极大似然法辨识

邢超

西北工业大学航天学院

基本思想

辨识准则 观测值的出现概率最大

似然函数 观察值的概率密度函数

方法特点

1. 适用于 $\xi(k)$ 相关情况;
2. 当系统信噪比较小时有较好的估计效果;
3. 算法稳定度好;
4. 有递推算法;
5. 实际工程中广泛使用

似然函数

设某离散随机过程 $\{V_k\}$ 与待辨识参数 θ 有关，其概率分布密度 $f(V_k|\theta)$ 已知，若测得 n 个独立的观测值 $V_1, V_2, \cdots, V_n$ ，其分布密度为：$f(V_1|\theta), \cdots, f(V_n|\theta)$ ，定义似然函数 L 为：

$$L(V_1, \cdots, V_n|\theta) = f(V_1|\theta) \cdot f(V_2|\theta) \cdots f(V_n|\theta)$$

极大似然估计

辨识 θ 的原则就是使得 L 达到极大值，即:

$$\frac{\partial \mathrm{L}}{\partial \theta} = 0$$

通常对 L 取对数：

$$\ln \mathrm{L} = \ln \mathrm{f}(\mathrm{V}_1|\theta) + \cdots + \ln \mathrm{f}(\mathrm{V}_\mathrm{n}|\theta)$$

求解：

$$\frac{\partial \ln \mathrm{L}}{\partial \theta}$$

所得 θ 即为极大似然估计 $\hat{\theta}_{\mathrm{ML}}$

差分方程的极大似然辨识：系统模型（白噪声情况）

系统差分方程：

$$a(z^{-1})y(k) = b(z^{-1})u(k) + \xi(k)$$

式中，$\xi(k)$ 为高斯白噪声序列且与 $u(k)$ 无关。以向量形式表示：

$$Y_N = \Phi_N \boldsymbol{\theta} + \boldsymbol{\xi}$$

差分方程的极大似然辨识：残差（白噪声情况）

系统估计残差：

$$
\begin{aligned}
\mathrm{e_N} &= \mathrm{Y_N} - \Phi_{\mathrm{N}}\hat{\boldsymbol{\theta}} \\
\mathrm{e_N} &= [\mathrm{e}(\mathrm{n}+1), \mathrm{e}(\mathrm{n}+2), \cdots, \mathrm{e}(\mathrm{n}+\mathrm{N})]^{\mathrm{T}}
\end{aligned}
$$

由于 $\xi(\mathrm{k})$ 为高斯白噪声，故假设 $\mathrm{e}(\mathrm{k})$ 也为高斯白噪声。设 $\mathrm{e}(\mathrm{k})$ 方差为 σ^2。概率密度函数为：

$$
\mathrm{p}(\mathrm{e}(\mathrm{k})|\hat{\theta}) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{\mathrm{e}^2(\mathrm{k})}{2\sigma^2}}
$$

似然函数为：

$$\begin{aligned}
L(Y_N|\hat{\theta}) &= \prod_{k=n+1}^{n+N} p(e(k)|\hat{\theta}) \\
&= \frac{1}{(2\pi\sigma^2)^{N/2}}\exp[-\frac{\sum e^2(k)}{2\sigma^2}] \\
&= \frac{1}{(2\pi\sigma^2)^{N/2}}\exp[-\frac{(Y_N-\Phi_N\hat{\theta})^T(Y_N-\Phi_N\hat{\theta})}{2\sigma^2}] \\
\ln L(Y_N|\hat{\theta}) &= -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln\sigma^2 - \frac{(Y_N-\Phi_N\hat{\theta})^T(Y_N-\Phi_N\hat{\theta})}{2\sigma^2}
\end{aligned}$$

差分方程的极大似然辨识：似然函数（白噪声情况）

则依极大似然辨识原理有：

$$\frac{\partial \ln L(Y_N|\hat{\theta})}{\partial \hat{\theta}} = \frac{\Phi_N^T Y_N - \Phi_N^T \Phi_N \hat{\boldsymbol{\theta}}}{\sigma^2} = 0$$

$$\frac{\partial \ln L(Y_N|\hat{\theta})}{\partial \hat{\sigma}^2} = -\frac{N}{2\sigma^2} + \frac{(Y_N - \Phi_N \hat{\boldsymbol{\theta}})^T (Y_N - \Phi_N \hat{\boldsymbol{\theta}})}{2\sigma^4} = 0$$

解上述方程有：

$$\hat{\boldsymbol{\theta}}_{ML} = (\boldsymbol{\Phi}_N^T \boldsymbol{\Phi}_N)^{-1} \boldsymbol{\Phi}_N^T Y_N$$

$$\sigma^2 = \frac{(Y_N - \Phi_N \hat{\boldsymbol{\theta}})^T (Y_N - \Phi_N \hat{\boldsymbol{\theta}})}{N}$$

可见在 $\xi(k)$ 为高斯白噪声序列这一特殊情况下，极大似然辨识与一般最小二乘法辨识具有相同结果。

差分方程的极大似然辨识：系统模型（有色噪声情况）

$$a(z^{-1})y(k) = b(z^{-1})u(k) + c(z^{-1})\varepsilon(k)$$

其中：

$$\begin{aligned} a(z^{-1}) &= 1 + a_1 z^{-1} + \cdots + a_n z^{-n} \\ b(z^{-1}) &= b_0 + b_1 z^{-1} + \cdots + b_n z^{-n} \\ c(z^{-1}) &= 1 + c_1 z^{-1} + \cdots + c_n z^{-n} \end{aligned}$$

差分方程的极大似然辨识：预测误差（有色噪声情况）

预测误差：

$$e(k) = y(k) - \hat{y}(k)$$

其向量形式为：

$$e_N = Y_N - \Phi_N \hat{\theta}$$

其中：

$$\begin{aligned}
\hat{\boldsymbol{\theta}} &= [\hat{a}_1, \cdots, \hat{a}_n, \hat{b}_0, \cdots, \hat{b}_n, \hat{c}_1, \cdots, \hat{c}_n]^T \\
Y_N &= [y(n+1), \cdots, y(n+N)]^T \\
e_N &= [e(n+1), \cdots, e(n+N)]^T \\
\Phi_N &= \begin{bmatrix} -y_n & \cdots & -y_1 & u_{n+1} & \cdots & u_1 & e_n & \cdots & e_1 \\ -y_{n+1} & \cdots & -y_2 & u_{n+2} & \cdots & u_2 & e_{n+1} & \cdots & e_2 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ -y_{n+N-1} & \cdots & -y_N & u_{n+N} & \cdots & u_N & e_{n+N-1} & \cdots & e_N \end{bmatrix}
\end{aligned}$$

因为 $\varepsilon(k)$ 为高斯白噪声，故而 $e(k)$ 可假设为零均值的高斯白噪声。则似然函数为：

$$L(Y_N|\hat{\theta}) = \frac{1}{(2\pi\sigma^2)^{N/2}}\exp[-\frac{(Y_N-\phi_N\hat{\boldsymbol{\theta}})^T(Y_N-\phi_N\hat{\boldsymbol{\theta}})}{2\sigma^2}]$$

$$\ln L(Y_N|\hat{\theta}) = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{k=n+1}^{n+N} e^2(k)$$

由 $\frac{\partial \ln L(Y_N|\hat{\theta})}{\partial\sigma^2} = 0$ 得：

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{k=n+1}^{n+N} e^2(k)$$

$$
\begin{aligned}
J &= \frac{1}{2}\sum_{k=n+1}^{n+N} e^2(k) \\
\sigma^2 &= \frac{2}{N}J \\
\ln L(Y_N|\hat{\theta}) &= -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln(\frac{2J}{N}) - \frac{N}{2}
\end{aligned}
$$

- J 为参数 $a_1, a_2, \cdots, a_n; b_0, b_1, \cdots, b_n; c_1, c_2, \cdots, c_n$ 的二次型函数。
- 使 L 最大的 $\hat{\theta}$, 等价于在约束条件

$$\hat{c}(z^{-1})e(k) = \hat{a}(z^{-1})y(k) - \hat{b}(z^{-1})u(k)$$

下求 $\hat{\theta}$，使 J 最小。

牛顿 -拉卜森法的迭代公式：

$$\hat{\theta}_1 = \hat{\theta}_0 - \left[\left(\frac{\partial^2 J}{\partial \theta^2} \right)^{-1} \frac{\partial J}{\partial \theta} \right] \Bigg|_{\theta = \hat{\theta}_0}$$

其中：

- $\frac{\partial J}{\partial \theta}$ 为梯度
- $\frac{\partial^2 J}{\partial \theta^2}$ 为海赛矩阵

Newton-Raphson 迭代步骤：初始值选定

$$\hat{\theta}_0 = [\hat{a}_1, \cdots, \hat{a}_n, \hat{b}_0, \cdots, \hat{b}_n, \hat{c}_1, \cdots, \hat{c}_n]^T$$

其中：

- $\hat{a}_1, \cdots, \hat{a}_n, \hat{b}_0, \cdots, \hat{b}_n$ 可用最小二乘法求得
- $\hat{c}_1, \cdots, \hat{c}_n$ 可取为零或任意假定某一组值

Newton-Raphson 迭代步骤：计算预测误差

- 预测误差，指标函数与误差方差估计值：

$$
\begin{aligned}
e(k) &= y(k) - \hat{y}(k) \\
J &= \frac{\sum_{k=n+1}^{n+N} e^2(k)}{2} \\
\sigma^2 &= \frac{2J}{N}
\end{aligned}
$$

Newton-Raphson 迭代步骤：计算梯度矩阵及海赛矩阵

$$\begin{aligned}
\frac{\partial J}{\partial \theta} &= \sum_{k=n+1}^{n+N} e(k)\frac{\partial e(k)}{\partial \theta} \\
\frac{\partial^2 J}{\partial \theta^2} &= \sum_{k=n+1}^{n+N} \frac{\partial e(k)}{\partial \theta}\left[\frac{\partial e(k)}{\partial \theta}\right]^T + \sum_{k=n+1}^{n+N} e(k)\frac{\partial^2 e(k)}{\partial \theta^2} \\
&\approx \sum_{k=n+1}^{n+N} \frac{\partial e(k)}{\partial \theta}\left[\frac{\partial e(k)}{\partial \theta}\right]^T
\end{aligned}$$

其中：

$$\frac{\partial e(k)}{\partial \theta} = \left[\frac{\partial e(k)}{\partial a_1}, \cdots, \frac{\partial e(k)}{\partial a_n}, \frac{\partial e(k)}{\partial b_0}, \cdots, \frac{\partial e(k)}{\partial b_n}, \frac{\partial e(k)}{\partial c_1}, \cdots, \frac{\partial e(k)}{\partial c_n}\right]^T$$

牛顿 -拉卜森迭代步骤：计算新的估计值

$$\hat{\boldsymbol{\theta}}_1 = \hat{\boldsymbol{\theta}}_0 - \left[\left(\frac{\partial^2 \mathrm{J}}{\partial \boldsymbol{\theta}^2} \right)^{-1} \frac{\partial \mathrm{J}}{\partial \boldsymbol{\theta}} \right] \Bigg|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_0}$$

停上条件：

$$\frac{\hat{\sigma}^2_{\mathrm{r}+1} - \hat{\sigma}^2_{\mathrm{r}}}{\hat{\sigma}^2_{\mathrm{r}}} < \delta$$

其中 δ 可取较小的数，如 $\delta = 10^{-4}$。

系统差分方程

$$a(z^{-1})y(k) = b(z^{-1})u(k) + c(z^{-1})\varepsilon(k)$$

其中：

$$\begin{aligned} a(z^{-1}) &= 1 + a_1 z^{-1} + \cdots + a_n z^{-n} \\ b(z^{-1}) &= b_0 + b_1 z^{-1} + \cdots + b_n z^{-n} \\ c(z^{-1}) &= 1 + c_1 z^{-1} + \cdots + c_n z^{-n} \end{aligned}$$

可写为：

$$\varepsilon(k) = c^{-1}(z^{-1})[a(z^{-1})y(k) - b(z^{-1})u(k)]$$

二次型指标函数

将指标函数用二次型近似表示：

$$\begin{aligned} J_N &= \sum_{k=n+1}^{n+N} \varepsilon^2(k) \\ &\approx (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_N)^T p_N^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_N) + \beta_N \end{aligned}$$

利用泰勒级数将 $\varepsilon(\mathrm{k})$ 在估值 $\hat{\theta}$ 展开：

$$\varepsilon(\mathrm{k}) \approx \varepsilon(\mathrm{k}, \hat{\boldsymbol{\theta}}) + \left[\frac{\partial \varepsilon(\mathrm{k}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]^{\mathrm{T}}\Bigg|_{\hat{\theta}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$$

其中：

$$\begin{aligned} \varepsilon(\mathrm{k}, \hat{\boldsymbol{\theta}}) &= \mathrm{e}(\mathrm{k}) \\ \mathrm{e}(\mathrm{k}) &= \hat{\mathrm{c}}^{-1}(\mathrm{z}^{-1})[\hat{\mathrm{a}}(\mathrm{z}^{-1})\mathrm{y}(\mathrm{k}) - \hat{\mathrm{b}}(\mathrm{z}^{-1})\mathrm{u}(\mathrm{k})] \end{aligned}$$

可得：

$$
\begin{aligned}
J_{N+1} &= \sum_{k=n+1}^{n+N+1} \varepsilon^2(k) \\
&\approx (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_N)^T p_N^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_N) \\
&\quad + \beta_N + [e_{N+1} + \boldsymbol{\psi}_{N+1}^T (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_N)]^2
\end{aligned}
$$

其中：

$$
\begin{aligned}
e_{N+1} &= e(n + N + 1) \\
\boldsymbol{\psi}_{N+1} &= \frac{\partial e_{N+1}}{\partial \hat{\boldsymbol{\theta}}}
\end{aligned}
$$

设：

$$\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}=\Delta$$

得：

$$\begin{aligned}
\mathrm{J}_{\mathrm{N}+1}(\boldsymbol{\theta}) &= \boldsymbol{\Delta}^{\mathrm{T}}(\mathrm{P}_{\mathrm{N}}^{-1}+\boldsymbol{\Psi}_{\mathrm{N}+1}\boldsymbol{\Psi}_{\mathrm{N}+1}^{\mathrm{T}})\boldsymbol{\Delta} \\
&\quad +2\mathrm{e}_{\mathrm{N}+1}\boldsymbol{\Psi}_{\mathrm{N}+1}^{\mathrm{T}}\boldsymbol{\Delta}+\mathrm{e}_{\mathrm{N}+1}^{2}+\beta_{\mathrm{N}} \\
&= (\boldsymbol{\Delta}+\mathrm{r}_{\mathrm{N}+1})^{\mathrm{T}}\mathrm{P}_{\mathrm{N}+1}^{-1}(\boldsymbol{\Delta}+\mathrm{r}_{\mathrm{N}+1})+\beta_{\mathrm{N}+1}
\end{aligned}$$

其中：

$$\begin{aligned}
\mathrm{P}_{\mathrm{N}+1}^{-1} &= \mathrm{P}_{\mathrm{N}}^{-1}+\boldsymbol{\Psi}_{\mathrm{N}+1}\boldsymbol{\Psi}_{\mathrm{N}+1}^{\mathrm{T}} \\
\mathrm{r}_{\mathrm{N}+1} &= \mathrm{P}_{\mathrm{N}+1}\boldsymbol{\Psi}_{\mathrm{N}+1}\mathrm{e}_{\mathrm{N}+1} \\
\beta_{\mathrm{N}+1} &= \mathrm{e}_{\mathrm{N}+1}^{2}+\beta_{\mathrm{N}}-\mathrm{e}_{\mathrm{N}+1}\boldsymbol{\Psi}_{\mathrm{N}+1}^{\mathrm{T}}\mathrm{P}_{\mathrm{N}+1}\boldsymbol{\Psi}_{\mathrm{N}+1}\mathrm{e}_{\mathrm{N}+1}
\end{aligned}$$

$\beta_{\mathrm{N}+1}$ 为已知值，所以

$$\hat{\boldsymbol{\theta}}_{\mathrm{N}+1}=\hat{\boldsymbol{\theta}}_{\mathrm{N}}-\mathrm{r}_{\mathrm{N}+1}$$

时使 $\mathrm{J}_{\mathrm{N}+1}$ 最小。

更新 $\mathrm{P_{N+1}}, \hat{\theta}_{\mathrm{N+1}}$

利用矩阵求逆引理，得：

$$
\begin{aligned}
\mathrm{P_{N+1}^{-1}} &= \mathrm{P_N^{-1}} + \boldsymbol{\Psi}_{\mathrm{N+1}}\boldsymbol{\Psi}_{\mathrm{N+1}}^{\mathrm{T}} \\
\mathrm{P_{N+1}} &= \mathrm{P_N}\left[\mathrm{I} - \frac{\Psi_{\mathrm{N+1}}\Psi_{\mathrm{N+1}}^{\mathrm{T}}\mathrm{P_N}}{1+\Psi_{\mathrm{N+1}}^{\mathrm{T}}\mathrm{P_N}\Psi_{\mathrm{N+1}}}\right] \\
\mathrm{r_{N+1}} &= \mathrm{P_{N+1}}\Psi_{\mathrm{N+1}}\mathrm{e_{N+1}} \\
&= \mathrm{P_N}\left[\mathrm{I} - \frac{\Psi_{\mathrm{N+1}}\Psi_{\mathrm{N+1}}^{\mathrm{T}}\mathrm{P_N}}{1+\Psi_{\mathrm{N+1}}^{\mathrm{T}}\mathrm{P_N}\Psi_{\mathrm{N+1}}}\right]\Psi_{\mathrm{N+1}}\mathrm{e_{N+1}} \\
&= \mathrm{P_N}\left[\frac{1+\Psi_{\mathrm{N+1}}\Psi_{\mathrm{N+1}}^{\mathrm{T}}\mathrm{P_N}\Psi_{\mathrm{N+1}} - \Psi_{\mathrm{N+1}}\Psi_{\mathrm{N+1}}^{\mathrm{T}}\mathrm{P_N}\Psi_{\mathrm{N+1}}}{1+\Psi_{\mathrm{N+1}}^{\mathrm{T}}\mathrm{P_N}\Psi_{\mathrm{N+1}}}\right]\mathrm{e_{N+1}} \\
&= \frac{\mathrm{P_N}\Psi_{\mathrm{N+1}}\mathrm{e_{N+1}}}{1+\Psi_{\mathrm{N+1}}^{\mathrm{T}}\mathrm{P_N}\Psi_{\mathrm{N+1}}} \\
\hat{\theta}_{\mathrm{N+1}} &= \hat{\theta}_{\mathrm{N}} - \mathrm{r_{N+1}} \\
&= \hat{\theta}_{\mathrm{N}} - \mathrm{P_N}\Psi_{\mathrm{N+1}}(1+\Psi_{\mathrm{N+1}}^{\mathrm{T}}\mathrm{P_N}\Psi\mathrm{N}+1)^{-1}\mathrm{e_{N+1}}
\end{aligned}
$$

更新 Ψ_{N+1}

$$\Psi_{N+1} = \begin{bmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{bmatrix} \Psi_N + D$$

其中：

$$A = \begin{bmatrix} -\hat{c}_1 & \cdots & \cdots & -\hat{c}_n \\ 1 & \cdots & \cdots & 0 \\ & \ddots & & \vdots \\ & & 1 & 0 \end{bmatrix}$$

$$
\begin{aligned}
B &= \begin{bmatrix} -\hat{c}_1 & \cdots & \cdots & -\hat{c}_n & 0 \\ 1 & \cdots & \cdots & 0 & 0 \\ & \ddots & & \vdots & \vdots \\ & & 1 & 0 & 0 \end{bmatrix} \\
C &= \begin{bmatrix} -\hat{c}_1 & \cdots & \cdots & -\hat{c}_n \\ 1 & \cdots & \cdots & 0 \\ & \ddots & & \vdots \\ & & 1 & 0 \end{bmatrix} \\
D &= [y_{n+N}, 0, \cdots, 0, -u_{n+N+1}, 0, \cdots, 0, -e_{n+N}, 0, \cdots, 0]^T
\end{aligned}
$$

A, B, C 的推导

从

$$e(k) = \hat{c}^{-1}(z^{-1})[\hat{a}(z^{-1})y(k) - \hat{b}(z^{-1})u(k)]$$

得：

$$\begin{aligned}
\frac{\partial e(k)}{\partial \hat{a}_i} &= \hat{c}^{-1}(z^{-1})y(k-i) \\
\frac{\partial e(k)}{\partial \hat{b}_i} &= -\hat{c}^{-1}(z^{-1})u(k-i) \\
\frac{\partial e(k)}{\partial \hat{c}_i} &= -\hat{c}^{-1}(z^{-1})e(k-i)
\end{aligned}$$

进一步有：

$$\begin{aligned}
\frac{\partial e(k)}{\partial \hat{a}_i} &= \frac{\partial e(k-i+j)}{\partial \hat{a}_j} \\
\frac{\partial e(k)}{\partial \hat{b}_i} &= \frac{\partial e(k-i+j)}{\partial \hat{b}_j} \\
\frac{\partial e(k)}{\partial \hat{c}_i} &= \frac{\partial e(k-i+j)}{\partial \hat{c}_j}
\end{aligned}$$

使用牛顿 -拉卜森方法的递推公式：系统差分方程

$$a(z^{-1})y(k) = b(z^{-1})u(k) + \frac{1}{d(z^{-1})}\varepsilon(k)$$

其中：

$$\begin{aligned}
a(z^{-1}) &= 1 + a_1 z^{-1} + \cdots + a_n z^{-n} \\
b(z^{-1}) &= b_0 + b_1 z^{-1} + \cdots + b_n z^{-n} \\
d(z^{-1}) &= 1 + d_1 z^{-1} + \cdots + d_n z^{-n}
\end{aligned}$$

参数向量为：

$$\begin{aligned}
a &= [a_1, a_2, \cdots, a_n]^T \\
b &= [b_o, b_1, \cdots, b_n]^T \\
d &= [d_1, d_2, \cdots, d_n]^T \\
\boldsymbol{\theta} &= [a^T, b^T, d^T]^T
\end{aligned}$$

计算 $\frac{\partial \varepsilon(k)}{\partial \boldsymbol{\theta}}$

将系统差分方程改写为：

$$\varepsilon(k) = d(z^{-1})[a(z^{-1})y(k) - b(z^{-1})u(k)]$$

可得：

$$\begin{aligned}
\frac{\partial \varepsilon(k)}{\partial a_j} &= d(z^{-1})y(k-j) = y^F_{k-j}, j = 1, 2, \cdots, n \\
\frac{\partial \varepsilon(k)}{\partial b_j} &= -d(z^{-1})u(k-j) = u^F_{k-j}, j = 0, 1, 2, \cdots, n \\
\frac{\partial \varepsilon(k)}{\partial d_j} &= a(z^{-1})y(k-j) - b(z^{-1})u(k-j) = -\mu_{k-j}, j = 1, 2, \cdots, n
\end{aligned}$$

计算 $\frac{\partial \varepsilon(k)}{\partial \boldsymbol{\theta}}$

$$\frac{\partial \varepsilon(k)}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \bar{y}^F_{(n)} \\ -\bar{u}^F_{(n+1)} \\ -\bar{\boldsymbol{\mu}}_{(n)} \end{bmatrix}$$

其中：

$$\begin{aligned} \bar{y}^F_{(n)} &= [y^F_{k-1}, y^F_{k-2}, \cdots, y^F_{k-n}]^T \\ -\bar{u}^F_{(n+1)} &= [u^F_k, u^F_{k-1}, \cdots, u^F_{k-n}]^T \\ -\bar{\boldsymbol{\mu}}_{(n)} &= [\mu_{k-1}, \mu_{k-2}, \cdots, \mu_{k-n}]^T \end{aligned}$$

计算 $\frac{\partial^2\varepsilon(k)}{\partial\boldsymbol{\theta}^2}$

$$\frac{\partial^2\varepsilon(k)}{\partial\theta^2}=\begin{bmatrix}\frac{\partial^2\varepsilon(k)}{\partial a^2} & \frac{\partial^2\varepsilon(k)}{\partial a\partial b} & \frac{\partial^2\varepsilon(k)}{\partial a\partial d}\\ \frac{\partial^2\varepsilon(k)}{\partial b\partial a} & \frac{\partial^2\varepsilon(k)}{\partial b^2} & \frac{\partial^2\varepsilon(k)}{\partial b\partial d}\\ \frac{\partial^2\varepsilon(k)}{\partial d\partial a} & \frac{\partial^2\varepsilon(k)}{\partial d\partial b} & \frac{\partial^2\varepsilon(k)}{\partial d^2}\end{bmatrix}$$

其中：

$$\frac{\partial^2\varepsilon(k)}{\partial a_j\partial d_m}=\frac{\partial^2\varepsilon(k)}{\partial d_m\partial a_j}=y(k-j-m)$$

$$\frac{\partial^2\varepsilon(k)}{\partial b_j\partial d_m}=\frac{\partial^2\varepsilon(k)}{\partial d_m\partial b_j}=-u(k-j-m)$$

估计准则

$$J = \frac{\sum_{k=n+1}^{n+N} e(k)}{2}$$

梯度：

$$\frac{\partial J}{\partial \hat{\theta}} = \sum_{k=n+1}^{n+N} e(k) \frac{\partial e(k)}{\partial \hat{\theta}} = q(N)$$

海赛矩阵：

$$\frac{\partial^2 J}{\partial \hat{\theta}^2} = \sum_{k=n+1}^{n+N} \left[\frac{\partial e(k)}{\partial \hat{\theta}} \left(\frac{\partial e(k)}{\partial \hat{\theta}} \right)^T + e(k) \frac{\partial^2 e(k)}{\partial \hat{\theta}^2} \right] = R(N)$$

迭代公式

牛顿 -拉卜森公式:

$$\hat{\boldsymbol{\theta}}_{\mathrm{r}} = \hat{\boldsymbol{\theta}}_{\mathrm{r}-1} - \mathrm{R}^{-1}(\mathrm{N})\mathrm{q}(\mathrm{N})$$

q 与 R 的递推公式:

$$\begin{aligned}
\mathrm{q}(\mathrm{N}) &= \mathrm{q}(\mathrm{N}-1) + \mathrm{e}(\mathrm{n}+\mathrm{N})\frac{\partial \mathrm{e}(\mathrm{n}+\mathrm{N})}{\partial \hat{\theta}} \\
\mathrm{R}(\mathrm{N}) &= \mathrm{R}(\mathrm{N}-1) + \frac{\partial \mathrm{e}(\mathrm{n}+\mathrm{N})}{\partial \hat{\theta}}\left(\frac{\partial \mathrm{e}(\mathrm{n}+\mathrm{N})}{\partial \hat{\theta}}\right)^{\mathrm{T}} \\
&\quad + \mathrm{e}(\mathrm{n}+\mathrm{N})\frac{\partial^2 \mathrm{e}(\mathrm{n}+\mathrm{N})}{\partial \hat{\theta}^2}
\end{aligned}$$