INT. J. CONTROL, 2003, VOL. 76, NO. 15, 1500–1507

Taylor & Francis
Taylor & Francis Group

Identification of FIR Wiener systems with unknown, non-invertible, polynomial non-linearities

SETH L. LACY†* and DENNIS S. BERNSTEIN‡

Wiener systems consist of a linear dynamic system whose output is measured through a static non-linearity. In this paper we study the identification of single-input single-output Wiener systems with finite impulse response dynamics and polynomial output non-linearities. Using multi-index notation, we solve a least squares problem to estimate products of the coefficients of the non-linearity and the impulse response of the linear system. We then consider four methods for extracting the coefficients of the non-linearity and impulse response: direct algebraic solution, singular value decomposition, multi-dimensional singular value decomposition and prediction error optimization.

1. Introduction

Non-linear system identification remains one of the most challenging and potentially useful problem areas in system theory. Numerous approaches have been developed for this problem, including black box and grey box techniques (Bayard and Eslami 1984, Pajunen 1985, Hunter and Korenberg 1986, Korenberg and Hunter 1986, Hasiewicz 1987, Chen and Fassois 1992, 1997, Greblicki 1992, 1994, 1997, 1998, Westwick and Kearney 1992, Wigren 1994, Westwick and Verhaegen 1996, Bai 1998, Lovera *et al*. 2000, Van Pelt and Bernstein 2000, Lacy and Bernstein 2001, Lacy *et al*. 2001, Nelles 2001). The grey box case includes the identification of block-structured models, such as the Hammerstein model (linear system with input non-linearity) and Weiner model (linear system with output non-linearity) (Haber and Keviczky 1999 a, b).

This paper is concerned with identifying Wiener systems under more general assumptions than have been previously considered. Many methods for Wiener system identification require the non-linearity to be known, invertible, differentiable, odd or require specially designed input sequences. In particular, the Wiener identification problem has been considered in Brillinger (1970), Pajunen (1985), Hasiewicz (1987), Greblicki (1992, 1994, 1997, 1998), Westwick and Kearney (1992), Wigren (1994), Westwick and Verhaegen (1996), Bai (1998), and Lovera *et al*. (2000) under the assumption that the non-linearity is unknown but one-to-one. This assumption simplifies the problem considerably since the inverse system can be viewed as a Hammerstein system wherein the input to the non-linearity is measured. If the non-linearity is known but not one-to-one, then identification is possible by first generating a candidate set of signals at the output of the linear system (Bayard and Eslami 1984, Lacy *et al*. 2001). These methods are applicable even if the output non-linearity is a step function in which case the output assumes at most two distinct values (Lacy *et al*. 2001). If the input sequence can be chosen freely, the frequency content of the input sequence can be selected such that the effect of the non-linearity can be derived from the frequency content of the output (Pintelon and Schoukens 2001).

In the present paper we consider Wiener system identification in which the output non-linearity is both unknown and not necessarily one-to-one. In this case the goal is to simultaneously identify both the linear system dynamics and the non-linearity despite the non-invertibility of the output non-linearity. To do this we assume that the non-linearity can be represented as a finite sum of polynomials. We use multi-index notation (Evans 1998, Dunkl and Xu 2001) to expand this polynomial and write the output as a linear-in-parameters sum of known terms with unknown coefficients. These coefficients consist of products of the system parameters. We then present several methods for extracting the system parameters. This approach of expanding the polynomial output non-linearity requires that the linear system output depend only on past inputs, that is, the linear dynamics are assumed to be FIR.

This paper is organized as follows. In § 2 we define the problem and list the assumptions. In § 3 we introduce notation used throughout the paper. In § 3.1 we present an algebraic solution. In § 3.2 we present a solution based on a singular value decomposition. In § 3.3 we present a solution based on a multi-dimensional singular value decomposition (Andersson and Bro 2000). In § 3.4 we present an optimality approach based on a prediction-error cost function. In § 4 we apply all of these methods to an example to illustrate their implementation and compare their effectiveness.

Received 1 August 2000. Revised 1 June 2003. Accepted 1 June 2003.

* Author for correspondence. seth.lacy@kirtland.af.mil
† Air Force Research Laboratory, Space Vehicles Directorate, Kirtland AFB, NM 87117, USA.
‡ Aerospace Engineering Department, University of Michigan, Ann Arbor, MI 48109, USA.

International Journal of Control ISSN 0020–7179 print/ISSN 1366–5820 online © 2003 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/00207170310001599122

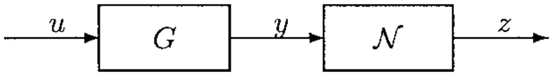


Figure 1. Wiener system.

2. Problem description

Here we study the identification of a single-input, single-output, linear time-invariant system with finite impulse response whose output is measured through a static non-linearity. This system, which is represented in figure 1, is modelled by

$$y(k) = (h_0 u(k) + h_1 u(k-1) + \cdots + h_m u(k-m) = \sum_{i=0}^{m} h_i u(k-i) \quad (1)$$

$$z(k) = \mathcal{N}(y(k)) \quad (2)$$

where u is the input to the system, y is the unmeasured output of the linear system, h_i are Markov parameters, and z is the measured output of the non-linearity. We assume that the non-linear function $\mathcal{N}: \mathbb{R} \to \mathbb{R}$ is a polynomial of the form

$$\mathcal{N}(y(k)) = \sum_{i=0}^{p} c_i y(k)^i \quad (3)$$

If $\mathcal{N}$ is not a polynomial, then (3) can be regarded as an approximation. We assume that the order m of the FIR dynamics and the degree p of the polynomial $\mathcal{N}$ are known. The non-linearity $\mathcal{N}$ is otherwise unknown and not necessarily one-to-one. In the practical situation m and p are not known. Also, it may be difficult to obtain satisfactory results if these parameters are under-estimated. However, if upper bounds on these parameters are known, then the bounds can be used in (1) and (3) at the expense of increasing the computational complexity. Alternatively, the values for p and $\mathcal{N}$ can be incremented until satisfactory performance is achieved, at the expense of increasing the computational load at each increment.

The identification problem is to estimate the coefficients h_i and c_i using ℓ measurements of u and z. We adopt a two-stage approach. First, we solve a least squares problem to obtain an estimate $\hat{\boldsymbol{\theta}}$ of a vector $\boldsymbol{\theta}$ whose entries are the unknown parameters and products of the unknown parameters. Next, we present several techniques that use $\hat{\boldsymbol{\theta}}$ to estimate the individual unknown parameters. In addition, we minimize a prediction error cost function to further refine the parameter estimates and compare to the direct approaches.

3. Wiener identification

Using (3), we rewrite equation (2) as

$$\begin{aligned} z(k) &= \sum_{i=0}^{p} c_i y(k)^i = \sum_{i=0}^{p} c_i \left(\sum_{j=0}^{m} h_j u(k-j) \right)^i \\ &= \sum_{i=0}^{p} c_i \sum_{|\alpha|=i} \frac{|\alpha|!}{\alpha!} h^\alpha v(k)^\alpha = \sum_{|\alpha|=0}^{p} \frac{|\alpha|!}{\alpha!} v(k)^\alpha c_{|\alpha|} h^\alpha \\ &= \phi(k)^{\mathrm{T}} \boldsymbol{\theta} \end{aligned} \quad (4)$$

where

$$h = [h_0 \quad h_1 \quad \cdots \quad h_m]^{\mathrm{T}} \in \mathbb{R}^{m+1} \quad (5)$$

$$\alpha = [\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_{m+1}]^{\mathrm{T}} \in \mathbb{N}_0^{m+1} \quad (6)$$

$$v(k) = [u(k) \quad u(k-1) \quad \cdots \quad u(k-m)]^{\mathrm{T}} \in \mathbb{R}^{m+1} \quad (7)$$

$$\boldsymbol{\theta} = [c_{|\alpha|} h^\alpha]_{|\alpha| \le p} \in \mathbb{R}^{D_p^{m+1}} \quad (8)$$

$$\phi(k) = \left[\frac{|\alpha|!}{\alpha!} v(k)^\alpha \right]_{|\alpha| \le p} \in \mathbb{R}^{D_p^{m+1}} \quad (9)$$

and $\alpha \in \mathbb{N}_0^{m+1}$ is a multi-index whose order is $m+1$, where $\mathbb{N}_0$ is the set of positive integers and zero. A multi-index is a vector whose components are non-negative integers (see Evans 1998, Dunkl and Xu 2001). We define

$$|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_{m+1} = \sum_{i=1}^{m+1} \alpha_i \quad (10)$$

$$\alpha! = (\alpha_1!)(\alpha_2!) \cdots (\alpha_{m+1}!) = \prod_{i=1}^{m+1} \alpha_i! \quad (11)$$

$$v(k)^\alpha = v_1(k)^{\alpha_1} v_2(k)^{\alpha_2} \cdots v_{m+1}(k)^{\alpha_{m+1}} = \prod_{i=1}^{m+1} v_i(k)^{\alpha_i} \quad (12)$$

$$h^\alpha = h_1^{\alpha_1} h_2^{\alpha_2} \cdots h_{m+1}^{\alpha_{m+1}} = \prod_{i=1}^{m+1} h_i^{\alpha_i} \quad (13)$$

The number of multi-indices of order $m+1$ of fixed absolute value i is given by

$$C_i^{m+1} = \binom{m+i}{i} = \binom{m+i}{m} = \frac{(m+i)!}{m!i!} \quad (14)$$

and the number of multi-indices of order $m+1$ of absolute value less than or equal to p is given by

$$D_p^{m+1} = \sum_{i=0}^{p} C_i^{m+1} = \sum_{i=0}^{p} \frac{(m+i)!}{m!i!} \quad (15)$$

We need to define an order relation for multi-indices of the same order. Let $\alpha,\ \beta \in \mathbb{N}_0^{m+1}$ be multi-indices. If $|\alpha| > |\beta|$ then $\alpha > \beta$. If α and β have the same order,

$|\alpha| = |\beta|$, then we choose the standard dictionary ordering. The notation

$$[f(\alpha)]_{|\alpha|\le p} = \begin{bmatrix} [f(\alpha)]_{|\alpha|=0} \\ [f[(\alpha)]_{|\alpha|=1} \\ \vdots \\ [f(\alpha)]_{|\alpha|=p} \end{bmatrix} \tag{16}$$

denotes the column vector whose components are f evaluated at every multi-index α such that $|\alpha| \le p$. The components are ordered according to the above ordering scheme. This vector has D_p^{m+1} components. Thus $\boldsymbol{\theta}$ and $\phi(k)$ have D_p^{m+1} components.

To estimate $\boldsymbol{\theta}$ we rewrite (4) as

$$\mathbf{z} = \Phi^{\mathrm{T}}\boldsymbol{\theta} \tag{17}$$

where

$$\mathbf{z} = [z(m+1) \quad \cdots \quad z(\ell)]^{\mathrm{T}} \in \mathbb{R}^{\ell-m} \tag{18}$$

$$\Phi = [\phi(m+1) \quad \cdots \quad \phi(\ell) \in \mathbb{R}^{D_p^{m+1}\times\ell-m} \tag{19}$$

We assume $\Phi\Phi^{\mathrm{T}}$ is non-singular, which is a persistency of excitation condition that requires $\ell \ge D_p^{m+1} + m$. Then we calculate the least squares estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ given by

$$\hat{\boldsymbol{\theta}} = (\Phi\Phi^{\mathrm{T}})^{-1}\Phi\mathbf{z} \tag{20}$$

Next we develop several methods for obtaining estimates $\hat{c}$ and $\hat{h}$ based on $\hat{\boldsymbol{\theta}}$. Note that an arbitrary scaling and its reciprocal can be applied to the linear system and the output non-linearity. We remove this ambiguity by introducing a normalization constraint, thereby selecting a single system from a class of equivalent systems. We can normalize $\hat{c}$ and $\hat{h}$ by setting $\hat{c}_i = a$, $\hat{h}_i = a$, $\|c\| = a$, $\|h\| = a$ or various other constraints.

3.1. *Direct solve*

We have $m + p + 2$ unknown parameters in h and c, one normalization constraint and D_p^{m+1} equations in terms of $\hat{\boldsymbol{\theta}}$. We can normalize as above, then choose $m + p + 1$ independent equations. These $m + p + 1$ equations must also be independent of the constraint equation, which constitutes equation number $m + p + 2$. A symbolic manipulator such as Mathematica can be used to invert these non-linear equations and obtain estimates $\hat{h}$ and $\hat{c}$ of h and c.

3.2. *SVD*

To begin, $\hat{c}_0$ can be estimated directly by

$$\hat{c}_0 = \hat{\boldsymbol{\theta}}(1) \tag{21}$$

To estimate the remaining components of $\hat{c}$ and $\hat{h}$, we arrange the components of $\boldsymbol{\theta}$ into the matrix $A(\boldsymbol{\theta}) \in \mathbb{R}^{m+1\times D_{p-1}^{m+1}}$, where

$$A(\boldsymbol{\theta}) = \psi h^{\mathrm{T}} \tag{22}$$

and

$$\psi = [c_{|\alpha|+1}h^{\alpha}]_{|\alpha|<p} \tag{23}$$

Then we calculate the singular value decomposition

$$A(\hat{\boldsymbol{\theta}}) = USV^{\mathrm{T}} \tag{24}$$

to obtain the estimates

$$\hat{h} = \sigma S(1,1)V(1,:) \tag{25}$$

$$\hat{\psi} = \frac{1}{\sigma}U(1,:) \tag{26}$$

where the scalar $\sigma \ne 0$ selects the normalization constraint. Finally, we extract the non-linearity coefficients $\hat{c}_i$ from $\hat{\psi}$. Specifically, $\hat{c}_1$ is given directly by $\hat{\psi}(1)$, while the remaining coefficients are calculated using least squares estimation and $\hat{h}$.

3.3. *Multi-dimensional SVD*

First, we define the tensors $A_0 \in \mathbb{R}$, $A_j \in \times_{i=1}^{j}\mathbb{R}^{n+1}$

$$A_0(\boldsymbol{\theta}) = c_0 \tag{27}$$

$$A_1(\boldsymbol{\theta}) = c_1 h = h_{A_1} \tag{28}$$

$$A_2(\boldsymbol{\theta}) = c_2 h \circ h = \circ_{i=1}^{2} h_{A_2} \tag{29}$$

$$A_3(\boldsymbol{\theta}) = c_3 h \circ h \circ h = \circ_{i=1}^{3} h_{A_3} \tag{30}$$

$$A_p(\boldsymbol{\theta}) = c_p \circ_{i=1}^{p} h = \circ_{i=1}^{p} h_{A_p} \tag{31}$$

where

$$h_{A_n} = c_n^{1/n} h \tag{32}$$

We use a multi-dimensional singular value decomposition (Andersson and Bro 2000) to obtain the estimate $\hat{h}_{A_i}$ of h_{A_i}. To do this we note that

$$[h_{A_1} \quad h_{A_2} \quad \cdots \quad h_{A_p}] = hd^{\mathrm{T}} \tag{33}$$

where

$$d = [c_1 \quad c_2^{1/2} \quad \cdots \quad c_p^{1/p}]^{\mathrm{T}} \tag{34}$$

Hence we compute the singular value decomposition

$$[\hat{h}_{A_i} \quad \cdots \quad \hat{h}_{A_p}] = USV^{\mathrm{T}} \tag{35}$$

to obtain the estimates

$$\hat{h} = \sigma S(1,1)U(:,1) \tag{36}$$

$$\hat{d} = \frac{1}{\sigma}V(:,1) \tag{37}$$

$$\hat{c}_i = \hat{d}_i^i \tag{38}$$

where $\sigma \neq 0$ selects the normalization constraint.

3.4. *Prediction error cost function*

Consider the prediction error cost function

$$J_{\text{pe}}(\hat{c}, \hat{h}) = \|\mathbf{z} - \hat{\mathbf{z}}\| \tag{39}$$

where $\hat{\mathbf{z}}$ is the response of the estimated system, that is

$$\hat{z}(k) = \sum_{i=0}^{p} \hat{c}_i \left(\sum_{j=0}^{m} \hat{h}_j u(k-j) \right) \tag{40}$$

Hence

$$J_{\text{pe}}^2(\hat{c}, \hat{h}) = \sum_{k=m+1}^{\ell} \left(z(k) - \sum_{|\alpha|=0}^{p} \frac{|\alpha|!}{\alpha!} v(k)^{\alpha} \hat{c}_{|\alpha|} \hat{h}^{\alpha} \right)^2 \tag{41}$$

The derivatives of $\hat{z}(k)$ are

$$\frac{\partial \hat{z}(k)}{\partial \hat{c}_i} = \sum_{|\alpha|=i} \frac{|\alpha|}{\alpha!} v(k)^{\alpha} \hat{h}^{\alpha} = i! \sum_{|\alpha|=i} \frac{\hat{h}^{\alpha}}{\alpha!} v(k)^{\alpha} \tag{42}$$

$$\frac{\partial \hat{z}(k)}{\partial \hat{h}_i} = \sum_{|\alpha|=0}^{p} \frac{|\alpha|!}{\alpha!} v(k)^{\alpha} \hat{c}_{|\alpha|} \alpha_i \hat{h}^{\alpha - e_i} \tag{43}$$

where e_i is the ith column of I_{m+1}. Hence

$$\begin{aligned} \frac{\partial J_{\text{pe}}^2}{\partial \hat{c}_i} &= 2 \sum_{k=m+1}^{\ell} (z(k) - \hat{z}(k)) \left(- \sum_{|\alpha|=i} \frac{i!}{\alpha!} v(k)^{\alpha} \hat{h}^{\alpha} \right) \\ &= -2i! \sum_{|\alpha|=i} \frac{\hat{h}^{\alpha}}{\alpha!} \sum_{k=m+1}^{\ell} v(k)^{\alpha} (z(k) - \hat{z}(k)) \end{aligned} \tag{44}$$

and

$$\begin{aligned} \frac{\partial J_{\text{pe}}^2}{\partial \hat{h}_i} &= 2 \sum_{k=m+1}^{\ell} (z(k) - \hat{z}(k)) \left(- \sum_{|\alpha|=0}^{p} \frac{|\alpha|!}{\alpha!} v(k)^{\alpha} \hat{c}_{|\alpha|} \alpha_i \hat{h}^{\alpha - e_i} \right) \\ &= -2 \sum_{|\alpha|=0}^{p} \frac{|\alpha|!}{\alpha!} \hat{c}_{|\alpha|} \alpha_i \hat{h}^{\alpha - e_i} \sum_{k=m+1}^{\ell} v(k)^{\alpha} (z(k) - \hat{z}(k)) \end{aligned} \tag{45}$$

The second derivatives are given by

$$\frac{\partial J_{\text{pe}}^2}{\partial \hat{c}_i \partial \hat{c}_j} = 2i!j! \sum_{|\alpha|=i} \frac{\hat{h}^{\alpha}}{\alpha!} \sum_{|\alpha|=j} \frac{\hat{h}^{\beta}}{\beta!} \sum_{k=m+1}^{\ell} v(k)^{\alpha} v(k)^{\beta} \tag{46}$$

$$\begin{aligned} \frac{\partial J_{\text{pe}}^2}{\partial \hat{c}_i \partial \hat{h}_j} &= -2i! \sum_{|\alpha|=i} \frac{1}{\alpha!} \sum_{k=m+1}^{\ell} v(k)^{\alpha} \\ &\quad \times \left(\alpha_j \hat{h}^{\alpha - e_j} (z(k) - \hat{z}(k)) - \hat{h}^{\alpha} \sum_{|\beta|=0}^{p} \frac{|\beta|!}{\beta!} v(k)^{\beta} \hat{c}_{|\beta|} \beta_j \hat{h}^{\beta - e_j} \right) \end{aligned} \tag{47}$$

and

$$\begin{aligned} \frac{\partial J_{\text{pe}}^2}{\partial \hat{h}_i \partial \hat{h}_j} &= -2 \sum_{|\alpha|=0}^{p} \frac{|\alpha|!}{\alpha!} \hat{c}_{|\alpha|} \alpha_i \sum_{k=m+1}^{\ell} v(k)^{\alpha} \\ &\quad \times \Bigg[(\alpha_j - \delta_{ij}) \hat{h}^{\alpha - e_i - e_j} (z(k) - \hat{z}(k)) - \hat{h}^{\alpha - e_i} \\ &\quad \times \sum_{|\beta|=0}^{p} \frac{|\beta|!}{\beta!} v(k)^{\beta} \hat{c}_{|\beta|} \beta_j \hat{h}^{\beta - e_j} \Bigg] \end{aligned} \tag{48}$$

where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases} \tag{49}$$

Using the above expressions, we implement a gradient-based optimization algorithm to minimize J_{pe} and obtain estimates $\hat{c}$ and $\hat{h}$. Assuming $c_1 \neq 0$, we normalize by letting $\hat{c}_1 = 1$, and thus remove it from the optimization problem.

4. Example

Let $m = 1$, $p = 3$, $\ell = 2^{13} = 8192$, $h = [2 \quad 1]^{\text{T}}$, $c = [-20 \quad 1 \quad 10 \quad 1]^{\text{T}}$, and $\mathcal{N}(y) = -20 + y + 10y - y^3$, which is shown in figure 2. Thus

$$y(k) = h_0 u(k) + h_1 u(k-1) \tag{50}$$

$$\begin{aligned} z(k) &= c_0 + c_1 h_1 u(k-1) + c_1 h_0 u(k) + c_2 h_1^2 u(k-1)^2 \\ &\quad + 2c_2 h_0 h_1 u(k) u(k-1) + c_2 h_0^2 u(k)^2 \\ &\quad + c_3 h_1^3 u(k-1)^3 + 3c_3 h_1 h_0^2 u(k) u(k-1)^2 \\ &\quad + 3c_3 h_0^2 h_1 u(k)^2 u(k-1) + c_3 h_0^3 u(k)^3 + w(k) \\ &= \phi(k)^{\text{T}} \boldsymbol{\theta} + w(k) \end{aligned} \tag{51}$$

where

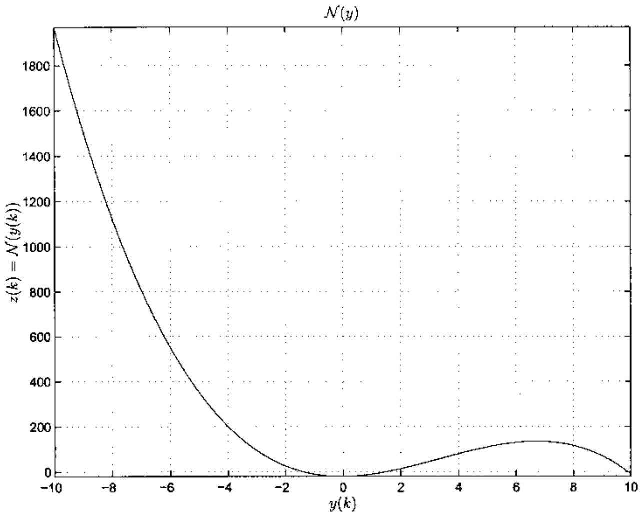


Figure 2. $\mathcal{N}(y)$.

$$\boldsymbol{\theta} = [c_{|\alpha|}h^{\alpha}]_{|\alpha|\le p} = [c_0 \quad c_1h_1 \quad c_1h_0 \quad c_2h_1^2 \quad c_2h_0h_1 \quad c_2h_0^2 \quad c_3h_1^3 \quad c_3h_0h_1^2 \quad c_3h_0^2h_1 \quad c_3h_0^3] \tag{52}$$

$$\phi(k) = \left[\frac{|\alpha|!}{\alpha!}v(k)^{\alpha}\right]_{|\alpha|\le p} = [1 \quad u(k-1) \quad u(k) \quad u(k-1)^2 \quad 2u(k)u(k-1) \quad u(k)^2 \quad u(k-1)^3 \quad 3u(k)u(k-1)^2 \quad 3u(k)^2u(k-1) \quad u(k)^3]^{\mathrm{T}} \tag{53}$$

and $w(k)$ is a realization of a zero-mean Gaussian process such that the signal to noise ratio

$$S/N = \frac{\|\mathbf{z}-\mathbf{w}\|}{\|\mathbf{w}\|} = 10 \tag{54}$$

Then (17) is replaced by

$$\mathbf{z} = \Phi^{\mathrm{T}}\boldsymbol{\theta} + \mathbf{w} \tag{55}$$

where

$$\mathbf{z} = [z(2) \quad \cdots \quad z(\ell)]^{\mathrm{T}} \tag{56}$$

$$\Phi = [\phi(2) \quad \cdots \quad \phi(\ell)] \tag{57}$$

$$\mathbf{w} = [w(m+1) \quad \cdots \quad w(\ell)]^{\mathrm{T}} = [w(2) \quad \cdots \quad w(\ell)]^{\mathrm{T}} \tag{58}$$

We assume $\Phi\Phi^{\mathrm{T}}$ is non-singular, and estimate the parameter vector using (20).

4.1. *Direct solve*

We have $m+p+2=6$ unknown parameters, one normalization constraint, and $D_p^{m+1}=10$ equations. We choose to normalize by setting $\hat{c}_1=1$, and solve $m+p+1=5$ equations. Using $\hat{\boldsymbol{\theta}}(1)$, $\hat{\boldsymbol{\theta}}(2)$, $\hat{\boldsymbol{\theta}}(3)$, $\hat{\boldsymbol{\theta}}(4)$, $\hat{\boldsymbol{\theta}}(7)$ and ignoring the rest of $\hat{\boldsymbol{\theta}}$ we obtain

$$\hat{c}_0 = \hat{\boldsymbol{\theta}}(1), \qquad \hat{c}_1 = 1, \qquad \hat{c}_2 = \frac{\hat{\boldsymbol{\theta}}(4)}{\hat{\boldsymbol{\theta}}(2)^2} \tag{59}$$

$$\hat{c}_3 = \frac{\hat{\boldsymbol{\theta}}(7)}{\hat{\boldsymbol{\theta}}(2)^3}, \qquad \hat{h}_0 = \boldsymbol{\theta}(2), \qquad \hat{h}_1 = \hat{\boldsymbol{\theta}}(3) \tag{60}$$

Figure 3(*a*) shows $\hat{h}$ for 100 simulations, each having a different realization of the input and noise sequences. Figure 3(*b*) shows the identified non-linearity.

4.2. *SVD*

Here we arrange the components of $\hat{\boldsymbol{\theta}}$ into a matrix that is also an outer product of h and ψ. Then we compute the singular value decomposition of this matrix to find $\hat{h}$ and $\hat{\psi}$. Finally, we extract $\hat{c}$ from $\hat{\psi}$. First, $\hat{c}_0$ can be estimated directly as in (21). To find the remaining parameters, we arrange the components of $\hat{\boldsymbol{\theta}}$ as

$$A(\boldsymbol{\theta}) = \psi h^{\mathrm{T}} = [c_{|\alpha|+1}h^{\alpha}]_{|\alpha|<p}h^{\mathrm{T}} = \begin{bmatrix} c_1 \\ c_2h_1 \\ c_2h_0 \\ c_3h_1^2 \\ c_3h_0h_1 \\ c_3h_0^2 \end{bmatrix} [h_0 \quad h_1] = \begin{bmatrix} \boldsymbol{\theta}(3) & \boldsymbol{\theta}(2) \\ \boldsymbol{\theta}(5) & \boldsymbol{\theta}(4) \\ \boldsymbol{\theta}(6) & \boldsymbol{\theta}(5) \\ \boldsymbol{\theta}(8) & \boldsymbol{\theta}(7) \\ \boldsymbol{\theta}(9) & \boldsymbol{\theta}(8) \\ \boldsymbol{\theta}(10) & \boldsymbol{\theta}(9) \end{bmatrix} \tag{61}$$

We calculate the singular value decomposition of $A(\hat{\boldsymbol{\theta}})$ as in (24). Then we obtain $\hat{h}$ and $\hat{\psi}$ from (25) and (26). Next, we extract the non-linearity coefficients c_i from $\hat{\psi}$. $\hat{c}_1$ is given directly by $\hat{\psi}(1)$, and we calculate the remaining c_i using least squares estimation. In this case the normalization constraint is selected by the choice of the scalar σ. We choose to normalize by setting $\sigma = U(1,1)$ such that $\hat{c}_1 = \hat{\psi}(1) = 1$. In figure 3(*c*) we plot $\hat{h}$ for 100 simulations. In figure 3(*d*) we plot the identified non-linearity.

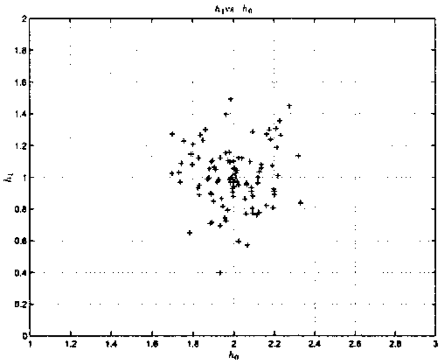
4.3. *Multi-dimensional SVD*

We arrange the elements of $\boldsymbol{\theta}$ into several matrices as in (27)–(31), corresponding to c_i and the ith power of h

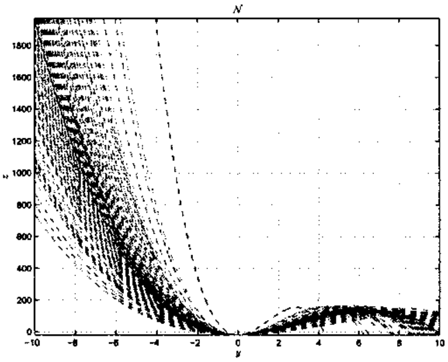
$$A_0 = c_0 = \boldsymbol{\theta}(1) \tag{62}$$

$$A_1 = c_1h = c_1\begin{bmatrix} h_0 \\ h_1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta}(3) \\ \boldsymbol{\theta}(2) \end{bmatrix} = h_{A_1} \tag{63}$$

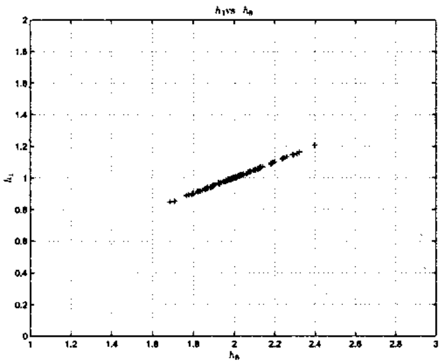
$$A_2 = c_2h \circ h = c_2\begin{bmatrix} h_0 \\ h_1 \end{bmatrix}[h_0 \quad h_1] = \begin{bmatrix} \boldsymbol{\theta}(6) & \boldsymbol{\theta}(5) \\ \boldsymbol{\theta}(5) & \boldsymbol{\theta}(4) \end{bmatrix} = h_{A_2} \circ h_{A_2} \tag{64}$$



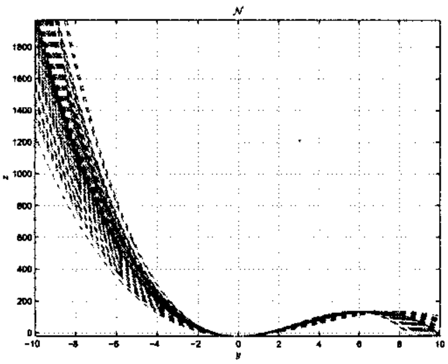
(a) Direct Solve Solutions: h_1 vs h_0



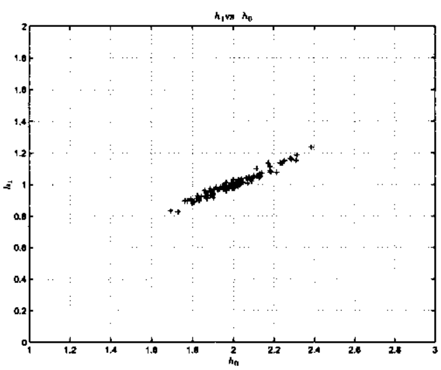
(b) Direct Solve Solutions: $\mathcal{N}$



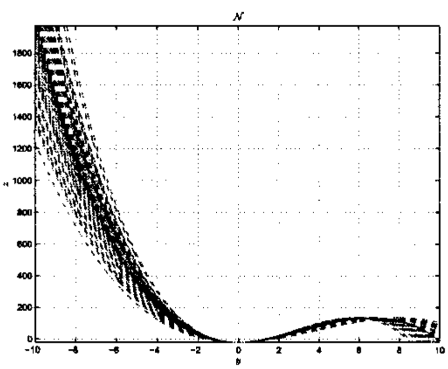
(c) SVD Solutions: h_1 vs h_0



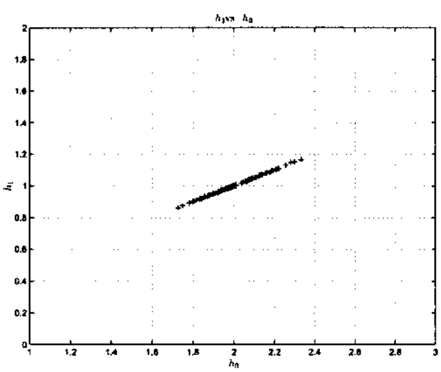
(d) SVD Solutions: $\mathcal{N}$



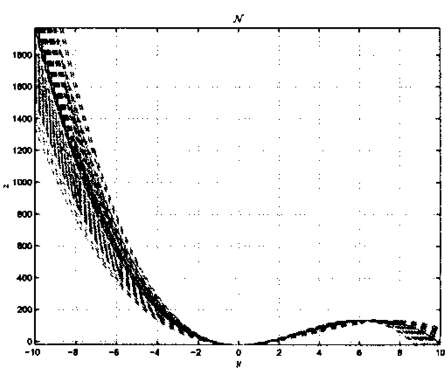
(e) Multi-Dimensional SVD Solutions: h_1 vs h_0



(f) Multi-Dimsional SVD Solutions: $\mathcal{N}$



(g) Optimization Solutions: h_1 vs h_0



(h) Optimization Solutions: $\mathcal{N}$

Figure 3. Simulation results.

$$A_3 = c_3 h \circ h \circ h = h_{A_3} \circ h_{A_3} \circ h_{A_3}$$

$$= c_3 \left[\begin{array}{cc} h_0^3 & h_0^2 h_1 \\ h_0^2 h_1 & h_0 h_1^2 \\ \hline h_0^2 h_1 & h_0 h_1^2 \\ h_0 h_1^2 & h_1^3 \end{array}\right] = \left[\begin{array}{cc} \boldsymbol{\theta}(10) & \boldsymbol{\theta}(9) \\ \boldsymbol{\theta}(9) & \boldsymbol{\theta}(8) \\ \hline \boldsymbol{\theta}(9) & \boldsymbol{\theta}(8) \\ \boldsymbol{\theta}(8) & \boldsymbol{\theta}(7) \end{array}\right] \qquad (65)$$

We use a multi-dimensional singular value decomposition (Andersson and Bro 2000) to estimate the scaled Markov vector h_{A_i}. Next, we arrange these results in a matrix and compute the singular value decomposition as in (35) to estimate $\hat{h}$, $\hat{d}$, and $\hat{c}$, see (36)–(38). In this case the normalization constraint is enforced by how we choose σ. We choose to normalize by setting $\sigma = V_{\mathcal{A}}(1,1)$ resulting in $\hat{c}_1 = 1$. In figure 3(*e*) we plot $\hat{h}$ for 100 simulations. In figure 3(*f*) we plot the identified non-linearity.

4.4. *Prediction error cost function*

We minimize the standard prediction error cost function, $J_{\rm pe}(\hat{c}, \hat{h}) = \|\mathbf{z} - \hat{\mathbf{z}}\|$, initialized with one of the three algorithms discussed earlier. Using the derivatives given previously we obtain $\hat{c}^*$ and $\hat{h}^*$. Although the results of the optimization are generally independent of the method used to obtain the initial estimate, the number of iterations needed by the optimization routine to converge usually depends on the accuracy of the initial estimate. We used `lsqnonlin` in the MATLAB optimization toolbox to minimize the function. In figure 3(*g*) we plot $\hat{h}$ for 100 simulations. In figure 3(*h*) we plot the identified non-linearity. Due to noise effects, the value of the cost function evaluated at the optimal parameters is generally less than the cost function evaluated at the true parameters $J_{\rm pe}(\hat{c}^*, \hat{h}^*) < J_{\rm pe}(c, h)$.

4.5. *Discussion*

The four methods presented, namely direct algebraic solution, singular value decomposition, multi-dimensional singular value decomposition and prediction-error minimization, produced solutions of varying accuracy. Solving some of the equations for the unknown parameters generally produced poor quality estimates, as compared to the other three. Since the direct algebraic solution method uses only a fraction of the entries of $\hat{\boldsymbol{\theta}}$ to compute the estimates, it is less robust than the other three methods. In addition, the direct algebraic solution method requires user interaction to select which equations to solve.

The singular value decomposition approch produced estimates that were close to the ones obtained by the prediction error minimization, but relied only on a singular value decomposition and some subsequent least squares steps. This method is the simplest to implement both for the user and numerically.

The multi-dimensional singular value decomposition approach produced estimates comparable to the first singular value decomposition approach, but the multi-dimensional singular value decomposition is more complex to implement than the two-dimensional singular value decomposition approach.

The prediction-error minimization approach is the most complex to implement. For best results, it should be initialized using one of the previous methods. However, it produced the best estimates of both the linear dynamics and the output non-linearity.

5. Conclusion

This paper presented four methods for identifying FIR Wiener systems with polynomial non-linearities. We presented three methods for simultaneous direct estimation of the non-linearity and linear dynamics, and a prediction error optimization method.

Many authors have studied Wiener system identification under the assumption that the non-linearity is unknown but one-to-one (Brillinger 1970, Pajunen 1985, Hasiewicz 1987, Greblicki 1992, 1994, 1997, Westwick and Kearney 1992, Wigren 1994, Westwick and Verhaegen 1996, Bai 1998, Lovera *et al*. 2000). Other methods for Wiener system identification require the non-linearity to be known, invertible, monotonic, odd, even, or require the use of specially designed input sequences. In this paper we require the non-linearity to be polynomial and the linear dynamics to have finite impulse response. These two assumptions are practical in that many non-linearities can be approximated with polynomials, and that many systems with infinite impulse response can be approximated with finite impulse response dynamics.

Future work will focus on extending the method in three directions: first, the identification of sandwich non-linear systems, i.e. systems with both input and output non-linearities; second, the identification of IIR Wiener systems; and finally, identification of Wiener systems with non-polynomial output non-linearities. While the identification of sandwich non-linear systems using this approach seems tractable, overcoming the FIR and polynomial assumptions on the linear dynamics and output non-linearity appears to be more challenging.

Acknowledgements

Supported in part by NASA under GSRP NGT4-52421 and AFOSR under grant F49620-01-1-0094 and laboratory research initiative 00VS17COR.

References

Andersson, C. A., and Bro, R., 2000, The n-way toolbox for matlab. *Chemometrics and Intelligent Laboratory Systems*, **52**, 1–4.

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

Bayard, D. S., and Eslami, M., 1984, Parameter identification of linear systems using nonlinear noninvertible measurements. In *Proceedings of the 23rd Conference on Decision and Control*, Las Vegas, NV, USA, pp. 348–352.

Brillinger, D. R., 1970, The identification of polynomial systems by means of higher order spectra. *Journal of Sound and Vibration*, **12**, 301–313.

Chen, C. H., and Fassois, S. D., 1992, Maximum likelihood identification of stochastic Wiener–Hammerstein-type nonlinear systems. *Mechanical Systems and Signal Processing*, **6**, 135–153.

Chen, C.-H., and Fassois, S. D., 1997, On the estimation of stochastic Wiener–Hammerstein-type systems with nonsmooth non-linearity. In *Proceedings of the 1997 American Control Conference*, Albuquerque, NM, USA.

Dunkl, C. F., and Xu, Y., 2001, *Orthogonal Polynomials of Several Variables*. Vol. 181 of *Encyclopedia of Mathematics and its Applications* (Cambridge, UK: Cambridge University Press).

Evans, L. C., 1998, *Partial Differential Equations*, Vol. 19 of *Graduate Studies in Mathematics*, American Mathematical Society, Providence, RI, USA.

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

Greblicki, W., 1994, Nonparametric identification of Wiener systems by orthogonal series. *IEEE Transactions on Automatic Control*, **39**, 2077–2086.

Greblicki, W., 1997, Nonparametric approach to Wiener system identification. *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications*, **44**, 538–545.

Greblicki, W., 1998, Continuous-time Wiener system identification. *IEEE Transactions on Automatic Control*, **43**, 1488–1493.

Haber, R., and Keviczky, L., 1999a, *Nonlinear System Identification—Input–Output Modeling Approach*. Vol 1: Nonlinear System Parameter Identification (Dordrecht: Kluwer Academic Publishers).

Haber, R., and Keviczky, L., 1999b, *Nonlinear System Identification—Input–Output Modeling Approach*. Vol. 2: Nonlinear System Structure Identification (Dordrecht: Kluwer Academic Publishers).

Hasiewicz, Z., 1987, Identification of a linear system observed through zero-memory non-linearity. *International Journal of Systems Science*, **18**, 1595–1607.

Hunter, I. W., and Kornberg, M. J., 1986, The identification of nonlinear and biological systems: Wiener and Hammerstein cascade models. *Biological Cybernetics*, **55**, 135–144.

Korenberg, M. J., and Hunter, I. W., 1986, The identification of nonlinear biological systems: LNL cascade models. *Biological Cybernetics*, **55**, 125–134.

Lacy, S. L., and Bernstein, D. S., 2001, Subspace identification for nonlinear systems that are linear in unmeasured states. In *Proceedings of the Conference on Decision and Control*, Orlando, FL, USA.

Lacy, S. L., Erwin, R. S., and Bernstein, D. S., 2001, Identification of Wiener systems with known noninvertible nonlinearities. *Journal of Dynamic Systems, Measurement, and Control*, **123**, 566–571.

Lovera, M., Gustafsson, T., and Verhaegen, M., 2000, Recursive subspace identification of linear and non-linear Wiener state-space models. *Automatica*, **36**, 1639–1650.

Nelles, O., 2001, *Nonlinear System Identification* (New York: Springer).

Pajunen, G. A., 1985, Recursive identification of Wiener type nonlinear systems. In *Proceedings of the 1985 American Control Conference*, Vol. 3, Boston, MA, USA, pp. 1365–1370.

Pintelon, R., and Schoukens, J., 2001, *System Identification: A Frequency Domain Approach* (Piscataway, NJ: IEEE Press).

Van Pelt, T., and Bernstein, D. S., 2000, Nonlinear system identification using hammerstein and nonlinear feedback models with piecewise linear static maps—part I: Theory—part II: Numerical examples. In *Proceedings of the ACC*, Chicago, IL, USA, pp. 225–229, 235–239.

Westwick, D. T., and Kearney, R. E., 1992, A new algorithm for the identification of multiple input Wiener systems. *Biological Cybernetics*, **68**, 75–85.

Westwick, D., and Verhaegen, M., 1996, Identifying MIMO Wiener systems using subspace model identification methods. *Signal Processing*, **52**, 235–258.

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