第7卷 第3期
1990年9月

控制理论与应用
CONTROL THEORY AND APPLICATIONS

Vol. 7, No. 3
Sept., 1990

非线性系统最小二乘估计的目标补偿算法

何建敏 达庆利
(东南大学管理学院，南京)

摘要 本文提出了求解非线性系统最小二乘估计问题的一种新的分解协调方法，该方法通过对各子问题的目标直接引入适当的补偿项进行协调，具有协调计算简单，子问题仍为最小二乘形式等优点. 文中对该方法收敛性进行了较为详细的研究. 仿真和实际应用表明，它的收敛速度快，优于整体算法和目标协调法.

关键词： 参数估计；非线性系统；目标补偿；分解一协调；最小二乘问题

1. 引 言

考察下列非线性系统模型

$$y(t) = f[x(t),\varphi] + u(t), \tag{1}$$

式中 $y(t)\in R$：被解释变量；$x(t)\in R^m$：解释向量；$\varphi\in R^n$：待估计参数向量；$u(t)\in R$：模型噪声，一般设为零均值白噪声. 若已得到样本集合 $S=\{y(t),x(t)\,|\,t=1,\cdots,t_f\}$，则模型(1)参数 φ 的最小二乘估计 $\hat\varphi^*$ 为下列问题的解

$$(\text{LS})\ \min_{\hat\varphi}\frac{1}{2}\sum_{t=1}^{t_f}\hat u^2(t) = \min_{\hat\varphi}\frac{1}{2}\sum_{t=1}^{t_f}\{y(t) - f[x(t),\varphi]\}^2. \tag{2}$$

许多求解最小二乘问题的较好方法，如 L—M 法，文献[1,2]中的诸方法，都可以用来求解(2). 然而，在较复杂的非线性系统中，一般 φ 的维数 n 较大，这就使得这些整体方法的计算量和存贮量剧增，从而降低了它们的有效性. 文献[3]针对这一困难提出了一种目标协调算法. 但它要求 $f[x(t),\varphi]$关于 φ 加性可分，同时未将样本集合 S 分解，这就限制了它的应用范围和进一步提高效率，但若将 S 分解后再使用目标协调法，则算法收敛时 Lagrange 乘子 $\lambda_i=0$不再成立[3]，这样既不利简化问题也不能使各子问题为最小二乘形式，也就不能利用上述较好方法求解它们，这些都严重妨碍了提高目标协调法的效率. 为此，本文提出了一种新的分解一协调方法，它通过对各子问题的目标直接引入适当的二次补偿项进行协调，具有协调级算法简单，第一级子问题均为最小二乘形式等特点，从而避免了目标协调法的不足. 理论分析表明该方法的收敛条件较弱，它的仿真和应用也得到了令人满意的结果.

2. 方法的推导

2.1 分解 将(2)中的变量 $\hat\varphi$ 分解为 $\hat\varphi=[\hat\varphi_1^T,\ldots,\hat\varphi_N^T]^T$；$\hat\varphi_i\in R^{n_i}$，$i=1,\ldots,N$，$\sum_{i=1}^{N}n_i=n$. 相

本文于1989年2月1日收到. 1989年9月11收到修改稿.

应地,样本集合 S 亦分解为 $S_i=\{y(t),x(t)\mid t=t_{i-1}+1,\ldots,t_i\}$,$i=1,\ldots,N$,且 $t_0=0$,$t_N=t_f$. 再引入预估参数向量 $\hat{\varphi}^o=[\hat{\varphi}_1^{oT},\ldots,\hat{\varphi}_N^{oT}]^T$($\hat{\varphi}_i^o\in R^{n_i}$,$i=1,\ldots,N$)和等式约束 $\hat{\varphi}^o=\hat{\varphi}$. 记 $f[x(t),\hat{\varphi}]=f[x(t),\hat{\varphi}_1,\ldots,\hat{\varphi}_N]$. 据此定义如下向量和函数

$$F_i\triangleq[f_i(t_{i-1}+1),\ldots,f_i(t_i)]^T\in R^{t_i-t_{i-1}},\tag{3.1}$$

$$f_i(t)\triangleq f[x(t),\hat{\varphi}_1^o,\ldots,\hat{\varphi}_{i-1}^o,\hat{\varphi}_i,\hat{\varphi}_{i+1}^o,\ldots,\hat{\varphi}_N^o],\quad t=1,\ldots,t_f\tag{3.2}$$

$$U_i\triangleq[\hat{u}_i(t_{i-1}+1),\ldots,\hat{u}_i(t_i)]^T\in R^{t_i-t_{i-1}},\tag{4.1}$$

$$\hat{u}_i(t)\triangleq y(t)-f_i(t),\quad t=1,\ldots,t_f,\tag{4.2}$$

$$F_i^o\triangleq[f^o(t_{i-1}+1),\ldots,f^o(t_i)]^T\in R^{t_i-t_{i-1}},\tag{5.1}$$

$$f^o(t)\triangleq f[x(t),\hat{\varphi}^o],\quad t=1,\ldots,t_f,\tag{5.2}$$

$$\hat{U}_i^o\triangleq[\hat{u}^o(t_{i-1}+1),\ldots,\hat{u}^o(t_i)]^T\in R^{t_i-t_{i-1}},\tag{6.1}$$

$$\hat{u}^o(t)\triangleq y(t)-f^o(t),\quad t=1,\ldots,t_f\tag{6.2}$$

$$Y_i=[y(t_{i-1}+1),\ldots,y(t_i)]^T\in R^{t_i-t_{i-1}},\tag{7}$$

$$i=1,\ldots,N,\text{和 } Y=[Y_1^T,\ldots,Y_N^T]^T\in R^{t_f},\tag{8}$$

则(2)转化为下列等价问题

$$\text{(PLS)}\quad \min_{\hat{\varphi}}\frac{1}{2}\sum_{i=1}^{N}\sum_{t=t_{i-1}+1}^{t_i}\{y(t)-f_i[x(t),\hat{\varphi}_i,\hat{\varphi}^o]\}^2=\min_{\hat{\varphi}}\frac{1}{2}\sum_{i=1}^{N}(Y_i-F_i)^T(Y_i-F_i).\tag{9}$$

s. t.

$$\hat{\varphi}^o=\hat{\varphi}.\tag{10}$$

现选 $\hat{\varphi}^o$ 为协调变量,它的值由协调级给定,相应的第一级 N 个子问题为($i=1,\ldots,N$)

$$\text{(spi)}\quad \min_{\hat{\varphi}_i}\frac{1}{2}[(Y_i-F_i)^T(Y_i-F_i)+(\hat{\varphi}_i-b_i^o)^T(\hat{\varphi}_i-b_i^o)],\tag{11}$$

s. t. $\hat{\varphi}^o$ 由协调级给定.

这里 $(\hat{\varphi}_i-b_i^o)^T(\hat{\varphi}_i-b_i^o)$ 即为目标补偿项,$b_i^o=b_i(\hat{\varphi}^o)$ 为其参数向量. 可见(spi)为最小二乘估计形式,它的最优性必要条件为

$$\hat{\varphi}_i-b_i^o-(\partial F_i/\partial\hat{\varphi}_i)^T\hat{U}_i=0.\tag{12}$$

由(2),(5)和(6),得(2)关于 $\hat{\varphi}_i$ 的最优性必要条件为

$$-\sum_{k=1}^{N}(\partial F_k^o/\partial\hat{\varphi}_i^o)^T\hat{U}_k^o\big|_{\hat{\varphi}^o=\hat{\varphi}}=0.\tag{13}$$

比较(12),(13),取

$$b_i^o=\hat{\varphi}_i^o+\sum_{k=1,\,k\neq i}^{N}(\partial F_k^o/\partial\hat{\varphi}_i^o)^T\hat{U}_k^o,\tag{14}$$

当协调算法收敛时($\hat{\varphi}^o=\hat{\varphi}$),即可保证其解满足原问题最优性必要条件.

2.2 协调　协调级的任务是实现 $\hat{\varphi}^o=\hat{\varphi}$,可采用直接迭代算法

$$\hat{\varphi}^{o(l+1)}=\hat{\varphi}^{o(l)}-sD^{(l)}[\hat{\varphi}^{o(l)}-\hat{\varphi}^{(l)}],\tag{15}$$

其中　s 为松驰因子,$D^{(l)}$ 为修正矩阵且 $\det(D^{(l)})\neq0$,它的部分形式将在收敛性一节中讨论. 两级分解—协调结构示于图1. 算法步骤为

1) 给出初值 $\hat{\varphi}^{o(0)}$,允许误差 $\varepsilon>0$,令 $l=0$.

2) 第一级求解$(\text{spi})^{(l)}$,得 $\hat{\varphi}_i^{(l)}$,$i=1,\ldots,N$.

3) 若 $\|\hat{\varphi}^{o(l)}-\hat{\varphi}^{(l)}\|\leqslant\varepsilon$,得 φ 的最小二乘估计 $\hat{\varphi}^*=\hat{\varphi}^{(l)}$,算法终止. 否则转4).

4) 协调级据(15)计算 $\hat{\varphi}^{o(l+1)}$,令 $l=l+1$,转回2).

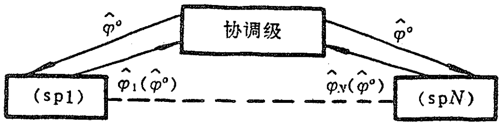

图 1 协调结构图

3. 收敛性研究

本节的讨论均假定模型(1)中 $f[x(t),\varphi]$关于 φ 二阶连续可微. 对于向量函数 $G=G(x):R^p\to R^q$,记 $\partial G/\partial x=(G)_x$,从而定义

$$A = \text{block}-\text{diag}\{A_{11},\ldots,A_{NN}\}\in R^{n\times n}, \tag{16.1}$$

$$A_{ii} = I+(F_i)^T_{\hat\varphi_i}(F_i)_{\hat\varphi_i}-\sum_{t=t_{i-1}+1}^{t_i}\hat u_i(t)\,[\nabla^2_{\hat\varphi_i}f_i]\in R^{n_i\times n_i}, \tag{16.2}$$

$$B=\begin{pmatrix}B_{11},\ldots,B_{1N}\\ B_{N1},\ldots,B_{NN}\end{pmatrix}\in R^{n\times n}, \tag{17.1}$$

$$B_{ij}=\partial b_i^o/\partial\hat\varphi_i^o=\begin{cases}\sum\limits_{\substack{k=1\\k\neq i}}^{N}\Big[\sum\limits_{t=t_{k-1}+1}^{t_k}\hat u^o(t)(\nabla_{\hat\varphi_i^o}f^o)_{\hat\varphi_j^o}-(F_k^o)^T_{\hat\varphi_i^o}(F_k^o)_{\hat\varphi_j^o}\Big], & i\neq j,\\ I+\sum\limits_{\substack{k=1\\k\neq i}}^{N}\Big[\sum\limits_{t=t_{k-1}+1}^{t_k}\hat u^o(t)(\nabla^2_{\hat\varphi_i^o}f_i)-(F_k^o)^T_{\varphi_i^o}(F_k^o)_{\hat\varphi_i^o}\Big], & i=j,\end{cases} \tag{17.2}$$

$$C=\begin{pmatrix}C_{11},\ldots,C_{1N}\\ C_{N1},\ldots,C_{NN}\end{pmatrix}\in R^{n\times n}, \tag{18.1}$$

$$C_{ij}=\begin{cases}(F_i)^T_{\hat\varphi_i}(F_i)_{\hat\varphi_j^o}-\sum\limits_{t=t_{i-1}+1}^{t_i}\hat u_i(t)(\nabla_{\hat\varphi_i}f_i)_{\hat\varphi_j^o}, & i\neq j,\\ 0 & i=j.\end{cases} \tag{18.2}$$

其它符号定义见(3)—(6)式.

1) $D^{(l)}=I,\quad s=1,\quad$ 即 $\quad\hat\varphi^{o(l+1)}=\hat\varphi^{(l)}.$ (19)

定理 1 设在开域 $D\subseteq R^n$ 上,矩阵 A 的逆存在,且矩阵 $A^{-1}(B-C)$的谱半径 $\rho[A^{-1}(B-C)]<1$. 则当 $\hat\varphi^o$ 的初值选取使迭代序列$\{\hat\varphi^{o(l)},l=l_m,l_m+1,\ldots\}\subset D$ 时,协调算法(19)收敛于函数 $\hat\varphi=\hat\varphi(\hat\varphi^o)$在 D 中的不动点 $\hat\varphi^*=\hat\varphi(\hat\varphi^*)$,这里,$l_m$ 为一有限的正整数.

证 对于协调级来说,(spi)的最优性必要条件(12)总是成立的,该式两端对 $\hat\varphi_j^o$ 求导可得

$$A_{ii}\cdot(\hat\varphi_i)_{\hat\varphi_j^o}-B_{ij}+C_{ij}=0.$$

又当 $l\geqslant l_m$ 时,A^{-1}存在即 A_{ii}^{-1}存在,故$(\hat\varphi_i)_{\hat\varphi_j^o}=A_{ii}^{-1}(B_{ij}-C_{ij})$,因此 $\hat\varphi'(\hat\varphi^o)=A^{-1}(B-C)$,再由定理条件得 $\rho[\hat\varphi'(\hat\varphi^o)]=\rho[A^{-1}(B-C)]<1$,据 Ostrouski 定理[4],(19)收敛于 $\hat\varphi^*=\hat\varphi(\hat\varphi^*)\in D$.
证毕.

下面的定理给出了判定矩阵 $A^{-1}(B-C)$的谱半径 $\rho[A^{-1}(B-C)]<1$的充分条件.

定理 2 若矩阵 $A-(B-C)$满足下列任一条件 a) $A-(B-C)$为严格对角优势矩阵;b) $A-(B-C)$为不可约对角优势矩阵;c) $A-(B-C)$为严格块对角优势矩阵[3]. 则 $\rho[A^{-1}(B-C)]<$

1.

证 仿文献[3]，这里从略.

2) $$D^{(l)}=A^{(l)},\quad \text{即}\ \hat{\varphi}^{\rho(l+1)}=\hat{\varphi}^{\rho(l)}-sA^{(l)}[\hat{\varphi}^{\rho(l)}-\hat{\varphi}^{(l)}]. \tag{20}$$

假定 $\hat{\varphi}^*$ 为 $\hat{\varphi}(\hat{\varphi}^\rho)$ 的不动点，即 $\hat{\varphi}(\hat{\varphi}^*)=\hat{\varphi}^*$，在 $\hat{\varphi}^*$ 处，因 $\hat{\varphi}^\circ=\hat{\varphi}^*=\hat{\varphi}$，所以 $[A-(B-C)]^*$ 为对称矩阵.

定理 3 设矩阵 $[A-(B-C)]$ 在函数 $\hat{\varphi}(\hat{\varphi}^\circ)$ 的不动点 $\hat{\varphi}^*$ 处正定，且 $0<s<2/\lambda^*_{\max}$，则协调算法(20)局部收敛于 $\hat{\varphi}^*$. 这里 $\lambda^*_{\max}$ 为 $[A-(B-C)]^*$ 的最大特征值.

证 令 $G(\hat{\varphi}^\circ)=\hat{\varphi}^\rho-sA(\hat{\varphi}^\rho-\hat{\varphi}(\hat{\varphi}^\rho))$，则

$G(\hat{\varphi}^*)=\hat{\varphi}^*-sA^*(\hat{\varphi}^*-\hat{\varphi}(\hat{\varphi}^*))=\hat{\varphi}^*$，即 $\hat{\varphi}^*$ 亦为 $G(\hat{\varphi}^\rho)$ 的不动点. 而

$G'(\hat{\varphi}^*)=I-sA^*(I-A^{-1}(B-C))^*=I-s(A-(B-C))^*$ 且 $\rho[G'(\hat{\varphi}^*)]=\max\limits_{\lambda^*}|1-s\lambda^*|$，

其中 λ^* 为 $[A-(B-C)]^*$ 的特征值. 由定理条件得 $0<s\lambda^*<\frac{2}{\lambda^*_{\max}}\cdot\lambda^*_{\max}=2$，故 $\max\limits_{\lambda^*}|1-s\lambda^*|<1$，即 $\rho[G'(\hat{\varphi}^*)]<1$，据 Ostrouski 定理[4]，算法(20)局部收敛于 $\hat{\varphi}^*$. 证毕.

注意到，对于 $\hat{\varphi}^\rho=\hat{\varphi}=\hat{\varphi}^*$，$[A-(B-C)]$ 恰为原问题(2)的 Hessian 矩阵，在最优点处常具有对称正定性质，因此该定理的条件是较弱的.

3) $$D^{(l)}=\{I-[A^{-1}(B-C)]^{(l)}\}^T. \tag{21}$$

即 $$\hat{\varphi}^{\rho(l+1)}=\hat{\varphi}^{\rho(l)}-s[I-[A^{-1}(B-C)]^{(l)}\}^T[\hat{\varphi}^{\rho(l)}-\hat{\varphi}^{(l)}]. \tag{22}$$

由于 A 为块对角矩阵，其逆较易计算.

定理 4 设矩阵 $[A-(B-C)]$ 在 $\hat{\varphi}(\hat{\varphi}^*)=\hat{\varphi}^*$ 处为可逆的，且 $0<s<2/\rho\{(I-[A^{-1}(B-C)]^{(l)})^T(I-[A^{-1}(B-C)]^{(l)})\}$，则协调算法(22)局部收敛于 $\hat{\varphi}^*$.

定理 5 设集合 $S(\hat{\varphi}^{\rho(0)})=\{\hat{\varphi}^\rho|e(\hat{\varphi}^\rho)\leqslant e(\hat{\varphi}^{\rho(0)})\}$ 有界，且 $[A-(B-C)]^{(l)}$ 可逆. 则当 s 足够小时，协调算法(22)收敛于 $\hat{\varphi}^*=\hat{\varphi}(\hat{\varphi}^*)$. 这里 $e(\hat{\varphi}^\rho)=\frac{1}{2}[\hat{\varphi}^\rho-\hat{\varphi}(\hat{\varphi}^\rho)]^T\cdot[\hat{\varphi}^\rho-\hat{\varphi}(\hat{\varphi}^\rho)]$.

定理4、5的证明类似于文献[3]中定理2、3的证明，这里从略.

由上述 $D^{(l)}$ 的几种取法及相应算法收敛性的讨论可见，算法(19)十分简单，算法(20)虽增加了部分计算量但易于满足收敛条件. 它们的 $D^{(l)}$ 均为分块对角形式，使得整个算法便于用 N 个处理机并行实现，从而大大缩短计算时间. 算法(22)复杂些，但全局收敛性较好. 下面的定理表明这些算法的解与一般非线性方法解的等价性.

定理 6 设协调算法(15)收敛于点 $\hat{\varphi}^*$，则点 $\hat{\varphi}^*$ 满足(2)的最优性必要条件.

证 $\because \det(D^{(l)})\neq 0$，$\therefore$ 收敛时有 $\hat{\varphi}^\rho=\hat{\varphi}=\hat{\varphi}^*$，将它们代入(14)、(12)立即得证. 证毕.

4. 仿真与应用

为检验本文所提算法的有效性，在 Dual System 83/20机上对它作了仿真和应用研究. 采用的仿真模型为

$$y(t)=\sum_{i=1}^{N}\varphi_{i1}x_{i2}^{\varphi_{i2}}(t)x_{i3}^{\varphi_{i3}}(t)+u(t), \tag{23}$$

这里，$u(t)$ 为正态白噪声序列，均值为零方差 $\sigma^2=25$，相应的信噪比约为1∶0.2，样本 S 的点数 $t_f=252$，并将它均分为 N 个子集 $S_i, i=1,\cdots,N$，参数的初值均为零，真值见[3]. 仿真结果见表1及表2.

表 1 $N=3$时的仿真结果

使用算法		收敛迭代次数	机时	$\max\limits_{i,j}\lvert\varphi_{ij}-\hat{\varphi}_{ij}^{*}\rvert/\lvert\varphi_{ij}\rvert$
L—M 法		28	16′41″	6.77×10^{-4}
[3]的目标协调法($C^{(i)}=\frac{1}{\omega}I$)		9	7′05″	3.64×10^{-4}
本文算法	$D^{(i)}=I, s=1$	10	4′55″	8.55×10^{-4}
	$D^{(i)}=A^{(i)}$	8	4′59″	0.952×10^{-5}
	$D^{(i)}$由(21)定义	9	8′36″	1.73×10^{-4}

表 2 $N=4$时的仿真结果

使用算法		收敛迭代次数	机时	$\max\limits_{i,j}\lvert\varphi_{ij}-\hat{\varphi}_{ij}^{*}\rvert/\lvert\varphi_{ij}\rvert$
L—M 法		37	25′51″	8.18×10^{-4}
[3]的目标协调法($C^{(i)}=\frac{1}{\omega}I$)		16	8′58″	2.15×10^{-4}
本文算法	$D^{(i)}=I, s=1$	15	5′32″	9.12×10^{-4}
	$D^{(i)}=A^{(i)}$	10	5′17″	7.35×10^{-4}
	$D^{(i)}$由(21)定义	9	6′03″	0.99×10^{-4}

(表 1 及表 2 中:φ_{ij}为真值,$\hat{\varphi}_{ij}^{*}$ 为估计值)

然后,这些算法被用于求解某地区环境—经济模型中 12 个部门的生产函数的参数估计问题.生产函数的数学形式为

$$G = a_1K_1^{\varphi_{11}}L_1^{\varphi_{12}}T_1^{\varphi_{13}} + a_2K_2^{\varphi_{21}}L_2^{\varphi_{22}}T_2^{\varphi_{23}} + a_3K_3^{\varphi_{31}}L_3^{\varphi_{32}}T_3^{\varphi_{33}}, \tag{24}$$

这里,G:工业总产值;K_i:固定资产原值;L_i:职工人数;T_i:工程技术人员数,而 $a_i,\varphi_{i1},\varphi_{i2},\varphi_{i3}$为待估计参数($i=1,2,3$).各算法所得的诸生产函数参数估计值相一致,而它们的平均迭代次数和平均机时见表 3.

表 3 平均迭代次数和平均机时

使用算法	收敛平均迭代次数	平均机时
L—M 法	52	11′32″
文献[3]算法(混合法)	11	6′55″
本文算法($D^{(i)}=A^{(i)}$)	8	3′59″

由表 1—3 可见,本文所提算法其收敛性确实优于整体算法和目标协调法。

5. 结　论

理论分析表明,本文利用目标补偿原理导出的递阶算法,其第一级便于利用现有的解最小二乘问题的高效算法,而协调级算法又十分简单,且具有较宽的适用范围和较好的收敛性.仿真及应用结果说明,它的计算量小,收敛速度快,既优于整体方法,又因实现了样本的分解,比

文献[3]的算法提高了一步,因此可以认为它是求解非线性系统最小二乘估计问题的一种更为有效的算法. 应当指出,本文的仿真及应用是在单机系统上实现的,若在多处理机系统上进行,可望获得更好的结果.

参 考 文 献

[1] Al—Baali,M. , and Fletcher,R. ,Variational Methods for Nonlinear Least—Squares, J. Opl Res. Soc. ,36,(1985),405—421.

[2] Ruhe, A. , Accelerated Gauss—Newton Algorithms for Nonlinear Least Squares Problems,BIT,19,(1979),356—367.

[3] He,J. M. ,and Da ,Q. L. ,The Multilevel Parameter Estimation Approach for a Kind of Nonlinear Models with Applications, Preprints of 8th IFAC/IFORS Symposium on Identification and System Parameter Estimation,1,(1988),627—632.

[4] Ortega,J. M. ,and Rheinboldt,W. C. ,Iterative Solution of Nonlinear Eguations in Several Variables, Academic Press, New York, (1970),300—301.

An Objective Compensated Approach least Squares Estimation in Nonlinear Systems

He Jianmin, Da Qingli

(School of Management,Southeast University,Nanjing)

Abstract: In this paper a new decomposition—coordination approach is proposed for solving the least—squares estimation problems in nonlinear systems. It coordinates the subproblems by introducing proper compensation terms into their objectivesand has the advantages of simplified calculation procedure and the least—squares forms of the subproblems. The convergence of the approach is proved in several patterns. The results of simulation and application show that the approach has fast convergent rate and is better than the global and the goal coordination methods.

Key words: parameter estimation; nonlinear systems; objective compensation; coordination approach; least — squares problems