第6卷 增刊1　　控制理论与应用　　Vol.6 Suppl.I. 1
1989年1月　　CONTROL THEORY AND APPLICATIONS　　Jan., 1989

有截距项指数模型的线性最小二乘辨识方法

达庆利　何建敏

(东南大学自动化研究所，南京)

摘　要

本文提出了有截距项指数模型的线性最小二乘辨识方法．证明了所提算法的收敛性和它与非线性规划算法的一致性．数值计算结果表明该方法具有简单易行，基本上无需猜测参数初值，收敛速度较快且与参数个数无关等一系列特点．

一、问题的提出

设有如下具有截距项的加性噪声指数模型

$$y_t = a + \prod_{j=1}^{m} x_{jt}^{\varphi_j} + u_t, \tag{1}$$

式中，a 为截距项，a 和 φ_j，$j=1,2,\cdots,m$ 均为待辨识的参数；$y_t>0, x_{1t}=e$ 且 $x_{jt}>0$，$j=1,2,\cdots,m$；u_t 为零均值正态白噪声．现要求利用它的 n 个样本($t=1,2,\cdots,n$)来辨识其中的参数．通常将其参数最优估计准则取作[1]（为简化表示，以下噪声和参数均指其估计值）：

$$J=\frac{1}{2}\sum_{t=1}^{n} u_t^2 . \tag{2}$$

容易推导该参数估计问题的最优性必要条件为

$$\frac{\partial J}{\partial \Phi} = -X_1^T \overline{W} U = 0, \tag{3.1}$$

$$\frac{\partial J}{\partial a} = -E^T U = 0, \tag{3.2}$$

本文于1986年9月23日收到．1987年1月16日收到修改稿．

式中

$$X_l=\begin{pmatrix}1 & \ln x_{21} & \cdots & \ln x_{m1}\\ 1 & \ln x_{22} & \cdots & \ln x_{m2}\\ \cdots & \cdots & \cdots & \cdots\\ 1 & \ln x_{2n} & \cdots & \ln x_{mn}\end{pmatrix}=\begin{pmatrix}x_l^1\\ \cdots\\ x_l^2\\ \cdots\\ \vdots\\ \cdots\\ x_l^n\end{pmatrix}\in R^{n\times m} \tag{4}$$

$$\overline{W}=\mathrm{diag}\{\hat{y}_1,\ \hat{y}_2,\cdots,\hat{y}_n\}\in R^{n\times n}, \tag{5}$$

$$\hat{y}_t=\prod_{j=1}^{m}x_{jt}^{\varphi_j},\ \ t=1,2,\cdots,n, \tag{6}$$

$$U=(u_1\ \ u_2\ \ \cdots\ \ u_n)^T\in R^n, \tag{7}$$

$$E=(1\ \ 1\ \ \cdots\ \ 1)^T\in R^n, \tag{8}$$

$$\Phi=(\varphi_1\ \ \varphi_2\ \ \cdots\ \ \varphi_m)^T\in R^m. \tag{9}$$

利用（3.1）和（3.2），可以形成各种非线性规划算法来进行搜索。当条件（3.1）和（3.2）得以满足时就认为找到了最优参数估计值。显然算法的有效性随着参数的增加而明显地下降。

为了避免使用这种非线性辨识方法，本文提出了一种与之相等价的线性辨识方法。这种方法的基本思想是，首先作平移变换消去模型中的截距项。接着引入被解释变量估计值 $\hat{y}_t$ 的预测变量 $\hat{y}_t^0$ 将模型（1）变换成线性形式。然后再用线性最小二乘迭代算法对变换后问题进行辨识。不断修正预测值 $\hat{y}_t^0$ 直至收敛，便可得到问题的解。可以认为本算法是文献[2]算法在 $a\neq 0$ 情况下的推广。

二、方法分析

令

$$v_t=a+u_t, \tag{10}$$

利用上式将（1）和（2）改写成

$$y_t=\sum_{j=1}^{m}x_{jt}^{\varphi_j}+v_t,\ \ t=1,\cdots,n \tag{11}$$

$$J=\frac{1}{2}\sum_{t=1}^{n}(v_t-a)^2. \tag{12}$$

引入预估变量 $\hat{y}_t^0$ 和相应的约束条件[3]

$$\hat{y}_t^0=\hat{y}_t=\prod_{j=1}^{m}x_{jt}^{\varphi_j}, \tag{13.1}$$

或者
$$\ln \hat{y}_t^0 = \sum_{j=1}^{m} \varphi_j \ln x_{jt}, \tag{13.2}$$

进而利用（13）将（11）改写为
$$w_t\left(\ln y_t - \sum_{j=1}^{m}\varphi_j \ln x_{jt}\right) = \nu_t, \tag{14}$$

式中，
$$w_t = \begin{cases} (y_t - \hat{y}_t^0)/(\ln y_t - \ln \hat{y}_t^0), & \hat{y}_t^0 \neq y_t \\ \lim\limits_{y_t^0 \to y_t} [(y_t - \hat{y}_t^0)/(\ln y_t - \ln \hat{y}_t^0)] = y_t, & \hat{y}_t^0 = y_t \end{cases} \tag{15}$$

不难证明
$$y_t^m \leqslant w_t \leqslant y_t^M, \tag{16}$$

式中，$y_t^m = \min(y_t, \hat{y}_t^0)$，$y_t^M = \max(y_t, \hat{y}_t^0)$。

一般地将（14）写成向量形式
$$W(Y_l - X_l\Phi) = V, \tag{17}$$

式中，
$$W = \mathrm{diag}\{w_1, w_2, \cdots, w_n\} \in R^{n\times n}, \tag{18}$$
$$Y_l = (\ln y_1 \quad \ln y_2 \quad \cdots \quad \ln y_n)^T \in R^n, \tag{19}$$
$$V = (\nu_1 \quad \nu_2 \quad \cdots \quad \nu_n)^T \in R^n. \tag{20}$$

于是，模型（1）在准则（2）下的参数估计问题就成为求如下拉格朗日函数极值问题
$$L(a, \Phi, \lambda, \hat{Y}_l^0) = \frac{1}{2}\|W(Y_l - X_l\Phi) - aE\|^2 + \lambda^T(\hat{Y}_l^0 - X_l\Phi), \tag{21}$$

式中，$\lambda = (\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_n)^T \in R^n$ 为拉格朗日乘子向量；
$$\hat{Y}_l^0 = (\ln \hat{y}_1^0 \ \ln \hat{y}_2^0 \cdots \ln \hat{y}_n^0)^T \in R^n. \tag{22}$$

其最优性必要条件
$$\frac{\partial L}{\partial \Phi} = -X_l^T[W^2(Y_l - X_l\Phi) - aWE + \lambda] = 0, \tag{23}$$
$$\frac{\partial L}{\partial \lambda} = \hat{Y}_l^0 - X_l\Phi = 0, \tag{24}$$
$$\frac{\partial L}{\partial \ln \hat{y}_t^0} = \begin{cases} \dfrac{w_t - \hat{y}_t^0}{\ln y_t - \ln \hat{y}_t^0}(\ln y_t - x_t^l\Phi)[w_t(\ln y_t - x_t^l\Phi) - a] + \lambda_t = 0, & \hat{y}_t^0 \neq y_t. \\ \dfrac{1}{2}w_t(\ln y_t - x_t^l\Phi)[w_t(\ln y_t - x_t^l\Phi) - a] + \lambda_t = 0, & \hat{y}_t^0 = y_t \end{cases}$$
$$t = 1, 2, \cdots, n. \tag{25}$$

$$\frac{\partial L}{\partial a}=-E^T[W(Y_l-X_l\Phi)-aE]=0, \tag{26}$$

即

$$a=\frac{1}{n}E^TW(Y_l-X_l\Phi). \tag{27}$$

考虑到(24)，由(25)得：

$$\frac{\partial L}{\partial \hat{Y}_l^0}=(W-\widetilde{W})[W(Y_l-X_l\Phi)-aE]+\lambda=0,$$

式中，

$$\widetilde{W}=\operatorname{diag}\{\widetilde{w}_1,\widetilde{w}_2,\cdots,\widetilde{w}_n\}=\operatorname{diag}\{\hat{y}_1^0,\hat{y}_2^0,\cdots,\hat{y}_n^0\}\in R^{n\times n}. \tag{28}$$

利用(23)和上式消去λ，再将(27)代入得

$$X_l^T\widetilde{W}\left(I-\frac{EE^T}{n}\right)W(Y_l-X_l\Phi)=0. \tag{29}$$

如果 $\operatorname{Rank} X_l=m$，则 $X_l^T\widetilde{W}\left(I-\frac{EE^T}{n}\right)WX_l$ 非奇异，可得

$$\Phi=\left[X_l^T\widetilde{W}\left(I-\frac{EE^T}{n}\right)WX_l\right]^{-1}X_l^T\widetilde{W}\left(I-\frac{EE^T}{n}\right)WY_l. \tag{30}$$

利用(27)和(30)形成迭代算法；利用(24)来判别算法的收敛性，再根据(16)来选择W的迭代初值便可得到如下节所述的线性最小二乘迭代算法。

三、迭代算法

第一步 选择a的迭代初值$a^{(0)}$. 并取

$$w_t^{(1)}=\widetilde{w}_t^{(1)}=\hat{y}_t^{(0)}=y_t-a^{(0)},\quad t=1,2,\cdots,n. \tag{31}$$

令$k=1$，转第三步.

第二步 计算

$$W^{(k)}=\operatorname{diag}\{w_t^{(k)}\}, \tag{32}$$

其中

$$w_t^{(k)}=\begin{cases}(y_t-\hat{y}_t^{(k-1)})/(\ln y_t-\ln\hat{y}_t^{(k-1)}), & \hat{y}_t^{(k-1)}\neq y_t\\ y_t, & \hat{y}_t^{(k-1)}=y_t\end{cases} \tag{33}$$

$$\widetilde{W}^{(k)}=\operatorname{diag}\{\widetilde{w}_t^{(k)}\}=\operatorname{diag}\{\hat{y}_t^{(k-1)}\} \tag{34}$$

第三步 计算

$$\Phi^{(k)}=\left[X_l^T\widetilde{W}^{(k)}\left(I-\frac{EE^T}{n}\right)W^{(k)}X_l\right]^{-1}X_l^T\widetilde{W}^{(k)}\left(I-\frac{EE^T}{n}\right)W^{(k)}Y_l, \tag{35}$$

$$\hat{Y}_l^{(k)}=X_l\Phi^{(k)}, \tag{36}$$

$$a^{(k)}=\frac{1}{n}E^TW^{(k)}(Y_l-\hat{Y}_l^{(k)}). \tag{37}$$

第四步　若 $k=1$，则令 $k=2$ 转第二步．否则，若

$$|a^{(k)}-a^{(k-1)}|<\varepsilon_1 \quad 且 \quad \max_j\left(|(\varphi_j^{(k)}-\varphi_j^{(k-1)})/\varphi_j^{(k)}|\right)<\varepsilon_2, \tag{38}$$

式中，ε_1 和 ε_2 是预先给定的小的正数，则转第五步；反之，令 $k=k+1$ 转第二步。

第五步　停机。

四、收　敛　性　研　究

定理 1　若 Rank $X_l=m$，且对于一切 $k>N$（N 为一有限的正整数）均有

$$A^{(k)}(A^{(k)})^T-B^{(k)}(B^{(k)})^T>0, \tag{39}$$

式中，

$$A^{(k)}=X_l^T\widetilde{W}^{(k)}\left(I-\frac{EE^T}{n}\right)W^{(k)}X_l, \tag{40}$$

$$B^{(k)}=X_l^T\widetilde{W}^{(k)}\left[\left(I-\frac{EE^T}{n}\right)\alpha^{(k)}+\beta^{(k)}\right]X_l, \tag{41}$$

$$\alpha^{(k)}=\mathrm{diag}\{\alpha_i^{(k)}\}\in R^{n\times n}, \tag{42}$$

$$\alpha_i^{(k)}=(w_i^{(k)}-\hat{y}_i^{(k-1)})(\ln y_i-\ln\hat{y}_i^{(k)})/(\ln y_i-\ln\hat{y}_i^{(k-1)}), \tag{43}$$

$$\beta^{(k)}=\mathrm{diag}\{\beta_i^{(k)}\}\in R^{n\times n}, \tag{44}$$

$$\beta_i^{(k)}=w_i^{(k)}(\ln y_i-\ln\hat{y}_i^{(k)})-a^{(k)}. \tag{45}$$

则上述迭代算法收敛。

注　1 证明参数向量序列 $\{\Phi^{(k)}\}$ 在定理条件下的收敛性

由（35）知

$$X_l^T\widetilde{W}^{(k)}\left(I-\frac{EE^T}{n}\right)W^{(k)}(Y_l-X_l\Phi^{(k)})=0, \tag{46}$$

则有

$$\frac{\partial\Phi^{(k)}}{\partial\ln\hat{y}_i^{(k-1)}}=(A^{(k)})^{-1}\left[(x_i^l)^T\widetilde{w}_i^{(k)}\beta_i^{(k)}+(x_i^l)^T\widetilde{w}_i^{(k)}\alpha_i^{(k)}\right.$$

$$\left.-X_l^T\widetilde{W}^{(k)}\frac{E}{n}\alpha_i^{(k)}\right].$$

秩条件 Rank $X_l=m$ 保证了上式中逆矩阵的存在。

那么，
$$\frac{\partial\Phi^{(k)}}{\partial(\hat{y}_l^{(k-1)})^T}=(A^{(k)})^{-1}\left[X_l^T\widetilde{W}^{(k)}\left(I-\frac{EE^T}{n}\right)\alpha^{(k)}+X_l^T\widetilde{W}^{(k)}\beta^{(k)}\right]. \tag{47}$$

于是

$$\frac{\partial\Phi^{(k)}}{\partial(\Phi^{(k-1)})^T}=\frac{\partial\Phi^{(k)}}{\partial(\hat{Y}_l^{(k-1)})^T}\frac{\partial\hat{Y}_l^{(k-1)}}{\partial(\Phi^{(k-1)})^T}=(A^{(k)})^{-1}B^{(k)}. \tag{48}$$

又由定理条件知，当 $k>N$ 时

$$A^{(k)}(A^{(k)})^T-B^{(k)}(B^{(k)})^T>0.$$

即
$$I-(A^{(k)})^{-1}B^{(k)}[(A^{(k)})^{-1}B^{(k)}]^T>0.$$

可见对于一切 $k>N$ 有

$$\left\|\frac{\partial\Phi^{(k)}}{\partial(\Phi^{(k-1)})^T}\right\|_2=\|(A^{(k)})^{-1}B^{(k)}\|_2<1. \tag{49}$$

考虑到定理中秩条件成立时 $A^{(N)}$ 和 $B^{(N)}$ 均有限，由压缩影射原理[4]知当 $k\to+\infty$ 时序列 $\{\Phi^{(k)}\}$ 收敛。

2. 证明序列 $\{a^{(k)}\}$ 的收敛性

由(37)知 $a^{(k)}$ 为 $\Phi^{(k)}$ 和 $\Phi^{(k-1)}$ 的连续函数，于是当 $\{\Phi^{(k)}\}$ 收敛时，序列 $\{a^{(k)}\}$ 也收敛。

总结1、2两点知定理成立，证毕。

定理 2 如果本算法收敛，则参数向量序列 $\{(a^{(k)},\Phi^{T(k)})^T\}$ 必收敛于(3.1)和(3.2)的解。

证 当算法收敛时

$$\lim_{k\to+\infty}\hat{y}_t^{(k)}=\hat{y}_t,\qquad t=1,2,\cdots,n$$

$$\lim_{k\to+\infty}\widetilde{W}^{(k)}=\mathrm{diag}\{\hat{y}_t\}=\overline{W} \tag{50}$$

则有 $a=\lim\limits_{k\to+\infty}a^{(k)}=\lim\limits_{k\to+\infty}\frac{E^T}{n}W^{(k)}(Y_l-\hat{Y}_l^{(k)})$

$$=\frac{E^T}{n}[(y_1-\hat{y}_1),\cdots,(y_n-\hat{y}_n)]^T.$$

或
$$\frac{E^T}{n}[(y_1-\hat{y}_1),\cdots,(y_n-\hat{y}_n)]^T-a=\frac{E^T}{n}[(y_1-a-\hat{y}_1),\cdots,(y_n-a-\hat{y}_n)]^T$$

$$=\frac{E^T}{n}U=0, \tag{51}$$

此外 $\lim\limits_{k\to+\infty}\left(I-\dfrac{EE^T}{n}\right)W^{(k)}(Y_l-\hat{Y}_l^{(k)})=U$. （52）

于是 $0=\lim\limits_{k\to+\infty}X_l^T\widetilde{W}^{(k)}\left(I-\dfrac{EE^T}{n}\right)W^{(k)}(Y_l-\hat{Y}_l^{(k)})=X_l^T\overline{W}U$. （53）

（53）和（51）分别与（3.1）和（3.2）相符，定理证毕。

由定理 1 容易得到如下推论：

推　论　若 Rank $X_l=m$，则保证最优点局部收敛性的充分条件是最优参数满足

$$AA^T-BB^T>0, \tag{54}$$

式中，

$$A=X_l^T\overline{W}\left(I-\frac{EE^T}{n}\right)WX_l, \tag{55}$$

$$B=X_l^T\overline{W}\left[\left(I-\frac{EE^T}{n}\right)(W-\overline{W})+\overline{U}\right]X_l, \tag{56}$$

$$\overline{U}=\mathrm{diag}\{u_1,u_2,\cdots,u_n\}. \tag{57}$$

进一步考察本节定理和推论可以发现：

1. 定理 2 实际上证明了本算法与各种非线性规划算法的一致性。

2. 当被辨识模型能够良好地拟合样本数据时，收敛条件是不难满足的。这可以通过考察一个极端情况来看出。设模型能够理想地拟合所有样本点，各点残差均为零，则推论中 $W=\overline{W}$，$\overline{U}=0$，于是 $B=0$；此外由秩条件知 AA^T 为正定阵，可见（54）必定成立。

五、实 例 计 算

某市某工业部门总产值 N，固定资产 K，职工人数 L 和科技人员数 T 如表 1 所示。

表 1

年份19××	72	73	74	75	76	77	78	79	80	81	82	83
序号 t	1	2	3	4	5	6	7	8	9	10	11	12
N_t 万元	1850	2600	2302	2485	2622	2889	3214	3617	4286	4298	4300	5166
K_t 万元	450	770.9	868.6	959.8	1045	1149	1264	1377	1502	1550	1911	2800
L_t ×10人	1600	1684	1651	1789	1878	1957	2200	2323	2501	2600	2810	2885
T_t 人	405	506	519	538	601	601	650	752	813	800	750	950

现利用本文算法建立如下五种生产函数：

(a) $N_t=a+e^{\varphi_1}K_t^{\varphi_2}+u_t$,

(b) $N_t=a+e^{\varphi_1}L_t^{\varphi_2}+u_t$,

(c) $N_t=a+e^{\varphi_1}K_t^{\varphi_2}\ L_t^{\varphi_3}+u_t$,

(d) $N_t=a+e^{\varphi_1}K_t^{\varphi_2}\ T_t^{\varphi_3}+u_t$,

(e) $N_t=a+e^{\varphi_1}K_t^{\varphi_2}\ L_t^{\varphi_3}\ T_t^{\varphi_4}+u_t$.

考虑到各参数对模型精度的影响，收敛条件选择为：$\varepsilon_1=0.5$，$\varepsilon_2=0.0005$。相应于不同初值 $a^{(0)}$ 下，其迭代收敛时性能准则 J，模型参数估计值，所需迭代次数如表 2 所示。

表 2

模型	性能准则最小值 J^*	初值 $a^{(0)}$ / 迭代次数 / 最优参数	−1000	0	1000	1500	1845
			$N_1-a^{(0)}=2850$	$N_1-a^{(0)}=1850$	$N_1-a^{(0)}=850$	$N_1-a^{(0)}=350$	$N_1-a^{(0)}=5$
(a)	458491	$a=-796$ $\varphi_1=4.687$ $\varphi_2=0.5090$	9	9	8	7	7
(b)	230113	$a=-97$ $\varphi_1=-2.315$ $\varphi_2=1.359$	6	6	5	5	4
(c)	150072	$a=-572$ $\varphi_1=0.7087$ $\varphi_2=0.1888$ $\varphi_3=0.8077$	6	6	6	5	5
(d)	196100.9	$a=203$ $\varphi_1=0.2782$ $\varphi_2=0.1912$ $\varphi_3=0.9835$	10	10	9	9	9
(e)	67981.0	$a=685$ ${}_1=-4.293$ $\varphi_2=0.04459$ $\varphi_2=0.9198$ $\varphi_4=0.7323$	7	7	6	6	6

表 3 给出了模型(e)在 $a^{(0)}=1500$ 时 J，$\left|\dfrac{\partial J}{\partial a}\right|$，$\max\limits_i\left(\left|\dfrac{\partial J}{\partial \varphi_i}\right|\right)$ 随迭代次数变化的情况。

表 3 （**计算机仅显示6位有效数字）

迭代次数 K	1	3	6	10	15
J	90291.5	69059.5	67981.0**	67981.0**	67981.0**
$\left\lvert\dfrac{\partial J}{\partial a}\right\rvert$	0.5880×10^3	0.1606×10^3	0.3539	0.2988×10^{-3}	0.8785×10^{-7}
$\max\limits_i\left(\left\lvert\dfrac{\partial J}{\partial \varphi_i}\right\rvert\right)$	0.1424×10^8	0.3207×10^7	0.5859×10^4	0.4502×10^1	0.6254×10^{-3}

由表 2 可见

1. 对于迭代初值的不同选择参数均收敛于唯一的最优点。

2. 所有算例收敛速度均相当快。

3. 收敛速度与模型中参数个数无关。

4. 初值 $a^{(0)}$ 的选择对算法收敛速度无明显影响。进一步分析表明大多数模型的收敛速度随着 $N_1 - a^{(0)}$ 的减小而略有提高。一般而言，可以认为选择 $a^{(0)}$ 使之略小于模型（1）中 $\min\limits_t(y_t)$ 是迭代初值比较合理的选择。

由表 3 可以看出准则值 J 以及它对于参数的梯度随着迭代次数的增加而迅速减小，因而也就从数值上验证了定理 2 的正确性。

六、结 束 语

本文提出了有截距项指数模型的线性辨识方法，该方法有如下特点：

1. 算法相当简单，便于使用。

2. 算法中仅要求猜测一个参数（截距项 a）的初值；算例又说明了初值 $a^{(0)}$ 的选择对收敛迭代次数的影响并不大；此外在上节讨论中还就一般情况建议了初值 $a^{(0)}$ 的统一选择方法。因此，可以认为算法给出了对于不同问题均为适用的初值选择方法。这就在较大程度上克服了通常非线性规划算法需要猜测初值的困难。

3. 收敛迭代次数与参数个数无关。因此，一般而论，迭代次数不会随着所辨识参数个数的增加而增加。相反，随着参数个数的增加，模型对样本的拟合精度的提高，反而更容易保证算法的收敛性。

总结 2，3 两点，可以认为，本算法在一定程度上克服了非线性规划算法在处理本文所研究的高维情况的大多数缺点。

4. 大量计算实例表明，只要解释变量选择确当，通常情况下经过 5～10 次迭代便可得到令人满意的精度。

5. 将本文和文献〔2〕相比较知，在本文算法中令 $a=0$ 便可直接得到文献〔2〕的算法。

致谢 本文投稿前承徐南荣教授审阅，谨表谢意。

参 考 文 献

〔1〕 Pindyck, R. S., Rubinfeld, D. L., Econometric Models and Economic Forecasts, McGraw-Hill Book Company, New York, (1981), 261—262.

〔2〕 达庆利，加性噪声指数模型的加权线性最小二乘辨识方法，系统工程学报，1，(1986)，52—62.

〔3〕 Hassan, M. F. and Singh, M. G., Hierarchical Successive Approximation Algorithms for Non-linear Systems. Part I. Generalisation of the Method of Takahara, Large Scale Systems Theory and Application, 2:2, (1981), 65—79.

〔4〕 王德义，非线性方程组解法与最优化方法，高等教育出版社，北京，(1984)，16—24.

A Linear Least-Squares Identification Algorithm for the Exponential Model with an Intersection Parameter

Da Qingli, He Jianmin

(Research Institute of Automation, Southeast University, Nanjing)

Abstract

In this paper a new linear least-squares identification algorithm for the exponential model with an intersection parameter is proposed. Its convergence and the consistency between it and nonlinear programming approaches are proved. Numerical calculation results demonstrate that the algorithm is quite simple, easy to be used, unnecessary to guess the initial values of the parameter, and that its convergence rate is very fast and independent of the number of the parameters identified, etc.