极大似然法辨识

邢超

<MLE.1>

1 简介

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

<MLE.2>

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

<MLE.3>

2 极大似然原理

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：

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

<MLE.4>

极大似然估计

辨识 θ 的原则就是使得 L 达到极大值，即:

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

通常对 L 取对数：

$$\ln L=\ln f(V_1|\theta)+\cdots+\ln f(V_n|\theta)$$

求解：

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

所得 θ 即为极大似然估计 $\hat{\theta}_{ML}$

<MLE.5>

3 极大似然辨识

3.1 白噪声情况

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：

$$Y_N = \Phi_N \theta + \xi$$

<MLE.6>

差分方程的极大似然辨识：残差（白噪声情况）

系统估计残差：

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

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

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

<MLE.7>

差分方程的极大似然辨识：似然函数（白噪声情况）

似然函数为：

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

<MLE.8>

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：

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

解上述方程有：

$$\begin{aligned} \hat{\theta}_{ML} &= (\Phi_N^T \Phi_N)^{-1} \Phi_N^T Y_N \\ \sigma^2 &= \frac{(Y_N - \Phi_N\hat{\theta})^T (Y_N - \Phi_N\hat{\theta})}{N} \end{aligned}$$

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。

<MLE.9>

3.2 有色噪声情况

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：

$$\begin{aligned}a(z^{-1}) &= 1+a_1z^{-1}+\cdots+a_nz^{-n}\\ b(z^{-1}) &= b_0+b_1z^{-1}+\cdots+b_nz^{-n}\\ c(z^{-1}) &= 1+c_1z^{-1}+\cdots+c_nz^{-n}\end{aligned}$$

<MLE.10>

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：

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

<MLE.11>

差分方程的极大似然辨识：似然函数（有色噪声情况）

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：

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

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

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

<MLE.12>

差分方程的极大似然辨识：准则（有色噪声情况）

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

- 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。

<MLE.13>

差分方程的极大似然辨识：牛顿 -拉卜森法

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

<MLE.14>

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

<MLE.15>

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：

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

<MLE.16>

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：

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

<MLE.17>

牛顿 -拉卜森迭代步骤：计算新的估计值

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

停上条件：

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

其中 δ 可取较小的数，如 $\delta = 10^{-4}$。

<MLE.18>

4 递推极大似然法

4.1 近似的递推极大似然法

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：

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

<MLE.19>

二次型指标函数

将指标函数用二次型近似表示：

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

利用泰勒级数将 $\varepsilon(k)$ 在估值 $\hat{\theta}$ 展开： <MLE.20>

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

其中：

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

<MLE.21>

可得：

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

其中：

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

<MLE.22>

设：

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

得：

$$
\begin{aligned}
J_{N+1}(\theta) &= \Delta^T(P_N^{-1}+\Psi_{N+1}\Psi_{N+1}^T)\Delta \\
&\quad +2e_{N+1}\Psi_{N+1}^T\Delta+e_{N+1}^2+\beta_N \\
&= (\Delta+r_{N+1})^T P_{N+1}^{-1}(\Delta+r_{N+1})+\beta_{N+1}
\end{aligned}
$$

其中：

$$
\begin{aligned}
P_{N+1}^{-1} &= P_N^{-1}+\Psi_{N+1}\Psi_{N+1}^T \\
r_{N+1} &= P_{N+1}\Psi_{N+1}e_{N+1} \\
\beta_{N+1} &= e_{N+1}^2+\beta_N-e_{N+1}\Psi_{N+1}^T P_{N+1}\Psi_{N+1}e_{N+1}
\end{aligned}
$$

β_{N+1} 为已知值，所以

$$\hat{\theta}_{N+1}=\hat{\theta}_N-r_{N+1}$$

时使 J_{N+1} 最小。

<MLE.23>

更新 $P_{N+1},\hat{\theta}_{N+1}$

利用矩阵求逆引理，得：

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

<MLE.24>

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：

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

<MLE.25>

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

<MLE.26>

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：

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

<MLE.27>

4.2 牛顿 -拉卜森递推公式

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：

$$\begin{aligned} a(z^{-1}) &= 1+a_1z^{-1}+\cdots+a_nz^{-n} \\ b(z^{-1}) &= b_0+b_1z^{-1}+\cdots+b_nz^{-n} \\ d(z^{-1}) &= 1+d_1z^{-1}+\cdots+d_nz^{-n} \end{aligned}$$

参数向量为：

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

<MLE.28>

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

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：

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

<MLE.29>

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

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

其中：

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

<MLE.30>

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

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

其中：

$$\begin{aligned}\frac{\partial^2\varepsilon(k)}{\partial a_j\partial d_m} &= \frac{\partial^2\varepsilon(k)}{\partial d_m\partial a_j}=y(k-j-m)\\ \frac{\partial^2\varepsilon(k)}{\partial b_j\partial d_m} &= \frac{\partial^2\varepsilon(k)}{\partial d_m\partial b_j}=-u(k-j-m)\end{aligned}$$

<MLE.31>

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：

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

<MLE.32>

迭代公式

牛顿 -拉卜森公式:

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

q 与 R 的递推公式：

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

<MLE.33>