Identification of Multivariable Hammerstein Systems using Rational Orthonormal Bases

Juan C. Gómez
Departamento de Electrónica
FCEIyA - Universidad Nacional de Rosario
Riobamba 245 Bis, 2000 Rosario
Argentina
e-mail:jgomez@eie.fceia.unr.edu.ar

Enrique Baeyens
Depto. de Ing. de Sistemas y Automática
ETSII - Universidad de Valladolid
Paseo del Cauce s/n, 47011 Valladolid
España
e-mail: enrbae@eis.uva.es

Abstract

In this paper, a non iterative algorithm for the simultaneous identification of the linear and nonlinear parts of multivariable Hammerstein systems is presented. The proposed algorithm is numerically robust, since it is based only on least squares estimation and singular value decomposition. Under weak assumptions on the persistency of excitation of the inputs, the algorithm provides consistent estimates even in the presence of coloured noise. Key in the derivation of the results is the use of rational orthonormal bases for the representation of the linear part of the system. An additional advantage of this is the possibility of incorporating prior information about the system in a typically *black-box* identification scheme.

1 Introduction

In the last decades, many research activities have been carried out on modelling, identification, and control design of nonlinear systems. Many dynamical systems can be better represented by nonlinear models, which are able to describe the global behaviour of the system over the whole operating range, rather than by linear ones that are only able to approximate the system around a given operating point. One of the most frequently studied nonlinear models is the Hammerstein model, which consists of a static (memoryless) nonlinearity followed by a linear time-invariant (LTI) system (see for instance [17], [16], and [15] for analysis, synthesis and structural theory of cascade interconnections of LTI and zero-memory nonlinear subsystems). These models have been successfully used to represent nonlinear systems in a number of practical applications in the areas of chemical processes [7], biological processes [10], signal processing, communications, and control.

Several techniques have been proposed in the literature for the identification of Hammerstein models. The reader is referred to [12, 4, 3, 2, 5, 19, 7, 6], and the references therein. For the purpose of putting into context the present work, we distinguish three parametric approaches for the identification of Hammerstein models. The first one is the traditional iterative algorithm proposed by Narendra and Gallman in [12]. In this algorithm, an appropriate parameterization of the system allows the prediction error to be separately linear in each set of parameters characterizing the linear and the nonlinear parts. The estimation is then carried out by minimizing alternatively with respect to each set of parameters, a quadratic criterion on the prediction errors. An analytical counterexample by Stoica [19] showed that the original algorithm could be divergent in some particular cases. A second approach, based on correlation techniques, is introduced in [4, 3, 2]. This method relies on a separation principle, but with the rather restrictive requirement on the input to be white noise. A more recent approach for the identification of Hammerstein-Wiener systems has been introduced by Bai in [1]. This algorithm is based on least squares estimation and singular value decomposition, however it only applies to the single-input/single-output (SISO) case, and consistency of the estimates can only be assured for the case of the disturbances being white noise, or in the noise-free case.

In this paper, a noniterative algorithm for the identification of **multivariable** Hammerstein models is presented. As in [1], the computational tools employed by the algorithm are Least Squares Estimation (LSE) and Singular Value Decomposition (SVD), which results in numerical robustness under very weak assumptions on the persistency of excitation of the inputs[1]. However, in contrast to [1], the algorithm presented here also applies to the multivariable case, and in addition, consistency of the estimates can be assured even in the presence of coloured noise. Key on the derivations of these results is the use of orthonormal basis functions for the representation of the linear part of the Hammerstein model.

[1] This is actually not a restriction, since it is clear that any identification algorithm requires some degree of persistency of excitation of the inputs. One can only identify the system modes that are sufficiently excited by the input and can be seen at the output.

In recent years, there has been a lot of research on the issue of how to introduce 'a priori' information in the identification of *black box* LTI model structures. A natural answer to this problem has been the use of rational orthonormal bases for the representation of the system. Choosing the poles of the bases closed to the (approximately known) system poles the accuracy of the estimate can be considerably improved (see [9] for a detailed review of the use of Orthonormal Bases in Identification of LTI Systems). It is not intended to give here a complete overview on Identification using Rational Orthonormal Bases, and the reader is referred to [9, 13, 21, 22, 23, 20], and the references therein. An advantage of using orthonormal bases to model LTI systems is that the input-output equation can be written as a linear regression. As a consequence, a parameter estimate can be obtained in closed form by minimizing a quadratic criterion on the prediction errors (viz, the Least Squares Estimate). In addition, since the regressors only depend on past inputs, the estimate is consistent even if the output is corrupted by coloured noise, under the assumption that the actual system belongs to the model class (i.e., there is no undermodelling).

In this paper, basis function expansions are used to represent both the linear and the nonlinear part of the system. This results in a linear regressor form so that least squares techniques can be used to estimate an oversized parameter matrix. Then, by recurring to Singular Value Decomposition and rank reduction, optimal estimates of the parameter matrices characterizing the linear and nonlinear parts can be obtained. In comparison with other works, the proposed algorithm has the following advantages

- It applies to multivariable Hammerstein models.
- It provides consistent estimates even in the presence of coloured noise.
- No special assumptions on the inputs, other than the standard persistency of excitation conditions, are required.

The rest of the paper is organized as follows. In Section 2, the multivariable Hammerstein model is introduced, and the identification problem is formulated. The optimal identification algorithm is derived in Section 3, while a simulation example illustrating the performance of the method is presented in Section 4. Finally, some concluding remarks are provided in Section 5.

2 Problem Formulation

A (multivariable) Hammerstein model is schematically represented in figure 1. The model consists of a zero-memory nonlinear element $\mathcal{N}(\cdot)$ in cascade with a LTI system with transfer function (matrix)[2] $G(q) \in H_2^{m\times n}(\mathbb{T})$. It is assumed that the measured output y_k contains an unknown additive noise component ν_k. The input-output relationship is then given by

$$y_k = G(q)\mathcal{N}(u_k) + \nu_k \tag{1}$$

where $y_k \in \mathbb{R}^m$, $u_k \in \mathbb{R}^n$, and $\nu_k \in \mathbb{R}^m$, are the system output, input, and measurement noise vectors at time k, respectively. It will be assumed that the nonlinear block can be described as

$$\mathcal{N}(u_k) = \sum_{i=1}^{r} a_i g_i(u_k) \tag{2}$$

where $g_i(\cdot) : \mathbb{R}^n \to \mathbb{R}^n$, $(i = 1, \cdots, r)$, are known smooth vector fields, and $a_i \in \mathbb{R}^{n\times n}$, $(i = 1, \cdots, r)$, are unknown matrix parameters.

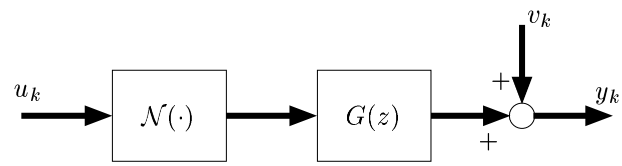

Figure 1: Hammerstein Model.

On the other hand, the LTI system will be represented using rational orthonormal bases as follows

$$G(q) = \sum_{\ell=0}^{p-1} b_\ell \mathcal{B}_\ell(q) \tag{3}$$

where $b_\ell \in \mathbb{R}^{m\times n}$ are unknown matrix parameters, and $\{\mathcal{B}_\ell(q)\}_{\ell=0}^{\infty}$ are rational orthonormal bases[3] on $H_2(\mathbb{T})$.

The identification problem is to estimate the unknown parameter matrices a_i, $(i = 1, \cdots, r)$, and b_ℓ, $(\ell = 0, \cdots, p-1)$, characterizing the nonlinear and the linear parts, respectively, from an N-point data set $\{u_k, y_k\}_{k=1}^{N}$ of observed input-output measurements.

In the following section, an optimal identification algorithm is presented, which is based only in Least Squares Estimation (LSE) and Singular Value Decomposition (SVD). Under weak conditions the algorithm delivers unbiased estimates of the parameter matrices.

[2] Here, q stands for the forward shift operator, and $H_2^{m\times n}(\mathbb{T})$ is the Hardy space of $(m\times n)$ transfer matrices whose elements are in $H_2(\mathbb{T})$, the Hardy space of functions that are square integrable on the unit circle $\mathbb{T}$, and analytic outside the unit disk. With some abuse of terminology we will refer to $H_2^{m\times n}(\mathbb{T})$ as the space of all stable, causal, discrete-time, $(m \times n)$ transfer matrices.

[3] The bases are orthonormal in the sense that

$$\langle \mathcal{B}_\ell, \mathcal{B}_k \rangle = \delta_{\ell k},$$

where $\delta_{\ell k}$ is the Kronecker delta, and $\langle \cdot, \cdot \rangle$ is the standard inner product in $L_2(\mathbb{T})$, defined as

$$\langle \mathcal{B}_\ell, \mathcal{B}_k \rangle \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathcal{B}_\ell(e^{j\omega})\overline{\mathcal{B}_k(e^{j\omega})}\mathrm{d}\omega$$

3 Optimal Estimation Algorithm

Substituting equations (2) and (3) in (1), the input-output relationship can be written as

$$
\begin{aligned}
y_k &= \left(\sum_{\ell=0}^{p-1} b_\ell \mathcal{B}_\ell(q)\right)\left(\sum_{i=1}^{r} a_i g_i(u_k)\right) + \nu_k \qquad (4)\\
&= \sum_{\ell=0}^{p-1}\sum_{i=1}^{r} b_\ell a_i \mathcal{B}_\ell(q) g_i(u_k) + \nu_k \qquad (5)
\end{aligned}
$$

It is clear from equation (5) that the parameterization (2)-(3) is not unique, since any parameter matrices αb_ℓ and $\alpha^{-1}a_i$, for some nonzero scalar α, provide the same input-output equation (5).

In other words, any identification experiment can not distinguish between the parameters (b_ℓ, a_i) and $(\alpha b_\ell, \alpha^{-1}a_i)$. As it is common in the literature [1, 14], these two sets of parameters will be called *equivalent*. To obtain a one-to-one parameterization, i.e. for the system to be identifiable, additional contraints must be imposed on the parameters. A technique that can be used to obtain uniqueness is to normalize the parameter matrices a_i (or b_ℓ), that is to assume that $\|a_i\|_2 = 1$ (or $\|b_\ell\|_2 = 1$). A similar methodology was employed in [1] for a scalar Hammerstein-Wiener model. Under this assumption the parameterization (2)-(3) is unique.

Defining now

$$
\begin{aligned}
\theta &= [b_0 a_1, \cdots, b_0 a_r, \cdots, b_{p-1}a_1, \cdots, b_{p-1}a_r]^T \qquad (6)\\
\phi_k &= \big[\mathcal{B}_0(q) g_1^T(u_k), \cdots, \mathcal{B}_0(q) g_r^T(u_k), \cdots,\\
&\quad \vdots\\
&\quad \mathcal{B}_{p-1}(q) g_1^T(u_k), \cdots, \mathcal{B}_{p-1}(q) g_r^T(u_k)\big]^T, \qquad (7)
\end{aligned}
$$

equation (5) can be written as

$$y_k = \theta^T \phi_k + \nu_k, \qquad (8)$$

which is in linear regression form. Considering the N-point data set, the last equation, and defining

$$
\begin{aligned}
Y_N &= \left[y_1^T, y_2^T, \cdots, y_N^T\right]^T, \qquad (9)\\
V_N &= \left[\nu_1^T, \nu_2^T, \cdots, \nu_N^T\right]^T, \qquad (10)\\
\Phi_N &= [\phi_1, \phi_2, \cdots, \phi_N], \qquad (11)
\end{aligned}
$$

the following equation can be written

$$Y_N = \Phi_N^T \theta + V_N. \qquad (12)$$

It is well known [11] that the estimate $\widehat{\theta}$ that minimizes a quadratic criterion on the prediction errors $\epsilon_N = Y_N - \Phi_N^T\theta$ (that is the least squares estimate) is given by

$$\widehat{\theta} = \left(\Phi_N \Phi_N^T\right)^{-1} \Phi_N Y_N = \Phi_N^\dagger Y_N, \qquad (13)$$

provided the indicated inverse exists[4] [18].

The problem is how to estimate the parameter matrices a_i, $(i = 1, \cdots, r)$, and b_ℓ, $(\ell = 0, \cdots, p-1)$ from the estimate $\widehat{\theta}$ in (13).

From the definition of the parameter matrix θ in (6), it is easy to see that

$$\theta = \mathsf{blockvec}\,(\Theta_{ab}), \qquad (14)$$

where $\mathsf{blockvec}\,(\Theta_{ab})$ is the block column matrix obtained by stacking the block columns of Θ_{ab} on top of each other, and where Θ_{ab} has been defined as

$$\Theta_{ab} = \begin{bmatrix} a_1^T b_0^T & a_1^T b_1^T & \cdots & a_1^T b_{p-1}^T \\ a_2^T b_0^T & a_2^T b_1^T & \cdots & a_2^T b_{p-1}^T \\ \vdots & \vdots & \cdots & \vdots \\ a_r^T b_0^T & a_r^T b_1^T & \cdots & a_r^T b_{p-1}^T \end{bmatrix} = ab^T, \qquad (15)$$

with the following definitions for the matrices a and b,

$$
\begin{aligned}
a &= [a_1, a_2, \cdots, a_r]^T, \qquad (16)\\
b &= \left[b_0^T, b_1^T, \cdots, b_{p-1}^T\right]^T. \qquad (17)
\end{aligned}
$$

An estimate $\widehat{\Theta}_{ab}$ of the matrix Θ_{ab} can then be obtained from the estimate $\widehat{\theta}$ in (13). The problem now is how to estimate the parameter matrices a and b from the estimate $\widehat{\Theta}_{ab}$. It is clear that the closest, in the 2-norm[5] sense, estimates $\widehat{a}$ and $\widehat{b}$ are such they minimize the norm

$$\left\| \widehat{\Theta}_{ab} - \widehat{a}\widehat{b}^T \right\|_2^2. \qquad (18)$$

That is,

$$\left(\widehat{a}, \widehat{b}\right) = \arg\min_{a,b} \left\{ \left\| \widehat{\Theta}_{ab} - ab^T \right\|_2^2 \right\}. \qquad (19)$$

The solution to this optimization problem is provided by the Singular Value Decomposition (SVD) [8] of the matrix $\widehat{\Theta}_{ab}$. The result is summarized in the following Theorem.

Theorem 3.1 Let $\widehat{\Theta}_{ab} \in \mathbb{R}^{nr \times mp}$ have rank $k > n$, and let the 'economy-size' SVD of $\widehat{\Theta}_{ab}$ be given by

$$\widehat{\Theta}_{ab} = U_k \Sigma_k V_k^T = \sum_{i=1}^{k} \sigma_i u_i v_i^T \qquad (20)$$

---

[4] The inverse exists, provided that the regressors ϕ_k are persistently exciting (PE) in the sense that there exist some integer ℓ_0, and positive constants α_1 and α_2 such that

$$\alpha_2 I \geq \sum_{k=k_0}^{k_0+\ell_0} \phi_k \phi_k^T \geq \alpha_1 I > 0.$$

[5] The 2-norm of a matrix $A = (a_{ij})_{(m\times n)}$ is the norm induced by the 2-norm (or Euclidean norm) of vectors

$$\|A\|_2 = \sup_{w \neq 0} \frac{\|Aw\|_2}{\|w\|_2}$$

where Σ_k is a diagonal matrix containing the k nonzero singular values ($\sigma_i, i = 1, \cdots, k$) of $\widehat{\Theta}_{ab}$ in nonincreasing order, and where the matrices $U_k = \begin{bmatrix} u_1 & u_2 & \cdots & u_k \end{bmatrix} \in \mathbb{R}^{nr \times k}$ and $V_k = \begin{bmatrix} v_1 & v_2 & \cdots & v_k \end{bmatrix} \in \mathbb{R}^{mp \times k}$ contain only the first k columns of the unitary matrices $U \in \mathbb{R}^{nr \times nr}$ and $V \in \mathbb{R}^{mp \times mp}$ provided by the full SVD of $\widehat{\Theta}_{ab}$,

$$\widehat{\Theta}_{ab} = U \Sigma V^T, \tag{21}$$

respectively[6]. Then, the matrices $\widehat{a} \in \mathbb{R}^{nr \times n}$ and $\widehat{b} \in \mathbb{R}^{mp \times n}$ that minimize the norm $\left\| \widehat{\Theta}_{ab} - ab^T \right\|_2^2$, are given by

$$\left(\widehat{a}, \widehat{b} \right) = \arg\min_{a,b} \left\{ \left\| \widehat{\Theta}_{ab} - ab^T \right\|_2^2 \right\} = \left(U_1, V_1 \Sigma_1 \right), \tag{22}$$

where $U_1 \in \mathbb{R}^{nr \times n}$, $V_1 \in \mathbb{R}^{mp \times n}$, and $\Sigma_1 = \mathsf{diag}\{\sigma_1, \sigma_2, \cdots, \sigma_n\}$ are given by the following partition of the 'economy size' SVD in (20),

$$\widehat{\Theta}_{ab} = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}, \tag{23}$$

and the approximation error is given by

$$\left\| \widehat{\Theta}_{ab} - \widehat{a}\widehat{b}^T \right\|_2^2 = \sigma_{n+1}^2. \tag{24}$$

Proof: See Appendix. ■

Based on this result, the nonlinear identification algorithm can then be summarized as follows.

Algorithm 3.1

Step 1: Compute the least squares estimate $\widehat{\theta}$ as in (13), and the matrix $\widehat{\Theta}_{ab}$ such that

$$\widehat{\theta} = \mathsf{blockvec}\left(\widehat{\Theta}_{ab} \right). \tag{25}$$

Step 2: Compute the 'economy size' SVD of $\widehat{\Theta}_{ab}$ as in Theorem 3.1 , and the partition of this decomposition as in equation (23).

Step 3: Compute the estimates of the parameter matrices a and b as

$$\begin{aligned} \widehat{a} &= U_1, & (26) \\ \widehat{b} &= V_1 \Sigma_1, & (27) \end{aligned}$$

respectively. □

[6] In equation (21), the matrix $\Sigma \in \mathbb{R}^{nr \times mp}$ is given by

$$\Sigma = \begin{bmatrix} \Sigma_{mp} \\ 0 \end{bmatrix}; \text{ for } \quad nr \geq mp,$$

or

$$\Sigma = \begin{bmatrix} \Sigma_{nr} & 0 \end{bmatrix}; \text{ for } \quad nr \leq mp.$$

An important issue in any identification method is that of the consistency of the estimates, i.e. the convergence of the estimated parameters to the 'true' values as the number of data points N tends to infinity. Suppose that the real system belongs to the model class (defined by equations (1)-(8)). Therefore, the observed data have actually been generated by

$$y_k = \theta_0^T \phi_k + \nu_k^0, \tag{28}$$

for some sequence $\{\nu_k^0\}$, where θ_0 can be considered as the 'true' parameter vector. Since the regressors ϕ_k depend only on past inputs, then they are uncorrelated from the noise. It is well known [11] that, under these conditions, the least squares estimate $\widehat{\theta}$ is strongly consistent, in the sense that $\widehat{\theta}$ converges (with probability one) to θ_0 as $N \to \infty$, under the assumption on persistency of excitation of the regressors (as expressed in footnote 4). Moreover, the consistency of the estimate $\widehat{\theta}$ holds even in the presence of coloured noise.

The convergence of the estimate $\widehat{\theta}$ implies that $\widehat{\Theta}_{ab} \to \Theta_{ab}$ with probability one as N tends to infinity (denoted $\widehat{\Theta}_{ab} \xrightarrow{\text{a.s.}} \Theta_{ab}$). Noting now that

$$\begin{aligned} \left\| \widehat{a}\widehat{b}^T - ab^T \right\|_2^2 &= \left\| \widehat{a}\widehat{b}^T - \widehat{\Theta}_{ab} + \widehat{\Theta}_{ab} - \Theta_{ab} \right\|_2^2, \\ &\leq \left\| \widehat{a}\widehat{b}^T - \widehat{\Theta}_{ab} \right\|_2^2 + \left\| \widehat{\Theta}_{ab} - \Theta_{ab} \right\|_2^2, \\ &= \sigma_{n+1}^2 + \left\| \widehat{\Theta}_{ab} - \Theta_{ab} \right\|_2^2, \end{aligned} \tag{29}$$

and taking into account that Θ_{ab} is a rank n matrix, then

$$\left\| \widehat{a}\widehat{b}^T - ab^T \right\|_2^2 \xrightarrow{\text{a.s.}} 0$$

as N tends to infinity. Now, from the uniqueness of the decomposition ab^T, it can be concluded that $\widehat{a} \xrightarrow{\text{a.s.}} a$, and $\widehat{b} \xrightarrow{\text{a.s.}} b$ as N tends to infinity. The result is summarized in the following theorem.

Theorem 3.2 Let $\widehat{a}$ and $\widehat{b}$ be computed using the identification Algorithm 3.1. Then, under the uniqueness condition, and the assumption on persistency of excitation of the regressors (as expressed in footnote 4), $\widehat{a} \xrightarrow{\text{a.s.}} a$, and $\widehat{b} \xrightarrow{\text{a.s.}} b$ as N tends to infinity. The result holds even in the presence of coloured noise. □

4 Simulation Example

To illustrate the proposed identification scheme, a simulation example is provided in this section. The nonlinear 'true' system consists of a third order linear discrete system with transfer function

$$G(z) = \frac{z^2 + 0.7z - 1.5}{z^3 + 0.9z^2 + 0.15z - 0.002}, \tag{30}$$

preceded by a static nonlinearity described by a fourth order polynomial of the form

$$\mathcal{N}(u_k) = 0.8585u_k + 0.0149u_k^2 - 0.5113u_k^3 - 0.0263u_k^4. \quad (31)$$

The nonlinear characteristic is shown as curve A (solid line) in figure 2. The system was excited with the de-

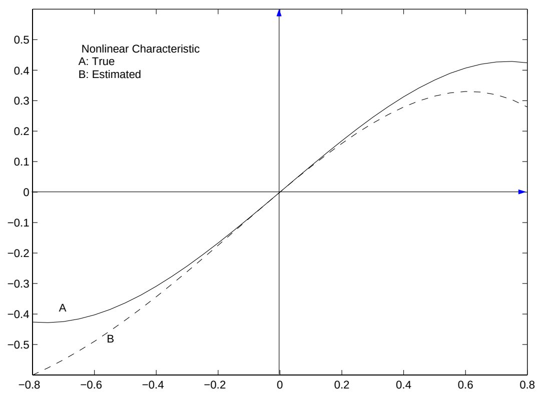

Figure 2: True (solid line) and Estimated (dashed line) nonlinear characteristic.

terministic input

$$\begin{aligned} u_k &= \sin(0.0005\pi k) + 0.5\sin(0.0015\pi k) \\ &+ 0.3\sin(0.0025\pi k) + 0.1\sin(0.0035\pi k), \end{aligned}$$

and the output was corrupted with zero-mean coloured noise with spectrum $\Phi_\nu(\omega) = \frac{0.64\times 10^{-8}}{1.2-0.4\cos(\omega)}$.

For the purposes of identification, the linear subsystem was represented using the rational Orthonormal Bases with Fixed Poles (OBFP) studied in [13, 9], that have the more common FIR, Laguerre [21, 23], and Kautz [22, 23] bases as special cases. The bases are defined as

$$\mathcal{B}_\ell(q) = \left(\frac{\sqrt{1-|\xi_\ell|^2}}{q-\xi_\ell}\right)\prod_{i=0}^{\ell-1}\left(\frac{1-\overline{\xi_i}q}{q-\xi_i}\right), \quad (32)$$

and they allow prior knowledge about an arbitrary number of system modes to be incorporated in the identification process. By choosing the poles of the bases $(\xi_0, \xi_1, \cdots, \xi_{p-1})$, closed to the (approximately known) dominant system poles, the accuracy of the estimation can be considerably improved with respect to the case of using FIR, Laguerre or Kautz bases, where the poles need all to be at the same fixed location [9]. In this example, the poles of the bases were chosen at $\{-0.01, -0.2, -0.7\}$, so that a third order linear model was identified. The estimated transfer function was (compare with the 'true' transfer function (30))

$$\widehat{G}(z) = \frac{1.0012z^2 + 0.6808z - 1.4832}{z^3 + 0.91z^2 + 0.149z + 0.0014}.$$

On the other hand, a fourth order polynomial was used to represent the nonlinear part of the model. The estimated nonlinear model was (compare with the 'true' nonlinearity (31))

$$\widehat{\mathcal{N}}(u_k) = 0.8829u_k - 0.0747u_k^2 - 0.4483u_k^3 - 0.1183u_k^4.$$

The estimated nonlinear characteristic is represented as curve B (dashed line) in figure 2.

Finally, the measured (solid line) and estimated (dashed line) outputs are plotted in figure 3, where a good agreement between them can be observed (they are almost indistinguishable one from the other).

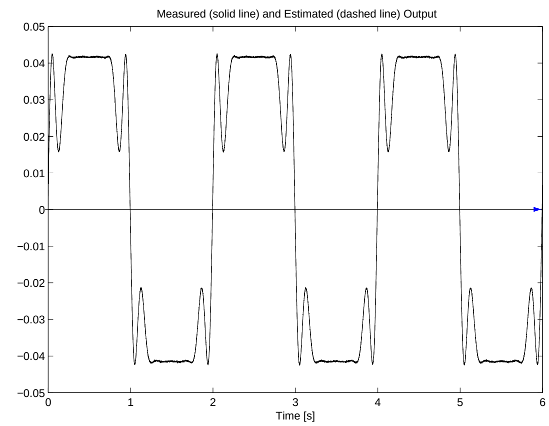

Figure 3: Measured (solid line) and Estimated (dashed line) Outputs.

5 Concluding Remarks

In this paper, a noniterative method for the identification of multivariable Hammerstein systems has been presented. The proposed method is numerically robust, since it relays only in Least Squares Estimation and Singular Value Decomposition, and provides consistent estimates under weak assumptions on the persistency of excitation of the inputs, even on the presence of coloured noise. The key issue here is the representation of the linear part of the system using orthonormal basis functions which allows to write the ouput equation in linear regressor form, with the (deterministic) regressors being uncorrelated from the noise. In addition, the use of rational orthonormal bases allows 'a priori' information one can have about the dominant dynamics of the linear part, to be incorporated in the identification process, to improve the estimation accuracy.

Acknowledgements.

The first author want to thank the people of the Departamento de Ingeniería de Sistemas y Automática, University of Valladolid, Spain, for their hospitality during his academic visit to that Department.

References

[1] E.-W. Bai. An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems. *Automatica*, 34(3):333–338, 1998.

[2] S. Billings. Identification of nonlinear systems - A survey. *Proc. of IEEE, Part D*, 127:272, 1980.

[3] S. Billings and S. Y. Fakhouri. Identification of a class of nonlinear systems using correlation analysis. *Proc. IEE, Corresp.*, 125:1307, 1978.

[4] S. Billings and S. Y. Fakhouri. Identification of nonlinear systems using correlation analysis. *Proc. IEE*, 125:691, 1978.

[5] S. Billings and S. Y. Fakhouri. Identification of systems containing linear dynamic and static nonlinear elements. *Automatica*, 18(1):15–26, 1982.

[6] M. Boutayeb and M. Darouach. Recursive identification method for MISO Wiener-Hammerstein model. *IEEE Trans. Automat. Contr.*, 40(2):287–291, February 1995.

[7] E. Eskinat, S. H. Johnson, and W. L. Luyben. Use of Hammerstein models in identification of nonlinear systems. *AIChE Journal*, 37(2):255–268, 1991.

[8] G. H. Golub and C. F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, Baltimore, 2nd edition, 1989.

[9] J. C. Gómez. *Analysis of Dynamic System Identification using Rational Orthonormal Bases.* PhD thesis, Dept. of Electrical and Computer Engineering, The University of Newcastle, Australia, PS file downloadable from http://www.eie.fceia.unr.edu.ar/~jgomez, July 1998.

[10] M. J. Korenberg. Identification of biological cascades of linear and static nonlinear systems. In *Proc. of the 16th. Midwest Symposium on Circuit Theory*, 1973.

[11] L. Ljung. *System Identification: Theory for the User.* Prentice-Hall, N.J., 2nd edition, 1999.

[12] K. S. Narendra and P. G. Gallman. An iterative method for the identification of nonlinear systems using a Hammerstein model. *IEEE Trans. Automat. Contr.*, AC-11:546–550, July 1966.

[13] B. Ninness and F. Gustafsson. A unifying construction of orthonormal bases for system identification. *IEEE Trans. Automat. Contr.*, AC-42(4):515–521, 1997.

[14] S. Rangan, G. Wolodkin, and K. Poolla. Identification methods for Hammerstein systems. In *Proc. of the IEEE Conference on Decision and Control*, pp. 697–702, New Orleans, Dec. 1995.

[15] K. S. Shanmugan and M. Lal. Analysis and synthesis of a class of nonlinear systems. *IEEE Trans. on Circuits and Systems*, CAS-23:17, 1976.

[16] H. B. Smets. Analysis and synthesis of nonlinear systems. *IRE Transactions on Circuit Theory*, CT-7:459, 1960.

[17] W. Smith and W. J. Rugh. On the structure of a class of nonlinear systems. *IEEE Trans. Automat. Contr.*, AC-19:701, 1974.

[18] T. Södeström and P. Stoica. *System Identification.* Prentice-Hall, N.J., 1989.

[19] P. Stoica. On the convergence of an iterative algorithm used for Hammerstein system identification. *IEEE Trans. Automat. Contr.*, AC-26(4):967–969, 1981.

[20] P. M. J. Van den Hof, P. S. C. Heuberger, and J. Bokor. System identification with generalized orthonormal basis functions. *Automatica*, 31(12):1821–1834, 1995.

[21] B. Wahlberg. System identification using Laguerre models. *IEEE Trans. Automat. Contr.*, AC-36(5):551–562, 1991.

[22] B. Wahlberg. System identification using Kautz models. *IEEE Trans. Automat. Contr.*, AC-39(6):1276–1282, 1994.

[23] B. Wahlberg and P. M. Mäkilä. On approximation of stable linear dynamical systems using Laguerre and Kautz functions. *Automatica*, 32(5):693–708, 1996.

Appendix

Proof of Theorem 3.1 Let the Singular Value Decomposition of the matrix $\widehat{\Theta}_{ab} \in \mathbb{R}^{nr \times mp}$ be given by

$$\widehat{\Theta}_{ab} = \sum_{i=1}^{k} \sigma_i u_i v_i^T. \tag{33}$$

where k is the rank of $\widehat{\Theta}_{ab}$. Appealing to Theorem 2.5.2 in [8], the rank-n matrix $\Theta \in \mathbb{R}^{nr \times mp}$ $(n < k)$ which is closest, in the 2-norm sense, to $\widehat{\Theta}_{ab}$ is given by

$$\Theta = \Theta_n \triangleq \sum_{i=1}^{n} \sigma_i u_i v_i^T, \tag{34}$$

and the approximation error is given by

$$\left\| \widehat{\Theta}_{ab} - \Theta_n \right\|_2^2 = \sigma_{n+1}^2. \tag{35}$$

Considering now the partition of the 'economy size' SVD of $\widehat{\Theta}_{ab}$ in (23), it is clear that

$$\Theta_n = U_1 \Sigma_1 V_1^T = (U_1)\left(V_1 \Sigma_1\right)^T,$$

what concludes the proof, by equating

$$\widehat{a} = U_1, \tag{36}$$
$$\widehat{b} = V_1 \Sigma_1. \tag{37}$$

■