统计信号处理

第六章

最小二乘估计

清华大学电子工程系
李洪 副教授

2023.3

内容概要

- 一、最小二乘估计概念
- 二、线性最小二乘估计
- 三、序贯最小二乘估计
- 四、约束最小二乘估计
- 五、非线性最小二乘估计
- 六、小结

引言

- 最小方差无偏估计（MVU）
  - Cramer-Rao Lower Bound (CRLB)
  - （一般）线性模型
  - Sufficient statistics (s.s.)
- 最大似然估计（MLE）
- 应用前提：似然函数
- 实际应用中条件是否满足？
- 若无似然函数该当如何？
- 最小二乘估计（Least Squares）

一、最小二乘估计概念

已知信号模型$s[n;\theta]$，观测数据为$x[n]=s[n;\theta]+w[n]$，θ的**最小二乘估计**为

$$\hat{\theta}=\arg\min_{\theta}\left\{\sum_{n=0}^{N-1}\left(x[n]-s[n;\theta]\right)^2\right\}$$

其中$J(\theta)=\sum_{n=0}^{N-1}\left(x[n]-s[n;\theta]\right)^2$称为**最小二乘误差**

最小二乘估计(LSE)

MVU

$$\begin{cases}\min\left\{\operatorname{var}\left(\hat{\theta}\right)\right\}\\ s.t.\ E\left(\hat{\theta}\right)=\theta\end{cases}$$

MLE

$$\hat{\theta}=\arg\max_{\theta}\left\{p\left(\mathbf{x};\theta\right)\right\}$$

例：噪声中电平估计问题：

$$x[n]=A+w[n],\ n=0,1,\dots,N-1$$

待估计参数为信号幅度 A，$w[n]$ 为噪声。

$$J(A)=\sum_{n=0}^{N-1}\left(x[n]-A\right)^2$$

$$\left.\begin{aligned}\frac{\partial J(A)}{\partial A}&=-2\sum_{n=0}^{N-1}\left(x[n]-A\right)\\ \frac{\partial J(A)}{\partial A}&=0\end{aligned}\right\}\Rightarrow \hat{A}=\frac{1}{N}\sum_{n=0}^{N-1}x[n]$$

二、线性最小二乘估计

若信号 $\mathbf{s}=\left[s[0],s[1],s[2],\ldots,s[N-1]\right]^T$ 与未知参数 $\boldsymbol{\theta}=\left[\theta_1,\theta_2,\theta_3,\ldots,\theta_p\right]^T$ 呈线性关系

$$\mathbf{s}=\mathbf{H}\boldsymbol{\theta}$$

$\mathbf{H}$ 是 $N\times p\,(N>p)$ 的观测矩阵，秩为 p。观测数据模型为

$$\mathbf{x}=\mathbf{H}\boldsymbol{\theta}+\mathbf{w}$$

求 $\boldsymbol{\theta}$ 的最小二乘估计？

$$J(\boldsymbol{\theta})=\sum_{n=0}^{N-1}\left(x[n]-s[n;\boldsymbol{\theta}]\right)^2$$

$$=(\mathbf{x}-\mathbf{s})^T(\mathbf{x}-\mathbf{s})$$

$$=(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})$$

$$=\mathbf{x}^T\mathbf{x}-2\mathbf{x}^T\mathbf{H}\boldsymbol{\theta}+\boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}$$

$$\frac{\partial J(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=-2\mathbf{H}^T\mathbf{x}+2\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}$$

$$\hat{\boldsymbol{\theta}}=\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}$$

相应的最小LS误差为：

$$J_{\min}=\mathbf{x}^T\left(\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\theta}}\right)$$

● 几何解释

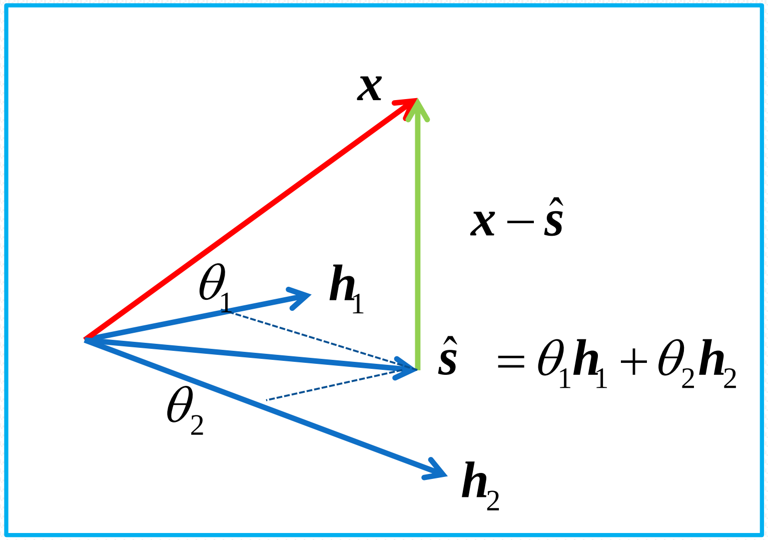
$$J(\boldsymbol{\theta})=(\boldsymbol{x}-\mathbf{H}\boldsymbol{\theta})^T(\boldsymbol{x}-\mathbf{H}\boldsymbol{\theta})$$
$$=\|\boldsymbol{x}-\mathbf{H}\boldsymbol{\theta}\|_2^2$$

观测矩阵按列向量可记为：

$$\mathbf{H}=[\boldsymbol{h}_1,\boldsymbol{h}_2,\ldots,\boldsymbol{h}_p]$$ ——信号矢量

$$\Rightarrow \boldsymbol{s}=\mathbf{H}\boldsymbol{\theta}=\sum_{i=1}^{p}\theta_i\boldsymbol{h}_i$$

$$\Rightarrow J(\boldsymbol{\theta})=\left\|\boldsymbol{x}-\sum_{i=1}^{p}\theta_i\boldsymbol{h}_i\right\|_2^2$$ ——观测数据到信号的距离的平方



欧式距离：$\|\xi\|=\sqrt{\sum_{i=1}^{N}\xi_i^2}$

$$\min\{J(\boldsymbol{\theta})\}\Rightarrow(\boldsymbol{x}-\hat{\boldsymbol{s}})\perp\boldsymbol{h}_i,\ i=1,2,\ldots,p\Rightarrow(\boldsymbol{x}-\hat{\boldsymbol{s}})^T\boldsymbol{h}_i=0\Rightarrow(\boldsymbol{x}-\hat{\boldsymbol{s}})^T\mathbf{H}=\mathbf{0}^T$$

$$\left.\begin{aligned}(\boldsymbol{x}-\hat{\boldsymbol{s}})^T\mathbf{H}&=\mathbf{0}^T\\ \hat{\boldsymbol{s}}&=\mathbf{H}\hat{\boldsymbol{\theta}}\end{aligned}\right\} \Rightarrow \left(\boldsymbol{x}-\mathbf{H}\hat{\boldsymbol{\theta}}\right)^T\mathbf{H}=\mathbf{0}^T \Rightarrow \boldsymbol{x}^T\mathbf{H}-\hat{\boldsymbol{\theta}}^T\mathbf{H}^T\mathbf{H}=\mathbf{0}^T$$

$$\Rightarrow \mathbf{H}^T\boldsymbol{x}-\mathbf{H}^T\mathbf{H}\hat{\boldsymbol{\theta}}=\mathbf{0} \Rightarrow \hat{\boldsymbol{\theta}}=\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\boldsymbol{x}$$ ——LSE估计量

其中

$\boldsymbol{\varepsilon}=\boldsymbol{x}-\mathbf{H}\hat{\boldsymbol{\theta}}$ ——称为**误差矢量**

$\boldsymbol{\varepsilon}^T\mathbf{H}=\mathbf{0}^T$ ——表示误差矢量与信号矢量是正交的！

——称为**正交原理**

若列矢量相互正交：

$$\boldsymbol{h}_i^T\boldsymbol{h}_j=\delta_{ij} \Rightarrow \left.\begin{aligned}&\mathbf{H}^T\mathbf{H}=\mathbf{I}\\ &\hat{\boldsymbol{\theta}}=\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\boldsymbol{x}\end{aligned}\right\} \Rightarrow \hat{\boldsymbol{\theta}}=\mathbf{H}^T\boldsymbol{x}$$

——合适的信号向量将大大简化LSE求解

加权最小二乘

$$J(\boldsymbol{\theta})=(\boldsymbol{x}-\mathbf{H}\boldsymbol{\theta})^T\,\mathbf{W}\,(\boldsymbol{x}-\mathbf{H}\boldsymbol{\theta})$$ （权系数 **W** 一般为对称矩阵）

$$=\boldsymbol{x}^T\mathbf{W}\boldsymbol{x}-\boldsymbol{x}^T\mathbf{W}\mathbf{H}\boldsymbol{\theta}-\boldsymbol{\theta}^T\mathbf{H}^T\mathbf{W}\boldsymbol{x}+\boldsymbol{\theta}^T\mathbf{H}^T\mathbf{W}\mathbf{H}\boldsymbol{\theta}$$

$$\left.\begin{aligned}\frac{\partial J(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}&=-\mathbf{H}^T\mathbf{W}^T\boldsymbol{x}-\mathbf{H}^T\mathbf{W}\boldsymbol{x}+2\mathbf{H}^T\mathbf{W}\mathbf{H}\boldsymbol{\theta}\\ \frac{\partial J(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}&=\mathbf{0}\end{aligned}\right\}\Rightarrow\hat{\boldsymbol{\theta}}=\left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\boldsymbol{x}$$

相应的最小LS误差为：$J_{\min}=\boldsymbol{x}^T\left(\mathbf{W}-\mathbf{W}\mathbf{H}\left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\right)\boldsymbol{x}$

若 $\mathbf{W}=\mathbf{C}^{-1}$?

$$\hat{\boldsymbol{\theta}}=\left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}^{-1}\boldsymbol{x}$$

例：不相关噪声中电平估计问题：

$$x[n]=A+w[n],\ n=0,1,\ldots,N-1$$

待估计参数为A，$w[n]$为不相关噪声且 $\mathrm{var}(w[n])=\sigma_n^2$,其加权LSE?

目标函数：$J(A)=\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}(x[n]-A)^2$

模型：$J(\boldsymbol{\theta})=(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T\mathbf{W}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})$ $\Rightarrow$ $\hat{\boldsymbol{\theta}}=(\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}\mathbf{x}$

参数：$\boldsymbol{\theta}=?\quad A$

$\mathbf{H}=?\quad [1,1,\ldots,1]^T$

$$\mathbf{W}=?\quad \begin{bmatrix} \frac{1}{\sigma_0^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_1^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_{N-1}^2} \end{bmatrix}$$

$$\hat{A}=\frac{\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}x[n]}{\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}}$$

三、序贯最小二乘估计

例：不相关噪声中电平估计问题：

$$x[n]=A+w[n],\ n=0,1,\ldots,N-1$$

待估计参数为A，$w[n]$为不相关噪声且$\mathrm{var}\left(w[n]\right)=\sigma_n^2$。

加权LSE：$J(A)=\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\left(x[n]-A\right)^2$

$$\hat{A}=\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}x[n]\Big/\left(\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\right)$$

$$\begin{cases}\hat{A}[N-1]=\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}x[n]\Big/\left(\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\right)\\ \hat{A}[N]=\sum_{n=0}^{N}\frac{1}{\sigma_n^2}x[n]\Big/\left(\sum_{n=0}^{N}\frac{1}{\sigma_n^2}\right)\end{cases}$$

能否序贯计算?

$$\hat{A}[N]=\left\{\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}x[n]+\frac{1}{\sigma_N^2}x[N]\right\}\Big/\sum_{n=0}^{N}\frac{1}{\sigma_n^2}$$

$$\hat{A}[N]=\left\{\left(\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\right)\hat{A}[N-1]+\frac{1}{\sigma_N^2}x[N]\right\}\Big/\left(\sum_{n=0}^{N}\frac{1}{\sigma_n^2}\right)$$

$$\hat{A}[N]=\left\{\left(\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}+\frac{1}{\sigma_N^2}-\frac{1}{\sigma_N^2}\right)\hat{A}[N-1]+\frac{1}{\sigma_N^2}x[N]\right\}\Big/\left(\sum_{n=0}^{N}\frac{1}{\sigma_n^2}\right)$$

$$\hat{A}[N]=\hat{A}[N-1]+\frac{\frac{1}{\sigma_N^2}}{\sum_{n=0}^{N}\frac{1}{\sigma_n^2}}\left(x[N]-\hat{A}[N-1]\right)$$

$K[N]$:增益因子

$$\hat{A}[N]=\hat{A}[N-1]+K[N]\left(x[N]-\hat{A}[N-1]\right)$$

- 估计量可以序贯计算

LSE的方差

$$\hat{A}[N-1]=\frac{\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}x[n]}{\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}}$$

$$\operatorname{var}\left(\hat{A}[N-1]\right)=\frac{\sum_{n=0}^{N-1}\left(\frac{1}{\sigma_n^2}\right)^2\operatorname{var}\left(x[n]\right)}{\left(\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\right)^2}=\frac{1}{\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}}$$

新的估计量：$\hat{A}[N]=\frac{\sum_{n=0}^{N}\frac{1}{\sigma_n^2}x[n]}{\sum_{n=0}^{N}\frac{1}{\sigma_n^2}}$

对应的方差为：

$$\operatorname{var}\left(\hat{A}[N]\right)=\frac{1}{\sum_{n=0}^{N}\frac{1}{\sigma_n^2}}$$

能否序贯计算?

$$\operatorname{var}\left(\hat{A}[N]\right)=\frac{\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\times\frac{1}{\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}}}{\sum_{n=0}^{N}\frac{1}{\sigma_n^2}}=\frac{\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\times\operatorname{var}\left(\hat{A}[N-1]\right)}{\sum_{n=0}^{N}\frac{1}{\sigma_n^2}}$$

$$=\frac{\left(\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}+\frac{1}{\sigma_N^2}-\frac{1}{\sigma_N^2}\right)\times\operatorname{var}\left(\hat{A}[N-1]\right)}{\sum_{n=0}^{N}\frac{1}{\sigma_n^2}}$$

$$=\left(1-\frac{\frac{1}{\sigma_N^2}}{\sum_{n=0}^{N}\frac{1}{\sigma_n^2}}\right)\operatorname{var}\left(\hat{A}[N-1]\right)$$

$$\operatorname{var}\left(\hat{A}[N]\right)=\left(1-K[N]\right)\operatorname{var}\left(\hat{A}[N-1]\right)$$

- 估计量的方差可以序贯计算
- 估计量的方差在不断减小

最小LS误差

$$J_{\min}=\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\left(x[n]-\hat{A}\right)^2 \Rightarrow \begin{cases} J_{\min}[N-1]=\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\left(x[n]-\hat{A}[N-1]\right)^2 \\ J_{\min}[N]=\sum_{n=0}^{N}\frac{1}{\sigma_n^2}\left(x[n]-\hat{A}[N]\right)^2 \end{cases}$$

能否序贯计算?

$$J_{\min}[N]=\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\left(x[n]-\hat{A}[N]\right)^2+\frac{1}{\sigma_N^2}\left(x[N]-\hat{A}[N]\right)^2$$

$$=\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\left(x[n]-\hat{A}[N-1]+\hat{A}[N-1]-\hat{A}[N]\right)^2+\frac{1}{\sigma_N^2}\left(x[N]-\hat{A}[N]\right)^2$$

$$=\sum_{n=0}^{N-1}\left\{\begin{array}{l}\underline{\frac{1}{\sigma_n^2}\left(x[n]-\hat{A}[N-1]\right)^2}+\underline{\frac{1}{\sigma_n^2}\left(\hat{A}[N-1]-\hat{A}[N]\right)^2}\\ +2\underline{\frac{1}{\sigma_n^2}\left(x[n]-\hat{A}[N-1]\right)}\left(\hat{A}[N-1]-\hat{A}[N]\right)\end{array}\right\}+\underline{\frac{1}{\sigma_N^2}\left(x[N]-\hat{A}[N]\right)^2}$$

$$\boxed{\hat{A}[N]=\hat{A}[N-1]+K[N]\left(x[N]-\hat{A}[N-1]\right)}$$

$$=J_{\min}[N-1]\ +\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}K^2[N]\left(x[N]-\hat{A}[N-1]\right)^2+\frac{1}{\sigma_N^2}\left(1-K[N]\right)^2\left(x[N]-\hat{A}[N-1]\right)^2$$

$$J_{\min}[N]=J_{\min}[N-1]+\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}K^2[N]\left(x[N]-\hat{A}[N-1]\right)^2+\frac{1}{\sigma_N^2}\left(1-K[N]\right)^2\left(x[N]-\hat{A}[N-1]\right)^2$$

$$\left\{\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}K^2[N]+\frac{1}{\sigma_N^2}\left(1-K[N]\right)^2\right\}\left(x[N]-\hat{A}[N-1]\right)^2$$

$$=\left\{\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\left(\frac{\frac{1}{\sigma_N^2}}{\sum_{n=0}^{N}\frac{1}{\sigma_n^2}}\right)^2+\frac{1}{\sigma_N^2}\left(1-\frac{\frac{1}{\sigma_N^2}}{\sum_{n=0}^{N}\frac{1}{\sigma_n^2}}\right)^2\right\}\left(x[N]-\hat{A}[N-1]\right)^2$$

$$=\frac{\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\left(\frac{1}{\sigma_N^2}\right)^2+\frac{1}{\sigma_N^2}\left(\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\right)^2}{\left(\sum_{n=0}^{N}\frac{1}{\sigma_n^2}\right)^2}\left(x[N]-\hat{A}[N-1]\right)^2$$

$$=\frac{\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\left(\frac{1}{\sigma_N^2}\right)\left\{\frac{1}{\sigma_N^2}+\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\right\}}{\left(\sum_{n=0}^{N}\frac{1}{\sigma_n^2}\right)^2}\left(x[N]-\hat{A}[N-1]\right)^2$$

$$=\frac{\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\left(\frac{1}{\sigma_N^2}\right)}{\left(\sum_{n=0}^{N}\frac{1}{\sigma_n^2}\right)}\left(x[N]-\hat{A}[N-1]\right)^2$$

$$=\frac{\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\left(\frac{1}{\sigma_N^2}\right)}{\left(\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}+\frac{1}{\sigma_N^2}\right)}\left(x[N]-\hat{A}[N-1]\right)^2$$

$$=\frac{1}{\left(\sigma_N^2+\frac{1}{\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}}\right)}\left(x[N]-\hat{A}[N-1]\right)^2$$

$$=\frac{1}{\left(\sigma_N^2+\operatorname{var}\left(\hat{A}[N-1]\right)\right)}\left(x[N]-\hat{A}[N-1]\right)^2$$

$$J_{\min}[N]=J_{\min}[N-1]+\frac{1}{\operatorname{var}\left(\hat{A}[N-1]\right)+\sigma_N^2}\left(x[N]-\hat{A}[N-1]\right)^2$$

$$=J_{\min}[N-1]+\frac{\sigma_N^2}{\operatorname{var}\left(\hat{A}[N-1]\right)+\sigma_N^2}\frac{\left(x[N]-\hat{A}[N-1]\right)^2}{\sigma_N^2}$$

$$K[N]=\frac{\frac{1}{\sigma_N^2}}{\sum_{n=0}^{N}\frac{1}{\sigma_n^2}}=\frac{\frac{1}{\sigma_N^2}}{\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}+\frac{1}{\sigma_N^2}}=\frac{\frac{1}{\sigma_N^2}}{\frac{1}{\operatorname{var}\left(\hat{A}[N-1]\right)}+\frac{1}{\sigma_N^2}}=\frac{\operatorname{var}\left(\hat{A}[N-1]\right)}{\operatorname{var}\left(\hat{A}[N-1]\right)+\sigma_N^2}$$

$$J_{\min}[N]=J_{\min}[N-1]+\left(1-K[N]\right)\frac{\left(x[N]-\hat{A}[N-1]\right)^2}{\sigma_N^2}$$

- 最小LS误差可以序贯计算
- 最小LS误差在不断增加

——因要拟合的数据在增加

例：不相关噪声中电平估计问题：

$$x[n]=A+w[n],\ n=0,1,\dots,N-1$$

待估计参数为A，$w[n]$为不相关噪声且$\operatorname{var}\left(w[n]\right)=\sigma_n^2$。

加权LSE：$J(A)=\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\left(x[n]-A\right)^2$

估计量更新：$\hat{A}[N]=\hat{A}[N-1]+\underline{K[N]}\left(x[N]-\hat{A}[N-1]\right)$

增益因子：$K[N]=\frac{\frac{1}{\sigma_N^2}}{\sum_{n=0}^{N}\frac{1}{\sigma_n^2}}=\frac{\frac{1}{\sigma_N^2}}{\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}+\frac{1}{\sigma_N^2}}=\frac{\frac{1}{\sigma_N^2}}{\frac{1}{\operatorname{var}\left(\hat{A}[N-1]\right)}+\frac{1}{\sigma_N^2}}$

新的信息量

总的信息量

LSE方差更新：$\operatorname{var}\left(\hat{A}[N]\right)=\left(1-\underline{K[N]}\right)\operatorname{var}\left(\hat{A}[N-1]\right)$

最小LS误差更新：$J_{\min}[N]=J_{\min}[N-1]+\left(1-\underline{K[N]}\right)\frac{\left(x[N]-\hat{A}[N-1]\right)^2}{\sigma_N^2}$

初始化：$\hat{A}[0]=x[0]\qquad \operatorname{var}\left(\hat{A}[0]\right)=\sigma_0^2$

● 推广至矢量参数情况

A. Giordano and Frank M. Hsu. *Least square estimation with application to digital signal processing*. Wiley, New York, 1985

信号模型： $\boldsymbol{x}=\mathbf{H}\boldsymbol{\theta}+\boldsymbol{w}$

$\boldsymbol{w}$ 为噪声，其协方差矩阵为 $\mathbf{C}_w$ 。

加权最小LS误差： $J(\boldsymbol{\theta})=(\boldsymbol{x}-\mathbf{H}\boldsymbol{\theta})^T\mathbf{C}_w^{-1}(\boldsymbol{x}-\mathbf{H}\boldsymbol{\theta})$

加权LSE： $\hat{\boldsymbol{\theta}}=\left(\mathbf{H}^T\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}_w^{-1}\boldsymbol{x}$

LSE的协方差阵： $\mathbf{C}_{\hat{\theta}}=\left(\mathbf{H}^T\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1}$

若 $\mathbf{C}_w$ 是**对角矩阵**，即噪声是不相关的，则可按如下方式序贯计算：

估计量更新： $\hat{\boldsymbol{\theta}}[n]=\hat{\boldsymbol{\theta}}[n-1]+\boldsymbol{k}[n]\left(x[n]-\boldsymbol{h}^T[n]\hat{\boldsymbol{\theta}}[n-1]\right)$

增益因子： $\boldsymbol{k}[n]=\dfrac{\mathbf{C}_{\hat{\theta}}[n-1]\boldsymbol{h}[n]}{\sigma_n^2+\boldsymbol{h}^T[n]\mathbf{C}_{\hat{\theta}}[n-1]\boldsymbol{h}[n]}$

$$\mathbf{H}[n]=\begin{bmatrix}\mathbf{H}[n-1]\\ \boldsymbol{h}^T[n]\end{bmatrix}$$

$$x[n]=\boldsymbol{h}^T[n]\boldsymbol{\theta}+w[n]$$

协方差更新： $\mathbf{C}_{\hat{\theta}}[n]=\left(\mathbf{I}-\boldsymbol{k}[n]\boldsymbol{h}^T[n]\right)\mathbf{C}_{\hat{\theta}}[n-1]$

最小LS误差更新： $J_{\min}[n]=J_{\min}[n-1]+\dfrac{\left(x[n]-\boldsymbol{h}^T[n]\hat{\boldsymbol{\theta}}[n-1]\right)^2}{\sigma_n^2+\boldsymbol{h}^T[n]\mathbf{C}_{\hat{\theta}}[n-1]\boldsymbol{h}[n]}$

四、约束最小二乘估计

信号模型：

$$\boldsymbol{x} = \mathbf{H}\boldsymbol{\theta} + \boldsymbol{w}$$

$\boldsymbol{w}$ 为噪声，其协方差矩阵为 $\mathbf{C}$。待估计参数需满足如下约束：

$$\mathbf{A}\boldsymbol{\theta} = \boldsymbol{b}$$

$$\begin{cases} \min\limits_{\theta} \left\{ (\boldsymbol{x} - \mathbf{H}\boldsymbol{\theta})^T (\boldsymbol{x} - \mathbf{H}\boldsymbol{\theta}) \right\} \\ s.t. \;\; \mathbf{A}\boldsymbol{\theta} = \boldsymbol{b} \end{cases}$$

$$\left. \begin{array}{r} J_c = (\boldsymbol{x} - \mathbf{H}\boldsymbol{\theta})^T (\boldsymbol{x} - \mathbf{H}\boldsymbol{\theta}) + \boldsymbol{\lambda}^T (\mathbf{A}\boldsymbol{\theta} - \boldsymbol{b}) \\ \dfrac{\partial J_c}{\partial \boldsymbol{\theta}} = \boldsymbol{0} \end{array} \right\}$$

$$\hat{\boldsymbol{\theta}}_c = \left(\mathbf{H}^T\mathbf{H}\right)^{-1} \mathbf{H}^T \boldsymbol{x} - \frac{1}{2}\left(\mathbf{H}^T\mathbf{H}\right)^{-1} \mathbf{A}^T \boldsymbol{\lambda}$$

$$\hat{\boldsymbol{\theta}}_c = \underline{\hat{\boldsymbol{\theta}}} - \left(\mathbf{H}^T\mathbf{H}\right)^{-1} \mathbf{A}^T \frac{\boldsymbol{\lambda}}{2}$$

约束条件： $\mathbf{A}\hat{\boldsymbol{\theta}}_c = \boldsymbol{b}$

$$\frac{\boldsymbol{\lambda}}{2} = \left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1} \left(\mathbf{A}\hat{\boldsymbol{\theta}} - \boldsymbol{b}\right)$$

$$\hat{\boldsymbol{\theta}}_c = \hat{\boldsymbol{\theta}} - \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T \left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1} \left(\mathbf{A}\hat{\boldsymbol{\theta}} - \boldsymbol{b}\right)$$

- 在无约束LSE的基础上加一“修正项”

无约束最小二乘解

例：信号模型：

$$x[n]=s[n]+w[n]$$

其中，信号为 $s[n]=\begin{cases}\theta_1, & n=0\\ \theta_2, & n=1\\ 0, & n=2\end{cases}$，观测数据为 $\{x[0],x[1],x[2]\}$。

若已知 $\theta_1=\theta_2$，对参数的估计为？

待估计参数：$\boldsymbol{\theta}=[\theta_1,\theta_2]^T$

第一种方法：无约束LSE

$$\left.\begin{aligned}&\hat{\boldsymbol{\theta}}=\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}\\ &\mathbf{H}=\begin{bmatrix}1&0\\0&1\\0&0\end{bmatrix}\end{aligned}\right\}\Rightarrow\hat{\boldsymbol{\theta}}=\begin{bmatrix}x[0]\\x[1]\end{bmatrix}$$

例：信号模型：

$$x[n]=s[n]+w[n]$$

其中，信号为 $s[n]=\begin{cases}\theta_1, & n=0\\ \theta_2, & n=1\\ 0, & n=2\end{cases}$，观测数据为 $\{x[0],x[1],x[2]\}$。

若已知 $\theta_1=\theta_2$，对参数的估计为？

待估计参数：$\boldsymbol{\theta}=[\theta_1,\theta_2]^T$

第二种方法：约束LSE

$$\hat{\boldsymbol{\theta}}_c=\hat{\boldsymbol{\theta}}-(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\left[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\right]^{-1}(\mathbf{A}\hat{\boldsymbol{\theta}}-\boldsymbol{b})$$

$$\hat{\boldsymbol{\theta}}=\begin{bmatrix}x[0]\\x[1]\end{bmatrix}\qquad \mathbf{H}=\begin{bmatrix}1&0\\0&1\\0&0\end{bmatrix}$$

约束 $\mathbf{A}\boldsymbol{\theta}=\boldsymbol{b}$ → $[1,-1]\boldsymbol{\theta}=0$

$$\Rightarrow\quad \hat{\boldsymbol{\theta}}_c=\begin{bmatrix}\frac{1}{2}(x[0]+x[1])\\ \frac{1}{2}(x[0]+x[1])\end{bmatrix}$$

五、非线性最小二乘估计

已知信号模型$s[n;\theta]$，观测数据为$x[n]=s[n;\theta]+w[n]$，θ的最小二乘估计为

$$\hat{\theta}=\arg\min_{\theta} J(\theta)$$

$$J(\theta)=\sum_{n=0}^{N-1}\left(x[n]-s[n;\theta]\right)^2$$

最小二乘估计

$$J(\boldsymbol{\theta})=\left(\mathbf{x}-\mathbf{s}(\boldsymbol{\theta})\right)^T\left(\mathbf{x}-\mathbf{s}(\boldsymbol{\theta})\right)$$ **非线性问题！**

1. 网格搜索法

一般适用于待搜索维数较小者

2. 参数变换法

$$J(\theta)=\left(\boldsymbol{x}-\boldsymbol{s}(\theta)\right)^{T}\left(\boldsymbol{x}-\boldsymbol{s}(\theta)\right)$$

核心思想：将非线性参数转换为线性参数来求解

对该估计参数 $\boldsymbol{\theta}$，进行某一对一变换：

$$\boldsymbol{\alpha}=\boldsymbol{g}(\boldsymbol{\theta})$$

使得非线性信号模型能转换为线性信号模型：

$$\boldsymbol{s}(\boldsymbol{\theta})=\mathbf{H}\boldsymbol{\alpha}$$

$$\boldsymbol{x}=\mathbf{H}\boldsymbol{\alpha}+\boldsymbol{w}$$

若 $\boldsymbol{\alpha}$ 的LSE为：

$$\hat{\boldsymbol{\alpha}}=\left(\mathbf{H}^{T}\mathbf{H}\right)^{-1}\mathbf{H}^{T}\boldsymbol{x}$$

则待估计参数 $\boldsymbol{\theta}$ 的LSE可由下式求出：

$$\hat{\boldsymbol{\theta}}=\boldsymbol{g}^{-1}(\hat{\boldsymbol{\alpha}})$$

——LSE的不变性！

例：信号模型：

$$x[n] = A\cos\left(2\pi f_0 n + \phi\right) + w[n]$$

f_0 已知，待估计参数为信号幅度（$A>0$）和载波相位 ϕ。

$$J = \sum_{n=0}^{N-1}\left(x[n] - A\cos\left(2\pi f_0 n + \phi\right)\right)^2$$ 非线性问题！

$$A\cos\left(2\pi f_0 n + \phi\right) = \underbrace{A\cos(\phi)}_{\alpha_1}\cos\left(2\pi f_0 n\right) \underbrace{- A\sin(\phi)}_{\alpha_2}\sin\left(2\pi f_0 n\right)$$

$$s[n] = \alpha_1 \cos\left(2\pi f_0 n\right) + \alpha_2 \sin\left(2\pi f_0 n\right)$$

$$\boldsymbol{s} = \underbrace{\begin{bmatrix} 1 & 0 \\ \cos\left(2\pi f_0\right) & \sin\left(2\pi f_0\right) \\ \vdots & \vdots \\ \cos\left(2\pi f_0 (N-1)\right) & \sin\left(2\pi f_0 (N-1)\right) \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}}_{\boldsymbol{\alpha}}$$

$$\boldsymbol{x} = \mathbf{H}\boldsymbol{\alpha} + \boldsymbol{w}$$

$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\boldsymbol{x}$$

$$A = \sqrt{\alpha_1^2 + \alpha_2^2}$$

$$\phi = \arctan\left(\frac{-\alpha_2}{\alpha_1}\right)$$

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{A} \\ \hat{\phi} \end{bmatrix} = \begin{bmatrix} \sqrt{\hat{\alpha}_1^2 + \hat{\alpha}_2^2} \\ \arctan\left(\frac{-\hat{\alpha}_2}{\hat{\alpha}_1}\right) \end{bmatrix}$$

3. 参数分离法

$$J(\theta)=\left(\mathbf{x}-\mathbf{s}(\theta)\right)^{T}\left(\mathbf{x}-\mathbf{s}(\theta)\right)$$

核心思想：将非线性参数尽量转换为线性参数，以减小复杂度

尽量将信号变换为如下形式

$$\mathbf{s}(\theta)=\mathbf{H}(\alpha)\beta$$

1. 参数分离

其中待估计参数为 $\theta=\begin{bmatrix}\alpha\\ \beta\end{bmatrix}$。此时，LS误差为

$$J(\alpha,\beta)=\left(\mathbf{x}-\mathbf{H}(\alpha)\beta\right)^{T}\left(\mathbf{x}-\mathbf{H}(\alpha)\beta\right)$$

对给定的 α，使LS误差最小的 β 为

2. 线性参数非线性表示

$$\hat{\beta}=\left(\mathbf{H}^{T}(\alpha)\mathbf{H}(\alpha)\right)^{-1}\mathbf{H}^{T}(\alpha)\mathbf{x}$$

4. 线性参数求解

此时，LS误差为

$$J(\alpha,\hat{\beta})=\mathbf{x}^{T}\left(\mathbf{I}-\mathbf{H}(\alpha)\left(\mathbf{H}^{T}(\alpha)\mathbf{H}(\alpha)\right)^{-1}\mathbf{H}^{T}(\alpha)\right)\mathbf{x}$$

此时， α 的LSE为

$$\max_{\alpha}\left\{\mathbf{x}^{T}\mathbf{H}(\alpha)\left(\mathbf{H}^{T}(\alpha)\mathbf{H}(\alpha)\right)^{-1}\mathbf{H}^{T}(\alpha)\mathbf{x}\right\}$$

3. 非线性参数求解

4. Gauss-Newton迭代法

$$J(\theta)=\sum_{n=0}^{N-1}\left(x[n]-s[n;\theta]\right)^2$$

核心思想：非线性问题线性化后迭代求解

某个标称的θ 附近对非线性函数进行线性化

$$s[n;\theta]\approx s[n;\theta_0]+\left.\frac{\partial s[n;\theta]}{\partial\theta}\right|_{\theta=\theta_0}(\theta-\theta_0)$$

此时，LS误差为

$$J(\theta)=\sum_{n=0}^{N-1}\left(x[n]-s[n;\theta]\right)^2$$

$$\approx\sum_{n=0}^{N-1}\left(x[n]-s[n;\theta_0]+\left.\frac{\partial s[n;\theta]}{\partial\theta}\right|_{\theta=\theta_0}\theta_0-\left.\frac{\partial s[n;\theta]}{\partial\theta}\right|_{\theta=\theta_0}\theta\right)^2$$

$$=\left(\underline{\mathbf{x}-\mathbf{s}(\theta_0)+\mathbf{H}(\theta_0)\theta_0}-\underline{\mathbf{H}(\theta_0)}\theta\right)^T\left(\mathbf{x}-\mathbf{s}(\theta_0)+\mathbf{H}(\theta_0)\theta_0-\mathbf{H}(\theta_0)\theta\right)$$

其中，$\left[\mathbf{H}(\theta)\right]_i=\dfrac{\partial s[i;\theta]}{\partial\theta}$

LSE为

$$\hat{\theta}=\left(\mathbf{H}^T(\theta_0)\mathbf{H}(\theta_0)\right)^{-1}\mathbf{H}^T(\theta_0)\left(\boldsymbol{x}-\mathbf{s}(\theta_0)+\mathbf{H}(\theta_0)\theta_0\right)$$
$$=\theta_0+\left(\mathbf{H}^T(\theta_0)\mathbf{H}(\theta_0)\right)^{-1}\mathbf{H}^T(\theta_0)\left(\boldsymbol{x}-\mathbf{s}(\theta_0)\right)$$

其迭代解法:

$$\theta_{k+1}=\theta_k+\left(\mathbf{H}^T(\theta_k)\mathbf{H}(\theta_k)\right)^{-1}\mathbf{H}^T(\theta_k)\left(\boldsymbol{x}-\mathbf{s}(\theta_k)\right)$$

进一步地，可推广至矢量参数时

$$\boldsymbol{\theta}_{k+1}=\boldsymbol{\theta}_k+\left(\mathbf{H}^T(\boldsymbol{\theta}_k)\mathbf{H}(\boldsymbol{\theta}_k)\right)^{-1}\mathbf{H}^T(\boldsymbol{\theta}_k)\left(\boldsymbol{x}-\mathbf{s}(\boldsymbol{\theta}_k)\right)$$

其中, $\left[\mathbf{H}(\boldsymbol{\theta})\right]_{ij}=\dfrac{\partial s[i;\boldsymbol{\theta}]}{\partial\theta_j}$

- **存在收敛问题**

Seber, G.A.F., Wild, C.J. *Nonlinear Regression*. J. Wiley, New York, 1989

其它类似方法：Newton-Raphson迭代法

5. 循环最小化方法

$$J(\boldsymbol{\theta})=\left(\boldsymbol{x}-\boldsymbol{s}(\boldsymbol{\theta})\right)^T\left(\boldsymbol{x}-\boldsymbol{s}(\boldsymbol{\theta})\right)$$

- 将待估计参数分为两部分：$\boldsymbol{\theta}=[\xi \quad \varsigma]^T$ 也可分为更多部分
- 给定 ς^0
- 计算 $\xi^1=\arg\min_{\xi}\left\{J\left(\left[\xi \quad \varsigma^0\right]^T\right)\right\}$ 和 $\varsigma^1=\arg\min_{\varsigma}\left\{J\left(\left[\xi^1 \quad \varsigma\right]^T\right)\right\}$
- 计算 $\xi^2=\arg\min_{\xi}\left\{J\left(\left[\xi \quad \varsigma^1\right]^T\right)\right\}$ 和 $\varsigma^2=\arg\min_{\varsigma}\left\{J\left(\left[\xi^2 \quad \varsigma\right]^T\right)\right\}$
- …
- 计算 $\xi^k=\arg\min_{\xi}\left\{J\left(\left[\xi \quad \varsigma^{k-1}\right]^T\right)\right\}$ 和 $\varsigma^k=\arg\min_{\varsigma}\left\{J\left(\left[\xi^k \quad \varsigma\right]^T\right)\right\}$

  直至收敛

6. 放松估计方法

例：正弦信号参数估计

$$x[n]=\sum_{p=1}^{P}\alpha_p e^{j2\pi f_p n}+w[n],\quad n=0,1,2,\dots N-1$$

其中信号幅度 α_p 和频率 f_p 未知。

$$\mathbf{x}=\left[x[0],x[1],\dots,x[N-1]\right]^T$$

$$\boldsymbol{\alpha}=\left[\alpha_1,\alpha_1,\dots,\alpha_p\right]^T$$

$$\boldsymbol{f}=\left[f_1,f_1,\dots,f_p\right]^T$$

$$\boldsymbol{h}(f_p)=\left[1,e^{j2\pi f_p},e^{j2\pi f_p 2}\dots,e^{j2\pi f_p(N-1)}\right]^T$$

$$\boldsymbol{w}=\left[w[0],w[1],\dots,w[N-1]\right]^T$$

$$\Rightarrow\quad \boldsymbol{x}=\sum_{p=1}^{P}\alpha_p\boldsymbol{h}(f_p)+\boldsymbol{w}$$

$$\mathbf{H}(\boldsymbol{f})=\left[\boldsymbol{h}(f_1),\boldsymbol{h}(f_2),\boldsymbol{h}(f_1),\dots,\boldsymbol{h}(f_p)\right]$$

$$\boldsymbol{x} = \mathbf{H}(\boldsymbol{f})\boldsymbol{\alpha} + \boldsymbol{w} \Rightarrow \begin{cases} \hat{\boldsymbol{\alpha}} = \left(\mathbf{H}^H(\boldsymbol{f})\mathbf{H}(\boldsymbol{f})\right)^{-1}\mathbf{H}^H(\boldsymbol{f})\boldsymbol{x} \\ \max\limits_{\boldsymbol{f}}\left\{\boldsymbol{x}^H\mathbf{H}(\boldsymbol{f})\left(\mathbf{H}^H(\boldsymbol{f})\mathbf{H}(\boldsymbol{f})\right)^{-1}\mathbf{H}^H(\boldsymbol{f})\boldsymbol{x}\right\} \end{cases}$$

问题：若 $\hat{f}_i \approx \hat{f}_j$ ？

将导致幅度估计性能较差

假定 $\left\{\hat{\alpha}_i, \hat{f}_i\right\}_{i=1,i\neq p}^{P}$ 已知或通过其它方式估计得到，那么可构建新的观测数据

$$\boldsymbol{x}_p = \boldsymbol{x} - \sum_{i=1,i\neq p}^{P} \hat{\alpha}_i \boldsymbol{h}\left(\hat{f}_i\right) \qquad \ldots\ldots(1)$$

$$\Rightarrow J\left(\alpha_p, f_p\right) = \left(\boldsymbol{x}_p - \alpha_p \boldsymbol{h}\left(f_p\right)\right)^H \left(\boldsymbol{x}_p - \alpha_p \boldsymbol{h}\left(f_p\right)\right)$$

$$\Rightarrow \hat{f}_p = \arg\max_{f_p} \left|\frac{\boldsymbol{h}^H\left(f_p\right)\boldsymbol{x}_p}{N}\right|^2 \qquad \hat{\alpha}_p = \left.\frac{\boldsymbol{h}^H\left(f_p\right)\boldsymbol{x}_p}{N}\right|_{f_p = \hat{f}_p} \qquad \ldots\ldots(2)$$

- *Li J, Stoica P, Efficient mixed-spectrum estimation with applications to target feature extraction, IEEE Transactions on signal processing, 1996, 44(2): 281-295*
- *吴仁彪等，通用鲁棒的放松估计方法，科学出版社，2017年*

➢ 假定信号数为1

- 利用观测数据 $\boldsymbol{x}$ 和式（2）估计得到 $\left\{\hat{\alpha}_1, \hat{f}_1\right\}$

➢ 假定信号数为2

- 利用式（1）和估计得到的 $\left\{\hat{\alpha}_1, \hat{f}_1\right\}$ 构建数据 $\boldsymbol{x}_2$ 。利用 $\boldsymbol{x}_2$ 和式（2）估计得到 $\left\{\hat{\alpha}_2, \hat{f}_2\right\}$
- 利用式（1）和估计得到的 $\left\{\hat{\alpha}_2, \hat{f}_2\right\}$ 构建数据 $\boldsymbol{x}_1$ 。利用 $\boldsymbol{x}_1$ 和式（2）估计得到 $\left\{\hat{\alpha}_1, \hat{f}_1\right\}$
- 重复上述步骤直至收敛

收敛: 残差小于门限或迭代次数达到门限

➢ 假定信号数为3，按类似方法反复估计直至收敛

……

➢ 假定信号数为 P，按类似方法反复估计直至收敛

六、小结

- 最小二乘估计
  - 无需似然函数，与MVU、MLE不同
  - 与同样也不需要似然函数的BLUE也不同
  - 核心思想：使由待估计参数构建的**信号**与**观测数据**“尽量接近”

体会各类估计理论与方法核心思想的差异

——信号模型至关重要

- 线性最小二乘估计
- 加权最小二乘估计
- 序贯最小二乘估计
- 约束最小二乘估计
- 非线性最小二乘估计
  - 网格搜索法、参数变换、参数分离法、迭代方法，等