A Blind Approach to Hammerstein Model Identification

Er-Wei Bai, *Senior Member, IEEE,* and Minyue Fu, *Senior Member, IEEE,*

***Abstract*—This paper discusses the Hammerstein model identification using a blind approach. By fast sampling at the output, it is shown that identification of the linear part can be achieved based only on the output measurements that makes the Hammerstein model identification possible without knowing the structure of the nonlinearity and the internal variables.**

***Index Terms*—Hammerstein systems, nonlinear systems, parameter estimation, system identification.**

I. Introduction

THE Hammerstein model is a special kind of nonlinear system where a nonlinear block is followed by a linear system. The Hammerstein model has applications in many engineering problems including control, signal processing and communication [8], [10], [14]. For instance, the Hammerstein model finds applications in modeling distortion in nonlinearly amplified digital communication signals (satellite and microwave links) followed by a linear channel [8], [14]. Therefore, the Hammerstein model identification has been an active research area for many years [4], [8], [14] in the control and signal processing communities. There exists a large number of research papers on the topic of the Hammerstein model identification in the literature. Existing methods can be roughly divided into four categories:

1) iterative method [12], [15], [19];
2) overparameterization method [2], [6], [7], [9];
3) stochastic method [5], [8], [13];
4) separable least squares method [1], [16].

The idea of the iterative method [12], [15], [19] is the alternative estimation of parameters. Although there are some variations, the parameter set is usually divided into two subsets. One finds the optimal values for the first set while the second set is fixed. Then, two sets are switched to find the optimal value for the second set while the first one is fixed. This method can often provide good results. However, convergence is a problem. In fact, it was shown in [17] that the method can be divergent. Some modifications were proposed in recent years [15] to overcome this problem. The overparameterization method [2], [6], [7], [9] is to overparameterize the Hammerstein system so that the overparameterized system is linear in the unknown parameters, and then, any linear estimation algorithm applies. The difficulty with this approach is that the dimension of the resulting linear system can be very large, and therefore, convergence or robustness becomes an issue. In general, this approach is limited to the Hammerstein system where the unknown nonlinear block is parameterized linearly by unknown parameters. The stochastic method [5], [8], [13] uses white noise properties to separate the nonlinear part from the linear part and works only if the input is white. The idea of the separable least squares [1], [16] is to write one set of variables as a function of the other set based on the first-order necessary and sufficient conditions. Thus, the dimension of the optimization space is reduced. This method is found particularly useful for hard or nonsmooth nonlinearities [1].

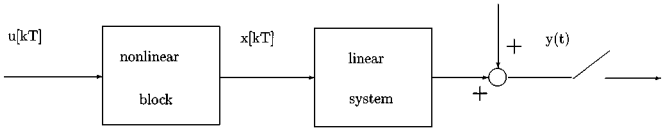

Fig. 1. Sampled Hammerstein system.

In this paper, we consider identification of a discrete-time Hammerstein system. We will mainly focus our study on sampled Hammerstein systems, as shown in Fig. 1, but we will also extend our results to nonsampled discrete-time Hammerstein systems.

A discrete-time Hammerstein system is shown in Fig. 1. The goal of the Hammerstein system identification is to estimate the transfer function of the equivalent sampled linear system for the given sampling interval T and to estimate the unknown nonlinear function f based only on the measurement of u and y. The internal signal x is not available. The order of the linear system is known *a priori*.

Our approach in this paper is different from all four methods discussed above and is based on the idea of our previous work on blind system identification [3]. We identify the linear part using the output measurements only, i.e., no information on the input $u[kT]$ and the interval variable $x[kT]$ are needed. In general, blind system identification is not possible only based on the output measurements because different systems coupled with appropriate inputs can produce identical outputs at the sampling instants kT. However, by fast sampling at the output, blind identification based on the output measurements is possible. Once the linear part is obtained, identification of the nonlinear part can be carried out in a number of ways.

We remark that the blind system identification techniques for IIR system were first developed in our early work [3] for linear systems. The use of fast sampling in identification of sampled

Manuscript received March 27, 2001; revised March 11, 2002. This work was supported in part by NSF ECS-9710297 and ECS-0098181. The associate editor coordinating the review of this paper and approving it for publication was Prof. Bjorn Ottersten.

E.-W. Bai is with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242 USA (e-mail: erwei@engineering.uiowa.edu).

M. Fu is with the Department of Electrical and Computer Engineering, The University of New Castle, Callaghan, Australia.

Publisher Item Identifier S 1053-587X(02)05648-9.

Hammerstein models was studied in a recent paper [18]. For a given sampling interval T, the transfer function at the sampling interval $T/(n+1)$, where n is the order of the system, was derived. However, it is not clear in [18] whether the information of the transfer function at $T/(n+1)$ is enough to derive the transfer function at T. In the current paper, this problem is completely solved. We show that the transfer function at the given sampling interval T can be identified based only on the output observations. The current paper also contains two additional minor contributions. The first one is that our proposed algorithm applies to a wide range of inputs, and moreover, the persistent excitation condition that guarantees convergence and robustness is obtained, whereas in [18], the input is restricted to white noises. The second minor extension is that [18] deals with control systems where the input is piecewise constant, and the current paper is in the digital signal processing setting where the input is a discrete pulse sequence.

The outline of the paper is as follows. Section II establishes some preliminary results. Some assumptions on the system are also given in this section. Section III studies blind identification of the linear part and convergence issues. Section IV devotes to identification of the nonlinear block and several methods are proposed. A numerical simulation is provided in Section V. Extension to nonsampled discrete Hammerstein systems is provided in Section VI. Section VII gives some final remarks.

II. Problem Statement and Preliminaries

As mentioned in the previous section, we will focus on the sampled Hammerstein system first. Extension to nonsampled Hammerstein systems will be made in Section VI. Consider a sampled Hammerstein model in Fig. 1, which consists of a nonlinear block and a continuous linear time-invariant system. For a given sampling interval T, the input $u[kT]$ is a discrete pulse sequence. The output of the nonlinear block x, which is the input to the linear system, is also a discrete sequence

$$\begin{aligned} x(t) =& 0,\ t \neq kT \\ x[kT] =& f\left(u[kT], \ldots, u[(k-m)T], \gamma\right) \end{aligned} \tag{1}$$

where f is a nonlinear function with known or unknown structure parameterized by an unknown parameter vector $\gamma \in R^l$. The model of (1) covers a large class of nonlinear functions. The most common static nonlinear model $x[kT] = f(u[kT])$ in the Hammerstein representation is obviously a special case of (1). Some nonlinearities with memory, e.g., hysteresis, also belong to (1). In such a case, the memory length m is assumed to be known.

Let the continuous linear time-invariant system be represented by a state-space equation

$$\begin{aligned} \dot{w}(t) =& Aw(t) + bx(t),\ x, y \in R,\ A \in R^{n\times n} \\ y(t) =& cw(t). \end{aligned} \tag{2}$$

It is a routine exercise to derive its equivalent discrete time equation for a given sampling interval T when the input is a discrete pulse sequence [3]

$$\begin{aligned} w[(k+1)T] =& \Phi w[kT] + \Gamma x[kT] \\ y[kT] =& cw[kT]. \end{aligned} \tag{3}$$

The discrete transfer function from $x[kT]$ to $y[kT]$ is, accordingly, given by

$$\begin{aligned} G(z) =& c(zI-\Phi)^{-1}\Gamma = \frac{\beta(z)}{\alpha(z)} \\ =& \frac{b_1 z^{-1} + b_2 z^{-2} + \cdots + b_n z^{-n}}{1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_n z^{-n}} \end{aligned} \tag{4}$$

form some a_is and b_js. The goal of the Hammerstein system identification is to estimate $G(z)$ in terms of its parameters a_is and b_js, as well as the unknown nonlinear function f based only on the measurement of u and y.

We now make an assumption on the sampled system (3) throughout the paper.

Assumption 1: It is assumed that the sampled system (3) at the sampling interval T is minimal (reachable and observable).

The following lemma can be easily verified [3].

Lemma II.1: Consider the continuous system (2) and the sampled system (3). Then, we have the following.

- The sampled system is minimal at the sampling interval $h = T/p$ for some positive integer $p \geq 1 \Longleftrightarrow Im(\lambda_i - \lambda_j) \neq 2k\pi/h$, $k = 0, \pm 1, \pm 2, \ldots$ whenever $Re(\lambda_i - \lambda_j) = 0$, where λ_is are the eigenvalues of the continuous system.
- The sampled system is minimal at any sampling interval $h = T/p, p \geq 1$ if it is minimal at the sampling interval T.

Our approach in this paper is based on blind system identification, i.e., to estimate $G(z)$ using only the output measurements. The idea is fast sampling at the output that results in a sampled system at a higher sampling rate or a smaller sampling interval. Let the output sampling interval be

$$h = \frac{T}{p},\ p \geq 1$$

for some positive integer p, which is referred to as the oversampling ratio. For given T and p, consider the following sequences:

$$\begin{aligned} \{y[kh]\} \Longleftrightarrow Y_p(z) &= \sum_{k=0}^{\infty} y[kh] z^{-k} \\ \{y[kT]\} \Longleftrightarrow Y(z) &= \sum_{k=0}^{\infty} y[kT] z^{-k} \\ \{y[kT+h]\} \Longleftrightarrow Y_{T,1}(z) &= \sum_{k=0}^{\infty} y[kT+h] z^{-k} \\ \{y[kT+2h]\} \Longleftrightarrow Y_{T,2}(z) &= \sum_{k=0}^{\infty} y[kT+2h] z^{-k} \\ &\cdots \\ \{y[kT+(p-1)h]\} \Longleftrightarrow Y_{T,p-1}(z) &= \sum_{k=0}^{\infty} y[kT+(p-1)h] z^{-k}. \end{aligned} \tag{5}$$

Although the input sampling interval is fixed at T and

$$\begin{aligned} \{u[kT]\} \Longleftrightarrow U(z) &= \sum_{k=0}^{\infty} u[kT] z^{-k} \\ \{x[kT]\} \Longleftrightarrow X(z) &= \sum_{k=0}^{\infty} x[kT] z^{-k} \end{aligned} \tag{6}$$

we can write $X(z)$ in terms of the output sampling interval h:

$$\begin{aligned}X_p(z)=&\sum_{k=0}^{\infty}x[kh]z^{-k}\\ =&x[0]+x[h]z^{-1}+\cdots+x[T-h]z^{-(p-1)}+x[T]z^{-p}\\ &+x[T+h]z^{-(p+1)}+\cdots+x[2T-h]z^{-(2p-1)}+\cdots\\ =&\sum_{k=0}^{\infty}x[kT]z^{-pk}=X\left(z^p\right).\end{aligned}\tag{7}$$

Denote by $G_p(z)$ the discrete transfer function from $x[kh]$ to $y[kh]$ and by $G_{T,i}(z)$, $0\le i\le(p-1)$ the discrete transfer functions from $x[kT]$ to $y[kT+ih]$, i.e.,

$$Y_p(z)=G_p(z)X_p(z),Y_{T,i}(z)=G_{T,i}(z)X(z).$$

The transfer functions of $G_{T,i}(z)$ can be easily derived [3]:

$$\begin{aligned}w[(k+1)T+ih]=&\Phi w[kT+ih]+\Gamma_{i1}x[kT]\\ &+\Gamma_{i2}x[(k+1)T]\\ y[kT+ih]=&cw[kT+ih]\end{aligned}$$

where $\Gamma_{i1}=0, 1\le i\le p-1$, and $\Gamma_{i2}=0,\ i=0$. Thus, the transfer function $G_{T,i}(z)$ is given by

$$\begin{aligned}G_{T,i}(z)=&\frac{\beta_i(z)}{\alpha_i(z)}=c(zI-\Phi)^{-1}\left(\Gamma_{i1}+\Gamma_{i2}z\right)\\ =&\frac{b_{i0}+b_{i1}z^{-1}+\cdots+b_{in}z^{-n}}{1-a_{i1}z^{-1}-a_{i2}z^{-2}-\cdots-a_{in}z^{-n}}.\end{aligned}\tag{8}$$

It is interesting to note the following.

- All $G_{T,i}(z)$s share the same denominator, i.e.,

  $$a_{i1}=a_1,a_{i2}=a_2,\ldots,a_{in}=a_n,\ 0\le i\le p-1.$$

- $G_{T,0}(z)=G(z)$ as in (4), and this implies

  $$b_{00}=0,b_{01}=b_1,\ldots,b_{0n}=b_n.$$

- $G_{T,0}(z)=G(z)$ is strictly proper and $G_{T,i}(z)$s, $i=1,\ldots,p-1$ are proper but not strictly proper. Moreover, $b_{in}=0$, and $i=1,\ldots,p-1$.

It will be shown later that by fast sampling at the output, $G_p(z)$ can be identified based only on the output measurements. The difficulty is that $G_p(z)$ is the transfer function at the sampling interval $h=T/p$, which is not the desired transfer function $G(z)$ at the sampling interval T. Thus, we have an identifiability problem, i.e., how to find $G(z)$ from $G_p(z)$. We have the following result.

Lemma II.2: Let $T=ph$ for some integer $p\ge1$. Suppose the transfer function $G_p(z)$ at the sampling interval $h=T/p$ is in the form of

$$G_p(z)=\frac{\beta_p(z)}{\alpha_p(z)}=\frac{\beta_p(z)}{(1-s_1z^{-1})(1-s_2z^{-1})\cdots(1-s_nz^{-1})}.$$

Write

$$\beta_p(z)\prod_{k=1}^{n}\left(\sum_{m=0}^{p-1}s_k^mz^{-m}\right)=\sum_{k=1}^{(n+1)p-1}q_kz^{-k}.$$

Then, the transfer function $G(z)$ at the sampling interval T is given by

$$G(z)=\frac{\beta(z)}{\alpha(z)}=\frac{q_pz^{-1}+q_{2p}z^{-2}+\cdots+q_{np}z^{-n}}{(1-s_1^pz^{-1})(1-s_2^pz^{-1})\cdots(1-s_n^pz^{-1})}.\tag{9}$$

Proof: Under the minimality assumption, it is clear that s is a pole of the continuous time system, $\Longleftrightarrow e^{sh}$ is a pole of $G_p(z)$, and e^{sT} is a pole of $G(z)$. This implies that the denominator of $G(z)$ is

$$\begin{aligned}\alpha(z)=&\left(1-s_1^pz^{-1}\right)\left(1-s_2^pz^{-1}\right)\cdots\left(1-s_n^pz^{-1}\right)\\ =&\prod_{k=1}^{n}\left(1-s_k^pz^{-1}\right).\end{aligned}$$

To determine the numerator $\beta(z)$, note

$$\begin{aligned}\alpha\left(z^p\right)=&\prod_{k=1}^{n}\left(1-s_k^pz^{-p}\right)\\ =&\prod_{k=1}^{n}\left(1-s_kz^{-1}\right)\left(1+s_kz^{-1}+\cdots+s_k^{p-1}z^{-(p-1)}\right)\\ =&\alpha_p(z)\prod_{k=1}^{n}\left(1+s_kz^{-1}+s_k^2z^{-2}+\cdots+s_k^{p-1}z^{-(p-1)}\right)\\ =&\alpha_p(z)\prod_{k=1}^{n}\left(\sum_{m=0}^{p-1}s_k^mz^{-m}\right)\end{aligned}\tag{10}$$

and recall

$$Y_p(z)=G_p(z)X_p(z),Y_{T,i}(z)=G_{T,i}(z)X(z).$$

In addition, we have the equation shown at the bottom of the page. The last equality is from (7). On the other hand, $Y_p(z)=G_p(z)X_p(z)$, and this implies

$$\frac{\beta_p(z)}{\alpha_p(z)}=\frac{\beta\left(z^p\right)}{\alpha\left(z^p\right)}+z^{-1}\frac{\beta_1\left(z^p\right)}{\alpha_1\left(z^p\right)}+\cdots+z^{-(p-1)}\frac{\beta_{p-1}\left(z^p\right)}{\alpha_{p-1}\left(z^p\right)}.$$

$$\begin{aligned}Y_p(z)=&y[0]+y[h]z^{-1}+\cdots+y[(p-1)h]z^{-(p-1)}\\ &+y[T]z^{-p}+y[T+h]z^{-(p+1)}+\cdots+y[T+(p-1)h]z^{-(2p-1)}+y[2T]z^{-2p}+\cdots\\ =&Y\left(z^p\right)+z^{-1}Y_{T,1}\left(z^p\right)+\cdots+z^{-(p-1)}Y_{T,p-1}\left(z^p\right)\\ =&\left(G\left(z^p\right)+z^{-1}G_{T,1}\left(z^p\right)+\cdots+z^{-(p-1)}G_{T,p-1}\left(z^p\right)\right)X\left(z^p\right)\\ =&\left(G\left(z^p\right)+z^{-1}G_{T,1}\left(z^p\right)+\cdots+z^{-(p-1)}G_{T,p-1}\left(z^p\right)\right)X_p(z).\end{aligned}$$

Now, from (10) and the fact that all $\alpha(z)$ and $\alpha_i(z)$ are the same, we have

$$\beta_p(z)\prod_{k=1}^{n}\left(\sum_{m=0}^{p-1}s_k^m z^{-m}\right)=\beta(z^p)+z^{-1}\beta_1(z^p)+\cdots+z^{-(p-1)}\beta_{p-1}(z^p).$$

This completes the proof.

Before closing this section, we observe that the parameterization of the Hammerstein model is actually not unique. Suppose the nonlinear block and the linear block are represented by some function f and the transfer function $G(z)$, respectively. Then, any pair of $c\cdot f$ and $c^{-1}G(z)$ for some nonzero constant c would produce the identical input–output measurements. In other words, any identification setting cannot distinguish between $(f, G(z))$ and $(cf, c^{-1}G(z))$. To obtain a unique parameterization, $G(z)$ needs to be normalized, e.g., set $b_1=1$. The problem with this approach is that it indirectly presumes $b_1\neq 0$, which may not be the case. To avoid this problem, we assume the following assumption throughout the paper.

Assumption 2: Consider $G(z)$ of (4). Assume that $\|(b_1,b_2,\ldots,b_n)\|_2=1$ and that the sign of the first nonzero element of b_i is positive.

III. Identification of the Linear Block $G(z)$

In this section, we will provide an algorithm for estimating $G(z)$ based only on the output measurements. The idea of our approach is to estimate $G_{n+1}(z)$ first and then to compute $G(z)$ using the result of Lemma II.2. To avoid unnecessary complications, we assume in Sections III-A and B that noise $v(\cdot)=0$. The convergence of the algorithm in the presence of noise will be discussed in Section III-D.

A. Estimation of the Denominator $\alpha_{n+1}(z)$

Given the input sampling interval T, let the output sampling interval be $h=T/(n+1)$. Write $G_{n+1}(z)$ as

$$G_{n+1}(z)=\frac{\beta_{n+1}(z)}{\alpha_{n+1}(z)}=\frac{\bar{b}_1z^{-1}+\bar{b}_2z^{-2}+\cdots+\bar{b}_nz^{-n}}{1-\bar{a}_1z^{-1}-\bar{a}_2z^{-2}-\cdots-\bar{a}_nz^{-n}}$$

for some unknown $\bar{b}_i$s and $\bar{a}_j$s. Its time domain expression is given by

$$y[kh]=\sum_{i=1}^{n}\bar{a}_iy[(k-i)h]+\sum_{j=1}^{n}\bar{b}_jx[(k-j)h].$$

The input sequence $x(t)$ is nonzero only if $t=kT=k(n+1)h$. In other words, $x[kh]$ is nonzero only if $k=l(n+1)$, and moreover

$$x[(l(n+1)-1)h]=\cdots=x[(l(n+1)-n)h]=0.$$

Now, at the sampling instants $kh=l(n+1)$, we have

$$y[l(n+1)h]=\sum_{i=1}^{n}\bar{a}_iy[(l(n+1)-i)h]. \tag{11}$$

Define

$$\phi_1'[l]=(y[lT-h],y[lT-2h],\ldots,y[lT-nh])$$
$$\bar{a}'=(\bar{a}_1,\bar{a}_2,\ldots,\bar{a}_n).$$

We have

$$y[lT]=\phi_1'[l]\bar{a}. \tag{12}$$

This equation is linear in the unknown $\bar{a}$. All other variables $y[lT]$ and $\phi_1'[l]$ consist of output measurements only, and therefore, the denominator coefficients $\bar{a}$ can be estimated by many standard algorithms, e.g., the recursive LMS or recursive least squares method.

B. Estimation of the Numerator $\beta_{n+1}(z)$

To estimate the numerator $\beta_{n+1}(z)$ of $G_{n+1}(z)$ at the sampling interval $h=T/(n+1)$, consider two sequences

$$\{y[kh]\}\Longleftrightarrow Y_{n+1}(z)=\sum_{k=1}^{\infty}y[kh]z^{-k}=G_{n+1}(z)X_{n+1}(z)$$
$$\left\{y\left[kh+\frac{h}{2}\right]\right\}\Longleftrightarrow \tilde{Y}_{n+1}(z)=\sum_{k=1}^{\infty}y\left[kh+\frac{h}{2}\right]z^{-k}=\tilde{G}_{n+1}(z)X_{n+1}(z)$$

where $X_{n+1}(z)=\sum_{k=0}^{\infty}x[kh]=X(z^{n+1})$. As discussed in (4) and (8), $G_{n+1}(z)$ is strictly proper, and $\tilde{G}_{n+1}(z)$ is proper but not strictly proper. In addition, $G_{n+1}(z)$ and $\tilde{G}_{n+1}(z)$ share the same denominator. Hence

$$\tilde{G}_{n+1}(z)=\frac{\tilde{b}_0+\tilde{b}_1z^{-1}+\cdots+\tilde{b}_{n-1}z^{-(n-1)}}{1-\bar{a}_1z^{-1}-\cdots-\bar{a}_nz^{-n}}=\frac{\tilde{\beta}_{n+1}(z)}{\alpha_{n+1}(z)}.$$

Clearly

$$\tilde{G}_{n+1}(z)Y_{n+1}(z)-G_{n+1}(z)\tilde{Y}_{n+1}(z)=0$$

and this implies

$$\tilde{\beta}_{n+1}(z)Y_{n+1}(z)-\beta_{n+1}\tilde{Y}_{n+1}(z)=0$$

or in the time domain

$$\left(y[kh],\ldots,y[kh-(n-1)h],-y\left[kh-h+\frac{h}{2}\right],\ldots,-y\left[kh-nh+\frac{h}{2}\right]\right)\cdot\left(\tilde{b}_0,\ldots,\tilde{b}_{n-1},\bar{b}_1,\ldots,\bar{b}_n\right)'=0,\ k=1,2,\ldots$$

Again, from (8), $\tilde{G}_{n+1}(z)$ is proper but not strictly proper, and this implies $\tilde{b}_0\neq 0$. Define

$$\phi_2'[k]=\left(y\left[kh-h+\frac{h}{2}\right],\ldots,y\left[kh-nh+\frac{h}{2}\right],-y[kh-h],\ldots,-y[kh-(n-1)h]\right)$$
$$\bar{b}'=\frac{\left(\bar{b}_1,\ldots,\bar{b}_n,\tilde{b}_1,\ldots,\tilde{b}_{n-1}\right)}{\tilde{b}_0}.$$

It follows that

$$y[kh]=\phi_2'[k]\bar{b}. \tag{13}$$

This is again linear in the unknown variable $\bar{b}$, and all other variables $y[kh]$ and $\phi_2'[k]$ are available. The unknown numerator coefficients $\bar{b}$ can thus be estimated by any standard algorithm.

C. Algorithm for Estimating $G(z)$

Now, we are in a position to provide the algorithm estimating $G(z)$ based only on the output measurements.

Blind Identification Algorithm for Estimating $G(z)$:

1) Given input sampling interval T. Set $h = T/(n+1)$.
2) Sample and collect output measurements $\{y[kh]\}$ and $\{y[kh+h/2]\}$. Define

$$\phi_1'[l] = (y[l(n+1)h-h], \ldots, y[l(n+1)h-nh])$$

$$\phi_2'[k] = \left(-y[kh-h], \ldots, -y[kh-(n-1)h], y\left[kh-h+\frac{h}{2}\right], \ldots, y\left[kh-nh+\frac{h}{2}\right]\right).$$

3) At each k, apply either the recursive least squares or the recursive LMS algorithm to estimate $\bar{b}$ using (13), for instance, the recursive LMS-type algorithm

$$\hat{\bar{b}}[k] = \hat{\bar{b}}[k-1] + \frac{\phi_2[k]}{1+\phi_2'[k]\phi_2[k]}\left(y[kh] - \phi_2'[k]\hat{\bar{b}}[k-1]\right) \tag{14}$$

where $\hat{\bar{b}}[k] = (\hat{\bar{b}}_1, \ldots, \hat{\bar{b}}_{2n-1})' \in R^{2n-1}$ is the estimate of $\bar{b}$ at time k, and at each $k = l(n+1)$, $l = 0, 1, 2, \ldots$, apply either the recursive least squares or the recursive LMS algorithm to estimate $\bar{a}$ using (12), for instance, the recursive LMS-type algorithm

$$\hat{\bar{a}}[l] = \hat{\bar{a}}[l-1] + \frac{\phi_1[l]}{1+\phi_1'[l]\phi_1[l]}\left(y[lT] - \phi_1'[l]\hat{\bar{a}}[l-1]\right) \tag{15}$$

where $\hat{\bar{a}}[k] = (\hat{\bar{a}}_1, \ldots, \hat{\bar{a}}_n)' \in R^n$ is the estimate of $\bar{a}$ at time $k = l(n+1)$. The estimate $\hat{G}_{n+1}(z)$ of $G_{n+1}(z)$ is defined as

$$\hat{G}_{n+1}(z) = \frac{\hat{\bar{b}}_1 z^{-1} + \hat{\bar{b}}_2 z^{-2} + \cdots + \hat{\bar{b}}_n z^{-n}}{1 - \hat{\bar{a}}_1 z^{-1} - \hat{\bar{a}}_2 z^{-2} - \cdots - \hat{\bar{a}}_n z^{-n}}.$$

4) By using Lemma II.2, compute the estimate $\hat{G}(z)$ of $G(z)$, which is the transfer function at the sampling interval T, in terms of its coefficient estimates $\hat{a} = (\hat{a}_1, \ldots, \hat{a}_n)'$ and $\hat{b} = (\hat{b}_1, \ldots, \hat{b}_n)'$ based on $\hat{G}_{n+1}(z)$. Because $\|b\|_2 = 1$ and the first nonzero element of b is positive, we normalize $\hat{b}$ by

$$\hat{b} = \frac{\hat{b}}{\left\|\hat{b}\right\|_2}$$

and set $\hat{b} = -\hat{b}$ if the first nonzero element of $\hat{b}$ is negative. Finally, the estimate $\hat{G}(z)$ of $G(z)$, at time k, is obtained.

$$\hat{G}(z) = \frac{\hat{b}_1 z^{-1} + \hat{b}_2 z^{-2} + \cdots + \hat{b}_n z^{-n}}{1 - \hat{a}_1 z^{-1} - \hat{a}_2 z^{-2} - \cdots - \hat{a}_n z^{-n}}.$$

5) Set $k = k+1$, and go to Step 2.

Only output measurements are needed to implement the algorithm. The algorithm is recursive and produces the estimate $\hat{G}(z)$ at each k.

D. Convergence Analysis

Whether $\hat{G}(z)$ converges to $G(z)$ depends on whether $\hat{G}_{n+1}(z)$ converges to $G_{n+1}(z)$. Therefore, it boils down to the parameter convergence of $\hat{\bar{a}}_i \to \bar{a}_i$ and $\hat{\bar{b}}_j \to \bar{b}_j$. It is well known that both parameter estimates converge asymptotically if $\phi_1[l]$ and $\phi_2[k]$ are persistently excited (PE) at least in the absence of noise. Therefore, it is important to establish the PE condition on $\phi_1[l]$ and $\phi_2[k]$. In fact, the PE condition on $\phi_2[k]$ has been developed in our early work on the subject of blind system identification [3]. However, the PE condition on $\phi_1[l]$ is new.

Lemma III.1: Consider the parameter estimation algorithms (14) and (15). Then, we have the following.

- Suppose the spectral measure of $x[lT]$ is not concentrated on $m < n$ points. Then, $\phi_1[l]$ is PE.
- Suppose the numerators of $G_{n+1}(z)$ and $\tilde{G}_{n+1}(z)$ do not share any common factor and that the spectral measure of $x[kT]$ is not concentrated on $m < 2n$ points. Then, $\phi_2[k]$ is PE.

Proof: The second part has been shown in [3]. We only provide the proof for the first part. By a simple calculation, we have

$$\phi_1[l] = \bar{A}^p \phi_1[l-1] + \bar{g}x[lT]$$

where

$$\bar{A} = \begin{pmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_{n-1} & \alpha_n \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}$$

and

$$\bar{g} = \bar{A}^{n-1}\begin{pmatrix} b_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \bar{A}^{n-2}\begin{pmatrix} b_2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \bar{A}\begin{pmatrix} b_{n-1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} b_n \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

From Assumption 1 and Lemma 1, the discrete time system is minimal, and therefore, each eigenvalue of the matrix $\bar{A}$, including the repeated ones, can only have one Jordan block. Let

$$M\bar{A}M^{-1} = \Lambda = \begin{pmatrix} \Lambda_1 & 0 & \ldots & 0 \\ 0 & \Lambda_2 & \ldots & 0 \\ & & \ddots & \\ \ldots & \ldots & & \ldots \\ 0 & 0 & \ldots & \Lambda_q \end{pmatrix}$$

where each Λ_i is a Jordan block with the dimension l_i. What we have to show is that the system is reachable; therefore, the sufficient richness of the input $x[lT]$ implies PE of $\phi_1[l]$. From Assumption 1 and Lemma II.1, we know that $(\bar{A}^p, \bar{g})$ is reachable if $(\bar{A}, \bar{g})$ is reachable. Note that $(\bar{A}, \bar{g})$ is reachable $\iff$ $(\Lambda, M\bar{b})$ is reachable $\iff$ $\operatorname{rank}(\lambda I - \Lambda, M\bar{b}) = n$ for all $\lambda = e^{h\lambda_1}, \ldots e^{h\lambda_n}$, where λ_is are the eigenvalues of the continuous-time system. Now, let the first column of the matrix M be

$$(t_1, \ldots, t_{l_1}, t_2, \ldots, t_{l_2}, \ldots, t_q, \ldots, t_{l_q})'.$$

It follows that we have the equation shown at the bottom of the page. From Assumption 1, $e^{\lambda_i h} \neq e^{\lambda_m h}$, if $i \neq m$. Hence, $(\Lambda, M\bar{g})$ is reachable $\Longleftrightarrow$ the last row of each

$$\Lambda_i^{n-1}\begin{pmatrix} t_i \\ \vdots \\ t_{l_i}\end{pmatrix} b_1 + \cdots + \Lambda_i \begin{pmatrix} t_i \\ \vdots \\ t_{l_i}\end{pmatrix} b_{n-1} + \begin{pmatrix} t_i \\ \vdots \\ t_{l_i}\end{pmatrix} b_n$$
$$= t_{l_i}\left(b_1\left(e^{\lambda_i h}\right)^{n-1} + b_2\left(e^{\lambda_i h}\right)^{n-2} + \cdots + b_{n-1}e^{\lambda_i h} + b_n\right)$$

is not zero. The term inside the bracket is not zero; otherwise, $e^{\lambda_i h}$ would be both a zero and a pole of the discrete-time system. This would contradict Assumption 1 and Lemma 1. Thus, $(\bar{A}, \bar{g})$ is reachable if t_{l_i} is not zero; $i = 1, 2, \ldots, q$. To show that those t_{l_i}s are not zero, observe that any eigenvalue $e^{\lambda_m h}$ of $\bar{A}$ is not zero, and moreover, if $e^{\lambda_m h}$ is an eigenvalue associated with the mth Jordan block Λ_m with multiplicity l_m, the corresponding eigenvector and generalized eigenvectors are the columns of the following matrix:

$$M_m = \begin{pmatrix} -e^{(n-1)\lambda_m h} & \cdots & -\begin{pmatrix} n-1 \\ l_m - 1\end{pmatrix} e^{(n-l_m)\lambda_m h} \\ \vdots & \vdots & \vdots \\ -e^{\lambda_m h} & \cdots & 0 \\ -1 & \cdots & 0 \end{pmatrix}.$$

Note that $M^{-1} = (M_1, M_2, \ldots, M_q)$ and

$$M = \left(M^{-1}\right)^{-1} = \frac{\text{adj}\left(M^{-1}\right)}{\det\left(M^{-1}\right)} = \frac{C}{\det\left(M^{-1}\right)}$$

where C' is the adjoint matrix formed by the cofactors c_{ij}. Thus, $t_{l_i} = c_{1l_i}/\det(M^{-1})$, where c_{1l_i} is the determinant of some nonsingular matrix formed by deleting the first row and l_i column of M^{-1}, and this implies $t_{l_i} \neq 0$ for all $i = 1, 2, \ldots, q$. This completes the proof.

Finally, we discuss the parameter convergence. In the presence of noise, (12) and (13) become, respectively

$$y[lT] = \phi_1'[l]\bar{a} + v_1[l]$$
$$y[kh] = \phi_2'[k]\bar{b} + v_2[k]$$

where

$$v_1[l] = v[lT]$$
$$v_2[k] = -\sum_{i=0}^{n} \tilde{b}_k v[kh - ih] + \sum_{i=1}^{n} \bar{b}_i v\left[kh - ih + \frac{h}{2}\right].$$

If the noise is bounded, i.e., $\sup_t |v(t)| \leq \epsilon$, then both $v_1[l]$ and $v_2[k]$ are bounded

$$\sup\left\{|v_1[l]|, |v_2[k]|\right\} \leq c_1\epsilon$$

for some constant $c_1 \geq 0$, and this leads to the following well-known result in the system identification literature [11].

Theorem III.1: Consider the parameter update algorithms (14) and (15) with bounded noise $\sup_t |v(t)| \leq \epsilon$. Suppose that $\phi_1[l]$ and $\phi_2[k]$ are PE. Then, the parameter estimation errors $(\hat{\bar{a}}[l] - \bar{a})$ and $(\hat{\bar{b}}[k] - \bar{b})$ converge exponentially to a ball centered at the origin with radius $c_2\epsilon$ for some constant $c_2 \geq 0$, where c_2 relies on the level of the PE.

Clearly, if noise is absent, the parameter estimates converge to the true values exponentially.

E. Sufficient Richness of $x[kT]$

From the convergence analysis, we see that parameter convergence depends on the PE conditions on $\phi_1[l]$ and $\phi_2[k]$, which rely on the spectral contents of $x[kT]$. This is often referred to as the sufficient richness condition in the system identification literature [3], [11]. However, x is the internal variable that is not measurable and directly controllable. It is desirable to have the richness conditions in terms of the input $u[kT]$ over which we may have control. Translation of the richness condition from x to u is actually difficult because of the nonlinearity f. For instance, let $u[kT]$ be a pseudo-random binary noise sequence (PRBS) taking values ± 1 and $x[kT] = u[kT]^4$. $u[kT]$ contains infinitely many spectral lines and $x[kT] \equiv 1$, containing only one spectral line. Thus, $x[kT]$ may not be sufficiently rich when $u[kT]$ is.

However, in the case of polynomial nonlinearities $x = \sum_{i=1}^{q} r_i u^i$ and sinusoidal inputs $u[kT] = \sum_{i=1}^{m} c_i \cos(\Omega_i k)$, the PE condition can be easily established. If the input has a single spectral line $u[kT] = e^{j\Omega k}$, $x[kT] = \sum_{i=1}^{q} r_i e^{j\Omega i k}$

$$M\bar{b} = \Lambda^{n-1}M\begin{pmatrix} b_1 \\ 0 \\ \vdots \\ 0\end{pmatrix} + \Lambda^{n-2}M\begin{pmatrix} b_2 \\ 0 \\ \vdots \\ 0\end{pmatrix} + \cdots + M\begin{pmatrix} b_n \\ 0 \\ \vdots \\ 0\end{pmatrix} = \begin{pmatrix} \Lambda_1^{n-1}\begin{pmatrix} t_1 \\ \vdots \\ t_{l_1}\end{pmatrix} b_1 + \cdots + \Lambda_1\begin{pmatrix} t_1 \\ \vdots \\ t_{l_1}\end{pmatrix} b_{n-1} + \begin{pmatrix} t_1 \\ \vdots \\ t_{l_1}\end{pmatrix} b_n \\ \Lambda_2^{n-1}\begin{pmatrix} t_2 \\ \vdots \\ t_{l_2}\end{pmatrix} b_1 + \cdots + \Lambda_2\begin{pmatrix} t_2 \\ \vdots \\ t_{l_2}\end{pmatrix} b_{n-1} + \begin{pmatrix} t_2 \\ \vdots \\ t_{l_2}\end{pmatrix} b_n \\ \vdots \\ \Lambda_q^{n-1}\begin{pmatrix} t_q \\ \vdots \\ t_{l_q}\end{pmatrix} b_1 + \cdots + \Lambda_q\begin{pmatrix} t_q \\ \vdots \\ t_{l_q}\end{pmatrix} b_{n-1} + \begin{pmatrix} t_q \\ \vdots \\ t_{l_q}\end{pmatrix} b_n \end{pmatrix}.$$

has q spectral lines unless some frequencies Ωis are the same module 2π. If

$$u[kT] = \sum_{i=1}^{m} 2c_i \cos(\Omega_i k) = \sum_{i=1}^{m} c_i \left(e^{j\Omega_i k} + e^{-j\Omega_i k}\right)$$

has $2m$ spectral lines, $x[kT]$ is in the form of the equation shown at the bottom of the page. Therefore, for $q \geq 2$, $x[kT]$ has all the frequencies

$$\pm\Omega_{i1} \pm \Omega_{i2} \pm \cdots \pm \Omega_{iq}, \quad \Omega_{il} \in [0, \Omega_1, \Omega_2, \ldots, \Omega_m]$$

unless it is in a pathological case, where either the coefficients are zeros or the frequencies are the same module 2π.

IV. Identification of the Nonlinear Block

A. Direct Approach

Once the linear part $G(z)$ is identified, we have the estimates $\hat{a}_i$s and $\hat{b}_j$s of a_is and b_js. The unknown parameter vector γ that parameterizes the nonlinear block can be estimated directly by minimizing

$$\hat{\gamma} = \arg\min_{\bar{\gamma}} \sum_{k} \left(y[kT] - \sum_{i=1}^{n} \hat{a}_i y[kT - iT] - \sum_{i=1}^{n} \hat{b}_i f\left(u[kT - iT], \ldots, u[kT - iT - mT], \bar{\gamma}\right) \right)^2. \quad (16)$$

The convergence and computational complexity of the minimization depends, of course, on the nonlinearity f. Here, we are particularly interested in the linear parameterization structure

$$f\left(u[kT], \ldots, u[kT - mT], \gamma\right) = \sum_{i=1}^{l} \gamma_i f_i\left(u[kT], \ldots, u[kT - mT]\right) \quad (17)$$

with known f_is and unknown γ_is. Then, by defining

$$\phi_3'[k] = \left(\sum_{i=1}^{n} \hat{b}_i f_1\left(u[kT - iT], \ldots, u[kT - iT - mT]\right), \ldots, \sum_{i=1}^{n} \hat{b}_i f_l\left(u[kT - iT], \ldots, u[kT - iT - mT]\right) \right)$$

$$\xi[k] = y[kT] - \sum_{i=1}^{n} \hat{a}_i y[kT - iT]$$

(16) can be rewritten as

$$\hat{\gamma} = \arg\min_{\bar{\gamma}} \sum_{k} \left(\xi[k] - \phi_3'[k]\bar{\gamma}\right)^2. \quad (18)$$

All variables $\phi_3[k]$ and $\xi[k]$ are available, and $\hat{\gamma}$ can be estimated by many standard algorithms. For example, using the recursive LMS algorithm, we get

$$\hat{\gamma}[k] = \hat{\gamma}[k-1] + \frac{\phi_3[k]}{1 + \phi_3'[k]\phi_3[k]} \left(\xi[k] - \phi_3'[k]\hat{\gamma}[k-1]\right).$$

We remark that the most common polynomial nonlinearity representation

$$x[kT] = \sum_{i=1}^{l} \gamma_i u^i[kT]$$

in the Hammerstein model literature is a special case of (17) with $f_j(u[kT], \ldots, u[kT - mT]) = u[kT]^j$.

B. Indirect Approach

In this approach, our goal is to recover the unknown internal signal $x[kT]$ first and then to estimate the nonlinear block using the information of $u[kT]$ and $x[kT]$. This approach is particularly useful when the nonlinear block is static $x[kT] = f(u[kT])$ but lacks structure. Because of unknown structure, it is not possible to estimate the nonlinear function f in terms of parameter estimation. However, if the data $u[kT]$ and $x[kT]$ become available, the complete picture of f can be easily graphed. This graphical picture provides us accurate

---

$$\begin{aligned} x &= \sum_{i=1}^{q} r_i \left[\sum_{l=1}^{m} c_i \left(e^{j\Omega_i k} + e^{-j\Omega_i k}\right) \right]^i \\ &= \sum_{i=1}^{q} r_i \sum_{\substack{p_{i1}+p_{i2}+\cdots+p_{im}=i \\ p_{i1}, p_{i2}, \ldots, p_{im} \geq 0}} \frac{i!}{p_{i1}! p_{i2}! \ldots p_{im}!} c_1^{p_{i1}} \ldots c_m^{p_{im}} \\ &\quad \cdot \left(e^{jp_{i1}\Omega_1 k} + p_{i1} e^{j(p_{i1}-1)\Omega_1 k} e^{-j\Omega_1 k} \right. \\ &\quad \left. + \frac{p_{i1}(p_{i1}-1)}{2!} \cdot e^{j(p_{i1}-2)\Omega_1 k} e^{-j2\Omega_1 k} + \cdots + e^{-jp_{i1}\Omega_1 k} \right) \\ &\quad \cdots \left(e^{jp_{im}\Omega_m k} + p_{im} e^{j(p_{im}-1)\Omega_m k} e^{-j\Omega_m k} + \frac{p_{im}(p_{im}-1)}{2!} \right. \\ &\quad \left. \cdot e^{j(p_{im}-2)\Omega_m k} e^{-j2\Omega_m k} + \cdots + e^{-jp_{im}\Omega_m k} \right). \end{aligned}$$

information on the unknown f as long as there is enough pair $(u[kT], x[kT])$ in the range of interest.

The first step of this indirect approach is to recover the unknown $x[kT]$. To simplify notation, suppose the transfer function $G(z)$ is available. If not, its estimate $\hat{G}(z)$ can be obtained by applying the blind identification algorithm presented in the previous section.

Now, recall

$$Y(z) = G(z)X(z)$$

where $Y(z)$ is the Z-transform of $\{y[kT]\}$, which is available. Supposing that $G(z)$ is minimum phase, $x[kT]$ can be recovered by taking the inverse

$$X(z) = G^{-1}(z)Y(z)$$

or, in the time domain

$$x[kT] = \frac{1}{b_1}\left(-b_2 x\left[(k-1)T\right] - \cdots - b_n x\left[(k-n+1)T\right] + y\left[(k+1)T\right] - a_1 y[kT] - \cdots - a_n y\left[(k-i-n)T\right]\right).$$

If, however, $G(z)$ is nonminimum phase, inversion becomes problematic. To this end, suppose $G_{n+1}(z)$ and $\tilde{G}_{n+1}(z)$ do not share any common zeros. Then, from the Bezout identity, there exist two stable transfer functions $F(z)$ and $\tilde{F}(z)$ such that

$$F(z)\tilde{G}_{n+1}(z) + \tilde{F}(z)G_{n+1}(z) = 1. \tag{19}$$

This implies

$$\begin{aligned} F(z)\tilde{Y}_{n+1}(z) + \tilde{F}(z)Y_{n+1}(z) &= \Big(F(z)\tilde{G}_{n+1}(z) \\ &\quad + \tilde{F}(z)G_{n+1}(z)\Big)X_{n+1}(z) \\ &= X_{n+1}(z). \end{aligned} \tag{20}$$

Therefore, $x[kh]$ and, consequently, $x[kT]$ can be obtained by filtering $y[kh]$ and $y[kh + (h/2)]$ using $\tilde{F}(z)$ and $F(z)$. Note that calculations of $F(z)$ and $\tilde{F}(z)$ are straightforward if $G(z)$ and $\tilde{G}(z)$ are available.

Once $x[kT]$ is obtained, the nonlinear block f can be estimated by using the information of $u[kT]$ and $x[kT]$. We consider two cases.

1) The nonlinear function f is static $x[kT] = f(u[kT])$ and nonparametric. In this case, the function f can be graphed using pairs of $u[kT]$s and the estimated $x[kT]$s. From the graph, the nonlinear function f can be estimated, provided that enough pairs of $(u[kT], x[kT])$ are availabe in the range of interest. Clearly, the pseudo-random binary noise sequence (PRBS), which can generate only two pairs of $(u[kT], x[kT])$, is certainly not a good choice. This is a well-known fact in the literature.
2) The nonlinear function is in the general form of (1), and then, γ can be estimated by minimizing

$$\hat{\gamma} = \arg\min_{\bar{\gamma}} \sum_k \left(x[kT] - f\left(u[kT], \ldots, u[kT - mT], \bar{\gamma}\right)\right)^2. \tag{21}$$

In particular, if f is linear in the unknown γ

$$f\left(u[kT], \ldots, u[kT - mT], \gamma\right) = \sum_{i=1}^{l} \gamma_i f_i\left(u[kT], \ldots, u[kT - mT]\right)$$

with known f_is and unknown γ_is, then, by defining

$$\phi_4'[k] = \left(f_1\left(u[kT], \ldots, u[kT - mT]\right) \ldots, f_l\left(u[kT], \ldots, u[kT - mT]\right)\right)$$

we have

$$\hat{\gamma} = \arg\min_{\bar{\gamma}} \sum_k \left(x[kT] - \phi_4'[k]\bar{\gamma}\right)^2 \tag{22}$$

and $\hat{\gamma}$ can be estimated using, e.g., the recursive LMS algorithm

$$\hat{\gamma}[k] = \hat{\gamma}[k-1] + \frac{\phi_4[k]}{1 + \phi_4'[k]\phi_4[k]}\left(x[kT] - \phi_4'[k]\hat{\gamma}[k-1]\right).$$

V. Simulations

In this section, we provide a simulated numerical example. Let the unknown transfer function of the continuous time system be

$$\frac{0.4095s + 1.0921}{s^2 + 0.32s + 0.02}$$

or in the state-space equation form

$$\begin{aligned} \dot{w}(t) &= \begin{pmatrix} 0 & 1 \\ -0.02 & -0.3 \end{pmatrix} w(t) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} x(t) \\ y(t) &= (1.0921, 0.4095) w(t). \end{aligned}$$

With the sampling interval $T = 0.6$, the corresponding discrete transfer function is given by

$$G(z) = \frac{0.9397z^{-1} - 0.3420z^{-2}}{1 - 1.8287z^{-1} + 0.8353z^{-2}}.$$

Note that the norm of numerator coefficient vector of $G(z)$ is normalized to 1. The unknown nonlinearity f is assumed to be a static second-order polynomial parameterized by the unknown γ

$$x[kT] = \gamma_1 u[kT] + \gamma_2 u[kT]^2, \quad \gamma_1 = \gamma_2 = 1.$$

The purpose of the identification is to estimate the unknown coefficient vectors of the numerator, denominator, and the polynomial

$$b = \begin{pmatrix} 0.9397 \\ -0.3420 \end{pmatrix}, \quad a = \begin{pmatrix} 1.8287 \\ -0.8353 \end{pmatrix}, \quad \gamma = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

For the simulation, 100 Monte Carlo runs were calculated. For each Monte Carlo run, the input $u[kT]$, $k = 1, \ldots, 300$ is uniformly distributed in $[-5,5]$, and the noise is uniformly distributed with magnitude 0.1. The estimate of $\hat{G}(z)$ at each Monte Carlo run is obtained by the blind identification algorithm proposed in the previous section. For the nonlinear part vector γ, we apply the indirect identification approach as discussed before, i.e., we first obtain the estimate $\hat{x}[kT]$ of $x[kT]$ and then find $\hat{\gamma}$ by minimizing

$$\hat{\gamma} = \arg\min_{\bar{\gamma}} \sum_{k=1}^{300} \left(\hat{x}[kT] - \bar{\gamma}_1 u[kT] - \bar{\gamma}_2 u[kT]^2\right)^2.$$

The normalized root mean square error (NRMSE) is used to show the performance of the proposed method. Let $\hat{a}^{(i)}$, $\hat{b}^{(i)}$,

TABLE I
SIMULATION RESULTS

	estimates	NRMSE error
$b = (0.9397, \ -0.3420)'$	$(0.9403, \ -0.3404)'$	0.0062
$a = (1.8287, \ -0.8353)'$	$(1.8260, \ -0.8326)'$	0.0043
$\gamma = (1, \ 1)'$	$(0.9990, \ 0.9976)'$	0.0062

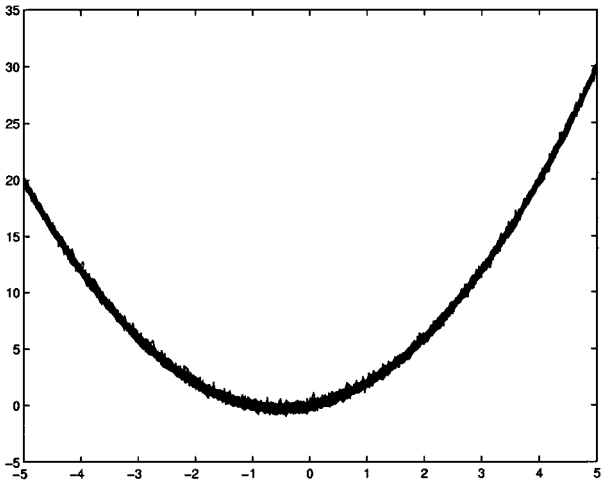

Fig. 2. Estimated nonlinearity drawn by the input $u[kT]$ and the estimated $x[kT]$.

and $\hat{\gamma}^{(i)}$ represent the estimates of a, b, and γ at the ith Monte Carlo run, respectively. The NRMSE error is defined as

$$a_{NRMSE} = \sqrt{\frac{1}{100\|a\|^2}\sum_{i=1}^{100}\left\|\hat{a}^{(i)} - a\right\|^2},$$

$$b_{NRMSE} = \sqrt{\frac{1}{100\|b\|^2}\sum_{i=1}^{100}\left\|\hat{b}^{(i)} - b\right\|^2}$$

$$\gamma_{NRMSE} = \sqrt{\frac{1}{100\|\gamma\|^2}\sum_{i=1}^{100}\left\|\hat{\gamma}^{(i)} - \gamma\right\|^2}.$$

Table I shows the mean values of the 100 Monte Carlo runs as well as the corresponding NRMSE errors.

Fig. 2 shows the true nonlinearity $x = u + u^2$ and 100 estimated nonlinearities by using the input $u[kT]$ and the estimated $\hat{x}[kT]$ for each Monte Carlo run.

Finally, we compare the proposed blind approach to other existing methods for the Hammerstein model identification, especially the popular iterative and stochastic methods. The advantage of the iterative method lies in its simplicity. Usually, the convergence rate of the iterative method is fast, provided that it converges. However, there is no guarantee for the convergence, and in fact, it can be divergent [17]. Moreover, it is impossible to check whether the method converges or not *a priori*. The stochastic method works in a similar way as the proposed blind approach, i.e., it identifies the linear part first. However, to achieve this, a white input assumption was imposed and used explicitly in the stochastic approach. Our approach does not require white inputs, and any input can apply. With the PE condition, e.g., a sinusoidal input with enough frequency content, convergence is guaranteed by using the proposed blind approach.

VI. NONSAMPLED HAMMERSTEIN SYSTEMS

In the previous sections, we have proposed identification algorithms for the sampled Hammerstein systems using blind techniques. Because we cannot do fast sampling for nonsampled systems, these algorithms are not directly applicable here. In this section, we show how to extend these results to nonsampled discrete-time Hammerstein systems. We will focus on the key ideas for brevity.

Consider a nonsampled discrete-time Hammerstein system with static polynomial nonlinearity $x = \sum_{i=1}^{l}\gamma_i u^i$

$$\begin{aligned} y[k] &= \sum_{i=1}^{n} a_i y[k-i] + \sum_{j=1}^{n} b_j x[k-j] + v[k] \\ &= \sum_{i=1}^{n} a_i y[k-i] + \sum_{j=1}^{n}\sum_{i=1}^{l} b_j\gamma_i u^i[k-j] + v[k]. \end{aligned}$$

Hold the input constant over the window of $(n+1)h$ so that for $l = 1, 2, \ldots$

$$\begin{aligned} u\left[(l-1)(n+1)\right] &= u\left[(l-1)(n+1)+1\right] \\ &= \cdots = u\left[(l-1)(n+1)+n\right]. \end{aligned}$$

This implies

$$\begin{aligned} x\left[(l-1)(n+1)\right] &= x\left[(l-1)(n+1)+1\right] \\ &= \cdots = x\left[(l-1)(n+1)+n\right]. \end{aligned}$$

Note that

$$\begin{aligned} y\left[l(n+1)\right] &= \sum_{i=1}^{n} a_i y\left[l(n+1)-i\right] \\ &\quad + \sum_{j=1}^{n} b_j x\left[l(n+1)-j\right] + v\left[l(n+1)\right] \\ y\left[l(n+1)-1\right] &= \sum_{i=1}^{n} a_i y\left[l(n+1)-1-i\right] \\ &\quad + \sum_{j=1}^{n} b_j x\left[l(n+1)-1-j\right] \\ &\quad + v\left[l(n+1)-1\right]. \end{aligned}$$

Therefore, at $l = 1, 2, \ldots$, we have

$$\begin{aligned} &y\left[l(n+1)\right] - y\left[l(n+1)-1\right] \\ &\quad = \sum_{i=1}^{n} a_i \left(y\left[l(n+1)-i\right] - y\left[l(n+1)-1-i\right]\right) \\ &\qquad + v\left[l(n+1)\right] - v\left[l(n+1)-1\right]. \end{aligned}$$

This equation is similar to (11), and a_is can be estimated by many algorithms. Once $\hat{a}_i$s are obtained, $\widehat{b_i\gamma_j}$ can be calculated

$$\begin{aligned} \left(\widehat{b_1\gamma_1}, \ldots, \widehat{b_n\gamma_l}\right) = \arg\min\Bigg\{ & y[k] \\ & - \left(\sum_{i=1}^{n}\hat{a}_i y[k-i] + \sum_{j=1}^{n}\sum_{i=1}^{l}\widehat{b_j\gamma_i}u^i[k-j]\right)\Bigg\}^2. \end{aligned}$$

Finally, $(\widehat{b_1\gamma_1},\ldots,\widehat{b_n\gamma_l})$ can be projected into $\hat{b}=(\hat{b}_1,\ldots,\hat{b}_n)'$ and $\hat{\gamma}=(\hat{\gamma}_1,\ldots,\hat{\gamma}_l)'$ by minimizing the Frobenius norm of

$$\left(\hat{b},\hat{\gamma}\right)=\arg\min\left\|\begin{pmatrix}\widehat{b_1\gamma_1} & \cdots & \widehat{b_1\gamma_l}\\ \vdots & \ddots & \vdots\\ \widehat{b_n\gamma_1} & \cdots & \widehat{b_n\gamma_l}\end{pmatrix}-\hat{b}\hat{\gamma}'\right\|_F.$$

This problem was solved in [2]. Let

$$\begin{pmatrix}\widehat{b_1\gamma_1} & \cdots & \widehat{b_1\gamma_l}\\ \vdots & \ddots & \vdots\\ \widehat{b_n\gamma_1} & \cdots & \widehat{b_n\gamma_l}\end{pmatrix}=\sum_{i=1}^{\min(n,l)}\sigma_i\xi_i\eta_i'$$

be the singular value decomposition(SVD), where σ_is are the singular values, and ξ_is and η_is are n- and l-dimensional orthonormal vectors, respectively. Then, a solution $\hat{b}$ and $\hat{\gamma}$ is

$$\hat{b}=s_\xi\xi_1,\quad \hat{\gamma}=\sigma_1 s_\xi\eta_1$$

where s_ξ is the sign of the first nonzero entry of ξ_1. This guarantees that $\|\hat{b}\|=1$ and that the first nonzero entry is positive.

VII. Concluding Remarks

In this paper, we have proposed blind approaches for Hammerstein model identification. The main interest of the paper is on sampled discrete-time systems where the linear part is originated from a continuous-time system. Using fast sampling at the output, the linear part can be obtained using only the output measurements. Convergence results in terms of PE conditions that apply to a large class of signals are also derived. We have also shown how to extend our results to nonsampled discrete-time models where fast sampling is not permitted.

Our focus in this paper has been on presenting the idea, and therefore, not much effort has been devoted to study the performance of the proposed algorithms under various type of model uncertainties and noises. This issue is certainly an interesting one for further study, and we expect that results will be quite different from the traditional linear system identification because of noise structure in the error equation. It will also be interesting to characterize conditions for sufficient richness for specific types of nonlinearities and inputs.

References

[1] E. W. Bai, "Identification of systems with hard input nonlinearities," in *Perspectives in Control*, R. Moheimani, Ed. New York: Springer-Verlag, 2001.

[2] ——, "An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems," *Automatica*, vol. 34, no. 3, pp. 333–338, 1998.

[3] E. W. Bai and M. Fu, "Blind system identification and channel equalization of IIR systems without statistical information," *IEEE Trans. Signal Processing*, vol. 47, pp. 1910–1921, July 1999.

[4] N. Bershad, S. Bouchired, and F. Castanie, "Stochastic analysis of adaptive gradient identification of Wiener-Hammerstein systems for Gaussian inputs," *IEEE Trans. Signal Processing*, vol. 48, pp. 557–559, Feb. 2000.

[5] S. A. Bilings and S. Y. Fakhouri, "Identification of a class of nonlinear systems using correlation analysis," *Proc. Inst. Elect. Eng.*, vol. 125, no. 7, pp. 691–697, 1978.

[6] M. Boutayeb, H. Rafaralahy, and M. Darouach, "A robust and recursive identification method for Hammerstein model," in *Proc. IFAC World Congr.*, San Francisco, CA, 1996, pp. 447–452.

[7] F. Chang and R. Luus, "A noniterative method for identification using Hammerstein model," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 464–468, 1971.

[8] W. Greblicki, "Nonlinearity estimation in Hammerstein systems based on ordered observations," *IEEE Trans. Signal Processing*, vol. 44, pp. 1224–1233, May 1996.

[9] T. Hsia, "A multi-stage least squares method for identifying Hammerstein model nonlinear systems," in *Proc. CDC*, Clearwater, FL, 1976, pp. 934–938.

[10] R. Haber and H. Unbehauen, "Structure identification of nonlinear dynamic systems-a survey of input/output approaches," *Automatica*, vol. 24, pp. 651–677, 1990.

[11] L. Ljung, *System Identification*. Englewood Cliffs, NJ: Prentice -Hall, 1987.

[12] K. S. Narendra and P. G. Gallman, "An iterative method for the identification of nonlinear systems using a Hammerstein model," *IEEE Trans. Automat. Contr.*, vol. AC-11, pp. 546–550, 1966.

[13] M. Pawlak, "On the series expansion approach to the identification of Hammerstein systems," *IEEE Trans. Automat. Contr.*, vol. 36, pp. 763–767, June 1991.

[14] Prakriya and D. Hatzinakos, "Blind identification of linear subsystems of LTI-ZMNL-LTI models with cyclostationary inputs," *IEEE Trans. Signal Processing*, vol. 45, pp. 2023–2036, Aug. 1997.

[15] S. Rangan, G. Wolodkin, and K. Polla, "Identification methods for Hammerstein systems," in *Proc. CDC*, New Orleans, LA, 1995, pp. 697–702.

[16] A. Ruhe and P. Wedin, "Algorithms for separable nonlinear least squares problems," *SIAM Rev.*, vol. 22, pp. 318–337, 1980.

[17] P. Stocia, "On the convergence of an iterative algorithm used for Hammerstein system identification," *IEEE Trans. Automat. Contr.*, vol. AC-26, pp. 967–969, 1981.

[18] L. Sun, W. Liu, and A. Sano, "Identification of dynamical system with input nonlinearity," *Proc. Inst. Elect. Eng., Contr. Theory Applicat.*, vol. 146, no. 1, pp. 41–51, 1998.

[19] J. Voros, "Parameter identification of discontinuous Hammerstein systems," *Automatica*, vol. 33, no. 6, pp. 1141–1146, 1997.

Er-Wei Bai (SM'00) was educated at Fudan University, Shanghai, China, Shanghai Jiaotong University, and the University of California, Berkeley.

He is Professor of electrical and computer engineering at the University of Iowa, Iowa City, where he teaches and conducts research in system identification and signal processing.

Dr. Bai serves the IEEE Control Systems Society and the IFAC in various capacities.

Minyue Fu (SM'94) received the B.Sc. degree in electrical engineering from the China University of Science and Technology, Hefei, in 1982 and the M.S. and Ph.D. degrees in electrical engineering from the University of Wisconsin, Madison (UWM), in 1983 and 1987, respectively.

From 1987 to 1989, he served as an Assistant Professor with the Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI. For the summer of 1989, he was with the Universite Catholoque de Louvain, Louvain, Belgium, as a Maitre de Conferences Invited. He joined the Department of Electrical and Computer Engineering, the University of Newcastle, Callaghan, Australia, in 1989. Currently, he is an Associate Professor with the University of Newcastle. He also visited the University of Iowa, Iowa City, from 1995 to 1996. He is currently visiting Nanyang Technological University, Singapore. His main research interests include control systems, signal processing, and telecommunications. He has been an Associate Editor of the *Journal of Optimization and Engineering.*

Dr. Fu was an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL.