PERGAMON

Automatica 38 (2002) 967–979

automatica

www.elsevier.com/locate/automatica

A blind approach to the Hammerstein–Wiener model identification ☆

Er-Wei Bai *

Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242, USA

Received 6 March 2001; received in revised form 6 August 2001; accepted 30 November 2001

Abstract

In this paper, we propose a blind approach to the sampled Hammerstein–Wiener model identification. By using the blind approach, it is shown that all internal variables can be recovered solely based on the output measurements. Then, identification of linear and nonlinear parts can be carried out. No a priori structural knowledge about the input nonlinearity is assumed and no white noise assumption is imposed on the input. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: System identification; Parameter estimation; Nonlinear systems; Hammerstein systems; Wiener systems; Hammerstein–Wiener systems

1. Introduction

The Hammerstein and Wiener models are special kinds of nonlinear systems where the nonlinear block is static and follows or is followed by a linear system. These models have applications in many engineering problems and therefore, identification of Hammerstein and Wiener models has been an active research area for many years. There exists a large number of research papers in the literature on the topics of Hammerstein and Wiener model identifications. There exists only scattered work reported in the literature on the Hammerstein–Wiener model identification (Bai, 1998; Haber & Unbehauen, 1990).

Existing methods for Hammerstein model identification can be roughly divided into four categories: the iterative method (Narendra & Gallman, 1966; Stoica, 1981), the over-parameterization method (Bai, 1998; Boutayeb, Rafaralahy, & Darouach, 1996; Chang & Luus, 1971; Hsia, 1976), the stochastic method (Bilings & Fakhouri, 1978; Greblicki, 1992; Pawlak, 1991) and the separable least-squares method (Bai, 2001). In most cases, the structure of the nonlinearity is assumed to be known. Otherwise, identification becomes a structural estimation problem (Haber & Unbehauen, 1990). Identification of Wiener models is more difficult. The reason is the lack of a good representation of the output nonlinearity for identification purpose. Unlike in the Hammerstein model case, the commonly used polynomial representation for the output nonlinearities makes identification in terms of unknown parameters very hard. The main technique used for Wiener model identification is the stochastic method (Bilings & Fakhouri, 1978; Wigren, 1994). Suppose the input is a zero mean white noise, identification of linear part and identification of nonlinear part are shown to be separable. Moreover, it was shown in Greblicki (1992) that even the structure of the nonlinearity may be estimated with white noise inputs under certain conditions. The difficulty with the stochastic method is that it is limited to white inputs. To overcome this difficulty, inverse representation method was proposed (Kalafatis, Wang, & Cluett, 1997; Pajunen, 1992) to model the inverse of the output nonlinearity allowing nonwhite inputs.

In this paper, we continue our work on the Hammerstein–Wiener model identification (Bai, 1998). Unlike in Bai (1998) where a very special structure is assumed, however, the blind approach in this paper allows a very general structure on the nonlinearities. In particular, the input nonlinearity structure can be arbitrary and is not assumed to be known. By using the blind identification approach, all the unknown internal variables can be recovered solely based on the output measurements. Once all interval variables are recovered, linear and nonlinear parts including the structure can be identified. Our scheme applies to either white or nonwhite inputs. The blind techniques adopted in this paper use our previous results presented for blind channel equalizations of IIR systems (Bai & Fu, 1999), and are also based on blind techniques developed for

☆ This work was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Johan Schoukens under the direction of Editor Torsten Söderström. This work was supported in part by NSF Grants ECS-9710297 and ECS-0098181.

* Tel.: +001-319-335-5197; fax: +001-319-6028.

E-mail address: erwei@engineering.uiowa.edu (E.-W. Bai).

0005-1098/02/$ - see front matter
PII: S0005-1098(01)00292-8

Hammerstein models (Bai & Fu, 2001; Sun, Liu, & Sano, 1998). Although the algorithm proposed in this paper is for the Hammerstein–Wiener model, it applies directly to the identification of either Wiener models or Hammerstein models with trivial modifications.

To close this section, we give an outline of the paper. Section 2 formulates the problem and derives some preliminary results. The main results are given in Section 3 along with some remarks and simulation examples. Additional remarks are included in Section 4.

2. Problem statement and preliminaries

Consider the sampled Hammerstein–Wiener model shown in Fig. 1, which consists of a zero-order-hold, an input nonlinearity, a scalar linear stable continuous time system and an output nonlinearity. It is assumed that

Assumption 1.

- The unknown continuous time system $P(s)$ possesses a state space representation

$$\dot{\eta}(t) = A\eta(t) + bu(t), \quad u, x \in R, \ A \in R^{n\times n},$$
$$x(t) = c\eta(t). \tag{2.1}$$

For a given sampling interval T, its discrete time state space equation and the corresponding transfer function are given, respectively, by

$$\eta[(k+1)T] = \underbrace{\mathrm{e}^{AT}}_{\Phi}\eta[kT] + \underbrace{\int_0^T \mathrm{e}^{At}\,\mathrm{d}t \cdot b}_{\Gamma} u[kT],$$
$$x[kT] = c\eta[kT] \tag{2.2}$$

and

$$G(z) = c(zI - \Phi)^{-1}\Gamma = \frac{\beta(z)}{\alpha(z)} = \frac{\beta_1 z^{-1} + \beta_2 z^{-2} + \cdots + \beta_n z^{-n}}{1 - \alpha_1 z^{-1} - \alpha_2 z^{-2} - \cdots - \alpha_n z^{-n}} \tag{2.3}$$

form some α_i's and β_j's.

- The input nonlinearity is static

$$u = g(w). \tag{2.4}$$

Its structure is not assumed to be known.

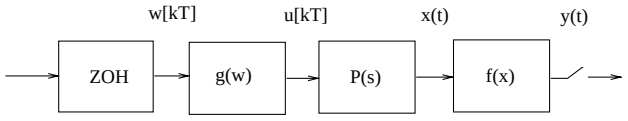

Fig. 1. The sampled Hammerstein–Wiener model.

- The output nonlinearity is also static

$$y = f(x). \tag{2.5}$$

However, it is assumed that $f(\cdot)$ is one-to-one so that the inverse $x = f^{-1}(y)$ exists and admits a polynomial representation

$$x = \sum_{i=1}^{m} r_i y^i. \tag{2.6}$$

In identification, we will first estimate the inverse coefficients r_i's and then, find the best forward function

$$y = \sum_{i=1}^{q} a_i x^i$$

of $x = \sum_{i=1}^{m} r_i y^i$ in the least squares sense using the observed data. Finally, we represent the output nonlinearity $y = f(x)$ by $y = \sum_{i=1}^{q} a_i x^i$. We remark that if $y = f(x)$ is one-to-one and continuous, its inverse $x = f^{-1}(y)$ exists and is also continuous. With a bounded input and a stable system, x and y are always bounded and this implies that the inverse $x = f^{-1}(y)$ can be approximated to any accuracy by a polynomial $x = \sum_{i=1}^{m} r_i y^i$. In particular, as the order m goes to infinite, $x = f^{-1}(y) = \sum_{i=1}^{m} r_i y^i$. This shows that the inverse representation of (2.6) is theoretically justified. Practically, of course, a high order m usually means a high sensitivity to noise and model uncertainty in the identification setting. Therefore, there is a limitation to the inverse approach. A discussion on this topic will be provided later.

For a given sampling interval T, *the goal of the Hammerstein–Wiener model identification is to estimate the transfer function $G(z)$ in terms of its parameters α_i's and β_j's, the output nonlinearity $x = \sum_{i=1}^{m} r_i y^i$ and its inverse $y = \sum_{i=1}^{q} a_i x^i$, and the input nonlinearity $u = g(w)$ solely based on the measurements of w and y*. No internal variables x and u are assumed available. Moreover, the structure of the input nonlinearity $u = g(w)$ is unknown.

Our idea of identification is the blind approach, i.e., to sample the output at a higher rate. Given the sampling interval T, let the output sampling interval be

$$h = T/l, \quad l \geqslant 1$$

for some positive integer l, referred to as the over-sampling ratio. The following lemma (Bai & Fu, 1999) will be useful.

Lemma 2.1. *Consider the continuous system* (2.1) *and its sampled system at the sampling interval $h = T/l$ for some $l \geqslant 1$. Then, the sampled system is minimal for any $l \geqslant 1$ if and only if the sampled system at the sampling interval T is minimal.*

Minimality is important in identification. Without minimality assumption, the transfer function $G(z)$ has pole-zero cancellations and this makes the parameterization nonunique. In other words, the identifiability is lost. To this end, we make the following assumption.

Assumption 2. The sampled system (2.2) is assumed to be minimal at the sampling interval T.

Before closing this section, we observe that the parameterization of the Hammerstein–Wiener model is actually not unique. Suppose the system is represented by three blocks $u=g(w)$, $G(z)$ and $y=f(x)$. Then, any triple $ag(w)$, $bG(z)$ and $cf(x)$, for some constants a,b,c, would produce the identical input–output measurements, provided that $abc=1$. In other words, any identification setting cannot distinguish between (g,G,f) and (ag,bG,cf). To obtain a unique parameterization, two blocks need to be normalized. Since the structure of the input nonlinearity is not assumed to be known, we normalize the linear block and the output nonlinearity.

Assumption 3. For the Hammerstein–Wiener model, it is assumed that $\beta_1 = 1$ and $r_1 = 1$.

With this normalization assumption and persistent exciting (PE) input, we will show that g, G and f can be uniquely identified. Note that there are other ways to normalize the system, Assumption 3.1 is the simplest one. The purpose is to avoid unnecessary complications so that our ideas can be presented clearly.

3. Identification of the Hammerstein–Wiener model

If the structure of the input nonlinearity $u = g(w)$ is unknown, identification of the Hammerstein–Wiener model is no longer a parameter estimation problem. Obviously, the identification involves structural estimation. It is clear, however, that if u were available, we would be able to estimate the input nonlinearity structure. At least, the complete picture of $u = g(w)$ can be graphed by using the pairs of (w,u) and this graphical picture provides us accuracy information on the unknown input nonlinearity $u = g(w)$. Therefore, the key is to estimate $G(z)$, $x = \sum r_i y^i$, and then to recover u solely based on the output measurements. We accomplish this goal in several steps, estimating the output nonlinearity, finding the linear part and then recovering u.

3.1. Output nonlinearity estimation

Given the input sampling interval T, let the output sampling interval be

$$h = T/(n+1),$$

where n is the order of $G(z)$. We remark that $h = T/(n+1)$ is not necessary but does make analysis and notation simple. In fact, $h = T/(\bar{n}+1)$ for any $\bar{n} \geqslant n$ will work, see remarks in Discussion section for details. Now, consider the sampled system at the sampling interval $h = T/(n+1)$. It is clear that the transfer function of the sampled system at the sampling interval h is also an nth order strictly proper rational function

$$\tilde{G}(z) = \frac{\tilde{\beta}(z)}{\tilde{\alpha}(z)} = \frac{\tilde{\beta}_1 z^{-1} + \tilde{\beta}_2 z^{-2} + \cdots + \tilde{\beta}_n z^{-n}}{1 - \tilde{\alpha}_1 z^{-1} - \tilde{\alpha}_2 z^{-2} - \cdots - \tilde{\alpha}_n z^{-n}} \tag{3.1}$$

for some unknown $\tilde{\alpha}_i$'s and $\tilde{\beta}_j$'s. Its time domain equation is accordingly given by

$$x[kh] = \sum_{i=1}^{n} \tilde{\alpha}_i x[kh - ih] + \sum_{i=1}^{n} \tilde{\beta}_i u[kh - ih].$$

Substituting the value $x[kh] = \sum_{j=1}^{m} r_j y^j[kh]$ into equation, it follows that

$$\begin{aligned} \sum_{j=1}^{m} r_j y^j[kh] = &\sum_{i=1}^{n} \tilde{\alpha}_i \sum_{j=1}^{m} r_j y^j[kh - ih] \\ &+ \sum_{i=1}^{n} \tilde{\beta}_i u[kh - ih] + v_1[kh], \\ &i = 1, \ldots, N, \end{aligned} \tag{3.2}$$

where $v_1[kh]$ denotes any discrepancy not counted by the model, e.g., the contribution of noise, model uncertainty and approximation errors.

By observing that $r_1 = 1$ from the normalization assumption, the above equation can be re-written as

$$y[kh] = \underbrace{\begin{matrix} (-y^2[kh], \ldots, -y^m[kh], y[(k-1)h], \\ \ldots, y[(k-n)h], \ldots, y^m[(k-1)h], \\ \ldots, y^m[(k-n)h]) \end{matrix}}_{\phi_1'[kh]} \times \underbrace{\begin{pmatrix} r_2 \\ \vdots, \\ r_m \\ \tilde{\alpha}_1 \\ \vdots \\ \tilde{\alpha}_n \\ \tilde{\alpha}_1 r_2 \\ \vdots \\ \tilde{\alpha}_n r_2 \\ \vdots \\ \tilde{\alpha}_1 r_m \\ \vdots \\ \tilde{\alpha}_n r_m \end{pmatrix}}_{\theta_1} + \sum_{i=1}^{n} \tilde{\beta}_i u[kh - ih] + v_1[kh].$$

This is the basic equation for the estimation of the output nonlinearity $x = \sum r_i y^i$ in terms of its coefficients r_i's. Now, consider two consecutive equations at $k = l(n+1)$

and $k = l(n+1)-1$,

$$y[l(n+1)h] = \phi_1'[l(n+1)h]\theta_1 + \sum_{i=1}^{n} \tilde{\beta}_i u[l(n+1)h - ih] + v_1[l(n+1)h],$$

$$y[l(n+1)h - h] = \phi_1'[l(n+1)h - h]\theta_1 + \sum_{i=1}^{n} \tilde{\beta}_i u[l(n+1)h - ih - h] + v_1[l(n+1)h - h]. \quad (3.3)$$

Since the input sampling interval is fixed at $T = (n+1)h$ where h is the output sampling interval, we have

$$w[(l-1)T] = w[(l-1)T + h] = \cdots = w[(l-1)T + nh], \quad l = 1, 2, \ldots$$

and this implies

$$u[(l-1)T] = u[(l-1)T + h] = \cdots = u[(l-1)T + nh], \quad l = 1, 2, \ldots$$

Therefore, we have

$$u[(l-1)T] = u[l(n+1)h - h - nh] = \cdots = u[l(n+1)h - h] = u[(l-1)T + nh],$$

and it follows that

$$\Delta y[l] = \Delta\phi_1'[l]\theta_1 + \Delta v[l] \quad (3.4)$$

with

$$\Delta y[l] = y[lT] - y[lT - h],$$

$$\Delta v[l] = v_1[lT] - v_1[lT - h],$$

$$\Delta\phi_1[l] = \phi_1[lT] - \phi_1[lT - h].$$

In Eq. (3.4), $\Delta y[l]$ and $\Delta\phi_1[l]$ consist of output measurements $y[kh]$ only and thus are available. Moreover, this equation is linear in the unknown parameter vector θ_1 which can be estimated by many standard methods, e.g., the least-squares method or the (normalized) LMS algorithm shown below

$$\hat{\theta}_1[l] = \hat{\theta}_1[l-1] + \frac{\Delta\phi_1[l]}{1 + \Delta\phi_1'[l]\Delta\phi_1[l]}(\Delta y[l] - \Delta\phi_1'[l]\hat{\theta}_1[l-1]).$$

Note that the estimates $\hat{r}_2, \ldots, \hat{r}_m, \hat{\tilde{\alpha}}_1, \ldots, \hat{\tilde{\alpha}}_n$ of $r_2, \ldots, r_m, \tilde{\alpha}_1, \ldots, \tilde{\alpha}_n$ are the first $(m-1+n)$ entries of $\hat{\theta}_1$. Also note that r_1 is normalized to be 1. Therefore, once $\hat{\theta}_1$ is obtained, we have the estimate of the inverse output nonlinearity

$$\hat{f}^{-1}(y) = \sum_{i=1}^{m} \hat{r}_i y^i[kh] \quad (3.5)$$

with $\hat{r}_1 = 1$, as well as the estimate of the denominator $\tilde{\alpha}(z)$ of $\tilde{G}(z)$.

The forward output nonlinearity

$$y[kh] = \sum_{i=1}^{q} \hat{a}_i x^i[kh]$$

can be constructed by minimizing

$$\hat{a} = \arg\min \sum_{k} \left\{ y[kh] - \sum_{i=1}^{q} a_i \hat{x}^i[kh] \right\}^2,$$

where $y[kh]$'s are the observed outputs and $\hat{x}[kh] = \sum_{i=1}^{m} \hat{r}_i y^i[kh]$'s are generated from (3.5). We comment that direct readings of $\hat{r}_i$ and $\hat{\tilde{\alpha}}_k$ from $\hat{\theta}_1$ may not be a good policy in a noisy situation because it ignores a large number of identified parameters $\tilde{\alpha}_k r_i$ without taking into account their contribution. A more robust way should consider their contribution. This can be done as follows. Let $\hat{\theta}_1$ be represented by

$$\hat{\theta}_1^T = (\bar{r}_2, \ldots, \bar{r}_m, \bar{\tilde{\alpha}}_1, \ldots, \bar{\tilde{\alpha}}_n, \overline{\tilde{\alpha}_1 r_2}, \ldots, \overline{\tilde{\alpha}_n r_2}, \ldots, \overline{\tilde{\alpha}_1 r_m}, \ldots, \overline{\tilde{\alpha}_n r_m}).$$

With $\hat{r}_1 = 1$, the estimates $\hat{r}_i$ and $\hat{\tilde{\alpha}}_k$ are defined as

$$(\hat{r}_i, \hat{\tilde{\alpha}}_k) = \arg\min \left\{ \sum_{k} (\hat{\tilde{\alpha}}_k - \bar{\tilde{\alpha}}_k)^2 + \sum_{i,k} (\hat{\tilde{\alpha}}_k \hat{r}_i - \overline{\tilde{\alpha}_k r_i})^2 \right\}$$

$$= \arg\min \left\| \underbrace{\begin{pmatrix} \bar{\tilde{\alpha}}_1 & \overline{\tilde{\alpha}_1 r_2} & \cdots & \overline{\tilde{\alpha}_1 r_m} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\tilde{\alpha}}_n & \overline{\tilde{\alpha}_n r_2} & \cdots & \overline{\tilde{\alpha}_n r_m} \end{pmatrix}}_{\Theta_1} - \begin{pmatrix} \hat{\tilde{\alpha}}_1 \\ \vdots \\ \hat{\tilde{\alpha}}_n \end{pmatrix} (\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_m) \right\|_F, \quad (3.6)$$

where F stands for the matrix Frobenius norm. This problem was solved in Bai (1998). Let

$$\sum_{i=1}^{\min(n,m)} \sigma_i \xi_i \eta_i' = \Theta_1 \quad (3.7)$$

be the singular value decomposition(SVD) of Θ_1, where σ_i's are the singular values and ξ_i and η_i are n- and m-dimensional orthonormal vectors, respectively. Then, a solution with $\hat{r}_1 = 1$ is given by

$$\begin{pmatrix} \hat{\tilde{\alpha}}_1 \\ \vdots \\ \hat{\tilde{\alpha}}_n \end{pmatrix} = \sigma_1 s_\eta \xi_1, \quad \begin{pmatrix} \hat{r}_1 \\ \hat{r}_2 \\ \vdots \\ \hat{r}_m \end{pmatrix} = \begin{pmatrix} 1 \\ \hat{r}_2 \\ \vdots \\ \hat{r}_m \end{pmatrix} = \frac{1}{s_\eta} \eta_1,$$

where s_η is the first entry of η_1.

3.2. Linear transfer function estimation

In this section, we propose a blind method to estimate $G(z)$ without requiring $u[kT]$. Recall that the denominator $1-\hat{\bar{\alpha}}_1 z^{-1}-\cdots-\hat{\bar{\alpha}}_n z^{-n}$ of $\bar{G}(z)$ was obtained as a result of the estimation of $\hat{\theta}_1$ in the previous section. Note $\tilde{G}(z)$ is the transfer function of the sampled system at the sampling interval $h=T/(n+1)$. Write

$$1-\hat{\tilde{\alpha}}_1 z^{-1}-\cdots-\hat{\tilde{\alpha}}_n z^{-n}=(1-\tilde{s}_1 z^{-1})\ldots(1-\tilde{s}_n z^{-1}),$$

where $\tilde{s}_i$ denotes the poles of $\tilde{G}(z)$. The sampled system is assumed to be minimal at the sampling interval T and is minimal at any sampling interval $h=T/l$, $l\geqslant 1$, from Lemma 2.1. Clearly, s is a pole of the continuous time system if and only if $e^{sT}=e^{sh(n+1)}$ is a pole of $G(z)$ if and only if e^{sh} is a pole of $\tilde{G}(z)$. In other words, if $\tilde{s}_i$'s are the poles of $\tilde{G}(z)$, then $\tilde{s}_i^{n+1}$'s are the poles of $G(z)$. This implies that in an estimate of $\alpha(z)$, the denominator of $G(z)$ is given by

$$\begin{aligned}\hat{\alpha}(z)&=(1-\tilde{s}_1^{n+1}z^{-1})\ldots(1-\tilde{s}_n^{n+1}z^{-1})\\&=1-\hat{\alpha}_1 z^{-1}-\cdots-\hat{\alpha}_n z^{-n}.\end{aligned}\tag{3.8}$$

Hence, an estimate of $\alpha(z)$ is already contained in $\hat{\theta}_1$ and what we have to estimate is only the numerator $\beta(z)$ of $G(z)$. To this end, consider two sequences

$$\begin{aligned}\{x[kT]\}\ &\Leftrightarrow\ X(z)=\sum_{k=0}^{\infty}x[kT]z^{-k}=G(z)U(z),\\ \{x[kT+T/2]\}\ &\Leftrightarrow\ \bar{X}(z)=\sum_{k=0}^{\infty}x[kT+T/2]z^{-k}\\&\qquad=\bar{G}(z)U(z),\end{aligned}\tag{3.9}$$

where

$$U(z)=\sum_{k=0}^{\infty}u[kT]z^{-k}$$

is the Z-transform of the sequence $u[kT]$ at the sampling interval T, and $G(z)$ and $\bar{G}(z)$ represent the transfer functions from $U(z)$ to $X(z)$ and $\bar{X}(z)$, respectively. The transfer function $G(z)$ is derived in (2.3) and is strictly proper. $\bar{G}(z)$ needs a special attention. From the continuous time state space equation (2.1), we have

$$\begin{aligned}\eta[(k+1)T+T/2]=&\underbrace{e^{AT}}_{\Phi}\eta[kT+T/2]\\&+\underbrace{\int_{T/2}^{T}e^{At}b\,dt}_{\Gamma_1}u[kT]\\&+\underbrace{\int_{0}^{T/2}e^{At}b\,dt}_{\Gamma_2}u[(k+1)T],\end{aligned}$$

$$x[kT+T/2]=c\eta[kT+T/2].$$

Thus, the transfer function $\bar{G}(z)$ is given by

$$\begin{aligned}\bar{G}(z)&=\frac{\bar{\beta}(z)}{\bar{\alpha}(z)}=c(zI-\Phi)^{-1}(\Gamma_1+\Gamma_2 z)\\&=\frac{\bar{\beta}_0+\bar{\beta}_1 z^{-1}+\cdots+\bar{\beta}_n z^{-n}}{1-\bar{\alpha}_1 z^{-1}-\bar{\alpha}_2 z^{-2}-\cdots-\bar{\alpha}_n z^{-n}}.\end{aligned}$$

It is interesting to note that $G(z)$ and $\bar{G}(z)$ share the same denominator but with different numerators and unlike $G(z)$, $\bar{G}(z)$ is proper but not strictly proper.

Now, consider again two sequences $\{x[kT]\}$, $\{x[kT+T/2]\}$ and their Z-transforms

$$X(z)=G(z)U(z),\quad \bar{X}(z)=\bar{G}(z)U(z).$$

Clearly,

$$\bar{G}(z)X(z)-G(z)\bar{X}(z)=0,\quad \bar{\beta}(z)X(z)-\beta(z)\bar{X}(z)=0$$

and this results in

$$\begin{aligned}&(\bar{\beta}_0+\bar{\beta}_1 z^{-1}+\cdots+\bar{\beta}_n z^{-n})X(z)\\&\quad-(\beta_1 z^{-1}+\cdots+\beta_n z^{-n})\bar{X}(z)=0.\end{aligned}$$

Its time domain equation is

$$\begin{aligned}&\beta_1 x[kT+T/2]\\&\quad=\underbrace{\begin{array}{l}(-x[kT-T+T/2],\ldots,-x[kT-(n-1)T+T/2],\\ x[kT+T],\ldots,x[kT-(n-1)T])\end{array}}_{\bar{\phi}_2'[k]}\\&\qquad\times\underbrace{\begin{pmatrix}\beta_2\\ \vdots\\ \beta_n\\ \bar{\beta}_0\\ \vdots\\ \bar{\beta}_n\end{pmatrix}}_{\theta_2}.\end{aligned}$$

By noting that $\beta_1=1$, it follows that

$$x[kT+T/2]=\bar{\phi}_2'[k]\theta_2.$$

In this equation, $\bar{\phi}_2$ is a function of $x[kT]$ and $x[kT+T/2]$ which are not available. However, their estimates $\hat{x}[kT]$ and $\hat{x}[kT+T/2]$ are readily available by using the estimated output nonlinearity $\hat{x}=\sum_{i=1}^{m}\hat{r}_i y^i$ and the observed outputs $y[kT]$ and $y[kT+T/2]$. Let $\hat{x}$ and ϕ_2 denote the estimates of x and $\bar{\phi}_2$ using $\hat{x}$ instead of x, respectively, we have

$$\hat{x}[kT+T/2]=\phi_2'[k]\theta_2+v_2[k],\tag{3.10}$$

where $v_2[k]$ indicates the contribution due to the error between x and $\hat{x}$. This equation is again linear in the unknown parameter vector θ_2 and many standard estimation

algorithms apply, e.g., the (normalized) LMS algorithm

$$\hat{\theta}_2[k] = \hat{\theta}_2[k-1] + \frac{\phi_2[k]}{1+\phi_2'[k]\phi_2[k]}(\hat{x}[kT+T/2] - \phi_2'[k]\hat{\theta}_2[k-1]).$$

$\hat{\theta}_2$ consists of the estimates $(\hat{\beta}_1,\ldots,\hat{\beta}_n,\hat{\bar{\beta}}_0,\ldots,\hat{\bar{\beta}}_n)$ with $\hat{\beta}_1 = 1$. Therefore, combining Eq. (3.8), we obtain the estimates $\hat{G}(z)$ and $\hat{\bar{G}}(z)$ of $G(z)$ and $\bar{G}(z)$, respectively,

$$\hat{G}(z) = \frac{\hat{\beta}_1 z^{-1} + \hat{\beta}_2 z^{-2} + \cdots + \hat{\beta}_n z^{-n}}{1 - \hat{\alpha}_1 z^{-1} - \hat{\alpha}_2 z^{-2} - \cdots - \hat{\alpha}_n z^{-n}},$$

$$\hat{\bar{G}}(z) = \frac{\hat{\bar{\beta}}_0 + \hat{\bar{\beta}}_1 z^{-1} + \cdots + \hat{\bar{\beta}}_n z^{-n}}{1 - \hat{\bar{\alpha}}_1 z^{-1} - \hat{\bar{\alpha}}_2 z^{-2} - \cdots - \hat{\bar{\alpha}}_n z^{-n}},$$

where $\hat{\alpha}_i = \hat{\bar{\alpha}}_i,\ i = 1,\ldots,n$.

3.3. Input nonlinearity estimation

Since the structure of the input nonlinearity is unknown, estimation of the input nonlinearity relies completely on the graph information determined by the pairs of $(w[kT], u[kT])$. The input $w[kT]$ is known, but not $u[kT]$. Therefore, estimation of $u[kT]$ becomes a key in determining the input nonlinearity.

Recall that the input sampling interval is T and thus $u[kT] = u[kT + T/2]$. Also recall

$$X(z) = G(z)U(z), \quad \bar{X}(z) = \bar{G}(z)U(z).$$

If either $G(z)$ or $\bar{G}(z)$ is minimum phase, $U(z)$ and consequently, $u[kT]$ can be recovered easily

$$U(z) = G^{-1}(z)X(z) \quad \text{or} \quad U(z) = \bar{G}^{-1}(z)\bar{X}(z).$$

In time domain, these equations are

$$u[kT] = -\beta_2 u[kT-T] - \cdots - \beta_n u[kT-(n-1)T] + x[kT+T] - \alpha_1 x[kT] - \cdots - \alpha_n x[kT-(n-1)T] \tag{3.11}$$

or

$$u[kT] = \frac{1}{\bar{\beta}_0}(-\bar{\beta}_1 u[kT-T] - \cdots - \bar{\beta}_n u[kT-(n-1)T] + x[kT+T/2] - \bar{\alpha}_1 x[kT+T/2-T] - \cdots - \bar{\alpha}_n x[kT+T/2-nT]). \tag{3.12}$$

In these equations, both the estimates $\hat{\beta}(z)$ and $\hat{\bar{\beta}}(z)$ have been obtained and so are the estimates of $\hat{x}[kT]$ and $\hat{x}[kT + T/2]$. Thus, $\hat{u}[kT]$ can be calculated.

In the case that both $G(z)$ and $\bar{G}(z)$ are nonminimum phase, recovery of $u[kT]$ becomes problematic. To overcome this difficulty, assume that $G(z)$ and $\bar{G}(z)$ do not share any common zeros. Then, from the Bezout identity, there exist two stable transfer functions $F(z)$ and $\bar{F}(z)$ such that

$$F(z)\bar{G}(z) + \bar{F}(z)G(z) = 1. \tag{3.13}$$

This implies that

$$F(z)\bar{X}(z) + \bar{F}(z)X(z) = [F(z)\bar{G}(z) + \bar{F}(z)G(z)]U(z) = U(z) \tag{3.14}$$

and recovery of $U(z)$ or equivalently $u[kT]$ can be easily implemented by using $\hat{x}[kT]$ and $\hat{x}[kT + T/2]$. Once the estimate $\hat{u}[kT]$ of $u[kT]$ is obtained, the input nonlinearity $u[kT] = g(w[kT])$ can be graphed using the pairs $(w[kT], \hat{u}[kT])$ that provides complete information about the unknown input nonlinearity.

3.4. Algorithm and simulations

We are now in a position to summarize the identification algorithm for the Hammerstein–Wiener model with unknown structure of the input nonlinearity.

Identification algorithm.

Step 1. Consider the sampled Hammerstein–Wiener model in Fig. 1 described by (2.1), (2.3)–(2.6). For a given sampling interval T, collect input and output measurements $w[kT]$, $y[kh]$, $y[kT]$ and $y[kT + T/2]$, where $h = T/(n+1)$.

Step 2: Construct $\Delta y[l]$ and $\Delta\phi_1[l]$ as in (3.4) and estimate θ_1 using, e.g., the LMS algorithm. From θ_1, determine the inverse output nonlinearity $x = \sum_{i=1}^{m} \hat{r}_i y^i$. The estimate $y = \sum_{i=1}^{q} \hat{a}_i x^i$ of the forward output nonlinearity $y = f(x)$ is the best inverse of $x = \sum_{i=1}^{m} \hat{r}_i y^i$. Denote by $\hat{x}[kT] = \sum_{i=1}^{m} \hat{r}_i y^i[kT]$ and $\hat{x}[kT + T/2] = \sum_{i=1}^{m} \hat{r}_i y^i[kT + T/2]$, the estimates of $x[kT]$ and $x[kT + T/2]$, respectively.

Step 3: From $\hat{\theta}_1$, calculate the estimate $\hat{\alpha}(z)$ of $\alpha(z)$ using (3.8). Construct $\phi_2[k]$ as in (3.10) and the estimate $\hat{\theta}_2$. From $\hat{\theta}_2$, determine the estimates $\hat{G}(z)$, $\hat{\bar{G}}(z)$ of $G(z)$, $\bar{G}(z)$, respectively.

Step 4: If either $\hat{G}(z)$ or $\hat{\bar{G}}(z)$ is minimum phase, calculate $\hat{u}[kT]$ using either (3.11) or (3.12). If both $\hat{G}(z)$ and $\hat{\bar{G}}(z)$ are nonminimum phase, calculate $F(z)$ and $\bar{F}(z)$ as in (3.13) and calculate the estimate $\hat{u}[kT]$ using (3.14).

Step 5: Graph the input nonlinearity $u = g(w)$ by using the pairs $(w[kT], \hat{u}[kT])$. Estimate the nonlinearity based on the information provided by the graph. If necessary, parameterize the input nonlinearity using some base functions.

To illustrate the idea, we provide two numerical simulations.

Example 1. The linear part in this example is a first-order system

$$\dot{\eta} = -0.8\eta + 2.4266u, \quad x = \eta.$$

With $T = 1$, the discrete transfer function is given by

$$G(z) = \frac{1}{1 - 0.6703z^{-1}}.$$

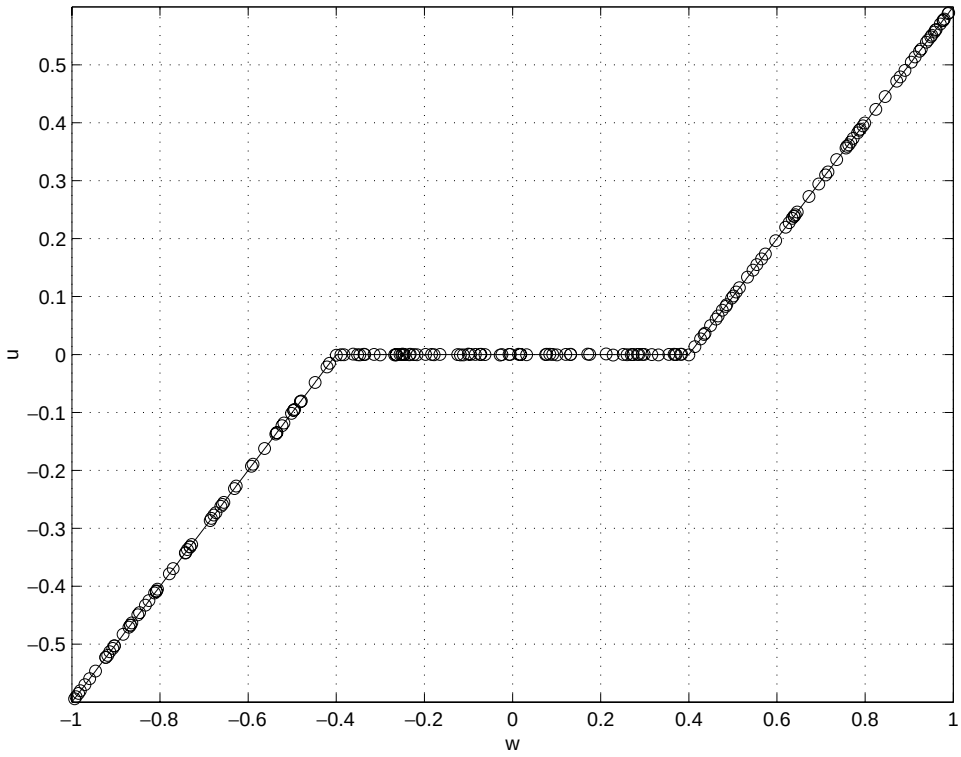

Fig. 2. Graph of the actual (solid) and estimated input deadzone (circle), Example 1.

Note that noninteger numbers are used in the model so that the normalization assumption $\beta_1 = 1$ and $r_1 = 1$ will be met.

The input nonlinearity is a deadzone with the threshold 0.4 shown in Fig. 2 in solid line. No a priori knowledge on the structure of the input nonlinearity is assumed in the simulation. The output nonlinearity is

$$y = f(x) = 0.18x + 1.37(e^{0.6x} - 1)$$

shown in Fig. 3 in solid line. Although nonpolynomial, the output nonlinearity can be approximated well by a second-order polynomial

$$y = 1.0287x + 0.2555x^2.$$

Of course, both the true output nonlinearity and its best polynomial approximation are unknown in the simulation. Instead, in the simulation, the inverse output nonlinearity $x = f^{-1}(y)$ is modeled by a fourth-order polynomial $x = \sum_{i=1}^{4} r_i y^i$. Then, the forward output nonlinearity $y = \sum_{i=1}^{2} a_i x^i$ is calculated by finding the best inverse of $x = \sum_{i=1}^{4} r_i y^i$.

For simulation, the input $w[kT],\ k = 1, \ldots, 500$, is a uniform i.i.d. random variable in $[-1, 1]$ and a uniform white noise distributed in $[-0.05, 0.05]$ is also added. Table 1 shows the simulation results. Fig. 2 shows the true input deadzone nonlinearity in solid line and the graph using the pairs of $(w[kT], \hat{u}[kT])$ in circles. Fig. 3 shows the true output nonlinearity in solid line and its estimate $y = 1.0291x + 0.2544x^2$ given by the proposed algorithm in circles. A satisfactory result is obtained.

Example 2. Consider a second-order linear continuous system

$$\dot{\eta} = \begin{pmatrix} 0 & 1 \\ -0.4 & -0.3 \end{pmatrix} \eta + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u,$$

$$x = [0.6556, 0.6556]\eta.$$

Both the input and output nonlinearities are polynomials

$$u = 1.5w - 1.2w^2 + 0.8w^3 + w^4,$$

$$y = 0.9962x + 0.0996x^2.$$

Again, noninteger numbers are used to normalize $\beta_1 = 1$ and $r_1 = 1$ for simulation purpose. The structure of the input nonlinearity is not assumed to be known. In the simulation, we model the inverse output nonlinearity $x = f^{-1}(y)$ by a fourth-order polynomial $x = \sum_{i=1}^{4} r_i y^i$. For simulation, the sampling interval $T = 1.2$ and input $w[kT],\ k = 1, \ldots, 500$, is a uniform i.i.d. random variable in $[-0.5, 0.5]$. A uniform white noise with magnitude 0.01 is also added in the simulation. With $T = 1.2$, the true but unknown $G(z)$ is

$$G(z) = \frac{1 - 0.2415z^{-1}}{1 - 1.2353z^{-1} + 0.7028z^{-2}}.$$

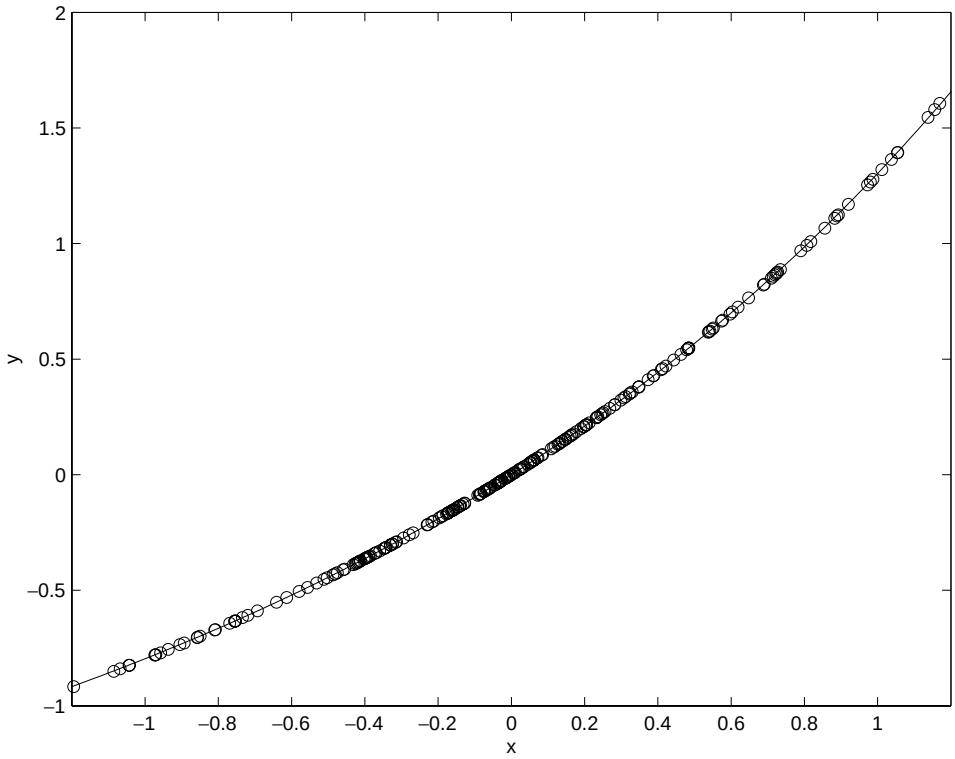

Fig. 3. Output nonlinearity (solid) and its estimate (circle).

Table 1
Simulation results of Example 1

	Linear part (numerator)	Linear part (denominator)	Output nonlinearity
True values	1	$[1, -0.6703]$	
Best approximation			$[1.0287, 0.2555]$
Estimates	1	$[1, -0.6713]$	$[1.0291, 0.2544]$

Table 2
Simulation results of Example 2

	Linear part (numerator)	Linear part (denominator)	Output nonlinearity
True values	$[1, -0.2415]$	$[1, -1.2353, 0.7028]$	$[0.0996, 0.9962]$
Estimates	$[1, -0.2445]$	$[1, -1.2367, 0.6977]$	$[0.097, 0.9986]$

Table 2 shows the simulation results using the proposed identification algorithm and Fig. 4 shows the true input nonlinearity in solid line and the graph using the pairs of $(w[kT], \hat{u}[kT])$ in circles.

4. Discussions

To avoid unnecessary complications so that the idea can be clearly conveyed, our attention was focused on presenting the basic algorithm. The algorithm can be in fact improved in several ways and we provide some discussions in this section.

4.1. Parameterization of the input nonlinearity $u = g(w)$

In the previous discussion, the structure of the input nonlinearity is assumed to be unknown and thus estimation relies on the graph given by the pairs $(w[kT], \hat{u}[kT])$. Once the picture of $u = g(w)$ is obtained, the structure of $u = g(w)$ can be determined. The next step is to parameterize this nonlinearity by using appropriate base functions, e.g.,

$$u = g(w) = \sum g_i(w, b_i)$$

for some known nonlinear functions g_i's and unknown coefficients b_i's. The choice of g_i's of course depends on the structure shown in the graph. For instance, in Example 1,

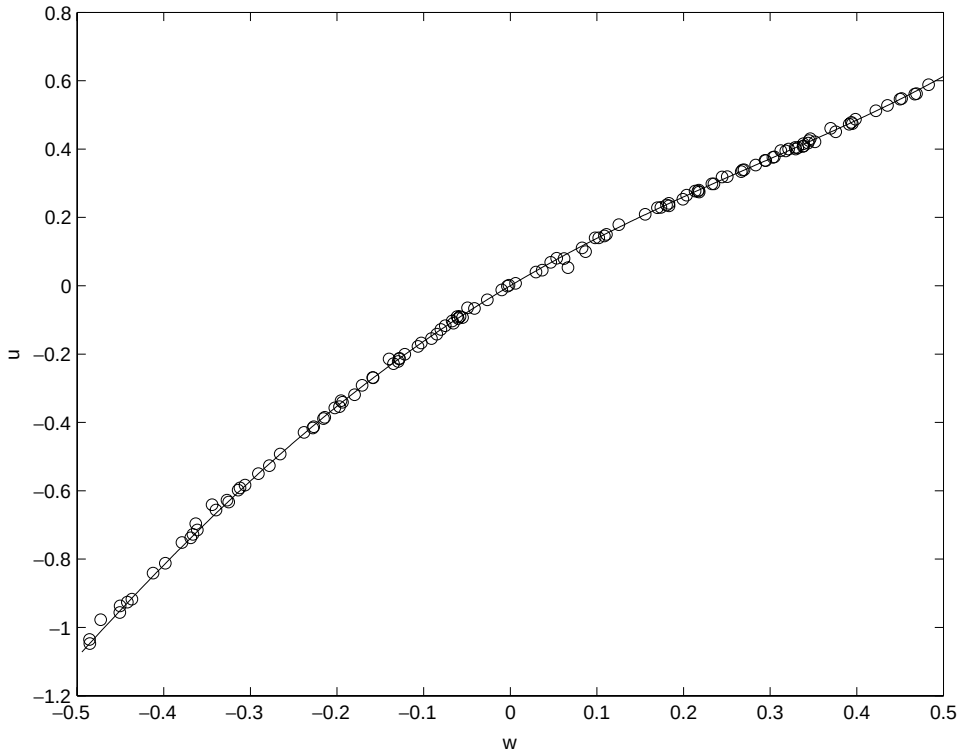

Fig. 4. Graph of the input nonlinearity, Example 2.

the graph clearly shows a deadzone input nonlinearity with some unknown threshold b. We can model this nonlinearity by the deadzone function

$$u = g(w) = w - b\,\mathrm{sgn}(w) - \frac{[1 + \mathrm{sgn}(b - |w|)]}{2} (w - b\,\mathrm{sgn}(w))$$

where sgn is the standard sign function and b is the unknown threshold which can be determined by

$$\hat{b} = \arg\min_b \sum_k \left\{ u[kT] - \left[w[kT] - b\,\mathrm{sgn}(w[kT]) - \frac{[1+\mathrm{sgn}(b-|w[kT]|)]}{2} (w[kT] - b\,\mathrm{sgn}(w[kT])) \right] \right\}^2.$$

Using the data generated in Example 1, the optimal estimate $\hat{b} = 0.3999$ is obtained and the actual $b = 0.4$.

In some cases, the input nonlinearity is assumed to be a polynomial,

$$u = g(w) = \sum_{i=1}^{p} b_i w^i.$$

Then, the coefficients b_i's can be determined by minimizing

$$b_i = \arg\min \sum_k \left(\hat{u}[kT] - \sum b_i w^i[kT] \right)^2.$$

This approach identifies the linear part and the output nonlinearity first. Then, using the estimated linear part, the estimated output nonlinearity and the observed output, we recover $u[kT]$ and identify the input nonlinearity. Another approach is to identify the input and output nonlinearities and the linear part together by over-parameterizing the whole system. To this end, recall the system equations. By substituting input and output nonlinearities into the system equation, we have

$$\sum_{i=1}^{m} r_i y^i[kT] = \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{m} r_j y^j[kT - iT] + \sum_{i=1}^{n} \beta_i \sum_{j=1}^{p} b_j w^j[kT - iT] + v[kT], \quad (3.15)$$

where $v[kT]$ is the contribution due to noise, model uncertainty and approximation errors. Since $r_1 = 1$, it follows that

$$\begin{aligned} y[kT] = (&-y^2[kT], \ldots, -y^m[kT], y[kT - T], \ldots, \\ &y^m[kT - T], \ldots, y[kT - nT], \ldots, y^m[kT - nT], \\ &w[kT - T], \ldots, w^p[kT - T], \ldots, w[kT - nT], \ldots, \\ &w^p[kT - nT]) \\ &\underbrace{(r_2, \ldots, r_m, \alpha_1, \alpha_1 r_2, \ldots, \alpha_1 r_m, \alpha_2, \alpha_2 r_2, \ldots, \alpha_2 r_m, \ldots, \\ &\beta_1 b_1, \ldots, \beta_1 b_p, \ldots, \beta_n b_1, \ldots, \beta_n b_p)'}_{\theta_3} \\ &+ v[kT] \end{aligned}$$

This equation involves only input–output measurements and is linear in θ_3. Therefore, any standard estimation algorithm applies.

It is clear that the estimates $\hat{r}_i$'s and $\hat{\alpha}_i$'s can be read directly from θ_3. However, the estimates $\widehat{\beta_i b_j}$'s, $i = 1,\ldots,n$, $j = 1,\ldots,p$, need to be projected into $\hat{\beta} = (\hat{\beta}_1,\ldots,\hat{\beta}_n)$ and $\hat{b} = (\hat{b}_1,\ldots,\hat{b}_p)$ with $\hat{\beta}_1 = 1$ such that

$$\sum_{i,j} (\widehat{\beta_i b_j} - \hat{\beta}_i \hat{b}_j)^2$$

is minimized. This is equivalent to

$$(\hat{\beta},\hat{b}) = \arg\min \left\| \underbrace{\begin{pmatrix} \widehat{\beta_1 b_1} & \ldots & \widehat{\beta_1 b_p} \\ \vdots & \ddots & \vdots \\ \widehat{\beta_n b_1} & \ldots & \widehat{\beta_n b_p} \end{pmatrix}}_{\Theta} - \hat{\beta}\hat{b}' \right\|_{\mathrm{F}} \quad \text{with } \hat{\beta}_1 = 1, \tag{3.16}$$

where F stands for the matrix Frobenius norm. This is exactly the same minimization problem as in (3.6) and the solution is given by

$$\hat{\beta} = \frac{1}{s_\xi}\xi_1, \quad \hat{b} = \sigma_1 s_\xi \eta_1,$$

where s_ξ is the first entry of ξ_1 and $\sum_{i=1}^{\min(n,p)} \sigma_i \xi_i \eta_i'$ is the singular value decomposition(SVD) of Θ.

The over-parameterization method is fairly straightforward. A disadvantage is that the dimension of identification is usually very high. To this end, the iterative method may be useful. The idea of the iterative method is reminiscent of the iterative method for the Hammerstein model (Narendra & Gallman, 1966) and for the Wiener model (Kalafatis et al., 1997). The parameter set is divided into two parts, linear part and nonlinear part. The linear part is estimated while the nonlinear part is fixed. Then, two sets are switched to evaluate the nonlinear part while the linear part is fixed.

Define the cost function

$$J(\hat{\alpha},\hat{\beta},\hat{r},\hat{b}) = \sum_k \left\{ \sum_{i=1}^m \hat{r}_i y^i[kT] - \left[\sum_{i=1}^n \hat{\alpha}_i \sum_{j=1}^m \hat{r}_j y^j[kT - iT] + \sum_{i=1}^n \hat{\beta}_i \sum_{j=1}^p \hat{b}_j w^j[kT - iT] \right] \right\}^2 \tag{3.17}$$

and the iterative method can be summarized as follows:

(a) Consider the system and collect data.
(b) Set initial value $\hat{r}(0)$ and $\hat{b}(0)$, and let $i = 1$.
(c) For fixed $\hat{r}(i-1)$ and $\hat{b}(i-1)$, find

$$\begin{pmatrix} \hat{\alpha}(i) \\ \hat{\beta}(i) \end{pmatrix} = \arg\min_{\alpha,\beta} J(\alpha,\beta,\hat{r}(i-1),\hat{b}(i-1)).$$

Normalize the estimates such that $\hat{\beta}_1(i) = 1$.
(d) For fixed $\hat{\alpha}(i)$ and $\hat{\beta}(i)$, determine

$$\begin{pmatrix} \hat{r}(i) \\ \hat{b}(i) \end{pmatrix} = \arg\min_{r,b} J(r,b,\hat{\alpha}(i),\hat{\beta}(i)).$$

Normalize the estimates such that $\hat{r}_1(i) = 1$.
(e) Let $J_i = J(\hat{\alpha}(i),\hat{\beta}(i),\hat{r}(i),\hat{b}(i))$. If $\Delta J_i = J_i - J_{i-1}$ is smaller than some prescribed threshold, go to next step. Otherwise, set $i = i + 1$ and go to Step 3.
(f) From $\hat{r}$, calculate $\hat{a}$. The final estimates are $\hat{a}$, $\hat{b}$, $\hat{\alpha}$ and $\hat{\beta}$.

Note that the cost function is bilinear and this implies that minimization at Step 3 or Step 4 is a simple linear least-squares problem and can be solved efficiently.

Although there is no guarantee for the global convergence, this iterative method is expected to work just like its counterpart for Hammerstein (Narendra & Gallman, 1966) or Wiener (Kalafatis et al., 1997) models where it has been demonstrated that the iterative method is usually very effective and converges quickly. Moreover, divergence is rare (Stoica, 1981).

4.2. Output nonlinearity order estimation

In the proposed algorithm, the order m of the inverse output nonlinearity $x = \sum_{i=1}^m r_i y^i$ is assumed to be known. In practice, m is unknown and needs to be estimated using the on-line data. A number of standard methods of order estimation for linear systems, e.g, rank test and the output error test, find their applications here. Interested readers can find details in (Ljung, 1987; Soderstrom & Stoica, 1989). We focus on the output error test method in this paper. Suppose that the output nonlinearity and the linear transfer function have been estimated in terms of their coefficients $\hat{r}(m) = (\hat{r}_1,\ldots,\hat{r}_m)$, $\hat{a}(m) = (\hat{a}_1,\ldots,\hat{a}_q)$, $\hat{\beta}(m) = (\hat{\beta}_1,\ldots,\hat{\beta}_n)$, $\hat{\alpha}(m) = (\hat{\alpha}_1,\ldots,\hat{\alpha}_n)$ respectively for a fixed order m. Here m in the brackets indicates the dependence of estimates on the order m. Further, suppose that the input nonlinearity $u = g(w)$ has been parameterized, say by a polynomial $u = \sum_{i=1}^p \hat{b}_i w^i$. Define the estimated output

$$\bar{u}[kT] = \sum_{i=1}^p \hat{b}_i w^i[kT],$$

$$\bar{x}[kT] = \sum_{j=1}^n \hat{\alpha}_j(m)\bar{x}[kT - jT] + \sum_{j=1}^n \hat{\beta}_j(m)\bar{u}[kT - jT],$$

$$\bar{y}_m[kT] = \sum_{i=1}^q \hat{a}_i(m)\bar{x}^i[kT]$$

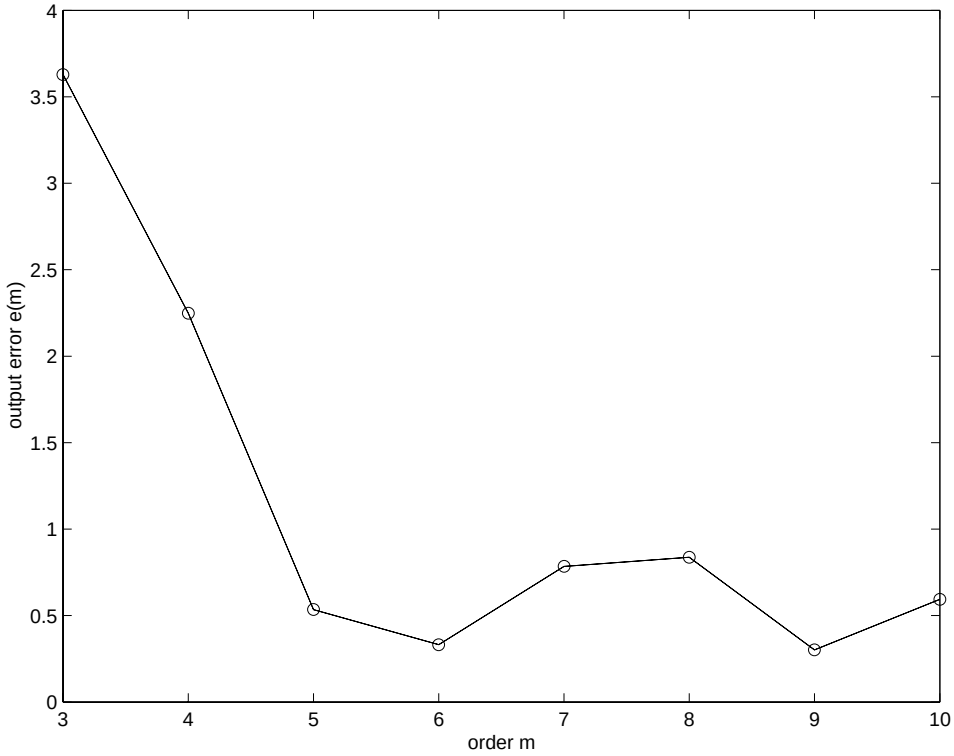

Fig. 5. Output error versus order, Example 3.

Table 3
Simulation results when $m = 5$, Example 3

	Linear part (denominator)	Output nonlinearity	Input nonlinearity
True values	$[1, -0.6703]$	$[1.059, 0.4236]$	$[1, 1, 2]$
Estimates	$[1, -0.6724]$	$[1.0410, 0.4146]$	$[1.0073, 1.0254, 1.9916]$

and the output error

$$e(m) = \sqrt{\frac{1}{N}\sum_{k=1}^{N} (\bar{y}_m[kT] - y[kT])^2},$$

where $y[kT]$ is the observed output. Now, consider the case where the order m increases. Note that $x = \sum_{i=1}^{m} r_i y^i$ approximates $x = f^{-1}(y)$. With more free parameters in the model, a better fit is expected, i.e., a smaller $e(m)$ is expected as m increases. The important thing is to investigate whether or not the improvement is significant. In a noise free and PE input case, if the order m is high enough to adequately describe the inverse function $x = f^{-1}(x)$, any increment in m only produce a small reduction in $e(m)$. In a noisy situation, the output error $e(m)$ may increase as m increases if m is already higher than the one that adequately describes the inverse function due to noise, computational error or the level of PE. Therefore, if $\Delta e(m) = e(m) - e(m-1)$ is small, the order m should be chosen. Otherwise, a higher order is preferred. We provide an example here.

Example 3. Let the continuous time system be a first order

$$\dot{x} = -0.88x + 2.4266u.$$

The unknown input and output nonlinearities are, respectively

$$u = w + w^2 + 2 * w^3, \quad y = 1.059x + 0.4236x^2.$$

With the same simulation conditions as in Example 2, Fig. 5 shows the output error $e(m)$ versus the order m. Clearly, $m = 5$ or 6 should be chosen. When $m = 5$, the true coefficients of $G(z)$, the input and output nonlinearities as well as their estimates are given in Table 3. Note that the numerator is normalized to $\beta(z) = 1$ because of the first-order transfer function. In this example, the unknown input nonlinearity is estimated and then is parameterized by a polynomial.

4.3. Inverse parameterization of the output nonlinearity

The actual output nonlinearity $y = f(x)$ is unknown and we use the inverse parameterization $x = \sum r_i y^i$. This

inverse approach has been used in the literature (Kalafatis et al., 1997; Pajunen, 1992) to model Wiener systems. Here, we adopt this approach to model Hammerstein–Wiener systems with unknown input nonlinearity structure. Because of the blind method, this inverse approach makes identification of the Hammerstein–Wiener model feasible even with the unknown structure of the input nonlinearity and non-white inputs. Clearly, the success of the proposed algorithm hinges on the accuracy between the approximation $x=\sum r_i y^i$ and the true $x=f^{-1}(y)$. Theoretically, as long as $y=f(x)$ is one-to-one and continuous, $x=\sum r_i y^i$ approximates $x=f^{-1}(y)$ to any accuracy as the order increases. Practically, however, a high order m introduces errors due to noise and model uncertainty and slows down the convergence rate. There is balance between the errors introduced by the approximation $x=\sum r_i y^i$ and the errors due to noise and model uncertainty. Therefore, the inverse approach is not a universal approach and probably would not work if the order of the approximation $x=\sum r_i y^i$ needs to be very high. The information whether or not the inverse approach is appropriate for a particular system which is unknown is in fact contained in the output error $e(m)$ discussed before. If $e(m)$ is uniformly large, the inverse approach does not work well. On the other hand, small $e(m)$ for some m indicates the success of the inverse approach together with the blind method.

4.4. Persistent excitation conditions

To have a robust identification algorithm in the presence of noise and model uncertainty, the regressors $\Delta\phi_1$ and ϕ_2 need to be PE. The conditions that ϕ_2 is PE are derived in Bai and Fu (1999), which basically say that ϕ_2 is PE if the spectral lines of $u[kT]$ is not concentrated on $<2n$ points, a richness condition. This richness condition can also be translated into conditions in terms of the input $w[kT]$. Suppose that $w[kT]$ is i.i.d. zero mean random variable and $u[kT]=g(w[kT])$ assumes at least two distinctive values with nonzero probability. Then, $u[kT]$ is also i.i.d. and has infinitely many spectral lines which implies that ϕ_2 is PE. The second scenario is the polynomial input nonlinearity $u=\sum b_i w^i$ and sinusoidal input $w[kT]=\sum c_i \sin(\Omega_i k)$. This is certainly the case if the input is periodic by the Fourier series representation. Then, $u[kT]$ has more than $2n$ spectral lines if $w[kT]$ has $2n$ spectral lines unless in a pathological case where either coefficients are zero or frequencies are of the same module 2π.

4.5. Choice of the over-sampling ratio l

In the algorithm, the over-sampling ratio $l=(n+1)$ is assumed, where n is of the order of $G(z)$. This seems to imply that the order of $G(z)$ has to be known a priori. In fact, l does not have to be $(n+1)$ and any $l>(n+1)$ suffices (Bai & Fu, 1999). Note that the key of the blind approach is to cancel the unknown signals u from Eq. (3.3) to form Eq. (3.4). This is possible for any $l\geqslant(n+1)$ because if $h=T/(\bar{n}+1)$ or $T=h(\bar{n}+1)$ for any $\bar{n}\geqslant n$, we have

$$u[(l-1)T]=u[(l-1)T+h]=\cdots=u[(l-1)T+\bar{n}h],\quad l=1,2,\ldots\ .$$

Hence, Eq. (3.4) follows. Of course, details of the algorithm including equations are not exactly the same when $l>(n+1)$ instead of $l=(n+1)$. However, the idea remains the same and all modifications are minor.

In theory, as long as $l\geqslant(n+1)$, blind identification is possible. In practice, however, there is a limit on how large l can be. For a very large l or equivalently a very small h, two consecutive samples will likely have similar values and this makes the blind identification numerically ill-conditioned. A good choice of h is

$$\frac{1}{2f_y}\leqslant h\leqslant\frac{T}{(n+1)},$$

where f_y is the bandwidth of the output y. Clearly from the sampling theorem, $h=1/2f_y$ implies that $y(t)$ can be completely determined from $y[kh]$ and further increasing l or equivalently reducing h will not provide any additional information and will only make the algorithm ill-conditioned.

The over-sampling approach is to fix the input sampling interval at T and over-sample the output at h so that $l=T/h=(\bar{n}+1)\geqslant(n+1)$. Another avenue to make $l=(\bar{n}+1)\geqslant(n+1)$ is the under-sampling approach. By letting the output sampling interval $h=T$ and keeping the input constant between $k(\bar{n}+1)T$ and $(k+1)(\bar{n}+1)T$ for each k, we have $l=(\bar{n}+1)$. This under-sampling approach would avoid the numerical instability problem at a price that (1) the utilization of time is less efficient prolonging the identification process and (2) the system may be excited only at low-frequency ranges. It is conceivable that an "optimal" way to achieve $l\geqslant(n+1)$ in some cases could be the one that combines both the over-sampling and the under-sampling approaches.

4.6. Relation with the step response identification

In a way, the blind technique presented in this paper may be considered as repeatedly applying piece-wise constant inputs. Conceptually, a number of step responses could be used to give information first on the output nonlinearity and the linear part, and then of the input nonlinearity. However, the blind technique works fundamentally different from the traditional step response identification method (Rake, 1980). The traditional step response method relies heavily on the steady-state value $y(t)$, $t\to\infty$, of the step response and would suffer from large noises at the end of transient. This is specially true in the setting of parametric identification and therefore, it is suggested to apply the step response identification method several times to average out the effect of noises (Rake, 1980).

Clearly, with only the output measurements $y[kT]$, blind identification is not possible if both the input and output sampling intervals are fixed at T. This is because two different sampled systems combined with properly chosen inputs could provide the identical output $y[kT]$. With additional intermediate values $y[kh]$ ($h \leqslant T/(n+1)$) between kT and $(k+1)T$, however, the choice of the system becomes unique. This is the basic observation that tells us why and how the blind technique works. Now, with the output observations over each T, an equation related to the unknown parameters (3.4) is derived. Obviously from (3.4), the blind technique does not rely heavily on any particular value of the output observation but depends on each one $y[kh]$ equally.

4.7. Identifiability

With the PE inputs and assumptions 1–3, the identifiability of the Hammerstein–Wiener model shown in Fig. 1 can be easily established. Identifiability here means that the representation of the system is unique in the absence of noise and model uncertainties. This can be seen easily. With the PE regressors, the solutions of (3.4) and (3.10) are unique. Moreover, the true but unknown system parameters are solutions in the absence of noises and model uncertainties. This establishes the identifiability.

5. Concluding remarks

In this paper, we proposed a blind identification approach to sampled Hammerstein–Wiener models with the structure of the input nonlinearity unknown. The idea is to recover the internal signals $u[kT]$ and $x[kT]$ solely based on the output measurements. This is essential because the input nonlinearity has an unknown structure. The purpose of the paper is to present the main idea and to illustrate the effectiveness of the proposed approach. Some important topics were not discussed in the paper, e.g., how exactly the noise would influence the estimates in blind identification? This is an interesting and difficult question, and it gets more difficult to separate the effects of the noise from the under-parameterization in inverting the output nonlinearity. We expect that the findings will be quite different from the existing results on the (nonblind) system identification. Another important issue is the application of the proposed method to a real world problem.

References

Bai, E. W. (2001). Identification of systems with hard input nonlinearities. In R. Moheimani (Ed.), *Perspectives in Control*. Berlin: Springer.

Bai, E. W. (1998). An optimal two-stage identification algorithm for Hammerstein–Wiener nonlinear systems. *Automatica*, *34*(3), 333–338.

Bai, E. W., & Fu, M. (1999). Blind system identification and channel equalization of IIR systems without statistical information. *IEEE Transactions on Signal Processing*, *47*(7), 1910–1921.

Bai, E. W., & Fu, M. (2001). *Hammerstein model identification: A blind approach*. Technical Report, Department of Electronics and Computer, University of Iowa.

Bilings, S. A., & Fakhouri, S. Y. (1978). Identification of a class of nonlinear systems using correlation analysis. *Proceedings of IEE*, *125*(7), 691–697.

Boutayeb, M., Rafaralahy, H., & Darouach, M. (1996). A robust and recursive identification method for Hammerstein model. *IFAC World Congress '96*, San Francisco (pp. 447–452).

Chang, F., & Luus, R. (1971). A non-iterative method for identification using Hammerstein model. *IEEE Transactions on Automatic Control*, *16*, 464–468.

Greblicki, W. (1992). Nonparametric identification of Wiener system. *IEEE Transactions on Information Theory*, *38*, 1487–1493.

Hsia, T. (1976). A multi-stage least squares method for identifying Hammerstein model nonlinear systems. *Proceedings of CDC*, Clearwater, FL (pp. 934–938).

Haber, R., & Unbehauen, H. (1990). Structure identification of nonlinear dynamic systems—A survey of input/output approaches. *Automatica*, *26*, 651–677.

Kalafatis, A. D., Wang, L., & Cluett, W. R. (1997). Identification of Wiener-type nonlinear systems in a noisy environment. *International Journal of Control*, *66*, 923–941.

Ljung, L. (1987). *System Identification*. Englewood Cliffs, NJ: Prentice-Hall.

Narendra, K. S., & Gallman, P. G. (1966). An iterative method for the identification of nonlinear systems using a Hammerstein model. *IEEE Transactions on Automatic Control*, *11*, 546–550.

Pajunen, G. (1992). Adaptive control of Wiener type nonlinear systems. *Automatica*, *28*, 781–785.

Pawlak, M. (1991). On the series expansion approach to the identification of Hammerstein systems. *IEEE Transactions on Automatic Control*, *36*, 763–767.

Rake, H. (1980). Step response and frequency response methods. *Automatica*, *16*, 519–526.

Soderstrom, T., & Stoica, P. (1989). *System Identification*. Englewood Cliffs, NJ: Prentice-Hall.

Stoica, P. (1981). On the convergence of an iterative algorithm used for Hammerstein system identification. *IEEE Transactions on Automatic Control*, *26*, 967–969.

Sun, L., Liu, W., & Sano, A. (1998). Identification of dynamical system with input nonlinearity. *IEE Proc-Control Theory Application*, *146*(1), 41–51.

Wigren, T. (1994). Convergence analysis of recursive identification algorithms based on the nonlinear Wiener model. *IEEE Transactions on Automatic Control*, *39*, 2191–2206.

Er-Wei Bai was educated in Fudan University, Shanghai Jiaotong University and the University of California at Berkeley. Dr. Bai holds the rank of Professor at the University of Iowa where he teaches and conducts Research in Controls and Signal Processing. He serves the IEEE CSS and the IFAC in various capacities.