Estimation of Generalised Hammerstein-Wiener Systems ⋆

Adrian Wills * **Brett Ninness** **

* *School of Electrical Engineering and Computer Science, University of Newcastle, Callaghan, NSW, 2308, Australia (Tel: +61 2 49216028; e-mail: adrian.wills@newcastle.edu.au).*
** *School of Electrical Engineering and Computer Science, University of Newcastle, Callaghan, NSW, 2308, Australia (Tel: +61 2 49216028; e-mail: brett.ninness@newcastle.edu.au).*

Abstract: This paper examines the use of a so-called "generalised Hammerstein–Wiener" model structure that is formed as the concatenation of an arbitrary number of Hammerstein systems. The latter are taken here to be memoryless non-linearities followed by linear time invariant dynamics. Hammerstein, Wiener, Hammerstein–Wiener and Wiener–Hammerstein models are all special cases of this structure. The parameter estimation of this model is investigated by using a standard prediction error criterion coupled with a robust gradient based search algorithm. This approach is profiled using the "silverbox" Wiener–Hammerstein system benchmark data, which illustrates it to be effective and, via Monte–Carlo simulation, relatively robust against capture in local minima.

Keywords: gradient-based search, output-error, Hammerstein, Wiener, black-box

1. INTRODUCTION

This paper considers the estimation of a class of non-linear models which are generalisations of Hammerstein–Wiener systems in that they allow for an arbitrary number of concatenations of linear dynamic sub-systems with memoryless non-linear sub-systems.

The motivation for this work is to address a benchmark non-linear system identification problem proposed in Schoukens et al. [2008]. This benchmark completely describes the underlying true physical system responsible for the provided experimental observations. This reveals that it is a Wiener–Hammerstein system comprising two third order linear systems sandwiching a memoryless non-linearity. It is dubbed "the silverbox" system in Schoukens et al. [2008], and has been the subject of previous benchmarking studies Session [2004].

At the same time, a stated purpose of the benchmark is "*to compare different black box identification methods to model non-linear systems" Schoukens et al. [2008].* This paper therefore considers the use of a black box non-linear model structure that encompasses the disclosed true system as a special case. More specifically, a structure is proposed which comprises an arbitrary number of concatenations of a Hammerstein model comprising a piece-wise linear memoryless non-linearity followed by linear dynamics modeled as an arbitrary order discrete time rational transfer function. Since this structure can incorporate all of the possibilities of Hammerstein and Wiener type models and combinations, it is called here a "generalised Hammerstein–Wiener" structure.

⋆ This work was supported by the Australian Research Council.

Due to their wide relevance, there has been a very large amount of work on the estimation of Hammerstein–Wiener type model structures, of which Bai [1998a], Goethals et al. [2005], Zhu [2002], Bai [1998b], Boutayeb and Darouach [1995], Vörös [2007], Lovera et al. [2000] represents a small selection of recent contributions. This paper investigates a simple method involving a standard prediction error framework coupled with a gradient-based search. Despite this being what could be considered a very straightforward approach, to the authors knowledge, it does not appear to be common practice.

This is possibly due to the fear that the non-convexity of the associated prediction error criterion implies that the method is susceptible to the gradient based search being attracted to local minima. This paper investigates this issue by considering a wide range of initial points for the gradient based search. This illustrates, via the silverbox example, that with careful design of the gradient based search algorithm the prediction error method presented here can provide an effective estimation approach.

2. MODEL STRUCTURE

As described in the introduction, this paper considers nonlinear model structures which are formed as concatenations of a Hammerstein "building block". For a given scalar input sequence $\{u_t\}$ and corresponding output sequence $\{y_t\}$ the k'th Hammerstein component is represented as an operator $\mathcal{T}_k(\boldsymbol{\theta}_k)$ on u_t

$$y_t = \mathcal{T}_k(\boldsymbol{\theta}_k)u_t \tag{1}$$

which is parametrized by a vector $\boldsymbol{\theta}_k = [\boldsymbol{\beta}_k, \boldsymbol{\lambda}_k]$ of real valued parameters, and defined via

$$y_t = G_k(q, \boldsymbol{\beta}_k)x_t^k, \qquad x_t^k = f_k(u_t, \boldsymbol{\lambda}_k). \tag{2}$$

This is graphically depicted in Figure 1.

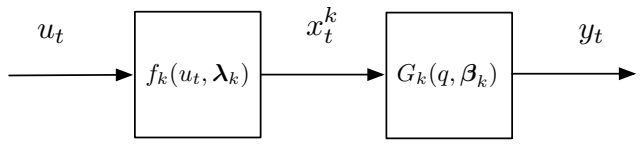

Fig. 1. *Hammerstein "building block".*

Here, the linear dynamics $G_k(q, \boldsymbol{\beta}_k)$ are represented by rational transfer function that is parametrized for given numerator and denominator orders m_{b_k},m_{a_k} by a vector $\boldsymbol{\beta}_k \in \mathbf{R}^{m_{a_k}+m_{b_k}+1}$ according to

$$G_k(q, \boldsymbol{\beta}_k) = \frac{b_0^k + b_1^k q^{-1} \dots b_{m_{b_k}}^k q^{-m_{b_k}}}{1 + a_1^k q^{-1} \dots a_{m_{a_k}}^k q^{-m_{a_k}}} \tag{3}$$

$$\boldsymbol{\beta}_k = \left[b_0^k, \dots, b_{m_{b_k}}^k, a_1^k, \dots, a_{m_{a_k}}^k\right]. \tag{4}$$

The function $f_k(\cdot, \boldsymbol{\lambda}_k)$ is a memoryless non-linear mapping, that in this paper is taken to be a piecewise linear function with an arbitrary number ℓ_k of transitions between linear sub-components. It is parametrized by a vector $\boldsymbol{\lambda}_k \in \mathbf{R}^{2(\ell_k+1)}$ that specifies a linear base together with ℓ_k "hinge" functions $h_j^k(\cdot, \boldsymbol{\lambda}_k)$ Breiman [1993]:

$$f_k(u_t, \boldsymbol{\lambda}_k) = \lambda_1^k + \lambda_2^k u_t + \sum_{j=1}^{\ell_k} h_j^k(u_t, \boldsymbol{\lambda}_k) \tag{5}$$

$$h_j^k(u_t, \boldsymbol{\lambda}_k) = \begin{cases} \lambda_{j+2}^k + \lambda_{j+3}^k u_t \ ; & u_t > -\dfrac{\lambda_{j+2}^k}{\lambda_{j+3}^k}, \\ \\ 0 \qquad ; & \text{Otherwise} \end{cases} \tag{6}$$

(7)

$$\boldsymbol{\lambda}_k = [\lambda_1^k, \lambda_2^k, \cdots, \lambda_{\ell+2}^k]. \tag{8}$$

The concatenation of m of these Hammerstein building blocks then delivers the non-linear model structure employed in this paper of

$$y_t = \left[\prod_{k=1}^{m} \mathcal{T}_k(\boldsymbol{\theta}_k)\right] u_t + \nu_t \tag{9}$$

and illustrated in Figure 2.

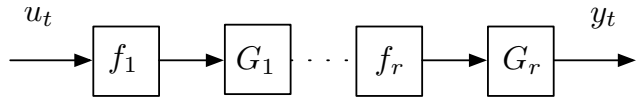

Fig. 2. *Generalised Hammerstein–Wiener model structure employed in this paper.*

Here, $\{\nu_t\}$ is a zero mean i.i.d. stochastic process representing measurement corruptions. This complete model structure is then parametrized by a vector $\boldsymbol{\theta}$ consisting of the sub-vectors $\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_m$ parametrizing the constituent Hammerstein building blocks

$$\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_m]. \tag{10}$$

Several common non-linear model structures are encompassed as specialisations of (9). Clearly, with $m = 1$ a Hammerstein model is realised. With $m = 2$ and $G_2(q, \beta_2) = 1$ (i.e. $m_{a_2} = m_{b_2} = 0$, $b_0^2 = 1$) a Hammerstein–Wiener model structure eventuates. If this is further restricted so that $f_1(u_t, \boldsymbol{\lambda}_1) = u_t$ (i.e. $\lambda_1 = 0, \lambda_1^1 = 0, \lambda_2^1 = 1$) then a Wiener model is obtained.

Because of this, the model (9), (10) is dubbed a "generalised Hammerstein Wiener" structure in this paper.

3. ESTIMATION METHOD

This paper employs a standard prediction error approach for computing an estimate $\widehat{\boldsymbol{\theta}}_N$ of the parameters $\boldsymbol{\theta}$ specifying the generalised Hammerstein–Wiener structure (9), (10) on the basis of N observations of input $\{u_t\}$ and response $\{y_t\}$. Specifically

$$\widehat{\boldsymbol{\theta}}_N = \arg\min_{\theta} V_N(\theta) \tag{11}$$

where

$$V_N(\theta) = \sum_{t=1}^{N} \varepsilon_t^2(\boldsymbol{\theta}), \qquad \varepsilon_t(\boldsymbol{\theta}) \triangleq y_t - \widehat{y}_{t|t-1}(\theta). \tag{12}$$

Here, $\widehat{y}_{t|t-1}(\boldsymbol{\theta})$ is the mean square optimal one step ahead predictor of y_t based on past observations and the model structure (9). Since the corruptions $\{\nu_t\}$ are an independent process, this is given by

$$\widehat{y}_{t|t-1}(\boldsymbol{\theta}) = \left[\prod_{k=1}^{m} \mathcal{T}_k(\boldsymbol{\theta}_k)\right] u_t. \tag{13}$$

4. ESTIMATION ALGORITHM

The formulation (11)-(13) specifies the estimate $\widehat{\boldsymbol{\theta}}_N$ as the solution of a non-convex and non-linearly parametrized optimisation problem. As such, a closed form solution is not possible, and this paper explores the use of an iterative gradient-based search for the computation of $\widehat{\boldsymbol{\theta}}_N$.

This is based on a standard Gauss–Newton search method, but with adaptive modifications to the rank of the Jacobian that is employed in computing the search direction in order to enhance robustness Wills and Ninness [2008]. The reader is referred to Wills and Ninness [2008] for full details and analysis of this method, but in the interests of a self contained presentation the essentials of the approach are presented here, which depends on a defining a vector $\mathbf{e}(\boldsymbol{\theta})$ of prediction error residuals $\{\varepsilon_t(\boldsymbol{\theta})\}$ according to

$$\mathbf{e}(\boldsymbol{\theta}) \triangleq \begin{bmatrix} \varepsilon_1(\boldsymbol{\theta}) \\ \vdots \\ \varepsilon_N(\boldsymbol{\theta}) \end{bmatrix} \tag{14}$$

and then forming a local linear approximation

$$\mathbf{e}(\boldsymbol{\theta} + \mathbf{p}) \approx \mathbf{e}(\boldsymbol{\theta}) + \mathbf{J}(\boldsymbol{\theta})\mathbf{p} \tag{15}$$

using the Jacobian $\mathbf{J}(\boldsymbol{\theta}) \in \mathbf{R}^{N \times m}$ of $\mathbf{e}(\boldsymbol{\theta})$ defined as

$$\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} \dfrac{\partial e_1(\boldsymbol{\theta})}{\partial \theta_1} & \cdots & \dfrac{\partial e_1(\boldsymbol{\theta})}{\partial \theta_m} \\ \vdots & & \vdots \\ \dfrac{\partial e_N(\boldsymbol{\theta})}{\partial \theta_1} & \cdots & \dfrac{\partial e_N(\boldsymbol{\theta})}{\partial \theta_m} \end{bmatrix}. \tag{16}$$

Therefore, locally around $\boldsymbol{\theta}$

$$\begin{aligned} V_N(\boldsymbol{\theta} + \mathbf{p}) &= \mathbf{e}^T(\boldsymbol{\theta} + \mathbf{p})\mathbf{e}(\boldsymbol{\theta} + \mathbf{p}) \\ &\approx \mathbf{e}^T(\boldsymbol{\theta})\mathbf{e}(\boldsymbol{\theta}) + 2\mathbf{e}^T(\boldsymbol{\theta})\mathbf{J}(\boldsymbol{\theta})\mathbf{p} + \mathbf{p}^T\mathbf{J}^T(\boldsymbol{\theta})\mathbf{J}(\boldsymbol{\theta})\mathbf{p}. \end{aligned}$$

The minimiser of this approximate cost satisfies

$$\mathbf{J}^T(\boldsymbol{\theta})\mathbf{J}(\boldsymbol{\theta})\mathbf{p} = \mathbf{J}^T(\boldsymbol{\theta})\mathbf{e}(\boldsymbol{\theta}). \tag{17}$$

However, there is no guarantee that such a $\mathbf{p}$, resulting from a local approximation of $V_N(\boldsymbol{\theta})$ will actually reduce the cost function in that $V_N(\boldsymbol{\theta}+\mathbf{p}) < V_N(\boldsymbol{\theta})$. As a result, $\mathbf{p}$ is treated as a search direction along which a point $\boldsymbol{\theta}+\alpha\mathbf{p}$ where $\alpha \in \mathbf{R}$ is sought such that $V_N(\boldsymbol{\theta}+\alpha\mathbf{p}) < V_N(\boldsymbol{\theta})$. Furthermore, there is no guarantee that the Jacobian $\mathbf{J}(\boldsymbol{\theta})$ has full column rank, so that $\mathbf{p}$ may not be uniquely defined by (17).

This paper employs the following algorithm developed in Wills and Ninness [2008] that addresses these issues to provide an iterative search method for computing a parameter estimate $\widehat{\boldsymbol{\theta}}_N$.

Algorithm 4.1. **Gauss–Newton based search**

Given an initial guess $\boldsymbol{\theta}_0 \in \mathbf{R}^m$, initialise parameters

$$\eta \in (0, 1/2), \quad \alpha_{\min} > 0, \quad \gamma \in (0, 1], \quad k = 0, \quad \epsilon > 0$$

Compute descent direction:

(1) Compute the prediction error vector $\mathbf{e}(\boldsymbol{\theta}_k)$, its Jacobian $\mathbf{J}(\boldsymbol{\theta}_k)$ and the gradient
$$\mathbf{g}_k = \mathbf{J}^T(\boldsymbol{\theta}_k)\mathbf{e}(\boldsymbol{\theta}_k);$$
(2) Compute the Singular Value Decomposition
$$\mathbf{J}(\boldsymbol{\theta}_k) = \mathbf{U}\mathbf{S}\mathbf{V}^T; \tag{18}$$
(3) Let $\{s_1, \ldots, s_m\}$ be the ordered singular values from $\mathbf{S}$ such that $s_i \geq s_{i+1}$;
(4) Perform the following:
  (a) Find the index r of the smallest singular value that satisfies $s_r \geq \gamma s_1$; let $\mathbf{U}_r$, $\mathbf{V}_r$ be the first r columns of $\mathbf{U}$ and $\mathbf{V}$ respectively and set $\mathbf{S}_r$ as a diagonal matrix
  $$\mathbf{S}_r = \mathrm{diag}\{s_1, \ldots, s_r\};$$
  (b) Compute a search direction $\mathbf{p}_k$ as
  $$\mathbf{p}_k = -\mathbf{V}_r\mathbf{S}_r^{-1}\mathbf{U}_r^T\mathbf{e}(\boldsymbol{\theta}_k); \tag{19}$$
  (c) If
  $$-\mathbf{p}_k^T\mathbf{g}_k \geq \eta\,||\mathbf{p}_k||_2||\mathbf{g}_k||_2, \tag{20}$$
  then goto Step 6;
  (d) Let $\gamma \mapsto 0.25\gamma$ and goto Step 4a;

Compute step length:

(6) Initialise the step length $\alpha = 1$ and perform the following

While $V_N(\boldsymbol{\theta}_k+\alpha\mathbf{p}_k) \geq V_N(\boldsymbol{\theta}_k)+\alpha\eta\,\mathbf{p}_k^T\mathbf{g}_k$ then update
$$\alpha \mapsto 0.5\alpha;$$
(7) If $\alpha = 1$ then update
$$\gamma \mapsto 0.25\gamma;$$
(8) If $\alpha \leq \alpha_{\min}$ then update $\gamma \mapsto \min\{1, 2\gamma\}$;
(9) Set $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \alpha\,\mathbf{p}_k$;
(10) Check termination conditions:

If
$$\mathbf{g}_k^T\left[\mathbf{J}^T(\boldsymbol{\theta}_k)\mathbf{J}(\boldsymbol{\theta}_k)\right]^{-1}\mathbf{g}_k \leq \epsilon$$
then stop.

Otherwise let $k \mapsto k+1$ and goto Step 1.

□

5. JACOBIAN CALCULATION

The estimation method examined here of using the standard prediction error method (11),(12) coupled with the gradient based search defined in Algorithm 4.1 is considered by the authors to be a quite straightforward approach.

The main obstacle in implementing the technique is the computation of the Jacobian $\mathbf{J}(\boldsymbol{\theta})$. To address this, denote the output of the j'th Hammerstein subsystem in the model structure (9) as $z_t^j(\boldsymbol{\theta})$; viz.

$$z_t^j(\boldsymbol{\theta}) = \left[\prod_{k=1}^{j}\mathcal{T}_k(\theta_k)\right]u_t \tag{21}$$

and suppose that the scalar element θ_i of the full parameter vector $\boldsymbol{\theta}$ falls within the j'th sub-block $\boldsymbol{\theta}_j$ defining the j'th Hammerstein-block. Then

$$\frac{\partial\mathbf{e}_t(\boldsymbol{\theta})}{\partial\theta_i} = -\frac{\partial}{\partial\theta_i}\widehat{y}_{t|t-1}(\boldsymbol{\theta}) \tag{22}$$

$$= -\left[\prod_{k=j+1}^{m}\mathcal{T}_k(\boldsymbol{\theta}_k)\right]\frac{\partial z_t^j(\boldsymbol{\theta})}{\partial\theta_i}. \tag{23}$$

Furthermore

$$z_t^j(\boldsymbol{\theta}) = G_j(q,\boldsymbol{\beta}_j)x_t^j(\boldsymbol{\lambda}_j), \qquad x_t^j(\boldsymbol{\lambda}_j) = f_j(z_t^{j-1},\boldsymbol{\lambda}_j) \tag{24}$$

where $z_t^0 = u_t$. Therefore, the required gradient depends on whether the component θ_i falls within the $\boldsymbol{\beta}_j$ sub-block parametrizing linear dynamics, or within the $\boldsymbol{\lambda}_j$ sub-block parametrizing the piece-wise linear memoryless non-linearity. The results in each case are now catalogued.

Suppose first that $\theta_i = b_r^j$, the co-efficient of q^{-r} in the numerator of the linear dynamics $G_j(q,\boldsymbol{\beta}_j)$. Then

$$\frac{\partial z_t^j(\boldsymbol{\theta})}{\partial\theta_i} = \frac{\partial}{\partial b_r^j}G(q,\boldsymbol{\beta}_j)x_t^j(\boldsymbol{\lambda}_j) \tag{25}$$

$$= \frac{q^{-r}}{1 + a_1^j q^{-1} \ldots a_{m_{a_k}}^j q^{-m_{a_k}}}x_t^j(\boldsymbol{\lambda}_j). \tag{26}$$

Suppose next that $\theta_i = a_r^j$, the co-efficient of q^{-r} in the denominator of the linear dynamics $G_j(q,\boldsymbol{\beta}_j)$. Then

$$\frac{\partial z_t^j(\boldsymbol{\theta})}{\partial\theta_i} = \frac{\partial}{\partial a_r^j}G(q,\boldsymbol{\beta}_j)x_t^j(\boldsymbol{\lambda}_j) \tag{27}$$

$$= -G(q,\boldsymbol{\beta}_j)\frac{q^{-r}}{1 + a_1^j q^{-1} \ldots a_{m_{a_k}}^j q^{-m_{a_k}}}x_t^j(\boldsymbol{\lambda}_j). \tag{28}$$

Finally, suppose that $\theta_i = \lambda_r^j$, the r'th co-efficient parametrizing the piece-wise linear memoryless non-linearity if the j'th Hammerstein component. Then

$$\frac{\partial z_t^j(\boldsymbol{\theta})}{\partial\theta_i} = G(q,\boldsymbol{\beta}_j)\frac{\partial}{\partial\lambda_r^j}x_t^j(\boldsymbol{\lambda}_j) = G(q,\boldsymbol{\beta}_j)\frac{\partial}{\partial\lambda_r^j}f_j(z_t^{j-1},\boldsymbol{\lambda}_j) \tag{29}$$

where for $r \leq 2$

$$\frac{\partial f_j(z,\boldsymbol{\lambda}_j)}{\partial\lambda_1^j} = 1, \qquad \frac{\partial f_j(z,\boldsymbol{\lambda}_j)}{\partial\lambda_2^j} = z \tag{30}$$

and for $r > 2$, with $s = r \bmod 2$

$$\frac{\partial f_j(z,\boldsymbol{\lambda}_j)}{\partial \lambda_r^j} = \frac{\partial h_{\lfloor (r-1)/2 \rfloor}(z,\boldsymbol{\lambda}_j)}{\partial \lambda_r^j} \tag{31}$$

$$= \begin{cases} 1\ ; & s=1 \text{ and } z > -\dfrac{\lambda_r^j}{\lambda_{r+1}^j}, \\ z\ ; & s=0 \text{ and } z > -\dfrac{\lambda_r^j}{\lambda_{r+1}^j}, \\ 0\ ; & z \le -\dfrac{\lambda_r^j}{\lambda_{r+1}^j}. \end{cases} \tag{32}$$

Here $\lfloor x \rfloor$ represents the "floor" operation returning the biggest integer no greater than x.

6. RESULTS

As explained in the introduction, the motivation for the model structure, estimation method and algorithm presented in sections 2-5 was to attempt the "Wiener–Hammerstein Benchmark" proposed in Schoukens et al. [2008]. This is based on data produced by a physical apparatus dubbed the "silverbox system" Schoukens et al. [2008]. It consists of two third order Chebychev filters (i.e. third order linear systems) which sandwich a memoryless non-linearity formed by a resistive divider with a diode across the output resistor.

It is therefore a "Wiener–Hammerstein" system which can be encompassed by employing $m = 2$ Hammerstein blocks in our more general structure (9). Using this, the estimation method and algorithm presented in sections 2-5 were employed on the first $N = 100000$ samples of input $\{u_t\}$ and output $\{y_t\}$ provided by the benchmark authors Schoukens et al. [2008].

The iterative search algorithm was initialised with the memoryless non-linearities being initialised as in fact linear with $\ell_1 = \ell_2 = 8$ possible "breakpoints":

$$\lambda_1^j = 0, \quad \lambda_2^j = 1, \quad \lambda_r^j = 0, r \in [2,8] \quad ; j = 1,2. \tag{33}$$

Furthermore, the linear blocks G_1 and G_2 were initialised as

$$G_1(q,\boldsymbol{\beta}_1) = G_2(q,\boldsymbol{\beta}_2) = \frac{(1-0.5q^{-1})^3}{(1-0.5q^{-1})^3} \tag{34}$$

so that the entire model is initialised as simply being a unit gain. Estimation then required 257 iterations from this initialisation, which consumed 10 minutes of computing time on a Mac 3GHz Intel Xeon.

The resulting estimates are profiled in Figures 3-6 as the solid lines. Also shown as dashed lines are the underlying "true" responses as disclosed in Schoukens et al. [2008].

In presenting these comparisons, it is important to note that the model structure we are employing is not system identifiable. Given an estimate, one of identical cost may be obtained by (for example) decreasing the gain of the dynamics G_1 and increasing the gain of the first memoryless non-linearity f_1 by an equal amount. In the figures following, in order to allow comparisons with the underlying system, the estimates of G_1 and G_2 have been normalised to have unit gain, with the associated memoryless blocks f_1 and f_2 then having their gain then altered by the same amount in the reverse direction.

With this in mind, Figures 3 and 4 show the bode magnitude frequency response of the estimates $G_1(e^{j\omega}, \widehat{\boldsymbol{\theta}}_N)$, $G_2(e^{j\omega}, \widehat{\boldsymbol{\theta}}_N)$ relative to the third order Chebychev and inverse Chebychev filters that are disclosed as the true linear dynamics in Schoukens et al. [2008]. The correspondence is considered quite close, and it interesting that the transmission zero in G_2 appears to have been accurately estimated.

The input-maps of the memoryless non-linearity estimates $f_1(\cdot, \widehat{\boldsymbol{\theta}}_N)$, $f_2(\cdot, \widehat{\boldsymbol{\theta}}_N)$ that precede the above mentioned linear estimates are shown in Figures 5, 6. In presenting the true non-linearity map in Figure 6, we have accounted for the IN4148 diode disclosed in Schoukens et al. [2008] by noting that from its data-sheet, at 25°C it has a voltage-current relationship that is approximately linear in the forward biased 0.1mA-100mA range of

$$\log_{10} i = 9.38v - 8.69. \tag{35}$$

As such, when included in the voltage divider circuit specified in Schoukens et al. [2008], the achieved input-voltage v_i to output voltage v_0 (assuming low impedance output for G_1 and high impedance input for G_2) satisfies

$$v_i = \frac{11}{10} v_0 + 10^{-8.69} \cdot 10^{9.38 v_0}. \tag{36}$$

This was solved numerically for v_0 for a range of given v_i in order to provide the dashed line "true system" plot shown in Figure 6.

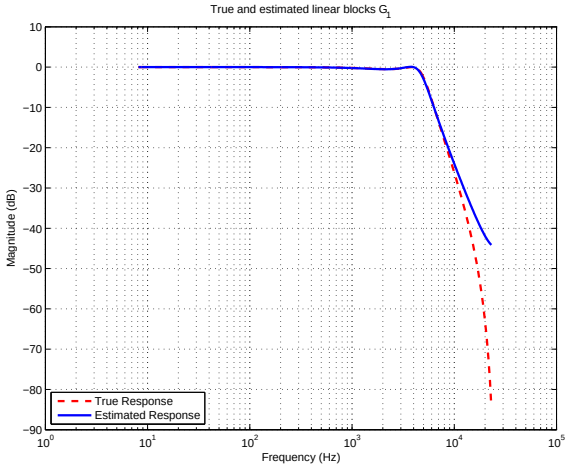

Fig. 3. *Bode magnitude frequency responses of true (dashed line) and estimated (solid line) linear dynamic block* G_1.

As for the linear blocks, these input-output maps illustrate what is considered accurate estimates of the underlying system.

To more comprehensively quantify the quality of these estimates, the benchmark proposers Schoukens et al. [2008] have defined four performance measures. The first two are (respectively), the mean and standard deviation of the simulation error (in our case, this is the same as the prediction error) over a subset of the validation data (i.e. data that was not part of the first $N = 10000$ samples used to derive the estimate $\widehat{\boldsymbol{\theta}}_N$:

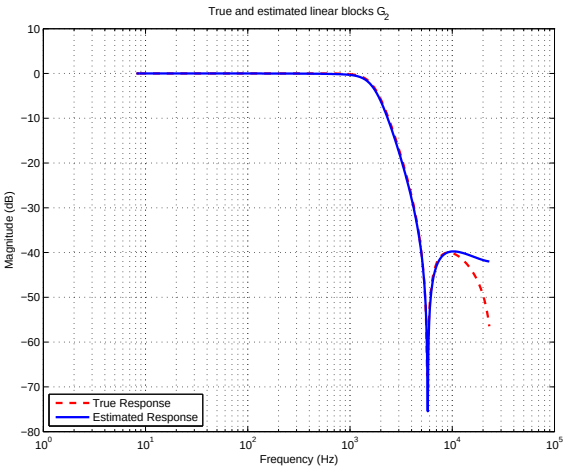

Fig. 4. *Bode magnitude frequency responses of true (dashed line) and estimated (solid line) linear dynamic block* G_2.

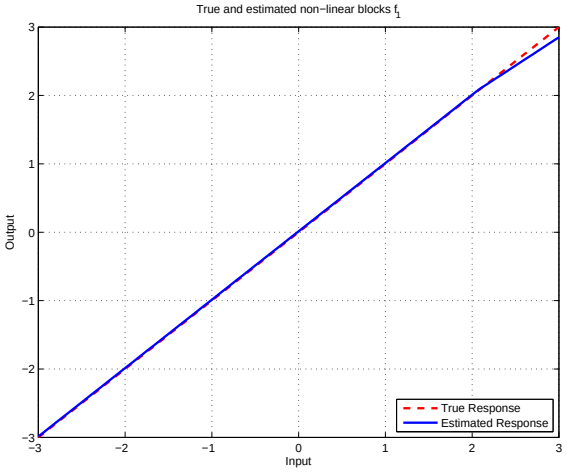

Fig. 5. *Input-output maps for true (dashed line) and estimated (solid line) memoryless non-linearity block* f_1.

$$\mu \triangleq \frac{1}{87000} \sum_{t=101001}^{188000} \varepsilon_t(\widehat{\boldsymbol{\theta}}_N), \tag{37}$$

$$s \triangleq \sqrt{\frac{1}{86999} \sum_{t=101001}^{188000} [\varepsilon_t(\widehat{\boldsymbol{\theta}}_N) - \mu]^2} \tag{38}$$

The second two proposed measures are the root mean square (RMS) prediction error on subsets (respectively) of the validation data, and the data used to actually derive the estimate $\widehat{\boldsymbol{\theta}}_N$:

$$e_{\mathrm{RMSt}} \triangleq \sqrt{\frac{1}{87000} \sum_{t=101001}^{188000} \varepsilon_t^2(\widehat{\boldsymbol{\theta}}_N)}, \tag{39}$$

$$e_{\mathrm{RMSe}} \triangleq \sqrt{\frac{1}{187000} \sum_{t=1001}^{100000} \varepsilon_t^2(\widehat{\boldsymbol{\theta}}_N)}. \tag{40}$$

These measures, for the estimate model profiled in Figures 3-6 are given in Table 1 following.

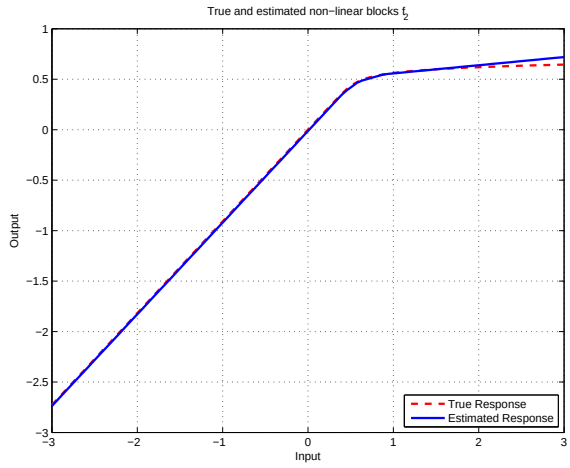

Fig. 6. *Input-output maps for true (dashed line) and estimated (solid line) memoryless non-linearity block* f_2.

Data set	μ	**s**	$\mathbf{e}_{\mathrm{RMS}}$
Estimation	-2.6735e-08	5.1230e-04	4.8846e-04
Validation	-2.1977e-05	4.8717e-04	4.8766e-04

Table 1. *Performance Measures.*

The benchmark specification Schoukens et al. [2008] also requests that the error between the observed response and that of the identified model be presented, both in the time and frequency domain. This is addressed in the time domain case in Figure 7 by plotting both the observed and model simulated response for the first 20ms of validation data. The Blackman–Tukey spectral density estimates of these same signals over the full 1.7 second long validation set are then provided in Figure 8.

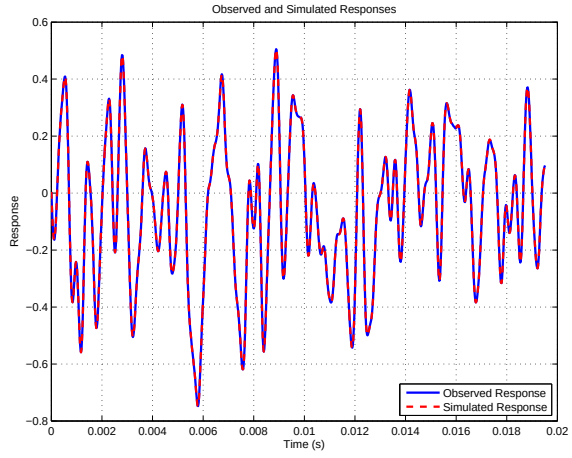

Fig. 7. *Observed and simulated response on first 20ms of validation data.*

Finally, the authors believe it is natural to question the robustness of the methods studied here to the search initialisation and the associated problem of possible capture in local minima.

To study this, the all-pass initialisation for the linear dynamics G_1, G_2 specified in (34) was replaced by a randomly chosen third order system using the MATLAB `poly(rand(1,3))` to specify inital numerator and denom-

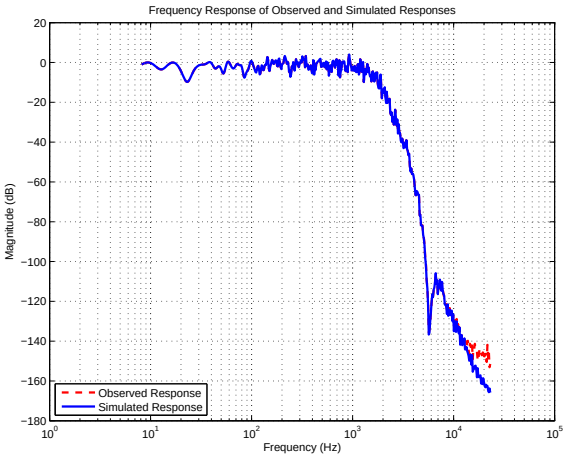

Fig. 8. Frequency response of measured and predicted outputs

inator polynomials (separately). This was done for 100 initial system realisations. The terminal costs of the gradient based search initialised at these 100 different points are plotted in Figure 9. Clearly, while there is some consistency in achieving a terminal cost near $10^{-6.8}$, there is a susceptibility to termination in local minima.

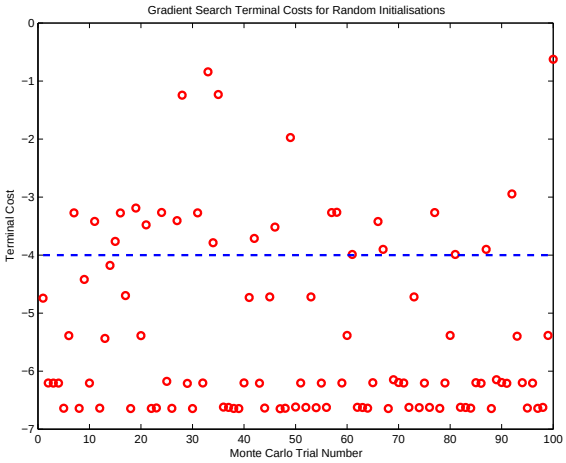

Fig. 9. *Gradient search terminal costs for 100 randomly chosen G_1 and G_2 initialisations.*

However, to gauge the importance of not achieving the global minimum, the estimated frequency responses $G_2(e^{j\omega}, \widehat{\boldsymbol{\theta}}_N)$ for all achieved costs less than 10^{-4} (indicated by the horizontal dashed line in Figure 9) are plotted relative to the true response in Figure 10. Similarly, the estimated input-output maps $f_2(\cdot, \widehat{\boldsymbol{\theta}})$ for these same achieved costs are plotted relative to the true input-output map in Figure 11. Clearly, while not achieving the global minimum can have a significant effect, the resulting model can still be considered informative of the general qualities of the underlying true system.

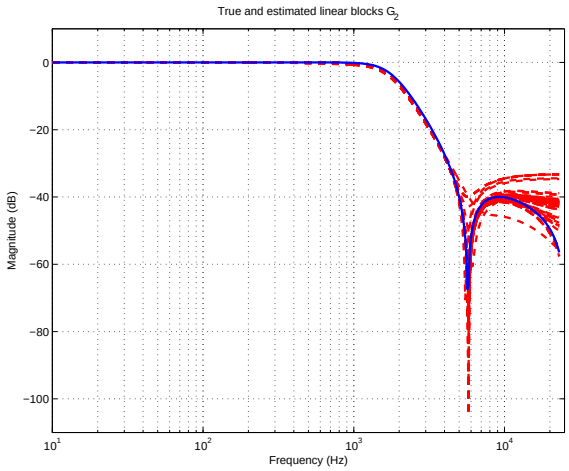

Fig. 10. *Estimated frequency responses $G_2(e^{j\omega}, \widehat{\boldsymbol{\theta}}_N)$ for all Monte–Carlo realisations achieving a terminal cost of 10^{-4} or less.*

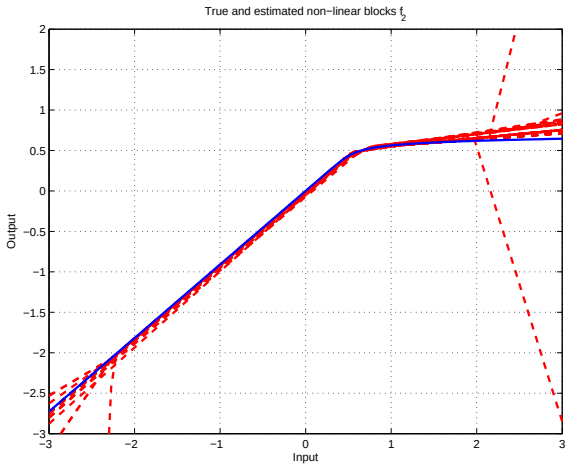

Fig. 11. *Estimated frequency responses $G_2(e^{j\omega}, \widehat{\boldsymbol{\theta}}_N)$ for all Monte–Carlo realisations achieving a terminal cost of 10^{-4} or less.*

REFERENCES

Er-Wei Bai. An optimal two stage identification algorithm for hammerstein–wiener nonlinear systems. *Automatica*, 34(3):333–338, 1998a.

Er-Wei Bai. An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems. *Automatica J. IFAC*, 34(3):333–338, 1998b. ISSN 0005-1098.

M. Boutayeb and M. Darouach. Recursive identification method for MISO Wiener-Hammerstein model. *IEEE Trans. Automat. Control*, 40(2):287–291, 1995. ISSN 0018-9286.

Leo Breiman. Hinging hyperplanes for regression, classification, and function approximation. *IEEE Trans. Inform. Theory*, 39(3):999–1013, 1993. ISSN 0018-9448.

I. Goethals, K. Pelckmans, J. A. K. Suykens, and B. De Moor. Supspace Identification of Hammerstein Systems Using Least Squares Support Vector Machines. *IEEE Transactions on Automatic Control*, 50(10):pp1509–1519, 2005.

Marco Lovera, Tony Gustafsson, and Michel Verhaegen. Recursive subspace identification of linear and non-linear Wiener state-space models. *Automatica J. IFAC*, 36(11):1639–1650, 2000. ISSN 0005-1098.

J. Schoukens, J. Suykens, and L. Ljung. Wiener–hammerstein benchmark (sysid 2009 special session). `wwwtw.vub.ac.be/elec/sysid09.htm`, 2008.

NOLCOS Special Session. Identification of nonlinear systems:the silverbox case study. In *Proceedings of the IFAC Symposium on Nonlinear Control Systems*, 2004.

Jozef Vörös. Parameter identification of Wiener systems with multisegment piecewise-linear nonlinearities. *Systems Control Lett.*, 56(2):99–105, 2007. ISSN 0167-6911.

Adrian Wills and Brett Ninness. On gradient-based search for multivariable system estimates. *IEEE Trans. Automat. Control*, 53(1):298–306, 2008. ISSN 0018-9286.

Yucai Zhu. Estimation of an N-L-N Hammerstein-Wiener model. *Automatica J. IFAC*, 38(9):1607–1614, 2002. ISSN 0005-1098.