Computers and Mathematics with Applications 59 (2010) 2615–2627

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

Convergence of stochastic gradient estimation algorithm for multivariable ARX-like systems[☆]

Yanjun Liu[a,*], Jie Sheng[b], Ruifeng Ding[a]

[a] School of Communication and Control Engineering, Jiangnan University, Wuxi, 214122, PR China
[b] Institute of Technology, University of Washington, Tacoma, 98402-3100, USA

ARTICLE INFO

Article history:
Received 17 August 2009
Received in revised form 13 January 2010
Accepted 13 January 2010

Keywords:
Recursive identification
Parameter estimation
Multivariable systems
Stochastic gradient
Convergence properties

ABSTRACT

This paper studies the convergence of the stochastic gradient identification algorithm of multi-input multi-output ARX-like systems (i.e., multivariable ARX-like systems) by using the stochastic martingale theory. This ARX-like model contains a characteristic polynomial and differs from the conventional multivariable ARX system. The results indicate that the parameter estimation errors converge to zero under the persistent excitation conditions. The simulation results validate the proposed convergence theorem.

1. Introduction

Parameter estimation is the key for building models. It has been widely discussed in many areas such as system identification [1–3], signal processing [4], probability distribution [5], economic activities [6,7]. Many estimation methods have been developed for various systems, e.g., the subspace identification methods for state–space models [8,9] and the particle filter algorithm for the discrete-time Heston models [7]. This paper considers the multivariable systems described by the following discrete-time state–space model [10,11]:

$$\begin{cases} \boldsymbol{x}(t+1) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{u}(t), \\ \boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{x}(t) + \boldsymbol{D}\boldsymbol{u}(t), \end{cases} \tag{1}$$

where $\boldsymbol{x}(t) \in \mathbb{R}^n$ is the state vector, $\boldsymbol{u}(t) \in \mathbb{R}^r$ the system input vector and $\boldsymbol{y}(t) \in \mathbb{R}^m$ the system output vector, $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times r}$, $\boldsymbol{C} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{D} \in \mathbb{R}^{m\times r}$ are constant matrices.

Although the subspace identification methods can be used to estimate the parameter matrices in (1), the computational complexity increases with the increase of the sizes of the singular value decomposition (SVD) matrices and QR factorization constructed by the input data and output data [10,11]. Recently, Ding and Chen proposed a hierarchical identification algorithm to simultaneously estimate the unknown parameters and states of the lifted state–space models for general dual-rate multivariable systems [12], but the convergence of their algorithm is still an open problem.

The difference equation model or the transfer matrix representation with the input–output relationship of the multivariable systems is more useful in practice and simpler than the state–space models [13]. In this literature, Ding and

☆ This work was supported in part by the National Natural Science Foundation of China.
* Corresponding author.
E-mail addresses: yanjunliu_1983@126.com (Y. Liu), shengj2@u.washington.edu (J. Sheng), rfding@yahoo.cn (R. Ding).

0898-1221/$ – see front matter
doi:10.1016/j.camwa.2010.01.030

Chen presented a hierarchical stochastic gradient algorithm and a hierarchical least squares algorithm for the input–output representation of multivariable systems [10,11]. Han and Ding presented a multi-innovation stochastic gradient algorithm for multi-input multi-output systems [14,15] and studied the identification problems for multirate multi-input systems using the multi-innovation identification theory [16–19].

This paper presents a stochastic gradient algorithm and studies its convergence for multivariable systems with the input–output representation. The proposed algorithm has less computational burden than the least squares algorithm [10,11,20] and is easier to implement.

The structure of this paper is as follows. Section 2 derives the input–output representation and the stochastic algorithm. Section 3 studies the convergence properties by using the martingale convergence theorem. Section 4 provides two illustrative examples for the results in this paper. Finally, we offer some concluding remarks in Section 5.

2. The input–output representation and basic algorithm

Let us introduce some notations first. The symbol $\boldsymbol{I}_m$ is an $m \times m$ identity matrix; the norm of the matrix $\boldsymbol{X}$ is defined by $\|\boldsymbol{X}\|^2 := \mathrm{tr}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$; $\lambda_{\max}[\boldsymbol{X}]$ and $\lambda_{\min}[\boldsymbol{X}]$ represent the maximum and minimum eigenvalues of the square matrix $\boldsymbol{X}$, respectively; $\otimes$ denotes the Kronecker product or direct product: if $\boldsymbol{A} = [a_{ij}] \in \mathbb{R}^{m\times n}$, $\boldsymbol{B} = [b_{ij}] \in \mathbb{R}^{p\times q}$, then $\boldsymbol{A} \otimes \boldsymbol{B} = [a_{ij}\boldsymbol{B}] \in \mathbb{R}^{mp\times nq}$; $\mathrm{col}[\boldsymbol{X}]$ denotes the vector formed by the column of the matrix $\boldsymbol{X}$, that is, if $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n] \in \mathbb{R}^{m\times n}$, then $\mathrm{col}[\boldsymbol{X}] = [\boldsymbol{x}_1^{\mathrm{T}}, \boldsymbol{x}_2^{\mathrm{T}}, \ldots, \boldsymbol{x}_n^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{mn}$.

Let $\boldsymbol{I}$ represent an identity matrix of appropriate sizes. By introducing a unit backward shift operator: $z^{-1}\boldsymbol{x}(t) = \boldsymbol{x}(t-1)$, the input–output relationship of the multivariable system in (1) can be represented as

$$\begin{aligned}\boldsymbol{y}(t) &= [\boldsymbol{C}(z\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{B}+\boldsymbol{D}]\boldsymbol{u}(t)\\ &= \left(\frac{z^{-n}\boldsymbol{C}\,\mathrm{adj}[z\boldsymbol{I}-\boldsymbol{A}]\boldsymbol{B}}{z^{-n}\det[z\boldsymbol{I}-\boldsymbol{A}]}+\boldsymbol{D}\right)\boldsymbol{u}(t)\\ &=: \frac{\boldsymbol{Q}(z)}{\alpha(z)}\boldsymbol{u}(t)\end{aligned}$$

or

$$\alpha(z)\boldsymbol{y}(t) = \boldsymbol{Q}(z)\boldsymbol{u}(t), \tag{2}$$

where $\alpha(z)$ is the characteristic polynomial in z^{-1} of the system (of degree n), $\boldsymbol{Q}(z)$ is the polynomial matrix in z^{-1}, and they can be expressed as

$$\begin{aligned}\alpha(z) &:= z^{-n}\det[z\boldsymbol{I}-\boldsymbol{A}]\\ &= 1+\alpha_1 z^{-1}+\alpha_2 z^{-2}+\cdots+\alpha_n z^{-n}, \quad \alpha_i \in \mathbb{R}^1,\\ \boldsymbol{Q}(z) &:= z^{-n}\boldsymbol{C}\,\mathrm{adj}[z\boldsymbol{I}-\boldsymbol{A}]\boldsymbol{B}+z^{-n}\det[z\boldsymbol{I}-\boldsymbol{A}]\boldsymbol{D}\\ &= \boldsymbol{Q}_0+\boldsymbol{Q}_1 z^{-1}+\boldsymbol{Q}_2 z^{-2}+\cdots+\boldsymbol{Q}_n z^{-n}, \quad \boldsymbol{Q}_i \in \mathbb{R}^{m\times r}.\end{aligned}$$

Taking into account disturbances in physical systems and introducing a noise vector $\boldsymbol{v}(t) \in \mathbb{R}^m$ based on (2), we get a multivariable ARX-like model:

$$\alpha(z)\boldsymbol{y}(t) = \boldsymbol{Q}(z)\boldsymbol{u}(t)+\boldsymbol{v}(t). \tag{3}$$

This multivariable ARX-like system in (3) differs from the multivariable CAR/ARX system in [21] because $\alpha(z)$ in (3) is a polynomial rather than a polynomial matrix.

Define the parameter matrix $\boldsymbol{\theta}$, parameter vector $\boldsymbol{\alpha}$, input information vector $\boldsymbol{\varphi}(t)$ and output information matrix $\boldsymbol{\psi}(t)$ as

$$\boldsymbol{\theta}^{\mathrm{T}} = [\boldsymbol{Q}_0, \boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_n] \in \mathbb{R}^{m\times(n+1)r},$$

$$\boldsymbol{\alpha} = \begin{bmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\\ \vdots\\ \alpha_n\end{bmatrix}, \qquad \boldsymbol{\varphi}(t) = \begin{bmatrix}\boldsymbol{u}(t)\\ \boldsymbol{u}(t-1)\\ \boldsymbol{u}(t-2)\\ \vdots\\ \boldsymbol{u}(t-n)\end{bmatrix} \in \mathbb{R}^{(n+1)r},$$

$$\boldsymbol{\psi}(t) = [\boldsymbol{y}(t-1), \boldsymbol{y}(t-2), \ldots, \boldsymbol{y}(t-n)] \in \mathbb{R}^{m\times n}.$$

Then Eq. (3) can be rewritten as [10,11]

$$\boldsymbol{y}(t)+\boldsymbol{\psi}(t)\boldsymbol{\alpha} = \boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\varphi}(t)+\boldsymbol{v}(t). \tag{4}$$