DOI:10.13878/j.cnki.jnuist.2016.06.001

丁锋[1,2,3]

辅助模型辨识方法(6):性能分析

摘要

辨识方法的性能分析是系统辨识领域的重要而困难的研究课题.新的辨识方法一诞生,就伴随着收敛性分析.辅助模型辨识是辨识的一个分支,辅助模型辨识方法已成为一大类辨识方法族,关于其收敛性也提出了许多课题.本文研究了输出误差系统辅助模型随机梯度算法、辅助模型递推最小二乘算法、辅助模型多新息随机梯度算法、变间隔辅助模型随机梯度算法和变间隔辅助模型递推最小二乘算法的一致收敛性,近似分析了 Box-Jenkins 系统辅助模型递推广义增广最小二乘算法的性能.

关键词

参数估计;递推辨识;最小二乘;辅助模型辨识思想;多新息辨识理论;递阶辨识原理;耦合辨识概念;滤波辨识理念;线性系统

中图分类号 TP273

文献标志码 A

收稿日期 2016-11-13

资助项目 国家自然科学基金(61273194);江苏省自然科学基金(BK2012549);高等学校学科创新引智"111 计划"(B12018)

作者简介

丁锋,男,博士,教授,博士生导师,主要从事系统辨识、过程建模、自适应控制方面的研究.fding@ jiangnan.edu.cn

1 江南大学 物联网工程学院,无锡,214122

2 江南大学 控制科学与工程研究中心,无锡,214122

3 江南大学 教育部轻工过程先进控制重点实验室,无锡,214122

0 引言

辨识方法性能分析是辨识领域最吸引人的永恒研究主题,是辨识输入信号设计、新辨识方法提出、辨识方法性能分析三大基本问题之一[1-7].新颖的辨识方法都是蕴藏在科学云端的璀璨明珠.新辨识思想、新辨识理论、新辨识原理、新辨识概念的诞生都是辨识史上的重要里程碑,都预示着大量新颖辨识方法的涌现.20 世纪下半叶是一个预示着系统辨识取得辉煌的时代,预示着控制科学和系统科学极大发展的年代,也是控制科学、信息技术对人类生活方式产生极大变革的时代.在 20 世纪与 21 世纪交接的年代,辅助模型辨识思想、多新息辨识理论、递阶辨识原理、耦合辨识概念的出现都标志着一个个辨识分支的诞生.在每一个辨识分支中都将产生各种各样的辨识方法,其收敛性为辨识科学提供了色彩斑斓的研究空间.其中一些辨识方法的收敛性,以及一些有待深入研究的科学问题和科学难题都是呈现在读者面前的耀眼明珠,期盼着辨识科学家来采撷[3-5].

系统辨识研究极其繁荣,处理不可测变量的辅助模型辨识方法、提高辨识精度的多新息辨识方法、处理复杂结构和减小计算量的递阶辨识方法、处理参数耦合问题的多变量耦合辨识方法、处理有色噪声干扰的滤波辨识方法等都各自形成了一个辨识分支.它们在线性系统、线性参数系统、双线性参数系统、双线性系统和非线性系统中的应用,产生了形形色色的辨识方法,一些方法可参见综述论文《系统辨识算法复杂性、收敛性、计算效率研究》[6],《Modeling and identification for multirate systems》(双率采样数据系统的建模和辨识)[7],《时变系统辨识方法及其收敛定理》[8].

作为辨识的一个分支,辅助模型辨识形成了一大类辨识方法族,对其的收敛性也产生了许多值得深入研究的课题.一些辅助模型辨识方法收敛性研究的相关成果与研究课题如下:

1) 双率采样输出误差系统的辅助模型随机梯度辨识算法与辅助模型递推最小二乘辨识算法的收敛性[9-10];

2) 输出误差(output-error)系统的辅助模型递阶最小二乘(辅助模型两阶段递推最小二乘)辨识算法的收敛性[11];

3) 输出误差系统的辅助模型多新息随机梯度辨识算法和辅助模型多新息最小二乘辨识算法的一致收敛性[12-13];

4) 输出误差系统的辅助模型梯度迭代辨识算法和辅助模型最小

二乘迭代辨识算法的收敛性分析[14];

5) 输出误差滑动平均(OEMA)系统的辅助模型增广随机梯度算法和辅助模型递推增广最小二乘算法的收敛性分析[15-16];

6) 输出误差滑动平均系统的辅助模型增广梯度迭代算法和辅助模型增广最小二乘迭代算法的收敛性分析[14,17];

7) 输出误差自回归(OEAR)系统的辅助模型广义随机梯度辨识算法和辅助模型递推广义最小二乘辨识算法及其收敛性;

8) 输出误差自回归系统的辅助模型广义梯度迭代辨识算法和辅助模型广义最小二乘迭代辨识算法及其收敛性;

9) 输出误差自回归系统的滤波辅助模型广义随机梯度辨识算法和滤波辅助模型递推广义最小二乘辨识算法的收敛性[18];

10) Box-Jenkins 系统的辅助模型广义增广随机梯度算法和辅助模型递推广义增广最小二乘算法的收敛性[1-2,19];

11) Box-Jenkins 系统的分解辅助模型广义增广随机梯度算法和分解辅助模型递推广义增广最小二乘算法的收敛性[3,19];

12) Box-Jenkins 系统的滤波辅助模型广义增广随机梯度算法和滤波辅助模型递推广义增广最小二乘算法的收敛性[3];

13) Box-Jenkins 系统的辅助模型广义增广梯度迭代算法和辅助模型广义增广最小二乘迭代算法的收敛性[1,20-21];

14) Box-Jenkins 系统的分解辅助模型广义增广梯度迭代算法和分解辅助模型广义增广最小二乘迭代算法及其收敛性;

15) Box-Jenkins 系统的滤波辅助模型广义增广梯度迭代算法和滤波辅助模型广义增广最小二乘迭代算法及其收敛性;

16) Box-Jenkins 系统的辅助模型多新息广义增广随机梯度辨识算法与辅助模型多新息广义增广最小二乘算法及其收敛性[22];

17) 线性参数输出误差 ARMA 系统的辅助模型多新息广义增广随机梯度算法和辅助模型多新息广义增广最小二乘算法的收敛性[23];

18) 输入非线性输出误差自回归(IN-OEAR)系统的辅助模型多新息广义随机梯度算法和滤波辅助模型多新息广义随机梯度算法的收敛性[24-25];

19) 输出非线性输出误差类系统的辅助模型随机梯度辨识算法和辅助模型递推最小二乘辨识算法及其收敛性;

20) 类多变量输出误差类系统的辅助模型递阶多新息随机梯度算法和辅助模型递阶多新息最小二乘算法的收敛性;

21) 反馈非线性输出误差系统的辅助模型随机梯度辨识算法和辅助模型递推最小二乘辨识算法及其收敛性[26];

22) 反馈非线性输出误差系统的辅助模型梯度迭代辨识算法和辅助模型最小二乘迭代辨识算法及其收敛性[27];

23) 损失数据输出误差系统的辅助模型随机梯度辨识算法和辅助模型递推最小二乘辨识算法的收敛性[28-29];

24) 稀少量测数据输出误差系统的变间隔辅助模型多新息随机梯度辨识算法和变间隔辅助模型递推最小二乘辨识算法的收敛性[28];

25) 多变量输出误差系统的辅助模型随机梯度辨识算法和辅助模型递推最小二乘辨识算法[30-32];

26) 多输入单输出输出误差系统的辅助模型随机梯度辨识算法和辅助模型多新息最小二乘辨识算法及其收敛性[33];

27) 周期非均匀采样数据系统的辅助模型随机梯度类算法、辅助模型多新息随机梯度类辨识方法、辅助模型递推最小二乘辨识算法的收敛性[34-35];

28) 基于滤波和(或)基于分解的多元输出误差类系统和多元自回归输出误差系统的辅助模型随机梯度类辨识算法和辅助模型递推最小二乘类多新息辨识方法及其收敛性.

《系统辨识——辨识方法性能分析》研究了一些基本辨识方法的一致收敛性、有界收敛性等[2],《系统辨识——多新息辨识理论与方法》第 6 章研究了线性回归系统的随机梯度算法、多新息随机梯度算法、遗忘因子多新息随机梯度算法、多新息增广随机梯度算法、变间隔辅助模型多新息随机梯度算法、多新息最小二乘算法的收敛性[3].本文研究输出误差系统的几个辅助模型随机梯度算法和辅助模型递推最小二乘算法的一致收敛性,以及 Box-Jenkins 系统的辅助模型递推广义增广最小二乘算法的收敛性.限于篇幅,一些定理省略了冗长证明公式,定理的详细证明参见给出的参考文献和即将出版的《系统辨识——辅助模型辨识思想与方法》第 6 章[4].

1 辅助模型随机梯度辨识方法

《系统辨识——辨识方法性能分析》[2]和《International Journal of Computer Mathematics》2013 年第 2 期上论文《Performance analysis of the auxiliary model-based stochastic gradient parameter estimation algorithm for state-space systems with one-step state delay》(一步状态迟延状态空间系统对应输入输出表达的辅助模型随机梯度辨识方法性能分析)[36]用随机过程理论近似研究了辅助模型随机梯度辨识方法的收敛性,它用鞅收敛定理详细推导了输出误差系统的辅助模型随机梯度辨识方法的收敛定理,主要内容可参考笔者等发表在国际期刊《IEEE Transactions on Automatic Control》2005 年第 9 期上的论文《Parameter estimation of dual-rate stochastic systems by using an output error method》(双率采样输出误差系统的辅助模型随机梯度参数估计及其损失输出估计的收敛性)[10]和《自动化学报》2010 年第 7 期上的论文《Auxiliary model based stochastic gradient algorithm for multivariable output error systems》(多变量输出误差系统的辅助模型随机梯度算法及其收敛性)[32],以及文献[4].

1.1 辅助模型随机梯度算法

考虑输出误差模型(Output Error model,OE 模型)描述的随机系统(参见图 1):

$$y(t)=\frac{B(z)}{A(z)}u(t)+v(t),\tag{1}$$

其中$\{u(t)\}$和$\{y(t)\}$分别是系统的输入和输出序列,$\{v(t)\}$是零均值、方差为σ^2的随机白噪声序列,$G(z):=\dfrac{B(z)}{A(z)}$为系统的传递函数,$A(z)$和$B(z)$是单位后移算子z^{-1}的多项式:

$$A(z):=1+a_1z^{-1}+a_2z^{-2}+\cdots+a_{n_a}z^{-n_a},$$
$$B(z):=b_1z^{-1}+b_2z^{-2}+\cdots+b_{n_b}z^{-n_b}.$$

设阶次n_a和n_b已知,记$n:=n_a+n_b$,且$t\leqslant 0$时,$y(t)=0,u(t)=0,v(t)=0.\{u(t),y(t)\}$是测量数据,$v(t)$是不可测噪声.

定义中间变量(真实输出,即无噪输出):

$$x(t):=\frac{B(z)}{A(z)}u(t)\in\mathbf{R}.\tag{2}$$

定义参数向量$\boldsymbol{\theta}$和信息向量$\boldsymbol{\varphi}(t)$如下:

$$\boldsymbol{\theta}:=[a_1,a_2,\cdots,a_{n_a},b_1,b_2,\cdots,b_{n_b}]^{\mathrm{T}}\in\mathbf{R}^n,$$
$$\boldsymbol{\varphi}(t):=[-x(t-1),-x(t-2),\cdots,-x(t-n_a),$$

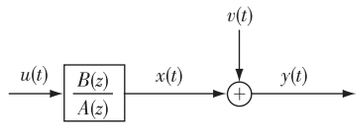


图 1 输出误差(OE)系统

Fig. 1 The output error system

$$u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}\in\mathbf{R}^n.$$

由式(2)和式(1),有

$$x(t)=[1-A(z)]x(t)+B(z)u(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta},\tag{3}$$
$$y(t)=x(t)+v(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t),\tag{4}$$

此即输出误差系统的辨识模型.

设$\hat{\boldsymbol{\theta}}(t)$是参数向量$\boldsymbol{\theta}$在时刻$t$的估计,用辅助模型的输出$x_{\mathrm{a}}(t-i)$构造未知信息向量$\boldsymbol{\varphi}(t)$的估计:

$$\hat{\boldsymbol{\varphi}}(t):=[-x_{\mathrm{a}}(t-1),-x_{\mathrm{a}}(t-2),\cdots,-x_{\mathrm{a}}(t-n_a),$$
$$u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}\in\mathbf{R}^n,$$

参照式(3),用$\hat{\boldsymbol{\varphi}}(t)$和$\hat{\boldsymbol{\theta}}(t)$构造辅助模型

$$x_{\mathrm{a}}(t)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t).$$

借助于上述辅助模型,在辨识算法中用$\hat{\boldsymbol{\varphi}}(t)$代替未知$\boldsymbol{\varphi}(t)$,我们能够获得辨识参数向量$\boldsymbol{\theta}$的辅助模型随机梯度算法(Auxiliary Model based Stochastic Gradient algorithm,AM-SG 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)}e(t),\quad \hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0,\tag{5}$$
$$e(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1),\tag{6}$$
$$r(t)=r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\quad r(0)=1,\tag{7}$$
$$x_{\mathrm{a}}(t)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\quad x_{\mathrm{a}}(-i)=1/p_0,$$
$$i=0,1,\cdots,n_a,\tag{8}$$
$$\hat{\boldsymbol{\varphi}}(t)=[-x_{\mathrm{a}}(t-1),\cdots,-x_{\mathrm{a}}(t-n_a),u(t-1),$$
$$\cdots,u(t-n_b)]^{\mathrm{T}}.\tag{9}$$

鞅收敛定理和鞅超收敛定理是研究时不变随机系统、时变随机系统辨识算法和自适应控制算法收敛性的有效工具.

1.2 AM-SG 辨识算法的收敛性

在《系统辨识——辨识方法性能分析》[2]的 4.6 节中,笔者在一个依赖于估计的信息向量$\hat{\boldsymbol{\varphi}}(t)$的强持续激励条件(SPE 条件)下,近似研究了AM-SG 算法的收敛性.因为 SPE 条件是否成立,依赖于算法计算出的参数估计,条件中的信息向量$\hat{\boldsymbol{\varphi}}(t)$是用参数估计计算构成的,因此无法肯定这个 SPE 条件成立.出于这一原因,这里运用鞅收敛定理证明 AM-SG 算法的一致收敛性,收敛条件依赖于系统的真实信息向量,这个结果具有普遍性,是 AM-SG 算法收敛性

的一个严格证明.这里的证明主要参考文献[4,10,32].

设 $\lambda_{\max}[\boldsymbol{X}]$ 为对称矩阵 $\boldsymbol{X}$ 的最大特征值;$\lambda_{\min}[\boldsymbol{X}]$ 为对称矩阵 $\boldsymbol{X}$ 的最小特征值;$\|\boldsymbol{X}\|$ 为矩阵 $\boldsymbol{X}$ 的范数,如定义为 $\|\boldsymbol{X}\|^2:=\mathrm{tr}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$.$f(t)=o(g(t))$ 表示 $g(t)>0,\lim\limits_{t\to\infty}\dfrac{f(t)}{g(t)}=0$;$f(t)=O(g(t))$ 表示 $g(t)\geqslant 0$,存在常数 $\delta_1>0$ 和 t_1 满足 $|f(t)|\leqslant\delta_1 g(t),t\geqslant t_1$.

假设 $\{v(t),\mathscr{F}_t\}$ 是定义在概率空间 $\{\Omega,\mathscr{F},P\}$ 上的鞅差序列,其中 $\{\mathscr{F}_t\}$ 是由直到时刻 t 的观测数据生成的 σ 代数序列.噪声序列 $\{v(t)\}$ 满足假设:

(A1) $E[v(t)|\mathscr{F}_{t-1}]=0$,a.s.,

(A2) $E[v^2(t)|\mathscr{F}_{t-1}]=\sigma^2 r^{\epsilon}(t-1)$,a.s.,

$\sigma^2<\infty$, $\epsilon<1$.

这里并没有假设干扰噪声 $\{v(t)\}$ 方差有界和高阶矩存在.

引理 1 对于 AM-SG 算法(5)—(9),下列不等式成立:

$$\sum_{j=1}^{t}\frac{\|\hat{\boldsymbol{\varphi}}(j)\|^2}{r(j)}\leqslant\ln r(t),\text{a.s.}.$$

定理 1[4,10] 对于输出误差系统(1)对应的辨识模型(4)和 AM-SG 辨识算法(5)—(9),定义数据乘积矩矩阵:

$$(\text{A3})\quad \boldsymbol{R}(t):=\sum_{j=1}^{t}\hat{\boldsymbol{\varphi}}(j)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(j)\in\mathbf{R}^{n\times n},$$

假设条件(A1)和(A2)成立,$A(z)$ 是严格正实的,$r(t):=\mathrm{tr}[\boldsymbol{R}(t)]\to\infty$,那么 AM-SG 算法给出的参数估计误差向量 $\hat{\boldsymbol{\theta}}(t)-\boldsymbol{\theta}$ 满足:

$$\|\hat{\boldsymbol{\theta}}(t)-\boldsymbol{\theta}\|^2=o\left(\frac{r(t)}{\lambda_{\min}[\boldsymbol{R}(t)]}\right),\text{a.s.}.$$

定理 2[4,10] 对于输出误差系统(1)对应的辨识模型(4)和 AM-SG 辨识算法(5)—(9),定理 1 的条件成立,定义数据乘积矩矩阵:

$$(\text{A4})\quad \boldsymbol{R}_0(t):=\sum_{j=1}^{t}\boldsymbol{\varphi}(j)\boldsymbol{\varphi}^{\mathrm{T}}(j)\in\mathbf{R}^{n\times n},$$

令 $r_0(t):=\mathrm{tr}[\boldsymbol{R}_0(t)]$,则有

$$\|\tilde{\boldsymbol{\theta}}(t)\|^2=o\left(\frac{r_0(t)}{\lambda_{\min}[\boldsymbol{R}_0(t)]}\right),$$

又假设 $r_0(t)=O(\lambda_{\min}[\boldsymbol{R}_0(t)])$,即

$$(\text{A5})\quad \limsup_{t\to\infty}\frac{r_0(t)}{\lambda_{\min}[\boldsymbol{R}_0(t)]}<\infty,$$

那么参数估计 $\hat{\boldsymbol{\theta}}(t)$ 一致收敛于真参数向量 $\boldsymbol{\theta}$,即 $\hat{\boldsymbol{\theta}}(t)\to\boldsymbol{\theta}$,a.s..

注 1 对于定理 2,假设对于大 t,存在正数 α 和 β 使得下列弱持续激励条件成立:

$$(\text{WPE})\quad \alpha\boldsymbol{I}_n\leqslant\frac{1}{t}\sum_{j=1}^{t}\boldsymbol{\varphi}(j)\boldsymbol{\varphi}^{\mathrm{T}}(j)\leqslant\beta\boldsymbol{I}_n,\text{a.s.},$$

我们有 $\alpha t\boldsymbol{I}_n\leqslant\boldsymbol{R}_0(t)\leqslant\beta t\boldsymbol{I}_n$,$\alpha nt\leqslant r_0(t)\leqslant\beta nt$,$\alpha t\leqslant\lambda_{\min}[\boldsymbol{R}(t)]\leqslant\beta t$,因此

$$\|\tilde{\boldsymbol{\theta}}(t)\|^2=o\left(\frac{r_0(t)}{\lambda_{\min}[\boldsymbol{R}_0(t)]}\right)=o\left(\frac{\beta nt}{\alpha t}\right)=o(1)\to 0,\text{a.s.},$$

即参数估计误差 $\|\tilde{\boldsymbol{\theta}}(t)\|^2$ 收敛于零.

注 2 在 AM-SG 辨识算法(5)—(9)中引入收敛指数(convergence index)ε,便得到修正辅助模型随机梯度算法(Modified AM-SG algorithm,M-AM-SG 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}(t)}{r^{\varepsilon}(t)}[y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],$$

$$\frac{1}{2}<\varepsilon\leqslant 1,\quad \hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0,\tag{10}$$

$$r(t)=r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\quad r(0)=1,\tag{11}$$

$$x_{\mathrm{a}}(t)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\quad x_{\mathrm{a}}(-i)=1/p_0,\quad i=0,1,\cdots,n_a,\tag{12}$$

$$\hat{\boldsymbol{\varphi}}(t)=[-x_{\mathrm{a}}(t-1),\cdots,-x_{\mathrm{a}}(t-n_a),u(t-1),\cdots,u(t-n_b)]^{\mathrm{T}}.\tag{13}$$

读者可以研究修正辅助模型随机梯度算法的收敛性.

注 3 在 AM-SG 辨识算法(5)—(9)中引入遗忘因子(Forgetting Factor,FF)λ,便得到遗忘因子辅助模型随机梯度算法(Forgetting Factor AM-SG algorithm,FF-AM-SG 算法)或称为辅助模型遗忘因子随机梯度算法(Auxiliary Model based Forgetting Factor Stochastic Gradient algorithm,AM-FFSG 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)}[y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],$$

$$\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0,\tag{14}$$

$$r(t)=\lambda r(t-1)+\|\hat{\boldsymbol{\varphi}}(t)\|^2,\quad 0\leqslant\lambda\leqslant 1,\quad r(0)=1,\tag{15}$$

$$x_{\mathrm{a}}(t)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\quad x_{\mathrm{a}}(-i)=1/p_0,\quad i=0,1,\cdots,n_a,\tag{16}$$

$$\hat{\boldsymbol{\varphi}}(t)=[-x_{\mathrm{a}}(t-1),\cdots,-x_{\mathrm{a}}(t-n_a),u(t-1),\cdots,u(t-n_b)]^{\mathrm{T}}.\tag{17}$$

遗忘因子辅助模型随机梯度算法的参数估计误差有界收敛性是辨识领域的研究难题.

2 辅助模型递推最小二乘辨识方法

《系统辨识——辨识方法性能分析》[2]和《Inter-

national Journal of Computer Mathematics》2012 年第 15 期上的论文《Performance analysis of the auxiliary model-based least-squares identification algorithm for one-step state-delay systems》(一步状态迟延系统的输入输出表达的辅助模型递推最小二乘辨识算法的性能分析)[37]用随机过程理论近似分析了输出误差系统的辅助模型递推最小二乘算法的收敛性,这里用鞅收敛定理详细推导辅助模型递推最小二乘算法的收敛定理,主要内容参考笔者等发表在国际期刊《Automatica》2004 年第 10 期上的论文《Combined parameter and output estimation of dual-rate systems using an auxiliary model》(使用辅助模型的双率系统联合参数与输出估计)[9],以及文献[4].

2.1 辅助模型递推最小二乘算法

考虑输出误差模型(1)描述的随机系统

$$y(t)=\frac{B(z)}{A(z)}u(t)+v(t), \tag{18}$$

其中各变量的定义同上.设阶次 n_a 和 n_b 已知,记 $n:=n_a+n_b$,且 $t\leqslant 0$ 时,$y(t)=0,u(t)=0,v(t)=0$.

定义中间变量

$$x(t):=\frac{B(z)}{A(z)}u(t). \tag{19}$$

定义参数向量 $\boldsymbol{\theta}$ 和信息向量 $\boldsymbol{\varphi}(t)$ 如下:

$$\boldsymbol{\theta}:=[a_1,a_2,\cdots,a_{n_a},b_1,b_2,\cdots,b_{n_b}]^{\mathrm{T}}\in\mathbf{R}^n,$$

$$\boldsymbol{\varphi}(t):=[-x(t-1),-x(t-2),\cdots,-x(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}\in\mathbf{R}^n.$$

式(19)和(18)可以写为

$$x(t)=[1-A(z)]x(t)+B(z)u(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}, \tag{20}$$

$$y(t)=x(t)+v(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t). \tag{21}$$

为了处理辨识模型(21)中信息向量 $\boldsymbol{\varphi}(t)$ 涉及的未知项 $x(t-i)$,使用辅助模型辨识思想,参照 AM-SG 辨识算法(5)—(9)构造辅助模型,在辨识算法中未知信息向量 $\boldsymbol{\varphi}(t)$ 用其估计 $\hat{\boldsymbol{\varphi}}(t)$ 代替,能够获得辨识参数向量 $\boldsymbol{\theta}$ 的辅助模型递推最小二乘算法(Auxiliary Model based Recursive Least Squares algorithm,AM-RLS 算法):

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

$$\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0, \tag{22}$$

$$\boldsymbol{P}^{-1}(t)=\boldsymbol{P}^{-1}(t-1)+\hat{\boldsymbol{\varphi}}(t)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}_n, \tag{23}$$

$$x_{\mathrm{a}}(t)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\quad x_{\mathrm{a}}(-i)=1/p_0,\ i=0,1,\cdots,n_a, \tag{24}$$

$$\hat{\boldsymbol{\varphi}}(t)=[-x_{\mathrm{a}}(t-1),\cdots,-x_{\mathrm{a}}(t-n_a),u(t-1),\cdots,u(t-n_b)]^{\mathrm{T}}. \tag{25}$$

定义增益向量 $\boldsymbol{L}(t):=\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)\in\mathbf{R}^n$.应用矩阵求逆引理

$$(\boldsymbol{A}+\boldsymbol{BC})^{-1}=\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{I}+\boldsymbol{CA}^{-1}\boldsymbol{B})^{-1}\boldsymbol{CA}^{-1}$$

于式(23),AM-RLS 算法可等价表示为

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)[y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],$$

$$\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0, \tag{26}$$

$$\boldsymbol{L}(t)=\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)[1+\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)]^{-1}, \tag{27}$$

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t-1),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}_n, \tag{28}$$

$$x_{\mathrm{a}}(t)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t),\quad x_{\mathrm{a}}(-i)=1/p_0,\ i=0,1,\cdots,n_a, \tag{29}$$

$$\hat{\boldsymbol{\varphi}}(t)=[-x_{\mathrm{a}}(t-1),\cdots,-x_{\mathrm{a}}(t-n_a),u(t-1),\cdots,u(t-n_b)]^{\mathrm{T}}. \tag{30}$$

算法的初始值选择为 $\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0$,$\boldsymbol{P}(0)=\boldsymbol{I}_n/p_0$,$x_{\mathrm{a}}(i)=1/p_0,i\leqslant 0$,$p_0$ 是一个大正数,如 $p_0=10^6$.

2.2 AM-RLS 辨识算法的收敛性

定义

$$\boldsymbol{P}_0^{-1}(t):=\sum_{j=1}^{t}\boldsymbol{\varphi}(j)\boldsymbol{\varphi}^{\mathrm{T}}(j)+\frac{1}{p_0}\boldsymbol{I}_n,$$

$$r_0(t):=\mathrm{tr}[\boldsymbol{P}_0^{-1}(t)],\quad r(t):=\mathrm{tr}[\boldsymbol{P}^{-1}(t)].$$

容易得到:

$$|\boldsymbol{P}^{-1}(t)|\leqslant r^n(t),\quad r(t)\leqslant n\lambda_{\max}[\boldsymbol{P}^{-1}(t)],$$

$$\ln|\boldsymbol{P}^{-1}(t)|=O(\ln r(t)).$$

假设$\{v(t),\mathscr{T}_t\}$是定义在概率空间$\{\Omega,\mathscr{T},P\}$上的鞅差序列,其中$\{\mathscr{T}_t\}$是由直到时刻 t 的观测数据生成的 σ 代数序列.噪声序列$\{v(t)\}$满足假设:

(B1) $E[v(t)|\mathscr{T}_{t-1}]=0$,a.s.,

(B2) $E[v^2(t)|\mathscr{T}_{t-1}]=\sigma_v^2(t)\leqslant\overline{\sigma}_v^2<\infty$,a.s.,

(B2′) $E[v^2(t)|\mathscr{T}_{t-1}]\leqslant\delta_\varepsilon r^\varepsilon(t)$,a.s.,

$0\leqslant\delta_\varepsilon<\infty$,$0\leqslant\varepsilon<1$.

(B2)说明干扰噪声$\{v(t)\}$的方差有界,(B2′)说明干扰噪声$\{v(t)\}$的方差可以是无界的,并没有假设噪声平稳和高阶矩存在.

引理 2 对于 AM-RLS 辨识算法(22)—(25),下列不等式成立:

1) $\sum_{j=1}^{t}\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(j)\boldsymbol{P}(j)\hat{\boldsymbol{\varphi}}(j)\leqslant\ln|\boldsymbol{P}^{-1}(t)|+n\ln p_0$,a.s.;

2) $\sum_{t=1}^{\infty}\frac{\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)}{\{\ln|\boldsymbol{P}^{-1}(t)|\}^c}<\infty$,a.s.,$c>1$.

3) $\sum_{t=1}^{\infty}\frac{\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)}{\ln|\boldsymbol{P}^{-1}(t)|\cdot\{\ln\ln|\boldsymbol{P}^{-1}(t)|\}^c}<\infty$,a.s.,$c>1$.

定义参数估计误差向量 $\tilde{\boldsymbol{\theta}}(t)$ 和非负定函数 $T(t)$ 如下:

$\tilde{\boldsymbol{\theta}}(t):=\hat{\boldsymbol{\theta}}(t)-\boldsymbol{\theta}$,

$T(t):=\tilde{\boldsymbol{\theta}}^{\mathrm{T}}(t)\boldsymbol{P}^{-1}(t)\tilde{\boldsymbol{\theta}}(t)$.

引理 3 对于辨识模型(21)和 AM-RLS 辨识算法(22)—(25),假设(B1)和(B2)成立,

(B3) $H(z):=\dfrac{1}{A(z)}-\dfrac{1}{2}$是严格正实的(strictly positive real),

那么下列不等式成立:

$$E[T(t)+S(t)\mid\mathscr{F}_{t-1}]\leqslant T(t-1)+S(t-1)+2\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)\bar{\sigma}_v^2,\text{a.s.},$$

式中

$$S(t):=2\sum_{j=1}^{t}\tilde{u}(j)\tilde{y}(j)\geqslant 0, \tag{31}$$

$$\tilde{u}(t):=-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\tilde{\boldsymbol{\theta}}(t), \tag{32}$$

$$\tilde{y}(t):=\frac{1}{2}\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\tilde{\boldsymbol{\theta}}(t)+[y(t)-x_{\mathrm{a}}(t)-v(t)]. \tag{33}$$

定理 3[2,4,9] 对于辨识模型(21)和 AM-RLS 辨识算法(22)—(25),假设引理 3 的条件成立,$A(z)$是严格稳定的(strictly stable),即 $A(z)$的零点都在单位圆内,

(B4) $[\ln r_0(t)]^c=O(\lambda_{\min}[\boldsymbol{P}_0^{-1}(t)])$,a.s.,$c>1$,

那么 AM-RLS 算法的参数估计误差满足:

1) $\|\hat{\boldsymbol{\theta}}(t)-\boldsymbol{\theta}\|^2=O\left(\dfrac{[\ln r_0(t)]^c}{\lambda_{\min}[\boldsymbol{P}_0^{-1}(t)]}\right)$,a.s.,

2) $\|\hat{\boldsymbol{\theta}}(t)-\boldsymbol{\theta}\|^2=O\left(\dfrac{\ln r_0(t)[\ln\ln r_0(t)]^c}{\lambda_{\min}[\boldsymbol{P}_0^{-1}(t)]}\right)$,a.s..

定理 3 说明:对于有界噪声方差,AM-RLS 算法的参数估计误差 $\|\tilde{\boldsymbol{\theta}}(t)\|^2$ 正比于 $\boldsymbol{P}_0^{-1}(t)$的对数与其最小特征值之比.从定理 3 可以得到下列推论.

推论 1 存在正常数 c_0,c_1,c_2 和 t_0 使得对于 $t\geqslant t_0$,下列广义持续激励条件(Generalized Persistent Excitation condition,GPE 条件)(无界条件数)成立:

(GPE) $c_1\boldsymbol{I}\leqslant\dfrac{1}{t}\sum_{j=1}^{t}\boldsymbol{\varphi}(j)\boldsymbol{\varphi}^{\mathrm{T}}(j)\leqslant c_2t^{c_0}\boldsymbol{I}$,a.s.

那么对任意 $c>1$,下列关系成立:

1) $\|\hat{\boldsymbol{\theta}}(t)-\boldsymbol{\theta}\|^2=O\left(\dfrac{(\ln t)^c}{t}\right)\to 0$,a.s.,

2) $\|\hat{\boldsymbol{\theta}}(t)-\boldsymbol{\theta}\|^2=O\left(\dfrac{\ln t(\ln\ln t)^c}{t}\right)\to 0$,a.s..

因为对任意小的 $\epsilon>0$,有$(\ln t)^c=o(t^{\epsilon})$,所以估计误差 $\|\hat{\boldsymbol{\theta}}(t)-\boldsymbol{\theta}\|^2$ 近似以 $1/t^{1-\epsilon}$的速度趋于零.

推论 2 假设信息向量 $\boldsymbol{\varphi}(t)$和协方差阵 $\boldsymbol{P}_0(t)$满足

(B5) $\|\boldsymbol{\varphi}(t)\|^2=O(t^{\varepsilon})$, $0\leqslant\varepsilon<1$,

(B6) $\lambda_{\min}[\boldsymbol{P}_0^{-1}(t)]=O(t)$,

那么 $r_0(t)=O(t^{1+\varepsilon})$,参数估计误差满足

$$\|\tilde{\boldsymbol{\theta}}(t)\|^2=O\left(\frac{(\ln t)^c}{t}\right)\to 0,\text{a.s.},c>1.$$

下面的定理说明对无界噪声方差,AM-RLS 算法的参数估计误差收敛于零.

定理 4 对于辨识模型(21)和 AM-RLS 辨识算法(22)—(25),假设(B1),(B2′)和(B3)成立,$A(z)$是严格稳定的(strictly stable),则参数估计误差满足

$$\|\hat{\boldsymbol{\theta}}(t)-\boldsymbol{\theta}\|^2=o\left(\frac{\lambda_{\max}[\boldsymbol{P}_0^{-1}(t)]}{\lambda_{\min}[\boldsymbol{P}_0^{-1}(t)]}\right),\text{a.s.}.$$

从定理 4 可以得到下列推论.

推论 3 假设存在正常数 c_3,c_4 和 t_0 使得当 $t>t_0$ 时,下列弱持续激励条件(Weak Persistent Excitation condition,WPE 条件)(条件数有界)成立:

(WPE) $c_3\boldsymbol{I}_n\leqslant\dfrac{1}{t}\sum_{j=1}^{t}\boldsymbol{\varphi}(j)\boldsymbol{\varphi}^{\mathrm{T}}(j)\leqslant c_4\boldsymbol{I}_n$,a.s.,

那么参数估计误差收敛于零:

$\lim\limits_{t\to\infty}\|\hat{\boldsymbol{\theta}}(t)-\boldsymbol{\theta}\|^2=0$,a.s..

当 $c_0=0$ 时,GPE 条件退化为 WPE 条件.

如果把辅助模型的输出 $x_{\mathrm{a}}(t)$作为未知真实输出 $x(t)$的估计,下面的定理说明输出估计接近真实输出的程度.

定理 5 对于辨识模型(21)和 AM-RLS 辨识算法(22)—(25),假设(B1),(B2)和(B3)成立,$A(z)$是严格稳定的,对于任意 $c>1$,存在整数 t_0 使得有界输入(即$|u(t)|\leqslant U<\infty$)使得输出估计误差 $x_{\mathrm{a}}(t)-x(t)$满足

1) $\sum_{j=t_0}^{t}[x_{\mathrm{a}}(j)-x(j)]^2=O([\ln t]^{c+1})$,a.s., $t>t_0$,

2) $\dfrac{1}{t}\sum_{j=1}^{t}[x_{\mathrm{a}}(j)-x(j)]^2=O\left(\dfrac{(\ln t)^{c+1}}{t}\right)$,a.s., $t>t_0$.

定理 5 说明输出估计误差在平均意义下收敛于零.

3 辅助模型多新息随机梯度辨识方法

本节研究输出误差系统的辅助模型多新息随机梯度辨识算法的收敛性,相关工作参见笔者指导的访问学者王冬青博士完成的论文《Performance analysis of the auxiliary models based multi-innovation stochastic gradient estimation algorithm for output error systems》(输出误差系统辅助模型多新息随机梯度算法的性能分析),该文发表在《Digital Signal

Processing》2010 年第 3 期上[13],以及文献[4].

3.1 辅助模型多新息随机梯度算法

设整数 $p\geqslant 1$ 表示新息长度.根据多新息辨识理论[3],将 AM-SG 算法(5)—(9)中输出 $y(t)$ 和信息向量 $\hat{\boldsymbol{\varphi}}(t)$ 扩展为堆积输出向量 $\boldsymbol{Y}(p,t)$ 和堆积信息矩阵 $\hat{\boldsymbol{\Phi}}(p,t)$ 如下:

$$\boldsymbol{Y}(p,t):=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}\in\mathbf{R}^{p},$$

$$\hat{\boldsymbol{\Phi}}(p,t):=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)]\in\mathbf{R}^{n\times p},$$

将标量新息 $e(t)\in\mathbf{R}$ 扩展为新息向量

$$\boldsymbol{E}(p,t):=\begin{bmatrix}y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)\\ y(t-1)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\theta}}(t-1)\\ \vdots\\ y(t-p+1)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{\theta}}(t-1)\end{bmatrix}=$$

$$\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)\in\mathbf{R}^{p},$$

便得到辨识输出误差系统参数向量 $\boldsymbol{\theta}$ 的辅助模型多新息随机梯度算法(Auxiliary Model based Multi-Innovation Stochastic Gradient algorithm, AM-MISG 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\hat{\boldsymbol{\Phi}}(p,t)}{r(t)}\boldsymbol{E}(p,t),\quad \hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0,\tag{34}$$

$$\boldsymbol{E}(p,t)=\boldsymbol{Y}(p,t)-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1),\tag{35}$$

$$r(t)=r(t-1)+\|\hat{\boldsymbol{\Phi}}(p,t)\|^2,\quad r(0)=1,\tag{36}$$

$$\boldsymbol{Y}(p,t)=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}},\tag{37}$$

$$\hat{\boldsymbol{\Phi}}(p,t)=[\hat{\boldsymbol{\varphi}}(t),\hat{\boldsymbol{\varphi}}(t-1),\cdots,\hat{\boldsymbol{\varphi}}(t-p+1)],\tag{38}$$

$$x_{\mathrm{a}}(t-i)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t-i)\hat{\boldsymbol{\theta}}(t),\quad i=p-1,p-2,\cdots,1,0,\tag{39}$$

$$\hat{\boldsymbol{\varphi}}(t)=[-x_{\mathrm{a}}(t-1),\cdots,-x_{\mathrm{a}}(t-n_a),u(t-1),\cdots,u(t-n_b)]^{\mathrm{T}}.\tag{40}$$

3.2 AM-MISG 辨识算法的收敛性

假设$\{v(t),\mathscr{F}_t\}$是定义在概率空间$\{\Omega,\mathscr{F},P\}$上的鞅差序列,其中$\{\mathscr{F}_t\}$是由直到时刻 t 的观测数据生成的 σ 代数序列.噪声序列$\{v(t)\}$满足假设:

(C1) $E[v(t)\,|\,\mathscr{F}_{t-1}]=0$,a.s.,

(C2) $E[v^2(t)\,|\,\mathscr{F}_{t-1}]=\sigma_v^2(t)\leqslant\overline{\sigma}^2<\infty$,a.s.,

(C2′) $E[v^2(t)\,|\,\mathscr{F}_{t-1}]\leqslant\delta_{\varepsilon}r^{\varepsilon}(t)$,a.s.,

$0\leqslant\delta_{\varepsilon}<\infty$,$0\leqslant\varepsilon<1$.

令

$$\boldsymbol{W}(t):=\boldsymbol{I}_n-\frac{\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\Phi}}(p,t)}{r(t)}\in\mathbf{R}^{n\times n}.$$

使用式(36)有

$$\boldsymbol{W}(t)=\frac{r(t)\boldsymbol{I}_n-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\Phi}}(p,t)}{r(t)}=$$

$$\frac{\{r(t-1)+\|\hat{\boldsymbol{\Phi}}(p,t)\|^2\}\boldsymbol{I}_n-\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\Phi}}(p,t)}{r(t)}\geqslant$$

$$\frac{r(t-1)}{r(t)}\boldsymbol{I}_n.$$

定义堆积输入向量 $\boldsymbol{U}(p,t)$,堆积输出向量$\boldsymbol{Y}(p,t)$,堆积噪声向量 $\boldsymbol{V}(p,t)$,堆积未知真实输出向量 $\boldsymbol{X}(p,t)$,堆积信息矩阵 $\boldsymbol{\Phi}(p,t)$ 如下:

$$\boldsymbol{U}(p,t):=[u(t),u(t-1),\cdots,u(t-p+1)]^{\mathrm{T}}\in\mathbf{R}^{p},$$

$$\boldsymbol{Y}(p,t):=[y(t),y(t-1),\cdots,y(t-p+1)]^{\mathrm{T}}\in\mathbf{R}^{p},$$

$$\boldsymbol{V}(p,t):=[v(t),v(t-1),\cdots,v(t-p+1)]^{\mathrm{T}}\in\mathbf{R}^{p},$$

$$\boldsymbol{X}(p,t):=[x(t),x(t-1),\cdots,x(t-p+1)]^{\mathrm{T}}=\boldsymbol{\Phi}^{\mathrm{T}}(p,t)\boldsymbol{\theta}\in\mathbf{R}^{p},$$

$$\boldsymbol{\Phi}(p,t):=[\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\cdots,\boldsymbol{\varphi}(t-p+1)]\in\mathbf{R}^{n\times p}.$$

于是有

$$\boldsymbol{X}(p,t)=\frac{B(z)}{A(z)}\boldsymbol{U}(p,t).$$

定理 6[4,13] 对于辨识模型(4)和 AM-MISG 辨识算法(34)—(40),令

$$\boldsymbol{R}(t):=\sum_{j=1}^{t}\hat{\boldsymbol{\Phi}}(p,j)\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,j),$$

$$\boldsymbol{R}_0(t):=\sum_{j=1}^{t}\boldsymbol{\Phi}(p,j)\boldsymbol{\Phi}^{\mathrm{T}}(p,j),$$

$$r_0(t):=\mathrm{tr}[\boldsymbol{R}_0(t)],$$

假设(C1)和(C2)成立,$A(z)$是严格正实的,且

(C3) $r_0(t)=O(\lambda_{\min}[\boldsymbol{R}_0(t)])$,

那么参数估计 $\hat{\boldsymbol{\theta}}(t)$一致收敛于真参数向量 $\boldsymbol{\theta}$.

4 变间隔辅助模型随机梯度辨识方法

在实际中会出现某些采样点的观测数据丢失的情形,这样的系统称为损失数据系统.也就是说,对每一个 $t=0,1,2,\cdots$,观测数据 $y(t)$ 不可能总是可得到,这样的系统称为损失数据系统.为了处理数据丢失情况,定义一个整数序列(integer sequence)$\{t_s:s=0,1,2,\cdots\}$满足

$$0=t_0<t_1<t_2<t_3<\cdots<t_{s-1}<t_s<\cdots,$$

且 $t_s^*:=t_s-t_{s-1}\geqslant 1$.假设当 $t=t_s(s=0,1,2,\cdots)$时,$y(t)$都可得到,即对所有 $s=0,1,2,\cdots$,观测 $y(t_s)$都可得到,也就是 $y(t_0),y(t_1),y(t_2),\cdots$都可得到.在这种情况下,数据集$\{y(t_s):s=0,1,2,\cdots\}$包含所有可得到的观测输出数据,而不可得到的输出数据$\{y(t_s+1),y(t_s+2),\cdots,y(t_{s+1}-1):s=0,1,2,\cdots\}$是损失数据.这是损失数据系统可得到数据的一般框架.当然包括对所有 $s,t_s^*=1$ 时所有观测输出可得到情形.

这里针对输出误差系统,假设系统的输入数据在每一个采样点都可得到,而只有系统输出存在损失数据,且损失数据点是随机的,是没有规则的.只有当采集到输出数据,我们才更新参数估计.这种方案可以使用变间隔辨识方法来跳过损失数据的点,从而获得系统的参数估计.本节研究输出误差系统的变间隔辅助模型随机梯度算法参数估计和损失输出估计的收敛性.该方法适用于损失数据系统和稀少量测数据系统.相关工作参见发表在国际期刊《Automatica》2011 年第 8 期上的论文《Parameter estimation with scarce measurements》(稀少量测数据系统的变间隔辅助模型多新息随机梯度参数估计)[28]和《系统辨识——多新息辨识理论与方法》6.6 节的变间隔辅助模型多新息随机梯度辨识算法的收敛性证明[3],以及文献[4].

设 $\hat{\boldsymbol{\theta}}(t)$ 是参数向量 $\boldsymbol{\theta}$ 在时刻 t 的估计,$\hat{\boldsymbol{\theta}}(t_s)$ 是参数向量 $\boldsymbol{\theta}$ 在时刻 $t=t_s$ 的估计.对于损失数据系统,观测输出 $y(t)$ 不是每一时刻都可得到,只是在时刻 $t=t_s(s=0,1,2,\cdots)$ 才有观测输出数据 $y(t_s)$,因此参数估计 $\hat{\boldsymbol{\theta}}(t)$ 只是在时刻 $t=t_s$ 才刷新,而在可得到的两个观测数据间保持估计不变,即:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t_{s-1}),\quad t\in T_s:=\{t_{s-1},t_{s-1}+1,\cdots,t_s-1\}.$$

这有几个等价的表示:

$$1)\ \hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t_{s-1}),\quad t_{s-1}\leqslant t<t_s,\quad s=1,2,3,\cdots,$$

$$2)\ \hat{\boldsymbol{\theta}}(t_{s-1}+i)=\hat{\boldsymbol{\theta}}(t_{s-1}),\quad i=1,2,\cdots,t_s^*-1,$$

$$3)\ \hat{\boldsymbol{\theta}}(t_s-i)=\hat{\boldsymbol{\theta}}(t_{s-1}),\quad i=1,2,\cdots,t_s^*-1.$$

这 4 种情形,都表示在区间 $[t_{s-1},t_s-1]$ 上,$\hat{\boldsymbol{\theta}}(t)$ 保持不变,即

$$\boldsymbol{\theta}(t)=\hat{\boldsymbol{\theta}}(t_{s-1}),\quad t\in[t_{s-1},t_s-1],\quad s=1,2,3,\cdots. \tag{41}$$

4.1 变间隔辅助模型随机梯度算法

考虑输出误差模型(1)描述的随机系统

$$y(t)=\frac{B(z)}{A(z)}u(t)+v(t), \tag{42}$$

其中各变量的定义同上.设阶次 n_a 和 n_b 已知,记 $n:=n_a+n_b$,且 $t\leqslant 0$ 时,$y(t)=0,u(t)=0,v(t)=0$.

定义中间变量:

$$x(t):=\frac{B(z)}{A(z)}u(t). \tag{43}$$

定义参数向量 $\boldsymbol{\theta}$ 和信息向量 $\boldsymbol{\varphi}(t)$ 如下:

$$\boldsymbol{\theta}:=[a_1,a_2,\cdots,a_{n_a},b_1,b_2,\cdots,b_{n_b}]^{\mathrm{T}}\in\mathbf{R}^n,$$

$$\boldsymbol{\varphi}(t):=[-x(t-1),-x(t-2),\cdots,-x(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}.$$

式(42)和(43)可以写为

$$x(t)=[1-A(z)]x(t)+B(z)u(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}, \tag{44}$$

$$y(t)=x(t)+v(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t). \tag{45}$$

用 t_s 代替式(44)和(45)中的 t,有

$$x(t_s)=\boldsymbol{\varphi}^{\mathrm{T}}(t_s)\boldsymbol{\theta}, \tag{46}$$

$$y(t_s)=\boldsymbol{\varphi}^{\mathrm{T}}(t_s)\boldsymbol{\theta}+v(t_s), \tag{47}$$

式中

$$\boldsymbol{\varphi}(t_s):=[-x(t_s-1),-x(t_s-2),\cdots,-x(t_s-n_a),u(t_s-1),u(t_s-2),\cdots,u(t_s-n_b)]^{\mathrm{T}}\in\mathbf{R}^n.$$

因为信息向量 $\boldsymbol{\varphi}(t_s)$ 包含了未知量 $x(t_s-i)$,解决方案是借助于辅助模型辨识思想,在辨识算法中用其估计代替.为此构造一个辅助模型,用辅助模型的输出 $x_a(t_s-i)$ 和系统输入 $u(t_s-i)$ 定义 $\boldsymbol{\varphi}(t_s)$ 的估计

$$\hat{\boldsymbol{\varphi}}(t_s):=[-x_a(t_s-1),-x_a(t_s-2),\cdots,-x_a(t_s-n_a),u(t_s-1),u(t_s-2),\cdots,u(t_s-n_b)]^{\mathrm{T}}\in\mathbf{R}^n,$$

根据式(46),用 $\hat{\boldsymbol{\varphi}}(t_s)$ 和 $\hat{\boldsymbol{\theta}}(t_s)$ 构造估算 $x(t_s)$ 的辅助模型:

$$x_a(t_s)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s)\hat{\boldsymbol{\theta}}(t_s).$$

在数据损失的情形下,为实现参数估计算法,两个观测数据间的辅助模型输出也应该计算.因为假设在区间 $[t_s,t_{s+1}-1]$ 上,参数估计 $\hat{\boldsymbol{\theta}}(t)$ 保持不变[参见式(41)],所以辅助模型应修改为

$$x_a(t_s+j)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s+j)\hat{\boldsymbol{\theta}}(t_s),\quad j=0,1,\cdots,t_{s+1}^*-1,$$

$$\hat{\boldsymbol{\varphi}}(t_s+j):=[-x_a(t_s+j-1),-x_a(t_s+j-2),\cdots,-x_a(t_s+j-n_a),u(t_s+j-1),u(t_s+j-2),\cdots,u(t_s+j-n_b)]^{\mathrm{T}}\in\mathbf{R}^n.$$

这有下列几种等价表达:

$$1)\ x_a(t)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t_s),\quad t\in[t_s,t_{s+1}-1],$$

$$2)\ x_a(t)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t_s),\quad t\in T_s:=\{t_s,t_s+1,\cdots,t_{s+1}-1\},$$

$$3)\ x_a(t)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t_s),\quad t=t_s,t_s+1,\cdots,t_{s+1}-1.$$

在式(41)规定的区间上保持参数估计 $\hat{\boldsymbol{\theta}}(t)$ 不变的条件下,辅助模型可以等价表示为

$$x_a(t)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \tag{48}$$

$$\hat{\boldsymbol{\varphi}}(t):=[-x_a(t-1),-x_a(t-2),\cdots,-x_a(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}\in\mathbf{R}^n. \tag{49}$$

设 $\mu(t_s)\geqslant 0$ 是收敛因子(convergence factor)或步长(step-size).根据辨识模型(47),定义梯度准则函数(gradient criterion function):

$$J_1(\boldsymbol{\theta}):=\frac{1}{2}[y(t_s)-\boldsymbol{\varphi}^{\mathrm{T}}(t_s)\boldsymbol{\theta}]^2,$$

使用梯度搜索,极小化 $J_1(\boldsymbol{\theta})$,可以得到下列梯度搜索算法(gradient search algorithm):

$$\hat{\boldsymbol{\theta}}(t_s)=\hat{\boldsymbol{\theta}}(t_{s-1})+\mu(t_s)\,\mathrm{grad}\{J_1[\hat{\boldsymbol{\theta}}(t_{s-1})]\}=$$

$$\hat{\boldsymbol{\theta}}(t_{s-1})+\mu(t_s)\boldsymbol{\varphi}(t_s)[y(t_s)-\boldsymbol{\varphi}^{\mathrm{T}}(t_s)\hat{\boldsymbol{\theta}}(t_{s-1})]. \quad (50)$$

因为 $\boldsymbol{\varphi}(t_s)$ 包含了未知量 $x(t_s-j)$,所以这个算法无法实现,解决的办法是借助于辅助模型辨识思想,式(50)中未知信息向量 $\boldsymbol{\varphi}(t_s)$ 用其估计 $\hat{\boldsymbol{\varphi}}(t_s)$ 代替,并取 $\mu(t_s):=1/r(t_s)$,$r(t_s)=r(t_{s-1})+\|\hat{\boldsymbol{\varphi}}(t_s)\|^2$,联立辅助模型(48)—(49),可以得到辨识输出误差系统参数向量 $\boldsymbol{\theta}$ 的变间隔辅助模型随机梯度算法(interval-Varying AM-SG algorithm,V-AM-SG 算法):

$$\hat{\boldsymbol{\theta}}(t_s)=\hat{\boldsymbol{\theta}}(t_{s-1})+\frac{\hat{\boldsymbol{\varphi}}(t_s)}{r(t_s)}[y(t_s)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s)\hat{\boldsymbol{\theta}}(t_{s-1})], \quad (51)$$

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t_{s-1}),\quad t\in[t_{s-1},t_s-1], \quad (52)$$

$$r(t_s)=r(t_{s-1})+\|\hat{\boldsymbol{\varphi}}(t_s)\|^2, \quad (53)$$

$$x_{\mathrm{a}}(t_s-j)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t_{s-1}),\quad j=0,1,2,\cdots,t_s^*-1, \quad (54)$$

$$\hat{\boldsymbol{\varphi}}(t):=[-x_{\mathrm{a}}(t-1),\cdots,-x_{\mathrm{a}}(t-n_a),u(t-1),$$
$$u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}, \quad (55)$$

$$\hat{\boldsymbol{\theta}}(t_s)=[\hat{a}_1(t_s),\hat{a}_2(t_s),\cdots,\hat{a}_{n_a}(t_s),\hat{b}_1(t_s),$$
$$\hat{b}_2(t_s),\cdots,\hat{b}_{n_b}(t_s)]^{\mathrm{T}}. \quad (56)$$

这里的 V-AM-SG 算法(51)—(56)不是简单将 AM-SG 算法(5)—(9)中 t 替换为 t_s 得到的.

V-AM-SG 算法(51)—(56)随 s 增加计算参数的步骤如下:

1) 初始化,令 $t=1$,$s=0$,$t_0=0$,置初值 $\hat{\boldsymbol{\theta}}(t_0)=\mathbf{1}_n/p_0$,$r(t_0)=1$,$x_{\mathrm{a}}(t-i)=1/p_0$,$i=1,2,\cdots,n_a$,$p_0=10^6$.$\mathbf{1}_n$ 是一个元均为 1 的 n 维列向量.

2) 采集输入输出数据 $u(t)$ 和 $y(t)$.保存输入数据 $u(t)$.

3) 如果输出数据 $y(t)$ 可得到,就跳到下一步;否则用式(55)构造信息向量 $\hat{\boldsymbol{\varphi}}(t)$,用式(54)计算辅助模型的输出 $x_{\mathrm{a}}(t)$,置 $\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t_s)$,$t$ 增加 1,即 $t:=t+1$,转到第 2)步.

4) 置 $s:=s+1$,$t_s:=t$,$t_s^*:=t_s-t_{s-1}$,保存输出数据 $y(t)$,用式(55)构造信息向量 $\hat{\boldsymbol{\varphi}}(t_s)$.

5) 用式(53)计算 $r(t_s)$.

6) 用式(54)计算辅助模型的输出 $x_{\mathrm{a}}(t)$.

7) 根据式(51)刷新参数估计向量 $\hat{\boldsymbol{\theta}}(t_s)$.根据式(52)设置参数估计向量 $\hat{\boldsymbol{\theta}}(t)$.

8) t 增 1,转到第 2)步.

V-AM-SG 算法的仿真例子参见文献[28]和《系统辨识——多新息辨识理论与方法》6.6 节的变间隔辅助模型多新息随机梯度(V-AM-MISG)算法的新息长度 $p=1$ 时的情形[3].

注 4 V-AM-SG 算法(51)—(56)是基于输出误差模型,而不是基于方程误差模型(即 CAR 模型或称 ARX 模型).事实上,这是一个联合参数与真实输出同时估计算法,利用系统的可得到数据 $\{u(t),y(t_s)\}$ 估计系统参数向量 $\boldsymbol{\theta}$ 和未知真实输出(无噪输出)$x(t)$.尽管提出的方法可以基于 CAR 模型,但是算法的收敛性证明会遇到特别的困难.

注 5 随机梯度算法收敛慢,基于 V-AM-SG 算法(51)—(56),在式(53)中引入遗忘因子 λ,即

$$r(t_s)=\lambda r(t_{s-1})+\|\hat{\boldsymbol{\varphi}}(t_s)\|^2,$$
$$0\leqslant\lambda\leqslant 1,\quad r(t_0)=1, \quad (57)$$

就得到遗忘因子 V-AM-SG 算法(51)—(52)和(54)—(57).它能够改善梯度算法的暂态性能.

注 6 为改善算法的暂态和稳态性能,基于 V-AM-SG 算法(51)—(56),可在式(51)中引入收敛指数 ε,即

$$\hat{\boldsymbol{\theta}}(t_s)=\hat{\boldsymbol{\theta}}(t_{s-1})+\frac{\hat{\boldsymbol{\varphi}}(t_s)}{r^{\varepsilon}(t_s)}[y(t_s)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s)\hat{\boldsymbol{\theta}}(t_{s-1})],$$
$$\frac{1}{2}<\varepsilon\leqslant 1. \quad (58)$$

就得到修正 V-AM-SG 算法.

4.2 V-AM-SG 辨识算法的收敛性

下面研究 V-AM-SG 算法的收敛性.V-AM-SG 算法的推广是变间隔辅助模型多新息随机梯度(V-AM-MISG)算法.相关研究可参考文献[3,28]

引理 4 对于 V-AM-SG 算法(51)—(56),下列不等式成立:

$$\sum_{s=1}^{\infty}\frac{\|\hat{\boldsymbol{\varphi}}(t_s)\|^2}{r^{\varepsilon}(t_s)}\leqslant\frac{1}{\varepsilon-1}<\infty,\quad \varepsilon>1.$$

引理 5 对于 V-AM-SG 算法(51)—(56),定义新息(innovation)$e(t_s)$ 和残差(residue)$\eta(t_s)$ 如下:

$$e(t_s):=y(t_s)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s)\hat{\boldsymbol{\theta}}(t_{s-1})\in\mathbf{R},$$
$$\eta(t_s):=y(t_s)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s)\hat{\boldsymbol{\theta}}(t_s)\in\mathbf{R}.$$

那么新息 $e(t_s)$ 与残差 $\eta(t_s)$ 有下列关系:

$$\eta(t_s)=\frac{r(t_{s-1})}{r(t_s)}e(t_s).$$

引理 6 对于辨识模型(47)和 V-AM-SG 算法(51)—(56),定义参数估计误差向量

$$\tilde{\boldsymbol{\theta}}(t_s):=\hat{\boldsymbol{\theta}}(t_s)-\boldsymbol{\theta}.$$

那么下列不等式成立:

$$A(z)[e(t_s)-v(t_s)]=-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s)\tilde{\boldsymbol{\theta}}(t_{s-1}).$$

定义

$$S(t_s):=\sum_{i=1}^{s}\frac{2}{r(t_i)}\left\{\tilde{y}(t_i)-\frac{1}{2}[e(t_i)-v(t_i)]\right\}[e(t_i)-v(t_i)]\geqslant 0.$$

定理 7[2,4] 对于辨识模型(47)和 V-AM-SG 算法(51)—(56),假设$\{v(t)\}$是一个零均值,方差为σ^2的不相关随机噪声,即:

(D1) $E[v(t)]=0$, $E[v^2(t)]=\sigma^2$,
$E[v(t)v(i)]=0, i\neq t$,

(D2) $A(z)-\frac{1}{2}$ 是严格正实的,

那么 $\|\tilde{\boldsymbol{\theta}}(t_s)\|^2+S(t_s)$ 均方(mean square, m.s.)收敛于一常数 C_7,即

$$\lim_{s\to\infty}[\|\tilde{\boldsymbol{\theta}}(t_s)\|^2+S(t_s)]=C_7<\infty\ ,\text{m.s.}.$$

这意味着参数估计误差 $\hat{\boldsymbol{\theta}}(t)-\boldsymbol{\theta}$ 在均方意义下是有界的.

下列定理是关于参数估计的收敛性.

定理 8[2,4] 对于辨识模型(47)和 V-AM-SG 算法(51)—(56),假设定理 7 的条件成立,存在正常数 C_8,使得 $t_s\to\infty$ 时,$r(t_s)\leqslant C_8\lambda_{\min}[\mathscr{R}(t_s)]$,其中

$$\mathscr{R}(t_s):=\sum_{i=1}^{s}\hat{\boldsymbol{\varphi}}(t_i)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_i),$$

那么参数估计误差向量 $\tilde{\boldsymbol{\theta}}(t_s)$ 收敛于零,即

$$\|\tilde{\boldsymbol{\theta}}(t_s)\|^2=o(1)\to 0, r(t_s)\to\infty.$$

下面的定理是关于不可得到真实输出估计的收敛性.

定理 9[2,4] 对于辨识模型(47)和 V-AM-SG 算法(51)—(56),假设定理 8 的条件成立,$A(z)$是严格稳定的,那么有界输入信号意味着真实输出估计误差 $x_{\mathrm{a}}(t)-x(t)$ 在均方意义(m.s.)下收敛于零,即

$$\lim_{t\to\infty}\frac{1}{t}\sum_{j=1}^{t}[x_{\mathrm{a}}(j)-x(j)]^2=0,\text{m.s.}.$$

此外,当 $t\to\infty$ 时,估计误差收敛于零,即

$$x_{\mathrm{a}}(t)-x(t)\to 0,\text{m.s.}.$$

注 7 定理 7 随着数据长度 t 增加,参数估计向量 $\hat{\boldsymbol{\theta}}(t)$ 收敛于真参数向量 $\boldsymbol{\theta}$.定理 8 说明:一旦参数估计收敛于真参数,真实输出估计 $x_{\mathrm{a}}(t)$ 就收敛于真值 $x(t)$.因为假设 $v(t)$ 是一个白噪声,且 $y(t)=x(t)+v(t)$,$x_{\mathrm{a}}(t)\to x(t)$,因此估计 $x_{\mathrm{a}}(t)$ 可以看作不可得到输出 $y(t)$ $(t\neq t_s)$ 的最优估计.

5 变间隔辅助模型递推最小二乘辨识方法

这里针对输出误差系统,假设系统的输入数据在每一个采样点都可得到,而只有系统输出存在损失数据,且损失数据点是随机的,是没有规则的.只有当采集到输出数据,才更新参数估计.这种方案可以使用变间隔辨识方法来跳过损失数据的点,从而获得系统的参数估计.本节研究输出误差系统的变间隔辅助模型递推最小二乘算法参数估计和损失输出估计的收敛性.相关工作参见发表在国际期刊《International Journal of Adaptive Control and Signal Processing》2010 年第 7 期上的论文《Least squares parameter estimation for systems with irregularly missing data》(基于输出误差模型的不规则损失输出数据系统递推最小二乘参数估计)[29],以及文献[4].

5.1 变间隔辅助模型递推最小二乘算法

考虑输出误差模型(1)描述的随机系统:

$$y(t)=\frac{B(z)}{A(z)}u(t)+v(t),\tag{59}$$

$$x(t):=\frac{B(z)}{A(z)}u(t),\tag{60}$$

$$\boldsymbol{\theta}:=[a_1,a_2,\cdots,a_{n_a},b_1,b_2,\cdots,b_{n_b}]^{\mathrm{T}}\in\mathbf{R}^n,$$

$$\boldsymbol{\varphi}(t):=[-x(t-1),-x(t-2),\cdots,-x(t-n_a),u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}},$$

$$x(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta},\tag{61}$$

$$y(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t).\tag{62}$$

其中各变量的定义同上.设阶次 n_a 和 n_b 已知,记 $n:=n_a+n_b$,且 $t\leqslant 0$ 时,$y(t)=0, u(t)=0, v(t)=0$.

用 t_s 代替式(61)和(62)中的 t,有

$$x(t_s)=\boldsymbol{\varphi}^{\mathrm{T}}(t_s)\boldsymbol{\theta},\tag{63}$$

$$y(t_s)=\boldsymbol{\varphi}^{\mathrm{T}}(t_s)\boldsymbol{\theta}+v(t_s),\tag{64}$$

对于辨识模型(63)—(64),极小化下列准则函数:

$$J_2(\boldsymbol{\theta}):=\sum_{j=1}^{s}\lambda^{s-j}[y(t_j)-\boldsymbol{\varphi}^{\mathrm{T}}(t_j)\boldsymbol{\theta}]^2,$$

借助于辅助模型,未知向量 $\boldsymbol{\varphi}(t)$ 用其估计 $\hat{\boldsymbol{\varphi}}(t)$ 代替,可以得到带遗忘因子的变间隔辅助模型递推最小二乘算法,即变间隔遗忘因子辅助模型递推最小二乘算法(V-FF-AM-RLS 算法或变间隔辅助模型遗忘因子递推最小二乘算法(V-AM-FF-RLS 算法):

$$\hat{\boldsymbol{\theta}}(t_s)=\hat{\boldsymbol{\theta}}(t_{s-1})+\boldsymbol{P}(t_s)\hat{\boldsymbol{\varphi}}(t_s)e(t_s),\tag{65}$$

$$e(t_s)=y(t_s)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s)\hat{\boldsymbol{\theta}}(t_{s-1}),\tag{66}$$

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t_{s-1}),\quad t\in T_s=\{t_{s-1},t_{s-1}+1,\cdots,t_s-1\},\tag{67}$$

$$\boldsymbol{P}^{-1}(t_s)=\lambda\boldsymbol{P}^{-1}(t_{s-1})+\hat{\boldsymbol{\varphi}}(t_s)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s),\quad 0<\lambda\leqslant 1,$$
$$\boldsymbol{P}(0)=p_0\boldsymbol{I}_n, \tag{68}$$
$$x_{\mathrm{a}}(t_s-j)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s-j)\hat{\boldsymbol{\theta}}(t_{s-1}),\quad j=1,2,\cdots,t_s^*, \tag{69}$$
$$\hat{\boldsymbol{\varphi}}(t_s-j)=[-x_{\mathrm{a}}(t_s-1-j),-x_{\mathrm{a}}(t_s-2-j),\cdots,-x_{\mathrm{a}}(t_s-n_a-j),$$
$$u(t_s-1-j),u(t_s-2-j),\cdots,u(t_s-n_b-j)]^{\mathrm{T}}, \tag{70}$$
$$\hat{\boldsymbol{\theta}}(t_s)=[\hat{a}_1(t_s),\hat{a}_2(t_s),\cdots,\hat{a}_{n_a}(t_s),\hat{b}_1(t_s),$$
$$\hat{b}_2(t_s),\cdots,\hat{b}_{n_b}(t_s)]^{\mathrm{T}}. \tag{71}$$

定义增益向量 $\boldsymbol{L}(t_s):=\boldsymbol{P}(t_s)\hat{\boldsymbol{\varphi}}(t_s)\in\mathbf{R}^n$,应用矩阵求逆引理于式(68),V-AM-FF-RLS 算法可以等价表示为

$$\hat{\boldsymbol{\theta}}(t_s)=\hat{\boldsymbol{\theta}}(t_{s-1})+\boldsymbol{L}(t_s)e(t_s), \tag{72}$$
$$e(t_s)=y(t_s)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s)\hat{\boldsymbol{\theta}}(t_{s-1}), \tag{73}$$
$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t_{s-1}),\quad t\in T_s=\{t_{s-1},t_{s-1}+1,\cdots,t_s-1\}, \tag{74}$$
$$\boldsymbol{L}(t_s)=\boldsymbol{P}(t_{s-1})\hat{\boldsymbol{\varphi}}(t_s)[\lambda+\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s)\boldsymbol{P}(t_{s-1})\hat{\boldsymbol{\varphi}}(t_s)]^{-1}, \tag{75}$$
$$\boldsymbol{P}(t_s)=\frac{1}{\lambda}\left[\boldsymbol{P}(t_{s-1})-\frac{\boldsymbol{P}(t_{s-1})\hat{\boldsymbol{\varphi}}(t_s)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s)\boldsymbol{P}(t_{s-1})}{\lambda+\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s)\boldsymbol{P}(t_{s-1})\hat{\boldsymbol{\varphi}}(t_s)}\right]=$$
$$\frac{1}{\lambda}[\boldsymbol{I}-\boldsymbol{L}(t_s)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s)]\boldsymbol{P}(t_{s-1}),\quad 0<\lambda\leqslant 1, \tag{76}$$
$$x_{\mathrm{a}}(t)=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \tag{77}$$
$$\hat{\boldsymbol{\varphi}}(t)=[-x_{\mathrm{a}}(t-1),\cdots,-x_{\mathrm{a}}(t-n_a),u(t-1),$$
$$\cdots,u(t-n_b)]^{\mathrm{T}}, \tag{78}$$
$$\hat{\boldsymbol{\theta}}(t_s)=[\hat{a}_1(t_s),\hat{a}_2(t_s),\cdots,\hat{a}_{n_a}(t_s),\hat{b}_1(t_s),$$
$$\hat{b}_2(t_s),\cdots,\hat{b}_{n_b}(t_s)]^{\mathrm{T}}. \tag{79}$$

$\boldsymbol{L}(t_s)\in\mathbf{R}^n$ 为增益向量(gain vector),$\boldsymbol{P}(t_s)\in\mathbf{R}^{n\times n}$ 为协方差矩阵(covariance matrix),$\hat{\boldsymbol{\theta}}(t_s)$ 为 $\boldsymbol{\theta}$ 在时刻 t_s 的估计.

V-AM-FF-RLS 算法(72)—(79)计算参数估计 $\hat{\boldsymbol{\theta}}(t_s)$ 的流程如图 2 所示,随 s 增加计算参数的步骤如下:

1)初始化,令 $t=1$,$s=0$,$t_0=0$,给定遗忘因子 λ,置初值 $\hat{\boldsymbol{\theta}}(t_0)=\mathbf{1}_n/p_0$,$\boldsymbol{P}(t_0)=p_0\boldsymbol{I}_n$,$x_{\mathrm{a}}(t-i)=1/p_0$,$i=1,2,\cdots,n_a$,$\mathbf{1}_n$ 是一个元均为 1 的 n 维列向量,$\boldsymbol{I}_n$ 是一个 n 阶单位阵,p_0 取一个很大的正数,如 $p_0=10^6$.

2)采集输入输出数据 $u(t)$ 和 $y(t)$.保存输入数据 $u(t)$.

3)如果输出数据 $y(t)$ 可得到,就跳到下一步;否则置 $\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t_s)$,用式(78)构造信息向量 $\hat{\boldsymbol{\varphi}}(t)$,用式(77)计算辅助模型的输出 $x_{\mathrm{a}}(t)$,t 增加 1,即 $t:=t+1$,转到第 2)步.

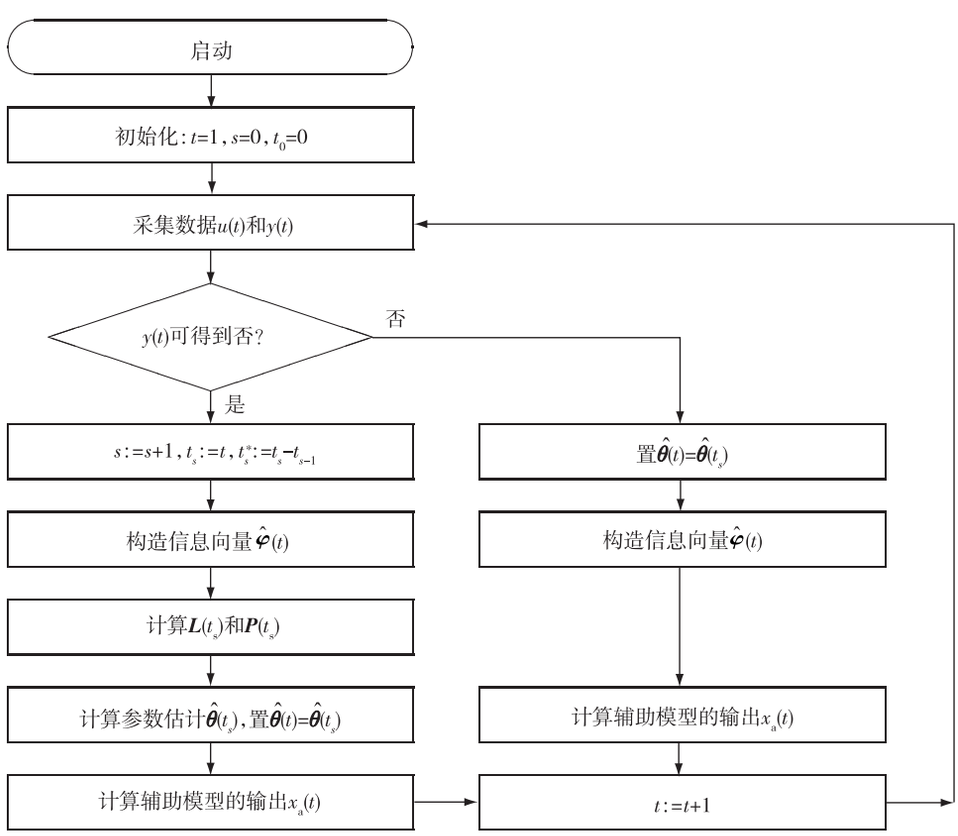


图 2 变间隔辅助模型遗忘因子递推最小二乘(V-AM-FF-RLS)算法流程

Fig. 2 The flowchart of the V-AM-FF-RLS algorithm

4)置 $s:=s+1,t_s:=t,t_s^*:=t_s-t_{s-1}$,保存输出数据 $y(t)$,用式(78)构造信息向量 $\hat{\boldsymbol{\varphi}}(t)$.

5)用式(75)计算增益向量 $\boldsymbol{L}(t_s)$,用式(76)计算协方差阵 $\boldsymbol{P}(t_s)$.

6)根据式(72)刷新参数估计向量 $\hat{\boldsymbol{\theta}}(t_s)$.根据式(74)设置参数估计向量 $\hat{\boldsymbol{\theta}}(t)$.

7)用式(77)计算辅助模型的输出 $x_a(t)$.

8)t 增 1,转到第 2)步.

当取遗忘因子 $\lambda=1$ 时,V-AM-FF-RLS 算法退化为变间隔辅助模型递推最小二乘算法(V-AM-RLS 算法).当取遗忘因子 $\lambda=1,t_s^*\equiv 1$ 时,V-AM-FF-RLS 算法退化为辅助模型递推最小二乘算法(AM-RLS 算法)(22)—(25).

本节分析 V-AM-RLS 算法(即单位遗忘因子时的 V-AM-FF-RLS 算法)的收敛性.遗忘因子 $0<\lambda<1$ 时的 V-AM-FF-RLS 算法的收敛性仍是该领域要解决的难题.

5.2 V-AM-RLS 辨识算法的收敛性

5.2.1 基本引理

定义残差(residue)

$$\eta(t_s):=y(t_s)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s)\hat{\boldsymbol{\theta}}(t_s)\in\mathbf{R}. \tag{80}$$

使用式(69)和(64)得到

$$\eta(t_s)=y(t_s)-x_a(t_s)=x(t_s)-x_a(t_s)+v(t_s). \tag{81}$$

定义

$$\boldsymbol{P}_0^{-1}(t_s):=\boldsymbol{P}_0^{-1}(t_{s-1})+\boldsymbol{\varphi}(t_s)\boldsymbol{\varphi}^{\mathrm{T}}(t_s),\quad \boldsymbol{P}_0(0)=p_0\boldsymbol{I},$$

$$r_0(t_s):=\mathrm{tr}[\boldsymbol{P}_0^{-1}(t_s)],\quad r(t_s):=\mathrm{tr}[\boldsymbol{P}^{-1}(t_s)].$$

根据矩阵行列式、矩迹与特征值的关系,我们有

$$|\boldsymbol{P}^{-1}(t_s)|\leqslant r^n(t_s),\quad r(t_s)\leqslant n\lambda_{\max}[\boldsymbol{P}^{-1}(t_s)]. \tag{82}$$

定义参数估计误差向量 $\tilde{\boldsymbol{\theta}}(t_s)$ 和非负定函数 $T(t_s)$ 如下:

$$\tilde{\boldsymbol{\theta}}(t_s)=\hat{\boldsymbol{\theta}}(t_s)-\boldsymbol{\theta}, \tag{83}$$

$$T(t_s)=\tilde{\boldsymbol{\theta}}^{\mathrm{T}}(t_s)\boldsymbol{P}^{-1}(t_s)\tilde{\boldsymbol{\theta}}(t_s). \tag{84}$$

由此可得

$$\|\tilde{\boldsymbol{\theta}}(t_s)\|^2\leqslant\frac{\tilde{\boldsymbol{\theta}}^{\mathrm{T}}(t_s)\boldsymbol{P}^{-1}(t_s)\tilde{\boldsymbol{\theta}}(t_s)}{\lambda_{\min}[\boldsymbol{P}^{-1}(t_s)]}\leqslant\frac{T(t_s)}{\lambda_{\min}[\boldsymbol{P}^{-1}(t_s)]}.$$

引理 7 对于输出误差系统(63)—(64)和 V-AM-RLS 算法,即遗忘因子 $\lambda=1$ 时的 V-AM-FF-RLS 算法(65)—(71),下列不等式成立:

$$1)\ \sum_{s=1}^{\infty}\frac{\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s)\boldsymbol{P}(t_s)\hat{\boldsymbol{\varphi}}(t_s)}{[\ln|\boldsymbol{P}^{-1}(t_s)|]^c}<\infty,\text{a.s.},c>1;$$

$$2)\ \sum_{s=1}^{\infty}\frac{\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s)\boldsymbol{P}(t_s)\hat{\boldsymbol{\varphi}}(t_s)}{r^c(t_s)}<\infty,\text{a.s.},c>0.$$

引理 8 对于输出误差系统(63)—(64)和 V-AM-RLS 算法,即遗忘因子 $\lambda=1$ 时的 V-AM-FF-RLS 算法(65)—(71),假设 $H(z):=A^{-1}(z)-1/2$ 是严格正实的,定义

$$\tilde{u}(t_s+i):=-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s+i)\tilde{\boldsymbol{\theta}}(t_s),i=0,1,\cdots,t_{s+1}^*-1, \tag{86}$$

$$\tilde{y}(t_s+i):=\frac{1}{2}\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t_s+i)\tilde{\boldsymbol{\theta}}(t_s)+[y(t_s+i)-x_a(t_s+i)-v(t_s+i)], \tag{87}$$

$$\tilde{y}_1(t_s):=\tilde{y}(t_s)-\frac{\rho}{2}\tilde{u}(t_s),$$

$$S_1(t_s):=2\sum_{i=1}^{s}\tilde{u}(t_i)\tilde{y}_1(t_i),$$

$$S(t_s):=2\sum_{i=1}^{s}\tilde{u}(t_i)\tilde{y}(t_i)=S_1(t)+\rho\sum_{i=1}^{s}\tilde{u}^2(t_i),$$

$S(0)=0$,

那么 $\tilde{y}(t_s)=H(z)\tilde{u}(t_s)$,$S(t_s)\geqslant 0$,且存在一个小正数 ρ 使得 $S_1(t_s)\geqslant 0$.

5.2.2 参数估计的收敛性

定理 10[4,29] 对于输出误差系统(63)—(64)和 V-AM-RLS 算法,即遗忘因子 $\lambda=1$ 时的 V-AM-FF-RLS 算法(65)—(71),假设 $\{v(t)\}$ 是一个与输入 $u(t)$ 无关的独立白噪声,满足:

(E1) $E[v(t)|\mathscr{F}_{t-1}]=0$,a.s.,

(E2) $E[v^2(t)|\mathscr{F}_{t-1}]\leqslant\sigma^2[\ln|\boldsymbol{P}^{-1}(t)|]^{\varepsilon_1}$,

a.s., $0\leqslant\sigma^2<\infty,0\leqslant\varepsilon_1<\infty$,

$\{\mathscr{F}_t\}$ 是由直到 t,且包括时刻 t 的观测生成的 σ 代数序列,假设 $A(z)$ 是严格稳定的(所有零点都在单位圆内),存在正数 c,α,β 和 k,以至于对所有 $s\geqslant k$,下列广义持续激励条件(Generalized Persistent Excitation condition,GPE 条件)成立(无界条件数):

$$(\mathrm{GPE})\quad \alpha\boldsymbol{I}\leqslant\frac{1}{s}\sum_{i=1}^{s}\boldsymbol{\varphi}(t_i)\boldsymbol{\varphi}^{\mathrm{T}}(t_i)\leqslant\beta s^c\boldsymbol{I},\text{a.s.},$$

那么对任意满足 $0\leqslant\varepsilon_1<\varepsilon<\infty$ 的 ε,参数估计误差满足

$$\|\hat{\boldsymbol{\theta}}(t_s)-\boldsymbol{\theta}\|^2=O\left(\frac{[\ln s]^{1+\varepsilon}}{s}\right)\to 0,\text{a.s.},s\to\infty.$$

注 8 定理 10 的条件是很松弛的——我们没有假设噪声方差有界,也没有平稳性,各态遍历和高阶矩存在假设.

定理 10 说明参数估计误差 $\|\hat{\boldsymbol{\theta}}(t_s)-\boldsymbol{\theta}\|^2$ 以 $O\left(\frac{[\ln s]^{1+\varepsilon}}{s}\right)$ 的速度收敛于零($0\leqslant\varepsilon_1<\varepsilon<\infty$),高噪声水平导致慢收敛速度(大噪声方差对应大 ε_1 和 ε).换句话说,大 ε_1 导致慢收敛速度.

注 9 如果将条件(E2)修改为

(E2′) $v^2(i)\leqslant\sigma^2[\ln|\boldsymbol{P}^{-1}(t)|]^{\varepsilon_1}$,a.s.,$i\leqslant t$,
$0\leqslant\sigma^2<\infty$,$0\leqslant\varepsilon_1<\infty$,

或

(E2″) $v^2(i)\leqslant\sigma^2[\ln|\boldsymbol{P}^{-1}(t)|+C_1]^{\varepsilon_1}+C_2$,a.s.,
$i\leqslant t$,$0\leqslant C_1,C_2<\infty$,

定理 10 的结论仍然成立.

注 10 满足条件(E1)和(E2)的噪声包含一大类随机过程,高斯白噪声,二阶矩过程和平稳过程.即使噪声方差使无界的,如(E2),(E2′)或(E2″),定理 10 仍然成立,可参见文献[38]中噪声方差无界的例子.

定理 11[4,29] 对于输出误差系统(63)—(64)和 V-AM-RLS 算法,即遗忘因子 $\lambda=1$ 时的 V-AM-FF-RLS 算法(65)—(71),假设$\{v(t)\}$是零均值,与系统输入 $u(t)$ 是统计独立的[意味(E1)成立],进一步假设:

(E3) $E[v^2(t)|\mathscr{F}_{t-1}]\leqslant\sigma^2 r^{\epsilon_1}(t)$,a.s.,
$0\leqslant\sigma^2<\infty$,$0\leqslant\epsilon_1<1$,

$A(z)$稳定,下列弱持续激励条件成立(有界条件数),即在 GPE 条件中取 $c=0$ 得到

(WPE) $\alpha\boldsymbol{I}\leqslant\dfrac{1}{s}\sum_{i=1}^{s}\phi(t_i)\phi^{\mathrm{T}}(t_i)\leqslant\beta\boldsymbol{I}$,a.s.,

对大 s,$0<\alpha\leqslant\beta<\infty$,

那么对于任意满足 $0\leqslant\epsilon_1<\epsilon<1$ 的 ϵ,参数估计误差满足:

$$\|\hat{\boldsymbol{\theta}}(t_s)-\boldsymbol{\theta}\|^2=O\left(\frac{1}{s^{1-\epsilon}}\right)\to 0,\text{a.s.},s\to\infty.$$

注 11 定理 11 说明参数估计误差 $\|\hat{\boldsymbol{\theta}}(t_s)-\boldsymbol{\theta}\|^2$ 以 $O\left(\dfrac{1}{s^{1-\epsilon}}\right)$的速度收敛于零($0\leqslant\epsilon_1<\epsilon<1$);小 ϵ_1(即低噪声水平)时,参数估计更快地收敛到其真值.

由定理 11 的证明可知$\{v(t)\}$是一个零均值独立随机噪声,且满足:

(E3′)$v^2(t)\leqslant\sigma^2[r(t)+C_1]^{\epsilon}+C_2$,a.s.,$0<C_1,C_2<\infty$,用其代替假设(E3),那么定理 11 的结论依然成立.

注 12 这里的损失数据系统辨识方法是基于输出误差模型的,其参数的估计和损失输出估计具有良好的收敛性质.然而,一些损失数据系统辨识方法基于方程误差模型,即受控自回归(CAR)模型,亦即 ARX 模型,其收敛性分析还没有取得令人满意的结果.

5.2.3 输出估计的收敛性

定理 12[4,29] 对于输出误差系统(63)—(64)和 V-AM-RLS 算法,即遗忘因子 $\lambda=1$ 时的 V-AM-FF-RLS 算法(65)—(71),定理 10 的条件成立,$A(z)$严格稳定,令 $\Delta_s:=\max[t_1^*,t_2^*,\cdots,t_s^*]$,假设存在常数 $0<\delta_1<\infty$ 和 $0\leqslant\mu<1$ 使得下列关系成立:

(E4) $t_{s+1}\leqslant\delta_1 t_s$,或 $t_{s+1}=O(t_s)$,

(E5) $\Delta_s\leqslant\delta_1 t_s^{\mu}$,或 $\Delta_s=O(t_s^{\mu})$,

那么有界输入意味着存在常数 k 使得输出估计误差 $x_a(t)-x(t)$ 满足:

1) $\sum_{i=t_k}^{t_s}[x_a(i)-x(i)]^2=O(t_s^{\mu}[\ln t_s]^{2+\varepsilon})$,a.s.,

2) $\dfrac{1}{t}\sum_{i=1}^{t}[x_a(i)-x(i)]^2=O\left(\dfrac{[\ln t]^{2+\varepsilon}}{t^{1-\mu}}\right)\to 0$,a.s.,

3) $x_a(t)-x(t)\to 0$.

注 13 条件(E4)和(E5)是合理的普通假设,它包括了连续两个可得到数据的长度为有限的情况,即 $\mu=0$.定理 12 说明从 $t=t_k$ 到 $t=t_s$ 输出误差 $[x_a(t)-x(t)]$的平方和以 $O(t_s^{\mu}[\ln t_s]^{2+\varepsilon})$为界.输出估计误差$[x_a(t)-x(t)]^2$ 在平均意义下以速度 $O\left(\dfrac{[\ln t]^{2+\varepsilon}}{t^{1-\mu}}\right)$收敛于零.一个小 μ(少损失数据)和 ε(低噪信比)导致快速收敛($\mu=0$ 包括了一般的规则数据可得到).输出估计 $x_a(t)$ 收敛于真实输出 $x(t)$.

定理 13[4,29] 对于输出误差系统(63)—(64)和 V-AM-RLS 算法,即遗忘因子 $\lambda=1$ 时的 V-AM-FF-RLS 算法(65)—(71),定理 11 的条件和(E4)—(E5)成立,且 $\mu+\epsilon<1$,那么有界输入意味着存在正整数 k 使得输出估计误差满足:

1) $\sum_{i=t_k}^{t_s}[x_a(i)-x(i)]^2=O(t_s^{\mu+\epsilon})$,a.s.,

2) $\dfrac{1}{t}\sum_{i=1}^{t}[x_a(i)-x(i)]^2=O\left(\dfrac{1}{t^{1-\mu-\epsilon}}\right)\to 0$,
a.s., $t\to\infty$.

注 14 从定理 12 和定理 13 可知,当参数估计收敛于其真值,式(69)中辅助模型的输出 $x_a(t)$ 就是 $x(t)$的一致估计.当 $v(t)$ 为白噪声时,$x_a(t)$ 可以看作系统输出 $y(t)$ 的最优估计.严格无源条件(strictly passive condition),即 $H(z)$ 是严格正实条件是保证辨识算法收敛的条件.

6 辅助模型递推广义增广最小二乘辨识方法

Box-Jenkins 系统的辅助模型递推广义增广最小二乘(AM-RGELS)辨识算法的收敛性已经在《系统辨识——辨识方法性能分析》2.7 节进行了讨论,研究思路借鉴于国际期刊《International Journal of Com-

puter Mathematics》2012 年第 15 期上的论文《Performance analysis of the auxiliary model-based least-squares identification algorithm for one-step state-delay systems》(一步状态迟延系统的输入输出表达的辅助模型递推最小二乘辨识算法的性能分析)中的方法[37],本节在此基础上研究自回归 Box-Jenkins 系统的 AM-RGELS 辨识算法的收敛性,相关内容参见文献[4].

6.1 辅助模型递推广义增广最小二乘算法

考虑下列 OEARMA 模型,即 Box-Jenkins 模型(BJ 模型)描述的随机系统:

$$y(t)=\frac{B(z)}{A(z)}u(t)+\frac{D(z)}{C(z)}v(t), \tag{88}$$

其中$\{u(t)\}$和$\{y(t)\}$分别为系统的输入和输出序列,$\{v(t)\}$是零均值、方差为σ^2的随机白噪声序列,$A(z),B(z),C(z)$和$D(z)$均为单位后移算子z^{-1}的多项式:

$$\begin{aligned}
&A(z):=1+a_1z^{-1}+a_2z^{-2}+\cdots+a_{n_a}z^{-n_a},\\
&B(z):=b_1z^{-1}+b_2z^{-2}+\cdots+b_{n_b}z^{-n_b},\\
&C(z):=1+c_1z^{-1}+c_2z^{-2}+\cdots+c_{n_c}z^{-n_c},\\
&D(z):=1+d_1z^{-1}+d_2z^{-2}+\cdots+d_{n_d}z^{-n_d}.
\end{aligned}$$

设阶次n_a,n_b,n_c和n_d已知,记$n:=n_a+n_b+n_c+n_d$,且$t\leqslant 0$时,$u(t)=0,y(t)=0,v(t)=0$.系统量测噪声$w(t):=\dfrac{D(z)}{C(z)}v(t)$为自回归滑动平均模型(AutoRegressive Moving Average model,ARMA 模型).

定义系统模型输出和噪声模型输出分别为

$$x(t):=\frac{B(z)}{A(z)}u(t), \tag{89}$$

$$w(t):=\frac{D(z)}{C(z)}v(t). \tag{90}$$

置系统参数向量$\boldsymbol{\vartheta}$,系统模型参数向量$\boldsymbol{\theta}$和噪声模型参数向量$\boldsymbol{\rho}$分别为

$$\begin{aligned}
&\boldsymbol{\vartheta}:=\begin{bmatrix}\boldsymbol{\theta}\\ \boldsymbol{\rho}\end{bmatrix}\in\mathbf{R}^{n_a+n_b+n_c+n_d},\\
&\boldsymbol{\theta}:=[a_1,a_2,\cdots,a_{n_a},b_1,b_2,\cdots,b_{n_b}]^{\mathrm{T}}\in\mathbf{R}^{n_a+n_b},\\
&\boldsymbol{\rho}:=[c_1,c_2,\cdots,c_{n_c},d_1,d_2,\cdots,d_{n_d}]^{\mathrm{T}}\in\mathbf{R}^{n_c+n_d},
\end{aligned}$$

信息向量$\boldsymbol{\varphi}(t)$,系统模型信息向量$\boldsymbol{\phi}(t)$和噪声模型信息向量$\boldsymbol{\psi}(t)$分别为

$$\begin{aligned}
&\boldsymbol{\varphi}(t):=\begin{bmatrix}\boldsymbol{\phi}(t)\\ \boldsymbol{\psi}(t)\end{bmatrix}\in\mathbf{R}^{n_a+n_b+n_c+n_d},\\
&\boldsymbol{\phi}(t):=[-x(t-1),-x(t-2),\cdots,-x(t-n_a),\\
&\qquad u(t-1),u(t-2),\cdots,u(t-n_b)]^{\mathrm{T}}\in\mathbf{R}^{n_a+n_b},\\
&\boldsymbol{\psi}(t):=[-w(t-1),-w(t-2),\cdots,-w(t-n_c),\\
&\qquad v(t-1),v(t-2),\cdots,v(t-n_d)]^{\mathrm{T}}\in\mathbf{R}^{n_c+n_d}.
\end{aligned}$$

式(89)—(90)和式(88)写成向量形式分别为

$$\begin{aligned}
x(t)&=[1-A(z)]x(t)+B(z)u(t)=\\
&\quad\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta},
\end{aligned} \tag{91}$$

$$\begin{aligned}
w(t)&=[1-C(z)]w(t)+D(z)v(t)=\\
&\quad\boldsymbol{\psi}^{\mathrm{T}}(t)\boldsymbol{\rho}+v(t),
\end{aligned} \tag{92}$$

$$\begin{aligned}
y(t)&=x(t)+w(t)=\boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}+\boldsymbol{\psi}^{\mathrm{T}}(t)\boldsymbol{\rho}+v(t)=\\
&\quad[\boldsymbol{\phi}^{\mathrm{T}}(t),\boldsymbol{\psi}^{\mathrm{T}}(t)]\begin{bmatrix}\boldsymbol{\theta}\\ \boldsymbol{\rho}\end{bmatrix}+v(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\vartheta}+v(t),
\end{aligned} \tag{93}$$

令$\hat{\boldsymbol{\vartheta}}(t):=\begin{bmatrix}\hat{\boldsymbol{\theta}}(t)\\ \hat{\boldsymbol{\rho}}(t)\end{bmatrix}$为$\boldsymbol{\vartheta}=\begin{bmatrix}\boldsymbol{\theta}\\ \boldsymbol{\rho}\end{bmatrix}$的估计.基于 BJ 系统辨识模型(93),信息向量$\boldsymbol{\varphi}(t)$中未知项$x(t-i)$,$w(t-i)$和$v(t-i)$分别用辅助模型(99)的输出$x_{\mathrm{a}}(t-i)$,估计的相关噪声项$\hat{w}(t-i)$和估计的白噪声项$\hat{v}(t-i)$代替,代替后的信息向量记作$\hat{\boldsymbol{\varphi}}(t)$,在辨识算法中使用$\hat{\boldsymbol{\varphi}}(t)$,我们能够获得辨识 BJ 系统的辅助模型递推广义增广最小二乘算法(Auxiliary Model based Recursive Generalized Extended Least Squares algorithm,AM-RGELS 算法):

$$\begin{aligned}
&\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)[y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)],\\
&\hat{\boldsymbol{\vartheta}}(0)=\mathbf{1}_n/p_0,
\end{aligned} \tag{94}$$

$$\boldsymbol{P}^{-1}(t)=\boldsymbol{P}^{-1}(t-1)+\hat{\boldsymbol{\varphi}}(t)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}_n, \tag{95}$$

$$\hat{\boldsymbol{\varphi}}(t)=\begin{bmatrix}\hat{\boldsymbol{\phi}}(t)\\ \hat{\boldsymbol{\psi}}(t)\end{bmatrix},\quad \hat{\boldsymbol{\vartheta}}(t)=\begin{bmatrix}\hat{\boldsymbol{\theta}}(t)\\ \hat{\boldsymbol{\rho}}(t)\end{bmatrix}, \tag{96}$$

$$\begin{aligned}
\hat{\boldsymbol{\phi}}(t)&=[-x_{\mathrm{a}}(t-1),\cdots,-x_{\mathrm{a}}(t-n_a),u(t-1),\\
&\quad\cdots,u(t-n_b)]^{\mathrm{T}},
\end{aligned} \tag{97}$$

$$\begin{aligned}
\hat{\boldsymbol{\psi}}(t)&=[-\hat{w}(t-1),\cdots,-\hat{w}(t-n_c),\hat{v}(t-1),\\
&\quad\cdots,\hat{v}(t-n_d)]^{\mathrm{T}},
\end{aligned} \tag{98}$$

$$x_{\mathrm{a}}(t)=\hat{\boldsymbol{\phi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \tag{99}$$

$$\hat{w}(t)=y(t)-x_{\mathrm{a}}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \tag{100}$$

$$\hat{v}(t)=\hat{w}(t)-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\rho}}(t)=y(t)-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t), \tag{101}$$

$$\begin{aligned}
\hat{\boldsymbol{\theta}}(t)&=[\hat{a}_1(t),\hat{a}_2(t),\cdots,\hat{a}_{n_a}(t),\hat{b}_1(t),\hat{b}_2(t),\\
&\quad\cdots,\hat{b}_{n_b}(t)]^{\mathrm{T}},
\end{aligned} \tag{102}$$

$$\begin{aligned}
\hat{\boldsymbol{\rho}}(t)&=[\hat{c}_1(t),\hat{c}_2(t),\cdots,\hat{c}_{n_c}(t),\hat{d}_1(t),\\
&\quad\hat{d}_2(t),\cdots,\hat{d}_{n_d}(t)]^{\mathrm{T}}.
\end{aligned} \tag{103}$$

6.2 AM-RGELS 辨识算法的收敛性

定理 14[2,4] 对于系统(88),辨识模型(93)和 AM-RGELS 辨识算法(94)—(103),假设$\{v(t)\}$是一个零均值、方差为σ^2的不相关噪声序列,即:

(F1) $E[v(t)]=0,\quad E[v^2(t)]=\sigma^2,$
$E[v(t)v(i)]=0,\quad t\neq i,$

存在正常数 α,β 和整数 t_0,使得当 $t\geqslant t_0$ 时,下列持续激励条件成立:

(F2) $\alpha\boldsymbol{I}_n\leqslant\frac{1}{t}\sum_{j=1}^{t}\hat{\boldsymbol{\varphi}}(j)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(j)\leqslant\beta\boldsymbol{I}_n$,a.s.,

那么参数估计误差 $\hat{\boldsymbol{\vartheta}}(t)-\boldsymbol{\vartheta}$ 收敛于零.

证明 定义参数估计误差向量 $\tilde{\boldsymbol{\vartheta}}(t):=\hat{\boldsymbol{\vartheta}}(t)-\boldsymbol{\vartheta}$.使用式(93),式(94)两边减去 $\boldsymbol{\vartheta}$,得到

$\tilde{\boldsymbol{\vartheta}}(t)=\tilde{\boldsymbol{\vartheta}}(t-1)+\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)[-\tilde{y}(t)+\xi(t)+v(t)]$,(104)

式中

$\tilde{y}(t):=\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\tilde{\boldsymbol{\vartheta}}(t-1)\in\mathbf{R},$
$\xi(t):=[\boldsymbol{\varphi}(t)-\hat{\boldsymbol{\varphi}}(t)]^{\mathrm{T}}\boldsymbol{\vartheta}\in\mathbf{R}.$

令

$T(t):=\tilde{\boldsymbol{\vartheta}}^{\mathrm{T}}(t)\boldsymbol{P}^{-1}(t)\tilde{\boldsymbol{\vartheta}}(t).$

使用式(104)和式(95),有

$$\begin{aligned}T(t)=&T(t-1)-[1-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)]\tilde{y}^2(t)+\\&2[1-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)]\tilde{y}(t)[\xi(t)+v(t)]+\\&\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)[v^2(t)+\xi^2(t)+2\xi(t)v(t)]\leqslant\\&T(t-1)+2[1-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)]\tilde{y}(t)[\xi(t)+v(t)]+\\&\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)[v^2(t)+\xi^2(t)+2\xi(t)v(t)].\end{aligned}\tag{105}$$

注意到

$1-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)=[1+\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\varphi}}(t)]^{-1}\geqslant0$,

$\{v(t)\}$是一个零均值白噪声,且独立于输入信号$\{u(t)\}$.对于给定的某常数 ε,假设 $\xi(t)$是有界的,即 $E[\xi^2(t)]\leqslant\varepsilon$.因为 $T(t-1),\tilde{y}(t),\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)$及 $\xi(t)$与 $v(t)$不相关,式(105)两边取期望,使用条件(F1)可得:

$$\begin{aligned}&E[T(t)]\boldsymbol{P}E[T(t-1)]+E[\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)](\sigma^2+\varepsilon)\leqslant\\&E[T(0)]+E\Big[\sum_{j=1}^{t}\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(j)\boldsymbol{P}(j)\hat{\boldsymbol{\varphi}}(j)\Big](\sigma^2+\varepsilon).\end{aligned}\tag{106}$$

由式(95)可得

$\boldsymbol{P}^{-1}(t-1)=\boldsymbol{P}^{-1}(t)[\boldsymbol{I}_n-\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)].$

取行列式得到

$$\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)=\frac{|\boldsymbol{P}^{-1}(t)|-|\boldsymbol{P}^{-1}(t-1)|}{|\boldsymbol{P}^{-1}(t)|}.$$

参考文献[2],用 j 代替 t,对 j 从 $j=1$ 到 $j=t$ 求和,可得

$$\sum_{j=1}^{t}\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(j)\boldsymbol{P}(j)\hat{\boldsymbol{\varphi}}(j)\leqslant\ln|\boldsymbol{P}^{-1}(t)|+n\ln p_0.\tag{107}$$

利用(F2),由式(95)可得

$(\alpha t+1/p_0)^n\leqslant|\boldsymbol{P}^{-1}(t)|\leqslant(\beta t+1/p_0)^n.$

代入式(107)可得

$$\begin{aligned}\sum_{j=1}^{t}\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(j)\boldsymbol{P}(j)\hat{\boldsymbol{\varphi}}(j)&\leqslant n\ln(\beta t+1/p_0)+n\ln p_0=\\&n\ln(p_0\beta t+1).\end{aligned}$$

根据 $T(t)$的定义,有不等式

$$\begin{aligned}E[T(t)]=&E[\tilde{\boldsymbol{\vartheta}}^{\mathrm{T}}(t)\boldsymbol{P}^{-1}(t)\tilde{\boldsymbol{\vartheta}}(t)]\geqslant\\&(\alpha t+1/p_0)E[\|\tilde{\boldsymbol{\vartheta}}(t)\|^2].\end{aligned}$$

由于 $T(0)=E[\tilde{\boldsymbol{\vartheta}}^{\mathrm{T}}(0)\boldsymbol{P}^{-1}(0)\tilde{\boldsymbol{\vartheta}}(0)]=n/p_0^3$,因此由式(106)有

$$\begin{aligned}&(\alpha t+1/p_0)E[\|\tilde{\boldsymbol{\vartheta}}(t)\|^2]n\leqslant T(t)\leqslant\\&n/p_0^3+[n\ln(\beta t+1/p_0)+n\ln p_0](\sigma^2+\varepsilon),\end{aligned}$$

取极限可得

$$\lim_{t\to\infty}\|\tilde{\boldsymbol{\vartheta}}(t)\|^2\leqslant\lim_{t\to\infty}\frac{n/p_0^3+[n\ln(\beta t+1/p_0)+n\ln p_0](\sigma^2+\varepsilon)}{\alpha t+1/p_0}=0,\quad\text{m.s.}$$

这说明参数估计误差 $\|\hat{\boldsymbol{\vartheta}}(t)-\boldsymbol{\vartheta}\|$ 随 t 增大均方收敛于零.

注 15 在证明中假设 $E[\xi^2(t)]\leqslant\varepsilon$,否则,令 $\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)$.在对式(105)取期望时,假设 $E\{[1-\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\hat{\boldsymbol{\varphi}}(t)]\tilde{y}(t)\xi(t)\}=0$,这是比较苛刻的.此外,条件(F2)是关于 $\hat{\boldsymbol{\varphi}}(t)$的弱持续激励条件,如果改为是关于 $\boldsymbol{\varphi}(t)$的弱持续激励条件,那么 AM-RGELS 算法在这更弱的条件下的收敛性仍是开放研究课题.相关结果参见文献[2,37].

7 结语

辨识算法的收敛性分析是辨识的永恒研究主题.每个辨识算法一诞生就伴随着收敛性分析和证明.辅助模型辨识算法的收敛性分析是极其困难的.本文针对白噪声干扰输出误差系统的辅助模型随机梯度算法、辅助模型递推最小二乘算法、辅助模型多新息随机梯度算法、变间隔辅助模型随机梯度算法、变间隔辅助模型递推最小二乘算法在理论上做了比较详细的分析.用一种很牵强的方式或者说很近似的方式研究了有色噪声干扰 Box-Jenkins 系统辅助模型递推广义增广最小二乘算法的收敛性.其他有色噪声干扰的线性系统和非线性系统的辅助模型辨识算法的一致收敛性和收敛条件都是国际控制领域的著名研究难题,如自回归输出误差误差类系统、线性参数自回归输出误差系统、输入非线性系统等的辅助模型辨识方法等.本文的收敛性研究方法可以延伸研究其他线性系统[39-43]、线性参数系统[44-48]、双线性系统[49]、非线性系统[50-55].

参考文献
References

[1] 丁锋.系统辨识新论[M].北京:科学出版社,2013
DING Feng. System identification: New theory and methods[M].Beijing:Science Press,2013
[2] 丁锋.系统辨识:辨识方法性能分析[M].北京:科学出版社,2014
DING Feng. System identification: Performance analysis for identification methods [M]. Beijing: Science Press,2014
[3] 丁锋.系统辨识:多新息辨识理论与方法[M].北京:科学出版社,2016
DING Feng.System identification:Multi-innovation identification theory and methods [M]. Beijing: Science Press,2016
[4] 丁锋.系统辨识:辅助模型辨识思想与方法[M].北京:科学出版社,2017
DING Feng.System identification:Auxiliary model identification idea and methods [M]. Beijing: Science Press,2017
[5] 丁锋.方程误差类系统的多新息辨识方法[J].南京信息工程大学学报(自然科学版),2015,7(5):385-407
DING Feng. Multi-innovation identification methods for equation-error type systems[J].Journal of Nanjing University of Information Science and Technology (Natural Science Edition),2015,7(5):385-407
[6] 丁锋.系统辨识算法复杂性、收敛性、计算效率研究[J].控制与决策,2016,31(10):1729-1741
DING Feng. Complexity, convergence and computational efficiency for system identification algorithms[J].Control and Decision,2016,31(10):1729-1741
[7] Ding F,Chen T.Modeling and identification for multirate systems[J]. Acta Automatica Sinica, 2005, 31(1):105-122
[8] 丁锋,杨慧中,纪志成.时变系统辨识方法及其收敛定理[J].江南大学学报(自然科学版),2006,5(1):115-126
DING Feng, YANG Huizhong JI Zhicheng. Time-varying system identification methods and convergence theorems [J].Journal of Southern Yangtze University (Natural Science Edition),2006,5(1):115-126
[9] Ding F, Chen T. Combined parameter and output estimation of dual-rate systems using an auxiliary model [J].Automatica,2004,40(10):1739-1748
[10] Ding F, Chen T. Parameter estimation of dual-rate stochastic systems by using an output error method[J]. IEEE Transactions on Automatic Control,2005,50(9):1436-1441
[11] Duan H H, Jia J, Ding R F. Two-stage recursive least squares parameter estimation algorithm for output error models[J].Mathematical and Computer Modelling,2012,55(3-4):1151-1159
[12] 丁锋.基于输出估计的多输入系统随机梯度估计算法[J].南京信息工程大学学报(自然科学版),2010,2(6):481-488
DING Feng. Stochastic gradient estimation algorithm for multiple-input systems based on the output estimation [J].Journal of Nanjing University of Information Science and Technology (Natural Science Edition),2010,2(6):481-488
[13] Wang D Q, Ding F. Performance analysis of the auxiliary models based multi-innovation stochastic gradient estimation algorithm for output error systems[J].Digital Signal Processing,2010,20(3):750-762
[14] Ding F, Liu X P, Liu G. Gradient based and least-squares based iterative identification methods for OE and OEMA systems[J]. Digital Signal Processing, 2010, 20(3):664-677
[15] 王冬青.基于辅助模型的递推增广最小二乘辨识方法[J].控制理论与应用,2009,26(1):51-56
WANG Dongqing.Recursive extended least squares identification method based on auxiliary models[J]. Control Theory and Applications,2009,26(1):51-56
[16] Ding F, Liu X P, Liu G. Auxiliary model based multi-innovation extended stochastic gradient parameter estimation with colored measurement noises[J]. Signal Processing,2009,89(10):1883-1890
[17] Wang D Q. Least squares-based recursive and iterative estimation for output error moving average systems using data filtering[J]. IET Control theory and Applications, 2011,5(14):1648-1657
[18] Ding F, Wang Y J, Ding J. Recursive least squares parameter identification for systems with colored noise using the filtering technique and the auxiliary model[J]. Digital Signal Processing,2015,37:100-108
[19] Ding F, Duan H H. Two-stage parameter estimation algorithms for Box-Jenkins systems [J]. IET Signal Processing,2013,7(2):176-184
[20] Liu Y J, Wang D Q, Ding F. Least-squares based iterative algorithms for identifying Box-Jenkins models with finite measurement data[J].Digital Signal Processing,2010,20(5):1458-1467
[21] Wang D Q, Yang G W, Ding F. Gradient-based iterative parameter estimation for Box-Jenkins systems with finite measurement data[J]. Computers and Mathematics with Applications,2010,60(5):1200-1208
[22] Wang X H, Ding F. Convergence of the auxiliary model based multi-innovation generalized extended stochastic gradient algorithm for Box-Jenkins systems[J]. Nonlinear Dynamics,2015,82(1-2):269-280
[23] 丁锋,郭兰杰.线性参数系统的多新息辨识方法[J].南京信息工程大学学报(自然科学版),2015,7(4):289-312
DING Feng, GUO Lanjie. Multi-innovation identification methods for linear-parameter systems[J].Journal of Nanjing University of Information Science and Technology (Natural Science Edition),2015,7(4):289-312
[24] Mao Y W, Ding F. Data filtering-based multi-innovation stochastic gradient algorithm for nonlinear output error autoregressive systems[J]. Circuits, Systems and Signal Processing,2016,35(2):651-667
[25] Mao Y W, Ding F. Parameter estimation for nonlinear systems by using the data filtering and the multi-innovation identification theory [J]. International Journal of

Computer Mathematics,2016,93(11):1869-1885
[26] Hu P P, Ding F, Sheng J. Auxiliary model based least squares parameter estimation algorithm for feedback nonlinear systems using the hierarchical identification principle[J].Journal of the Franklin Institute,2013,350(10),3248-3259
[27] Hu P P, Ding F. Multistage least squares based iterative estimation for feedback nonlinear systems with moving average noises using the hierarchical identification principle[J].Nonlinear Dynamics,2013,73(1-2),583-592
[28] Ding F, Liu G, Liu X P. Parameter estimation with scarce measurements[J].Automatica,2011,47(8):1646-1655
[29] Ding F, Ding J. Least squares parameter estimation with irregularly missing data [J]. International Journal of Adaptive Control and Signal Processing, 2010, 24(7): 540-553
[30] 丁锋,谢新民.多变量系统的辅助模型辨识算法[J].清华大学学报(自然科学版),1992,32(4):100-106
DING Feng, XIE Xinmin. Auxiliary model identification method for multivariable systems[J].Journal of Tsinghua University: Science and Technology, 1992, 32(4): 100-106
[31] 丁锋.多变量系统的辅助模型辨识方法的收敛性分析[J].控制理论与应用,1997,14(2):192-200
DING Feng. Convergence analysis of auxiliary model identification algorithms for multivariable systems [J]. Control Theory and Applications,1997,14(2):192-200
[32] Ding F, Liu X P. Auxiliary model based stochastic gradient algorithm for multivariable output error systems [J].Acta Automatica Sinica,2010,36(7):993-998
[33] Ding F, Chen H B, Li M. Multi-innovation least squares identification methods based on the auxiliary model for MISO systems [J]. Applied Mathematics and Computation,2007,187(2):658-668
[34] Liu Y J, Xie L, Ding F. An auxiliary model recursive least squares parameter estimation algorithm for non-uniformly sampled multirate systems[J].Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering,2009,223(4):445-454
[35] Xie L, Liu Y J, Yang H Z, et al. Modeling and identification for non-uniformly periodically sampled-data systems[J].IET Control Theory and Applications,2010,4(5):784-794
[36] Ding F, Gu Y. Performance analysis of the auxiliary model-based stochastic gradient parameter estimation algorithm for state space systems with one-step state delay [J]. Circuits, Systems and Signal Processing, 2013, 32(2):585-599
[37] Ding F, Gu Y. Performance analysis of the auxiliary model-based least-squares identification algorithm for one-step state-delay systems[J].International Journal of Computer Mathematics,2012,89(15):2019-2028
[38] Ding F, Chen T, Qiu L. Bias compensation based recursive least squares identification algorithm for MISO systems [J].IEEE Transactions on Circuits and Systems, II: Express Briefs,2006,53(5):349-353
[39] Xu L. A proportional differential control method for a time-delay system using the Taylor expansion approximation[J]. Applied Mathematics and Computation, 2014, 236:391-399
[40] Xu L. Application of the Newton iteration algorithm to the parameter estimation for dynamical systems[J].Journal of Computational and Applied Mathematics, 2015, 288: 33-43
[41] Xu L, Chen L, Xiong W L. Parameter estimation and controller design for dynamic systems from the step responses based on the Newton iteration[J].Nonlinear Dynamics, 2015,79(3):2155-2163
[42] Xu L. The damping iterative parameter identification method for dynamical systems based on the sine signal measurement[J].Signal Processing,2016,120:660-667
[43] Xie L, Yang H Z, Huang B. FIR model identification of multirate processes with random delays using EM algorithm[J].AIChE Journal,2013,59(11):4124-4132
[44] Wang C, Tang T. Recursive least squares estimation algorithm applied to a class of linear-in-parameters output error moving average systems[J]. Applied Mathematics Letters,2014,29:36-41
[45] Wang C, Tang T. Several gradient-based iterative estimation algorithms for a class of nonlinear systems using the filtering technique [J]. Nonlinear Dynamics, 2014,77(3):769-780
[46] Wang C, Zhu L. Parameter identification of a class of nonlinear systems based on the multi-innovation identification theory[J].Journal of the Franklin Institute, 2015,352(10):4624-4637
[47] Guo L J, Wang Y J, Wang C. A recursive least squares algorithm for pseudo-linear arma systems using the auxiliary model and the filtering technique[J].Circuits, Systems and Signal Processing,2016,35(7):2655-2667
[48] Wang F F, Liu Y J, Yang E F. Least squares-based iterative identification methods for linear-in-parameters systems using the decomposition technique[J].Circuits, Systems and Signal Processing, 2016, 35(11): 3863-3881
[49] Meng D D. Recursive least squares and multi-innovation gradient estimation algorithms for bilinear stochastic systems[J].Circuits, Systems and Signal Processing, 2017. DOI:10.1007/s00034-016-0337-0
[50] Chen J, Ni Y X. Parameter identification methods for an additive nonlinear system [J]. Circuits, Systems and Signal Processing,2014,33(10):3053-3064
[51] Chen J, Wang X P. Identification of Hammerstein systems with continuous nonlinearity [J]. Information Processing Letters,2015,115(11):822-827
[52] Wang D Q, Zhang W. Improved least squares identification algorithm for multivariable Hammerstein systems[J].Journal of the Franklin Institute, 2015, 352(11):5292-5307
[53] Wang D Q. Hierarchical parameter estimation for a class of MIMO Hammerstein systems based on the reframed models[J].Applied Mathematics Letters,2016,57:13-19
[54] Li J H. Parameter estimation for Hammerstein CARARMA systems based on the Newton iteration[J].Applied Mathematics Letters,2013,26(1):91-96
[55] Hu Y B, Liu B L, Zhou Q. A multi-innovation generalized

extended stochastic gradient algorithm for output nonlinear autoregressive moving average systems[J].Applied Mathematics and Computation,2014,247:218-224

Auxiliary model based identification methods.Part F:Performance Analysis

DING Feng[1,2,3]

1 School of Internet of Things Engineering,Jiangnan University,Wuxi 214122

2 Control Science and Engineering Research Center,Jiangnan University,Wuxi 214122

3 Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education),Jiangnan University,Wuxi 214122

Abstract Performance analysis of identification methods is the important and difficult projects in the area of system identification. Once one new identification method is born, its convergence analysis appears. The auxiliary model identification is a branch of system identification and has become a large family of identification methods, their convergence brings many projects. This paper studies the consistent convergence of the auxiliary model (AM) based stochastic gradient (SG) algorithm, the AM recursive least squares (RLS) algorithm, the AM multi-innovation SG algorithm, the interval-varying AM SG algorithm and the interval-varying AM RLS algorithm for output-error systems, and analyzes approximately the convergence of the AM recursive generalized extended least squares algorithm for Box-Jenkins systems.

Key words parameter estimation; recursive identification; least squares; auxiliary model identification idea; multi-innovation identification theory; hierarchical identification principle; coupling identification concept; filtering identification idea; linear system