Nonlinear Dynamics manuscript No.
(will be inserted by the editor)

Modeling a nonlinear process using the exponential autoregressive time series model

Huan Xu · Feng Ding · Erfu Yang

Submitted: June 23, 2018

Abstract The parameter estimation methods for the nonlinear exponential autoregressive (ExpAR) model are investigated in this work. Combining the hierarchical identification principle with the negative gradient search, we derive a hierarchical stochastic gradient algorithm. Inspired by the multi-innovation identification theory, we develop a hierarchical-based multi-innovation identification algorithm for the ExpAR model. Introducing two forgetting factors, a variant of the hierarchical-based multi-innovation identification algorithm is proposed. Moreover, to compare and demonstrate the serviceability of these algorithms, a nonlinear ExpAR process is taken as an example in the simulation.

Keywords Nonlinear ExpAR model · Parameter estimation · Hierarchical identification · Multi-innovation identification · Negative gradient search

1 Introduction

Nonlinear time series models can reveal nonlinear features of many practical processes, and they are widely used in finance, ecology and some other fields [1]. The exponential autoregressive (ExpAR) model [2] is a significant kind of nonlinear time series models. In the early days, the ExpAR model is applied to the statistical analysis of the Canadian lynx data [3,4], and then it shows the appropriateness in describing certain nonlinear behaviors, such as amplitude-dependent frequency, jump phenomena and limit cycle, and in conducting accurate multistep-ahead predictions [5]. In recent years, a good deal of publications are devoted to studying the stationarity, estimation and application of the ExpAR model. For example, Chen *et al.* discussed the stationary conditions of several generalized ExpAR models, developed a variable projection based estimation algorithm, and adopted the generalized ExpAR models to model and predict the monthly mean thickness ozone column [6].

Analyzing and controlling a nonlinear time series process relies on an appropriate dynamical model. System identification is a common tool to construct the mathematical models of dynamical systems, **parameter estimation is generating the unknown system parameters via a set of observations.** System identification and parameter estimation are widely used in areas of fault diagnosis [7], network communication [8,9] and so on. Many identification methods such

H. Xu · F. Ding (Corresponding author)
School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, People's Republic of China
E-mail: fding@jiangnan.edu.cn
ORCID: https://orcid.org/0000-0002-2721-2025

H. Xu
E-mail: hxu1992@126.com

E.F. Yang
Space Mechatronic Systems Technology Laboratory, University of Strathclyde, Glasgow G1 1XJ, Scotland, United Kingdom
E-mail: erfu.yang@strath.ac.uk

as the maximum likelihood [10], the genetic algorithm [11], the blind identification [12] and the subspace identification [13] have been developed for decades. The gradient-based methods are a class of fundamental system identification methods. Combining with recursive and iterative techniques, the gradient-based methods can be provided for identifying many kinds of systems. However, the gradient-based methods have poor parameter estimation accuracies. By introducing the forgetting factor, some variants of the gradient-based identification algorithms are derived, which have improved parameter estimation accuracies. For instance, Chen and Jiang developed a gradient-based identification method with several forgetting factors for nonlinear two-variable difference systems [14].

In the area of system identification, many techniques have been exploited to improve the identification results. For example, the hierarchical identification principle has been developed as a significant branch of system identification [15]. Recently, a hierarchical gradient based iterative algorithm was used to simultaneously estimate the unknown amplitudes and angular frequencies of multi-frequency signals [16]. In addition, the multi-innovation identification has shown the effectiveness in nonlinear system identification [17]. By expanding a scalar innovation into a multi-dimensional vector, a multi-innovation stochastic gradient (SG) algorithm was derived for Wiener-Hammerstein systems with backlash [18]; a multi-innovation fractional order SG algorithm was developed for Hammerstein nonlinear ARMAX systems [19]. However, there is few research on the nonlinear time series model identification using these novel identification techniques.

This communique investigates the recursive identification algorithms for the ExpAR model. Applying the hierarchical identification principle, the ExpAR model is decomposed into two sub-identification (Sub-ID) models, one of which contains the unknown parameter vector of the linear subsystem, and the other contains the unknown parameter of the nonlinear part. With the negative gradient search, two unknown parameter sets are estimated interactively. In order to make the most of the information, the scalar innovations are expanded into innovation vectors. Moreover, two forgetting factors are introduced into the multi-innovation algorithm, so that we can present a new recursive identification algorithm with improved parameter estimation accuracy. In brief, we list the following contributions provided in this paper.

• Considering the difficulty of the nonlinear optimal problem arising in identifying the ExpAR model, we combine the hierarchical identification principle with negative gradient search so as to derive a hierarchical stochastic gradient (H-SG) algorithm for the ExpAR model.

• Using the multi-innovation identification theory, a hierarchical multi-innovation stochastic gradient (H-MISG) algorithm is presented for the ExpAR model. Introducing two forgetting factors, we obtain a modified H-MISG algorithm.

• Comparing the parameter estimation accuracies of the proposed hierarchical algorithms, we find that the modified version of the H-MISG algorithm has improved parameter estimation accuracy and can be effectively used to identify the ExpAR model.

2 Problem description

Some notations used throughout this paper are first introduced in Table 1.

Given a time series $\{x_k, x_{k-1}, x_{k-2}, \cdots\}$, an ExpAR model can be expressed as

$$\begin{aligned} x_k = &\left(\alpha_1 + \beta_1 \mathrm{e}^{-\xi x_{k-1}^2}\right) x_{k-1} + \left(\alpha_2 + \beta_2 \mathrm{e}^{-\xi x_{k-1}^2}\right) x_{k-2} + \cdots \\ &+ \left(\alpha_n + \beta_n \mathrm{e}^{-\xi x_{k-1}^2}\right) x_{k-n} + \varepsilon_k, \end{aligned} \tag{1}$$

where ε_k is a white noise with zero mean, n denotes the system degree, α_i, β_i and ξ are the model parameters to be estimated.

When the parameters $\beta_i = 0$, $i = 1, 2, \cdots, n$, Equation (1) reduces to an autoregressive (AR) model which has no nonlinear dynamics.

The form in (1) is the classic ExpAR model, some modified versions have been presented. For instance, in order to give a more sophisticated specification of the dynamics of the characteristic roots of AR models, Ozaki derived a variant of the ExpAR model in [3] using the Hermite type

Table 1 The notations used throughout this paper

Item	Notations	Descriptions
1	$x_k \in \mathbb{R}$	Measurement data
2	$\varepsilon_k \in \mathbb{R}$	Stochastic white noise
3	$\boldsymbol{\alpha} \in \mathbb{R}^n$, $\boldsymbol{\beta} \in \mathbb{R}^n$, $\boldsymbol{\Theta} \in \mathbb{R}^{2n}$, $\xi \in \mathbb{R}$	Parameters to be estimated
4	$\hat{\boldsymbol{\alpha}}_k \in \mathbb{R}^n$, $\hat{\boldsymbol{\beta}}_k \in \mathbb{R}^n$, $\hat{\boldsymbol{\Theta}}_k \in \mathbb{R}^{2n}$, $\hat{\xi}_k \in \mathbb{R}$	Parameter estimates at time k
5	$\boldsymbol{X}_k \in \mathbb{R}^n$, $\boldsymbol{\phi}(\xi, k) \in \mathbb{R}^{2n}$	Information vectors
6	$\psi(\boldsymbol{\beta}) \in \mathbb{R}$	Information item
7	$x_{1,k} \in \mathbb{R}$	Intermediate variable
8	$\boldsymbol{\phi}'(\xi, k) \in \mathbb{R}^{2n}$	Derivative of $\boldsymbol{\phi}(\xi, k)$
9	$e_{1,k} \in \mathbb{R}$, $e_{2,k} \in \mathbb{R}$	Innovations
10	$\mu_{1,k} \in \mathbb{R}$, $\mu_{2,k} \in \mathbb{R}$	Step-sizes
11	$r_{1,k} \in \mathbb{R}$, $r_{2,k} \in \mathbb{R}$	Reciprocals of the step-sizes
12	$\boldsymbol{E}_1(l) \in \mathbb{R}^l$, $\boldsymbol{E}_2(l) \in \mathbb{R}^l$	Innovation vectors
13	$\boldsymbol{X}(l) \in \mathbb{R}^l$	Stacked information vector
14	$\boldsymbol{\Phi}(l, \hat{\xi}_{k-1}) \in \mathbb{R}^{(2n)\times l}$	Stacked information matrix
15	$\boldsymbol{\Phi}'(l, \hat{\xi}_{k-1}) \in \mathbb{R}^{(2n)\times l}$	Derivative of $\boldsymbol{\Phi}(l, \hat{\xi}_{k-1})$

polynomials:

$$x_k = \sum_{i=1}^{n} \Big[\alpha_i + \Big(\beta_{i0} + \sum_{j=1}^{m_i} \beta_{ij} x_{k-1}^j\Big) \mathrm{e}^{-\xi x_{k-1}^2}\Big] x_{k-i} + \varepsilon_k.$$

Introducing a time-delay d and a scalar parameter ζ, Teräsvirta developed a different variant of the ExpAR model in [4]:

$$x_k = \Big[\alpha_0 + \beta_0 \mathrm{e}^{-\xi(x_{k-d}-\zeta)^2}\Big] + \sum_{i=1}^{n} \Big[\alpha_i + \beta_i \mathrm{e}^{-\xi(x_{k-d}-\zeta)^2}\Big] x_{k-i} + \varepsilon_k.$$

Some other generalized ExpAR models were summarized by Chen [6]. After parametrization, we can derive the corresponding identification models, which have different parameter and information vectors, for the ExpAR family. This paper copes with the recursive identification for the classic ExpAR model. The proposed hierarchical algorithms are also appropriate for other ExpAR models.

Assume that the degree n is known, the data x_k is measurable. The initial values are taken as $x_k = 0$ and $\varepsilon_k = 0$ for $t \le 0$.

It is obvious that x_k is linear with respect to the parameters α_i and β_i, and is nonlinear with respect to the parameter ξ. Define the parameter vectors of the linear subsystem

$$\boldsymbol{\alpha} := [\alpha_1, \alpha_2, \cdots, \alpha_n]^{\mathrm{T}} \in \mathbb{R}^n,$$
$$\boldsymbol{\beta} := [\beta_1, \beta_2, \cdots, \beta_n]^{\mathrm{T}} \in \mathbb{R}^n,$$

and the information vector

$$\boldsymbol{X}_k := [x_{k-1}, x_{k-2}, \cdots, x_{k-n}]^{\mathrm{T}} \in \mathbb{R}^n.$$

Then Equation (1) can be transformed into

$$\begin{aligned} x_k &= \sum_{i=1}^{n} \alpha_i x_{k-i} + \mathrm{e}^{-\xi x_{k-1}^2} \sum_{i=1}^{n} \beta_i x_{k-i} + \varepsilon_k \\ &= \boldsymbol{X}_k^{\mathrm{T}} \boldsymbol{\alpha} + \mathrm{e}^{-\xi x_{k-1}^2} \boldsymbol{X}_k^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_k. \end{aligned} \tag{2}$$

Furthermore, define the following vectors:

$$\boldsymbol{\Theta} := [\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{2n},$$
$$\boldsymbol{\phi}(\xi, k) := [\boldsymbol{X}_k^{\mathrm{T}}, \mathrm{e}^{-\xi x_{k-1}^2} \boldsymbol{X}_k^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{2n}.$$

Then Equation (2) can be equivalently transformed into the identification model

$$x_k = \boldsymbol{\phi}^{\mathrm{T}}(\xi, k) \boldsymbol{\Theta} + \varepsilon_k. \tag{3}$$

Since the unknown parameter of the nonlinear subsystem ξ exists in $\boldsymbol{\phi}(\xi,k)$, the identification problem becomes a complex nonlinear optimization problem and the least squares method cannot be used for parameter estimation. The previous work aims to explore new recursive identification methods for the ExpAR model.

3 The hierarchical stochastic gradient algorithm

Hierarchical identification is the decomposition based identification. The key idea is to decompose the identification model into several subsystems, such that the scale of the optimization problem becomes small [20]. In this section, by the hierarchical identification principle, the ExpAR model is decomposed into two subsystems, one of which contains $\boldsymbol{\Theta}$, and the other contains ξ, both these two parameter sets are to be estimated. **In addition, the negative gradient search is widely adopted to deal with some optimization problems and to determine the extreme point of the objective function.** Applying the negative gradient search, an H-SG algorithm is proposed for the ExpAR model.

Define the information item $\psi(\boldsymbol{\beta})$ and the intermediate variable $x_{1,k}$ as

$$\psi(\boldsymbol{\beta}) := \boldsymbol{X}_k^{\mathrm{T}}\boldsymbol{\beta} \in \mathbb{R},$$
$$x_{1,k} := x_k - \boldsymbol{X}_k^{\mathrm{T}}\boldsymbol{\alpha} \in \mathbb{R}.$$

From (2), we can see that the ExpAR model is decomposed into these two Sub-ID models:

$$S_1 : x_k = \boldsymbol{\phi}^{\mathrm{T}}(\xi,k)\boldsymbol{\Theta} + \varepsilon_k, \tag{4}$$
$$S_2 : x_{1,k} = \psi(\boldsymbol{\beta})\mathrm{e}^{-\xi x_{k-1}^2} + \varepsilon_k. \tag{5}$$

The parameter sets $\boldsymbol{\Theta}$ and ξ in Sub-ID models (4) and (5) contain all the parameters to be estimated. The parameter ξ in $\boldsymbol{\phi}(\xi,k)$ and the parameter vector $\boldsymbol{\beta}$ in $\psi(\boldsymbol{\beta})$ are the associate terms between these two Sub-ID models. Decomposing the identification model in (2) or (3) into the above fictitious subsystems, we can obtain a hierarchical structure which is demonstrated in Figure 1.

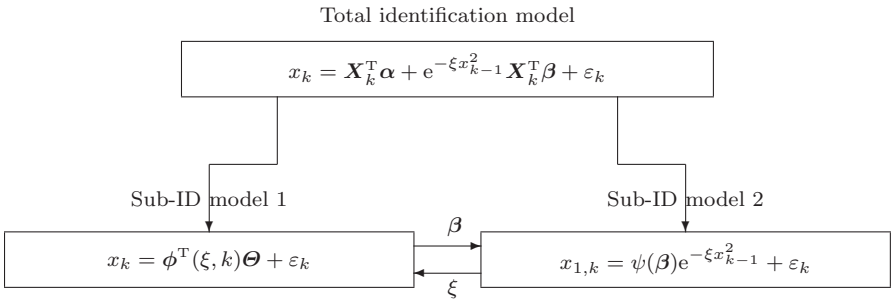


Fig. 1 The hierarchical structure of the identification models for the ExpAR model

Define two criterion functions

$$J_1(\boldsymbol{\Theta}) := \frac{1}{2}[x_k - \boldsymbol{\phi}^{\mathrm{T}}(\xi,k)\boldsymbol{\Theta}]^2, \tag{6}$$
$$J_2(\xi) := \frac{1}{2}[x_{1,k} - \psi(\boldsymbol{\beta})\mathrm{e}^{-\xi x_{k-1}^2}]^2. \tag{7}$$

Computing the gradients of $J_1(\boldsymbol{\Theta})$ and $J_2(\xi)$, we have

$$\mathrm{grad}[J_1(\boldsymbol{\Theta})] = \frac{\partial J_1(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} = -\boldsymbol{\phi}(\xi,k)[x_k - \boldsymbol{\phi}^{\mathrm{T}}(\xi,k)\boldsymbol{\Theta}],$$
$$\begin{aligned}\mathrm{grad}[J_2(\xi)] &= \frac{\partial J_2(\xi)}{\partial \xi} = x_{k-1}^2\psi(\boldsymbol{\beta})\mathrm{e}^{-\xi x_{k-1}^2}[x_{1,k} - \psi(\boldsymbol{\beta})\mathrm{e}^{-\xi x_{k-1}^2}] \\ &= x_{k-1}^2\psi(\boldsymbol{\beta})\mathrm{e}^{-\xi x_{k-1}^2}[x_k - \boldsymbol{X}_k^{\mathrm{T}}\boldsymbol{\alpha} - \psi(\boldsymbol{\beta})\mathrm{e}^{-\xi x_{k-1}^2}]\end{aligned}$$

$$
\begin{aligned}
&= -\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\phi}'(\xi,k)[x_k - \boldsymbol{X}_k^{\mathrm{T}}\boldsymbol{\alpha} - \psi(\boldsymbol{\beta})\mathrm{e}^{-\xi x_{k-1}^2}] \\
&= -\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\phi}'(\xi,k)[x_k - \boldsymbol{\phi}^{\mathrm{T}}(\xi,k)\boldsymbol{\Theta}],
\end{aligned}
$$

where

$$
\boldsymbol{\phi}'(\xi,k) := \frac{\partial\boldsymbol{\phi}(\xi,k)}{\partial\xi} = [\mathbf{0}_n^{\mathrm{T}}, -x_{k-1}^2\mathrm{e}^{-\xi x_{k-1}^2}\boldsymbol{X}_k^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{2n}.
$$

Let $\hat{\boldsymbol{\Theta}}_k$ and $\hat{\xi}_k$ signify the estimates of $\boldsymbol{\Theta}$ and ξ at time k, $\mu_{1,k}$ and $\mu_{2,k}$ represent the step-sizes to be given later. Employing the negative gradient search, we have:

$$
\begin{aligned}
\hat{\boldsymbol{\Theta}}_k &= \hat{\boldsymbol{\Theta}}_{k-1} - \mu_{1,k}\mathrm{grad}[J_1(\hat{\boldsymbol{\Theta}}_{k-1})] \\
&= \hat{\boldsymbol{\Theta}}_{k-1} + \mu_{1,k}\boldsymbol{\phi}(\xi,k)[x_k - \boldsymbol{\phi}^{\mathrm{T}}(\xi,k)\hat{\boldsymbol{\Theta}}_{k-1}], \qquad (8) \\
\hat{\xi}_k &= \hat{\xi}_{k-1} - \mu_{2,k}\mathrm{grad}[J_2(\hat{\xi}_{k-1})] \\
&= \hat{\xi}_{k-1} + \mu_{2,k}\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\phi}'(\hat{\xi}_{k-1},k)[x_k - \boldsymbol{\phi}^{\mathrm{T}}(\hat{\xi}_{k-1},k)\boldsymbol{\Theta}]. \qquad (9)
\end{aligned}
$$

The following finds the optimal step-sizes $\mu_{1,k}$ and $\mu_{2,k}$. One method is to apply the one-dimensional search, that is, to solve the optimization problems

$$
\begin{aligned}
&\min_{\mu_{1,k}\geq 0} J_1\{\hat{\boldsymbol{\Theta}}_{k-1} - \mu_{1,k}\mathrm{grad}[J_1(\hat{\boldsymbol{\Theta}}_{k-1})]\}, \\
&\min_{\mu_{2,k}\geq 0} J_2\{\hat{\xi}_{k-1} - \mu_{2,k}\mathrm{grad}[J_2(\hat{\xi}_{k-1})]\}.
\end{aligned}
$$

Remark 1: The one-dimensional search method is a fundamental method of finding the optimal step-size in the minimization problem. The key idea is to determine the negative gradient direction (i.e., the direction where the criterion function descends fastest) and to compute the step-size, which makes the criterion function minimal, by the one-dimensional search of the negative gradient direction.

For the sake of convenience, we define the innovations $e_{1,k}$ and $e_{2,k}$ as

$$
e_{1,k} := x_k - \boldsymbol{\phi}^{\mathrm{T}}(\xi,k)\hat{\boldsymbol{\Theta}}_{k-1} \in \mathbb{R}, \qquad (10)
$$

$$
e_{2,k} := x_k - \boldsymbol{\phi}^{\mathrm{T}}(\hat{\xi}_{k-1},k)\boldsymbol{\Theta} \in \mathbb{R}. \qquad (11)
$$

Substituting $\boldsymbol{\Theta} = \hat{\boldsymbol{\Theta}}_k$ into (6) gives

$$
\begin{aligned}
g_1[\mu_{1,k}] := J_1[\hat{\boldsymbol{\Theta}}_k] &= \frac{1}{2}[x_k - \boldsymbol{\phi}^{\mathrm{T}}(\xi,k)\hat{\boldsymbol{\Theta}}_k]^2 \\
&= \frac{1}{2}\{x_k - \boldsymbol{\phi}^{\mathrm{T}}(\xi,k)[\hat{\boldsymbol{\Theta}}_{k-1} + \mu_{1,k}\boldsymbol{\phi}(\xi,k)e_{1,k}]\}^2 \\
&= \frac{1}{2}\{x_k - \boldsymbol{\phi}^{\mathrm{T}}(\xi,k)\hat{\boldsymbol{\Theta}}_{k-1} - \mu_{1,k}\|\boldsymbol{\phi}(\xi,k)\|^2 e_{1,k}\}^2 \\
&= \frac{1}{2}\{e_{1,k} - \mu_{1,k}\|\boldsymbol{\phi}(\xi,k)\|^2 e_{1,k}\}^2 \\
&= \frac{1}{2}\{1 - \mu_{1,k}\|\boldsymbol{\phi}(\xi,k)\|^2\}^2 e_{1,k}^2.
\end{aligned}
$$

In order to make $J_1[\hat{\boldsymbol{\Theta}}_k]$ minimum, we take the optimal step-size $\mu_{1,k}$ as

$$
\mu_{1,k} = \frac{1}{\|\boldsymbol{\phi}(\xi,k)\|^2}. \qquad (12)
$$

To avoid the denominator being zero, the above equation can be modified to

$$
\mu_{1,k} = \frac{1}{1 + \|\boldsymbol{\phi}(\xi,k)\|^2}. \qquad (13)
$$

Substituting (12) or (13) into (8), we obtain the gain vector $\frac{\boldsymbol{\phi}(\xi,k)}{\|\boldsymbol{\phi}(\xi,k)\|^2}$ or $\frac{\boldsymbol{\phi}(\xi,k)}{1+\|\boldsymbol{\phi}(\xi,k)\|^2}$. Neither of these two gain vectors approaches zero with increasing k. From (8), we can see that when $\hat{\boldsymbol{\Theta}}_{k-1}$

is close to $\boldsymbol{\Theta}$, the large gain vector $\mu_{1,k}\boldsymbol{\phi}(\xi,k)$ will make $\hat{\boldsymbol{\Theta}}_k$ deviate from $\boldsymbol{\Theta}$. To address this problem, we let the step-size $\mu_{1,k}$ tend to zero with increasing k. Therefore, $\mu_{1,k}$ is taken as

$$\begin{aligned}\mu_{1,k} &:= \frac{1}{r_{1,k}},\\ r_{1,k} &= r_{1,k-1} + \|\boldsymbol{\phi}(\xi,k)\|^2.\end{aligned} \tag{14}$$

Similarly, substituting $\xi = \hat{\xi}_k$ into (7) gives

$$\begin{aligned}g_2[\mu_{2,k}] := J_2[\hat{\xi}_k] &= \frac{1}{2}[x_{1,k} - \psi(\boldsymbol{\beta})\mathrm{e}^{-\hat{\xi}_k x_{k-1}^2}]^2\\ &= \frac{1}{2}[x_k - \boldsymbol{X}_k^{\mathrm{T}}\boldsymbol{\alpha} - \psi(\boldsymbol{\beta})\mathrm{e}^{-\hat{\xi}_k x_{k-1}^2}]^2\\ &= \frac{1}{2}[x_k - \boldsymbol{\phi}^{\mathrm{T}}(\hat{\xi}_k,k)\boldsymbol{\Theta}]^2.\end{aligned}$$

Plugging the first-order Taylor expansion of $\boldsymbol{\phi}(\xi,k)$ at $\xi = \hat{\xi}_{k-1}$ into the above equation, we have

$$\begin{aligned}g_2[\mu_{2,k}] &= \frac{1}{2}\{x_k - [\boldsymbol{\phi}^{\mathrm{T}}(\hat{\xi}_{k-1},k) + [\boldsymbol{\phi}'(\hat{\xi}_{k-1},k)]^{\mathrm{T}}(\hat{\xi}_k - \hat{\xi}_{k-1}) + o(\hat{\xi}_k - \hat{\xi}_{k-1})]\boldsymbol{\Theta}\}^2\\ &= \frac{1}{2}\{x_k - [\boldsymbol{\phi}^{\mathrm{T}}(\hat{\xi}_{k-1},k) + [\boldsymbol{\phi}'(\hat{\xi}_{k-1},k)]^{\mathrm{T}}[\mu_{2,k}\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\phi}'(\hat{\xi}_{k-1},k)e_{2,k}] + o(\hat{\xi}_k - \hat{\xi}_{k-1})]\boldsymbol{\Theta}\}^2\\ &= \frac{1}{2}[x_k - \boldsymbol{\phi}^{\mathrm{T}}(\hat{\xi}_{k-1},k)\boldsymbol{\Theta} - [\boldsymbol{\phi}'(\hat{\xi}_{k-1},k)]^{\mathrm{T}}[\mu_{2,k}\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\phi}'(\hat{\xi}_{k-1},k)e_{2,k}]\boldsymbol{\Theta} + o(\hat{\xi}_k - \hat{\xi}_{k-1})]^2\\ &= \frac{1}{2}[e_{2,k} - \mu_{2,k}\|\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\phi}'(\hat{\xi}_{k-1},k)\|^2 e_{2,k} + o(\hat{\xi}_k - \hat{\xi}_{k-1})]^2\\ &= \frac{1}{2}[1 - \mu_{2,k}\|\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\phi}'(\hat{\xi}_{k-1},k)\|^2]^2 e_2^2(k) + o(\hat{\xi}_k - \hat{\xi}_{k-1})^2.\end{aligned}$$

The optimal $\mu_{2,k}$ can be obtained by minimizing $g_2[\mu_{2,k}]$, i.e., by solving the equation

$$1 - \mu_{2,k}\|\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\phi}'(\hat{\xi}_{k-1},k)\|^2 = 0.$$

Thus, the step-size $\mu_{2,k}$ can be chosen as

$$\mu_{2,k} = \frac{1}{\|\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\phi}'(\hat{\xi}_{k-1},k)\|^2}.$$

Similarly, considering the stability of the identification algorithm, the above equation can be modified to

$$\begin{aligned}\mu_{2,k} &:= \frac{1}{r_{2,k}},\\ r_{2,k} &= r_{2,k-1} + \|\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\phi}'(\hat{\xi}_{k-1},k)\|^2.\end{aligned} \tag{15}$$

Plugging (10), (14) into (8), and (11), (15) into (9), we obtain the following recursive relations:

$$\hat{\boldsymbol{\Theta}}_k = \hat{\boldsymbol{\Theta}}_{k-1} + \frac{1}{r_{1,k}}\boldsymbol{\phi}(\xi,k)e_{1,k}, \tag{16}$$

$$e_{1,k} = x_k - \boldsymbol{\phi}^{\mathrm{T}}(\xi,k)\hat{\boldsymbol{\Theta}}_{k-1}, \tag{17}$$

$$r_{1,k} = r_{1,k-1} + \|\boldsymbol{\phi}(\xi,k)\|^2, \tag{18}$$

$$\hat{\xi}_k = \hat{\xi}_{k-1} + \frac{1}{r_{2,k}}\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\phi}'(\hat{\xi}_{k-1},k)e_{2,k}, \tag{19}$$

$$e_{2,k} = x_k - \boldsymbol{\phi}^{\mathrm{T}}(\hat{\xi}_{k-1},k)\boldsymbol{\Theta}, \tag{20}$$

$$r_{2,k} = r_{2,k-1} + \|\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\phi}'(\hat{\xi}_{k-1},k)\|^2. \tag{21}$$

Here, a difficulty arises. Since the parameter sets $\boldsymbol{\Theta}$ and ξ, existing in the right-hand sides of (16)–(21), are to be estimated later, the algorithm in (16)–(21) cannot be realized. Inspired by

the hierarchical identification principle, we replace the unknown parameters ξ in (16)–(18) and $\boldsymbol{\Theta}$ in (19)–(21) with the estimates $\hat{\xi}_{k-1}$ and $\hat{\boldsymbol{\Theta}}_k$. It follows that

$$\hat{\boldsymbol{\Theta}}_k = \hat{\boldsymbol{\Theta}}_{k-1} + \frac{1}{r_{1,k}}\boldsymbol{\phi}(\hat{\xi}_{k-1}, k)e_{1,k}, \tag{22}$$

$$e_{1,k} = x_k - \boldsymbol{\phi}^{\mathrm{T}}(\hat{\xi}_{k-1}, k)\hat{\boldsymbol{\Theta}}_{k-1}, \tag{23}$$

$$r_{1,k} = r_{1,k-1} + \|\boldsymbol{\phi}(\hat{\xi}_{k-1}, k)\|^2, \tag{24}$$

$$\boldsymbol{\phi}(\hat{\xi}_{k-1}, k) = [\boldsymbol{X}_k^{\mathrm{T}}, \mathrm{e}^{-\hat{\xi}_{k-1}x_{k-1}^2}\boldsymbol{X}_k^{\mathrm{T}}]^{\mathrm{T}}, \tag{25}$$

$$\boldsymbol{X}_k = [x_{k-1}, x_{k-2}, \cdots, x_{k-n}]^{\mathrm{T}}, \tag{26}$$

$$\hat{\boldsymbol{\Theta}}_k = [\hat{\boldsymbol{\alpha}}_k^{\mathrm{T}}, \hat{\boldsymbol{\beta}}_k^{\mathrm{T}}]^{\mathrm{T}}, \tag{27}$$

$$\hat{\xi}_k = \hat{\xi}_{k-1} + \frac{1}{r_{2,k}}\hat{\boldsymbol{\Theta}}_k^{\mathrm{T}}\boldsymbol{\phi}'(\hat{\xi}_{k-1}, k)e_{2,k}, \tag{28}$$

$$e_{2,k} = x_k - \boldsymbol{\phi}^{\mathrm{T}}(\hat{\xi}_{k-1}, k)\hat{\boldsymbol{\Theta}}_k, \tag{29}$$

$$r_{2,k} = r_{2,k-1} + \|\hat{\boldsymbol{\Theta}}_k^{\mathrm{T}}\boldsymbol{\phi}'(\hat{\xi}_{k-1}, k)\|^2, \tag{30}$$

$$\boldsymbol{\phi}'(\hat{\xi}_{k-1}, k) = [\boldsymbol{0}_n^{\mathrm{T}}, -x_{k-1}^2\mathrm{e}^{-\hat{\xi}_{k-1}x_{k-1}^2}\boldsymbol{X}_k^{\mathrm{T}}]^{\mathrm{T}}. \tag{31}$$

The above computational process forms the H-SG algorithm for the ExpAR model.

The process of computing $\hat{\boldsymbol{\Theta}}_k$ and $\hat{\xi}_k$ by the H-SG algorithm is exhibited in the following list.

1. To initialize, let $k = 1$, $\hat{\boldsymbol{\Theta}}_0 = [\hat{\boldsymbol{\alpha}}_0^{\mathrm{T}}, \hat{\boldsymbol{\beta}}_0^{\mathrm{T}}]^{\mathrm{T}} = \boldsymbol{1}_{2n}/p_0$, $\hat{\xi}_0 = 1/p_0$, $p_0 = 10^6$, $r_{1,0} = 1$ and $r_{2,0} = 1$, give an error tolerance $\eta > 0$.
2. Collect the measurement data x_k, form the information vectors $\boldsymbol{X}_k$ and $\boldsymbol{\phi}(\hat{\xi}_{k-1}, k)$ by (26) and (25).
3. Compute the reciprocal of the step-size $r_{1,k}$ by (24) and the innovation $e_{1,k}$ by (23).
4. Update the parameter estimation vector $\hat{\boldsymbol{\Theta}}_k$ by (22), and read out $\hat{\boldsymbol{\alpha}}_k$ and $\hat{\boldsymbol{\beta}}_k$ from $\hat{\boldsymbol{\Theta}}_k$ in (27).
5. Form the derivative of $\boldsymbol{\phi}(\hat{\xi}_{k-1}, k)$ with respect to $\hat{\xi}_{k-1}$ by (31).
6. Compute the reciprocal of the step-size $r_{2,k}$ by (30) and the innovation $e_{2,k}$ by (29).
7. Update the parameter estimate $\hat{\xi}_k$ by (28).
8. Compare $\{\hat{\boldsymbol{\Theta}}_k, \hat{\xi}_k\}$ with $\{\hat{\boldsymbol{\Theta}}_{k-1}, \hat{\xi}_{k-1}\}$: if $\|\hat{\boldsymbol{\Theta}}_k - \hat{\boldsymbol{\Theta}}_{k-1}\| + \|\hat{\xi}_k - \hat{\xi}_{k-1}\| > \eta$, increase k by 1 and return to Step 2; otherwise, terminate this computational process.

The H-SG algorithm in (22)–(31) estimates the parameter sets $\boldsymbol{\Theta}$ and ξ in an interactive way. The innovations $e_{1,k}$ and $e_{2,k}$ in (23) and (29) are scalars. In order to make the most of the information, we derive an interactive multi-innovation parameter estimation method in the next section.

4 The hierarchical multi-innovation stochastic gradient algorithm

The innovation is the useful information which can improve the parameter and state estimation accuracy. The multi-innovation identification is the innovation expansion based identification [21]. Applying the multi-innovation identification theory, we expand the scalar innovations $e_{1,k}$ and $e_{2,k}$ in (23) and (29), and develop an H-MISG algorithm for the ExpAR model in this section.

Let l denote the innovation length. Expand the scalar innovations in (23) and (29) into the l-dimensional vectors:

$$\boldsymbol{E}_1(l) := \begin{bmatrix} x_k - \boldsymbol{\phi}^{\mathrm{T}}(\hat{\xi}_{k-1}, k)\hat{\boldsymbol{\Theta}}_{k-1} \\ x_{k-1} - \boldsymbol{\phi}^{\mathrm{T}}(\hat{\xi}_{k-1}, k-1)\hat{\boldsymbol{\Theta}}_{k-1} \\ \vdots \\ x_{k-l+1} - \boldsymbol{\phi}^{\mathrm{T}}(\hat{\xi}_{k-1}, k-l+1)\hat{\boldsymbol{\Theta}}_{k-1} \end{bmatrix} \in \mathbb{R}^l,$$

$$\boldsymbol{E}_2(l) := \begin{bmatrix} x_k - \boldsymbol{\phi}^{\mathrm{T}}(\hat{\xi}_{k-1}, k)\hat{\boldsymbol{\Theta}}_k \\ x_{k-1} - \boldsymbol{\phi}^{\mathrm{T}}(\hat{\xi}_{k-1}, k-1)\hat{\boldsymbol{\Theta}}_k \\ \vdots \\ x_{k-l+1} - \boldsymbol{\phi}^{\mathrm{T}}(\hat{\xi}_{k-1}, k-l+1)\hat{\boldsymbol{\Theta}}_k \end{bmatrix} \in \mathbb{R}^l.$$

Define the following stacked vector and matrix:

$$\boldsymbol{X}(l) := \begin{bmatrix} x_k \\ x_{k-1} \\ \vdots \\ x_{k-l+1} \end{bmatrix} \in \mathbb{R}^l, \quad \boldsymbol{\Phi}(l, \hat{\xi}_{k-1}) := \begin{bmatrix} \boldsymbol{\phi}^{\mathrm{T}}(\hat{\xi}_{k-1}, k) \\ \boldsymbol{\phi}^{\mathrm{T}}(\hat{\xi}_{k-1}, k-1) \\ \vdots \\ \boldsymbol{\phi}^{\mathrm{T}}(\hat{\xi}_{k-1}, k-l+1) \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{(2n)\times l}.$$

Then the innovation vectors can be equivalently transformed into

$$\boldsymbol{E}_1(l) = \boldsymbol{X}(l) - \boldsymbol{\Phi}^{\mathrm{T}}(l, \hat{\xi}_{k-1})\hat{\boldsymbol{\Theta}}_{k-1},$$
$$\boldsymbol{E}_2(l) = \boldsymbol{X}(l) - \boldsymbol{\Phi}^{\mathrm{T}}(l, \hat{\xi}_{k-1})\hat{\boldsymbol{\Theta}}_k.$$

Since $\boldsymbol{E}_1(l) = e_{1,k}$, $\boldsymbol{\Phi}(l, \hat{\xi}_{k-1}) = \boldsymbol{\phi}(\hat{\xi}_{k-1}, k)$ and $\boldsymbol{X}(l) = x_k$ for $l = 1$, Equation (22) can be written as

$$\hat{\boldsymbol{\Theta}}_k = \hat{\boldsymbol{\Theta}}_{k-1} + \frac{1}{r_{1,k}}\boldsymbol{\Phi}(l, \hat{\xi}_{k-1})\boldsymbol{E}_1(l).$$

Similarly, Equation (28) can be transformed into

$$\hat{\xi}_k = \hat{\xi}_{k-1} + \frac{1}{r_{2,k}}\hat{\boldsymbol{\Theta}}_k^{\mathrm{T}}\boldsymbol{\Phi}'(l, \hat{\xi}_{k-1})\boldsymbol{E}_2(l),$$

where

$$\boldsymbol{\Phi}'(l, \hat{\xi}_{k-1}) := [\boldsymbol{\phi}'(\hat{\xi}_{k-1}, k), \boldsymbol{\phi}'(\hat{\xi}_{k-1}, k-1), \cdots, \boldsymbol{\phi}'(\hat{\xi}_{k-1}, k-l+1)] \in \mathbb{R}^{(2n)\times l}.$$

In summary, the H-MISG algorithm for the ExpAR model can be derived as follows,

$$\hat{\boldsymbol{\Theta}}_k = \hat{\boldsymbol{\Theta}}_{k-1} + \frac{1}{r_{1,k}}\boldsymbol{\Phi}(l, \hat{\xi}_{k-1})\boldsymbol{E}_1(l), \tag{32}$$
$$\boldsymbol{E}_1(l) = \boldsymbol{X}(l) - \boldsymbol{\Phi}^{\mathrm{T}}(l, \hat{\xi}_{k-1})\hat{\boldsymbol{\Theta}}_{k-1}, \tag{33}$$
$$r_{1,k} = r_{1,k-1} + \|\boldsymbol{\phi}(\hat{\xi}_{k-1}, k)\|^2, \tag{34}$$
$$\boldsymbol{X}(l) = [x_{k-1}, x_{k-2}, \cdots, x_{k-l+1}]^{\mathrm{T}}, \tag{35}$$
$$\boldsymbol{\Phi}(l, \hat{\xi}_{k-1}) = [\boldsymbol{\phi}(\hat{\xi}_{k-1}, k), \boldsymbol{\phi}(\hat{\xi}_{k-1}, k-1), \cdots, \boldsymbol{\phi}(\hat{\xi}_{k-1}, k-l+1)], \tag{36}$$
$$\boldsymbol{\phi}(\hat{\xi}_{k-1}, k) = [\boldsymbol{X}_k^{\mathrm{T}}, \mathrm{e}^{-\hat{\xi}_{k-1}x_{k-1}^2}\boldsymbol{X}_k^{\mathrm{T}}]^{\mathrm{T}}, \tag{37}$$
$$\boldsymbol{X}_k = [x_{k-1}, x_{k-2}, \cdots, x_{k-n}]^{\mathrm{T}}, \tag{38}$$
$$\hat{\boldsymbol{\Theta}}_k = [\hat{\boldsymbol{\alpha}}_k^{\mathrm{T}}, \hat{\boldsymbol{\beta}}_k^{\mathrm{T}}]^{\mathrm{T}}, \tag{39}$$
$$\hat{\xi}_k = \hat{\xi}_{k-1} + \frac{1}{r_{2,k}}\hat{\boldsymbol{\Theta}}_k^{\mathrm{T}}\boldsymbol{\Phi}'(l, \hat{\xi}_{k-1})\boldsymbol{E}_2(l), \tag{40}$$
$$\boldsymbol{E}_2(l) = \boldsymbol{X}(l) - \boldsymbol{\Phi}^{\mathrm{T}}(l, \hat{\xi}_{k-1})\hat{\boldsymbol{\Theta}}_k, \tag{41}$$
$$r_{2,k} = r_{2,k-1} + \|\hat{\boldsymbol{\Theta}}_k^{\mathrm{T}}\boldsymbol{\phi}'(\hat{\xi}_{k-1}, k)\|^2, \tag{42}$$
$$\boldsymbol{\phi}'(\hat{\xi}_{k-1}, k) = [\boldsymbol{0}_n^{\mathrm{T}}, -x_{k-1}^2\mathrm{e}^{-\hat{\xi}_{k-1}x_{k-1}^2}\boldsymbol{X}_k^{\mathrm{T}}]^{\mathrm{T}}, \tag{43}$$
$$\boldsymbol{\Phi}'(l, \hat{\xi}_{k-1}) = [\boldsymbol{\phi}'(\hat{\xi}_{k-1}, k), \boldsymbol{\phi}'(\hat{\xi}_{k-1}, k-1), \cdots, \boldsymbol{\phi}'(\hat{\xi}_{k-1}, k-l+1)]. \tag{44}$$

When $l = 1$, the H-MISG degenerates into the H-SG algorithm.

The H-MISG algorithm in (32)–(44) can be implemented by the following steps.

1. Set the innovation length l and initialize: let $k = 1$, $\hat{\boldsymbol{\Theta}}_0 = [\hat{\boldsymbol{\alpha}}_0^{\mathrm{T}}, \hat{\boldsymbol{\beta}}_0^{\mathrm{T}}]^{\mathrm{T}} = \boldsymbol{1}_{2n}/p_0$, $\hat{\xi}_0 = 1/p_0$, $p_0 = 10^6$, $r_{1,0} = 1$ and $r_{2,0} = 1$, give an error tolerance $\eta > 0$.

2. Collect the measurement data x_k, form the stacked information vector $\boldsymbol{X}(l)$ by (35), the information vectors $\boldsymbol{X}_k$ and $\boldsymbol{\phi}(\hat{\xi}_{k-1}, k)$ by (38) and (37), and $\boldsymbol{\Phi}(l, \hat{\xi}_{k-1})$ by (36).
3. Compute the reciprocal of the step-size $r_{1,k}$ by (34) and the innovation vector $\boldsymbol{E}_1(l)$ by (33).
4. Update the parameter estimation vector $\hat{\boldsymbol{\Theta}}_k$ by (32), and read out $\hat{\boldsymbol{\alpha}}_k$ and $\hat{\boldsymbol{\beta}}_k$ from (39).
5. Form the derivative of $\boldsymbol{\phi}(\hat{\xi}_{k-1}, k)$ by (43), and $\boldsymbol{\Phi}'(l, \hat{\xi}_{k-1})$ by (44).
6. Compute the reciprocal of the step-size $r_{2,k}$ by (42) and the innovation vector $\boldsymbol{E}_2(l)$ by (41).
7. Update the parameter estimate $\hat{\xi}_k$ by (40).
8. Compare $\{\hat{\boldsymbol{\Theta}}_k, \hat{\xi}_k\}$ with $\{\hat{\boldsymbol{\Theta}}_{k-1}, \hat{\xi}_{k-1}\}$: if $\|\hat{\boldsymbol{\Theta}}_k - \hat{\boldsymbol{\Theta}}_{k-1}\| + \|\hat{\xi}_k - \hat{\xi}_{k-1}\| > \eta$, increase k by 1 and return to Step 2; otherwise, stop this computational process.

Remark 2: In order to obtain more accurate parameter estimates but not increase the computational cost of the H-MISG algorithm, we introduce the forgetting factors (FF) λ_1 and λ_2 into (34) and (42):

$$r_{1,k} = \lambda_1 r_{1,k-1} + \|\boldsymbol{\phi}(\hat{\xi}_{k-1}, k)\|^2, \quad 0 \leq \lambda_1 < 1, \tag{45}$$

$$r_{2,k} = \lambda_2 r_{2,k-1} + \|\hat{\boldsymbol{\Theta}}_k^{\mathrm{T}} \boldsymbol{\phi}'(\hat{\xi}_{k-1}, k)\|^2, \quad 0 \leq \lambda_2 < 1. \tag{46}$$

Replacing (34) and (42) in the H-MISG algorithm with (45) and (46), we obtain the variant of the H-MISG, i.e., the FF-H-MISG algorithm for the ExpAR model. When $\lambda_1 = 1$ and $\lambda_2 = 1$, the FF-H-MISG degenerates into the H-MISG algorithm.

Remark 3: Before using the proposed algorithms to identify the ExpAR model, we need to determine the order from input-output data by using the order estimation methods, such as the orthogonalization procedure and the correlation analysis in [22].

At each recursion, the H-SG algorithm involves the current measurement data and innovation, the H-MISG or the FF-H-MISG algorithm applies all the current and the preceding $(l-1)$ measurement data and innovations, which makes the latter has a higher parameter estimation accuracy.

5 Example

Consider the following ExpAR time series

$$\begin{aligned} x_k &= \left(\alpha_1 + \beta_1 \mathrm{e}^{-\xi x_{k-1}^2}\right) x_{k-1} + \left(\alpha_2 + \beta_2 \mathrm{e}^{-\xi x_{k-1}^2}\right) x_{k-2} + \cdots \\ &\quad + \left(\alpha_n + \beta_n \mathrm{e}^{-\xi x_{k-1}^2}\right) x_{k-n} + \varepsilon_k \\ &= \left(1.25 + 2.00\mathrm{e}^{-2.30 x_{k-1}^2}\right) x_{k-1} + \left(-0.28 + 1.85\mathrm{e}^{-2.30 x_{k-1}^2}\right) x_{k-2} + \varepsilon_k. \end{aligned}$$

The parameters to be identified are

$$\boldsymbol{\Theta} = [\alpha_1, \alpha_2, \beta_1, \beta_2]^{\mathrm{T}} = [1.25, -0.28, 2.00, 1.85]^{\mathrm{T}}, \quad \xi = 2.30.$$

In simulation, the variance of the white noise $\{\varepsilon_k\}$ is set to be σ^2, the measurement data length is taken as $L_e = 3000$. For simplicity, we define $\boldsymbol{\vartheta} := [\boldsymbol{\Theta}^{\mathrm{T}}, \xi]^{\mathrm{T}}$.

Taking $\sigma^2 = 0.20^2$ and using the H-SG algorithm, H-MISG algorithm with $l = 5$ and FF-H-MISG algorithm with $l = 5$, $\lambda_1 = 0.91$ and $\lambda_2 = 1.00$ to identify this ExpAR model, respectively, the parameter estimates and their errors are shown in Tables 2–4, the parameter estimation errors $\delta := \|\hat{\boldsymbol{\vartheta}}_k - \boldsymbol{\vartheta}\|/\|\boldsymbol{\vartheta}\| \times 100\%$ versus k are shown in Figure 2.

To illustrate the advantage of the proposed multi-innovation identification algorithms, we fix the noise variance $\sigma^2 = 0.20^2$, the forgetting factors $\lambda_1 = 0.91$ and $\lambda_2 = 1.00$, and adopt the FF-H-MISG algorithm to identify this ExpAR model with the innovation length $l = 5$, $l = 6$ and $l = 7$. The corresponding results are demonstrated in Table 5 and Figure 3.

To demonstrate how the performance of the proposed FF-H-MISG algorithm depends on the forgetting factors, we fix the noise variance $\sigma^2 = 0.20^2$, the innovation length $l = 7$, the forgetting factor $\lambda_2 = 1.00$, and adopt the FF-H-MISG algorithm to identify this ExpAR model with the forgetting factor $\lambda_1 = 0.91$, $\lambda_1 = 0.97$ and $\lambda_1 = 0.99$. The corresponding results are exhibited in Table 6 and Figure 4.

To show the influence of the noise level on the proposed FF-H-MISG algorithm, we fix the innovation length $l = 7$, the forgetting factors $\lambda_1 = 0.91$, $\lambda_2 = 1.00$, and adopt the FF-H-MISG algorithm to identify this ExpAR model with the noise variance $\sigma^2 = 0.20^2$, $\sigma^2 = 0.23^2$ and $\sigma^2 = 0.26^2$. The results are shown in Table 7 and Figure 5.

Table 2 The H-SG estimates and errors ($\sigma^2 = 0.20^2$)

k	α_1	α_2	β_1	β_2	ξ	δ (%)
100	0.23837	0.24173	0.23671	0.23995	-0.00786	92.65856
200	0.24113	0.24281	0.23950	0.24102	-0.00885	92.60908
500	0.24379	0.24329	0.24217	0.24145	-0.00850	92.54419
1000	0.24513	0.24309	0.24350	0.24120	-0.00738	92.49927
2000	0.24640	0.24283	0.24477	0.24090	-0.00610	92.45321
3000	0.24725	0.24263	0.24562	0.24068	-0.00538	92.42506
True values	1.25000	-0.28000	2.00000	1.85000	2.30000	

Table 3 The H-MISG estimates and errors ($\sigma^2 = 0.20^2$, $l = 5$)

k	α_1	α_2	β_1	β_2	ξ	δ (%)
100	0.48056	0.44219	1.12466	1.13314	1.54026	45.49540
200	0.51411	0.45249	1.12335	1.13063	1.54783	45.17986
500	0.52195	0.43228	1.12051	1.12670	1.58769	44.49311
1000	0.53864	0.42754	1.11815	1.12327	1.63957	43.76030
2000	0.54514	0.41393	1.11625	1.12065	1.72130	42.77483
3000	0.55058	0.40671	1.11540	1.11946	1.76471	42.26100
True values	1.25000	-0.28000	2.00000	1.85000	2.30000	

Table 4 The FF-H-MISG estimates and errors ($\sigma^2 = 0.20^2$, $l = 5$, $\lambda_1 = 0.91$, $\lambda_2 = 1.00$)

k	α_1	α_2	β_1	β_2	ξ	δ (%)
100	0.49534	0.27397	1.13043	1.13351	1.53723	43.59349
200	0.69813	0.22973	1.13034	1.12075	1.53837	41.16763
500	0.99277	0.03144	1.12374	1.09265	1.53952	38.09418
1000	1.16897	-0.17448	1.13851	1.08685	1.52340	36.81494
2000	1.28966	-0.33550	1.20027	1.14395	1.48566	35.47039
3000	1.31692	-0.39613	1.26166	1.20153	1.45519	34.38786
True values	1.25000	-0.28000	2.00000	1.85000	2.30000	

From Tables 2–7 and Figures 2–5, we draw the following conclusions.

• The parameter estimation errors decrease as the data length k increases for all the algorithms proposed in this paper. The FF-H-MISG algorithm has the highest parameter estimation accuracy among these three algorithms – see Tables 2–4 and Figure 2.

• The parameter estimation accuracy becomes higher with the innovation length l increasing and the forgetting factor decreasing for the FF-H-MISG algorithm – see Tables 5–6 and Figures 3–4.

• The estimation errors of the FF-H-MISG algorithm tend to zero with the decreasing of noise levels – see Table 7 and Figure 5.

• The proposed FF-H-MISG algorithm with appropriate innovation length and forgetting factors is effective to identify the nonlinear ExpAR process – see Tables 5–6 and Figures 3–4.

For the model validation, we use $L_r = 200$ observations from $k = L_e + 1$ to $k = L_e + L_r$ and the predicted model gave by the FF-H-MISG algorithm with $\lambda_1 = 0.91$, $\lambda_2 = 1.00$ and $l = 7$. The predicted data $\hat{x}_k$ and the measurement data x_k are plotted in Figure 6. To evaluate the prediction performance, we define and compute the mean square error (MSE)

$$MSE := \left[\frac{1}{L_r} \sum_{k=L_e+1}^{L_e+L_r} (\hat{x}_k - x_k)^2 \right]^{1/2} = 0.19635.$$

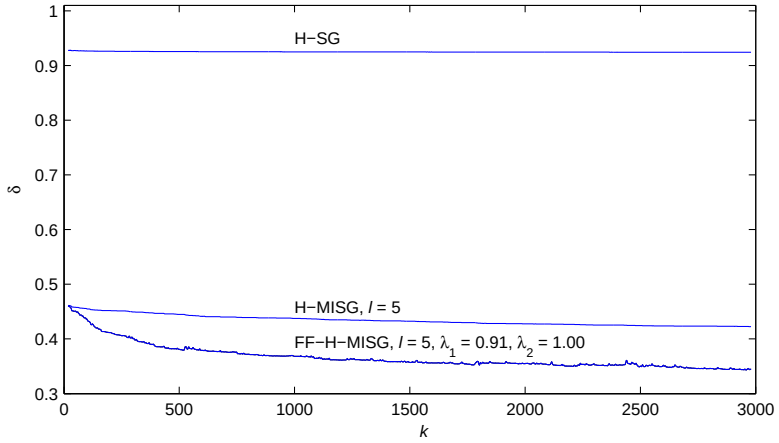


Fig. 2 The H-SG, H-MISG and FF-H-MISG estimation errors δ versus k

Table 5 The FF-H-MISG estimates and errors ($\sigma^2 = 0.20^2$, $\lambda_1 = 0.91$, $\lambda_2 = 1.00$)

l	k	α_1	α_2	β_1	β_2	ξ	δ (%)
5	100	0.49534	0.27397	1.13043	1.13351	1.53723	43.59349
	200	0.69813	0.22973	1.13034	1.12075	1.53837	41.16763
	500	0.99277	0.03144	1.12374	1.09265	1.53952	38.09418
	1000	1.16897	-0.17448	1.13851	1.08685	1.52340	36.81494
	2000	1.28966	-0.33550	1.20027	1.14395	1.48566	35.47039
	3000	1.31692	-0.39613	1.26166	1.20153	1.45519	34.38786
6	100	0.53946	0.26806	1.36935	1.36071	1.92855	33.18161
	200	0.73563	0.20753	1.37115	1.35494	1.92911	29.86680
	500	1.02242	0.02122	1.37425	1.34767	1.92839	25.37927
	1000	1.16564	-0.15212	1.39993	1.36572	1.91305	23.12858
	2000	1.23194	-0.27820	1.48535	1.45777	1.87609	20.42404
	3000	1.24631	-0.31193	1.57122	1.54571	1.84419	18.38652
7	100	0.54603	0.29172	1.60914	1.61667	2.36236	26.83406
	200	0.73972	0.20569	1.61202	1.61533	2.36188	22.17354
	500	1.03169	0.02289	1.62143	1.62111	2.35862	15.35780
	1000	1.14722	-0.12360	1.65573	1.65741	2.34266	11.57955
	2000	1.18572	-0.20994	1.75055	1.76906	2.30523	7.36433
	3000	1.18065	-0.23155	1.84794	1.87601	2.27055	4.70866
True values		1.25000	-0.28000	2.00000	1.85000	2.30000	

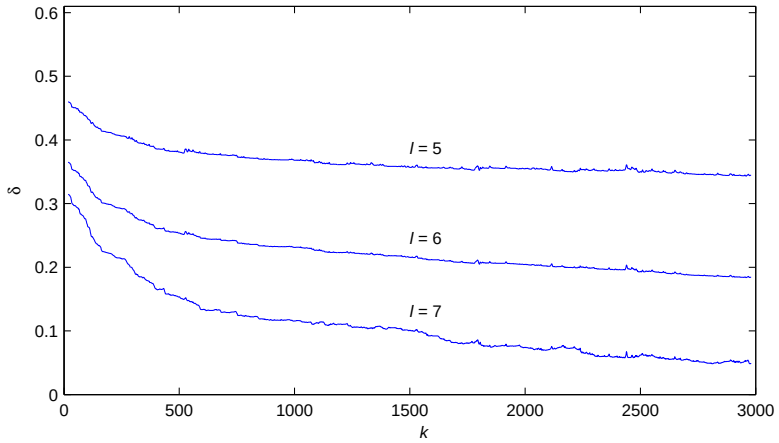


Fig. 3 The FF-H-MISG estimation errors δ versus k ($\sigma^2 = 0.20^2$, $\lambda_1 = 0.91$, $\lambda_2 = 1.00$)

From Figure 6, we can see that the predicted data is close to the measurement data, which means the predicted model can reveal the dynamics of this ExpAR process.

Table 6 The FF-H-MISG estimates and errors ($\sigma^2 = 0.20^2$, $l = 7$, $\lambda_2 = 1.00$)

λ_1	k	α_1	α_2	β_1	β_2	ξ	δ (%)
0.91	100	0.54603	0.29172	1.60914	1.61667	2.36236	26.83406
	200	0.73972	0.20569	1.61202	1.61533	2.36188	22.17354
	500	1.03169	0.02289	1.62143	1.62111	2.35862	15.35780
	1000	1.14722	-0.12360	1.65573	1.65741	2.34266	11.57955
	2000	1.18572	-0.20994	1.75055	1.76906	2.30523	7.36433
	3000	1.18065	-0.23155	1.84794	1.87601	2.27055	4.70866
0.97	100	0.50182	0.39921	1.60333	1.61150	2.35988	29.37939
	200	0.59836	0.37113	1.60352	1.60999	2.35995	27.26498
	500	0.75346	0.24373	1.60370	1.60784	2.35917	22.70653
	1000	0.91250	0.07740	1.60520	1.60535	2.35349	17.90619
	2000	1.08223	-0.08715	1.62537	1.62689	2.32268	13.35437
	3000	1.15777	-0.18392	1.65491	1.65908	2.28378	10.99554
0.99	100	0.49891	0.44383	1.60125	1.60958	2.35939	30.19217
	200	0.54791	0.43635	1.60106	1.60860	2.36015	29.23940
	500	0.59940	0.36804	1.60001	1.60648	2.36314	27.25883
	1000	0.68393	0.28638	1.59785	1.60200	2.36847	24.60749
	2000	0.81932	0.17078	1.59786	1.60014	2.36203	20.72777
	3000	0.93413	0.05726	1.60147	1.60275	2.34243	17.41429
True values		1.25000	-0.28000	2.00000	1.85000	2.30000	

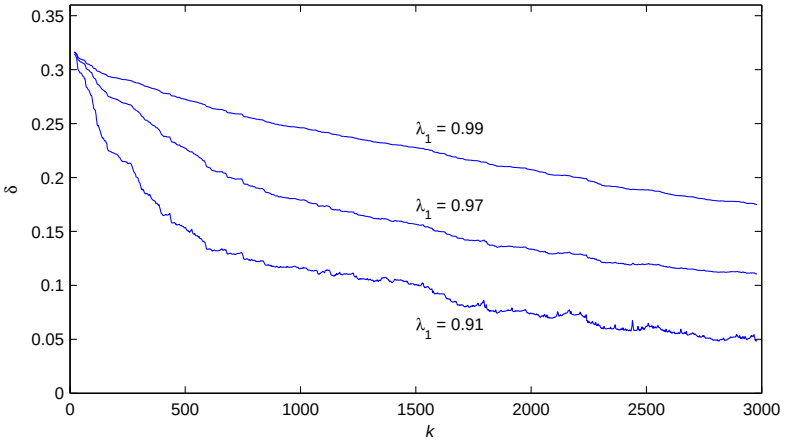


Fig. 4 The FF-H-MISG estimation errors δ versus k ($\sigma^2 = 0.20^2$, $l = 7$, $\lambda_2 = 1.00$)

Table 7 The FF-H-MISG estimates and errors ($l = 7$, $\lambda_1 = 0.91$, $\lambda_2 = 1.00$)

σ^2	k	α_1	α_2	β_1	β_2	ξ	δ (%)
0.20^2	100	0.54603	0.29172	1.60914	1.61667	2.36236	26.83406
	200	0.73972	0.20569	1.61202	1.61533	2.36188	22.17354
	500	1.03169	0.02289	1.62143	1.62111	2.35862	15.35780
	1000	1.14722	-0.12360	1.65573	1.65741	2.34266	11.57955
	2000	1.18572	-0.20994	1.75055	1.76906	2.30523	7.36433
	3000	1.18065	-0.23155	1.84794	1.87601	2.27055	4.70866
0.23^2	100	0.56746	0.24224	1.59956	1.60044	1.97636	27.25173
	200	0.80713	0.13856	1.59557	1.58461	1.97839	22.22129
	500	1.13045	-0.08102	1.58406	1.55834	1.98164	16.96918
	1000	1.22336	-0.21425	1.59756	1.56192	1.97465	15.74527
	2000	1.26659	-0.29252	1.66707	1.63962	1.95573	13.81780
	3000	1.25020	-0.30111	1.73947	1.71097	1.94113	12.27984
0.26^2	100	0.59214	0.19257	1.58259	1.57361	1.67381	30.08145
	200	0.87582	0.07124	1.56283	1.52826	1.67857	25.64775
	500	1.21452	-0.17371	1.51219	1.44412	1.68690	23.47748
	1000	1.27296	-0.28121	1.49179	1.40197	1.68458	24.16677
	2000	1.31721	-0.34485	1.51491	1.42746	1.67419	23.81738
	3000	1.28663	-0.33779	1.54898	1.45054	1.66876	23.10347
True values		1.25000	-0.28000	2.00000	1.85000	2.30000	

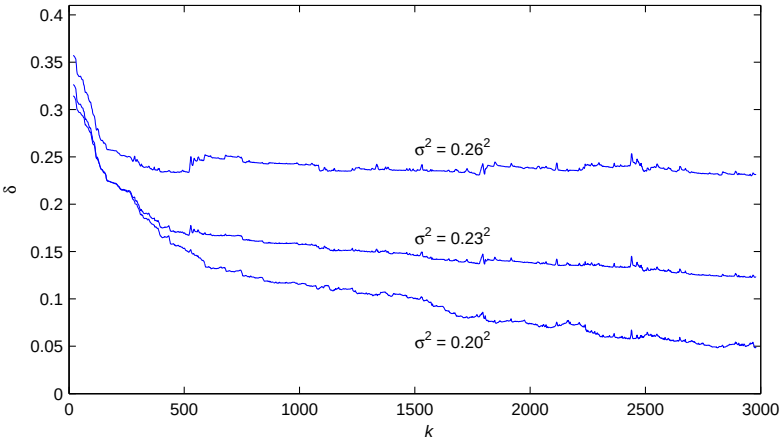
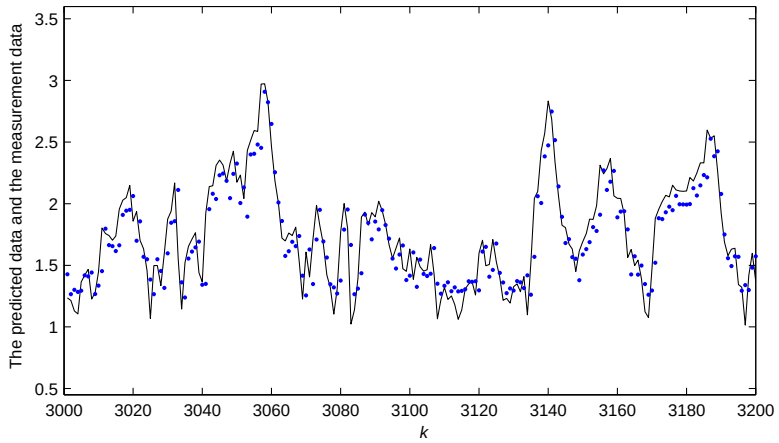


Fig. 5 The FF-H-MISG estimation errors δ versus k ($l = 7$, $\lambda_1 = 0.91$, $\lambda_2 = 1.00$)



Solid line: the measurement data, dots line: the predicted data

Fig. 6 The predicted data $\hat{x}_k$ and the measurement data x_k for the FF-H-MISG algorithm

6 Conclusions

Applying the hierarchical identification principle and the multi-innovation identification theory, this paper derives an H-SG algorithm and an H-MISG algorithm for the ExpAR model. For the sake of the improved estimation accuracy, two forgetting factors are introduced into the H-MISG, and a variant of the H-MISG, i.e., the FF-H-MISG algorithm is presented in this work. The simulation results demonstrate that the FF-H-MISG algorithm with appropriate innovation length and forgetting factors is effective to identify the ExpAR model. Jointing the neural network [23, 24], the kernel collocation [25,26] and other mathematical tools [27], the algorithms proposed in this paper can be exploit to study parameter identification of different systems and can be applied to other fields.

Acknowledgements This work was supported by the 111 Project (B12018).

Compliance with ethical standards

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

References

1. Gan, M., Li, H.X., Peng, H.: A variable projection approach for efficient estimation of RBF-ARX model. IEEE Transactions on Cybernetics **45**(3), 462-471 (2015)
2. Ozaki, T.: Non-linear time series models for non-linear random vibrations. Journal of Applied Probability **17**(1), 84-93 (1980)
3. Ozaki, T.: The statistical analysis of perturbed limit cycle processes using nonlinear time series models. Journal of Time Series Analysis **3**(1), 29-41 (1982)
4. Teräsvirta, T.: Specification, estimation, and evaluation of smooth transition autoregressive models. Journal of the American Statistical Association **89**(425), 208-218 (1994)
5. Merzougui, M., Dridi, H., Chadli, A.: Test for periodicity in restrictive EXPAR models. Communications in Statistics – Theory and Methods **45**(9), 2770-2783 (2016)
6. Chen, G.Y., Gan, M., Chen, G.L.: Generalized exponential autoregressive models for nonlinear time series: Stationarity, estimation and applications. Information Sciences **438**, 46-57 (2018)
7. Zhou, Z.P., Liu, X.F.: State and fault estimation of sandwich systems with hysteresis. International Journal of Robust and Nonlinear Control **28**(13), 3974-3986 (2018)
8. Yu, C.P., Verhaegen, M., Hansson, A.: Subspace identification of local systems in one-dimensional homogeneous networks. IEEE Transactions on Automatic Control **63**(4), 1126-1131 (2018)
9. Li, B., Wu, N., Wang, H., Tseng, P.H., Kuang, J.M.: Gaussian message passing-based cooperative localization on factor graph in wireless networks. Signal Processing **111**, 1-12 (2015)
10. Schoukens, M., Tiels, K.: Identification of block-oriented nonlinear systems starting from linear approximations: A survey. Automatica **85**, 272-292 (2017)
11. Arqub, O.A., Abo-Hammour, Z.: Numerical solution of systems of second-order boundary value problems using continuous genetic algorithm. Information Sciences **279**, 396-415 (2014)
12. Yu, C.P., Verhaegen, M.: Blind multivariable ARMA subspace identification. Automatica **66**, 3-14 (2016)
13. Yu, C.P., Verhaegen, M.: Data-driven fault estimation of non-minimum phase LTI systems. Automatica **92**, 181-187 (2018)
14. Chen, J., Jiang, B.: Modified stochastic gradient parameter estimation algorithms for a nonlinear two-variable difference system. International Journal of Control, Automation and Systems **14**(6), 1493-1500 (2016)
15. Chen, F.W., Garnier, H., Gilson, M.: Robust identification of continuous-time models with arbitrary time-delay from irregularly sampled data. Journal of Process Control **25**, 19-27 (2015)
16. Ding, F., Xu, L., Liu, X.M.: Signal modeling – Part F: Hierarchical iterative parameter estimation for multi-frequency signal models. Journal of Qingdao University of Science and Technology (Natural Science Edition) **38**(6), 1-13 (2017)
17. Cao, P.F., Luo, X.L.: Performance analysis of multi-innovation stochastic Newton recursive algorithms. Digital Signal Processing **56**, 15-23 (2016)
18. Li, L.W., Ren, X.M., Guo, F.M.: Modified multi-innovation stochastic gradient algorithm for Wiener-Hammerstein systems with backlash. Journal of the Franklin Institute **355**(9), 4050-4075 (2018)
19. Cheng, S.S., Wei, Y.H., Sheng, D., Chen, Y.Q., Wang, Y.: Identification for Hammerstein nonlinear ARMAX systems based on multi-innovation fractional order stochastic gradient. Signal Processing **142**, 1-10 (2018)
20. Ding, F., Xu L., Liu X.M.: Signal modeling – Part E: Hierarchical parameter estimation for multi-frequency signal models. Journal of Qingdao University of Science and Technology (Natural Science Edition) **38**(5), 1-15 (2017)
21. Ding, F.: System Identification – Multi-Innovation Identification Theory and Methods. Science Press, Beijing (2016)
22. Zhang, B., Billings, S.A.: Identification of continuous-time nonlinear systems: the nonlinear difference equation with moving average noise (NDEMA) framework. Mechanical Systems and Signal Processing **60-61**, 810-835 (2015)
23. Li, X., Zhu, D.Q.: An improved SOM neural network method to adaptive leader-follower formation control of AUVs. IEEE Transactions on Industrial Electronics **65**(10), 8260-8270 (2018)
24. Chen, M.Z., Zhu, D.Q.: A workload balanced algorithm for task assignment and path planning of inhomogeneous autonomous underwater vehicle system. IEEE Transactions on Cognitive and Developmental Systems (2018). https://doi.org/10.1109/TCDS.2018.2866984
25. Geng, F.Z., Qian, S.P.: An optimal reproducing kernel method for linear nonlocal boundary value problems. Applied Mathematics Letters **77**, 49-56 (2018)
26. Li, X.Y., Wu, B.Y.: A new reproducing kernel collocation method for nonlocal fractional boundary value problems with non-smooth solutions. Applied Mathematics Letters **86**, 194-199 (2018)
27. El-Ajou, A., Arqub, O.A., Al-Smadi, M.: A general form of the generalized Taylor's formula with some applications. Applied Mathematics and Computation **256**, 851-859 (2015)