Generalised Hammerstein-Wiener System Estimation and a Benchmark Application[☆]

Adrian Wills[a], Brett Ninness[a]

[a]*School of Electrical Engineering and Computer Science, University of Newcastle, Callaghan, NSW, 2308, Australia (Tel: +61 2 49216028; e-mail: {adrian.wills, 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 a Wiener–Hammerstein benchmark example, which illustrates it to be effective and, via Monte–Carlo simulation, relatively robust against capture in local minima.

Keywords: gradient-based search, prediction error method, Hammerstein, Wiener, block oriented

1. Introduction

The benchmark system identification problem formulated in (Schoukens et al., 2008) is based on data collected from a physical apparatus which is a diode-dependent non-linear circuit sandwiched by two linear ones.

This formed the basis for two special sessions at the 15th IFAC Symposium on System identification (Schoukens et al., 2008) where a wide range of techniques for estimating this mixed linear/non-linear system was presented. These include non-parametric approaches (Pillonetto and Chiuso,

[☆]This work was supported by the Australian Research Council.

2009; Truong and Wang, 2009), the use of support vector machines (Marconato and Schoukens, 2009; Falck et al., 2009; Brabanter et al., 2009) and subspace-based methods (dos Santos et al., 2009; Ase et al., 2009) amongst others (Piroddi et al., 2009; Paduart et al., 2009; Mulders et al., 2009; Ase et al., 2009; Lauwers et al., 2009), including the paper (Wills and B.Ninness, 2009) by the current authors.

The paper here is an expansion of the work (Wills and B.Ninness, 2009) where a prediction error approach (Ljung, 1999) is examined. Associated with this, the employed model structure is a block-oriented one consisting of the concatenation of an arbitrary number of Hammerstein structures, with the latter defined here as a memoryless non-linear system preceding a linear transfer function. This is termed a "generalised Hammerstein–Wiener system" structure since it can incorporate many of the possibilities of Hammerstein and Wiener type models and combinations.

Relative to (Wills and B.Ninness, 2009), this paper is more general in its consideration of the non-linear block, allows for correlated measurement noise, and provides an expanded study of the benchmark study. In particular, the paper provides details of commonly used nonlinearities and have expanded the simulation study to include two other commonly used nonlinearities.

There is a already an extensive body of work related to the estimation of Hammerstein–Wiener type model structures, including the recent contributions (Bai, 1998a; Goethals et al., 2005; Zhu, 2002; Bai, 1998b; Boutayeb and Darouach, 1995; Vörös, 2007; Lovera et al., 2000). The current paper investigates a method involving a standard prediction error framework coupled with a gradient-based search.

There are two main factors motivating this approach. First is the availability of a very rich body of supporting theory (Ljung, 1999; L.Ljung and P.E.Caines, 1979) that establishes quantifiable statistical properties of these prediction-error (PE) estimates. Second, is the attraction of scientific consistency in which a standard and proven PE framework is tested before alternative options are considered.

This lack of prior PE study is possibly due to the concern 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 application to the benchmark problem (Schoukens et al., 2008), that with careful design of the gra-

dient 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". More specifically, 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 according to

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

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 depicted in Figure 1.

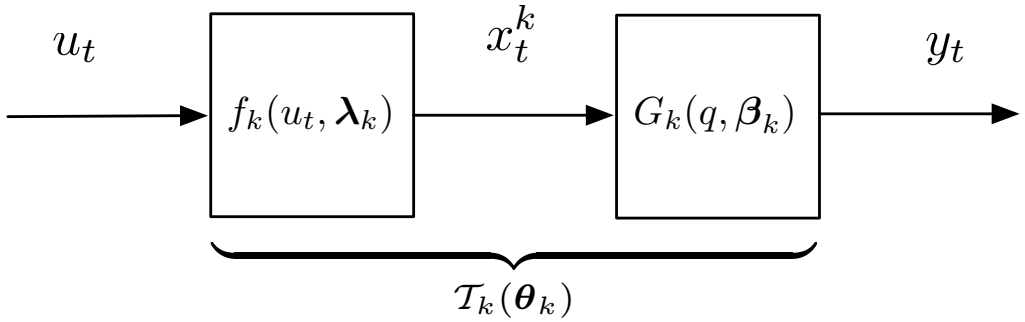

Figure 1: *Hammerstein "building block".*

Here, the linear dynamics $G_k(q, \boldsymbol{\beta}_k)$ are represented by a 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

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

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

The function $f_k(\cdot, \boldsymbol{\lambda}_k)$ is a memoryless non-linear mapping, for which the only restriction placed here is that the derivatives

$$\frac{\partial}{\partial \lambda_k^i} f_k(u_t, \boldsymbol{\lambda}_k) \tag{5}$$

exist for all elements λ_k^i of the vector $\boldsymbol{\lambda}_k$ which parametrizes f_k. Some specific examples will be detailed later in section 6

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

$$y_t = \left[\prod_{k=1}^{m} \mathcal{T}_k(\boldsymbol{\theta}_k)\right] u_t + H(q, \boldsymbol{\gamma}) e_t \tag{6}$$

illustrated in Figure 2.

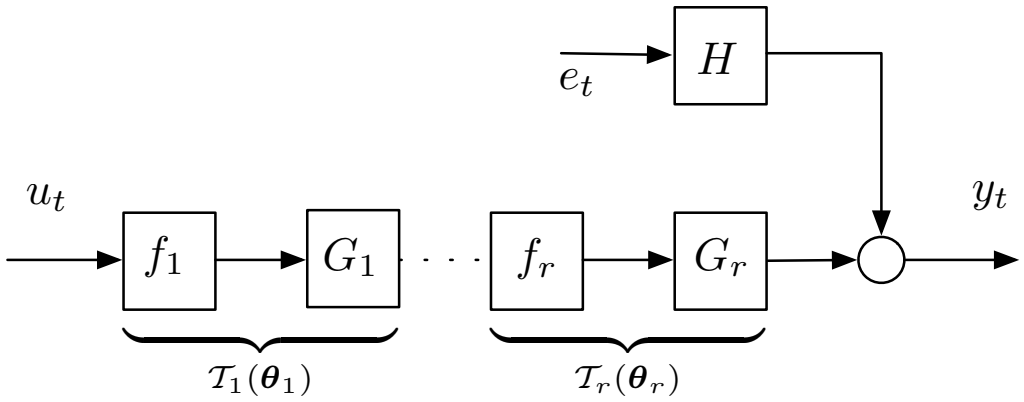

Figure 2: *Generalised Hammerstein–Wiener model structure employed in this paper.*

Here, $\{e_t\}$ is a zero mean i.i.d. stochastic process representing measurement corruptions that are possibly correlated according to the rational noise model

$$H(q, \boldsymbol{\gamma}) = \frac{1 + c_1 q^{-1} + \cdots, c_{m_c} q^{-m_c}}{1 + d_1 q^{-1} + \cdots, d_{m_d} q^{-m_d}} = \frac{C(q, \boldsymbol{\gamma})}{D(q, \boldsymbol{\gamma})} \tag{7}$$

$$\boldsymbol{\gamma} \triangleq [c_1, \cdots, c_{m_c}, d_1, \cdots, d_{m_d}]. \tag{8}$$

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

$$\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T, \cdots, \boldsymbol{\theta}_m^T, \boldsymbol{\gamma}^T]^T. \tag{9}$$

Several common non-linear model structures are encompassed as specialisations of (6). 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 (6), (9) is dubbed a "generalised Hammerstein Wiener" structure in this paper. In what follows, this structure will be assumed known. That is, the model order for the linear dynamic block and the structure of the nonlinear block will be assumed known throughout the remainder of this paper. The method detailed below does not identify the structure or model orders.

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 (6), (9) on the basis of N observations of input $\{u_t\}$ and response $\{y_t\}$. Specifically

$$\widehat{\boldsymbol{\theta}}_N = \arg\min_{\theta} V_N(\theta) \tag{10}$$

where V_N is a "least squares" cost

$$V_N(\theta) = \sum_{t=1}^{N} \varepsilon_t^2(\boldsymbol{\theta}), \qquad \varepsilon_t(\boldsymbol{\theta}) \triangleq y_t - \widehat{y}_{t|t-1}(\theta). \tag{11}$$

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 (6). Since the corruptions $\{e_t\}$ are an independent process, this is given by

$$\widehat{y}_{t|t-1}(\boldsymbol{\theta}) = H^{-1}(q, \boldsymbol{\gamma}) \left[\prod_{k=1}^{m} \mathcal{T}_k(\boldsymbol{\theta}_k) \right] u_t + \left[1 - H^{-1}(q, \boldsymbol{\gamma}) \right] y_t. \tag{12}$$

4. Estimation Algorithm

The formulation (10)-(12) specifies the estimate $\widehat{\boldsymbol{\theta}}_N$ as the solution of a possibly 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. 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

$$\mathbf{e}(\boldsymbol{\theta}) \triangleq \begin{bmatrix} \varepsilon_1(\boldsymbol{\theta}) \\ \vdots \\ \varepsilon_N(\boldsymbol{\theta}) \end{bmatrix} \tag{13}$$

and then forming a local linear approximation

$$\mathbf{e}(\boldsymbol{\theta} + \mathbf{p}) \approx \mathbf{e}(\boldsymbol{\theta}) + \mathbf{J}(\boldsymbol{\theta})\mathbf{p} \tag{14}$$

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

$$\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial e_1(\boldsymbol{\theta})}{\partial \theta_1} & \cdots & \frac{\partial e_1(\boldsymbol{\theta})}{\partial \theta_m} \\ \vdots & & \vdots \\ \frac{\partial e_N(\boldsymbol{\theta})}{\partial \theta_1} & \cdots & \frac{\partial e_N(\boldsymbol{\theta})}{\partial \theta_m} \end{bmatrix}. \tag{15}$$

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 $\mathbf{p}$ of this approximate cost satisfies

$$\mathbf{J}^T(\boldsymbol{\theta})\mathbf{J}(\boldsymbol{\theta})\mathbf{p} = -\mathbf{J}^T(\boldsymbol{\theta})\mathbf{e}(\boldsymbol{\theta}). \tag{16}$$

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 (16).

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*
$$\mathbf{J}(\boldsymbol{\theta}_k) = \mathbf{U}\mathbf{S}\mathbf{V}^T; \tag{17}$$
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*
   $$\mathbf{p}_k = -\mathbf{V}_r\mathbf{S}_r^{-1}\mathbf{U}_r^T\mathbf{e}(\boldsymbol{\theta}_k); \tag{18}$$
   (c) *If*
   $$-\mathbf{p}_k^T\mathbf{g}_k \geq \eta\,||\mathbf{p}_k||_2||\mathbf{g}_k||_2, \tag{19}$$
   then go to Step 6;
   (d) *Otherwise let* $\gamma \mapsto 0.25\gamma$ *and return to 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 go to Step 1.*

□

5. Jacobian Calculation

The estimation method examined here of using the standard prediction error method (10),(11) coupled with the gradient based search defined in Algorithm 4.1 is a very straightforward approach.

Nevertheless, the main obstacle in implementing the technique is the computation of the Jacobian $\mathbf{J}(\boldsymbol{\theta})$. To address this, suppose first that the scalar element θ_i of the full parameter vector $\boldsymbol{\theta}$ falls within the sub-block $\boldsymbol{\gamma}$ parametrizing the noise model $H(q, \boldsymbol{\gamma})$. Then for some index j

$$\frac{\partial\, \varepsilon_t(\boldsymbol{\theta})}{\partial \theta_i} = \frac{\partial}{\partial \gamma_j} H^{-1}(q, \boldsymbol{\gamma}) w_t, \quad w_t \triangleq y_t - z_t^m, \qquad z_t^m \triangleq \left[\prod_{k=1}^{m} \mathcal{T}_k(\boldsymbol{\theta}_k)\right] u_t \quad (20)$$

and therefore if $\theta_i = d_r$, the co-efficient of q^{-r} in the denominator of the noise model

$$\frac{\partial\, \varepsilon_t(\boldsymbol{\theta})}{\partial \theta_i} = \frac{q^{-r}}{C(q, \boldsymbol{\gamma})}\, w_t \quad (21)$$

while if $\theta_i = c_r$, the co-efficient of q^{-r} in the numerator of the noise model then

$$\frac{\partial\, \varepsilon_t(\boldsymbol{\theta})}{\partial \theta_i} = -H^{-1}(q, \boldsymbol{\gamma})\, \frac{q^{-r}}{C(q, \boldsymbol{\gamma})}\, w_t. \tag{22}$$

Alternatively 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

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

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 $z_t^0 \triangleq u_t$. 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

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

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

$$\frac{\partial z_t^j(\boldsymbol{\theta})}{\partial \theta_i} = \frac{\partial}{\partial a_r^j} G(q, \boldsymbol{\beta}_j) x_t^j(\boldsymbol{\lambda}_j) = -G(q, \boldsymbol{\beta}_j) \frac{q^{-r}}{A_j(q, \boldsymbol{\beta}_j)} x_t^j(\boldsymbol{\lambda}_j). \tag{26}$$

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

$$\frac{\partial z_t^j(\boldsymbol{\theta})}{\partial \theta_i} = G(q, \boldsymbol{\beta}_j) \frac{\partial}{\partial \lambda_r^j} x_t^j(\boldsymbol{\lambda}_j) = G(q, \boldsymbol{\beta}_j) \frac{\partial}{\partial \lambda_r^j} f_j(z_t^{j-1}, \boldsymbol{\lambda}_j). \tag{27}$$

The derivative term $\partial f_j / \partial \lambda$ will then depend on the nature of the non-linearity f_j, and clearly a necessary restriction is that the derivative exist.

6. Memoryless Non-linearities

Several commonly useful non-linearities which satisfy the requirement of derivative existence are now made explicit.

6.1. Saturation

This non-linearity is illustrated in figure 3 and mathematically described

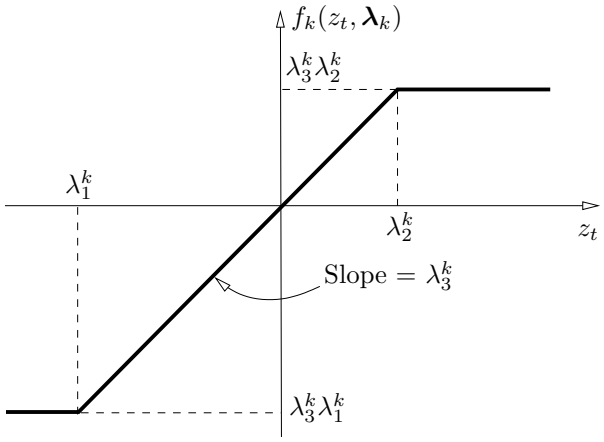

Figure 3: *Saturation Non-Linearity*

as

$$f_k(z_t, \boldsymbol{\lambda}_k) = \begin{cases} \lambda_3^k z_t & ; z_t \in [\lambda_1^k, \lambda_2^k], \\ \lambda_2^k \lambda_3^k & ; z_t > \lambda_2^k, \\ \lambda_1^k \lambda_3^k & ; z_t < \lambda_1^k. \end{cases} \tag{28}$$

The necessary gradients exist and are given by (assuming without loss of generality that $\lambda_2^k \geq \lambda_1^k$)

$$\frac{\partial f_k(z_t, \boldsymbol{\lambda}_k)}{\partial \lambda_1^k} = \begin{cases} \lambda_3^k & ; z_t < \lambda_1^k \\ 0 & ; \text{Otherwise} \end{cases} \tag{29}$$

$$\frac{\partial f_k(z_t, \boldsymbol{\lambda}_k)}{\partial \lambda_2^k} = \begin{cases} \lambda_3^k & ; z_t \geq \lambda_2^k \\ 0 & ; \text{Otherwise} \end{cases} \tag{30}$$

$$\frac{\partial f_k(z_t, \boldsymbol{\lambda}_k)}{\partial \lambda_3^k} = \begin{cases} z_t & ; z_t \in [\lambda_1^k, \lambda_2^k] \\ \lambda_2^k & ; z_t > \lambda_2^k \\ \lambda_1^k & ; z_t < \lambda_1^k. \end{cases} \tag{31}$$

6.2. Deadzone

This is illustrated in figure 4. with associated mathematical description

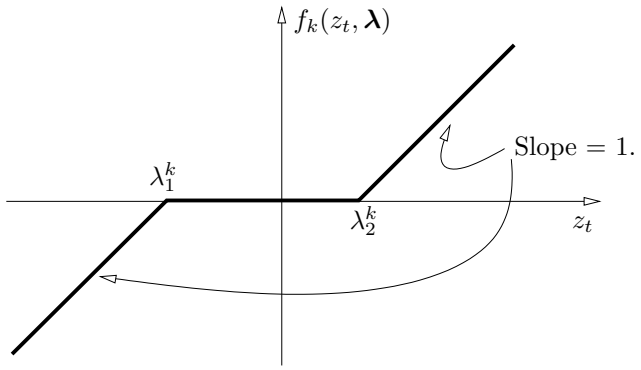

Figure 4: *Dead-zone Non-Linearity*

given by

$$f_k(z_t, \boldsymbol{\lambda}_k) = \begin{cases} z_t - \lambda_2^k & ; z_t > \lambda_2^k \\ z_t - \lambda_1^k & ; z_t < \lambda_1^k \\ 0 & ; z_t \in [\lambda_1^k, \lambda_2^k]. \end{cases} \tag{32}$$

Note that, in the generalised Hammerstein–Wiener setting of this paper, there is no loss of generality by assuming the gain to be one in the linear region, since non-unity gains can be accounted for by gains in a subsequent linear block.

Importantly, this is *not* the case for the previously discussed saturation. A saturation with slope λ_3^k as illustrated in figure 3 is not equivalent to one of slope 1 followed by an amplifier of gain λ_3^k, since this latter case would involve altering the saturation limits λ_1^k and λ_2^k proportional to the gain λ_3^k.

The required gradients associated with this non-linearity again exist and are given by (again assuming $\lambda_1^k \leq \lambda_2^k$)

$$\frac{\partial f_k(z_t, \boldsymbol{\lambda}_k)}{\partial \lambda_1^k} = \begin{cases} -1 & ; z_t < \lambda_1^k \\ 0 & ; \text{Otherwise} \end{cases} \tag{33}$$

$$\frac{\partial f_k(z_t, \boldsymbol{\lambda}_k)}{\partial \lambda_2^k} = \begin{cases} -1 & ; z_t \geq \lambda_2^k \\ 0 & ; \text{Otherwise} \end{cases} \tag{34}$$

6.3. Polynomial

The preceding memoryless non-linearities of saturation and deadzone are generally specifically tailored to describe situations that commonly occur in practice (e.g. clipping of amplifiers). In other cases, where either different non-linearities are involved, or the form is not suspected before hand, an alternative approach is to model the non-linearity in a non-specific manner.

One example of this approach, which is very commonly used, is that of polynomial description. This is motivated by the fact that if the non-linearity $f_k(\cdot)$ is smooth enough to be differentiable an arbitrary number of times, then it has a Taylor series representation about an operating point (say 0) of

$$f_k(z_t) = f_k(0) + z_t \cdot \frac{\partial}{\partial z_t} f_k(z_t)\bigg|_{z_t=0} + z_t^2 \cdot \frac{\partial^2}{\partial z_t^2} f_k(z_t)\bigg|_{z_t=0} + z_t^3 \cdot \frac{\partial^3}{\partial z_t^3} f_k(z_t)\bigg|_{z_t=0} + \cdots \tag{35}$$

Therefore, in cases where the non-linearity $f_k(\cdot)$ does not produce a constant offset so that $f_k(0) = 0$, then by truncating this expansion at a fixed order (say ℓ_k) the following polynomial description is obtained

$$f_k(z_t, \boldsymbol{\lambda}_k) = \lambda_1^k z_t + \lambda_2^k z_t^2 + \lambda_3^k z_t^3 + \cdots + \lambda_{\ell_k}^k z_t^{\ell_k}. \tag{36}$$

However, note that for a non-smooth non-linearity like a saturation or a dead-zone, the expansion order ℓ_k may need to be quite large before accurate approximation is possible. The necessary derivatives clearly exist and are given by

$$\frac{\partial}{\partial \lambda_j^k} = z_t^j. \tag{37}$$

6.4. Piecewise linear

A further example of memoryless non-linearity modelling in a non-specific manner is that of piecewise linear modelling. In this situation, the non-linearity $f_k(z_t, \boldsymbol{\lambda}_k)$ is described by a a set of 'one-dimensional hinge' building blocks $h_1^k(z_t, \boldsymbol{\lambda}_k), \cdots, h_{\ell_k}^k(z_t, \boldsymbol{\lambda}_k)$ of the form

$$h_j^k(z_t, \boldsymbol{\lambda}_k) = \begin{cases} \lambda_{j+2}^k + \lambda_{j+3}^k z_t & ; \quad z_t > -\dfrac{\lambda_{j+2}^k}{\lambda_{j+3}^k}, \\ \\ 0 & ; \quad \text{Otherwise} \end{cases} \tag{38}$$

(39)

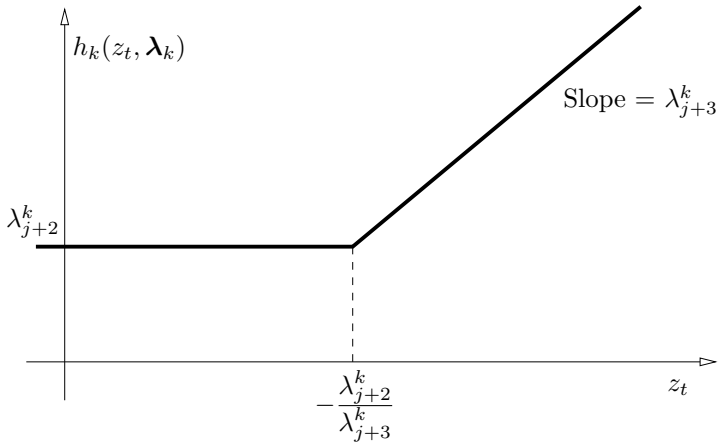

Figure 5: *Hinge building block for piecewise linear description*

which are illustrated in figure 5. The complete piecewise linear description $f_k(z_t, \boldsymbol{\lambda}_k)$ is then built as a sum of (say ℓ_k) hinges (Breiman, 1993):

$$f_k(z_t, \boldsymbol{\lambda}_k) = \lambda^k_1 + \lambda^k_2 z_t + \sum_{j=1}^{\ell_k} h^k_j(z_t, \boldsymbol{\lambda}_k) \tag{40}$$

which is then parametrized according to

$$\boldsymbol{\lambda}_k = [\lambda^k_1, \lambda^k_2, \cdots, \lambda^k_{\ell+2}]. \tag{41}$$

Again, the tradeoff for the generality of this modelling approach is that more parameters are required than would be for a tailored approach (if the latter is possible). For example, to model the deadzone shown in figure 4 would require an initial linear term followed by two hinges for a total of six parameters. Nevertheless (as will be illustrated) it can be very useful in situations where detailed knowledge of the non-linearity type is not known.

The associated gradients are computed as follows. First

$$\frac{\partial f_k(z_t, \boldsymbol{\lambda}_k)}{\partial \lambda^k_1} = 1, \qquad \frac{\partial f_k(z_t, \boldsymbol{\lambda}_k)}{\partial \lambda^k_2} = z_t. \tag{42}$$

Furthermore for $r > 2$

$$\frac{\partial f_k(z_t, \boldsymbol{\lambda}_j)}{\partial \lambda_r^k} = \frac{\partial h_{\lfloor (r-1)/2 \rfloor}(z_t, \boldsymbol{\lambda}_k)}{\partial \lambda_r^k} \tag{43}$$

$$= \begin{cases} 1 \quad ; \quad r \text{ odd and } z_t > -\dfrac{\lambda_r^k}{\lambda_{r+1}^k}, \\ z_t \quad ; \quad r \text{ even and } z_t > -\dfrac{\lambda_r^k}{\lambda_{r+1}^k}, \\ 0 \quad ; \quad z_t \leq -\dfrac{\lambda_r^k}{\lambda_{r+1}^k}. \end{cases} \tag{44}$$

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

7. Results

As explained in the introduction, a key motivation for studying the model structure, estimation method and algorithm presented in sections 2-5 was to attempt the "Wiener–Hammerstein Benchmark" originally proposed in (Schoukens et al., 2008).

The physical apparatus responsible for this benchmark is revealed in (Schoukens et al., 2008) to consist 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 (6). 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).

Three choices for the structure of the non-linear blocks f_1 and f_2 where studied. Namely, the polynomial structure defined in section 6.3, the saturation defined in section 6.1 and the piecewise linear form specified in section 6.4.

In the polynomial case, the order $\ell_k = 8$ was chosen for both blocks $k = 1, 2$, and the co-efficients were initialised to give linear response:

$$\lambda_1^k = 1, \quad \lambda_r^k = 0,\ r \in [2, 8], k = 1, 2. \tag{45}$$

In the saturation case, the initialisation was chosen as

$$-\lambda_1^k = \lambda_2^k = 0.1, \quad \lambda_3^k = 1, \quad k = 1, 2. \tag{46}$$

Finally, for the piecewise linear situation, the number of possible "break-points" was set at $\ell_1 = \ell_2 = 8$. These potential non-linearities were then in fact initialised to be linear according to

$$\lambda_1^k = 0, \quad \lambda_2^k = 1, \quad \lambda_r^k = 0, r \in [3, 8] \quad ; k = 1, 2. \tag{47}$$

The linear blocks G_1 and G_2 were initialised as

$$G_1(q, \boldsymbol{\beta}_1) = G_2(q, \boldsymbol{\beta}_2) = \frac{(1 - 0.5q^{-1})^3}{(1 - 0.5q^{-1})^3}. \tag{48}$$

Finally, in all cases the noise filter H was chosen as $H = 1$. With these choices of model structure and initialisation, an estimate $\widehat{\boldsymbol{\theta}}_N$ was computed using the methods described in sections 3-6 and using the first $N = 10000$ measured data samples. In order to evaluate it, the performance measures defined by the benchmark proposers (Schoukens et al., 2008) are presented in Table 1. The measures μ and s are the mean and standard deviation of

Nonlinearity	**Data set**	μ	**s**	$\mathbf{e}_{\mathrm{RMS}}$
Polynomial	**Estimation**	8.83×10^{-5}	2.07×10^{-3}	2.08×10^{-3}
Polynomial	**Validation**	8.87×10^{-5}	2.13×10^{-3}	2.14×10^{-3}
Saturation	**Estimation**	-2.10×10^{-3}	6.02×10^{-3}	6.38×10^{-3}
Saturation	**Validation**	-2.16×10^{-5}	6.20×10^{-3}	6.56×10^{-3}
Piecewise Linear	**Estimation**	4.14×10^{-7}	4.73×10^{-4}	4.73×10^{-4}
Piecewise Linear	**Validation**	-2.37×10^{-5}	4.80×10^{-4}	4.81×10^{-4}

Table 1: *Performance Measures.*

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 $\widehat{\boldsymbol{\theta}}_N$:

$$\mu \triangleq \frac{1}{87000} \sum_{t=101001}^{188000} \varepsilon_t(\widehat{\boldsymbol{\theta}}_N), \quad s \triangleq \sqrt{\frac{1}{87000} \sum_{t=101001}^{188000} [\varepsilon_t(\widehat{\boldsymbol{\theta}}_N) - \mu]^2} \tag{49}$$

An additional measure is the root mean square (RMS) prediction error, which was calculated on subsets (respectively) of the validation data, and the data used to actually derive the estimate $\widehat{\boldsymbol{\theta}}_N$:

$$e_{\mathrm{RMSt}} \triangleq \sqrt{\frac{1}{87000} \sum_{t=101001}^{188000} \varepsilon_t^2(\widehat{\boldsymbol{\theta}}_N)}, \quad e_{\mathrm{RMSe}} \triangleq \sqrt{\frac{1}{99000} \sum_{t=1001}^{100000} \varepsilon_t^2(\widehat{\boldsymbol{\theta}}_N)}. \tag{50}$$

Consideration of these measures indicates that the piecewise linear modelling delivers the most accurate estimate. Consequently, the paper will now consider this choice more closely by first presenting the associated estimated frequency responses $G_1(e^{j\omega}, \widehat{\boldsymbol{\theta}}_N)$, $G_2(e^{j\omega}, \widehat{\boldsymbol{\theta}}_N)$ in Figures 6 and 7 as the solid lines. Also shown as dashed lines are the underlying "true" responses as disclosed 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.

In presenting these comparisons, it is important to note that the model structure that are being employed are 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.

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 8 and 9. In presenting the true non-linearity map in Figure 9, in was necessary to account for the IN4148 diode disclosed in (Schoukens et al., 2008) by noting that from its data-sheet, at 25°C it has a voltage versus log-current relationship that in the forward biased 0.1mA-100mA range approximately obeys

$$\log_{10} i = 9.38v - 8.69. \tag{51}$$

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

$$v_i = \frac{11}{10} v_0 + 10^{-8.69} \cdot 10^{9.38 v_0}. \tag{52}$$

This was solved numerically for v_0 for a range of v_i in order to provide the dashed line "true system" plot shown in Figure 9. As for the linear blocks, these input-output maps illustrate what is considered accurate estimates of the underlying system.

The benchmark specification (Schoukens et al., 2008) also requests that the error between the observed response and that of the identified model

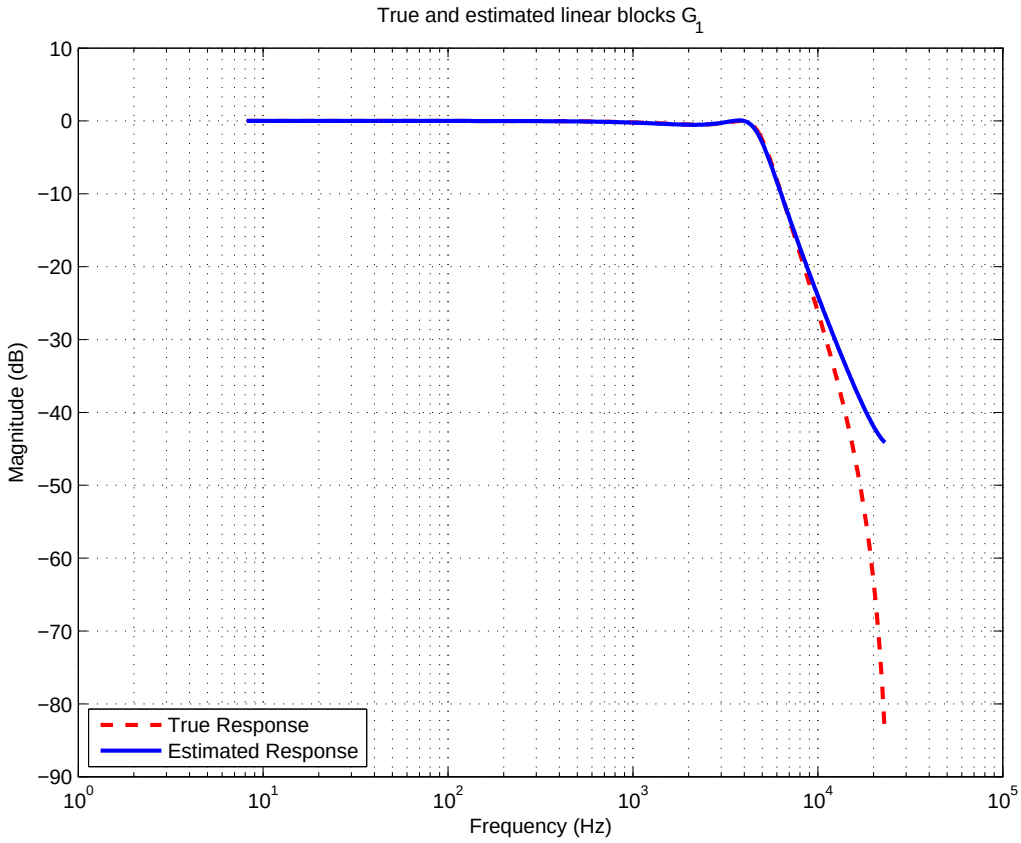

Figure 6: *Bode magnitude frequency responses of true (dashed line) and estimated (solid line) linear dynamic block* G_1.

be presented, both in the time and frequency domain. This is addressed in the time domain case in Figure 10 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 11.

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 (48) was replaced by a randomly chosen third order system using the MATLAB command `poly(rand(1,3))` to specify initial numerator and

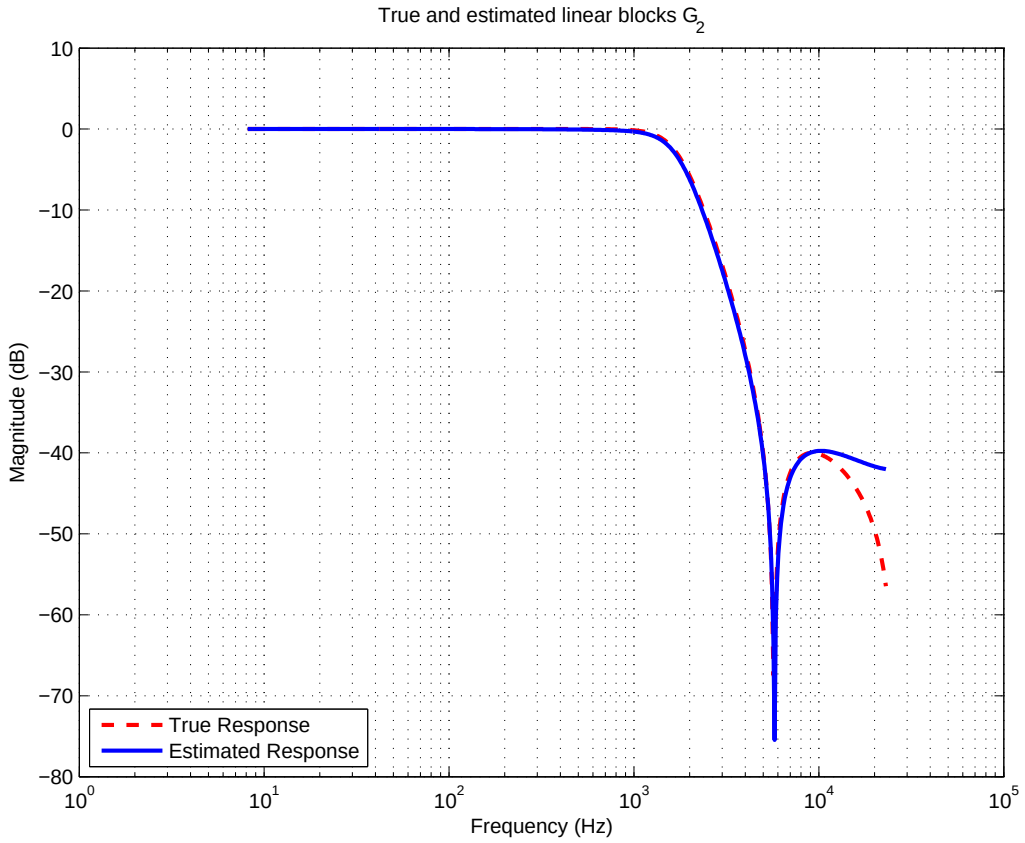

Figure 7: *Bode magnitude frequency responses of true (dashed line) and estimated (solid line) linear dynamic block* G_2.

denominator polynomials (separately). This was done for 100 initial system realisations. The final cost value of the gradient based search Algorithm 4.1 (i.e. the cost value when it terminates) initialised at these 100 different points are plotted in Figure 12. Clearly, while there is some consistency in achieving a terminal cost near $10^{-6.8}$, there is a susceptibility to termination in local minima.

However, to gauge the importance of not achieving the global minimum, the estimated frequency responses $G_1(e^{j\omega}, \widehat{\boldsymbol{\theta}}_N)$ and $G_2(e^{j\omega}, \widehat{\boldsymbol{\theta}}_N)$ for all achieved costs less than 10^{-4} (indicated by the horizontal dashed line in Figure 12) are plotted relative to the true response in Figures 13 and 14, respectively. Similarly, the estimated input-output maps $f_1(\cdot, \widehat{\boldsymbol{\theta}})$ and $f_2(\cdot, \widehat{\boldsymbol{\theta}})$

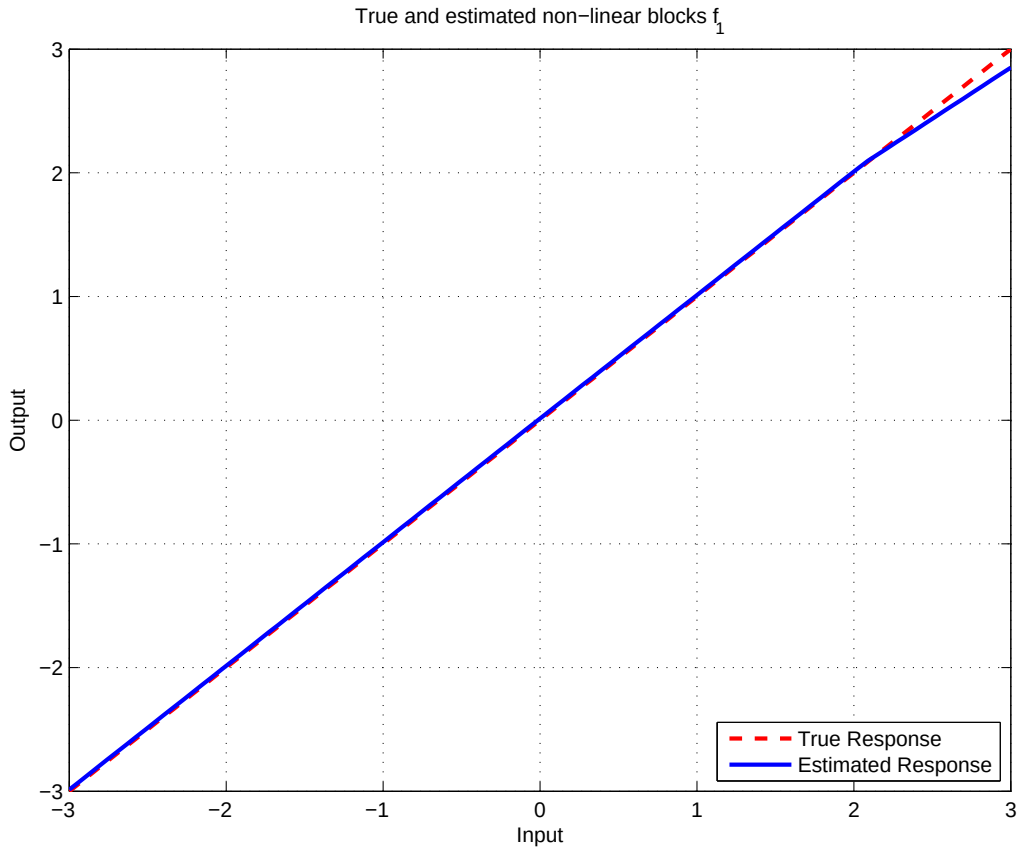

Figure 8: *Input-output maps for true (dashed line) and estimated (solid line) memoryless non-linearity block* f_1.

for these same achieved costs are plotted relative to the true input-output map in Figures 15 and 16, respectively. Consideration of these plots shows that 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.

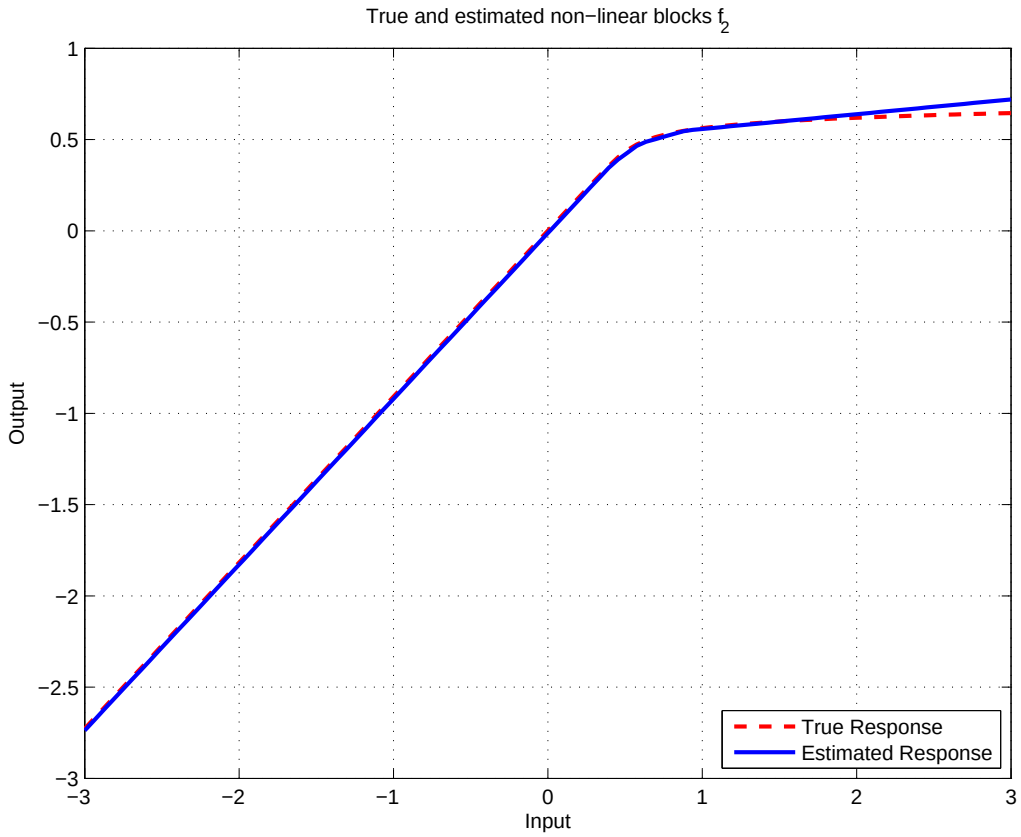

Figure 9: *Input-output maps for true (dashed line) and estimated (solid line) memoryless non-linearity block* f_2.

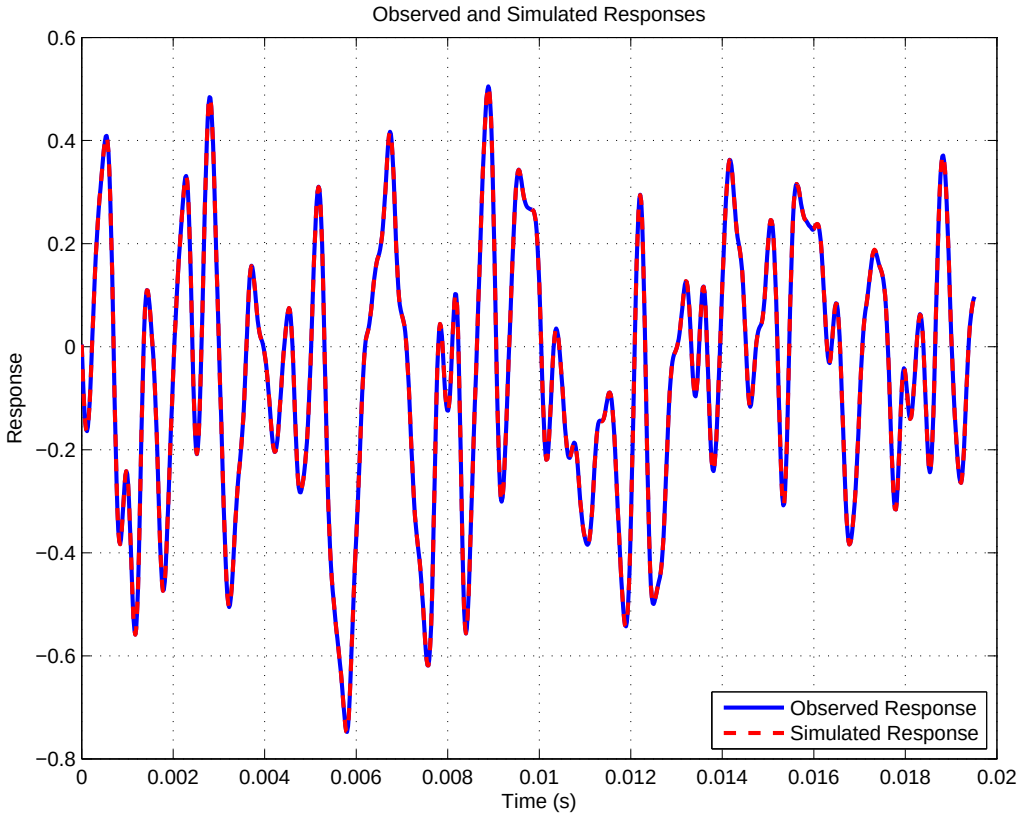

Figure 10: *Observed and simulated response on first 20ms of validation data.*

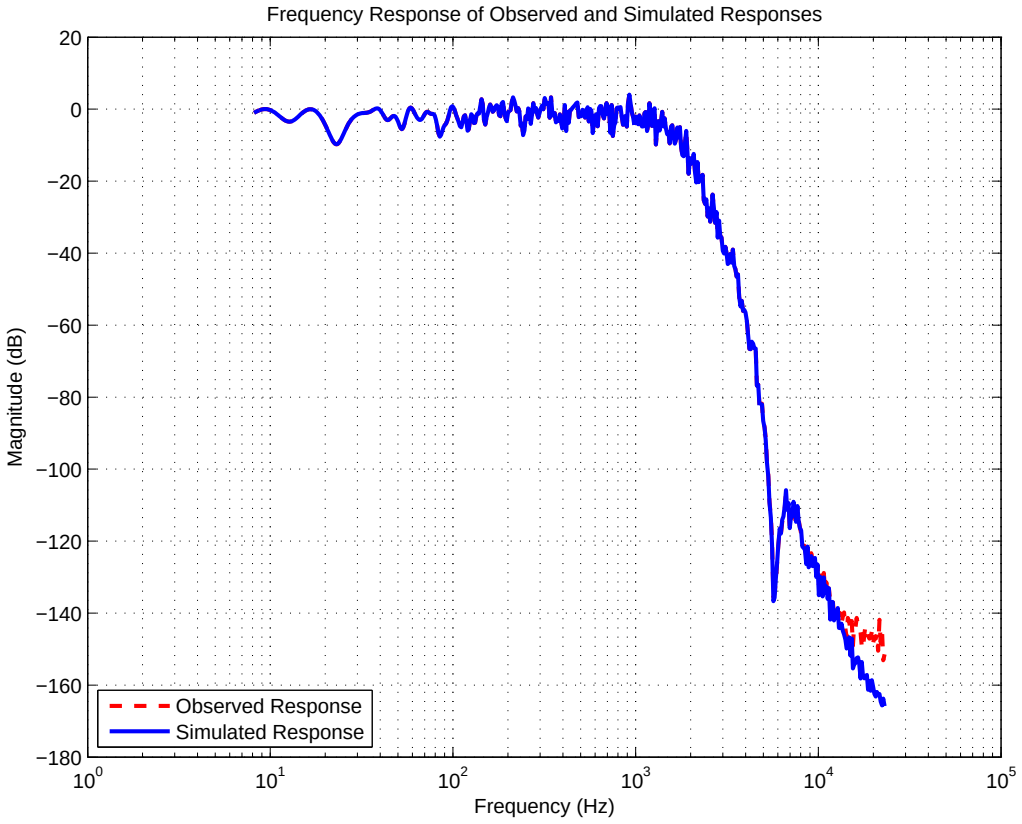

Figure 11: Frequency response of measured and predicted outputs

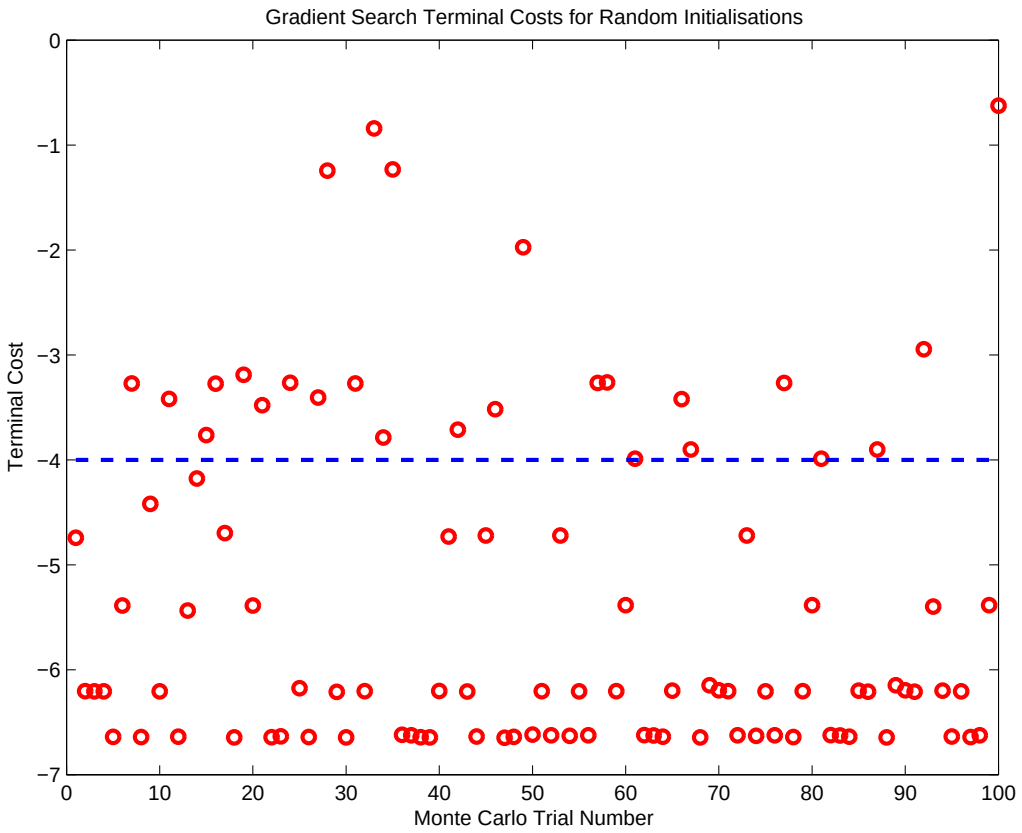

Figure 12: *Gradient search terminal costs (final cost value at the termination of Algorithm 4.1) for 100 randomly chosen G_1 and G_2 initialisations.*

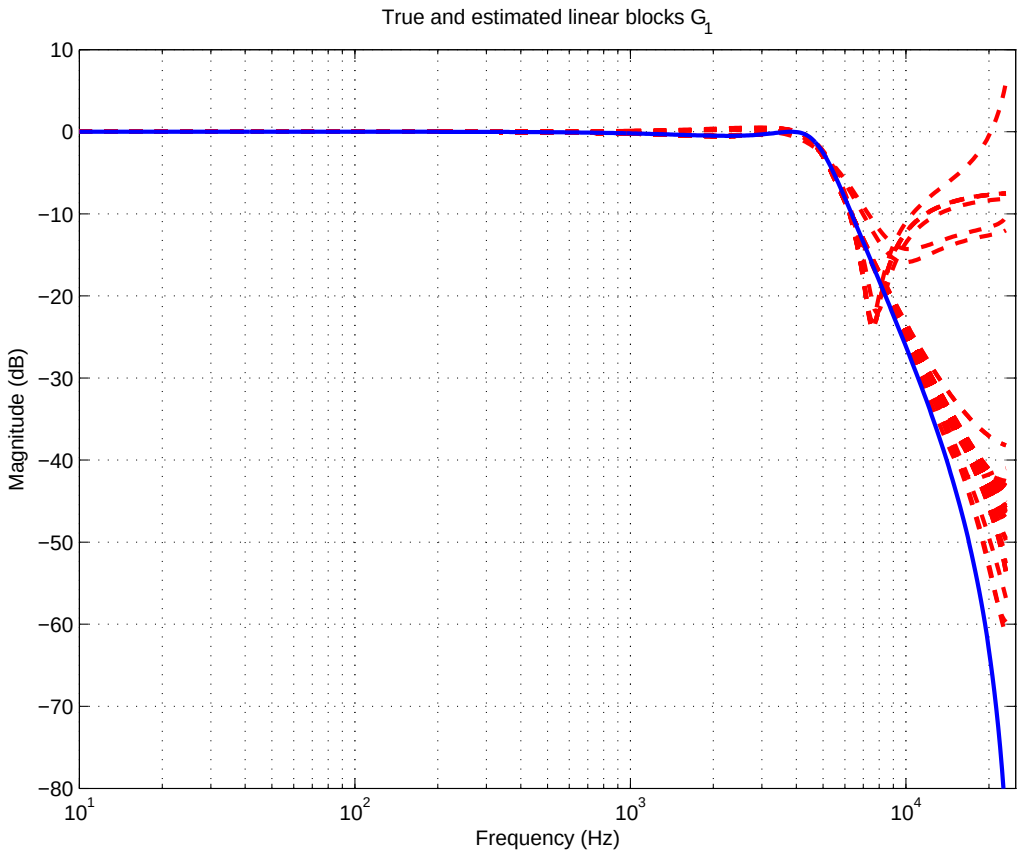

Figure 13: *Estimated frequency responses* $G_1(e^{j\omega}, \widehat{\boldsymbol{\theta}}_N)$ *for all Monte–Carlo realisations achieving a terminal cost of* 10^{-4} *or less.*

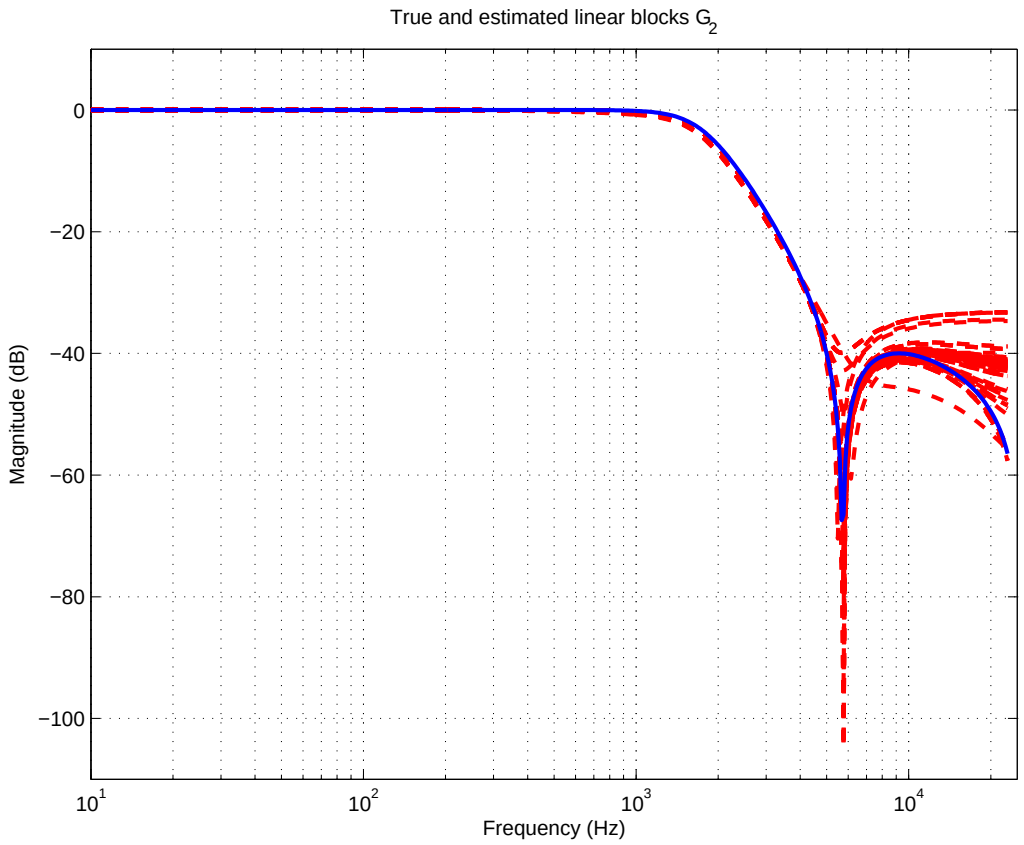

Figure 14: *Estimated frequency responses* $G_2(e^{j\omega}, \widehat{\boldsymbol{\theta}}_N)$ *for all Monte–Carlo realisations achieving a terminal cost of* 10^{-4} *or less.*

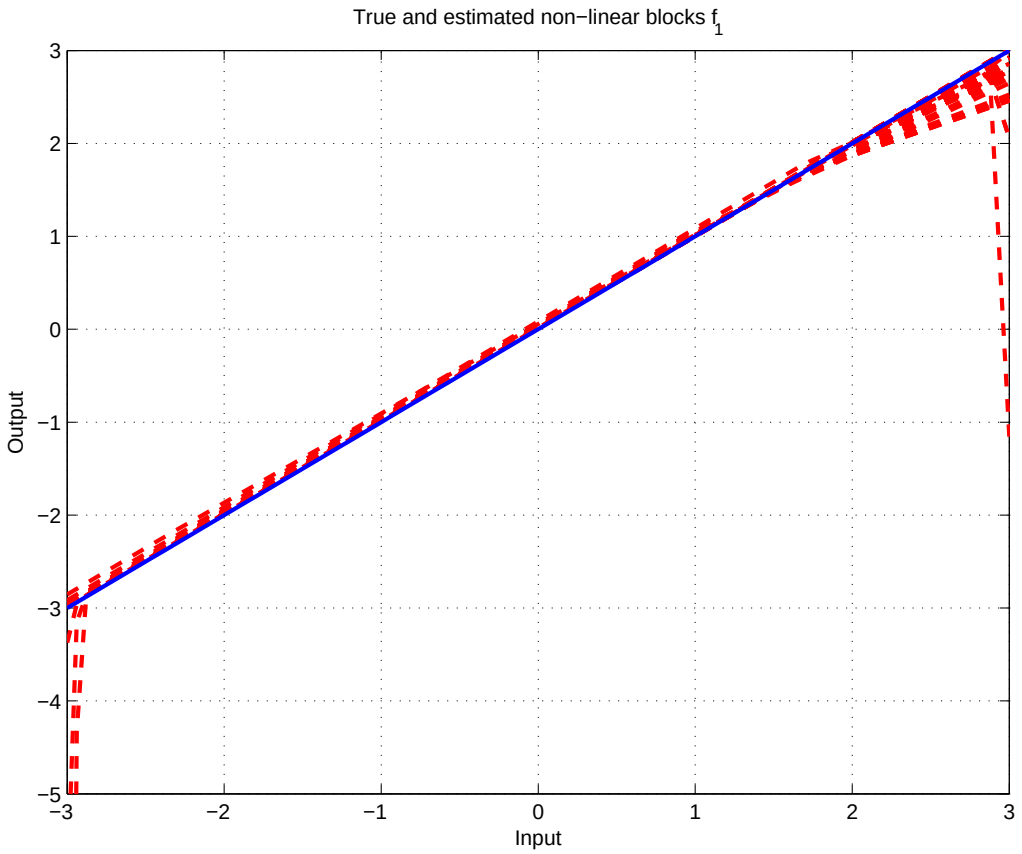

Figure 15: *Estimated input-output map of* $f_1(\cdot, \widehat{\boldsymbol{\theta}})$ *for all Monte–Carlo realisations achieving a terminal cost of* 10^{-4} *or less.*

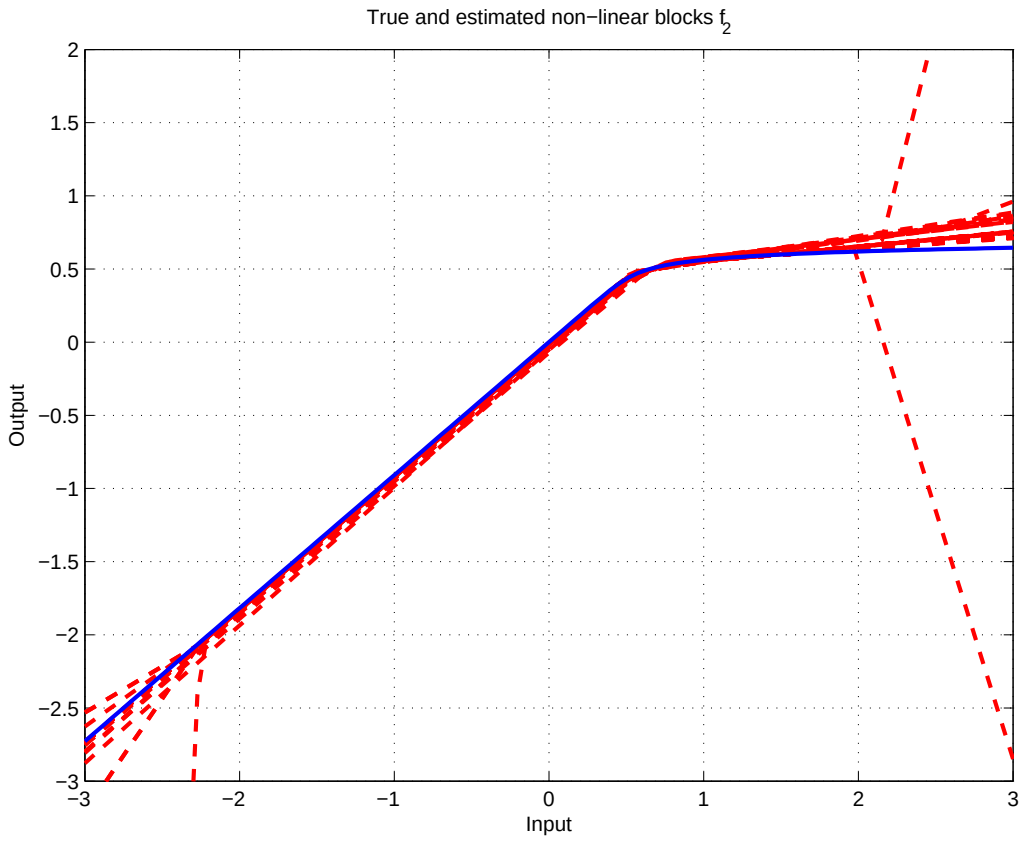

Figure 16: *Estimated input-output map of* $f_2(\cdot, \widehat{\boldsymbol{\theta}})$ *for all Monte–Carlo realisations achieving a terminal cost of* 10^{-4} *or less.*

8. Conclusion

This paper has studied the application of the widely accepted prediction error framework to a block-oriented non-linear model structure that can encompass Hammerstein and Wiener type models together with their combinations. The utility of this approach has been profiled via a widely studied benchmark example, namely that proposed at the 15th IFAC Symposium on System identification (Schoukens et al., 2008).

References

Ase, H., Katayama, T., Tanaka, H., July 6–8 2009. A State-Space Approach to Identification of Wiener-Hammerstein Benchmark Model. In: 15th IFAC Symposium on System Identification. Saint-Malo, France, pp. 1092–1097.

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

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

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

Brabanter, K. D., Dreesen, P., Karsmakers, P., Pelckmans, K., Brabanter, J. D., Suykens, J., Moor, B. D., July 6–8 2009. Fixed-Size LS-SVM Applied to the Wiener-Hammerstein Benchmark. In: 15th IFAC Symposium on System Identification. Saint-Malo, France, pp. 826–831.

Breiman, L., 1993. Hinging hyperplanes for regression, classification, and function approximation. IEEE Transactions on Information Theory 39 (3), 999–1013.

dos Santos, P. L., Ramos, J., de Carvalho, J. L. M., July 6–8 2009. Identification of a Benchmark Wiener-Hammerstein System by Bilinear and Hammerstein-Bilinear Models. In: 15th IFAC Symposium on System Identification. Saint-Malo, France, pp. 832–837.

Falck, T., Pelckmans, K., Suykens, J., Moor, B. D., July 6–8 2009. Identification of Wiener-Hammerstein Systems using LS-SVMs. In: 15th IFAC Symposium on System Identification. Saint-Malo, France, pp. 820–825.

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

Lauwers, L., Pintelon, R., Schoukens, J., July 6–8 2009. Modelling of Wiener-Hammerstein Systems via the Best Linear Approximation. In: 15th IFAC Symposium on System Identification. Saint-Malo, France, pp. 1098–1103.

Ljung, L., 1999. System Identification: Theory for the User, (2nd edition). Prentice-Hall, Inc., New Jersey.

L.Ljung, P.E.Caines, 1979. Asymptotic Normality of prediction error estimators for approximate system models. Stochastics 3, 29–46.

Lovera, M., Gustafsson, T., Verhaegen, M., 2000. Recursive subspace identification of linear and non-linear Wiener state-space models. Automatica J. IFAC 36 (11), 1639–1650.

Marconato, A., Schoukens, J., July 6–8 2009. Identification of Wiener-Hammerstein Benchmark Data by Means of Support Vector Machines. In: 15th IFAC Symposium on System Identification. Saint-Malo, France, pp. 816–819.

Mulders, A. V., Schoukens, J., Volckaert, M., Diehl, M., July 6–8 2009. Two Nonlinear Optimization Methods for Black Box Identification Compared. In: 15th IFAC Symposium on System Identification. Saint-Malo, France, pp. 1086–1091.

Paduart, J., Lauwers, L., Pintelon, R., Schoukens, J., July 6–8 2009. Identification of a Wiener-Hammerstein System Using the Polynomial Nonlinear State Space Approach. In: 15th IFAC Symposium on System Identification. Saint-Malo, France, pp. 1080–1085.

Pillonetto, G., Chiuso, A., July 6–8 2009. Gaussian Processes for Wiener-Hammerstein system identification. In: 15th IFAC Symposium on System Identification. Saint-Malo, France, pp. 838–843.

Piroddi, L., Farina, M., Lovera, M., July 6–8 2009. Polynomial NARX Model Identification: a Wiener-Hammerstein Benchmark. In: 15th IFAC Symposium on System Identification. Saint-Malo, France, pp. 1074–1079.

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

Truong, N., Wang, L., July 6–8 2009. Benchmark Nonlinear System Identification using Wavelet based SDP Models. In: 15th IFAC Symposium on System Identification. Saint-Malo, France, pp. 844–849.

Vörös, J., 2007. Parameter identification of Wiener systems with multisegment piecewise-linear nonlinearities. Systems Control Lett. 56 (2), 99–105.

Wills, A., B.Ninness, July 6–8 2009. Estimation of Generalised Hammerstein-Wiener Systems. In: 15th IFAC Symposium on System Identification. Saint-Malo, France, pp. 1104–1109.

Wills, A., Ninness, B., 2008. On gradient-based search for multivariable system estimates. IEEE Trans. Automat. Control 53 (1), 298–306.

Zhu, Y., 2002. Estimation of an N-L-N Hammerstein-Wiener model. Automatica J. IFAC 38 (9), 1607–1614.