



# CHALMERS

## Chalmers Publication Library

### **Identification of Wiener–Hammerstein models: Two algorithms based on the best split of a linear model applied to the SYSID'09 benchmark problem**

This document has been downloaded from Chalmers Publication Library (CPL). It is the author's version of a work that was accepted for publication in:

**Control Engineering Practice (ISSN: 0967-0661)**

Citation for the published paper:

Sjöberg, J. ; Lieve, L. ; Schoukens, J. (2012) "Identification of Wiener–Hammerstein models: Two algorithms based on the best split of a linear model applied to the SYSID'09 benchmark problem". *Control Engineering Practice*, vol. 20(11), pp. 1119–1125.

<http://dx.doi.org/10.1016/j.conengprac.2012.07.001>

Downloaded from: <http://publications.lib.chalmers.se/publication/164072>

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source. Please note that access to the published version might require a subscription.

Chalmers Publication Library (CPL) offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all types of publications: articles, dissertations, licentiate theses, masters theses, conference papers, reports etc. Since 2006 it is the official tool for Chalmers official publication statistics. To ensure that Chalmers research results are disseminated as widely as possible, an Open Access Policy has been adopted. The CPL service is administrated and maintained by Chalmers Library.

(article starts on next page)

# Identification of Wiener-Hammerstein Models: Two Algorithms Based on the Best Split of a Linear Model Applied to the SYSID'09 Benchmark Problem

J. Sjöberg<sup>a</sup> L. Lauwers<sup>b</sup> J. Schoukens<sup>b</sup>

<sup>a</sup>*Chalmers University of Technology, Department of Signals & Systems, SE412 96  
Gothenburg, Sweden.*

<sup>b</sup>*Vrije Universiteit Brussel, Faculty of Engineering, Department of Fundamental  
Electricity and Instrumentation, Pleinlaan 2, 1050 Brussels, Belgium*

---

## Abstract

This paper describes the identification of Wiener-Hammerstein models and two recently suggested algorithms are applied to the SYSID'09 benchmark data. The most difficult step in the identification process of such block-oriented models is to generate good initial values for the linear dynamic blocks so that local minima are avoided. Both of the considered algorithms obtain good initial estimates by using the Best Linear Approximation (BLA) which can easily be estimated from data. Given the BLA, the two algorithms differ in the way the dynamics are separated into two linear parts. The first algorithm simply considers all possible splits of the dynamics. Each of the splits is used to initialize one Wiener-Hammerstein using least-squares and the best performing model is selected. In the second algorithm, both linear blocks are initialized with the entire BLA model using basis function expansions of the poles and zeros of the BLA. This gives over-parameterized linear blocks and their order is decreased in a model reduction step. Both algorithms are explained and their properties are discussed. They both give good, comparable, models on the benchmark data.

*Key words:* Wiener-Hammerstein systems, Hammerstein systems, Wiener systems, nonlinear system identification, initial estimates, best linear approximation

---

---

*Email addresses:* Jonas.Sjoberg@chalmers.se (J. Sjöberg),  
Lieve.Lauwers@vub.ac.be (L. Lauwers), Johan.Schoukens@vub.ac.be  
(J. Schoukens).

## 1 Introduction

This paper considers the identification of Wiener-Hammerstein models. Two recently suggested, similar, algorithms, [8] and [15,16] are applied to the data from the benchmark session at SYSID 2009, [13].

General in system identification, the prediction error estimate gives an asymptotic efficient estimator when the number of estimation data goes to infinity, see, eg, [9,17]. However, in most cases, except when the model can be expressed as linear regression, the computation of the estimate requires an iterative search of the minimum of the cost function. The cost function can have multiple minima and it is a main challenge in system identification research to invent algorithms which can guarantee, or at least increase the chances that the estimate converges to the global minimum. Hence, for each type of model structure there is a need of good initialization algorithms and the two algorithms considered in this paper deliver initialization for the Wiener-Hammerstein model.

The Wiener-Hammerstein model consists of two linear dynamic systems with a static nonlinearity sandwiched in between, see Figure 1. Early results about the identification of these models can be found in [2] and [4]. More recent work is reported in [5] and [3]. Few papers give hints how to obtain good initial estimates. In [6], an iterative initialization procedure is proposed which requires specially designed periodic excitation signals. This experimental requirement is loosened in [12]. Other methods circumvent the need for initial values by performing a large number of experiments [22], or by restricting the allowed model complexity [1], [18], [19].

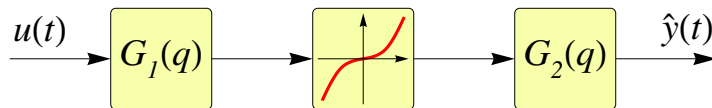


Fig. 1. A Wiener-Hammerstein model structure.

The considered algorithms, [8] and [15,16], rely on the fact that the BLA model is a consistent estimate of the concatenation of the two linear blocks, see [11]. This result assumes that the data indeed was generated by a Wiener-Hammerstein process and some additional assumption on the data and the system, for instance that the linear parts need to be stable. Hence, given the BLA, the remaining problem is to divide the dynamics into two parts, and to estimate the nonlinearity.

The first algorithm, [15,16], is a kind of brute-force method where the dynamics are split in all possible ways, and for all possible splits a Wiener-

Hammerstein model is estimated by fitting the (linear) parameters of the nonlinear part using least-squares. The number of splits might be high, but since least-squares is used in the initialization, fairly large number of splits can be handled within a reasonable time. In [15,16] it is argued that problems where the order of the BLA model is up to about 10 can be handled. Moreover, this initialization procedure is asymptotic consistent in the number of data.

The second algorithm, [8], circumvent the problem with many splits of the BLA by offering the whole BLA model to both blocks. The initialization is formulated as a Total-Least-Squares problem (TLS) *in between the linear blocks*. This is done by filtering the input signal through the basis functions of the first linear block, based on the poles of the BLA, and filtering the output through the basis functions describing the inverse of the BLA, based on the zeros of the BLA. This approach does not have the drawback with exponential increasing computational time with respect to the model order, as the first algorithm has. However, it has the drawback that the linear blocks will have higher order than necessary, but model order reduction techniques can be applied. Also, the solution is in general not consistent if there is noise on the output.

In their general forms, none of the algorithms can handle the case where there is pole-zero cancellation in the BLA models i.e., when the pole and the zero come from different linear blocks.

The paper is organized as follows. A description of the problem setting and the considered model structure is given in Section 2. This is followed by an overview of the two algorithms in Section 3. Then, in Section 4 and 5 the main steps of both algorithms are explained, respectively. In Section 6 the result on the benchmark data is given and in Section 7 the paper is concluded.

## 2 Problem Formulation

The problem formulation is divided into three steps, the definition of the model structure, the assumptions on the data, and the definition and computation of the estimate.

### 2.1 Model structure

The concerned model structure is of the Wiener-Hammerstein type and is described by

$$\begin{aligned}
z(t) &= G_1(q^{-1}, \alpha) u(t) \\
x(t) &= f(\beta, z(t)) \\
\hat{y}(t) &= G_2(q^{-1}, \gamma) x(t)
\end{aligned} \tag{1}$$

where  $G_1(q^{-1}, \alpha)$  and  $G_2(q^{-1}, \gamma)$  are linear time invariant transfer functions in the delay operator  $q^{-1}$ , and parameterized with  $\alpha$  and  $\gamma$ , respectively.  $\hat{y}(t)$  is the model output, the prediction of the system output  $y(t)$ , and  $z(t)$  and  $x(t)$  are model internal variables. The function  $f$  is a static nonlinearity parameterized with  $\beta$ .

All parameters of the model structure are stored in a common parameter vector

$$\theta = [\alpha, \beta, \gamma]. \tag{2}$$

The first linear part of the model can be described as

$$G_1(q^{-1}, \alpha) = \frac{b_0^1 + b_1^1 q^{-1} + \dots + b_{m_{b_1}}^1 q^{-m_{b_1}}}{1 + a_1^1 q^{-1} + \dots + a_{m_{a_1}}^1 q^{-m_{a_1}}} \tag{3}$$

where  $\alpha = [b_0^1, \dots, b_{m_{b_1}}^1, a_1^1, \dots, a_{m_{a_1}}^1]$ , and  $G_2(q^{-1}, \gamma)$  is described similarly with  $\gamma = [b_0^2, \dots, b_{m_{b_2}}^2, a_1^2, \dots, a_{m_{a_2}}^2]$ .

The static nonlinearity is described as a basis function expansion

$$\begin{aligned}
f(\beta, z) &= \sum_{k=1}^n \beta_k^1 f_k(\beta_k^2, z) \\
\beta &= [\beta^1, \beta^2]^T = \\
& \quad [\beta_1^1, \dots, \beta_n^1, \beta_1^2, \dots, \beta_n^2]^T
\end{aligned} \tag{4}$$

where  $f_k$  are basis functions, and  $\beta$  has been divided into  $\beta^1$ , which enters linearly in  $f$ , and  $\beta^2$ , which enters non-linearly in  $f$ . Often there are several parameters entering nonlinearly for each basis function, i.e., each  $\beta_k^2$ , contains several parameters. With this general description of the static nonlinearity most specific basis function expansions can be described with a specific choice of the basis  $f_k$ . If, for example, a polynomial model is chosen, then

$$f(\beta, z) = \beta_0^1 + \beta_1^1 z + \beta_2^1 z^2 + \dots + \beta_n^1 z^n$$

and in this case there are no parameters in  $\beta^2$ .

To define a model in this model structure, not only the parameters need to be determined but also the orders of the sub-models, and the type of basis function expansion in  $f$ .

## 2.2 Data

For the estimation of the parameters in the model (1) a data set is assumed to be available,  $\{u(t), y(t)\}_{t=1}^N$  of  $N$  input  $u(t)$  and output  $y(t)$  samples.

## 2.3 Estimation

A standard prediction error approach is assumed to be used to define the estimate  $\hat{\theta}_N$  of the parameter vector  $\theta$  for the model (1) based on the data set  $\{u(t), y(t)\}_{t=1}^N$ . It is based on minimizing the prediction error

$$\varepsilon(t, \theta) = y(t) - \hat{y}(t, \theta), \quad (5)$$

i.e., the difference between the measured output  $y(t)$  and the prediction according to (1). This is done by using a criterion of fit

$$V_N(\theta) = \frac{1}{N} \sum_{t=1}^N \varepsilon^2(t, \theta) \quad (6)$$

and then defining the estimate as

$$\hat{\theta}_N = \arg \min_{\theta} V_N(\theta). \quad (7)$$

Other criteria than the sum of squared errors can be used and in, eg, [9] it is explained how  $V_N(\theta)$  can be defined so that the Maximum Likelihood estimate is obtained.

After defining estimate (7), it remains to compute it. This must be done using a gradient based iterative algorithm since the model is not a linear regression model. That is, given a start value  $\theta^{(0)}$ , iterate

$$\theta^{(i+1)} = \theta^{(i)} - R_i \frac{dV_N(\theta)}{d\theta} \quad (8)$$

until convergence. The matrix  $R_i$  is to modify the search direction and the step size in order to assure downhill steps. Depending on how  $R_i$  is chosen, (8) describes a wide class of well-known standard algorithms like Gauss-Newton and Levenberg-Marquardt algorithms. Also the algorithm used in [23] can be obtained by choosing  $R_i$  to be the Gauss-Newton approximation with some of the smallest eigenvalues truncated.

All three blocks of the model structure (1) contain a gain parameter, and two of them are typically fixed in the iterative minimization, e.g.,  $b_0^1$  and  $b_0^2$ .

Typically,  $V_N(\theta)$  can have many minima and the success of the minimization depends on the initial estimate  $\theta^{(0)}$ . The considered algorithms deliver such initial estimates and on the benchmark example it is shown that among many local minima they find good ones.

### 3 The Algorithms

Here, the main steps of the considered algorithms are given, for more detailed explanations, see [8] and [15,16].

#### 3.1 Best Linear Approximation

Both algorithms start by estimating a linear model. Theory for linear system identification is a fairly mature area, well covered in books like, e.g., [9,17], focusing on time domain methods and [11], focusing on frequency domain methods.

Only the plant model is of interest. One can, hence, constrain to evaluate output error models. If Box-Jenkins models, or ARMAX models are used, only the estimated plant model is retained. Also frequency domain methods can be used to obtain the model. In that case a non-parametric noise weighting can be used to improve the quality of the initial estimate. Hence, this step gives a linear model described as

$$\hat{G}_{BLA}(q, \theta) = \frac{B(q)}{A(q)} = \frac{\sum_{i=0}^{n_b} b_i q^{-i}}{\sum_{j=0}^{n_a} a_j q^{-j}} \quad (9)$$

where  $a_0 = 1$ .

#### 3.2 Initialization of the Model Structure

This is the step where the two algorithms differ. The two following sections describe the initialization for each of the algorithms.

#### 3.3 Fit All Parameters Using Prediction Error

The final step of both algorithms is to apply the iterative minimization (8) to all parameters.

#### 4 Initialization: Algorithm 1

The algorithm consists of the following steps.

1. Split  $\hat{G}_{BLA}$  model into all possible  $\hat{G}_1(q^{-1})$  and  $\hat{G}_2(q^{-1})$  so that  $\hat{G}_{BLA}(q^{-1}) = \hat{G}_1(q^{-1})\hat{G}_2(q^{-1})$ .

Poles and zeros of the linear model need to be calculated. These are then divided in all possible ways into two sub-models  $\hat{G}_1$  and  $\hat{G}_2$ .

2. For all partitions of the linear model,  $\{\hat{G}_1, \hat{G}_2\}$ , use  $u(t)$  and  $\hat{G}_1$  to decide values for  $\beta^2$  and then LS to fit the linear parameters,  $\beta^1$ , in the nonlinearity as initialization.

The position parameters  $\beta^2$  for the basis functions are decided using the distribution of the input to the nonlinearity  $\{z(t) = \hat{G}_1(q^{-1})u(t)\}_{t=1}^N$ .

Minimizing (6) with respect to the parameters  $\beta^1$  in (4) is straightforward by first writing (1),  $\hat{y}(t, \theta)$ ,

$$\hat{y}(t, \theta) = \sum_{k=1}^n \beta_k^1 \hat{G}_2(q^{-1}, \gamma) f_k(\beta_k^2, z(t)) = \beta^{1T} \varphi(t) \quad (10)$$

where

$$\varphi^T(t) = [\hat{G}_2(q^{-1}, \gamma) f_0(\beta_0^2, z(t)), \dots, \hat{G}_2(q^{-1}, \gamma) f_n(\beta_n^2, z(t))]. \quad (11)$$

Since (10) is a linear regression, the LS estimate is given by

$$\hat{\beta}^1 = \left( \frac{1}{N} \sum_{t=1}^N \varphi(t) \varphi^T(t) \right)^{-1} \frac{1}{N} \sum_{t=1}^N \varphi^T(t) y(t). \quad (12)$$

3. Order the initialized models with respect to their initial fit and select the best one.

This means that  $V_N(\theta)$ , (6) is calculated for all initializations and the models are ranked using this measure.

#### 5 Initialization: Algorithm 2

Now, instead of considering all possible splits of the poles and zeros of  $G_{BLA}$  into the two linear blocks, as in Algorithm 1, all poles and zeros will initially



be “offered” to both linear blocks, and then it will be estimated which ones are needed.

### 5.1 Constructing the basis functions from $G_{BLA}$

Estimates of the internal signals  $z$  and  $x$  can be expressed as

$$\begin{aligned} z(t) &= G_1 u(t) \\ x(t) &= G_2^{-1} y(t). \end{aligned} \tag{13}$$

Further, assuming that the static nonlinearity can be written as a concatenation of two static nonlinear functions  $f_1$  and  $f_2$  such that

$$f(x) = f_2^{-1}(f_1(x)). \tag{14}$$

Then, neglecting the influence of disturbances one obtains

$$f_1(G_1 u(t)) = f_2(G_2^{-1} y(t)) \tag{15}$$

and this algorithm is based on the fact that (15) can be approximately solved as a linear-in-the-parameters TLS problem.  $G_2$  is inverted in the second subsystem, since this system is approached from the output  $y$ . The aim is to write  $\hat{G}_1$  and  $\hat{G}_2^{-1}$  as a linear combination of basis functions containing the poles and the zeros of  $G_{BLA}$ , respectively.

From the partial fraction expansion of  $\hat{G}_{BLA}(z, \theta)$  and  $\hat{G}_{BLA}^{-1}(z, \theta)$ , the following basis functions for  $\hat{G}_1$  and  $\hat{G}_2^{-1}$  are deduced, respectively,

$$\hat{G}_1 : \{W_i\}_{i=1}^r = \left\{ \frac{1}{1 - \rho_i q^{-1}}, q^{-j} \right\} \tag{16}$$

$$\hat{G}_2^{-1} : \{H_i\}_{i=1}^s = \left\{ \frac{1}{1 - \eta_k q^{-1}}, q^{-j} \right\} \tag{17}$$

with  $\rho_i$  the poles and  $\eta_k$  the zeros of the BLA, (9), and  $i = 0, \dots, n_a$ ,  $k = 0, \dots, n_b$ , and  $j = 0, \dots, \max(n_a, n_b)$ . For convenience, the symbols  $W_i$  and  $H_i$  for the basis functions are numbered with a single index up to  $r$  and  $s$  which definitions follows from the definition of the basis functions. It is important to note that in both sets of basis functions extra delay terms are included up to the maximum order which was used to estimate  $G_{BLA}$  parametrically. It can easily be shown that generally it is not sufficient to take only the basis functions deduced from the partial fraction expansion.

Note that for simplicity we restrict ourselves to simple poles/zeros, in order to obtain a real-valued estimate for  $G_1$  and  $G_2^{-1}$ . The basis functions with

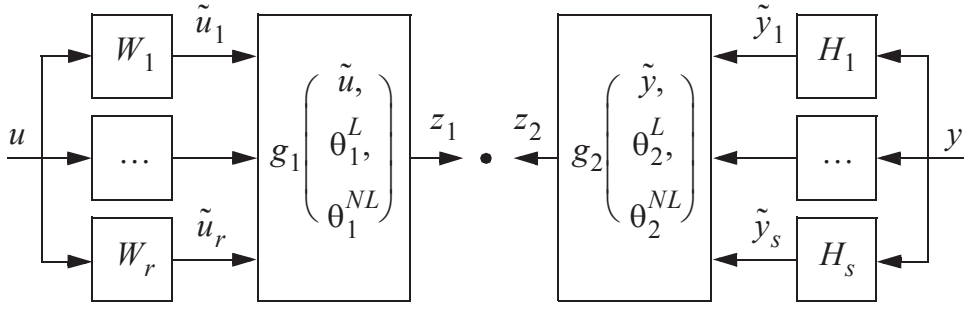


Fig. 2. Model structure leading to a problem linear-in-the-parameters: each subsystem consists of basis functions and a multiple input, single output static nonlinearity. complex conjugate poles/zeros are merged together, resulting in a second order fraction.

To summarize,  $\hat{G}_1$  and  $\hat{G}_2^{-1}$  are written as a linear combination of first and (possibly) second order fractions and pure delay terms,

$$\begin{cases} \hat{G}_1 = \sum_{i=1}^r \hat{\theta}_1^L(i) W_i \\ \hat{G}_2^{-1} = \sum_{i=1}^s \hat{\theta}_2^L(i) H_i \end{cases} \quad (18)$$

with  $W_i$  and  $H_i$  representing the basis functions as described in (16) and (17), respectively. The goal is now to find the proper coefficients  $\theta^L$  of these basis functions.

## 5.2 Solving a problem linear-in-the-parameters

We now consider the model structure in Figure 2 to model the relation in (15). By making a proper choice for the nonlinearities  $g_1$  and  $g_2$ , this model structure leads to a problem that is linear-in-the-parameters ( $\theta^L, \theta^{NL}$ ) which can easily be solved.

To obtain expressions for  $z_1$  and  $z_2$ , the nonlinearities  $g_1$  and  $g_2$  in Figure 2 are approximated by a multi-variable polynomial consisting of a linear and a nonlinear part:

$$\begin{aligned} g_1(\tilde{u}, \theta_1^L, \theta_1^{NL}) &\approx P^L(\tilde{u})\theta_1^L + P_d^{NL}(\tilde{u})\theta_1^{NL} \\ g_2(\tilde{y}, \theta_2^L, \theta_2^{NL}) &\approx P^L(\tilde{y})\theta_2^L + P_d^{NL}(\tilde{y})\theta_2^{NL} \end{aligned} \quad (19)$$

with  $\tilde{u} = [\tilde{u}_1, \dots, \tilde{u}_r]$  and  $\tilde{y} = [\tilde{y}_1, \dots, \tilde{y}_s]$ , the output of the first and second set of basis functions, respectively. The column vectors  $\theta_i^L$  and  $\theta_i^{NL}$  (with  $i = 1, 2$ )

are the unknown parameters, corresponding to the linear and nonlinear part of the multi-variable polynomial, respectively. The row vector  $P_d^{NL}(\cdot)$  consists of all the distinct nonlinear monomials up to a certain degree  $d > 1$  which is chosen by the user.

Using the following definitions

$$\theta_i = \begin{bmatrix} \theta_i^L \\ \theta_i^{NL} \end{bmatrix} \quad i = 1, 2 \quad (20)$$

and

$$P_{\tilde{x}} = \begin{bmatrix} P^L(\tilde{x}) & P_d^{NL}(\tilde{x}) \end{bmatrix} \quad \tilde{x} = \tilde{u}, \tilde{y} \quad (21)$$

the equality requirement  $z_1 = z_2$  can be formulated as a TLS problem [20],

$$\begin{bmatrix} P_{\tilde{u}} & -P_{\tilde{y}} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = P\theta = 0 \quad (22)$$

with  $P = [P_{\tilde{u}} \quad -P_{\tilde{y}}]$  and  $\theta = [\theta_1 ; \theta_2]$ . The unknown parameters  $\theta$  are estimated by performing a Singular Value Decomposition (SVD) [7] of the total regressor matrix  $P$ . The parameter estimate  $\hat{\theta}$  is then given by the right singular vector corresponding to the smallest singular value of  $P$ .

### 5.3 Orthogonality in the regressor matrix $P$

In practice, the nonlinear terms in (19) will also contribute to the linear blocks  $\hat{G}_1$  and  $\hat{G}_2^{-1}$  in (18) via their contribution to the best linear approximation. Indeed, a part of the linear system information is captured by the coefficients  $\theta_i^{NL}$  of the nonlinear regressors. In order to concentrate the linear behavior in the coefficients corresponding to the linear regressors, the nonlinear regressors in the matrices  $P_d^{NL}(\tilde{u})$  and  $P_d^{NL}(\tilde{y})$  are made orthogonal to the linear regressors in  $P^L(\tilde{u})$  and  $P^L(\tilde{y})$ , respectively. This issue is tackled numerically using a thin QR decomposition [7], prior to the SVD, on the regressor matrix corresponding to the left and the right subsystem separately (i.e.,  $P_{\tilde{u}}$  and  $P_{\tilde{y}}$ , respectively).

#### 5.4 Composing the initial estimates

The estimated parameters  $\hat{\theta}^L$  are the coefficients of the linear basis functions  $W_i$  and  $H_i$ . Consequently,  $\hat{G}_1$  and  $\hat{G}_2$  can be composed parametrically to the form (1) by calculating the linear combination of the basis functions in (18). We can also calculate a non-parametric estimate of the system's intermediate signals (up to a scale factor),

$$\begin{cases} \hat{z} = P^L(\tilde{u})\hat{\theta}_1^L \\ \hat{x} = P^L(\tilde{y})\hat{\theta}_2^L. \end{cases} \quad (23)$$

Between  $\hat{z}$  and  $\hat{x}$ , the same nonlinear relation exists as between the true internal signals  $z$  and  $x$  (again, up to a scale factor). To obtain also a parametric initial estimate for the static nonlinearity  $f(\cdot)$ , any basis function expansion of the form (4) can be fit to the  $\{\hat{z}, \hat{x}\}$  data.

## 6 Experimental Results

The presented identification procedures are now applied to input/output measurements obtained from the benchmark data and evaluation follows the specifications in [13].

### 6.1 Description of the System

The device under test is an electric nonlinear circuit with a Wiener-Hammerstein structure (see Figure 3), designed by Gerd Vandersteen [21]. The system is composed of a static nonlinear block, sandwiched between two linear dynamic blocks. The first linear dynamic system is a third order Chebyshev low-pass filter with a 0.5 dB ripple and a pass-band up to 4.4 kHz. The static nonlinearity is realized by two resistors and a diode. The second linear dynamic system is a third order inverse Chebyshev low-pass filter with a -40 dB stop-band, starting at 5 kHz. This filter is designed to have a zero in the frequency band of interest.

### 6.2 Description of the Data

The system was excited with a filtered Gaussian excitation signal with a cut-off frequency of 10 kHz. This noise sequence consisted of  $N = 188000$  samples

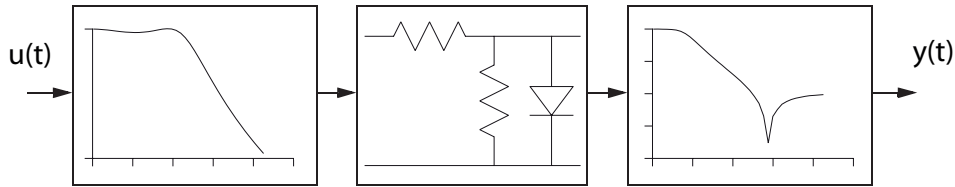


Fig. 3. Experimental Wiener-Hammerstein system.

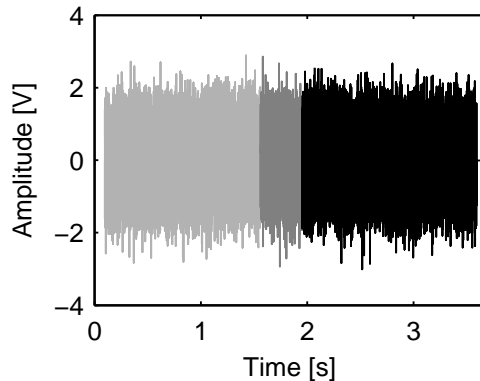


Fig. 4. Gaussian noise excitation signal containing an estimating set (light grey) , a validation set (dark grey) and a benchmark test set (black).

as shown in Figure 4. The input and output signal were measured with a sample frequency of 51.2 kHz. The output signal had an RMS value of 242.3 mV.

We discarded the first 5000 data samples of the input since they only consist of quantization noise; no excitation was present here. The remaining part is split into two data sets: the estimation data ( $N = 5001, \dots, 100000$ ) and the test data ( $N = 100001, \dots, 188000$ ) which is used to benchmark the quality of the identified Wiener-Hammerstein models.

### 6.3 Best Linear Model

Best linear model, obtained by standard system identification algorithms, is a 6th order model. It gives an RMSE of 56.2 mV on the raw test data, and an RMSE of 43.7 mV if the DC-offset is first removed from the test data. The poles and zeros of the model are depicted in Figure 5. Two pairs of zeros lay close to the unit circle and potentially they can both explain the transmission zero. The placement of the poles shows the low pass character of the model.

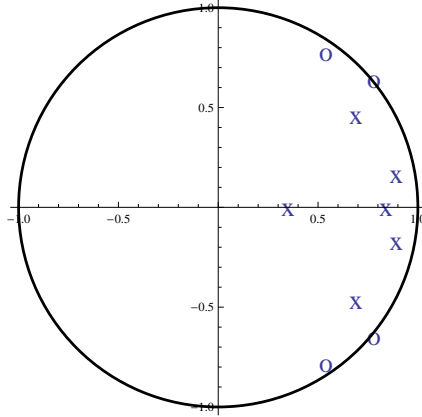


Fig. 5. Poles and zeros of the initial 6th order linear model. One zero, at 3.5, is outside the plot.

#### 6.4 Algorithm 1

The linear model is now split into two sub-models in all possible ways with the constraint that the order of each linear sub-model should be at least one, and it should be proper. This gives 42 pairs of linear sub-models. For each pair of sub-models a first order spline with eight knots is initialized between the linear subsystems. This gives a local linear nonlinearity with nine segments. The knot position, corresponding to the parameters  $\beta^2$  in (4), are chosen so that equally many data points are placed in each segment. The linear spline parameters,  $\beta^1$  in (4), are then fitted using least squares for all of the models. The resulting fit of the 42 initialized models is depicted in Figure 6 with dots marks. The best initialized model obtained RMSE 6.4 mV, more than 6 times better than the linear model.

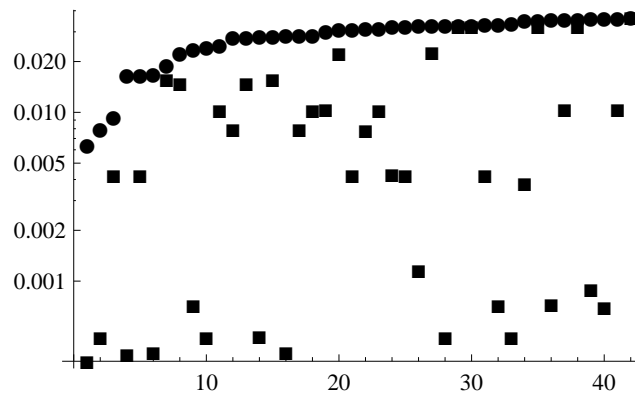


Fig. 6. RMSE computed on test data for all 42 initialized models sorted according to the fit after initialization. The points indicates the RMSE (in Volts) after initialization and the squares indicates the value after that all parameters have been fitted.

Normally the best initialized model is selected and only for that one, all parameters are fitted simultaneously. Here, for illustrative reasons the general prediction error minimization is applied to all the parameters in the models, ie, the equations in Section 2.3 are applied to all 42 models. In (8) a Levenberg-Marquardt algorithm is used, and the implementation is described in [14]. Figure 6 shows the fit, RMSE, after training with large squares. For each of the 42 partitions of the BLA model, the fit after initialization, and after fitting all parameters are shown. The best initialized model also gives the best fit after minimizing with respect to all parameters, with RMSE 0.31 mV. It can also be seen in the figure that the final fit is very different for some models with similar fit at the initializations. This happens when the linear blocks contain the wrong number of poles and zeros since this cannot be compensated for by adapting all parameters.

Consider now the quality of the initialized linear blocks. Figure 7 shows this for the best initialized model, and it is also shown how these change after that all parameters have been fitted. The poles in the first linear part have changed, but the ones in the second part seem to have been quite accurate from the beginning. Also, only one zero pair remains close to the unit circle, responsible for the transmission zero.

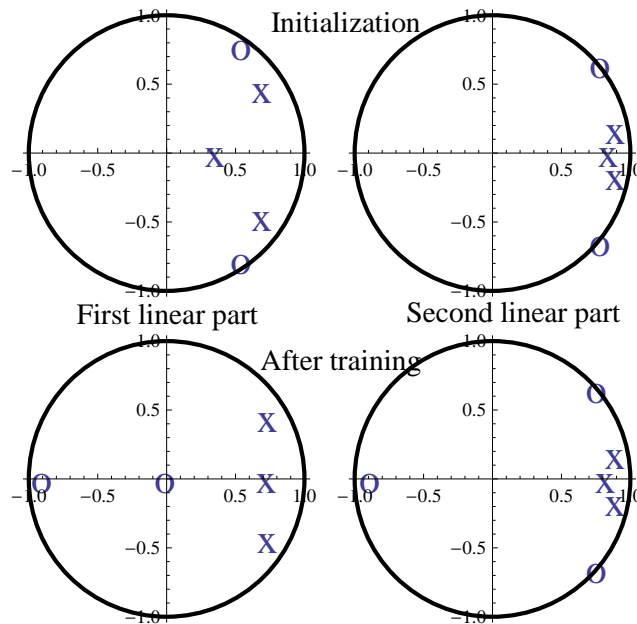


Fig. 7. Upper two plots: Partitioning of the poles (x) and the zeros (o) of the linear model which gave the best initialization. Lower two plots: Positions of the poles and zeros after that all parameters have been fitted.

Are there any local minima? Yes, this can be seen by looking at a close up of Figure 6 for the 15 best models of the original 42 models. It is shown shown in Figure 8. Apparently, all of these 15 models represent different local minima

although they all have impressive good fit on the data. Hence, there are plenty of local minima.

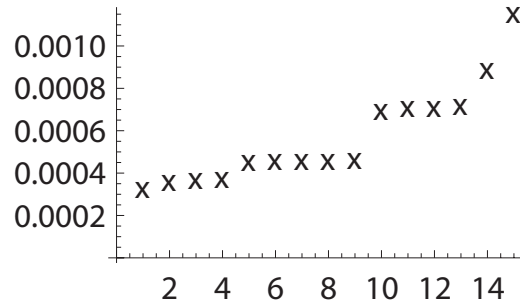


Fig. 8. RMSE of the 15 best models.

Including more flexibility in the nonlinear block improved the fit slightly but without any remarkable change. The RMSE decreased from 31 for 8-knots to 27 for the 24 knot-nonlinearity on the test data. In Figure 9 the 24 knot-nonlinearity is shown together with the distribution of the input data to the nonlinearity. In that plot it is also illustrated how the input data to the nonlinearity is distributed.

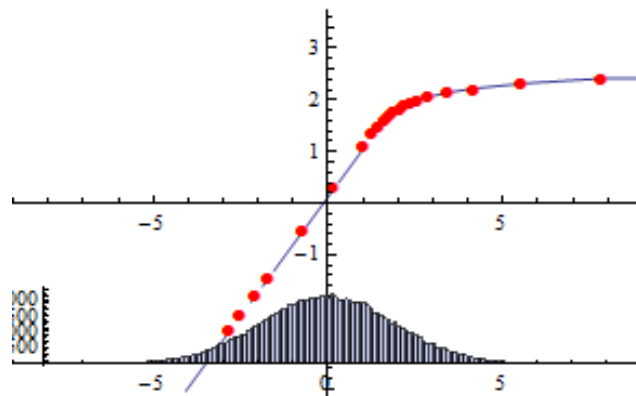


Fig. 9. Estimated nonlinearity, final model with 24 knots together with the distribution of input data to the nonlinearity.

The best 8-knot model contains 30 parameters, 17 of them for the nonlinear spline. The 24 knot-nonlinearity contains 62 parameters, of them 49 in the nonlinearity.

### 6.5 Algorithm 2

The algorithm starts from the linear model described in Section 6.3. The static nonlinearities  $g_1$  and  $g_2$  are approximated by a multi-variable polynomial of degree 3 ( $d = 3$  in (19)). For the parametrisation of the static nonlinearity, a



median method [12] was used. In this method, the  $\hat{z}, \hat{x}$  data is divided into 20 vertical slices such that each slice consists of an equal amount of data samples. Per slice, the median value of the data is calculated (in the  $\hat{z}$  direction). Next, a spline is fit through the calculated median values in order to parameterize the static nonlinearity.

To reduce the number of regressors in the initialization method, the assumption was made that in each linear dynamic block at least one pole/zero is present. As a consequence, in the sets of basis functions delay terms  $q^{-j}$  are included only up to  $j = \max(n_b, n_a) - 1$  instead of  $j = \max(n_b, n_a)$ . This decreases the number of linear regressors by two (and hence the number of nonlinear regressors), without compromising on the flexibility of the model.

### 6.6 Reduction of linear regressors

In order to reduce the number of model parameters, linear basis functions can be removed from (18). Unfortunately, it is unknown in advance which poles, zeros or delay terms should remain in the model. To tackle this problem, a scan is performed to verify the effect of removing a basis function (and all related nonlinear regressors). During the first scan run, each basis function is alternately removed, and the best performing model (in RMS sense) is then selected. A second scan run is then performed to find the following regressor that should be omitted, and so on. After each scan run, a linear regressor is permanently deleted, resulting in an initialized WH model. All these models are then ranked with respect to their RMSE.

The RMSE of the obtained initialized WH models as a function of the number of discarded linear regressors is shown in Figure 10 using the 6th order linear model as a starting point in the initialization. It can be seen that up to 12 linear regressors can be discarded without affecting the RMSE of the model too much.

### 6.7 Nonlinear Optimization

Next, a selection of the initialized models given in Figure 10 are optimized, in the quest for the best nonlinear model.

Up to now, the static nonlinearity was parameterized using the median method, followed by a spline fit. In the nonlinear optimization, another parameterization for the static nonlinearity is considered: a piece-wise linear approximation realized by so-called hinge functions [23] of various knots (4 to 8). Furthermore, the order of the numerator and denominator of the linear dynamic blocks is

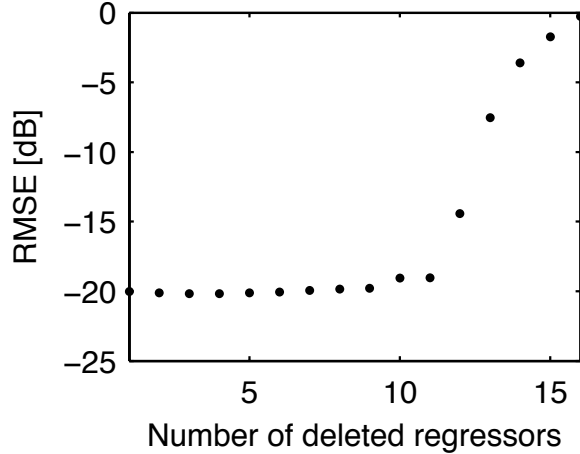


Fig. 10. RMSE of the initialized models as a function of the number of deleted linear regressors, using the 6th order linear model.

always set equal, in order to give the model enough flexibility. For instance, if the initialization method yields a 2/4 block (i.e.,  $n_b = 2$  and  $n_a = 4$ ), then a 4/4 block (with zero coefficients where necessary) will be given as an input to the optimization algorithm. In practice, it turned out that this flexibility yielded much better results at the cost of higher model orders for the linear blocks.

Given the different options described above, a number of candidate nonlinear models are obtained.

The best nonlinear model was obtained when deleting 8 linear regressors and using hinge functions with 8 knots for the parameterization of the static non-linearity. This model has a validation RMSE of 0.30 mV and contains 64 model parameters. The simulation error (light grey) of this model is shown in Figure 11, together with the modeled output signal (black).

Figure 12 shows the spectra of the modeled output signal (black), the linear simulation error (dark grey), and the nonlinear simulation error (light grey). In the pass-band of the DUT, the nonlinear model error is more than 20 dB lower than the linear model error.

### 6.8 Discussion

The best obtained nonlinear models resulting from the two initialization algorithms have an RMSE around 0.30 mV on the test data. Although this is an impressive improvement compared to the linear model, this is still about 30% above the noise level, estimated to about RMS 0.19 mV on the first part of the data where the input signal is constant.

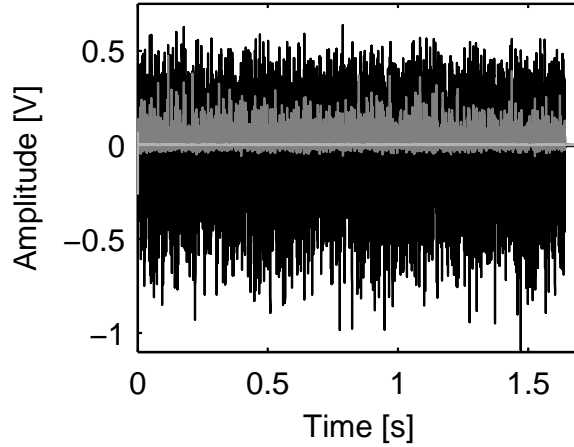


Fig. 11. Benchmark validation result: modeled output (black) and model error of the best linear model (dark grey) and model error of the best nonlinear model (light grey). All good models gave a result like this.

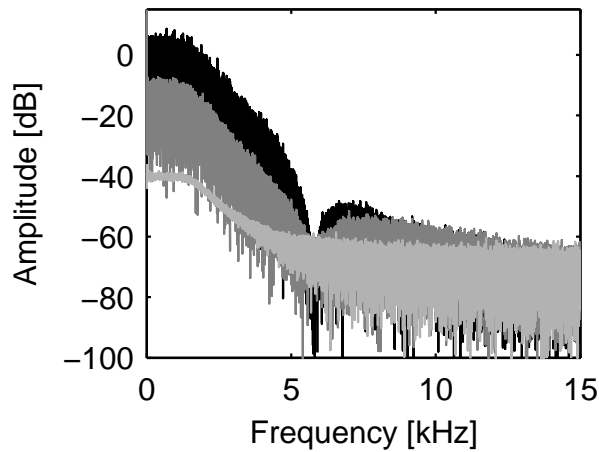


Fig. 12. DFT spectra of the modeled output signal (black); linear model error (dark grey); and nonlinear model error (light grey). All good models gave a result like this.

The performance of these models can be compared to the best results at SYSID'09. In [10] a best fit  $e_{RMSt} = 0.42\text{mV}$  was obtained with a polynomial state space model containing 797 parameters. The same structure as used in this paper was also used in [23] and they obtained an RMSE 0.49 mV. However, they did not have any initialization algorithm for the parameters so their solution is probably a local minimum. They also study how the solution converges to different minima depending on the randomly chosen initialization.

## 7 Conclusions

Two recently suggested initialization algorithms for Wiener-Hammerstein models have been applied to the benchmark problem from SYSID 2009. They both start with a linear model and then divide the dynamics into two linear blocks in different ways. Given good initializations of the linear blocks, it is straight-forward to obtain an initial estimate of the static nonlinear block.

The result of the two algorithms have been compared and they both give good modelling results. It has been shown that there are many local minima corresponding to good models. However, one of them is better than the other, and both algorithms find that minimum. Still, of course, there is no guarantee that the global minimum has been found.

## Acknowledgments

This work was funded by the Fund for Scientific Research (FWO-Vlaanderen), the Methusalem grant of the Flemish Government (METH-1), and the Belgian program on interuniversity poles of attraction initiated by the Belgian State, Prime Minister's Office, Science Policy programming (IUAP-VI/4).

## References

- [1] N. J. Bershad, P. Celka, and S. McLaughlin. Analysis of stochastic gradient identification of Wiener-Hammerstein systems for nonlinearities with hermite polynomial expansions. *IEEE Transactions on Signal Processing*, 49(5):1060–1072, 2001.
- [2] S. A. Billings and S. Y. Fakhouri. Identification of systems containing linear dynamic and static non-linear elements. *Automatica*, 18(1):15–26, 1982.
- [3] H. Boutayeb and M. Darouach. Recursive identification method for miso Wiener-Hammerstein model. *IEEE Transactions on Automatic Control*, 40(2):287–291, 1995.
- [4] D. R. Brillinger. Identification of a particular nonlinear time-series system. *Biometrika*, 64(3):509–515, 1977.
- [5] C. H. Chen and S. D. Fassois. Maximum-likelihood identification of stochastic Wiener-Hammerstein-type nonlinear systems. *Mechanical Systems and Signal Processing*, 6(2):135–153, 1992.

- [6] P. Crama and J. Schoukens. Computing an initial estimate of a Wiener-Hammerstein system with a random phase multisine excitation. *IEEE Transactions on Instrumentation and Measurement*, 54(1):117–122, 2005.
- [7] G. H. Golub and C. F. Van Loan. *Matrix Computations*. John Hopkins University Press, Baltimore, 3rd edition, 1996.
- [8] L. Lauwers. *Some practical applications of the best linear approximation in nonlinear block-oriented modelling*. PhD thesis, ELEC, Vrije Universiteit Brussels, 2011.
- [9] L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ, 2nd edition, 1999.
- [10] J. Paduart, L. Lauwers, R. Pintelo, and J. Schoukens. Identification of a Wiener-Hammerstein systems using the polynomial nonlinear state space approach. In *Preprint, 15th IFAC Symposium on System Identification Saint-Malo, France*, pages 1080–1085, July 2009.
- [11] R. Pintelon and J. Schoukens. *System Identification: A Frequency Domain Approach*. IEEE-press, Piscataway, 2001.
- [12] J. Schoukens, R. Pintelon, J. Paduart, and G. Vandersteen. Nonparametric initial estimates for Wiener-Hammerstein systems. In *14th IFAC Symposium System Identification (SYSID), Newcastle, Australia, 29-31 March 2006*, pages 1033–1037, 2006b.
- [13] J. Schoukens, J. Suykens, and L. Ljung. Wiener-Hammerstein benchmark. (SYSID 2009 special session, <http://wwwtw.vub.ac.be/elec/sysid09.htm>, 2009.
- [14] J. Sjöberg and H. Hjalmarsson. A system, signal, and identification toolbox in Mathematica with symbolic capabilities. In *Preprint, 15th IFAC Symposium on System Identification Saint-Malo, France*, pages 747–751, July 2009.
- [15] J. Sjöberg and J. Schoukens. Initializing Wiener-Hammerstein models based on partitioning of the best linear approximation. In *Conference: 18th IFAC World Congress, Milano, Italy*, 2011.
- [16] J. Sjöberg and J. Schoukens. Initializing Wiener-Hammerstein models based on partitioning of the best linear approximation. *Automatica*, 48(2):353–359, 2012.
- [17] T. Söderström and P. Stoica. *System Identification*. Prentice-Hall International, Hemel Hempstead, Hertfordshire, 1989.
- [18] A. H. Tan and K. Godfrey. Identification of Wiener-Hammerstein models using linear interpolation in the frequency domain (lifred). *IEEE Transactions on Instrumentation and Measurement*, 51(3):509–521, 2002.
- [19] A. H. Tan and K. Godfrey. Identification of Wiener-Hammerstein models with cubic nonlinearity using lifred. In *Proc. of 13th IFAC Symposium on System Identification*, pages 1339–1344, Rotterdam, 2003.

- [20] S. Van Huffel and J. Vandewalle. *The Total Least Squares Problem: Computational Aspects and Analysis*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1991.
- [21] G. Vandersteen. Identification of linear and nonlinear systems in an errors-in-variables least squares and total least squares framework. Master's thesis, Vrije Universiteit Brussel, Brussels, Belgium, 1997.
- [22] G. Vandersteen, Y. Rolain, and J. Schoukens. Non-parametric estimation of the frequency-response functions of the linear blocks of a Wiener-Hammerstein model. *Automatica*, 33(7):1351–1355, 1997.
- [23] A. Wills and B. Ninness. Estimation of generalised Hammerstein-Wiener systems. In *Preprint, 15th IFAC Symposium on System Identification, Saint-Malo, France*, pages 1104–1109, July 2009.