldsymbol{\theta}} = \sigma^2 (\mathbf{\Phi}_N^T \mathbf{\Phi}_N)^{-1}
\end{cases}
$$

记：$\mathbf{P}_N=(\mathbf{\Phi}_N^T\mathbf{\Phi}_N)^{-1}$，则 $\hat{\boldsymbol{\theta}}_N$ 可写成：

已有的估计值 $$\hat{\boldsymbol{\theta}}_N=\mathbf{P}_N\mathbf{\Phi}_N{}^{\mathbf{T}}\mathbf{Y}_N$$

现获得了一组新的I/O数据值：u(n+N+1)、y(n+N+1)。需推导出 $\hat{\boldsymbol{\theta}}_{N+1}$ 的计算公式，即

$$\hat{\boldsymbol{\theta}}_{N+1}=f(\hat{\boldsymbol{\theta}}_N,\quad u(n+N+1),\quad y(n+N+1))$$

(n+N+1)时刻的观测值y(n+N+1)可表示为：

$$\underbrace{y(n+N+1)}_{y_{N+1}}=\underbrace{\left[-y(n+N)\quad\cdots\quad -y(N+1)\quad u(n+N+1)\quad\cdots\quad u(N+1)\right]}_{\boldsymbol{\psi}_{N+1}^T}\boldsymbol{\theta}+\underbrace{\xi(n+N+1)}_{\xi_{N+1}}$$

则上式可写为：

第N+1个等式 $$y_{N+1}=\boldsymbol{\psi}_{N+1}^T\boldsymbol{\theta}+\xi_{N+1}$$

则输入输出方程可写成分块矩阵形式：

原有数据

新数据

$$\begin{bmatrix} \mathbf{Y}_N \\ \cdots\cdots \\ y_{N+1} \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_N \\ \cdots\cdots \\ \mathbf{\psi}_{N+1}^T \end{bmatrix} \mathbf{\theta} + \begin{bmatrix} \xi_N \\ \cdots\cdots \\ \xi_{N+1} \end{bmatrix} \quad (3.4)$$

共N+1行

则由最小二乘(LS)可得(n+N+1)时刻的参数估计值为：

$\downarrow \Phi_{N+1}$

$$\hat{\mathbf{\theta}}_{N+1} = \left\{ \begin{bmatrix} \mathbf{\Phi}_N \\ \cdots\cdots \\ \mathbf{\psi}_{N+1}^T \end{bmatrix}^T \begin{bmatrix} \mathbf{\Phi}_N \\ \cdots\cdots \\ \mathbf{\psi}_{N+1}^T \end{bmatrix} \right\}^{-1} \begin{bmatrix} \mathbf{\Phi}_N \\ \cdots\cdots \\ \mathbf{\psi}_{N+1}^T \end{bmatrix}^T \begin{bmatrix} \mathbf{Y}_N \\ \cdots\cdots \\ y_{N+1} \end{bmatrix}$$

$\Phi_{N+1} = \begin{bmatrix} \Phi_N \\ \cdots \\ \psi_{N+1}^T \end{bmatrix}$

$$= (\mathbf{\Phi}_N^T \mathbf{\Phi}_N + \mathbf{\psi}_{N+1} \mathbf{\psi}_{N+1}^T)^{-1} (\mathbf{\Phi}_N^T \mathbf{Y}_N + \mathbf{\psi}_{N+1} y_{N+1})$$

上式记为： 记： $\mathbf{P}_N = (\mathbf{\Phi}_N^T \mathbf{\Phi}_N)^{-1}$

$$\begin{cases} \hat{\mathbf{\theta}}_{N+1} = \mathbf{P}_{N+1} (\mathbf{\Phi}_N^T \mathbf{Y}_N + \mathbf{\psi}_{N+1} y_{N+1}) \\ \mathbf{P}_{N+1} = (\mathbf{P}_N^{-1} + \mathbf{\psi}_{N+1} \mathbf{\psi}_{N+1}^T)^{-1} \end{cases}$$

现在的主要任务是求解矩阵的逆

矩阵求逆引理：若相应矩阵的逆均存在，则有

$$(\mathbf{A}+\mathbf{B}\mathbf{C}^T)^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{I}+\mathbf{C}^T\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}^T\mathbf{A}^{-1}$$

对照 $\mathbf{P}_{N+1}=(\mathbf{P}_N^{-1}+\boldsymbol{\psi}_{N+1}\boldsymbol{\psi}_{N+1}^T)^{-1}$ ，令：

$$\mathbf{A}=\mathbf{P}_N^{-1},\quad \mathbf{B}=\boldsymbol{\psi}_{N+1},\quad \mathbf{C}^T=\boldsymbol{\psi}_{N+1}^T$$

则有

$$\mathbf{P}_{N+1}=\mathbf{P}_N-\mathbf{P}_N\boldsymbol{\psi}_{N+1}(I+\boldsymbol{\psi}_{N+1}^T\mathbf{P}_N\boldsymbol{\psi}_{N+1})^{-1}\boldsymbol{\psi}_{N+1}^T\mathbf{P}_N$$

实际上是数字!

考查矩阵 $\boldsymbol{\psi}_{N+1}^T\mathbf{P}_N\boldsymbol{\psi}_{N+1}$ 的维数：

$$[1\times(2\mathrm{n}+1)]\times[(2\mathrm{n}+1)\times(2\mathrm{n}+1)]\times[(2\mathrm{n}+1)\times 1]=1\times 1$$

可见，矩阵求逆运算转化为求标量的倒数运算。

将 $\mathbf{P}_{N+1}$ 代入 $\hat{\boldsymbol{\theta}}_{N+1}$ 的表达式 $\hat{\boldsymbol{\theta}}_{N+1}=\mathbf{P}_{N+1}(\boldsymbol{\Phi}_N^T\mathbf{Y}_N+\boldsymbol{\psi}_{N+1}y_{N+1})$，经整理有：

标量

$$\hat{\boldsymbol{\theta}}_{N+1}=\hat{\boldsymbol{\theta}}_N+\underbrace{\mathbf{P}_N\boldsymbol{\psi}_{N+1}(1+\boldsymbol{\psi}_{N+1}^T\mathbf{P}_N\boldsymbol{\psi}_{N+1})^{-1}(y_{N+1}-\boldsymbol{\psi}_{N+1}^T\hat{\boldsymbol{\theta}}_N)}_{\text{来新数据后的估计量的修正值}}\qquad(3.5)$$

第N+1个等式 $y_{N+1}=\boldsymbol{\psi}_{N+1}^T\boldsymbol{\theta}+\xi_{N+1}$

解 将 $P_{N+1}=P_N-P_N\psi_{N+1}(1+\psi_{N+1}^T P_N\psi_{N+1})^{-1}\psi_{N+1}^T P_N$

代入 $\hat{\theta}_{N+1}=P_{N+1}(\Phi_N^T Y_N+\psi_{N+1}y_{N+1})$.

有. $\hat{\theta}_{N+1}=\left[P_N-P_N\psi_{N+1}(1+\psi_{N+1}^T P_N\psi_{N+1})^{-1}\psi_{N+1}^T P_N\right](\Phi_N^T Y_N+\psi_{N+1}y_{N+1})$.

$$
\begin{aligned}
&=\underbrace{P_N\Phi_N^T Y_N}_{\hat{\theta}_N}+P_N\psi_{N+1}y_{N+1}-P_N\psi_{N+1}(1+\psi_{N+1}^T P_N\psi_{N+1})^{-1}\psi_{N+1}^T P_N\Phi_N^T Y_N\\
&\qquad -P_N\psi_{N+1}(1+\psi_{N+1}^T P_N\psi_{N+1})^{-1}\psi_{N+1}^T P_N\psi_{N+1}y_{N+1}\\
&=\hat{\theta}_N+P_N\psi_{N+1}y_{N+1}(1+\psi_{N+1}^T P_N\psi_{N+1})^{-1}(1+\psi_{N+1}^T P_N\psi_{N+1})\\
&\qquad -P_N\psi_{N+1}(1+\psi_{N+1}^T P_N\psi_{N+1})^{-1}\psi_{N+1}^T P_N\Phi_N^T Y_N\\
&\qquad -P_N\psi_{N+1}(1+\psi_{N+1}^T P_N\psi_{N+1})^{-1}\psi_{N+1}^T P_N\psi_{N+1}y_{N+1}\\
&=\hat{\theta}_N+P_N\psi_{N+1}(1+\psi_{N+1}^T P_N\psi_{N+1})^{-1}\left[y_{N+1}+\underline{y_{N+1}\psi_{N+1}^T P_N\psi_{N+1}}-\psi_{N+1}^T\underbrace{P_N\Phi_N^T Y_N}_{\hat{\theta}_N}-\underline{\psi_{N+1}^T P_N\psi_{N+1}y_{N+1}}\right]\\
&=\hat{\theta}_N+\underbrace{P_N\psi_{N+1}(1+\psi_{N+1}^T P_N\psi_{N+1})^{-1}(y_{N+1}-\psi_{N+1}^T\hat{\theta}_N)}_{\text{修正值}}
\end{aligned}
$$

由(3.5)式可得LS的递推算法:

$$\hat{\boldsymbol{\theta}}_{N+1}=\hat{\boldsymbol{\theta}}_N+\mathbf{P}_N\boldsymbol{\psi}_{N+1}(1+\boldsymbol{\psi}_{N+1}^T\mathbf{P}_N\boldsymbol{\psi}_{N+1})^{-1}(y_{N+1}-\boldsymbol{\psi}_{N+1}^T\hat{\boldsymbol{\theta}}_N)$$

↓增益系数矩阵

$$\Rightarrow\begin{cases}\hat{\boldsymbol{\theta}}_{N+1}=\hat{\boldsymbol{\theta}}_N+\mathbf{K}_{N+1}(y_{N+1}-\boldsymbol{\psi}_{N+1}^T\hat{\boldsymbol{\theta}}_N)\\ \mathbf{K}_{N+1}=\mathbf{P}_N\boldsymbol{\psi}_{N+1}(1+\boldsymbol{\psi}_{N+1}^T\mathbf{P}_N\boldsymbol{\psi}_{N+1})^{-1}\\ \mathbf{P}_{N+1}=\mathbf{P}_N-\mathbf{P}_N\boldsymbol{\psi}_{N+1}(1+\boldsymbol{\psi}_{N+1}^T\mathbf{P}_N\boldsymbol{\psi}_{N+1})^{-1}\boldsymbol{\psi}_{N+1}^T\mathbf{P}_N\end{cases}\qquad(3.6)$$

上述递推算法的运行需获取两个初值: $\hat{\boldsymbol{\theta}}_0$, $\mathbf{P}_0$

递推过程:

$$\hat{\boldsymbol{\theta}}_0,\mathbf{P}_0\rightarrow\mathbf{K}_1,\mathbf{P}_1,\hat{\boldsymbol{\theta}}_1\rightarrow\mathbf{K}_2,\mathbf{P}_2,\hat{\boldsymbol{\theta}}_2\rightarrow\mathbf{K}_3,\mathbf{P}_3,\hat{\boldsymbol{\theta}}_3\rightarrow\cdots\cdots$$

初值获取方法:

(1) 记录一组少量的I/O数据$(N_0>(2n+1))$, 采用LS方法估计出$\hat{\boldsymbol{\theta}}_0,\mathbf{P}_0$;

(2) 直接取$\hat{\boldsymbol{\theta}}_0=\mathbf{0}$, $\mathbf{P}_0=c^2\mathbf{I}_{(2n+1)\times(2n+1)}$, c为充分大的数。⇒无需求逆计算

证明：基本最小二乘离线算法与递推算法的结果完全一致.

$$\mathbf{P}_0 = c^2\mathbf{I}, \quad \hat{\boldsymbol{\theta}}_0 = 0$$

$$\Rightarrow \mathbf{P}_1 = (\mathbf{P}_0^{-1} + \boldsymbol{\psi}_1\boldsymbol{\psi}_1^T)^{-1}, \quad \hat{\boldsymbol{\theta}}_1 = \mathbf{P}_1\boldsymbol{\psi}_1 y_1$$

$$\Rightarrow \mathbf{P}_2 = (\mathbf{P}_1^{-1} + \boldsymbol{\psi}_2\boldsymbol{\psi}_2^T)^{-1} = (\mathbf{P}_0^{-1} + \boldsymbol{\psi}_1\boldsymbol{\psi}_1^T + \boldsymbol{\psi}_2\boldsymbol{\psi}_2^T)^{-1}, \quad \hat{\boldsymbol{\theta}}_2 = \mathbf{P}_2(\boldsymbol{\psi}_1 y_1 + \boldsymbol{\psi}_2 y_2)$$

$$\Rightarrow \mathbf{P}_3 = (\mathbf{P}_2^{-1} + \boldsymbol{\psi}_3\boldsymbol{\psi}_3^T)^{-1} = (\mathbf{P}_0^{-1} + \boldsymbol{\psi}_1\boldsymbol{\psi}_1^T + \boldsymbol{\psi}_2\boldsymbol{\psi}_2^T + \boldsymbol{\psi}_3\boldsymbol{\psi}_3^T)^{-1}, \quad \hat{\boldsymbol{\theta}}_3 = \mathbf{P}_3(\boldsymbol{\psi}_1 y_1 + \boldsymbol{\psi}_2 y_2 + \boldsymbol{\psi}_3 y_3)$$

$$\vdots$$

$$\Rightarrow \mathbf{P}_N = (\mathbf{P}_0^{-1} + \boldsymbol{\psi}_1\boldsymbol{\psi}_1^T + \cdots + \boldsymbol{\psi}_N\boldsymbol{\psi}_N^T)^{-1}, \quad \hat{\boldsymbol{\theta}}_N = \mathbf{P}_N(\boldsymbol{\psi}_1 y_1 + \cdots + \boldsymbol{\psi}_N y_N)$$

$$\lim_{c\to\infty}\mathbf{P}_N = \lim_{c\to\infty}\left(\frac{\boldsymbol{I}}{c^2} + \boldsymbol{\psi}_1\boldsymbol{\psi}_1^T + \boldsymbol{\psi}_2\boldsymbol{\psi}_2^T + \cdots + \boldsymbol{\psi}_N\boldsymbol{\psi}_N^T\right)^{-1}$$

$$= (\boldsymbol{\psi}_1\boldsymbol{\psi}_1^T + \boldsymbol{\psi}_2\boldsymbol{\psi}_2^T + \cdots + \boldsymbol{\psi}_N\boldsymbol{\psi}_N^T)^{-1} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}$$

$$\Rightarrow \hat{\boldsymbol{\theta}}_N = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\mathbf{Y}$$

$$\Phi = \begin{bmatrix} \psi_1^T \\ \psi_2^T \\ \vdots \\ \psi_N^T \end{bmatrix}$$

注:
$$\begin{cases} \hat{\boldsymbol{\theta}}_{N+1} = \mathbf{P}_{N+1}(\boldsymbol{\Phi}_N^T\mathbf{Y}_N + \boldsymbol{\psi}_{N+1}y_{N+1}) \\ \mathbf{P}_{N+1} = (\mathbf{P}_N^{-1} + \boldsymbol{\psi}_{N+1}\boldsymbol{\psi}_{N+1}^T)^{-1} \end{cases}$$

$$\boldsymbol{\Phi}^T = [\boldsymbol{\psi}_1 \quad \boldsymbol{\psi}_2 \quad \cdots \quad \boldsymbol{\psi}_N]$$