INT. J. CONTROL, 2001, VOL. 74, NO. 18, 1807–1823

Non-linear system identification using Hammerstein and non-linear feedback models with piecewise linear static maps

TOBIN H. VAN PELT†* and DENNIS S. BERNSTEIN‡

We consider the identification of Hammerstein/non-linear feedback models by approximating internal non-linearities using piecewise linear static maps. The resulting method utilizes a point-slope parameterization that leads to a computationally tractable optimization problem. The computational appeal of this technique is derived from the fact that the method only requires a matrix inverse and singular value decomposition. Furthermore, the identification method simultaneously identifies the linear dynamic and static non-linear blocks without requiring prior assumptions on the form of the static non-linearity.

1. Introduction

Empirical or data-based modelling, generally referred to as system identification, plays an essential role in control systems engineering as well as many other branches of science and engineering. System identification using linear model structures has been extensively developed and the theory is mature. Issues such as model order selection, consistency and optimal input selection have been discussed at length in several landmark texts (Ljung and Soderstrom 1983, Soderstrom and Stoica 1989, Ljung 1999).

In practice, however, all real systems possess some non-linearity, and this non-linearity can degrade the effectiveness of linear system identification methods. Accordingly, there has been significant effort during the past several decades to develop techniques for non-linear system identification (Billings 1980, Bendat 1989, Juditsky *et al.* 1995, Sjober *et al.* 1995, Wemhoff *et al.* 1999, Fitzgerald *et al.* 2000, Young 2000). While a large variety of techniques have been proposed for this problem, most research has focused on either black-box modelling (i.e. a purely input–output framework), or alternatively, physical modelling, which is based on scientific principles that retain parametric interpretation and correspondence with the physical world.

From an input-output point of view, the Volterra series representation provides a general framework for representing a large class of non-linear systems. A highly readable account of this modelling formalism is given in Rugh (1981), whose last chapter provides a summary of non-linear system identification methods based on this model representation.

Received for publication 16 August 2001.

*Author for correspondence. e-mail: vanpelt@huntersystech.com

†Hunter System Technologies, LLC, PO Box 975, Sausalito, CA 94966, USA.

‡Department of Aerospace Engineering, The University of Michigan, Ann Arbor, MI 48109, USA.

An alternative model representation for non-linear system identification involves realizations consisting of dynamic linear blocks ($\mathcal{L}$) and static non-linear blocks ($\mathcal{N}$). Interconnection of these blocks in various configurations allows identification under different model structures. The most frequently used structures are the Hammerstein model ($\mathcal{N} \to \mathcal{L}$) (Narendra and Gallman 1966, Stoica 1981, Hunter and Korenberg 1986, Greblicki 1989, Krzyzak 1989, Eskinat *et al.* 1991, Pawlak 1991, Chen and Fassois 1992, Rangan *et al.* 1995), the Wiener model ($\mathcal{L} \to \mathcal{N}$) (Billings and Fakhovri 1977, Pajunen 1985, Hunter and Korenberg 1986, Billings and Voon 1987, Chen and Fassois 1992, Greblicki 1999), the sandwich model ($\mathcal{L}_1 \to \mathcal{N} \to \mathcal{L}_2$ or $\mathcal{N}_1 \to \mathcal{L} \to \mathcal{N}_2$) (Sandor and Williamson 1978, Billings and Fakhouri 1982, Boyd and Chua 1983, Korenberg and Hunter 1986, Shi and Sun 1990, Bai 1998, Vandersteen and Schoukens 1998), and the non-linear feedback model ($\mathcal{L} \to \mathcal{N} \to \mathcal{L}$) (Wemhoff *et al.* 1999), as well as various combinations of these basic configurations. Some of these techniques assume that one or more of the subsystems is known, while others develop models for all subsystems, either iteratively or simultaneously.

A link between the input–output representation and the realization model structures was established in Boyd and Chua (1985), where it was shown that a fading memory non-linear operator can be approximated by a Volterra series operator which in turn can be realized by a Wiener model structure with polynomial non-linearity.

Another class of methods for non-linear system identification is based on neural network functional approximation techniques (Chen *et al.* 1990, Narendra and Parthasarathy 1990, Lu and Basar 1998). These techniques are especially suitable for multivariable non-linearities through the use of radial basis function approximations. Yet another approach involves non-stationary stochastic time series methods (Young 2000, Young *et al.* 2001). These techniques assume that parameters within a linear model structure vary in time or

International Journal of Control ISSN 0020–7179 print/ISSN 1366–5820 online
http://www.tandf.co.uk/journals
DOI: 10.1080/00207170110089798

depend on other system measurements. This non-parametric technique yields information regarding the form of the non-linearities within the system, such that identification can be implemented using a parsimonious parameterization.

In the present paper we develop a novel technique for non-linear system identification using block-structured models consisting of Hammerstein, non-linear feedback, and Hammerstein/non-linear feedback model structures. One of the key components of our approach is the use of piecewise linear approximations for the static non-linear blocks. Piecewise linear approximations have been widely studied in the circuit theory literature (Kahlert and Chua 1990, 1992) as well as in non-linear control theory (Sontag 1991, Petit 1985, Ahmed 1995, Hwang 1995). In addition, these non-linearities have been used for non-linear system identification in Pajunen (1985), Billings and Voon (1987) and Shi and Sun (1990). In the present paper, we use a 'point-slope' representation of the piecewise linear non-linearity. We note that non-linear feedback models have been extensively studied in classical absolute stability theory. See, for example, Narendra and Taylor (1973) and Haddad and Bernstein (1993, 1994).

It was shown in Bai (1998) that for a specific type of sandwich model ($\mathcal{N}_1 \to \mathcal{L} \to \mathcal{N}_2$), the non-linear least squares optimization problem can be replaced by a pair of standard computational procedures, namely, linear least squares optimization and fixed rank approximation in the Frobenius norm. This procedure provides simultaneous approximations of the linear and non-linear blocks of the system. The attractiveness of this procedure lies in the relative computational simplicity of the technique and its suitability for online implementation.

In this paper the work in Bai (1998) is extended in the following ways. First, the computational procedure is motivated by means of a bounding technique that entails a suboptimal, but computationally tractable, approximation. Next, the use of piecewise linear approximations is considered. A potential problem associated with this parameterization arises for non-linear feedback models. When parameterizing the piecewise linear feedback non-linearity it is difficult to guarantee the invertibility of the regression matrix that arises in the linear least squares optimization problem. This difficulty is overcome by implementing a 'point-slope' representation for this non-linearity. Furthermore, the work of Bai (1998) is extended by considering non-linear feedback and Hammerstein/non-linear feedback model structures.

The contents of the paper are as follows. In § 2 we present the Hammerstein/non-linear feedback model. In § 3 we present the 'point-slope' parameterization for piecewise linear functions. In § 4 we develop the piecewise linear least squares (PLLS) identification technique using a Hammerstein model, a non-linear feedback model and a Hammerstein/non-linear feedback model. Finally, in § 5 we present two numerical examples to illustrate this technique. Although the PLLS procedure does not account for the effects of noise *a priori*, we consider numerical examples involving noisy data to illustrate its performance.

1.1. *Notation*

We define the following notation. I_n is the $n \times n$ identity matrix, $0_{n\times m}$ is the $n \times m$ zero matrix, 1_n is the ones column vector of length n, $\|M\|_{\mathrm{F}}^2 \triangleq tr(M^{\mathrm{T}}M)$ is the Frobenius norm of the matrix M, $\sigma_{\max}(M)$ is the largest singular value of M, $\mathrm{vec}\,(\cdot)$ and $\mathrm{vec}^{-1}\,(\cdot)$ are the column stacking operation and its inverse and $\mathbf{q}^{-1}$ is the backward shift operator.

2. Hammerstein/non-linear feedback modelling

Consider the single-input single-output *Hammerstein/non-linear feedback model* shown in Figure 1, where $u(k)$ is the *model input* and $y(k)$ is the *model output*. This model consists of a linear time-invariant (LTI) block and two static non-linearities $f_0 : \mathbb{R} \to \mathbb{R}$ and $h_0 : \mathbb{R} \to \mathbb{R}$. One or both non-linearities can be absent by setting $f_0(u) = u$ or $h_0(y) = 0$. When $h_0(y) = 0$ the model becomes a *Hammerstein* model, and when $f_0(u) = u$ the model becomes a *non-linear feedback model*. The LTI block is represented by the nth-order strictly proper transfer function

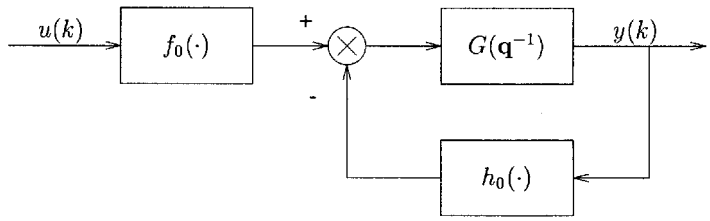

Figure 1. Hammerstein/non-linear feedback model.

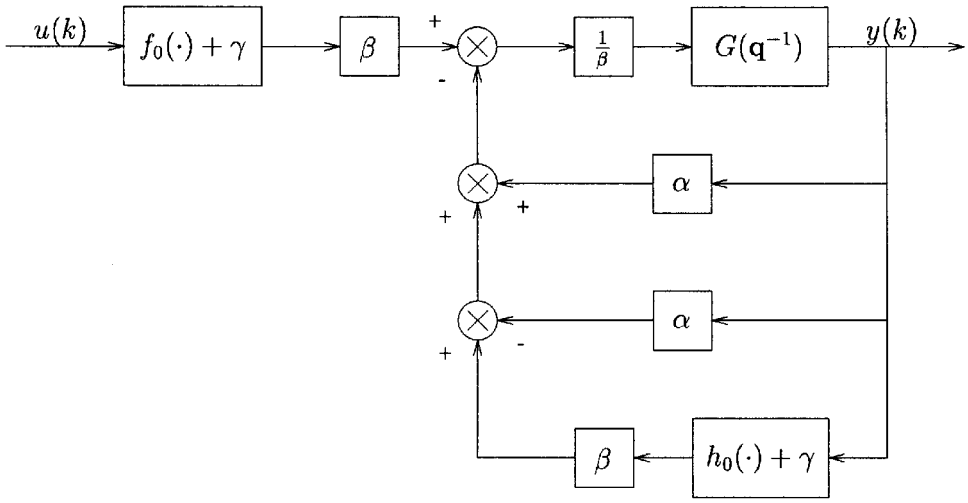

Figure 2. Representation of Hammerstein/non-linear feedback model.

$$G(\mathbf{q}^{-1}) = \frac{B(\mathbf{q}^{-1})}{A(\mathbf{q}^{-1})} = \frac{b_1\mathbf{q}^{-1} + \cdots + b_n\mathbf{q}^{-n}}{1 + a_1\mathbf{q}^{-1} + \cdots + a_n\mathbf{q}^{-n}} \quad (1)$$

The model output $y(k)$ is then given by

$$y(k) = \frac{B(\mathbf{q}^{-1})}{A(\mathbf{q}^{-1})}[f_0(u(k)) - h_0(y(k))] \quad (2)$$

The representation of G, f_0 and h_0 in the Hammerstein/non-linear feedback model is non-unique. Specifically, letting $\alpha, \beta, \gamma \in \mathbb{R}$, where $\beta \neq 0$, it can be seen that Figure 1 is equivalent to Figure 2 in the sense that both systems have the same input–output response. Hence, $u(k)$ and $y(k)$ satisfy

$$y(k) = \frac{B(\mathbf{q}^{-1})}{\beta A(\mathbf{q}^{-1}) + \alpha B(\mathbf{q}^{-1})} \times [\beta(f_0(u(k)) + \gamma) - (\beta(h_0(y(k)) + \gamma) - \alpha y(k))] \quad (3)$$

which is equivalent to (2). Thus, $f_0(u)$, $h_0(y)$ and $G(\mathbf{q}^{-1})$ can be replaced by $\beta(f_0(u) + \gamma)$, $\beta(h_0(y) + \gamma) - \alpha y$ and $G(\mathbf{q}^{-1})/(\beta + \alpha G(\mathbf{q}^{-1}))$, respectively.

The *input gain parameter* β scales the gain of the LTI block $G(\mathbf{q}^{-1})$. Furthermore, for a given β, the *stability parameter* α determines the stability of the LTI block. Specifically, in some cases α can be chosen to ensure stability of the LTI block by examining the root locus of the denominator polynomial

$$A(\mathbf{q}^{-1}) + \frac{\alpha}{\beta}B(\mathbf{q}^{-1}) \quad (4)$$

The *offset parameter* γ affects the offset of the non-linearities $f_0(u)$ and $h_0(y)$.

3. Parameterization of piecewise linear functions

Henceforth we approximate f_0 and h_0 by continuous piecewise linear functions f and h. To represent these functions we use the parameterization illustrated in Figure 3. This parameterization is characterized by the function value $\kappa = f(c_r)$ and the slope parameters $\mu_1, \ldots, \mu_{p+1}$ defined over a partitioning $(-\infty, c_1]$, $[c_1, c_2], \ldots, [c_p, \infty)$ of the domain of f. Let $c_1 < c_2 < \cdots < c_p$ be real numbers, let $c = [c_1 \ \cdots \ c_p]^{\mathrm{T}}$ be the *partition* of the domain of f, let $\mu_i \in \mathbb{R}$, $i = 1, \ldots, p+1$, $\kappa \in \mathbb{R}$, and let $r \in \{1, \ldots, p\}$ be the *primary index* at which the value $\kappa = f(c_r)$ is specified. Then f is represented by

$$f(u) = \begin{cases} \mu_{\delta(u)}(u - c_{\delta(u)}) - \displaystyle\sum_{j=\delta(u)+1}^{r} \mu_j(c_j - c_{j-1}) + \kappa, & \\ & \delta(u) < r \\ \mu_{\delta(u)}(u - c_r) + \kappa, & r \leq \delta(u) \leq r+1 \\ \mu_{\delta(u)}(u - c_{\delta(u)-1}) + \displaystyle\sum_{j=r+1}^{\delta(u)-1} \mu_j(c_j - c_{j-1}) + \kappa, & \\ & \delta(u) > r+1 \end{cases} \quad (5)$$

where

$$\delta(u) \triangleq \begin{cases} 1, & u \leq c_1 \\ i, & c_{i-1} < u \leq c_i, \quad i = 2, \ldots, p \\ p+1, & c_p < u \end{cases} \quad (6)$$

By defining the slope vector

$$\mu \triangleq [\mu_1 \ \cdots \ \mu_{p+1}]^{\mathrm{T}} \in \mathbb{R}^{p+1} \quad (7)$$

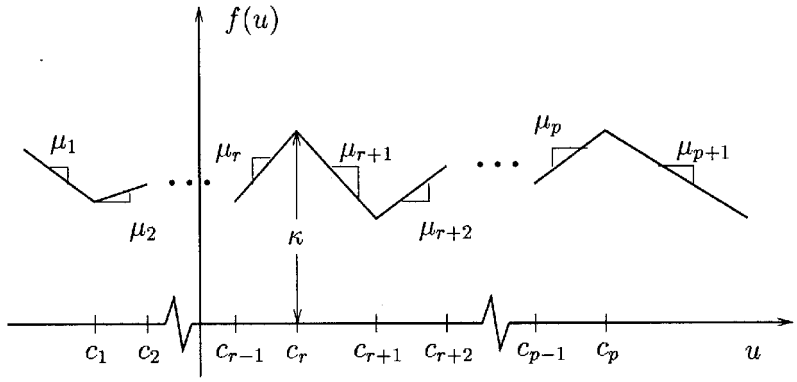

Figure 3. Parameterization of the piecewise linear function f.

$f(u)$ can be written as

$$f(u) = \mu^{\mathrm{T}}\eta(u) + \kappa \tag{8}$$

where

$$\eta(u) \triangleq \begin{cases} \eta_1(u), & \delta(u) < r+1 \\ \eta_2(u), & \delta(u) \geq r+1 \end{cases} \tag{9}$$

and where

$$\eta_1(u) \triangleq [0_{1\times(\delta(u)-1)} \quad u - c_{\delta(u)} \quad c_{\delta(u)} - c_{\delta(u)+1} \quad \cdots \quad c_{r-1} - c_r \quad 0_{1\times(p+1-r)}]^{\mathrm{T}} \in \mathbb{R}^{p+1} \tag{10}$$

and

$$\eta_2(u) \triangleq [0_{1\times r} \quad c_{r+1} - c_r \quad \cdots \quad c_{\delta(u)-1} - c_{\delta(u)-2} \quad u - c_{\delta(u)-1} \quad 0_{1\times(p-r+1)}]^{\mathrm{T}} \in \mathbb{R}^{p+1} \tag{11}$$

By a slight modification of these definitions we can restrict the domain of f to be compact. The specific choice of index parameter r and partitions $c_1, c_2, \ldots, c_p$ will typically depend on the application considered. Further discussion of this point appears in § 5.

4. Piecewise linear least squares identification

This section presents the piecewise linear least squares (PLLS) identification method. For clarity, the case of the Hammerstein model is considered first and then the method is extended to non-linear feedback and Hammerstein/non-linear feedback.

4.1. *Hammerstein model identification*

Consider the Hammerstein model with $h_0(y) = 0$ in (2). The output $y(k)$, using the piecewise linear approximation f of f_0, is then given by

$$y(k) = G(\mathbf{q}^{-1})f(u(k)) \tag{12}$$

which has the time series representation

$$y(k) = (1 - A(\mathbf{q}^{-1}))y(k) + B(\mathbf{q}^{-1})f(u(k)) \tag{13}$$

Next define

$$\begin{aligned} a &\triangleq [a_1 \quad \cdots \quad a_n]^{\mathrm{T}} \in \mathbb{R}^n \\ b &\triangleq [b_1 \quad \cdots \quad b_n]^{\mathrm{T}} \in \mathbb{R}^n \end{aligned} \tag{14}$$

and expand (13) as

$$\begin{aligned} y(k) = &-a_1 y(k-1) - \cdots - a_n y(k-n) \\ &+ b_1 f(u(k-1)) + \cdots + b_n f(u(k-n)) \end{aligned} \tag{15}$$

Substituting (8) into (15) yields

$$y(k) = \theta^{\mathrm{T}}\phi(k) \tag{16}$$

where

$$\theta \triangleq \begin{bmatrix} a \\ \mathrm{vec}\,(\mu b^{\mathrm{T}}) \\ \kappa 1_n^{\mathrm{T}} b \end{bmatrix} \in \mathbb{R}^{n+n(p+1)+1} \tag{17}$$

and

$$\phi(k) \triangleq [-\phi_y^{\mathrm{T}}(k) \quad \phi_\eta^{\mathrm{T}}(k) \quad 1]^{\mathrm{T}} \in \mathbb{R}^{n+n(p+1)+1} \tag{18}$$

where

$$\phi_y(k) \triangleq [y(k-1) \quad \cdots \quad y(k-n)]^{\mathrm{T}} \in \mathbb{R}^n \tag{19}$$

and

$$\phi_\eta(k) \triangleq [\eta^{\mathrm{T}}(u(k-1)) \quad \cdots \quad \eta^{\mathrm{T}}(u(k-n))]^{\mathrm{T}} \in \mathbb{R}^{n(p+1)} \tag{20}$$

Next consider input–output measurements $u(k)$ and $y(k)$ for $k = 0, \ldots, l$, where $l \geq n$, and form the least squares cost

$$J(\theta) = \|Y - \Phi\theta\|_2 \tag{21}$$

where

$$Y \triangleq [y(n) \quad \cdots \quad y(l)]^{\mathrm{T}} \in \mathbb{R}^{l-n+1} \tag{22}$$

and

$$\Phi \triangleq \begin{bmatrix} \phi^{\mathrm{T}}(n) \\ \vdots \\ \phi^{\mathrm{T}}(l) \end{bmatrix} = [\, -\Phi_y \quad \Phi_\eta \quad 1_{l-n+1} \,] \in \mathbb{R}^{(l-n+1)\times(n+n(p+1)+1)} \tag{23}$$

where

$$\Phi_y \triangleq \begin{bmatrix} \phi_y^{\mathrm{T}}(n) \\ \vdots \\ \phi_y^{\mathrm{T}}(l) \end{bmatrix} \in \mathbb{R}^{(l-n+1)\times n}$$

$$\Phi_\eta \triangleq \begin{bmatrix} \phi_\eta^{\mathrm{T}}(n) \\ \vdots \\ \phi_\eta^{\mathrm{T}}(l) \end{bmatrix} \in \mathbb{R}^{(l-n+1)\times n(p+1)} \tag{24}$$

We note that the model output (16) is not linear in b, μ and κ. Although non-linear least squares techniques can be used to minimize $J(\theta)$, we proceed by bounding $J(\theta)$. First, let $\theta_{\mathcal{B}} \in \mathbb{R}^{n(p+1)}$ and rewrite (21) as

$$\begin{aligned} J(\theta) &= \|Y - \Phi\theta + \Phi_\eta\theta_{\mathcal{B}} - \Phi_\eta\theta_{\mathcal{B}}\|_2 \\ &= \|Y + \Phi_y a - \Phi_\eta\theta_{\mathcal{B}} - \kappa 1_{l-n+1}1_n^{\mathrm{T}} b \\ &\quad + \Phi_\eta\theta_{\mathcal{B}} - \Phi_\eta \mathrm{vec}\,(\mu b^{\mathrm{T}})\|_2 \\ &= \|Y - \Phi\tilde{\theta} + \Phi_\eta(\theta_{\mathcal{B}} - \mathrm{vec}\,(\mu b^{\mathrm{T}}))\|_2 \end{aligned} \tag{25}$$

where

$$\tilde{\theta} \triangleq \begin{bmatrix} a \\ \theta_{\mathcal{B}} \\ \theta_\kappa \end{bmatrix} \in \mathbb{R}^{n+(p+1)n+1} \tag{26}$$

and

$$\theta_\kappa \triangleq \kappa 1_n^{\mathrm{T}} b \tag{27}$$

By invoking the triangle inequality, we obtain

$$\begin{aligned} J(\theta) &\le \|Y - \Phi\tilde{\theta}\|_2 + \|\Phi_\eta(\theta_{\mathcal{B}} - \mathrm{vec}\,(\mu b^{\mathrm{T}}))\|_2 \\ &\le \|Y - \Phi\tilde{\theta}\|_2 + \sigma_{\max}(\Phi_\eta)\|\theta_{\mathcal{B}} - \mathrm{vec}\,(\mu b^{\mathrm{T}})\|_2 \\ &= \|Y - \Phi\tilde{\theta}\|_2 + \sigma_{\max}(\Phi_\eta)\|\mathrm{vec}^{-1}(\theta_{\mathcal{B}}) - \mu b^{\mathrm{T}}\|_{\mathrm{F}} \end{aligned} \tag{28}$$

Therefore

$$J(\theta) \le J_{\mathrm{LS}}(\tilde{\theta}) + \sigma_{\max}(\Phi_\eta) J_{\mathcal{B}}(\mu, b, \theta_{\mathcal{B}}) \tag{29}$$

where

$$J_{\mathrm{LS}}(\tilde{\theta}) \triangleq \|Y - \Phi\tilde{\theta}\|_2 \tag{30}$$

and

$$J_{\mathcal{B}}(\mu, b, \theta_{\mathcal{B}}) \triangleq \|\mathrm{vec}^{-1}(\theta_{\mathcal{B}}) - \mu b^{\mathrm{T}}\|_{\mathrm{F}} \tag{31}$$

where $\mathrm{vec}^{-1}(\theta_{\mathcal{B}}) \in \mathbb{R}^{(p+1)\times n}$.

We proceed by sequentially minimizing $J_{\mathrm{LS}}(\tilde{\theta})$ and $J_{\mathcal{B}}(\mu, b, \theta_{\mathcal{B}})$. To do this, we first determine $\hat{\tilde{\theta}}$ that minimizes the linear least squares cost $J_{\mathrm{LS}}(\tilde{\theta})$. Writing

$$\hat{\tilde{\theta}} = \begin{bmatrix} \hat{a} \\ \hat{\theta}_{\mathcal{B}} \\ \hat{\theta}_\kappa \end{bmatrix} \tag{32}$$

we then extract $\hat{\theta}_{\mathcal{B}}$ from $\hat{\tilde{\theta}}$ and minimize $J_{\mathcal{B}}(\mu, b, \hat{\theta}_{\mathcal{B}}) = \|\mathrm{vec}^{-1}(\hat{\theta}_{\mathcal{B}}) - \mu b^{\mathrm{T}}\|_{\mathrm{F}}$ with respect to μ and b to obtain an optimal rank 1 approximation of $\mathrm{vec}^{-1}(\hat{\theta}_{\mathcal{B}})$. We note that both of these optimization problems are standard, and they are solvable by computationally tractable procedures.

Assuming that $(\Phi^{\mathrm{T}}\Phi)^{-1}$ exists, $J_{\mathrm{LS}}(\tilde{\theta})$ is minimized by

$$\hat{\tilde{\theta}} = (\Phi^{\mathrm{T}}\Phi)^{-1}\Phi^{\mathrm{T}} Y \tag{33}$$

and thus the minimum value of $J_{\mathrm{LS}}(\tilde{\theta})$ is given by

$$J_{\mathrm{LS}}(\hat{\tilde{\theta}}) = Y^{\mathrm{T}}(I_{l-n+1} - \Phi(\Phi^{\mathrm{T}}\Phi)^{-1}\Phi^{\mathrm{T}}) Y \tag{34}$$

In practice $\hat{\tilde{\theta}}$ is computed using a QR factorization (Ljung 1999, p. 318) to ensure numerical robustness.

Next, by extracting $\hat{\theta}_{\mathcal{B}}$ from (32), $J_{\mathcal{B}}(\mu, b, \hat{\theta}_{\mathcal{B}})$ is minimized by (Stewart and Sun 1990, p. 208)

$$\hat{\mu}\hat{b}^{\mathrm{T}} = \sigma_{\max}(\mathrm{vec}^{-1}(\hat{\theta}_{\mathcal{B}}))u_1 v_1^{\mathrm{T}} \tag{35}$$

where σ_i, $i = 1, \ldots, \min\{(p+1), n\}$ are the singular values of $\mathrm{vec}^{-1}(\hat{\theta}_{\mathcal{B}})$, u_1 is the first left singular vector of $\mathrm{vec}^{-1}(\hat{\theta}_{\mathcal{B}})$, and v_1 is the first right singular vector of $\mathrm{vec}^{-1}(\hat{\theta}_{\mathcal{B}})$. The minimum value of $J_{\mathcal{B}}(\mu, b, \hat{\theta}_{\mathcal{B}})$ is thus given by

$$J_{\mathcal{B}}(\hat{\mu}, \hat{b}, \hat{\theta}_{\mathcal{B}}) = \sqrt{\sum_{i=2}^{\min\{(p+1),n\}} \sigma_i^2} \tag{36}$$

It follows from (35) that $\hat{b}$ and $\hat{\mu}$ are given by

$$\hat{\mu} = \beta\sigma_{\max}(\mathrm{vec}^{-1}(\hat{\theta}_{\mathcal{B}}))u_1, \qquad \hat{b} = \frac{1}{\beta} v_1 \tag{37}$$

for an arbitrary input gain parameter β. Finally, $\hat{a}$ is given by direct extraction from (32), and $\hat{\kappa}$ is given by

$$\hat{\kappa} = \frac{\hat{\theta}_\kappa}{1_n^{\mathrm{T}}\hat{b}} \tag{38}$$

PLLS identification with a Hammerstein model is summarized as follows:

(1) Collect input–output measurements $u(k)$ and $y(k)$, $k = 0, \ldots, l$, where $l \ge n$.
(2) Form the regression matrix Φ in (23) and the output vector Y in (22).
(3) Obtain $\hat{\tilde{\theta}}$ in (32) by solving the linear least squares problem given by (33).
(4) Extract $\hat{\theta}_{\mathcal{B}}$ from $\hat{\tilde{\theta}}$.

(5) Compute the singular value decomposition of $\mathrm{vec}^{-1}(\hat{\theta}_{\mathcal{B}})$ and obtain u_1, v_1 and $\sigma_{\max}(\mathrm{vec}^{-1}(\hat{\theta}_{\mathcal{B}}))$.
(6) Set the input gain parameter β.
(7) Compute the parameter estimates $\hat{b}$ and $\hat{\mu}$ using (37).
(8) Extract the parameter estimate $\hat{a}$ from $\hat{\tilde{\theta}}$.
(9) Compute the parameter estimate $\hat{\kappa}$ using (38).

4.2. *Non-linear feedback model identification*

Consider the non-linear feedback model where $f_0(u) = u$ in (2). The model output $y(k)$, using the piecewise linear approximation h of h_0, is then given by

$$y(k) = G(\mathbf{q}^{-1})[u(k) - h(y(k))] \tag{39}$$

which has the time series representation

$$\begin{aligned} y(k) = {} & b_1 u(k-1) + \cdots + b_n u(k-n) \\ & - b_1 h(y(k-1)) - \cdots - b_n h(y(k-n)) \\ & - a_1 y(k-1) - \cdots - a_n y(k-n) \end{aligned} \tag{40}$$

Representing h by replacing u, c, p, r, δ, μ, κ, η with y, d, q, s, ι, ν, λ, ζ, respectively, and noting that

$$1_{q+1}^{\mathrm{T}} \zeta(y) = y - d_s \tag{41}$$

(40) can further be written as

$$\begin{aligned} y(k) = \sum_{j=1}^{n} & (b_j u(k-j) - (b_j \nu^{\mathrm{T}} + a_j 1_{q+1}^{\mathrm{T}}) \\ & \times \zeta(y(k-j)) - a_j d_s - \lambda b_j) \end{aligned} \tag{42}$$

Furthermore, (42) can be expressed as in (16), where

$$\theta \triangleq \begin{bmatrix} \mathrm{vec}\,(\nu b^{\mathrm{T}} + 1_{q+1} a^{\mathrm{T}}) \\ b \\ 1_n^{\mathrm{T}}(\lambda b + d_s a) \end{bmatrix} \in \mathbb{R}^{n(q+1)+n+1} \tag{43}$$

and

$$\phi(k) \triangleq [\, -\phi_{\zeta}^{\mathrm{T}}(k) \quad \phi_u^{\mathrm{T}}(k) \quad -1 \,]^{\mathrm{T}} \in \mathbb{R}^{n(q+1)+n+1} \tag{44}$$

where

$$\phi_{\zeta}(k) \triangleq [\, \zeta^{\mathrm{T}}(y(k-1)) \quad \cdots \quad \zeta^{\mathrm{T}}(y(k-n)) \,]^{\mathrm{T}} \in \mathbb{R}^{n(q+1)} \tag{45}$$

and

$$\phi_u(k) \triangleq [\, u(k-1) \quad \cdots \quad u(k-n) \,]^{\mathrm{T}} \in \mathbb{R}^{n} \tag{46}$$

Next consider input–output measurements $u(k)$ and $y(k)$ for $k = 0, \ldots, l$, where $l \geq n$, and form the least squares cost in (21), where Y is define as in (22) and

$$\begin{aligned} \Phi & \triangleq \begin{bmatrix} \phi(n)^{\mathrm{T}} \\ \vdots \\ \phi(l)^{\mathrm{T}} \end{bmatrix} \\ & = \begin{bmatrix} -\Phi_{\zeta} & \Phi_u & -1_{l-n+1} \end{bmatrix} \in \mathbb{R}^{(l-n+1)\times n(q+1)+n+1} \end{aligned} \tag{47}$$

where

$$\begin{aligned} \Phi_{\zeta} & = \begin{bmatrix} \phi_{\zeta}^{\mathrm{T}}(n) \\ \vdots \\ \phi_{\zeta}^{\mathrm{T}}(l) \end{bmatrix} \in \mathbb{R}^{(l-n+1)\times n(q+1)} \\ \Phi_u & = \begin{bmatrix} \phi_u^{\mathrm{T}}(n) \\ \vdots \\ \phi_u^{\mathrm{T}}(l) \end{bmatrix} \in \mathbb{R}^{(l-n+1)\times n} \end{aligned} \tag{48}$$

In a similar manner as was done in § 4.1, we proceed by bounding $J(\theta)$. First, let $\theta_{\mathcal{A}} \in \mathbb{R}^{n(q+1)}$ and rewrite (21) as

$$J(\theta) = \| Y - \Phi\tilde{\theta} - \Phi_{\zeta}(\theta_{\mathcal{A}} - \mathrm{vec}\,(\nu b^{\mathrm{T}} + 1_{q+1} a^{\mathrm{T}})) \|_2$$

where

$$\tilde{\theta} \triangleq \begin{bmatrix} \theta_{\mathcal{A}} \\ b \\ \theta_{\lambda} \end{bmatrix} \in \mathbb{R}^{n(q+1)+n+1} \tag{49}$$

and

$$\theta_{\lambda} \triangleq 1_n^{\mathrm{T}}(\lambda b + d_s a) \tag{50}$$

As before, by invoking the triangle inequality we obtain

$$J(\theta) \leq J_{\mathrm{LS}}(\tilde{\theta}) + \sigma_{\max}(\Phi_{\zeta}) J_{\mathcal{A}}(\nu, a, b, \theta_{\mathcal{A}}) \tag{51}$$

where $J_{\mathrm{LS}}(\tilde{\theta})$ is defined in (30) and

$$J_{\mathcal{A}}(\nu, a, b, \theta_{\mathcal{A}}) \triangleq \| \mathrm{vec}^{-1}(\theta_{\mathcal{A}}) - (\nu b^{\mathrm{T}} + 1_{q+1} a^{\mathrm{T}}) \|_{\mathrm{F}} \tag{52}$$

where $\mathrm{vec}^{-1}(\theta_{\mathcal{A}}) \in \mathbb{R}^{(q+1)\times n}$.

Again we proceed by sequentially minimizing $J_{\mathrm{LS}}(\tilde{\theta})$ and $J_{\mathcal{A}}(\nu, a, b, \theta_{\mathcal{A}})$. To do this, we write

$$\hat{\tilde{\theta}} = \begin{bmatrix} \hat{\theta}_{\mathcal{A}} \\ \hat{b} \\ \hat{\theta}_{\kappa} \end{bmatrix} \in \mathbb{R}^{n(q+1)+n+1} \tag{53}$$

and determine $\hat{\tilde{\theta}}$ that minimizes the least squares cost $J_{\mathrm{LS}}(\tilde{\theta})$. Then we extract $\hat{\theta}_{\mathcal{A}}$ and $\hat{b}$ from $\hat{\tilde{\theta}}$ and minimize $J_{\mathcal{A}}(\nu, a, \hat{b}, \hat{\theta}_{\mathcal{A}}) = \| \mathrm{vec}^{-1}(\hat{\theta}_{\mathcal{A}}) - (\nu \hat{b}^{\mathrm{T}} + 1_{q+1} a^{\mathrm{T}}) \|_{\mathrm{F}}$ to obtain an optimal approximation of $\mathrm{vec}^{-1}(\hat{\theta}_{\mathcal{A}})$. Once again, both of these optimization problems are computationally tractable. Assuming that $(\Phi^{\mathrm{T}}\Phi)^{-1}$ exists, the minimizer of $J_{\mathrm{LS}}(\tilde{\theta})$ is given in (33). To compute the minimum of $J_{\mathcal{A}}(\nu, a, \hat{b}, \hat{\theta}_{\mathcal{A}})$ we have the following result.

Proposition 1: *Let $A \in \mathbb{R}^{n\times m}$, $r \in \mathbb{R}^m$ and $s \in \mathbb{R}^n$, and define V: $\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ by*

$$V(x,y) \triangleq ||A-(xr^{\mathrm{T}}+sy^{\mathrm{T}})||_{\mathrm{F}} \tag{54}$$

Then, for all $\alpha \in \mathbb{R}$,

$$\begin{aligned} V(x_\alpha, y_\alpha) &= \min_{\mathbb{R}^n \times \mathbb{R}^m} V(x,y) \\ &= \left\| \left(I_n - \frac{ss^{\mathrm{T}}}{s^{\mathrm{T}}s} \right) A \left(I_m - \frac{rr^{\mathrm{T}}}{r^{\mathrm{T}}r} \right) \right\|_{\mathrm{F}} \end{aligned} \tag{55}$$

where

$$x_\alpha = \left(I_n - \frac{ss^{\mathrm{T}}}{s^{\mathrm{T}}s} \right) A \frac{r}{r^{\mathrm{T}}r} - \alpha s \tag{56}$$

and

$$y_\alpha = \frac{A^{\mathrm{T}}s}{s^{\mathrm{T}}s} + \alpha r \tag{57}$$

Proof: First we note that the gradient of $V^2(x,y)$ is

$$\nabla V^2(x,y) = 2 \begin{bmatrix} r^{\mathrm{T}}rx^{\mathrm{T}} + r^{\mathrm{T}}ys^{\mathrm{T}} - r^{\mathrm{T}}A^{\mathrm{T}} \\ s^{\mathrm{T}}sy^{\mathrm{T}} + s^{\mathrm{T}}xr^{\mathrm{T}} - s^{\mathrm{T}}A \end{bmatrix}$$

Thus

$$\begin{aligned} &\nabla V^2(x_\alpha, y_\alpha) \\ &= 2 \begin{bmatrix} r^{\mathrm{T}}r \left(\left(I_n - \frac{ss^{\mathrm{T}}}{s^{\mathrm{T}}s} \right) A \frac{r}{r^{\mathrm{T}}r} - \alpha s \right)^{\mathrm{T}} + r^{\mathrm{T}} \left(\frac{A^{\mathrm{T}}s}{s^{\mathrm{T}}s} + \alpha r \right) s^{\mathrm{T}} - r^{\mathrm{T}}A^{\mathrm{T}} \\ s^{\mathrm{T}}s \left(\frac{A^{\mathrm{T}}s}{s^{\mathrm{T}}s} + \alpha r \right)^{\mathrm{T}} + s^{\mathrm{T}} \left(\left(I_n - \frac{ss^{\mathrm{T}}}{s^{\mathrm{T}}s} \right) A \frac{r}{r^{\mathrm{T}}r} - \alpha s \right) r^{\mathrm{T}} - s^{\mathrm{T}}A \end{bmatrix} \\ &= 2 \begin{bmatrix} r^{\mathrm{T}}A^{\mathrm{T}} \left(I_n - \frac{ss^{\mathrm{T}}}{s^{\mathrm{T}}s} \right) + \frac{r^{\mathrm{T}}A^{\mathrm{T}}ss^{\mathrm{T}}}{s^{\mathrm{T}}s} - r^{\mathrm{T}}A^{\mathrm{T}} \\ s^{\mathrm{T}} \left(I_n - \frac{ss^{\mathrm{T}}}{s^{\mathrm{T}}s} \right) A \frac{rr^{\mathrm{T}}}{r^{\mathrm{T}}r} \end{bmatrix} \\ &= 0_{(m+n)\times 1} \end{aligned}$$

Furthermore, the Hessian of V^2 is given by

$$\nabla^2 V^2(x,y) = 2 \begin{bmatrix} r^{\mathrm{T}}rI_n & sr^{\mathrm{T}} \\ rs^{\mathrm{T}} & s^{\mathrm{T}}sI_m \end{bmatrix} \tag{58}$$

Since

$$r^{\mathrm{T}}rI_n - sr^{\mathrm{T}}(s^{\mathrm{T}}sI_m)^{-1}rs^{\mathrm{T}} = r^{\mathrm{T}}r \left(I_n - \frac{ss^{\mathrm{T}}}{s^{\mathrm{T}}s} \right) \geq 0 \tag{59}$$

$\nabla^2 V^2$ is non-negative definite and it follows from Bazaraa *et al.* (1993, p. 91) that V^2 is convex. Since $\nabla V^2(x_\alpha, y_\alpha) = 0$, it follows from Bazaraa *et al.* (1993, p. 134) that x_α and y_α are global minimizers of V^2. Since minimizing V^2 is equivalent to minimizing V, x_α and y_α are global minimizers of V. Finally, the minimum cost in (55) is obtained by substituting (56) and (57) into (54). □

The estimates of $\hat{a}$ and $\hat{\nu}$ that minimize $J_{\mathcal{A}}(\nu, a, \hat{b}, \hat{\theta}_{\mathcal{A}})$ are given by

$$\hat{a} = \mathrm{vec}^{-\mathrm{T}}(\hat{\theta}_{\mathcal{A}}) \frac{1_{q+1}}{q+1} + \alpha \hat{b} \tag{60}$$

and

$$\hat{\nu} = \left(I_{q+1} - \frac{1_{q+1}1_{q+1}^{\mathrm{T}}}{q+1} \right) \mathrm{vec}^{-1}(\hat{\theta}_{\mathcal{A}}) \frac{\hat{b}}{\hat{b}^T\hat{b}} - \alpha 1_{q+1} \tag{61}$$

for an arbitrary stability parameter α. Furthermore, $\hat{\lambda}$ is given by

$$\hat{\lambda} = \frac{\hat{\theta}_\lambda - d_s 1_n^{\mathrm{T}} \hat{a}}{1_n^{\mathrm{T}} \hat{b}} \tag{62}$$

PLLS identification with a non-linear feedback model is summarized as follows:

(1) Collect input–output measurements $u(k)$ and $y(k)$, $k = 0, \ldots, l$, where $l \geq n$.
(2) Form the regression matrix Φ in (47) and the output vector Y in (22).
(3) Obtain $\hat{\bar{\theta}}$ by solving the linear least squares problem given by (33).
(4) Extract $\hat{\theta}_{\mathcal{A}}$ and $\hat{b}$ from $\hat{\bar{\theta}}$.
(5) Set the stability parameter α.
(6) Compute the parameter estimates $\hat{a}$, $\hat{\nu}$ and $\hat{\lambda}$ using (60), (61) and (62).

4.3. *Hammerstein/non-linear feedback model identification*

Consider the single-input single-output Hammerstein/non-linear feedback model in (2). Representing f and h as in §§ 4.1 and 4.2, and noting (41), the time series can be written as

$$\begin{aligned} y(k) = \sum_{j=1}^{n} \{ &b_j \mu^{\mathrm{T}} \eta(u(k-j)) - (b_j \nu^{\mathrm{T}} + a_j 1_{q+1}^{\mathrm{T}}) \zeta(y(k-j)) \\ &- a_j d_s + b_j(\kappa - \lambda) \} \end{aligned} \tag{63}$$

and (63) can further be expressed as in (16), where

$$\theta \triangleq \begin{bmatrix} \mathrm{vec}\,(\nu b^{\mathrm{T}} + 1_{q+1} a^{\mathrm{T}}) \\ \mathrm{vec}\,(\mu b^{\mathrm{T}}) \\ 1_n^{\mathrm{T}}((\lambda - \kappa) b + d_s a) \end{bmatrix} \in \mathbb{R}^{n(q+1)+n(p+1)+1} \tag{64}$$

and

$$\phi(k) \triangleq \begin{bmatrix} -\phi_\zeta^{\mathrm{T}}(k) & \phi_\eta^{\mathrm{T}}(k) & -1 \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{n(q+1)+n(p+1)+1} \tag{65}$$

where $\phi_\eta(k)$ and $\phi_\zeta(k)$ are defined in (20) and (45), respectively.

Next consider input–output measurements $u(k)$ and $y(k)$ for $k = 0, \ldots, l$, where $l \geq n$, and form the least squares cost in (21), where Y is defined as in (22) and

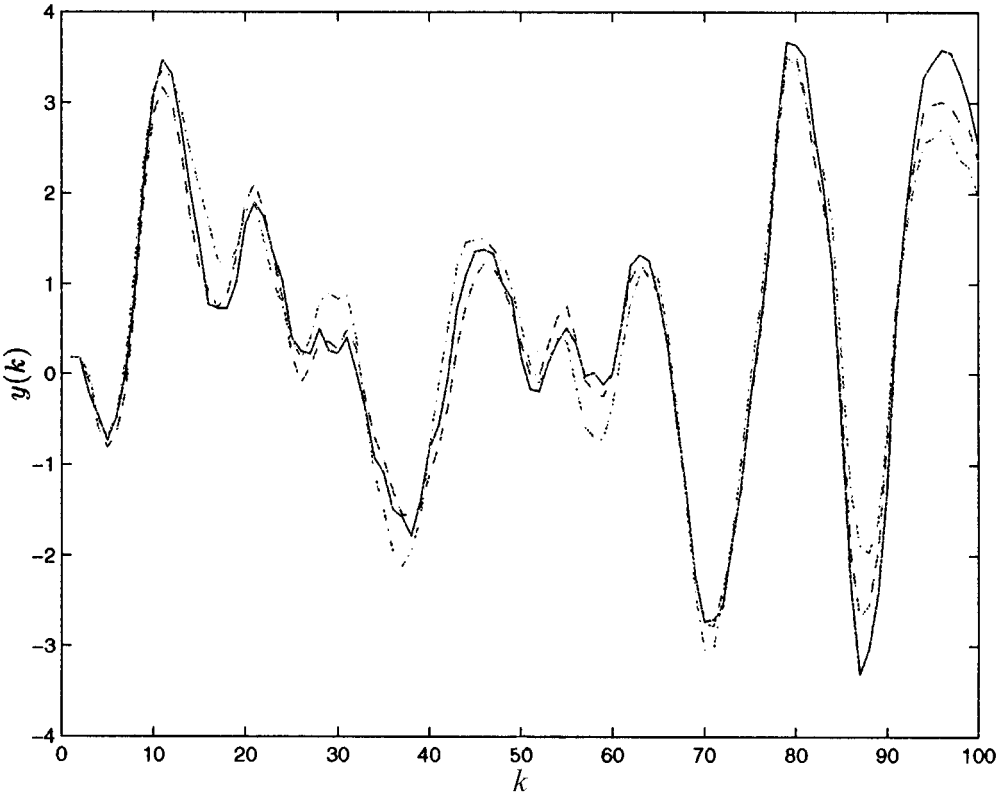

Figure 4. Simulation of the Hammerstein and linear models using the identification input signal (——, actual output; – – – –, Hammerstein model output (RMS error = 0.54); – · – · –, linear model output (RMS error = 0.59)).

$$\Phi \triangleq \begin{bmatrix} \phi(n)^{\mathrm{T}} \\ \vdots \\ \phi(l)^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} -\Phi_\zeta & \Phi_\eta & -1_{n-l+1} \end{bmatrix} \in \mathbb{R}^{(l-n+1)\times(n(q+1)+n(p+1)+1)} \tag{66}$$

where Φ_η and Φ_ζ are defined as in (24) and (48), respectively.

By combining the bounding steps from §§ 4.1 and 4.2

$$J(\theta) \leq J_{\mathrm{LS}}(\tilde{\theta}) + \sigma_{\max}(\Phi_\zeta) J_{\mathcal{A}}(\nu, a, b, \theta_{\mathcal{A}}) + \sigma_{\max}(\Phi_\eta) J_{\mathcal{B}}(\mu, b, \theta_{\mathcal{B}}) \tag{67}$$

where $J_{\mathrm{LS}}(\tilde{\theta})$ is defined in (30), $J_{\mathcal{B}}(\mu, b, \theta_{\mathcal{B}})$ is defined in (31), $J_{\mathcal{A}}(\nu, a, b, \theta_{\mathcal{A}})$ is defined in (52), and where

$$\tilde{\theta} \triangleq \begin{bmatrix} \theta_{\mathcal{A}} \\ \theta_{\mathcal{B}} \\ \theta_{\lambda\kappa} \end{bmatrix} \in \mathbb{R}^{n(q+1)+n(p+1)+1} \tag{68}$$

and

$$\theta_{\lambda\kappa} \triangleq 1_n^{\mathrm{T}}((\lambda - \kappa)b + d_s a) \tag{69}$$

To sequentially minimize $J_{\mathrm{LS}}(\tilde{\theta})$, $J_{\mathcal{B}}(\mu, b, \theta_{\mathcal{B}})$ and $J_{\mathcal{A}}(\nu, a, b, \theta_{\mathcal{A}})$, we write

$$\hat{\tilde{\theta}} = \begin{bmatrix} \hat{\theta}_{\mathcal{A}} \\ \hat{\theta}_{\mathcal{B}} \\ \hat{\theta}_{\lambda\kappa} \end{bmatrix} \in \mathbb{R}^{q(n+1)+p(n+1)+1} \tag{70}$$

and determine $\hat{\tilde{\theta}}$ that minimizes the least squares cost $J_{\mathrm{LS}}(\tilde{\theta})$. Then we extract $\hat{\theta}_{\mathcal{B}}$ from $\hat{\tilde{\theta}}$ and determine $\hat{b}$ and $\hat{\mu}$ that minimize $J_{\mathcal{B}}(\mu, b, \hat{\theta}_{\mathcal{B}})$. Finally, we extract $\hat{\theta}_{\mathcal{A}}$ from $\hat{\tilde{\theta}}$ and obtain $\hat{a}$ and $\hat{\nu}$ that minimize $J_{\mathcal{A}}(\nu, a, \hat{b}, \hat{\theta}_{\mathcal{A}})$.

Again, these optimization problems are computationally tractable. Assuming that $(\Phi^{\mathrm{T}}\Phi)^{-1}$ exists, the minimizer of $J_{\mathrm{LS}}(\tilde{\theta})$ is given by (33), the minimizer of $J_{\mathcal{B}}(\mu, b, \hat{\theta}_{\mathcal{B}})$ is given by (37) and the minimizer of $J_{\mathcal{A}}(\nu, a, \hat{b}, \hat{\theta}_{\mathcal{A}})$ is given by (60) and (61). Furthermore, by setting $\hat{\lambda} = \gamma$ the parameter estimate $\hat{\kappa}$ is given by

$$\hat{\kappa} = \gamma - \frac{\hat{\theta}_{\lambda\kappa} - d_s 1_n^{\mathrm{T}} \hat{a}}{1_n^{\mathrm{T}} \hat{b}} \tag{71}$$

for an arbitrary offset parameter γ.

PLLS identification with a Hammerstein/non-linear feedback model is summarized as follows:

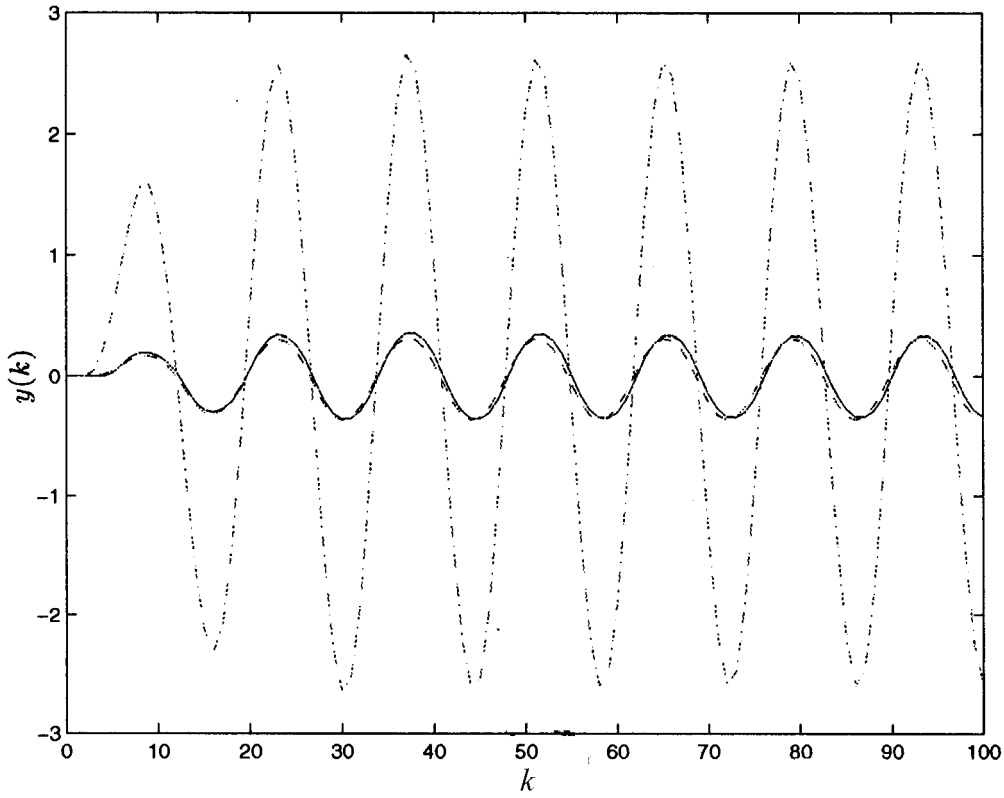

Figure 5. Simulation of the Hammerstein and linear models using the validation signal (——, actual output; – – – –, Hammerstein model output (RMS error = 0.045); – · – · –, linear model output (RMS error = 1.55)).

(1) Collect input–output measurements $u(k)$ and $y(k)$, $k = 0, \ldots, l$, where $l \geq n$.

(2) Form the regression matrix Φ in (66) and the output vector Y in (22).

(3) Obtain $\hat{\tilde{\theta}}$ from (70) by solving the linear least squares problem given by (33).

(4) Extract $\hat{\theta}_{\mathcal{B}}$ from $\hat{\tilde{\theta}}$.

(5) Compute the singular value decomposition of $\mathrm{vec}^{-1}(\hat{\theta}_{\mathcal{B}})$ and obtain u_1, v_1 and $\sigma_{\max}(\mathrm{vec}^{-1}(\hat{\theta}_{\mathcal{B}}))$.

(6) Set the input gain parameter β.

(7) Compute the parameter estimates $\hat{b}$ and $\hat{\mu}$ using (37).

(8) Extract $\hat{\theta}_{\mathcal{A}}$ from $\hat{\tilde{\theta}}$.

(9) Set the stability parameter α.

(10) Compute the parameter estimates $\hat{a}$, $\hat{\nu}$, using (60) and (61).

(11) Set the offset parameter γ.

(12) Set $\hat{\lambda} = \gamma$ and compute the parameter estimate $\hat{\kappa}$ using (71).

5. Numerical examples

In this section we illustrate the PLLS identification technique through two numerical examples. In each example, we simulate a non-linear system to obtain input-output data. We then use this data to identify the system using PLLS identification. The results are compared to least squares identification using a linear model structure. This comparison demonstrates the degree of non-linearity present in the system. Additionally, a separate input signal is used to validate the models. For purposes of examining the individual blocks within the identified models, the input gain parameter β, the stability parameter α, and the offset parameter γ, were chosen to match the identified non-linearities with the components of the simulated system.

5.1. *Example 1*

We consider a discrete-time Hammerstein system containing a deadband, where

$$G(\mathbf{q}^{-1}) = \frac{0.5992 + 0.5679\mathbf{q}^{-1}}{1 - 1.706\mathbf{q}^{-1} + 0.8521\mathbf{q}^{-2}} \qquad (72)$$

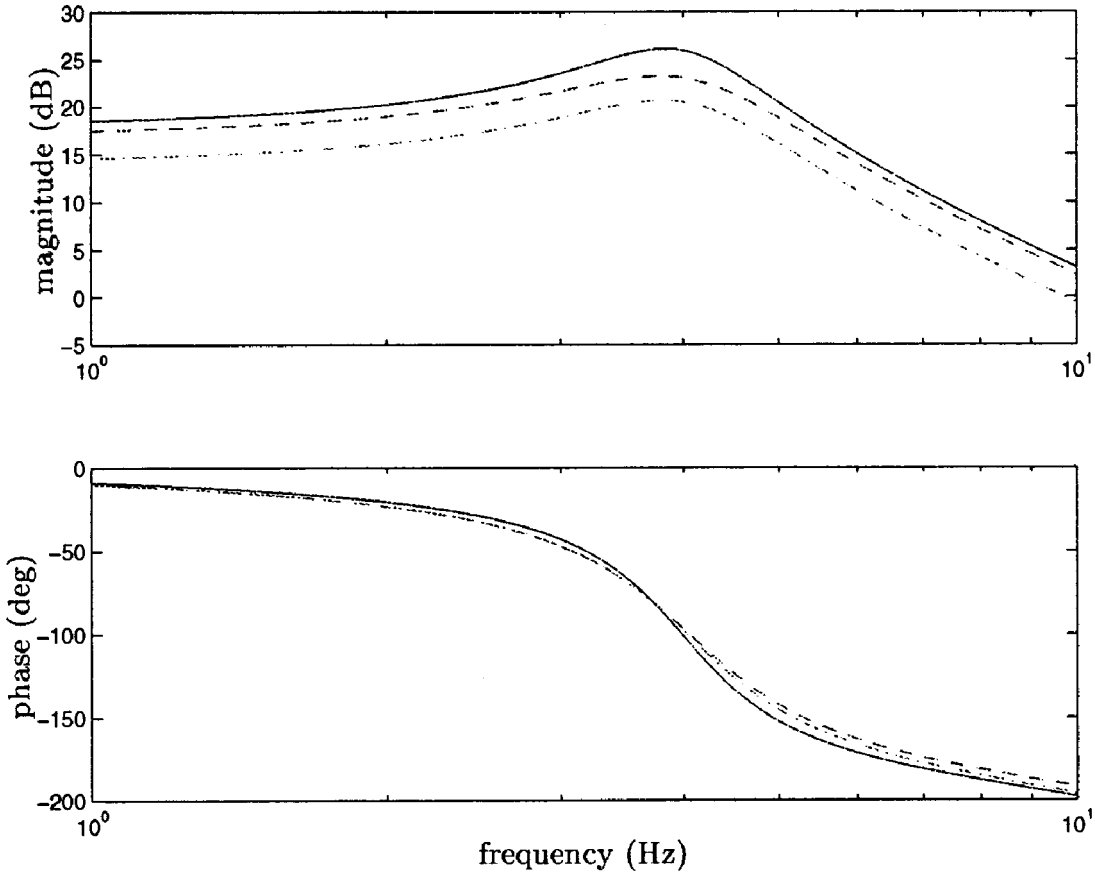

Figure 6. Average frequency response of linear blocks (——, linear block of actual system; – – – –, estimate of linear block of Hammerstein model; – · – · –, estimate of linear model).

and

$$f_0(u) = \begin{cases} u + 0.25, & u \leq -0.25 \\ 0, & -0.25 < u < 0.25 \\ u - 0.25, & u \geq 0.25 \end{cases} \quad (73)$$

The identification input signal is generated as the realization of a zero-mean, white noise sequence that is uniformly distributed between -1 and 1. Noisy output data is generated for 1000 samples, where the measurement noise is a zero-mean white noise Gaussian sequence. The resulting signal-to-noise ratio (SNR) in terms of standard deviation is 22.4.

PLLS identification using a Hammerstein model is performed using $u(k)$ and $y(k)$, $k = 0, \ldots, l$, where $l = 1000$, $n = 2$, $p = 10$, $r = 5$ and an equally spaced domain partition

$$c = [-0.9 \quad -0.7 \quad -0.5 \quad \cdots \quad 0.5 \quad 0.7 \quad 0.9]^{\mathrm{T}} \quad (74)$$

For comparison, linear least squares identification is performed using the same input–output measurements with model order $n = 2$. In order to estimate the uncertainty in the parameter estimates, 100 Monte Carlo runs are conducted using different noise realizations.

Figure 4 shows the model outputs for the identification input signal of a typical run, and Figure 5 shows the model outputs for the validation input signal $u(k) = 0.3 \sin(4.5k)$. For clarity only the first 100 samples are shown. Examination of Figure 4 reveals that both the linear model and the Hammerstein model reasonably fit the identification data. The root mean squared (RMS) error of the model outputs for this case is 0.54 for the Hammerstein model and 0.59 for the linear model. For this example a linear model is capable of fitting the identification data of the non-linear system. In contrast, when examining the model outputs for the validation signal, the linear model performs poorly. The RMS error for this case is 0.045 for the Hammerstein model and 1.55 for the linear model.

A linear model of order $n = 8$ is also identified based on the identification data. The RMS error of the model

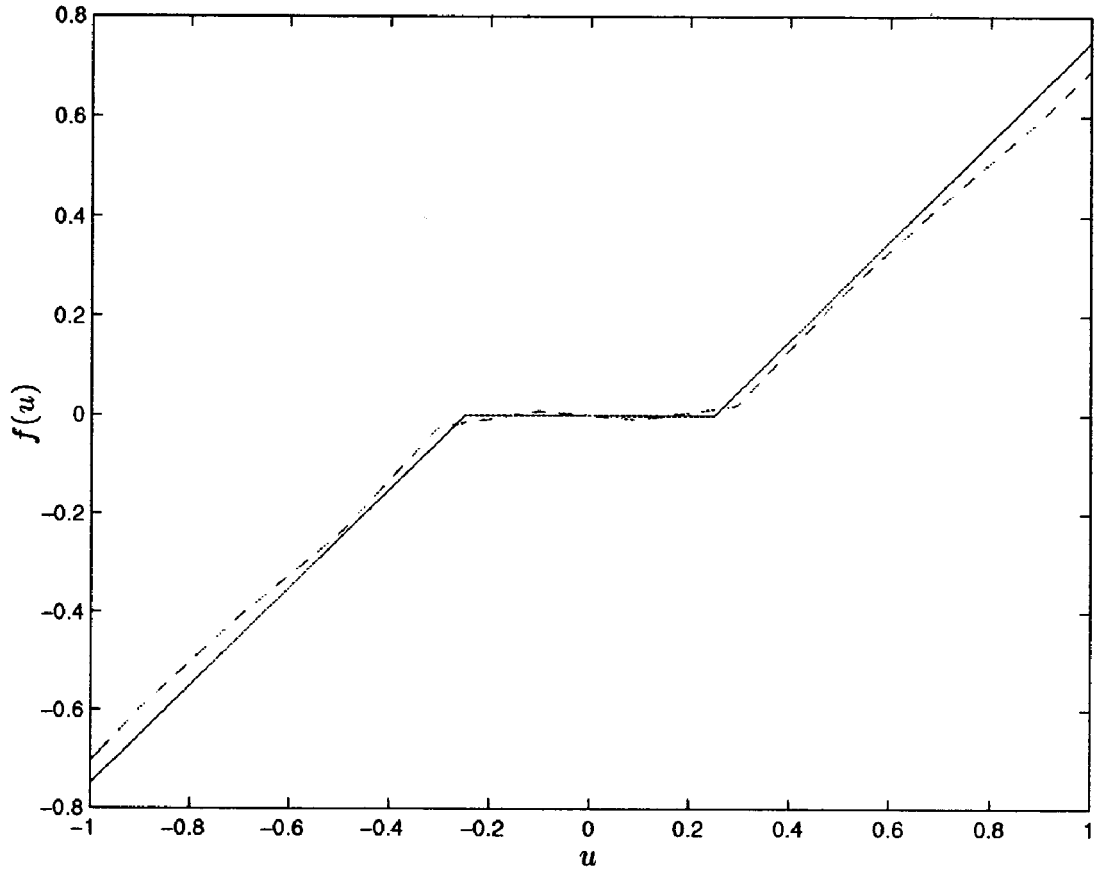

Figure 7. Average estimate of non-linearity (——, actual non-linearity $f_0(u)$; – – – –, estimate of non-linearity $f(u)$).

output for the identification data is 0.52, while the RMS error for the validation output is 1.78. For this example, a higher order linear model offers no improvement in fitting the Hammerstein system.

Figure 6 shows the frequency response of the linear block $G(\mathbf{q}^{-1})$ of the Hammerstein system, the average frequency response of the linear block in the Hammerstein model and the average frequency response of the linear model. Figure 7 shows the non-linearity in the Hammerstein system and the average non-linearity in the Hammerstein model.

The mean parameter estimates and the associated 95% confidence interval (plus or minus two standard deviations) are given below. For the Hammerstein model the parameter estimates of the linear block are

$$\hat{a} = \begin{bmatrix} -1.6677 \pm 0.0080 \\ 0.8146 \pm 0.0076 \end{bmatrix}, \qquad \hat{b} = \begin{bmatrix} 0.5913 \pm 0.2917 \\ 0.4570 \pm 0.6423 \end{bmatrix} \tag{75}$$

while the estimated parameters of the non-linear block are

$$\hat{\mu} = \begin{bmatrix} 1.0855 \pm 1.8683 \\ 0.9359 \pm 0.6411 \\ 0.8384 \pm 0.7071 \\ 1.1005 \pm 0.8063 \\ 0.1604 \pm 0.4516 \\ -0.0818 \pm 0.4131 \\ 0.1356 \pm 0.3524 \\ 1.0689 \pm 0.8178 \\ 0.9257 \pm 0.7231 \\ 0.8616 \pm 0.7642 \\ 0.9705 \pm 1.9314 \end{bmatrix}, \quad \hat{\kappa} = 0.0084 \pm 0.0391 \tag{76}$$

The parameter estimates for the linear model are

$$\hat{a} = \begin{bmatrix} -1.6746 \pm 0.0078 \\ 0.8222 \pm 0.0074 \end{bmatrix}, \qquad \hat{b} = \begin{bmatrix} 0.3798 \pm 0.0228 \\ 0.3711 \pm 0.0246 \end{bmatrix} \tag{77}$$

Examination of the parameter estimates of the piecewise linear static non-linearity reveals that parameter bias and uncertainty increases for slope parameter indices with greater distance from the primary index

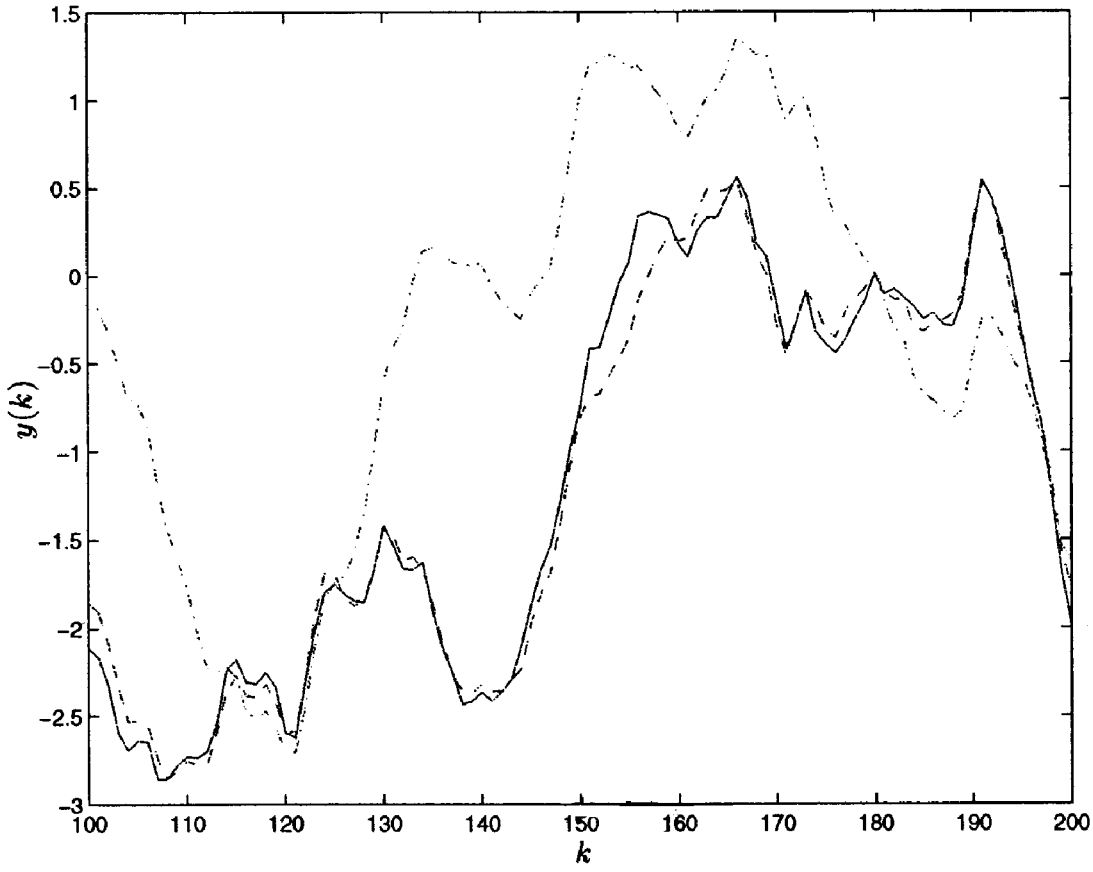

Figure 8. Simulation of the non-linear feedback and linear models using the identification input signal (——, actual output; – – – –, non-linear feedback model output (RMS error = 0.82); –·–·–, linear model output (RMS error = 1.73)).

parameter $r = 5$ at which the value of f is specified. This is consistent with the form of the parameterization, where the error in approximating f grows as the error due to individual slope parameters accumulates for ordinates increasingly farther from r. This observation suggests that r should be chosen to lie in the centre of the region in the domain of f where the greatest accuracy is desired.

For additional model accuracy, a bootstrapping technique can be used by iteratively refining the partitioning c to accommodate the estimated non-linearity. In this example, the specific partitioning c is chosen to be equally spaced with $p = 10$. This allows visual inspection of the estimated non-linearity to suggest a more parsimonious choice of c. The identified non-linearity clearly resembles a deadband with breakpoints in the intervals $[-0.3, -0.2]$ and $[0.2, 0.3]$. For comparison purposes, the more parsimonious parameterization given by $n = 2$, $p = 3$, $r = 1$, and partitioning $c = [-0.25\ \ 0.25]^{\mathrm{T}}$ is considered as well. The resulting RMS error for the model output from the experimental signal decreases to 0.33, and the resulting parameter estimate certainty increases.

The parameter estimates of the linear block of the Hammerstein model are

$$\hat{a} = \begin{bmatrix} -1.6700 \pm 0.0072 \\ 0.8169 \pm 0.0070 \end{bmatrix}, \qquad \hat{b} = \begin{bmatrix} 0.5903 \pm 0.0525 \\ 0.5760 \pm 0.0549 \end{bmatrix} \tag{78}$$

while the parameter estimates for the non-linear block are

$$\hat{\mu} = \begin{bmatrix} 1.0236 \pm 0.0616 \\ 0.0046 \pm 0.0707 \\ 1.0135 \pm 0.0512 \end{bmatrix}, \qquad \hat{\kappa} = -0.0007 \pm 0.0214 \tag{79}$$

The specific choice of partitioning c may affect the invertibility of the matrix $\Phi^{\mathrm{T}}\Phi$. For linear model identification, the invertibility of $\Phi^{\mathrm{T}}\Phi$ is guaranteed by using a sufficiently exciting input signal. An interpretation of this condition can be viewed in terms of sufficient frequency content to excite all of the dynamics of the linear system. When considering PLLS identification the magnitude of the input signal affects the invertibility of this

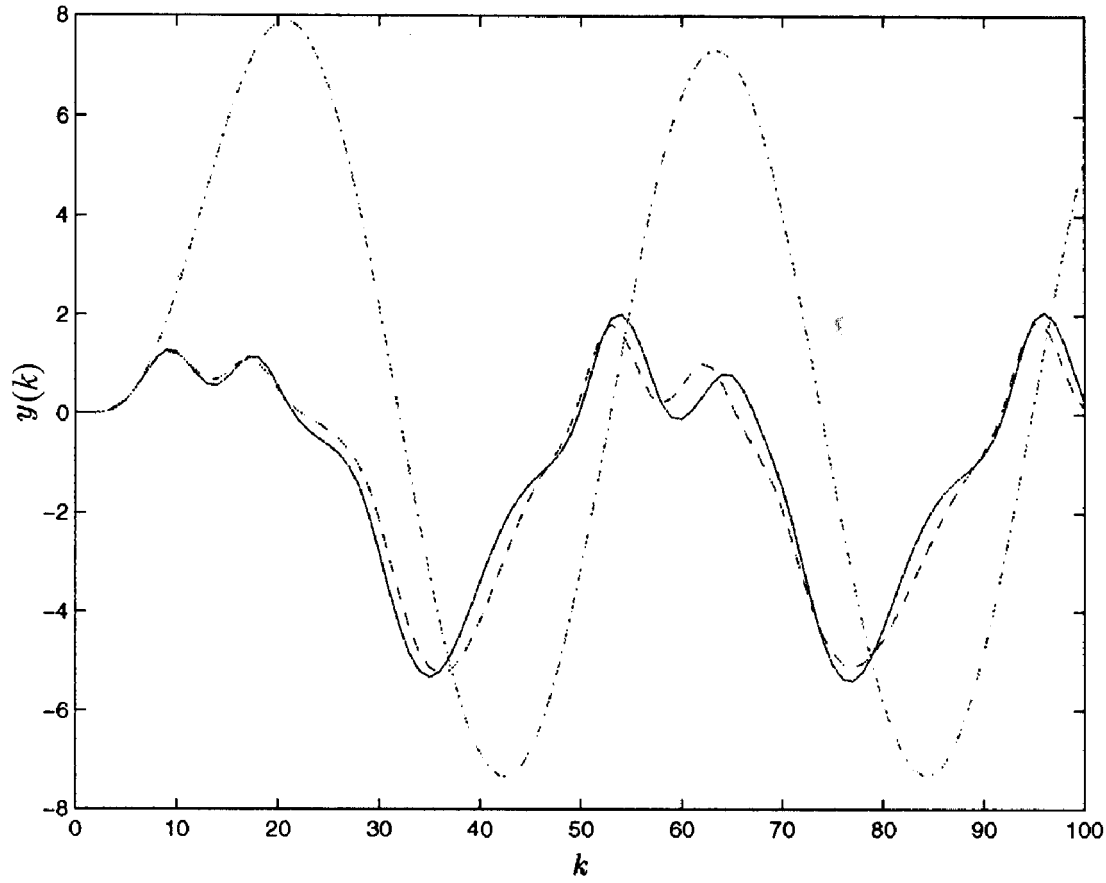

Figure 9. Simulation of the non-linear feedback and linear models using the validation signal (——, actual output; – – – –, non-linear feedback model output (RMS error = 0.39); –·–·–, linear model output (RMS error = 4.70)).

matrix as well. It is necessary that each partition in c contain at least one element of the input time series $u(k)$ in order to guarantee the invertibility of $\Phi^{\mathrm{T}}\Phi$. Typically, the partition c must be chosen to satisfy this requirement. Sufficient conditions on the input signal and partitioning in order to guarantee the invertibility of $\Phi^{\mathrm{T}}\Phi$ is left for future work.

Next we compared the results of sequentially minimizing (30) and (31) in the least squares bound to the results of numerically optimizing (21) using Newton's method. Monte Carlo runs were performed over various signal-to-noise ratios. The cost J_{SMB} was then computed by using the estimates $\hat{a}$, $\hat{b}$, $\hat{\mu}$ and $\hat{\kappa}$ from sequentially minimizing (30) and (31), while the cost J_{N} was computed by using an interior-reflective Newton method in the Matlab Optimization Toolbox. The initial parameter vector for the Newton method was chosen randomly.

The values of J_{SMB} and J_{N} were computed using 25 trials at each signal-to-noise ratio for $k = 0, \ldots, 1000$. Results showed that J_{SMB} approached J_{N} as SNR increased, while for small SNR the difference was minimal. Although J_{SMB} is a bound on the true cost, J_{N} was larger for some SNRs. These results revealed slow convergence when using Newton's method which was further demonstrated by large standard deviations of J_{N} as compared to J_{SMB}.

A more serious drawback of the numerical optimization was that some initial choices of the parameter vector θ converged incorrectly. For example, by choosing $\theta = 0$ as the initial parameter vector, the converged parameter vector had only two non-zero elements resulting in $B(\mathbf{q}^{-1}) = 0$ and $f(u) = 0$. Consequently, we choose the initial parameter vector randomly for each of the runs in this comparison.

Furthermore, J_{SMB} was compared to J_{N} for the case of no noise but varying input sequences. In this case, J_{SMB} was always slightly smaller than J_{N}, which suggests that sequential minimization of the least squares bound

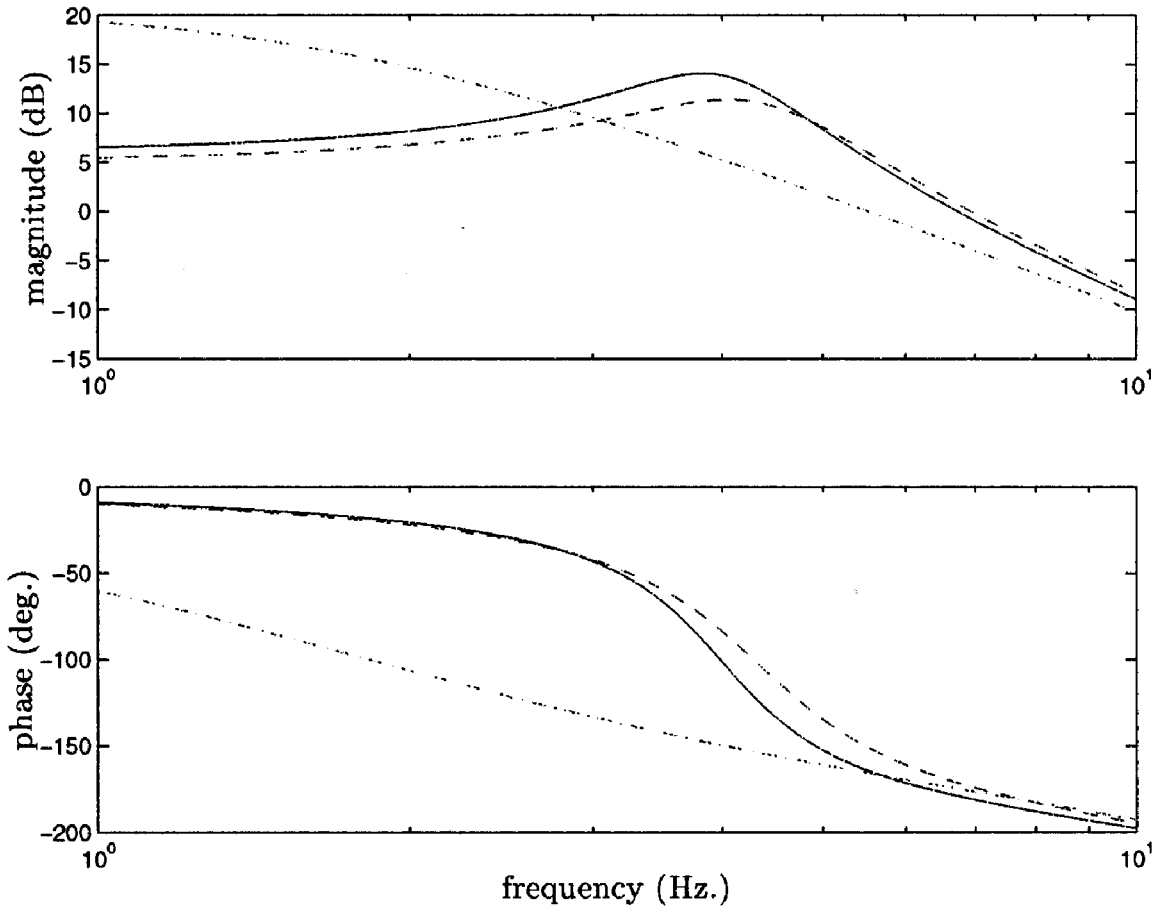

Figure 10. Average frequency response of linear blocks (——, actual linear block of system; – – – –, estimate of linear block of non-linear feedback model; – · – · –, estimate of linear model).

yields the global minimum of (21) in the case of no noise.

5.2. *Example 2*

We consider a discrete-time system with non-linear feedback structure, where

$$G(\mathbf{q}^{-1}) = \frac{0.1498 + 0.142\mathbf{q}^{-1}}{1 - 1.706\mathbf{q}^{-1} + 0.8521\mathbf{q}^{-2}} \tag{80}$$

and

$$h_0(y) = 1 - \mathrm{e}^{-y^4} \tag{81}$$

The identification input–output data is generated as in Example 1, where a uniformly distributed Gaussian input sequence is used, and measurement noise over 100 Monte Carlo runs is considered. The resulting signal-to-noise ratio (std.) is SNR = 30.0.

PLLS identification using a non-linear feedback model is performed using $u(k)$ and $y(k)$, $k = 0, \ldots, l$, where $l = 1000$, $n = 2$, $q = 17$, $s = 8$ and an equally spaced domain partition

$$d = [-3.0 \quad -2.75 \quad -2.5 \quad \cdots \quad 0.50 \quad 0.75 \quad 1.00]^{\mathrm{T}} \tag{82}$$

As in Example 1, linear least squares identification is performed using the same input–output measurements with model order $n = 2$.

The mean parameter estimates and the associated 95% confidence interval are given below. For the non-linear feedback model the parameter estimates of the linear block are

$$\hat{a} = \begin{bmatrix} -1.6406 \pm 0.0241 \\ 0.8072 \pm 0.0169 \end{bmatrix}, \qquad \hat{b} = \begin{bmatrix} 0.1499 \pm 0.0084 \\ 0.1490 \pm 0.0093 \end{bmatrix} \tag{83}$$

while the estimated parameters of the non-linear block are

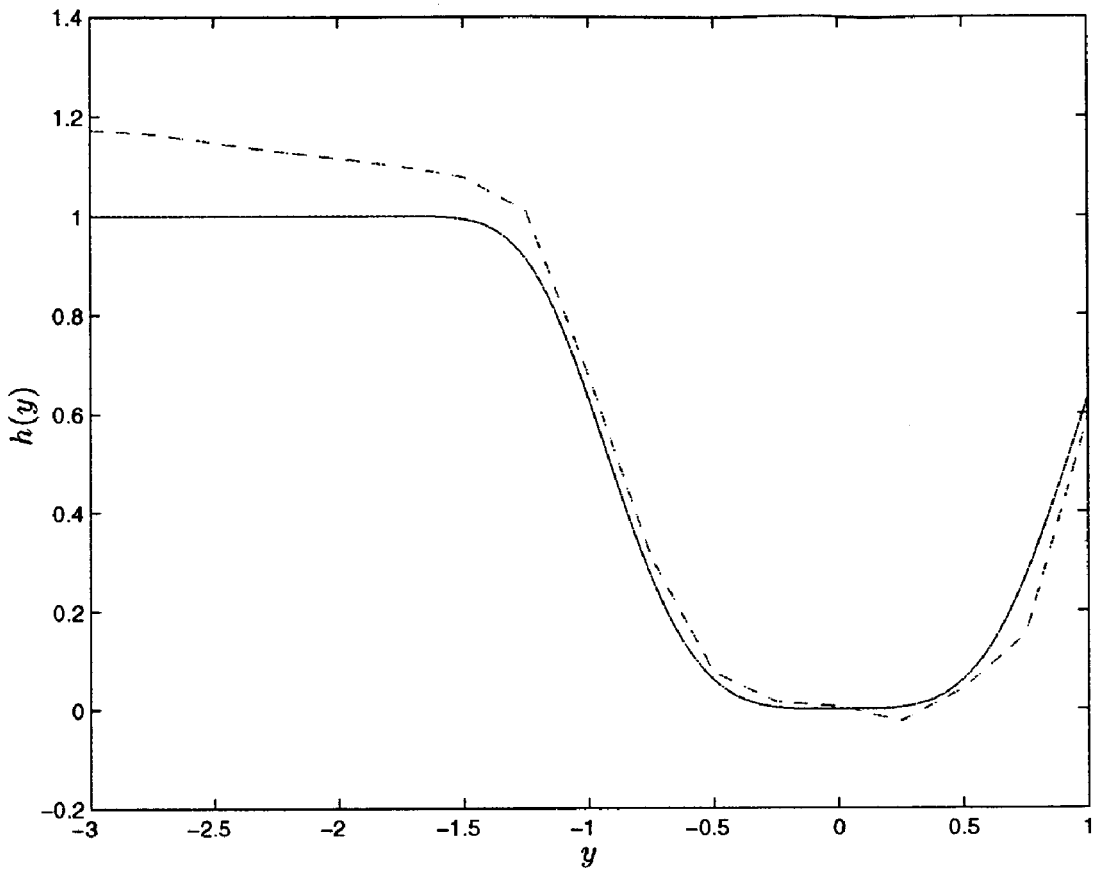

Figure 11. Average estimate of non-linearity (——, actual non-linearity $h_0(y)$; – · – · –, estimate of non-linearity $h(y)$).

$$\hat{\nu} = \begin{bmatrix} -0.0790 \pm 0.0868 \\ -0.0326 \pm 0.3605 \\ -0.0712 \pm 0.5029 \\ -0.0682 \pm 0.4612 \\ -0.0568 \pm 0.4070 \\ -0.0671 \pm 0.4093 \\ -0.0835 \pm 0.4355 \\ -0.2722 \pm 0.3942 \\ -1.3227 \pm 0.4676 \\ -1.4768 \pm 0.4627 \\ -0.9392 \pm 0.3653 \\ -0.2420 \pm 0.3111 \\ -0.0369 \pm 0.3670 \\ -0.1198 \pm 0.3950 \\ 0.2646 \pm 0.3774 \\ 0.4398 \pm 0.5607 \\ 1.7733 \pm 0.9522 \\ 0.8983 \pm 1.4650 \end{bmatrix}, \quad \hat{\lambda} = 1.0101 \pm 0.0845 \quad (84)$$

The parameter estimates for the linear model are

$$\hat{a} = \begin{bmatrix} -1.7181 \pm 0.0094 \\ 0.7432 \pm 0.0093 \end{bmatrix}, \qquad \hat{b} = \begin{bmatrix} 0.1426 \pm 0.0083 \\ 0.1395 \pm 0.0094 \end{bmatrix} \quad (85)$$

Figure 8 shows the model outputs for the identification input signal of a typical run, and Figure 9 shows the model outputs for the validation input signal $u(k) = \sin(1.5k)$. For clarity, only 100 samples are shown. Examination of these figures reveals that the estimate of the non-linear feedback model fits the data better than the estimate of the linear model. The root mean squared (RMS) error of the model outputs for the case of the identification signal is 0.82 for the non-linear feedback model and 1.73 for the linear model. The RMS error of the model outputs for the case of the validation signals is 0.39 for the non-linear feedback model and 4.70 for the linear model.

Figure 10 shows the frequency response of the linear block $G(\mathbf{q}^{-1})$ of the non-linear feedback system, the average frequency response of the linear block in the non-linear feedback model and the average frequency response of the linear model. Figure 11 shows the non-linearity in the non-linear feedback system and the average non-linearity in the non-linear feedback model. We note that although the non-linearity h_0 is not piecewise linear, the results of the identification approximate h_0 well. Furthermore, the accuracy of the non-linearity h

is greater for partition intervals near the primary index $s = 8$ as compared to partition intervals farther away. Once again, the choice of the primary index affects the distribution of the bias in estimating h.

6. Conclusions

We considered non-linear identification using a Hammerstein/non-linear feedback model with piecewise linear static maps. Our method used a point-slope parameterization that leads to a computationally tractable optimization problem. This identification method simultaneously approximated the linear dynamic and static non-linear blocks, and did not require prior information about the form of the non-linearity. Two numerical examples were investigated, and the effects of noisy output data was considered. These examples revealed that the PLLS identification technique produced reasonable models of the system in question. Future work will involve the application of this technique to real data.

Acknowledgements

This research was supported in part by the Air Force Office of Scientific Research under grants F49620-98-1-0037 and F49620-97-1-0406.

References

Ahmed, A. M., 1995, BPD computation and model reference adaptive control (MRAC) of Hammerstein plants. *IEE Proceedings—Control Theory Applications*, **142**, 475–485.

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

Bazaraa, M. S., Sherali, H. D., and Shetty, C. M., 1993, *Nonlinear Programming Theory and Algorithms* (New York: Wiley), 2nd edition.

Bendat, J. S., 1989, *Nonlinear Systems Techniques and Applications* (New York: Wiley).

Billings, S. A., 1980, Identification of nonlinear systems—A survey. *IEE Proceedings*, **127**, 272–285.

Billings, S. A., and Fakhouri, S. Y., 1977, Identification of nonlinear systems using the Wiener model. *Electronic Letters*, **13**, 502–504.

Billings, S. A., and Fakhouri, S. Y., 1982, Identification of systems containing linear dynamic and static nonlinear elements. *Automatica*, **18**, 15–26.

Billings, S. A., and Voon, W. S. F., 1987, Piecewise linear identification of non-linear systems. *International Journal of Control*, **46**, 215–235.

Boyd, S., and Chua, L. O., 1983, Uniqueness of a basic nonlinear structure. *IEEE Transactions on Circulation and Systems*, **cas–30**, 648–651.

Boyd, S., and Chua, L. O., 1985, Fading memory and the problem of approximating nonlinear operators with Voletta series. *IEEE Transactions on Circular Systems*, **SAC–32**, 1150–1161.

Chen, S., Billings, S. A., Cowan, C. F. N., and Grant, P. M., 1990, Non-linear systems identification using radial basis functions. *International Journal of Systems Science*, **21**, 2513–2539.

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

Eskinat, E., Johnson, S. H., and Luyben, W. L. 1991, Use of Hammerstein models in identification of nonlinear systems. *AIChE Journal*, **37**, 255–268.

Fitzgerald, W. J., Walden, A., Smith, R., and Young, P. C., (Eds), 2000, *Nonstationary and Nonlinear Signal Processing* (Cambridge: Cambridge University Press).

Greblicki, W., 1989, Non-parametric orthogonal series identification of Hammerstein systems. *International Journal of Systems Science*, **20**, 2355–2367.

Greblicki, W., 1999, Recursive identification of continuous-time Wiener systems. *International Journal of Control*, **72**, 981–989.

Haddad, W. M., and Bernstein, D. S., 1993, Explicit construction of quadratic Lyapunov functions for the small gain, positivity, circle, and Popov theorems and their application to robust stability. Part I: Continuous-time theory. *International Journal of Robust Nonlinear Control*, **3**, 313–339.

Haddad, W. M., and Bernstein, D. S., 1994, Explicit construction of quadratic Lyapunov functions for the small gain, positivity, circle, and Popov theorems and their application to robust stability. Part II: Discrete-time theory. *International Journal of Robust Nonlinear Control*, **4**, 249–265.

Hunter, I. W., and Korenberg, J., 1986, The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biol. Cybern.*, **55**, 135–144.

Hwang, C. L., 1995, Nonlinear control design for a Hammerstein model system. *IEE Proceedings—Control Theory Applications*, **142**, 277–285.

Juditsky, A., Hjalmarsson, H., Benveniste, A., Delyon, B., Ljung, L., Sjoberg, J., and Zhang, Q., 1995, Nonlinear black-box models in system identification: Mathematical foundations. *Automatica*, **31**, 1725–1750.

Kahlert, C., and Chua, L. O., 1990, A generalized canonical piecewise-linear representation. *IEEE Transactions on Circulation and Systems*, **37**, 373–383.

Kahlert, C., and Chua, L. O., 1992, The complete canonical piecewise-linear representation—Part 1: The geometry of the domain space. *IEEE Transactions on Circulation and System*, **39**, 222–236.

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

Krzyzak, A., 1989, Identification of discrete Hammerstein systems by the Fourier series regression estimate. *International Journal of Systems Science*, **20**, 1729–1744.

Ljung, L., 1999, *System Identification: Theory for the User*, 2nd edition (Upper Saddle River, NJ: Prentice Hall).

Ljung, L., and Soderstrom, T., 1983, *Theory and Practice of Recursive Identification* (MIT Press).

Lu, S., and Basar, T., 1998, Robust nonlinear system identification using neural-network models. *IEEE Neural Networks*, **9**, 407–429.

Narendra, K. S., and Gallman, P. G., 1966, An iterative method for the identification of nonlinear systems using a Hammerstein model. *IEEE Transactions on Automatic Control*, **12**, 546–550.

Narendra, K. S., and Parthasarathy, K., 1990, Identification and control of dynamical systems using neural networks. *IEEE Transactions on Neural Networks*, **1**, 4–27.

Narendra, K. S., and Taylor, J. H., 1973, *Frequency Domain Criteria for Absolute Stability* (New York: Academic Press).

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

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

Petit, N. B. O. L., 1985, *Analysis of Piecewise Linear Dynamical Systems* (New York: Wiley).

Rangan, S., Wolodkin, G., and Poola, K., 1995, New results for Hammerstein system identification. In *Proceedings of the IEEE Conference on Decision and Control*, New Orleans, LA, USA, pp. 697–702.

Rugh, W. J. 1981, *Nonlinear System Theory: The Volterra/Wiener Approach* (Johns Hopkins University Press).

Sandor, J., and Williamson, D., 1978, Identification and analysis of non-linear systems by tensor techniques. *International Journal of Control*, **27**, 853–878.

Shi, J., and Sun, H. H., 1990, Nonlinear system identification for cascaded block model: An application to electrode polarization impedance. *IEEE Transactions on Biomedical Engineering*, **37**, 574–587.

Sjober, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P., Hjalmarsson, H., and Juditsky, A., 1995, Nonlinear black-box modeling in system identification: A unified overview. *Automatica*, **31**, 1691–1724.

Soderstrom, T., and Stoica, P., 1989, *System Identification* (Upper Saddle River, NJ: Prentice Hall).

Sontag, E., 1981, Nonlinear regulation: The piecewise linear approach. *IEEE Transactions on Automatic Control*, **AC–26**, 346–358.

Stewart, G. W., and Sun, J., 1990, *Matrix Perturbation Theory* (Academic Press).

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

Vandersteen, G., and Schoukens J., 1999, Measurement and identification of nonlinear systems consisting of linear dynamic blocks and one static nonlinearity. *IEEE Tractions Automatic Control*, **44**, 1266–1271.

Wemhoff, E., Packard, A., and Poola, K., 1999, On the identification of nonlinear maps in a general interconnected system. In *Proceedings of the American Control Conference*, San Diego, CA, USA, June, pp. 3456–3461.

Young, P. C., 2000, Stocastic, dynamic modelling and signal processing: time variable and state dependent parameter estimation. In W. J. Fitzgerald, A. Walden, R. Smith and P. C. Young (Eds) *Nonstationary and Nonlinear Signal Processing* (Cambridge: Cambridge University Press).

Young, P. C., McKenna, P., and Bruun, J., 2001, Identification of nonlinear stochastic systems by state dependent parameter estimation. *International Journal of Control:* appears in this Special issue.