Signal Processing 128 (2016) 212–221

Contents lists available at ScienceDirect

Signal Processing

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

The auxiliary model based hierarchical gradient algorithms and convergence analysis using the filtering technique☆

Yanjiao Wang, Feng Ding*

Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, PR China

ARTICLE INFO

Article history:
Received 8 February 2016
Received in revised form
24 March 2016
Accepted 30 March 2016
Available online 1 April 2016

Keywords:
Filtering
Parameter estimation
Gradient search
Convergence analysis
Hierarchical principle

ABSTRACT

On the basis of the auxiliary model identification idea, this paper studies the filtering based parameter estimation issues for a class of multivariable control systems with colored noise. An auxiliary model based hierarchical stochastic gradient (AM-HSG) algorithm is given for comparison and a data filtering AM-HSG identification algorithm is derived by using the data filtering technique. Its main key is to decompose a multivariable system into two subsystems and to coordinate the associate items between two subsystem identification algorithms. The convergence analysis indicates that the parameter estimates given by the presented algorithms converge to the true values under proper conditions by using the stochastic process theory. The simulation results show that the proposed hierarchical stochastic gradient estimation algorithms are effective.

1. Introduction

Parameter estimation of multivariable systems is important in various fields such as process control [1–3], signal processing and system modeling [4,5], and network communication [6,7]. Many identification methods of the multivariable systems have been reported for the past decades, e.g., the hierarchical identification methods [8–10], the recursive robust methods [11,12], the maximum likelihood estimation methods [13,14], and the subspace identification methods [15,16]. Most of the above-mentioned identification approaches focus on the multivariable systems with white noise disturbance. However, the case with colored noise disturbance has not been fully investigated and still needs more attention. Therefore, this paper is aimed to solve the identification issues of the multivariable systems with colored noise.

In the area of parameter estimation, the hierarchical identification principle based on the decomposition technique can deal with parameter estimation for multivariable systems [17–19]. The auxiliary model identification idea can handle the identification problems in the presence of the unmeasurable variables in the information vector [20,21]. In this aspect, Liu et al. studied multi-innovation stochastic gradient algorithms for multiple-input single-output systems using the auxiliary model [22]; Wang and Zhang gave an improved least squares identification algorithm for multivariable Hammerstein output error moving average systems [23]; Jafari et al. used the hierarchical identification principle to solve the identification problems of the multi-input multi-output nonlinear systems with colored noises and presented an iterative hierarchical least squares algorithm [24].

The filtering technique based least squares algorithm and gradient algorithm can be used to reduce the computational burdens of identification algorithms and to improve the estimation accuracy through filtering the input–output data of the systems [25–27]. In this literature, Wang and Tang used the data filtering technique to study the identification problem for a class of nonlinear systems and presented the gradient-based iterative estimation algorithms [28]; Ding et al. presented an auxiliary model based least squares algorithm for a dual-rate state space system with time-delay by means of the data filtering [29]; Li and Shi investigated the H-infinity filtering problem for a class of general nonlinear stochastic systems considering model uncertainties and random time delays governed by Markov chains simultaneously [30].

Recently, Vladimir et al. studied robust Kalman filtering for nonlinear multivariable stochastic systems in the presence of non-Gaussian noise [31] and presented a joint state and parameter robust estimation of stochastic nonlinear systems [32]. For non-Gaussian disturbance noise, this paper considers the identification problem of multivariable Box–Jenkins systems with autoregressive moving average noise and the main contributions are as follows.

- An auxiliary model based hierarchical stochastic gradient (AM-HSG) algorithm and a forgetting factor based AM-HSG

☆This work was supported by the National Natural Science Foundation of China (No. 61273194).

* Corresponding author.
E-mail addresses: yjwang12@126.com (Y. Wang), fding@jiangnan.edu.cn (F. Ding).

http://dx.doi.org/10.1016/j.sigpro.2016.03.027
0165-1684/

algorithm are presented for the multivariable Box–Jenkins systems using the auxiliary model identification idea and the hierarchical identification principle.

- A data filtering based AM-HSG algorithm is derived to improve the parameter estimation accuracy for the multivariable Box–Jenkins systems by means of the data filtering technique.
- The convergence theorems of the proposed algorithms are established by using the stochastic process theory and assuming that the disturbance is a zero-mean stochastic noise with bounded variances.

The rest of this paper is organized as follows. Section 2 introduces the problem formulation for multivariable Box–Jenkins system identification. Section 3 presents an AM-HSG algorithm for multivariable Box–Jenkins systems and Section 4 analyzes its convergence performance. Section 5 employs the data filtering technique and presents a data filtering based AM-HSG (F-AM-HSG) identification algorithm. Section 6 gives an example for testing the AM-HSG and the F-AM-HSG algorithms. Finally, the concluding remarks are given in Section 7.

2. The problem formulation

Consider a multivariable Box–Jenkins system,

$$\boldsymbol{y}(k)=\frac{\boldsymbol{\beta}(q)}{\alpha(q)}\boldsymbol{u}(k)+\frac{D(q)}{C(q)}\boldsymbol{v}(k), \tag{1}$$

where $\boldsymbol{y}(k)\in\mathbb{R}^m$ is the system output vector, $\boldsymbol{u}(k)\in\mathbb{R}^r$ is the system input vector, $\boldsymbol{v}(k)\in\mathbb{R}^m$ is a stochastic white noise vector with zero mean, $\alpha(q)$, $C(q)$ and $D(q)$ are the monic polynomials in the unit backward shift operator q^{-1} $[q^{-1}\boldsymbol{y}(k)=\boldsymbol{y}(k-1)]$, $\boldsymbol{\beta}(q)$ is a matrix polynomial in q^{-1}, and are defined by

$$\alpha(q):=1+\alpha_1q^{-1}+\alpha_2q^{-2}+\cdots+\alpha_nq^{-n},\quad \alpha_i\in\mathbb{R},$$
$$\boldsymbol{\beta}(q):=\boldsymbol{\beta}_1q^{-1}+\boldsymbol{\beta}_2q^{-2}+\cdots+\boldsymbol{\beta}_nq^{-n},\quad \boldsymbol{\beta}_i\in\mathbb{R}^{m\times r},$$
$$C(q):=1+c_1q^{-1}+c_2q^{-2}+\cdots+c_{n_c}q^{-n_c},\quad c_i\in\mathbb{R},$$
$$D(q):=1+d_1q^{-1}+d_2q^{-2}+\cdots+d_{n_d}q^{-n_d},\quad d_i\in\mathbb{R}.$$

Define two intermediate variables

$$\boldsymbol{x}(k):=\frac{\boldsymbol{\beta}(q)}{\alpha(q)}\boldsymbol{u}(k)=-\sum_{i=1}^{n}\alpha_i\boldsymbol{x}(k-i)+\sum_{i=1}^{n}\boldsymbol{\beta}_i\boldsymbol{u}(k-i)\in\mathbb{R}^m, \tag{2}$$

$$\boldsymbol{w}(k):=\frac{D(q)}{C(q)}\boldsymbol{v}(k)=-\sum_{i=1}^{n_c}c_i\boldsymbol{w}(k-i)+\sum_{i=1}^{n_d}d_i\boldsymbol{v}(k-i)+\boldsymbol{v}(k)\in\mathbb{R}^m. \tag{3}$$

Define the parameter vectors $\boldsymbol{\vartheta}_s$, $\boldsymbol{\vartheta}_n$ and ϑ, the parameter matrix θ, and the information matrices $\boldsymbol{\psi}_s(k)$, $\boldsymbol{\psi}_n(k)$, $\boldsymbol{\psi}(k)$, the information vector $\boldsymbol{\varphi}(k)$ as

$$\boldsymbol{\vartheta}_s:=[\alpha_1,\alpha_2,\ldots,\alpha_n]^{\mathrm{T}}\in\mathbb{R}^n,$$
$$\boldsymbol{\vartheta}_n:=[c_1,c_2,\ldots,c_{n_c},d_1,d_2,\ldots,d_{n_d}]^{\mathrm{T}}\in\mathbb{R}^{n_c+n_d},$$
$$\boldsymbol{\theta}^{\mathrm{T}}:=[\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\ldots,\boldsymbol{\beta}_n]\in\mathbb{R}^{m\times(nr)},$$
$$\boldsymbol{\psi}_s(k):=[-\boldsymbol{x}(k-1),-\boldsymbol{x}(k-2),\ldots,-\boldsymbol{x}(k-n)]\in\mathbb{R}^{m\times n},$$
$$\boldsymbol{\psi}_n(k):=[-\boldsymbol{w}(k-1),-\boldsymbol{w}(k-2),\ldots,-\boldsymbol{w}(k-n_c),$$
$$\boldsymbol{v}(k-1),\boldsymbol{v}(k-2),\ldots,\boldsymbol{v}(k-n_d)]\in\mathbb{R}^{m\times(n_c+n_d)},$$
$$\boldsymbol{\varphi}(k):=[\boldsymbol{u}^{\mathrm{T}}(k-1),\boldsymbol{u}^{\mathrm{T}}(k-2),\ldots,\boldsymbol{u}^{\mathrm{T}}(k-n)]^{\mathrm{T}}\in\mathbb{R}^{nr},$$
$$\boldsymbol{\vartheta}:=\begin{bmatrix}\boldsymbol{\vartheta}_s\\ \boldsymbol{\vartheta}_n\end{bmatrix}\in\mathbb{R}^{n+n_c+n_d},$$
$$\boldsymbol{\psi}(k):=[\boldsymbol{\psi}_s(k),\boldsymbol{\psi}_n(k)]\in\mathbb{R}^{m\times(n+n_c+n_d)}.$$

Using the above definitions, Eqs. (2) and (3) can be written as

$$\boldsymbol{x}(k)=\boldsymbol{\psi}_s(k)\boldsymbol{\vartheta}_s+\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\varphi}(k), \tag{4}$$

$$\boldsymbol{w}(k)=\boldsymbol{\psi}_n(k)\boldsymbol{\vartheta}_n+\boldsymbol{v}(k). \tag{5}$$

Thus, Eq. (1) can be rewritten as

$$\boldsymbol{y}(k)=\boldsymbol{\psi}_s(k)\boldsymbol{\vartheta}_s+\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\varphi}(k)+\boldsymbol{\psi}_n(k)\boldsymbol{\vartheta}_n+\boldsymbol{v}(k)=\boldsymbol{\psi}(k)\boldsymbol{\vartheta}+\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\varphi}(k)+\boldsymbol{v}(k). \tag{6}$$

Eq. (6) is the identification model of the multivariable Box–Jenkins systems. The objective of this paper is to study novel identification methods for estimating the parameter vectors/matrix in (6) by means of the filtering technique and the auxiliary model identification idea and to analyze the convergence performances of the presented algorithms.

3. The auxiliary model HSG algorithm

Let $\hat{\vartheta}(k):=\begin{bmatrix}\hat{\vartheta}_s(k)\\ \hat{\vartheta}_n(k)\end{bmatrix}\in\mathbb{R}^{n+n_c+n_d}$ be the estimate of ϑ at time k, and $\hat{\boldsymbol{\theta}}(k)$ be the estimate of $\boldsymbol{\theta}$ at time k. Define two quadratic cost functions $J_1(\boldsymbol{\vartheta}):=J_2(\boldsymbol{\theta})=\|\boldsymbol{y}(k)-\boldsymbol{\psi}(k)\boldsymbol{\vartheta}-\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\varphi}(k)\|^2$. Using the negative gradient search [33], and minimizing the cost functions $J_1(\boldsymbol{\vartheta})$ and $J_2(\boldsymbol{\theta})$ yield the following equations:

$$\hat{\boldsymbol{\vartheta}}(k)=\hat{\boldsymbol{\vartheta}}(k-1)+\frac{\boldsymbol{\psi}^{\mathrm{T}}(k)}{r_1(k)}[\boldsymbol{y}(k)-\boldsymbol{\psi}(k)\hat{\boldsymbol{\vartheta}}(k-1)-\hat{\boldsymbol{\theta}}^{\mathrm{T}}(k-1)\boldsymbol{\varphi}(k)], \tag{7}$$

$$r_1(k):=r_1(k-1)+\|\boldsymbol{\psi}(k)\|^2,\quad r_1(0)=1, \tag{8}$$

$$\hat{\boldsymbol{\theta}}(k)=\hat{\boldsymbol{\theta}}(k-1)+\frac{\boldsymbol{\varphi}(k)}{r_2(k)}[\boldsymbol{y}(k)-\boldsymbol{\psi}(k)\hat{\boldsymbol{\vartheta}}(k-1)-\hat{\boldsymbol{\theta}}^{\mathrm{T}}(k-1)\boldsymbol{\varphi}(k)]^{\mathrm{T}}, \tag{9}$$

$$r_2(k):=r_2(k-1)+\|\boldsymbol{\varphi}(k)\|^2,\quad r_2(0)=1. \tag{10}$$

Like the method in [33], $1/r_1(k)$ and $1/r_2(k)$ are called the convergence factors.

Remark 1. Notice that the information vector $\boldsymbol{\psi}(k)$ contains the unknown inner vectors $\boldsymbol{x}(k-i)$, and the unknown noise vectors $\boldsymbol{w}(k-i)$ and $\boldsymbol{v}(k-i)$, so the algorithm in (7)–(10) is impossible to implement. The solution here is based on the auxiliary model identification idea and to construct an auxiliary model: the unknown $\boldsymbol{x}(k-i)$ is replaced with the output $\hat{\boldsymbol{x}}(k-i)$ of the auxiliary model, the unknown noise terms $\boldsymbol{w}(k-i)$ and $\boldsymbol{v}(k-i)$ are replaced with their estimates $\hat{\boldsymbol{w}}(k-i)$ and $\hat{\boldsymbol{v}}(k-i)$, respectively (see Fig. 1).

Define the estimates

$$\hat{\boldsymbol{\psi}}(k):=[\hat{\boldsymbol{\psi}}_s(k),\hat{\boldsymbol{\psi}}_n(k)]\in\mathbb{R}^{m\times(n+n_c+n_d)}, \tag{11}$$

$$\hat{\boldsymbol{\psi}}_s(k):=[-\hat{\boldsymbol{x}}(k-1),-\hat{\boldsymbol{x}}(k-2),\ldots,-\hat{\boldsymbol{x}}(k-n)]\in\mathbb{R}^{m\times n}, \tag{12}$$

$$\hat{\boldsymbol{\psi}}_n(k):=[-\hat{\boldsymbol{w}}(k-1),-\hat{\boldsymbol{w}}(k-2),\ldots,-\hat{\boldsymbol{w}}(k-n_c),\hat{\boldsymbol{v}}(k-1),$$
$$\hat{\boldsymbol{v}}(k-2),\ldots,\hat{\boldsymbol{v}}(k-n_d)]\in\mathbb{R}^{m\times(n_c+n_d)}. \tag{13}$$

From (4)–(6), the estimates of $\boldsymbol{x}(k)$, $\boldsymbol{w}(k)$ and $\boldsymbol{v}(k)$ can be computed through