RESEARCH ARTICLE

Iterative Estimation Algorithm for Bilinear Stochastic Systems by Using the Newton Search

Siyu Liu[1] **and Feng Ding**[2,*]

[1] College of Engineering, Zhejiang Normal University, Jinhua 321004, China
[2] School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China

Abstract

This study addresses the challenge of estimating parameters iteratively in bilinear state-space systems affected by stochastic noise. A Newton iterative (NI) algorithm is introduced by utilizing the Newton search and iterative identification theory for identifying the system parameters. Following the estimation of the unknown parameters, we create a bilinear state observer (BSO) using the Kalman filtering principle for state estimation. Subsequently, we propose the BSO-NI algorithm for simultaneous parameter and state estimation. An iterative algorithm based on gradients is given for comparisons to illustrate the effectiveness of the proposed algorithms.

Keywords: newton search, bilinear system, parameter estimation, system identification, iterative method.

Academic Editor:
Seifedine Kadry

Submitted: 02 November 2024
Accepted: 26 March 2025
Published: 14 April 2025

Vol. 2, **No.** 2, 2025.
10.62762/TIS.2024.155941

***Corresponding author:**
✉ Feng Ding
fding@jiangnan.edu.cn

1 Introduction

The design of intelligent systems—including autonomous control agents, industrial IoT platforms, and data-driven decision-support systems—fundamentally depends on the availability of accurate mathematical models of the underlying physical processes. System identification provides the theory and techniques used to construct such models from observation data [1, 2], forming an essential preprocessing layer for any intelligent systematics pipeline. These systems can be linear, bilinear, or nonlinear [3], with nonlinear systems frequently appearing in industrial applications. For instance, Hammerstein-type nonlinear models have been applied to wind power prediction using deep belief networks, demonstrating the practical relevance of nonlinear system identification in energy engineering [4]. As a result, the need for effective modeling techniques has grown significantly in fields such as signal processing, system analysis, and intelligent control. Bilinear models, in particular, are advantageous for intelligent industrial applications, as they more accurately capture the nonlinear characteristics of systems compared to linear models. This capability makes them essential in various intelligent process control scenarios, including heat exchangers, nuclear reactors, and chemical operations, where precise model knowledge directly governs the quality of automated decision-making.

Some classical identification methods have been

Citation
Liu, S., & Ding, F. (2025). Iterative Estimation Algorithm for Bilinear Stochastic Systems by Using the Newton Search. *ICCK Transactions on Intelligent Systematics*, 2(2), 76–84.

© 2025 ICCK (Institute of Central Computation and Knowledge)

applied in the parameter estimation of bilinear systems. For instance, gradient identification techniques have been developed by minimizing criterionfunctions, including greedy search methods combining gradient descent and least squares for separable nonlinear models [5], as well asmultidirection gradient iterative approaches applicable to general parameter estimation problems [6]. An et al. [8] proposed a maximum likelihood basedmulti-innovation stochastic gradient identification algorithm for bilinearstochastic systems affected by colored ARMA noise. Least squares methods are foundational for identifying linear-parameter systems and have been widely utilized across various fields. An auxiliary model based maximum likelihood least squares iterative algorithm has been developed to estimate unknown parameters for multivariable autoregressive output-error systems [9].

Recently, several new ideas, theories, and principles have emerged in system identification for constructing mathematical models and determining model parameters. Notable among these are the multi-innovation identification theory, hierarchical identification principle, and filtering identification concept, all of which contribute to the advancement of system identification. The multi-innovation identification theory enhances the accuracy of estimating parameters. The hierarchical identification principle significantly improves the computational efficiency of identification algorithms, particularly for large-scale complex systems. Meanwhile, the filtering identification concept addresses parameter estimation challenges in systems with colored noise, thereby enhancing estimation accuracy. These approaches are applicable to bilinear systems. Recently, An et al. [10] explored various parameter estimation algorithms for bilinear systems utilizing the maximum likelihood gradient-based iterative and hierarchical principles. Additionally, Gu examined an identification algorithm based on multi-innovation stochastic gradient that incorporates filtering methods for bilinear systems utilizing data filtering theory [11].

The Newton method, a classical optimization tool, is a root-finding algorithm that leverages the initial terms of a function's Taylor series expansion near an estimated root [12]. This method has been extensively studied over the years and has diverse applications, including minimization and maximization problems, solving transcendental equations, and numerically verifying solutions to nonlinear equations. For instance, Xu proposed a separable multi-innovation Newton iterative modeling algorithm for multi-frequency signals based on the sliding measurement window [12]. Furthermore, a separable Newton recursive estimation method through system responses based on dynamically discrete measurements with increasing data length was also developed [13].

Notably, bilinear systems pose a greater challenge than general linear ones due to the increased number of parameters that need to be estimated. This complexity makes standard parameter estimation methods inadequate for achieving high-precision estimates. Additionally, the characteristics of bilinear systems prevent the direct application of conventional state estimation techniques. While Kalman filtering has been widely adopted for linear systems—for example, combined with multi-step gradient iterative algorithms for ARX models with random missing outputs [16]—alternative frameworks such as confidence set-membership estimation for LPV systems [14] and constrained zonotope-based set-membership estimation for discrete time-varying systems [15] address broader classes of systems but are not directly applicable to bilinear dynamics. The finite impulse response (FIR) filter [17] similarly cannot be directly applied to bilinear systems. Consequently, there is a need to develop a state estimator specifically tailored to bilinear systems. Although nonlinear state estimation can be approached through methods such as Koopman operator-based set-membership estimation [18], such general nonlinear techniques are not directly optimized for the bilinear structure, motivating the design of a Kalman filtering-based bilinear state observer in this work.

In this paper, we present a novel identification algorithm aimed at identifying bilinear stochastic systems. Our approach introduces a Newton-based iterative algorithm for parameter estimation. Subsequently, we employ the bilinear state observer-based NI (BSO-NI) method to achieve simultaneous estimation of states and parameters. The availability of the scheme is validated by comparing estimation errors with those of gradient-based algorithms. Below, we summarize the main contributions of our work.

- For the system with unknown states, a bilinear state observer is designed based on the Kalman filtering principle for state estimation.
- To tackle the challenge of estimating unknown parameters, we propose a Newton iterative

algorithm founded on the Newton search, which utilizes all sampling data to estimate the parameters of the bilinear system.

- The performance of the proposed algorithm is evaluated against a gradient-based iterative algorithm through a numerical simulation.

2 System description

Consider a bilinear system in state-space representation featuring autoregressive noise, described by an observability canonical model:

$$\boldsymbol{x}(t+1) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{x}(t)u(t) + \boldsymbol{C}u(t) + w(t), \quad (1)$$

$$y(t) = \boldsymbol{D}\boldsymbol{x}(t) + v(t), \quad (2)$$

where $\boldsymbol{x}(t) \in \mathbb{R}^m$, $y(t) \in \mathbb{R}$, and $u(t) \in \mathbb{R}$ are the state vector, output and input variable, and $\boldsymbol{A} \in \mathbb{R}^{m\times m}$, $\boldsymbol{B} \in \mathbb{R}^{m\times m}$, $\boldsymbol{C} \in \mathbb{R}^m$, and $\boldsymbol{D} \in \mathbb{R}^{1\times m}$ are the matrices/vectors of the system parameters:

$$\boldsymbol{A} := \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I}_{m-1} \\ -a_m & \boldsymbol{a} \end{bmatrix}, \quad (3)$$

$$\boldsymbol{a} := [-a_{m-1}, -a_{m-2}, \cdots, -a_1] \in \mathbb{R}^{1\times(m-1)}, \quad (4)$$

$$\boldsymbol{B} := \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \vdots \\ \boldsymbol{b}_m \end{bmatrix} \in \mathbb{R}^{m\times m}, \quad (5)$$

$$\boldsymbol{b}_l := [b_{l1}, b_{l2}, \cdots, b_{lm}] \in \mathbb{R}^{1\times m},\ l = 1, 2, \cdots, m, \quad (6)$$

$$\boldsymbol{C} := \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} \in \mathbb{R}^m, \quad (7)$$

$$\boldsymbol{D} := [1, 0, \ldots, 0] \in \mathbb{R}^{1\times m}, \quad (8)$$

where $\boldsymbol{I}_{m-1}$ represents an identity matrix of size $(m-1)\times(m-1)$, $w(t)$ is the process noise, $v(t)$ is the measurement noise, let the known dimension of the state vector be m, and assume that $y(t) = 0$ and $v(t) = 0$ for $t \leqslant 0$. Bilinear systems can be interpreted as linear systems with time-varying characteristics, exhibiting stability that changes over time. The stability of these systems is determined by the time-varying matrix $\boldsymbol{A} + \boldsymbol{B}u(t)$. Stability can be ensured by designing an input sequence such that all eigenvalues of $\boldsymbol{A} + \boldsymbol{B}u(t)$ remain within the unit circle.

Based on Equations (1)–(8), the subsequent relationships exist:

$$x_p(t+1) = x_{p+1}(t) + \boldsymbol{b}_p\boldsymbol{x}(t)u(t) + c_p u(t), \quad (9)$$

$$x_m(t+1) = -\sum_{p=1}^{m} a_p x_{m-p+1}(t) + \boldsymbol{b}_m\boldsymbol{x}(t)u(t) + c_m u(t), \quad (10)$$

where $p = 1, \cdots, m-1$. To multiply both sides of Equation (9) by z^{-p}, and sum from $p = 1$ to $p = m-1$, and then add this to Equation (10), where both sides are multiplied by z^{-m}, one can derive

$$x_1(t) = -\sum_{p=1}^{m} a_p x_{m-p+1}(t-m) + \sum_{p=1}^{m} \boldsymbol{b}_p\boldsymbol{x}(t-p)u(t-p) + \sum_{p=1}^{m} c_p u(t-p). \quad (11)$$

Inserting (11) into (2) gives:

$$y(t) = -\sum_{p=1}^{m} a_p x_{m-p+1}(t-m) + \sum_{p=1}^{m} \boldsymbol{b}_p\boldsymbol{x}(t-p)u(t-p) + \sum_{p=1}^{m} c_p u(t-p) + v(t). \quad (12)$$

After that, one defines the parameter vector $\boldsymbol{\theta}$, along with the data vector $\boldsymbol{\varphi}(t)$ as:

$$\begin{aligned} \boldsymbol{\theta} &:= [a_1, a_2, \cdots, a_m, \boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_m, \\ &\quad c_1, c_2, \cdots, c_m]^{\mathrm{T}} \in \mathbb{R}^{m^2+2m}, \\ \boldsymbol{\varphi}(t) &:= [-x_m(t-m), -x_{m-1}(t-m), \cdots, -x_1(t-m), \\ &\quad \boldsymbol{x}^{\mathrm{T}}(t-1)u(t-1), \cdots, \boldsymbol{x}^{\mathrm{T}}(t-m)u(t-m), \\ &\quad u(t-1), u(t-2), \cdots, u(t-m)]^{\mathrm{T}} \in \mathbb{R}^{m^2+2m}, \end{aligned} \quad (13)$$

The expression for $y(t)$ in Equation (12) can be expressed as follows:

$$y(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta} + v(t). \quad (14)$$

The model for identification is established in Equation (14). The information vector $\boldsymbol{\varphi}(t)$ includes $\boldsymbol{x}(t-l)$ $(l = 1, 2, \cdots, m)$, $u(t-l)$, and the system parameters a_l, $\boldsymbol{b}_l$, c_l which together form the parameter vector $\boldsymbol{\theta}$. The goal of this study is to explore novel approaches for estimating these states and parameters using $\{u(t)\}$, $\{y(t)\}$, which is affected by white noise $v(t)$. Equation (14) serves as the basis for deriving the iterative estimation approach.

3 The NI algorithm for parameter estimation

Parameter estimation for different types of systems has been tackled using stochastic gradient methods,

iterative methods based on gradients, and various other gradient-based approaches [19]. In order to enhance the precision of parameter estimation further, the use of Newton search has been suggested. Based on the iterative identification idea, a Newton iterative (NI) algorithm has been developed specifically for bilinear systems.

Considering the data length h, we define the criterion function

$$\begin{aligned} J_1(\boldsymbol{\theta}) :=& \frac{1}{2}\sum_{r=1}^{h}[y(r)-\boldsymbol{\varphi}^{\mathrm{T}}(r)\boldsymbol{\theta}]^2 \\ =& \frac{1}{2}\|\boldsymbol{Y}(h)-\boldsymbol{\Phi}(h)\theta\|^2, \end{aligned} \tag{15}$$

where the stacked observed data $\boldsymbol{Y}(h)$ and the stacked information matrix $\boldsymbol{\Phi}(h)$ are defined as:

$$\boldsymbol{Y}(h) := [y(1), y(2), \cdots, y(h)]^{\mathrm{T}} \in \mathbb{R}^h, \tag{16}$$

$$\boldsymbol{\Phi}(h) := [\boldsymbol{\varphi}(1), \boldsymbol{\varphi}(2), \cdots, \boldsymbol{\varphi}(h)]^{\mathrm{T}} \in \mathbb{R}^{h\times(m^2+2m)}. \tag{17}$$

Calculating the partial derivative of $J_1(\boldsymbol{\theta})$ in relation to $\boldsymbol{\theta}$ produces

$$\begin{aligned} \mathrm{grad}[J_1(\boldsymbol{\theta})] :=& \frac{\partial J_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ =& \left[\frac{\partial J_1(\boldsymbol{\theta})}{\partial a_l}, \frac{\partial J_1(\boldsymbol{\theta})}{\partial \boldsymbol{b}_l}, \frac{\partial J_1(\boldsymbol{\theta})}{\partial c_l}\right]^{\mathrm{T}} \in \mathbb{R}^{m^2+2m}. \end{aligned} \tag{18}$$

Define $\hat{\boldsymbol{\theta}}_k$ as the approximation of $\boldsymbol{\theta}$ during iteration k, where $k = 1, 2, 3, \cdots$ represents an iterative parameter. Set the estimate of the stacked information matrix $\boldsymbol{\Phi}(h)$ at iteration k as follows:

$$\hat{\boldsymbol{\Phi}}_k(h) = [\hat{\boldsymbol{\varphi}}_k(1), \hat{\boldsymbol{\varphi}}_k(2), \cdots, \hat{\boldsymbol{\varphi}}_k(h)]^{\mathrm{T}}, \tag{19}$$

where

$$\begin{aligned} \hat{\boldsymbol{\varphi}}_k(t) =& [-\hat{x}_{m,k-1}(t-m), \cdots, -\hat{x}_{1,k-1}(t-m), \\ & \hat{\boldsymbol{x}}_{k-1}^{\mathrm{T}}(t-1)u(t-1), \cdots, \hat{\boldsymbol{x}}_{k-1}^{\mathrm{T}}(t-m)u(t-m), \\ & u(t-1), u(t-2), \cdots, u(t-m)]^{\mathrm{T}}, \end{aligned} \tag{20}$$

From Equation (15), we can calculate the gradient of $J_1(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{k-1}$:

$$\mathrm{grad}[J_1(\hat{\boldsymbol{\theta}}_{k-1})] := \hat{\boldsymbol{\Phi}}_k^{\mathrm{T}}(h)[\boldsymbol{Y}(h) - \hat{\boldsymbol{\Phi}}_k(h)\hat{\boldsymbol{\theta}}_{k-1}]. \tag{21}$$

The gradient method relies on the first derivative, whereas the Newton method exhibits quadratically convergence. To enhance the accuracy of parameter estimation, we calculate the second partial derivative of the criterion function $J_1(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, resulting in the Hessian matrix

$$\begin{aligned} \boldsymbol{H}(\boldsymbol{\theta}) =& \frac{\partial^2 J_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\mathrm{T}}} = \frac{\partial \mathrm{grad}[J_1(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \\ =& \boldsymbol{\Phi}^{\mathrm{T}}(h)\boldsymbol{\Phi}(h). \end{aligned} \tag{22}$$

Using the Newton search and iterative identification idea to minimize $J_1(\boldsymbol{\theta})$, we can obtain the following iterative relation of computing $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_{k-1} + H^{-1}[J_1(\hat{\boldsymbol{\theta}}_{k-1})]\mathrm{grad}[J_1(\hat{\boldsymbol{\theta}}_{k-1})]. \tag{23}$$

Substituting (21) into the above equation, we obtain:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_k =& \hat{\boldsymbol{\theta}}_{k-1} + H^{-1}[J_1(\hat{\boldsymbol{\theta}}_{k-1})] \\ & \times \hat{\boldsymbol{\Phi}}_k^{\mathrm{T}}(h)[\boldsymbol{Y}(h) - \hat{\boldsymbol{\Phi}}_k(h)\hat{\boldsymbol{\theta}}_{k-1}]. \end{aligned} \tag{24}$$

From Equations (16)–(24), we can summarize the Newton iterative (NI) algorithm for the bilinear stochastic systems as follows:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_k =& \hat{\boldsymbol{\theta}}_{k-1} + H^{-1}[J_1(\hat{\boldsymbol{\theta}}_{k-1})]\hat{\boldsymbol{\Phi}}_k^{\mathrm{T}}(h) \\ & \times [\boldsymbol{Y}(h) - \hat{\boldsymbol{\Phi}}_k(h)\hat{\boldsymbol{\theta}}_{k-1}], \end{aligned} \tag{25}$$

$$H[J_1(\hat{\boldsymbol{\theta}}_{k-1})] = \hat{\boldsymbol{\Phi}}_k^{\mathrm{T}}(h)\hat{\boldsymbol{\Phi}}_k(h), \tag{26}$$

$$\hat{\boldsymbol{\Phi}}_k(h) = [\hat{\boldsymbol{\varphi}}_k(1), \hat{\boldsymbol{\varphi}}_k(2), \cdots, \hat{\boldsymbol{\varphi}}_k(h)]^{\mathrm{T}}, \tag{27}$$

$$\begin{aligned} \hat{\boldsymbol{\varphi}}_k(t) =& [-\hat{x}_{m,k-1}(t-m), \cdots, -\hat{x}_{1,k-1}(t-m), \\ & \hat{\boldsymbol{x}}_{k-1}^{\mathrm{T}}(t-1)u(t-1), \cdots, \\ & \hat{\boldsymbol{x}}_{k-1}^{\mathrm{T}}(t-m)u(t-m), u(t-1), \\ & u(t-2), \cdots, u(t-m)]^{\mathrm{T}}, \end{aligned} \tag{28}$$

$$\boldsymbol{Y}(h) = [y(1), y(2), \cdots, y(h)]^{\mathrm{T}} \in \mathbb{R}^h, \tag{29}$$

$$\begin{aligned} \hat{\boldsymbol{\theta}}_k =& [\hat{a}_{1,k}, \hat{a}_{2,k}, \cdots, \hat{a}_{m,k}, \hat{\boldsymbol{b}}_{1,k}, \cdots, \hat{\boldsymbol{b}}_{m,k}, \\ & \hat{\boldsymbol{c}}_{1,k}, \hat{\boldsymbol{c}}_{2,k}, \cdots, \hat{\boldsymbol{c}}_{m,k}]^{\mathrm{T}} \in \mathbb{R}^{m^2+2m}. \end{aligned} \tag{30}$$

In the above NI algorithm for the parameter estimation, we assume that the system states in the information vector are known. Thus, we design the state estimator to estimate them.

4 Bilinear state observer based NI algorithm

Let $\boldsymbol{xi}(t) := \boldsymbol{B}\boldsymbol{x}(t) + \boldsymbol{C}$. This type of bilinear state-space framework can be represented in the form of a linear time-varying model:

$$\boldsymbol{x}(t+1) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{xi}(t)u(t) + w(t), \tag{31}$$

$$y(t) = \boldsymbol{D}\boldsymbol{x}(t) + v(t), \tag{32}$$

According to the NI algorithm, the obtained parameter estimates $\hat{a}_{l,k}$, $\hat{\boldsymbol{b}}_{l,k}$ and $\hat{c}_{l,k}$ from $\hat{\boldsymbol{\theta}}_k$ can be used to construct the estimates $\hat{\boldsymbol{A}}_k$, $\hat{\boldsymbol{B}}_k$ and $\hat{\boldsymbol{C}}_k$ of the system

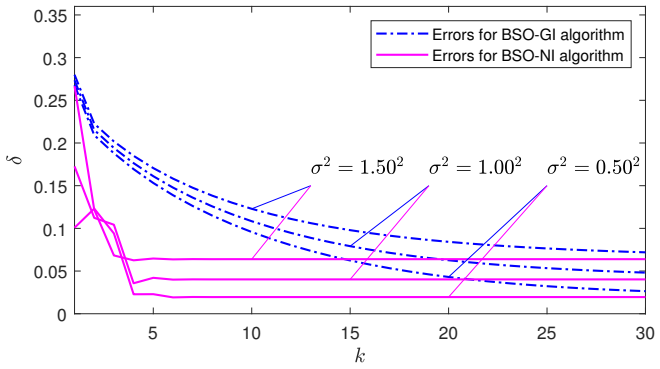


Figure 1. The estimation errors δ versus k for the BSO-NI and BSO-GI algorithms.

matrices/vector $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$, respectively. According to the Kalman filtering principle and referencing to [7], we can design the bilinear state estimator as follows:

$$\hat{\boldsymbol{x}}_k(t+1) = \hat{\boldsymbol{A}}_k\hat{\boldsymbol{x}}_k(t) + \hat{\boldsymbol{xi}}_k(t)u(t) + \boldsymbol{L}_k(t)[y(t) - \boldsymbol{D}\hat{\boldsymbol{x}}_k(t)], \tag{33}$$

$$\boldsymbol{L}_k(t) = \hat{\boldsymbol{A}}_k\boldsymbol{P}_k(t)\boldsymbol{D}^{\mathrm{T}}[1 + \boldsymbol{D}\boldsymbol{P}_k(t)\boldsymbol{D}^{\mathrm{T}}]^{-1}, \tag{34}$$

$$\boldsymbol{P}_k(t+1) = \hat{\boldsymbol{A}}_k\boldsymbol{P}_k(t)\hat{\boldsymbol{A}}_k^{\mathrm{T}} - \boldsymbol{L}_k(t)\boldsymbol{D}\boldsymbol{P}_k(t)\hat{\boldsymbol{A}}_k^{\mathrm{T}}, \tag{35}$$

$$\hat{\boldsymbol{A}}_k = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I}_{m-1} \\ -\hat{a}_{m,k} & -\hat{\boldsymbol{a}}_k \end{bmatrix}, \tag{36}$$

$$\hat{\boldsymbol{a}}_k = [\hat{a}_{m-1,k}, \hat{a}_{m-2,k}, \cdots, \hat{a}_{1,k}], \tag{37}$$

$$\hat{\boldsymbol{xi}}_k(t) = \hat{\boldsymbol{B}}_k\hat{\boldsymbol{x}}_k(t) + \hat{\boldsymbol{C}}_k, \tag{38}$$

$$\hat{\boldsymbol{B}}_k = [\hat{\boldsymbol{b}}_{1,k}^{\mathrm{T}}, \hat{\boldsymbol{b}}_{2,k}^{\mathrm{T}}, \cdots, \hat{\boldsymbol{b}}_{m,k}^{\mathrm{T}}]^{\mathrm{T}}, \tag{39}$$

$$\hat{\boldsymbol{C}}_k = [\hat{c}_{1,k}, \hat{c}_{2,k}, \cdots, \hat{c}_{m,k}]^{\mathrm{T}}. \tag{40}$$

Equations (33)–(40) form the state estimator for bilinear systems, enabling the computation of the state estimation vector denoted by $\hat{\boldsymbol{x}}_k(t)$. The initial values $\hat{\boldsymbol{x}}_k(1)$ and $\boldsymbol{P}_k(1)$ can be chosen arbitrarily. For instance, we can set $\hat{\boldsymbol{x}}_k(1) = \boldsymbol{1}_m$ and $\boldsymbol{P}_k(1) = \boldsymbol{I}_m$.

Based on the above preparations, Equations (25)–(30) and (33) to (40) form the bilinear state estimator-based NI (BSO-NI) algorithm, and the joint state and parameter estimation is realized. Here, we can summarize the BSO-NI algorithm in Algorithm 1.

Algorithm 1: The BSO-NI algorithm for bilinear stochastic systems

Data: $u(t)$ and $y(t)$, $t = 1, 2, \cdots, h$
Result: The BSO-NI estimates $\hat{\boldsymbol{\theta}}_k$ and $\hat{\boldsymbol{x}}_k(t)$
Initialization: Data length h, parameter estimation accuracy ε, the maximum iteration $k_{\max}$, $\hat{\boldsymbol{a}}_0 = \hat{\boldsymbol{c}}_0 = \boldsymbol{1}_m/p_0$, $\hat{\boldsymbol{b}}_0 = \boldsymbol{1}_{m^2}/p_0$, $\hat{\boldsymbol{x}}_0(t) = \boldsymbol{1}_m/p_0$, $\boldsymbol{P}_0(1) = \boldsymbol{I}_m$, $p_0 = 10^6$.
for $k = 1 : k_{\max}$ **do**
  Form $\hat{\boldsymbol{\varphi}}_k(t)$ using Equation (28);
  Construct $\hat{\boldsymbol{\Phi}}_k(h)$ and $\boldsymbol{Y}(h)$ using Equation (27) and Equation (29);
  Compute $H[J_1(\hat{\boldsymbol{\theta}}_{k-1})]$ using Equation (26);
  Update $\hat{\boldsymbol{\theta}}_k$ using Equation (25);
  Read out $\hat{a}_{l,k}$, $\hat{\boldsymbol{b}}_{l,k}$ and $\hat{c}_{l,k}$ from $\hat{\boldsymbol{\theta}}_k$ using Equation (30);
  Construct $\hat{\boldsymbol{A}}_k$, $\hat{\boldsymbol{B}}_k$ and $\hat{\boldsymbol{C}}_k$ using Equations (36)–(40);
  for $t = 1 : h$ **do**
    Calculate $\boldsymbol{L}_k(t)$ and $\boldsymbol{P}_k(t+1)$ using Equation (34) and Equation (35);
    Calculate $\hat{\boldsymbol{x}}_k(t+1)$ using Equation (33) ;
  end
  if $\|\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_{k-1}\| > \varepsilon$ **then**
    $k = k + 1$;
  else
    Obtain $\hat{\boldsymbol{\theta}}_k$ and $\hat{\boldsymbol{x}}_k(t)$, break;
  end
end

5 Bilinear state estimator based GI algorithm

To highlight the advantages of the proposed BSO-NI algorithm, this section presents the bilinear state observer based GI (BSO-GI) algorithm for comparison.

When measurements contain missing inputs, advanced iterative strategies such as variable fractional order Nesterov accelerated gradient algorithms with interpolated models have been proposed to handle incomplete data effectively [20]. In the present work, complete measurement data collected from the system outputs are utilized to update the parameter estimates. Consider the $\boldsymbol{Y}(h)$ and the $\boldsymbol{\Phi}(h)$ defined in Section 3. Gradient-based iterative methods provide a natural starting point for parameter estimation; accelerated variants such as the Nesterov gradient algorithm have been shown to improve convergence for ARX-type models with random lost measurements [21]. Following the standard negative gradient search and minimizing $J_1(\boldsymbol{\theta})$ gives

$$\begin{aligned}\hat{\boldsymbol{\theta}}_k &= \hat{\boldsymbol{\theta}}_{k-1} - \mu_k \mathrm{grad}[J(\hat{\boldsymbol{\theta}}_{k-1})] \\ &= \hat{\boldsymbol{\theta}}_{k-1} + \mu_k\boldsymbol{\Phi}^{\mathrm{T}}(h)[\boldsymbol{Y}(h) - \boldsymbol{\Phi}(h)\hat{\boldsymbol{\theta}}_{k-1}],\end{aligned} \tag{41}$$

where the iterative step-size μ_k satisfies

$$\mu_k \leqslant 2\lambda_{\max}^{-1}[\boldsymbol{\Phi}^{\mathrm{T}}(h)\boldsymbol{\Phi}(h)].$$

Nevertheless, similar issues occur. The information matrix $\boldsymbol{\Phi}(h)$ includes the unknown terms $x_i(t-m)$

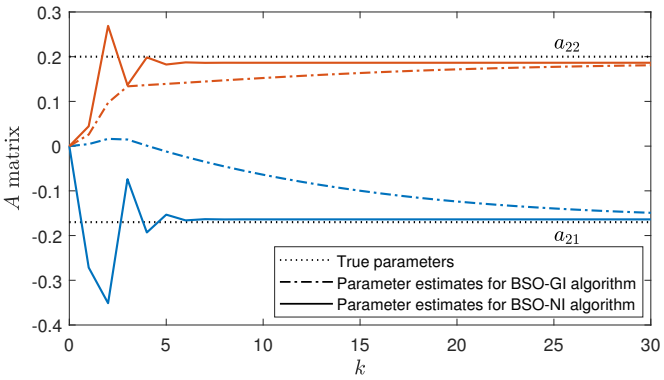


Figure 2. The $\boldsymbol{A}$ vector parameter estimates versus k for the BSO-NI and BSO-GI algorithms.

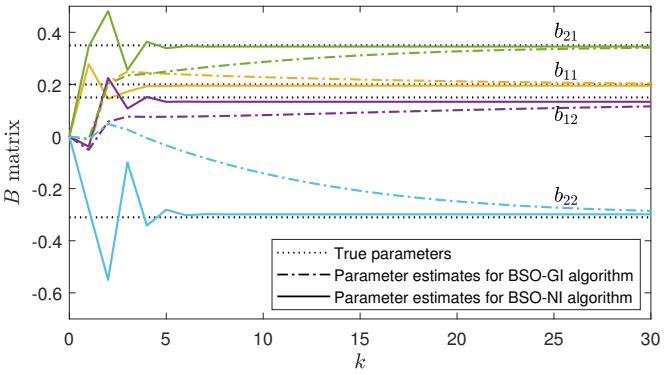


Figure 3. The $\boldsymbol{B}$ matrix parameter estimates versus k for the BSO-NI and BSO-GI algorithms.

and $\boldsymbol{x}(t-i)$. Following the similar method from the BSO-NI algorithm, we replace these unknowns with their estimates $\hat{x}_{i,k}(t-m)$ and $\hat{\boldsymbol{x}}_k(t-i)$. Following this, we can derive the estimated information matrix $\hat{\boldsymbol{\Phi}}_k(h)$. By substituting the unknown information matrix $\boldsymbol{\Phi}(h)$ in (41) with its estimate $\hat{\boldsymbol{\Phi}}_k(h)$, and referring to the bilinear state estimator in (33)–(40), the BSO-GI algorithm for bilinear systems, based on BSO, can be outlined as follows:

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_{k-1} + \mu_k \hat{\boldsymbol{\Phi}}_k^{\mathrm{T}}(h)\boldsymbol{E}_k(h), \tag{42}$$

$$\boldsymbol{E}_k(h) = \boldsymbol{Y}(h) - \hat{\boldsymbol{\Phi}}_k(h)\hat{\boldsymbol{\theta}}_{k-1}, \tag{43}$$

$$\boldsymbol{Y}(h) = [y(1), y(2), \cdots, y(h)]^{\mathrm{T}}, \tag{44}$$

$$\hat{\boldsymbol{\Phi}}_k(h) = [\hat{\boldsymbol{\varphi}}_k(1), \hat{\boldsymbol{\varphi}}_k(2), \cdots, \hat{\boldsymbol{\varphi}}_k(h)]^{\mathrm{T}}, \tag{45}$$

$$\begin{aligned}\hat{\boldsymbol{\varphi}}_k(t) = [&-\hat{x}_{m,k-1}(t-m), \cdots, -\hat{x}_{1,k-1}(t-m),\\ &\hat{\boldsymbol{x}}_{k-1}^{\mathrm{T}}(t-1)u(t-1), \cdots,\\ &\hat{\boldsymbol{x}}_{k-1}^{\mathrm{T}}(t-m)u(t-m), u(t-1),\\ &u(t-2), \cdots, u(t-m)]^{\mathrm{T}},\end{aligned} \tag{46}$$

$$0 < \mu_k \leqslant 2\lambda_{\max}^{-1}[\hat{\boldsymbol{\Phi}}_k^{\mathrm{T}}(h)\hat{\boldsymbol{\Phi}}_k(h)], \tag{47}$$

$$\begin{aligned}\hat{\boldsymbol{x}}_k(t+1) = &\hat{\boldsymbol{A}}_k\hat{\boldsymbol{x}}_k(t) + \hat{\boldsymbol{x}\boldsymbol{i}}_k(t)u(t)\\ &+ \boldsymbol{L}_k(t)[y(t) - \boldsymbol{D}\hat{\boldsymbol{x}}_k(t)],\end{aligned} \tag{48}$$

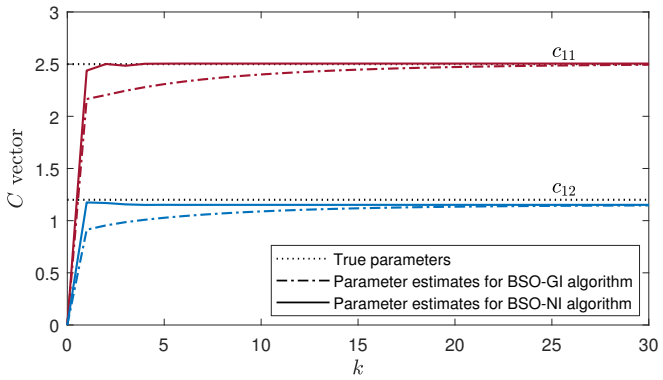


Figure 4. The $\boldsymbol{C}$ vector parameter estimates versus k for the BSO-NI and BSO-GI algorithms.

$$\boldsymbol{L}_k(t) = \frac{\hat{\boldsymbol{A}}_k\boldsymbol{P}_k(t)\boldsymbol{D}^{\mathrm{T}}}{1 + \boldsymbol{D}\boldsymbol{P}_k(t)\boldsymbol{D}^{\mathrm{T}}}, \tag{49}$$

$$\boldsymbol{P}_k(t+1) = \hat{\boldsymbol{A}}_k\boldsymbol{P}_k(t)\hat{\boldsymbol{A}}_k^{\mathrm{T}} - \boldsymbol{L}_k(t)\boldsymbol{D}\boldsymbol{P}_k(t)\hat{\boldsymbol{A}}_k^{\mathrm{T}}, \tag{50}$$

$$\hat{\boldsymbol{A}}_k = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & 1\\ -\hat{a}_{m,k} & -\hat{a}_{m-1,k} & -\hat{a}_{m-2,k} & \cdots & -\hat{a}_{1,k} \end{bmatrix}, \tag{51}$$

$$\hat{\boldsymbol{x}\boldsymbol{i}}_k(t) = \hat{\boldsymbol{B}}_k\hat{\boldsymbol{x}}_k(t) + \hat{\boldsymbol{C}}_k, \tag{52}$$

$$\hat{\boldsymbol{B}}_k = \begin{bmatrix} \hat{\boldsymbol{b}}_{1,k}\\ \hat{\boldsymbol{b}}_{2,k}\\ \vdots\\ \hat{\boldsymbol{b}}_{m,k} \end{bmatrix}, \tag{53}$$

$$\hat{\boldsymbol{C}}_k = \begin{bmatrix} \hat{c}_{1,k}\\ \hat{c}_{2,k}\\ \vdots\\ \hat{c}_{m,k} \end{bmatrix}. \tag{54}$$

Equations (42)–(54) form the BSO-GI algorithm. The algorithmic procedure for bilinear systems are outlined below:

1. Initialization: Set the data length h, the parameter estimation accuracy ε and the maximum iteration $k_{\max}$. Let $\hat{x}_{i,0}(t-m) = 1/p_0$, and $\hat{\boldsymbol{x}}_0(t-i) = \boldsymbol{1}_m/p_0$, $i = 1, 2, \ldots, m$, $\hat{\boldsymbol{\vartheta}}_0 = \boldsymbol{1}_{m^2+2m}/p_0$, $p_0 = 10^6$.
2. Set $k = 1$, gather the input-output data $u(t)$ and $y(t)$, where $t = 1, 2, \cdots, h$. Construct the stacked output vector $\boldsymbol{Y}(h)$ using Equation (44).
3. Construct $\hat{\boldsymbol{\varphi}}_k(t)$ using Equation (46), and form $\hat{\boldsymbol{\Phi}}_k(h)$ using Equation (45). Compute $\boldsymbol{E}_k(h)$ using Equation (43).
4. Update $\hat{\boldsymbol{\theta}}_k$ by Equation (42).

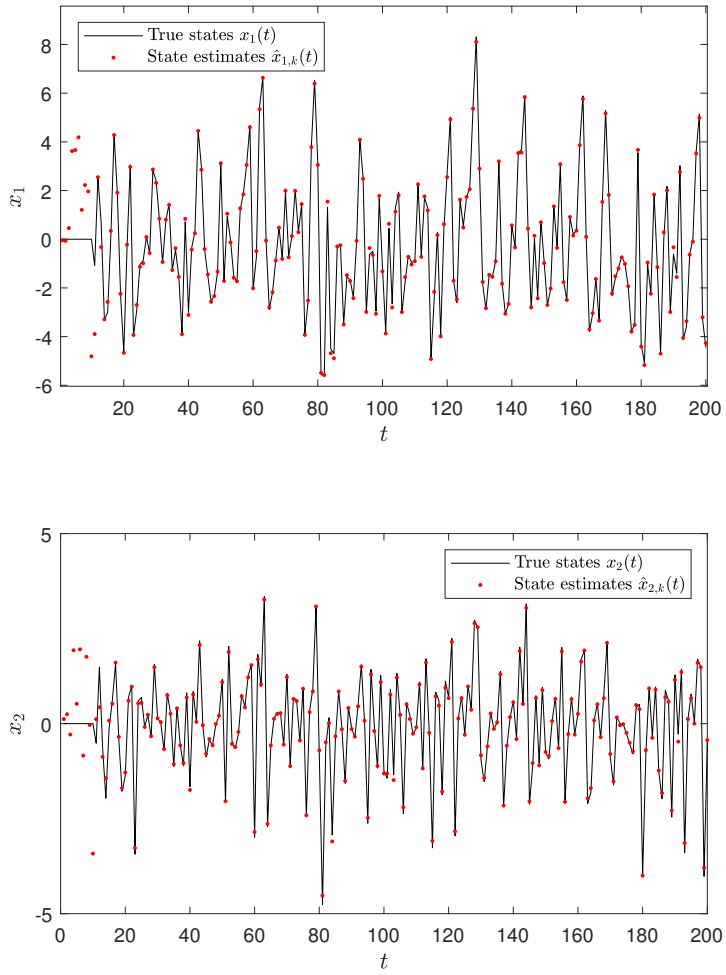


Figure 5. The true states and their estimates for the BSO-NI algorithm.

5. Extract $\hat{a}_{i,k}$, $\hat{\boldsymbol{b}}_{i,k}$ and $\hat{c}_{i,k}$ from $\hat{\boldsymbol{\theta}}_k$, construct $\hat{\boldsymbol{A}}_k$, $\hat{\boldsymbol{B}}_k$ and $\hat{\boldsymbol{C}}_k$ using Equations (51)–(54).
6. Set $t = 1$ and initialize $\hat{\boldsymbol{x}}_k(1) = \boldsymbol{1}_m$, $\boldsymbol{P}_k(1) = \boldsymbol{I}_m$.
7. Calculate $\boldsymbol{L}_k(t)$ and $\boldsymbol{P}_k(t+1)$ through Equations (49)–(50).
8. Compute $\hat{\boldsymbol{x}}_k(t+1)$ according to Equation (48).
9. If $t \leqslant h-1$, increment t by 1, return to Step 7.
10. If $\|\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_{k-1}\| > \varepsilon$, then, increment k by 1, return to Step 3; otherwise, obtain $\hat{\boldsymbol{\theta}}_k$ and $\hat{\boldsymbol{x}}_k(t)$, terminate.

6 Simulations

The following consider a second-order bilinear system defined by the equations:

$$\begin{aligned}\boldsymbol{x}(t+1) =& \begin{bmatrix} 0 & 1 \\ -0.30 & 0.17 \end{bmatrix} \boldsymbol{x}(t) + \begin{bmatrix} 2.00 \\ 1.20 \end{bmatrix} u(t) \\ &+ \begin{bmatrix} 0.10 & 0.08 \\ 0.35 & -0.21 \end{bmatrix} \boldsymbol{x}(t)u(t) + w(t), \\ y(t) =& [1,\ 0]\boldsymbol{x}(t) + v(t),\end{aligned}$$

The system parameters will be estimated are

$$\begin{aligned}\boldsymbol{\vartheta} &= [a_1, a_2, b_{11}, b_{12}, b_{21}, b_{22}, c_1, c_2]^{\mathrm{T}} \\ &= [-0.17, 0.30, 0.10, 0.08, 0.35, -0.21, 2.00, 1.20]^{\mathrm{T}}.\end{aligned}$$

For the simulation, the sequence $\{u(t)\}$ is set as an independent persistent excitation signal that has a mean of zero and unit variance, while $\{v(t)\}$ is modeled as a white noise sequence, exhibiting zero mean and variances $\sigma^2 = 0.50^2$, $\sigma^2 = 1.00^2$ and $\sigma^2 = 1.50^2$, respectively. We define h as 500 for the data length and establish a maximum of 30 iterations, represented by $k_{\max}$.

For the state estimation of this bilinear system, the BSO-NI algorithm is utilized to estimate each state.

The true state values and their estimates produced by the BSO-NI algorithm are shown in Figure 5. With the length of the data increasing, the state estimates tend toward the actual values.

Apply the BSO-NI algorithm Equations (25)–(30) and Equations (33)–(40) to estimate the parameters $\boldsymbol{\theta}$ and the states $\boldsymbol{x}$. For comparison, we also utilize the BSO-GI algorithm Equations (42)–(54) for the bilinear system estimation. The parameter estimation errors $\delta = \|\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}\|/\|\boldsymbol{\theta}\|$ for both algorithms are presented in Figure 1. From it, it is evident that the parameter estimation errors δ gradually decrease with the increase of k. Additionally, as the noise levels diminish, both algorithms yield more accurate parameter estimates. Under the same noise conditions, the BSO-NI algorithm demonstrates greater parameter estimation accuracy compared to the BSO-GI algorithm.

In order to better demonstrate the estimation effect of each parameter for the bilinear system, we use two algorithms to predict each parameter matrix or vector respectively, and show the comparison results in Figures 2–4. Figure 2 shows true parameters and the estimation results of the parameters in the $\boldsymbol{A}$ matrix. Figure 3 displays the parameter estimation values for two algorithms in the $\boldsymbol{B}$ matrix, and the parameter estimates in the $\boldsymbol{C}$ vector are clearly presented in Figure 4. These figures clearly illustrate that the BSO-NI algorithm outperforms the BSO-GI algorithm in terms of parameter estimation precision. Furthermore, the BSO-NI algorithm converges to the parameter estimates more rapidly.

7 Conclusion

This paper studies the combined estimation of states and parameters for bilinear systems. By means of the Newton search and batch data, we derive a Newton iterative (NI) algorithm to estimate the parameters that are not known. Drawing inspiration from the Kalman filtering, the bilinear state estimator (BSO) is created for estimating the unknown states, and a BSO based NI (BSO-NI) algorithm is proposed for the bilinear systems. In comparison to gradient-based algorithm, this NI algorithm achieves higher estimation accuracy. Simulation results illustrate the validity of the proposed method. The proposed BSO-NI framework provides a reliable model acquisition foundation for intelligent systems built upon bilinear processes, with direct applicability to model-based predictive control, adaptive fault detection, and data-driven decision support in industrial IoT and cyber-physical system environments. Future work will extend the algorithm to multi-input multi-output bilinear systems and investigate its integration with reinforcement learning-based intelligent controllers, further advancing its utility within the intelligent systematics paradigm.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported by the Young Scientists Fund of the National Natural Science Foundation of Zhejiang Province, China under Grant LQN25F030017.

Conflicts of Interest

Feng Ding served as an Associate Editor of *ICCK Transactions on Intelligent Systematics* at the time of manuscript submission. To ensure the integrity of the peer-review process, Feng Ding was not involved in the editorial handling, peer review, or decision-making process for this manuscript, which was handled independently by another editor. The remaining authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Ljung, L. (1999). *System identification: Theory for the user*. (2nd ed.). Prentice Hall, Englewood Cliffs, New Jersey. [CrossRef]

[2] Miao, G., Ding, F., Liu, Q., & Yang, E. (2023). Iterative parameter identification algorithms for transformed dynamic rational fraction input–output systems. *Journal of Computational and Applied Mathematics, 434*, 115297. [CrossRef]

[3] Benesty, J., Paleologu, C., Dogariu, L. M., & Ciochină, S. (2021). Identification of linear and bilinear systems: A unified study. *Electronics, 10*(15), 1790. [CrossRef]

[4] Li, F., Zhang, M., Yu, Y., & Li, S. (2024). Deep belief network-based hammerstein nonlinear system for wind power prediction. *IEEE Transactions on Instrumentation and Measurement, 73*, 1-12. 2024. 73:6505912. [CrossRef]

[5] Chen, J., Mao, Y., Gan, M., Wang, D., & Zhu, Q. (2023). Greedy search method for separable nonlinear models using stage Aitken gradient descent and least squares algorithms. *IEEE Transactions on Automatic Control, 68*(8), 5044-5051. [CrossRef]

[6] Chen, J., Ma, J., Gan, M., & Zhu, Q. (2022). Multidirection gradient iterative algorithm: A unified framework for gradient iterative and least squares algorithms. *IEEE Transactions on Automatic Control, 67*(12), 6770-6777. [CrossRef]

[7] Khodarahmi, M., & Maihami, V. (2023). A review on Kalman filter models. *Archives of Computational Methods in Engineering, 30*(1), 727-747. [CrossRef]

[8] An, S., He, Y., & Wang, L. (2023). Maximum likelihood based multi-innovation stochastic gradient identification algorithms for bilinear stochastic systems with ARMA noise. *International Journal of Adaptive Control and Signal Processing, 37*(10), 2690-2705. [CrossRef]

[9] Zhang, Q., Wang, H., & Liu, X. (2025). Auxiliary model maximum likelihood least squares-based iterative algorithm for multivariable autoregressive output-error autoregressive moving average systems. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 09596518241280323. [CrossRef]

[10] An, S., Wang, L., & He, Y. (2023). Filtering-based maximum likelihood hierarchical recursive identification algorithms for bilinear stochastic systems. *Nonlinear Dynamics, 111*(13), 12405-12420. [CrossRef]

[11] Gu, Y., Dai, W., Zhu, Q., & Nouri, H. (2023). Hierarchical multi-innovation stochastic gradient identification algorithm for estimating a bilinear state-space model with moving average noise. *Journal of Computational and Applied Mathematics, 420*, 114794. [CrossRef]

[12] Xu, L. (2021). Separable multi-innovation Newton iterative modeling algorithm for multi-frequency signals based on the sliding measurement window. *Circuits, Systems, and Signal Processing, 41*(2), 805-830. [CrossRef]

[13] Xu, L. (2022). Separable Newton recursive estimation method through system responses based on dynamically discrete measurements with increasing data length. *International Journal of Control, Automation and Systems, 20*(2), 432-443. [CrossRef]

[14] Pan, Z., Luan, X., & Liu, F. (2022). Confidence set-membership state estimation for LPV systems with inexact scheduling variables. *ISA Transactions, 122*, 38-48. [CrossRef]

[15] Pan, Z., Luan, X., & Liu, F. (2022). Set-membership state and parameter estimation for discrete time-varying systems based on the constrained zonotope. *International Journal of Control, 96*(12), 3226-3238. [CrossRef]

[16] Chen, J., Zhu, Q., & Liu, Y. (2020). Modified Kalman filtering based multi-step-length gradient iterative algorithm for ARX models with random missing outputs. *Automatica, 118*, 109034. [CrossRef]

[17] Pan, Z., Zhao, S., Huang, B., & Liu, F. (2023). Confidence set-membership FIR filter for discrete time-variant systems. *Automatica, 157*, 111231. [CrossRef]

[18] Pan, Z., & Liu, F. (2022). Nonlinear set-membership state estimation based on the Koopman operator. *International Journal of Robust and Nonlinear Control, 33*(4), 2703-2721. [CrossRef]

[19] Yang, D., Liu, Y., Ding, F., & Yang, E. (2023). Hierarchical gradient-based iterative parameter estimation algorithms for a nonlinear feedback system based on the hierarchical identification principle. *Circuits, Systems, and Signal Processing, 43*(1), 124-151. [CrossRef]

[20] Xu, F., Chen, J., Cheng, L., & Zhu, Q. (2025). Variable fractional order Nesterov accelerated gradient algorithm for rational models with missing inputs using interpolated model. *Mechanical Systems and Signal Processing, 224*, 111944. [CrossRef]

[21] Xu, F., Cheng, L., Chen, J., & Zhu, Q. (2024). The Nesterov accelerated gradient algorithm for auto-regressive exogenous models with random lost measurements: Interpolation method and auxiliary model method. *Information Sciences, 659*, 120055. [CrossRef]



Siyu Liu was born in Jilin Province, China. She received her B. Sc. degree from Jiangnan University (Wuxi, China) in 2017, and received her Ph. D. degree in Control Science and Engineering from Jiangnan University in 2024. She was a visiting Ph.D. student in the Department of Chemical and Materials Engineering at the University of Alberta from March 2021 to March 2023. She is currently an Associate Professor at Zhejiang Normal University. Her interests include system identification, process control and deep learning. (Email: siyu.liu@zjnu.edu.cn)



Feng Ding was born in Guangshui, Hubei Province. He received the B. Sc. degree in Electrical Engineering from the Hubei University of Technology (Wuhan, China) in 1984, and the M. Sc. and Ph. D. degrees in Automatic Control Theory and Applications from Tsinghua University (Beijing) in 1991 and 1994, respectively. He has been a Professor with Jiangnan University (Wuxi, China) since 2004. He has authored or coauthored the books Adaptive Control Systems (Tsinghua University Press, Beijing, 2002) and Modern Control Theory (Tsinghua University Press, Beijing, 2018), and five books on System Identification–New Theory and Methods 2013), System Identification–Performance Analysis for Identification Methods (2014), System Identification–Multi-Innovation Identification Theory and Methods (2016), System Identification–Auxiliary Model Identification Idea and Methods (2017), System Identification–Iterative Search Principle and Identification Methods (2018) at Science Press, Beijing, China. His research interests include model identification and adaptive control. (Email: fding@jiangnan.edu.cn)