

ITERATIVE LEARNING-BASED GRADIENT ESTIMATION FOR NONLINEAR THREE-BLOCK CASCADE SYSTEMS WITH QUANTIZED OBSERVATIONS

HAOZHE CAO, ZILONG CHEN AND LIHUA LI*

School of National Security
People's Public Security University of China
No. 1, Muxidi Nanli, Xicheng District, Beijing 100038, P. R. China
caohaozhe@ppsuc.edu.cn; 2024212010@stu.ppsuc.edu.cn
*Corresponding author: lilihua@ppsuc.edu.cn

Received September 2025; revised January 2026

ABSTRACT. *In this study, an iterative learning-based gradient estimation is proposed for identifying the parameters in nonlinear three-block cascade systems via quantized output. Different from existing modified MI-G algorithms, which solve the performance problem of multi-information length suppression estimation by designing some complex additional algorithms, this paper proposes an internal sub-innovation iterative updating gradient algorithm, the problem of multi-information length is solved from the dynamic data utilization viewpoints, and the additional methods are not required. First, a particular regression model, in which redundant parameter estimation problem is avoided, is constructed. Second, an iterative learning idea is incorporated into the update of the multi-innovation gradient algorithm adaptive law, so that the problem of the batch introduction of noise signals is solved. Then, the parameter estimation value is continuously updated and optimized by resorting to an activation mechanism. In addition, the convergence of estimation error to zero is strictly proven. Finally, the simulation example and experiment on motor-drive system results demonstrate the feasibility and applicability of the proposed estimation scheme.*

Keywords: Adaptive parameter estimation, Multi-innovation estimator, Parameter identification, Quantized data

1. Introduction. Quantization mechanisms are exploited to save communication resources in the age of interconnected systems and massive data. They collect data driven by a specific sensor, rather than the precise conventional sensors, to reduce the data communication rate on the premise of guaranteeing control system performance [1]. Thus, the quantized system has been widely used in practical industry applications, such as, network communications [2], control systems [3], and bio-pharmaceutical systems [4].

In modern industries, their components are diverse due to the complexity and strong nonlinearity of actual systems [5]. For this reason, using cascaded systems, i.e., block-oriented systems is a natural modeling approach for researchers and engineers to represent the characteristics of actual systems [6]. In recent years, quantized block-oriented systems have been studied and have become a research hotspot in the fields of system identification and automatic control [7]. Guo and Zhao [8] combined the empirical measure method to derive a multistage identification method for quantized nonlinear sandwich systems based on the gradient optimization technique. Ding et al. [9] examined the parameter estimation of binary-valued Hammerstein systems by proposing a kernel estimation approach based

on the mean square error method. Figwer [10] used a randomized search algorithm to estimate the sandwich systems under quantized measurements.

Along with control technology's fast development, numerous estimation algorithms integrating new theory and principle [11, 14, 15], such as the hierarchical identification methods [12], and multi-innovation methods [13], have been proposed and applied to actual systems. Among the aforementioned new identification versions, the Multi-Innovation Gradient algorithm (MI-G) has attracted attention because of its low computational complexity and ease of online implementation. Xu et al. [16] proposed a separable synchronous MI-G based on the parameter separation idea for nonlinear systems, in which high-accuracy performance and low complexity are ensured. Lughofer and Škrjanc [17] presented an improved MI-G estimator of fuzzy model, which gave a high estimator robustness.

Although the versions of the abovementioned MI-G are effective for nonlinear systems, the focus of research on the gathered data is precise data, and literature on quantifying nonlinear systems is limited. By contrast, the improved versions are all designed by adding additional methods, and the problems existing in MI-G have not been effectively solved. For instance, when the initial value of the MI-G is inappropriate, the convergence speed is slow, and the estimator requires additional samples to achieve parameter convergence. In the parameter updates of the MI-G, the batch introduction of noise leads to an increase in estimation performance within a certain range, followed by a sharp decline. Such phenomenon is the so-called innovation length limits estimation performance problem, and it has not yet been solved. In this study, we propose a novel MI-G, handle the above shortages and then apply the proposed algorithm to the quantitative three-block cascade systems identification problem, as shown in Figure 1.

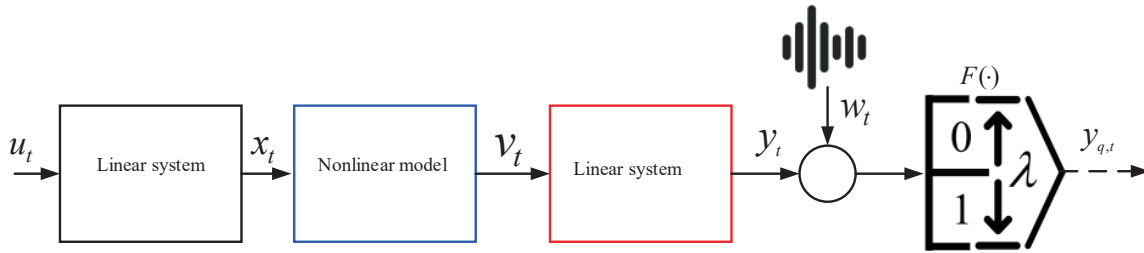


FIGURE 1. Quantized three-block cascade systems

Based on the above analysis and discussion, this study considers the identification problem under the framework of three-block cascade systems with quantized measurements. First, with the help of the half-decomposition technique, the quantized identification model is constructed, in which the computational cost is reduced. Second, by considering the iterative update principle, the parameter adaptive law batch update is decomposed into multiple subparameter adaptive law updates within a given multi-innovation length. Then, based on an activation mechanism, the estimated value achieves suboptimal estimation update before the next triggering action. The convergence performance of the estimator is rigorously proven. Finally, a simulation example and practical system are used to confirm the effectiveness of the proposed estimator.

The major contributions of the study are summarized into the following items.

1) Compared with the existing improved MI-G method [13, 18], the batch update of noise information is converted into a single innovation update means, where the multi-innovation length is resolved from the perspective of data utilization, and extra methods are not designed.

2) Different from the initial optimization in the existing improved MI-G method [20, 21], the initial value of the estimated parameter is continuously optimized based on value-triggered mechanism, in which the convergence speed is accelerated.

3) A reconstructed identification model for separating parameters of linear and nonlinear subsystems is introduced, where the redundant parameter estimation is escaped and computational complexity is decreased.

The remaining sections of this study are organized as follows. Section 2 introduces the system dynamic. Section 3 shows the design framework of the proposed algorithm. Section 4 provides the convergence of the developed method. Section 5 executes the simulation and experiment. Section 6 is a summary and discussion on future works.

2. Problem Statement. Considering the three-block cascade systems with binary-valued observations [22]

$$B(q^{-1})x_t = A(q^{-1})u_t, \quad (1)$$

$$D(q^{-1})y_t = C(q^{-1})v_t + w_t, \quad (2)$$

$$v_t = \begin{cases} K_z(x_t + d_z) & \text{when } x_t \leq d_z \\ 0 & \text{when } -d_z < x_t < d_r \\ K_r(x_t - d_r) & \text{when } x_t \geq d_r, \end{cases} \quad (3)$$

$$y_{q,t} = F(y_t), \quad (4)$$

$$F(y_t) = \begin{cases} 1 & \text{when } y_t < 1 \\ -1 & \text{when } y_t > 1, \end{cases} \quad (5)$$

where u_t , y_t , and w_t are system input, output and noise signal, respectively. $F(y_t)$ is quantizer with threshold 1, and $y_{q,t}$ represents the quantized data by using quantizer $F(y_t)$. $A(q^{-1})$, $B(q^{-1})$, $C(q^{-1})$ and $D(q^{-1})$ are polynomials with the backward operator q . Their equations are $A(q^{-1}) = \sum_{i=1}^{n_a} a_i q^{-i}$, $B(q^{-1}) = 1 + \sum_{j=1}^{n_b} b_j q^{-j}$, $C(q^{-1}) = \sum_{i=1}^{n_c} c_i q^{-i}$, and $D(q^{-1}) = 1 + \sum_{j=1}^{n_d} d_j q^{-j}$. v_t is the output of deadzone nonlinearity with the slopes K_z , K_r and cross points d_z , d_r .

Assuming that the polynomial coefficients a_1 and c_1 are set to one, which is used to gain the parameter unique model.

By using half-substitution technology [23] and combining (1)-(4), one can obtain

$$y_t = \varphi_t^T \theta + w_t, \quad (6)$$

where the data vector φ_t and parameter vector θ are

$$\varphi_t = [h_{1,t-1}u_{t-2}, h_{1,t-1}u_{t-3}, \dots, h_{1,t-1}u_{t-n_a-1}, -h_{1,t-1}x_{t-2}, \dots, -h_{1,t-1}x_{t-n_b-1}, \\ h_{1,t-1}, h_{2,t-1}x_{t-1}, v_{t-2}, v_{t-3}, \dots, v_{t-n_c}, -y_{t-1}, -y_{t-2}, \dots, -y_{t-n_d}]^T, \quad (7)$$

and

$$\theta = [K_z c_1 a_1, K_z c_1 a_2, \dots, K_z c_1 a_{n_a}, K_z c_1 b_1, \dots, K_z c_1 b_{n_b}, K_z c_1 d_z, K_r c_1, K_r d_r c_1, \\ c_2, \dots, c_{n_c}, d_1, d_2, \dots, d_{n_d}]^T, \quad (8)$$

where $h_{1,t} = s[x_t - d_z]$ and $h_{2,t} = g[d_r - x_t]$ describe two switching functions, and $s(x)$ is given as

$$s(x) = \begin{cases} 0 & \text{if } x > 0 \\ 1 & \text{if } x \leq 0. \end{cases} \quad (9)$$

This paper aims at developing an improved MI-G method to estimate the parameter information of considered systems based on the collected data $\{u_t, y_{q,t}\}$, to discuss the convergence of the estimator, and to validate the rationality of the proposed method.

3. Estimation Design. In this section, to address with noise problem, we modify the batch addition of noise data to a scalar feeding way, and avoid the defect of noise addition. On the other hand, in order to resolve initial value issue, the particular method is that when the sub-innovation length reaches the maximum of multi-innovation, the next sub-innovation is activated, and the estimated value at the previous time is used as the initial value for parameter update at the current time.

By using the batch data L , the typical stochastic gradient algorithm is modified based on multi-innovation theory, which is the MI-G algorithm, and its expression is as follows:

$$\Xi_{L,t} = Y_{L,t} - \Phi_{L,t}^T \hat{\theta}_{t-1}, \quad (10)$$

$$R_{L,t} = R_{L,t} + \|\Phi_{L,t}\|^2, \quad (11)$$

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{1}{R_{L,t}} \Phi_{L,t} \Xi_{L,t}, \quad (12)$$

where L is also known as the multi-innovation length, $\Phi_{L,t}$ is the multi-innovation data matrix, which is given by $\Phi_{L,t} = [\varphi_t, \varphi_{t-1}, \dots, \varphi_{t-L+1}]$, $\Xi_{L,t}$ is the multi-innovation vector, which is defined as $\Xi_{L,t} = [e_t, e_{t-1}, \dots, e_{t-L+1}]^T$, and the multi-innovation output vector is denoted by $Y_{L,t}$, whose form is $Y_{L,t} = [y_{q,t}, y_{q,t-1}, \dots, y_{q,t-L+1}]^T$.

Remark 3.1. According to multi-innovation theory, estimation performance should continuously improve with the increase in L , but the identification performance first lifts and then decreases. The reason is that when $\Xi_{L,t}$ is used for parameter estimation updates, batch noise signals continue to be fed, deteriorating its performance.

To address the aforementioned problem, this paper considers the iterative learning idea, and uses an activation mechanism to solve the initial value issue. Define the variables $y_{t,j} = y_{t-L+j}$, $j = 1, 2, \dots, L$, $\varphi_{t,j} = \varphi_{t-L+j}$ and $e_{t,j} = e_{t-L+j}$, the proposed Iterative Learning-based Gradient estimation (IL-G) is written as

$$e_{t,j} = y_{t,j} - \varphi_{t,j}^T \theta_{t,j-1}, \quad (13)$$

$$R_{t,j} = R_{t,j-1} + \|\varphi_{t,j}\|^2, \quad (14)$$

$$\hat{\theta}_{t,j} = \hat{\theta}_{t,j-1} + \frac{\varphi_{t,j}}{R_{t,j}} (y_{t,j} - \varphi_{t,j}^T \theta_{t,j-1}), \quad (15)$$

where $e_{t,j}$ is the j th sub-innovation variable, and $\hat{\theta}_{t,j}$ denotes the j th sub-innovation update of the parameter estimation.

The initial value triggering activation update is defined as

$$R_{t,0} = R_{t-1,L}, \quad (16)$$

$$\hat{\theta}_{t,0} = \hat{\theta}_{t-1,L}, \quad (17)$$

where if $j = L$, we use $R_{t-1,L}$ as the initial value for $R_{t,0}$ at the time t , and apply $\hat{\theta}_{t-1,L}$ as the initial value of $\hat{\theta}_{t,0}$ at the time t . $j = L$ denotes the triggering activation condition. Equations (16) and (17) are named the initial value-triggered mechanism. $R_{0,1} = 1$.

Remark 3.2. Compared to the existing MI-G methods, at the current time $t' = t$ and the j th sub-innovation, IL-G uses the scalar innovation $e_{t,j}$ to achieve the adaptive law update, so that noise data are not introduced in batch manner. When $j = L$, the parameter estimation of the next time is activated.

Remark 3.3. When $t' = t$ and $j = L$, the step and initial value of the estimated parameter by using the activation mechanism update (16) and (17) will be conducted. The optimization value at the time t via L iterations is chosen as the initial values of step and parameter vector at time $t = t + 1$, improving the convergence rate.

As displayed in Figure 1, to address the unknown variable problems of v_t and x_t , the reference model method is an effective way. Its fundamental principle is to use the raw system to construct the reference models and to replace the signals based on the outputs of the reference models. Interested readers can refer to [21].

4. Convergence Analysis. Let us suppose, $\{\mathcal{F}(t)\}$ is established by using w_t , we named $\{w_t, \mathcal{F}(t)\}$ as a bounded martingale sequence. Suppose the additive noise w_t satisfies

- (F1) $E[w_t | \mathcal{F}(t-1)] = 0$,
- (F2) $E[\|w_t\|^2 | \mathcal{F}(t-1)] = \sigma_{w,t}^2 \leq \sigma_w^2 < \infty$,
- (F3) $\limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{g=1}^t \|w_g\|^2 \leq \sigma_w^2 < \infty$.

Theorem 4.1. *For the quantized three-block cascade systems (1)-(4) and the presented method (13)-(17), when u_t is PE condition, then $R_t = O\left(\lambda_{\min}\left[\sum_{m=1}^t \sum_{j=1}^L \hat{\varphi}_{m,j} \hat{\varphi}_{m,j}^T\right]\right)$, $t \rightarrow \infty$, and (F1), (F2) and (F3) hold. Then, the estimated parameters can converge to zero, that is, $\|\tilde{\theta}_t\|^2 = \|\hat{\theta}_{t,L} - \theta\|^2 \rightarrow 0$, $t \rightarrow \infty$.*

Proof: Based on (15) with subtracting θ , it can be written as follows:

$$\tilde{\theta}_{t,j} = \tilde{\theta}_{t,j-1} + \frac{1}{R_{t,j}} \hat{\varphi}_{t,j} [-\tilde{y}(t, j-1) + w_{t,j}], \quad (18)$$

where $\tilde{y}(t, j-1) = \hat{\varphi}_{t,j}^T \tilde{\theta}_{t,j-1}$.

By using the definition of norm and combining (18), it yields

$$\begin{aligned} \|\tilde{\theta}_{t,j}\|^2 &= \|\tilde{\theta}_{t,j-1}\|^2 + \frac{-2R_{t,j} + \|\hat{\varphi}_{t,j}\|^2}{R_{t,j}^2} \|\tilde{y}(t, j-1)\|^2 + \frac{\|\hat{\varphi}_{t,j}\|^2 \|w_{t,j}\|^2}{R_{t,j}^2} \\ &\quad + \frac{2R_{t,j} - 2\|\hat{\varphi}_{t,j}\|^2}{R_{t,j}^2} (\tilde{y}(t, j-1)w_{t,j}). \end{aligned} \quad (19)$$

Define $T_{L,t} = \|\tilde{\theta}_{t,j}\|^2$, by using the scaling method to (19), we get

$$T_{L,t} \leq T_{L,t-1} + \sum_{j=1}^L \frac{\|\hat{\varphi}_{t,j}\|^2 \|w_{t,j}\|^2}{R_{t,j}^2} - \sum_{j=1}^L \frac{1}{R_{t,j}^2} \|\tilde{y}(t, j-1)\|^2 + \frac{2R_{t,j-1}}{R_{t,j}^2} [\tilde{y}(t, j-1)w_{t,j}]. \quad (20)$$

Because

$$\sum_{j=1}^L \sum_{t=1}^{\infty} \frac{\|\hat{\varphi}_{t,j}\|^2}{R_{t,j}^2} \leq \sum_{j=1}^L \sum_{t=1}^{\infty} \frac{R_{t,j} - R_{t,j-1}}{R_{t,j-1}R_{t,j}} = \sum_{j=1}^L \left[\frac{1}{R_{0,j}} - \frac{1}{R_{\infty,j}} \right] < \infty, \quad (21)$$

we can see that (21) is bounded. By using the definition of the expectation and combining (F1)-(F3), (20) has the following expression:

$$E[T_{L,t} | \mathcal{F}(t-1)] \leq T_{L,t-1} + \frac{L\sigma_w^2}{R_{t,j}^2} \sum_{j=1}^L \|\hat{\varphi}_{t,j}\|^2 - \sum_{j=1}^L \frac{1}{R_{t,j}^2} \|\tilde{y}(t, j-1)\|^2. \quad (22)$$

Based on (21), the following inequality holds:

$$L\sigma_w^2 \sum_{j=1}^L \sum_{t=1}^{\infty} \frac{\|\hat{\varphi}_{t,j}\|^2}{R_{t,j}^2} < \infty. \quad (23)$$

By applying martingale convergence theorem to (22) and using (23), we have

$$E \left[\left\| \tilde{\theta}_{t,L} \right\|^2 \right] \rightarrow T_0, t \rightarrow \infty, \quad (24)$$

and

$$\lim_{t \rightarrow \infty} \sum_{j=1}^L \sum_{m=1}^t \frac{1}{R_{m,j}} \|\tilde{y}(m, j-1)\|^2 < \infty, a.s., \quad (25)$$

where T_0 represents a random variable.

By exploiting Kronecker Lemma to (25), we have

$$\sum_{j=1}^L \frac{1}{R_{m,j}} \sum_{m=1}^t \|\tilde{y}(m, j-1)\|^2 \rightarrow 0, \quad t \rightarrow \infty. \quad (26)$$

Because of $\tilde{y}(t, j-1) = \hat{\varphi}_{t,j}^T \tilde{\theta}_{t,j-1}$, the following expression holds:

$$\tilde{y}(t, j-1) = \hat{\varphi}_{t,j}^T \left(\tilde{\theta}_{t,j} - \frac{1}{R_{t,j}} \hat{\varphi}_{t,j} (-\tilde{y}(t, j) + w_{t,j}) \right). \quad (27)$$

By using the definition of the determinant, for (27), we have

$$\|\tilde{y}(t, j-1)\| = \left\| \hat{\varphi}_{t,j}^T \tilde{\theta}_{t,j} + \frac{\hat{\varphi}_{t,j}^T \hat{\varphi}_{t,j}}{R_{t,j}} \tilde{y}(t, j) - \frac{\hat{\varphi}_{t,j}^T \hat{\varphi}_{t,j}}{R_{t,j}} w_{t,j} \right\|. \quad (28)$$

Squaring on both sides of (28) yields

$$\begin{aligned} \|\tilde{y}(t, j-1)\|^2 &= \left(\hat{\varphi}_{t,j}^T \tilde{\theta}_{t,j} \right)^2 + \left(\frac{\hat{\varphi}_{t,j}^T \hat{\varphi}_{t,j}}{R_{t,j}} \tilde{y}(t, j) \right)^2 + \left(\frac{\hat{\varphi}_{t,j}^T \hat{\varphi}_{t,j}}{R_{t,j}} w_{t,j} \right)^2 \\ &\quad + 2 \hat{\varphi}_{t,j}^T \tilde{\theta}_{t,j} \frac{\hat{\varphi}_{t,j}^T \hat{\varphi}_{t,j}}{R_{t,j}} \tilde{y}(t, j) - 2 \frac{\hat{\varphi}_{t,j}^T \hat{\varphi}_{t,j}}{R_{t,j}} \tilde{y}(t, j) \frac{\hat{\varphi}_{t,j}^T \hat{\varphi}_{t,j}}{R_{t,j}} w_{t,j} \\ &\quad - 2 \hat{\varphi}_{t,j}^T \tilde{\theta}_{t,j} \frac{\hat{\varphi}_{t,j}^T \hat{\varphi}_{t,j}}{R_{t,j}} w_{t,j} \\ &= \|\tilde{y}(t, j)\|^2 + \|\hat{\varphi}_{t,j}\|^2 \frac{\|\hat{\varphi}_{t,j}\|^2}{\|R_{t,j}\|^2} \|\tilde{y}(t, j)\|^2 + \|\hat{\varphi}_{t,j}\|^2 \frac{\|\hat{\varphi}_{t,j}\|^2}{\|R_{t,j}\|^2} \|w_{t,j}\|^2 \\ &\quad + 2 \|\tilde{y}(t, j)\|^2 \frac{\|\hat{\varphi}_{t,j}\|^2}{R_{t,j}} - 2 \|\hat{\varphi}_{t,j}\|^2 \frac{\|\hat{\varphi}_{t,j}\|^2}{\|R_{t,j}\|^2} \tilde{y}(t, j) w_{t,j} \\ &\quad - 2 \frac{\|\hat{\varphi}_{t,j}\|^2}{R_{t,j}} \tilde{y}(t, j) w_{t,j} \\ &\leq \|\tilde{y}(t, j)\|^2 + \|\hat{\varphi}_{t,j}\|^2 \frac{\|\hat{\varphi}_{t,j}\|^2}{\|R_{t,j}\|^2} \|\tilde{y}(t, j)\|^2 + \|\hat{\varphi}_{t,j}\|^2 \frac{\|\hat{\varphi}_{t,j}\|^2}{\|R_{t,j}\|^2} \|w_{t,j}\|^2 \\ &\quad - 2 \|\hat{\varphi}_{t,j}\|^2 \frac{\|\hat{\varphi}_{t,j}\|^2}{\|R_{t,j}\|^2} \tilde{y}(t, j) w_{t,j}. \end{aligned} \quad (29)$$

Based on Equation (29), we obtain

$$\begin{aligned} \left\| \hat{\varphi}_{t,j}^T \tilde{\theta}_{t,j-1} \right\|^2 &\leq \|\tilde{y}(t, j)\|^2 + \|\hat{\varphi}_{t,j}\|^2 \frac{\|\hat{\varphi}_{t,j}\|^2}{\|R_{t,j}\|^2} \|\tilde{y}(t, j)\|^2 + \|\hat{\varphi}_{t,j}\|^2 \frac{\|\hat{\varphi}_{t,j}\|^2}{\|R_{t,j}\|^2} \|w_{t,j}\|^2 \\ &\quad - 2 \|\hat{\varphi}_{t,j}\|^2 \frac{\|\hat{\varphi}_{t,j}\|^2}{\|R_{t,j}\|^2} \tilde{y}(t, j) w_{t,j} \end{aligned}$$

$$\begin{aligned}
&\leq 2 \|\tilde{y}(t, j)\|^2 + 2 \|\hat{\varphi}_{t,j}\|^2 \frac{\|\hat{\varphi}_{t,j}\|^2}{\|R_{t,j}\|^2} \|\tilde{y}(t, j)\|^2 + 2 \|\hat{\varphi}_{t,j}\|^2 \frac{\|\hat{\varphi}_{t,j}\|^2}{\|R_{t,j}\|^2} \|w_{t,j}\|^2 \\
&\quad - 4 \|\hat{\varphi}_{t,j}\|^2 \frac{\|\hat{\varphi}_{t,j}\|^2}{\|R_{t,j}\|^2} \tilde{y}(t, j) w_{t,j}.
\end{aligned} \tag{30}$$

Summing the given length of the innovation with $i = 1, 2, \dots, j-1$, (30) can be rewritten as

$$\begin{aligned}
\left\| \hat{\varphi}_{t,j}^T \tilde{\theta}_{t,i-1} \right\|^2 &\leq 2 \|\tilde{y}(t, j)\|^2 + 2 \|\hat{\varphi}_{t,j}\|^2 \sum_{i=0}^{j-1} \frac{\|\varphi_{t-i,j}\|^2}{\|R_{t-i,j}\|^2} \|\tilde{y}(t-i, j)\|^2 \\
&\quad - 4 \|\hat{\varphi}_{t,j}\|^2 \sum_{i=0}^{j-1} \frac{\|\varphi_{t-i,j}\|^2}{\|R_{t-i,j}\|^2} \tilde{y}(t-i, j) w_{t-i,j} \\
&\quad + 2 \|\hat{\varphi}_{t,j}\|^2 \sum_{i=0}^{j-1} \frac{\|\varphi_{t-i,j}\|^2}{\|R_{t-i,j}\|^2} \|w_{t-i,j}\|^2.
\end{aligned} \tag{31}$$

By using a method similar to (20), it has

$$\begin{aligned}
E[Z_{L,t} | \mathcal{F}(t-1)] &\leq \sum_{j=1}^L 2 \|\tilde{y}(t, j)\|^2 + 2 \sum_{j=1}^L \|\hat{\varphi}_{t,j}\|^2 \sum_{i=0}^{j-1} \frac{\|\hat{\varphi}_{t-i,j}\|^2}{\|R_{t-i,j}\|^2} \|\tilde{y}(t-i, j)\|^2 \\
&\quad + 2L \|\hat{\varphi}_{t,j}\|^2 \sum_{i=0}^{j-1} \frac{\|\hat{\varphi}_{t-i,j}\|^2}{\|R_{t-i,j}\|^2} \sigma_w^2,
\end{aligned} \tag{32}$$

where $Z_{L,t} = \left\| \hat{\varphi}^T(t, j) \tilde{\theta}_{t,j} \right\|^2$.

By using $R_{t,j}$ to (32), we have

$$\frac{E[Z_{L,t} | \mathcal{F}(t-1)]}{R_{t,j}} \leq \frac{2}{R_{t,j}} \sum_{i=1}^t \sum_{j=1}^L \|\tilde{y}(i, j)\|^2 + \frac{2}{R_{t,j}} \sum_{i=2}^t \sum_{j=1}^L \frac{\|\hat{\varphi}_{i,j}\|^2}{R_{i,j}} [\|\tilde{y}(i, j)\|^2 + L\sigma_w^2]. \tag{33}$$

Because of $\sum_{j=1}^L \sum_{i=1}^t \frac{\|\hat{\varphi}_{i,j}\|^2}{R_{i,j}} \leq \ln R_{t,j}$, (33) has the following form:

$$\frac{E[Z_{L,t} | \mathcal{F}(t-1)]}{R_{t,j}} \leq \frac{4}{R_{t,j}} \sum_{i=1}^t \sum_{j=1}^L \|\tilde{y}(i, j)\|^2 + \frac{2 \ln R_{t,j}}{R_{t,j}} (L\sigma_w^2) \rightarrow 0, \text{ as } t \rightarrow \infty. \tag{34}$$

It is a fact that

$$\lambda_{\min} \left[\sum_{j=1}^L \sum_{i=1}^t \hat{\varphi}_{i,j} \hat{\varphi}_{i,j}^T \right] \|\tilde{\theta}_t\|^2 \leq \tilde{\theta}_t^T \left[\sum_{j=1}^L \sum_{i=1}^t \hat{\varphi}_{i,j} \hat{\varphi}_{i,j}^T \right] \tilde{\theta}_t \leq \lambda_{\max} \left[\sum_{j=1}^L \sum_{i=1}^t \hat{\varphi}_{i,j} \hat{\varphi}_{i,j}^T \right] \|\tilde{\theta}_t\|^2,$$

and it has

$$\frac{\lambda_{\min} \left[\sum_{j=1}^L \sum_{i=1}^t \hat{\varphi}_{i,j} \hat{\varphi}_{i,j}^T \right] \|\tilde{\theta}_t\|^2}{R_{i,j}} \leq \frac{E[Z_{L,t} | \mathcal{F}(t-1)]}{R_{t,j}} \rightarrow 0, \text{ } t \rightarrow \infty. \tag{35}$$

Because of $R_t = O \left(\lambda_{\min} \left[\sum_{m=1}^t \sum_{j=1}^L \hat{\varphi}_{m,j} \hat{\varphi}_{m,j}^T \right] \right)$, $t \rightarrow \infty$, Equation (35) is written as

$$\|\tilde{\theta}_t\|^2 = O \left(R_t / \left(\lambda_{\min} \left[\sum_{m=1}^t \sum_{j=1}^L \hat{\varphi}_{m,j} \hat{\varphi}_{m,j}^T \right] \right) \right) \rightarrow 0, \text{ a.s. } t \rightarrow \infty. \tag{36}$$

To this end, Theorem 4.1 has been completed. $\square$

5. Examples. To compare the performance of the proposed algorithm, two improved versions of MI-G are used as comparison algorithms, such as Filtering Data-Multi-Innovation Gradient method (FD-MI-G) [18], and Separable Synchronous Multi-Innovation Gradient method (SS-MI-G) [16]. Before giving the examples, we provide the parameter selection criteria for the proposed algorithm. 1) Initial value of the estimated vector: when the true value of the estimated value is unknown, the common operation is to set the estimated initial value to a very small value, e.g., 0.001, 0.001. 2) Step size and innovation length: Based on the parameter selection criteria of multi-innovation gradient algorithm, step size $R_{0,1} = 1$, $L = 4$.

5.1. Simulation example. Consider the following quantized three-block cascade systems

$$(0.5q^{-1} + 0.45q^{-2})x_t = (q^{-1} + 0.35q^{-2})u_t, \quad (37)$$

$$(0.4q^{-1} + 0.3q^{-2})y_t = (q^{-1} + 0.1q^{-2})v_t + w_t, \quad (38)$$

$$v_t = \begin{cases} 0.4(x_t + 0.1) & \text{when } x_t \leq 0.1 \\ 0 & \text{when } -0.1 < x_t < 0.1 \\ 0.4(x_t - 0.1) & \text{when } x_t \geq 0.1, \end{cases} \quad (39)$$

$$y_{q,t} = F(y_t), \quad (40)$$

$$F(y_t) = \begin{cases} 1 & \text{when } y_t < 1 \\ -1 & \text{when } y_t > 1. \end{cases} \quad (41)$$

In simulation, $\{u_t\}$ is sequence of random signal with Gaussian distribution, and $\{w_t\}$ is signal of white noise with zero mean and standard deviation $\sigma = 0.1$. The main parameters of the proposed method are $\hat{\theta}_{0,0} = I/p_0$, $L = 4$, $p_0 = 1e6$, $R_{0,1} = 1$, $N = 1200$. The four methods are used to estimate the parameters of system, and comparative statistical results are shown in Figures 2-4. The estimated parameters using the proposed IL-G have a remarkably faster convergence rate compared with that using FD-MI-G and SS-MI-G because the proposed IL-G requires less time for searching true values. In estimating the accuracy, the developed IL-G can converge to true value as the addition of samples, while FD-MI-G and SS-MI-G oscillate around the expected value with different accuracies, demonstrating the enhanced performance of the proposed estimator. Figure 5 shows the comparison estimation errors. It can be seen that the error curve of the proposed algorithm is closer to the x -axis than the other two algorithms, indicating the high estimation accuracy of the designed method. At the beginning stage of parameter estimation, the error curve shown by the proposed scheme tends to be closer to the original point and reaches stability, indicating the fast convergence speed of the proposed scheme.

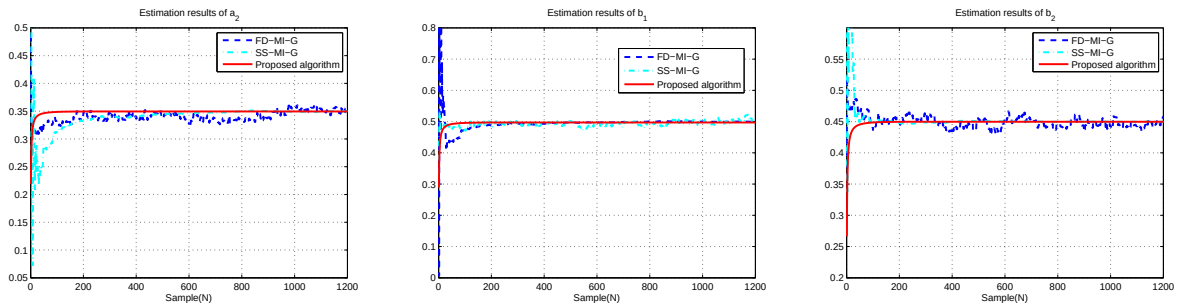


FIGURE 2. Estimates of first linear system

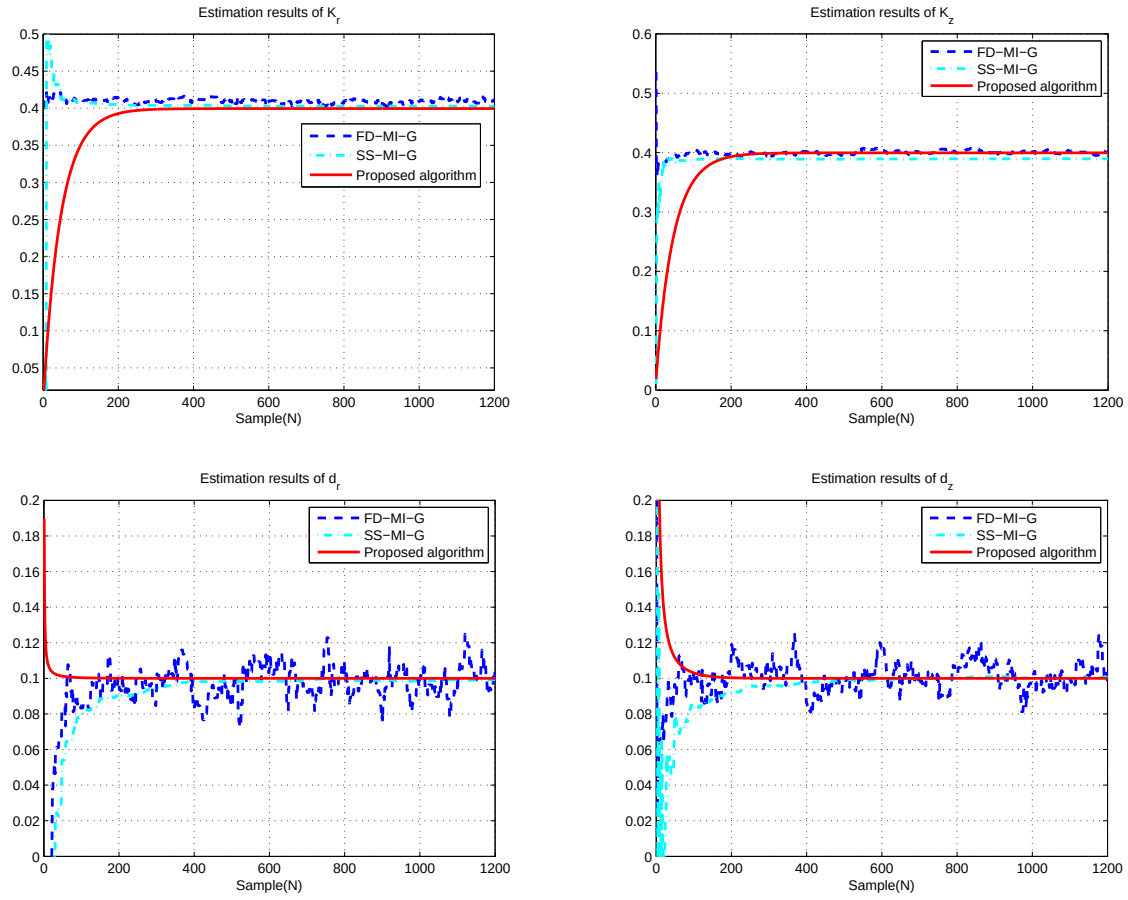


FIGURE 3. Estimates of deadzone model

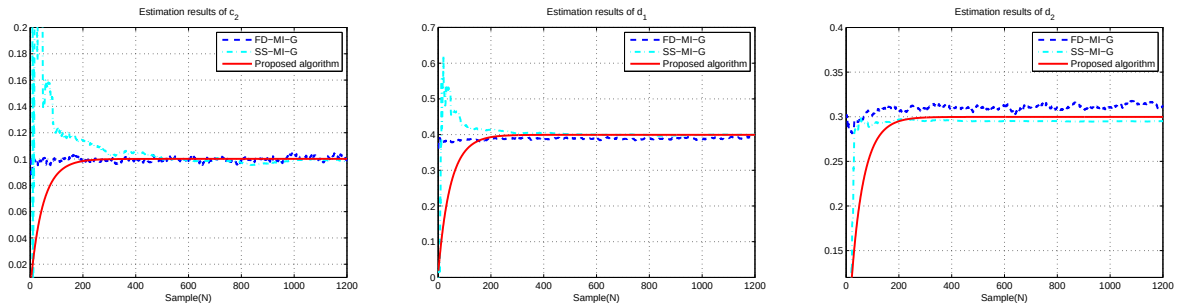


FIGURE 4. Estimates of second linear system

To show the tracking performance of the established model, we compare our method with the FD-MI-G and SS-MI-G methods by conducting model validation, as plotted in Figure 6. The three estimation models can track the dynamic outputs of the actual system, which implies the effectiveness of the three estimators and the rationality of the established estimation models. Further analysis reveals that our method yields more accurate model tracking performance with minimal output error compared with the other two comparison approaches, evidencing the positive performance of our scheme. Figure 7 exhibits the estimation performance using our method based on different noise intensities. As shown in Figure 7, the weaker noise intensity indicates the more accurate estimation performance, and a stronger case degrades identification ability. From $\delta^2 = 0.01$ to $\delta^2 = 1$,

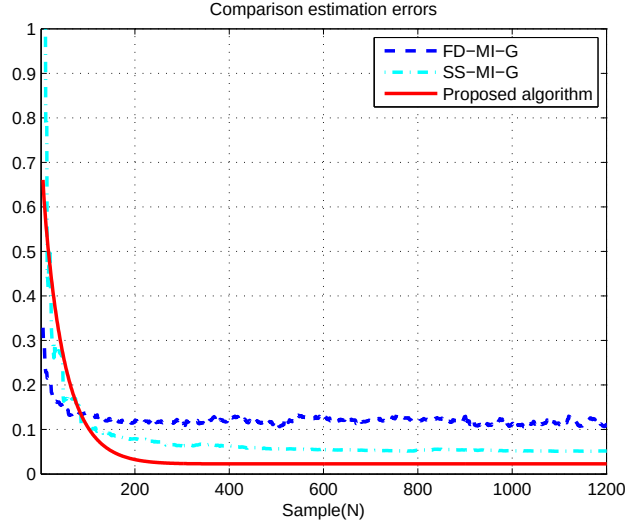


FIGURE 5. Estimation error results

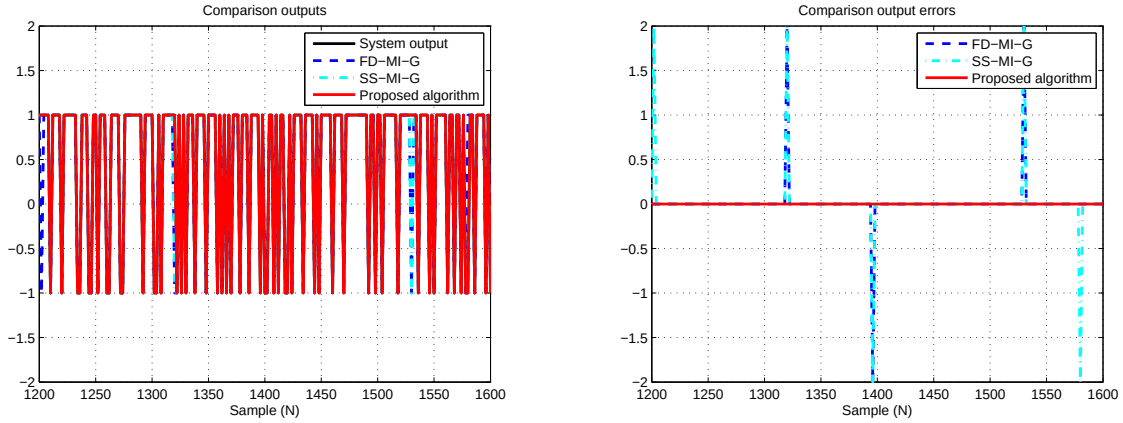


FIGURE 6. Model validation

for quantified nonlinear systems, the estimation performance is clearly satisfactory due to the strong coupling of the system and the destructive nature of quantitative data, which suggests that our algorithm can achieve insensitivity to noise signals within a certain range. Even though the noise intensity continues to increase ($\delta^2 = 4$), no sharp decrease in estimation performance occurs. The reason for analyzing this phenomenon is that when the parameter estimation vector $\hat{\theta}_{t,j}$ in (15) is updated by using the scalar introduction of noise of $e_{t,j}$ in (13) rather than the batch noise data $\Xi_{L,t}$ in Equation (10), so that the utilization rate of noise data is much lower than that of useful system data, resulting in a gradual decline in estimation performance, which is consistent with the analysis of Remark 3.2.

5.2. Experimental confirmation. Experiment validation is conducted to test the practicality of the designed estimator. The test platform consists of a PWM drive motor, a control card, digital encoders, and a PC, as displayed in Figure 8. The controller environment is set in C program, the graphical results are shown on the PC, and the mode of the drive motor is torque control manner. The quantized output data are gathered by using the output of motor and a digital binary-value quantizer. The sampling time is chosen as $T_s = 0.01\text{sec}$. The parameters of the motor drive system are listed in Table 1.

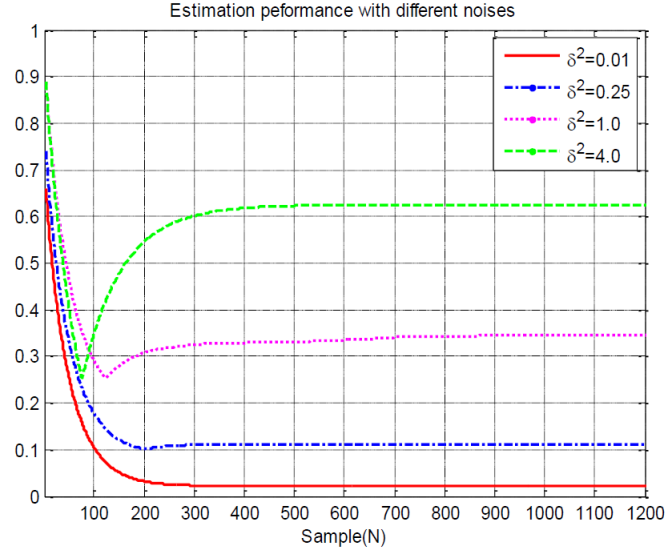


FIGURE 7. Estimation performance with different noises



FIGURE 8. Experiment validation platform

TABLE 1. Parameter and meaning

Variables	Units	Meaning
J_m	kg/m^2	Rotational inertia
L_c	H	Inductance
R_c	Ω	Resistance
I_c	A	Armature current
K_0	/	Back-electromotive coefficient
K_c	/	Electrical time constant

The parameter optimization results of three estimators are plotted in Figure 9. Three estimation curves transition from oscillation to stationarity, indicating that all estimators can optimize the parameter information of the system, but the optimization performance is various. Compared with the curves gained using the FD-MI-G, the SS-MI-G has smaller

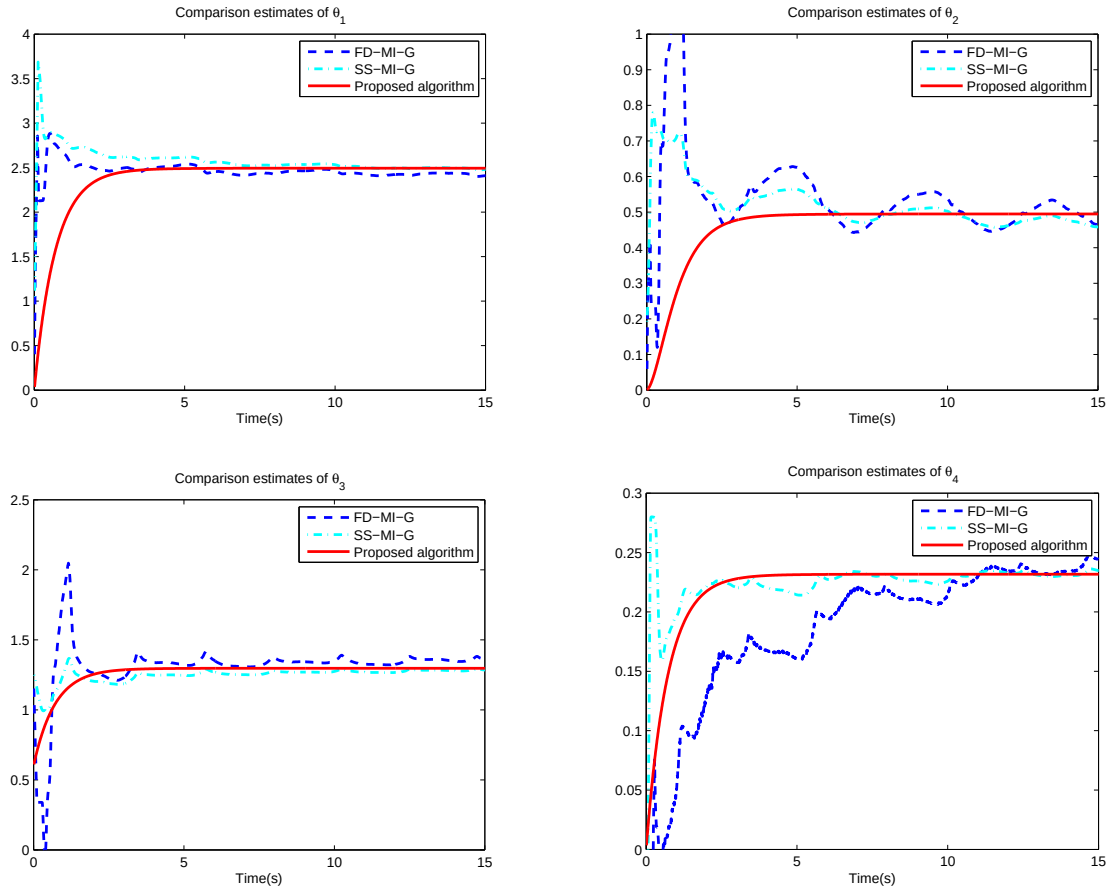


FIGURE 9. System parameter optimizations

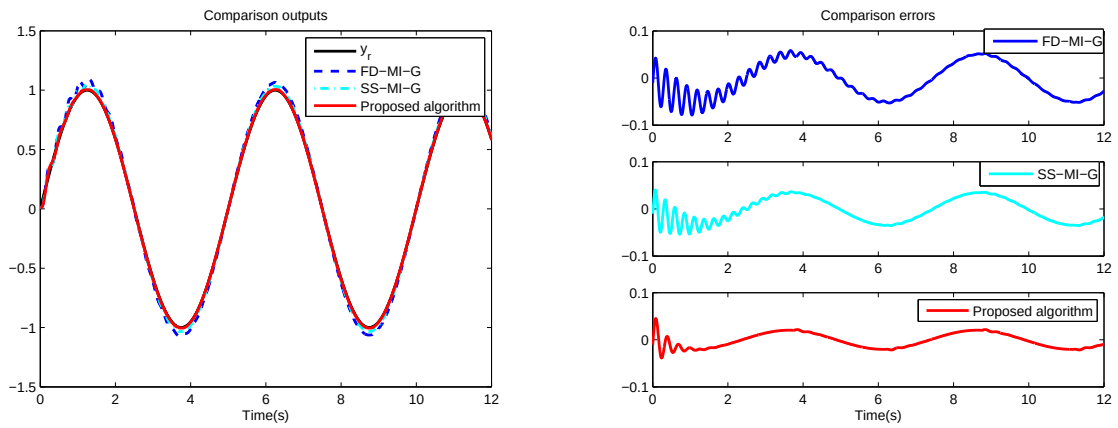


FIGURE 10. Comparison outputs

vibration, while the proposed algorithm is relatively smoother and exhibits no oscillation. In terms of optimization time, the proposed method can reach steady state in approximately 4 seconds, while the other two algorithms clearly require more time, verifying that the developed approach has faster convergence performance.

The comparison outputs are shown in Figure 10 to illustrate the output performance of the estimated model based on three estimation results. All model outputs can represent real-life system output with different output errors, suggesting the usefulness of the three

estimators. Moreover, the smallest model error is obtained by the proposed IL-G method, and the presented scheme is more capable of accurately tracking the actual system output.

6. Conclusions. An improved MI-G scheme has been introduced for nonlinear three-block cascade systems with quantized observations in this paper. By considering the iterative learning principle into the parameter estimation adaptive law, we solve the problem of batch noise addition. In addition, to improve the convergence rate, an activation mechanism is introduced to obtain the initial values for parameter estimation under given innovation length. Simulation results have verified the effectiveness of the proposed identifier, and experiment outcomes have been demonstrated the usefulness of the estimator in this paper. Theoretical analysis also proved that parameter error can converge to zero. Future work will focus on exploiting the proposed estimator for linear and nonlinear multi-input multi-output system subject to quantized data, and nonlinear three-block cascade systems with time-delay system. In addition, by using fractional order theory to enhance the utilization of system data and applying the hierarchical principle to reducing algorithmic complexity, the estimation performance can be further improved.

Acknowledgment. This paper is supported by the National Social Science Fund of China under Grant 22ZDA122, and the Fundamental Research Funds for the Central Universities under Grant 2023JKF01ZK06.

REFERENCES

- [1] D. Xu, Y. Ma, W. Yang, T. Pan and Z. Dou, Sliding mode observer-based sensor fault diagnosis for lithium-ion battery packs, *International Journal of Innovative Computing, Information and Control*, vol.19, no.5, pp.1455-1470, 2023.
- [2] A. Afshari, M. Raeispour, M. Davari et al., Finite-rate distributed secondary control over digital communication networks using an event-triggered quantized algorithm for islanded modern microgrids utilizing inverter-based resources, *IEEE Transactions on Industrial Informatics*, vol.21, no.2, pp.1002-1016, 2025.
- [3] L. Li, Y. Wang, X. Wang et al., An error learning scenario-based scheme to quantized identification in Wiener-Hammerstein systems subject to deadzone nonlinearity, *IEEE Transactions on Automation Science and Engineering*, vol.22, pp.5375-5387, 2025.
- [4] W. Mahdi, A. Alhowyan and A. Obaidullah, Optimization of recombinant antibody fragment production via machine learning models: Model development and validation, *Chemical Engineering Research and Design*, vol.216, pp.25-34, 2025.
- [5] D. Xu, Z. Cheng, W. Yang and W. Zhang, Robust nonlinear control for virtual synchronous generator based on exact feedback linearization, *International Journal of Innovative Computing, Information and Control*, vol.18, no.4, pp.1133-1145, 2022.
- [6] V. Corlay, A new method for the identification of a Wiener-Hammerstein model in a communication context, *IEEE Transactions on Communications*, vol.73, no.7, pp.5065-5078, 2025.
- [7] F. Fotiadis, A. Kanellopoulos, K. G. Vamvoudakis et al., On the effect of clock offsets and quantization on learning-based adversarial games, *Automatica*, vol.167, 111762, 2024.
- [8] J. Guo and Y. Zhao, Identification for Wiener-Hammerstein systems under quantized inputs and quantized output observations, *Asian Journal of Control*, vol.23, no.1, pp.118-127, 2021.
- [9] J. Ding, W. Zhang and J. Zhao, A kernel-based identification approach for a class of nonlinear systems with quantized output data, *Digital Signal Processing*, vol.22, 103595, 2022.
- [10] J. Figwer, Random time-series model identification from binary-valued observations and quantized measurements, *Archives of Control Sciences*, vol.35, no.1, pp.19-46, 2025.
- [11] W. Yang, C. Zhao, G. Hu, D. Xu and T. Pan, Hierarchical distributed model predictive control for multiple hybrid energy storage systems in a DC microgrid, *International Journal of Innovative Computing, Information and Control*, vol.19, no.6, pp.1735-1751, 2023.
- [12] F. Ding, L. Xu, X. Zhang et al., Recursive identification methods for general stochastic systems with colored noises by using the hierarchical identification principle and the filtering identification idea, *Annual Reviews in Control*, vol.57, 100942, 2024.

- [13] H. Xu, L. Xu and B. Shen, A novel three-stage filtering identification algorithm for the exponential autoregressive time-series model, *IEEE Signal Processing Letters*, vol.32, pp.2089-2093, 2025.
- [14] R. Guo, Y. Xu, C. Hu et al., Self-adaptive neural network-based fractional-order nonlinear observer design for state of charge estimation of lithium-ion batteries, *IEEE/ASME Transactions on Mechatronics*, vol.29, no.3, pp.1761-1772, 2024.
- [15] L. Li, J. Zhang and X. Ren, Adaptive learning-based quantized parameter identification of nonlinear sandwich-like systems with memory nonlinearity, *IEEE Transactions on Industrial Electronics*, DOI: 10.1109/TIE.2025.3621657, 2025.
- [16] L. Xu, F. Ding and Q. Zhu, Separable synchronous multi-innovation gradient-based iterative signal modeling from on-line measurements, *IEEE Transactions on Instrumentation and Measurement*, vol.71, pp.1-13, 2022.
- [17] E. Lughofer and I. Škrjanc, Online active learning for evolving error feedback fuzzy models within a multi-innovation context, *IEEE Transactions on Fuzzy Systems*, vol.32, no.11, pp.5998-6011, 2024.
- [18] M. Li, X. Liu and Y. Fan, Multi-innovation gradient identification methods for bilinear output-error systems, *Optimal Control Applications and Methods*, vol.76, no.4, pp.1569-1586, 2025.
- [19] Y. Gu, Y. Hou and Q. Zhu, Separable synchronous redundant rule-based multi-innovation predictive gradient algorithms and convergence analysis for nonlinear ExpARX models, *Nonlinear Dynamics*, vol.113, no.9, pp.9685-9707, 2025.
- [20] L. Xu, H. Xu and F. Ding, Adaptive multi-innovation gradient identification algorithms for a controlled autoregressive autoregressive moving average model, *Circuits, Systems, and Signal Processing*, vol.43, no.6, pp.3718-3747, 2024.
- [21] W. Liu and W. Xiong, Auxiliary model-based continuous mixed p-norm algorithm for output-error moving average systems using the multi-innovation optimization, *International Journal of Control, Automation and Systems*, vol.22, no.1, pp.217-227, 2024.
- [22] J. Guo and Y. Zhao, Identification for Wiener-Hammerstein systems under quantized inputs and quantized output observations, *Asian Journal of Control*, vol.23, no.1, pp.118-127, 2021.
- [23] J. Vörös, New application of the key term separation principle, *Journal of Electrical Engineering*, vol.76, no.6, pp.437-441, 2022.

Author Biography



Haozhe Cao received the Ph.D. degree from Beijing Institute of Technology, China, in 2019. He is currently working at the School of National Security, People's Public Security University of China, China. His main research focuses on national security risk monitoring and assessment, and nonlinear system control.



Zilong Chen received his undergraduate degree in 2022 from People's Public Security University of China, China. He is a postgraduate student at the School of National Security, People's Public Security University of China. His main research focuses on national security risk monitoring.



Lihua Li received the Ph.D. degree from the Tsinghua University, China, in 2016. She is currently working at the School of National Security, People's Public Security University of China, China. Her main research focuses on national security risk monitoring and assessment.