International Journal of Control, Automation, and Systems 20(8) (2022) 2606-2615
http://dx.doi.org/10.1007/s12555-021-0367-7
ISSN:1598-6446 eISSN:2005-4092
http://www.springer.com/12555

Maximum Likelihood Recursive Generalized Extended Least Squares Estimation Methods for a Bilinear-parameter Systems with ARMA Noise Based on the Over-parameterization Model

Haibo Liu*, Junwei Wang, and Yan Ji

Abstract: Maximum likelihood methods have wide applications in system modeling and parameter estimation. For the purpose of improving the precision of parameter estimation, this paper presents a maximum likelihood recursive generalized extended least squares (ML-RLS) algorithm for a bilinear-parameter system with autoregressive moving average noise based on the over-parameterization identification model. An over-parameterization-based recursive generalized extended least squares algorithm is presented to show the effectiveness of the proposed ML-RLS algorithm for comparison. The simulation test shows that the proposed algorithm has a higher estimation accuracy than the recursive least squares algorithm.

Keywords: Bilinear-parameter system, least squares, maximum likelihood, parameter estimation, recursive identification.

1. INTRODUCTION

System identification constitutes a crucial part in controller designs if the model parameters of a system are unknown [1–3]. Parameter estimation is the foundation of dynamic systems modeling [4–8], and has been widely used in the fields of science and engineering. For decades, the parameter estimation methods have received much attention [9–11]. Some parameter estimation methods such as the auxiliary model identification idea, the multi-innovation identification theory and the data filtering technique have been developed to identify linear and nonlinear systems [12–14]. Block-oriented nonlinear models consisting of linear dynamic blocks and static nonlinear parts are generally used for describing all sorts of nonlinear processes [15,16]. When the output of nonlinear block can be represented a linear combination of the unknown parameters and known basis functions, such a nonlinear system can be transformed into a bilinear-parameter system. These systems involve two product terms of the parameter vectors, and the outputs are linear with respect to any one of two parameter vectors [17]. For example, Ji *et al*. considered the parameter estimation problems of block-oriented nonlinear systems using the hierarchial identification principle and the multi-innovation identification theory for improving the parameter estimation accuracy [18].

The maximum likelihood principle is widely used in the field of stochastic system identification such as linear systems, bilinear systems [19–22], nonlinear systems [23–25] and multivariable systems [26–31], because it has good statistical properties. The maximum likelihood identification method needs to construct a likelihood function with respect to the observed data and the unknown parameters, and the parameter estimation can be obtained by maximizing the likelihood function [32]. The objective of maximum likelihood principle is to construct a likelihood function with respect to the observed data and the unmeasurable parameters, and to obtain the parameter estimation by maximizing the likelihood function. The main contributions of this paper are as follows:

- The identification model of the bilinear-parameter systems is obtained based on the over-parameterization method.
- Based on the over-parameterization model, a maximum likelihood recursive least squares (ML-RLS) algorithm is derived for identifying the parameters. Moreover, am over-parameterization-based recursive least squares (RLS) algorithm is provided as a comparison.
- The simulation example showed the effectiveness of the ML-RLS algorithm, which can give more accurate parameter estimates than the RLS algorithm.

Manuscript received June 11, 2021; revised September 14, 2021; accepted October 27, 2021. Recommended by Associate Editor Aldo Jonathan Munoz-Vazquez under the direction of Editor Jay H. Lee. This work was supported by the National Natural Science Foundation of China (Nos. 61472195, 61773356) and the Natural Science Foundation of Shandong Province (ZR2020MF160).

Haibo Liu, Junwei Wang, and Yan Ji are with the College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266061, China (e-mails: liuhaibo@qust.edu.cn, Junwei1666@126.com, yjichina@163.com).
* Corresponding author.

Briefly, the rest of this paper is organized as follows: Section 2 introduces the identification model of a bilinear-parameter system with autoregressive moving average noise based on the over-parameterization method. A ML-RLS algorithm is introduced in Section 3. In Section 4, a RLS algorithm is presented. Simulation examples are provided to verify the effectiveness in Section 5. Finally, some concluding remarks are given in Section 6.

2. SYSTEM DESCRIPTION AND IDENTIFICATION MODEL

Let us introduce some symbols used in this paper. A=:X stands for X is defined as A, or X:=A stands for A is defined as X, the symbol $\boldsymbol{I}_n$ denotes an identity matrix of appropriate size ($n\times n$); $\mathbf{1}_n$ stands for an n-dimensional column vector whose elements are 1, $\mathbf{1}_{m\times n}$ represents a matrix of size ($m\times n$) whose elements are 1; the superscript T stands for the vector/matrix transpose; the norm of a matrix $\boldsymbol{X}$ is defined by $\|\boldsymbol{X}\|^2 := \mathrm{tr}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$; $\mathrm{sgn}(x)$ denotes the sign of x.

Consider a bilinear-parameter system with autoregressive moving average noise [17,33],

$$\begin{aligned} y(t) &= \boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\beta} + \frac{D(z)}{C(z)}v(t), \\ D(z) &:= 1 + d_1 z^{-1} + d_2 z^{-2} + \cdots + d_{n_d} z^{-n_d}, \\ C(z) &:= 1 + c_1 z^{-1} + c_2 z^{-2} + \cdots + c_{n_c} z^{-n_c}, \end{aligned} \tag{1}$$

where $y(t)$ is the output, and $v(t)$ is a Gaussian distributed white noise with zero mean and variance σ^2, $\boldsymbol{F}(t) := \boldsymbol{F}(u(t)) \in \mathbb{R}^{m\times n}$ consists of available input $u(t-j)$, $j=1$, 2, $\cdots$, n, $\boldsymbol{a}\in\mathbb{R}^m$ and $\boldsymbol{\beta}\in\mathbb{R}^n$ are the parameter vectors to be identified, z^{-1} is a unit backward shift operator, i.e., z^{-1} $[z^{-1}y(t) = y(t-1)]$. Let $\boldsymbol{a}$ and $\boldsymbol{\beta}$ and the information matrix $\boldsymbol{F}(t)$ as

$$\begin{aligned} \boldsymbol{a} &:= [a_1,\ a_2,\ \ldots,\ a_m]^{\mathrm{T}} \in \mathbb{R}^m, \\ \boldsymbol{\beta} &:= [\beta_1,\ \beta_2,\ \ldots,\ \beta_n]^{\mathrm{T}} \in \mathbb{R}^n, \\ \boldsymbol{F}(t) &:= \begin{bmatrix} f_1(u(t-1)) & \cdots & f_1(u(t-n)) \\ f_2(u(t-1)) & \cdots & f_2(u(t-n)) \\ \vdots & & \vdots \\ f_m(u(t-1)) & \cdots & f_m(u(t-n)) \end{bmatrix}, \end{aligned}$$

where $u(t)$ is the input of the system, $\boldsymbol{f}_i(u(t-i))$'s are the known nonlinear basis functions. Here, introduce the intermediate variable

$$w(t) := \frac{D(z)}{C(z)}v(t), \tag{2}$$

and define the parameter vector $\boldsymbol{\rho}$ and the noise vector $\boldsymbol{\psi}(t)$ as

$$\begin{aligned} \boldsymbol{\rho} &:= [c_1,\ c_2,\ \ldots,\ c_{n_c},\ d_1,\ d_2,\ \ldots,\ d_{n_d}]^{\mathrm{T}} \in \mathbb{R}^{n_c+n_d}, \\ \boldsymbol{\psi}(t) &:= [-w(t-1),\ -w(t-2),\ \ldots,\ -w(t-n_c), \\ &\qquad v(t-1),\ v(t-2),\ \ldots,\ v(t-n_d)]^{\mathrm{T}} \in \mathbb{R}^{n_c+n_d}, \end{aligned}$$

where $w(t)$ is an autoregressive moving average process. Then, (2) can be rewritten as

$$\begin{aligned} w(t) &= [1-C(z)]w(t) + [D(z)-1]v(t) + v(t) \\ &= \boldsymbol{\psi}^{\mathrm{T}}(t)\boldsymbol{\rho} + v(t). \end{aligned} \tag{3}$$

Let $\boldsymbol{F}_i(t)$ be the ith row of the information matrix $\boldsymbol{F}(t)$. Define the parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ and the information vectors $\boldsymbol{\phi}(t)$ and $\boldsymbol{\Phi}(t)$ as

$$\begin{aligned} \boldsymbol{\theta} &:= [a_1\boldsymbol{\beta}^{\mathrm{T}},\ a_2\boldsymbol{\beta}^{\mathrm{T}},\ \ldots,\ a_m\boldsymbol{\beta}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{mn}, \\ \boldsymbol{\phi}(t) &:= [\boldsymbol{F}_1(t),\ \boldsymbol{F}_2(t),\ \ldots,\ \boldsymbol{F}_m(t)]^{\mathrm{T}} \in \mathbb{R}^{mn}, \\ \boldsymbol{\vartheta} &:= [\boldsymbol{\theta}^{\mathrm{T}},\ \boldsymbol{\rho}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{mn+n_c+n_d}, \\ \boldsymbol{\Phi}(t) &:= [\boldsymbol{\phi}^{\mathrm{T}}(t),\ \boldsymbol{\psi}^{\mathrm{T}}(t)]^{\mathrm{T}} \in \mathbb{R}^{mn+n_c+n_d}. \end{aligned}$$

Then referring to the method in [17], equation (1) can be written as

$$\begin{aligned} y(t) &= \boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta} + w(t) \\ &= \boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta} + \boldsymbol{\psi}^{\mathrm{T}}(t)\boldsymbol{\rho} + v(t) \\ &= \boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{\vartheta} + v(t). \end{aligned} \tag{4}$$

For identifiability, we adopt the normalization constraint on $\boldsymbol{a}$ or $\boldsymbol{\beta}$. This paper uses the assumption: $\|\boldsymbol{\beta}\| = 1$, and the first nonzero element of the parameter vector $\boldsymbol{\beta}$ is positive, i.e., $\beta_1 > 0$.

3. THE MAXIMUM LIKELIHOOD RECURSIVE RECURSIVE GENERALIZED LEAST SQUARES ALGORITHM

According to the maximum likelihood principle, define

$$J(\boldsymbol{\vartheta}) := \frac{1}{2}\sum_{j=1}^{t} v^2(j), \tag{5}$$

$$v(t) = y(t) - \boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{\vartheta}. \tag{6}$$

The objective function can be written as a recursive form

$$J(\boldsymbol{\vartheta},t) := J(\boldsymbol{\vartheta},t-1) + \frac{1}{2}v^2(t). \tag{7}$$

Using the first-order Taylor expansion, $v(t)$ at $\boldsymbol{\vartheta} = \hat{\boldsymbol{\vartheta}}(t-1)$ can be approximately expressed as

$$v(t) \approx v(t)|_{\boldsymbol{\vartheta}(t-1)} + \left[\frac{\partial v(t)}{\partial \boldsymbol{\vartheta}}\right]^{\mathrm{T}}_{\hat{\boldsymbol{\vartheta}}(t-1)} [\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}(t-1)].$$

Define the filtered information vectors

$$\begin{aligned} \hat{\boldsymbol{\Phi}}_f(t) &:= -\frac{\partial v(t)}{\partial \boldsymbol{\vartheta}}\bigg|_{\hat{\boldsymbol{\vartheta}}(t-1)} \\ &= \begin{bmatrix} \hat{\boldsymbol{\phi}}_f(t) \\ \hat{\boldsymbol{\psi}}_f(t) \end{bmatrix} \in \mathbb{R}^{mn+n_c+n_d}, \end{aligned}$$

$$\hat{\boldsymbol{\phi}}_f(t) := -\frac{\partial v(t)}{\partial \boldsymbol{\theta}}\bigg|_{\hat{\boldsymbol{\theta}}(t-1)} = -\begin{bmatrix} \dfrac{\partial v(t)}{a_1\boldsymbol{\beta}} \\ \dfrac{\partial v(t)}{a_2\boldsymbol{\beta}} \\ \vdots \\ \dfrac{\partial v(t)}{a_m\boldsymbol{\beta}} \end{bmatrix}_{\hat{\boldsymbol{\theta}}(t-1)}$$
$$= [\hat{f}_{1,f}(u(t-1)),\ \ldots,\ \hat{f}_{1,f}(u(t-n)),\ \ldots,\ \hat{f}_{m,f}(u(t-1)),\ \ldots,\ \hat{f}_{m,f}(u(t-n))]^{\mathrm{T}},$$
$$\hat{\boldsymbol{\psi}}_f(t) := -\frac{\partial v(t)}{\partial \boldsymbol{\rho}}\bigg|_{\hat{\boldsymbol{\rho}}(t-1)} = [-\hat{w}_f(t-1),\ -\hat{w}_f(t-2),\ \ldots,\ -\hat{w}_f(t-n_c),\ \hat{v}_f(t-1),\ \hat{v}_f(t-2),\ \ldots,\ \hat{v}_f(t-n_d)]^{\mathrm{T}},$$

where $\hat{f}_{i,f}(u(t))$, $\hat{w}_f(t)$ and $\hat{v}_f(t)$ are the filtered values of $f_i(u(t))$, $w(t)$ and $v(t)$, respectively, and defined as

$$\begin{aligned}\hat{f}_{i,f}(u(t)) &:= f_{i,f}(u(t)) + \hat{c}_1(t-1)f_{i,f}(u(t-1)) + \cdots + \hat{c}_{n_c}(t-1)f_{i,f}(u(t-n_c)) \\ &\quad - \hat{d}_1(t-1)\hat{f}_{i,f}(u(t-1)) - \cdots - \hat{d}_{n_d}(t-1)\hat{f}_{i,f}(u(t-n_d)),\\ \hat{w}_f(t) &:= \hat{w}(t) - \hat{d}_1(t-1)\hat{w}_f(t-1) - \cdots - \hat{d}_{n_d}(t-1)\hat{w}_f(t-n_d)),\\ \hat{v}_f(t) &:= \hat{v}(t) - \hat{d}_1(t-1)\hat{v}_f(t-1) - \cdots - \hat{d}_{n_d}(t-1)\hat{v}_f(t-n_d)).\end{aligned}$$

Define the estimated information vector

$$\hat{\boldsymbol{\Phi}}(t) = [\boldsymbol{\phi}^{\mathrm{T}}(t),\ -\hat{w}(t-1),\ -\hat{w}(t-2),\ \ldots,\ -\hat{w}(t-n_c),\ \hat{v}(t-1),\ \hat{v}(t-2),\ \ldots,\ \hat{v}(t-n_d)]^{\mathrm{T}}.$$

The the estimate $\hat{w}(t)$ and $\hat{v}(t)$ can be computed by

$$\hat{w}(t) = y(t) - \boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1),$$
$$\hat{v}(t) = y(t) - \hat{\boldsymbol{\Phi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1).$$

Applying the Taylor series expansion to $J(\boldsymbol{\vartheta},t-1)$ gives

$$\begin{aligned}J(\boldsymbol{\vartheta},t-1) \approx& \frac{\partial J(\hat{\boldsymbol{\vartheta}}(t-1),t-1)}{\partial \boldsymbol{\vartheta}}[\boldsymbol{\vartheta}-\hat{\boldsymbol{\vartheta}}(t-1)] \\ &+ \frac{1}{2}[\boldsymbol{\vartheta}-\hat{\boldsymbol{\vartheta}}(t-1)]^{\mathrm{T}}\frac{\partial^2 J(\hat{\boldsymbol{\vartheta}}(t-1),t-1)}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}^{\mathrm{T}}} \\ &\times [\boldsymbol{\vartheta}-\hat{\boldsymbol{\vartheta}}(t-1)] + \frac{1}{2}\eta(t),\end{aligned}$$

where the variable $\eta(t)$ is the residual of the Taylor expansion of $J(\boldsymbol{\vartheta},t-1)$. Since the first-order derivative of $J(\boldsymbol{\vartheta},t-1)$ at $\boldsymbol{\vartheta} = \hat{\boldsymbol{\vartheta}}(t-1)$ approximately equals zero, and

$$\boldsymbol{P}_f^{-1}(t) := \frac{\partial^2 J(\boldsymbol{\vartheta},t-1)}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}^{\mathrm{T}}}\bigg|_{\hat{\boldsymbol{\vartheta}}(t-1)}$$

is a positive-definite matrix, equation (7) can be written as

$$J(\boldsymbol{\vartheta},t) \approx \frac{1}{2}[\boldsymbol{\vartheta}-\hat{\boldsymbol{\vartheta}}(t-1)]^{\mathrm{T}}\boldsymbol{P}_f^{-1}(t)[\boldsymbol{\vartheta}-\hat{\boldsymbol{\vartheta}}(t-1)] + \frac{1}{2}\eta(t) + \frac{1}{2}v^2(t). \tag{8}$$

From (8) and (7), we have

$$\begin{aligned}2J(\boldsymbol{\vartheta},t) \approx& [\boldsymbol{\vartheta}-\hat{\boldsymbol{\vartheta}}(t-1)]^{\mathrm{T}}\boldsymbol{P}_f^{-1}(t)[\boldsymbol{\vartheta}-\hat{\boldsymbol{\vartheta}}(t-1)] + \eta(t) + v^2(t)\\ =& [\boldsymbol{\vartheta}-\hat{\boldsymbol{\vartheta}}(t-1)]^{\mathrm{T}}\boldsymbol{P}_f^{-1}(t)[\boldsymbol{\vartheta}-\hat{\boldsymbol{\vartheta}}(t-1)] \\ &+ \eta(t)\left[v(t)|_{\hat{\boldsymbol{\vartheta}}(t-1)} + \left[\frac{\partial v(t)}{\partial \boldsymbol{\vartheta}}\right]^{\mathrm{T}}_{\hat{\boldsymbol{\vartheta}}(t-1)} \times [\boldsymbol{\vartheta}-\hat{\boldsymbol{\vartheta}}(t-1)]\right]^2\\ =& [\boldsymbol{\vartheta}-\hat{\boldsymbol{\vartheta}}(t-1)]^{\mathrm{T}}[\boldsymbol{P}_f^{-1}(t) + \hat{\boldsymbol{\Phi}}_f(t)\hat{\boldsymbol{\Phi}}_f^{\mathrm{T}}(t)] \\ &\times [\boldsymbol{\vartheta}-\hat{\boldsymbol{\vartheta}}(t-1)] - 2\hat{v}(t)\hat{\boldsymbol{\Phi}}_f^{\mathrm{T}}(t) \\ &\times [\boldsymbol{\vartheta}-\hat{\boldsymbol{\vartheta}}(t-1)] + v^2(t) + \eta(t).\end{aligned}$$

The second-order derivative of $J(\boldsymbol{\vartheta},t)$ in (7) with respect to $\boldsymbol{\vartheta}$ at $\boldsymbol{\vartheta} = \hat{\boldsymbol{\vartheta}}(t)$ is

$$\boldsymbol{P}_f^{-1}(t) = \frac{\partial^2 J(\boldsymbol{\vartheta},t)}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}^{\mathrm{T}}} = \frac{\partial^2 J(\boldsymbol{\vartheta},t-1)}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}^{\mathrm{T}}} + \hat{v}(t)\frac{\partial^2 \hat{v}(t)}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}^{\mathrm{T}}} + \frac{\partial \hat{v}(t)}{\partial \boldsymbol{\vartheta}}\left[\frac{\partial \hat{v}(t)}{\partial \boldsymbol{\vartheta}}\right]^{\mathrm{T}}. \tag{9}$$

Since the second-order derivative of $v(t)$ with respect to $\boldsymbol{\vartheta}$ at $\boldsymbol{\vartheta} = \hat{\boldsymbol{\vartheta}}(t-1)$ is zero, equation (9) can be written as

$$\boldsymbol{P}_f^{-1}(t) = \boldsymbol{P}_f^{-1}(t-1) + \hat{\boldsymbol{\Phi}}_f(t)\hat{\boldsymbol{\Phi}}_f^{\mathrm{T}}(t). \tag{10}$$

Furthermore, applying the matrix inversion

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

to (10) gives

$$\boldsymbol{P}_f(t) = \boldsymbol{P}_f(t-1) - \frac{\boldsymbol{P}_f(t-1)\hat{\boldsymbol{\Phi}}_f(t)\hat{\boldsymbol{\Phi}}_f^{\mathrm{T}}(t)\boldsymbol{P}_f(t-1)}{1+\hat{\boldsymbol{\Phi}}_f^{\mathrm{T}}(t)\boldsymbol{P}_f(t-1)\hat{\boldsymbol{\Phi}}_f(t)}. \tag{11}$$

Define the filtered gain vector $\boldsymbol{L}_f(t) := \boldsymbol{P}_f(t)\hat{\boldsymbol{\Phi}}_f(t)$. Then

$$\begin{aligned}\boldsymbol{L}_f(t) =& \boldsymbol{P}_f(t-1)\hat{\boldsymbol{\Phi}}_f(t) - \frac{\boldsymbol{P}_f(t-1)\hat{\boldsymbol{\Phi}}_f(t)\hat{\boldsymbol{\Phi}}_f^{\mathrm{T}}(t)\boldsymbol{P}_f(t-1)\hat{\boldsymbol{\Phi}}_f(t)}{1+\hat{\boldsymbol{\Phi}}_f^{\mathrm{T}}(t)\boldsymbol{P}_f(t-1)\hat{\boldsymbol{\Phi}}_f(t)}\\ =& \frac{\boldsymbol{P}_f(t-1)\hat{\boldsymbol{\Phi}}_f(t)}{1+\hat{\boldsymbol{\Phi}}_f^{\mathrm{T}}(t)\boldsymbol{P}_f(t-1)\hat{\boldsymbol{\Phi}}_f(t)}.\end{aligned}$$

Bring $\boldsymbol{L}_f(t)$ into (11), we have

$$\boldsymbol{P}_f(t) = \boldsymbol{P}_f(t-1) - \boldsymbol{L}(t)\hat{\boldsymbol{\Phi}}_f^{\mathrm{T}}(t)\boldsymbol{P}_f(t-1)$$

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

The objective function $J(\boldsymbol{\vartheta}, t)$ can be written as

$$\begin{aligned}2J(\boldsymbol{\vartheta}, t) \approx & [\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}(t-1) - \boldsymbol{L}(t)\hat{v}(t)]^{\mathrm{T}}\boldsymbol{P}_f^{-1}(t) \\ & \times [\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}(t-1) - \boldsymbol{L}(t)\hat{v}(t)] \\ & + [\boldsymbol{L}(t)\hat{v}(t)]^{\mathrm{T}}\boldsymbol{P}_f^{-1}(t)[\boldsymbol{L}(t)\hat{v}(t)] + \hat{v}^2(t) \\ & + \eta(t).\end{aligned}$$

Minimizing $2J(\boldsymbol{\vartheta}, t)$ gives the estimate $\hat{\boldsymbol{\vartheta}}(t)$ of $\boldsymbol{\vartheta}$ as

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \boldsymbol{L}(t)\hat{v}(t).$$

Then, we can summarize the maximum likelihood recursive generalized extended least squares (ML-RLS) identification algorithm based on the over-parameterization model as

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \boldsymbol{L}_f(t)\hat{v}(t), \tag{12}$$

$$\boldsymbol{L}_f(t) = \frac{\boldsymbol{P}_f(t-1)\hat{\boldsymbol{\Phi}}_f(t)}{1 + \hat{\boldsymbol{\Phi}}_f^{\mathrm{T}}(t)\boldsymbol{P}_f(t-1)\hat{\boldsymbol{\Phi}}_f(t)}, \tag{13}$$

$$\boldsymbol{P}_f(t) = [\boldsymbol{I} - \boldsymbol{L}_f(t)\hat{\boldsymbol{\Phi}}_f^{\mathrm{T}}(t)]\boldsymbol{P}_f(t-1), \tag{14}$$

$$\hat{v}(t) = y(t) - \hat{\boldsymbol{\Phi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1), \tag{15}$$

$$\hat{\boldsymbol{\Phi}}_f(t) = \begin{bmatrix}\hat{\boldsymbol{\phi}}_f(t) \\ \hat{\boldsymbol{\psi}}_f(t)\end{bmatrix}, \tag{16}$$

$$\hat{\boldsymbol{\Phi}}(t) = \begin{bmatrix}\boldsymbol{\phi}(t) \\ \hat{\boldsymbol{\psi}}(t)\end{bmatrix}, \tag{17}$$

$$\begin{aligned}\hat{\boldsymbol{\phi}}_f(t) = & [\hat{f}_{1,f}(u(t-1)),\ \ldots,\ \hat{f}_{1,f}(u(t-n)),\ \ldots, \\ & \hat{f}_{m,f}(u(t-1)),\ \ldots,\ \hat{f}_{m,f}(u(t-n))]^{\mathrm{T}},\end{aligned} \tag{18}$$

$$\begin{aligned}\hat{\boldsymbol{\psi}}_f(t) = & [-\hat{w}_f(t-1),\ \ldots,\ -\hat{w}_f(t-n_c), \\ & \hat{v}_f(t-1),\ \ldots,\ \hat{v}_f(t-n_d)]^{\mathrm{T}},\end{aligned} \tag{19}$$

$$\boldsymbol{\phi}(t) = [\boldsymbol{F}_1(t),\ \boldsymbol{F}_2(t),\ \ldots,\ \boldsymbol{F}_m(t)]^{\mathrm{T}}, \tag{20}$$

$$\begin{aligned}\hat{\boldsymbol{\psi}}(t) = & [-\hat{w}(t-1),\ \ldots,\ -\hat{w}(t-n_c), \\ & \hat{v}(t-1),\ \ldots,\ \hat{v}(t-n_d)]^{\mathrm{T}},\end{aligned} \tag{21}$$

$$\begin{aligned}\hat{f}_{i,f}(u(t)) = & f_{i,f}(u(t)) + \hat{c}_1(t-1)f_{i,f}(u(t-1)) + \cdots \\ & + \hat{c}_{n_c}(t-1)f_{i,f}(u(t-n_c)) \\ & - \hat{d}_1(t-1)\hat{f}_{i,f}(u(t-1)) - \cdots \\ & - \hat{d}_{n_d}(t-1)\hat{f}_{i,f}(u(t-n_d)),\end{aligned} \tag{22}$$

$$\begin{aligned}\hat{w}_f(t) = & \hat{w}(t) - \hat{d}_1(t-1)\hat{w}_f(t-1) \\ & - \cdots - \hat{d}_{n_d}(t-1)\hat{w}_f(t-n_d)),\end{aligned} \tag{23}$$

$$\begin{aligned}\hat{v}_f(t) = & \hat{v}(t) - \hat{d}_1(t-1)\hat{v}_f(t-1) \\ & - \cdots - \hat{d}_{n_d}(t-1)\hat{v}_f(t-n_d)),\end{aligned} \tag{24}$$

$$\hat{w}(t) = y(t) - \boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1), \tag{25}$$

$$\hat{\boldsymbol{\vartheta}}(t) = [\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t), \hat{\boldsymbol{\rho}}^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{26}$$

$$\hat{\boldsymbol{\theta}}(t) = [\widehat{a_1\boldsymbol{\beta}^{\mathrm{T}}(t)},\ \widehat{a_2\boldsymbol{\beta}^{\mathrm{T}}(t)},\ \ldots,\ \widehat{a_m\boldsymbol{\beta}^{\mathrm{T}}(t)}]^{\mathrm{T}}, \tag{27}$$

$$\hat{\boldsymbol{\rho}}(t) = [\hat{c}_1(t),\ \ldots,\ \hat{c}_{n_c}(t),\ \hat{d}_1,\ \ldots,\ \hat{d}_{n_d}(t)]^{\mathrm{T}}. \tag{28}$$

The parameter vector $\hat{\boldsymbol{\theta}}(t)$ contains the products of a_i and β_l ($i = 1, 2, \ldots, m$, $l = 1, 2, \ldots, n$). Then we have

$$\begin{aligned}\sum_{l=1}^{m}\widehat{a_i\beta_l(t)}^2 &= \widehat{a_i\beta_1(t)}^2 + \widehat{a_i\beta_2(t)}^2 + \cdots + \widehat{a_i\beta_n(t)}^2 \\ &= (\hat{a}_i(t)\hat{\beta}_1(t))^2 + \cdots + (\hat{a}_m(t)\hat{\beta}_n(t))^2 \\ &= \hat{a}_i^2(t)\|\hat{\boldsymbol{\beta}}(t)\|^2.\end{aligned}$$

From the normalization hypothesis $\|\boldsymbol{\beta}\| = 1$, $\beta_1 > 0$. The estimate $\hat{a}_i(t)$ can be computed by

$$\hat{a}_i(t) = \mathrm{sgn}[\widehat{a_i\beta_1(t)}]\sqrt{\sum_{l=1}^{m}\widehat{a_i\beta_l(t)}}. \tag{29}$$

Then we can obtain the estimates $\hat{\beta}_{1,i}(t), \hat{\beta}_{2,i}(t), \ldots, \hat{\beta}_{n,i}(t)$ of $\beta_1, \beta_2, \ldots, \beta_n$ as

$$\hat{\beta}_{1,i}(t) = \frac{\widehat{a_i\beta_1(t)}}{\hat{a}_i(t)},\ \ldots,\ \hat{\beta}_{n,i}(t) = \frac{\widehat{a_i\beta_n(t)}}{\hat{a}_i(t)}. \tag{30}$$

In order to improve the estimation accuracy, take their average as the estimates of $\beta_1, \beta_2, \ldots, \beta_n$, that is

$$\hat{\beta}_1(t) = \frac{1}{m}\sum_{i=1}^{m}\hat{\beta}_{1,i}(t),\ \ldots,\ \hat{\beta}_n(t) = \frac{1}{m}\sum_{i=1}^{m}\hat{\beta}_{n,i}(t). \tag{31}$$

The steps for implementing the ML-RLS algorithm in (12)-(31) are as follows:

1) To initialize: set the parameter estimation accuracy ε. Set the initial values $\hat{\boldsymbol{\vartheta}}(0) = \mathbf{1}_{(mn+n_c+n_d)\times 1}/p_0$, $\hat{f}_{i,f}(u(t-j)) = 1/p_0$, $\hat{w}_f(t-j) = 1/p_0$, $\hat{w}(t-j) = 1/p_0$, $\hat{v}_f(t-j) = 1/p_0$, $\hat{v}(t-j) = 1/p_0$ and $\boldsymbol{P}_f(0) = p_0\boldsymbol{I}_{(mn+n_c+n_d)}$, $j = 1, 2, \ldots, \max[n_c, n_d]$, $i = 1, 2, \ldots, m$, $p_0 = 10^6$.
2) Collect the observation data $u(t)$ and $y(t)$, form the information vector $\hat{\boldsymbol{\Phi}}(t)$ by (17) and form the filtered information vector $\hat{\boldsymbol{\Phi}}_f(t)$ by (16).
3) Compute the gain vector $\boldsymbol{L}_f(t)$ by (13) and the covariance matrix $\boldsymbol{P}_f(t)$ by (14).
4) Update the parameter estimation vector $\hat{\boldsymbol{\vartheta}}(t)$ by (12).
5) Compute the residual $\hat{v}(t)$ and $\hat{w}(t)$ by (15), (25). Compute the filtered variables $\hat{f}_{i,f}(u(t))$, $\hat{w}_f(t)$ and $\hat{v}_f(t)$ by (22)-(24). Compute $\hat{a}_i(t)$, $\hat{\beta}_l(t)$ in (29)-(31).
6) If $\|\hat{\boldsymbol{\vartheta}}(t) - \hat{\boldsymbol{\vartheta}}(t-1)\| > \varepsilon$, increase t by 1 go to Step 4; Otherwise, terminate the recursive calculation procedure and obtain the parameter estimate $\hat{\boldsymbol{\vartheta}}(t)$.

4. THE RECURSIVE GENERALIZED EXTENDED LEAST SQUARES ALGORITHM

In order to illustrate the advantage of the ML-RLS identification algorithm proposed in Section 3, we derive an over-parameterization-based recursive generalized extended least squares identification algorithm as a comparison.

Define a quadratic criterion function

$$J(\boldsymbol{\vartheta}) = \frac{1}{2}\sum_{j=1}^{t}[y(t) - \boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{\vartheta}(t-1)]^2.$$

Minimizing $J(\boldsymbol{\vartheta})$ and letting the derivatives of $J(\boldsymbol{\vartheta})$ with respect to $\boldsymbol{\vartheta}$ be zero, the estimate of the parameter vector $\boldsymbol{\vartheta}$ at time t is given by

$$\hat{\boldsymbol{\vartheta}}(t) = \left[\sum_{j=1}^{t}\boldsymbol{\Phi}(t)\boldsymbol{\Phi}^{\mathrm{T}}(t)\right]^{-1}\sum_{j=1}^{t}\boldsymbol{\Phi}(t)y(t). \tag{32}$$

The parameter estimation vector $\hat{\boldsymbol{\vartheta}}(t)$ in (32) can be written as a recursive least squares algorithm

$$\begin{aligned}
\hat{\boldsymbol{\vartheta}}(t) &= \hat{\boldsymbol{\vartheta}}(t-1) + \boldsymbol{L}(t)[y(t) - \boldsymbol{\Phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)],\\
\boldsymbol{L}(t) &= \frac{\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(t)}{1+\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(t)},\\
\boldsymbol{P}(t) &= [\boldsymbol{I} - \boldsymbol{L}(t)\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(t)]\boldsymbol{P}(t-1).
\end{aligned}$$

Similarly, replacing the unknown information $\boldsymbol{\psi}(t)$ in $\boldsymbol{\Phi}(t)$ with its estimate $\hat{\boldsymbol{\psi}}(t)$, we can obtain the recursive generalized extended least squares (RLS) parameter estimation algorithm based on the over-parameterization model as follows:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \boldsymbol{L}(t)[y(t) - \hat{\boldsymbol{\Phi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1)], \tag{33}$$

$$\boldsymbol{L}(t) = \frac{\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(t)}{1+\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\hat{\boldsymbol{\Phi}}(t)}, \tag{34}$$

$$\boldsymbol{P}(t) = [\boldsymbol{I} - \boldsymbol{L}(t)\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(t)]\boldsymbol{P}(t-1), \tag{35}$$

$$\hat{\boldsymbol{\Phi}}(t) = \begin{bmatrix}\boldsymbol{\phi}(t)\\ \hat{\boldsymbol{\psi}}(t)\end{bmatrix}, \tag{36}$$

$$\boldsymbol{\phi}(t) = [\boldsymbol{F}_1(t), \boldsymbol{F}_2(t), \ldots, \boldsymbol{F}_m(t)]^{\mathrm{T}}, \tag{37}$$

$$\begin{aligned}\hat{\boldsymbol{\psi}}(t) = [&-\hat{w}(t-1),\ \ldots,\ -\hat{w}(t-n_c),\\ &\hat{v}(t-1), \ldots, \hat{v}(t-n_d)]^{\mathrm{T}},\end{aligned} \tag{38}$$

$$\hat{w}(t) = y(t) - \boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1), \tag{39}$$

$$\hat{v}(t) = y(t) - \hat{\boldsymbol{\Phi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1), \tag{40}$$

$$\hat{\boldsymbol{\vartheta}}(t) = [\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t), \hat{\boldsymbol{\rho}}^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{41}$$

$$\hat{\boldsymbol{\theta}}(t) = [\widehat{a_1\boldsymbol{\beta}^{\mathrm{T}}(t)},\ \widehat{a_2\boldsymbol{\beta}^{\mathrm{T}}(t)},\ \ldots,\ \widehat{a_m\boldsymbol{\beta}^{\mathrm{T}}(t)}]^{\mathrm{T}}, \tag{42}$$

$$\hat{\boldsymbol{\rho}}(t) = [\hat{c}_1(t),\ \ldots,\ \hat{c}_{n_c}(t),\ \hat{d}_1(t),\ \ldots,\ \hat{d}_{n_d}(t)]^{\mathrm{T}}. \tag{43}$$

The steps for implementing the RLS algorithm in (33)-(43) are as follows:

1) To initialize: Set the parameter estimation accuracy ε and the initial values $\hat{w}(t-j) = 1/p_0$, $\hat{v}(t-j) = 1/p_0$ and $\boldsymbol{P}(0) = p_0\boldsymbol{I}_{(mn+n_c+n_d)}$, $j = 1, 2, \ldots, \max[n_c, n_d]$, $p_0 = 10^6$.
2) Collect the observation data $u(t)$ and $y(t)$, form the information vector $\hat{\boldsymbol{\Phi}}(t)$ by (36).
3) Compute the gain vector $\boldsymbol{L}(t)$ by (34) and the covariance matrix $\boldsymbol{P}(t)$ by (35).
4) Update the parameter estimation vector $\hat{\boldsymbol{\vartheta}}(t)$ by (33).
5) Compute the residual $\hat{v}(t)$ and $\hat{w}(t)$ by (39) and (40). Compute $\hat{a}_i(t)$ and $\hat{\beta}_l(t)$ using (29)-(31).
6) Compare $\hat{\boldsymbol{\vartheta}}(t)$ with $\hat{\boldsymbol{\vartheta}}(t-1)$: if $\|\hat{\boldsymbol{\vartheta}}(t) - \hat{\boldsymbol{\vartheta}}(t-1)\| > \varepsilon$, increase t by 1 go to Step 3; Otherwise, terminate the recursive calculation procedure and obtain the parameter estimate $\hat{\boldsymbol{\vartheta}}(t)$.

The proposed algorithms in this paper can combine other methods [34–38] to study parameter identification of different systems [39–41] and can be applied to other fields [42–47] such as chemical engineering systems. In practice, the design of many fault detection and control algorithms assume that the parameters of the considered system models are known [48–51]. If the case is not so, then some identification methods can be used first for obtaining the parameters of the systems from observation information.

5. EXAMPLE

Consider the following simulation system

$$\begin{aligned}
y(t) &= \boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\beta} + \frac{D(z)}{C(z)}v(t),\\
\boldsymbol{F}(t) &= \begin{bmatrix}u(t-1) & u^2(t-1)\\ u(t-2) & u^2(t-2)\end{bmatrix},\\
\boldsymbol{a} &= [a_1,\ a_2]^{\mathrm{T}} = [-2.1,\ 1]^{\mathrm{T}},\\
\boldsymbol{\beta} &= [\beta_1,\ \beta_2]^{\mathrm{T}} = [0.912,\ -0.41]^{\mathrm{T}},\\
D(z) &= 1 + d_1z^{-1} = 1 + 0.1z^{-1},\\
C(z) &= 1 + c_1z^{-1} = 1 - 0.4z^{-1},\\
\boldsymbol{\Upsilon} &= [a_1,\ a_2,\ \beta_1,\ \beta_2,\ d_1,\ c_1]^{\mathrm{T}}\\
&= [-2.1,\ 1.0,\ 0.912,\ -0.41,\ -0.4,\ 0.1]^{\mathrm{T}}.
\end{aligned}$$

The inputs $\{u(t)\}$ is taken as an independent persistent signal sequence with zero mean and unit variance, $\{v(t)\}$ is an uncorrelated noise sequence with zero mean and variance $\sigma^2 = 0.50^2$, $\sigma^2 = 1.00^2$ and $\sigma^2 = 1.50^2$, respectively. Applying the ML-RLS algorithm and the RLS algorithm to estimate the parameters of this example system, the parameter estimates and their estimation errors $\delta := \|\hat{\boldsymbol{\Upsilon}}_t - \boldsymbol{\Upsilon}\|/\|\boldsymbol{\Upsilon}\|$ of the RLS algorithm and the ML-RLS algorithm are shown in Tables 1-3 and Figs. 1-4. The ML-RLS parameter estimates versus k with $\sigma^2 = 0.50^2$ are shown in Fig. 5.

From Tables 1-3 and Figs. 1-5, we can draw the following conclusions.

1) It can be seen that the estimation errors are becoming smaller as t increases, and then tends to be stationary. This shows that the ML-RLS algorithm and the RLS algorithm are effective.
2) The estimation errors are becoming smaller as the noise variance decreases in the ML-RLS algorithm and the RLS algorithm.

Table 1. The RLS and ML-RLS estimates and their errors with $\sigma^2 = 0.50^2$.

Algorithms	t	a_1	a_2	β_1	β_2	c_1	d_1	$\delta(\%)$
RLS	100	-1.97966	0.75807	0.94943	-0.28672	-0.12451	0.27310	17.23563
	200	-2.05863	0.91699	0.92993	-0.35685	-0.25117	0.21996	8.56679
	500	-2.03982	0.98246	0.91952	-0.38971	-0.33661	0.18206	4.79855
	1000	-2.05026	1.01979	0.92039	-0.39000	-0.37145	0.02228	3.93572
	2000	-2.07973	1.00311	0.91645	-0.40001	-0.39202	0.03507	2.70655
ML-RLS	100	-2.12771	0.99331	0.92520	-0.37917	-0.39068	0.23794	5.65630
	200	-2.09912	1.04162	0.91904	-0.39403	-0.40103	0.15360	2.73205
	500	-2.07329	0.99781	0.91751	-0.39737	-0.38119	0.15704	2.61963
	1000	-2.08792	1.00071	0.91734	-0.39795	-0.38776	0.04826	2.18695
	2000	-2.09465	1.00653	0.91438	-0.40486	-0.40718	0.08380	0.79657
True values		-2.10000	1.00000	0.91200	-0.41000	-0.40000	0.10000	

Table 2. The RLS and ML-RLS estimates and their errors with $\sigma^2 = 1.00^2$.

Algorithms	t	a_1	a_2	β_1	β_2	c_1	d_1	$\delta(\%)$
RLS	100	-1.94222	0.69348	0.95564	-0.22540	-0.02833	0.32741	22.88745
	200	-2.04614	0.89668	0.93394	-0.33637	-0.19965	0.25234	11.21966
	500	-2.01989	0.97964	0.92129	-0.38283	-0.31323	0.19420	6.04699
	1000	-2.03343	1.02786	0.92278	-0.38356	-0.35932	-0.02114	5.82843
	2000	-2.07277	1.00459	0.91789	-0.39662	-0.38620	0.00474	3.94526
ML-RLS	100	-2.22207	0.95829	0.94742	-0.31815	-0.33599	0.24764	8.90927
	200	-2.08815	1.12630	0.93050	-0.36568	-0.38978	0.19292	6.42219
	500	-2.01288	0.99549	0.92769	-0.37003	-0.32236	0.15786	5.34948
	1000	-2.06217	1.00590	0.92703	-0.37344	-0.34383	-0.01816	5.53326
	2000	-2.08319	1.02103	0.91891	-0.39446	-0.40130	0.01899	3.39379
True values		-2.10000	1.00000	0.91200	-0.41000	-0.40000	0.10000	

Table 3. The RLS and ML-RLS estimates and their errors with $\sigma^2 = 1.50^2$.

Algorithms	t	a_1	a_2	β_1	β_2	c_1	d_1	$\delta(\%)$
RLS	100	-2.28337	0.92833	0.96289	-0.26585	-0.28844	0.00378	11.28758
	200	-2.07573	1.20737	0.93803	-0.34525	-0.38274	0.00203	9.41775
	500	-1.96215	1.00186	0.93436	-0.34745	-0.27382	0.00445	8.58289
	1000	-2.04112	1.01266	0.93471	-0.35123	-0.30415	-0.03099	7.18253
	2000	-2.07471	1.03439	0.92286	-0.38512	-0.39596	-0.01823	5.01563
ML-RLS	100	-2.25222	0.92296	0.96124	-0.27244	-0.28908	0.01824	10.27323
	200	-2.06935	1.18723	0.93696	-0.34829	-0.38224	0.01431	8.54847
	500	-1.97019	0.99942	0.93333	-0.35149	-0.28283	0.01613	7.94086
	1000	-2.04466	1.01043	0.93295	-0.35650	-0.31297	0.00157	6.00622
	2000	-2.07584	1.03129	0.92195	-0.38731	-0.39718	0.00385	4.16785
True values		-2.10000	1.00000	0.91200	-0.41000	-0.40000	0.10000	

3) Under the same noise variances, compared with the RLS algorithm, the ML-RLS algorithm produces higher parameter estimation accuracy. Thus, the ML-RLS algorithm shows better performance.

6. CONCLUSIONS

This paper derives an ML-RLS algorithm and an RLS algorithm for identifying the bilinear-parameter systems with ARMA noise using the over-parameterization method and the maximum likelihood principle. The simulation results show that the ML-RLS algorithm has a higher estimation accuracy compared with the RLS algorithm. The proposed identification algorithm for bilinear-parameter stochastic systems with ARMA noises in this paper can joint some recirsive and iterative schemes [52–60] to explored new estimation algorithms for linear and nonlinear systems and can be applied to other control and

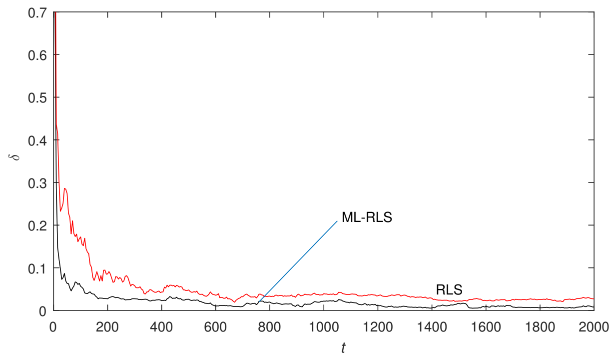


Fig. 1. The ML-RLS and RLS estimation errors versus t ($\sigma^2 = 0.50^2$).

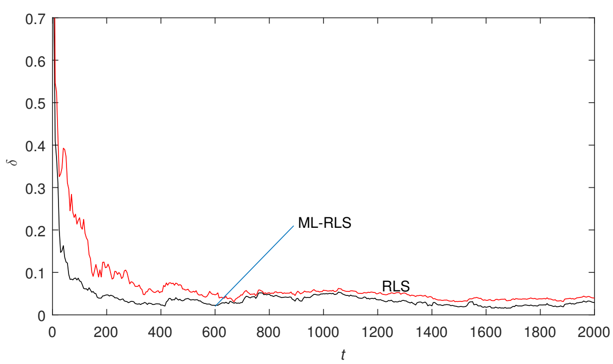


Fig. 2. The ML-RLS and RLS estimation errors versus t ($\sigma^2 = 1.00^2$).

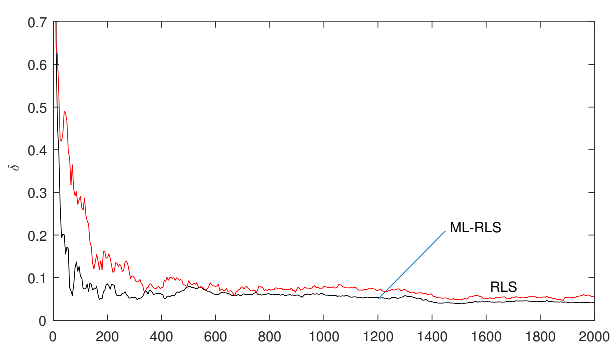


Fig. 3. The ML-RLS and RLS estimation errors versus t ($\sigma^2 = 1.50^2$).

schedule areas [61–68] such as the information processing and engineering systems by means of some mathematical tools [69–74].

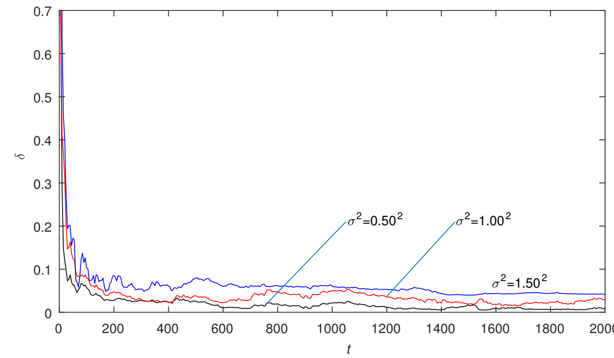


Fig. 4. The ML-RLS parameter estimation errors versus t with different σ^2.

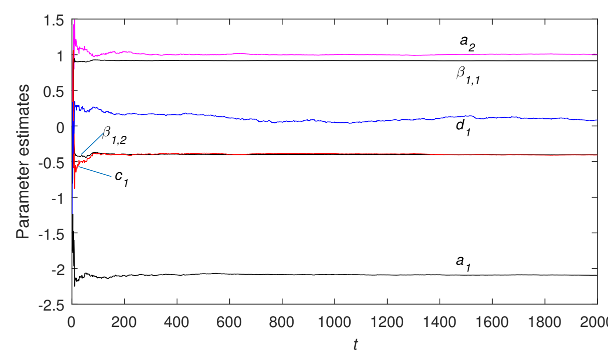


Fig. 5. The ML-RLS parameter estimates versus t with $\sigma^2 = 0.50^2$.

REFERENCES

[1] Y. W. Mao, S. Liu, and J. F. Liu, “Robust economic model predictive control of nonlinear networked control systems with communication delays,” *International Journal of Adaptive Control and Signal Processing*, vol. 34, no. 5, pp. 614-637, May 2020.

[2] F. Ding and T. Chen, “Combined parameter and output estimation of dual-rate systems using an auxiliary model,” *Automatica*, vol. 40, no. 10, pp. 1739-1748, 2004.

[3] F. Ding and T. Chen, “Parameter estimation of dual-rate stochastic systems by using an output error method,” *IEEE Transactions on Automatic Control*, vol. 50, no. 9, pp. 1436-1441, September 2005.

[4] L. Xu, “Separable Newton recursive estimation method through system responses based on dynamically discrete measurements with increasing data length,” *International Journal of Control, Automation, and Systems*, vol. 20, pp. 432-443, 2022.

[5] L. Xu and G. L. Song, “A recursive parameter estimation algorithm for modeling signals with multi-frequencies,” *Circuits Systems and Signal Processing*, vol. 39, no. 8, pp. 4198-4224, August 2020.

[6] L. Xu, “Separable multi-innovation Newton iterative modeling algorithm for multi-frequency signals based on the

sliding measurement window," *Circuits Systems and Signal Processing*, vol. 41, pp. 805-830, 2022.

[7] D. Q. Wang, Y. R. Yan, Y. J. Liu, and J. H. Ding, "Model recovery for Hammerstein systems using the hierarchical orthogonal matching pursuit method," *Journal of Computational and Applied Mathematics*, vol. 345, pp. 135-145, January 2019.

[8] D. Q. Wang, S. Zhang, M. Gan, and J. L. Qiu, "A novel EM identification method for Hammerstein systems with missing output data," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2500-2508, April 2020.

[9] M. H. Li and X. M. Liu, "Maximum likelihood hierarchical least squares-based iterative identification for dual-rate stochastic systems," *International Journal of Adaptive Control and Signal Processing*, vol. 35, no. 2, pp. 240-261, February 2021.

[10] M. H. Li and X. M. Liu, "Iterative identification methods for a class of bilinear systems by using the particle filtering technique," *International Journal of Adaptive Control and Signal Processing*, vol. 35, no. 11, pp. 2056-2074, November 2021.

[11] M. H. Li and X. M. Liu, "The filtering-based maximum likelihood iterative estimation algorithms for a special class of nonlinear systems with autoregressive moving average noise using the hierarchical identification principle," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 7, pp. 1189-1211, July 2019.

[12] L. J. Wang, Y. Ji, H. L. Yang, and L. Xu, "Decomposition-based multiinnovation gradient identification algorithms for a special bilinear system based on its input-output representation," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 9, pp. 3607-3623, June 2020.

[13] Y. Ji, X. K. Jiang, and L. J. Wan, "Hierarchical least squares parameter estimation algorithm for two-input Hammerstein finite impulse response systems," *Journal of the Franklin Institute*, vol. 357, no. 8, pp. 5019-5032, May 2020.

[14] J. Wang, Y. Ji, and C. Zhang, "Iterative parameter and order identification for fractional-order nonlinear finite impulse response systems using the key term separation," *International Journal of Adaptive Control and Signal Processing*, vol. 35, no. 8, pp. 1562-1577, August 2021.

[15] Y. M. Fan and X. M. Liu, "Two-stage auxiliary model gradient-based iterative algorithm for the input nonlinear controlled autoregressive system with variable-gain nonlinearity," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 14, pp. 5492-5509, September 2020.

[16] X. M. Liu and Y. M. Fan, "Maximum likelihood extended gradient-based estimation algorithms for the input nonlinear controlled autoregressive moving average system with variable-gain nonlinearity," *International Journal of Robust and Nonlinear Control*, vol. 31, no. 9, pp. 4017-4036, June 2021.

[17] M. T. Chen and T. Hayat, "Iterative identification algorithms for bilinear-parameter systems by using the over-parameterization model and the decomposition," *International Journal of Control, Automation, and Systems*, vol. 16, no. 6, pp. 2634-2643, December 2018.

[18] Y. Ji, C. Zhang, Z. Kang, and T. Yu, "Parameter estimation for block-oriented nonlinear systems using the key term separation," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 9, pp. 3727-3752, June 2020.

[19] X. Zhang and E. F. Yang, "State estimation for bilinear systems through minimizing the covariance matrix of the state estimation errors," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 7, pp. 1157-1173, July 2019.

[20] X. Zhang and L. Xu, "Recursive parameter estimation methods and convergence analysis for a special class of nonlinear systems," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 4, pp. 1373-1393, March 2020.

[21] M. H. Li and X. M. Liu, "The least squares based iterative algorithms for parameter estimation of a bilinear system with autoregressive noise using the data filtering technique," *Signal Processing*, vol. 147, pp. 23-34, June 2018.

[22] M. H. Li and X. M. Liu, "Maximum likelihood least squares based iterative estimation for a class of bilinear systems using the data filtering technique," *International Journal of Control, Automation, and Systems*, vol. 18, no. 6, pp. 1581-1592, June 2020.

[23] F. Ding, X. G. Liu, and J. Chu, "Gradient-based and least-squares-based iterative algorithms for Hammerstein systems using the hierarchical identification principle," *IET Control Theory and Applications*, vol. 7, no. 2, pp. 176-184, January 2013.

[24] F. Ding, "Hierarchical multi-innovation stochastic gradient algorithm for Hammerstein nonlinear system modeling," *Applied Mathematical Modelling*, vol. 37, no. 4, pp. 1694-1704. April 2013.

[25] Y. H. Zhou, "Modeling nonlinear processes using the radial basis function-based state-dependent autoregressive models," *IEEE Signal Processing Letters*, vol. 27, pp. 1600-1604, 2020.

[26] Y. Ji and Z. Kang, "Three-stage forgetting factor stochastic gradient parameter estimation methods for a class of nonlinear systems," *International Journal of Robust and Nonlinear Control*, vol. 31, no. 3, pp. 871-987, February 2021.

[27] F. Ding, "Coupled-least-squares identification for multivariable systems," *IET Control Theory and Applications*, vol. 7, no. 1, pp. 68-79, January 2013.

[28] Y. Ji, Z. Kang, and X. M. Liu, "The data filtering based multiple-stage Levenberg-Marquardt algorithm for Hammerstein nonlinear systems," *International Journal of Robust and Nonlinear Control*, vol. 31, no. 15, pp. 7007-7025, October 2021.

[29] F. Ding, Y. J. Liu, and B. Bao, "Gradient based and least squares based iterative estimation algorithms for multi-input multi-output systems," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 226, no. 1, pp. 43-55, February 2012.

[30] Y. Ji, Z. Kang, and C. Zhang, "Two-stage gradient-based recursive estimation for nonlinear models by using the data filtering," *International Journal of Control, Automation, and Systems*, vol. 19, no. 8, pp. 2706-2715, August 2021.

[31] Y. J. Wang, "Novel data filtering based parameter identification for multiple-input multiple-output systems using the auxiliary model," *Automatica*, vol. 71, pp. 308-313, September 2016.

[32] D. Q. Wang, Q. H. Fan, and Y. Ma, "An interactive maximum likelihood estimation method for multivariable Hammerstein systems," *Journal of the Franklin Institute*, vol. 357, no. 17, pp. 12986-13005, 2020.

[33] H. B. Liu, J. W. Wang, and X. X. Meng, "Hierarchical maximum likelihood generalized extended stochastic gradient algorithms for bilinear-in-parameter systems," *Optimal Control Applications and Methods*, vol. 43, no. 2, pp. 402-417, 2022.

[34] P. Ma and L. Wang, "Filtering-based recursive least squares estimation approaches for multivariate equation-error systems by using the multiinnovation theory," *International Journal of Adaptive Control and Signal Processing*, vol. 35, no. 9, pp. 1898-1915, 2021.

[35] J. Chen, B. Huang, M. Gan, and C. L. P. Chen, "A novel reduced-order algorithm for rational models based on Arnoldi process and Krylov subspace," *Automatica*, vol. 129, Article Number: 109663, July 2021.

[36] J. Pan, X. Jiang, X. K. Wan, and W. Ding, "A filtering based multi-innovation extended stochastic gradient algorithm for multivariable control systems," *International Journal of Control, Automation, and Systems*, vol. 15, no. 3, pp. 1189-1197, June 2017.

[37] L. Xu, F. Y. Chen, and T. Hayat, "Hierarchical recursive signal modeling for multi-frequency signals based on discrete measured data," *International Journal of Adaptive Control and Signal Processing*, vol. 35, no. 5, pp. 676-693, May 2021.

[38] L. Xu, "Separable multi-innovation stochastic gradient estimation algorithm for the nonlinear dynamic responses of systems," *International Journal of Adaptive Control and Signal Processing*, vol. 34, no. 7, pp. 937-954, July 2020.

[39] J. Pan, W. Li, and H. P. Zhang, "Control algorithms of magnetic suspension systems based on the improved double exponential reaching law of sliding mode control," *International Journal of Control, Automation, and Systems*, vol. 16, no. 6, pp. 2878-2887, December 2018.

[40] C. Xu, H. Xu, and C. Liu, "Adaptive bipartite consensus of competitive linear multi-agent systems with asynchronous intermittent communication," *International Journal of Robust and Nonlinear Control*, vol. 32, no. 9, pp. 5120-5140, 2022.

[41] J. L. Ding and W. H. Zhang, "Finite-time adaptive control for nonlinear systems with uncertain parameters based on the command filters," *International Journal of Adaptive Control and Signal Processing*, vol. 35, no. 9, pp. 1754-1767, 2021.

[42] F. Ding, G. Liu, and X. P. Liu, "Partially coupled stochastic gradient identification methods for non-uniformly sampled systems," *IEEE Transactions on Automatic Control*, vol. 55, no. 8, pp. 1976-1981, August 2010.

[43] G. Zhao, T. H. Cao, Y. D. Wang, H. R. Zhou, C. Zhang, and C. X. Wan, "Optimal sizing of isolated microgrid containing photovoltaic/photothermal/wind/diesel/battery," *International Journal of Photoenergy*, vol. 2021, Article ID 5566597, 2021.

[44] X. G. Wang, M. Zhao, Y. Zhou, Z. W. Wan, and W. Xu, "Design and analysis for multi-disc coreless axial-flux permanent-magnet synchronous machine," *IEEE Transactions on Applied Superconductivity*, vol. 31, no. 8, Article Number: 5203804, November 2021.

[45] F. Ding, G. Liu, and X. P. Liu, "Parameter estimation with scarce measurements," *Automatica*, vol. 47, no. 8, pp. 1646-1655, August 2011.

[46] X. G. Wang, Z. W. Wan, L. Tang, W. Xu, and M. Zhao, "Electromagnetic performance analysis of an axial flux hybrid excitation motor for HEV drives," *IEEE Transactions on Applied Superconductivity*, vol. 31, no. 8, Article Number: 5205605, November 2021.

[47] Y. H. Zhou, "Hierarchical estimation approach for RBF-AR models with regression weights based on the increasing data length," *IEEE Transactions on Circuits and Systems-II: Express Briefs*, vol. 68, no. 12, pp. 3597-3601, 2021.

[48] W. X. Zhou, J. Fu, H. C. Yan, X. Du, Y. Y. Wang, and H. Zhou, "Event-triggered approximate optimal path-following control for unmanned surface vehicles with state constraints," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1-15, 2021. DOI: 10.1109/TNNLS.2021.3090054

[49] Y. Y. Wang, H. Y. Pu, P. Shi, C. K. Ahn, and J. Luo, "Sliding mode control for singularly perturbed Markov jump descriptor systems with nonlinear perturbation," *Automatica*, vol. 127, Article Number: 109515, May 2021.

[50] N. Zhao, A. Wu, and D. Niyato, "Spatial-temporal aggregation graph convolution network for efficient mobile cellular traffic prediction," *IEEE Communications Letters*, vol. 26, no. 3, pp. 587-591, 2022.

[51] J. Hou, F. W. Chen, P. H. Li, and Z. Q. Zhu, "Gray-box parsimonious subspace identification of Hammerstein-type systems," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 10, pp. 9941-9951, October 2022.

[52] J. Ding and G. Liu, "Hierarchical least squares identification for linear SISO systems with dual-rate sampled-data," *IEEE Transactions on Automatic Control*, vol. 56, no. 11, pp. 2677-2683, November 2011.

[53] J. Xiong, J. Pan, and G. Chen, "Sliding mode dual-channel disturbance rejection attitude control for a quadrotor," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 10, pp. 10489-10499, 2022.

[54] Y. Gu, J. Liu, X. Li, Y. Chou, and Y. Ji, "State space model identification of multirate processes with time-delay using the expectation maximization," *Journal of the Franklin Institute*, vol. 356, no. 3, pp. 1623-1639, February 2019.

[55] Y. H. Zhou, "Partially-coupled nonlinear parameter optimization algorithm for a class of multivariate hybrid models," *Applied Mathematics and Computation*, vol. 414, p. 126663, Februray 2022.

[56] Y. Gu, Y. Chou, J. Liu, and Y. Ji, "Moving horizon estimation for multirate systems with time-varying time-delays," *Journal of the Franklin Institute*, vol. 356, no. 4, pp. 2325-2345, March 2019.

[57] L. Xu, "Hierarchical Newton and least squares iterative estimation algorithm for dynamic systems by transfer functions based on the impulse responses," *International Journal of Systems Science*, vol. 50, no. 1, pp. 141-151, 2019.

[58] X. Zhang and E. F. Yang, "Highly computationally efficient state filter based on the delta operator," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 6, pp. 875-889, June 2019.

[59] Y. An, Y. Zhang, and Z. He, "A lightweight and practical anonymous authentication protocol based on bit-self-test PUF," *Electronics*, vol. 11, no. 5, p. 772, 2022.

[60] X. Zhang, "Adaptive parameter estimation for a general dynamical system with unknown states," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 4, pp. 1351-1372, March 2020.

[61] Y. Cao, Y. K. Sun, G. Xie, and P. Li, "A sound-based fault diagnosis method for railway point machines based on two-stage feature selection strategy and ensemble classifier," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1-10, 2021. DOI: 10.1109/TITS.2021.3109632

[62] Y. Cao, L. C. Ma, S. Xiao, X. Zhang, and W. Xu, "Standard analysis for transfer delay in CTCS-3," *Chinese Journal of Electronics*, vol. 26, no. 5, pp. 1057-1063, September 2017.

[63] Y. Cao, J. K. Wen, and L. C. Ma, "Tracking and collision avoidance of virtual coupling train control system," *Alexandria Engineering Journal*, vol. 60, no. 2, pp. 2115-2125, April 2021.

[64] Y. Chen, C. Zhang, and X. Wan, "Atrial fibrillation detection using feedforward neural network," *Journal of Medical and Biological Engineering*, vol. 242, no. 1, pp. 63-73, 2022.

[65] S. Su, T. Tang, J. Xun, F. Cao, and Y. H. Wang, "Design of running grades for energy-efficient train regulation: A case study for Beijing Yizhuang line," *IEEE Intelligent Transportation Systems Magazine*, vol. 13, no. 2, pp. 189-200, February 2021.

[66] S. Su, X. K. Wang, Y. Cao, and J. T. Yin, "An energy-efficient train operation approach by integrating the metro timetabling and eco-driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 10, pp. 4252-4268, October 2020.

[67] Y. Cao, Z. Wang, F. Liu, P. Li, and G. Xie, "Bio-inspired speed curve optimization and sliding mode tracking control for subway trains," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 7, pp. 6331-6342, July 2019.

[68] Y. Cao, Y. K. Sun, G. Xie, and T. Wen, "Fault diagnosis of train plug door based on a hybrid criterion for IMFs selection and fractional wavelet package energy entropy," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7544-7551, August 2019.

[69] X. Li, H. Wang, and B. Wu, "A stable and efficient technique for linear boundary value problems by applying kernel functions," *Applied Numerical Mathematics*, vol. 172, pp. 206-214, 2022.

[70] X. Li and B. Wu, "A kernel regression approach for identification of first order differential equations based on functional data," *Applied Mathematics Letters*, vol. 127, p. 107832, 2022.

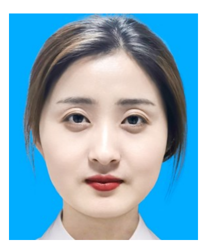
[71] F. Geng and X. Wu, "Reproducing kernel functions based univariate spline interpolation," *Applied Mathematics Letters*, vol. 122, p. 107525, 2021.

[72] X. Y. Li and B. Y. Wu, "Superconvergent kernel functions approaches for the second kind Fredholm integral equations," *Applied Numerical Mathematics*, vol. 167, pp. 202-210, 2021.

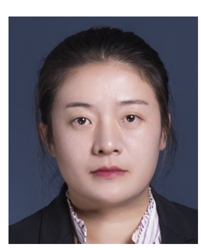
[73] H. J. Wang, H. D. Fan, and J. Pan, "Complex dynamics of a four-dimensional circuit system," *International Journal of Bifurcation and Chaos*, vol. 31, no. 14, p. 2150208, 2021.

[74] H. Wang, G. Ke, and H. Fan, "Multitudinous potential hidden Lorenz-like attractors coined," *European Physical Journal-Special Topics*, vol. 231, no. 3, pp. 359-368, 2022.

Haibo Liu was born in Weifang, Shandong Province in May 1989. He received his MA.Sc. degree at the Shandong University (Jinan, China), in 2014. Since then he is with the Qingdao University of Science and Technology. He is currently a Ph.D. student at the Qingdao University of Science and Technology. His research interests include system identification.



Junwei Wang was born in Zibo, Shangdong Province in September 1996. She received his B.E. degree from Yantai Nanshan University (Yantai, China), in 2019. She is currently a Master student in the College of Automation and Electronic Engineering at the Qingdao University of Science and Technology. Her research interests include system identification and parameter estimation.



Yan Ji received her Ph.D. degree in the School of Communication and Control Engineering from Jiangnan University (Wuxi, China), in 2010. She has been an Associate Professor in the College of Automation and Electronic Engineering, Qingdao University of Science and Technology. Her research interests include control of complex networks, system identification, and process control.