Maximum likelihood identification

Xing Chao

Schoool of Astronautics, Northwestern Polytechnical University

Identification criterion The probability of the observed value is the largest

likelihood function Probability density function of observed values

Features of Method

1. It can be applied to correlated noises $\xi(\mathrm{k})$;
2. When the system SNR is relatively small, it has a better estimation effect ;
3. The algorithm has good stability ;
4. There is recursive calculation method ;
5. It is widely used in practical engineering

Likelihood function

A discrete stochastic process $\{\mathrm{V_k}\}$ is related to θ, parameters to be identified. The probability distribution density $\mathrm{f(V_k|\theta)}$ is known. If n independent observations are measured as $\mathrm{V_1, V_2, \cdots, V_n}$, Its distribution density is $\mathrm{f(V_1|\theta), \cdots, f(V_n|\theta)}$, Define likelihood function L as:

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

Maximum likelihood estimation

The principle of identifying $\theta$is to make the Lreach the maximum:

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

Usually the logarithm of the L is taken:

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

result is:

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

Obtained θ is maximum likelihood estimation $\hat{\theta}_{\mathrm{ML}}$

Maximum likelihood identification of difference equations: system model (white noise case)

System difference equation:

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

where $\xi(k)$ is Gauss white noise sequence and is uncorrelated to u(k). Represented in vector form:

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

Maximum likelihood identification of difference equations: residual (white noise)

System estimation residual:

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

Since $\xi(\mathrm{k})$ is gaussion white noise, it is assumed that e(k) is also the Gauss white noise. Let e(k) variance be σ^2. Probability density function is:

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

likelihood function is,

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

Maximum likelihood identification of difference equations: likelihood function (white noise case)

According to the principle of maximum likelihood identification:

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

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

solve equations :

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

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

It is shown that in the special case of $\xi(k)$ being Gauss white noise sequence, the maximum likelihood identification has the same result as the general least squares identification.

Introduction
Maximum likelihood principle
Maximum likelihood identification
White noise
Colored noise situation
Recursive Maximum Likelihood Method
Approximately Recursive Maximum Likelihood Method
Newton-Raphson resursive formular

Maximum likelihood identification of difference equations: a system model (colored noise case)

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

其中：

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

Maximum likelihood identification of difference equations: prediction error (colored noise case)

prediction error:

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

vector form is:

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

where:

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

Maximum likelihood identification of difference equations: likelihood function (colored noise case)

Since $\varepsilon(k)$ is Gauss white noise, e(k) can be assumed to be the zero mean Gauss white noise. Then the likelihood function is:

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

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

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

Maximum likelihood identification of difference equations: criteria (colored noise case)

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

- J is a quadratic form function of parameter $a_1, a_2, \cdots, a_n; b_0, b_1, \cdots, b_n; c_1, c_2, \cdots, c_n$ ◦
- The maximum L of $\hat{\theta}$, is equivalent to find $\hat{\theta}$ with minimum J under the constraint conditions,

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

The difference of maximum likelihood identification equation: Newton-Raphson

Newton-Raphson iteration:

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

where:

- $\frac{\partial \mathrm{J}}{\partial \theta}$ is gradient
- $\frac{\partial^2 \mathrm{J}}{\partial \theta^2}$is Hessian matrix

Newton-Raphson iterative procedure: initial value selection

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

where:

- $\hat{a}_1, \cdots, \hat{a}_n, \hat{b}_0, \cdots, \hat{b}_n$ can be obtained by least squares method
- $\hat{c}_1, \cdots, \hat{c}_n$ can be zeros or arbitrarily assumed

Newton-Raphson iterative procedure: prediction errors

- Prediction error, cost function and estimation of error variance:

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

Newton-Raphson iterative procedure: gradient matrix and Hessian matrix

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

where:

$$
\frac{\partial \mathrm{e}(\mathrm{k})}{\partial \theta} = \left[\frac{\partial \mathrm{e}(\mathrm{k})}{\partial \mathrm{a}_1}, \cdots, \frac{\partial \mathrm{e}(\mathrm{k})}{\partial \mathrm{a}_\mathrm{n}}, \frac{\partial \mathrm{e}(\mathrm{k})}{\partial \mathrm{b}_0}, \cdots, \frac{\partial \mathrm{e}(\mathrm{k})}{\partial \mathrm{b}_\mathrm{n}}, \frac{\partial \mathrm{e}(\mathrm{k})}{\partial \mathrm{c}_1}, \cdots, \frac{\partial \mathrm{e}(\mathrm{k})}{\partial \mathrm{c}_\mathrm{n}}\right]^{\mathrm{T}}
$$

Newton-Raphson iterative procedure:new estimates

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

stop criterion：

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

where δ can be small number，i.e. $\delta = 10^{-4}$.

System difference equation

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

where:

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

can be writen as:

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

The cost function is approximately represented as a quadratic form:

$$\begin{aligned} \mathrm{J_N} &= \sum_{\mathrm{k=n+1}}^{\mathrm{n+N}} \varepsilon^2(\mathrm{k}) \\ &\approx (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\mathrm{N}})^{\mathrm{T}} \mathrm{p}_{\mathrm{N}}^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\mathrm{N}}) + \beta_{\mathrm{N}} \end{aligned}$$

Using Taylor series expand $\varepsilon(\mathrm{k})$ at estimated $\hat{\theta}$:

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

where:

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

Result:

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

where:

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

Let：

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

result：

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

where：

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

$\beta_{\mathrm{N+1}}$ is known, so $\mathrm{J}_{\mathrm{N+1}}$ is minimum when:

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

update $\mathrm{P}_{\mathrm{N}+1}, \hat{\theta}_{\mathrm{N}+1}$

By using the inverse lemma of matrices, we obtain:

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

update Ψ_{N+1}

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

where:

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

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

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

$$D = [y_{n+N}, 0, \cdots, 0, -u_{n+N+1}, 0, \cdots, 0, -e_{n+N}, 0, \cdots, 0]^T$$

deduction of A, B, C

from

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

get:

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

then:

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

Resursive formular using Newton-Raphson method: system difference equation

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

where:

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

parameter vectors are:

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

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

rewrite system difference equation as:

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

result:

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

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

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

where:

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

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

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

where:

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

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

Estimation criterion

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

gradient:

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

Hessian Matrix:

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

iteration formular

Newton-Raphson formular:

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

iteration formular for q and R :

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