高校应用数学学报A辑 Appl. Math. J. Chinese Univ. Ser. A
2000,15(1):113～118

递推阻尼最小二乘法

林茂琼　陈增强　袁著祉

（南开大学计算机与系统科学系，天津 300071）

摘　要　递推最小二乘法是参数辨识中最常用的方法，但容易产生参数爆发现象. 本文推导了一种更稳定的辨识方法——阻尼最小二乘法的递推求解算法.

关键词　系统辨识，最小二乘法，递推算法，矩阵求逆引理.

分类号　（中图）O241.2；（1991MR）93B30.

§1　引　言

在参数辨识中，递推最小二乘法(RLS)是用得最多的一种算法. 但最小二乘法存在一些缺点[1]，如随着协方差矩阵的减小，参数易产生爆发现象；参数向量和协方差矩阵的初值选择不当会使得辨识过程在参数收敛之前结束；在存在随机噪声的情况下，参数易产生漂移，出现不稳定等. 为防止参数产生爆发现象，Levenberg[2]提出在参数优化算法中增加一个阻尼项以增加算法的稳定性. 本文在通常递推最小二乘法的目标函数上增加了对参数变化量的阻尼项，并推导了其递推算法.

§2　阻尼最小二乘法(DLS)

考虑单输入-单输出系统：

$$y(t)+\sum_{i=1}^{na}a_iy(t-i)=\sum_{i=1}^{nb}b_iu(t-i). \tag{1}$$

我们的问题是根据输入、输出数据来确定未知参数 $a_1,a_2,\cdots,a_{na},b_1,b_2,\cdots,b_{nb}$. 但是在实际的测量中，测量到的数据总是有误差的. 这包含了测量噪声、模型误差等. 因此，实际测得的输入、输出数据之间的关系应修正为：

$$y(t)+\sum_{i=1}^{na}a_iy(t-i)=\sum_{i=1}^{nb}b_iu(t-i)+e(t), \tag{2}$$

其中 $e(t)$ 称为模型误差或残差.

若令

收稿：1998-06-09. 修回：1998-11-27.

$$\boldsymbol{\varphi}^{\mathrm{T}}(t)=[-y(t-1),\cdots,-y(t-na),u(t-1),\cdots,u(t-nb)], \tag{3}$$

$$\boldsymbol{\theta}^{\mathrm{T}}=[a_1,a_2,\cdots,a_{na},b_1,b_2,\cdots,b_{nb}], \tag{4}$$

则(2)式可写成向量形式：

$$y(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\theta+e(t). \tag{5}$$

这时，对于给定的阶次 n，基于模型(1)式的最小二乘估计问题可以表述如下：通过对系统测量到的输入输出测量值 $\{u(k),y(k),k=t-N,\cdots,t\}$，按照如下优化准则函数确定 θ 的估计量：

$$J=\sum_{k=t-N}^{t}\beta^{t-k}[y(k)-\boldsymbol{\varphi}^{\mathrm{T}}(k)\hat{\boldsymbol{\theta}}(t)]^2+\mu\|\hat{\boldsymbol{\theta}}(t)-\hat{\boldsymbol{\theta}}(t-1)\|^2, \tag{6}$$

其中，$0<\beta\leqslant1$ 为遗忘因子. 选择 β 的不同值就可得到不同的遗忘效果. β 越小，遗忘速度越快，或者说记忆越短. $\mu>0$ 为阻尼因子，其大小标志自变量增量与函数值之间在 J 取极小时的相对重要性. 即如果模型的线性程度较大，那么对于很小的 μ，由(6)求得的 $\hat{\boldsymbol{\theta}}(t)$ 可以对 $\boldsymbol{\theta}$ 有较好的修正；如果模型的线性程度较差，那么必须 μ 较大，才能保证由(6)求得的 $\boldsymbol{\theta}(t)$ 对 $\boldsymbol{\theta}$ 有较好的修正.

指标 J 可以写成向量的形式：

$$J=[\boldsymbol{Y}(t)-\boldsymbol{\Phi}(t)\hat{\boldsymbol{\theta}}(t)]^{\mathrm{T}}\boldsymbol{W}(t)[\boldsymbol{Y}(t)-\boldsymbol{\Phi}(t)\hat{\boldsymbol{\theta}}(t)]+\mu[\hat{\boldsymbol{\theta}}(t)-\hat{\boldsymbol{\theta}}(t-1)]^{\mathrm{T}}[\hat{\boldsymbol{\theta}}(t)-\hat{\boldsymbol{\theta}}(t-1)], \tag{7}$$

其中

$$\boldsymbol{\Phi}(t)=[\boldsymbol{\varphi}^{\mathrm{T}}(t-N+1),\boldsymbol{\varphi}^{\mathrm{T}}(t-N+2),\cdots,\boldsymbol{\varphi}^{\mathrm{T}}(t)]^{\mathrm{T}}$$
$$=\begin{bmatrix}-y(t-N) & \cdots & -y(t-N+1-na) & u(t-N) & \cdots & u(t-N+1-nb)\\ -y(t-N+1) & \cdots & -y(t-N+2-na) & u(t-N+1) & \cdots & u(t-N+2-nb)\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ -y(t-1) & \cdots & -y(t-na) & u(t-1) & \cdots & u(t-nb)\end{bmatrix} \tag{8}$$

$$\boldsymbol{Y}(t)=[y(t-N+1),y(t-N+2),\cdots,y(t)]^{\mathrm{T}},$$
$$\boldsymbol{W}(t)=\mathrm{diag}(\beta^{N-1},\beta^{N-2},\cdots,\beta^{0}).$$

§3 递推算法

递推算法的基本思想是：每取得一个新数据，就根据新数据对原估计量进行修正，从而得到改善的新估计量，而不是将新数据加到老数据里重新计算. 现在来分析增加新数据后，最小二乘估计量的变化.

假定根据 t 时刻采样数据得到的最小二乘估计量为 $\hat{\boldsymbol{\theta}}(t)$.

定理1 以(6)为准则函数对模型(1)进行参数估计可由以下递推公式得到：

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\beta\mu\boldsymbol{P}(t)[\hat{\boldsymbol{\theta}}(t-1)-\hat{\boldsymbol{\theta}}(t-2)]+\boldsymbol{P}(t)\boldsymbol{\varphi}(t)[y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)], \tag{9}$$

$$\boldsymbol{P}(t)=\frac{1}{\beta}\left[\boldsymbol{P}'(t)-\frac{\boldsymbol{P}'(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}'(t)}{\beta+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}'(t)\boldsymbol{\varphi}(t)}\right], \tag{10}$$

$$\boldsymbol{P}'(t)=\boldsymbol{P}(t-1)-\sum_{i=1}^{n}\frac{\boldsymbol{P}'_{i-1}(t-1)\boldsymbol{r}_i\boldsymbol{r}_i^{\mathrm{T}}\boldsymbol{P}'_{i-1}(t-1)\mu'}{1+\boldsymbol{r}_i^{\mathrm{T}}\boldsymbol{P}'_{i-1}(t-1)\boldsymbol{r}_i\mu'},\tag{11}$$

$$\boldsymbol{P}'_i(t-1)=\boldsymbol{P}'_{i-1}(t-1)-\frac{\boldsymbol{P}'_{i-1}(t-1)\boldsymbol{r}_i\boldsymbol{r}_i^{\mathrm{T}}\boldsymbol{P}'_{i-1}(t-1)\mu'}{1+\boldsymbol{r}_i^{\mathrm{T}}\boldsymbol{P}'_{i-1}(t-1)\boldsymbol{r}_i\mu'},\tag{12}$$

$$\boldsymbol{P}'_0(t-1)=\boldsymbol{P}(t-1),\tag{13}$$

其中 $\boldsymbol{r}_i$ 是 $\boldsymbol{r}_1$ 的后继向量,$\boldsymbol{r}_1=[1,0,\cdots,0]^{\mathrm{T}}$, (14)

$$\mu'=(1-\beta)\mu/\beta.\tag{15}$$

证　将(7)式关于 $\hat{\theta}$ 求偏导数并令其为零,得

$$[\mu\boldsymbol{I}+\boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Phi}(t)]\hat{\theta}(t)=\mu\hat{\theta}(t-1)+\boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{W}(t)\boldsymbol{Y}(t),\tag{16}$$

因为 $\mu\boldsymbol{I}+\boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Phi}(t)$ 可逆,所以(16)式存在唯一解:

$$\hat{\theta}(t)=[\mu\boldsymbol{I}+\boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Phi}(t)]^{-1}[\mu\hat{\theta}(t-1)+\boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{W}(t)\boldsymbol{Y}(t)],\tag{17}$$

称 $\hat{\theta}(t)$ 为 $\theta(t)$ 的阻尼最小二乘估计. 令

$$\boldsymbol{P}^{-1}(t-1)=[\mu\boldsymbol{I}+\boldsymbol{\Phi}^{\mathrm{T}}(t-1)\boldsymbol{W}(t-1)\boldsymbol{\Phi}(t-1)],\tag{18}$$

则

$$\boldsymbol{P}(t)=[\mu\boldsymbol{I}+\boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Phi}(t)]^{-1},\tag{19}$$

$$\hat{\theta}(t)=\boldsymbol{P}(t)[\mu\hat{\theta}(t-1)+\boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{W}(t)\boldsymbol{Y}(t)],\tag{20}$$

其中

$$\boldsymbol{\Phi}(t)=\begin{bmatrix}\boldsymbol{\Phi}(t-1)\\ \boldsymbol{\varphi}^{\mathrm{T}}(t)\end{bmatrix},\boldsymbol{W}(t)=\begin{bmatrix}\beta\boldsymbol{W}(t-1) & 0\\ 0 & 1\end{bmatrix},\boldsymbol{Y}(t)=\begin{bmatrix}\boldsymbol{Y}(t-1)\\ y(t)\end{bmatrix},\tag{21}$$

将(21)式代入(20)式,得

$$\hat{\theta}(t)=\boldsymbol{P}(t)[\mu\hat{\theta}(t-1)+\beta\boldsymbol{\Phi}^{\mathrm{T}}(t-1)\boldsymbol{W}(t-1)\boldsymbol{Y}(t-1)+\boldsymbol{\varphi}(t)y(t)],\tag{22}$$

考虑到

$$\hat{\theta}(t-1)=\boldsymbol{P}(t-1)[\mu\hat{\theta}(t-2)+\boldsymbol{\Phi}^{\mathrm{T}}(t-1)\boldsymbol{W}(t-1)\boldsymbol{Y}(t-1)],\tag{23}$$

可得:

$$\boldsymbol{\Phi}^{\mathrm{T}}(t-1)\boldsymbol{W}(t-1)\boldsymbol{Y}(t-1)=\boldsymbol{P}^{-1}(t-1)\hat{\theta}(t-1)-\mu\hat{\theta}(t-2).\tag{24}$$

将(18)式和(24)式代入(22)式,得

$$\begin{aligned}\hat{\theta}(t)=&\boldsymbol{P}(t)[\mu\hat{\theta}(t-1)+\beta\boldsymbol{P}^{-1}(t-1)\hat{\theta}(t-1)-\beta\mu\hat{\theta}(t-2)+\boldsymbol{\varphi}(t)y(t)]=\\ &\boldsymbol{P}(t)[\mu\hat{\theta}(t-1)+\boldsymbol{P}(t)^{-1}\hat{\theta}(t-1)-\mu\hat{\theta}(t-1)+\beta\mu\hat{\theta}(t-1)-\\ &\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\theta}(t-1)-\beta\mu\hat{\theta}(t-2)+\boldsymbol{\varphi}(t)y(t)]=\\ &\hat{\theta}(t-1)+\beta\mu\boldsymbol{P}(t)[\hat{\theta}(t-1)-\hat{\theta}(t-2)]+\boldsymbol{P}(t)\boldsymbol{\varphi}(t)[y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\theta}(t-1)],\end{aligned}\tag{25}$$

到此,(9)式得证.

将(21)式和(18)式代入(19)式得:

$$\begin{aligned}\boldsymbol{P}(t)=&[\mu\boldsymbol{I}+\beta\boldsymbol{\Phi}^{\mathrm{T}}(t-1)\boldsymbol{W}(t-1)\boldsymbol{\Phi}(t-1)+\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)]^{-1}=\\ &[\mu\boldsymbol{I}-\beta\mu\boldsymbol{I}+\beta\boldsymbol{P}^{-1}(t-1)+\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)]^{-1}.\end{aligned}\tag{26}$$

记

$$\boldsymbol{P}(t)=\beta[\mu\boldsymbol{I}-\beta\mu\boldsymbol{I}+\beta\boldsymbol{P}^{-1}(t-1)]^{-1},\tag{27}$$

将(27)式代入(26)式,得:

$$\boldsymbol{P}(t)=[\beta\boldsymbol{P}'(t)^{-1}+\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)]^{-1}= \\ \frac{1}{\beta}[\boldsymbol{P}'(t)-\boldsymbol{P}'(t)\boldsymbol{\varphi}(t)[\beta+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}'(t)\boldsymbol{\varphi}(t)]^{-1}\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}'(t)], \tag{28}$$

利用矩阵求逆引理(见附录),可得:

$$\boldsymbol{P}(t)=\frac{1}{\beta}\left[\boldsymbol{P}'(t)-\frac{\boldsymbol{P}'(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}'(t)}{\beta+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}'(t)\boldsymbol{\varphi}(t)}\right], \tag{29}$$

由此(10)式得证.记

$$\mu'=(1-\beta)\mu/\beta, \tag{30}$$

则由(27)式得:

$$\boldsymbol{P}'(t)^{-1}=\mu'\boldsymbol{I}+\boldsymbol{P}^{-1}(t-1)=\sum_{i=1}^{n}\boldsymbol{r}_i\boldsymbol{r}_i^{\mathrm{T}}\mu'+\boldsymbol{P}^{-1}(t-1), \tag{31}$$

其中,$\boldsymbol{r}_i$ 是 $\boldsymbol{r}_1$的后继向量,$\boldsymbol{r}_1=[1,0,\cdots,0]^{\mathrm{T}}$.令

$$\boldsymbol{P}'_n(t-1)=\boldsymbol{P}'(t), \tag{32}$$

则

$$\boldsymbol{P}'_n(t-1)=\left[\boldsymbol{P}^{-1}(t-1)+\sum_{i=1}^{n}\boldsymbol{r}_i\boldsymbol{r}_i^{\mathrm{T}}\mu'\right]^{-1}. \tag{33}$$

利用矩阵求逆引理可得:

$$\begin{aligned}\boldsymbol{P}'_n(t-1)=&[\boldsymbol{P}'_{n-1}(t-1)^{-1}+\boldsymbol{r}_n\boldsymbol{r}_n^{\mathrm{T}}\mu]^{-1}= \\ &\boldsymbol{P}'_{n-1}(t-1)-\frac{\boldsymbol{P}'_{n-1}(t-1)\boldsymbol{r}_n\boldsymbol{r}_n^{\mathrm{T}}\boldsymbol{P}'_{n-1}(t-1)\mu}{1+\boldsymbol{r}_n^{\mathrm{T}}\boldsymbol{P}'_{n-1}(t-1)\boldsymbol{r}_n\mu}= \\ &\boldsymbol{P}'_{n-2}(t-1)-\frac{\boldsymbol{P}'_{n-2}(t-1)\boldsymbol{r}_{n-1}\boldsymbol{r}_{n-1}^{\mathrm{T}}\boldsymbol{P}'_{n-2}(t-1)\mu}{1+\boldsymbol{r}_{n-1}^{\mathrm{T}}\boldsymbol{P}'_{n-1}(t-1)\boldsymbol{r}_{n-1}\mu}- \\ &\frac{\boldsymbol{P}'_{n-2}(t-1)\boldsymbol{r}_n\boldsymbol{r}_n^{\mathrm{T}}\boldsymbol{P}'_{n-2}(t-1)\mu}{1+\boldsymbol{r}_n^{\mathrm{T}}\boldsymbol{P}'_{n-1}(t-1)\boldsymbol{r}_n\mu}=\cdots= \\ &\boldsymbol{P}'_0(t-1)-\sum_{i=1}^{n}\frac{\boldsymbol{P}'_{i-1}(t-1)\boldsymbol{r}_i\boldsymbol{r}_i^{\mathrm{T}}\boldsymbol{P}'_{i-1}(t-1)\mu}{1+\boldsymbol{r}_i^{\mathrm{T}}\boldsymbol{P}'_{i-1}(t-1)\boldsymbol{r}_i\mu},\end{aligned} \tag{34}$$

其中,

$$\boldsymbol{P}'_i(t-1)=\boldsymbol{P}'_{i-1}(t-1)-\frac{\boldsymbol{P}'_{i-1}(t-1)\boldsymbol{r}_i\boldsymbol{r}_i^{\mathrm{T}}\boldsymbol{P}'_{i-1}(t-1)\mu'}{1+\boldsymbol{r}_i^{\mathrm{T}}\boldsymbol{P}'_{i-1}(t-1)\boldsymbol{r}_i\mu'},\quad i=1,\cdots,n, \tag{35}$$

$$\boldsymbol{P}'_0(t-1)=\boldsymbol{P}(t-1). \tag{36}$$

到此,(11),(12),(13)式得证.

§4 仿真研究

分别用递推最小二乘法和递推阻尼最小二乘法对下列系统进行仿真研究:

$$y(t)-1.6y(t-1)+0.7y(t-2)=u(t-1)+5.5u(t-2)+\xi(t). \tag{37}$$

为了检验辨识算法的鲁棒稳定性,辨识时选择降阶模型:

$$y(t)+a_1y(t-1)=b_0u(t-1)+b_1u(t-2). \tag{38}$$

用递推最小二乘法辨识时选择参数:$\beta=0.95$;

用递推阻尼最小二乘法辨识时选择参数:$\beta=0.95$,$\mu=0.95$.得到的辨识参数曲线分别如以下4个图所示.

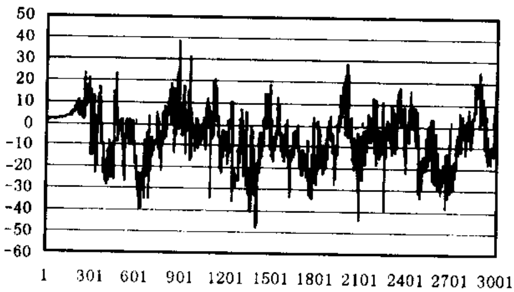

图1 用 RLS 辨识参数 b_0 的曲线图

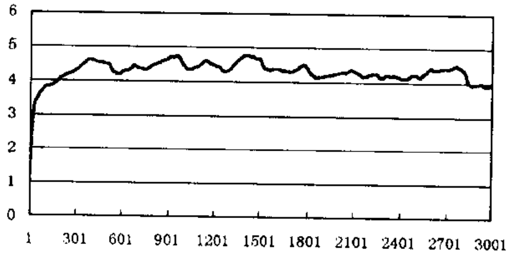

图2 用 DLS 辨识参数 b_0 的曲线图

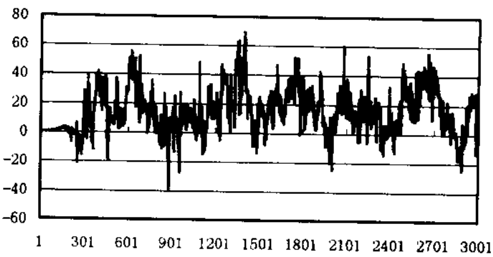

图3 用 RLS 辨识参数 b_1 的曲线图

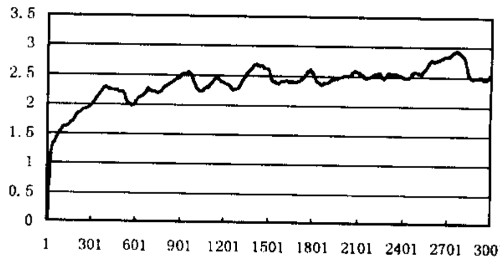

图4 用 DLS 辨识参数 b_1 的曲线图

从以上四图可以看出,递推阻尼最小二乘法能有效地防止参数爆发现象.

参 考 文 献

1 Spripada, N. R., Fisher, D. G., Improved least squares identification, Internat. J. Control, 1987, 46(6):1889～1913.

2 Levenberg, K., A method for the solution of certain non-linear problems in least squares, Quart. Appl. Math., 1944, 2:164～168.

附录 矩阵求逆引理

若 $P_2^{-1}=P_1^{-1}+HR^{-1}H^{\mathrm{T}}$,则 $P_2=P_1-P_1H(H^{\mathrm{T}}P_1H+R)^{-1}H^{\mathrm{T}}P_1$.

RECURSIVE DAMPED LEAST SQUARE ALGORITHM

Lin Maoqiong Chen Zengqiang and Yuan zhuzhi

(*Department of Computer & System Science, Nankai University, Tianjin* 300071)

Abstract The recursive least square is widely used in parameter identification. But it is easy to bring about parameters' burst-off. The recursive process of a more stable identification algorithm—— damped least square are proposed.

Keywords System Identification, Least Square, Recursive Algorithm, Inverse Matrix Theorem.

Subject Classification (CL)O241.2;(1991MR)93B30.