第12卷 第4期 自动化学报 Vol. 12, No. 4
1986年10月 ACTA AUTOMATICA SINICA Oct., 1986

近似极大似然法的一个注记[1)]

袁 著 祉
(南 开 大 学)

摘　　要

本文证明了近似极大似然法与文献[2]提出的改进增广最小二乘法的等价性，计算机仿真结果表明了改进增广最小二乘法的优点.

一、引　　言

在建立实际对象的数学模型时，经常遇到受有色噪声干扰的系统

$$A(q^{-1})y(k) = B(q^{-1})u(k) + e(k). \tag{1}$$

其中 $e(k)$ 是有色噪声，$A(q^{-1})$ 与 $B(q^{-1})$ 是向后移位算子 q^{-1} 的多项式

$$A(q^{-1}) = 1 + a_1q^{-1} + \cdots + a_nq^{-n},$$
$$B(q^{-1}) = b_1q^{-1} + \cdots + b_nq^{-n}.$$

在估计差分方程(1)的参数 $a_1, \cdots, a_n, b_1, \cdots, b_n$ 时，用最小二乘法会得到有偏估计. 多年来，人们提出了许多方法估计有色噪声干扰下的系统参数，如广义最小二乘法、辅助变量法、相关最小二乘法、增广最小二乘法、近似极大似然法[1]与改进增广最小二乘法[2]等. 如此众多的方法，使用户难以选择. 因此，学者们用理论分析与数值仿真将这些方法进行了比较，表明了增广最小二乘法简单易行、应用广泛. 七十年代中期，瑞典学者 Ljung 等人用常微分方程稳定性法 (ODE)，分析了随机递推算法的收敛性，给出了增广最小二乘法不收敛的例子，提出了既保证收敛性又简单的改进增广最小二乘法[2].

本文指出了 Goodwin 与 Payne 合著的书[1] 中关于近似极大似然法论述的两处小的错误.

二、近似极大似然法与改进增广最小二乘法的等价性

当有色噪声 $e(k)$ 可以写成 $e(k) = C(q^{-1})\varepsilon(k)$ 时，系统模型(1)改写为

$$A(q^{-1})y(k) = B(q^{-1})u(k) + C(q^{-1})\varepsilon(k). \tag{2}$$

其中 $\{\varepsilon(k)\}$ 是白噪声序列，$C(q^{-1})$ 是 q^{-1} 的多项式

$$C(q^{-1}) = 1 + c_1q^{-1} + \cdots + c_nq^{-n}.$$

本文于1984年6月26日收到.
1) 中国自然科学基金资助.

为了证明等价性的方便，简述增广最小二乘法、近似极大似然法与改进增广最小二乘法.

1. 增广最小二乘法

1968 年 Panuska 提出了增广最小二乘法，同时估计系统参数 $A(q^{-1}),B(q^{-1})$ 与噪声参数 $C(q^{-1})$.

若噪声 $\varepsilon(k)$ 可以量测，则模型(2)可以写成

$$y(k)=\boldsymbol{x}^T(k)\boldsymbol{\theta}+\varepsilon(k). \tag{3}$$

其中

$$\boldsymbol{x}^T(k)=[-y(k-1),\cdots,-y(k-n),u(k-1),\cdots,u(k-n),\varepsilon(k-1),\cdots,\varepsilon(k-n)], \tag{4}$$

$$\boldsymbol{\theta}^T=[a_1,\cdots,a_n,b_1,\cdots,b_n,c_1,\cdots,c_n]. \tag{5}$$

由于 $\varepsilon(k)$ 是白噪声，用最小二乘法可以得到增广参数向量 $\boldsymbol{\theta}$ 的无偏估计.

实际上，序列 $\{\varepsilon(k)\}$ 是未知的，Panuska 给出一个计算 $\varepsilon(k)$ 的方法

$$\hat{\varepsilon}(k)=y(k)-\hat{\boldsymbol{x}}^T(k)\hat{\boldsymbol{\theta}}_{k-1}, \tag{6}$$

$$\hat{\boldsymbol{x}}^T(k)=[-y(k-1),\cdots,-y(k-n),u(k-1),\cdots,u(k-n),\hat{\varepsilon}(k-1),\cdots,\hat{\varepsilon}(k-n)]. \tag{7}$$

增广最小二乘法递推公式为

$$\hat{\boldsymbol{\theta}}_{k+1}=\hat{\boldsymbol{\theta}}_k+\boldsymbol{P}_k\hat{\boldsymbol{x}}(k+1)[1+\hat{\boldsymbol{x}}^T(k+1)\boldsymbol{P}_k\hat{\boldsymbol{x}}(k+1)]^{-1}\hat{\varepsilon}(k+1), \tag{8}$$

$$\boldsymbol{P}_{k+1}=\boldsymbol{P}_k-\boldsymbol{P}_k\hat{\boldsymbol{x}}(k+1)[1+\hat{\boldsymbol{x}}^T(k+1)\boldsymbol{P}_k\hat{\boldsymbol{x}}(k+1)]^{-1}\boldsymbol{P}_k. \tag{9}$$

这种方法在大量的实际应用中没有发现不收敛的情况，但是 Ljung 等人从收敛性理论出发，给出了不收敛的例子[2].

2. 近似极大似然法[1]

为了估计模型(2)的参数 $A(q^{-1}),B(q^{-1})$ 与 $C(q^{-1})$，当 $\{\varepsilon(k)\}$ 是高斯白噪声时，可以用极大似然法寻求参数 $\boldsymbol{\theta}$ 的极大似然估计，使得目标函数

$$J_N(\boldsymbol{\theta})=\sum_{k=1}^{N}[w_k(\boldsymbol{\theta})]^2 \tag{10}$$

达到极小，这里 $\{w_k(\boldsymbol{\theta})\}$ 是

$$w_k(\boldsymbol{\theta})=[C(q^{-1})]^{-1}[A(q^{-1})y(k)-B(q^{-1})u(k)]. \tag{11}$$

事实上，这里把 $\varepsilon(k)$ 视为 $\boldsymbol{\theta}$ 的函数 $w_k(\boldsymbol{\theta})$. 由上式可见，$w_k(\boldsymbol{\theta})$ 不是 $\boldsymbol{\theta}$ 的线性函数，因而 $J_N(\boldsymbol{\theta})$ 不是 $\boldsymbol{\theta}$ 的二次函数. 对于这样的 $J_N(\boldsymbol{\theta})$ 计算极小值是困难的. 于是，把 $w_k(\boldsymbol{\theta})$ 泰乐展开取 $\boldsymbol{\theta}$ 的一次项，$J_N(\boldsymbol{\theta})$ 就可以用 $\boldsymbol{\theta}$ 的二次函数近似. 在假设

$$\boldsymbol{\phi}(k)=\partial\hat{w}(k)/\partial\boldsymbol{\theta} \tag{12}$$

与

$$\hat{w}(k)\simeq\hat{\varepsilon}(k)=y(k)-\hat{\boldsymbol{x}}^T(k)\hat{\boldsymbol{\theta}}_{k-1} \tag{13}$$

条件下，可以导出近似极大似然法递推公式

$$\hat{\boldsymbol{\theta}}_{k+1}=\hat{\boldsymbol{\theta}}_k-\boldsymbol{P}_k\boldsymbol{\phi}(k+1)[1+\boldsymbol{\phi}^T(k+1)\boldsymbol{P}_k\boldsymbol{\phi}(k+1)]^{-1}\hat{w}(k+1), \tag{14}$$

$$\boldsymbol{P}_{k+1}=\boldsymbol{P}_k-\boldsymbol{P}_k\boldsymbol{\phi}(k+1)[1+\boldsymbol{\phi}^T(k+1)\boldsymbol{P}_k\boldsymbol{\phi}(k+1)]^{-1}\boldsymbol{\phi}^T(k+1)\boldsymbol{P}_k. \tag{15}$$

其中，向量 $\boldsymbol{\phi}(k+1)$ 由下列递推公式计算

$$\boldsymbol{\phi}(k+1)=\left[\begin{array}{ccc|ccc|ccc} -\hat{c}_1 & \cdots & -\hat{c}_n & & & & & & \\ 1 & \ddots & \vdots & & & & & & \\ & 1 & 0 & & & & & & \\ \hline & & & -\hat{c}_1 & \cdots & -\hat{c}_n & & & \\ & & & 1 & \ddots & \vdots & & & \\ & & & & 1 & 0 & & & \\ \hline & & & & & & -\hat{c}_1 & \cdots & -\hat{c}_n \\ & & & & & & 1 & \ddots & \vdots \\ & & & & & & & 1 & 0 \end{array}\right]\boldsymbol{\phi}(k)+\left[\begin{array}{c} y(k) \\ 0 \\ \vdots \\ \hline -u(k) \\ 0 \\ \vdots \\ \hline -\hat{\varepsilon}(k) \\ 0 \\ \vdots \end{array}\right]. \tag{16}$$

注. 这里的公式(14)与文献[1]中公式(7.4.59)不同,纠正了式(7.4.59)的错误,详见附录.

3. 改进增广最小二乘法[2]

Ljung 等人导出的改进增广最小二乘法为

$$\hat{\boldsymbol{\theta}}_{k+1}=\hat{\boldsymbol{\theta}}_k+\boldsymbol{P}_k\hat{\boldsymbol{z}}(k+1)[1+\hat{\boldsymbol{z}}^T(k+1)\boldsymbol{P}_k\hat{\boldsymbol{z}}(k+1)]^{-1}\hat{\varepsilon}(k+1), \tag{17}$$

$$\boldsymbol{P}_{k+1}=\boldsymbol{P}_k-\boldsymbol{P}_k\hat{\boldsymbol{z}}(k+1)[1+\hat{\boldsymbol{z}}^T(k+1)\boldsymbol{P}_k\hat{\boldsymbol{z}}(k+1)]^{-1}\hat{\boldsymbol{z}}^T(k+1)\boldsymbol{P}_k. \tag{18}$$

其中 $\hat{\varepsilon}(k)=y(k)-\hat{\boldsymbol{z}}^T(k)\hat{\boldsymbol{\theta}}_{k-1}$,

$$\begin{aligned}\hat{\boldsymbol{z}}^T(k)=[&-\tilde{y}(k-1),\cdots,-\tilde{y}(k-n),\tilde{u}(k-1),\cdots,\tilde{u}(k-n),\\ &\tilde{\varepsilon}(k-1),\cdots\tilde{\varepsilon}(k-n)].\end{aligned} \tag{19}$$

式(19)中的 $\tilde{y}(k)$,$\tilde{u}(k)$ 与 $\tilde{\varepsilon}(k)$ 由下列公式计算

$$\begin{aligned}\tilde{y}(k)&=y(k)-\sum_{i=1}^{n}\hat{c}_i\tilde{y}(k-i),\\ \tilde{u}(k)&=u(k)-\sum_{i=1}^{n}\hat{c}_i\tilde{u}(k-i),\\ \tilde{\varepsilon}(k)&=\hat{\varepsilon}(k)-\sum_{i=1}^{n}\hat{c}_i\tilde{\varepsilon}(k-i).\end{aligned} \tag{20}$$

这里 $\hat{c}_i$ 是 $C(q^{-1})$ 在时刻 k 上的估计值.

4. 近似极大似然法与改进增广最小二乘法等价性的证明

由(13)式可知 $\hat{w}(k)\simeq\hat{\varepsilon}(k)$,如果能够证明 $\hat{\boldsymbol{z}}(k)=-\boldsymbol{\phi}(k)$,那末,两个方法的等价性得证.

事实上,计算 $\boldsymbol{\phi}(k)$ 的公式(16)可以写作

$$\begin{aligned}\phi_{k+1}(1)&=y(k)-\sum_{i=1}^{n}\hat{c}_i\phi_k(i),\\ \phi_{k+1}(n+1)&=-u(k)-\sum_{i=1}^{n}\hat{c}_i\phi_k(n+i),\\ \phi_{k+1}(2n+1)&=-\hat{\varepsilon}(k)-\sum_{i=1}^{n}\hat{c}_i\phi_k(2n+i),\\ \phi_{k+1}(i)&=\phi_k(i-1),\quad i\neq 1,n+1,\ 2n+1.\end{aligned} \tag{21}$$

不失一般性，从 $k=n+1$ 开始递推计算，分别置 $\hat{\boldsymbol{z}}$ 与 $\boldsymbol{\phi}$ 的初始值为

$$\hat{\boldsymbol{z}}^T(n+1)=[-\tilde{y}(n),\cdots-\tilde{y}(1),\tilde{\boldsymbol{u}}(n),\cdots,\tilde{\boldsymbol{u}}(1),\tilde{\varepsilon}(n),\cdots,\tilde{\varepsilon}(1)], \tag{22}$$

$$\boldsymbol{\phi}^T(n+1)=[\tilde{y}(n),\cdots,\tilde{y}(1),-\tilde{\boldsymbol{u}}(n),\cdots,-\tilde{\boldsymbol{u}}(1),-\tilde{\varepsilon}(n),\cdots,-\tilde{\varepsilon}(1)]. \tag{23}$$

对于任意 $k\geqslant n+1$，由式(19)，(20)与(22)可得(24)式，由式(21)与(23)可得(25)式：

$$\hat{\boldsymbol{z}}(k+1)=\begin{bmatrix} -y(k)+\sum_{i=1}^{n}\hat{c}_i\tilde{y}(k-i) \\ -\tilde{y}(k-1) \\ \vdots \\ u(k)-\sum_{i=1}^{n}\hat{c}_i\tilde{\boldsymbol{u}}(k-i) \\ \tilde{\boldsymbol{u}}(k-1) \\ \vdots \\ \tilde{\varepsilon}(k)-\sum_{i=1}^{n}\hat{c}_i\tilde{\varepsilon}(k-i) \\ \tilde{\varepsilon}(k-1) \\ \vdots \end{bmatrix}. \tag{24}$$

$$\boldsymbol{\phi}(k+1)=\begin{bmatrix} y(k)-\sum_{i=1}^{n}\hat{c}_i\tilde{y}(k-i) \\ \tilde{y}(k-1) \\ \vdots \\ -u(k)+\sum_{i=1}^{n}\hat{c}_i\tilde{\boldsymbol{u}}(k-i) \\ -\tilde{\boldsymbol{u}}(k-1) \\ \vdots \\ -\tilde{\varepsilon}(k)+\sum_{i=1}^{n}\hat{c}_i\tilde{\varepsilon}(k-i) \\ -\tilde{\varepsilon}(k-1) \\ \vdots \end{bmatrix}. \tag{25}$$

由(24)与(25)两式可知 $\hat{\boldsymbol{z}}(k)=-\boldsymbol{\phi}(k)$. 从而证明了由(13)—(16)式组成的近似极大似然法与由(17)—(20)式组成的改进增广最小二乘法是等价的.

三、数 字 仿 真

(1) 对文献[2]中的两个例子，分别用增广最小二乘法与改进增广最小二乘法估计模型参数，得到了下述结果：在递推辨识 2000 步以后，增广最小二乘法得到的 $\hat{c}(q^{-1})$ 参数仍然远离真值，而改进增广最小二乘法加快了 $\hat{c}(q^{-1})$ 参数向真值收敛的速度.

(2) 用文献[1]中有错误的近似极大似然法辨识的参数远离真值，改错后就能收敛到真值.

(3) 在大港发电厂，建立过热蒸汽温度模型时，分别采用增广最小二乘法与改进增广

最小二乘法进行辨识,得到的模型分别是

$$y(k+2) = 0.7y(k+1) + 0.25y(k) - 0.24 \times 10^{-9}u(k+1) - 0.915 \times 10^{-10}u(k), \tag{26}$$

$$y(k+2) = 0.2637y(k+1) + 0.7367y(k) - 0.0046u(k+1) - 0.00113u(k). \tag{27}$$

容易看出,用增广最小二乘法得到的(26)式不符合实际工况,式中控制作用 u 对输出 y 的影响太小;而用改进增广最小二乘法得到的(27)式比较适合工艺情况,详见文献[3].

附 录

文献[1]中,(7.4.52)式与(7.4.53)式有错.

$$J_{N+1}(\theta) \simeq \Delta^T(P_N^{-1} + \phi_{N+1}\phi_{N+1}^T)\Delta + 2\hat{\omega}_{N+1}\phi_{N+1}^T\Delta + \beta_N + \hat{\omega}_{N+1}^2. \tag{7.4.52}$$

(右方最后一项 $\hat{\omega}_{N+1}^2$ 原文中遗漏). (7.4.52)式配完全平方得

$$J_{N+1}(\theta) \simeq (\Delta + r_{N+1})^T P_{N+1}^{-1}(\Delta + r_{N+1}) + \beta_{N+1}. \tag{7.4.53}$$

(原文中误把 $(\Delta + r_{N+1})$ 写作 $(\Delta - r_{N+1})$),其中

$$P_{N+1}^{-1} = P_N^{-1} + \phi_{N+1}\phi_{N+1}^T, \tag{7.4.54}$$

$$r_{N+1} = P_{N+1}\phi_{N+1}\hat{\omega}_{N+1}. \tag{7.4.55}$$

将式(7.4.54),(7.4.55)代入(7.4.53)式可得(7.4.52)式,可见修改后的(7.4.53)式是正确的.

于是(7.4.53)式在 Δ 为下述值时达到极小值.

$$\Delta_{N+1} = -r_{N+1}, \tag{7.4.58}$$

即

$$\hat{\theta}_{N+1} = \hat{\theta}_N - P_N\phi_{N+1}(1 + \phi_{N+1}^T P_N \phi_{N+1})^{-1}\hat{\omega}_{N+1}. \tag{7.4.59}$$

参 考 文 献

[1] Goodwin, G.G., Payne, R.L., Dynamic System Identification, Academic Press. (1977).

[2] Ljung, L., et al., Counterxamples to General Convergence of a Commonly Used Recursive Identification Method, *IEEE Trans. Auto. Control*, AC-**20**(1975), 643—652.

[3] 袁著祉等,热电厂过热器建模与适应控制器,控制理论与应用,第1卷第4期(1984),135—139.

A NOTE FOR APPROXIMATE MAXIMUM LIKELIHOOD METHOD

YUAN ZHUZHI

(Nankai University)

ABSTRACT

It is shown in this paper that the approximate maximum likelihood method is equivalent to the improved extended least square method[2]. Advantages of the improved extended least square method and the extended least square method have been demonstrated by simulation.