第6卷 第4期　　控制理论与应用　　Vol. 6, No. 4
1989年12月　　CONTROL THEORY AND APPLICATIONS　　Dec., 1989

多层递阶预报参数估值初值选择的循环方法

郑宇红　　　韩光文

（华中理工大学系统工程研究所，武汉）（武汉工业大学管理工程系）

摘　　要

本文提出了多层递阶预报参数估值初值选择的新方法——循环法，并对其收敛性及不动点性质进行了初步探讨。实验结果表明，该方法应用效果较好。

关键词：多层递阶预报；初值选择；循环法；收敛性；不动点。

一、问 题 的 提 出

文[1]讨论了多层递阶预报参数估值的初值选择问题，指出在一定条件下该方法的预报精度不受参数估值初值选择的影响，而依赖于时变参数的预报算法的精度。然而实验结果表明，初值选择并非随心所欲，也不象文[1，2]提出的初值选择方法那样仅使参数估值序列具有某种唯一性。实际上，初值的选择在一定程度上决定了在参数估值序列基础上建模的难易程度，影响了时变参数的预报精度，最终将影响多层递阶预报的预报精度。

我们认为，初值最理想的选择是使参数估值序列真实地反映系统时变参数的变化规律。该提法将为后面的时变参数预报提供很大的方便，但在理论分析上会造成一定的困难。不过初值选择至少应做到这一点，即若系统参数为非时变时，应使参数估值序列反映出非时变的特性。

基于上述思想，我们提出了初值选择的循环方法。

二、初值选择的循环方法

对线性单输出系统

$$y(k)=\phi^T(k)\theta(k)+v(k)$$

其中　$\theta(k)$为参数列向量，$v(k)$为随机噪声。

由公式

$$\hat{\theta}(k)=\hat{\theta}(k-1)+\frac{1}{\|\phi(k)\|_2^2}\phi(k)\{y(k)-\phi^T(k)\hat{\theta}(k-1)\} \tag{1}$$

$$k=2,\cdots,N$$

本文于1988年6月1日收到。1989年3月10日收到修改稿。

从$\hat{\theta}(1)$向前递推到$\hat{\theta}(N)$。然后由公式

$$\hat{\theta}(k)=\hat{\theta}(k+1)+\frac{1}{\|\phi(k)\|_2^2}\phi(k)\{y(k)-\phi^T(k)\hat{\theta}(k+1)\} \tag{2}$$

$$k=N-1,\cdots,1$$

从$\hat{\theta}(N)$向后递推到$\hat{\theta}(1)$，这样便完成了一次循环。

多次重复此循环，若相继两次循环所得出参数估值序列的偏离度小于预定的值，则终止运算，最后一次循环向前递推的结果作为参数的估值序列。

文〔3〕曾提出一种使参数估值序列唯一的初值选择方法，它和本文提出的方法类似。值得指出的是，该法仅作为一种经验方法提出，并未作进一步的理论分析。

下面我们将分别对循环方法的收敛问题和不动点的性质作一些理论分析。

三、循环方法的收敛问题

记 $A(k)=\dfrac{\phi(k)\phi^T(k)}{\|\phi(k)\|_2^2}$， $B(k)=\dfrac{\phi(k)y(k)}{\|\phi(k)\|_2^2}$

则式（1）、（2）可变为

$$\hat{\theta}(k)=(1-A(k))\hat{\theta}(k-1)+B(k) \tag{3}$$

$$\hat{\theta}(k)=(1-A(k))\hat{\theta}(k+1)+B(k) \tag{4}$$

给定$\hat{\theta}^l(1)$，注意到式（3）有

$$\hat{\theta}(N)=\prod_{k=N}^{2}(1-A(k))\hat{\theta}^l(1)+\sum_{k=2}^{N-1}\prod_{i=N}^{k+1}(1-A(i))B(k)+B(N) \tag{5}$$

同理由式（4）有

$$\hat{\theta}^{l+1}(1)=\prod_{k=1}^{N-1}(1-A(k))\hat{\theta}(N)+\sum_{k=2}^{N-1}\prod_{i=1}^{k-1}(1-A(i))B(k)+B(1) \tag{6}$$

由式（5）、（6）有

$$\hat{\theta}^{l+1}(1)=E(N\ \hat{\theta}^l(1)+F(N) \tag{7}$$

其中 $E(N)=\prod_{k=1}^{N-1}(1-A(k))\prod_{k=N}^{2}(1-A(k))$

$$F(N)=\prod_{k=1}^{N-1}(1-A(k))\sum_{k=2}^{N-1}\prod_{i=N}^{k+1}(1-A(i))B(k)+\prod_{k=1}^{N-1}(1-A(k))B(N)$$

$$+\sum_{k=2}^{N-1}\prod_{i=1}^{k-1}(1-A(i))B(k)+B(1)$$

显然式（7）是从R^n到R^n的一个变换（n为参数个数），记为T：$R^n\longrightarrow R^n$，

在欧几里德空间R^n中定义距离$d(X_1, X_2)=\|X_1-X_2\|_p$，易证度量空间（$R^n$, d）构成完备空间。

$$d(TX_1, TX_2)=\|E(N)(X_1-X_2)\|_p\leqslant\|E(N)\|_p d(X_1, X_2)$$

若$\|E(N)\|_p<1$，则T是R^n上的压缩映射。由压缩映射原理有如下结论：

若$\|E(N)\|_p<1$，则变换T有唯一的不动点$\hat{\theta}(1)$，$\hat{\theta}(1)$满足方程$\hat{\theta}(1)=E(N)\hat{\theta}(1)+F(N)$。因而从任意选定的初值出发，逐次进行上述循环，必收敛到不动点，且$\|E(N)\|_p$越小收敛速度越快。

下面讨论$\|E(N)\|_p<1$成立的条件。

设　$\phi^T(k)=(x_1(k),\cdots, x_n(k))_{1\times n}$，$A'(k)=1-A(k)\quad k=1,\cdots,N$

$$A(k)\text{的秩}\operatorname{rank}(A(k))=\operatorname{rank}\left(\frac{\phi(k)\phi^T(k)}{\|\phi(k)\|_2^2}\right)=1$$

$$A(k)\text{的迹}\operatorname{trace}(A(k))=\operatorname{trace}\left(\frac{\phi(k)\phi^T(k)}{\|\phi(k)\|_2^2}\right)=1$$

故　$A(k)$的特征值 $\lambda_i(A(k))=\begin{cases}0 & i=1,\cdots,n-1\\ 1 & i=n\end{cases}$

$\because$　　$\lambda(A'(k))=1-\lambda(A(k))$且$A'(k)$为实对称阵

$\therefore$　　$\|A'(k)\|_2=\max\limits_i\{\lambda_i(A'(k))\}=1\qquad k=1,\cdots,N$

$\therefore$　$\|E(N)\|_2=\|\prod\limits_{k=1}^{N-1}A'(k)\prod\limits_{k=N}^{2}A'(k)\|_2\leqslant\prod\limits_{k=1}^{N-1}\|A'(k)\|_2\prod\limits_{k=N}^{2}\|A'(k)\|_2=1$

下面就$n=2$的特殊情形作进一步的分析。

此时　$$A'(k)=\frac{1}{x_1^2(k)+x_2^2(k)}\begin{pmatrix}x_2^2(k) & -x_1(k)x_2(k)\\ -x_1(k)x_2(k) & x_1^2(k)\end{pmatrix}$$

若令　$\phi'^T(k)=(x_2(k), -x_1(k))$，则有$A'(k)=\dfrac{\phi'(k)\phi'^T(k)}{\|\phi'(k)\|_2^2}$

而　$$E(N)=\prod_{k=1}^{N-1}A'(k)\prod_{k=N}^{2}A'(k)=\frac{\prod\limits_{k=1}^{N-1}\phi'(k)\phi'^T(k)\prod\limits_{k=N}^{2}\phi'(k)\phi'^T(k)}{\|\phi'(1)\|_2^2\|\phi'(N)\|_2^2\prod\limits_{k=2}^{N-1}\|\phi'(k)\|_2^4}$$

注意到　$$\phi'^T(i)\phi'(j)=x_1(i)x_1(j)+x_2(i)x_2(j)$$

有 $$E(N)=\frac{[x_1(1)x_1(2)+x_2(1)x_2(2)]\prod_{k=2}^{N-1}[x_1(k)x_1(k+1)+x_2(k)x_2(k+1)]^2}{[x_1^2(1)+x_2^2(1)][x_1^2(N)+x_2^2(N)]\prod_{k=2}^{N-1}[x_1^2(k)+x_2^2(k)]^2}$$

$$\cdot\phi'(1)\phi'^T(2)$$

$\because$ $$\|\phi'(1)\phi'^T(2)\|_2=\sqrt{\max_i\{\lambda_i[(\phi'(1)\phi'^T(2))^T(\phi'(1)\phi'^T(2))]\}}$$

$$=\sqrt{[x_1^2(1)+x_2^2(1)]\max_i\{\lambda_i[\phi'(2)\phi'^T(2)]\}}$$

$$=\sqrt{[x_1^2(1)+x_2^2(1)][x_1^2(2)+x_2^2(2)]}$$

$\therefore$ $\|E(N)\|_2=$

$$\frac{|x_1(1)x_1(2)+x_2(1)x_2(2)|\prod_{k=2}^{N-1}[x_1(k)x_1(k+1)+x_2(k)x_2(k+1)]^2}{[x_1^2(1)+x_2^2(1)]^{\frac{1}{2}}[x_1^2(2)+x_2^2(2)]^{\frac{3}{2}}[x_1^2(N)+x_2^2(N)]\prod_{k=3}^{N-1}[x_1^2(k)+x_2^2(k)]^2}$$

记 $$\rho^2(k)=\frac{[x_1(k)x_1(k+1)+x_2(k)x_2(k+1)]^2}{[x_1^2(k)+x_2^2(k)][x_1^2(k+1)+x_2^2(k+1)]}\qquad k=1,\cdots,N-1$$

而 $$1-\rho^2(k)=\frac{[x_1(k)x_2(k+1)-x_2(k)x_1(k+1)]^2}{[x_1^2(k)+x_2^2(k)][x_1^2(k+1)+x_2^2(k+1)]}\geqslant 0$$

$\therefore$ $$0\leqslant\rho^2(k)\leqslant 1$$

$\because$ $$\|E(N)\|_2=|\rho(1)|\prod_{k=2}^{N-1}\rho^2(k)$$

$\therefore$ $$0\leqslant\|E_N\|_2\leqslant 1$$

当且仅当$x_1(k)x_2(k+1)=x_2(k)x_1(k+1)$，即当$\phi(k)$与$\phi(k+1)$线性相关时，$|\rho(k)|=1$。$(k=1,\cdots,N-1)$

当且仅当对所有的$k\in\{1,\cdots,N-1\}$有$|\rho(k)|=1$时，$\|E_N\|_2=1$。

综上所述，可得出如下的结论：

在一般情况下循环方法收敛。当参数个数为2时，收敛的充分条件为至少能找到一个$k^*\in\{1,\cdots,N-1\}$使$\phi(k^*)$与$\phi(k^*+1)$不线性相关。

三、不动点的性质

我们分三种情况进行讨论。

1。定常确定性模型 $y(k)=\phi^T(k)\theta$

若取 $\hat{\theta}^l(1)=\theta$，显然 $\hat{\theta}^{l+1}(1)=\theta$，故$\theta$为不动点．由不动点的唯一性可知循环法收敛于系统的真实参数．

2．定常随机性模型 $y(k)=\phi^T(k)\theta+v(k)$

此时
$$B(k)=\frac{\phi(k)y(k)}{\|\phi(k)\|_2^2}=A(k)\theta+B^v(k)$$

其中 $B^v(k)=\dfrac{\phi(k)v(k)}{\|\phi(k)\|_2^2}$

而
$$F(N)=\left\{\prod_{k=1}^{N-1}(1-A(k))\left[\sum_{k=2}^{N-1}\prod_{i=N}^{k+1}(1-A(i))A(k)+A(N)\right]\right.$$
$$\left.+\sum_{k=2}^{N-1}\prod_{i=1}^{k-1}(1-A(i))A(k)+A(1)\right\}\theta+F^v(N)$$

其中 $F^v(N)$表示$F(N)$中的$B(k)$以$B^v(k)$代替

∴
$$(F(N)-F^v(N))/\theta=\prod_{k=1}^{N-1}(1-A(k))\left[\sum_{k=2}^{N-1}\left(\prod_{i=N}^{k+1}(1-A(i))-\prod_{i=N}^{k}(1-A(i))\right)\right.$$
$$\left.+A(N)\right]+\sum_{k=2}^{N-1}\left(\prod_{i=1}^{k-1}(1-A(i))-\prod_{i=1}^{k}(1-A(i))\right)+A(1)=1-E(N)$$

故不动点　$\hat{\theta}(1)=\theta+(1-E(N))^{-1}F^v(N)$．

不难理解，$(1-E(N))^{-1}F^v(N)$相当于在$\theta=0$时进行参数估计的估值 $\hat{\theta}^v(1)$，显然其均值为零．因而循环法能保持非时变性．

3．时变参数的随机性模型 $y(k)=\phi^T(k)\theta(k)+v(k)$

该情形下的理论分析比较困难，我们进行了几例仿真计算，并和离线方法比较（见附录），发现循环法对时变参数的跟踪效果比离线方法的好．当然一般性的结论还需从理论上论证．

四、结　束　语

循环法的优点在于通过多次向前、向后递推，“扩大”了样本容量，充分利用了信息，从而在一定程度上克服了因样本容量不大引起的问题．我们在武汉市人才需求总量预测中采用了该方法，收到了良好的效果．

参　考　文　献

〔1〕 韩志刚，多层递阶预报方法参数估值的初值选择，控制与决策，1，1，（1986），25—29．

〔2〕 韩志刚，天气系统的建模与预报，控制理论与应用，3，2，（1986），64—71．

〔3〕 罗新红等，多层递阶预报方法的改进及其应用，信息与控制，15，1，(1986)，1—5。
〔4〕 夏道行等著，实变函数与泛函分析（下册），高等教育出版社，北京，（1979）。
〔5〕 Assem S, Deif, Advanced Matrix Theory For Scientists and Engineers, Abacus Press, London, (1982).

附 录

1. 线性随机性模型 $y(k)=\theta_1(k)+\theta_2(k)x(k)+v(k)$。

$$\theta_1(k)=35,\ \theta_2(k)=2+0.1(k-1)。$$

通过Monte Carlo模拟产生下述10组数据：

y(k)	43.2	54.5	83.9	70.4	121.6	161.4	152.0	116.6	197.6	94.8
x(k)	4	9	23	15	36	50	45	30	58	20

（1）循环法

$\hat{\theta}_2(k)$	2.088	2.180	2.131	2.369	2.410	2.529	2.602	2.721	2.804	2.992
$\hat{\theta}_1(k)$	34.88	34.89	34.89	34.91	34.91	34.91	34.91	34.91	34.92	34.93

（2）离线方法

$\hat{\theta}_2(k)$	10.04	5.77	3.54	4.52	3.31	3.18	3.32	3.80	3.36	4.61
$\hat{\theta}_1(k)$	3.089	2.615	2.518	2.583	2.550	2.547	2.550	2.566	2.559	2.621

2. 线性随机性模型 $y(k)=\theta_1(k)+\theta_2(k)x(k)+v(k)$、

$$\theta_1(k)=35,\ \theta_2(k)=2-0.2(k-1)。$$

通过Monte Carlo 模拟产生下述10组数据：

y(k)	44.0	50.6	71.4	54.7	78.2	84.3	70.3	52.3	59.9	37.6
x(k)	4	9	23	15	36	50	45	30	58	20

（1）循环法

$\hat{\theta}_2(k)$	1.603	1.455	1.472	1.147	1.131	0.935	0.728	0.494	0.385	0.006
$\hat{\theta}_1(k)$	37.58	37.56	37.56	37.54	37.54	37.53	37.53	37.52	37.52	37.50

（2）离线方法

$\hat{\theta}_2(k)$	10.21	5.340	2.997	3.481	2.104	1.636	1.507	1.662	0.990	1.756
$\hat{\theta}_1(k)$	3.136	2.596	2.492	2.525	2.486	2.477	2.474	2.479	2.468	2.506

Recursive Method for the Initial Value Choice of Multi-layer Hierarchical Prediction Estimated Parameters

Zheng Yuhong

(Institute of System Engineering, Huazhong University of Science and Technology, Wuhan)

Han Guangwen

(Department of Management Engineering, Wuhan University of Technology)

Abstract

This paper presents a new method for the initial value choice of multi-layer hierarchical prediction estimated parameters——Recursive method, analyses its convergence and the property of the fixing points. Simulation shows that the method is satisfactory.

Key Words——Multi-layer hierarchical prediction; Initial value choice; Recurring method; Convergence; Fixing point.