

## Article

# A Robust Hammerstein-Wiener Model Identification Method for Highly Nonlinear Systems

Lijie Sun <sup>1</sup>, Jie Hou <sup>2,\*</sup>, Chuanjun Xing <sup>3,\*</sup> and Zhewei Fang <sup>1</sup><sup>1</sup> School of Electronics and Information Engineering, Taizhou University, Taizhou 318000, China<sup>2</sup> College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China<sup>3</sup> College of Computer Science and Technology, Heilongjiang Institute of Technology, Harbin 150050, China

\* Correspondence: jiehou.phd@hotmail.com (J.H.); xcj\_hljit@163.com (C.X.); Tel.: +86-158-2608-9865 (J.H.); +86-131-9945-5545 (C.X.)

**Abstract:** The existing results show the applicability of the Over-Parameterized Model based Hammerstein-Wiener model identification methods. However, it requires to estimate extra parameters and perform a low rank approximation step. Therefore, it may give rise to unnecessarily high variance in parameter estimates for highly nonlinear systems, especially using a small and noisy data set. To overcome this corruptive phenomenon. To overcome this corruptive phenomenon, in this paper, a robust Hammerstein-Wiener model identification method is developed for highly nonlinear systems when using a small and noisy data set, where two parsimonious parametrization models with fewer parameters are used, and an iteration method is then used to retrieve the true system parameters from the parametrization models. Such modification can improve the parameter estimation performance in terms of accuracy and variance compared with the over-parametrization model based identification methods. All the above-mentioned developments are analyzed with variance analysis, along with a simulation example to confirm the effectiveness.

**Keywords:** Hammerstein–Wiener model; iteration method; nonlinear system identification



**Citation:** Sun, L.; Hou, J.; Xing, C.; Fang, Z. A Robust Hammerstein-Wiener Model Identification Method for Highly Nonlinear Systems. *Processes* **2022**, *10*, 2664. <https://doi.org/10.3390/pr10122664>

Academic Editor: Mohd Azlan Hussain

Received: 21 November 2022

Accepted: 8 December 2022

Published: 11 December 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The Hammerstein–Wiener (H-W) models, also called N-L-N models, are known as a particular class of block-oriented nonlinear models where a Linear (L) subsystem is embedded in two static Nonlinear subsystems (N). Due to the use of two nonlinear elements instead of one, H-W models offer convenient higher modeling capabilities for nonlinear systems, especially for nonlinear systems with both actuator and sensor nonlinearities [1]. H-W models identification has attracted considerable attention in the past years and a large amount of works have been explored in different frameworks, see, for example, iterative methods [2,3], and Least-Squares (LS) based Two-Step Algorithms (TSA) [4].

In this paper, the emphasis is put on the LS issue for H-W models with over-parametrization methods, which was perhaps first studied by Bai [4] where first characterizes the original H-W model by an over-parametrization model and then estimates the parameters of this over-parametrization model (OPM) by using a LS method. In consequence, the original system parameters are extracted from the estimated over parametrization model parameters by performing low-rank approximation using Singular Value Decompositions (SVDs). The TSA is attractive because of its numerical simplicity. Thus, following from this pioneering work [4], a number of TSA-based methods have been proposed to identify H-W models under different conditions: For instance, in [5], a auxiliary model multi-innovation stochastic gradient (AM-MISG) based TSA was proposed for H-W models online identification, which has fast convergence rate with robust characteristics and its convergence was proved by using the stochastic process theory. In [6], an instrumental variable based TSA was proposed for H-W continuous-time models in the presence of colored noise. In [7], a modified Bias-Eliminating Least-Squares (BELS) based TSA was proposed for H-W

errors-in-variables models. The TSA has also been extensively used for the identification of other types of block-oriented nonlinear systems, e.g., Hammerstein models [8,9], and Wiener models [10].

In these contributions, all existing TSAs provide asymptotically consistent unbiased estimations for H-W models when the data length tends to infinity. However, it is well known that the existing TSAs related to the use of over-parametrization model and low-rank approximation step may result in a degradation of the accuracy and variance of the estimated quantities for highly nonlinear systems, especially for short noisy data length sequences. One reason is that the over-parametrization model has much more parameters than the actual Hammerstein-Wiener model, the extra parameters in TSAs may lead to larger estimation errors and variances, and the number of the extra parameters increase with the degree of nonlinearity. The other is that the low-rank approximation step in TSAs inevitably brings some truncation errors, which will result in some loss of model quality.

In this paper, inspired by the parsimonious models based identification methods [11–14] for block-oriented nonlinear Hammerstein models can improve the variance properties of the over-parametrization model based methods, a robust consistent identification formulation for H-W models is thus proposed to overcome the above-mentioned drawbacks for highly nonlinear systems when using short noisy data length sequences, without the necessity of estimating extra parameters and performing low-rank approximation. The main contributions of the paper are stated below:

- Two parsimonious models with fewer parameters constructed to characterize the H-W system instead of an over-parametrization model, and a projection based LS iterative method is proposed to estimate all system parameters in a parallel fashion. The proposed method avoids the estimating extra parameters and a low rank approximation step in classical over-parametrization model based methods, which leads to improved results compared to existing over-parametrization model based methods, because that estimating extra parameters and performing low rank approximation may result into a degradation of the accuracy of the estimated quantities.
- The variance analysis is given to demonstrate that the new method generally gives smaller variance compared to conventional TSA method. Note that the results of variance analysis explains why the proposed identification methods can lead to improved results compared to existing over-parametrization model based methods for H-W systems.

Note that Hammerstein models and Wiener models are two special cases of Hammerstein–Wiener models, that is, when  $n_c = 1$  and  $f_1(u(t)) = u(t)$  in following (1), the Hammerstein–Wiener models become Wiener models, and when  $n_d = 1$  and  $g_1(y(t)) = y(t)$  in following (1), the Hammerstein–Wiener models become Hammerstein models. The presented methodology in this paper is general for N-L-N models, L-N models and N-L models.

The paper is organized as follows. In Section 2, the identification problem is presented. In Section 3, the proposed method is presented, followed by the Variance analysis in Section 4. Section 5 presents a simulation example to validate and evaluate the performance of the new algorithm. Finally, some conclusions are drawn in Section 6.

## 2. Problem Statement and Analysis

Consider a particular class of nonlinear H-W model with polynomial nonlinearities, see also [4]

$$y(t) = \sum_{j=1}^{n_b} b_j \left\{ \sum_{m=1}^{n_c} c_m f_m[u(t-j)] \right\} + \sum_{i=1}^{n_a} a_i \left\{ \sum_{l=1}^{n_d} d_l g_l[y(t-i)] \right\} + e(t) \quad (1)$$

where system parameters are  $\mathbf{a} = [a_1, \dots, a_{n_a}]$ ,  $\mathbf{b} = [b_1, \dots, b_{n_b}]$ ,  $\mathbf{c} = [c_1, \dots, c_{n_c}]$ , and  $\mathbf{d} = [d_1, \dots, d_{n_d}]$ ,  $\{u(t)\}$  and  $\{y(t)\}$  are the system input and output sequences, respectively,  $\{e(t)\}$  is the noise sequence, and  $f_m(\bullet)$  and  $g_l(\bullet)$  are nonlinear functions. The following general assumptions are made in this paper:

- A1 The first nonzero elements of  $\mathbf{a}$  and  $\mathbf{b}$  are positive and  $\|\mathbf{a}\|_2 = 1$  and  $\|\mathbf{b}\|_2 = 1$ .  
 A2 The functions  $\{f_m(\bullet)\}_{m=1}^{n_c}$  and  $g_l(\bullet)_{l=1}^{n_d}$  and degrees  $n_a, n_b, n_c$  and  $n_d$  are known.  
 A3  $\{e(i)\}$  is a zero-mean white noise sequence with finite variance  $\delta_e^2$  uncorrelated with  $\{u(j)\}$  for all  $i, j$ , and uncorrelated with  $\{y(j)\}$  for  $i \neq j$ .

Assumption 1 ensures uniqueness of the N-L-N nonlinear system model, as the introduction of a factor  $q$  ( $q \neq 0$ ) to the pair  $(q\mathbf{a}, \mathbf{d}/q)$ , the input-output relationship will not change. The terms  $\|\mathbf{a}\|_2$  and  $\|\mathbf{b}\|_2$  can be set as any other positive number, the proposed method can be directly used without any change.

Assumption 2 are related to a priori knowledge about the structure of the true system. In this context the static nonlinearity of the N-L-N model is approximated using a linear expansion in terms of basis functions. In generally, the basis functions and system orders lie in the fact that tuning them are more directly related to a priori knowledge about the true system. However, since the unknown coefficients of parameter vectors  $\mathbf{a}, \mathbf{b}, \mathbf{c}$  and  $\mathbf{d}$  can give the degrees of freedom to accurately modeling the nonlinear systems, the basis functions and system orders can be assumed to known simply for having not directly related knowledge about the nonlinear function in practice. More specifically, the system orders can be set high enough to provide enough degrees of freedom to accurately describing the nonlinear N-L-N systems. The basis functions can be chosen as polynomial functions simply, which is realistic in many applications where the nonlinearities characterizing the systems are smooth enough [9]. Note that the identification method developed under those assumptions in which the nonlinear basis functions as well as the system orders in polynomial model supposed to be known can be regarded as a grey-box method compared with the black-box model approach without any assumption about the structure of the N-L-N model.

Assumption 3 requires the noises to be white, which is realistic in many applications when the main objective is on the plant model.

Based on (1), an over-parametrization model can be formulated as

$$\mathbf{y}(t) = \boldsymbol{\theta}\Phi(t) + e(t) \quad (2)$$

where

$$\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2] \in \mathbb{R}^{1 \times (n_b n_c + n_a n_d)} \quad (3a)$$

$$\boldsymbol{\theta}_1 = [b_1 c_1, \dots, b_1 c_{n_c}, \dots, b_{n_b} c_1, \dots, b_{n_b} c_{n_c}] \in \mathbb{R}^{1 \times n_b n_c} \quad (3b)$$

$$\boldsymbol{\theta}_2 = [a_1 d_1, \dots, a_1 d_{n_d}, \dots, a_{n_a} d_1, \dots, a_{n_a} d_{n_d}] \in \mathbb{R}^{1 \times n_a n_d} \quad (3c)$$

$$\Phi(t) = [f(t), g(t)]^\top \in \mathbb{R}^{(n_b n_c + n_a n_d) \times 1} \quad (3d)$$

$$f(t) = [f_1[u(t-1)], \dots, f_{n_c}[u(t-1)], \dots, f_1[u(t-n_b)], \dots, f_{n_c}[u(t-n_b)]]^\top \in \mathbb{R}^{n_b n_c \times 1} \quad (3e)$$

$$g(t) = [g_1[y(t-1)], \dots, g_{n_d}[y(t-1)], \dots, g_1[y(t-n_b)], \dots, g_{n_d}[y(t-n_a)]]^\top \in \mathbb{R}^{n_a n_d \times 1}. \quad (3f)$$

Define

$$Y = [y(1), y(2), \dots, y(N)] \quad (4a)$$

$$E = [e(1), e(2), \dots, e(N)] \quad (4b)$$

$$F = [f(1), f(2), \dots, f(N)] \quad (4c)$$

$$G = [g(1), g(2), \dots, g(N)] \quad (4d)$$

$$\Phi = [F^\top, G^\top]^\top. \quad (4e)$$

It follows from (1)–(4) that

$$Y = \theta_1 F + \theta_2 G + E = \theta \Phi + E. \quad (5)$$

For the estimation of the unknown system parameters in (5), the following criterion is considered in the TSA:

$$\min_{a,b,c,d} \frac{1}{2} (Y - \theta \Phi)^\top W (Y - \theta \Phi) \quad (6a)$$

$$\text{s.t. rank}(bc) = 1, \text{rank}(ad) = 1 \quad (6b)$$

where  $W$  is a symmetric weighting matrix, e.g., an identity matrix of appropriate dimension  $I$ .

The TSA is a relaxation method by solving the unconstrained LS problem (6a) without considering the rank constraint (6b), where the system parameters are extracted via SVDs, see [4] for more details. Clearly, the number of extra parameters in the TSA method [4] is  $n_a n_d + n_b n_c - n_a - n_d - n_b - n_c$ , and two SVDs have to be used to extract the system parameters from  $\theta$ . These may deteriorate the accuracy and variance of estimated quantities. Note that the number of the extra parameters in the TSA will increase as the degree of nonlinearity of H-W models increases. The performance of the TSA method in terms of the accuracy and variance of estimated quantities will be deteriorated for identifying highly nonlinear systems from a limited number of noisy data.

This paper aims to seek a method to estimate unknown parameter vectors  $a, b, c$  and  $d$  from the known observed input-output data  $\{u(t), y(t)\}$  collected at instants  $t = 1, \dots, N$  without estimating extra parameters and performing SVDs.

### 3. The Proposed Method

#### 3.1. Parsimonious Parametrization Model Formulation

We decouple  $\theta$  into two parts  $\theta_1$  and  $\theta_2$ . The over-parametrization model in (2) can be then represented as following two parsimonious parametrization models (PPMs):

$$y(t) = a\Phi_1(t) + b\Phi_2(k) + e(t) = d\Phi_3(t) + c\Phi_4(k) + e(t) \quad (7)$$

where

$$\Phi_1(t) = (d_1 g_1[y(t-1)] + \dots + d_{n_d} g_{n_d}[y(t-1)], \dots, d_1 g_1[y(t-n_a)] + \dots + d_{n_d} g_{n_d}[y(t-n_a)])^\top \quad (8a)$$

$$\Phi_2(t) = (c_1 f_1[u(t-1)] + \dots + c_{n_c} f_{n_c}[u(t-1)], \dots, c_1 f_1[u(t-n_b)] + \dots + c_{n_c} f_{n_c}[u(t-n_b)])^\top \quad (8b)$$

$$\Phi_3(t) = \left( \sum_{i=1}^{n_a} (a_i g_1[y(t-i)]), \dots, \sum_{i=1}^{n_a} (a_i g_{n_d}[y(t-i)]) \right)^\top \quad (8c)$$

$$\Phi_4(t) = \left( \sum_{i=1}^{n_b} (b_i f_1[u(t-i)]), \dots, \sum_{i=1}^{n_b} (b_i f_{n_c}[u(t-i)]) \right)^\top. \quad (8d)$$

Define

$$\Phi_1 = [\Phi_1(1), \Phi_1(2), \dots, \Phi_1(N)] = \underbrace{\begin{bmatrix} \mathbf{d} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{d} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{d} \end{bmatrix}}_{\eta_1} \mathbf{G} \quad (9a)$$

$$\Phi_2 = [\Phi_2(1), \Phi_2(2), \dots, \Phi_2(N)] \quad (9b)$$

$$\Phi_3 = [\Phi_3(1), \Phi_3(2), \dots, \Phi_3(N)] \quad (9c)$$

$$\Phi_4 = [\Phi_4(1), \Phi_4(2), \dots, \Phi_4(N)]. \quad (9d)$$

It follows from (7) that

$$Y = \mathbf{a}\Phi_1 + \mathbf{b}\Phi_2 + E = \mathbf{d}\Phi_3 + \mathbf{c}\Phi_4 + E. \quad (10)$$

### 3.2. Parsimonious Identification Method

We consider an iterative estimation procedure to estimate unknown parameters from (10). Projecting the row space of  $Y$  in (10) onto the orthogonal complement of the row space of the matrix  $F$ , we have

$$Y\Pi_F^\perp = (\mathbf{a}\Phi_1 + \mathbf{b}\Phi_2 + E)\Pi_F^\perp = (\mathbf{d}\Phi_3 + \mathbf{c}\Phi_4 + E)\Pi_F^\perp \quad (11)$$

where  $\Pi_F^\perp = I_N - F^\top(F F^\top)^{-1}F$ . Due to  $\mathbf{b}\Phi_2 = \boldsymbol{\theta}_1 F$  and  $\mathbf{c}\Phi_4 = \boldsymbol{\theta}_1 F$ , we have

$$Y\Pi_F^\perp = \mathbf{a}\Phi_1\Pi_F^\perp + E\Pi_F^\perp = \mathbf{d}\Phi_3\Pi_F^\perp + E\Pi_F^\perp. \quad (12)$$

Once  $\mathbf{d}$  is assumed to be known, the parameter vector  $\mathbf{a}$  can be estimated without knowing  $\mathbf{c}$  and  $\mathbf{b}$ , while the parameter vector  $\mathbf{d}$  can be estimated with only knowing  $\mathbf{a}$ , which allows an iterative method to obtain  $\mathbf{a}$  and  $\mathbf{d}$  based (12). The details of estimation of  $\mathbf{a}$  and  $\mathbf{d}$  are listed as below.

- An LS estimate  $\hat{\mathbf{a}}^{(k)}$  can be obtained by substituting  $\hat{\mathbf{d}}^{(k-1)}$  into (12) as

$$\hat{\mathbf{a}}^{(k)} = Y\Pi_F^\perp \hat{\Phi}_1^{\top(k-1)} (\hat{\Phi}_1^{(k-1)} \Pi_F^\perp \hat{\Phi}_1^{\top(k-1)})^{-1} \quad (13)$$

where the estimate  $\hat{\Phi}_1^{(k-1)}$  of  $\Phi_1$  is a function of  $\hat{\mathbf{d}}^{(k-1)}$ .

- The unique solution  $\hat{\mathbf{a}}^{(k)}$  can be obtained by performing a normalization operation as follows:

$$\hat{\mathbf{a}}^{(k)} = \hat{\mathbf{a}}^{(k)} / \|\hat{\mathbf{a}}^{(k)}\|_2. \quad (14)$$

- An LS estimate  $\hat{\mathbf{d}}^{(k)}$  can be obtained by substituting  $\hat{\mathbf{a}}^{(k)}$  into (12) as

$$\hat{\mathbf{d}}^{(k)} = Y\Pi_F^\perp \hat{\Phi}_3^{\top(k)} (\hat{\Phi}_3^{(k)} \Pi_F^\perp \hat{\Phi}_3^{\top(k)})^{-1} \quad (15)$$

where the estimate  $\hat{\Phi}_3^{(k)}$  of  $\Phi_3$  is a function of  $\hat{\mathbf{a}}^{(k)}$ .

In similar arguments to estimate  $\mathbf{a}$  and  $\mathbf{d}$ , by projecting the row space of  $Y$  in (10) onto the orthogonal complement of the row space of the matrix  $G$ , we get

$$Y\Pi_G^\perp = \mathbf{b}\Phi_2\Pi_G^\perp + E\Pi_G^\perp = \mathbf{c}\Phi_4\Pi_G^\perp + E\Pi_G^\perp. \quad (16)$$

- An LS estimate  $\hat{\mathbf{b}}^{(k)}$  can be estimated by substituting  $\hat{\mathbf{c}}^{(k-1)}$  into (16) as

$$\hat{\mathbf{b}}^{(k)} = Y\Pi_G^\perp \hat{\Phi}_2^{\top(k-1)} (\hat{\Phi}_2^{(k-1)} \Pi_G^\perp \hat{\Phi}_2^{\top(k-1)})^{-1} \quad (17)$$

where the estimate  $\hat{\Phi}_2^{(k-1)}$  of  $\Phi_2$  is a function of  $\hat{\mathbf{c}}^{(k-1)}$ .

- The unique solution  $\hat{\mathbf{b}}^{(k)}$  can be obtained by performing a normalization operation as follows:

$$\hat{\mathbf{b}}^{(k)} = \hat{\mathbf{b}}^{(k)} / \|\hat{\mathbf{b}}^{(k)}\|_2. \quad (18)$$

- An LS estimate  $\hat{\mathbf{c}}^{(k)}$  can be estimated by substituting  $\hat{\mathbf{b}}^{(k)}$  into (16) as

$$\hat{\mathbf{c}}^{(k)} = Y \Pi_G^\perp \hat{\Phi}_4^\top(k) (\hat{\Phi}_4^{(k)} \Pi_G^\perp \hat{\Phi}_4^\top(k))^{-1} \quad (19)$$

where the estimate  $\hat{\Phi}_4^{(k)}$  of  $\Phi_4$  is a function of  $\hat{\mathbf{b}}^{(k)}$ .

The following general assumption A4 about persistently exciting (PE) conditions is made for system identifiability.

A4 Matrices  $F$  and  $G$  are full row rank.

The assumption A4 is a general assumption for all identification methods, which can be easily guaranteed by constructing  $F$  and  $G$  based on random signals. Based on which, once the initial values  $\hat{\mathbf{d}}(0)$  and  $\hat{\mathbf{c}}(0)$  are nonzero, it follows from (10) that matrices  $(\hat{\Phi}_1(0), \hat{\Phi}_2(0))$  will be full row rank, due to that the rank of matrices  $\hat{\eta}_1$  and  $\hat{\eta}_2$  with nonzero parameters  $\hat{\mathbf{d}}(0)$  and  $\hat{\mathbf{c}}(0)$  are full row rank, such that the estimates  $\hat{\mathbf{a}}(1)$  in (13) and  $\hat{\mathbf{b}}(1)$  in (17) are ensured nonzero. Once the nonzero estimated parameters  $\hat{\mathbf{a}}(1)$  and  $\hat{\mathbf{b}}(1)$  are obtained, and then  $\hat{\mathbf{c}}(1)$  and  $\hat{\mathbf{d}}(1)$  are ensured nonzero, due to that the rank of matrices  $\hat{\eta}_3$  and  $\hat{\eta}_4$  with nonzero parameters  $\hat{\mathbf{a}}(1)$  and  $\hat{\mathbf{b}}(1)$  are also full row rank. Following above arguments, we conclude that once the PE matrices and the nonzero initial values are given, the proposed algorithm can provide consistent estimated parameters.

The parameters are estimated asynchronously from the PPMs by two independent projection based iterative methods, thus the proposed method, referred to as Asynchronous Parsimonious Method (APM), is summarized as Algorithm 1.

---

#### Algorithm 1:

---

1. Initialization:  $\hat{\mathbf{d}}(0)$  and  $\hat{\mathbf{c}}(0)$ : arbitrary nonzero values.
  2. Iteration:  
for  $k = 0$  : convergence
    - a1. Estimate  $\hat{\mathbf{a}}^{(k)}$  as (13) and (14).
    - a2. Estimate  $\hat{\mathbf{d}}^{(k)}$  as (15).
    - a3. If the stopping criterion is satisfied, and break, else go to step (a1).
 end
  3. Iteration:  
for  $k = 0$  : convergence
    - b1. Estimate  $\hat{\mathbf{b}}^{(k)}$  as (17) and (18).
    - b2. Estimate  $\hat{\mathbf{c}}^{(k)}$  as (19).
    - b3. If the stopping criterion is satisfied, and break, else go to step (b1).
 end
  4. Let  $\hat{\mathbf{a}}^\top = \hat{\mathbf{a}}^\top(k) \text{sgn}(\hat{\mathbf{a}}_1^\top(k))$ ,  $\hat{\mathbf{b}}^\top = \hat{\mathbf{b}}^\top(k) \text{sgn}(\hat{\mathbf{b}}_1^\top(k))$ ,  $\hat{\mathbf{d}}_2 = \hat{\mathbf{d}}_2(k+1) \text{sgn}(\hat{\mathbf{a}}_1^\top(k))$ , and  $\hat{\mathbf{c}}_2 = \hat{\mathbf{c}}_2(k+1) \text{sgn}(\hat{\mathbf{b}}_1^\top(k))$ , where  $\text{sgn}()$  is a sign function and  $\hat{\mathbf{a}}_1^\top(k)$  and  $\hat{\mathbf{b}}_1^\top(k)$  are the first element of  $\hat{\mathbf{a}}^\top(k)$  and  $\hat{\mathbf{b}}^\top(k)$ , respectively.
- 

#### 4. Variance Analysis

Following the Theorem 1 in [11], one can easily prove that the proposed method with arbitrary nonzero initial values  $\hat{\mathbf{d}}(0)$  and  $\hat{\mathbf{c}}(0)$  provides consistent estimates for H-W models. This section, we analyze the variance of the AMP estimates relative to that of conventional TSA algorithm. Due to all parameters are estimated in a similar way, we only analyze the

variances of estimated parameter  $\hat{a}$ . The LS estimation error variance of  $\hat{\theta}$  in TSA [4] can be computed as,

$$\text{cov}(\hat{\theta}_1, \hat{\theta}_2) = \delta_e^2 (\Phi \Phi^\top)^{-1} \quad (20)$$

Substituting (4e) into (20), we have

$$\text{cov}(\hat{\theta}_1, \hat{\theta}_2) = \delta_e^2 \begin{bmatrix} \mathbf{F}\mathbf{F}^\top & \mathbf{F}\mathbf{G}^\top \\ \mathbf{G}\mathbf{F}^\top & \mathbf{G}\mathbf{G}^\top \end{bmatrix}^{-1} \quad (21)$$

Based the block matrix inversion relation in [15], we have

$$\text{cov}(\hat{\theta}_2) = \delta_e^2 (\mathbf{G}\mathbf{G}^\top)^{-1} + \delta_e^2 (\mathbf{G}\mathbf{G}^\top)^{-1} \mathbf{G}\mathbf{F}^\top \mathbf{Q}^{-1} (\mathbf{G}\mathbf{F}^\top)^\top (\mathbf{G}\mathbf{G}^\top)^{-1} \quad (22)$$

where  $\mathbf{Q} = \mathbf{F}\mathbf{F}^\top - \mathbf{F}\mathbf{G}^\top (\mathbf{G}\mathbf{G}^\top)^{-1} \mathbf{G}\mathbf{F}^\top$ .

Since the solutions of  $\hat{a}$ ,  $\hat{b}$ ,  $\hat{c}$  and  $\hat{d}$  are provided by SVD, decompositions of  $\hat{\theta}_1$ , and  $\hat{\theta}_2$ . Omitting the error terms more than first order which do not contribute to the estimation error variance, deriving based on first-order Taylor expansion of SVD in [16], the variance of the parameter estimation errors of  $\hat{a}$  can be computed in approximately as following,

$$\begin{aligned} \text{cov}(\hat{a}_{TSA}) &= \delta_e^2 (\boldsymbol{\eta}_1 \mathbf{F}\mathbf{F}^\top \boldsymbol{\eta}_1^\top)^{-1} + \\ &\delta_e^2 (\boldsymbol{\eta}_1 \mathbf{F}\mathbf{F}^\top)^{-1} \mathbf{F}\mathbf{G}^\top \mathbf{Q}^{-1} (\mathbf{F}\mathbf{G}^\top)^\top (\mathbf{F}\mathbf{F}^\top \boldsymbol{\eta}_1^\top)^{-1} \end{aligned} \quad (23)$$

Note that the second terms at the right-hand side of (52) are positive definite, we have

$$\text{cov}(\hat{a}_{TSA}) \geq \delta_e^2 (\boldsymbol{\eta}_1 \mathbf{F}\mathbf{F}^\top \boldsymbol{\eta}_1^\top)^{-1} \quad (24)$$

In a similar way as (20), the estimation error variances of  $\hat{a}$ ,  $\hat{b}$ ,  $\hat{c}$  and  $\hat{d}$  in APM are computed as following

$$\text{cov}(\hat{a}_{APM}) = \delta_e^2 (\hat{\boldsymbol{\eta}}_1 \mathbf{F}\mathbf{F}^\top \hat{\boldsymbol{\eta}}_1^\top)^{-1} \quad (25)$$

Since estimates from AMP are consistent, neglecting the high-order terms in above equations for asymptotic variance, and only considering the noise terms for estimation error variances, we have

$$\text{cov}(\hat{a}_{APM}) = \delta_e^2 (\boldsymbol{\eta}_1 \mathbf{F}\mathbf{F}^\top \boldsymbol{\eta}_1^\top)^{-1} \quad (26)$$

Compared (24) and (26), we conclude that the APM estimates generally have a smaller variance than those of the TSA estimates by neglecting high-order terms in estimation error.

## 5. Case Studies

In this section, we use a benchmark Hammerstein–Wiener model with highly nonlinearities modified from [4] to evaluate the proposed APM and compare it with the TSA. The system model of the benchmark example is presented in the following:

$$\begin{aligned} \mathbf{a} &= [0.1569, 0.0934, 0.5477, 0, -0.7303, -0.3651]^\top \\ \mathbf{b} &= [0.1569, 0.0934, 0.5477, 0, 0.7303, 0.3651]^\top \\ \mathbf{c} &= [1, 4, 3, 7, 1, 3.4]^\top \\ \mathbf{d} &= [0.4, 0.25]^\top \\ f(t) &= [u(t), u^2(t), u^3(t), u^4(t), u^5(t), u^6(t)]^\top \\ g(t) &= [y(t), \sin(y(t))]^\top. \end{aligned}$$

The input  $\{u(t)\}$  and the measurement noise  $\{e(t)\}$  are chosen as two zero-mean Gaussian random sequences. The variance of  $\{u(t)\}$  is fixed as one, while the variance of  $\{e(t)\}$  is adjusted to achieve a given Signal-to-Noise Ratio (SNR). In this example, three SNR values will be considered, i.e., 15, 20, and 25 dB. For each experimental condition, we conduct a Monte-Carlo (MC) simulation of 100 runs, each with  $N = 500$  data points. The performance of the methods is evaluated with respect to the accuracy and standard deviations in tracking the system parameters  $\theta = (a, b, c, d)$  and output.

More specifically, the quality of the estimates is evaluated by investigating the Parameter Estimation Errors (PEE), Variance Accounted For (VAF) on a validation data set which is different from the one used for model identification. The PEE and VAF are defined as follows:

$$PEE = \frac{\|\theta - \hat{\theta}\|_2}{\|\theta\|_2} 100\%$$

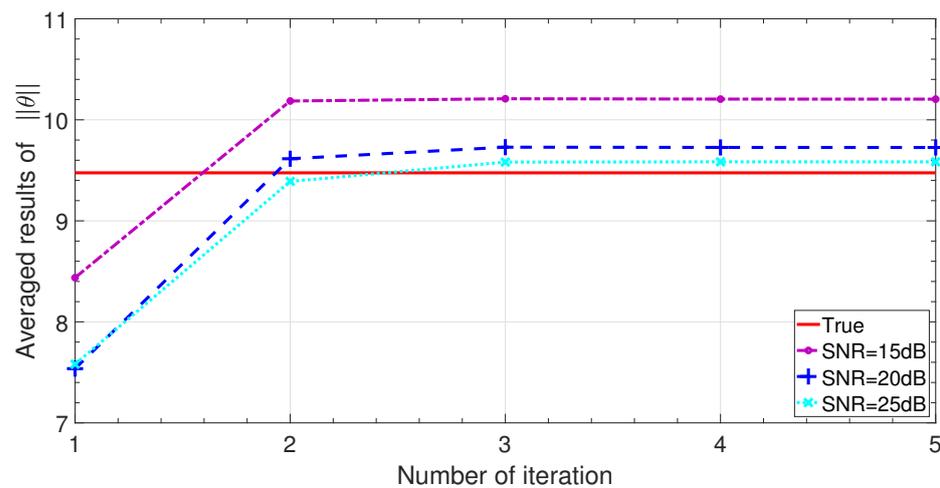
$$VAF = \max\left\{0, 1 - \frac{\text{var}(\bar{y}_k - \hat{y})}{\text{var}(\bar{y}_k)}\right\} 100\%$$

where the operator  $\text{var}(\cdot)$  denotes the variance of its argument,  $\bar{y}_k$  denotes the noise-free simulated output of the system, and  $\hat{y}_k$  denotes the the noise-free simulated model output. The VAF is allowed to reach 100%, and PPE is allowed to reach 0% when the model is equal to the true system.

Theoretically, the computational complexity of the proposed APM per iteration is  $O(N^2)$ , which is obviously more computationally demanding than  $O(N)$  in the TSA, because of the projection operations [17]. Fortunately, from the averaged estimation results of Euclidean norm  $\|\theta\|$  for 100 MC simulations at each iteration in APM under different SNRs shown in Figure 1, we observe that APM converges very quickly within 4 iterations for high SNRs (e.g., 20 and 25 dB), and also provides accurate estimation results even for a low SNR of 15 dB. It is concluded that the APM can provide accurate estimation results for highly nonlinear systems from a limited number of noisy samples. The estimated parameters in  $\theta$  along with the mean values and standard deviations and the PEE are listed in Table 1. From the results shown in Table 1, we can conclude that

**Table 1.** Estimated results of  $\theta$  and PEE.

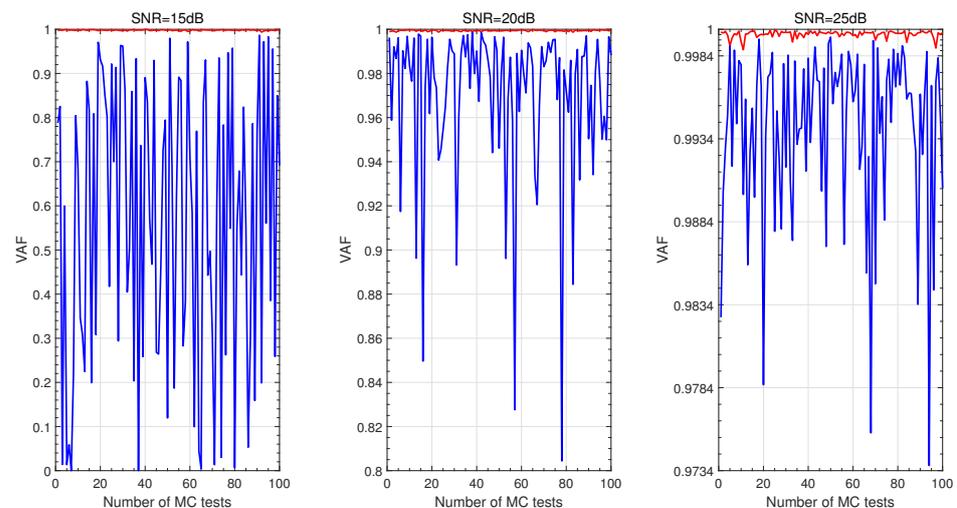
True	APM (SNR = 15)	APM (SNR = 20)	APM (SNR = 25)	TSA (SNR = 15)	TSA (SNR = 20)	TSA (SNR = 25)
$a_1 = 0.1569$	0.1604 ( $\pm 0.0793$ )	0.1654 ( $\pm 0.0617$ )	0.1477 ( $\pm 0.0423$ )	0.1643 ( $\pm 0.0840$ )	0.1669 ( $\pm 0.0597$ )	0.1470 ( $\pm 0.0467$ )
$a_2 = 0.0934$	0.0912 ( $\pm 0.0632$ )	0.0906 ( $\pm 0.0443$ )	0.0979 ( $\pm 0.0326$ )	0.0855 ( $\pm 0.0836$ )	0.0907 ( $\pm 0.0485$ )	0.0928 ( $\pm 0.0367$ )
$a_3 = 0.5477$	0.5442 ( $\pm 0.1472$ )	0.5480 ( $\pm 0.1187$ )	0.5465 ( $\pm 0.0277$ )	0.5162 ( $\pm 0.1961$ )	0.5407 ( $\pm 0.0413$ )	0.5474 ( $\pm 0.0280$ )
$a_4 = 0$	0.0009 ( $\pm 0.0500$ )	0.0032 ( $\pm 0.0357$ )	0.0002 ( $\pm 0.0267$ )	0.0014 ( $\pm 0.00826$ )	0.0041 ( $\pm 0.0413$ )	0.0026 ( $\pm 0.0333$ )
$a_5 = -0.7303$	-0.7218 ( $\pm 0.0481$ )	-0.7178 ( $\pm 0.1473$ )	-0.7271 ( $\pm 0.0252$ )	-0.6758 ( $\pm 0.2514$ )	-0.7317 ( $\pm 0.0419$ )	-0.7249 ( $\pm 0.0257$ )
$a_6 = -0.3651$	-0.3597 ( $\pm 0.07317$ )	-0.3478 ( $\pm 0.0839$ )	-0.3690 ( $\pm 0.0228$ )	-0.3164 ( $\pm 0.1460$ )	-0.3510 ( $\pm 0.0398$ )	-0.3713 ( $\pm 0.0267$ )
$b_1 = 0.1569$	0.1572 ( $\pm 0.0128$ )	0.1577 ( $\pm 0.0068$ )	0.1558 ( $\pm 0.0040$ )	0.2757 ( $\pm 0.2053$ )	0.1515 ( $\pm 0.0770$ )	0.1620 ( $\pm 0.0377$ )
$b_2 = 0.0934$	0.0930 ( $\pm 0.0121$ )	0.0931 ( $\pm 0.0060$ )	0.0939 ( $\pm 0.0038$ )	0.0039 ( $\pm 0.3265$ )	0.0692 ( $\pm 0.1037$ )	0.0913 ( $\pm 0.0367$ )
$b_3 = 0.5477$	0.5465 ( $\pm 0.0155$ )	0.5478 ( $\pm 0.0112$ )	0.5468 ( $\pm 0.0079$ )	0.1405 ( $\pm 0.4996$ )	0.4736 ( $\pm 0.2627$ )	0.5469 ( $\pm 0.0313$ )
$b_4 = 0.7303$	0.7293 ( $\pm 0.0114$ )	0.7306 ( $\pm 0.0074$ )	0.7299 ( $\pm 0.0050$ )	0.1432 ( $\pm 0.5629$ )	0.6178 ( $\pm 0.3892$ )	0.7220 ( $\pm 0.0511$ )
$b_5 = 0.3651$	0.3676 ( $\pm 0.0176$ )	0.3634 ( $\pm 0.0156$ )	0.3675 ( $\pm 0.0093$ )	0.1299 ( $\pm 0.4004$ )	0.3028 ( $\pm 0.2003$ )	0.3720 ( $\pm 0.1045$ )
$c_1 = 1$	1.0774 ( $\pm 0.2146$ )	0.9999 ( $\pm 0.1701$ )	1.0022 ( $\pm 0.0967$ )	0.2651 ( $\pm 0.8143$ )	0.8495 ( $\pm 0.5122$ )	0.9914 ( $\pm 0.3376$ )
$c_2 = 4$	4.0237 ( $\pm 0.8490$ )	3.9649 ( $\pm 0.4590$ )	4.0077 ( $\pm 0.2340$ )	1.1182 ( $\pm 3.0569$ )	3.3149 ( $\pm 1.7629$ )	3.9646 ( $\pm 0.4040$ )
$c_3 = 3$	2.6377 ( $\pm 1.3845$ )	2.9785 ( $\pm 0.6425$ )	3.0432 ( $\pm 0.3804$ )	0.9981 ( $\pm 2.5809$ )	2.5448 ( $\pm 1.6301$ )	3.0505 ( $\pm 0.3848$ )
$c_4 = 7$	7.1772 ( $\pm 1.6055$ )	7.0295 ( $\pm 1.4038$ )	7.0416 ( $\pm 0.7487$ )	1.9601 ( $\pm 10.3455$ )	6.1131 ( $\pm 4.6544$ )	7.1346 ( $\pm 0.8611$ )
$c_5 = 1$	1.0199 ( $\pm 1.3840$ )	1.0168 ( $\pm 0.5661$ )	0.9645 ( $\pm 0.3401$ )	0.0885 ( $\pm 1.6670$ )	0.8553 ( $\pm 0.7665$ )	0.9455 ( $\pm 0.3612$ )
$c_6 = 3.4$	3.5277 ( $\pm 1.3443$ )	3.4013 ( $\pm 1.0777$ )	3.3914 ( $\pm 0.5844$ )	0.7933 ( $\pm 5.1063$ )	2.7925 ( $\pm 1.8014$ )	3.2967 ( $\pm 0.6655$ )
$d_1 = 0.4$	0.5043 ( $\pm 0.1304$ )	0.4927 ( $\pm 0.1017$ )	0.5014 ( $\pm 0.0040$ )	0.4731 ( $\pm 0.1723$ )	0.5027 ( $\pm 0.0071$ )	0.5064 ( $\pm 0.0045$ )
$d_2 = 0.25$	0.2551 ( $\pm 0.0675$ )	0.2481 ( $\pm 0.0595$ )	0.2500 ( $\pm 0.0211$ )	0.2518 ( $\pm 0.1244$ )	0.2549 ( $\pm 0.0330$ )	0.2501 ( $\pm 0.0216$ )
PPE (%)	38.7442 ( $\pm 21.5544$ )	18.7887 ( $\pm 10.3743$ )	10.8245 ( $\pm 5.8194$ )	130.4982 ( $\pm 71.5442$ )	36.6713 ( $\pm 48.9617$ )	11.8245 ( $\pm 7.0916$ )



**Figure 1.** Averaged estimation results of  $\|\theta\|_2$  at each iteration for APM under different SNRs.

- Under the same noise levels, the proposed APM provides better estimation results in terms of accuracy and standard deviations than TSA.
- The two methods yield similar results for SNR = 25 dB (the corresponding PEEs are 10.8245 ( $\pm 5.8194$ ) and 11.8245 ( $\pm 7.0916$ ), respectively). The performance of APM is obviously better than the performance of TSA for  $c_2, c_4, c_6$ , PEE for SNR = 20 dB, and  $c_1, c_2, c_3, c_4, c_5, c_6$ , VAF, PPE for SNR = 15 dB. The attenuation of the signal levels does not obviously affect the APM while it deteriorates the performance of TSA significantly, showing that proposed APM is more robust to noise and data size.

The calculated VAFs for the proposed APM and TSA are given in Figure 2, while the corresponding histograms of the VAF values are shown in Figure 3. In Table 2 the mean values and standard deviations of VAF are given. The results show that the VAFs of APM are better than that of TSA under the same noise levels. The results verify again our argument: as the avoidance of introducing extra parameters and performing SVD to estimate the parameters of  $(a, b, c, d)$  from the products  $ad$  and  $bc$ , the proposed APM based on PPMs with fewer parameters can enhance the accuracy of the estimated quantities and decrease the variance of estimates, for the identification of highly nonlinear Hammerstein-Wiener systems given short noisy data length sequences.



**Figure 2.** VAF of APM (left, red line) and TSA (right, blue line) under different SNRs.

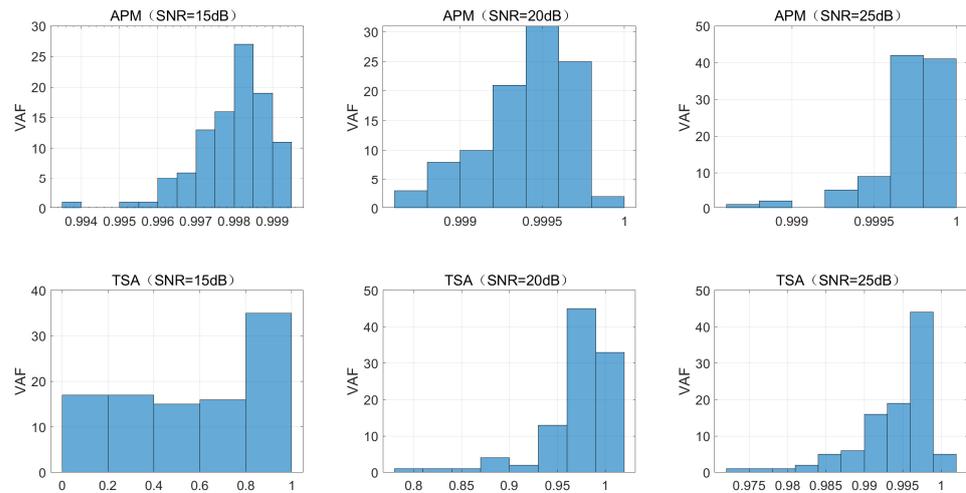


Figure 3. Histogram of VAF values of APM and TSA under different SNRs.

Table 2. Estimated Results of VAF under Different SNRs.

Methods	VAF
APM (SNR = 15 dB)	99.79 ( $\pm 0.0977$ )
APM (SNR = 20 dB)	99.94 ( $\pm 0.0270$ )
APM (SNR = 25 dB)	99.97 ( $\pm 0.0214$ )
TSA (SNR = 15 dB)	56.81 ( $\pm 31.1700$ )
TSA (SNR = 20 dB)	97.03 ( $\pm 3.5800$ )
TSA (SNR = 25 dB)	99.40 ( $\pm 0.5200$ )

To evaluate the sensitivity of the proposed APM under different sample lengths, the mean values of PEE and VAF for 100 Monte-Carlo simulations with respect to different  $N$  and SNRs using proposed APM and TSA are shown in Figures 4 and 5, respectively. It can be seen that the proposed method provides quite good performance when  $N > 100$ . The performance of the proposed APM in terms of PEE and VAF is better than the TSA especially for identifying N-L-N systems from a limited number of noisy data with a low SNR, see PEE and VAF when data point  $N < 1000$  and SNR = 25 dB. These results show that proposed APM is more robust to noise and data size.

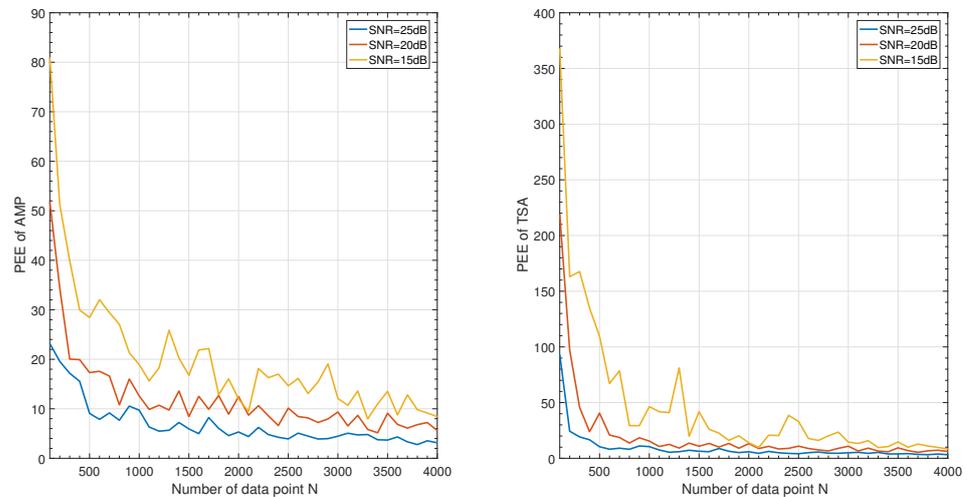
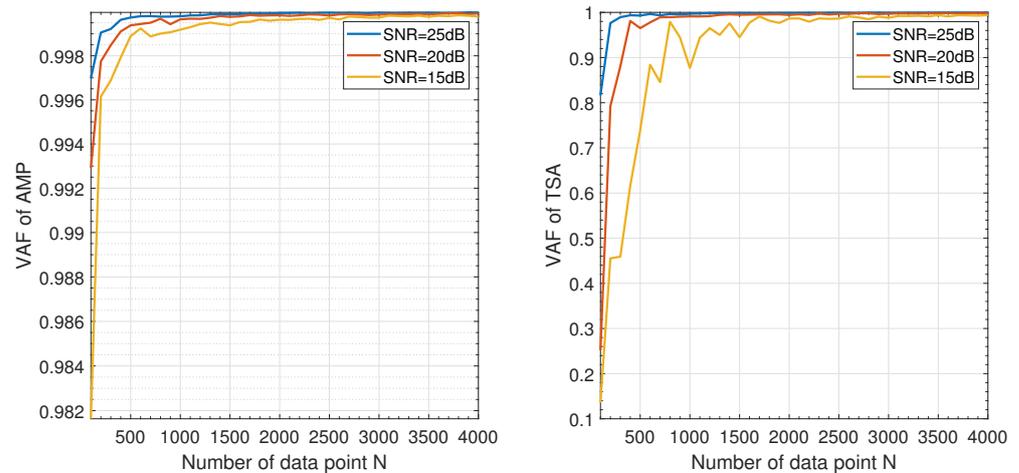


Figure 4. Averaged results of PEE for APM (left, red line) and TSA (right, blue line) under different  $N$  and SNRs, where  $N \in [100, 4000]$ .



**Figure 5.** Averaged results of VAF for APM (left, red line) and TSA (right, blue line) under different  $N$  and SNRs, where  $N \in [100, 4000]$ .

## 6. Conclusions

A new identification formulation for H-W models, referred to as APM, is presented. The avoidance of using over-parametrization models leads to a number of improvements, i.e., reducing the number of unknown parameters to be estimated and avoiding a truncated SVD step, so improving the estimated quantities in terms of accuracy and variance compared to TSA. These improvements especially hold for highly nonlinear H-W model identification from a limited number of noisy samples. The reasons why the new method results in an improvement model over the previous LS method have been demonstrated by variance analysis. The application to an illustrative example has well demonstrated the effectiveness and merit of the proposed method in comparison with TSA. Future work should focus on identifying other nonlinear systems by taking into account some additional prior knowledge about the systems.

**Author Contributions:** Conceptualization, L.S. and J.H.; methodology, L.S. and J.H.; software, L.S. and J.H.; validation, L.S. and J.H.; formal analysis, L.S. and J.H.; investigation, L.S. and J.H.; writing—original draft preparation, L.S. and J.H.; writing—review and editing, J.H. and C.X.; visualization, J.H., Z.F. and C.X.; supervision, J.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the Chongqing Natural Science Foundation under Grant CSTB2022NSCQ-MSX1225; in part by the China Postdoctoral Science Foundation under Grant 2022MD713688; in part by the Chongqing Postdoctoral Science Foundation under Grant 2021XM3079; in part by the Science and Technology Research Program of Chongqing Municipal Education Commission under Grant KJQN202000602.

**Institutional Review Board Statement:** This study does not involve any institutional review issues.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data is given by The system model of the benchmark example is presented in Case Studies in Section 3 and simulated by matlab, which can be obtained by emailing jiehou.phd@hotmail.com.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

H-W	Hammerstein–Wiener
N-L-N	Nonlinear-Linear-Nonlinear
BELS	Bias-Eliminating Least-Squares
TSA	Two-Step Algorithms
OPM	over-parametrization model
SVDs	Singular Value Decompositions
AM-MISG	auxiliary model multi-innovation stochastic gradient
APM	Asynchronous Parsimonious Method
TSA	Two-Step Algorithms
SNR	Signal-to-Noise Ratio
PEE	Parameter Estimation Errors
VAF	Variance Accounted For

## References

- Schoukens, M.; Tiels, K. Identification of block-oriented nonlinear systems starting from linear approximations: A survey. *Automatica* **2017**, *85*, 272–292. [[CrossRef](#)]
- Wang, D.; Ding, F. Hierarchical least squares estimation algorithm for Hammerstein-Wiener systems. *IEEE Signal Process. Lett.* **2012**, *19*, 825–828. [[CrossRef](#)]
- Li, F.; Jia, L. Parameter estimation of Hammerstein-Wiener nonlinear system with noise using special test signals. *Neurocomputing* **2019**, *344*, 37–48. [[CrossRef](#)]
- Bai, E. An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems. *Automatica* **1998**, *34*, 333–338. [[CrossRef](#)]
- Xu, L.; Ding, F.; Yang, E.F. Auxiliary model multiinnovation stochastic gradient parameter estimation methods for nonlinear sandwich systems. *Int. J. Robust Nonlinear Control* **2020**, *31*, 148–165. [[CrossRef](#)]
- Ni, B.; Gilson, M.; Garnier, H. Refined instrumental variable method for Hammerstein-Wiener continuous-time model identification. *IET Control Theory Appl.* **2013**, *7*, 1276–1286. [[CrossRef](#)]
- Wang, Z.; Wang, Y.; Ji, Z. A novel two-stage estimation algorithm for nonlinear Hammerstein-Wiener systems from noisy input and output data. *J. Frankl. Inst.* **2017**, *354*, 1937–1944. [[CrossRef](#)]
- Wang, J.; Zhang, Q.; Ljung, L. Revisiting Hammerstein system identification through the two-stage algorithm for bilinear parameter estimation. *Automatica* **2009**, *45*, 2627–2633. [[CrossRef](#)]
- Hou, J.; Su, H.; Yu, C.; Chen, F.; Li, P. Bias-Correction Errors-in-Variables Hammerstein Model Identification. *IEEE Trans. Ind. Electron.* **2022**. [[CrossRef](#)]
- Gomez, J.; Baeyens, E. Identification of block-oriented nonlinear systems using orthonormal bases. *J. Process Control* **2004**, *14*, 685–697. [[CrossRef](#)]
- Hou, J.; Chen, F.; Li, P.; Zhu, Z. Fixed point iteration-based subspace identification of Hammerstein state-space models. *IET Control Theory Appl.* **2019**, *13*, 1173–1181. [[CrossRef](#)]
- Hou, J.; Chen, F.; Li, P.; Shu, J.; Zhao, F. Recursive Parsimonious Subspace Identification for Closed-Loop Hammerstein Nonlinear Systems. *IEEE Access* **2019**, *7*, 173515–17352. [[CrossRef](#)]
- Hou, J.; Chen, F.; Li, P.; Zhu, Z. Gray-Box Parsimonious Subspace Identification of Hammerstein-Type Systems. *IEEE Trans. Ind. Electron.* **2021**, *68*, 9941–9951. [[CrossRef](#)]
- Hou, J. Parsimonious model based consistent subspace identification of Hammerstein systems under periodic disturbances. *Int. J. Control. Autom. Syst.* **2022**. [[CrossRef](#)]
- Ding, F. Decomposition based fast least squares algorithm for output error systems. *Signal Process.* **2013**, *93*, 1235–1242. [[CrossRef](#)]
- Hou, J.; Su, H.; Yu, C.P.; Chen, F.; Li, P.; Xie, H.; Li, T. Consistent Subspace Identification of Errors-In-Variables Hammerstein Systems. *IEEE Trans. Syst. Man Cybern. Syst.* **2022**. [[CrossRef](#)]
- Ding, F. Computational efficiency of the identification methods. PartB: Iterative algorithm. *J. Nanjing Univ. Inf. Sci. Technol. Nat. Sci. Ed.* **2021**, *4*, 385–401.