A nonlinear recursive instrumental variables identification method of Hammerstein ARMAX system

Liang Ma · Xinggao Liu

Abstract Identification of nonlinear Hammerstein models has received much attention due to its ability to describe a wide variety of nonlinear systems. This paper considers the problem of the parameter estimation of the ARMAX models for the Hammerstein systems. A novel nonlinear recursive instrumental variables method, which is simple and easy for practical applications, is proposed to deal with the problem. In order to make the instrumental variables uncorrelated with the colored noise and to obtain better identification effect, three approaches for choosing the instrumental variables usually used in the linear RIV method are introduced. Furthermore, the procedure of the nonlinear RIV method and its property of the mean square convergence of the nonlinear RIV method are rigorously derived. Finally, an example is carried out as illustration, where the ARMAX-RLS method is compared as the basis, and the results show that the nonlinear RIV method is superior to ARMAX-RLS method in terms of identification accuracy and convergence speed under colored noise, which reveals the effectiveness of the proposed method.

Keywords Hammerstein ARMAX system · The nonlinear RIV method · Mean square convergence · ARMAX-RLS

L. Ma · X. Liu (✉)
State Key Laboratory of Industrial Control Technology,
Department of Control Science and Engineering, Zhejiang
University, Hangzhou 310027, People's Republic of China
e-mail: lxg@zjuem.zju.edu.cn

List of symbols

$y(t)$	System output
$x(t)$	True output, not corrupted by noise
$v(t)$	White noise
$e(t)$	Colored noise
$\bar{u}(t)$	Inner variable
a_i, b_i, c_i, d_i	System parameters
$f(u(t))$	Nonlinear function
$w_1, w_2, \ldots, w_{n_c}$	Basis of a nonlinear function
σ_{v}^2	Variance of white noise
$h^*(t)$	The instrumental variable
$\hat{\boldsymbol{\theta}}_{\mathrm{IV}}$	The identification result using instrumental variable method
$\hat{\boldsymbol{\theta}}_{\mathrm{RIV}}$	The identification result using nonlinear RIV method
$\hat{\boldsymbol{\theta}}_{\mathrm{ARMAX\text{-}RLS}}$	The identification result using ARMAX-RLS method
NSR	Noise-to-Signal-Ratio, defined as $\mathrm{NSR} = \sqrt{\frac{\mathrm{var}[v(t)]}{\mathrm{var}[u(t)]}} \times 100\,\%$
P	Probability

1 Introduction

Most real-life systems can be modeled quite well with a linear model, but in practice, almost all processes are nonlinear if they are considered not merely in a small vicinity of their working points, so better results can be obtained by using a nonlinear model. As is known, nonlinear models are commonly used to describe the

behavior of many industrial processes because of the increased accuracy and performances of the system behavior, which has drawn great attention of many researchers. And the nonlinear time series analysis is usually a very useful and widely used tool to study the observed nonlinear behavior [1–5]. Among the nonlinear models, the so-called block-oriented models such as the Hammerstein model, Wiener model and Hammerstein–Wiener model have turned out to be very useful for the estimation of nonlinear systems. A great deal of work on the identification of these models has been proposed [6–20]. The Hammerstein model, which consists of a static nonlinear block followed by a linear time-invariant subsystem, is the focus of this paper.

The identification of Hammerstein systems has been extensively studied in the recent years, so there exist a large number of studies on this topic in the literature, which can be roughly classified into several categories, namely the over-parameterization method [21–24], the nonparametric method [25–28], the iterative method [29–32], the gradient algorithm [33–35], the kernel machine and space projection method [36–38] and the recursive algorithm [23,39–42]. Recently, Ding et al. [43] put forward decomposition-based Newton iterative method for the identification of a Hammerstein nonlinear FIR system with ARMA noise. Li [44] developed an algorithm based on Newton iteration for the identification of Hammerstein model. Hong and Mitchell [45] used the Bezier–Bernstein approximation of the nonlinear static function to identify the model. Han and Raymond [46] extended the rank minimization approach to Hammerstein system identification by using reconstructing the models. And Zhang et al. [47] proposed a hierarchical gradient-based iterative parameter estimation algorithm for multivariable output error moving average systems. Chen and Chen [41] designed a weighted least squares (WLSs)-based adaptive tracker for a class of Hammerstein systems. Vanbeylen et al. [48] showed a method about the identification of discrete-time Hammerstein systems from output measurements only. Ding et al. [35] proposed a modified stochastic gradient-based parameter estimation algorithm for dual-rate sampled-data systems. Sun and Liu [49] proposed a novel APSO-aided maximum-likelihood identification method for Hammerstein systems. Furthermore, some new nonparametric identification methods are presented, such as the using of neural network to model the static nonlinear part [50,51] and identification without explicit

parameterization of nonlinearity driven by piecewise constant inputs [52].

Ding and Chen [23] presented an excellent recursive least squares (RLSs) method to deal with the Hammerstein models, which has good identification accuracy and convergence rate. And many researchers take Ding's method and models as a basis to compare with their methods [23,49,53]. However, Ding's [23] RLS approach is not applicable, when the noise model is unknown. And better identification accuracy and faster convergence rate still remain the targets both in theory and applications.

In this paper, the nonlinear recursive instrumental variables (RIVs) method is proposed to the identification of a classic Hammerstein model. Better identification results can be obtained in comparison with RLS method, and the method can be applied when the noise structure is unknown, which has shown that the proposed method is more flexible than RLS method. Furthermore, the property of the mean square convergence of the nonlinear RIV method is rigorously proved.

This paper is organized as follows: Sect. 2 describes the problem formulation based on Hammerstein models. Section 3 depicts the instrumental variable method for the Hammerstein models and three approaches to choose instrumental variables in linear RIV method. Section 4 presents the derivation of the nonlinear RIV method and proves its property of the mean square convergence, respectively. Section 5 illustrates the proposed approach with a classic model based on Hammerstein systems, where the nonlinear RIV method and the ARMAX-RLS method reported in the open literature [23] are compared in detail to show the effectiveness of the proposed method, and finally some conclusions from the above analysis are achieved in Sect. 6.

2 Problem statement

In a deterministic setting, the linear part of the system is characterized by a rational transfer function, and the system output $y(t)$ is exactly observed. However, in practice, the system itself may be random and the observations may be corrupted by noise. So, it is of practical importance to consider stochastic Hammerstein systems as shown in Fig. 1, which is composed of a nonlinear memoryless block $f(\cdot)$ followed by a linear subsystem. $u(t)$ is the system input, $y(t)$ is the system output and $v(t)$ is a white noise sequence, respectively.
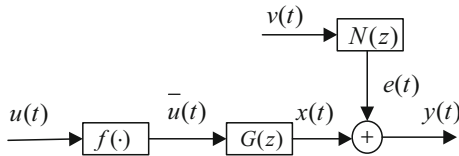
Fig. 1 The discrete-time SISO Hammerstein system

The true output $x(t)$, colored noise $e(t)$ and the inner variable $\bar{u}(t)$, which is the output of the nonlinear block, are unmeasurable. $N(z)$ is the transfer function of the noise model, and $G(z)$ is the transfer function of the linear part in the model [49].

The linear dynamical block in Fig. 1 is an ARMAX subsystem, the nonlinear model in Fig. 1 has the following input–output relationship [23,33]:

$$y(t) = x(t) + e(t) \tag{1}$$

$$x(t) = G(z)\bar{u}(t) = \frac{B(z)}{A(z)}\bar{u}(t) \tag{2}$$

$$e(t) = N(z)v(t) = \frac{D(z)}{A(z)}v(t) \tag{3}$$

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_{n_a} z^{-n_a} \tag{4}$$

$$B(z) = b_1 z^{-1} + b_2 z^{-2} + \cdots + b_{n_b} z^{-n_b} \tag{5}$$

$$D(z) = 1 + d_1 z^{-1} + d_2 z^{-2} + \cdots + d_{n_d} z^{-n_d} \tag{6}$$

This can be transformed as

$$y(t) = -\sum_{i=1}^{n_a} a_i y(t-i) + \sum_{i=1}^{n_b} b_i \bar{u}(t-i)$$
$$+ \sum_{i=1}^{n_d} d_i v(t-i) + v(t) \tag{7}$$

where $v(t)$ is a white noise sequence with the normal distribution $v(t) \sim N(0, \sigma_v^2)$ for the nonparametric $f(\cdot)$, the value $f(u)$ is estimated for any fixed u. In the parametric case, $f(\cdot)$ either is expressed by a linear combination of known basis functions with unknown coefficients, or is a piecewise linear function with unknown joints and slopes, and hence, identification of the nonlinear block in this case is equivalent to estimating unknown parameters. The nonlinear part is considered as a known basis $(\omega_1, \omega_2, \ldots, \omega_{n_c})$ with coefficients $(c_1, c_2, \ldots, c_{n_c})$ in this paper:

$$f(u(t)) = \bar{u}(t) = c_1 \omega_1(u(t)) + c_2 \omega_2(u(t))$$
$$+ \ldots + c_{n_c} \omega_{n_c}(u(t)) = \sum_{i=1}^{n_c} c_i \omega_i(u(t)) \tag{8}$$

Notice that the parameterization is actually not unique. In order to get a unique parameter estimate,

without loss of generality, one of the gains of $f(\cdot)$ must be fixed. Here, the first coefficient of the nonlinear function is assumed to equal 1, i. e., $c_1 = 1$ [12,54].

Substitute Eq. (8) into Eq. (7) gives

$$y(t) = -\sum_{i=1}^{n_a} a_i y(t-i) + \sum_{j=1}^{n_b} b_j \sum_{i=1}^{n_c} c_i \omega_i(u(t-j))$$
$$+ \sum_{i=1}^{n_d} d_i v(t-i) + v(t) \tag{9}$$

This paper is aimed at presenting the nonlinear RIV method to get the estimation of the unknown parameters $a_i(i = 1, 2, \ldots, n_a)$, $b_j(j = 1, 2, \ldots, n_b)$ and $c_k(k = 1, 2, \ldots, n_c)$ of the nonlinear ARMAX model by using the input and output data $\{u(t)\}$, $\{y(t)\}$ and to find the property of the algorithm.

3 Instrumental variables

Define the parameter vector, referring to [23,33], as

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{a} \\ c_1\boldsymbol{b} \\ c_2\boldsymbol{b} \\ \cdots \\ c_{n_c}\boldsymbol{b} \end{bmatrix} \in R^{n_0}, \quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_{n_a} \end{bmatrix} \in R^{n_a},$$

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_{n_b} \end{bmatrix} \in R^{n_b}, \quad e(t) = \sum_{i=1}^{n_d} d_i v(t-i) + v(t) \tag{10}$$

Equation (9) can be written as

$$y(t) = \mathbf{h}^T(t)\boldsymbol{\theta} + e(t) \tag{11}$$

or

$$\mathbf{Y}_L = \mathbf{H}_L \boldsymbol{\theta} + \mathbf{e}_L \tag{12}$$

where

$$\mathbf{h}(t) = \begin{bmatrix} \boldsymbol{\psi}_0(t) \\ \boldsymbol{\psi}_1(t) \\ \boldsymbol{\psi}_2(t) \\ \cdots \\ \boldsymbol{\psi}_{n_c}(t) \end{bmatrix} \in R^{n_0},$$

$$\boldsymbol{\psi}_0(t) = \begin{bmatrix} -y(t-1) \\ -y(t-2) \\ \cdots \\ -y(t-n_a) \end{bmatrix} \in R^{n_a},$$

$$\mathbf{e}_L = \begin{bmatrix} e^T(1) \\ e^T(2) \\ \cdots \\ e^T(L) \end{bmatrix}, \quad \mathbf{Y}_L = \begin{bmatrix} y(1) \\ y(2) \\ \cdots \\ y(L) \end{bmatrix},$$

$$\mathbf{H}_L = \begin{bmatrix} \mathbf{h}^T(1) \\ \mathbf{h}^T(2) \\ \cdots \\ \mathbf{h}^T(L) \end{bmatrix}, \tag{13}$$

$$\boldsymbol{\psi}_j(t) = \begin{bmatrix} \omega_j(u(t-1)) \\ \omega_j(u(t-2)) \\ \cdots \\ \omega_j(u(t-n_b)) \end{bmatrix} \in R^{n_b},$$

$$j = 1, 2, \ldots, n_c$$

The least squares estimation formula is as follows:

$$\hat{\boldsymbol{\theta}}_{\mathrm{LS}} = (\mathbf{H}_L^T \mathbf{H}_L)^{-1} \mathbf{H}_L \mathbf{Y}_L = \boldsymbol{\theta} + \left(\frac{1}{L} \mathbf{H}_L^T \mathbf{H}_L \right)^{-1} \left(\frac{1}{L} \mathbf{H}_L^T \mathbf{e}_L \right) \tag{14}$$

where

$$\begin{cases} \frac{1}{L} \mathbf{H}_L^T \mathbf{H}_L = \frac{1}{L} \sum_{t=1}^{L} \mathbf{h}(t)\mathbf{h}^T(t) \xrightarrow[L\to\infty]{W.P.1} E\left\{\mathbf{h}(t)\mathbf{h}^T(t)\right\} \\ \frac{1}{L} \mathbf{H}_L \mathbf{e}_L = \frac{1}{L} \sum_{t=1}^{L} \mathbf{h}(t)e(t) \xrightarrow[L\to\infty]{W.P.1} E\left\{\mathbf{h}(t)e(t)\right\} \end{cases} \tag{15}$$

The Frechet Theorem [55] is introduced here.

Theorem 1 *Assume that $\{x(t)\}$ is a sequence of random variables which converges to a constant x_0, then, there are*

$$f(x(t)) \xrightarrow[t\to\infty]{W.P.1} f(x_0) \tag{16}$$

or

$$P \lim_{t\to\infty} f(x(t)) = f(x_0) \tag{17}$$

where $f(\cdot)$ is continuous scalar function.

Theorem 2 *Assume that matrices A_t and B_t exist the probability limit, and the dimension of them does not change with the increase of t, apply Theorem 1, it gives*

$$\begin{cases} P \lim_{t\to\infty} (A_t B_t) = \left(P \lim_{t\to\infty} A_t \right) \left(P \lim_{t\to\infty} B_t \right) \\ P \lim_{t\to\infty} \left(A_t^{-1} \right) = \left(P \lim_{t\to\infty} A_t \right)^{-1} \end{cases} \tag{18}$$

Apply the two convergence theorems above, there are

$$\hat{\boldsymbol{\theta}}_{LS} \xrightarrow[L\to\infty]{W.P.1} \boldsymbol{\theta}_0 + [E\{\mathbf{h}(t)\mathbf{h}(t)\}]^{-1} E\{\mathbf{h}(t)e(t)\} \tag{19}$$

If $e(t)$ is white noise, $E\{\mathbf{h}(t)e(t)\} = \mathbf{0}$, thus $\hat{\boldsymbol{\theta}}_{LS} \xrightarrow[L\to\infty]{W.P.1} \boldsymbol{\theta}$, then the unbiased estimation can be obtained.

If $e(t)$ is not white noise, $E\{\mathbf{h}(t)e(t)\} \neq \mathbf{0}$. In order to get the unbiased estimation of the parameters, or to say $\hat{\boldsymbol{\theta}}_{LS} \xrightarrow[L\to\infty]{W.P.1} \boldsymbol{\theta}$ holds, an instrumental matrix is defined as

$$\mathbf{H}_L^* = \begin{bmatrix} \mathbf{h}^{*T}(1) \\ \mathbf{h}^{*T}(2) \\ \cdots \\ \mathbf{h}^{*T}(L) \end{bmatrix} \tag{20}$$

Two conditions are given as follows:

(a) $\frac{1}{L} \mathbf{H}_L^{*T} \mathbf{H}_L \xrightarrow[L\to\infty]{W.P.1} E\{\mathbf{h}^*(t)\mathbf{h}^T(t)\}$ is a nonsingular matrix;

(b) $\mathbf{h}^*(t)$ is independent of $e(t)$, which means $\frac{1}{L} \mathbf{H}_L^{*T} \mathbf{e}_L \xrightarrow[L\to\infty]{W.P.1} E\{\mathbf{h}^*(t)e(t)\} = \mathbf{0}$, where $\mathbf{h}^*(t)$ are the instrumental variables.

If the instrumental variables meet the two conditions above, it gives

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{IV} &= (\mathbf{H}_L^{*T} \mathbf{H}_L)^{-1} \mathbf{H}_L^{*T} \mathbf{Y}_L \\ &= \boldsymbol{\theta} + \left(\frac{1}{L} \mathbf{H}_L^{*T} \mathbf{H}_L \right)^{-1} \left(\frac{1}{L} \mathbf{H}_L^{*T} \mathbf{e}_L \right) \\ &\xrightarrow[L\to\infty]{W.P.1} \boldsymbol{\theta} + (E\{\mathbf{h}^*(t)\mathbf{h}^T(t)\})^{-1} E\{\mathbf{h}^*(t)e(t)\} = \boldsymbol{\theta} \end{aligned} \tag{21}$$

where $\hat{\boldsymbol{\theta}}_{IV}$ is the parameter estimation with instrumental variables method.

It can be seen that if the chosen instrumental variables are suitable to satisfy the two conditions above, then unbiased and consistent parameter estimation can be obtained. So, how to choose suitable instrumental variables is crucial to instrumental variables method, the following three approaches, usually used in the linear RIV method [56], are introduced first to choose instrumental variables.
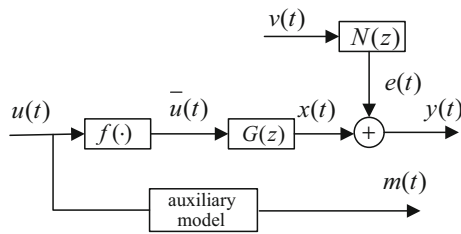
Fig. 2 The choice of the instrumental variables

$m(t)$ is noted as the instrumental variables in Fig. 2. $\mathbf{h}^*(t)$ is given as follows:

$$\mathbf{h}^*(t) = \begin{bmatrix} \boldsymbol{\psi}_0^*(t) \\ \boldsymbol{\psi}_1(t) \\ \boldsymbol{\psi}_2(t) \\ \cdots \\ \boldsymbol{\psi}_{n_c}(t) \end{bmatrix} \in R^{n_0},$$

$$\boldsymbol{\psi}_0^*(t) = \begin{bmatrix} -m(t-1) \\ -m(t-2) \\ \cdots \\ -m(t-n_a) \end{bmatrix} \in R^{n_a} \quad (22)$$

1. Adaptive filtering method

When $u(t)$ is persistent excitation signal, then $E\{\mathbf{h}^*(t)\mathbf{h}^T(t)\}$ is nonsingular. Since $m(t)$ is only related to $u(t)$, that is to say $\mathbf{h}^*(t)$ must be uncorrelated with the noise, hence $E\{\mathbf{h}^*(t)e(t)\} = 0$.

The instrumental model can be regarded as an adaptive filter. The instrumental variables can be obtained by the following method:

$$m(t) = \mathbf{h}^{*T}(t)\hat{\boldsymbol{\theta}}(t)$$

or

$$\begin{cases} m(t) = \mathbf{h}^{*T}(t)\bar{\boldsymbol{\theta}}(t) \\ \bar{\boldsymbol{\theta}}(t) = (1-\alpha)\bar{\boldsymbol{\theta}}(t-1) + \alpha\hat{\boldsymbol{\theta}}(t-d) \end{cases} \quad (23)$$

where $\alpha \in (0.01, 0.1)$, $d \in (0, 10)$, $\hat{\boldsymbol{\theta}}(t)$ is the parameter estimation at moment t with instrumental variable method, it can be calculated by recursive method. If the two conditions (a) and (b) are satisfied, then Eqs. (23) and (24) are equivalent.

2. Tally principle

If noise $e(t)$ can be regarded as following model:

$$e(t) = D\left(z^{-1}\right)v(t) \quad (24)$$

where $w(t)$ is uncorrelated stochastic noise with zero mean value and

$$D\left(z^{-1}\right) = 1 + d_1 z^{-1} + d_2 z^{-2} + \cdots + d_{n_d} z^{-n_d} \quad (25)$$

The instrumental variables can be chosen as

$$m(t) = y(t - n_d) \quad (26)$$

3. Pure lag method

The instrumental model can be regarded as pure lag segment. The instrumental variables can be chosen as follows:

$$m(t) = u(t - n_b) \quad (27)$$

where n_b is the order of polynomial $B(z^{-1})$. It is evident that if $u(t)$ is persistent excitation signal and is unrelated to $e(t)$, the instrumental variables will meet the two conditions above.

Detailed description about the three methods to choose instrumental variables can be found in [56].

4 The nonlinear recursive instrumental variables method

4.1 The nonlinear RIV algorithm

Following gives the derivation of the nonlinear RIV algorithm.

Replace $\mathbf{h}(t)$ by $\mathbf{h}^*(t)$ in Eq. (14) gives

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{\text{IV}} &= (\mathbf{H}_L^{*T}\mathbf{H}_L)^{-1}\mathbf{H}_L^{*T}\mathbf{Y}_L \\ &= \left[\sum_{i=1}^L \mathbf{h}^*(i)\mathbf{h}^T(i)\right]^{-1}\left[\sum_{i=1}^L \mathbf{h}^*(i)y(i)\right] \end{aligned} \quad (28)$$

Define

$$\begin{cases} \boldsymbol{P}^{-1}(t) = \sum_{i=1}^t \mathbf{h}^*(i)\mathbf{h}^T(i) \\ \boldsymbol{K}(t) = \boldsymbol{P}(t)\mathbf{h}^*(t) \end{cases} \quad (29)$$

Then,

$$\begin{aligned} \boldsymbol{P}^{-1}(t) &= \sum_{i=1}^{t-1} \mathbf{h}^*(i)\mathbf{h}^T(i) + \mathbf{h}^*(t)\mathbf{h}^T(t) \\ &= \boldsymbol{P}^{-1}(t-1) + \mathbf{h}^*(t)\mathbf{h}^T(t) \end{aligned} \quad (30)$$

$$P^{-1}(t-1)\hat{\theta}_{\mathrm{IV}}(t-1) = \sum_{i=1}^{t-1} \mathbf{h}^*(i)y(i) \tag{31}$$

Substitute Eqs. (30)–(31) into Eq. (28) gives

$$
\begin{aligned}
\hat{\theta}_{\mathrm{IV}}(t) &= P(t)\left[\sum_{i=1}^{t} \mathbf{h}^*(i)y(i)\right] \\
&= \hat{\theta}_{\mathrm{IV}}(t-1) + P(t)\mathbf{h}^*(t)[y(t) - \mathbf{h}^T(t)\hat{\theta}_{\mathrm{IV}}(t-1)].
\end{aligned}
\tag{32}
$$

Apply the inversion formula of matrix $(A+BC)^{-1} = A^{-1} - A^{-1}B(I + CA^{-1}B)^{-1}CA^{-1}$ to (30) yields

$$
\begin{aligned}
P(t) = \{I &- P(t-1)\mathbf{h}^*(t)\mathbf{h}^T(t)[1 \\
&+ \mathbf{h}^T(t)P(t-1)\mathbf{h}^*(t)]^{-1}\}P(t-1)
\end{aligned}
\tag{33}
$$

Substitute Eq. (33) into Eq. (29) gives

$$
\begin{aligned}
K(t) &= P(t)\mathbf{h}^*(t) \\
&= \left[1 + \mathbf{h}^T(t)P(t-1)\mathbf{h}^*(t)\right]^{-1} P(t-1)\mathbf{h}^*(t)
\end{aligned}
\tag{34}
$$

And finally the nonlinear RIV algorithm is obtained as follows:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t)[y(t) - \mathbf{h}^T(t)\hat{\theta}(t-1)] \tag{35}$$

$$K(t) = P(t-1)\mathbf{h}^*(t)[1 + \mathbf{h}^T(t)P(t-1)\mathbf{h}^*(t)]^{-1} \tag{36}$$

$$P(t) = [I - K(t)\mathbf{h}^T(t)]P(t-1) \tag{37}$$

where $\mathbf{h}^*(t)$ are the instrumental variables shown in Eq. (20)–(21), the choice of $\mathbf{h}^*(t)$ can use the three methods described above. If the instrumental variables are in adaptive filtering form, it needs least squares method to calculate a few steps to obtain the initial parameter estimation $\hat{\theta}(t)$ as the initial state of the nonlinear RIV method, to initialize the algorithm, p_0 is taken as a large positive real number, e.g., $p_0 = 10^6 I$, and $\hat{\theta}(0) = 10^{-6}I_{n_0 \times 1}$.

4.2 Mean square convergence of the nonlinear RIV method

Firstly, two lemmas [57] are introduced.

Lemma 1 *Assume the eigenvalues of matrix $A \in R^{n \times n}$ are $\lambda_i[A], i = 1, 2, \ldots, n$, then the eigenvalues of matrix $A + sI$ are $\lambda_i[A + sI] = \lambda_i[A] + s, i = 1, 2, \ldots, n, s$ is a constant*

Lemma 2 *Assume the eigenvalues of matrix $A \in R^{n \times n}$ are $\lambda_i[A], i = 1, 2, \ldots, n, \min_i\{\lambda_i[A]\} = \alpha$, then $A^T A \geq \alpha^2 I$, $(A + sI)^T(A + sI) \geq (\alpha - s)^2 I$, where $0 < s < \alpha$.*

Theorem 3 *Assume that $\{e(t)\}$ is a random noise vector sequence with zero mean and bounded variance, namely $E[||e(t)||^2] = \sigma_e^2(t) \leq \sigma^2 < \infty$; the input vectors $\{u(t)\}$ and the instrumental vectors $\{\mathbf{h}^*(t)\}$ are uncorrelated with $\{e(t)\}$ and the system meets the weak persistence of excitation, which is to ensure that the matrix $\frac{1}{t}\mathbf{H}_t^*\mathbf{H}_t$ is nonsingular, that is*

$$
\begin{aligned}
A_1 &: \min\left\{\left|\left|\lambda_i\left[\frac{1}{t}\mathbf{H}_t^*\mathbf{H}_t\right]\right|\right|\right\} \\
&= \min\left\{\left|\left|\lambda_i\left[\frac{1}{t}\sum_{i=1}^{t}\mathbf{h}^*(i)\mathbf{h}(i)\right]\right|\right|\right\} \geq \alpha > 0, \ a.s.
\end{aligned}
$$

$$A_2 : \frac{1}{t}\sum_{i=1}^{t}\mathbf{h}^*(i)\mathbf{h}(i) = \frac{1}{t}\mathbf{H}_t^*\mathbf{H}_t \leq \beta I < \infty, \ a.s.$$

Define $||X||^2 = tr[XX^T]$.

Assume $E[||\hat{\theta}(0) - \theta||^2] \leq M_0 < \infty$, and $\hat{\theta}(0)$ is uncorrelated with $\{e(t)\}$, then the parameter estimation error using the nonlinear RIV method converges to zero at the rate of $O(\frac{1}{\sqrt{t}})$, that is

$$
\begin{aligned}
E[||\hat{\theta}(t) - \hat{\theta}||^2] &\leq 2\left[\frac{||P^{-1}(0)||^2 M_0}{(\alpha t - a)^2} + \frac{n\beta\sigma^2 t}{(\alpha t - a)^2}\right] \\
&\triangleq f(t) \ \text{or} \ \lim_{t \to \infty} E[||\hat{\theta}(t) - \theta||]^2 = 0
\end{aligned}
\tag{38}
$$

where n is the rank of matrix $\mathbf{H}_t^*P^T(t)P(t)\mathbf{H}_t^{*T}$, β is the largest eigenvalue of matrix $\mathbf{H}_t^*P^T(t)P(t)\mathbf{H}_t^{*T}$, $P(0) = P_0 = \frac{1}{a}I, 0 < a < 1$, say $a = 10^{-6}$.

Refer to [58,59], the following gives the proof of Theorem 3.

Define

$$\tilde{\theta}(t) = \hat{\theta}(t) - \theta \tag{39}$$

Combine Eqs. (30) and (35) gives

$$\tilde{\theta}(t) = [I - P(t)\mathbf{h}^*(t)\mathbf{h}^T(t)]\tilde{\theta}(t-1) + P(t)\mathbf{h}^*(t)e(t) \tag{40}$$

Multiplying both sides of the Eq. (30) by $P^{-1}(t)$ gives

$$I - P(t)\mathbf{h}^*(t)\mathbf{h}^T(t) = P(t)P^{-1}(t-1) \tag{41}$$

Substituting Eq. (41) in Eq. (40) gives

$$\tilde{\boldsymbol{\theta}}(t) = \boldsymbol{P}(t)\boldsymbol{P}^{-1}(0)\tilde{\boldsymbol{\theta}}(0) + \boldsymbol{P}(t)\mathbf{H}_t^{*T} e_t = \gamma_1(t) + \gamma_2(t) \tag{42}$$

where $\mathbf{H}_t^{*T} e_t = \sum_{i=1}^{t} \mathbf{h}^*(t)e^T(t)$, $\boldsymbol{P}^{-1}(t) = \mathbf{H}_t^*\mathbf{H}_t + \boldsymbol{P}^{-1}(0)$, $\gamma_1(t) = \boldsymbol{P}(t)\boldsymbol{P}^{-1}(0)\tilde{\boldsymbol{\theta}}(0)$, $\gamma_2(t) = \boldsymbol{P}(t)\mathbf{H}_t^{*T} e_t$.

Apply Lemmas 1 and 2, then for any $t \to \infty$, it gives $\boldsymbol{P}^{-T}(t) = (\mathbf{H}_t^*\mathbf{H}_t)^T + \boldsymbol{P}^{-1}(0)\boldsymbol{I}$.

Define

$$\frac{1}{t}\mathbf{H}_t^*\mathbf{H}_t = A, \ \boldsymbol{P}^{-1}(0) = s\boldsymbol{I}, \quad \min_i \left\{ |\lambda_i[\mathbf{H}_t^*\mathbf{H}_t^T]| \right\} = \alpha \tag{43}$$

Then,

$$\min_i \left\{ \left| \lambda_i \left[\frac{1}{t}\mathbf{H}_t^*\mathbf{H}_t^T \right] \right| \right\} = \alpha \tag{44}$$

$$\left[\frac{1}{t}\boldsymbol{P}^{-T}(t) \right]^T \left[\frac{1}{t}\boldsymbol{P}^{-T}(t) \right] = (A + s\boldsymbol{I})^T(A + s\boldsymbol{I})$$

$$\geq \left(\alpha - \frac{1}{t}a \right)^2$$

$$\boldsymbol{P}^T(t)\boldsymbol{P}(t) \leq \frac{1}{(\alpha t - a)^2} \tag{45}$$

Thus, it gives

$$0 \leq E[||\gamma_1(t)||^2]$$
$$= E\left[\tilde{\boldsymbol{\theta}}^T(0)\boldsymbol{P}^{-T}(0)\boldsymbol{P}^T(t)\boldsymbol{P}(t)\boldsymbol{P}^{-1}(0)\tilde{\boldsymbol{\theta}}(0) \right]$$
$$\leq \frac{||\boldsymbol{P}^{-1}(0)||^2 M_0}{(\alpha t - a)^2} \tag{46}$$

$$0 \leq E[||\gamma_2(t)||^2] = E\left[e_t^T \mathbf{H}_t^* \boldsymbol{P}^T(t)\boldsymbol{P}(t)\mathbf{H}_t^{*T} e_t \right]$$
$$\leq \frac{n\beta\sigma^2 t}{(\alpha t - a)^2} \tag{47}$$

Substitute Eq. (46)–(47) into Eq. (42)

$$0 \leq E||\tilde{\boldsymbol{\theta}}(t)||^2 = \mathrm{E}||\gamma_1(t) + \gamma_2(t)||^2 \leq 2E\left[||\gamma_1(t)||^2 \right.$$
$$\left. + ||\gamma_2(t)||^2 \right]$$
$$= 2\left(\frac{||\boldsymbol{P}^{-1}(0)||^2 M_0}{(\alpha t - a)^2} + \frac{n\beta\sigma^2 t}{(\alpha t - a)^2} \right)$$
$$\text{or} \lim_{t \to \infty} E[||\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}||]^2 = 0 \tag{48}$$

Then the mean square convergence of the nonlinear RIV method is completely proved.

From what is discussed above, it can be seen that $e(t)$ is colored noise, but as long as the system is persistently excited, and the noise $e(t)$ is zero mean and bounded variance and uncorrelated with $\hat{\boldsymbol{\theta}}(0)$, which means condition A_1 and A_2 are satisfied, then the nonlinear RIV method has the property of mean square convergence for identification of Hammerstein models, that is to say the parameter estimation error $\tilde{\boldsymbol{\theta}}(t)$ using the proposed method converges to zero at the rate of $O(\frac{1}{\sqrt{t}})$, which guarantees that the nonlinear RIV method has good capability against colored noise.

5 Example

Due to the commonly recognized effectiveness of Ding's RLS algorithm, Ding's example [23] is hereby taken as the model to demonstrate the improved identification performance of the new algorithm. This is a Hammerstein ARMAX system as follows:

$$A(z)y(t) = B(z)\bar{u}(t) + D(z)v(t) \tag{49}$$
$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} = 1 - 1.60z^{-1}$$
$$+ 0.80z^{-2} \tag{50}$$
$$B(z) = b_1 z^{-1} + b_2 z^{-2} = 0.85z^{-1} + 0.65z^{-2} \tag{51}$$
$$D(z) = 1 + d_1 z^{-1} = 1 - 0.64z^{-1} \tag{52}$$
$$\bar{u}(t) = f(u(t)) = c_1 u(t) + c_2 u^2(t) + c_3 u^3(t)$$
$$= u(t) + 0.5u^2(t) + 0.25u^3(t) \tag{53}$$

$\boldsymbol{\theta} = [a_1, a_2, b_1, b_2, c_2, c_3] = [-1.60, 0.8, 0.85, 0.65, 0.5, 0, 25]$ are the parameters to be identified. $\{u(t)\}$ is taken as a persistent excitation signal sequence with zero mean and unit variance, and $\{v(t)\}$ as a white noise sequence with zero mean and constant variance σ_v^2. The noise-to-signal-ratio (NSR) is defined by the standard deviation of the ratio of input-free output and noise-free output, namely NSR $= \sqrt{\frac{\mathrm{var}[v(t)]}{\mathrm{var}[u(t)]}} \times 100\%$.

When $\sigma_v^2 = 0.3^2$, $\sigma_v^2 = 0.5^2$ and $\sigma_v^2 = 0.7^2$, the corresponding NSRs are 16.34, 25.45 and 35.70%, respectively. The accuracy of identification of the proposed models is assessed by comparing overall output response of estimated model and the true output, and also the relative parameter estimation error, which is

$$\delta = ||\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}||/||\boldsymbol{\theta}|| \times 100\% \tag{54}$$

For the model discussed above, the nonlinear RIV method is chosen to be compared with ARMAX-RLS

[23] under three NSRs mentioned above and sampling data length 3,000 to get the results. The comparison of the relative parameter estimation errors between real parameters and identified parameters of the nonlinear RIV and ARMAX-RLS methods with NSR 16.34, 25.45 and 37.50 % is shown in Figs. 3, 4 and 5 (dashed line for the errors by RLS method and solid line for the errors by the nonlinear RIV method), respectively. And Tables 1, 3 and 5 are the records of the parameter estimation and the related relative error of nonlinear RIV method with NSR 16.34, 25.45 and 37.50 %, respectively. And Tables 2, 4 and 6 keep record of the parameter estimation and the related relative error of ARMAX-RLS method with NSR 16.34, 25.45 and 37.50 %, respectively.

Figure 3 shows the changing process of relative error between nonlinear RIV method and ARMAX-RLS method when NSR is 16.34 %. And the relative error is obtained through Eq. (54). When the sampling time is between 0 and 400, the identification error of ARMAX-RLS method is smaller than that of the non-linear RIV method. However, the identification error of nonlinear RIV method shows decline trend and is lower than that of the ARMAX-RLS method afterward. This has shown that the advantage of the proposed method may not be obvious at first, but with the increase of the sampling time, the advantage becomes more and more obvious.

Tables 1 and 2 have shown that the concrete values of the parameters with 16.34 %. $\theta_1, \theta_2, \theta_3, \theta_4$, which can be directly obtained through Eqs. (35)–
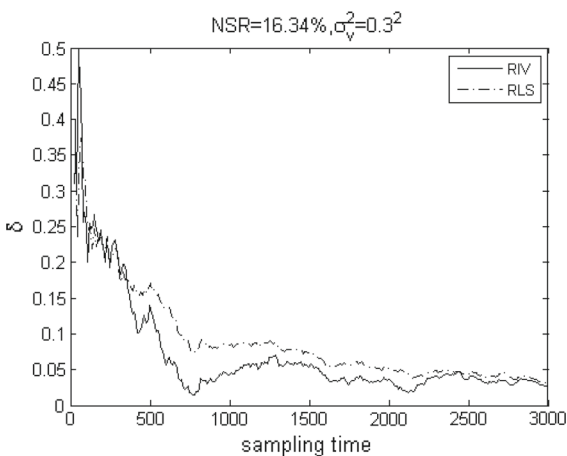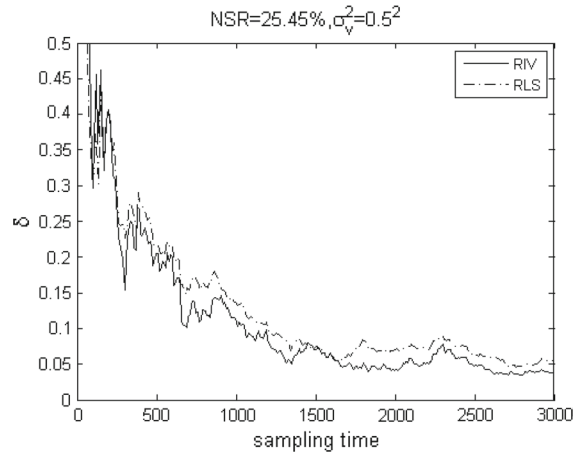
Fig. 4 Comparison of relative error between the nonlinear RIV method and ARMAX-RLS method (NSR = 25.45 %)
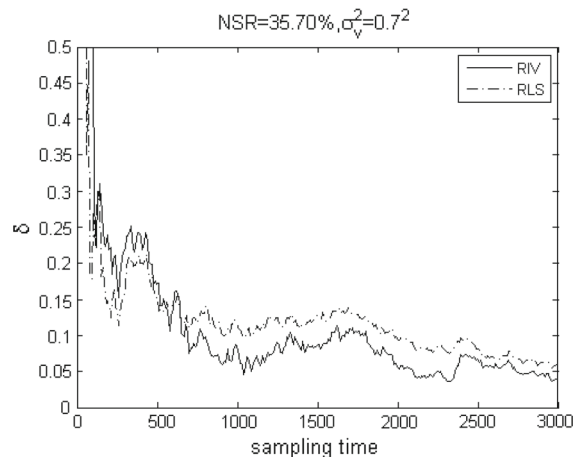
Fig. 5 Comparison of relative error between the nonlinear RIV method and ARMAX-RLS method (NSR = 35.70 %)

(37), represent the parameter a_1, a_2, b_1, b_2, respectively. $\theta_5, \theta_6, \theta_7, \theta_8$ are not the parameters of the real system but obtained by linear transform and they represent $b_1c_2, b_2c_2, b_1c_3, b_2c_3$, respectively. And c_2, c_3 can be obtained by the following method:

$$c_2 = (\theta_5 + \theta_6)/(\theta_3 + \theta_4)$$
$$c_3 = (\theta_7 + \theta_8)/(\theta_3 + \theta_4) \qquad (55)$$

After iteration of 3,000, the final error of the nonlinear RIV method is 2.54 %, the final error of ARMAX-RLS method is 3.32 %.

Fig. 3 Comparison of relative error between the nonlinear RIV method and ARMAX-RLS method (NSR = 16.34 %)

Table 1 Identification results of the nonlinear RIV method (NSR = 16.34 %)

Iteration	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	δ
100	−1.5853	0.7788	1.0862	0.4640	0.4789	0.2191	0.0550	0.4460	0.2115
200	−1.5869	0.7890	1.1467	0.4666	0.4668	0.2927	−0.0652	0.3518	0.2248
500	−1.5984	0.7968	0.9744	0.5028	0.4572	0.2795	0.0741	0.3441	0.1402
1,000	−1.5960	0.7968	0.8771	0.5956	0.4166	0.3560	0.2141	0.2307	0.0447
1,500	−1.5985	0.7982	0.9137	0.5964	0.4174	0.3443	0.1613	0.2263	0.0547
2,000	−1.5978	0.7985	0.9037	0.6342	0.4308	0.3341	0.1610	0.1935	0.0382
2,500	−1.5987	0.7998	0.8967	0.6370	0.4290	0.3405	0.1617	0.1903	0.0356
3,000	−1.5971	0.7987	0.8835	0.6613	0.4241	0.3445	0.1755	0.1681	0.0254
Real value	−1.6000	0.8000	0.8500	0.6500	0.4250	0.3250	0.2125	0.1625	0

Table 2 Identification results of the ARMAX-RLS method (NSR = 16.34%)

Iteration	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	δ
100	−1.5772	0.7754	1.1378	0.4437	0.5121	0.2195	−0.1141	0.4544	0.2677
200	−1.5859	0.7910	1.0962	0.4256	0.4601	0.2894	−0.0408	0.3940	0.2219
500	−1.5952	0.7908	0.9684	0.4342	0.4352	0.2778	0.0805	0.4030	0.1715
1,000	−1.6045	0.8025	0.9084	0.5354	0.4040	0.3534	0.1620	0.2807	0.0855
1,500	−1.6055	0.8025	0.9214	0.5631	0.3967	0.3422	0.1461	0.2476	0.0735
2,000	−1.6068	0.8057	0.9078	0.5993	0.4079	0.3391	0.1483	0.2091	0.0521
2,500	−1.6050	0.8055	0.8889	0.6005	0.4077	0.3479	0.1680	0.2127	0.0446
3,000	−1.6036	0.8050	0.8821	0.6249	0.4057	0.3496	0.1705	0.1899	0.0332
Real value	−1.6000	0.8000	0.8500	0.6500	0.4250	0.3250	0.2125	0.1625	0

The final parameter identification result is:

$$\hat{\theta}_{RIV} = [-1.5971 \quad 0.7987 \quad 0.8835 \quad 0.6613 \\ 0.4975 \quad 0.2224]$$

$$\hat{\theta}_{ARMAX\text{-}RLS} = [-1.6036 \quad 0.8050 \quad 0.8821 \quad 0.6249 \\ 0.5012 \quad 0.2392]$$

Figure 4 depicts the changing process of relative error between predicted output and real output by the nonlinear RIV method and ARMAX-RLS method with 25.45 % NSR. From Fig. 4, it can be seen that the advantage of the nonlinear RIV method is still significant and the error curve of this method is sharply declining when the sampling time is about 300 and keeps under the error curve of the ARMAX-RLS method afterward.

Tables 3 and 4 have shown that the concrete values of the parameters with 25.45 %. After iteration of 3,000, the final error of the nonlinear RIV method is 3.73 %, the final error of ARMAX-RLS method is 5.51 %.

Based on Eq. (55), the final parameter identification result is:

$$\hat{\theta}_{RIV} = [-1.6030 \quad 0.8022 \quad 0.8957 \quad 0.6428 \\ 0.4920 \quad 0.2302]$$

$$\hat{\theta}_{ARMAX\text{-}RLS} = [-1.5965 \quad 0.7973 \quad 0.8901 \\ 0.6126 \quad 0.5140 \quad 0.2891].$$

Figure 5 depicts the changing process of relative error between predicted output and real output by nonlinear RIV method and ARMAX-RLS method when NSR is 35.70 %. From Fig. 5, it can be shown that the advantage of the proposed method is not significant at first. The overlapping of the two curves when the sampling time is between 0 and 700 indicates that the two methods show their advantages at different sampling times, and when the sampling time is over 700, the error curve of the nonlinear RIV method keeps under the error curve of the ARMAX-RLS method all through. The identification results of the nonlinear RIV method

Table 3 Identification results of the nonlinear RIV method (NSR = 25.45 %)

Iteration	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	δ
100	−1.5854	0.7821	1.0437	0.4610	0.3247	0.3127	−0.1957	0.5875	0.3028
200	−1.5937	0.7949	1.0415	0.3731	0.2997	0.4466	−0.2891	0.7940	0.4110
500	−1.5987	0.7983	0.9910	0.5779	0.3713	0.3464	−0.1355	0.3744	0.2033
1,000	−1.6037	0.8013	0.9588	0.5833	0.3613	0.3657	0.0386	0.2560	0.1139
1,500	−1.6046	0.8029	0.9262	0.6640	0.3537	0.3737	0.1068	0.1356	0.0735
2,000	−1.6039	0.8016	0.8965	0.6433	0.3686	0.3652	0.1842	0.1723	0.0410
2,500	−1.6055	0.8036	0.9104	0.6325	0.3646	0.3745	0.1859	0.1656	0.0479
3,000	−1.6030	0.8022	0.8957	0.6428	0.3829	0.3740	0.2007	0.1534	0.0373
Real value	−1.6000	0.8000	0.8500	0.6500	0.4250	0.3250	0.2125	0.1625	0.0000

Table 4 Identification results of the ARMAX-RLS method (NSR = 25.45%)

Iteration	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	δ
100	−1.5744	0.7728	1.1382	0.4891	0.3474	0.3121	−0.2512	0.4822	0.3037
200	−1.5911	0.7924	1.0572	0.3439	0.2878	0.4670	−0.2634	0.7349	0.3939
500	−1.5958	0.7970	1.0070	0.5399	0.3447	0.3668	−0.1302	0.4059	0.2171
1,000	−1.5958	0.7938	0.9364	0.5355	0.3787	0.3773	0.0813	0.3713	0.1355
1,500	−1.6001	0.7991	0.9219	0.6232	0.3848	0.3692	0.1168	0.2303	0.0703
2,000	−1.5975	0.7953	0.8956	0.6022	0.3768	0.3651	0.1800	0.2762	0.0689
2,500	−1.5975	0.7967	0.9157	0.6140	0.3802	0.3770	0.1660	0.2418	0.0632
3,000	−1.5965	0.7973	0.8901	0.6126	0.3840	0.3884	0.1987	0.2357	0.0551
Real value	−1.6000	0.8000	0.8500	0.6500	0.4250	0.3250	0.2125	0.1625	0.0000

are all better than those of ARMAX-RLS method with the three different NSRs.

Tables 5 and 6 have shown that the concrete values of the parameters with 35.70 %. After iteration of 3,000, the final error of the nonlinear RIV method is 4.25 %, the final error of ARMAX-RLS method is 6.17 %.

The final parameter identification result is:

$$\hat{\theta}_{RIV} = [-1.5948 \quad 0.7951 \quad 0.9036 \quad 0.6498$$
$$0.4918 \quad 0.1995]$$

$$\hat{\theta}_{ARMAX\text{-}RLS} = [-1.5943 \quad 0.7961 \quad 0.9224$$
$$0.6530 \quad 0.4865 \quad 0.1907].$$

From the identification results above, the following conclusions can be obtained:

(1) The proposed nonlinear RIV method for Hammerstein system identification is effective, and the identification errors with three NSRs are lower than that of ARMAX-RLS method after sampling time about 400, 300 and 700, respectively. From Figs. 3, 4 and 5, we get the identification error curve of the

nonlinear RIV method is basically under that of ARMAX-RLS method in the whole identification course. When NSR is 16.34 %, the identification error of the nonlinear RIV method is lower than that of ARMAX-RLS by 30.7 % after 3,000 iteration, when NSR is 25.45 %, the identification error of the proposed method is lower than that of ARMAX-RLS by 47.7 %. When NSR is 35.70 %, the identification error of the nonlinear RIV is lower than that of ARMAX-RLS by 45.2 %. So, the larger NSR is, the better advantage of the proposed method over ARMAX-RLS method is.

(2) With the increase of the NSR, the deviation of identification error of predicted output with the real output is enlarging, which means the proposed method is applicable with small noise signals. When the noise signal is getting larger, the identification results are getting worse or the system may not be identified. But the nonlinear RIV method is superior to ARMAX-RLS method with noise interference.

Table 5 Identification results of the nonlinear RIV method (NSR = 35.70%)

Iteration	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	δ
100	−1.5912	0.7870	0.3941	0.6365	0.6252	0.2422	0.5913	0.3240	0.3005
200	−1.5729	0.7732	0.5995	0.5908	0.6146	0.1707	0.5140	0.3085	0.2258
500	−1.5852	0.7794	0.9016	0.5027	0.5775	0.1451	0.1426	0.3837	0.1691
1,000	−1.6001	0.7981	0.8919	0.5550	0.4334	0.2894	0.1499	0.2925	0.0837
1,500	−1.5969	0.7956	0.9112	0.6129	0.4346	0.2872	0.0910	0.2808	0.0868
2,000	−1.6011	0.7997	0.9183	0.6466	0.4400	0.2975	0.1093	0.1679	0.0589
2,500	−1.5967	0.7960	0.9388	0.6571	0.4480	0.3079	0.1012	0.1348	0.0683
3,000	−1.5948	0.7951	0.9036	0.6498	0.4465	0.3175	0.1417	0.1682	0.0425
Real value	−1.6000	0.8000	0.8500	0.6500	0.4250	0.3250	0.2125	0.1625	0.0000

Table 6 Identification results of the ARMAX-RLS method (NSR = 35.70%)

Iteration	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	δ
100	−1.4886	0.7026	0.7978	0.9050	0.6864	0.2796	0.0112	0.0068	0.2186
200	−1.5390	0.7422	0.7757	0.7344	0.5985	0.1417	0.2374	0.2153	0.1359
500	−1.5812	0.7716	0.9352	0.6231	0.6040	0.1269	0.0740	0.2655	0.1531
1,000	−1.6086	0.8067	0.9499	0.6152	0.5213	0.2216	0.0511	0.2229	0.1139
1,500	−1.5933	0.7934	0.9764	0.6687	0.4822	0.2343	−0.0017	0.2100	0.1272
2,000	−1.5990	0.7975	0.9571	0.6798	0.4720	0.2588	0.0565	0.1450	0.0963
2,500	−1.5946	0.7953	0.9508	0.6818	0.4705	0.2860	0.0824	0.1305	0.0835
3,000	−1.5943	0.7961	0.9224	0.6530	0.4688	0.2977	0.1155	0.1850	0.0617
Real value	−1.6000	0.8000	0.8500	0.6500	0.4250	0.3250	0.2125	0.1625	0.0000

(3) As is shown in the error curves, at the beginning of identification, the parameter identification results converge very quickly; the cutoff point appears almost at 1,000 iterations. After 1,000 iterations, with the increase of iteration number, the speed of convergence is getting slower, and there exists some steady-state error between identification parameters and real parameters under different NSRs.

6 Conclusions

A nonlinear recursive instrumental variable method in identification of nonlinear system is obtained. It is applied to Hammerstein ARMAX model under different NSRs and is compared with ARMAX-RLS method in detail. The results show that the nonlinear RIV method is not only an effective method but also superior to ARMAX-RLS method in terms of identification accuracy and convergence speed especially under colored noise. ARMAX-RLS method is not applicable when noise model is unknown, while the proposed method is far more appropriate. In other words, the nonlinear RIV method is more flexible in identification of nonlinear systems with colored noise.

With the increase of the NSR, the deviation of identification error of predicted output with the real output is enlarging, which means the proposed method is applicable with small noise signals. When the noise signal is getting larger, the identification results are getting worse or the system may not be identified. But the nonlinear RIV method is superior to ARMAX-RLS method with noise interference.

The procedure of the proposed method and its mean square convergence for identification of Hammerstein models are also established, and its convergence analysis is worth further research.

References

1. Perc, M.: Nonlinear time series analysis of the human electrocardiogram. Eur. J. Phys. **26**, 757–768 (2005)
2. Perc, M.: The dynamics of human gait. Eur. J. Phys. **26**, 525–534 (2005)
3. Kodba, S., Perc, M., Marhl, M.: Detecting chaos from a time series. Eur. J. Phys. **26**, 205–215 (2005)
4. Gao, J., Hu, J., Mao, X., Perc, M.: Culturomics meets random fractal theory: insights into long-range correlations of social and natural phenomena over the past two centuries. J. R. Soc. Interface **9**, 1956–1964 (2012)
5. Perc, M.: Self-organization of progress across the century of physics. Sci. Rep. **3**, 1720 (2013)
6. Li, J.H., Ding, F., Hua, L.: Maximum likelihood Newton recursive and the Newton iterative estimation algorithms for Hammerstein CARAR systems. Nonlinear Dyn. **75**, 235–245 (2014)
7. Westwick, D., Verhaegen, M.: Identifying MIMO Wiener systems using subspace model identification methods. Sig. Process. **52**, 235–258 (1999)
8. Shi, Y., Fang, H.: Kalman filter based identification for systems with randomly missing measurements in a network environment. Int. J. Control **83**, 538–551 (2010)
9. MacArthur, J.W.: A new approach for nonlinear process identification using orthonormal bases and ordinal splines. J. Process Contr. **22**, 375–389 (2012)
10. Ding, F., Liu, X.P., Liu, G.J.: Identification methods for Hammerstein nonlinear systems. Digit. Signal Proc. **21**, 215–238 (2012)
11. Liu, X.G., Chen, L., Hu, Y.Q.: Solution of chemical dynamic optimization using the simultaneous strategies. Chin. J. Chem. Eng. **21**, 55–63 (2013)
12. Bai, E.W.: A blind approach to the Hammerstein–Wiener model identification. Automatica **38**, 967–979 (2002)
13. Michalkiewicz, J.: Modified kolmogorov's neural network in the identification of Hammerstein and Wiener systems. IEEE Trans. Neural Netw. Learn. **23**, 657–662 (2012)
14. Cerone, V., Piga, D., Regruto, D.: Bounded error identification of Hammerstein systems through sparse polynomial optimization. Automatica **48**, 2693–2698 (2012)
15. Gallman, P.G.: A comparison of two Hammerstein model identification algorithms. IEEE Trans. Autom. Control **21**, 124–126 (1976)
16. Li, G.Q., Wen, C.Y.: Convergence of normalized iterative identification of Hammerstein systems. Syst. Control Lett. **60**, 919–935 (2011)
17. Liu, X.G., Qian, J.: Modeling, control and optimization of ideal internal thermally coupled distillation columns. Chem. Eng. Technol. **23**, 235–241 (2000)
18. Liu, Y.J., Sheng, J., Ding, R.F.: Convergence of stochastic gradient estimation algorithm for multivariable ARX-like systems. Comput. Math. Appl. **59**, 2615–2627 (2010)
19. Maarten, S., Rik, P., Yves, R.: Parametric identification of parallel Hammerstein systems. IEEE Trans. Instrum. Meas. **60**, 3931–3938 (2011)
20. Vazquez Feijoo, J.A., Worden, K., Stanway, R.: System identification using associated linear equations. Mech. Syst. Signal Process. **18**, 431–455 (2004)
21. Chang, F., Luns, R.: A noniterative method for identification using Hammerstein model. IEEE Trans. Autom. Control **16**, 464–468 (1971)
22. Greblicki, W.: Recursive identification of continuous-time Hammerstein systems. Int. J. Syst. Sci. **33**, 969–977 (2002)
23. Ding, F., Chen, T.W.: Identification of Hammerstein nonlinear ARMAX systems. Automatica **41**, 1479–1489 (2005)
24. Ding, F., Shi, Y., Chen, T.W.: Auxiliary model based least-square identification methods for Hammerstein output error systems. Syst. Control Lett. **56**, 373–380 (2007)
25. Crama, P., Schoukens, J.: Initial estimates of Wiener and Hammerstein systems using multisine excitation. Syst. Control Lett. **50**, 1791–1795 (2001)
26. Greblicki, W.: Stochastic approximation in nonparametric identification of Hammerstein systems. IEEE Trans. Automat. Control **47**, 1800–1810 (2002)
27. Greblicki, W., Pawlak, M.: Nonparametric System Identification. Cambridge University Press, Cambridge (2008)
28. Hasiewicz, Z., Mzyk, G.: Hammerstein system identification by non-parametric instrumental variables. Int. J. Control **82**, 440–455 (2009)
29. Narendra, K., Gallman, P.: An iterative method for the identification of nonlinear systems using a Hammerstein model. IEEE Trans. Automat. Control **11**, 546–550 (1966)
30. Vörös, J.: Iterative algorithm for parameter identification of Hammerstein systems with two-segment nonlinearities. IEEE Trans. Automat. Control **44**, 2145–2149 (1999)
31. Liu, Y., Bai, E.W.: Iterative identification of Hammerstein systems. Automatica **43**, 346–354 (2007)
32. Liu, X.G., Bai, E.W.: Iterative identification of Hammerstein systems. Automatica **46**, 549–554 (2010)
33. Ding, F., Shi, Y., Chen, T.W.: Gradient-based identification methods for Hammerstein nonlinear ARMAX models. Nonlinear Dyn. **45**, 31–43 (2005)
34. Wang, D.Q., Ding, F.: Extended stochastic gradient identification algorithms for Hammerstein–Wiener ARMAX systems. Comput. Math. Appl. **56**, 3157–3164 (2008)
35. Ding, J., Shi, Y., Wang, H., Ding, F.: A modified stochastic gradient based parameter estimation algorithm for dual-rate sampled-data systems. Digit. Signal Proc. **20**, 1238–1247 (2010)
36. Greblicki, W., Pawlak, M.: Identification of discrete Hammerstein systems using kernel regression estimates. IEEE Trans. Automat. Control **31**, 74–77 (1986)
37. Li, G., Wen, C., Zheng, W., Chen, Y.: Identification of a class of nonlinear autoregressive with exogenous inputs models based on kernel machines. IEEE Trans. Signal Process. **59**, 2146–2158 (2011)
38. Huang, B., Ding, S.X., Qin, S.J.: Closed-loop subspace identification: an orthogonal projection approach. J. Process Control **15**, 53–66 (2005)
39. Vörös, J.: Recursive identification of Hammerstein systems with discontinuous nonlinearities containing dead-zones. IEEE Trans Autom. Control **48**, 2203–2206 (2003)

40. Zhao, W.X., Chen, H.F.: Adaptive tracking and recursive identification for Hammerstein systems. Automatica **45**, 2773–2783 (2009)
41. Chen, X.M., Chen, H.M.: Recursive identification for MIMO Hammerstein systems. IEEE Trans. Autom. Control **56**, 895–902 (2011)
42. Söderström, T., Ljung, L., Gustavsson, I.: A theoretical analysis of recursive identification methods. Automatica **14**, 231–244 (1978)
43. Ding, F., Deng, K.P., Liu, X.M.: Decomposition based Newton iterative identification method for a Hammerstein nonlinear FIR system with ARMA noise. Circ. Syst. Signal Proc. **33**, 2881–2893 (2014)
44. Li, J.H.: Parameter estimation for Hammerstein CARARMA systems based on the Newton iteration. Appl. Math. Lett. **26**, 91–96 (2013)
45. Hong, X., Mitchell, R.J.: Hammerstein model identification algorithm using Bezier–Bernstein approximation. IET Control Theory Appl. **1**, 1149–1159 (2007)
46. Han, Y., Raymond, A.: Hammerstein system identification using nuclear norm minimization. Automatica **48**, 2189–2193 (2012)
47. Zhang, Z.N., Ding, F., Liu, X.G.: Hierarchical gradient based iterative parameter estimation algorithm for multivariable output error moving average systems. Comput. Math. Appl. **61**, 672–682 (2011)
48. Vanbeylen, L., Pintelon, R., Schoukens, J.: Blind maximum likelihood identification of Hammerstein systems. Automatica **44**, 3139–3146 (2008)
49. Sun, J.L., Liu, X.G.: A novel APSO-aided maximum likelihood identification method for Hammerstein systems. Nonlinear Dyn. **73**, 449–462 (2013)
50. LL-Duwaish, H., Karim, M.: A new method for the identification of Hammerstein model. Automatica **33**, 1871–1875 (1997)
51. Lang, Z.: A nonparametric polynomial identification algorithm for the Hammerstein system. IEEE Trans. Autom. Control **42**, 1435–1441 (1997)
52. Wang, J.D., Sano, A., Chen, T.W., Huang, B.: Identification of Hammerstein systems without explicit parameterization of non-linearity. Int. J. Control **82**, 937–952 (2009)
53. Wang, D.Q., Ding, F., Chu, Y.Y.: Data filtering based recursive least squares algorithm for Hammerstein systems using the key-term separation principle. Inform. Sciences. **222**, 203–212 (2013)
54. Gallman, P.G.: A comparison of two Hammerstein model identification algorithms. IEEE Trans. Autom. Control **21**, 124–126 (1976)
55. Golfberger, A.S.: Econometric Theory. Wiley, New York (1964)
56. Fang, C.Z.: Process Identification. Tsinghua University Press, Beijing (1988)
57. Ding, F.: System Identification-New Theory and Methods. Science Press, Beijing (2013)
58. Ding, F., Chen, T.W.: Performance bounds of forgetting factor least squares algorithm for time-varying systems with finite measurement data. IEEE Trans. Circuits **52**, 555–566 (2005)
59. Ding, F., Chen, T.W.: Parameter estimation for dual-rate system with finite measurement data. Dyn. Cont. Dis. Ser. B **11**, 101–121 (2004)