

# 基于输出估计的多输入系统随机梯度估计算法

丁锋<sup>1</sup>

## 摘要

对于输出误差模型描述的多输入单输出系统,辨识的困难在于辨识模型信息向量中包含系统未知输出量(真实输出或无噪输出),以致标准辨识算法无法应用.提出了利用输出估计代替系统真实输出的辨识思想,即通过估计模型预测(估算)系统输出,利用这个估计输出来递推计算系统参数,进而提出了基于输出估计的随机梯度辨识算法,并研究了算法的收敛性,给出了仿真例子.

## 关键词

系统辨识;递推辨识;参数估计;随机梯度;多新息辨识;多变量系统

中图分类号 TP273

文献标志码 A

## 0 引言

### Introduction

2008年,西安交通大学万百五<sup>[1]</sup>教授在《控制理论与应用》上撰文《控制论创立六十年》,综述了控制论提出60年来的光辉发展历程,包括历史遭遇、对哲学引起的变革;更新了它的定义并归纳了它的方法论;对控制论的形成及其奠基人之争发表了看法,对控制论的推广、应用及其未来进行了评论;充分肯定了控制论在社会和经济发展中的贡献,也说明了控制理论对综合国力提升的重要性<sup>[2]</sup>.现代宇航升天、太空探索,高集成度计算机芯片的生产和广泛应用,无不说明控制科学的辉煌成就与贡献.

系统建模与模型辨识(模型参数估计)是一切控制问题的基础.

大千世界每一事物都有其运动规律,不同学科领域的研究对象,其运动规律用方程描述,就是数学模型.因此,可以说不同学科的发展过程就是建立它的数学模型的过程<sup>[3]</sup>.系统辨识是研究建立(动态)系统数学模型的理论与方法.

辨识输入信号的设计,系统辨识新方法的提出,辨识方法的性能分析(辨识精度,参数估计收敛速度,辨识算法的收敛性)是辨识领域的三大基本研究课题<sup>[4]</sup>.

系统辨识的基础是最小二乘原理.最小二乘方法可以追溯到18世纪末,德国数学家高斯在观测和计算天体运行的轨道时提出的,基本原理是极小化误差平方和准则函数.最小二乘方法后来被广泛用于系统辨识领域.随着电子计算机的普及,系统辨识方法(参数估计方法)有了飞速的发展<sup>[3,5]</sup>.

数学建模有很长的历史,但是形成系统辨识学科才几十年的历史.有关论述系统辨识的重要著作有东南大学冯纯伯等<sup>[6]</sup>的《自适应控制》,清华大学方崇智等<sup>[5]</sup>的《过程辨识》,澳大利亚学者 Goodwin 等<sup>[7]</sup>的《Adaptive Filtering, Prediction and Control》,瑞典学者 Ljung<sup>[8]</sup>的《System Identification: Theory for the User》以及本文作者即将出版的《系统辨识理论与方法》<sup>[3]</sup>.

最近几十年,系统辨识在理论和应用上都得到了很大发展,许多方法相继问世,如最小二乘辨识方法、增广最小二乘辨识方法、广义增广最小二乘辨识方法<sup>[9-10]</sup>以及投影辨识方法、随机梯度辨识<sup>[7,8,11-13]</sup>、极大似然辨识方法<sup>[5]</sup>等.近年来,本文作者提出了一些原创性辨识方法,如

收稿日期 2010-10-18

资助项目 国家自然科学基金(60974043)

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

<sup>1</sup> 江南大学 物联网工程学院,无锡,214122

辅助模型辨识方法<sup>[14-18]</sup>、多新息辨识方法<sup>[4,19-24]</sup>、递阶辨识方法<sup>[25-29]</sup>、迭代辨识方法<sup>[30-32]</sup>、多变量系统的耦合辨识方法<sup>[33]</sup>等. 这些新型的系统辨识方法都发表在国内外重要学术学术期刊上, 如《Automatica》、《IEEE Transactions》、《Digital Signal Processing》、《自动化学报》、《控制理论与应用》等. 这些对丰富和推动系统辨识学科的发展有重要作用.

系统辨识方法源于一种思想. 例如, 辅助模型辨识方法是利用辅助模型的输出代替系统的不可测真实输出(无噪输出), 多新息辨识方法是通过扩展新息长度, 递阶辨识方法是借助于辨识模型分解和辨识算法协调原理等.

本文将借助于线性回归模型的多新息辨识理论<sup>[9]</sup>, 通过堆积系统输出、堆积信息矩阵, 以及使用未知输出估计的思想, 来研究多输入单输出系统的辨识问题, 提出了基于输出估计的多新息随机梯度辨识算法, 这是对基于辅助模型的多输入单输出系统多新息最小二乘辨识方法的发展<sup>[34]</sup>.

## 1 问题构成

### Problem formulation

考虑下列状态空间模型描述的多输入单输出系统<sup>[34]</sup>

$$\begin{cases} \mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \\ y_0(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t). \end{cases} \quad (1)$$

其中:  $\mathbf{x}(t) \in \mathbf{R}^n$  是系统状态向量;  $\mathbf{u}(t) = [u_1(t), u_2(t), \dots, u_r(t)]^T \in \mathbf{R}^r$  是系统输入向量;  $y_0(t) \in \mathbf{R}^1$  是系统输出(真实输出或无噪输出);  $\mathbf{A} \in \mathbf{R}^{n \times n}$ ,  $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_r] \in \mathbf{R}^{n \times r}$ ,  $\mathbf{C} \in \mathbf{R}^{1 \times n}$  和  $\mathbf{D} = [d_1, d_2, \dots, d_r] \in \mathbf{R}^{1 \times r}$  是系统参数矩阵.

令  $z^{-1}$  为单位后移算子 [ $z^{-1}\mathbf{x}(t) = \mathbf{x}(t-1)$ ], 由式(1)可得

$$\begin{aligned} y_0(t) &= [\mathbf{C}(z\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}]\mathbf{u}(t) = \\ &= \left[ \frac{\text{Cadj}[\mathbf{I} - \mathbf{A}z^{-1}]\mathbf{B}}{z^{-n}\det[\mathbf{I} - \mathbf{A}z^{-1}]} + \mathbf{D} \right] \mathbf{u}(t) = \\ &= \frac{1}{A(z)} \sum_{i=1}^r B_i(z)u_i(t). \end{aligned} \quad (2)$$

式(2)中  $\mathbf{I}$  为适当维数的单位阵; “ $A = :X$ ” 或 “ $X = :A$ ” 表示 “ $A$  记作(定义为)  $X$ ” 之意;  $A(z)$  是系统特征多项式(阶次为  $n$ ). 多项式  $A(z)$  和  $B_i(z)$  可以表示为

$$\begin{aligned} A(z) &:= z^{-n}\det[z\mathbf{I} - \mathbf{A}] = \\ &= 1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \dots + \alpha_n z^{-n}, \quad \alpha_i \in \mathbf{R}^1, \end{aligned}$$

丁锋. 基于输出估计的多输入系统随机梯度估计算法.

$$\begin{aligned} B_i(z) &:= \text{Cadj}[\mathbf{I} - \mathbf{A}z^{-1}]\mathbf{b}_i + d_i z^{-n}\det[z\mathbf{I} - \mathbf{A}] = \\ &= \beta_{i0} + \beta_{i1}z^{-1} + \beta_{i2}z^{-2} + \dots + \beta_{in}z^{-n}, \quad \beta_{ij} \in \mathbf{R}^1. \end{aligned}$$

假设系统阶次为  $n$ ; 系统输入  $\{u_i(t)\}$  通常是由计算机产生的、也是已知的; 系统真实输出  $y_0(t)$  是未知的, 它的测量值  $y(t)$  包含了一个零均值量测白噪声  $v(t)$ , 即

$$y(t) = y_0(t) + v(t). \quad (3)$$

方程误差模型(CAR/ARX)和CARMA/ARMAX模型的参数估计问题, 在辨识领域得到较多的研究<sup>[3,7-8]</sup>, 而多输入系统输出误差模型(2)、(3)不同于多输入系统ARX模型

$$A(z)y(t) = \sum_{i=1}^r B_i(z)u_i(t) + v(t),$$

所以标准辨识算法不能直接用于估计模型(2)、(3)的参数. 模型(2)、(3)可以化成一个特殊的多输入单输出CARMA/ARMAX模型:

$$\begin{aligned} A(z)y(t) &= \sum_{i=1}^r B_i(z)u_i(t) + D(z)v(t); \\ D(z) &:= A(z). \end{aligned}$$

尽管有许多方法, 包括增广最小二乘算法<sup>[5,7-8]</sup>能够用来估计这个模型的参数, 但是需要估计更多的参数, 因为这个ARMAX模型比模型(2)、(3)多  $n$  个参数(多项式  $D(z)$  的系数), 算法的计算量加大. 通常的做法是把模型写为形式

$$A(z)[y(t) - v(t)] = \sum_{i=1}^r B_i(z)u_i(t)$$

进行辨识, 这种思想可扩展为本文的方法. 本文的目标就是利用输入输出测量数据  $\{u_1(t), u_2(t), \dots, u_r(t), y(t)\}$ , 提出一个计算有效的辨识方法来估计模型(2)、(3)的参数  $(\alpha_i, \beta_{ij})$ , 并研究算法的性能.

许多出版物研究了ARX和ARMAX模型辨识算法的性能, 然而很少研究输出误差模型辨识算法的收敛性问题. 输出误差模型辨识的困难在于辨识模型信息向量中存在系统未知真实输出(也称无噪输出), 而本文正是提出基于未知输出估计的系统参数辨识算法. 它的基本思想是通过估计模型预测(估算)系统输出, 再利用这个估计输出来递推计算系统参数, 并研究提出算法的性能.

## 2 算法推导

### The derivation of algorithm

为了下面的叙述方便, 先定义一些符号:  $\hat{\boldsymbol{\theta}}(t)$  为  $t$  时刻参数向量  $\boldsymbol{\theta}$  的估计;  $\mathbf{1}_n$  是一个元均为 1 的  $n$  维列向量; 上标 T 表示矩阵转置;  $\|\mathbf{X}\|^2 = \text{tr}[\mathbf{X}\mathbf{X}^T]$

表示矩阵的范数; $\lambda_{\max}[\mathbf{X}]$ 和 $\lambda_{\min}[\mathbf{X}]$ 分别表示矩阵 $\mathbf{X}$ 的最大和最小特征值; $\mathbf{f}(t) = \mathbf{o}(\mathbf{g}(t))$ 表示当 $t \rightarrow \infty$ 时, $\mathbf{f}(t)/\mathbf{g}(t) \rightarrow 0$ .

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

$$\boldsymbol{\theta} :=$$

$$[\alpha_1, \alpha_2, \dots, \alpha_n, \beta_{10}, \beta_{11}, \dots, \beta_{1n}, \dots, \beta_{r0}, \beta_{r1}, \dots, \beta_{rn}]^T \in \mathbf{R}^{n_0};$$

$$\boldsymbol{\varphi}(t) := [\boldsymbol{\psi}^T(t), \boldsymbol{\phi}^T(t)]^T \in \mathbf{R}^{n_0}, n_0 := n + r(n + 1).$$

其中,

$$\boldsymbol{\psi}(t) :=$$

$$[-y_0(t-1), -y_0(t-2), \dots, -y_0(t-n)]^T \in \mathbf{R}^n,$$

$$\boldsymbol{\phi}(t) :=$$

$$[u_1(t), \dots, u_1(t-n), u_2(t), \dots, u_2(t-n), \dots, u_r(t), \dots, u_r(t-n)]^T \in \mathbf{R}^{r(n+1)}.$$

从式(2)和(3)可以得到辨识模型:

$$y_0(t) = \boldsymbol{\varphi}^T(t)\boldsymbol{\theta}, y(t) = \boldsymbol{\varphi}^T(t)\boldsymbol{\theta} + v(t). \quad (4)$$

令 $p$ 是数据长度( $p \gg n_0$ ),定义堆积输出向量 $\mathbf{Y}(t)$ 、堆积信息矩阵 $\boldsymbol{\Phi}(t)$ 和堆积噪声向量 $\mathbf{V}(t)$ 如下:

$$\mathbf{Y}(t) := \begin{bmatrix} y(t) \\ y(t-1) \\ \vdots \\ y(t-p+1) \end{bmatrix},$$

$$\boldsymbol{\Phi}(t) := \begin{bmatrix} \boldsymbol{\varphi}^T(t) \\ \boldsymbol{\varphi}^T(t-1) \\ \vdots \\ \boldsymbol{\varphi}^T(t-p+1) \end{bmatrix},$$

$$\mathbf{V}(t) := \begin{bmatrix} v(t) \\ v(t-1) \\ \vdots \\ v(t-p+1) \end{bmatrix}. \quad (5)$$

从式(4)和(5)可得

$$\mathbf{Y}(t) = \boldsymbol{\Phi}(t)\boldsymbol{\theta} + \mathbf{V}(t). \quad (6)$$

构造准则函数

$$J(\boldsymbol{\theta}) := \|\mathbf{Y}(t) - \boldsymbol{\Phi}(t)\boldsymbol{\theta}\|^2. \quad (7)$$

因为 $\boldsymbol{\Phi}(t)$ 包含了未知的无噪声输出 $y_0(t-i), i=1, 2, \dots, n$ ,所以不可能通过极小化准则函数 $J(\boldsymbol{\theta})$ ,运用标准最小二乘方法直接获得参数向量 $\boldsymbol{\theta}$ 的最小二乘估计 $[\boldsymbol{\Phi}^T(t)\boldsymbol{\Phi}(t)]^{-1}\boldsymbol{\Phi}^T(t)\mathbf{Y}(t)$ .这里采用递阶辨识的交互估计原理: $y_0(t-i)$ 用其估计 $\hat{y}(t-i)$ 代替,信息向量 $\boldsymbol{\varphi}(t)$ 中的 $y_0(t-i)$ 用 $\hat{y}(t-i)$ 代替后记作 $\hat{\boldsymbol{\varphi}}(t)$ .那么根据式(4),估计 $\hat{y}(t-i)$ 可通过下式计算

$$\hat{y}(t) = \hat{\boldsymbol{\varphi}}^T(t)\hat{\boldsymbol{\theta}}(t),$$

$$\hat{\boldsymbol{\varphi}}(t) = [\hat{\boldsymbol{\psi}}^T(t), \hat{\boldsymbol{\phi}}^T(t)]^T,$$

$$\hat{\boldsymbol{\psi}}(t) = [-\hat{y}(t-1), -\hat{y}(t-2), \dots, -\hat{y}(t-n)] \in \mathbf{R}^n.$$

进一步 $\boldsymbol{\Phi}(t)$ 中的 $\boldsymbol{\varphi}(t-i)$ 用 $\hat{\boldsymbol{\varphi}}(t-i)$ 代替后记作 $\hat{\boldsymbol{\Phi}}(t)$ ,即

$$\hat{\boldsymbol{\Phi}}(t) = \begin{bmatrix} \hat{\boldsymbol{\varphi}}^T(t) \\ \hat{\boldsymbol{\varphi}}^T(t-1) \\ \vdots \\ \hat{\boldsymbol{\varphi}}^T(t-p+1) \end{bmatrix}. \quad (8)$$

用 $\hat{\boldsymbol{\Phi}}(t)$ 代替式(7)中的 $\boldsymbol{\Phi}(t)$ ,极小化 $J(\boldsymbol{\theta})$ ,使用梯度搜索原理容易得到估计参数向量 $\boldsymbol{\theta}$ 的递推算法:

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\Phi}}^T(t)}{r(t)}[\mathbf{Y}(t) - \hat{\boldsymbol{\Phi}}(t)\hat{\boldsymbol{\theta}}(t-1)],$$

$$t = 1, 2, 3, \dots$$

这里 $\frac{1}{r(t)}$ 是步长或收敛因子.上式也可写作

$$\hat{\boldsymbol{\theta}}(t) = \left[ \mathbf{I} - \frac{\hat{\boldsymbol{\Phi}}^T(t)\hat{\boldsymbol{\Phi}}(t)}{r(t)} \right] \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\Phi}}^T(t)}{r(t)}\mathbf{Y}(t).$$

为了保证 $\hat{\boldsymbol{\theta}}(t)$ 的收敛性, $\hat{\boldsymbol{\theta}}(t-1)$ 的系数矩阵的特征值必须在单位圆内.如果 $r(t)$ 按下面式(10)选择,那么就可以得到基于未知输出估计的随机梯度算法:

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\Phi}}^T(t)}{r(t)}[\mathbf{Y}(t) - \hat{\boldsymbol{\Phi}}(t)\hat{\boldsymbol{\theta}}(t-1)]; \quad (9)$$

$$r(t) = r(t-1) + \text{tr}[\hat{\boldsymbol{\Phi}}(t)\hat{\boldsymbol{\Phi}}^T(t)], r(0) = 1; \quad (10)$$

$$\mathbf{Y}(t) = [y(t), y(t-1), \dots, y(t-p+1)]^T; \quad (11)$$

$$\hat{\boldsymbol{\Phi}}(t) = [\hat{\boldsymbol{\varphi}}(t), \hat{\boldsymbol{\varphi}}(t-1), \dots, \hat{\boldsymbol{\varphi}}(t-p+1)]^T; \quad (12)$$

$$\hat{\boldsymbol{\varphi}}(t) = \begin{bmatrix} \hat{\boldsymbol{\psi}}(t) \\ \hat{\boldsymbol{\phi}}(t) \end{bmatrix}; \quad (13)$$

$$\hat{\boldsymbol{\psi}}(t) = [-\hat{y}(t-1), -\hat{y}(t-2), \dots, -\hat{y}(t-n)]^T; \quad (14)$$

$$\hat{y}(t) = \hat{\boldsymbol{\varphi}}^T(t)\hat{\boldsymbol{\theta}}(t); \quad (15)$$

$$\hat{\boldsymbol{\phi}}(t) = [u_1(t), \dots, u_1(t-n), u_2(t), \dots, u_2(t-n), \dots, u_r(t), \dots, u_r(t-n)]^T. \quad (16)$$

参数向量初始值选择为一个很小实向量,如 $\hat{\boldsymbol{\theta}}(0) = 10^{-6}\mathbf{1}_{n_0}$ .

因为 $e(t) = y(t) - \hat{\boldsymbol{\varphi}}^T(t)\hat{\boldsymbol{\theta}}(t-1)$ 为新息(Innovation),而 $\mathbf{E}(t) = \mathbf{Y}(t) - \hat{\boldsymbol{\Phi}}(t)\hat{\boldsymbol{\theta}}(t-1) \in \mathbf{R}^p$ 为新息向量,即多新息(Multi-Innovation),所以算法(9)~(16)又称为基于输出估计的多新息随机梯度辨识算法,这是对ARX模型多新息辨识方法的发展<sup>[4]</sup>.

### 3 算法性能分析

Performance analysis of the algorithm

定理 对于系统(4)或(6)和算法(9)~(16),

假设  $\{v(t), \mathcal{F}_t\}$  是定义在概率空间  $\{\Omega, \mathcal{F}, P\}$  上的鞅差向量序列, 其中  $\{\mathcal{F}_t\}$  是直到且包括  $t$  时刻的观测生成的 sigma 代数序列, 噪声序列  $\{v(t)\}$  满足

$$(A1) \quad E[v(t) | \mathcal{F}_{t-1}] = 0; \quad (17)$$

$$(A2) \quad E[v^2(t) | \mathcal{F}_{t-1}] \leq \sigma^2 < \infty, \quad (18)$$

且当  $r(t) \rightarrow \infty$  时, 存在一个整数序列  $[t_0, t_1, t_2, \dots, t_s, \dots]$ , 满足  $t_0 = 0, t_s^* = t_{s+1} - t_s \geq \dim \theta$  以及一个与  $t$  无关的正常数  $c$ , 使得下列强持续激励条件成立:

$$(A3) \quad \sum_{t=t_s+1}^{t_s+1} \frac{\hat{\Phi}^T(t) \hat{\Phi}(t)}{r(t)} \geq cI, \quad s = 1, 2, 3, \dots,$$

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

鞅收敛定理是研究辨识算法的一个重要工具<sup>[7]</sup>, 鞅收敛定理相当于 Lyapunov 稳定性理论在随机系统中的应用<sup>[35]</sup>. 文献[26-27]已经用它研究了多变量系统递阶辨识方法的收敛性. 这里通过构造一个鞅过程, 进而用文献[24, 36]的方法, 利用鞅收敛定理来证明这个定理.

**证明** 定义参数估计误差向量  $\tilde{\theta}(t)$ , 新息向量  $E(t)$  和残差向量  $\eta(t)$  如下:

$$\tilde{\theta}(t) := \hat{\theta}(t) - \theta; \quad (19)$$

$$E(t) := Y(t) - \hat{\Phi}(t) \hat{\theta}(t-1); \quad (20)$$

$$\eta(t) := Y(t) - \hat{\Phi}(t) \hat{\theta}(t). \quad (21)$$

使用式(9)和(6)可得

$$\eta(t) = \left[ I - \frac{\hat{\Phi}^T(t) \hat{\Phi}(t)}{r(t)} \right] E(t) =: W(t) E(t),$$

$$W(t) := I - \frac{\hat{\Phi}^T(t) \hat{\Phi}(t)}{r(t)}. \quad (22)$$

令

$$\Delta \tilde{\theta}(t) := \frac{\hat{\Phi}^T(t)}{r(t)} E(t), \quad \tilde{Y}(t) := -\hat{\Phi}(t) \tilde{\theta}(t). \quad (23)$$

由式(9)和(19)可得

$$\tilde{\theta}(t) = \tilde{\theta}(t-1) + \frac{\hat{\Phi}^T(t)}{r(t)} E(t) = \tilde{\theta}(t-1) + \Delta \tilde{\theta}(t). \quad (24)$$

式(24)取范数, 并使用式(23)和(24)可得

$$\|\tilde{\theta}(t)\|^2 =$$

$$\|\tilde{\theta}(t-1)\|^2 - \frac{2\tilde{Y}^T(t) W^{-1}(t) [\eta(t) - V(t)]}{r(t)} +$$

$$\frac{2\tilde{\theta}^T(t) \hat{\Phi}^T(t) W^{-1}(t) V(t)}{r(t)} - \|\Delta \tilde{\theta}(t)\|^2.$$

式(24)两边转置后, 右乘以  $\hat{\Phi}^T(t) W^{-1}(t) V(t)$  可得

$$\tilde{\theta}^T(t) \hat{\Phi}^T(t) W^{-1}(t) V(t) =$$

$$\tilde{\theta}^T(t-1) \hat{\Phi}^T(t) W^{-1}(t) V(t) +$$

$$[E(t) - V(t) + V(t)]^T \frac{\hat{\Phi}(t) \hat{\Phi}^T(t)}{r(t)} W^{-1}(t) V(t) =$$

$$\tilde{\theta}^T(t-1) \hat{\Phi}^T(t) W^{-1}(t) V(t) +$$

$$[E(t) - V(t)]^T \frac{\hat{\Phi}(t) \hat{\Phi}^T(t)}{r(t)} W^{-1}(t) V(t) +$$

$$V^T(t) \frac{\hat{\Phi}(t) \hat{\Phi}^T(t)}{r(t)} W^{-1}(t) V(t). \quad (25)$$

于是有

$$\|\tilde{\theta}(t)\|^2 =$$

$$\|\tilde{\theta}(t-1)\|^2 - \frac{2\tilde{Y}^T(t) W^{-1}(t) [\eta(t) - V(t)]}{r(t)} +$$

$$\frac{2\tilde{\theta}^T(t-1) \hat{\Phi}^T(t) W^{-1}(t) V(t)}{r(t)} +$$

$$2[E(t) - V(t)]^T \frac{\hat{\Phi}(t) \hat{\Phi}^T(t)}{r^2(t)} W^{-1}(t) V(t) +$$

$$2V^T(t) \frac{\hat{\Phi}(t) \hat{\Phi}^T(t)}{r^2(t)} W^{-1}(t) V(t) - \|\Delta \tilde{\theta}(t)\|^2. \quad (26)$$

根据  $\eta(t)$  的定义, 参考文献[24]中引理的证明, 不难得到

$$A(z) [\eta(t) - V(t)] = -\hat{\Phi}(t) \tilde{\theta}(t) = \tilde{Y}(t). \quad (27)$$

因为  $A(z)$  是严格正实的, 利用文献[7]附录 C 离散正实引理可知下列不等式成立:

$$S(t) := \sum_{i=1}^t \frac{2\tilde{Y}^T(i) W^{-1}(i) [\eta(i) - V(i)]}{r(i)} \geq 0,$$

令  $Z(t) := \|\tilde{\theta}(t)\|^2 + S(t)$ . 式(26)两边加上  $S(t)$  得到

$$Z(t) = Z(t-1) + \frac{2\tilde{\theta}^T(t-1) \hat{\Phi}^T(t) W^{-1}(t) V(t)}{r(t)} +$$

$$2[E(t) - V(t)]^T \frac{\hat{\Phi}(t) \hat{\Phi}^T(t)}{r^2(t)} W^{-1}(t) V(t) +$$

$$2V^T(t) \frac{\hat{\Phi}(t) \hat{\Phi}^T(t)}{r^2(t)} W^{-1}(t) V(t) - \|\Delta \tilde{\theta}(t)\|^2.$$

上式两边对  $\mathcal{F}_{t-1}$  取条件期望, 利用不相关假设和 (A1) - (A2) 可得

$$E[Z(t) | \mathcal{F}_{t-1}] \leq$$

$$Z(t-1) + 2\text{tr} \left\{ \frac{\hat{\Phi}(t) \hat{\Phi}^T(t)}{r^2(t)} W^{-1}(t) \right\} \sigma^2 -$$

$$E[\|\Delta \tilde{\theta}(t)\|^2 | \mathcal{F}_{t-1}].$$

上式右边第 2 项从  $t=1$  到  $t=\infty$  的和是有限的<sup>[24]</sup>. 上式应用鞅收敛定理(见文献[7]中 Lemma D.5.3) 可以得出结论  $Z(t)$  收敛于于一有限随机变量  $C_1$ , 即:

$$\lim_{t \rightarrow \infty} Z(t) = \lim_{t \rightarrow \infty} \|\tilde{\theta}(t)\|^2 + S(t) = C_1 < \infty, \quad (28)$$

且

$$\sum_{t=1}^{\infty} \|\Delta \tilde{\theta}(t)\|^2 = \sum_{t=1}^{\infty} \frac{\|\hat{\Phi}^T(t) E(t)\|^2}{r^2(t)} =: C_2 < \infty. \quad (29)$$

由式(28)和  $S(t)$  的定义, 不难得到

$$S(\infty) = \sum_{t=1}^{\infty} \frac{[\boldsymbol{\eta}(t) - \mathbf{V}(t)]^T \mathbf{W}^{-1}(t) [\boldsymbol{\eta}(t) - \mathbf{V}(t)]}{r(t)} =: C_3 < \infty. \quad (30)$$

因为  $\mathbf{W}(t) \leq \mathbf{I}$ , 即  $\mathbf{W}^{-1}(t) \geq \mathbf{I}$ , 所以有

$$\sum_{t=1}^{\infty} \frac{\|\boldsymbol{\eta}(t) - \mathbf{V}(t)\|^2}{r(t)} \leq$$

$$\sum_{t=1}^{\infty} \frac{[\boldsymbol{\eta}(t) - \mathbf{V}(t)]^T \mathbf{W}^{-1}(t) [\boldsymbol{\eta}(t) - \mathbf{V}(t)]}{r(t)} =$$

$$C_3 < \infty.$$

使用 Kronecker 引理(见文献[7]中 Lemma D. 5.5) 可得

$$\lim_{t \rightarrow \infty} \frac{1}{r(t)} \sum_{i=1}^t \|\boldsymbol{\eta}(i) - \mathbf{V}(i)\|^2 = 0.$$

因为  $A(z)$  是严格稳定的, 应用文献[7]中 Lemma B.

3.3 于式(27), 并利用上式可得

$$\lim_{t \rightarrow \infty} \frac{1}{r(t)} \sum_{i=1}^t \|\bar{\mathbf{Y}}(i)\|^2 = 0, \text{ 或} \quad (31)$$

$$\sum_{i=1}^t \|\bar{\mathbf{Y}}(i)\|^2 = O(r(t)),$$

式(28)说明算法给出的参数估计误差一致有界, 这并不要求强持续激励条件(A3)成立. 下面证明估计误差  $\|\bar{\boldsymbol{\theta}}(t)\|^2$  一致收敛于零.

由式(9)和(20)可得

$$\bar{\boldsymbol{\theta}}(t+j) = \bar{\boldsymbol{\theta}}(t) + \sum_{i=1}^j \frac{\hat{\boldsymbol{\Phi}}^T(t+i)}{r(t+i)} \mathbf{E}(t+i). \quad (32)$$

由于  $\hat{\boldsymbol{\Phi}}(t+j)\bar{\boldsymbol{\theta}}(t+j) = -\bar{\mathbf{Y}}(t+j)$ , 利用式(32)可得

$$\hat{\boldsymbol{\Phi}}(t+j)\bar{\boldsymbol{\theta}}(t) = -\bar{\mathbf{Y}}(t+j) - \hat{\boldsymbol{\Phi}}(t+j) \sum_{i=1}^j \frac{\hat{\boldsymbol{\Phi}}^T(t+i)}{r(t+i)} \mathbf{E}(t+i).$$

上式两边取范数并同时除以  $r(t+j)$  得到

$$\frac{\bar{\boldsymbol{\theta}}^T(t) \hat{\boldsymbol{\Phi}}(t+j) \hat{\boldsymbol{\Phi}}(t+j) \bar{\boldsymbol{\theta}}(t)}{r(t+j)} = \frac{1}{r(t+j)} \left\| \bar{\mathbf{Y}}(t+j) + \hat{\boldsymbol{\Phi}}(t+j) \sum_{i=1}^j \frac{\hat{\boldsymbol{\Phi}}^T(t+i)}{r(t+i)} \mathbf{E}(t+i) \right\|^2 \leq \frac{2\|\bar{\mathbf{Y}}(t+j)\|^2 + 2\|\boldsymbol{\Phi}(t+j)\|^2}{r(t+j)} \left\| \sum_{i=1}^j \frac{\hat{\boldsymbol{\Phi}}^T(t+i)}{r(t+i)} \mathbf{E}(t+i) \right\|^2.$$

令  $t = t_s$ , 并对  $j$  从 1 到  $t_s^*$  求和可得

$$\bar{\boldsymbol{\theta}}(t_s) \left[ \sum_{t=t_s+1}^{t_s^*} \frac{\hat{\boldsymbol{\Phi}}^T(t) \hat{\boldsymbol{\Phi}}(t)}{r(t)} \right] \bar{\boldsymbol{\theta}}(t_s) \leq \sum_{j=1}^{t_s^*} \left[ \frac{2\|\bar{\mathbf{Y}}(t_s+i)\|^2}{r(t_s+i)} + \frac{2\|\hat{\boldsymbol{\Phi}}(t_s+j)\|^2}{r(t_s+j)} \right] \left\| \sum_{i=1}^j \frac{\hat{\boldsymbol{\Phi}}^T(t_s+i)}{r(t_s+i)} \mathbf{E}(t_s+i) \right\|^2 \leq$$

$$2 \sum_{j=1}^{t_s^*} \left[ \frac{\|\bar{\mathbf{Y}}(t_s+j)\|^2}{r(t_s+j)} + \frac{t_s^* \|\hat{\boldsymbol{\Phi}}(t_s+j)\|^2}{r(t_s+j)} \sum_{i=1}^{t_s^*} \left\| \frac{\hat{\boldsymbol{\Phi}}^T(t_s+i)}{r(t_s+i)} \mathbf{E}(t_s+i) \right\|^2 \right].$$

如果假设  $s \rightarrow \infty$  和  $t_s \rightarrow \infty$  时,

$$\frac{t_s^* \|\hat{\boldsymbol{\Phi}}(t_s+j)\|^2}{r(t_s+j)} < \infty,$$

则由式(29)和式(30)可得

$$\lim_{s \rightarrow \infty} \bar{\boldsymbol{\theta}}^T(t_s) \left[ \sum_{t=t_s+1}^{t_s^*} \frac{\hat{\boldsymbol{\Phi}}^T(t) \hat{\boldsymbol{\Phi}}(t)}{r(t)} \right] \bar{\boldsymbol{\theta}}(t_s) = 0,$$

由条件(A3)可得  $\bar{\boldsymbol{\theta}}(t_s) \rightarrow 0$ , 且根据式(30)可知

$$\bar{\boldsymbol{\theta}}(t_s+j) = \bar{\boldsymbol{\theta}}(t_s) + \sum_{i=1}^j \frac{\hat{\boldsymbol{\Phi}}^T(t_s+i)}{r(t_s+i)} \mathbf{E}(t_s+i), \quad j=1, 2, \dots, t_s^* - 1.$$

根据级数余项引理(文献[36]中 Lemma 1)以及式(29)可知, 当  $t \rightarrow \infty$  时,  $\bar{\boldsymbol{\theta}}(t) = \bar{\boldsymbol{\theta}}(t_s+j) \rightarrow 0$ , 证毕.

与最小二乘算法相比, 基于输出估计的多新息梯度算法(9)–(16)有较小的计算量, 但是它的收敛速度慢. 为了提高算法的跟踪性能, 在算法中引入一个遗忘因子  $\lambda$ , 得到基于输出估计的多新息遗忘因子随机梯度算法, 简称基于输出估计的多新息遗忘梯度算法:

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\Phi}}^T(t)}{r(t)} [Y(t) - \hat{\boldsymbol{\Phi}}(t) \hat{\boldsymbol{\theta}}(t-1)]; \quad (33)$$

$$r(t) = \lambda r(t-1) + \text{tr}[\hat{\boldsymbol{\Phi}}(t) \hat{\boldsymbol{\Phi}}^T(t)], \quad r(0) = 1, 0 \leq \lambda < 1; \quad (34)$$

$$Y(t) = [y(t), y(t-1), \dots, y(t-p+1)]^T; \quad (35)$$

$$\hat{\boldsymbol{\Phi}}(t) = [\hat{\boldsymbol{\varphi}}(t), \hat{\boldsymbol{\varphi}}(t-1), \dots, \hat{\boldsymbol{\varphi}}(t-p+1)]; \quad (36)$$

$$\hat{\boldsymbol{\varphi}}(t) = \begin{bmatrix} \hat{\boldsymbol{\psi}}(t) \\ \boldsymbol{\phi}(t) \end{bmatrix}; \quad (37)$$

$$\hat{\boldsymbol{\psi}}(t) = [-\hat{\boldsymbol{\varphi}}^T(t-1) \hat{\boldsymbol{\theta}}(t-1), -\hat{\boldsymbol{\varphi}}^T(t-2) \hat{\boldsymbol{\theta}}(t-2), \dots, -\hat{\boldsymbol{\varphi}}^T(t-n) \hat{\boldsymbol{\theta}}(t-n)]^T; \quad (38)$$

$$\boldsymbol{\phi}(t) = [u_1(t), \dots, u_1(t-n), u_2(t), \dots, u_2(t-n), \dots, u_r(t), \dots, u_r(t-n)]^T. \quad (39)$$

当  $\lambda = 1$  时, 上述算法就是基于输出估计的多新息梯度算法; 当  $\lambda = 0$ , 上述算法就是基于输出估计的多新息投影算法.

## 4 仿真试验

### Simulation test

下面通过一个例子来说明提出算法的有效性. 考虑 2 输入 1 输出系统:

$$y_0(t) = \frac{1}{A(z)} [B_1(z) u_1(t) + B_2(z) u_2(t)];$$

$$y(t) = y_0(t) + v(t);$$

$$A(z) = 1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} = 1 - 0.84z^{-1} + 0.16z^{-2};$$

$$B_1(z) = \beta_{11} z^{-1} + \beta_{12} z^{-2} = 0.12z^{-1} + 0.324z^{-2};$$

$$B_2(z) = \beta_{21} z^{-1} + \beta_{22} z^{-2} = 0.25z^{-1} + 0.186z^{-2};$$

$$\theta = [\alpha_1, \alpha_2, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}]^T.$$

仿真时,  $\{u_1(t)\}$  和  $\{u_2(t)\}$  采用零均值、单位方差、不相关持续激励信号序列,  $\{v(t)\}$  采用零均值方差为  $\sigma^2 = 0.20^2$  的白噪声序列. 应用提出的算法估计这个系统的参数, 不同遗忘因子下的参数估计和估计误差如表 1 所示, 参数估计误差  $\delta: = \|\hat{\theta}(t) - \theta\| / \|\theta\|$  随  $t$  变化曲线如图 1 所示.

从表 1 和图 1 可以看出: 当遗忘因子较小时, 算法的跟踪性能快, 但参数估计波动厉害. 因此, 一个折中方案是在算法起始阶段选择较小的遗忘因子,

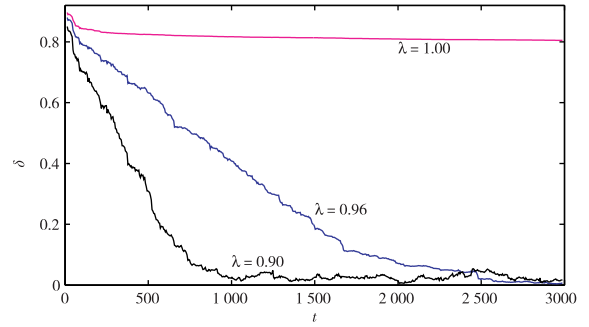


图 1 不同遗忘因子时参数估计误差随  $t$  变化曲线  
Fig. 1 The estimation errors with different  $\lambda$  versus  $t$

随后增加遗忘因子直到趋于 1, 就可取得较快的暂态跟踪性能, 又可保证一定稳态精度. 对于适当的遗忘因子, 参数估计精度是令人满意的.

表 1 参数估计及其误差 ( $\sigma^2 = 0.20^2$ )

Table 1 The parameter estimates and errors ( $\sigma^2 = 0.20^2$ )

$\lambda$	$t$	$\alpha_1$	$\alpha_2$	$\beta_{11}$	$\beta_{12}$	$\beta_{21}$	$\beta_{22}$	$\delta/\%$
1.00	100	-0.527 92	-0.307 25	0.204 76	0.290 29	0.245 21	0.406 15	84.373 70
	200	-0.517 57	-0.276 46	0.197 16	0.297 21	0.252 04	0.410 71	83.597 26
	300	-0.534 53	-0.270 09	0.210 89	0.325 34	0.260 00	0.418 98	82.812 22
	500	-0.564 99	-0.292 18	0.196 35	0.328 44	0.246 07	0.412 21	82.425 30
	1 000	-0.563 87	-0.270 28	0.178 41	0.319 11	0.251 68	0.424 76	81.683 21
	1 500	-0.571 20	-0.269 48	0.167 49	0.315 67	0.248 01	0.424 72	81.360 19
	2 000	-0.571 15	-0.258 06	0.168 98	0.322 57	0.252 73	0.427 84	80.952 46
	2 500	-0.572 61	-0.253 08	0.170 14	0.327 13	0.253 14	0.428 03	80.711 95
3 000	-0.576 23	-0.250 74	0.172 66	0.332 88	0.252 27	0.427 65	80.490 00	
0.96	100	-0.599 44	-0.216 81	0.24209	0.395 70	0.351 74	0.467 20	79.227 30
	200	-0.563 13	-0.101 81	0.183 45	0.364 23	0.338 13	0.443 05	75.893 78
	300	-0.717 87	-0.117 03	0.294 30	0.508 98	0.320 17	0.413 27	71.370 90
	500	-0.705 90	0.068 27	0.094 64	0.399 47	0.171 07	0.307 90	63.145 51
	1 000	-1.119 81	0.238 17	0.074 72	0.356 00	0.201 83	0.312 45	40.760 56
	1 500	-1.359 91	0.580 69	0.086 94	0.406 19	0.221 48	0.257 02	18.698 21
	2 000	-1.518 67	0.714 34	0.141 17	0.364 63	0.281 32	0.205 25	7.137 06
	2 500	-1.585 57	0.776 62	0.127 65	0.307 26	0.255 52	0.172 63	1.952 85
3 000	-1.603 69	0.799 83	0.119 76	0.324 97	0.251 01	0.178 60	0.453 79	
0.90	100	-0.661 69	-0.003 95	0.337 64	0.542 16	0.334 37	0.410 03	70.103 55
	200	-0.781 15	0.009 19	0.157 41	0.361 75	0.324 51	0.378 12	62.654 99
	300	-0.958 80	0.183 71	0.394 49	0.595 03	0.247 64	0.284 14	52.714 78
	500	-1.154 09	0.529 42	0.041 09	0.478 77	0.101 44	0.148 33	30.876 86
	1 000	-1.593 86	0.774 94	0.100 36	0.314 77	0.271 71	0.172 78	2.284 36
	1 500	-1.584 95	0.776 12	0.111 97	0.354 08	0.262 80	0.186 36	2.376 77
	2 000	-1.600 44	0.786 05	0.117 88	0.322 62	0.261 27	0.173 49	1.190 92
	2 500	-1.616 22	0.812 71	0.151 92	0.285 79	0.264 61	0.246 79	4.465 48
3 000	-1.592 81	0.776 87	0.125 45	0.339 71	0.255 29	0.185 07	1.616 22	
True values		-1.600 00	0.800 00	0.120 00	0.324 00	0.250 00	0.186 00	

## 5 结束

## Conclusions

针对输出误差模型描述的多输入单输出系统,提出了一个基于输出估计的多新息随机梯度辨识方法. 基于鞅收敛定理的算法收敛性分析表明:在一定条件下,提出的算法能够给出一致参数估计,仿真例子证实了提出的收敛性定理.

## 参考文献

## References

- [1] 万百五. 控制论创立六十年[J]. 控制理论与应用, 2008, 25(4): 597-602  
WAN Baiwu. Sixty years of cybernetics since founding [J]. Control Theory and Applications, 2008, 25(4): 597-602
- [2] 廖晓昕. 综合国力非线性扩散模型稳定性分析[J]. 南京信息工程大学学报:自然科学版, 2009, 1(3): 247-251  
LIAO Xiaoxin. Stability analysis of nonlinear diffusion model describing comprehensive national power [J]. Journal of Nanjing University of Information Science & Technology: Natural Science Edition, 2009, 1(3): 247-251
- [3] 丁锋. 系统辨识理论与方法[M]. 北京: 电力出版社, 2011  
DING Feng. System identification theory and methods [M]. Beijing: China Electric Power Press, 2011
- [4] 丁锋, 萧德云, 丁韬. 多新息随机梯度辨识方法[J]. 控制理论与应用, 2003, 20(6): 870-874  
DING Feng, XIAO Deyun, DING Tao. Multi-innovation stochastic gradient identification methods [J]. Control Theory and Application, 2003, 20(6): 870-874
- [5] 方崇智, 萧德云. 过程辨识[M]. 北京: 清华大学出版社, 1988  
FANG Chongzhi, XIAO Deyun. Process identification [M]. Beijing: Tsinghua University Press, 1988
- [6] 冯纯伯, 史维. 自适应控制[M]. 北京: 电子工业出版社, 1986  
FENG Chunbo, SHI Wei. Adaptive Control [M]. Beijing: Electronics Industry Press, 1986
- [7] Goodwin G C, Sin K S. Adaptive filtering, prediction and control [M]. Prentice-Hall: Englewood Cliffs, NJ, 1984
- [8] Ljung L. System identification: Theory for the user (2nd edn) [M]. Prentice-Hall: Englewood Cliffs, NJ, 1999
- [9] 丁锋. 辨识 Box-Jenkins 模型参数的递推广义增广最小二乘法[J]. 控制与决策, 1990, 5(6): 53-56  
DING Feng. Recursive generalized extended least squares (RGELS) method of identifying the Box-Jenkin model [J]. Control and Decision, 1990, 5(6): 53-56
- [10] Ding F, Xiao Y S. A finite-data-window least squares algorithm with a forgetting factor for dynamical modeling [J]. Applied Mathematics and Computation, 2007, 186(1): 184-192
- [11] 丁锋, 杨慧中, 刘飞. 弱条件下随机梯度算法性能分析[J]. 中国科学 E 辑: 信息科学, 2008, 38(12): 2173-2184
- [12] Ding F, Yang H Z, Liu F. Performance analysis of stochastic gradient algorithms under weak conditions [J]. Science in China Series F: Information Sciences, 2008, 51(9): 1269-1280
- [13] Ding F, Liu P X, Yang H Z. Parameter identification and intersample output estimation for dual-rate systems [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, 2008, 38(4): 966-975
- [14] Ding F, Chen T. Identification of dual-rate systems based on finite impulse response models [J]. International Journal of Adaptive Control and Signal Processing, 2004, 18(7): 589-598
- [15] Ding F, Chen T. Combined parameter and output estimation of dual-rate systems using an auxiliary model [J]. Automatica, 2004, 40(10): 1739-1748
- [16] Ding F, Chen T. Modeling and identification for multirate systems [J]. Acta Automatica Sinica, 2005, 31(1): 105-122
- [17] 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
- [18] 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
- [19] Ding F, Chen T. Performance analysis of multi-innovation gradient type identification methods [J]. Automatica, 2007, 43(1): 1-14
- [20] Ding F, Liu P X, Liu G. Auxiliary model based multi-innovation extended stochastic gradient parameter estimation with colored measurement noises [J]. Signal Processing, 2009, 89(10): 1883-1890
- [21] Xie L, Yang H Z, Ding F. Modeling and identification for non-uniformly periodically sampled-data systems [J]. IET Control Theory & Applications, 2010, 4(5): 784-794
- [22] Ding F. Several multi-innovation identification methods [J]. Digital Signal Processing, 2010, 20(4): 1027-1039
- [23] Ding F, Liu P X, Liu G. Multi-innovation least squares identification for linear and pseudo-linear regression models [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2010, 40(3): 767-778
- [24] 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
- [25] 丁锋, 杨家本. 大系统的递阶辨识[J]. 自动化学报, 1999, 25(5): 647-654  
DING Feng, YANG Jiaben. Hierarchical identification of large scale systems [J]. Acta Automatica Sinica, 1999, 25(5): 647-654
- [26] Ding F, Chen T. Hierarchical gradient-based identification of multivariable discrete-time systems [J]. Automatica, 2005, 41(2): 315-325
- [27] Ding F, Chen T. Hierarchical least squares identification methods for multivariable systems [J]. IEEE Transactions on Automatic Control, 2005, 50(3): 397-402
- [28] 丁锋, 萧德云. 多变量系统状态空间模型的递阶辨识[J]. 控制与决策, 2005, 20(8): 848-853, 859  
DING Feng, XIAO Deyun. Hierarchical identification of state space models for multivariable systems [J]. Control and Decision, 2005, 20(8): 848-853, 859
- [29] Ding F, Chen T. Hierarchical identification of lifted state-space models for general dual-rate systems [J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2005, 52(6): 1179-1187
- [30] Liu X G, Lu J. Least squares based iterative identification for a class of multirate systems [J]. Automatica, 2010, 46(3): 549-554
- [31] 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
- [32] Ding F, Liu P X, 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
- [33] Ding F, Liu G, Liu P X. Partially coupled stochastic gradient identification methods for non-uniformly sampled systems [J]. IEEE Transactions on Automatic Control, 2010, 55(8): 1976-1981
- [34] 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
- [35] 廖晓昕. 漫谈 Lyapunov 稳定性的理论、方法和应用[J]. 南京信息工程大学学报:自然科学版, 2009, 1(1): 1-15

LIAO Xiaoxin. Talking on the theory, methods and applications of Lyapunov stability [J]. Journal of Nanjing University of Information Science & Technology: Natural Science Edition, 2009, 1(1): 1-15

[36] Liu Y J, Sheng J, Ding R F. Convergence of stochastic gradient algorithm for multivariable ARX-like systems [J]. Computers & Mathematics with Applications, 2010, 59(8): 2615-2627

## Stochastic gradient estimation algorithm for multiple-input systems based on the output estimation

DING Feng<sup>1</sup>

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

**Abstract** For multiple-input single-output output-error systems, a difficulty is that the information vector in the identification model contains the unknown system outputs (true outputs or noise-free outputs), thus the standard identification algorithm cannot be applied directly. This paper presents a stochastic gradient identification algorithm based on the unknown output estimation. The basic idea is to replace the true output with the output estimate which is predicted/estimated by the estimated model, and also to compute the system parameter estimates by using the output estimates. Convergence of the proposed algorithm is studied and a simulation example is provided.

**Key words** system identification; recursive identification; parameter estimation; stochastic gradient; multi-innovation identification; multivariable systems