Computers and Mathematics with Applications 60 (2010) 1200–1208

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

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

Gradient-based iterative parameter estimation for Box–Jenkins systems☆

Dongqing Wang [a], Guowei Yang [a], Ruifeng Ding [b,*]

[a] College of Automation Engineering, Qingdao University, Qingdao 266071, PR China

[b] School of Communication and Control Engineering, Jiangnan University, Wuxi 214122, PR China

ARTICLE INFO

Article history:
Received 20 March 2009
Received in revised form 31 May 2010
Accepted 1 June 2010

Keywords:
Signal processing
Recursive dentification
Parameter estimation
Stochastic gradient
Iterative algorithms
Box–Jenkins models

ABSTRACT

This paper presents a gradient-based iterative identification algorithms for Box–Jenkins systems with finite measurement input/output data. Compared with the pseudo-linear regression stochastic gradient approach, the proposed algorithm updates the parameter estimation using all the available data at each iterative computation (at each iteration), and thus can produce highly accurate parameter estimation. An example is given.

1. Introduction

The least-squares and stochastic gradient parameter estimation methods are two classes of basic identification algorithms. They have received much attention in many areas, e.g., signal processing, adaptive control and system identification [1–9]. These two methods are used in studying different types of systems, e.g., multivariable systems [2,10–14], dual-rate and multirate sampled-data systems [14–18], nonlinear block-oriented systems [19–22], and the performances of these two classes of identification methods for different systems were analyzed in [18,23–26].

The recursive prediction error least-squares method can identify the parameters of Box–Jenkins systems [27], but the stochastic gradient (SG) identification algorithm has low computational load and slow convergence rates [28]. Recently, Liu, Wang and Ding presented a least-square-based iterative identification algorithm for Box–Jenkins models [29]. On the basis of their work in [29], the objective of this paper is to develop new identification algorithms using the iterative techniques and to present a gradient-based iterative identification algorithm for Box–Jenkins systems to improve the parameter estimation accuracy.

The paper is organized as follows. Section 2 simply introduces the prediction error stochastic gradient algorithm for Box–Jenkins models and Section 3 derives a gradient-based iterative identification algorithm for Box–Jenkins systems. Section 4 gives an illustrative example. Finally, concluding remarks are given in Section 5.

2. The stochastic gradient algorithms

Consider the following Box–Jenkins systems in [29],

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

☆ This work was supported by the National Natural Science Foundation of China (60973048).

* Corresponding author.
E-mail addresses: dqwang64@163.com (D. Wang), ygw_ustb@163.com (G. Yang), rfding@yahoo.cn (R. Ding).

0898-1221/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2010.06.001

where $\{u(t)\}$ and $\{y(t)\}$ are the input and output sequences, respectively, $\{v(t)\}$ is a white noise sequence with zero mean, and $A(z)$, $B(z)$, $C(z)$ and $D(z)$ are the polynomials, of known orders (n_a, n_b, n_c, n_d), in the unit backward shift operator z^{-1} [i.e., $z^{-1}y(t) = y(t-1)$], defined by

$$\begin{aligned}
A(z) &= 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_{n_a} z^{-n_a},\\
B(z) &= b_1 z^{-1} + b_2 z^{-2} + \cdots + b_{n_b} z^{-n_b},\\
C(z) &= 1 + c_1 z^{-1} + c_2 z^{-2} + \cdots + c_{n_c} z^{-n_c},\\
D(z) &= 1 + d_1 z^{-1} + d_2 z^{-2} + \cdots + d_{n_d} z^{-n_d}.
\end{aligned}$$

Without loss of generality, assume that $u(t) = 0$, $y(t) = 0$ and $v(t) = 0$ as $t \leqslant 0$, and $n := n_a + n_b + n_c + n_d$.

Like in [29], define two intermediate variables,

$$x(t) := \frac{B(z)}{A(z)} u(t), \tag{2}$$

$$w(t) := \frac{D(z)}{C(z)} v(t). \tag{3}$$

Define the parameter vectors,

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

and the information vectors,

$$\begin{aligned}
\boldsymbol{\varphi}(t) &:= \begin{bmatrix} \boldsymbol{\varphi}_{\mathrm{s}}(t) \\ \boldsymbol{\varphi}_{\mathrm{n}}(t) \end{bmatrix} \in \mathbb{R}^{n_a+n_b+n_c+n_d},\\
\boldsymbol{\varphi}_{\mathrm{s}}(t) &:= [-x(t-1), -x(t-2), \ldots, -x(t-n_a), u(t-1), u(t-2), \ldots, u(t-n_b)]^{\mathrm{T}} \in \mathbb{R}^{n_a+n_b},\\
\boldsymbol{\varphi}_{\mathrm{n}}(t) &:= [-w(t-1), -w(t-2), \ldots, -w(t-n_c), v(t-1), v(t-2), \ldots, v(t-n_d)]^{\mathrm{T}} \in \mathbb{R}^{n_c+n_d},
\end{aligned}$$

where subscripts Roman s and n denote the first letters of the words 'system' and 'noise', respectively. Eqs. (1)–(3) can be written as

$$x(t) = \boldsymbol{\varphi}_{\mathrm{s}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{s}}, \tag{4}$$

$$w(t) = \boldsymbol{\varphi}_{\mathrm{n}}^{\mathrm{T}}(t)\boldsymbol{\theta}_{\mathrm{n}} + v(t), \tag{5}$$

$$y(t) = x(t) + w(t) \tag{6}$$

$$= \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta} + v(t). \tag{7}$$

In order to show the advantages of the iterative identification methods proposed in the next section, the following is simply to discuss the comparable pseudo-linear regression or prediction error identification approaches [27].

Since $x(t-i)$, $w(t-i)$ and $v(t-i)$ in the information vector $\boldsymbol{\varphi}(t)$ are unknown, so the stochastic gradient algorithm [28]:

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\boldsymbol{\varphi}(t)}{r(t)}[y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)], \tag{8}$$

$$r(t) = r(t-1) + \|\boldsymbol{\varphi}(t)\|^2, \qquad r(0) = 1 \tag{9}$$

cannot generate the estimate $\hat{\boldsymbol{\theta}}(t)$ of the parameter vector $\boldsymbol{\theta}$ in (7). The solution is to use the prediction error method or so-called Bootstrap method [27]: replacing the unknown variables $x(t-i)$, $w(t-i)$ and $v(t-i)$ in $\boldsymbol{\varphi}(t)$ with their estimates $\hat{x}(t-i)$, $\hat{w}(t-i)$ and $\hat{v}(t-i)$, respectively, and $\boldsymbol{\varphi}(t)$ in (8)–(9) with $\hat{\boldsymbol{\varphi}}(t)$ leads to the following generalized extended stochastic gradient algorithm for the Box–Jenkins systems (the BJ–GESG algorithm for short):

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\varphi}}(t)}{r(t)}[y(t) - \hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)], \tag{10}$$

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

$$\hat{\boldsymbol{\varphi}}(t) = \begin{bmatrix} \hat{\boldsymbol{\varphi}}_{\mathrm{s}}(t) \\ \hat{\boldsymbol{\varphi}}_{\mathrm{n}}(t) \end{bmatrix}, \qquad \hat{\boldsymbol{\theta}}(t) = \begin{bmatrix} \hat{\boldsymbol{\theta}}_{\mathrm{s}}(t) \\ \hat{\boldsymbol{\theta}}_{\mathrm{n}}(t) \end{bmatrix}, \tag{12}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{s}}(t) = [-\hat{x}(t-1), -\hat{x}(t-2), \ldots, -\hat{x}(t-n_a), u(t-1), u(t-2), \ldots, u(t-n_b)]^{\mathrm{T}}, \tag{13}$$