Nonlinear Dyn
https://doi.org/10.1007/s11071-024-10613-y

RESEARCH

Separable synchronous redundant rule-based multi-innovation predictive gradient algorithms and convergence analysis for nonlinear ExpARX models

Ya Gu · Yuting Hou · Quanmin Zhu

Received: 5 July 2024 / Accepted: 5 November 2024

Abstract This paper studies the parameter identification and time-delay estimation problem for the nonlinear exponential autoregressive with exogenous input (ExpARX) models. To overcome the limitations of the traditional gradient algorithms, which have slow convergence and low identification accuracy, this paper proposes a modified predictive gradient algorithm through using the multi-innovation theory. Due to the extensive number of parameters, the time-delay ExpARX model is segmented into two subsystems by using the hierarchical principle. On the basis of the detached parameters, a modified separable synchronous predictive gradient algorithm is proposed. Moreover, the convergence of the proposed algorithm is proved. Through analyzing the computational efficiency, it has been demonstrated that the decomposition principle reduces computational workload and enhances computational efficiency. Finally, a simulation example and a real-life example of piezoelectric ceramics are used to verify the effectiveness of proposed algorithms.

Keywords ExpARX models · Convergence analysis · Redundant rule · Computational efficiency · Predictive gradient

Yuting Hou and Quanmin Zhu have contributed equally to this work.

corresponding author: Ya Gu.

Y. Gu (✉) · Y. Hou
College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 201418, People's Republic of China
e-mail: guya@shnu.edu.cn

Y. Hou
e-mail: 1000527989@smail.shnu.edu.cn

Q. Zhu
Department of Engineering Design and Mathematics, University of the West of England, Bristol BS16 1QY, UK
e-mail: Quan.Zhu@uwe.ac.uk

1 Introduction

In control systems, precise mathematical models are essential for analysis and optimization, but the complexity presents a significant challenge [1–3]. System identification technology uses measurement data to solve optimization problems, minimizing input and output errors to construct parametric models [4–6]. With advancements in automation technology, system identification theory is widely applied in power systems, adaptive control systems, fault diagnosis, and wireless communication systems [7–9]. Recently, various identification algorithms have been proposed for different systems, including bilinear and nonlinear systems [10–12] and nonlinear systems [13–15].

In piezoelectric ceramic flow control systems, system identification technology plays a crucial role in constructing highly accurate nonlinear models [16]. By employing advanced system identification algorithms, measurement data can be effectively utilized

Published online: 18 November 2024

to extract the core parameters of the system, thereby allowing for the construction of a nonlinear model that better adapts to flow fluctuations and external disturbances [17]. Given the unique dynamic characteristics of piezoelectric ceramics and the complex nonlinear behavior during the flow control process, traditional linear models often struggle to adequately and accurately depict the actual response of the system [18].

In recent years, several nonlinear systems capable of accurately capturing random vibrations have been successfully derived [19–21], among which the most representative is the exponential autoregressive with exogenous input (ExpARX) model. The ExpARX model extends the traditional ARX framework and is suitable for time series data with exponential growth or decay. It effectively describes dynamic systems with delay and step characteristics, particularly capturing nonlinear behaviors like limit cycles and amplitude-dependent frequency [22]. For exponential-type systems, various identification algorithms have been studied, including maximum likelihood, stochastic gradient, and variable projection algorithms [23,24]. However, most of contributions overlook the inevitable intrinsical time-delay in the data transmission process, which increases the complexity.

The separable least squares method is crucial for identifying linear and nonlinear models, especially in complex model identification. It decomposes parameters into linear and nonlinear parts, reducing complexity and improving computational efficiency. Thus, it is also known as hierarchical identification [25–27]. By separating parameters or models based on their characteristics, this method reduces model scale and enables more accurate joint estimation [28].

Traditional gradient identification algorithms, which include stochastic gradient algorithms, gradient iterative algorithms and gradient descent algorithms, are widely used in the system identification [29,30]. Traditional gradient algorithms may get stuck in local optima, with identification accuracy significantly influenced by initial parameters and learning rate settings [31]. Compared to previous work [32,33], the proposed algorithms effectively overcome the limitations of traditional gradient methods and address parameter identification for the time-delay ExpARX model. The main contributions are as follows.

- To overcome the limitations of the traditional gradient algorithms with slow convergence speed, an innovative threshold-free algorithm is proposed for estimating the time-delay, without relying on the empirical threshold selection used in the traditional algorithms.
- To improve the identification accuracy, a redundant rule-based multi-innovation predictive gradient (RR-MIPG) algorithm is proposed. Unlike the traditional gradient algorithms, the RR-MIPG algorithm effectively avoids local optima and addresses parameter identification issues in nonlinear ExpARX models.
- To reduce computational workload and enhance computational efficiency, a separable synchronous redundant rule-based multi-innovation predictive gradient (SS-RR-MIPG) algorithm is proposed.

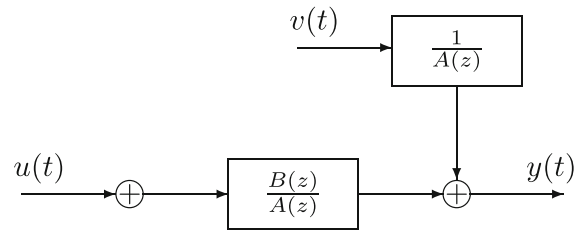

Fig. 1 The ExpARX model

The structure of this paper is outlined as follows. Section 2 gives the system description. Section 3 presents the RR-MIPG algorithm. Section 4 proposes the SS-RR-MIPG algorithm. Section 5 analyzes the computational amounts. The convergence property is analyzed in Sect. 6. In Sect. 7, two examples are provided. Finally, some conclusions are given in Sect. 8.

2 System description

Consider the nonlinear time-delay exponential autoregressive with exogenous input (ExpARX) model depicted in Fig. 1,

where $u(t)$ and $y(t)$ are the input and output of the system, $v(t)$ is a white noise with zero mean, and $A(z)$ and $B(z)$ are the polynomials of the operator z^{-1}, defined as

$$A(z) = 1 + \sum_{i=1}^{n_a} (a_{i,1} + a_{i,2}\mathrm{e}^{-\gamma y^2(t-1)}) z^{i-n_a-1},$$

$$B(z) = \sum_{j=1}^{n_b} (b_{j,1} + b_{j,2}\mathrm{e}^{-\gamma y^2(t-1)}) z^{j-n_a-1}.$$

The time-delay ExpARX model is

$$\begin{aligned} y(t) = &-\sum_{i=1}^{n_\xi}(\xi_{i,1}+\xi_{i,2}\mathrm{e}^{-\gamma y^2(t-1)}) \\ &\times y(t+i-n_\xi-1-\tau) \\ &+\sum_{j=1}^{n_\eta}(\eta_{j,1}+\eta_{j,2}\mathrm{e}^{-\gamma y^2(t-1)}) \\ &\times u(t+j-n_\eta-1-\tau)+v(t), \end{aligned} \tag{1}$$

where $\xi_{i,j}$ are linear parameters, $\eta_{i,j}$ are nonlinear parameters, γ is an exponential parameter, n_ξ and n_η are the model orders, and τ is an unknown time-delay.

Define the parameters as

$$\begin{aligned} \boldsymbol{\vartheta} &:= [\boldsymbol{\theta}^{\mathrm{T}}, \gamma]^{\mathrm{T}} \in \mathbb{R}^{2n_\xi+2n_\eta+1}, \\ \boldsymbol{\theta} &:= [\boldsymbol{\theta}_y^{\mathrm{T}}, \boldsymbol{\theta}_u^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{2n_\xi+2n_\eta}, \\ \boldsymbol{\theta}_y &:= [\xi_{1,1}, \xi_{2,1}, \ldots, \xi_{n_\xi,1}, \xi_{1,2}, \xi_{2,2}, \ldots, \xi_{n_\xi,2}]^{\mathrm{T}} \\ &\quad \in \mathbb{R}^{2n_\xi}, \\ \boldsymbol{\theta}_u &:= [\eta_{1,1}, \eta_{2,1}, \ldots, \eta_{n_\eta,1}, \eta_{1,2}, \eta_{2,2}, \ldots, \eta_{n_\eta,2}]^{\mathrm{T}} \\ &\quad \in \mathbb{R}^{2n_\eta}. \end{aligned}$$

Define the information vectors as

$$\begin{aligned} \boldsymbol{\phi}(t,\gamma,\tau) &:= [\boldsymbol{\varphi}_y^{\mathrm{T}}(t,\gamma,\tau), \mathrm{e}^{-\gamma y^2(t-1)}\boldsymbol{\varphi}_y^{\mathrm{T}}(t,\gamma,\tau), \\ &\quad \boldsymbol{\varphi}_u^{\mathrm{T}}(t,\gamma,\tau), \mathrm{e}^{-\gamma y^2(t-1)}\boldsymbol{\varphi}_u^{\mathrm{T}}(t,\gamma,\tau)]^{\mathrm{T}} \\ &\quad \in \mathbb{R}^{2n_\xi+2n_\eta}, \\ \boldsymbol{\varphi}_y(t,\gamma,\tau) &:= [-y(t-n_\xi-\tau), \ldots, -y(t-1-\tau)]^{\mathrm{T}} \\ &\quad \in \mathbb{R}^{n_\xi}, \\ \boldsymbol{\varphi}_u(t,\gamma,\tau) &:= [u(t-n_\eta-\tau), \ldots, \\ &\quad u(t-n_\xi+n_\eta-1-\tau)]^{\mathrm{T}} \in \mathbb{R}^{n_\eta}. \end{aligned}$$

Then, Eq. (1) can be compactly written as the identification model

$$y(t) = \boldsymbol{\phi}^{\mathrm{T}}(t,\gamma,\tau)\boldsymbol{\theta} + v(t). \tag{2}$$

Note that the information vector $\boldsymbol{\phi}(t,\gamma,\tau)$ contains the variables $y(t+i-n_a-1-\tau)$, $i=1,2,\ldots,n_a$ and $u(t+j-n_b-1-\tau)$, $j=1,2,\ldots,n_b$, which cannot be determined due to the unknown time-delay τ. Hence, the traditional algorithms cannot be used for the time-delay ExpARX model. To address the difficulty, this paper sets a maximum regression length N, extends the redundant parameters to be identified together with the parameter vector, and employs the redundant rule to derive an augmented identification model.

Define the augmented parameter vectors

$$\begin{aligned} \boldsymbol{\vartheta}_a &:= [\boldsymbol{\theta}_a^{\mathrm{T}}, \gamma]^{\mathrm{T}} \in \mathbb{R}^{4N+1}, \\ \boldsymbol{\theta}_a &:= [\boldsymbol{\theta}_{ya}^{\mathrm{T}}, \boldsymbol{\theta}_{ua}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{4N}, \\ \boldsymbol{\theta}_{ya} &:= [\alpha_1, \ldots, \alpha_\tau, \xi_{1,1}, \xi_{2,1}, \ldots, \\ &\quad \xi_{n_\xi,1}, \xi_{1,2}, \xi_{2,2}, \ldots, \xi_{n_\xi,2}, \\ &\quad \alpha_{\tau+1}, \ldots, \alpha_{2N-2n_a}]^{\mathrm{T}} \in \mathbb{R}^{2N}, \\ \boldsymbol{\theta}_{ua} &:= [\beta_1, \ldots, \beta_\tau, \eta_{1,1}, \eta_{2,1}, \ldots, \\ &\quad \eta_{n_\eta,1}, \eta_{1,2}, \eta_{2,2}, \ldots, \eta_{n_\eta,2}, \\ &\quad \beta_{\tau+1}, \ldots, \beta_{2N-2n_b}]^{\mathrm{T}} \in \mathbb{R}^{2N}, \end{aligned}$$

and the augmented information vectors

$$\begin{aligned} \boldsymbol{\phi}_a(t,\gamma) &:= [\boldsymbol{\varphi}_{ya}^{\mathrm{T}}(t,\gamma), \mathrm{e}^{-\gamma y^2(t-1)}\boldsymbol{\varphi}_{ya}^{\mathrm{T}}(t,\gamma), \\ &\quad \boldsymbol{\varphi}_{ua}^{\mathrm{T}}(t,\gamma), \mathrm{e}^{-\gamma y^2(t-1)}\boldsymbol{\varphi}_{ua}^{\mathrm{T}}(t,\gamma)]^{\mathrm{T}} \in \mathbb{R}^{4N}, \\ \boldsymbol{\varphi}_{ya}(t,\gamma) &:= [-y(t-N), \ldots, -y(t-1)]^{\mathrm{T}} \in \mathbb{R}^{N}, \\ \boldsymbol{\varphi}_{ua}(t,\gamma) &:= [u(t-N), \ldots, u(t-1)]^{\mathrm{T}} \in \mathbb{R}^{N}. \end{aligned}$$

Then, Eq. (2) can be written as

$$y(t) = \boldsymbol{\phi}_a^{\mathrm{T}}(t,\gamma)\boldsymbol{\theta}_a + v(t). \tag{3}$$

The proposed parameter estimation algorithms in this paper are based on the parameter identification model in (3). Many identification methods are derived based on the identification models of systems [34–37] and can be used to estimate the parameters of other linear stochastic systems and nonlinear stochastic systems [38–41] and can be applied to other fields [42–46] such as information processing and process control systems. In practical applications, the observation model inevitably includes a non-zero noise term. The characteristic prevents the estimations of redundant parameters from converging precisely to zero. The phenomenon can lead to erroneous delay estimations because the interference from non-zero noise disrupts the accurate identification of delay parameters. As a result, it affects the overall predictive accuracy and reliability of the model.

3 Redundant rule-based multi-innovation predictive gradient algorithm

In dealing with the highly nonlinear dependence on the parameter γ and the complexity introduced by the unknown time-delay τ in (3), traditional algorithms struggle to find direct solutions. To address the challenge innovatively, this section proposes the RR-MIPG algorithm. The redundant rule-based multi-innovation predictive gradient (RR-MIPG) algorithm not only effectively tackles the nonlinearities within the model, but also accurately captures and manages the unknown

delay, thereby achieving precise estimation of model parameters.

Define the output vector as

$$\boldsymbol{Y}(p) := [y(t), y(t-1), \ldots, y(t-p+1)]^{\mathrm{T}} \in \mathbb{R}^{p},$$

where p is the innovation length.

Define the information matrices as

$$\begin{aligned}
\boldsymbol{\Psi}_a(p) &:= [\boldsymbol{\phi}_a(t,\gamma), \boldsymbol{\phi}_a(t-1,\gamma), \ldots, \\
&\quad \boldsymbol{\phi}_a(t-p+1,\gamma)] \in \mathbb{R}^{4N\times p}, \\
\boldsymbol{\Phi}_{ya}(p) &:= [\boldsymbol{\varphi}_{ya}(t,\gamma), \boldsymbol{\varphi}_{ya}(t-1,\gamma), \ldots, \\
&\quad \boldsymbol{\varphi}_{ya}^{\mathrm{T}}(t-p+1,\gamma)] \in \mathbb{R}^{2N\times p}, \\
\boldsymbol{\Phi}_{ua}(p) &:= [\boldsymbol{\varphi}_{ua}(t,\gamma), \boldsymbol{\varphi}_{ua}(t-1,\gamma), \ldots, \\
&\quad \boldsymbol{\varphi}_{ua}(t-p+1,\gamma)] \in \mathbb{R}^{2N\times p}.
\end{aligned}$$

To determine the strategy of the RR-MIPG algorithm [47], define the criterion function

$$J(\boldsymbol{\vartheta}_a) = \frac{1}{2}\boldsymbol{\chi}^2(t) + \frac{\lambda(t)}{2} \parallel \Delta\boldsymbol{\vartheta}_a(t) \parallel^2, \tag{4}$$

where $\boldsymbol{\chi}(t)$ is the output error, $\Delta\boldsymbol{\vartheta}_a(t) = \boldsymbol{\vartheta}_a(t) - \boldsymbol{\vartheta}_a(t-1)$, and $\lambda(t)$ is the weighting factor of model based compensation.

Let the gradient of the criterion function with respect to $\boldsymbol{\vartheta}_a(t)$ equal zero yields to get

$$\boldsymbol{\vartheta}_a(t) = \boldsymbol{\vartheta}(t-1) - r(t)\boldsymbol{\chi}(t)\boldsymbol{\chi}'(t), \tag{5}$$

where $r(t) = 1/\lambda(t)$ and $\boldsymbol{\chi}'(t) = \frac{\partial \boldsymbol{\chi}(t)}{\partial \boldsymbol{\vartheta}_a}$.

Define the gradient $\mathbf{g}(t)$ as

$$\mathbf{g}(t) := -\boldsymbol{\Psi}_a(p)[\boldsymbol{Y}(p) - \boldsymbol{\Psi}_a^{\mathrm{T}}(p)\boldsymbol{\theta}_a]. \tag{6}$$

Similar to the memory gradient algorithm [48], define the descent direction as

$$\boldsymbol{\psi}(t) = \begin{cases} -\mathbf{g}(t), & t = 1, \\ -\mathbf{g}(t) + \zeta(t)[\boldsymbol{\psi}(t-1) & \\ -\mathbf{g}(t-1)], & \text{otherwise}, \end{cases} \tag{7}$$

where

$$\zeta(t) = \frac{\sigma_\zeta \parallel \mathbf{g}(t) \parallel}{\parallel \boldsymbol{\psi}(t-1) - \mathbf{g}(t-1) \parallel},$$

and $\sigma_\zeta > 0$ is a constant.

Based on the above equation, the predictive gradient is constructed by the following autoregressive formula

$$\varepsilon(t) = H(z^{-1})\mathbf{g}(t), \tag{8}$$

where

$$H(z^{-1}) = 1 + h_1 z^{-1} + \cdots + h_n z^{-n},$$

where $n = n_\xi = n_\eta$, and the coefficients of $H(z^{-1})$ are determined by $h_1 = 1$, $h_2 = 2\zeta(t)$, …, $h_n = 2\prod_{k=0}^{n-2}\zeta(t-k)$.

Then, introducing the Diophantine equation

$$H(z^{-1})P(z^{-1}) + z^{-k}Q(z^{-1}) = 1, \tag{9}$$

where

$$\begin{aligned}
P(z^{-1}) &= p_0 + p_1 z^{-1} + \cdots + p_{d-1} z^{-k+1}, \\
Q(z^{-1}) &= q_0 + q_1 z^{-1} + \cdots + q_{n-1} z^{-n+1}.
\end{aligned}$$

Then, substituting (8) into (9) to derive the k-step-ahead predictive gradient

$$\hat{\mathbf{g}}(t+k) = \sum_{l=0}^{n-1} q_l \hat{\mathbf{g}}(t-l). \tag{10}$$

From (6)–(7) and (10), it can be deduced that the gradient at time $t+k$ is predicted solely based on the gradient values acquired before time t. The predictive descent direction can be derived as

$$\hat{\boldsymbol{\psi}}(t+k) = -\hat{\mathbf{g}}(t+k) + \sum_{l=1}^{n-1} \zeta_l(t)\hat{\boldsymbol{\psi}}(t+k-l), \tag{11}$$

$$\zeta_l(t) = \frac{\sigma_\zeta \parallel \hat{\mathbf{g}}(t+k) \parallel}{\parallel \hat{\boldsymbol{\psi}}(t+k-l) - \hat{\mathbf{g}}(t+k-l) \parallel}. \tag{12}$$

Remark 1 Compared with the traditional gradient algorithms, the RR-MIPG algorithm can quickly escape the current search path and flexibly adjust its direction when it detects that the current trajectory may lead to a local optima by predicting the gradient direction $t+k$ steps ahead. The gradient prediction capability allows the RR-MIPG algorithm to demonstrate greater efficiency and stronger global search ability in complex and dynamic optimization problems.

An optimization step-size is employed based on the Goldstein inexact search.

$$r(t) = \frac{-\rho\hat{\mathbf{g}}^{\mathrm{T}}(t+k)\hat{\boldsymbol{\psi}}(t+k)}{L \parallel \hat{\boldsymbol{\psi}}(t+k) \parallel^2}, \tag{13}$$

where $\rho > 0$ is the convergence factor,

$$L = \max\left\{L^*, \frac{\parallel \hat{\mathbf{g}}(t+k) - \hat{\mathbf{g}}(t+k-1) \parallel}{\parallel \hat{\boldsymbol{\vartheta}}_a(t) - \hat{\boldsymbol{\vartheta}}_a(t-1) \parallel + \Lambda}\right\} \tag{14}$$

is the *Lipschitz* constant, L^* is a positive constant used to prevent L from being zero, and Λ is a very small positive number to ensure the denominator not to be zero.

When $0 < t \leqslant k$, the proposed algorithm fails to effectively compute $\hat{\psi}(t+k)$ and $\hat{\mathbf{g}}(t+k)$, resulting in inaccurate outcomes. To address the limitation and

ensure data stability and integrity, simplified processing measures have been implemented

$$r(t) = \frac{-\rho \hat{\mathbf{g}}^{\mathrm{T}}(t)\hat{\boldsymbol{\psi}}(t)}{L^* \parallel \hat{\boldsymbol{\psi}}(t) \parallel^2}. \tag{15}$$

Remark 2 The aforementioned adjustments prevent potential risks caused by data loss in the early stage of the algorithm, enhancing reliability in practical applications and ensuring stable operation.

For the time-delay threshold estimation algorithms, the identification performance is affected by the choice of threshold, but currently it is no universally accepted method for selecting the threshold.

Define

$$\beth(\tau) := \sum_{s=\tau+1}^{\tau+n} [\hat{\boldsymbol{\theta}}_{a,s}^2(L) + \hat{\boldsymbol{\theta}}_{a,s+2N}^2(L)].$$

The time-delay estimation is

$$\hat{\tau} = \underset{\tau \in 0,\ldots,\tau_{\max}}{\operatorname{argmax}} \beth(\tau), \tag{16}$$

where $\tau_{\max} = N - n$.

In summary, the RR-MIPG algorithm is

$$\hat{\boldsymbol{\vartheta}}_a(t) = \hat{\boldsymbol{\vartheta}}_a(t-1) - r(t)\hat{\mathbf{g}}(t+k), \tag{17}$$

$$\hat{\mathbf{g}}(t+k) = \sum_{l=0}^{n-1} q_l \hat{\mathbf{g}}(t-l), \tag{18}$$

$$\hat{\boldsymbol{\psi}}(t+k) = -\hat{\mathbf{g}}(t+k) + \sum_{l=1}^{n-1} \zeta_l(t)\hat{\boldsymbol{\psi}}(t+k-l), \tag{19}$$

$$\zeta_l(t) = \frac{\sigma_\zeta \parallel \hat{\mathbf{g}}(t+k) \parallel}{\parallel \hat{\boldsymbol{\psi}}(t+k-l) - \hat{\mathbf{g}}(t+k-l) \parallel}, \tag{20}$$

$$r(t) = \frac{-\rho \hat{\mathbf{g}}^{\mathrm{T}}(t+k)\hat{\boldsymbol{\psi}}(t+k)}{L \parallel \hat{\boldsymbol{\psi}}(t+k) \parallel^2}, \tag{21}$$

$$L = \max\left\{L^*, \frac{\parallel \hat{\mathbf{g}}(t+k) - \hat{\mathbf{g}}(t+k-1) \parallel}{\parallel \hat{\boldsymbol{\vartheta}}_a(t) - \hat{\boldsymbol{\vartheta}}_a(t-1) \parallel + \Lambda}\right\}, \tag{22}$$

$$\hat{\tau} = \underset{\tau \in 0,\ldots,\tau_{\max}}{\operatorname{argmax}} \beth(\tau). \tag{23}$$

Remark 3 Before using algorithms proposed in this paper, it is necessary to accurately ascertain the order of the system through methods such as orthogonalization process, correlation analysis, and other order estimation techniques to guarantee the accuracy and effectiveness of the algorithm.

Algorithm 1 summarizes the steps for computing the estimates $\hat{\boldsymbol{\vartheta}}_a(t)$ by the RR-MIPG algorithm.

Algorithm 1 RR-MIPG algorithm

Require: Set the maximum iteration t_{max}, time-delay τ, prediction horizon $k = 2$.
Ensure: $\hat{\boldsymbol{\vartheta}}_a(t)$

```
for t = 1 do
  if 0 < t ⩽ d then
    Compute ĝ(t) by (6)
    Update ψ̂(t) by (7)
    Compute r(t) by (15)
  else[t > d]
    Compute ĝ(t+k), ψ̂(t+k) and ζ_l(t) by (18), (19) and (20)
    Update r(t) and L by (21) and (22)
  end if
  Update ϑ̂_a(t) by (17)
  if ‖ϑ̂(t) − ϑ̂(t−1)‖ > ε_1 then
    t := t + 1
  else
    break
  end if
  Compute τ̂(t) by (23)
  if ‖τ̂ − τ‖ = 0 then
    for ι = 1 do
      α_ι = 0
      β_ι = 0
      if ι < τ then
        ι := ι + 1
      else
        break
      end if
    end for
  end if
end for
Update ϑ̂_a(t)
```

4 Separable synchronous redundant rule-based multi-innovation predictive gradient algorithm

In this section, the time-delay the exponential autoregressive with exogenous input (ExpARX) model is separated into two subsystems. Based on the obtained subsystems, the separable synchronous redundant rule-based multi-innovation predictive gradient (SS-RR-MIPG) algorithm is derived.

Define the augmented parameters

$$\begin{aligned}\boldsymbol{\theta}_{1a} &:= [\alpha_1, \ldots, \alpha_\tau, \xi_{1,1}, \ldots, \\ &\xi_{n_\xi,1}, \alpha_{\tau+1}, \ldots, \alpha_{2N-n_\xi-n_\eta}, \eta_{1,1}, \ldots, \eta_{n_\eta,1}]^{\mathrm{T}} \in \mathbb{R}^{2N}, \\ \boldsymbol{\theta}_{2a} &:= [\xi_{1,2}, \ldots, \xi_{n_\xi,2}, \beta_1, \ldots, \\ &\beta_\tau, \eta_{1,2}, \ldots, \eta_{n_\eta,2}, \beta_{\tau+1}, \ldots, \beta_{2N-n_\xi-n_\eta}]^{\mathrm{T}} \in \mathbb{R}^{2N}.\end{aligned}$$

Define the augmented information vector

$$\begin{aligned}\boldsymbol{\varphi}_{1a}(t) := [&-y(t-N), \ldots, -y(t-1), u(t-N), \\ &\ldots, u(t-1)]^{\mathrm{T}} \in \mathbb{R}^{2N},\end{aligned}$$

Define

$$\boldsymbol{\Psi}_{1a}(p) := [\boldsymbol{\phi}_{1a}(t,\gamma), \boldsymbol{\phi}_{1a}(t-1,\gamma), \ldots, \boldsymbol{\phi}_{1a}(t-p+1,\gamma)] \in \mathbb{R}^{2N\times p}.$$

Eq. (2) can be written as

$$\boldsymbol{Y}(p) = \boldsymbol{\Phi}_{1a}^{\mathrm{T}}(p)\boldsymbol{\theta}_{1a} + \mathrm{e}^{-\gamma y^2(t-1)}\boldsymbol{\Phi}_{1a}^{\mathrm{T}}(p)\boldsymbol{\theta}_{2a} + v(t). \tag{24}$$

Define the fictitious output $\boldsymbol{Y}_1(t)$ and $\boldsymbol{\Gamma}(\boldsymbol{\theta}_{2a})$ as

$$\begin{aligned} \boldsymbol{Y}_1(p) &:= \boldsymbol{Y}(t) - \boldsymbol{\Phi}_{1a}^{\mathrm{T}}(p)\boldsymbol{\theta}_{1a}, \\ \boldsymbol{\Gamma}(\boldsymbol{\theta}_{2a}) &:= \boldsymbol{\Phi}_{1a}^{\mathrm{T}}(p)\boldsymbol{\theta}_{2a}. \end{aligned} \tag{25}$$

From (24), the ExpARX model is separated into two subsystems

$$\begin{aligned} \boldsymbol{Y}(p) &= \boldsymbol{\Psi}_a^{\mathrm{T}}(p)\boldsymbol{\theta}_a + v(t), \\ \boldsymbol{Y}_1(p) &= \mathrm{e}^{-\gamma y^2(t-1)}\boldsymbol{\Gamma}(\boldsymbol{\theta}_{2a}) + v(t). \end{aligned}$$

The time-delay ExpARX model is shown in Fig. 2.

Define two criterion functions

$$\begin{aligned} J(\boldsymbol{\theta}_a) &= \frac{1}{2}\boldsymbol{\chi}_1^2(t) + \frac{\lambda_1(t)}{2} \parallel \Delta\boldsymbol{\theta}_a(t) \parallel^2, \\ J(\gamma) &= \frac{1}{2}\boldsymbol{\chi}_2^2(t) + \frac{\lambda_2(t)}{2} \parallel \Delta\gamma(t) \parallel^2. \end{aligned}$$

Remark 4 Due to numerous parameters that need to be estimated, it is possible to classify all parameters into linear and nonlinear groups and redefine the criterion function based on the classification. The separable methods are advantageous for handling models with both linear and nonlinear relationships by reducding parameter dimensions and algorithmic complexity. The methods are particularly effective in lowering computational complexity for Newton algorithms and least squares algorithms.

Similarly, the gradients of $J_1(\boldsymbol{\theta}_a)$ and $J_2(\gamma)$ are

$$\mathbf{g}_1(t) := -\boldsymbol{\Psi}_a(p)[\boldsymbol{Y}(p) - \boldsymbol{\Psi}_a^{\mathrm{T}}(p)\boldsymbol{\theta}_a], \tag{26}$$

$$g_2(t) := -\boldsymbol{\theta}_a^{\mathrm{T}}\boldsymbol{\Psi}_a'(p)[\boldsymbol{Y}(p) - \boldsymbol{\Psi}_a^{\mathrm{T}}(p)\boldsymbol{\theta}_a]. \tag{27}$$

Define the descent direction separately as

$$\boldsymbol{\psi}_1(t) = \begin{cases} -\mathbf{g}_1(t), & t = 1, \\ -\mathbf{g}_1(t) + \zeta_1(t)[\boldsymbol{\psi}_1(t-1) \\ -\mathbf{g}_1(t-1)], & \text{otherwise}, \end{cases} \tag{28}$$

$$\psi_2(t) = \begin{cases} -g_2(t), & t = 1, \\ -g_2(t) + \zeta_2(t)[\psi_2(t-1) \\ -g_2(t-1)], & \text{otherwise}, \end{cases} \tag{29}$$

where

$$\zeta_{l_1}(t) = \frac{\sigma_1 \parallel \hat{\mathbf{g}}_1(t) \parallel}{\parallel \hat{\boldsymbol{\psi}}_1(t-1) - \hat{\mathbf{g}}_1(t-1) \parallel},$$

$$\zeta_{l_2}(t) = \frac{\sigma_2 \parallel \hat{g}_2(t) \parallel}{\parallel \hat{\psi}_2(t-1) - \hat{g}_2(t-1) \parallel}.$$

The associated predictive descent direction can be determined separately as

$$\hat{\mathbf{g}}_1(t+k) = \sum_{l_1=0}^{n-1} q_{l_1}\hat{\mathbf{g}}_1(t-l_1),$$

$$\hat{\boldsymbol{\psi}}_1(t+k) = -\hat{\mathbf{g}}_1(t+k) + \sum_{l_1=1}^{n-1} \zeta_{l_1}(t)\hat{\boldsymbol{\psi}}_1(t+k-l_1),$$

$$\zeta_{l_1}(t) = \frac{\sigma_1 \parallel \hat{\mathbf{g}}_1(t+k) \parallel}{\parallel \hat{\boldsymbol{\psi}}_1(t+k-l) - \hat{\mathbf{g}}_1(t+k-l_1) \parallel},$$

$$\hat{g}_2(t+k) = \sum_{l_2=0}^{n-1} q_{l_2}\hat{g}_2(t-l_2),$$

$$\hat{\psi}_2(t+k) = -\hat{\mathbf{g}}_1(t+k) + \sum_{l_2=1}^{n-1} \zeta_{l_2}(t)\hat{\psi}_2(t+k-l_2),$$

$$\zeta_{l_2}(t) = \frac{\sigma_2 \parallel \hat{g}_2(t+k) \parallel}{\parallel \hat{\psi}_2(t+k-l_2) - \hat{g}_2(t+k-l_2) \parallel}.$$

The step-sizes are the same as (13)

$$r_1(t) = \frac{-\rho_1\hat{\mathbf{g}}_1^{\mathrm{T}}(t+k)\hat{\boldsymbol{\psi}}_1(t+k)}{L_1 \parallel \hat{\boldsymbol{\psi}}_1(t+k) \parallel^2},$$

$$L_1 = \max\left\{L_1^*, \frac{\parallel \hat{\mathbf{g}}_1(t+k) - \hat{\mathbf{g}}_1(t+k-1) \parallel}{\parallel \hat{\boldsymbol{\theta}}_a(t) - \hat{\boldsymbol{\theta}}_a(t-1) \parallel + \Lambda_1}\right\},$$

$$r_2(t) = \frac{-\rho_2\hat{g}_2(t+k)\hat{\psi}_2(t+k)}{L_2 \parallel \hat{\psi}_2(t+k) \parallel^2},$$

$$L_2 = \max\left\{L_2^*, \frac{\parallel \hat{g}_2(t+k) - \hat{g}_2(t+k-1) \parallel}{\parallel \hat{\gamma}(t) - \hat{\gamma}(t-1) \parallel + \Lambda_2}\right\}.$$

Similarly, when $0 < t \leqslant d$,

$$r_1(t) = \frac{-\rho_1\hat{\mathbf{g}}_1^{\mathrm{T}}(t)\hat{\boldsymbol{\psi}}_1(t)}{L_1^* \parallel \hat{\boldsymbol{\psi}}_1(t) \parallel^2}, \tag{30}$$

$$r_2(t) = \frac{-\rho_2\hat{g}_2(t)\hat{\psi}_2(t)}{L_2^* \parallel \hat{\psi}_2(t) \parallel^2}. \tag{31}$$

Remark 5 $\boldsymbol{\vartheta}_a$ and γ are separated by the hierarchical principle, which not only significantly reduces decreases the computational load associated with the algorithm, but also improves the identification accuracy. By accurately capturing the respective variation rules of $\boldsymbol{\vartheta}_a$ and γ, the dynamic features of the system can be described with greater accuracy.

For the SS-RR-MIPG algorithm, the estimation of time-delay τ is very important. It is found that τ is only

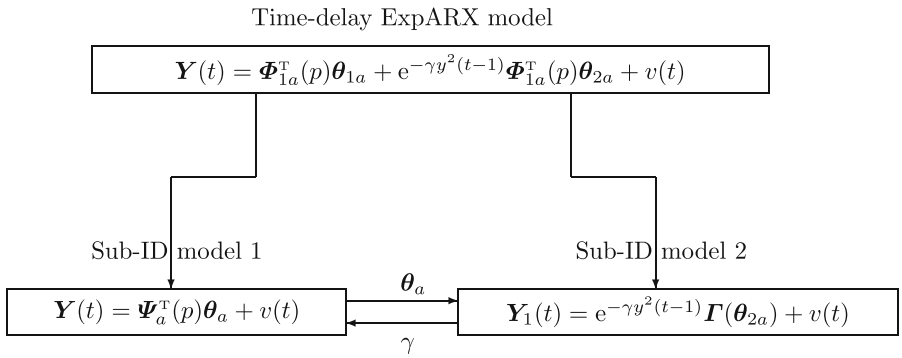

Fig. 2 The hierarchical identification structure of the time-delay ExpARX model

related to $\boldsymbol{\vartheta}_a$, but not to γ. Therefore, it is unnecessary to consider the time-delay of γ.

In summary, the SS-RR-MIPG algorithm is

$$\hat{\boldsymbol{\theta}}_a(t)=\hat{\boldsymbol{\theta}}_a(t-1)-r_1(t)\hat{\mathbf{g}}_1(t+k), \tag{32}$$

$$\hat{\mathbf{g}}_1(t+k)=\sum_{l_1=0}^{n-1}q_{l_1}\hat{\mathbf{g}}_1(t-l_1), \tag{33}$$

$$\hat{\boldsymbol{\psi}}_1(t+k)=-\hat{\mathbf{g}}_1(t+k)+\sum_{l_1=1}^{n-1}\zeta_{l_1}(t)\hat{\boldsymbol{\psi}}_1(t+k-l_1), \tag{34}$$

$$\zeta_{l_1}(t)=\frac{\sigma_1\parallel\hat{\mathbf{g}}_1(t+k)\parallel}{\parallel\hat{\boldsymbol{\psi}}_1(t+k-l)-\hat{\mathbf{g}}_1(t+k-l_1)\parallel}, \tag{35}$$

$$r_1(t)=\frac{-\rho_1\hat{\mathbf{g}}_1^{\mathrm{T}}(t+k)\hat{\boldsymbol{\psi}}_1(t+k)}{L_1\parallel\hat{\boldsymbol{\psi}}_1(t+k)\parallel^2}, \tag{36}$$

$$L_1=\max\left\{L_1^*,\frac{\parallel\hat{\mathbf{g}}_1(t+k)-\hat{\mathbf{g}}_1(t+k-1)\parallel}{\parallel\hat{\boldsymbol{\theta}}_a(t)-\hat{\boldsymbol{\theta}}_a(t-1)\parallel+\Lambda_1}\right\}, \tag{37}$$

$$\hat{\tau}=\underset{\tau\in0,\ldots,\tau_{\max}}{\operatorname{argmax}}\ \beth(\tau), \tag{38}$$

$$\hat{\gamma}(t)=\hat{\gamma}(t-1)-r_2(t)\hat{g}_2(t+k), \tag{39}$$

$$\hat{g}_2(t+k)=\sum_{l_2=0}^{n-1}q_{l_2}\hat{g}_2(t-l_2), \tag{40}$$

$$\hat{\psi}_2(t+k)=-\hat{g}_2(t+k)+\sum_{l_2=1}^{n-1}\zeta_{l_2}(t)\hat{\psi}_2(t+k-l_2), \tag{41}$$

$$\zeta_{l_2}(t)=\frac{\sigma_2\parallel\hat{g}_2(t+k)\parallel}{\parallel\hat{\psi}_2(t+k-2\,l)-\hat{g}_2(t+k-l_2)\parallel}, \tag{42}$$

$$r_2(t)=\frac{-\rho_2\hat{g}_2(t+k)\hat{\psi}_2(t+k)}{L_2\parallel\hat{\psi}_2(t+k)\parallel^2}, \tag{43}$$

$$L_2=\max\left\{L_2^*,\frac{\parallel\hat{g}_2(t+k)-\hat{g}_2(t+k-1)\parallel}{\parallel\hat{\gamma}(t)-\hat{\gamma}(t-1)\parallel+\Lambda_2}\right\}. \tag{44}$$

Remark 6 Using the hierarchical principle, the SS-RR-MIPG algorithm has higher flexibility and adaptability than the RR-MIPG algorithm. By co-processing $\boldsymbol{\theta}_a$ and γ simultaneously, the relationship between $\boldsymbol{\theta}_a$ and γ can be better coordinated to ensure that the algorithm can maintain stable performance during runtime. At the same time, it is possible to adjust a parameter individually according to the actual demand to satisfy specific needs.

The proposed SS-RR-MIPG parameter estimation algorithm in this article can combine other parameter identification algorithms [49–56] to present new parameter estimation methods for various dynamic stochastic systems, and can be applied to other control and schedule areas [57–63] such as the transportation communication systems [64–69] and dynamical systems and so on. The procedure of computing $\hat{\boldsymbol{\theta}}_a(t)$ and $\hat{\gamma}(t)$ by the SS-RR-MIPG algorithm is listed in Algorithm 2.

5 The analysis of the computational efficiency

The efficiency of algorithm is determined by computational complexity, quantified by computation volume, measured through counting floating-point operations-multiplications and additions combined. The computational amounts of the RR-MIPG and SS-RR-MIPG algorithms are shown in Tables 1 and 2.

From Tables 1 and 2, it has

$$N_1:=16nN+32pN+72N+8np+2n^2+4n-4p+20,$$
$$N_2:=16nN+24pN+72N+8np+3n-4p+26.$$

Table 1 Computational amounts of the RR-MIPG algorithm

Expressions	Multiplications	Additions
$\hat{\boldsymbol{\vartheta}}_a(t) = \hat{\boldsymbol{\vartheta}}_a(t-1) - r(t)\hat{\mathbf{g}}(t+k)$	$4N+1$	$4N+1$
$r(t) = \frac{-\rho\hat{\mathbf{g}}^{\mathrm{T}}(t+k)\hat{\boldsymbol{\psi}}(t+k)}{L\|\hat{\boldsymbol{\psi}}(t+k)\|^2}$	$12N+5$	$8N$
$L = \frac{\|\hat{\mathbf{g}}(t+k)-\hat{\mathbf{g}}(t+k-1)\|}{\|\hat{\boldsymbol{\vartheta}}_a(t)-\hat{\boldsymbol{\vartheta}}_a(t-1)\|+\Lambda}$	$8N+5$	$16N+3$
$\hat{\mathbf{g}}(t+k) = \sum_{l=0}^{n-1} q_l\hat{\mathbf{g}}(t-l)$	$4nN+n$	$4nN-4N+n-1$
$\hat{\boldsymbol{\psi}}(t+k) = -\hat{\mathbf{g}}(t+k) + \sum_{l=1}^{n-1}\zeta_l(t)\hat{\boldsymbol{\psi}}(t+k-l)$	$4nN-4N+n-1$	$4nN-4N+n-1$
$\zeta_l(t) = \frac{\sigma_\zeta\|\hat{\boldsymbol{\psi}}(t+k)\|}{\|\hat{\boldsymbol{\psi}}(t+k-l)-\hat{\mathbf{g}}(t+k-l)\|}$	$8N+6$	$12N+1$
$\beth(\tau) := \sum_{s=\tau+1}^{\tau+n}[\hat{\boldsymbol{\theta}}_{a,s}^2(L) + \hat{\boldsymbol{\theta}}_{a,s+N}^2(L)]$	$2n^2$	$2n-1$
$\hat{\mathbf{g}}(t) = \frac{\partial J(\hat{\boldsymbol{\vartheta}}_a(t))}{\partial\hat{\boldsymbol{\vartheta}}_a(t)}$	$16pN+4pn+p$	$16pN+4pn-5p-1$
$\hat{\boldsymbol{\psi}}(t) = -\hat{\mathbf{g}}(t) + \zeta(t)[\hat{\boldsymbol{\psi}}(t-1) - \hat{\mathbf{g}}(t-1)]$	$4N+1$	$8N+2$
Total flops $N_1 = 16nN+32pN+72N+8np$	$+2n^2+4n-4p+20$	

Table 2 Computational amounts of the SS-RR-MIPG algorithm

Expressions	Multiplications	Additions
$\hat{\boldsymbol{\theta}}_a(t) = \hat{\boldsymbol{\theta}}_a(t-1) - r_1(t)\hat{\mathbf{g}}_1(t+k)$	$4N$	$4N$
$r_1(t) = \frac{-\rho_1\hat{\mathbf{g}}_1^{\mathrm{T}}(t+k)\hat{\boldsymbol{\psi}}_1(t+k)}{L_1\|\hat{\boldsymbol{\psi}}_1(t+k)\|^2}$	$12N+3$	$8N-2$
$L_1 = \frac{\|\hat{\mathbf{g}}_1(t+k)-\hat{\mathbf{g}}_1(t+k-1)\|}{\|\hat{\boldsymbol{\theta}}_a(t)-\hat{\boldsymbol{\theta}}_a(t-1)\|+\Lambda_1}$	$8N+3$	$16N-1$
$\hat{\mathbf{g}}_1(t+k) = \sum_{l_1=0}^{n-1} q_{l_1}\hat{\mathbf{g}}_1(t-l_1)$	$4nN$	$4nN-4N-n-1$
$\hat{\boldsymbol{\psi}}_1(t+k) = -\hat{\mathbf{g}}_1(t+k) + \sum_{l_1=1}^{n-1}\zeta_{l_1}(t)$ $\times\hat{\boldsymbol{\psi}}_1(t+k-l_1)$	$4nN-4N$	$4nN-4N$
$\zeta_{l_1}(t) = \frac{\sigma_1\|\hat{\boldsymbol{\psi}}_1(t+k)\|}{\|\hat{\boldsymbol{\psi}}_1(t+k-l_1)-\hat{\mathbf{g}}_1(t+k-l_1)\|}$	$8N+4$	$12N-2$
$\beth(\tau) = \sum_{s=\tau+1}^{\tau+n}[\hat{\boldsymbol{\theta}}_{a,s}^2(L) + \hat{\boldsymbol{\theta}}_{a,s+N}^2(L)]$	$2n^2$	$2n-1$
$\hat{\mathbf{g}}_1(t) = \frac{\partial J(\hat{\boldsymbol{\theta}}_a(t))}{\partial\hat{\boldsymbol{\theta}}_a(t)}$	$8pN+4pn$	$8pN+4pn-2p$
$\hat{\boldsymbol{\psi}}_1(t) = -\hat{\mathbf{g}}_1(t) + \zeta_1(t)[\hat{\boldsymbol{\psi}}_1(t-1) - \hat{\mathbf{g}}_1(t-1)]$	$4N$	$8N$
$\hat{\gamma}(t) = \hat{\gamma}(t-1) - r_2(t)\hat{g}_2(t+k)$	1	2
$r_2(t) = \frac{-\rho_2\hat{g}_2(t+k)\hat{\psi}_2(t+k)}{L_2\|\hat{\psi}_2(t+k)\|^2}$	n	$n-1$
$L_2 = \frac{\|\hat{g}_2(t+k)-\hat{g}_2(t+k-1)\|}{\|\gamma(t)-\gamma(t-1)\|+\Lambda_2}$	7	0
$\hat{g}_2(t+k) = \sum_{l_2=0}^{n-1} q_{l_2}\hat{g}_2(t-l_2)$	3	3
$\hat{\psi}_2(t+k) = -\hat{g}_2(t+k) + \sum_{l_2=1}^{n-1}\zeta_{l_2}(t)$ $\times\hat{\psi}_2(t+k-l_2)$	$n-1$	$n-1$
$\zeta_{l_2}(t) = \frac{\sigma_2\|\hat{\psi}_2(t+k)\|}{\|\hat{\psi}_2(t+k-l_2)-\hat{g}_2(t+k-l_2)\|}$	4	1
$\hat{g}_2(t) = \frac{\partial J(\hat{\gamma}(t))}{\partial\hat{\gamma}(t)}$	$2pN+p$	$8pN-3p-1$
$\hat{\psi}_2(t) = -\hat{g}_2(t) + \zeta_1(t)[\hat{\psi}_2(t-1) - \hat{g}_2(t-1)]$	1	3
Total flops $N_2 = 16nN+24pN+72N+8np$	$+8np+3n-4p+26$	

Algorithm 2 SS-RR-MIPG algorithm

Require: Set the maximum iteration t_{max}, time-delay τ, prediction horizon $k = 2$.
Ensure: $\hat{\boldsymbol{\theta}}_a(t)$ and $\hat{\gamma}(t)$
for $t = 1$ **do**
  if $0 < t \leqslant d$ **then**
    Compute $\hat{\mathbf{g}}_1(t)$ and $\hat{g}_2(t)$ by (26) and (27)
    Update $\hat{\boldsymbol{\psi}}_1(t)$ and $\hat{\psi}_2(t)$ by (28) and (29)
    Compute $r_1(t)$ and $r_2(t)$ by (30) and (31)
  else[$t > d$]
    Compute $\hat{\mathbf{g}}_1(t+k)$, $\hat{\boldsymbol{\psi}}_1(t+k)$ and $\zeta_{l_1}(t)$ by (33), (34) and (35)
    Update $r_1(t)$ and L_1 by (36) and (37)
    Compute $\hat{g}_2(t+k)$, $\hat{\psi}_2(t+k)$ and $\zeta_{l_2}(t)$ by (40), (41) and (42)
    Update $r_2(t)$ and L_2 by (43) and (44)
  end if
  Update $\hat{\boldsymbol{\theta}}_a(t)$ and $\hat{\gamma}(t)$ by (32) and (39)
  if $\|\hat{\boldsymbol{\theta}}(t) - \hat{\boldsymbol{\theta}}(t-1)\| + \|\hat{\gamma}(t) - \hat{\gamma}(t-1)\| > \epsilon_1$ **then**
    $t := t + 1$
  else
    break
  end if
  Compute $\hat{\tau}(t)$ by (38)
  if $\|\hat{\tau} - \tau\| = 0$ **then**
    for $\iota = 1$ **do**
      $\alpha_\iota = 0$
      $\beta_\iota = 0$
      if $\iota < \tau$ **then**
        $\iota := \iota + 1$
      else
        break
      end if
    end for
  end if
end for
Update $\hat{\boldsymbol{\theta}}_a(t)$ and $\hat{\gamma}(t)$

When $n,\ p \geqslant 1$, the difference in computational amounts between the RR-MIPG algorithm and the SS-RR-MIPG algorithms is

$$N_1 - N_2 = 8pN + 2n^2 + n - 6 > 0.$$

Obviously, $N_1 > N_2$. Compared to the RR-MIPG algorithm, the SS-RR-MIPG algorithm leverages hierarchical principles to significantly enhance computational efficiency. For example, let $n = 10$, $N = 40$, $p = 40$, it yields

$$\begin{aligned} N_1 &= 6.378 \times 10^4, \\ N_2 &= 5.0776 \times 10^4, \\ N_1 - N_2 &= 1.3004 \times 10^4. \end{aligned}$$

With increasing the system dimension, the computational difference becomes more noticeable, highlighting the superior computational efficiency of the SS-RR-MIPG algorithm over the RR-MIPG algorithm. For a clear visualization of the computational disparity between the RR-MIPG and SS-RR-MIPG algorithms, $N = 200$ is fixed, and the number of operations of n and p is drawn in Fig. 3.

6 The main convergence results

Suppose that the gradient $\mathbf{g}_1(t)$ is $Lipschitz$ continuous on open convex set Ξ containing $L_c(\boldsymbol{\theta}_{a0})$. The criterion function $J(\boldsymbol{\theta}_a)$ and the gradient vectors $\mathbf{g}_1(t)$ satisfy the following assumptions:

(A1) $L_{c_1}(\boldsymbol{\theta}_{a0}) = \{\boldsymbol{\theta}_{a0} \in R^{4N} | J(\boldsymbol{\vartheta})_a \leqslant J(\boldsymbol{\vartheta})_{a0}\}$,

(A2) $\| \mathbf{g}_1(\boldsymbol{\theta}_a(t)) - \mathbf{g}_1(\boldsymbol{\theta}_a(t-1)) \| \leqslant L_{c_1} \| \Delta\boldsymbol{\theta}_a(t) \|$.

Lemma 1 *Assume (A1) and (A2) are satisfied, then for all t*

$$\hat{\mathbf{g}}_1^{\mathrm{T}}(t+k)\hat{\boldsymbol{\psi}}_1(t+k) \leqslant -(1-\sigma_1 N) \| \hat{\mathbf{g}}_1(t+k) \|^2, \tag{45}$$

$$\| \hat{\boldsymbol{\psi}}_1(t+k) \| \leqslant (1+\sigma_1 N) \| \hat{\mathbf{g}}_1(t+k) \|. \tag{46}$$

Theorem 1 *For all t, if* $\rho_1 = \frac{1-\sigma_1 N}{2}$, *the infinite sequence* $\boldsymbol{\theta}_a$ *generated by the separable synchronous redundant rule-based multi-innovation predictive gradient (SS-RR-MIPG) algorithm satisfies*

$$\lim_{t\to\infty} \| \hat{\mathbf{g}}_1(t+k) \| = 0.$$

Proof In accordance with the mean value theorem

$$J(\boldsymbol{\theta}_a(t)) - J(\boldsymbol{\theta}_a(t-1)) = \mathbf{g}_1^{\mathrm{T}}(\tilde{\boldsymbol{\theta}}_a)\Delta\boldsymbol{\theta}_a(t),$$

where $\tilde{\boldsymbol{\theta}}_a \in [\boldsymbol{\theta}_a(t), \boldsymbol{\theta}_a(t-1)]$.

It can obtain that

$$\begin{aligned} \mathbf{g}_1^{\mathrm{T}}\tilde{\boldsymbol{\theta}}_a\Delta\boldsymbol{\theta}_a(t) \leqslant & -r_1(t) \| \hat{\mathbf{g}}_1^{\mathrm{T}}(t+k) \|^2 \\ & + L_{c_1} r_1^2(t) \| \hat{\mathbf{g}}_1(t+k) \|^2. \end{aligned}$$

Combining (45)–(46) and $Cauchy-Schwarz$ inequality, it yields

$$\begin{aligned} & J(\boldsymbol{\theta}_a(t)) - J(\boldsymbol{\theta}_a(t-1)) \\ \leqslant & \frac{\rho_1 \hat{\mathbf{g}}_1^{\mathrm{T}}(t+d)\hat{\boldsymbol{\psi}}_1(t+d)}{L_{c_1} \| \hat{\boldsymbol{\psi}}_1(t+d) \|}^2 \| \\ & + L_{c_1} r_1^2(t) \| \hat{\boldsymbol{\psi}}_1(t+d) \|^2 \\ \leqslant & \frac{-\rho_1(1-\sigma_1 N) \| \hat{\mathbf{g}}_1(t+d) \|^4}{L_{c_1} \| \hat{\boldsymbol{\psi}}_1(t+d) \|^2} \end{aligned}$$

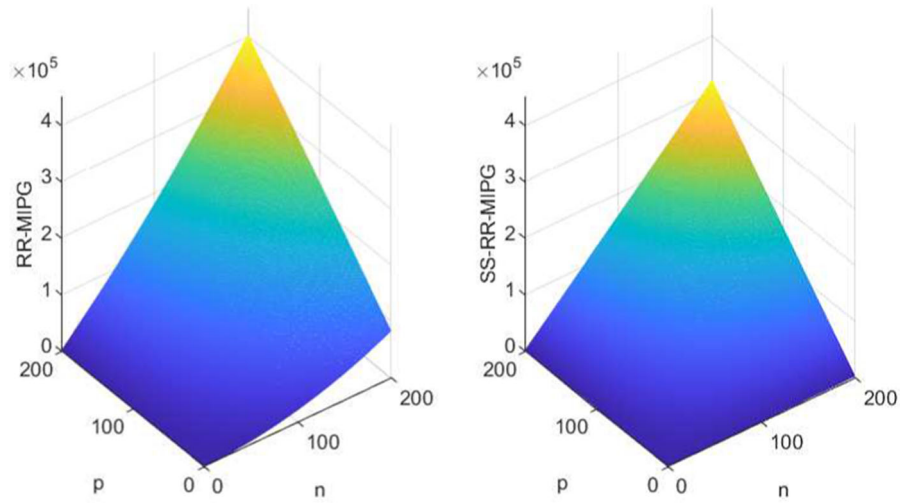

Fig. 3 The RR-MIPG and SS-RR-MIPG flop numbers versus p and n

$$+\frac{L_{c_1}\rho_1^2 \parallel \hat{\mathbf{g}}_1(t+d) \parallel^4 \parallel \hat{\boldsymbol{\psi}}_1(t+d) \parallel^2}{L_{c_1}^2 \parallel \hat{\boldsymbol{\psi}}_1(t+d) \parallel^4}$$
$$=\frac{[\rho_1^2-(1-\sigma_1 N)] \parallel \hat{\mathbf{g}}_1(t+d) \parallel^4}{L_{c_1} \parallel \hat{\boldsymbol{\psi}}_1(t+d) \parallel^2}, \quad (47)$$

Considering $\rho_1 = \frac{1-\sigma_1 N}{2}$, which is also solutions to the quadratic equations $\rho_1^2 - (1-\sigma_1 N)\rho_1$, Eq. (47) will be

$$J(\boldsymbol{\theta}_a(t)) - J(\boldsymbol{\theta}_a(t-1)) \leqslant \frac{-(1-\sigma_1 N)^2 \parallel \hat{\mathbf{g}}_1(t+k) \parallel^4}{4L_{c_1} \parallel \hat{\boldsymbol{\psi}}_1(t+k) \parallel^2}. \quad (48)$$

Thus, $J(\boldsymbol{\theta}_a(t))$ and $J(\gamma)$ are decrease sequences, and

$$\sum_{t=1}^{\infty} \frac{-(1-\sigma_1 N)^2 \parallel \hat{\mathbf{g}}_1(t+k) \parallel^4}{4L_{c_1} \parallel \hat{\boldsymbol{\psi}}_1(t+k) \parallel^2} < \infty.$$

Based on (45)–(46), it has

$$0 < \frac{(1-\sigma_1 N)^2}{(1+\sigma_1 N)^2} \sum_{t=1}^{\infty} \parallel \hat{\mathbf{g}}_1(t+k) \parallel^2$$
$$\leqslant \sum_{t=1}^{\infty} \frac{\parallel \hat{\mathbf{g}}_1(t+k) \parallel^4}{\parallel \hat{\boldsymbol{\psi}}_1(t+k) \parallel^2} < \infty.$$

Therefore, it results in

$$\lim_{t\to\infty} \parallel \hat{\mathbf{g}}_1(t+k) \parallel = 0.$$

It is assumed that $J(\boldsymbol{\theta}_a)$ is locally convex except for non-smooth extremum. □

Theorem 2 *Suppose $J(\boldsymbol{\theta}_a)$ has a minima $\boldsymbol{\theta}_a^* \in \mathbb{R}^{4N}$. It is also supposed that it exists $0 < \aleph_{11} < \aleph_{21}$, such that*

$$\frac{\aleph_{11}}{2} \parallel \boldsymbol{\theta}_a - \boldsymbol{\theta}_a^* \parallel^2 \leqslant J(\boldsymbol{\theta}_a(t)) - J(\boldsymbol{\theta}_a^*)$$
$$\leqslant \frac{\aleph_{21}}{2} \parallel \boldsymbol{\theta}_a - \boldsymbol{\theta}_a^* \parallel^2, \quad (49)$$
$$\aleph_{11} \parallel \boldsymbol{\theta}_a - \boldsymbol{\theta}_a^* \parallel \leqslant \parallel \mathbf{g}_1(\boldsymbol{\theta}_a) \parallel \leqslant \aleph_{21} \parallel \boldsymbol{\theta}_a - \boldsymbol{\theta}_a^* \parallel \quad (50)$$

Then, the infinite sequence $\{\boldsymbol{\theta}_a(t)\}$ generated by the SS-RR-MIPG algorithm lead to $\boldsymbol{\theta}_a(t) \to \boldsymbol{\theta}_a^*$ as $t \to \infty$.

Proof As the level set $\Upsilon_1 = \{\boldsymbol{\theta}_a \in \mathbb{R}^{4N}\}$ is bounded. Let $\aleph_2 = 4L_{c_1}$. According to (48)–(50) and $0 < \aleph_{11} < \aleph_{21}$. It can be obtained that

$$\frac{-(1-\sigma_1 N) \parallel \hat{\mathbf{g}}_1(t+k) \parallel^4}{4L_{c_1} \parallel \hat{\boldsymbol{\psi}}_1(t+k) \parallel^2} \leqslant -\frac{\aleph_{11}(1-\sigma_1 N)^2}{\aleph_{21}(1+\sigma_1 N)^2}$$
$$\times [J(\boldsymbol{\theta}_a(t)) - J(\boldsymbol{\theta}_a^*)].$$

Define

$$\Omega_1 = \frac{\aleph_{11}(1-\sigma_1 N)^2}{\aleph_{21}(1+\sigma_1 N)^2} \in (0,1).$$

Therefore, it has

$$\frac{\aleph_{11}}{2} \parallel \boldsymbol{\theta}_a - \boldsymbol{\vartheta}_a^* \parallel^2 \leqslant J(\boldsymbol{\theta}_a(t)) - J(\boldsymbol{\theta}_a^*)$$
$$\leqslant (1-\Omega_1)[J(\boldsymbol{\theta}_a(t)) - J(\boldsymbol{\theta}_a^*)]$$
$$\leqslant \cdots$$
$$\leqslant (1-\Omega_1)[J(\boldsymbol{\theta}_a(1)) - J(\boldsymbol{\theta}_a^*)].$$

It means that $\boldsymbol{\theta}_a(t) \to \boldsymbol{\theta}_a^*$. □

Theorem 3 *Suppose the identification strategy described by (32) and (39) are applied to the system described by (24). Let $\varpi_1 = [\boldsymbol{\chi}_1'(t)]^2 + \nu_1$, $0 < \nu_1 < \infty$. If the optimization step-size meets $0 < r_1(t) < \frac{2}{\varpi_1}$, the SS-RR-MIPG sub-algorithm of $\boldsymbol{\theta}_a$ is stable.*

Proof Consider

$$\boldsymbol{\chi}_1(t) = \boldsymbol{\chi}_1(t-1) + [\boldsymbol{\chi}_1'(t)]^{\mathrm{T}} \Delta\boldsymbol{\theta}_a(t). \tag{51}$$

Similarly, it has

$$\Delta\boldsymbol{\theta}_a(t) = -r_1(t)\boldsymbol{\chi}_1(t)\boldsymbol{\chi}_1'(t). \tag{52}$$

Substituting (51) into (52) yields

$$\begin{aligned}\boldsymbol{\chi}_1^2(t) &= \boldsymbol{\chi}_1^2(t-1) + r_1^2(t) \parallel \boldsymbol{\chi}_1'(t) \parallel^4 \boldsymbol{\chi}_1^2(t) \\ &\quad -2r_1(t) \parallel \boldsymbol{\chi}_1'(t) \parallel^2 \boldsymbol{\chi}_1^2(t-1)\boldsymbol{\chi}_1^2(t), \end{aligned} \tag{53}$$

$$\boldsymbol{\chi}_1(t-1)\boldsymbol{\chi}_1(t) = \boldsymbol{\chi}_1^2(t) + r_1(t) \parallel \boldsymbol{\chi}_1'^2(t) \parallel \boldsymbol{\chi}_1^2(t). \tag{54}$$

Define the *Lyapunov's* function as

$$\Upsilon_1(t) = \boldsymbol{\chi}_1^2(t) + 2(\parallel \boldsymbol{\chi}_1'(t) \parallel^2 + \frac{\nu_1}{2}) \parallel \Delta\boldsymbol{\theta}_a(t) \parallel^2 \tag{55}$$

In accordance with (51)–(52)

$$\begin{aligned}&\Upsilon_1(t) - \Upsilon_1(t-1) \\ &= \boldsymbol{\chi}_1^2(t-1) + r_1^2(t) \parallel \boldsymbol{\chi}_1'(t) \parallel^4 \boldsymbol{\chi}_1^2(t) \\ &\quad -2r_1(t)\boldsymbol{\chi}_1'^2(t)\boldsymbol{\chi}_1^2(t-1)\boldsymbol{\chi}_1^2(t) \\ &\quad +2r_1^2(t) \parallel \boldsymbol{\chi}_1'(t) \parallel^4 \boldsymbol{\chi}_1^2(t) \\ &\quad +\nu_1 r_1^2(t) \parallel \boldsymbol{\chi}_1'^2(t) \parallel \\ &\quad -\boldsymbol{\chi}_1^2(t-1) - 2 \parallel \boldsymbol{\chi}_1'^2(t-1) \parallel^2 \parallel \Delta\boldsymbol{\theta}_a(t-1) \parallel^2 \\ &\leqslant r_1(t) \parallel \boldsymbol{\chi}_1'(t) \parallel^2 [-2\boldsymbol{\chi}_1(t-1)\boldsymbol{\chi}_1(t) \\ &\quad +3r_1(t)\boldsymbol{\chi}_1^2(t) \parallel \boldsymbol{\chi}_1'(t) \parallel^2 \\ &\quad +\nu_1 r_1(t)\boldsymbol{\chi}_1^2(t)]. \end{aligned}$$

Suppose

$$\begin{aligned}0 < r_1(t) &\leqslant 2\boldsymbol{\chi}_1(t-1)\boldsymbol{\chi}_1(t)\{3\boldsymbol{\chi}_1^2(t) \parallel \boldsymbol{\chi}_1'(t) \parallel^2 \\ &\quad +\nu_1\boldsymbol{\chi}_1^2(t)\}^{-1}. \end{aligned} \tag{56}$$

Substituting (54) into (56)results in

$$0 < r_1(t) < \frac{2}{\parallel \boldsymbol{\chi}_1'^2 \parallel + \nu_1}. \tag{57}$$

In the case, $\Upsilon_1(t) - \Upsilon_1(t-1) \leqslant 0$.

Thus, the stability of the SS-RR-MIPG sub-algorithm of $\boldsymbol{\theta}_a$ is proved. □

Theorem 4 *Assume* $(\boldsymbol{\chi}_1^*)^2 = \min\{\boldsymbol{\chi}_1^2(t)\} = 0$. *If the optimizing step-size satisfies* $0 < r_1(t) < \frac{2}{\varpi_1}$ *and the SS-RR-MIPG sub-algorithm of* $\boldsymbol{\theta}_a$ *is stable, the time-delay exponential autoregressive with exogenous input (ExpARX) model is also asymptotically stable, which leads to*

$$\lim_{t\to\infty} \boldsymbol{\chi}_1(t) \to 0.$$

Proof If $r_1(t)$ satisfies $0 < r_1(t) < \frac{2}{\varpi_1}$, $\Upsilon_1(t)$ is decreasing sequence. Obviously,

$$\tilde{\boldsymbol{\chi}}_1(t) = \tilde{\boldsymbol{\chi}}_1(t-1) - r_1(t) \parallel \boldsymbol{\chi}_1'(t) \parallel^2 \boldsymbol{\chi}_1(t),$$

where $\tilde{\boldsymbol{\chi}}_1(t) = \boldsymbol{\chi}_1(t) - \boldsymbol{\chi}_1^*$.

Thus, it has

$$\begin{aligned}\tilde{\boldsymbol{\chi}}_1^2(t) &= \tilde{\boldsymbol{\chi}}_1^2(t-1) - 2\left[1 - [\boldsymbol{\chi}_1(t)]^{-1}\boldsymbol{\chi}_1^*\right] r_1(t) \\ &\quad \parallel \boldsymbol{\chi}_1'(t) \parallel^2 \boldsymbol{\chi}_1^2(t) \\ &\quad +r_1^2(t) \parallel \boldsymbol{\chi}_1'(t) \parallel^4 \boldsymbol{\chi}_1^2(t). \end{aligned} \tag{58}$$

It means

$$\tilde{\boldsymbol{\chi}}_1^2(t) \leqslant \tilde{\boldsymbol{\chi}}_1^2(t-1). \tag{59}$$

According to (58)–(59), $\{\tilde{\boldsymbol{\chi}}_1^2(t)\}$ is decreasing sequence. Thus, it exists nonnegative scalars $(\tilde{\boldsymbol{\chi}}_1^*)^2$ that satisfy

$$\lim_{t\to\infty} \tilde{\boldsymbol{\chi}}_1^2(t) \to (\tilde{\boldsymbol{\chi}}_1^*)^2.$$

Then, based on (57) and the definition of $\tilde{\boldsymbol{\chi}}_1^2(t)$, it yields

$$\begin{aligned}\tilde{\boldsymbol{\chi}}_1(t-1)[\tilde{\boldsymbol{\chi}}_1(t)]^{-1} &= 1 + r_1(t) \parallel \boldsymbol{\chi}_1'(t) \parallel^2 \\ &\quad \left\{1 - \boldsymbol{\chi}_1^*[\boldsymbol{\chi}_1(t)]^{-1}\right\}^{-1}. \end{aligned}$$

Thus,

$$\begin{aligned}\tilde{\boldsymbol{\chi}}_1(t)[\tilde{\boldsymbol{\chi}}_1(t-1)]^{-1} &= \left\{1 + r_1(t) \parallel \boldsymbol{\chi}_1'(t) \parallel^2 \right. \\ &\quad \left. \left[1 - \boldsymbol{\chi}_1^*[\boldsymbol{\chi}_1(t)]^{-1}\right]^{-1}\right\}^{-2}. \end{aligned}$$

Define

$$\mu_1(t) = \left\{1 + r_1(t) \parallel \boldsymbol{\chi}_1'(t) \parallel^2 \left[1 - \boldsymbol{\chi}_1^*[\boldsymbol{\chi}_1(t)]^{-1}\right]^{-1}\right\}^{-2}.$$

Let $\mu_1^* = \max\{\mu_1(t)\}$, which results in $0 < \mu_1^* < 1$. It can be obtained that

$$0 \leqslant \tilde{\boldsymbol{\chi}}_1^2(t) \leqslant \mu_1^*\tilde{\boldsymbol{\chi}}_1^2(t) \leqslant \cdots \leqslant \mu_1^*\tilde{\boldsymbol{\chi}}_1^2(0).$$

$\tilde{\boldsymbol{\chi}}_1(0)$ is bounded number and

$$\lim_{t\to\infty} (\mu_1^*)^{\mathrm{t}} \to 0.$$

From the above equations, it has

$$\lim_{t\to\infty} \tilde{\boldsymbol{\chi}}_1^2(t) \to 0.$$

Note that $(\tilde{\boldsymbol{\chi}}_1^*)^2 \to 0$ means $\tilde{\boldsymbol{\chi}}_1^* \to 0$. Therefore, it leads to

$$\lim_{t\to\infty} \boldsymbol{\chi}_1(t) \to \boldsymbol{\chi}_1^* = 0.$$

□

Table 3 The parameter estimation and errors of the RR-MIPG and SS-RR-MIPG algorithms ($\sigma^2 = 0.10^2$)

Algorithm	p	α_1	$\xi_{1,1}$	$\xi_{1,2}$	$\xi_{2,1}$	$\xi_{2,2}$	α_2	β_1	$\eta_{1,1}$	$\eta_{1,2}$	$\eta_{2,1}$	$\eta_{2,2}$	β_2	γ	$\delta(\%)$
	1	0.000	0.269	0.341	0.271	0.263	0.000	0.000	0.529	0.493	0.408	0.418	0.000	0.206	86.133
RR-MIPG	5	0.000	0.351	0.416	0.275	0.211	0.000	0.000	0.644	0.584	0.342	0.405	0.000	0.555	74.814
	10	0.000	0.444	0.509	0.275	0.089	0.000	0.000	0.787	0.726	0.212	0.347	0.000	2.942	3.021
	1	0.000	0.391	0.458	0.246	0.166	0.000	0.000	0.762	0.647	0.187	0.418	0.000	0.825	66.067
SS-RR-MIPG	5	0.000	0.410	0.468	0.259	0.160	0.000	0.000	0.773	0.685	0.209	0.405	0.000	1.277	52.336
	10	0.000	0.454	0.490	0.242	0.157	0.000	0.000	0.803	0.746	0.178	0.334	0.000	2.987	0.695
True values		0.000	0.450	0.500	0.250	0.150	0.000	0.000	0.800	0.750	0.170	0.330	0.000	3.000	

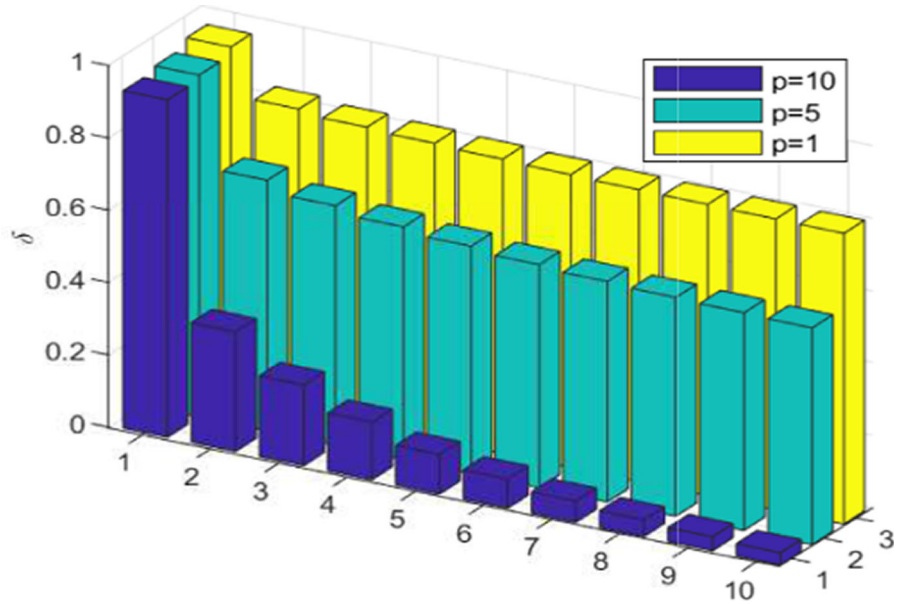

Fig. 4 The estimation errors versus t of RR-MIPG under different p

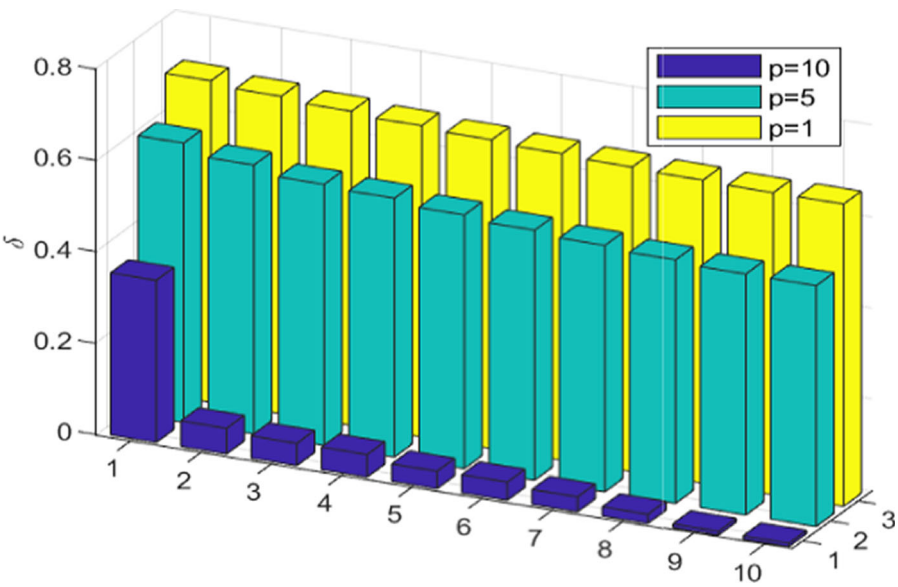

Fig. 5 The estimation errors versus t of SS-RR-MIPG under different p

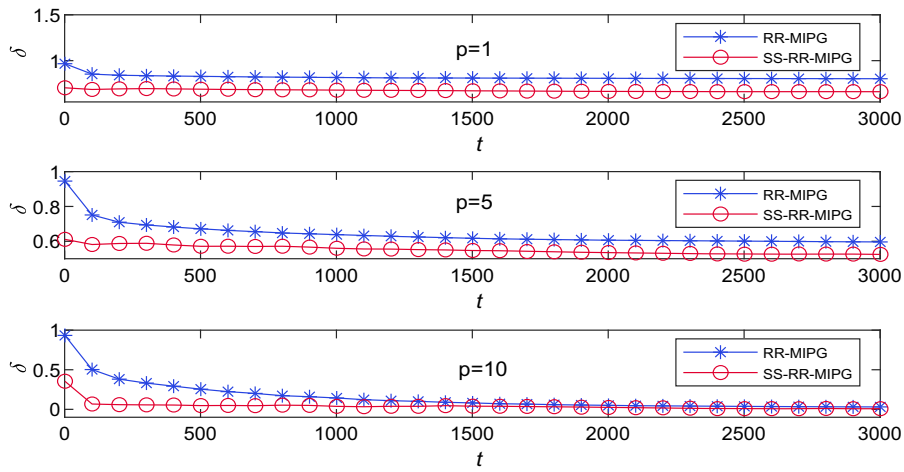

Fig. 6 The estimation errors versus t of RR-MIPG and SS-RR-MIPG

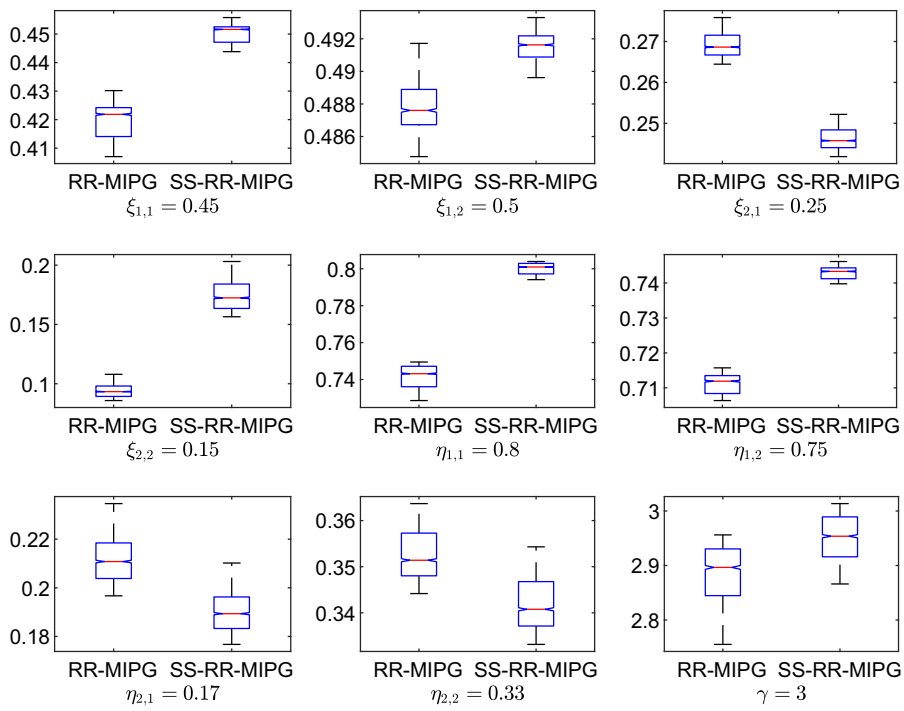

Fig. 7 The parameter estimation of $\boldsymbol{\theta}_a$ by RR-MIPG and SS-RR-MIPG

Remark 7 The convergence of the proposed algorithm is demonstrated by showing the gradient and the output error approach zero as $t \to \infty$. The global convergence of the proposed algorithm is proven by establishing the existence of $\boldsymbol{\theta}_a(t)$ that minimizes the criterion function. The stability of the proposed system is proven by determining the minimum of the Lyapunov function composed of the output error.

The convergence of the SS-RR-MIPG sub-algorithm of $\boldsymbol{\theta}_a$ is proved. Similarly, the SS-RR-MIPG sub-algorithm of γ is convergent.

7 Numerical results

This section employs numerical simulations and a real-life example of piezoelectric ceramics to validate the effectiveness of the redundant rule-based multi-innovation predictive gradient (RR-MIPG) and separable synchronous redundant rule-based multi-innovation predictive gradient (SS-RR-MIPG) algorithms.

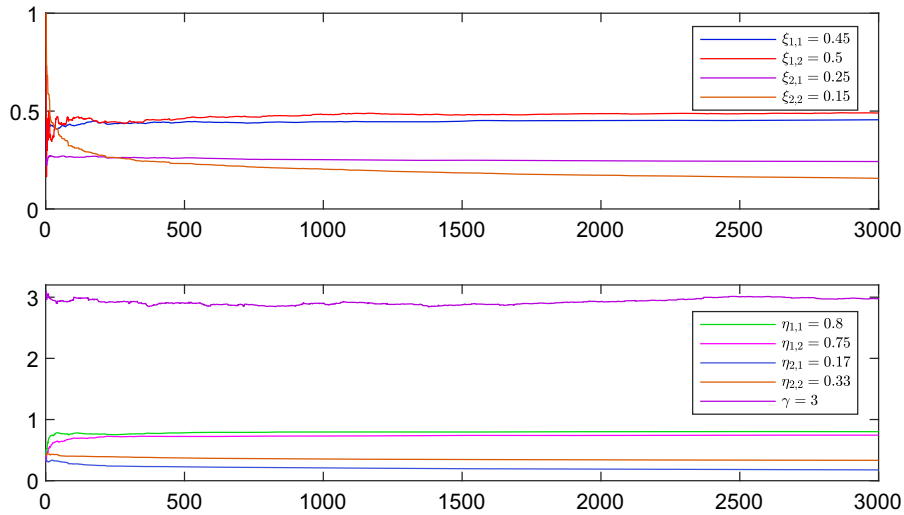

Fig. 8 The parameter estimation by SS-RR-MIPG versus t

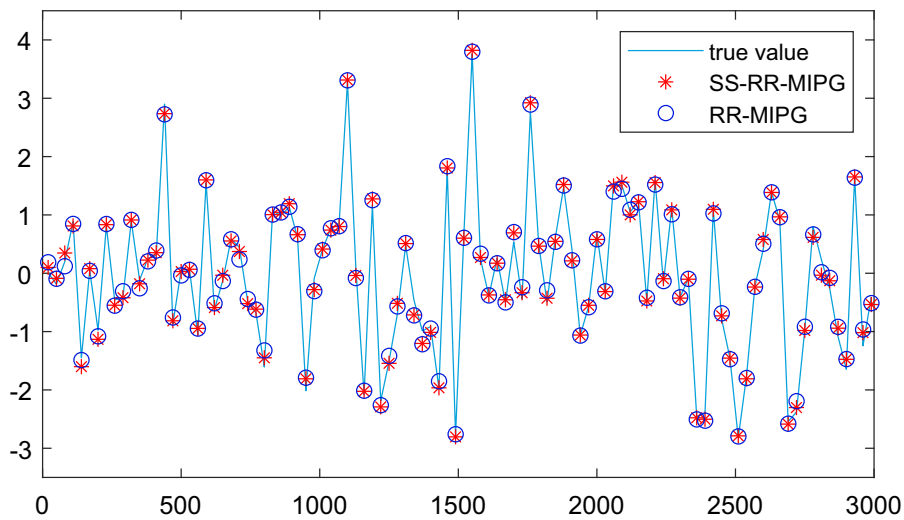

Fig. 9 The true outputs and the estimated outputs versus t

7.1 Numerical simulation

Consider the exponential autoregressive with exogenous input (ExpARX) model with order $n = 2$, $\tau = 1$, and $k = 2$,

$$\begin{aligned}\boldsymbol{\vartheta} &= [\xi_{1,1}, \xi_{1,2}, \xi_{2,1}, \xi_{2,2}, \eta_{1,1}, \eta_{1,2}, \eta_{2,1}, \eta_{2,2}, \gamma]^{\mathrm{T}} \\ &= [0.45, 0.5, 0.25, 0.15, 0.8, 0.75, 0.17, 0.33, 3.00]^{\mathrm{T}},\end{aligned}$$

The augmented parameter is

$$\begin{aligned}\boldsymbol{\vartheta}_a &= [\alpha_1, \xi_{1,1}, \xi_{1,2}, \xi_{2,1}, \xi_{2,2}, \alpha_2, \beta_1, \eta_{1,1}, \eta_{1,2}, \eta_{2,1}, \\ &\quad \eta_{2,2}, \beta_2, \gamma]^{\mathrm{T}} \\ &= [0, 0.45, 0.5, 0.25, 0.15, 0, 0, 0.8, 0.75, \\ &\quad 0.17, 0.33, 0, 3.00]^{\mathrm{T}}.\end{aligned}$$

The parameter estimation and errors at time $t = L = 3000$ are shown in Table 3 and Figs. 4, 5 and 6, where $\delta(t) := \|\hat{\boldsymbol{\vartheta}}_a(t) - \boldsymbol{\vartheta}_a\| \times 100\%$ is the parameter estimation error. Figures 7 and 8 show the estimation of parameters by the separable synchronous redundant rule-based multi-innovation predictive gradient (SS-RR-MIPG) and redundant rule-based multi-innovation predictive gradient (RR-MIPG) algorithms. The actual outputs $y(t)$ is shown in Figs. 9 and 10. Table 4 and Fig. 11 and 12 illustrate the parameter estimation under the different σ^2. The mean values and estimation deviations of the SS-RR-MIPG algorithm are illustrated in Tables 5 and 6 and Figs. 13 and 14. The result comparison between the SS-RR-MIPG algorithm and the traditional H-MIGI algorithm is shown in Fig. 15.

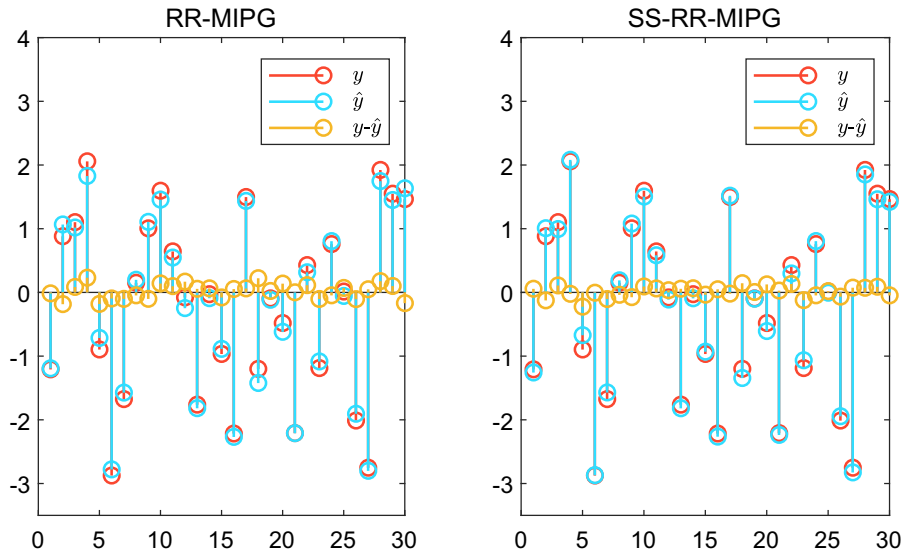

Fig. 10 The errors of the outputs

Table 4 The parameter estimation and errors of the RR-MIPG algorithm ($p = 10$)

Algorithm	σ^2	α_1	$\xi_{1,1}$	$\xi_{1,2}$	$\xi_{2,1}$	$\xi_{2,2}$	α_2	β_1	$\eta_{1,1}$	$\eta_{1,2}$	$\eta_{2,1}$	$\eta_{2,2}$	β_2	γ	$\delta(\%)$
	0.60^2	0.000	0.440	0.499	0.254	0.084	0.000	0.000	0.782	0.720	0.214	0.350	0.000	2.712	9.104
RR-MIPG	0.40^2	0.000	0.441	0.502	0.252	0.086	0.000	0.000	0.784	0.721	0.213	0.349	0.000	2.778	7.298
	0.20^2	0.000	0.442	0.506	0.251	0.087	0.000	0.000	0.786	0.723	0.212	0.348	0.000	2.845	4.879
	0.10^2	0.000	0.444	0.509	0.250	0.089	0.000	0.000	0.787	0.726	0.212	0.347	0.000	2.942	3.021
	0.60^2	0.000	0.458	0.494	0.245	0.150	0.000	0.000	0.741	0.720	0.168	0.338	0.000	3.258	7.833
SS-RR-MIPG	0.40^2	0.000	0.457	0.492	0.244	0.153	0.000	0.000	0.742	0.721	0.173	0.341	0.000	3.112	3.617
	0.20^2	0.000	0.456	0.491	0.243	0.155	0.000	0.000	0.744	0.723	0.174	0.343	0.000	3.077	2.380
	0.10^2	0.000	0.454	0.490	0.242	0.157	0.000	0.000	0.803	0.746	0.178	0.334	0.000	2.987	0.695
True values		0.000	0.450	0.500	0.250	0.150	0.000	0.000	0.800	0.750	0.170	0.330	0.000	3.000	

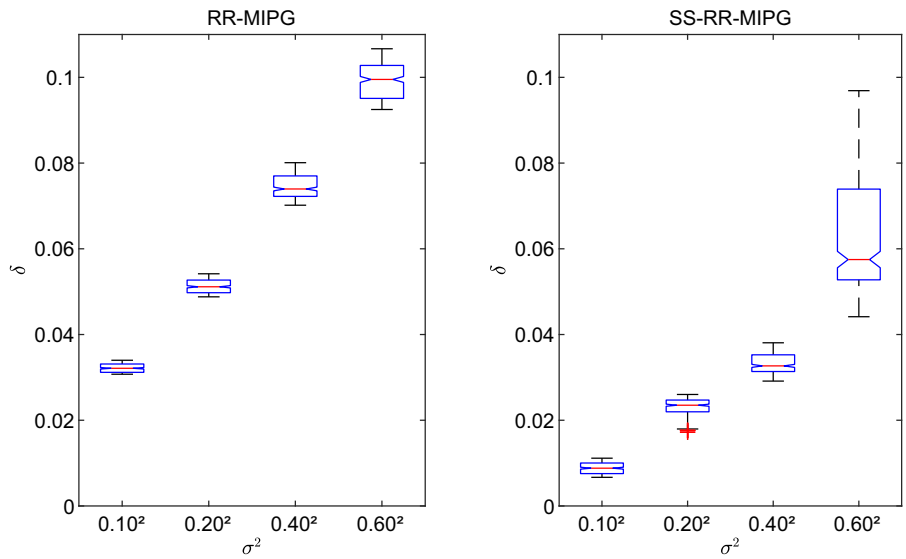

Fig. 11 The estimation errors of RR-MIPG and SS-RR-MIPG

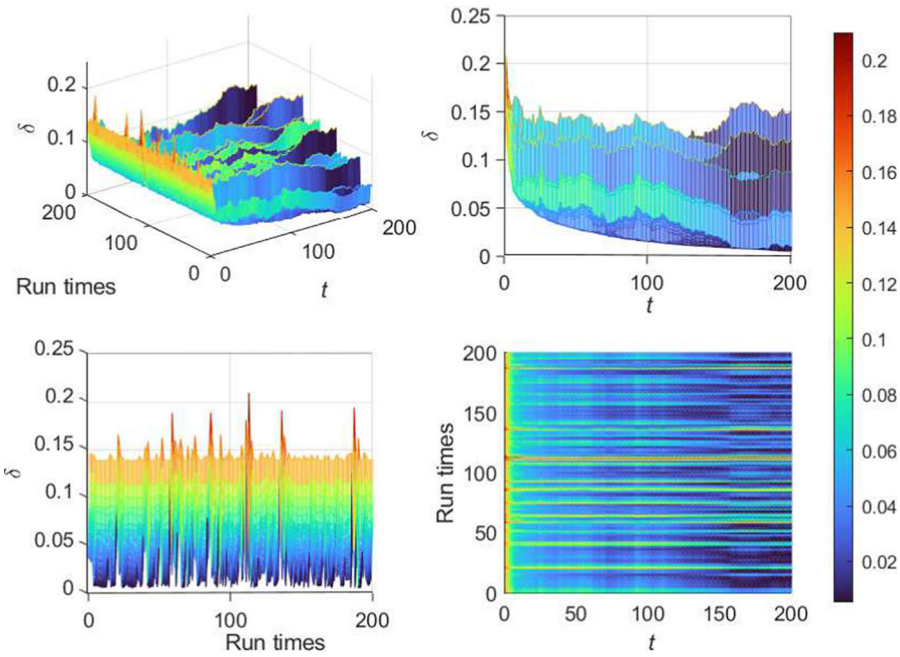

Fig. 12 The Monte Carlo estimation errors t and Run Times ($p = 10$)

Table 5 Mean values and deviations of parameter ξ ($\sigma^2 = 0.10^2$)

Run	$\xi_{1,1}$	$\xi_{1,2}$	$\xi_{2,1}$	$\xi_{2,2}$
10	0.4424 ± 0.0234	0.4905 ± 0.0849	0.2538 ± 0.0178	0.1692 ± 0.0928
20	0.4456 ± 0.0266	0.4924 ± 0.0837	0.2532 ± 0.0175	0.1668 ± 0.0912
30	0.4469 ± 0.0279	0.4949 ± 0.0238	0.2517 ± 0.0153	0.1615 ± 0.0896
40	0.4477 ± 0.0287	0.4963 ± 0.0252	0.2511 ± 0.0175	0.1599 ± 0.0844
50	0.4475 ± 0.0285	0.4960 ± 0.0249	0.2504 ± 0.0148	0.1583 ± 0.0483
True values	0.4500	0.500	0.250	0.150

Based on the simulation results, it can be drawn.

- As t and p increase, the errors of the RR-MIPG and SS-RR-MIPG algorithms decrease, and the SS-RR-MIPG algorithm is more accurate than the RR-MIPG algorithm—see Table 3 and Figs. 4, 5 and 6.
- The parameter identification results of the SS-RR-MIPG algorithm are closer to the true values compared to the RR-MIPG algorithm—see Figs. 7 and 8.
- The predicted outputs by proposed algorithms closely match the actual outputs, and the SS-RR-MIPG algorithm has a smaller prediction error compared to the RR-MIPG algorithm—see Figs. 9 and 10.
- As noise variance decreases, the errors decrease for both RR-MIPG and SS-RR-MIPG algorithms, and the SS-RR-MIPG algorithm exhibits superior estimation accuracy under equivalent σ^2—see Table 4 and Fig. 11.
- The Monte Carlo results indicate that the SS-RR-MIPG algorithm is effective for identifying the ExpARX model—see Tables 5 and 6 and Figs. 12 and 14.
- The SS-RR-MIPG algorithm has smaller identification error and faster convergence speed compared to the H-MIGI algorithm—see Fig. 15.

Table 6 Mean values and deviations of parameters η and γ ($\sigma^2 = 0.10^2$)

Run	$\eta_{1,1}$	$\eta_{1,2}$	$\eta_{2,1}$	$\eta_{2,2}$	γ
10	0.8044 ± 0.0233	0.7583 ± 0.0545	0.1981 ± 0.1277	0.3341 ± 0.0347	3.0021 ± 0.1242
20	0.7996 ± 0.0415	0.7571 ± 0.0540	0.1833 ± 0.1260	0.3315 ± 0.0321	2.9955 ± 0.2036
30	0.7986 ± 0.0405	0.7558 ± 0.0537	0.1799 ± 0.1244	0.3284 ± 0.0378	3.0132 ± 0.2198
40	0.7998 ± 0.0417	0.7555 ± 0.0523	0.1783 ± 0.1209	0.3289 ± 0.0368	3.0066 ± 0.1655
50	0.8002 ± 0.0420	0.7550 ± 0.0489	0.1766 ± 0.1161	0.3290 ± 0.0386	3.0025 ± 0.1614
True values	0.8000	0.7500	0.1700	0.3300	3.0000

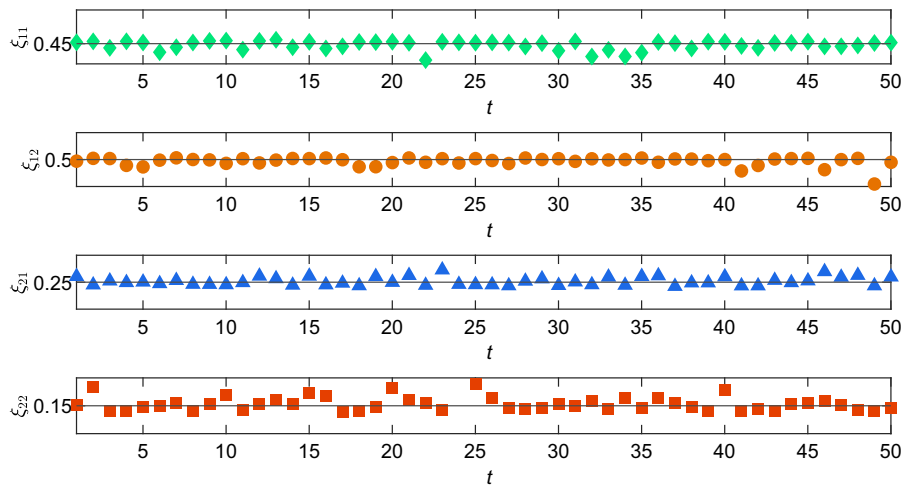

Fig. 13 The Monte Carlo estimation mean values of parameter ξ

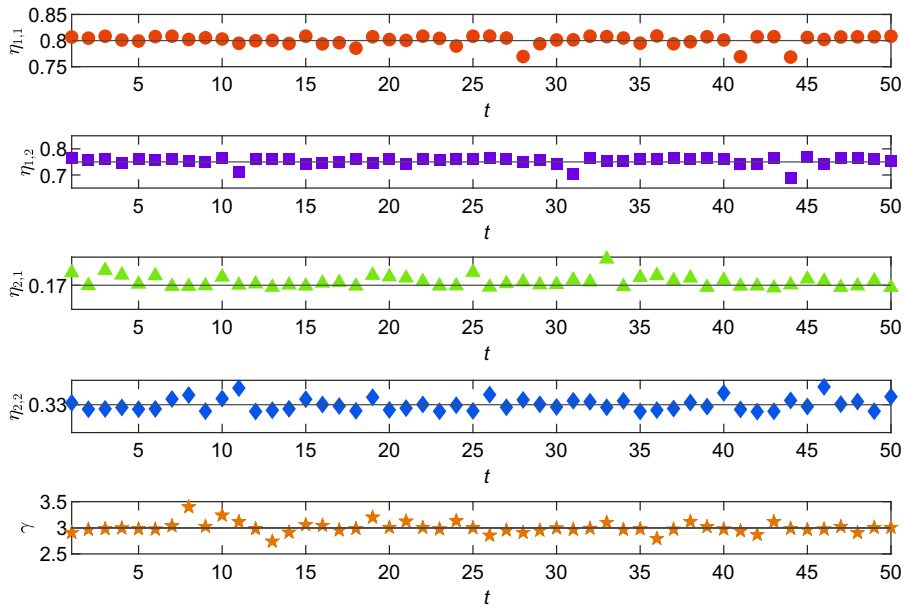

Fig. 14 The Monte Carlo estimation mean values of parameters η and γ

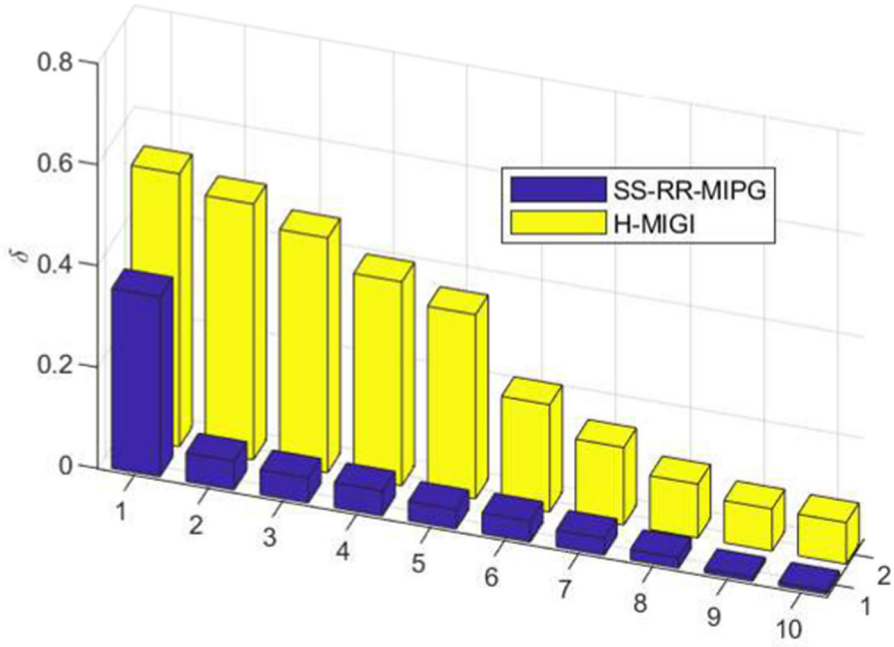

Fig. 15 The estimation errors versus t of SS-RR-MIPG and H-MIGI

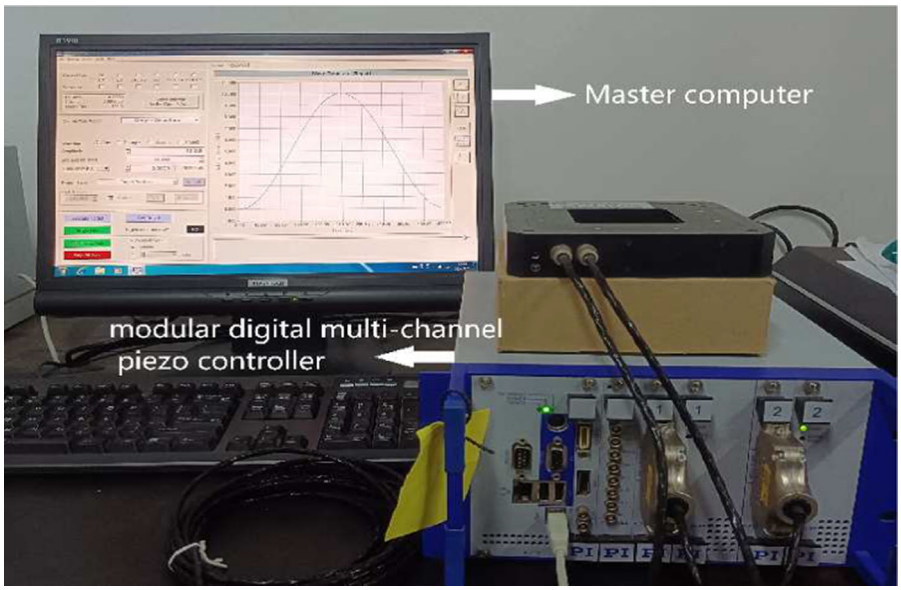

Fig. 16 The schematic diagram of the modular digital multi-channel piezo controller structure

7.2 Application to a piezoelectric ceramic system

The piezoelectric ceramic flow control system utilizes a piezoelectric ceramic actuator as its core to achieve precise gas flow control. Piezoelectric ceramics possess unique physical properties, generating voltage when subjected to external forces, while also deforming when voltage is applied. The characteristic allows the piezoelectric ceramic actuator to respond to minute electrical signal changes, enabling precise adjustments to the gas flow. By monitoring the flow rate and adjusting the voltage applied to the piezoelectric ceramic, the system ensures stable, high-precision control of gas flow. The piezoelectric ceramic flow control system features fast response and high control accuracy, making it suitable for various applications requiring precise flow control [70].

Table 7 The functional list of the modular digital multi-channel piezo controller structure

	Function	The modular digital multi-channel piezo controller
	Axes	3
	Processor	PC-based
	Sampling rate, sensor	20 KHz
	Sensor channels	3
	Sensor bandwidth	5.6 Khz
	Sensor resolution	18bit
	Output voltage	−30–135 V
	Amplifier channels	4
	Peak output power per channel	25 w
	Resolution DAC	24bit

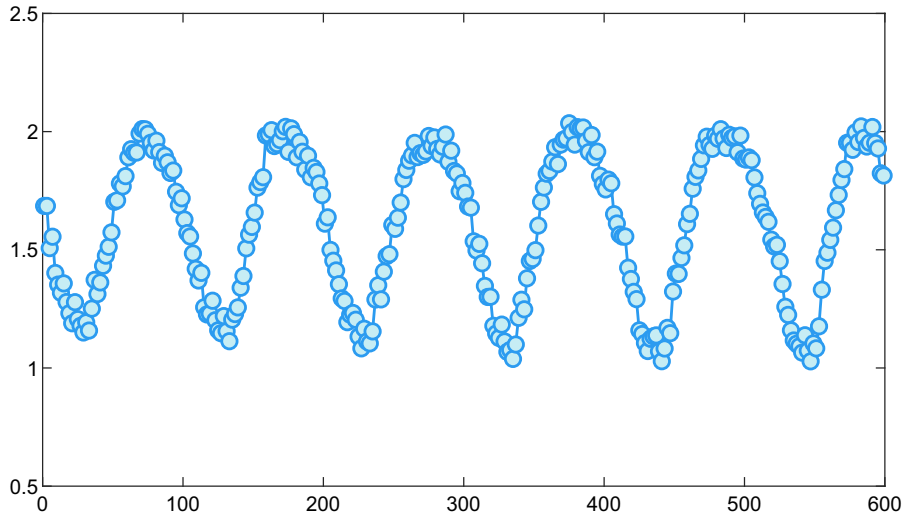

Fig. 17 Response under sine wave excitation

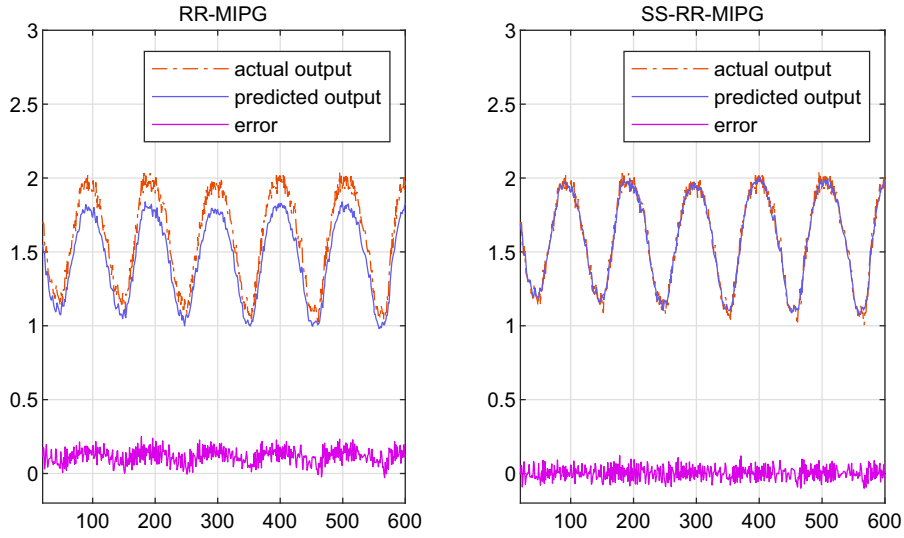

Fig. 18 The predicted results by RR-MIPG and SS-RR-MIPG ($p = 10$)

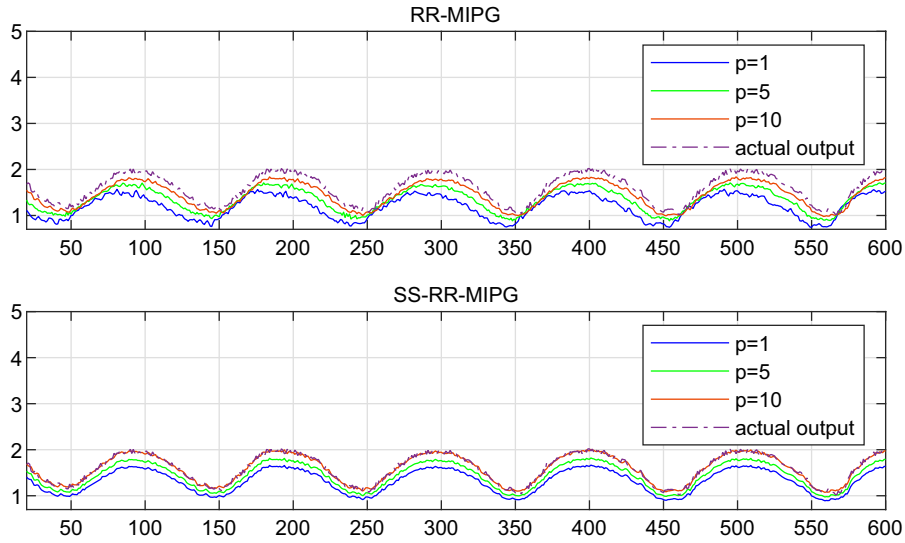

Fig. 19 The predicted results by RR-MIPG and SS-RR-MIPG under different innovation lengths ($\sigma^2 = 0.10^2$)

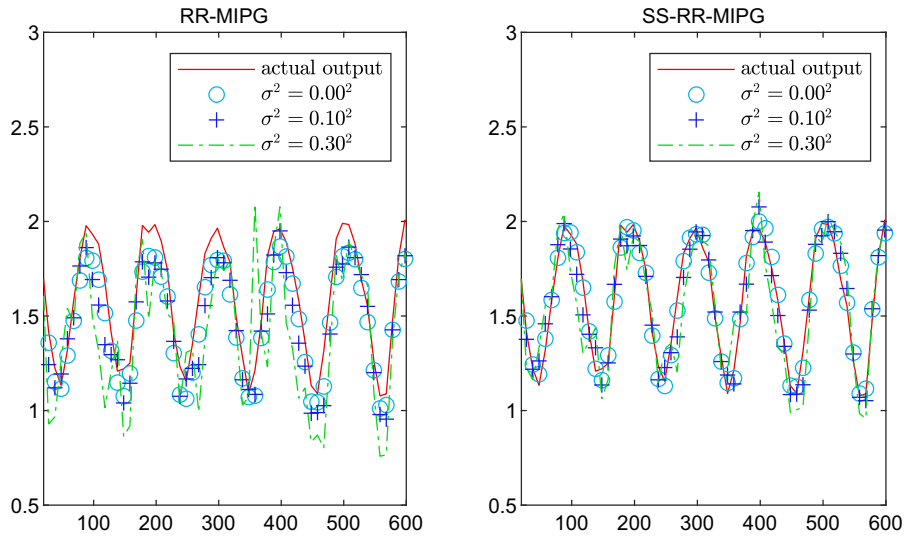

Fig. 20 The predicted results by RR-MIPG and SS-RR-MIPG under different noise variances ($p = 10$)

The modular digital multi-channel piezo controller structure is depicted in Fig. 16, with its functional list provided in Table 7. When the input signal is sinusoidal, the response of the piezoelectric ceramics is illustrated in Fig. 17. The predicted piezoelectric results using the redundant rule-based multi-innovation predictive gradient (RR-MIPG) and separable synchronous redundant rule-based multi-innovation predictive gradient (SS-RR-MIPG) algorithms are shown in Figs. 18, 19 and 20.

Based on the simulation results, the following conclusions are derived.

- Both the RR-MIPG and SS-RR-MIPG algorithms demonstrate predictive capabilities for the output of the piezoelectric ceramic flow control system. Upon further comparison, the SS-RR-MIPG algorithm demonstrates superior prediction accuracy over the RR-MIPG algorithm—see Fig. 18.
- The predictive capabilities given by the RR-MIPG and SS-RR-MIPG algorithms improve with t and p increasing—see Fig. 19.
- The RR-MIPG and SS-RR-MIPG algorithms have better predictive capabilities when the noise variance is smaller—see Fig. 20.

8 Conclusions

This paper focuses on the estimation problem of the nonlinear time-delay the exponential autoregressive with exogenous input (ExpARX) model and proposes the redundant rule-based multi-innovation predictive gradient (RR-MIPG) and separable synchronous redundant rule-based multi-innovation predictive gradient (SS-RR-MIPG) algorithms, which effectively overcome the shortcoming of traditional gradient algorithms easily falling into the local optima. Besides, the convergence analysis of the SS-RR-MIPG algorithm is presented. The computational efficiency of the RR-MIPG and SS-RR-MIPG algorithms is analyzed in detail. The results indicate that the SS-RR-MIPG algorithm is more computationally efficient than the RR-MIPG algorithm, and advantages gradually enhance with the increase of the vector dimension. Finally, simulation examples confirm the effectiveness of the RR-MIPG and the SS-RR-MIPG algorithms, revealing that the SS-RR-MIPG algorithm offers higher accuracy compared to the RR-MIPG algorithm. The RR-MIPG and SS-RR-MIPG algorithms can be integrated and applied with other identification theories in industries such as communication systems, dispensing systems, and intelligent power systems. However, considering the possible issue of data loss in practical applications, future research will delve into the estimation of nonlinear systems with missing data.

Funding This work was supported by the National Natural Science Foundation of China (No. 61903050) and the Natural Science Foundation of Shanghai (No. 22ZR1445300, No. 23ZR1446100).

Data availibility All data generated or analyzed during this study are included in this article.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Wang, Y.J., Liu, Y.T., Chen, J.H., Tang, S.H., Deng, M.Q.: Online identification of Hammerstein systems with B-spline networks. Int. J. Adapt. Control Signal Process. **38**(6), 2074–2092 (2024)
2. Xu, L.: Parameter estimation for nonlinear functions related to system responses. Int. J. Control Autom. Syst. **21**(6), 1780–1792 (2023)
3. Gu, Y., Dai, W., Zhu, Q.M., Nouri, H.: Hierarchical multi-innovation stochastic gradient identification algorithm for estimating a bilinear state-space model with moving average noise. J. Comput. Appl. Math. **420**, 114794 (2023)
4. Brouri, A., Giri, F.: Identification of series-parallel systems composed of linear and nonlinear blocks. Int. J. Adapt. Control Signal Process. **37**(8), 2021–2040 (2023)
5. Ding, F.: Least squares parameter estimation and multi-innovation least squares methods for linear fitting problems from noisy data. J. Comput. Appl. Math. **426**, 115107 (2023)
6. Xu, L.: Separable multi-innovation Newton iterative modeling algorithm for multi-frequency signals based on the sliding measurement window. Circuits Syst. Signal Process. **41**(2), 805–830 (2022)
7. Ji, Y., Liu, J., Liu, H.B.: An identification algorithm of generalized time-varying systems based on the Taylor series expansion and applied to a pH process. J. Process. Control **128**, 103007 (2023)
8. Ji, Y., Jiang, A.N.: Filtering-based accelerated estimation approach for generalized time-varying systems with disturbances and colored noises. IEEE Trans. Circuits Syst. II Express Briefs. **70**(1), 206–210 (2023)
9. Ji, Y., Kang, Z., Zhang, X.: Model recovery for multi-input signal-output nonlinear systems based on the compressed sensing recovery theory. J. Frankl. Inst. **359**(5), 2317–2339 (2022)
10. Li, M.H., Liu, X.M.: Maximum likelihood hierarchical least squares-based iterative identification for dual-rate stochastic systems. Int. J. Adapt. Control Signal Process. **35**(2), 240–261 (2021)
11. Li, M., Liu, X.: Iterative identification methods for a class of bilinear systems by using the particle filtering technique. Int. J. Adapt. Control Signal Process. **35**(10), 2056–2074 (2021)
12. Li, M.H., Liu, X.M.: 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. Int. J. Adapt. Control Signal Process. **33**(7), 1189–1211 (2019)
13. Wang, Y., Tang, S.: Parameter estimation for nonlinear Volterra systems by using the multi-innovation identification theory and tensor decomposition. J. Frankl. Inst. **359**(2), 1782–1802 (2022)
14. Wang, Y., Yang, L.: An efficient recursive identification algorithm for multilinear systems based on tensor decomposition. Int. J. Robust Nonlinear Control **31**(11), 7920–7936 (2021)
15. Wang, Y., Tang, S., Deng, M.: Modeling nonlinear systems using the tensor network B-spline and the multi-innovation identification theory. Int. J. Robust Nonlinear Control **32**(13), 7304–7318 (2022)
16. Liu, Y.F., Wang, Y., Chen, X.: Online hysteresis identification and compensation for piezoelectric actuators. IEEE Trans. Ind. Electron. **67**(7), 5595–5603 (2020)
17. Wang, Z.Y., Zhang, Z., Mao, J.Q.: Precision tracking control of piezoelectric actuator based on Bouc-Wen hysteresis compensator. Electron. Lett. **48**(23), 1459–1460 (2012)
18. Xie, Y.Q., Tan, Y.H., Dong, R.L.: Nonlinear modeling and decoupling control of XY micropositioning stages with piezoelectric actuators. IEEE/ASME Trans. Mechatron. **18**(3), 821–832 (2013)

19. Gan, M., Li, H.X., Peng, H.: A variable projection approach for efficient estimation of RBF-ARX model. IEEE Trans. Cybern. **45**(3), 462–471 (2015)
20. Xu, H., Xu, L., Shen, S.: Online identification methods for a class of Hammerstein nonlinear systemsusing the adaptive particle filtering. Chaos Solitons Fractals **186**, 115181 (2024)
21. Xu, H., Champagne, B.: Joint parameter and time-delay estimation for a class of nonlinear time-series models. IEEE Signal Process. Lett. **28**, 628–632 (2021)
22. Ozaki, T.: Non-linear time-series models for non-linear random vibrations. J. Appl. Probab. **17**(1), 84–93 (1980)
23. Liu, L.J., Xia, H.F., Ma, J.X., Li, F.: Auxiliary model-based maximum likelihood gradient iterative identification for feedback nonlinear systems. Optim. Control Appl. Methods **45**(5), 2346–2363 (2024)
24. Liu, L.J.: Decomposition-based maximum likelihood gradient iterative algorithm for multivariate systems with colored noise. Int. J. Robust Nonlinear Control **34**(11), 7265–7284 (2024)
25. Shahriari, F., Arefi, M., Luo, H., Yin, S.: Multistage parameter estimation algorithms for identification of bilinear systems. Nonlinear Dyn. **110**(3), 2635–2655 (2022)
26. Ding, F., Xu, L., Zhang, X., Ma, H.: Hierarchical gradient- and least-squares-based iterative estimation algorithms for input-nonlinear output-error systems by using the over-parameterization. Int. J. Robust Nonlinear Control **34**(2), 1120–1147 (2024)
27. Ding, F., Xu, L., Zhang, X., Zhou, Y., Luan, X.: Recursive identification methods for general stochastic systems with colored noises by using the hierarchical identification principle and the filtering identification idea. Annu. Rev. Control. **57**, 100942 (2024)
28. Xu, L.: Separable Newton recursive estimation method through system responses based on dynamically discrete measurements with increasing data length. Int. J. Control Autom. Syst. **20**(2), 432–443 (2022)
29. Ding, F., Yang, H.Z., Liu, F.: Performance analysis of stochastic gradient algorithms under weak conditions. Sci. China Ser. F-Inf. Sci. **51**(9), 1269–1280 (2008)
30. Xu, L.: Separable synthesis estimation methods and convergence analysis for multivariable systems. J. Comput. Appl. Math. **427**, 115104 (2023)
31. Ma, H.: A novel multi-innovation gradient support vector machine regression method. ISA Trans. **130**, 343–359 (2022)
32. Yang, D.: Multi-innovation gradient-based iterative identification methods for feedback nonlinear systems by using the decomposition technique. Int. J. Robust Nonlinear Control **33**(13), 7755–7773 (2023)
33. Yang, D., Yang, E.F.: Hierarchical gradient-based iterative parameter estimation algorithms for a nonlinear feedback system based on the hierarchical identification principle. Circuits Syst. Signal Process. **43**(1), 124–151 (2024)
34. Mao, Y.W., Xu, C., Chen, J.: Regularization based reweighted estimation algorithms for nonlinear systems in presence of outliers. Nonlinear Dyn. **112**(15), 13131–13146 (2024)
35. Xu, L.: Decomposition and composition modeling algorithms for control systems with colored noises. Int. J. Adapt. Control Signal Process. **38**(1), 255–278 (2024)
36. Ding, F., Shao, X.L., Xu, L., Zhang, X., Xu, H., Zhou, Y.H.: Filtered generalized iterative parameter identification for equation-error autoregressive models based on the filtering identification idea. Int. J. Adapt. Control Signal Process. **38**(4), 1363–1385 (2024)
37. Hu, C., Ji, Y., Ma, C.Q.: Joint two-stage multi-innovation recursive least squares parameter and fractional-order estimation algorithm for the fractional-order input nonlinear output-error autoregressive model. Int. J. Adapt. Control Signal Process. **37**(7), 1650–1670 (2023)
38. An, S., He, Y., Wang, L.J.: Maximum likelihood based multi-innovation stochastic gradient identification algorithms for bilinear stochastic systems with ARMA noise. Int. J. Adapt. Control Signal Process. **37**(10), 2690–2705 (2023)
39. Bi, Y.Q., Ji, Y.: Parameter estimation of fractional-order Hammerstein state space system based on the extended Kalman filter. Int. J. Adapt. Control Signal Process. **37**(7), 1827–1846 (2023)
40. Ding, F., Ma, H., Pan, J., Yang, E.F.: Hierarchical gradient- and least squares-based iterative algorithms for input nonlinear output-error systems using the key term separation. J. Frankl. Inst. **358**(9), 5113–5135 (2021)
41. Xu, L., Xu, H.: Adaptive multi-innovation gradient identification algorithms for a controlled autoregressive autoregressive moving average model. Circuits Syst. Signal Process. **43**(6), 3718–3747 (2024)
42. Xu, L., Zhu, Q.M.: Novel parameter estimation method for the systems with colored noises by using the filtering identification idea. Syst. Control Lett. **186**, 105774 (2024)
43. Ding, F., Chen, T., Qiu, L.: Bias compensation based recursive least squares identification algorithm for MISO systems. IEEE Trans. Circuits Syst. II Express Briefs **53**(5), 349–353 (2006)
44. Xu, N., Xu, L.: Convergence analysis of a synchronous gradient estimation scheme for time-varying parameter systems. J. Comput. Appl. Math. **443**, 115724 (2024)
45. Xu, N.: Recursive estimation algorithms based on the least squares and their convergence for a class of time-varying systems. Nonlinear Dyn. **111**(19), 18191–18213 (2023)
46. Hu, C., Liu, H.B., Ji, Y.: Parameter and order estimation algorithms and convergence analysis for lithium-ion batteries. Int. J. Robust Nonlinear Control **33**(18), 11411–11433 (2023)
47. Chai, G., Tan, Y.H., Dong, R.L., Long, X.C.: Predictive Gradient Based Control Using Hammerstein Model for MEMS Micromirrors. IEEE/ASME Trans. Mechatron. **29**(3), 1–13 (2023)
48. Miele, A., Cantrell, J.W.: Study on a memory gradient method for the minimization of functions. J. Optim. Theory Appl. **3**, 459–470 (1969)
49. Xing, H.M., Yang, E.F.: Highly-efficient filtered hierarchical identification algorithms for multiple-input multiple-output systems with colored noises. Syst. Control Lett. **186**, 105762 (2024)
50. Sun, S.Y., Sheng, J.: Filtered multi-innovation-based iterative identification methods for multivariate equation-error ARMA systems. Int. J. Adapt. Control Signal Process. **37**(3), 836–855 (2023)
51. Sun, S.Y., Wang, X.: Hierarchical iterative identification algorithms for a nonlinear system with dead-zon and sat-

uration nonlinearity based on the auxiliary model. Int. J. Adapt. Control Signal Process. **37**(7), 1866–1892 (2023)
52. Wang, X.Y., Ma, J.X., Xiong, W.L.: Expectation-maximization algorithm for bilinear state-space models with time-varying delays under non-Gaussian noise. Int. J. Adapt. Control Signal Process. **37**(10), 2706–2724 (2023)
53. Xing, H.M., Yang, E.F.: Hierarchical recursive least squares parameter estimation methods for multiple-input multiple-output systems by using the auxiliary models. Int. J. Adapt. Control Signal Process. **37**(11), 2983–3007 (2023)
54. Jin, Y.: A coupled recursive least squares algorithm for multivariable systems and its computational amount analysis by using the coupling identification concept. Int. J. Adapt. Control Signal Process. **38**(2), 513–533 (2024)
55. Zhang, X., Ding, F.: Adaptive parameter estimation for a general dynamical system with unknown states. Int. J. Robust Nonlinear Control **30**(4), 1351–1372 (2020)
56. Zhang, X., Ding, F., Xu, L.: Recursive parameter estimation methods and convergence analysis for a special class of nonlinear systems. Int. J. Robust Nonlinear Control **30**(4), 1373–1393 (2020)
57. Cao, Y., Sun, Y., Li, P.: Vibration-based fault diagnosis for railway point machines using multi-domain features, ensemble feature selection and SVM. IEEE Trans. Veh. Technol. **73**(1), 176–184 (2024)
58. Cao, Y., Liu, Y.: Enhancing rail safety through real-time defect detection: A novel lightweight network approach. Accid. Anal. Prevent. **203**, 107617 (2024)
59. Su, S., Zhu, Q., Liu, J.: A data-driven iterative learning approach for optimizing the train control strategy. IEEE Trans. Ind. Inf. **19**(7), 7885–7893 (2023)
60. Cao, Y., Su, S.: Fractional gradient descent algorithms for systems with outliers: a matrix fractional derivative or a scalar fractional derivative. Chaos Solitons Fractals **174**, 113881 (2023)
61. Cao, Y., An, Y.T., Su, S.: Is the safety index of modern safety integrity level (SIL) truly appropriate for the railway? Accid. Anal. Prevent. **192**, 107267 (2023)
62. Cao, Y., An, Y., Su, S., et al.: A statistical study of railway safety in China and Japan 1990–2020. Accidt. Anal. Prevent. **175**, 106764 (2022)
63. Su, S., She, J., Li, K., et al.: A nonlinear safety equilibrium spacing based model predictive control for virtually coupled train set over gradient terrains. IEEE Trans. Transp. Electr. **8**(2), 2810–2824 (2022)
64. Cao, Y., Yang, R., Ma, L.: Research on virtual coupled train control method based on GPC & VAPF. Chinese J. Electron. **31**(5), 897–905 (2022)
65. Cao, Y., Ji, Y., Sun, Y., Su, S.: The fault diagnosis of a switch machine based on deep random forest fusion. IEEE Intell. Transp. Syst. Mag. **15**(1), 437–452 (2023)
66. Su, S., Wang, X., Cao, Y., et al.: An energy-efficient train operation approach by integrating the metro timetabling and eco-driving. IEEE Trans. Intell. Transp. Syst. **21**(10), 4252–4268 (2020)
67. Cao, Y., Wang, Z., Liu, F., et al.: Bio-inspired speed curve optimization and sliding mode tracking control for subway trains. IEEE Trans. Veh. Technol. **68**(7), 6331–6342 (2019)
68. Su, S., Tang, T., Xu, J., et al.: Design of running grades for energy-efficient train regulation: A case study for Beijing Yizhuang line. IEEE Intell. Transp. Syst. Mag. **13**(2), 189–200 (2021)
69. Cao, Y., Sun, Y.K., Xie, G., et al.: Fault diagnosis of train plug door based on a hybrid criterion for IMFs selection and fractional wavelet package energy entropy. IEEE Trans. Veh. Technol. **68**(8), 7544–7551 (2019)
70. Tan, Y.H., Cheng, W.L., Dong, R.L.: Online optimizing positioning control with model error compensator for LEGRV system. IEEE/ASME Trans. Mechatron. **25**(2), 594–603 (2020)