第 11 卷第 3 期
1994 年 6 月

控制理论与应用
CONTROL THEORY AND APPLICATIONS

Vol. 11, No. 3
Jun., 1994

改进的非线性系统最小二乘算法

侯忠生　韩志刚
(黑龙江大学应用数学研究所·哈尔滨,150080)

摘要: 本文给出了可以统一处理线性系统,非线性系统参数辨识的改进最小二乘算法,包括批量形式及递推形式,它是文[1,2]中算法的综合及推广,改进了收敛速度,能克服病态,算法简单,易于应用,并且给出了算法的收敛性证明.

关键词: 线性系统; 非线性系统; 系统辨识; 收敛性; 最小二乘问题

1 引　言

最小二乘算法及其各种变形在系统辨识中已经得到了广泛的应用. 它的基本思想是通过极小化系统估计的输出与系统实际输出的误差的平方和来求得系统未知参数的估值. 众所周知,在应用最小二乘算法时,要求系统的观测数据要满足一定的条件,比如持续激励等,初始方差的选取也得适当以便能增加收敛速度,尤其在病态情况下. 文[2]给出了在线性 ARMR 模型情况下,从最小二乘算法的准则的直接修改,得到了递推形式的改进最小二乘算法,给出了类似[1]的收敛性证明,证明了改进的最小二乘算法的收敛速度比最小二乘算法收敛速度快,且能克服病态. 另一类著名的参数估计算法是投影算法[1],其基本思想则是在约束条件下,求解参数变化率的极小而获得参数辨识. 本文给出的算法综合了最小二乘算法及投影算法两种形式的优点,从多目标的观点推导出一般非线性系统改进的最小二乘辨识算法,它是[2]的推广. 它可以统一处理线性及一般非线性系统的辨识问题,改进了收敛速度,并且能克服病态.

2 一般非线性系统的辨识算法推导

2.1 一般非线性系统的离线辨识

考虑下述一般带有白噪声及多维控制输入的非线系统模型

$$y(k+1) = f[Y(k), U(k), \theta(k+1), k] + \varepsilon_k. \tag{1}$$

其中 $Y(k)$, $U(k)$ 分别表示到时刻 k 为止的系统输出集合和系统输入向量集合, θ 表示系统的未知参数向量, $f[\cdots]$ 表示已知形式的非线性函数, ε_k 是干扰,一般的最小二乘准则为

$$J_1(\theta) = \sum_{i=1}^{N+1} [f_i(Y(i-1), U(i-1), \theta(i-1), i-1) - y(i)]^2.$$

其中 $N+1$ 表示数据长度. 求 $J_1(\theta)$ 对 θ 的极小值,并用某些近似的线性化处理,就可以得到关于 θ 的批量非线性最小二乘辨识算法. 当系统是线性的时候,即

$$y(k) = \Phi^{\mathrm{T}}(k-1)\theta.$$

那么上述过程所得到的就是著名的线性最小二乘算法,但上述两种情形可能出现系统是

本文于1992年11月20日收到. 1993年9月15日收到修改稿.

病态情况，或者使算法无法进行或收敛速度过慢，或者估计出来的参数变化过快，此部分原因可能由于噪声或传感器误差使得检测数据出现个别的严重不准[3]. 特别在非线性情况下，为了得到简单，易于计算的算法，都要在工作点处进行线性替代，从而由于参数变化过大，或者说方差过大，使得线性替代的范围过大，超过了合理的范围，精度下降，有时甚至发散，为此将目标函数 $J_1(\theta)$ 改写成为如下形式

$$J_2(\theta)=\sum_{i=1}^{N+1}[f(Y(i-1),U(i-1),\theta,i-1)-y(i)]^2+\mu\|\theta-\hat{\theta}(k)\|^2. \tag{2}$$

上式的 $\mu\|\theta-\hat{\theta}(k)\|^2$ 的作用就是限制参数 θ 的变化，也即限制了非线性系统用动态的线性系统替代的范围. 假定已选定了初始点 $\theta(0)$，经过迭代已求得 $\hat{\theta}(k)$，现在考虑 $\hat{\theta}(k+1)$ 的求法. 上述算法准则折衷了最小二乘算法与投影算法两种准则的优点，当系统是线性时就变成文[2]的结果.

为了能够求解准则函数(2)的极小点，避免繁杂的非线性求解的数值运算，将非线性系统在当前点 $\hat{\theta}(k)$ 处线性化，即在每一个迭代点上用切平面来替代原来的非线性系统，从而使得非线性系统用时变的动态线性系统来逼近，即

$$\begin{aligned} f_i(Y(i-1),U(i-1),\theta,i-1)\simeq& f_i(Y(i-1),U(i-1),\hat{\theta}(k),i-1)\\ &+\nabla f_i^{\mathrm{T}}(\hat{\theta}(k))\Delta\hat{\theta}(k). \end{aligned} \tag{3}$$

其中 $\Delta\hat{\theta}(k)=\theta-\hat{\theta}(k)$，$\nabla f_i(\hat{\theta}(k))$ 表示

$$\left.\frac{\partial f_i(Y(i-1),U(i-1),\theta,i-1)}{\partial\theta}\right|_{\theta=\hat{\theta}(k)},\quad i=1,2,\cdots,N+1.$$

如果令

$$Y(N+1)=[y(1),\cdots,y(N+1)]^{\mathrm{T}},$$

$$F_{N+1}=[f_1(Y(0),U(0),\theta(1),0),\cdots,f_{N+1}(Y(N),U(N),\theta(N+1),N)]^{\mathrm{T}}.$$

则可将(2)写成向量形式，即

$$J_2(\theta)=\|Y(N+1)-F_{N+1}\|^2+\mu\|\theta-\hat{\theta}(k)\|^2. \tag{4}$$

(3)式写成

$$F_{N+1}(Y(N),U(N),\theta,N)\simeq F_{N+1}(Y(N),U(N),\hat{\theta}(k),N)+\Psi(\hat{\theta}(k))(\theta-\hat{\theta}(k)). \tag{5}$$

其中 $\Psi(\hat{\theta}(k))$ 是 F_{N+1} 在 $\hat{\theta}(k)$ 处的 Jacobi 矩阵，即

$$\Psi(\hat{\theta}(k))=\begin{bmatrix}\nabla f_1(\cdots,\hat{\theta}(k),\cdot)^{\mathrm{T}}\\ \vdots\\ \nabla f_{N+1}(\cdots,\hat{\theta}(k),\cdot)^{\mathrm{T}}\end{bmatrix}$$

$$=\begin{bmatrix}\dfrac{\partial\nabla f_1(\cdots,\hat{\theta}(k),\cdot)^{\mathrm{T}}}{\partial\theta_1} & \cdots & \dfrac{\partial\nabla f_1(\cdots,\hat{\theta}(k),\cdot)^{\mathrm{T}}}{\partial\theta_n}\\ \vdots & & \vdots\\ \dfrac{\partial\nabla f_{N+1}(\cdots,\hat{\theta}(k),\cdot)^{\mathrm{T}}}{\partial\theta_1} & \cdots & \dfrac{\partial\nabla f_{N+1}(\cdots,\hat{\theta}(k),\cdot)^{\mathrm{T}}}{\partial\theta_n}\end{bmatrix}.$$

将(5)式代入(4)式，并求关于 θ 的极小值，整理可得

$$\begin{aligned}\hat{\theta}(k+1)=&\hat{\theta}(k)+[\mu I+\Psi^{\mathrm{T}}(\hat{\theta}(k))\Psi^{\mathrm{T}}(\hat{\theta}(k))]^{-1}\Psi(\hat{\theta}(k))^{\mathrm{T}}\\ &\cdot[Y(N+1)-F_{N+1}(Y(N),U(N),\hat{\theta}(k),N)].\end{aligned} \tag{6}$$

上述(6)式就是对一般非线性系统(1)式的批量改进的最小二乘算法.

注 1　算法(6)中的形式就是优化中的阻尼最小二乘法，而阻尼最小二乘法中的 μI 项是仅从计算角度加进去的，并且证明了此项的引入改进了算法的数值计算条件，见文[4]，没有具体的物理意义. 由于 $\mu(\hat{\theta}(k))$ 与 $\hat{\theta}(k)$ 有关，故不能写成递推形式.

注 2　$[\mu I+\Psi^{T}(\hat{\theta}(k))\Psi(\hat{\theta}(k))]$ 这个矩阵，即使 $\Psi^{T}(\hat{\theta}(k)\Psi(\hat{\theta}(k))$ 是奇异矩阵，整个矩阵也将是非奇异矩阵，只要选取 μ 适当地大. 为了避免求逆可以用其它的数值方法求解方程

$$[\mu I+\Psi^{T}(\hat{\theta}(k))\Psi(\hat{\theta}(k))]P=\Psi^{T}(\hat{\theta}(k))[Y(N+1)-F_{N+1}(Y(N),U(N),\hat{\theta}(k),N)].$$

注 3　当 $\mu=0$ 时算法变成

$$\hat{\theta}(k+1)=\hat{\theta}(k)+(\Psi^{T}(\hat{\theta}(k))\Psi(\hat{\theta}(k)))^{-1}\Psi^{T}(\hat{\theta}(k))[Y(N+1)-F_{N+1}(Y(N),U(N),\hat{\theta}(k),N)].$$

即 Gauss-Newton 算法，$\Psi^{T}(\hat{\theta}(k)\Psi(\hat{\theta}(k))$ 可以看成 Hessian 矩阵的一种近似.

注 4　当 $\mu\to\infty$ 时，$[\mu I+\Psi^{T}(\hat{\theta}(k))\Psi(\hat{\theta}(k))]^{-1}\simeq(\mu I)^{-1}$ 上述算法就变成最速下降方法.

注 5　在算法(6)中，取 $\mu I=P^{-1}(0)$ 时，系统取成线性系统时，此算法就变成文[2]中的算法.

2.2　一般非线性系统(1)的在线递推辨识算法

算法(6)的注 1 中说明了上述算法不能写成递推形式，然而在实践中，我们经常用到的是递推辨识算法[3]. 假定 $\hat{\theta}(k)$ 已经得到，并且它是准则(2)的最优解，现在又假设我们得到另一组数据 $y(N+1)$, $U(N)$，我们希望在线求出 $\hat{\theta}(k+1)$.

令 $J_N(\theta)=\sum_{i=1}^{N}[f(Y(i-1),U(i-1),\theta(i),i-1)-y(i)]^2+\mu(\theta-\hat{\theta}(N-1))^2$. 简记 $f(Y(i-1),U(i-1),\theta(i),i-1)$ 为 $\hat{y}(i,\theta)$，那么上式就变成

$$J_N(\theta)=\sum_{i=1}^{N}[y(i)-\hat{y}(i,\theta)]^2+\mu(\theta-\hat{\theta}(N-1))^2. \tag{7}$$

而

$$J_{N+1}(\theta)=J_N(\theta)+(y(N+1)-\hat{y}(N+1,\theta))^2+\mu[(\theta-\hat{\theta}(N))^2-(\theta-\hat{\theta}(N-1))^2]. \tag{8}$$

用 2.1 中一样的思想，用时变的动态线性切平面即线性模型来逼近一般的非线性系统，即在当前点 $\hat{\theta}(N)$ 处对 $\hat{y}(N+1,\theta)$ 进行一阶台劳展开得

$$\hat{y}(N+1,\theta)\simeq\hat{y}(N+1,\hat{\theta}(N))+\hat{y}'(N+1,\hat{\theta}(N))^{T}(\theta-\hat{\theta}(N)). \tag{9}$$

将(9)式代入(8)式，并令

$$\hat{y}'(N+1,\hat{\theta}(N))\triangleq\Psi(N),$$

$$Z(N+1)\triangleq y(N+1)-\hat{y}(N+1,\hat{\theta}(N))+\Psi^{T}(N)\hat{\theta}(N).$$

得

$$J_{N+1}(\theta)=J_N(\theta)+(Z(N+1)-\Psi^{T}(N)\theta]^2+\mu[(\theta-\hat{\theta}(N))^2-(\theta-\hat{\theta}(N+1))^2]. \tag{10}$$

对(10)式两边关于 θ 求导得

$$\dot{J}_{N+1}(\theta)\simeq\dot{J}_N(\theta)-\Psi(N)[Z(N+1)-\Psi^{T}(N)\theta]-\mu[\hat{\theta}(N)-\hat{\theta}(N-1)]. \tag{11}$$

将 $\dot{J}_N(\theta)$ 在 $\hat{\theta}(N)$ 处一阶台劳展开得

$$\dot{J}_N(\theta)\simeq\dot{J}_N(\hat{\theta}(N))+\ddot{J}_N(\hat{\theta}(N))(\theta-\hat{\theta}(N)). \tag{12}$$

将(12)式代入(11)式,并利用 $\hat{\theta}(N)$是 $J_N(\theta)$的极小值这个条件得

$$\dot{J}_{N+1}(\theta) \simeq \ddot{J}_N(\hat{\theta}(N))(\theta - \hat{\theta}(N)) - \Psi(N)[Z(N+1) - \Psi^T(N)\theta] - \mu[\hat{\theta}(N) - \hat{\theta}(N-1)].$$

设 $\hat{\theta}(N+1)$是 $J_{N+1}(\theta)$的极小值,并记 $\Delta\hat{\theta}(N-1)=\hat{\theta}(N)-\hat{\theta}(N-1)$,则由上式整理可得

$$\hat{\theta}(N+1) = \hat{\theta}(N) + P(N)\Psi(N)(Z(N+1) - \Psi^T(N)\hat{\theta}(N)) + \mu P(N)\Delta\hat{\theta}(N-1). \tag{13}$$

其中

$$P(N) = [\ddot{J}_N(\hat{\theta}(N)) + \Psi(N)\Psi^T(N)]^{-1}. \tag{14}$$

为了推导关于 $P(N)$的递推公式,对(8)式两边关于 θ 求导得

$$\ddot{J}_{N+1}(\theta) = \ddot{J}_N(\theta) + \Psi(N)\Psi^T(N).$$

故

$$P^{-1}(N) = \ddot{J}_{N+1}(\theta).$$

由(14)式,利用矩阵求逆引理可以得到

$$P(N) = P(N-1) - \frac{P(N-1)\Psi(N)\Psi^T(N)P(N-1)}{1 + \Psi^T(N)P(N)\Psi(N)}.$$

从而就可以得到改进的最小二乘算法的递推形式

$$\hat{\theta}(k+1) = \hat{\theta}(k) + P(k)\Psi(k)[y(k+1) - f[Y(k),U(k),\hat{\theta}(k),k)] + \mu P(k)\Delta\hat{\theta}(k-1), \tag{15}$$

$$P(k) = P(k-1) - \frac{P(k-1)\Psi(k)\Psi^T(k)P(k-1)}{1 + \Psi^T(k)P(k)\Psi(k)}, \tag{16}$$

$$\Psi(k) = \left.\frac{\partial f[Y(k),U(k),\theta,k)]}{\partial\theta}\right|_{\theta} = \hat{\theta}(k). \tag{17}$$

注 此算法仅是在一般的非线性最小二乘算法的基础上增加了一项,$\mu P(k)\Delta\hat{\theta}(k-1)$,计算上几乎不增加计算量,形式简单,当 $\mu=0$ 时,就变成普通的非线性最小二乘算法[1].

3 一类对参数线性的非线性系统

下面考虑一类对参数线性的非线性系统,它具有比较广泛的意义,许多非线性模型均可以化成此类,如著名的 NARMAX 模型,Hammerstein 模型,输出仿射模型等,此时模型可以写成

$$y(k+1) = \Phi^T(k)\theta, \tag{18}$$

其中

$$\Phi(k) = (f_1(Y(k),U(k),k),\cdots,f_n(Y(k),U(k),k))^T,$$

$$\theta = (\theta_1,\cdots,\theta_n)^T,$$

$f_i(\cdots)$, $i=1,2,\cdots,n$ 是与未知参数 θ 无关的已知形式的非线性函数,在此种情况下,经过类似的推导,(15)～(17)就变成

$$\hat{\theta}(k+1) = \hat{\theta}(k) + P(k)\Phi(k)[y(k+1) - \Phi^T(k)\hat{\theta}(k)] + \mu P(k)\Delta\hat{\theta}(k-1),$$

$$P(k) = P(k-1) - \frac{P(k-1)\Phi(k)\Phi^T(k)P(k-1)}{1 + \Phi^T(k)P(k-1)\Phi(k)}. \tag{19}$$

4 算法的性质分析

下面我们讨论算法(6)的收敛性,对模型(1),假设 $f[\cdots]$对 θ 有连续的一阶偏导数,且存在常数向量 θ_0,使得

$$y(k+1)=f[Y(k),U(k),\theta_0,k]. \tag{20}$$

令 $\tilde{\theta}(k)=\hat{\theta}(k)-\theta_0$，由算法(6)可得

$$\tilde{\theta}(k+1)=\tilde{\theta}(k)+[\mu I+\Psi^{\mathrm{T}}(\hat{\theta}(k))\Psi(\hat{\theta}(k))]^{-1}\Psi^{\mathrm{T}}(\hat{\theta}(k))$$
$$\cdot[Y(N+1)-F_{N+1}(Y(N),U(N),\hat{\theta}(k),k)].$$

将(20)代入上式，并利用微分学中值定理可得

$$\tilde{\theta}(k+1)=\tilde{\theta}(k)+[\mu I+\Psi^{\mathrm{T}}(\hat{\theta}(k))\Psi(\hat{\theta}(k))]^{-1}\Psi^{\mathrm{T}}(\hat{\theta}(k))\Psi(\hat{\theta}_{0k})\tilde{\theta}(k). \tag{21}$$

其中 $\Psi(\hat{\theta}_{0k})$表示 f 在 θ_0 到 $\hat{\theta}(k)$之间某点处的 Jacobi 矩阵.

设 $F(k)\triangleq I-[\mu I+\Psi^{\mathrm{T}}(\hat{\theta}(k))\Psi(\hat{\theta}(k))]^{-1}\Psi^{\mathrm{T}}(\hat{\theta}(k))\Psi(\bar{\theta}_{0k})$，则(21)式可以写成如下形式

$$\tilde{\theta}(k+1)=F(k)\tilde{\theta}(k). \tag{22}$$

反复利用(22)式，对某一正整数 j，置

$$G(k+j,k)=\prod_{t=0}^{j-1}F(k+t).$$

得

$$\tilde{\theta}(k+j)=G(k+j,k)\tilde{\theta}(k).\quad j\geqslant 0. \tag{23}$$

定理 1 设 $M(k,\bar{\theta}_{0k})=I-F^{\mathrm{T}}(k)F(k)$，$N_l(k)=\sum_{i=1}^{l-1}G(k+i,k)^{\mathrm{T}}M(k+i,\bar{\theta}_{0k+i})G(k+i,k)$，$l$ 是某一正整数，假定 $N_l(k)\geqslant cI$，$0<c<1$，则参数辨识算法(6)按指数收敛于 θ_0.

证 定义 $V(k)=\tilde{\theta}^{\mathrm{T}}(k)\tilde{\theta}(k)$，则由(22)式得

$$V(k+1)=\tilde{\theta}^{\mathrm{T}}(k+1)\tilde{\theta}(k+1)=\tilde{\theta}^{\mathrm{T}}(k)F^{\mathrm{T}}(k)F(k)\tilde{\theta}(k).$$

由 $M(k,\bar{\theta}_{0k})$的定义可知上式可写成

$$V(k+1)=V(k)-\tilde{\theta}(k)^{\mathrm{T}}M(k,\bar{\theta}_{0k})\tilde{\theta}(k). \tag{24}$$

对某一正的整数 l，将(24)式相加 l 次，并利用(23)式可得

$$V(k+l)=V(k)-\sum_{i=0}^{l-1}\tilde{\theta}^{\mathrm{T}}(k)G(k+i,k)^{\mathrm{T}}M(k+i,\bar{\theta}_{0k+i})G(k+i,k)\tilde{\theta}(k)$$
$$=V(k)-\tilde{\theta}^{\mathrm{T}}(k)N_l(k)\tilde{\theta}(k). \tag{25}$$

容易验证 $M(k,\bar{\theta}_{0k})$及 $N_l(k)$是实对称矩阵，由于实对称矩阵可以用正交变换化成对角形矩阵，故假设 $N_l(k)\geqslant cI$ 是合理的. 从而由假设条件及(25)式可得

$$V(k+l)=V(k)-\tilde{\theta}^{\mathrm{T}}(k)N_l(k)\tilde{\theta}(k)\leqslant(1-c)V(k).$$

由于当 $c<1$ 时有$(1-c)<e^{-c}$，故有

$$V(k+l)\leqslant e^{-c}V(k).$$

此即意味着参数的指数收敛性.

对于模型(18)，算法(19)，我们可以完全类似文[1]中第 60 页或文[2]中定理 1 的证明，得到如下定理，证明见[1,2].

定理 2 对模型(18)，算法(19)，有

i) $\|\hat{\theta}(k)-\theta_0\|^2\leqslant\|\hat{\theta}_{(0)}-\theta_0\|^2$，对 $k>0$.

ii) a) $\lim_{k\to\infty}\|P(k)\varphi(k)e(k)\|=0$，其中 $e(k)=y(k)-\varphi^{\mathrm{T}}(k)\hat{\theta}(k-1)$.

b) $\lim_{k\to\infty}\|\hat{\theta}(k)-\hat{\theta}(k-l)\|=0$，对 $l<\infty$.

定理 3 如果普通的最小二乘法与改进的最小二乘法的初值一样，那么我们有

$$\|\tilde{\theta}(k)\|^2\leqslant\|\tilde{\theta}'(k)\|^2.$$

其中 $\tilde{\theta}(k)$和 $\tilde{\theta}'(k)$分别表示算法(19)与普通的最小二乘法的算法估值与真值的差.

证 与文[2]中的定理2类似,略.

此结果将文[2]中的结论推广到更广泛的模型类(18)上.对一般的模型(1)的结论证明有待于研究.

5 结 论

本文给出了非线性最小二乘算法的改进形式,并且给出了收敛性证明.说明了改进形式与普通非线性最小二乘算法的区别及其改进形式的优越性,算法简单,几乎不增加计算量,推广文[2]的结果到非线性系统上来.

参 考 文 献

[1] Goodwin, G. C. and Sin, K. S.. Adaptive Filtering, Prediction and Control. Englewood Cliffs, New Jersey; Prentice-Hall Inc. 1984

[2] Chen Yen-Ming and Wu Yungchun. Modified Recursive Least-Squares Algorithm for Parameter Identification. Int. J. System Sci., 1992, 23(2):187—205

[3] Ljung, L. and Soderstrom, T.. Theory and Practive of Recursive Identification. MIT Prese, 1983

[4] 邓乃扬.无约束最优化计算方法.北京:科学出版社,1982

Modified Nonlinear System Least-Squares Algorithm

HOU Zhongsheng and HAN Zhigang

(The Institute of Applied Mathematics, Heilongjiang University · Harbin, 150080, PRC)

Abstract: This paper presents a modified nonlinear system least-squares algorithm which can deal with the linear and the nonlinear system parameter identification unitedly, it concludes batch algorithm and recursive algorithm. It is an extension and combination of the results of refernces [1,2], and of higher convergence speed, and also applicable for the ill-conditioned situations. This algorithm is simple, easy to use, and its' convergences are discussed, too.

Key words: linear system; nonlinear system; system identification; convergence; least-squares problem

本文作者简介

侯忠生 1962年生.1982年获理学学士学位,1988年获理学硕士学位.现为东北大学自动控制系91级博士研究生.研究方向为非线性系统的辨识,自适应控制及无模型学习自适应控制.

韩志刚 1934年生.1958年毕业于吉林大学,现任黑龙江大学教授.博士生导师.主要从事系统辨识,自适应预报与控制的研究.曾提出多层递阶方法.目前研究领域为结构与参数皆时变的系统的辨识及自适应控制和无模型控制.