12th IFAC International Symposium on
Advanced Control of Chemical Processes
July 14-17, 2024. Toronto, Canada

Hierarchical extended parameter estimation algorithms for finite impulse response moving average models

Feng Ding [*], **Xiao Zhang** [*], **Ling Xu** [**] **Xingchen Liu** [***]

[*] *School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, PR China (e-mail: fding@jiangnan.edu.cn, xzhang@jiangnan.edu.cn)*
[**] *School of Microelectronics and Control Engineering, Changzhou University, Changzhou 213164, China (email: lingxu0848@163.com)*
[***] *College of Computer and Control Engineering, Qiqihar University, Qiqihar 161006, China*

Abstract: This paper explores some hierarchical extended parameter estimation algorithms for finite impulse response moving average (FIR-MA) model from observation data, including the hierarchical extended stochastic gradient algorithm, the hierarchical multi-innovation extended stochastic gradient algorithm, the hierarchical extended gradient algorithm, the hierarchical multi-innovation extended gradient algorithm, the hierarchical extended least squares algorithm and the hierarchical multi-innovation extended least squares algorithm. The proposed hierarchical algorithms for the FIR-MA systems can be extended to other stochastic systems with colored noises.

Keywords: Recursive identification; Hierarchical identification; Multi-innovation identification

[1]

1. INTRODUCTION

System identification is the theory and methods of exploring and constructing the mathematical models of static and dynamical systems from observation data (Ljung, 1999; Xu et al., 2024; Miao, 2023). Such systems include linear systems, bilinear systems and nonlinear systems (Liu et al., 2022). The gradient identification methods are derived through minimizing the criterion functions by using the negative gradient search (Xu et al., 2020; Liu and Chen, 2023; Liu et al., 2024). The least squares is the basic methods of investigating the identification of linear-parameter systems and has been applied to many fields (Pan et al., 2023). The basic idea of separable identification is to decompose the system parameters into two parameter sets and to use the existing gradient and least squares identification methods for interactively estimating two parameter vectors. Xu proposed a separable Newton recursive estimation method through system responses based on dynamically discrete measurements with increasing data length (Xu, 2022). Xu et al developed a separable stochastic gradient estimation method for multivariable systems (Xu and Ding, 2023).

Recently, some identification ideas and identification principles have been proposed for building the mathematical models and determining the model parameters, such as the multi-innovation identification theory (Ding, 2013), the hierarchical identification principle (Ding, 2024; Ding et al., 2024), and the coupling identification concept (Ding, 2013). These promote the development and prosperity of system identification. The multi-innovation identification theory can enhance parameter estimation accuracy (Xu et al., 2024; Yang, 2023). The hierarchical identification principle can greatly enhance computational efficiency of identification algorithms especially for large-scale complex systems (Ding et al., 2024, 2023; Yang, 2024; Xing et al., 2024). The coupling identification concept can reduce computational amount of identification algorithms for multivariable systems. They can be used for linear systems and nonlinear systems. The separable projection algorithms, the separable least squares algorithms, the separable gradient algorithms and the separable Newton algorithms belong to the category of hierarchical identification.

2. FINITE IMPULSE RESPONSE MOVING AVERAGE SYSTEMS

Consider the following finite impulse response moving average (FIR-MA) system described by

$$y(\tau) = B(z)u(\tau) + D(z)v(\tau), \tag{1}$$

where $\{u(\tau)\}$ and $\{y(\tau)\}$ are the input-output sequences of the system, $\{v(\tau)\}$ is a stochastic white noise sequence with zero mean and variance σ^2, and $B(z)$ and $D(z)$ are the polynomials in the unit backward shift operator z^{-1} ($z^{-1}y(\tau) = y(\tau-1)$ or $zy(\tau) = y(\tau+1)$) and defined as

$$B(z) := b_1 z^{-1} + b_2 z^{-2} + \cdots + b_{n_b} z^{-n_b},$$
$$D(z) := 1 + d_1 z^{-1} + d_2 z^{-2} + \cdots + d_{n_d} z^{-n_d}.$$

Define the parameter vectors $\boldsymbol{b}$ and $\boldsymbol{d}$ and the information vectors $\boldsymbol{\phi}(\tau)$ and $\boldsymbol{\psi}(\tau)$ as

[1] This work was supported by the National Natural Science Foundation of China (No. 62273167).

$$\boldsymbol{b} := [b_1, b_2, \cdots, b_{n_b}]^{\mathrm{T}} \in \mathbb{R}^{n_b},$$
$$\boldsymbol{d} := [d_1, d_2, \cdots, d_{n_d}]^{\mathrm{T}} \in \mathbb{R}^{n_d},$$
$$\boldsymbol{\phi}(\tau) := [u(\tau-1), u(\tau-2), \cdots, u(\tau-n_b)]^{\mathrm{T}} \in \mathbb{R}^{n_b}, \quad (2)$$
$$\boldsymbol{\psi}(\tau) := [v(\tau-1), v(\tau-2), \cdots, v(\tau-n_d)]^{\mathrm{T}} \in \mathbb{R}^{n_d}. \quad (3)$$

Suppose that $u(\tau)$ and $y(\tau)$ are available input-output data. Note that the information vector $\boldsymbol{\phi}(\tau)$ consisting of the system input $u(\tau-i)$ is known, and the noise vector $\boldsymbol{\psi}(\tau)$ is unknown. The objective of identification in this paper is to investigate some hierarchical parameter estimation algorithms for identifying the system parameters b_i and d_i from collected input-output data $u(\tau)$ and $y(\tau)$.

Define the disturbance noise

$$w(\tau) := D(z)v(\tau). \quad (4)$$

Suppose that the orders n_b and n_d are known, and let $n := n_b + n_d$. The initial values of all variables are set to zero, i.e., $u(\tau) = 0$ and $y(\tau) = 0$ for $\tau \leq 0$.

Equations (4) and (1) can be written as

$$w(\tau) = \boldsymbol{\psi}^{\mathrm{T}}(\tau)\boldsymbol{d} + v(\tau), \quad (5)$$
$$y(\tau) = \boldsymbol{\phi}^{\mathrm{T}}(\tau)\boldsymbol{b} + w(\tau) \quad (6)$$
$$= \boldsymbol{\phi}^{\mathrm{T}}(\tau)\boldsymbol{b} + \boldsymbol{\psi}^{\mathrm{T}}(\tau)\boldsymbol{d} + v(\tau). \quad (7)$$

Define the fictitious output

$$y_1(\tau) := y(\tau) - \boldsymbol{\psi}^{\mathrm{T}}(\tau)\boldsymbol{d} \in \mathbb{R}.$$

From (7), we have

$$y_1(\tau) = \boldsymbol{\phi}^{\mathrm{T}}(\tau)\boldsymbol{b} + v(\tau). \quad (8)$$

Equations (8) and (5) form the decomposition-based hierarchical identification model of the FIR-MA system in (1). Such a decomposition leads to the coupled variables $\boldsymbol{b}$ and $\boldsymbol{d}$ between two subsystems in (8) and (5). Therefore, when exploring their identification methods, we must use the hierarchical identification principle to coordinate the associate items between them.

3. HESG ESTIMATION ALGORITHMS

The decomposition-based identification methods are called the hierarchical identification methods. Here discusses the hierarchical recursive extended stochastic gradient (HRESG) algorithm for FIR-MA systems. The HRESG algorithm is called the hierarchical extended stochastic gradient (HESG) algorithm for short.

For the hierarchical identification models in (8) and (5), define two gradient criterion functions:

$$J_1(\boldsymbol{b}) := \frac{1}{2}[y_1(\tau) - \boldsymbol{\phi}^{\mathrm{T}}(\tau)\boldsymbol{b}]^2,$$
$$J_2(\boldsymbol{d}) := \frac{1}{2}[w(\tau) - \boldsymbol{\psi}^{\mathrm{T}}(\tau)\boldsymbol{d}]^2.$$

Let $\hat{\boldsymbol{b}}(\tau) \in \mathbb{R}^{n_b}$ and $\hat{\boldsymbol{d}}(\tau) \in \mathbb{R}^{n_d}$ be the estimates of $\boldsymbol{b}$ and $\boldsymbol{d}$ at time τ. Using the negative gradient search, we can obtain the gradient-based recursive relations:

$$\hat{\boldsymbol{b}}(\tau) = \hat{\boldsymbol{b}}(\tau-1) + \frac{\boldsymbol{\phi}(\tau)}{r_1(\tau)} \times [y(\tau) - \boldsymbol{\psi}^{\mathrm{T}}(\tau)\boldsymbol{d} - \boldsymbol{\phi}^{\mathrm{T}}(\tau)\hat{\boldsymbol{b}}(\tau-1)], \quad (9)$$
$$r_1(\tau) = r_1(\tau-1) + \|\boldsymbol{\phi}(\tau)\|^2, \quad r_1(0) = 1, \quad (10)$$
$$\hat{\boldsymbol{d}}(\tau) = \hat{\boldsymbol{d}}(\tau-1) + \frac{\boldsymbol{\psi}(\tau)}{r_2(\tau)} \times [y(\tau) - \boldsymbol{\phi}^{\mathrm{T}}(\tau)\boldsymbol{b} - \boldsymbol{\psi}^{\mathrm{T}}(\tau)\hat{\boldsymbol{d}}(\tau-1)], \quad (11)$$
$$r_2(\tau) = r_2(\tau-1) + \|\boldsymbol{\psi}(\tau)\|^2, \quad r_2(0) = 1. \quad (12)$$

The algorithm in (9)–(12) involves the unknown $\boldsymbol{\psi}(\tau)$, $\boldsymbol{d}$ and $\boldsymbol{b}$. Here uses the hierarchical identification principle to realize the parameter estimation.

Let $\hat{v}(\tau)$ be the estimate of the noise $v(\tau)$. According to the definition of $\boldsymbol{\psi}(\tau)$ in (3), use the residual $\hat{v}(\tau-i)$ to define the estimate of $\boldsymbol{\psi}(\tau)$ as

$$\hat{\boldsymbol{\psi}}(\tau) := [\hat{v}(\tau-1), \hat{v}(\tau-2), \cdots, \hat{v}(\tau-n_d)]^{\mathrm{T}} \in \mathbb{R}^{n_d}. \quad (13)$$

From (7), we have

$$v(\tau) = y(\tau) - \boldsymbol{\phi}^{\mathrm{T}}(\tau)\boldsymbol{b} - \boldsymbol{\psi}^{\mathrm{T}}(\tau)\boldsymbol{d}.$$

Replacing the unknown vectors $\boldsymbol{\psi}(\tau)$, $\boldsymbol{b}$ and $\boldsymbol{d}$ with their corresponding estimates $\hat{\boldsymbol{\psi}}(\tau)$, $\hat{\boldsymbol{b}}(\tau)$ and $\hat{\boldsymbol{d}}(\tau)$, the estimate of $v(\tau)$ can be computed through

$$\hat{v}(\tau) := y(\tau) - \boldsymbol{\phi}^{\mathrm{T}}(\tau)\hat{\boldsymbol{b}}(\tau) - \hat{\boldsymbol{\psi}}^{\mathrm{T}}(\tau)\hat{\boldsymbol{d}}(\tau). \quad (14)$$

According to the hierarchical identification principle, replacing the unknown vectors $\boldsymbol{\psi}(\tau)$, $\boldsymbol{d}$ and $\boldsymbol{b}$ in (9)–(12) with their estimates $\hat{\boldsymbol{\psi}}(\tau)$, $\hat{\boldsymbol{d}}(\tau-1)$ and $\hat{\boldsymbol{b}}(\tau-1)$, getting (15)–(19) and combining (2) and (13)–(14) obtain the **hierarchical extended stochastic gradient (HESG) algorithm** for estimating $\boldsymbol{b}$ and $\boldsymbol{d}$ of the FIR-MA systems:

$$\hat{\boldsymbol{b}}(\tau) = \hat{\boldsymbol{b}}(\tau-1) + \frac{\boldsymbol{\phi}(\tau)}{r_1(\tau)}e(\tau), \quad (15)$$
$$e(\tau) = y(\tau) - \boldsymbol{\phi}^{\mathrm{T}}(\tau)\hat{\boldsymbol{b}}(\tau-1) - \hat{\boldsymbol{\psi}}^{\mathrm{T}}(\tau)\hat{\boldsymbol{d}}(\tau-1), \quad (16)$$
$$r_1(\tau) = r_1(\tau-1) + \|\boldsymbol{\phi}(\tau)\|^2, \quad r_1(0) = 1, \quad (17)$$
$$\hat{\boldsymbol{d}}(\tau) = \hat{\boldsymbol{d}}(\tau-1) + \frac{\hat{\boldsymbol{\psi}}(\tau)}{r_2(\tau)}e(\tau), \quad (18)$$
$$r_2(\tau) = r_2(\tau-1) + \|\hat{\boldsymbol{\psi}}(\tau)\|^2, \quad r_2(0) = 1, \quad (19)$$
$$\boldsymbol{\phi}(\tau) = [u(\tau-1), u(\tau-2), \cdots, u(\tau-n_b)]^{\mathrm{T}}, \quad (20)$$
$$\hat{\boldsymbol{\psi}}(\tau) = [\hat{v}(\tau-1), \hat{v}(\tau-2), \cdots, \hat{v}(\tau-n_d)]^{\mathrm{T}}, \quad (21)$$
$$\hat{v}(\tau) = y(\tau) - \boldsymbol{\phi}^{\mathrm{T}}(\tau)\hat{\boldsymbol{b}}(\tau) - \hat{\boldsymbol{\psi}}^{\mathrm{T}}(\tau)\hat{\boldsymbol{d}}(\tau), \quad (22)$$
$$\hat{\boldsymbol{\theta}}(\tau) = [\hat{\boldsymbol{b}}^{\mathrm{T}}(\tau), \hat{\boldsymbol{d}}^{\mathrm{T}}(\tau)]^{\mathrm{T}}. \quad (23)$$

In the HESG algorithm in (15)–(23), $e(\tau) \in \mathbb{R}$ is called the innovation. The noise vector $\hat{\boldsymbol{\psi}}(\tau)$ in (21) is composed of the residual $\hat{v}(\tau-i)$, so the HESG algorithm in (15)–(23) is called the residual-based HESG algorithm. If we make the modification:

$$\hat{\boldsymbol{\psi}}(\tau) = [e(\tau-1), e(\tau-2), \cdots, e(\tau-n_d)]^{\mathrm{T}}, \quad (24)$$

which is composed of the innovation $e(\tau-i)$, then we obtain the innovation-based HESG algorithm in (15)–(20) and (22)–(24).

(1) To initialize, let $\tau = 1$, $\hat{\boldsymbol{b}}(0) = \mathbf{1}_{n_b}/p_0$, $\hat{\boldsymbol{d}}(0) = \mathbf{1}_{n_d}/p_0$, $r_1(0) = 1$, $r_2(0) = 1$, $u(\tau-i) = 1/p_0$, $\hat{v}(\tau-i) = 1/p_0$, $i = 1, 2, \cdots, \max[n_b, n_d]$, $p_0 = 10^6$. Set a small positive number ε.
(2) Collect the input-output data $u(\tau)$ and $y(\tau)$. Form the information vector $\boldsymbol{\phi}(\tau)$ using (20) and the noise vector $\hat{\boldsymbol{\psi}}(\tau)$ using (21).

(3) Compute the innovation $e(\tau)$ using (16), $r_1(\tau)$ using (17) and $r_2(\tau)$ using (19).
(4) Update the parameter estimation vectors $\hat{\boldsymbol{b}}(\tau)$ using (15) and $\hat{\boldsymbol{d}}(\tau)$ using (18). Compute the residual $\hat{v}(\tau)$ using (22).
(5) If $\|\hat{\boldsymbol{b}}(\tau)-\hat{\boldsymbol{b}}(\tau-1)\|+\|\hat{\boldsymbol{d}}(\tau)-\hat{\boldsymbol{d}}(\tau-1)\|>\varepsilon$, then increase τ by 1, and go to Step 2; otherwise, output the parameter estimates $\hat{\boldsymbol{b}}(\tau)$ and $\hat{\boldsymbol{d}}(\tau)$ and terminate this recursive calculation process.

For the HESG algorithm in (15)–(23), let $r_1(\tau)$ in (15) equal $r_2(\tau)$ in (18), and take

$$r_1(\tau)=r_2(\tau)=r(\tau)=r(\tau-1)+\|\boldsymbol{\phi}(\tau)\|^2+\|\hat{\boldsymbol{\psi}}(\tau)\|^2,$$

with $r(0)=1$, then the HESG algorithm reduces to the extended stochastic gradient (ESG) algorithm for the FIR-MA systems.

4. HMI-ESG ESTIMATION ALGORITHMS

Let the positive integer p denote the innovation length. According to the multi-innovation identification theory (Ding, 2013), define the stacked output vector $\boldsymbol{Y}(p,\tau)$ and the stacked information matrices $\boldsymbol{\Phi}(p,\tau)$ and $\hat{\boldsymbol{\Psi}}(p,\tau)$ as

$$\boldsymbol{Y}(p,\tau):=[y(\tau),y(\tau-1),\cdots,y(\tau-p+1)]^{\mathrm{T}}, \quad (25)$$

$$\boldsymbol{\Phi}(p,\tau):=[\boldsymbol{\phi}(\tau),\boldsymbol{\phi}(\tau-1),\cdots,\boldsymbol{\phi}(\tau-p+1)], \quad (26)$$

$$\hat{\boldsymbol{\Psi}}(p,\tau):=[\hat{\boldsymbol{\psi}}(\tau),\hat{\boldsymbol{\psi}}(\tau-1),\cdots,\hat{\boldsymbol{\psi}}(\tau-p+1)]. \quad (27)$$

Based on the hierarchical extended stochastic gradient algorithm in (15)–(23), expanding the scalar innovation $e(\tau)=y(\tau)-\boldsymbol{\phi}^{\mathrm{T}}(\tau)\hat{\boldsymbol{b}}(\tau-1)-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(\tau)\hat{\boldsymbol{d}}(\tau-1)\in\mathbb{R}$ in (15) and (18) to an innovation vector

$$\begin{aligned}\boldsymbol{E}(p,\tau):=\;&\boldsymbol{Y}(p,\tau)-\boldsymbol{\Phi}^{\mathrm{T}}(p,\tau)\hat{\boldsymbol{b}}(\tau-1)\\&-\hat{\boldsymbol{\Psi}}^{\mathrm{T}}(p,\tau)\hat{\boldsymbol{d}}(\tau-1)\in\mathbb{R}^p, \quad (28)\end{aligned}$$

expanding the information vectors $\boldsymbol{\phi}(\tau)$ and $\hat{\boldsymbol{\psi}}(\tau)$ in (15)–(19) to the information matrices $\boldsymbol{\Phi}(p,\tau)$ and $\hat{\boldsymbol{\Psi}}(p,\tau)$, obtaining (29)–(33) and combining (25)–(28) and (20)–(23), we obtain the **hierarchical multi-innovation extended stochastic gradient (HMI-ESG) algorithm** for identifying the parameter vectors $\boldsymbol{b}$ and $\boldsymbol{d}$ as follows:

$$\hat{\boldsymbol{b}}(\tau)=\hat{\boldsymbol{b}}(\tau-1)+\frac{\boldsymbol{\Phi}(p,\tau)}{r_1(\tau)}\boldsymbol{E}(p,\tau), \quad (29)$$

$$\begin{aligned}\boldsymbol{E}(p,\tau)=\;&\boldsymbol{Y}(p,\tau)-\boldsymbol{\Phi}^{\mathrm{T}}(p,\tau)\hat{\boldsymbol{b}}(\tau-1)\\&-\hat{\boldsymbol{\Psi}}^{\mathrm{T}}(p,\tau)\hat{\boldsymbol{d}}(\tau-1), \quad (30)\end{aligned}$$

$$r_1(\tau)=r_1(\tau-1)+\|\boldsymbol{\Phi}(p,\tau)\|^2, \quad (31)$$

$$\hat{\boldsymbol{d}}(\tau)=\hat{\boldsymbol{d}}(\tau-1)+\frac{\hat{\boldsymbol{\Psi}}(p,\tau)}{r_2(\tau)}\boldsymbol{E}(p,\tau), \quad (32)$$

$$r_2(\tau)=r_2(\tau-1)+\|\hat{\boldsymbol{\Psi}}(p,\tau)\|^2, \quad (33)$$

$$\boldsymbol{Y}(p,\tau)=[y(\tau),y(\tau-1),\cdots,y(\tau-p+1)]^{\mathrm{T}}, \quad (34)$$

$$\boldsymbol{\Phi}(p,\tau)=[\boldsymbol{\phi}(\tau),\boldsymbol{\phi}(\tau-1),\cdots,\boldsymbol{\phi}(\tau-p+1)], \quad (35)$$

$$\hat{\boldsymbol{\Psi}}(p,\tau)=[\hat{\boldsymbol{\psi}}(\tau),\hat{\boldsymbol{\psi}}(\tau-1),\cdots,\hat{\boldsymbol{\psi}}(\tau-p+1)], \quad (36)$$

$$\boldsymbol{\phi}(\tau)=[u(\tau-1),u(\tau-2),\cdots,u(\tau-n_b)]^{\mathrm{T}}, \quad (37)$$

$$\hat{\boldsymbol{\psi}}(\tau)=[\hat{v}(\tau-1),\hat{v}(\tau-2),\cdots,\hat{v}(\tau-n_d)]^{\mathrm{T}}, \quad (38)$$

$$\hat{v}(\tau)=y(\tau)-\boldsymbol{\phi}^{\mathrm{T}}(\tau)\hat{\boldsymbol{b}}(\tau)-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(\tau)\hat{\boldsymbol{d}}(\tau). \quad (39)$$

This is the residual-based HMI-ESG algorithm. Letting $p=1$, the HMI-ESG algorithm reduces to the HESG algorithm in (15)–(23). The steps of recursively computing the parameter estimation vectors using the HMI-ESG algorithm in (29)–(39) are listed here.

(1) To initialize, let $\tau=1$, set the innovation length p and the data length L_e, $\hat{\boldsymbol{b}}(0)=\mathbf{1}_{n_b}/p_0$, $\hat{\boldsymbol{d}}(0)=\mathbf{1}_{n_d}/p_0$, $r_1(0)=1$, $r_2(0)=1$, $y(\tau-j)=0$, $u(\tau-j)=0$, $\hat{v}(\tau-j)=1/p_0$, $j=1,2,\cdots,p+\max[n_b,n_d]$. p_0 is a large positive number, e.g., $p_0=10^6$.
(2) Collect the input-output data $u(\tau)$ and $y(\tau)$.
(3) Construct the information vectors $\boldsymbol{\phi}(\tau)$ and $\hat{\boldsymbol{\psi}}(\tau)$ using (37) and (38).
(4) Form the stacked output vector $\boldsymbol{Y}(p,\tau)$ using (34) and the stacked information matrices $\boldsymbol{\Phi}(p,\tau)$ and $\hat{\boldsymbol{\Psi}}(p,\tau)$ using (35)–(36).
(5) Compute $r_1(\tau)$ and $r_2(\tau)$ using (31) and (33), and the innovation vector $\boldsymbol{E}(p,\tau)$ using (30).
(6) Update the parameter estimation vectors $\hat{\boldsymbol{b}}(\tau)$ and $\hat{\boldsymbol{d}}(\tau)$ using (29) and (32). Compute the residual $\hat{v}(\tau)$ using (39).
(7) If $\tau<L_e$, then increase τ by 1 and go to Step 2; otherwise, obtain the estimates $\hat{\boldsymbol{b}}(L_e)$ and $\hat{\boldsymbol{d}}(L_e)$ and terminate this procedure.

5. HEG ESTIMATION ALGORITHMS

According to the hierarchical identification models in (8) and (5), define two criterion functions:

$$J_3(\boldsymbol{b}):=\frac{1}{2}\|\boldsymbol{Y}_1(\tau)-\boldsymbol{\Phi}(\tau)\boldsymbol{b}\|^2,$$

$$J_4(\boldsymbol{d}):=\frac{1}{2}\|\boldsymbol{W}(\tau)-\boldsymbol{\Psi}(\tau)\boldsymbol{d}\|^2,$$

where the stacked output vector $\boldsymbol{Y}(\tau)$, the stacked information matrix $\boldsymbol{\Phi}(\tau)$, the stacked fictitious output $\boldsymbol{Y}_1(\tau)$, the stacked noise vector $\boldsymbol{W}(\tau)$ and the stacked noise matrix $\boldsymbol{\Psi}(\tau)$ are defined as

$$\begin{aligned}\boldsymbol{Y}(\tau)&:=[y(1),y(2),\cdots,y(\tau)]^{\mathrm{T}}\in\mathbb{R}^{\tau},\\ \boldsymbol{\Phi}(\tau)&:=[\boldsymbol{\phi}(1),\boldsymbol{\phi}(2),\cdots,\boldsymbol{\phi}(\tau)]^{\mathrm{T}}\in\mathbb{R}^{\tau\times n_b},\\ \boldsymbol{Y}_1(\tau)&:=[y_1(1),y_1(2),\cdots,y_1(\tau)]^{\mathrm{T}}\in\mathbb{R}^{\tau},\\ \boldsymbol{W}(\tau)&:=[w(1),w(2),\cdots,w(\tau)]^{\mathrm{T}}\in\mathbb{R}^{\tau},\\ \boldsymbol{\Psi}(\tau)&:=[\boldsymbol{\psi}(1),\boldsymbol{\psi}(2),\cdots,\boldsymbol{\psi}(\tau)]^{\mathrm{T}}\in\mathbb{R}^{\tau\times n_d}.\end{aligned}$$

Let $\hat{\boldsymbol{b}}(\tau)\in\mathbb{R}^{n_b}$ and $\hat{\boldsymbol{d}}(\tau)\in\mathbb{R}^{n_d}$ be the estimates of the parameter vectors $\boldsymbol{b}$ and $\boldsymbol{d}$ at time τ. Minimizing $J_3(\boldsymbol{b})$ and $J_4(\boldsymbol{d})$, we can obtain the **hierarchical extended gradient (HEG) algorithm** for identifying the parameter vectors $\boldsymbol{b}$ and $\boldsymbol{d}$ of the FIR-MA systems:

$$\hat{\boldsymbol{b}}(\tau)=\hat{\boldsymbol{b}}(\tau-1)+\frac{1}{r_1(\tau)}[\boldsymbol{\xi}_1(\tau)-\boldsymbol{R}_1(\tau)\hat{\boldsymbol{b}}(\tau-1)], \quad (40)$$

$$r_1(\tau)=r_1(\tau-1)+\|\boldsymbol{\phi}(\tau)\|^2, \quad (41)$$

$$\boldsymbol{\xi}_1(\tau)=\boldsymbol{\xi}_1(\tau-1)+\boldsymbol{\phi}(\tau)[y(\tau)-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(\tau)\hat{\boldsymbol{d}}(\tau-1)], \quad (42)$$

$$\boldsymbol{R}_1(\tau)=\boldsymbol{R}_1(\tau-1)+\boldsymbol{\phi}(\tau)\boldsymbol{\phi}^{\mathrm{T}}(\tau), \quad (43)$$

$$\hat{\boldsymbol{d}}(\tau)=\hat{\boldsymbol{d}}(\tau-1)+\frac{1}{r_2(\tau)}[\boldsymbol{\xi}_2(\tau)-\boldsymbol{R}_2(\tau)\hat{\boldsymbol{d}}(\tau-1)], \quad (44)$$
$$r_2(\tau)=r_2(\tau-1)+\|\hat{\boldsymbol{\psi}}(\tau)\|^2, \quad (45)$$
$$\boldsymbol{\xi}_2(\tau)=\boldsymbol{\xi}_2(\tau-1)+\hat{\boldsymbol{\psi}}(\tau)[y(\tau)-\boldsymbol{\phi}^{\rm T}(\tau)\hat{\boldsymbol{b}}(\tau-1)], \quad (46)$$
$$\boldsymbol{R}_2(\tau)=\boldsymbol{R}_2(\tau-1)+\hat{\boldsymbol{\psi}}(\tau)\hat{\boldsymbol{\psi}}^{\rm T}(\tau), \quad (47)$$
$$\boldsymbol{\phi}(\tau)=[u(\tau-1),u(\tau-2),\cdots,u(\tau-n_b)]^{\rm T}, \quad (48)$$
$$\hat{\boldsymbol{\psi}}(\tau)=[\hat{v}(\tau-1),\hat{v}(\tau-2),\cdots,\hat{v}(\tau-n_d)]^{\rm T}, \quad (49)$$
$$\hat{v}(\tau)=y(\tau)-\boldsymbol{\phi}^{\rm T}(\tau)\hat{\boldsymbol{b}}(\tau)-\hat{\boldsymbol{\psi}}^{\rm T}(\tau)\hat{\boldsymbol{d}}(\tau). \quad (50)$$

(1) To initialize, let $\tau=1$, $\hat{\boldsymbol{b}}(0)=\mathbf{1}_{n_b}/p_0$, $r_1(0)=1$, $\boldsymbol{\xi}_1(0)=\mathbf{1}_{n_b}/p_0$, $\boldsymbol{R}_1(0)=\boldsymbol{I}_{n_b}/p_0$, $\hat{\boldsymbol{d}}(0)=\mathbf{1}_{n_d}/p_0$, $r_2(0)=1$, $\boldsymbol{\xi}_2(0)=\mathbf{1}_{n_d}/p_0$, $\boldsymbol{R}_2(0)=\boldsymbol{I}_{n_d}/p_0$, $\hat{v}(\tau-i)=1/p_0$, $u(\tau-i)=0$, $i=1,2,\cdots,\max[n_b,n_d]$, $p_0=10^6$. Give a small positive number ε.
(2) Collect the input-output data $u(\tau)$ and $y(\tau)$. Form the information vectors $\boldsymbol{\phi}(\tau)$ and $\hat{\boldsymbol{\psi}}(\tau)$ using (48) and (49).
(3) Compute $r_1(\tau)$ using (41), the vector $\boldsymbol{\xi}_1(\tau)$ using (42), and the matrix $\boldsymbol{R}_1(\tau)$ using (43). Update the parameter estimation vector $\hat{\boldsymbol{b}}(\tau)$ using (40).
(4) Compute $r_2(\tau)$ using (45), the vector $\boldsymbol{\xi}_2(\tau)$ using (46), and the matrix $\boldsymbol{R}_2(\tau)$ using (47). Update the parameter estimation vector $\hat{\boldsymbol{d}}(\tau)$ using (44). Compute the residual $\hat{v}(\tau)$ using (50).
(5) If $\|\hat{\boldsymbol{b}}(\tau)-\hat{\boldsymbol{b}}(\tau-1)\|+\|\hat{\boldsymbol{d}}(\tau)-\hat{\boldsymbol{d}}(\tau-1)\|>\varepsilon$, then increase τ by 1, and go to Step 2; otherwise, output the parameter estimates $\hat{\boldsymbol{b}}(\tau)$ and $\hat{\boldsymbol{d}}(\tau)$, and terminate this recursive calculation process.

6. HMI-EG ESTIMATION ALGORITHMS

Let the positive integer p denote the innovation length. Based on the HEG algorithm in (40)–(50), replacing $y(\tau)$, $\boldsymbol{\phi}(\tau)$ and $\hat{\boldsymbol{\psi}}(\tau)$ on the right-hand sides of (40)–(47) with $\boldsymbol{Y}(p,\tau)$, $\boldsymbol{\Phi}(p,\tau)$ and $\hat{\boldsymbol{\Psi}}(p,\tau)$, getting (51)–(58) and combining (25)–(27) and (48)–(50), we can obtain the **hierarchical multi-innovation extended gradient (HMI-EG) algorithm** for identifying the parameter vectors $\boldsymbol{b}$ and $\boldsymbol{d}$ of the FIR-MA systems:

$$\hat{\boldsymbol{b}}(\tau)=\hat{\boldsymbol{b}}(\tau-1)+\frac{1}{r_1(\tau)}[\boldsymbol{\xi}_1(\tau)-\boldsymbol{R}_1(\tau)\hat{\boldsymbol{b}}(\tau-1)], \quad (51)$$
$$r_1(\tau)=r_1(\tau-1)+\|\boldsymbol{\Phi}(p,\tau)\|^2, \quad r_1(0)=1, \quad (52)$$
$$\boldsymbol{\xi}_1(\tau)=\boldsymbol{\xi}_1(\tau-1)+\boldsymbol{\Phi}(p,\tau)[\boldsymbol{Y}(p,\tau)-\hat{\boldsymbol{\Psi}}^{\rm T}(p,\tau)\hat{\boldsymbol{d}}(\tau-1)], \quad (53)$$
$$\boldsymbol{R}_1(\tau)=\boldsymbol{R}_1(\tau-1)+\boldsymbol{\Phi}(p,\tau)\boldsymbol{\Phi}^{\rm T}(p,\tau), \quad (54)$$
$$\hat{\boldsymbol{d}}(\tau)=\hat{\boldsymbol{d}}(\tau-1)+\frac{1}{r_2(\tau)}[\boldsymbol{\xi}_2(\tau)-\boldsymbol{R}_2(\tau)\hat{\boldsymbol{d}}(\tau-1)], \quad (55)$$
$$r_2(\tau)=r_2(\tau-1)+\|\hat{\boldsymbol{\Psi}}(p,\tau)\|^2, \quad (56)$$
$$\boldsymbol{\xi}_2(\tau)=\boldsymbol{\xi}_2(\tau-1)+\hat{\boldsymbol{\Psi}}(p,\tau)[\boldsymbol{Y}(p,\tau)-\boldsymbol{\Phi}^{\rm T}(p,\tau)\hat{\boldsymbol{b}}(\tau-1)], \quad (57)$$
$$\boldsymbol{R}_2(\tau)=\boldsymbol{R}_2(\tau-1)+\hat{\boldsymbol{\Psi}}(p,\tau)\hat{\boldsymbol{\Psi}}^{\rm T}(p,\tau), \quad (58)$$
$$\boldsymbol{Y}(p,\tau)=[y(\tau),y(\tau-1),\cdots,y(\tau-p+1)]^{\rm T}, \quad (59)$$
$$\boldsymbol{\Phi}(p,\tau)=[\boldsymbol{\phi}(\tau),\boldsymbol{\phi}(\tau-1),\cdots,\boldsymbol{\phi}(\tau-p+1)], \quad (60)$$
$$\hat{\boldsymbol{\Psi}}(p,\tau)=[\hat{\boldsymbol{\psi}}(\tau),\hat{\boldsymbol{\psi}}(\tau-1),\cdots,\hat{\boldsymbol{\psi}}(\tau-p+1)], \quad (61)$$
$$\boldsymbol{\phi}(\tau)=[u(\tau-1),u(\tau-2),\cdots,u(\tau-n_b)]^{\rm T}, \quad (62)$$
$$\hat{\boldsymbol{\psi}}(\tau)=[\hat{v}(\tau-1),\hat{v}(\tau-2),\cdots,\hat{v}(\tau-n_d)]^{\rm T}, \quad (63)$$
$$\hat{v}(\tau)=y(\tau)-\boldsymbol{\phi}^{\rm T}(\tau)\hat{\boldsymbol{b}}(\tau)-\hat{\boldsymbol{\psi}}^{\rm T}(\tau)\hat{\boldsymbol{d}}(\tau). \quad (64)$$

If one takes $p=1$, the HMI-EG algorithm in (51)–(64) reduces to the HEG algorithm in (40)–(50). The procedures of computing the parameter estimation vectors $\hat{\boldsymbol{b}}(\tau)$ and $\hat{\boldsymbol{d}}(\tau)$ using the HMI-EG algorithm in (51)–(64) are as follows.

(1) To initialize, let $\tau=1$, set the innovation length p and the data length L_e, $\hat{\boldsymbol{b}}(0)=\mathbf{1}_{n_b}/p_0$, $r_1(0)=1$, $\boldsymbol{\xi}_1(0)=\mathbf{1}_{n_b}/p_0$, $\boldsymbol{R}_1(0)=\boldsymbol{I}_{n_b}/p_0$, $\hat{\boldsymbol{d}}(0)=\mathbf{1}_{n_d}/p_0$, $r_2(0)=1$, $\boldsymbol{\xi}_2(0)=\mathbf{1}_{n_d}/p_0$, $\boldsymbol{R}_2(0)=\boldsymbol{I}_{n_d}/p_0$, $\hat{v}(\tau-i)=1/p_0$, $u(\tau-i)=0$, $i=1,2,\cdots,p+\max[n_b,n_d]$, $p_0=10^6$.
(2) Collect the input-output data $u(\tau)$ and $y(\tau)$. Form $\boldsymbol{\phi}(\tau)$ and $\hat{\boldsymbol{\psi}}(\tau)$ using (62) and (63).
(3) Form $\boldsymbol{Y}(p,\tau)$ using (59) and form $\boldsymbol{\Phi}(p,\tau)$ and $\hat{\boldsymbol{\Psi}}(p,\tau)$ using (60) and (61).
(4) Compute $r_1(\tau)$ using (52), the vector $\boldsymbol{\xi}_1(\tau)$ using (53), and the matrix $\boldsymbol{R}_1(\tau)$ using (54). Update the parameter estimation vector $\hat{\boldsymbol{b}}(\tau)$ using (51).
(5) Compute $r_2(\tau)$, $\boldsymbol{\xi}_2(\tau)$, $\boldsymbol{R}_2(\tau)$ using (56)–(58). Update $\hat{\boldsymbol{d}}(\tau)$ using (55). Compute $\hat{v}(\tau)$ using (64).
(6) If $\tau<L_e$, then increase τ by 1, and go to Step 2; otherwise, output $\hat{\boldsymbol{b}}(\tau)$ and $\hat{\boldsymbol{d}}(\tau)$, and terminate.

7. HELS ESTIMATION ALGORITHMS

Here discusses the hierarchical extended least squares (HELS) algorithm. Letting $\boldsymbol{b}=\hat{\boldsymbol{b}}(\tau)$ and $\boldsymbol{d}=\hat{\boldsymbol{d}}(\tau)$ make $J_3(\boldsymbol{b})=\min$ and $J_4(\boldsymbol{d})=\min$, and the gradients of $J_3(\boldsymbol{b})$ and $J_4(\boldsymbol{d})$ with respect to $\boldsymbol{b}$ and $\boldsymbol{d}$ at $\boldsymbol{b}=\hat{\boldsymbol{b}}(\tau)$ and $\boldsymbol{d}=\hat{\boldsymbol{d}}(\tau)$ be zero, and we obtain the **hierarchical extended least squares (HELS) algorithm** for identifying the parameter vectors $\boldsymbol{b}$ and $\boldsymbol{d}$ of the FIR-MA systems:

$$\hat{\boldsymbol{b}}(\tau)=\hat{\boldsymbol{b}}(\tau-1)+\boldsymbol{L}_1(\tau)e(\tau), \quad (65)$$
$$e(\tau)=y(\tau)-\boldsymbol{\phi}^{\rm T}(\tau)\hat{\boldsymbol{b}}(\tau-1)-\hat{\boldsymbol{\psi}}^{\rm T}(\tau)\hat{\boldsymbol{d}}(\tau-1), \quad (66)$$
$$\boldsymbol{L}_1(\tau)=\frac{\boldsymbol{P}_1(\tau-1)\boldsymbol{\phi}(\tau)}{1+\boldsymbol{\phi}^{\rm T}(\tau)\boldsymbol{P}_1(\tau-1)\boldsymbol{\phi}(\tau)}, \quad (67)$$
$$\boldsymbol{P}_1(\tau)=[\boldsymbol{I}_{n_b}-\boldsymbol{L}_1(\tau)\boldsymbol{\phi}^{\rm T}(\tau)]\boldsymbol{P}_1(\tau-1), \quad (68)$$
$$\hat{\boldsymbol{d}}(\tau)=\hat{\boldsymbol{d}}(\tau-1)+\boldsymbol{L}_2(\tau)e(\tau), \quad (69)$$
$$\boldsymbol{L}_2(\tau)=\frac{\boldsymbol{P}_2(\tau-1)\hat{\boldsymbol{\psi}}(\tau)}{1+\hat{\boldsymbol{\psi}}^{\rm T}(\tau)\boldsymbol{P}_2(\tau-1)\hat{\boldsymbol{\psi}}(\tau)}, \quad (70)$$
$$\boldsymbol{P}_2(\tau)=[\boldsymbol{I}_{n_d}-\boldsymbol{L}_2(\tau)\hat{\boldsymbol{\psi}}^{\rm T}(\tau)]\boldsymbol{P}_2(\tau-1), \quad (71)$$
$$\boldsymbol{\phi}(\tau)=[u(\tau-1),u(\tau-2),\cdots,u(\tau-n_b)]^{\rm T}, \quad (72)$$
$$\hat{\boldsymbol{\psi}}(\tau)=[\hat{v}(\tau-1),\hat{v}(\tau-2),\cdots,\hat{v}(\tau-n_d)]^{\rm T}, \quad (73)$$
$$\hat{v}(\tau)=y(\tau)-\boldsymbol{\phi}^{\rm T}(\tau)\hat{\boldsymbol{b}}(\tau)-\hat{\boldsymbol{\psi}}^{\rm T}(\tau)\hat{\boldsymbol{d}}(\tau). \quad (74)$$

The steps of the HELS algorithm in (65)–(74) are listed.

(1) To initialize, let $\tau=1$, $\hat{\boldsymbol{b}}(0)=\mathbf{1}_{n_b}/p_0$, $\hat{\boldsymbol{d}}(0)=\mathbf{1}_{n_d}/p_0$, $\boldsymbol{P}_1(0)=p_0\boldsymbol{I}_{n_b}$, $\boldsymbol{P}_2(0)=p_0\boldsymbol{I}_{n_d}$, $\hat{v}(\tau-i)=1/p_0$,

$u(\tau-i)=0$, $i=1, 2, \cdots, \max[n_b, n_d]$, $p_0=10^6$. Give a small positive number ε.

(2) Collect the input-output data $u(\tau)$ and $y(\tau)$. Form $\boldsymbol{\phi}(\tau)$ and $\hat{\boldsymbol{\psi}}(\tau)$ using (72) and (73).

(3) Compute the innovation $e(\tau)$ using (66). Compute the gain vectors $\boldsymbol{L}_1(\tau)$ and $\boldsymbol{L}_2(\tau)$ using (67) and (70), and the covariance matrices $\boldsymbol{P}_1(\tau)$ and $\boldsymbol{P}_2(\tau)$ using (68) and (71).

(4) Update the parameter estimation vectors $\hat{\boldsymbol{b}}(\tau)$ using (65) and $\hat{\boldsymbol{d}}(\tau)$ using (69). Compute the residual $\hat{v}(\tau)$ using (74).

(5) If $\|\hat{\boldsymbol{b}}(\tau)-\hat{\boldsymbol{b}}(\tau-1)\|+\|\hat{\boldsymbol{d}}(\tau)-\hat{\boldsymbol{d}}(\tau-1)\|>\varepsilon$, then increase τ by 1, and go to Step 2; otherwise, obtain the parameter estimates $\hat{\boldsymbol{b}}(\tau)$ and $\hat{\boldsymbol{d}}(\tau)$, and terminate.

8. HMI-ELS ESTIMATION ALGORITHMS

Based on the HELS identification algorithm in (65)–(74), expanding the scalar innovation $e(\tau)\in\mathbb{R}$ in (65) and (69) to an innovation vector

$$\boldsymbol{E}(p,\tau):=\boldsymbol{Y}(p,\tau)-\boldsymbol{\Phi}^{\mathrm{T}}(p,\tau)\hat{\boldsymbol{b}}(\tau-1)-\hat{\boldsymbol{\Psi}}^{\mathrm{T}}(p,\tau)\hat{\boldsymbol{d}}(\tau-1), \tag{75}$$

expanding the information vectors $\boldsymbol{\phi}(\tau)$ and $\hat{\boldsymbol{\psi}}(\tau)$ in (67)–(68) and (70)–(71) to the information matrices $\boldsymbol{\Phi}(p,\tau)$ and $\hat{\boldsymbol{\Psi}}(p,\tau)$, getting (76) and (78)–(80) and combining (75) and (72)–(74), we get the **hierarchical multi-innovation extended least squares (HMI-ELS) algorithm** for identifying $\boldsymbol{b}$ and $\boldsymbol{d}$ as follows:

$$\hat{\boldsymbol{b}}(\tau)=\hat{\boldsymbol{b}}(\tau-1)+\boldsymbol{L}_1(\tau)\boldsymbol{E}(p,\tau), \tag{76}$$

$$\boldsymbol{E}(p,\tau)=\boldsymbol{Y}(p,\tau)-\boldsymbol{\Phi}^{\mathrm{T}}(p,\tau)\hat{\boldsymbol{b}}(\tau-1)-\hat{\boldsymbol{\Psi}}^{\mathrm{T}}(p,\tau)\hat{\boldsymbol{d}}(\tau-1), \tag{77}$$

$$\boldsymbol{L}_1(\tau)=\boldsymbol{P}_1(\tau)\boldsymbol{\Phi}(p,\tau)$$

$$\boldsymbol{P}_1(\tau)=\boldsymbol{P}_1(\tau-1)-\boldsymbol{L}_1(\tau)\boldsymbol{\Phi}^{\mathrm{T}}(p,\tau)\boldsymbol{P}_1(\tau-1), \tag{78}$$

$$\hat{\boldsymbol{d}}(\tau)=\hat{\boldsymbol{d}}(\tau-1)+\boldsymbol{L}_2(\tau)\boldsymbol{E}(p,\tau), \tag{79}$$

$$\boldsymbol{L}_2(\tau)=\boldsymbol{P}_2(\tau)\hat{\boldsymbol{\Psi}}(p,\tau)$$

$$\boldsymbol{P}_2(\tau)=\boldsymbol{P}_2(\tau-1)-\boldsymbol{L}_2(\tau)\hat{\boldsymbol{\Psi}}^{\mathrm{T}}(p,\tau)\boldsymbol{P}_2(\tau-1), \tag{80}$$

$$\boldsymbol{Y}(p,\tau)=[y(\tau),y(\tau-1),\cdots,y(\tau-p+1)]^{\mathrm{T}}, \tag{81}$$

$$\boldsymbol{\Phi}(p,\tau)=[\boldsymbol{\phi}(\tau),\boldsymbol{\phi}(\tau-1),\cdots,\boldsymbol{\phi}(\tau-p+1)], \tag{82}$$

$$\hat{\boldsymbol{\Psi}}(p,\tau)=[\hat{\boldsymbol{\psi}}(\tau),\hat{\boldsymbol{\psi}}(\tau-1),\cdots,\hat{\boldsymbol{\psi}}(\tau-p+1)], \tag{83}$$

$$\boldsymbol{\phi}(\tau)=[u(\tau-1),u(\tau-2),\cdots,u(\tau-n_b)]^{\mathrm{T}}, \tag{84}$$

$$\hat{\boldsymbol{\psi}}(\tau)=[\hat{v}(\tau-1),\hat{v}(\tau-2),\cdots,\hat{v}(\tau-n_d)]^{\mathrm{T}}, \tag{85}$$

$$\hat{v}(\tau)=y(\tau)-\boldsymbol{\phi}^{\mathrm{T}}(\tau)\hat{\boldsymbol{b}}(\tau)-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(\tau)\hat{\boldsymbol{d}}(\tau). \tag{86}$$

The steps of the HMI-ELS algorithm in (76)–(86) with τ increasing are listed here.

(1) To initialize, let $\tau=1$, set the innovation length p and the data length L_e, $\hat{\boldsymbol{b}}(0)=\mathbf{1}_{n_b}/p_0$, $\hat{\boldsymbol{d}}(0)=\mathbf{1}_{n_d}/p_0$, $\boldsymbol{P}_1(0)=p_0\boldsymbol{I}_{n_b}$, $\boldsymbol{P}_2(0)=p_0\boldsymbol{I}_{n_d}$, $y(\tau-i)=0$, $u(\tau-i)=0$, $\hat{v}(\tau-i)=1/p_0$, $i=1,2,\cdots,p+n_b+n_d$. p_0 is a large positive number, e.g., $p_0=10^6$.

(2) Collect the input-output data $u(\tau)$ and $y(\tau)$. Construct $\boldsymbol{\phi}(\tau)$ and $\hat{\boldsymbol{\psi}}(\tau)$ using (84)–(85).

(3) Form $\boldsymbol{Y}(p,\tau)$ using (81) and form $\boldsymbol{\Phi}(p,\tau)$ and $\hat{\boldsymbol{\Psi}}(p,\tau)$ using (82) and (83).

(4) Compute the gain matrices $\boldsymbol{L}_1(\tau)$ and $\boldsymbol{L}_2(\tau)$ using (78) and (80), and the covariance matrices $\boldsymbol{P}_1(\tau)$ and $\boldsymbol{P}_2(\tau)$ using (78) and (80).

(5) Compute the innovation vector $\boldsymbol{E}(p,\tau)$ using (77), and update the parameter estimation vectors $\hat{\boldsymbol{b}}(\tau)$ and $\hat{\boldsymbol{d}}(\tau)$ using (76) and (79).

(6) Compute the residual $\hat{v}(\tau)$ using (86).

(7) If $\tau<L_e$, then increase τ by 1 and go to Step 2; otherwise, obtain the estimates $\hat{\boldsymbol{b}}(L_e)$ and $\hat{\boldsymbol{d}}(L_e)$, and terminate this procedure.

9. SIMULATION EXAMPLE

Assume that the system considered has the following third-order finite impulse response moving average representation:

$$y(\tau)=B(z)u(\tau)+D(z)(\tau),$$
$$B(z)=1.32z^{-1}+1.68z^{-2}+0.76z^{-3},$$
$$D(z)=1+d_1z^{-1}+d_2z^{-1}=1+1.10z^{-1}+0.30z^{-2}.$$

The parameter vector to be identified is given by

$$\boldsymbol{\theta}=[b_1,b_2,b_3,d_1,d_2]^{\mathrm{T}}=[1.32,1.68,0.76,1.10,0.30]^{\mathrm{T}}.$$

In simulation, the input $\{u(\tau)\}$ is taken as an uncorrelated uniform distribution random signal sequence with zero mean and unit variance, $\{v(\tau)\}$ is taken as a normal distribution white noise sequence with zero mean and variance $\sigma^2=0.50^2$. The corresponding noise-to-signal ratio is $\delta_{\rm ns}=33.44\%$. We use the example parameters and the input signal to generate the output sequence $\{y(\tau)\}$.

Taking the data length $L_e=5000$, applying the HESG, HMI-ESG, HREG, HMI-REG, RELS and HMI-ELS algorithms and the input-output data $\{u(\tau),y(\tau)\colon \tau=1,2,\cdots,L_e\}$ to estimate the parameters of this system, the HMI-ESG estimation errors $\delta:=\|\hat{\boldsymbol{\theta}}(\tau)-\boldsymbol{\theta}\|/\|\boldsymbol{\theta}\|$ versus τ under different innovation lengths $p=1$, $p=2$ and $p=5$ are shown in Figure 1, the HESG, HREG and HELS estimation errors δ versus τ are shown in Figure 2, and the HMI-ESG, HMI-REG and HMI-ELS estimation errors δ versus τ are shown in Figure 3 with $p=10$.

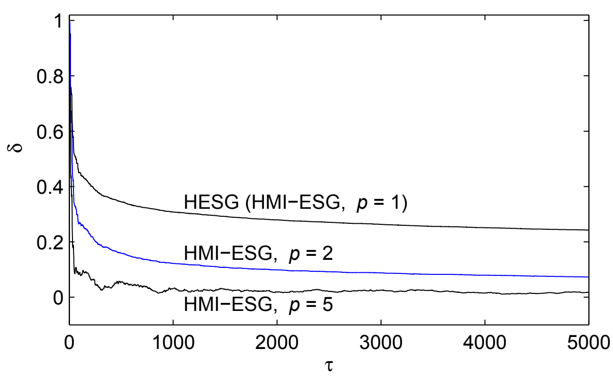

Fig. 1. The HESG and HMI-ESG estimation errors δ versus τ under the innovation lengths $p=2$ and $p=5$

From Figures 1–3, we draw the following conclusions.

- From Figure 1, it can be seen that the HMI-ESG estimation accuracies become high as the innovation length p increases under the same data lengths.

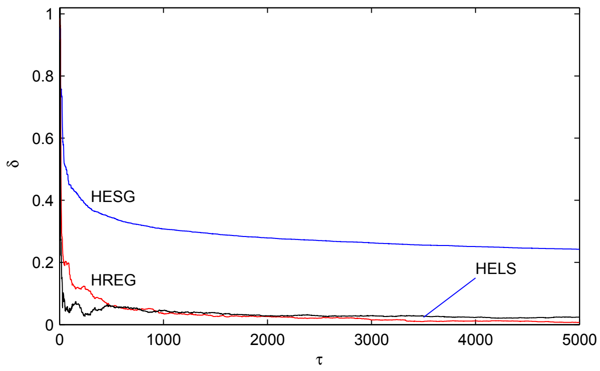

Fig. 2. The HESG, HREG and HELS estimation errors δ versus τ

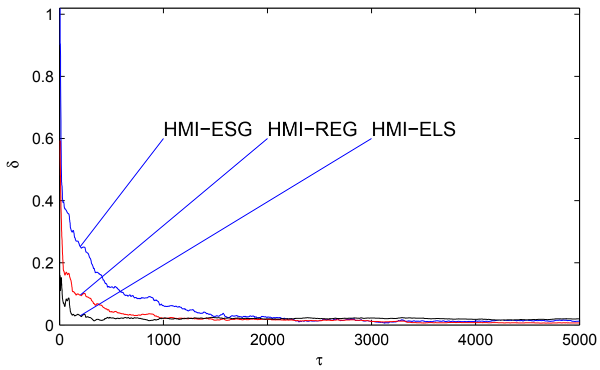

Fig. 3. The HMI-ESG, HMI-REG and HMI-ELS estimation errors δ versus τ for the innovation length $p = 10$

- From Figure 2, it is clear that the HREG and HELS algorithms have higher estimation accuracies than the HESG algorithm, and the HREG and HELS estimation accuracies are close for large data lengths.
- Figure 3 shows that the HMI-ESG, HMI-REG and HMI-ELS estimation errors become small and their estimation accuracies are close for large data lengths.

REFERENCES

Ding F. *System Identification-New Theory and Methods*. Science Press, Beijing, 2013.

Ding F. *System Identification-Hierarchical Identification Principle and Methods*. Science Press, Beijing, 2024.

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*. 2024;34(2):1120-1147.

Ding F, Xu L, Zhang X, Zhou YH. Filtered auxiliary model recursive generalized extended parameter estimation methods for Box-Jenkins systems by means of the filtering identification idea. *Int J Robust Nonlinear Control*. 2023;33(10):5510-5535.

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*. 2023;57:100942.

Liu QY, Chen FY. Model transformation based distributed stochastic gradient algorithm for multivariate output-error systems. *Int J Syst Sci*. 2023;54(7):1484-1502.

Liu SY, Wang, YJ, Ding F, et al. Joint iterative state and parameter estimation for bilinear systems with autoregressive noises via the data filtering. *ISA Trans*. 2024. https://doi.org/10.1016/j.isatra.2024.01.035

Liu SY, Zhang X, Xu L, Ding F. Expectation-maximization algorithm for bilinear systems by using the Rauch-Tung-Striebel smoother. *Automatica*. 2022;142:110365.

Ljung L. *System Identification: Theory for the User. 2nd ed.*, Prentice Hall, Englewood Cliffs, New Jersey, 1999.

Miao GQ. Iterative parameter identification algorithms for transformed dynamic rational fraction input-output systems. *J Comput Appl Math*. 2023;434:115297.

Pan J, Zhang H, Guo H, Liu S, Liu Y. Multivariable CAR-like system identification with multi-innovation gradient and least squares algorithms. *Int J Control Autom Syst*. 2023; 21(5):1455-1464

Xing HM, Ding F, Zhang X, Luan XL, Yang EF. Highly-efficient filtered hierarchical identification algorithms for multiple-input multiple-output systems with colored noises. *Syst Control Lett*. 2024; 186:105762.

Xu L. Separable Newton recursive estimation method through system responses based on dynamically discrete measurements with increasing data length. *Int J Control Autom Syst*. 2022; 20(2)432-443.

Xu L. Parameter estimation for nonlinear functions related to system responses. *Int J Control Autom Syst*. 2023;21(6):1780-1792.

Xu L, Ding F. Separable synthesis estimation methods and convergence analysis for multivariable systems. *J Comput Appl Math*. 2023; 427: 115104.

Xu L, Ding F, Yang EF. Separable recursive gradient algorithm for dynamical systems based on the impulse response signals. *Int J Control Autom Syst*. 2020; 18(12):3167-3177.

Xu L, Ding F, Zhang X, Zhu QM. Novel parameter estimation method for the systems with colored noises by using the filtering identification idea. *Syst Control Lett*. 2024; 186:105774.

Xu L, Xu H, Ding F. Adaptive multi-innovation gradient identification algorithms for a controlled autoregressive autoregressive moving average model. *Circuits Syst Signal Process*. 2024;43. https://doi.org/10.1007/s00034-024-02627-z

Yang D. Multi-innovation gradient-based iterative identification methods for feedback nonlinear systems by using the decomposition technique. *Int J Robust Nonlinear Control*. 2023;33(13):7755-7773.

Yang D, Liu YJ. Hierarchical gradient-based iterative parameter estimation algorithms for a nonlinear feedback system based on the hierarchical identification principle. *Circuits Syst Signal Process*. 2024;43(1):124-151.