Research Article

Data-driven Learning Algorithm of Neural Fuzzy Based Hammerstein-Wiener System

Feng Li ¹, Yinsheng Luo,¹ Naibao He,¹ Ya Gu,² and Qingfeng Cao³

¹College of Electrical and Information Engineering, Jiangsu University of Technology, Changzhou 213001, China

²School of Electrical Engineering and Automation, Changshu Institute of Technology, Changshu, 215500 Jiangsu, China

³College of Electrical, Energy and Power Engineering, Yangzhou University, Yangzhou 225127, China

Correspondence should be addressed to Feng Li; lifeng@jsut.edu.cn

Received 22 April 2021; Accepted 6 July 2021; Published 31 July 2021

Academic Editor: Yunze He

Copyright © 2021 Feng Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A novel data-driven learning approach of nonlinear system represented by neural fuzzy Hammerstein-Wiener model is presented. The Hammerstein-Wiener system has two static nonlinear blocks represented by two independent neural fuzzy models surrounding a dynamic linear block described by finite impulse response model. The multisignal theory is designed for employing Hammerstein-Wiener system to separate parameter learning issues. To begin with, the output nonlinearity parameters are learned utilizing separable signal with different amplitudes. Furthermore, correlation analysis method is implemented for estimating linear block parameters using separable signal inputs and outputs; thereby, the interference of process noise is effectively handled. Finally, multi-innovation learning technology is introduced to improve system learning accuracy, and then, multi-innovation extended stochastic gradient algorithm is obtained for optimizing input nonlinearity and noise model using multi-innovation technique and gradient search method. The simulation results display that presented data-driven learning approach has the availability of learning Hammerstein-Wiener system.

1. Introduction

The real industrial processes are almost nonlinear systems to some extent, and linear approximation means are usually unacceptable, and nonlinear models should be taken into account that they can present the nonlinearity successfully. For this, block-oriented nonlinear models which are composed of linear dynamic block and static nonlinear functions for instance Hammerstein model and Wiener model have been performed on account of their simple structures. The two nonlinear models can approximate nonlinear dynamics of many practical industrial processes applications [1–7].

In the past few years, many theoretical researchers and engineers have been performed for extensions of the Hammerstein and Wiener models to improve approximation capabilities of nonlinear systems for instance Hammerstein-Wiener system. In the existing literatures, a lot of optimization techniques have been developed to research the Hammerstein-

Wiener system [8–15]. For Hammerstein-Wiener system, the least square algorithm and blind identification method are put forward by Bai in [8, 9]. In literature [10], recursive parameter learning method are developed for a special nonlinear form described by a Hammerstein-Wiener nonlinear system including dead-zone input nonlinear function. Vörös [11] applied least square-based iterative technique to research Hammerstein-Wiener model parameters using measured input-output data. Xu et al. [12] used two extreme learning machine networks to approximate nonlinear functions of Hammerstein-Wiener system, and parameter estimation method of extreme learning machine-based Hammerstein-Wiener system is developed for large-scale complex nonlinear dynamic systems. The major drawback of the above-analyzed literatures is that the unmodeled dynamic or stochastic disturbances of the Hammerstein-Wiener process is not taken into account, which is an important factor for designing significant parameters learning algorithms [16, 17].

The stochastic gradient estimation methods have attracted much attention due to its less computational load in parameter learning. In recent years, the stochastic gradient-based algorithms have also been implemented to optimize Hammerstein-Wiener models corrupted by stochastic disturbances. For the Hammerstein-Wiener ARMAX system, Wang and Ding [18] investigated extended stochastic gradient estimation. Mansouri et al. [19] developed parameter estimation method through extended Kalman filter theory. Based on data filtering technique, a data filtering-based generalized extended stochastic gradient algorithm is derived for estimating Hammerstein-Wiener system parameters for improving computational efficiency in [20]. It is recognizable that two main problems should be considered in these proposed methods. On the one hand, methods mentioned above assume that unknown nonlinearities in systems are modeled by polynomial functions; if these nonlinearities are not polynomial functions or nonsmooth, methods mentioned do not converge [21]. On the other hand, the parameter cross-products of estimated system are included in learned system, thereby separating each block parameters from obtained parameter estimation of the cross-product terms is required, which increases computation load of learning algorithms [22].

Although many contributions in existing literatures have developed to learn nonlinear system represented by Hammerstein-Wiener model, the problem of stochastic disturbances is not fully considered. This paper focuses attention on a three-stage parameter learning approach of the Hammerstein-Wiener nonlinear systems with stochastic disturbances using multisignal data. In the first stage, the output nonlinearity are estimated depending on separable signal with different amplitudes. In phase two, correlation analysis method is implemented for estimating the linear dynamic block parameter according to one of separable signal inputs and outputs. In the third stage, in order to achieve a fast convergence rate of stochastic gradient algorithm, multi-innovation-based extended stochastic gradient scheme by expanding the scalar innovation to an innovation vector is used to learn parameters of input nonlinearity and noise model. The contributions of developed learning approach lies in:

- (1) Multisignal theory is designed to employ the Hammerstein-Wiener system to separate parameter learning issues, thereby avoiding redundant parameters
- (2) The unmeasurable problems of Hammerstein-Wiener system are well settled by using correlation analysis method
- (3) The multi-innovation-based extended stochastic gradient scheme by expanding the scalar innovation to an innovation vector is used to achieve a fast convergence rate

The paper is organized as follows. Section 2 introduces problem statement of the neural fuzzy Hammerstein-Wiener system. Section 3 analyzes parameter learning based on multisignal data for the Hammerstein-Wiener systems

with stochastic disturbances. Section 4 presents simulation cases of presented learning method. Lastly, the concluding remark is approached.

2. Preliminaries and Problem Statements

As described Figure 1, the nonlinear system represented by Hammerstein-Wiener model with disturbance is modeled by two neural fuzzy networks and finite impulse response model, which is formulated by

$$v(k) = f(u(k)), \quad (1)$$

$$x(k) = B(z)v(k), \quad (2)$$

$$w(k) = D(z)e(k), \quad (3)$$

$$z(k) = x(k) + w(k), \quad (4)$$

$$y(k) = g(z(k)), \quad (5)$$

where $u(k)$ and $y(k)$ denote input and output, $v(k)$ and $x(k)$ represent outputs of input nonlinearity and linear block, $e(k)$ indicates white noise sequence, $f(\cdot)$ shows input nonlinearity, $g(\cdot)$ is output nonlinearity, $B(z)$ is finite impulse response model with $B(z) = b_1z^{-1} + \dots + b_{n_b}z^{-n_b}$, and $D(z) = 1 + d_1z^{-1} + \dots + d_{n_d}z^{-n_d}$ is noise model.

For the given parameter ε , the establishment of the Hammerstein nonlinear system is to seek parameters satisfying the following conditions:

$$\begin{aligned} & E\left(\hat{f}(u(k)), \hat{b}_1, \dots, \hat{b}_{n_b}, \hat{d}_1, \dots, \hat{d}_{n_d}, \hat{g}(\hat{z}(k))\right) \\ &= \frac{1}{2N} \sum_{k=1}^N [y(k) - \hat{y}(k)]^2 \leq \varepsilon, \\ & \text{subject to} \quad \hat{v}(k) = \hat{f}(u(k)), \\ & \hat{z}(k) = \hat{B}(z)\hat{v}(k) + \hat{D}(z)e(k), \\ & \hat{y}(k) = \hat{g}(\hat{z}(k)), \end{aligned} \quad (6)$$

where “ $\wedge$ ” is estimate and N represents length of measured data. From the perspective of easy analysis, the output nonlinearity is expressed by $\hat{z}(k) = \hat{g}^{-1}(y(k))$.

In this research, input nonlinear function and output nonlinear function are modeled using two independent neural fuzzy networks [23]. Figure 1 exhibits the neural fuzzy network, and its output is expressed as

$$\hat{v}(k) = \hat{f}(u(k)) = \sum_{l=1}^L \phi_l(u(k))w_l, \quad (7)$$

where

$$\phi_l(u(k)) = \frac{\mu_l(u(k))}{\sum_{l=1}^L \mu_l(u(k))}, \quad (8)$$

FIGURE 1: Mathematical model of neural fuzzy Hammerstein-Wiener system.

where $\mu_l = \exp(-(u(k) - c_l)^2 / \sigma_l^2)$, w_l represents weights of neural fuzzy network, c_l and σ_l are the center and width, and L is the number of fuzzy rules.

Moreover, expressions of input-output nonlinear blocks are provided

$$\hat{v}(k) = \hat{f}(u(k)) = \sum_{l=1}^{L_{\text{input}}} \phi_l^{\text{input}}(u(k)) w_l^{\text{input}}, \quad (9)$$

$$\hat{z}(k) = \hat{g}^{-1}(y(k)) = \sum_{l=1}^{L_{\text{output}}} \phi_l^{\text{output}}(y(k)) w_l^{\text{output}}, \quad (10)$$

where “input” refers to input nonlinearity and “output” means output nonlinearity.

3. Learning Approach of Neural Fuzzy Hammerstein-Wiener System with Moving Average Noise

The tasks of parameter learning method are to estimate Hammerstein-Wiener system parameters, that is, two nonlinear blocks, linear block, and noise model. Previous research [24] pointed out that the separable signals are employed to realize separation identification of nonlinear block and linear block for the Hammerstein model. Inspired by this work, the separable signals are extended to present Hammerstein-Wiener system with unknown disturbance.

Theorem 1. *Considering a type of Hammerstein-Wiener system, when the separable signals are used as input signal, then the following expression maintains.*

$$R_{vu}(\tau) = b_0 R_u(\tau), \quad \forall \tau \in \mathbb{Z}, \quad (11)$$

where $R_u(\tau) = E(u(k)u(k-\tau))$ is the autocorrelation function, $R_{vu}(\tau) = E(v(k)u(k-\tau))$ is the cross-correlation function, and $b_0 = E(v(k)u(k))/E(u(k)u(k))$ is a constant.

The proof can be done by referring to previous method in [23], hence omitted here.

According to Theorem 1, cross-correlation function $R_{vu}(\tau)$ is taken over by autocorrelation function $R_u(\tau)$ utilizing

separable signal. Therefore, the unknown variable $v(k)$ in Hammerstein-Wiener system is solved.

3.1. Learning Parameters of Output Nonlinearity. The parameters of output nonlinearity are computed using separable signals with multiple relation. In the light of description, the output nonlinearity is modeled by neural fuzzy network, thereby the centre c_l^{output} , the width σ_l^{output} , and the weights w_l^{output} need to be estimated. The centre c_l^{output} and width σ_l^{output} are learned using previous cluster method [24]. Now, a crucial problem needs to be solved for learning parameters w_l^{output} .

Under the condition of two groups of separable signal with multiple relation, we can obtain following output nonlinearities:

$$z_1(k) = \hat{g}^{-1}(y_1(k)) = \sum_{l=1}^{L_{\text{output}}} \phi_l^{\text{output}}(y_1(k)) w_l^{\text{output}}, \quad (12)$$

$$z_2(k) = \hat{g}^{-1}(y_2(k)) = \sum_{l=1}^{L_{\text{output}}} \phi_l^{\text{output}}(y_2(k)) w_l^{\text{output}}. \quad (13)$$

From Equation (2) to Equation (4), we derive

$$z_1(k) = \sum_{j=1}^{n_b} b_j v_1(k-j) + \sum_{m=1}^{n_d} d_m e(k-m) + e(k), \quad (14)$$

$$z_2(k) = \sum_{j=1}^{n_b} b_j v_2(k-j) + \sum_{m=1}^{n_d} d_m e(k-m) + e(k). \quad (15)$$

Using $u_1(k-\tau)$ and $u_2(k-\tau)$ to multiply Equation (14) and Equation (15), respectively, the relation of correlation function are as below.

$$R_{z_1 u_1}(\tau) = \sum_{j=1}^{n_b} b_j R_{v_1 u_1}(\tau-j) + \sum_{m=1}^{n_d} d_m R_{e u_1}(\tau-m) + R_{e u_1}(\tau),$$

$$R_{z_2 u_2}(\tau) = \sum_{j=1}^{n_b} b_j R_{v_2 u_2}(\tau-j) + \sum_{m=1}^{n_d} d_m R_{e u_2}(\tau-m) + R_{e u_2}(\tau). \quad (16)$$

It should be noted that $e(k)$ is uncorrelated with $u(k)$, thereby $R_{eu}(\tau) = 0$, and $R_{eu}(\tau - m) = 0$. Thus,

$$\begin{aligned} R_{z_1 u_1}(\tau) &= \sum_{j=1}^{n_b} b_j R_{v_1 u_1}(\tau - j), \\ R_{z_2 u_2}(\tau) &= \sum_{j=1}^{n_b} b_j R_{v_2 u_2}(\tau - j). \end{aligned} \quad (17)$$

Using Theorem 1, we have

$$R_{z_1 u_1}(\tau) = \sum_{j=1}^{n_b} b_{01} b_j R_{u_1}(\tau - j), \quad (18)$$

$$R_{z_2 u_2}(\tau) = \sum_{j=1}^{n_b} b_{02} b_j R_{u_2}(\tau - j) = \lambda^2 \sum_{j=1}^{n_b} b_{02} b_j R_{u_1}(\tau - j), \quad (19)$$

where $b_{01} = E(v_1(k)u_1(k))/E(u_1(k)u_1(k))$ and $b_{02} = E(v_2(k)u_2(k))/E(u_2(k)u_2(k))$.

Using Equation (18) and Equation (19) acquires

$$R_{z_2 u_2}(\tau) = \beta R_{z_1 u_1}(\tau), \quad (20)$$

where $\beta = (\lambda^2 b_{01})/b_{02}$ and $\lambda = u_2/u_1$.

Equation (20) is given by

$$E(z_2(k)u_2(k - \tau)) = \beta E(z_1(k)u_1(k - \tau)). \quad (21)$$

According to Equation (12), Equation (13) and Equation (21) yields

$$\begin{aligned} \sum_{l=1}^{L_{\text{output}}} w_l^{\text{output}} E(\phi_l^{\text{output}}(y_2(k))u_2(k - \tau)) \\ = \beta \sum_{l=1}^{L_{\text{output}}} w_l^{\text{output}} E(\phi_l^{\text{output}}(y_1(k))u_1(k - \tau)). \end{aligned} \quad (22)$$

Let $\phi_{l,i}^{\text{output}}(k) = \phi_l^{\text{output}} y_i(k) (l = 1, \dots, L_{\text{output}}; i = 1, 2)$, we have

$$\sum_{l=1}^{L_{\text{output}}} w_l^{\text{output}} R_{\phi_{l,2}^{\text{output}} u_2}(\tau) = \beta \sum_{l=1}^{L_{\text{output}}} w_l^{\text{output}} R_{\phi_{l,1}^{\text{output}} u_1}(\tau). \quad (23)$$

Equation (23) is divided by w_1^{output} gets

$$\begin{aligned} R_{\phi_{l,2}^{\text{output}} u_2}(\tau) - \beta R_{\phi_{l,1}^{\text{output}} u_1}(\tau) \\ = \frac{\beta \sum_{l=2}^{L_{\text{output}}} w_l^{\text{output}} R_{\phi_{l,1}^{\text{output}} u_1}(\tau) - \sum_{l=2}^{L_{\text{output}}} w_l^{\text{output}} R_{\phi_{l,2}^{\text{output}} u_2}(\tau)}{w_1^{\text{output}}}. \end{aligned} \quad (24)$$

Let $\tilde{w}_l^{\text{output}} = w_l^{\text{output}}/w_1^{\text{output}}$, we have

$$\begin{aligned} R_{\phi_{l,2}^{\text{output}} u_2}(\tau) - \beta R_{\phi_{l,1}^{\text{output}} u_1}(\tau) \\ = \left[\sum_{l=2}^{L_{\text{output}}} \beta w_l^{\text{output}} R_{\phi_{l,1}^{\text{output}} u_1}(\tau) - R_{\phi_{l,2}^{\text{output}} u_2}(\tau) \right] \tilde{w}_l^{\text{output}}. \end{aligned} \quad (25)$$

Assuming $\tau = 1, 2, \dots, P (P \geq L_{\text{output}} - 1)$, and defining the following cost function according to Equation (20),

$$E(\tilde{w}_l^{\text{output}}) = \frac{1}{2} \sum_{\tau=1}^P [R_{z_2 u_2}(\tau) - \beta R_{z_1 u_1}(\tau)]^2. \quad (26)$$

Based on least square method, parameter θ is estimated

$$\hat{\theta} = (X^T X)^{-1} X^T Y, \quad (27)$$

where $\hat{\theta} = [\tilde{w}_2^{\text{output}}, \tilde{w}_3^{\text{output}}, \dots, \tilde{w}_{L_{\text{output}}}^{\text{output}}]^T$ is estimation, and

$$\begin{aligned} X = [x_1, x_2, \dots, x_{L_{\text{output}}-1}], x_{l-1} = \begin{bmatrix} \beta R_{\phi_{l,1}^{\text{output}} u_1}(1) - R_{\phi_{l,2}^{\text{output}} u_2}(1) \\ \beta R_{\phi_{l,1}^{\text{output}} u_1}(2) - R_{\phi_{l,2}^{\text{output}} u_2}(2) \\ \vdots \\ \beta R_{\phi_{l,1}^{\text{output}} u_1}(P) - R_{\phi_{l,2}^{\text{output}} u_2}(P) \end{bmatrix}, \\ Y = \begin{bmatrix} R_{\phi_{1,2}^{\text{output}} u_2}(1) - \beta R_{\phi_{1,1}^{\text{output}} u_1}(1) \\ R_{\phi_{1,2}^{\text{output}} u_2}(2) - \beta R_{\phi_{1,1}^{\text{output}} u_1}(2) \\ \vdots \\ R_{\phi_{1,2}^{\text{output}} u_2}(P) - \beta R_{\phi_{1,1}^{\text{output}} u_1}(P) \end{bmatrix}. \end{aligned} \quad (28)$$

The correlation functions $R_{\phi_{l,1}^{\text{output}} u_1}(\tau)$ and $R_{\phi_{l,2}^{\text{output}} u_1}(\tau)$ are given by

$$R_{\phi_{l,1}^{\text{output}} u_1}(\tau) = \frac{1}{N} \sum_{k=1}^N \sum_{l=2}^{L_{\text{output}}} \phi_{l,1}^{\text{output}}(y_1(k))u_1(k - \tau), \quad (29)$$

$$R_{\phi_{l,2}^{\text{output}} u_2}(\tau) = \frac{1}{N} \sum_{k=1}^N \sum_{l=2}^{L_{\text{output}}} \phi_{l,2}^{\text{output}}(y_2(k))u_2(k - \tau).$$

Taking the derivative of Equation (26) obtains

$$\beta = \frac{\sum_{\tau=1}^P R_{z_1 u_1}(\tau) R_{z_2 u_2}(\tau)}{\sum_{\tau=1}^P (R_{z_1 u_1}(\tau))^2}. \quad (30)$$

3.2. Learning Parameters of the Linear Block. The measured input-output data of separable signal are implemented to optimize linear block relying on correlation analysis method.

Using Equation (4) gets

$$z_1(k) = \sum_{j=1}^{n_b} b_j v_1(k-j) + \sum_{m=1}^{n_d} d_m e(k-m) + e(k). \quad (31)$$

According to Equation (14)–Equation (18), we have

$$R_{z_1 u_1}(\tau) = \sum_{j=1}^{n_b} \tilde{b}_j R_{u_1}(\tau-j), \quad (32)$$

where $\tilde{b}_j = b_{01} b_j$, $b_{01} = E(v_1(k)u_1(k))/E(u_1(k)u_1(k))$.

Using Equation (32) gets

$$R = \theta_1 \psi, \quad (33)$$

where

$$\begin{aligned} \theta_1 &= [\tilde{b}_1, \tilde{b}_2, \dots, \tilde{b}_{n_b}], R = [R_{z_1 u_1}(1), R_{z_1 u_1}(2), \dots, R_{z_1 u_1}(P)], \psi \\ &= \begin{bmatrix} R_{u_1}(0) & R_{u_1}(1) & R_{u_1}(2) & \dots & R_{u_1}(P-1) \\ 0 & R_{u_1}(0) & R_{u_1}(1) & \dots & R_{u_1}(P-2) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & R_{u_1}(P-n_b) \end{bmatrix}. \end{aligned} \quad (34)$$

Defining the following criterion function:

$$E(\theta_1) = \|R - \theta_1 \psi\|^2. \quad (35)$$

Taking derivative of Equation (43) obtains

$$\frac{\partial E(\theta_1)}{\partial \theta_1} = \frac{\partial [(R - \theta_1 \psi)^T (R - \theta_1 \psi)]}{\partial \theta_1} = 2\theta_1 \psi \psi^T - 2R\psi. \quad (36)$$

Let $\partial E(\theta_1)/\partial \theta_1 = 0$, we get

$$\theta_1 \psi \psi^T = R\psi^T. \quad (37)$$

Equation (37) is multiplied by $(\psi \psi^T)^{-1}$ achieves

$$\hat{\theta}_1 = R\psi^T (\psi \psi^T)^{-1}. \quad (38)$$

The correlation functions $R_{z_1 u_1}(\tau)$ and $R_{u_1}(\tau)$ are presented by

$$\begin{aligned} R_{z_1 u_1}(\tau) &= \frac{1}{N} \sum_{k=1}^N \sum_{l=2}^{L^{\text{output}}} \phi_{l,1}^{\text{output}}(y_1(k)) u_1(k-\tau), \\ R_{u_1}(\tau) &= \frac{1}{N} \sum_{k=1}^N u_1(k) u_1(k-\tau). \end{aligned} \quad (39)$$

3.3. Learning Parameters of Input Nonlinearity and Noise Model. Based on the measured data of random signals, parameters of input nonlinearity and noise model, that is, c_l^{input} , σ_l^{input} , w_l^{input} , and d_m , are learned. Parameters c_l^{input} and σ_l^{input} of input neural fuzzy network are learned using cluster method. Therefore, we need to learn parameters w_l^{input} and d_m .

Using Equation (1)–Equation (4) and Equation (9) gets

$$z(k) = \sum_{j=1}^{n_b} \sum_{l=1}^{L^{\text{input}}} b_j \phi_l(u(k)) w_l^{\text{input}} + \sum_{m=1}^{n_d} d_m e(k-m) + e(k). \quad (40)$$

For convenience, the above equation is described as below:

$$z(k) = \varphi^T(k) \theta_2 + e(k), \quad (41)$$

where

$$\begin{aligned} \theta_2 &= [\theta_s, \theta_e]^T, \theta_s = [b_1 \tilde{w}_2^{\text{input}}, b_1 \tilde{w}_3^{\text{input}}, \dots, b_1 \tilde{w}_{L^{\text{input}}}^{\text{input}}, \dots, b_{n_b} \tilde{w}_2^{\text{input}}, \dots, b_{n_b} \tilde{w}_{L^{\text{input}}}^{\text{input}}]^T, \\ \theta_e &= [d_1, d_2, \dots, d_{n_d}]^T, \varphi_s(k) = [\phi_1(u(k-1)), \dots, \phi_{L^{\text{input}}}(u(k-1)), \dots, \\ &\phi_1(u(k-n_b)), \dots, \phi_{L^{\text{input}}}(u(k-n_b))]^T, \varphi(k) = [\varphi_s(k), \varphi_e(k)]^T, \varphi_e(k) \\ &= [e(k-1), \dots, e(k-n_d)]^T. \end{aligned} \quad (42)$$

The quadratic cost function is defined as

$$J(\theta_2) = \sum_{k=1}^N \|z(k) - \varphi^T(k) \theta_2\|^2. \quad (43)$$

Based on negative search theory, minimizing $J(\theta_2)$ draws

$$\begin{aligned} \hat{\theta}_2(k) &= \hat{\theta}_2(k-1) - \frac{1}{2r(k)} \text{grad} [J(\hat{\theta}_2(k-1))] \\ &= \hat{\theta}_2(k-1) + \frac{\varphi(k)}{r(k)} [z(k) - \varphi^T(k) \hat{\theta}_2(k-1)], \end{aligned} \quad (44)$$

$$r(k) = r(k-1) + \|\varphi(k)\|^2. \quad (45)$$

It is worth emphasizing that the algorithm in Equation (44) and Equation (45) is not carried out due to unknown noise terms $e(k)$ in $\varphi(k)$. In order to solve this issue, a feasible method is to use noise estimation, that is, replacing unmeasurable noise terms $e(k)$ with corresponding estimates $\hat{e}(k)$.

The estimate $\hat{e}(k)$ is expressed as

$$\hat{e}(k) = z(k) - \hat{\varphi}^T(k) \hat{\theta}_2(k), \quad (46)$$

where

$$\begin{aligned} \hat{\varphi}(k) &= [\varphi_s(k), \hat{\varphi}_e(k)]^T, \hat{\varphi}_e(k) \\ &= [\hat{e}(k-1), \dots, \hat{e}(k-n_d)]^T, \hat{\theta}_2 = [\hat{\theta}_s, \hat{\theta}_e]^T. \end{aligned} \quad (47)$$

FIGURE 2: The flowchart of developed learning method.

As a consequence, the following algorithm is obtain:

$$\begin{aligned}
 \hat{\theta}_2(k) &= \hat{\theta}_2(k-1) + \frac{\hat{\varphi}(k)}{r(k)} \left[z(k) - \hat{\varphi}^T(k) \hat{\theta}_2(k-1) \right], \\
 r(k) &= r(k-1) + \|\hat{\varphi}(k)\|^2, \\
 \hat{\varphi}(k) &= [\varphi_s(k), \hat{\varphi}_e(k)]^T, \quad \hat{\theta}_2 = [\hat{\theta}_s, \hat{\theta}_e]^T, \\
 \hat{\varphi}_e(k) &= [\hat{e}(k-1), \dots, \hat{e}(k-n_d)]^T, \\
 \hat{e}(k) &= z(k) - \hat{\varphi}^T(k) \hat{\theta}_2(k).
 \end{aligned} \tag{48}$$

As is known to all, stochastic gradient algorithm has poor convergence rate. To improve convergence rate, an effective method is to use multi-innovation learning theory by expanding the scalar innovation to an innovation vector

FIGURE 3: The estimation of the output nonlinearity.

[25], which uses not only the current data but also past data at each recursive computation.

Set the length of p from $t = k - p + 1$ to $t = k$ and define cost function as below.

$$J(\theta_2) = \sum_{t=0}^{p-1} \|z(k-t) - \varphi^T(k-t) \theta_2\|^2. \tag{49}$$

Using stochastic gradient and minimizing $J(\theta_2)$ gets

$$\hat{\theta}_2(k) = \hat{\theta}_2(k-1) + \frac{1}{r(k)} \sum_{t=0}^{p-1} \varphi(k-t) \left[z(k-t) - \varphi^T(k-t) \hat{\theta}_2(k-1) \right], \tag{50}$$

where p is innovation length.

It is similar to extend stochastic gradient method, replacing unknown variables $\varphi(k-t)$ in Equation (50) by their estimates $\hat{\varphi}(k-t)$, and then, the following approach combining multi-innovation theory with stochastic gradient technique is accomplished:

$$\begin{aligned}
 \hat{\theta}_2(k) &= \hat{\theta}_2(k-1) + \frac{1}{r(k)} \sum_{t=0}^{p-1} \hat{\varphi}(k-t) \\
 &\quad \cdot \left[z(k-t) - \hat{\varphi}^T(k-t) \hat{\theta}_2(k-1) \right], \\
 r(k) &= r(k-1) + \sum_{t=0}^{p-1} \|\hat{\varphi}(k-t)\|^2, \\
 \hat{\theta}_2 &= [\hat{\theta}_s, \hat{\theta}_e]^T, \quad \hat{\varphi}(k) = [\varphi_s(k), \hat{\varphi}_e(k)]^T, \\
 \hat{\varphi}_e(k) &= [\hat{e}(k-1), \dots, \hat{e}(k-n_d)]^T, \\
 \hat{e}(k) &= z(k) - \hat{\varphi}^T(k) \hat{\theta}_2(k).
 \end{aligned} \tag{51}$$

FIGURE 4: Error comparisons using CA method and RELS method.

From the above analysis, the flowchart of developed data-driven learning method is shown in Figure 2.

Remark 2. The proposed three-stage parameter learning approach estimates independently each block parameters of identified Hammerstein-Wiener system using designed multisignals, which avoids the redundant parameters of the system. In contrast, other algorithms like blind parameter identification method [9], extended stochastic gradient identification algorithm [18], and modified bias-eliminating least square algorithm [26] estimate parameters in the product term form, and they need another algorithms such as singular value decomposition method and average method to separate the hybrid parameters. Therefore, the computational complexity of these approaches increases.

4. Numerical Examples

For the developed learning approach, two kinds of multisignals are designed, and numerical cases of nonlinear system represented by Hammerstein-Wiener model with disturbance are applied into certifying the availability.

4.1. Numerical Example 1. The Hammerstein-Wiener system corrupted by noise is concerned with, where the input non-linearity is polynomial.

$$\begin{aligned}
 v(k) &= 0.98u(k) + 0.2u(k)^2, \\
 x(k) &= 0.2v(k-1) + 0.5v(k-2), \\
 z(k) &= x(k) + w(k), \\
 w(k) &= e(k) + 0.5e(k-1), \\
 y(k) &= \begin{cases} 0.1z(k) & z(k) \leq 1.5, \\ 0.15 \exp(z(k) - 1.5) & z(k) > 1.5, \end{cases}
 \end{aligned} \tag{52}$$

where $e(k)$ is stochastic white noise.

Define the noise-to-signal ratios as $\delta_{ns} = \sqrt{\text{var}[w(k)]/\text{var}[x(k)]} \times 100\%$ and parameter estimation error $\delta = \|\hat{\theta}_1(k) - \theta_1\|/\|\theta_1\|$ at sample time k .

The designed multisignal data consist of two sets of Gaussian signals and random signals, including Gaussian signals with mean value of 0 and variance of 1, the mean

FIGURE 5: Approximation of the input nonlinearity using different innovation length.

FIGURE 6: Estimate of noise model.

value of Gaussian signal is 0, and the variance is 0.5, and the range of random signal is -3 to 3.

To begin with, parameters of output nonlinear block are learned with the aid of collected input-output data of two sets of Gaussian signals using least square method. Set the parameters as below: $S_0 = 0.99$, $\rho = 1$, and $\lambda = 0.01$. The estimation of output nonlinearity is depicted in Figure 3. From Figure 3, neural fuzzy networks can well approximate the output nonlinearity by means of developed parameter learning approach.

Moreover, based on input-output data of Gaussian signals with variance of 1, the CA (correlation analysis) algorithm and RELS (recursive extended least square) algorithm [20] are implemented for optimizing linear block. Figure 4 shows error comparisons using CA method and RELS method of different noise-to-signal ratios. The CA algorithm

FIGURE 7: Estimate of output nonlinearity in example 2.

uses cross-covariance function between input and output variables and auto covariance function of input variables to learn the model parameters, which can effectively handle noise interference and improve learning accuracy. From Figure 4, with noise-to-signal ratio increases, the CA algorithm has higher precision than RELS method.

Finally, on the basis of measured input-output data of random signals, parameters of input nonlinearity and noise model are learned adopting $S_0 = 0.9$, $\lambda = 0.01$, and $\rho = 1$. Figure 5 displays the approximation of the input nonlinearity with different innovation length. Figure 6 shows estimate of noise model.

According to Figure 5, it is evident that presented learning method can effectively model input nonlinearity and obtain small approximation error with p increases. According to Figure 6, with the increase of p , the noise model estimate is closer to real value. As a consequence, the introduction of innovation length in developed algorithm can obtain fast convergence rate. This demonstrates that presented three-stage method can accurately learn the Hammerstein-Wiener system.

4.2. Numerical Example 2. In view of a class of Hammerstein-Wiener system with disturbance whose input nonlinearity is discontinuous function:

$$v(k) = \begin{cases} 2 - \cos(3u(k)) - \exp(-u(k)) & u(k) \leq 3.15, \\ 3 & u(k) > 3.15, \end{cases}$$

$$x(k) = 0.9v(k-1) + 0.6v(k-2) + 0.3v(k-3) + 0.1v(k-4),$$

$$z(k) = x(k) + w(k),$$

$$w(k) = e(k) + 0.5e(k-1),$$

$$y(k) = \begin{cases} 0.25 \exp(z(k) - 2.5) & z(k) > 2.5, \\ 0.1z(k) & z(k) \leq 2.5, \end{cases}$$

(53)

where $e(k)$ is noise sequence.

FIGURE 8: Error comparisons using two methods with different noise-to-signal ratios.

The designed multisignal data consist of two sets of binary signals and random signals, including the amplitudes of the two binary signals are 2 and 4, respectively, and interval of random signal is 0 to 5.

The parameters of output nonlinearity are calculated with the aid of collected input-output data of two sets of binary signals using least square method. Set the parameters as below: $S_0 = 0.99$, $\rho = 1$, and $\lambda = 0$. The estimation of output nonlinearity is described in Figure 7. From Figure 7, the neural fuzzy networks can well approximate the output nonlinearity with the help of developed parameter learning approach.

In addition, using data of binary signals whose amplitude is 4, the CA algorithm and RELS algorithm are used. Figure 8 gives error comparisons using two methods in presence of different noise-to-signal ratios. The CA method can effectively deal with the process noise disturbance, so it achieves good parameters learning results. As can be evidently seen from Figure 8, the CA method can more effectively obtain linear block parameters and have better robustness than RELS method.

Lastly, on the basis of measured input-output data of random signals, parameters of input nonlinearity and noise model are learned adopting $S_0 = 0.92$, $\lambda = 0.01$, and

$\rho = 1$. Figure 9 displays the approximation of the input nonlinearity with different innovation length. Figure 10 lists estimate of moving average noise model for different innovation length.

Multi-innovation learning theory is combined with stochastic gradient technique to jointly improve convergence rate by expanding the scalar innovation to an innovation vector. According to Figure 9, it is recognizable that presented learning method can effectively model input nonlinearity and obtain small approximation error with p increases. According to Figure 10, the noise mode estimate is closer to real value with larger innovation length.

Remark 3. For more complex Hammerstein-Wiener system with unknown disturbance in example 2, its input nonlinearity is a discontinuous function; the learning accuracy of parameter learning method proposed is reduced. In addition, it is a common knowledge that convergence rate of stochastic gradient algorithm is poor, the parameter estimation results fluctuate greatly owing to the less information in data used. With data length increases, more data information are used in parameter learning; thus, the fluctuation decreases gradually.

FIGURE 9: Approximation of input nonlinearity with different innovation length.

FIGURE 10: Estimate noise model in example 2.

5. Conclusions

A novel data-driven learning approach of the nonlinear system represented by neural fuzzy Hammerstein-Wiener model with stochastic disturbances is presented. The idea of the developed method is to model the structure of two nonlinear blocks and a linear block at first and then learn the unknown parameters of each block. In the process of Hammerstein-Wiener model modeling, two nonlinear functions are approximated utilizing two neural fuzzy networks, the linear block is modeled applying impulse response model, and stochastic disturbances are described by means of moving average noise.

Multisignal theory is designed for implementing the Hammerstein-Wiener system to segregate parameter learning issues of each block, simplifying parameter learning process. Firstly, the output nonlinear block parameters are learned utilizing separable signal with different amplitudes. Secondly, the correlation analysis algorithm is used; thereby, the interference of process disturbance is effectively settled when estimating linear block. In the end, multi-innovation learning technique is combined with stochastic gradient theory to jointly learn parameters of input nonlinearity and noise model by expanding the scalar innovation to an innovation vector; the convergence rate of the system is improved. This demonstrates the availability of presented Hammerstein-Wiener system with stochastic disturbances using developed learning method.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (62003151 and 61903050), the Natural Science Foundation of Jiangsu Province (BK20191035), and the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (19KJB120002).

References

- [1] F. Li, L. Jia, D. Peng, and C. Han, "Neuro-fuzzy based identification method for Hammerstein output error model with colored noise," *Neurocomputing*, vol. 244, pp. 90–101, 2017.
- [2] J. C. Jeng and Y. W. Lin, "Data-driven nonlinear control design using virtual reference feedback tuning based on the block-oriented modeling of nonlinear systems," *Industrial and Engineering Chemistry Research*, vol. 57, no. 22, pp. 7583–7599, 2018.
- [3] S. I. Biagiola and J. L. Figueroa, "Identification of uncertain MIMO Wiener and Hammerstein models," *Computers and Chemical Engineering*, vol. 35, no. 12, pp. 2867–2875, 2011.
- [4] M. Lawrynczuk, "Nonlinear predictive control of dynamic systems represented by Wiener-Hammerstein models," *Nonlinear Dynamics*, vol. 86, no. 2, pp. 1193–1214, 2016.
- [5] O. Naeem and A. E. M. Huesman, "Non-linear model approximation and reduction by new input-state Hammerstein block structure," *Computers and Chemical Engineering*, vol. 35, no. 5, pp. 758–773, 2011.
- [6] F. le, I. Markovsky, C. Freeman, and E. Rogers, "Recursive identification of Hammerstein systems with application to electrically stimulated muscle," *Control Engineering Practice*, vol. 20, no. 4, pp. 386–396, 2012.
- [7] J. G. Smith, S. Kamat, and K. P. Madhavan, "Modeling of PH process using wavenet based Hammerstein model," *Journal of Process Control*, vol. 17, no. 6, pp. 551–561, 2007.

- [8] E. W. Bai, "An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems," *Automatica*, vol. 34, no. 3, pp. 333–338, 1998.
- [9] E. W. Bai, "A blind approach to the Hammerstein-Wiener model identification," *Automatica*, vol. 38, no. 6, pp. 967–979, 2002.
- [10] F. Yu, Z. Mao, and M. Jia, "Recursive identification for Hammerstein-Wiener systems with dead-zone input nonlinearity," *Journal of Process Control*, vol. 23, no. 8, pp. 1108–1115, 2013.
- [11] J. Vörös, "Iterative identification of nonlinear dynamic systems with output backlash using three-block cascade models," *Nonlinear Dynamics*, vol. 79, no. 3, pp. 2187–2195, 2015.
- [12] K. K. Xu, H. D. Yang, and C. J. Zhu, "A novel extreme learning machine-based Hammerstein-Wiener model for complex nonlinear industrial processes," *Neurocomputing*, vol. 358, pp. 246–254, 2019.
- [13] F. Li, K. Yao, B. Li, and L. Jia, "A novel learning algorithm of the neuro-fuzzy based Hammerstein-Wiener model corrupted by process noise," *Journal of the Franklin Institute*, vol. 358, no. 3, pp. 2115–2137, 2021.
- [14] F. Li, L. Chen, S. Wo, S. Li, and Q. Cao, "Modeling and parameter learning method for the Hammerstein-Wiener model with disturbance," *Measurement and Control*, vol. 53, no. 5–6, pp. 971–982, 2020.
- [15] F. Li and L. Jia, "Parameter estimation of Hammerstein-Wiener nonlinear system with noise using special test signals," *Neurocomputing*, vol. 344, pp. 37–48, 2019.
- [16] J. Wang, H. W. Wang, and H. Gu, "A novel recursive subspace identification approach of closed-loop systems," *Mathematical and Computer Modelling of Dynamical Systems*, vol. 19, no. 6, pp. 526–539, 2013.
- [17] A. Mahmoudi, M. Karimi, and H. Amindavar, "Parameter estimation of autoregressive signals in presence of colored AR(1) noise as a quadratic eigenvalue problem," *Signal Processing*, vol. 92, no. 4, pp. 1151–1156, 2012.
- [18] D. Wang and F. Ding, "Extended stochastic gradient identification algorithms for Hammerstein-Wiener ARMAX systems," *Computers and Mathematics with Applications*, vol. 56, no. 12, pp. 3157–3164, 2008.
- [19] M. Mansouri, H. Tolouei, and M. A. Shoorehdeli, "Identification of Hammerstein-Wiener ARMAX systems using extended kalman filter," in *2011 Chinese Control and Decision Conference (CCDC)*, pp. 1110–1114, Mianyang, China, May 2011.
- [20] D. Wang and F. Ding, "Recursive least squares algorithm and gradient algorithm for Hammerstein-Wiener systems using the data filtering," *Nonlinear Dynamics*, vol. 84, no. 2, pp. 1045–1053, 2016.
- [21] Z. Q. Lang, "On identification of the controlled plants described by the Hammerstein system," *IEEE Transactions on Automatic Control*, vol. 39, no. 3, pp. 569–573, 1994.
- [22] F. Ding and T. Chen, "Identification of Hammerstein nonlinear ARMAX systems," *Automatica*, vol. 41, no. 9, pp. 1479–1489, 2005.
- [23] F. Li and L. Jia, "Correlation analysis-based error compensation recursive least-square identification method for the Hammerstein model," *Journal of Statistical Computation and Simulation*, vol. 88, no. 1, pp. 56–74, 2018.
- [24] F. Li, J. Li, and D. Peng, "Identification method of neuro-fuzzy-based Hammerstein model with coloured noise," *IET Control Theory and Applications*, vol. 11, no. 17, pp. 3026–3037, 2017.
- [25] X. Wang and F. Ding, "Modelling and multi-innovation parameter identification for Hammerstein nonlinear state space systems using the filtering technique," *Mathematical and Computer Modelling of Dynamical Systems*, vol. 22, no. 2, pp. 113–140, 2016.
- [26] Z. Wang, Y. Wang, and Z. Ji, "A novel two-stage estimation algorithm for nonlinear Hammerstein-Wiener systems from noisy input and output data," *Journal of the Franklin Institute*, vol. 354, no. 4, pp. 1937–1944, 2017.