CrossMark

ORIGINAL PAPER

Hierarchical multi-innovation extended stochastic gradient algorithms for input nonlinear multivariable OEMA systems by the key-term separation principle

Qianyan Shen · Feng Ding

Abstract This paper focuses on the identification of input nonlinear multivariable systems (i.e., Hammerstein multi-input multi-output nonlinear systems) described by output-error moving average models. Based on the key-term separation principle, we separate a proper key term in the nonlinear system and transform a complex nonlinear optimization problem into a pseudo-linear optimization problem which does not involve the products of the parameters between the linear parts and the nonlinear parts. Thus, a hierarchical extended stochastic gradient (H-ESG) algorithm is given and a hierarchical multi-innovation extended stochastic gradient (H-MI-ESG) algorithm is derived for the nonlinear systems. The proposed H-MI-ESG algorithm is an extension of the H-ESG algorithms. Compared with the over-parameterization-based least-squares identification algorithm, the H-ESG and H-MI-ESG algorithms have high computational efficiency. The simulation results show the effectiveness of the proposed algorithms.

Keywords System identification · Multi-innovation identification · Nonlinear system · Recursive identification · Key-term separation

Q. Shen
Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122, People's Republic of China
e-mail: shenqianyan1990@163.com

F. Ding (✉)
Control Science and Engineering Research Center, Jiangnan University, Wuxi 214122, People's Republic of China
e-mail: fding@jiangnan.edu.cn

1 Introduction

The parameter estimation theory and identification technique have been studied for decades [1–3] and have many applications in the field of networked control [4,5], signal processing and filtering [6,7], system modeling [8–10] and so on. In general, existing parameter identification methods can be roughly divided into two categories: the recursive algorithms [11–13] which compute the parameter estimates by using new information at each step as time increases and the iterative algorithms [14–16] which use a batch of data to update the parameter estimates. For example, Hu et al. [17] proposed two recursive extended least-squares parameter estimation algorithms for Wiener nonlinear systems with moving average noises by means of the over-parameterization principle; Vörös [18] presented a least squares-based iterative algorithm for three-block models with nonlinear static, linear dynamic and nonlinear dynamic blocks; Chen and Ding developed a hierarchical least-squares identification algorithm for Ham-

merstein nonlinear controlled autoregressive systems by using the hierarchical identification principle [19].

Finding an appropriate and simple model for a nonlinear system is a complex problem in the area of control. The typical nonlinear systems include the Wiener nonlinear system and the Hammerstein nonlinear system, which consist of a linear time-invariant block following (followed by) a memoryless nonlinear block [20–22] and the identification problems of nonlinear systems are vibrant [23–25]. A large amount of work has been published in this research field [26–28]. Recently, Wang [29] presented a filtering and auxiliary model-based recursive least-squares algorithm and a filtering and auxiliary model-based least-squares iterative identification algorithm for Hammerstein nonlinear systems; Zhang [30] derived a recursive least-squares identification algorithm based on the bias compensation technique for multi-input single-output systems with colored noises; Ding et al. [31] proposed a recursive least-squares algorithm for estimating the parameters of the nonlinear systems based on the model decomposition.

The over-parameterization algorithms are common and useful for the identification of nonlinear systems [32–34]. In [34], Ding and Chen transformed the nonlinear system into a pseudo-linear regressive identification model by the over-parameterization method and proposed an iterative least-squares algorithm and a recursive least-squares algorithm for Hammerstein nonlinear ARMAX systems. However, the drawback is that there exist redundant parameters and the over-parameterization-based least-squares algorithms have heavy computational cost. Motivated by the over-parameterization algorithms, this paper develops a multi-innovation algorithm using the key-term separation principle for Hammerstein multi-input multi-output output-error moving average (H-MIMO-OEMA) systems, i.e.,

$$
\begin{aligned}
y(t) &= A^{-1}(z)B(z)\bar{u}(t) + D(z)v(t), \\
A(z) &:= I + A_1 z^{-1} + A_2 z^{-2} \\
&\quad + \cdots + A_{n_a} z^{-n_a} \in \mathbb{R}^{m \times m}, \ A_l = [a_{ij}^l] \\
&\quad \in \mathbb{R}^{m \times m}, \\
B(z) &:= B_0 + B_1 z^{-1} + B_2 z^{-2} \\
&\quad + \cdots + B_{n_b} z^{-n_b} \in \mathbb{R}^{m \times m}, \ B_l = [b_{ij}^l] \\
&\quad \in \mathbb{R}^{m \times m}, \\
D(z) &:= I + D_1 z^{-1} + D_2 z^{-2}
\end{aligned}
$$

$$
\begin{aligned}
&+ \cdots + D_{n_d} z^{-n_d} \in \mathbb{R}^{m \times m}, \ D_l = [d_{ij}^l] \\
&\in \mathbb{R}^{m \times m},
\end{aligned} \tag{1}
$$

where $\bar{u}(t) = [\bar{u}_1(t), \bar{u}_2(t), \ldots, \bar{u}_m(t)]^{\mathrm{T}} \in \mathbb{R}^m$ is the output vector of the nonlinear part, $y(t) \in \mathbb{R}^m$ is the output vector, $v(t) \in \mathbb{R}^m$ is an additive noise with zero mean, $\bar{u}_i(t)$ is a linear combination with unknown coefficients c_{ij} of a known basis $(f_1, f_2, \ldots, f_{n_c})$, i.e.,

$$
\begin{aligned}
\bar{u}_i(t) &= c_{i1} f_1(u_i(t)) + c_{i2} f_2(u_i(t)) \\
&\quad + \cdots + c_{in_c} f_{n_c}(u_i(t)).
\end{aligned} \tag{2}
$$

$u_i(t)$ is the ith element of the system input vector $\bar{u}(t)$, $A(z)$, $B(z)$ and $D(z)$ are polynomial matrices in the unit backward shift operator z^{-1}: $z^{-1} y(t) = y(t-1)$. Assume that the orders n_a, n_b, n_c and n_d are known and $y(t) = 0$, $u(t) = 0$ and $v(t) = 0$ for $t \leqslant 0$.

Recently, Shen and Ding [35] considered the identification problem of Hammerstein multi-input multi-output output-error (H-MIMO-OE) systems with white noise. Based on the work in [35], this paper extends the method in [35] from H-MIMO-OE systems to H-MIMO-OEMA systems with moving average noise, presents the multi-innovation extended stochastic gradient algorithm for H-MIMO-OEMA systems and analyzes the performance of the involved algorithm for different innovation length p. The main contributions of in this paper lie in the following.

- For H-MIMO-OEMA systems, this paper derives a hierarchical extended stochastic gradient (H-ESG) algorithm. In order to improve convergence rates, a hierarchical multi-innovation extended stochastic gradient (H-MI-ESG) algorithm is presented by increasing the innovation length.
- Different form the over-parameterization identification methods, the parameter vector/matrix in the proposed algorithm does not involve the products of the parameters between the linear parts and the nonlinear parts and has no redundant parameters to be estimated.

The rest of this paper is organized as follows. Section 2 introduces the identification problems for H-MIMO-OEMA systems using the key-term separation principle. Sections 3 and 4 discuss the hierarchical extended stochastic gradient identification algorithm and the hierarchical multi-innovation extended stochastic gradient algorithm for the H-MIMO-OEMA systems. Section 5 gives an illustrative example to show the effectiveness of the proposed algorithm. Finally, we offer some concluding remarks in Sect. 6.

2 Problem formulation

Let $\boldsymbol{I}$ represent an identity matrix of appropriate sizes; the norm of a matrix $\boldsymbol{X}$ is defined by $\|\boldsymbol{X}\|^2 := \text{tr}[\boldsymbol{X}\boldsymbol{X}^{\text{T}}]$. Define the parameter vector $\boldsymbol{\theta}$ and the information matrix $\boldsymbol{\Psi}(t)$ of the nonlinear part as

$$\boldsymbol{\theta} := [c_{11}, c_{12}, \ldots, c_{1n_c}, c_{21}, c_{22}, \ldots,$$
$$c_{2n_c}, \ldots, c_{m1}, c_{m2}, \ldots, c_{mn_c}]^{\text{T}} \in \mathbb{R}^{mn_c},$$
$$\boldsymbol{\Psi}(t) := \text{blockdiag}[\boldsymbol{f}(u_1(t)), \boldsymbol{f}(u_2(t)), \ldots,$$
$$\boldsymbol{f}(u_m(t))] \in \mathbb{R}^{(mn_c) \times m}.$$

According to Eq. (2), the output vector $\bar{\boldsymbol{u}}(t)$ of the nonlinear part can be expressed as

$$\bar{\boldsymbol{u}}(t) = \boldsymbol{\Psi}^{\text{T}}(t)\boldsymbol{\theta}.$$

Define an inner vector or the true output vector

$$\boldsymbol{x}(t) = \boldsymbol{A}^{-1}(z)\boldsymbol{B}(z)\bar{\boldsymbol{u}}(t)$$
$$= [\boldsymbol{I} - \boldsymbol{A}(z)]\boldsymbol{x}(t) + \boldsymbol{B}(z)\bar{\boldsymbol{u}}(t)$$
$$= [\boldsymbol{I} - \boldsymbol{A}(z)]\boldsymbol{x}(t) + \boldsymbol{B}_0\boldsymbol{\Psi}^{\text{T}}(t)\boldsymbol{\theta}$$
$$+ \boldsymbol{B}_1\boldsymbol{\Psi}^{\text{T}}(t-1)\boldsymbol{\theta} + \cdots + \boldsymbol{B}_{n_b}\boldsymbol{\Psi}^{\text{T}}(t-n_b)\boldsymbol{\theta}. \quad (3)$$

According to Eq. (3), it is clear that there exist the products of parameters b_{ij}^l and c_{ij}. Therefore, to obtain unique parameter estimates, we take $\bar{\boldsymbol{u}}(t)$ as a key term, fix $\boldsymbol{B}_0 = \boldsymbol{I}$ and define the system parameter matrices $\boldsymbol{\vartheta}_s$ and $\boldsymbol{\vartheta}$ as

$$\boldsymbol{\vartheta}_s^{\text{T}} := [\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_{n_a}, \boldsymbol{B}_1, \boldsymbol{B}_2, \ldots, \boldsymbol{B}_{n_b}]$$
$$\in \mathbb{R}^{m \times (mn_a + mn_b)},$$
$$\boldsymbol{\vartheta}^{\text{T}} := [\boldsymbol{\vartheta}_s^{\text{T}}, \boldsymbol{D}_1, \boldsymbol{D}_2, \ldots, \boldsymbol{D}_{n_d}]$$
$$\in \mathbb{R}^{m \times mn}, \ n := n_a + n_b + n_d,$$

and the information vectors $\boldsymbol{\phi}_s(t)$ and $\boldsymbol{\phi}(t)$ as

$$\boldsymbol{\phi}_s(t) := [-\boldsymbol{x}^{\text{T}}(t-1), -\boldsymbol{x}^{\text{T}}(t-2), \ldots,$$
$$-\boldsymbol{x}^{\text{T}}(t-n_a), \bar{\boldsymbol{u}}^{\text{T}}(t-1), \bar{\boldsymbol{u}}^{\text{T}}(t-2), \ldots,$$
$$\bar{\boldsymbol{u}}^{\text{T}}(t-n_b)]^{\text{T}} \in \mathbb{R}^{(mn_a + mn_b)},$$
$$\boldsymbol{\phi}(t) := [\boldsymbol{\phi}_s^{\text{T}}(t), \boldsymbol{v}^{\text{T}}(t-1), \boldsymbol{v}^{\text{T}}(t-2), \ldots,$$
$$\boldsymbol{v}^{\text{T}}(t-n_d)]^{\text{T}} \in \mathbb{R}^{mn}.$$

Then, Eqs. (1) and (3) can be rewritten as

$$\boldsymbol{x}(t) = \boldsymbol{\vartheta}_s^{\text{T}}\boldsymbol{\phi}_s(t) + \bar{\boldsymbol{u}}(t), \quad (4)$$
$$\boldsymbol{y}(t) = \boldsymbol{x}(t) + [\boldsymbol{D}(z) - \boldsymbol{I}]\boldsymbol{v}(t) + \boldsymbol{v}(t)$$

$$= \boldsymbol{\vartheta}^{\text{T}}\boldsymbol{\phi}(t) + \boldsymbol{\Psi}^{\text{T}}(t)\boldsymbol{\theta} + \boldsymbol{v}(t). \quad (5)$$

Compared with the over-parameterization identification models in [34], the identification model in (4)–(5) does not involve the products of the parameters and avoids the redundant parameters to be estimated. Next, we investigate the hierarchical extended stochastic gradient identification algorithm and the hierarchical multi-innovation stochastic gradient algorithm for the H-MIMO-OEMA system.

3 The hierarchical extended stochastic gradient identification algorithm

According to Eq. (5), define two fictitious output vectors,

$$\boldsymbol{\xi}_1(t) := \boldsymbol{y}(t) - \boldsymbol{\Psi}^{\text{T}}(t)\boldsymbol{\theta} \in \mathbb{R}^m,$$
$$\boldsymbol{\xi}_2(t) := \boldsymbol{y}(t) - \boldsymbol{\vartheta}^{\text{T}}\boldsymbol{\phi}(t) \in \mathbb{R}^m,$$

and obtain two subsystems:

$$S_1 : \boldsymbol{\xi}_1(t) = \boldsymbol{\vartheta}^{\text{T}}\boldsymbol{\phi}(t) + \boldsymbol{v}(t),$$
$$S_2 : \boldsymbol{\xi}_2(t) = \boldsymbol{\Psi}^{\text{T}}(t)\boldsymbol{\theta} + \boldsymbol{v}(t).$$

Notice that the associate parameters between two subsystems are c_{ij} in $\boldsymbol{\phi}(t)$ and $\boldsymbol{\theta}$. Define two cost functions:

$$J_1(\boldsymbol{\vartheta}) := \frac{1}{2}\|\boldsymbol{\xi}_1(t) - \boldsymbol{\vartheta}^{\text{T}}\boldsymbol{\phi}(t)\|^2,$$
$$J_2(\boldsymbol{\theta}) := \frac{1}{2}\|\boldsymbol{\xi}_2(t) - \boldsymbol{\Psi}^{\text{T}}(t)\boldsymbol{\theta}\|^2.$$

The difficulty is that $\boldsymbol{\phi}(t)$ contains the unknown vectors $\boldsymbol{x}(t-i)$, $\bar{\boldsymbol{u}}(t-i)$ and $\boldsymbol{v}(t-i)$, and the output vectors $\boldsymbol{\xi}_1(t)$ and $\boldsymbol{\xi}_2(t)$ contain the unknown parameter matrix/vector $\boldsymbol{\vartheta}$ and $\boldsymbol{\theta}$, respectively. The solution is to apply the hierarchical identification principle and to replace the unknown vectors $\boldsymbol{x}(t-i)$, $\bar{\boldsymbol{u}}(t-i)$ and $\boldsymbol{v}(t-i)$ with their corresponding estimates $\hat{\boldsymbol{x}}(t-i)$, $\hat{\bar{\boldsymbol{u}}}(t-i)$ and $\hat{\boldsymbol{v}}(t-i)$, and define the information vectors $\hat{\boldsymbol{\phi}}_s(t)$ and $\hat{\boldsymbol{\phi}}(t)$ as

$$\hat{\boldsymbol{\phi}}_s(t) := [-\hat{\boldsymbol{x}}^{\text{T}}(t-1), -\hat{\boldsymbol{x}}^{\text{T}}(t-2), \ldots, -\hat{\boldsymbol{x}}^{\text{T}}(t-n_a),$$
$$\hat{\bar{\boldsymbol{u}}}^{\text{T}}(t-1), \hat{\bar{\boldsymbol{u}}}^{\text{T}}(t-2), \ldots,$$
$$\hat{\bar{\boldsymbol{u}}}^{\text{T}}(t-n_b)]^{\text{T}} \in \mathbb{R}^{(mn_a + mn_b)}, \quad (6)$$
$$\hat{\boldsymbol{\phi}}(t) := [\hat{\boldsymbol{\phi}}_s^{\text{T}}(t), \hat{\boldsymbol{v}}^{\text{T}}(t-1), \hat{\boldsymbol{v}}^{\text{T}}(t-2), \ldots,$$
$$\hat{\boldsymbol{v}}^{\text{T}}(t-n_d)]^{\text{T}} \in \mathbb{R}^{mn}. \quad (7)$$

Using the estimate $\hat{\boldsymbol{\theta}}(t)$ of $\boldsymbol{\theta}$, the estimate $\hat{\bar{\boldsymbol{u}}}(t)$ of $\bar{\boldsymbol{u}}(t)$ and $\hat{\boldsymbol{\xi}}_1(t)$ of $\boldsymbol{\xi}_1(t)$ can be computed by

$$\hat{\bar{\boldsymbol{u}}}(t) = \boldsymbol{\Psi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t), \tag{8}$$

$$\hat{\boldsymbol{\xi}}_1(t) = \boldsymbol{y}(t) - \boldsymbol{\Psi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1), \tag{9}$$

$$\boldsymbol{\Psi}(t) = \mathrm{blockdiag}[\boldsymbol{f}(u_1(t)), \boldsymbol{f}(u_2(t)), \ldots, \boldsymbol{f}(u_m(t))], \tag{10}$$

$$\begin{aligned}
\hat{\boldsymbol{\theta}}(t) &:= [\hat{c}_{11}(t), \hat{c}_{12}(t), \ldots, \hat{c}_{1n_c}(t), \hat{c}_{21}(t), \\
&\quad \hat{c}_{22}(t), \ldots, \hat{c}_{2n_c}(t), \ldots, \\
&\quad \hat{c}_{m1}(t), \hat{c}_{m2}(t), \ldots, \hat{c}_{mn_c}(t)]^{\mathrm{T}} \in \mathbb{R}^{mn_c}.
\end{aligned} \tag{11}$$

Replacing $\boldsymbol{\phi}_s(t)$, $\bar{\boldsymbol{u}}(t)$ and $\boldsymbol{\vartheta}_s$ in (4) with their estimates $\hat{\boldsymbol{\phi}}_s(t)$, $\hat{\bar{\boldsymbol{u}}}(t)$ and $\hat{\boldsymbol{\vartheta}}_s(t)$, the estimate $\hat{\boldsymbol{x}}(t)$ of $\boldsymbol{x}(t)$ can be computed by

$$\hat{\boldsymbol{x}}(t) = \hat{\boldsymbol{\vartheta}}_s^{\mathrm{T}}(t)\hat{\boldsymbol{\phi}}_s(t) + \hat{\bar{\boldsymbol{u}}}(t), \tag{12}$$

$$\begin{aligned}
\hat{\boldsymbol{\vartheta}}_s^{\mathrm{T}}(t) &:= [\hat{\boldsymbol{A}}_1(t), \hat{\boldsymbol{A}}_2(t), \ldots, \hat{\boldsymbol{A}}_{n_a}(t), \hat{\boldsymbol{B}}_1(t), \hat{\boldsymbol{B}}_2(t), \ldots, \\
&\quad \hat{\boldsymbol{B}}_{n_b}(t)] \in \mathbb{R}^{m \times (mn_a + mn_b)}.
\end{aligned} \tag{13}$$

Replacing $\boldsymbol{\vartheta}$, $\boldsymbol{\phi}(t)$ and $\bar{\boldsymbol{u}}(t)$ in (5) with their estimates $\hat{\boldsymbol{\vartheta}}(t)$, $\hat{\boldsymbol{\phi}}(t)$ and $\hat{\bar{\boldsymbol{u}}}(t)$, the estimate $\hat{\boldsymbol{v}}(t)$ of $\boldsymbol{v}(t)$ and $\hat{\boldsymbol{\xi}}_2(t)$ of $\boldsymbol{\xi}_2(t)$ can be computed by

$$\hat{\boldsymbol{v}}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{\vartheta}}^{\mathrm{T}}(t)\hat{\boldsymbol{\phi}}(t) - \hat{\bar{\boldsymbol{u}}}(t), \tag{14}$$

$$\hat{\boldsymbol{\xi}}_2(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{\vartheta}}^{\mathrm{T}}(t)\hat{\boldsymbol{\phi}}(t), \tag{15}$$

$$\begin{aligned}
\hat{\boldsymbol{\vartheta}}^{\mathrm{T}}(t) &:= [\hat{\boldsymbol{\vartheta}}_s^{\mathrm{T}}(t), \hat{\boldsymbol{D}}_1(t), \hat{\boldsymbol{D}}_2(t), \ldots, \\
&\quad \hat{\boldsymbol{D}}_{n_d}(t)] \in \mathbb{R}^{m \times mn}.
\end{aligned} \tag{16}$$

Using the negative gradient search and minimizing the cost functions $J_1(\boldsymbol{\vartheta})$ and $J_2(\boldsymbol{\theta})$, we can obtain the ESG algorithm:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\hat{\boldsymbol{\phi}}(t)}{r_1(t)}\boldsymbol{e}_1^{\mathrm{T}}(t), \tag{17}$$

$$\boldsymbol{e}_1(t) = \hat{\boldsymbol{\xi}}_1(t) - \hat{\boldsymbol{\vartheta}}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\phi}}(t), \tag{18}$$

$$r_1(t) = r_1(t-1) + \|\hat{\boldsymbol{\phi}}(t)\|^2, \ r_1(0) = 1, \tag{19}$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\boldsymbol{\Psi}(t)}{r_2(t)}\boldsymbol{e}_2(t), \tag{20}$$

$$\boldsymbol{e}_2(t) = \hat{\boldsymbol{\xi}}_2(t) - \boldsymbol{\Psi}(t)\hat{\boldsymbol{\theta}}(t-1), \tag{21}$$

$$r_2(t) = r_2(t-1) + \|\boldsymbol{\Psi}(t)\|^2, \ r_2(0) = 1. \tag{22}$$

Here, $\boldsymbol{e}_1(t) \in \mathbb{R}^m$ and $\boldsymbol{e}_2(t) \in \mathbb{R}^m$ represent innovation vectors and each element of $\boldsymbol{e}_1(t)$ and $\boldsymbol{e}_2(t)$ is a scalar innovation at the current time. Equations (6)–(22) form the hierarchical extended stochastic gradient (H-ESG) identification algorithm for the H-MIMO-OEMA sys-

tem. The steps of computing the estimates $\hat{\boldsymbol{\vartheta}}(t)$ and $\hat{\boldsymbol{\theta}}(t)$ by the H-ESG algorithm are as follows.

1. Let $t = 1$, set the initial values $\hat{\boldsymbol{\vartheta}}(0) = \boldsymbol{1}_{m \times (mn)}/p_0$, $\hat{\boldsymbol{\theta}}(0) = \boldsymbol{1}_{mn_c}/p_0$, $\hat{\boldsymbol{x}}(0) = \boldsymbol{0}$, $\hat{\bar{\boldsymbol{u}}}(0) = \boldsymbol{0}$, $\hat{\boldsymbol{v}}(0) = \boldsymbol{0}$ for $t \leqslant 0$ (p_0 is a large number, e.g., $p_0 = 10^6$).
2. Collect the input–output data $u_i(t)$ and $y_i(t)$, and form $\boldsymbol{\Psi}(t)$ using (10).
3. Form $\hat{\boldsymbol{\phi}}_s(t)$ using (6) and $\hat{\boldsymbol{\phi}}(t)$ using (7).
4. Compute $\hat{\boldsymbol{\xi}}_1(t)$, $\boldsymbol{e}_1(t)$ and $r_1(t)$ using (9), (18) and (19), and update $\hat{\boldsymbol{\vartheta}}(t)$ using (17).
5. Compute $\hat{\boldsymbol{\xi}}_2(t)$, $\boldsymbol{e}_1(t)$ and $r_2(t)$ using (15), (21) and (22), and update $\hat{\boldsymbol{\theta}}(t)$ using (20).
6. Compute $\hat{\bar{\boldsymbol{u}}}(t)$ using (8), $\hat{\boldsymbol{x}}(t)$ using (12) and $\hat{\boldsymbol{v}}(t)$ using (14).
7. Increase t by 1 and go to Step 2.

4 The hierarchical multi-innovation extended stochastic gradient algorithm

The H-ESG algorithm only uses the current data and thus has slow convergence rates. In order to enhance the convergence rate, we derive an H-MI-ESG algorithm with computational efficiency and high accuracy for subsystems S_1 and S_2 by expanding the single-innovation vectors $\boldsymbol{e}_1(t)$ and $\boldsymbol{e}_2(t)$ to an innovation matrix $\boldsymbol{E}_1(p, t)$ and an innovation vector $\boldsymbol{E}_2(p, t)$ as

$$\begin{aligned}
&\boldsymbol{E}_1(p, t) \\
&:= \begin{bmatrix}
\hat{\boldsymbol{\xi}}_1^{\mathrm{T}}(t) - \hat{\boldsymbol{\phi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1) \\
\hat{\boldsymbol{\xi}}_1^{\mathrm{T}}(t-1) - \hat{\boldsymbol{\phi}}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\vartheta}}(t-1) \\
\vdots \\
\hat{\boldsymbol{\xi}}_1^{\mathrm{T}}(t-p+1) - \hat{\boldsymbol{\phi}}^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{\vartheta}}(t-1)
\end{bmatrix} \\
&\quad \in \mathbb{R}^{p \times m},
\end{aligned} \tag{23}$$

$$\begin{aligned}
&\boldsymbol{E}_2(p, t) \\
&:= \begin{bmatrix}
\hat{\boldsymbol{\xi}}_2(t) - \boldsymbol{\Psi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1) \\
\hat{\boldsymbol{\xi}}_2(t-1) - \boldsymbol{\Psi}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\theta}}(t-1) \\
\vdots \\
\hat{\boldsymbol{\xi}}_2(t-p+1) - \boldsymbol{\Psi}^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{\theta}}(t-1)
\end{bmatrix} \\
&\quad \in \mathbb{R}^{mp},
\end{aligned} \tag{24}$$

where p represents the innovation length.

Define the information matrices $\boldsymbol{\Phi}(p, t)$ and $\boldsymbol{\Omega}(p, t)$ and the stacking output matrix/vector $\boldsymbol{Y}_1(p, t)$ and $\boldsymbol{Y}_2(p, t)$ as

$$\boldsymbol{\Phi}(p, t) := [\boldsymbol{\phi}(t), \boldsymbol{\phi}(t-1), \ldots,$$
$$\boldsymbol{\phi}(t-p+1)] \in \mathbb{R}^{mn \times p},$$
$$\boldsymbol{\Omega}(p, t) := [\boldsymbol{\Psi}(t), \boldsymbol{\Psi}(t-1), \ldots,$$
$$\boldsymbol{\Psi}(t-p+1)] \in \mathbb{R}^{(mn_c) \times (mp)},$$
$$\boldsymbol{Y}_1(p, t) := [\boldsymbol{\xi}_1(t), \boldsymbol{\xi}_1(t-1), \ldots,$$
$$\boldsymbol{\xi}_1(t-p+1)]^\mathrm{T} \in \mathbb{R}^{p \times m},$$
$$\boldsymbol{Y}_2(p, t) := [\boldsymbol{\xi}_2^\mathrm{T}(t), \boldsymbol{\xi}_2^\mathrm{T}(t-1), \ldots,$$
$$\boldsymbol{\xi}_2^\mathrm{T}(t-p+1)]^\mathrm{T} \in \mathbb{R}^{mp}. \tag{25}$$

The similar difficulties arise that $\boldsymbol{\Phi}(p, t)$ contains the unknown vectors $\boldsymbol{\phi}(t)$, $\boldsymbol{Y}_1(p, t)$ and $\boldsymbol{Y}_2(p, t)$ contain the unknown fictitious output vectors $\boldsymbol{\xi}_1(t)$ and $\boldsymbol{\xi}_2(t)$, respectively. Our approach is to replace the unmeasurable inner vectors $\boldsymbol{\phi}(t-i)$, $\boldsymbol{\xi}_1(t-i)$ and $\boldsymbol{\xi}_2(t-i)$ with their estimates $\hat{\boldsymbol{\phi}}(t-i)$, $\hat{\boldsymbol{\xi}}_1(t-i)$ and $\hat{\boldsymbol{\xi}}_2(t-i)$, and define

$$\hat{\boldsymbol{\Phi}}(p, t) := [\hat{\boldsymbol{\phi}}(t), \hat{\boldsymbol{\phi}}(t-1), \ldots,$$
$$\hat{\boldsymbol{\phi}}(t-p+1)] \in \mathbb{R}^{(mn) \times p}, \tag{26}$$
$$\hat{\boldsymbol{Y}}_1(p, t) := [\hat{\boldsymbol{\xi}}_1(t), \hat{\boldsymbol{\xi}}_1(t-1), \ldots,$$
$$\hat{\boldsymbol{\xi}}_1(t-p+1)]^\mathrm{T} \in \mathbb{R}^{p \times m}, \tag{27}$$
$$\hat{\boldsymbol{Y}}_2(p, t) := [\hat{\boldsymbol{\xi}}_2^\mathrm{T}(t), \hat{\boldsymbol{\xi}}_2^\mathrm{T}(t-1), \ldots,$$
$$\hat{\boldsymbol{\xi}}_2^\mathrm{T}(t-p+1)]^\mathrm{T} \in \mathbb{R}^{mp}. \tag{28}$$

Referring to [36] and according to the ESG algorithm in (17)–(22) and Eqs. (23) and (24), the MI-ESG algorithm with the innovation length p can be expressed as

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\hat{\boldsymbol{\Phi}}(p, t)}{r_1(t)} \boldsymbol{E}_1(p, t), \tag{29}$$
$$\boldsymbol{E}_1(p, t) = \hat{\boldsymbol{Y}}_1(p, t) - \hat{\boldsymbol{\Phi}}^\mathrm{T}(p, t)\hat{\boldsymbol{\vartheta}}(t-1), \tag{30}$$
$$r_1(t) = r_1(t-1) + \|\hat{\boldsymbol{\phi}}(t)\|^2, \ r_1(0) = 1, \tag{31}$$
$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\boldsymbol{\Omega}(p, t)}{r_2(t)} \boldsymbol{E}_2(p, t), \tag{32}$$
$$\boldsymbol{E}_2(p, t) = \hat{\boldsymbol{Y}}_2(p, t) - \boldsymbol{\Omega}^\mathrm{T}(p, t)\hat{\boldsymbol{\theta}}(t-1), \tag{33}$$
$$r_2(t) = r_2(t-1) + \|\boldsymbol{\Psi}(t)\|^2, \ r_2(0) = 1. \tag{34}$$

Equations (6)–(16) and (25)–(34) form the hierarchical multi-innovation extended stochastic gradient (H-MI-ESG) algorithm for computing $\hat{\boldsymbol{\vartheta}}(t)$ and $\hat{\boldsymbol{\theta}}(t)$. The procedures of computing $\hat{\boldsymbol{\vartheta}}(t)$ and $\hat{\boldsymbol{\theta}}(t)$ in the H-MI-ESG algorithm are listed in the following.
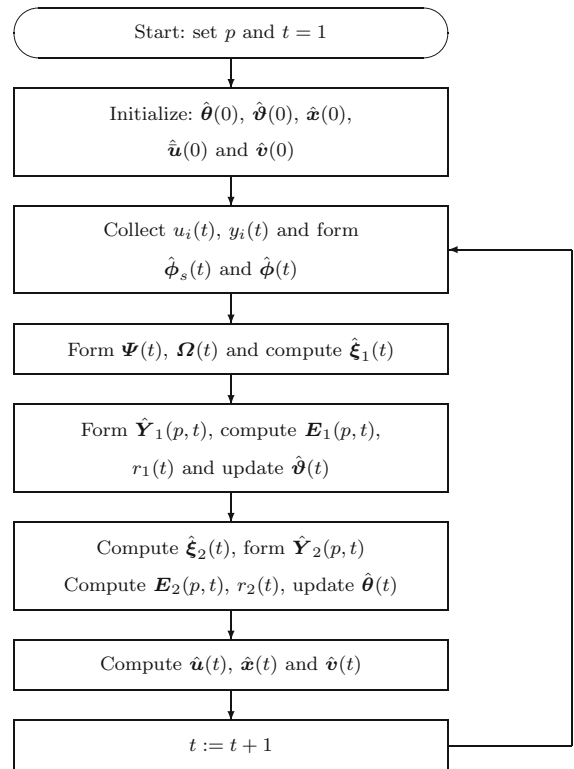
Fig. 1 The flowchart of computing the estimates $\hat{\boldsymbol{\theta}}(t)$ and $\hat{\boldsymbol{\vartheta}}(t)$

1. Let $t = 1$, set the initial values $\hat{\boldsymbol{\vartheta}}(0) = \mathbf{1}_{m \times (mn)}/p_0$, $\hat{\boldsymbol{\theta}}(0) = \mathbf{1}_{mn_c}/p_0$, $\hat{\boldsymbol{x}}(0) = \mathbf{0}$, $\hat{\bar{\boldsymbol{u}}}(0) = \mathbf{0}$, $\hat{\boldsymbol{v}}(0) = \mathbf{0}$ for $t \leqslant 0$ (p_0 is a large number, e.g., $p_0 = 10^6$).
2. Collect the input–output data $u_i(t)$ and $y_i(t)$, form $\hat{\boldsymbol{\phi}}_s(t)$ using (6), $\hat{\boldsymbol{\phi}}(t)$ using (7), $\hat{\boldsymbol{\Phi}}(p, t)$ using (26), $\boldsymbol{\Psi}(t)$ using (10) and $\boldsymbol{\Omega}(p, t)$ using (25).
3. Compute $\hat{\boldsymbol{\xi}}_1(t)$ using (9) and form $\hat{\boldsymbol{Y}}_1(p, t)$ using (27), compute $\boldsymbol{E}_1(p, t)$ and $r_1(t)$ using (30) and (31) and update $\hat{\boldsymbol{\vartheta}}(t)$ using (29).
4. Compute $\hat{\boldsymbol{\xi}}_2(t)$ using (15) and form $\hat{\boldsymbol{Y}}_2(p, t)$ using (28), compute $\boldsymbol{E}_2(p, t)$ and $r_2(t)$ using (33) and (34) and update $\hat{\boldsymbol{\theta}}(t)$ using (32).
5. Compute $\hat{\bar{\boldsymbol{u}}}(t)$ using (8), $\hat{\boldsymbol{x}}(t)$ using (12) and $\hat{\boldsymbol{v}}(t)$ using (14).
6. Increase t by 1 and go to Step 2.

The flowchart of computing the parameter estimates $\hat{\boldsymbol{\theta}}(t)$ and $\hat{\boldsymbol{\vartheta}}(t)$ in the H-MI-ESG algorithm is shown in Fig. 1.

Moreover, in order to improve the convergence rate of the H-MI-ESG algorithm further, we can introduce a forgetting factor λ, so Eqs. (31) and (34) can be rewritten as

Table 1 The H-ESG estimates and errors

t	100	200	500	1000	2000	3000
$a_{11} = 0.49$	0.75689	0.74095	0.54226	0.39886	0.40113	0.44236
$a_{12} = -0.10$	0.60103	0.40076	0.20331	−0.09078	−0.17821	−0.19922
$b_{11} = 1.26$	0.13204	0.16089	0.36970	0.62481	0.88656	1.06500
$b_{12} = 2.79$	0.75857	0.83806	0.99856	1.25881	1.61487	1.89961
$d_{11} = -0.20$	−0.38137	−0.30757	−0.23716	−0.36615	−0.55190	−0.63334
$d_{12} = 0$	0.20664	0.25457	0.29324	0.22774	0.12794	0.08468
$a_{21} = 0.35$	−0.06956	−0.06460	−0.06821	−0.06261	0.04465	0.10561
$a_{22} = -0.80$	−0.77384	−0.80396	−0.84364	−0.80959	−0.76893	−0.77005
$b_{21} = 0.79$	0.10562	0.12125	0.17475	0.24985	0.31651	0.41977
$b_{22} = 1.65$	0.70220	0.73071	0.80188	0.92645	1.11623	1.26803
$d_{21} = 0$	−0.32212	−0.26530	−0.26066	−0.38093	−0.50538	−0.56851
$d_{22} = 0.39$	0.58827	0.61927	0.61617	0.58524	0.53519	0.50028
$c_{11} = -2.21$	−0.79593	−0.83266	−1.65541	−1.94048	−2.24349	−2.27225
$c_{12} = 1.21$	1.23283	1.13020	1.19911	1.32668	1.30840	1.23642
$c_{21} = 0.76$	0.76993	0.74787	0.58549	0.65408	0.75194	0.76214
$c_{22} = 1.29$	0.80176	0.97154	1.17119	1.27766	1.27115	1.29565
$\delta\,(\%)$	66.52201	63.22889	51.32285	42.68745	33.75840	27.65244

$r_1(t) = \lambda r_1(t - 1) + \|\hat{\boldsymbol{\phi}}(t)\|^2$, $r_1(0) = 1$, $0 < \lambda \leqslant 1$,
$r_2(t) = \lambda r_2(t - 1) + \|\boldsymbol{\Psi}(t)\|^2$, $r_2(0) = 1$.
As $p = 1$, the H-MI-ESG algorithm reduces to the H-ESG algorithm in (6)–(22).

5 Example

Consider the following H-MIMO-OEMA system,
$\boldsymbol{y}(t) = \boldsymbol{x}(t) + \boldsymbol{D}(z)\boldsymbol{v}(t)$,
$\boldsymbol{x}(t) = \boldsymbol{A}^{-1}(z)\boldsymbol{B}(z)\bar{\boldsymbol{u}}(t)$,
$\boldsymbol{y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}$, $\boldsymbol{u}(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$,
$\boldsymbol{v}(t) = \begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix}$,
$\boldsymbol{A}(z) = \boldsymbol{I} + \boldsymbol{A}_1 z^{-1}$
$= \begin{bmatrix} 1 + 0.49z^{-1} & -0.10z^{-1} \\ 0.35z^{-1} & 1 - 0.80z^{-1} \end{bmatrix}$,
$\boldsymbol{B}(z) = \boldsymbol{I} + \boldsymbol{B}_1 z^{-1}$
$= \begin{bmatrix} 1 + 1.26z^{-1} & 2.79z^{-1} \\ 0.79z^{-1} & 1 + 1.65z^{-1} \end{bmatrix}$,
$\boldsymbol{D}(z) = \boldsymbol{I} + \boldsymbol{D}_1 z^{-1}$
$= \begin{bmatrix} 1 - 0.20z^{-1} & 0 \\ 0 & 1 + 0.39z^{-1} \end{bmatrix}$,

$\bar{\boldsymbol{u}}(t) = \begin{bmatrix} \bar{u}_1(t) \\ \bar{u}_2(t) \end{bmatrix} = \begin{bmatrix} c_{11} f_1(u_1(t)) + c_{12} f_2(u_1(t)) \\ c_{21} f_1(u_2(t)) + c_{22} f_2(u_2(t)) \end{bmatrix}$
$= \begin{bmatrix} -2.21 f_1(u_1(t)) + 1.21 f_2(u_1(t)) \\ 0.76 f_1(u_2(t)) + 1.29 f_2(u_2(t)) \end{bmatrix}$,

where $f_1(u_i(t)) = u_i^2(t)$ and $f_2(u_i(t)) = u_i(t)$, the parameter matrix $\boldsymbol{\vartheta}$ and the parameter vector $\boldsymbol{\theta}$ are

$\boldsymbol{\vartheta}^{\mathrm{T}} = [\boldsymbol{A}_1, \boldsymbol{B}_1, \boldsymbol{D}_1]$
$= \begin{bmatrix} 0.49 & -0.10 & 1.26 & 2.79 & -0.20 & 0 \\ 0.35 & -0.80 & 0.79 & 1.65 & 0 & 0.39 \end{bmatrix}$,

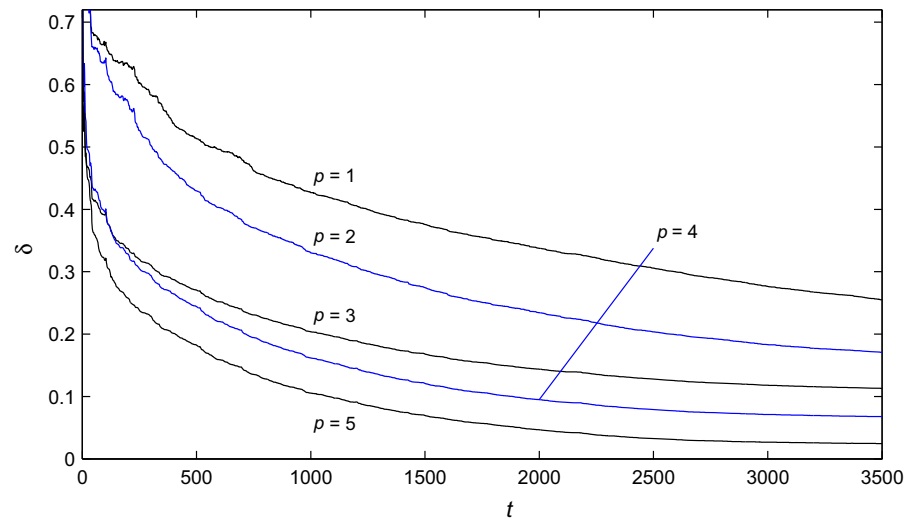$\boldsymbol{\theta} = [c_{11}, c_{12}, c_{21}, c_{22}]^{\mathrm{T}} = [-2.21, 1.21, 0.76, 1.29]^{\mathrm{T}}$.

The input vector $\boldsymbol{u}(t)$ is taken as an uncorrelated persistent excitation signal vector with zero mean and unit variances, and $\boldsymbol{v}(t)$ as a white noise vector with zero mean and variances $\sigma_1^2 = 0.10^2$ for $v_1(t)$ and $\sigma_2^2 = 0.10^2$ for $v_2(t)$. Taking the forgetting factor $\lambda = 0.99$ and applying the H-MI-ESG algorithm to estimate parameters of the example system, the parameter estimates, and the estimation errors δ are shown in Tables 1, 2 with the innovation lengths $p = 1$ and $p = 5$ and the estimation errors versus t are shown in Fig. 2, where

$$\delta := \sqrt{\frac{\|\hat{\boldsymbol{\vartheta}}(t) - \boldsymbol{\vartheta}\|^2 + \|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|^2}{\|\boldsymbol{\vartheta}\|^2 + \|\boldsymbol{\theta}\|^2}} \times 100\,\%.$$

Table 2 The H-MI-ESG estimates and errors with $p = 5$

t		100	200	500	1000	2000	3000
$a_{11} =$	0.49	0.58128	0.58212	0.60026	0.56062	0.52877	0.50158
$a_{12} =$	−0.10	−0.16805	−0.20361	−0.21335	−0.16956	−0.14479	−0.11458
$b_{11} =$	1.26	1.17078	1.21495	1.28530	1.34248	1.30293	1.27752
$b_{12} =$	2.79	1.63001	1.81132	2.09725	2.38318	2.63660	2.72962
$d_{11} =$	−0.20	−0.18453	−0.17924	−0.19600	−0.19846	−0.20386	−0.20539
$d_{12} =$	0	−0.05823	−0.02891	−0.04360	−0.03664	−0.03719	−0.03654
$a_{21} =$	0.35	0.18292	0.21538	0.32481	0.36189	0.36833	0.35382
$a_{22} =$	−0.80	−0.69112	−0.77752	−0.82646	−0.82416	−0.82149	−0.81047
$b_{21} =$	0.79	0.34002	0.45868	0.62851	0.77082	0.80391	0.79665
$b_{22} =$	1.65	0.92404	1.04696	1.23779	1.42435	1.56753	1.61981
$d_{21} =$	0	−0.03413	−0.03751	−0.08532	−0.09519	−0.09608	−0.09527
$d_{22} =$	0.39	0.38200	0.41466	0.37503	0.37589	0.37721	0.37929
$c_{11} =$	−2.21	−2.06742	−2.21027	−2.26588	−2.17296	−2.19095	−2.22214
$c_{12} =$	1.21	1.45748	1.21311	1.36462	1.21656	1.19714	1.21918
$c_{21} =$	0.76	0.88035	0.87460	0.67676	0.72887	0.74178	0.76564
$c_{22} =$	1.29	1.08758	1.37270	1.33729	1.34249	1.26327	1.29522
δ (%)		31.80743	25.76417	18.23396	10.53547	4.66747	2.68456

Fig. 2 The H-MI-ESG estimation errors δ versus t

The computational efficiency of the H-MI-ESG algorithm is shown in Table 3.

From Tables 1, 2 and Fig. 2, we can draw the following conclusions.

- The H-MI-ESG algorithm has faster convergence rates than the H-ESG algorithm.
- The H-MI-ESG estimation errors become smaller with the innovation length p increasing, and the

parameter estimates given by the H-MI-ESG algorithm can converge to their true values with the increasing of t.

From Table 3, it is clear that although a larger p leads to a model with high accuracy, the price we have to pay is a large computational effort. Moreover, the proposed algorithm can reduce computational load by decomposing a large-scale system into two subsystems

Table 3 The computational efficiency of the H-MI-ESG algorithm

Computational sequences	Number of multiplications	Number of additions
$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \zeta_1(p,t)/r_1(t) \in \mathbb{R}^{(mn) \times m}$	$m^2 n$	$m^2 n$
$\boldsymbol{\zeta}_1(p,t) := \hat{\boldsymbol{\Phi}}(p,t) \boldsymbol{E}_1(p,t) \in \mathbb{R}^{(mn) \times m}$	$m^2 np$	$m^2 n(p-1)$
$\boldsymbol{E}_1(p,t) = \hat{\boldsymbol{Y}}_1(p,t) - \hat{\boldsymbol{\Phi}}(p,t)\hat{\boldsymbol{\vartheta}}(t-1) \in \mathbb{R}^{p \times m}$	$m^2 np$	$m^2 np$
$\hat{\boldsymbol{\xi}}_1(t-i) = \boldsymbol{y}(t-i) - \boldsymbol{\Psi}^{\mathrm{T}}(t-i)\hat{\boldsymbol{\theta}}(t-1) \in \mathbb{R}^m$	$mn_c p$	$mn_c p$
$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \zeta_2(p,t)/r_2(t) \in \mathbb{R}^{mn_c}$	mn_c	mn_c
$\boldsymbol{\zeta}_2(p,t) := \boldsymbol{\Omega}(p,t) \boldsymbol{E}_2(p,t) \in \mathbb{R}^{mn_c}$	$mn_c p$	$mn_c(p-1)$
$\boldsymbol{E}_2(p,t) = \hat{\boldsymbol{Y}}_2(p,t) - \boldsymbol{\Omega}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1) \in \mathbb{R}^{mp}$	$mn_c p$	$mn_c p$
$\hat{\boldsymbol{\xi}}_2(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{\vartheta}}^{\mathrm{T}}(t)\hat{\boldsymbol{\phi}}(t) \in \mathbb{R}^m$	$m^2 np$	$m^2 np$
$r_1(t) = r_1(t-1) + \|\hat{\boldsymbol{\phi}}(t)\|^2 \in \mathbb{R}$	mn	mn
$r_2(t) = r_2(t-1) + \|\boldsymbol{\Psi}(t)\|^2 \in \mathbb{R}$	$m^2 n_c$	$(n_c-1)m^2 + m + 1$
Sum	$(3m^2 n + 3mn_c)p + (m^2 + m)(n + n_c)$	$(3m^2 n + 3mn_c)p + (n_c-1)m^2 + (n+1)m + 1$
Total flops	$(6m^2 n + 6mn_c)p + (n + 2n_c - 1)m^2 + (2n + n_c + 1)m + 1$	

and avoid the complexity of identifying the products of parameters in contrast to the over-parameterization methods.

6 Conclusions

This paper uses the key term separation principle to study identification problems of the H-MIMO-OEMA system and proposes a hierarchical multi-innovation stochastic gradient algorithm with a forgetting factor. By expanding the single-innovation vectors, the proposed H-MI-ESG algorithm has higher estimation accuracy than the H-ESG algorithm. And separating a proper key term from the nonlinear system, we can obtain a pseudo-linear model which avoids the redundant parameters. So the proposed algorithm has high computational efficiency. However, this algorithm cannot be used to identify the nonlinear systems with unknown time delay. The proposed algorithm in this paper can be extended to study identification algorithms for other linear and nonlinear systems with colored noises [37–39].

References

1. Vörös, J.: Modeling and identification of nonlinear cascade and sandwich systems with general backlash. J. Electr. Eng. Slovak **65**(2), 104–110 (2014)

2. Erturk, A., Inman, D.J.: Parameter identification and optimization in piezoelectric energy harvesting: analytical relations, asymptotic analyses, and experimental validations. J. Syst. Control Eng. **225**(4), 485–496 (2014)

3. Shen, Q.Y., Ding, F.: Iterative estimation methods for Hammerstein controlled autoregressive moving average systems based on the key-term separation principle. Nonlinear Dyn. **75**(4), 709–716 (2014)

4. Shi, Y., Huang, J., Yu, B.: Robust tracking control of networked control systems: application to a networked DC motor. IEEE Trans. Ind. Electr. **60**(12), 5864–5874 (2013)

5. Shi, Y., Yu, B.: Robust mixed H_2/H_infinity control of networked control systems with random time delays in both forward and backward communication links. Automatica **47**(4), 754–760 (2011)

6. Li, H., Shi, Y.: Robust H-infty filtering for nonlinear stochastic systems with uncertainties and random delays modeled by Markov chains. Automatica **48**(1), 159–166 (2012)

7. Liu, Y., Wang, Z.D., He, X., Zhou, D.H.: Filtering and fault detection for nonlinear systems with polynomial approximation. Automatica **54**, 348–359 (2015)

8. Xu, L., Chen, L., Xiong, W.L.: Parameter estimation and controller design for dynamic systems from the step responses based on the Newton iteration. Nonlinear Dyn. **79**(3), 2155–2163 (2015)

9. Xu, L.: A proportional differential control method for a time-delay system using the Taylor expansion approximation. Appl. Math. Comput. **236**, 391–399 (2014)

10. Li, G.Q., Wen, C.Y., Zheng, W.X., Zhao, G.S.: Iterative identification of block-oriented nonlinear systems based on biconvex optimization. Syst. Control Lett. **79**, 68–75 (2015)

11. Ding, F., Wang, Y.J., Ding, J.: Recursive least squares parameter identification for systems with colored noise using the filtering technique and the auxiliary model. Digit. Signal Process. **37**, 100–108 (2015)

12. Ding, J., Fan, C.X., Lin, J.X.: Auxiliary model based parameter estimation for dual-rate output error systems with colored noise. Appl. Math. Model. **37**(6), 4051–4058 (2013)
13. Mu, B.Q., Chen, H.F.: Recursive identification of MIMO Wiener systems. IEEE Trans. Autom. Control **58**(3), 802–808 (2013)
14. Hu, Y.B.: Iterative and recursive least squares estimation algorithms for moving average systems. Simul. Model. Pract. Theory **34**, 12–19 (2013)
15. Wang, C., Tang, T.: Several gradient-based iterative estimation algorithms for a class of nonlinear systems using the filtering technique. Nonlinear Dyn. **77**(3), 769–780 (2014)
16. Shen, Q.Y., Ding, F.: Iterative identification methods for input nonlinear multivariable systems using the key-term separation principle. J. Frankl. Inst. **352**(7), 2847–2865 (2015)
17. Hu, Y.B., Liu, B.L., Zhou, Q., Yang, C.: Recursive extended least squares parameter estimation for Wiener nonlinear systems with moving average noises. Circuits Syst. Signal Process. **33**(2), 655–664 (2014)
18. Vörös, J.: Iterative identification of nonlinear dynamic systems with output backlash using three-block cascade models. Nonlinear Dyn. **79**(3), 2187–2195 (2015)
19. Chen, H.B., Ding, F.: Hierarchical least squares identification for hammerstein nonlinear controlled autoregressive systems. Circuits Syst. Signal Process. **34**(1), 61–75 (2015)
20. Ikonen, E., Najim, K.: Non-linear process modelling based on a Wiener approach. J. Syst. Control Eng. **215**(1), 15–27 (2001)
21. Hashemizadeh, E., Rostami, M.: Numerical solution of Hammerstein integral equations of mixed type using the Sinc-collocation method. J. Comput. Appl. Math. **279**, 31–39 (2015)
22. Mu, B.Q., Chen, H.F.: Recursive identification of multi-input multi-output errors-in-variables Hammerstein systems. IEEE Trans. Autom. Control **60**(3), 843–849 (2015)
23. Lu, M.M., Zhou, X.Q., Lin, J.Q.: Improved memetic algorithm for nonlinear identification of a three-dimensional elliptical vibration cutting system. J. Syst. Control Eng. **228**(7), 449–460 (2014)
24. Abrishambaf, R., Akbari, A., Hashemipour, M.: Comparison of wireless sensor network and radio frequency identification for the process control of distributed industrial systems. J. Syst. Control Eng. **228**(5), 316–329 (2014)
25. Nauclér, P., Söderström, T.: Unbalance estimation using linear and nonlinear regression. Automatica **46**(11), 1752–1761 (2010)
26. Tiels, K., Schoukens, M., Schoukens, J.: Initial estimates for Wiener–Hammerstein models using phase-coupled multisines. Automatica **60**, 201–209 (2015)
27. Chen, H.B., Xiao, Y.S., Ding, F.: Hierarchical gradient parameter estimation algorithm for Hammerstein nonlinear systems using the key term separation principle. Appl. Math. Comput. **247**, 1202–1210 (2014)
28. Pouliquen, M., Pigeon, E., Gehan, O.: Identification scheme for Hammerstein output error models with bounded noise. IEEE Trans. Autom. Control (2015). doi:10.1109/TAC.2015.2444151
29. Wang, D.Q.: Least squares-based recursive and iterative estimation for output error moving average systems using data filtering. IET Control Theory Appl. **5**(14), 1648–1657 (2011)
30. Zhang, Y.: Unbiased identification of a class of multi-input single-output systems with correlated disturbances using bias compensation methods. Math. Comput. Model. **53**(9–10), 1810–1819 (2011)
31. Ding, F., Wang, X.H., Chen, Q.J., Xiao, Y.S.: Recursive least squares parameter estimation for a class of output nonlinear systems based on the model decomposition. Circuits Syst. Signal Process. (2016). doi:10.1007/s00034-015-0190-6
32. Bai, E.W.: An optimal two-Stage identification algorithm for Hammerstein–Wiener nonlinear system. Automatica **34**(3), 333–338 (1998)
33. Wang, X.H., Ding, F.: Convergence of the recursive identification algorithms for multivariate pseudo-linear regressive systems. Int. J. Adapt. Control Signal Process. (2016). doi:10.1002/acs.2642
34. Ding, F., Chen, T.: Identification of Hammerstein nonlinear ARMAX systems. Automatica **41**(9), 1479–1489 (2005)
35. Shen, Q.Y., Ding, F.: Multi-innovation parameter estimation for Hammerstein MIMO output-error systems based on the key-term separation. IFAC-PapersOnLine **48**(8), 457–462 (2015)
36. Hu, Y.B., Liu, B.L., Zhou, Q.: A multi-innovation generalized extended stochastic gradient algorithm for output nonlinear autoregressive moving average systems. Appl. Math. Comput. **247**, 218–224 (2014)
37. Wang, Y.J., Ding, F.: The filtering based iterative identification for multivariable systems. IET Control Theory Appl. (2016). doi:10.1049/iet-cta.2015.1195
38. Wang, Y.J., Ding, F.: Iterative estimation for a nonlinear IIR filter with moving average noise by means of the data filtering technique. IMA J. Math. Control Inf. (2016). doi:10.1093/imamci/dnv067
39. Wang, Y.J., Ding, F.: Recursive least squares algorithm and gradient algorithm for Hammerstein-Wiener systems using the data filtering. Nonlinear Dyn. (2016). doi:10.1007/s11071-015-2548-5