应用数学和力学编委会编
重 庆 出 版 社 出 版

文章编号: 1000_0887(2000) 02_0209_06

递推阻尼最小二乘法的收敛性与稳定性*

陈增强, 林茂琼, 袁著祉

(南开大学 计算机与系统科学系, 天津 300071)

(王彪推荐)

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· 因此对一种更稳定的辨识方法——递推阻尼最小二乘法进行了收敛特性的分析· 在使用算法之前先归一化测量向量,结果表明, 参数化距离收敛于一个零均值随机变量, 并且在持续激励条件下, 适应增益矩阵的条件数有界· 参数化距离的方差有界·

关 键 词: 系统辨识; 阻尼最小二乘法; 递推算法; 收敛性; 稳定性
中图分类号: O231; O241 文献标识码: A

引 言

系统辨识是一种数学建模的新方法,这种方法在工程上已获得日益广泛的应用· 其最主要的应用背景是工业过程自动控制· 我们可以在不知道复杂工业过程内部机理的条件下,利用过程的输入输出信息, 通过参数辨识来建立过程的数学模型· 因而研究参数辨识方法的收敛性具有重要的理论意义和实际意义·

递推最小二乘法是参数辨识中用得最多的一种算法, G. C. Goodwin、R. Lozano、L. Ljung 等曾对最小二乘法进行了收敛性分析[1]~[3]· 但它存在一些缺点[4], 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, K. Levenberg 提出在参数优化算法中增加一个阻尼项增加算法的稳定性[5]· 本文推导了带有阻尼项的递推最小二乘法辨识算法, 并给出了其收敛性与稳定性分析·

1 递推阻尼最小二乘法

考虑单输入_单输出系统:

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

其中, $w(t)$ 为高斯白噪声·

若令

$$\varphi^{\mathrm{T}}(t)=[-y(t-1),\cdots,-y(t-na),u(t-1),\cdots,u(t-nb)], \tag{2}$$

* 收稿日期: 1998_09_03; 修订日期: 1999_01_04
基金来源: 国家 863CIMS 计划资助课题(863_511_945_010)
作者简介: 陈增强(1964~), 男, 天津人, 教授, 工学博士.

$$\theta_0^{\mathrm{T}} = [a_1, a_2, \cdots, a_{na}, b_1, b_2, \cdots, b_{nb}], \tag{3}$$

则(1)式可写成向量形式:

$$y(t) = \varphi^{\mathrm{T}}(t)\theta_0 + w(t), \tag{4}$$

将系统归一化,则

$$\|\varphi(t)\|^2 \leqslant 1. \tag{5}$$

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:通过对系统量测到的输入输出量测值 $\{y(k), u(k), k = t-T, \cdots, t\}$,按照如下优化准则函数确定 θ_0 的估计量:

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

其中,$0 < \beta \leqslant 1$ 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. $\mu \geqslant 0$ 为阻尼因子,调整 μ,控制参数 $\hat{\theta}$ 的变化率.

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

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:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \beta\mu \boldsymbol{P}(t)[\hat{\theta}(t-1) - \hat{\theta}(t-2)] + \boldsymbol{P}(t)\varphi(t)[y(t) - \varphi^{\mathrm{T}}(t)\hat{\theta}(t-1)], \tag{7}$$

$$\boldsymbol{P}(t)^{-1} = \mu(1-\beta)I + \beta\boldsymbol{P}(t-1)^{-1} + \varphi(t)\varphi(t)^{\mathrm{T}}. \tag{8}$$

对 $\boldsymbol{P}(t)$ 进行递推求解,得

$$\boldsymbol{P}(t) = \frac{1}{\beta}\left[\boldsymbol{P}'(t) - \frac{\boldsymbol{P}'(t)\varphi(t)\varphi^{\mathrm{T}}(t)\boldsymbol{P}'(t)}{\beta + \varphi^{\mathrm{T}}(t)\boldsymbol{P}'(t)\varphi(t)}\right], \tag{9}$$

$$\boldsymbol{P}'(t) = \boldsymbol{P}(t-1) - \sum_{i=1}^{n} \frac{\boldsymbol{P}'_{i-1}(t-1)\boldsymbol{r}_i\boldsymbol{r}_i^{\mathrm{T}}\boldsymbol{P}'_{i-1}(t-1)\mu'}{1 + \boldsymbol{r}_i^{\mathrm{T}}\boldsymbol{P}'_{i-1}(t-1)\boldsymbol{r}_i\mu'}, \tag{10}$$

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

$$\boldsymbol{P}'_0(t-1) = \boldsymbol{P}(t-1), \tag{12}$$

其中 $\boldsymbol{r}_i$ 是 $\boldsymbol{r}_1$ 的后继向量,

$$\boldsymbol{r}_1 = [1, 0, \cdots, 0]^{\mathrm{T}}, \tag{13}$$

$$\mu' = (1-\beta)\mu/\beta. \tag{14}$$

证明:(略)

2 收敛性与稳定性

定义 $\{F(t)\}$ 为 $\{y(t), u(t)\}$ 生成的子 σ 代数的增序列,并且引进下述关于序列 $\{w(t)\}$ 的基本假定:

$$E\{w(t) \mid F(t-1)\} = 0 \quad \text{a.s.}, \tag{15}$$

$$E\{w(t)^2 \mid F(t-1)\} = \sigma^2 < \infty \quad \text{a.s.}. \tag{16}$$

定理 2 考虑用于归一化系统(1)的在(15)、(16)式假定下的算法(7)~(14). 若

(a) 持续激励

$$\sum_{i=t+1}^{t+N} \varphi(i)\varphi(i)^{\mathrm{T}} \geqslant \delta_1 I > 0 \quad \forall t \ \text{a.s.}, \tag{17}$$

其中 N 是 $\varphi(t)$ 的维数,$N = na + nb$.

(b) $0<\beta\leqslant 1-\varepsilon,\ \varepsilon>0,\ \mu^2\leqslant\delta_1^2\beta^{2N+3}$, (18)

则算法(7)~(14)具有如下特性:

(i) $\lambda_{\min}P(t)^{-1}\geqslant p_{\min}=\delta_1\beta^N>0,\qquad t\geqslant N\quad \text{a.s.}$, (19)

(ii) $\lambda_{\max}P(t)^{-1}\leqslant p_{\max}=1/\varepsilon<\infty$, (20)

(iii) $\lim\limits_{t\to\infty}E\left\{\theta(t)\mid F(t-1)\right\}=0\qquad \text{a.s.}$, (21)

(iv) $\lim\limits_{t\to\infty}\sup E\left\{\theta(t)^{\mathrm{T}}\theta(t)\right\}\leqslant\dfrac{\beta p_{\max}p_{\min}}{\beta p_{\min}^2-\mu^2}\sigma^2<\infty\qquad \text{a.s.}$, (22)

其中,$\lambda_{\min}P(t)^{-1}$ 和 $\lambda_{\max}P(t)^{-1}$ 分别表示 $P(t)^{-1}$ 的最小和最大的特征值,$\theta(t)=\theta_0-\hat{\theta}(t)$ 表示参数估计值与真值间的距离•

证明

(i) (8)式可写成:

$$P(t)^{-1}=\mu(1-\beta^t)I+\beta^tP(0)^{-1}+\sum_{i=0}^{t-1}\beta^i\varphi(t-i)\varphi(t-i)^{\mathrm{T}},\tag{23}$$

则

$$\lambda_{\min}P(t)^{-1}\geqslant\lambda_{\min}\left[\sum_{i=0}^{t-1}\beta^i\varphi(t-i)\varphi(t-i)^{\mathrm{T}}\right]\geqslant\lambda_{\min}\left[\sum_{i=t-N}^{t-1}\beta^i\varphi(t-i)\varphi(t-i)^{\mathrm{T}}\right]•\tag{24}$$

由于 $t-N\leqslant i\leqslant t-1$, 所以

$$\beta^{t-i}\geqslant\beta^N•\tag{25}$$

由(17)、(18)得

$$\lambda_{\min}P(t)^{-1}\geqslant\delta_1\beta^N=p_{\min}>0\qquad t\geqslant N\ \text{a.s.},\tag{26}$$

(i) 式得证•

(ii) 设矩阵 $Q(t)^{-1}$ 由下式给出:

$$Q(t)^{-1}=\mu(1-\beta)I+\beta Q(t-1)^{-1}+I/N,\quad Q(0)^{-1}=P(0)^{-1},\tag{27}$$

因此 $\quad \mathrm{tr}Q(t)^{-1}=N\mu(1-\beta)+\beta\,\mathrm{tr}Q(t-1)^{-1}+1,\qquad (28)$

其中,$\mathrm{tr}Q(t)^{-1}$ 表示 $Q(t)^{-1}$ 的迹•

由(8)式得:

$$\mathrm{tr}P(t)^{-1}\leqslant N\mu(1-\beta)+\beta\,\mathrm{tr}P(t-1)^{-1}+\|\varphi(t)\|^2,\tag{29}$$

因为 $\|\varphi(t)\|^2\leqslant 1$, 假定 $t-1$ 时刻成立:

$$\mathrm{tr}P(t-1)^{-1}\leqslant\mathrm{tr}Q(t-1)^{-1},\tag{30}$$

则由归纳法可得:

$$\mathrm{tr}P(t)^{-1}\leqslant\mathrm{tr}Q(t)^{-1}•\tag{31}$$

现在的目的是要建立 $\mathrm{tr}Q(t)^{-1}$ 的如下形式的上界:

$$\mathrm{tr}Q(t)^{-1}\leqslant 1/\varepsilon•\tag{32}$$

用归纳法加以证明• 假定 $t-1$ 时刻成立,即:

$$\mathrm{tr}Q(t-1)^{-1}\leqslant 1/\varepsilon,\tag{33}$$

要得到(32)式,由(28)式得:

$$N\mu(1-\beta)+\beta\,\mathrm{tr}Q(t-1)^{-1}+1\leqslant 1/\varepsilon•\tag{34}$$

由于 $N\mu(1-\beta)\geqslant 0$, 所以

$$\beta \operatorname{tr} \boldsymbol{Q}(t-1)^{-1}+1 \leqslant 1 / \varepsilon \cdot \tag{35}$$

β 应满足下式:

$$\beta \leqslant [\operatorname{tr} \boldsymbol{Q}(t-1)^{-1}]^{-1}(1-\varepsilon) / \varepsilon, \tag{36}$$

因此, 可选择如下的 β:

$$\beta \leqslant 1-\varepsilon=\left(\frac{1}{\varepsilon}\right)^{-1} \frac{1-\varepsilon}{\varepsilon} \leqslant [\operatorname{tr} \boldsymbol{Q}(t-1)^{-1}]^{-1} \frac{1-\varepsilon}{\varepsilon}, \tag{37}$$

可得

$$\lambda_{\max} \boldsymbol{P}(t)^{-1} \leqslant \operatorname{tr} \boldsymbol{P}(t)^{-1} \leqslant \operatorname{tr} \boldsymbol{Q}(t)^{-1} \leqslant 1 / \varepsilon, \tag{38}$$

(ii) 式得证•

从(28) 式可得

$$\operatorname{tr} \boldsymbol{Q}(t)^{-1}=\beta^{t} \operatorname{tr} \boldsymbol{Q}(0)^{-1}+[1+N \mu(1-\beta)] \sum_{i=0}^{t-1} \beta^{i} \leqslant \frac{1}{\varepsilon}, \tag{39}$$

所以

$$\sum_{i=0}^{t-1} \beta^{i} \leqslant \frac{1}{\varepsilon}=p_{\max }, \tag{40}$$

(40) 式将在证明(iv) 时用到•

(iii) 令

$$\theta(t)=\theta_{0}-\hat{\theta}(t), \tag{41}$$

由(4) 、(7) 和(41) 式可得:

$$\begin{aligned} \theta(t)= & [\boldsymbol{I}+\beta \mu \boldsymbol{P}(t)-\boldsymbol{P}(t) \varphi(t) \varphi(t)^{\mathrm{T}}] \theta(t-1)- \\ & \beta \mu \boldsymbol{P}(t) \theta(t-2)-\boldsymbol{P}(t) \varphi(t) w(t) \cdot \end{aligned} \tag{42}$$

由(8) 式得:

$$\boldsymbol{I}+\beta \mu \boldsymbol{P}(t)-\boldsymbol{P}(t) \varphi(t) \varphi(t)^{\mathrm{T}}=\mu \boldsymbol{P}(t)+\beta \boldsymbol{P}(t) \boldsymbol{P}(t-1)^{-1} \cdot \tag{43}$$

将(43) 式代入(42) 式, 且等式两边同乘以 $\boldsymbol{P}(t)^{-1}$ 得:

$$\begin{aligned} & \boldsymbol{P}(t)^{-1} \theta(t)-\mu \theta(t-1)= \\ & \quad \beta[\boldsymbol{P}(t-1)^{-1} \theta(t-1)-\mu \theta(t-2)]-\varphi(t) w(t)= \\ & \quad \beta^{t-1}[\boldsymbol{P}(1)^{-1} \theta(1)-\mu \theta(0)]-\sum_{i=0}^{t-2} \beta^{i} \varphi(t-i) w(t-i) \cdot \end{aligned} \tag{44}$$

令

$$\boldsymbol{P}(1)^{-1} \theta(1)-\mu \theta(0)=\boldsymbol{a}, \tag{45}$$

代入(44) 式, 且等式两边同乘以 $\boldsymbol{P}(t)$ 得:

$$\theta(t)=\mu \boldsymbol{P}(t) \theta(t-1)+\beta^{t-1} \boldsymbol{P}(t) \boldsymbol{a}-\boldsymbol{P}(t) \sum_{i=0}^{t-2} \beta^{i} \varphi(t-i) w(t-i) \cdot \tag{46}$$

因为 $\varphi(t)$ 与 $w(t)$ 无关, 所以

$$\begin{aligned} & E\{\theta(t) \mid F(t-1)\}=\mu^{t} \prod_{i=0}^{t-1} \boldsymbol{P}(t-i) E\{\theta(0)\}+ \\ & \quad \sum_{i=0}^{t-1} \mu^{i} \beta^{t-i-1} \prod_{j=0}^{i} \boldsymbol{P}(t-j) \boldsymbol{a} \cdot \end{aligned} \tag{47}$$

由(19) 、(47) 式知:

$$\begin{aligned} & \| E\{\theta(t) \mid F(t-1)\} \| \leqslant \mu^{t} \prod_{i=0}^{t-1} \| \boldsymbol{P}(t-i) \| \| E\{\theta(0)\} \|+ \\ & \quad \beta^{t-1} \sum_{i=0}^{t-1}\left(\frac{\mu}{\beta}\right)^{i} \prod_{j=0}^{i} \| \boldsymbol{P}(t-j) \| \| \boldsymbol{a} \| \leqslant \end{aligned}$$

$$\mu^t(\lambda_{\max}\boldsymbol{P}(t))^t\|E\{\theta(0)\}\|+\beta^{t-1}\sum_{i=0}^{t-1}\left(\frac{\mu}{\beta}\right)^i(\lambda_{\max}\boldsymbol{P}(t))^i\|\boldsymbol{a}\|\leqslant$$
$$\left(\frac{\mu}{\delta_1\beta^N}\right)^t\|E\{\theta(0)\}\|+\beta^{t-1}\sum_{i=0}^{t-1}\left(\frac{\mu}{\delta_1\beta^{N+1}}\right)^i\|\boldsymbol{a}\|\cdot \tag{48}$$

对(48)式两端取极限,并由(18)式,可得:

$$\lim_{t\to\infty}E\{\theta(t)\mid F(t-1)\}=0\qquad \text{a.s.}, \tag{49}$$

(iii)式得证•

(iv) 根据(46)式,可得:

$$\theta(t)^{\mathrm T}\boldsymbol{P}(t)^{-1}\theta(t)=\mu^2\theta(t-1)^{\mathrm T}\boldsymbol{P}(t)\theta(t-1)+2\mu\beta^{t-1}\boldsymbol{a}^{\mathrm T}\boldsymbol{P}(t)\theta(t-1)+$$
$$\sum_{i=0}^{t-2}z(t-i-1)w(t-i)+$$
$$\sum_{i=0}^{t-2}\beta^i\varphi(t-i)^{\mathrm T}w(t-i)\boldsymbol{P}(t)\sum_{i=0}^{t-2}\beta^i\varphi(t-i)w(t-i)+\beta^{2t-2}\boldsymbol{a}^{\mathrm T}\boldsymbol{P}(t)\boldsymbol{a}=$$
$$\mu^2\theta(t-1)^{\mathrm T}\boldsymbol{P}(t)\theta(t-1)+2\mu\beta^{t-1}\boldsymbol{a}^{\mathrm T}\boldsymbol{P}(t)\theta(t-1)+$$
$$\sum_{i=0}^{t-2}z_1(t-i-1)w(t-i)+$$
$$\sum_{i=0}^{t-2}\beta^{2i}\varphi(t-i)^{\mathrm T}\boldsymbol{P}(t)\varphi(t-i)w^2(t-i)+\beta^{2t-2}\boldsymbol{a}^{\mathrm T}\boldsymbol{P}(t)\boldsymbol{a}, \tag{50}$$

其中,$z(t-i-1)$, $z_1(t-i-1)$ 表示所有与 $w(t-i)$ 相乘的项的系数•

由(19)知 $\boldsymbol{P}(t)$ 有上界,所以:

$$\boldsymbol{a}^{\mathrm T}\boldsymbol{P}(t)\boldsymbol{a}\leqslant M\qquad (M\text{ 为常数}), \tag{51}$$
$$\boldsymbol{a}^{\mathrm T}\boldsymbol{P}(t)E\{\theta(t-1)\}\leqslant R\qquad (R\text{ 为常数})\cdot \tag{52}$$

由(20)得:

$$\lambda_{\max}\boldsymbol{P}(t)\leqslant 1/p_{\min}\cdot \tag{53}$$

由(9)、(10)式知:

$$\boldsymbol{P}(t)\leqslant\frac{1}{\beta}\boldsymbol{P}(t-1), \tag{54}$$

所以,由(50)、(51)、(52)、(54)式得:

$$\lambda_{\min}\boldsymbol{P}(t)^{-1}E\{\|\theta(t)\|^2\}\leqslant E\{\theta(t)^{\mathrm T}\boldsymbol{P}(t)^{-1}\theta(t)\}\leqslant$$
$$\frac{\mu^2}{\beta}E\{\theta(t-1)^{\mathrm T}\boldsymbol{P}(t-1)\theta(t-1)\}+2\mu\beta^{t-1}R+$$
$$\sum_{i=0}^{t-2}\beta^i\varphi(t-i)^{\mathrm T}\boldsymbol{P}(t-i)\varphi(t-i)E\{w(t-i)^2\}+\beta^{2t-2}M\cdot \tag{55}$$

考虑到:

$$\varphi(t)^{\mathrm T}\boldsymbol{P}(t)\varphi(t)=\varphi(t)^{\mathrm T}\frac{1}{\beta}\left[\boldsymbol{P}'(t)-\frac{\boldsymbol{P}'(t)\varphi(t)\varphi(t)^{\mathrm T}\boldsymbol{P}'(t)}{\beta+\varphi(t)^{\mathrm T}\boldsymbol{P}'(t)\varphi(t)}\right]\varphi(t)=$$
$$\frac{\varphi(t)^{\mathrm T}\boldsymbol{P}'(t)\varphi(t)}{\beta+\varphi(t)^{\mathrm T}\boldsymbol{P}'(t)\varphi(t)}\leqslant 1, \tag{56}$$

所以,由(19)、(40)、(53)和(56)式得:

$$E\{\|\theta(t)\|^2\}\leqslant\frac{\mu^2}{\beta p_{\min}^2}E\{\|\theta(t-1)\|^2\}+\frac{2\mu\beta^{t-1}}{p_{\min}}R+\frac{p_{\max}}{p_{\min}}\sigma^2+\frac{\beta^{2t-2}}{p_{\min}}M\leqslant$$

$$\left(\frac{\mu^2}{\beta p_{\min}^2}\right)^t E\left\{\|\theta(0)\|^2\right\} + \frac{2\mu R}{p_{\min}}\beta^{t-1}\sum_{i=0}^{t-1}\left(\frac{\mu}{\beta p_{\min}}\right)^{2i} +$$
$$\frac{p_{\max}}{p_{\min}}\sigma^2\sum_{i=0}^{t-1}\left(\frac{\mu^2}{\beta p_{\min}^2}\right)^i + \frac{\beta^{2t-2}M}{p_{\min}}\sum_{i=0}^{t-1}\left(\frac{\mu^2}{\beta^3 p_{\min}^2}\right)^i. \tag{57}$$

对(57)式两端取极限,并由(18)式得:

$$\lim_{t\to\infty} E\left\{\|\theta(t)\|^2\right\} \leqslant \frac{p_{\max}}{p_{\min}}\sigma^2 \lim_{t\to\infty}\sum_{i=0}^{t-1}\left(\frac{\mu^2}{\beta p_{\min}^2}\right)^i = \frac{\beta p_{\max}p_{\min}}{\beta p_{\min}^2-\mu^2}\sigma^2 \qquad \text{a. s.}, \tag{58}$$

(iv)式得证.

定理证毕.

3 结束语

本文对另一种辨识算法——递推阻尼最小二乘法进行了收敛特性分析. 结果表明,在持续激励条件下,参数化距离收敛于一个零均值、方差有界的随机变量,并且适应增益矩阵有界且正定.

[参 考 文 献]

[1] Goodwin G C. Adaptive Filtering Prediction and Control[M]. Englewood Cliffs: Prentice Hall, 1984.

[2] Lozanno R. Convergence analysis of recursive identification algorithms with forgetting factor[J]. Automatica, 1983, **19**(1): 95~97.

[3] Ljung L. Analysis of a general recursive prediction error identification algorithm[J]. Automatica, 1981, **17**(1): 89~99.

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

[5] Levenberg K. A method for the solution of certain non_linear problems in least squares[J]. Quart Appl Math, 1944, **26**(2): 164~168.

Convergence and Stability of Recursive Damped Least Square Algorithm

Chen Zengqiang, Lin Maoqiong, Yuan Zhuzhi

(Department of Computer and System Science, Nankai University, Tianjin 300071, P R China)

Abstract: The recursive least square is widely used in parameter identification. But it is easy to bring about the phenomena of parameters burst_off. A convergence analysis of a more stable identification algorithm_recursive damped least square is proposed. This is done by normalizing the measurement vector entering into the identification algorithm. It is shown that the parametric distance converges to a zero mean random variable. It is also shown that under persistent excitation condition, the condition number of the adaptation gain matrix is bounded, and the variance of the parametric distance is bounded.

Key words: system identification; damped least square; recursive algorithm; convergence; stability