

HAMMERSTEIN AND WIENER MODEL IDENTIFICATION USING RATIONAL ORTHONORMAL BASES

J. C. GÓMEZ[†] and E. BAEYENS[‡]

[†]*Laboratory for System Dynamics, FCEIA, Universidad Nacional de Rosario
Riobamba 245 Bis, 2000 Rosario, Argentina*

e-mail: jcgomez@fceia.unr.edu.ar

[‡]*Depto. de Ing. de Sistemas y Automática, ETSII, Universidad de Valladolid
Paseo del Cauce s/n, 47011 Valladolid, Spain*

e-mail: enrbae@eis.uva.es

Abstract— In this paper, non iterative algorithms for the identification of (multivariable) Hammerstein and Wiener systems are presented. The proposed algorithms are numerically robust, since they are based only on least squares estimation and singular value decomposition. For the Hammerstein model, the algorithm provides consistent estimates even in the presence of coloured output noise, under weak assumptions on the persistency of excitation of the inputs. For the Wiener model, consistency of the estimates can only be guaranteed in the noise free case. Key in the derivation of the results is the use of rational orthonormal bases for the representation of the linear part of the systems.

Keywords— Hammerstein and Wiener models, Nonlinear Identification, Singular Value Decomposition.

I. INTRODUCTION

In the last decades, a considerable amount of research has been carried out on modelling, identification, and control of nonlinear systems. Most 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 classes of nonlinear models are the so called *block-oriented nonlinear models* (Pearson and Pottmann, 2000), which consist of the interconnection of Linear Time Invariant (LTI) systems and static (memoryless) nonlinearities. Within this class, two of the more common model structures are:

- the **Hammerstein** model, which consists of the cascade connection of a static (memoryless) nonlinearity followed by a LTI system (see for instance (Eskinat *et al.*, 1991) for a review on identification of Hammerstein models), and

- the **Wiener** model, in which the order of the linear and the nonlinear blocks in the cascade connection is reversed (see for instance (Greblicki, 1994), (Wigren, 1993), (Wigren, 1994) for different methods for the identification of Wiener models).

These model structures have been successfully used to represent nonlinear systems in a number of practical applications in the areas of chemical processes (Eskinat *et al.*, 1991), (Pearson and Pottmann, 2000), (Kalafatis *et al.*, 1995), (Chou *et al.*, 2000), biological processes (Korenberg, 1978), signal processing, communications, and control (Fruzzetti *et al.*, 1997).

Several techniques have been proposed in the literature for the identification of Hammerstein and Wiener models. The reader is referred to (Narendra and Gallman, 1966), (Billings, 1980), (Billings and Fakhouri, 1982), (Eskinat *et al.*, 1991), (Greblicki and Pawlak, 1989), and the references therein, for identification of Hammerstein models; and to (Billings, 1980), (Wigren, 1993), (Wigren, 1994), (Greblicki, 1994), (Hagenblad and Ljung, 2000), and the references therein, for identification of Wiener models. For the purpose of putting into context the present work, three main approaches for the identification of Hammerstein and Wiener models will be distinguished. The first one is the traditional iterative algorithm proposed by Narendra and Gallman in (Narendra and Gallman, 1966). 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. A second approach, based on correlation techniques, is introduced in (Billings, 1980). 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 single-input/single-output (SISO) Hammerstein-Wiener systems has been introduced by

Bai in (Bai, 1998). This algorithm is based on Least Squares Estimation (LSE) and Singular Value Decomposition (SVD), but consistency of the estimates can only be guaranteed for the case of the disturbances being white noise, or in the noise-free case. Inspired by the work in (Bai, 1998), Gómez and Baeyens (Gómez and Baeyens, 2000) proposed a noniterative algorithm for the identification of Hammerstein models, which, in contrast to (Bai, 1998), applies also to multivariable systems and where the consistency of the estimates is guaranteed even in the presence of coloured output noise. As in (Bai, 1998), the main computational tools employed by the algorithm are LSE and SVD, which results in numerical robustness under weak assumptions on the persistency of excitation of the inputs¹. 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.

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 close to the (approximately known) system poles the accuracy of the estimate can be considerably improved (see (Gómez, 1998) 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 (Gómez, 1998), (Ninness and Gustafsson, 1997), (Wahlberg, 1991), (Wahlberg, 1994), (Van den Hof *et al.*, 1995), and the references therein. In addition, the use of orthonormal bases leads to a linear regressor model, so that least squares techniques can be used for the parameter estimation. Furthermore, 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 parts of Hammerstein and Wiener systems. For the Hammerstein model, this parameterization 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. For the Wiener model, the parameterization also results in a linear regressor from where the parameters characterizing the linear and

the nonlinear parts can be estimated using only least squares techniques.

In comparison with other works, the proposed algorithms have the following advantages: **1.** They apply to **multivariable** Hammerstein and Wiener models, **2.** No special assumptions on the inputs, other than the standard persistency of excitation conditions, are required, **3.** For the case of the Hammerstein model, the algorithm provides consistent estimates even in the presence of coloured noise, while for the Wiener model, the algorithm provides consistent estimates only in the noise free case.

The rest of the paper is organized as follows. In Section II, the multivariable Hammerstein model is introduced, the identification problem is formulated, and the optimal identification algorithm is derived. The same is done in Section III for multivariable Wiener models. Simulation examples illustrating the performance of the algorithms are presented in Section IV, and finally, some concluding remarks are provided in Section V.

II. HAMMERSTEIN MODEL IDENTIFICATION

A. Problem Formulation

A (multivariable) Hammerstein model is schematically represented in Fig. 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

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

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

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

where $g_i(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^n$, ($i = 1, \dots, r$), are known (nonlinear) basis functions, and $a_i \in \mathbb{R}^{n \times n}$, ($i = 1, \dots, r$), are unknown matrix parameters. Typically, the nonlinear basis functions $g_i(\cdot)$ are polynomials that allows the representation of smooth nonlinearities³, but they can also be Radial Basis Functions (RBF) or basis functions generated from a mother function (*e.g.*, wavelets). It is not the intention of this paper to give

¹This 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 that can be observed from the output.

²Here, q stands for the forward shift operator defined by $qx_k = x_{k+1}$, 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 $H_2^{m \times n}(\mathbb{T})$ will be referred as the space of all stable, causal, discrete-time, $(m \times n)$ transfer matrices.

³Any smooth function in an interval can be represented with arbitrary accuracy by a polynomial of sufficiently high order.

a complete overview of nonlinear approximation using basis functions, and the reader is referred to the survey papers (Sjöberg *et al.*, 1995), (Juditsky *et al.*, 1995), and the references therein.

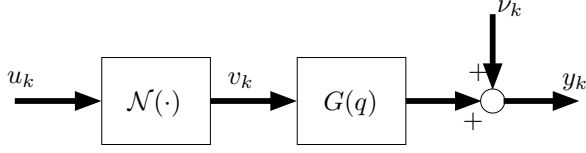


Figure 1: Multivariable Hammerstein Model.

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

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

where $b_{\ell} \in \mathbb{R}^{m \times n}$ are unknown matrix parameters, and $\{\mathcal{B}_{\ell}(q)\}_{\ell=0}^{\infty}$ are rational orthonormal bases⁴ on $H_2(\mathbb{T})$.

The identification problem is to estimate the unknown parameter matrices a_i , ($i = 1, \dots, r$), and b_{ℓ} , ($\ell = 0, \dots, p-1$), 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.

B. Identification Algorithm

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

$$y_k = \left(\sum_{\ell=0}^{p-1} b_{\ell} \mathcal{B}_{\ell}(q) \right) \left(\sum_{i=1}^r a_i g_i(u_k) \right) + \nu_k, \quad (4)$$

$$= \sum_{\ell=0}^{p-1} \sum_{i=1}^r b_{\ell} a_i \mathcal{B}_{\ell}(q) g_i(u_k) + \nu_k. \quad (5)$$

It is clear from equation (5) that the parameterization (2)-(3) is not unique, since any parameter matrices $b_{\ell} \alpha$ and $\alpha^{-1} a_i$, for some nonsingular matrix $\alpha \in \mathbb{R}^{n \times n}$, provide the same input-output equation (5). In other words, any identification experiment can not distinguish between the parameters (b_{ℓ}, a_i) and $(b_{\ell} \alpha, \alpha^{-1} a_i)$. As it is common in the literature (Bai, 1998), 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 constraints must be imposed on the parameters. A technique that is often used to obtain uniqueness is to normalize the parameter matrices a_i (or b_{ℓ}), that is to assume for

instance that $\|a_i\|_2 = 1$ (or equivalently $\|b_{\ell}\|_2 = 1$). A similar methodology was employed in (Bai, 1998) for a scalar Hammerstein-Wiener model. Under this assumption the parameterization (2)-(3) is unique.

Defining now

$$\theta = [b_0 a_1, \dots, b_0 a_r, \dots, b_{p-1} a_1, \dots, b_{p-1} a_r]^T \quad (6)$$

$$\begin{aligned} \phi_k &= [\mathcal{B}_0(q) g_1^T(u_k), \dots, \mathcal{B}_0(q) g_r^T(u_k), \dots, \\ &\vdots \\ &\mathcal{B}_{p-1}(q) g_1^T(u_k), \dots, \mathcal{B}_{p-1}(q) g_r^T(u_k)]^T, \end{aligned} \quad (7)$$

equation (5) can be written as

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

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

$$Y_N \triangleq [y_1, y_2, \dots, y_N]^T, \quad (9)$$

$$V_N \triangleq [\nu_1, \nu_2, \dots, \nu_N]^T, \quad (10)$$

$$\Phi_N \triangleq [\phi_1, \phi_2, \dots, \phi_N]^T, \quad (11)$$

the following equation can be written

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

It is well known (Ljung, 1999) that the estimate $\hat{\theta}$ 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

$$\hat{\theta} = (\Phi_N \Phi_N^T)^{-1} \Phi_N Y_N = \Phi_N^{\dagger} Y_N, \quad (13)$$

provided the indicated inverse exists⁵ (Ljung, 1999), (Söderström and Stoica, 1989).

The problem is how to estimate the parameter matrices a_i , ($i = 1, \dots, r$), and b_{ℓ} , ($\ell = 0, \dots, p-1$) from the estimate $\hat{\theta}$ in (13). From the definition of θ in (6), it is easy to see that $\theta = \text{blockvec}(\Theta_{ab})$, where $\text{blockvec}(\Theta_{ab})$ 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

$$\Theta_{ab} = \begin{bmatrix} a_1^T b_0^T & a_1^T b_1^T & \dots & a_1^T b_{p-1}^T \\ a_2^T b_0^T & a_2^T b_1^T & \dots & a_2^T b_{p-1}^T \\ \vdots & \vdots & \dots & \vdots \\ a_r^T b_0^T & a_r^T b_1^T & \dots & a_r^T b_{p-1}^T \end{bmatrix} = ab^T, \quad (14)$$

with the following definitions for the matrices a and b ,

$$a = [a_1, a_2, \dots, a_r]^T, \quad (15)$$

$$b = [b_0^T, b_1^T, \dots, b_{p-1}^T]^T. \quad (16)$$

⁴The 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

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

⁵The 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

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

An estimate $\hat{\Theta}_{ab}$ of the matrix Θ_{ab} can then be obtained from the estimate $\hat{\theta}$ in (13). The problem now is how to estimate the parameter matrices a and b from the estimate $\hat{\Theta}_{ab}$. It is clear that the closest, in the 2-norm ⁶ sense, estimates \hat{a} and \hat{b} are such they minimize the norm

$$\left\| \hat{\Theta}_{ab} - \hat{a}\hat{b}^T \right\|_2^2. \quad (17)$$

That is,

$$(\hat{a}, \hat{b}) = \arg \min_{a, b} \left\{ \left\| \hat{\Theta}_{ab} - ab^T \right\|_2^2 \right\}. \quad (18)$$

The solution to this optimization problem is provided by the Singular Value Decomposition (SVD) (Golub and Van Loan, 1989) of the matrix $\hat{\Theta}_{ab}$. The result is summarized in the following Theorem.

Theorem 1 Let $\hat{\Theta}_{ab} \in \mathbb{R}^{nr \times mp}$ have rank $k > n$, and let the *economy-size* SVD of $\hat{\Theta}_{ab}$ be given by

$$\hat{\Theta}_{ab} = U_k \Sigma_k V_k^T = \sum_{i=1}^k \sigma_i u_i v_i^T \quad (19)$$

where Σ_k is a diagonal matrix containing the k nonzero singular values ($\sigma_i, i = 1, \dots, k$) of $\hat{\Theta}_{ab}$ in nonincreasing order, and where the matrices $U_k = [u_1 \ u_2 \ \dots \ u_k] \in \mathbb{R}^{nr \times k}$ and $V_k = [v_1 \ v_2 \ \dots \ v_k] \in \mathbb{R}^{mp \times k}$ 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 $\hat{\Theta}_{ab}$,

$$\hat{\Theta}_{ab} = U \Sigma V^T, \quad (20)$$

respectively ⁷. Then, the matrices $\hat{a} \in \mathbb{R}^{nr \times n}$ and $\hat{b} \in \mathbb{R}^{mp \times n}$ that minimize the norm $\left\| \hat{\Theta}_{ab} - ab^T \right\|_2^2$, are given by

$$(\hat{a}, \hat{b}) = \arg \min_{a, b} \left\{ \left\| \hat{\Theta}_{ab} - ab^T \right\|_2^2 \right\} = (U_1, V_1 \Sigma_1), \quad (21)$$

where $U_1 \in \mathbb{R}^{nr \times n}$, $V_1 \in \mathbb{R}^{mp \times n}$, and $\Sigma_1 = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_n\}$ are given by the following partition of the *economy-size* SVD in (19),

$$\hat{\Theta}_{ab} = [U_1 \ U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}, \quad (22)$$

⁶The 2-norm of a matrix $A = (a_{ij})_{(m \times n)}$ is the norm induced by the 2-norm (or Euclidean norm) of vectors

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

⁷In equation (20), the matrix $\Sigma \in \mathbb{R}^{nr \times mp}$ is given by

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

or

$$\Sigma = [\Sigma_{nr} \ 0]; \text{ for } nr \leq mp.$$

and the approximation error is given by

$$\left\| \hat{\Theta}_{ab} - \hat{a}\hat{b}^T \right\|_2^2 = \sigma_{n+1}^2. \quad (23)$$

Proof: See APPENDIX. ■

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

Algorithm 1

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

$$\hat{\theta} = \text{blockvec} \left(\hat{\Theta}_{ab} \right). \quad (24)$$

Step 2: Compute the *economy-size* SVD of $\hat{\Theta}_{ab}$ as in Theorem 1, and the partition of this decomposition as in equation (22).

Step 3: Compute the estimates of the parameter matrices a and b as $\hat{a} = U_1$ and $\hat{b} = V_1 \Sigma_1$, respectively. □

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

$$y_k = \theta_0^T \phi_k + \nu_k^0, \quad (25)$$

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 (Ljung, 1999) that, under these conditions, the least squares estimate $\hat{\theta}$ is strongly consistent, in the sense that $\hat{\theta}$ converges (with probability one) to θ_0 as $N \rightarrow \infty$, under the assumption on persistency of excitation of the regressors. Moreover, the consistency of the estimate $\hat{\theta}$ holds even in the presence of coloured noise. The convergence of the estimate $\hat{\theta}$ implies that of \hat{a} and \hat{b} . The result is summarized in the following Theorem.

Theorem 2 Let \hat{a} and \hat{b} be computed using the identification Algorithm 1. Then, under the uniqueness condition, and the assumption on persistency of excitation of the regressors, $\hat{a} \xrightarrow{\text{a.s.}} a$, and $\hat{b} \xrightarrow{\text{a.s.}} b$ as N tends to infinity. The result holds even in the presence of coloured noise.

Proof: See APPENDIX. ■

III. WIENER MODEL IDENTIFICATION

A. Problem Formulation

A (multivariable) Wiener model is schematically depicted in Fig. 2. The model consists of the cascade of a LTI system with transfer function (matrix)

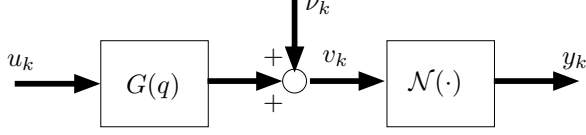


Figure 2: Multivariable Wiener Model.

$G(z) \in H_2^{m \times n}(\mathbb{T})$, followed by a zero-memory nonlinear element with input-output characteristic given by $\mathcal{N}(\cdot)$. In this case, $y_k \in \mathbb{R}^m$, $u_k \in \mathbb{R}^n$, and $\nu_k \in \mathbb{R}^m$, represent the system output, input, and process noise vectors at time k , respectively.

As in the case of Hammerstein models, it will be assumed that the LTI subsystem is represented as an orthonormal basis expansion of the form (3). On the other hand, the nonlinear function $\mathcal{N}(\cdot) : \mathbb{R}^m \rightarrow \mathbb{R}^m$ will be assumed to be invertible^{8,9}, and that its inverse $\mathcal{N}^{-1}(\cdot)$ can be described as

$$\mathcal{N}^{-1}(y_k) = \sum_{i=1}^r a_i g_i(y_k) \quad (26)$$

where $g_i(\cdot) : \mathbb{R}^m \rightarrow \mathbb{R}^m$, ($i = 1, \dots, r$), are known basis functions, and $a_i \in \mathbb{R}^{m \times m}$, ($i = 1, \dots, r$), are unknown matrix parameters. Without loss of generality, it will also be assumed that $a_1 = I_m$, with I_m standing for the identity matrix of dimensions ($m \times m$).

The identification problem is to estimate the unknown parameter matrices a_i , ($i = 2, \dots, r$), and b_ℓ , ($\ell = 0, \dots, p-1$), 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.

B. Identification Algorithm

The intermediate variable v_k in Fig. 2, can be written as

$$v_k = G(q)u_k + \nu_k, \quad (27)$$

and also as

$$v_k = \mathcal{N}^{-1}(y_k). \quad (28)$$

Equating the right hand sides of the above two equations, and considering the parameterizations (3) and (26) of the linear and the nonlinear subsystems, respectively, the following equation is obtained

$$g_1(y_k) = - \sum_{i=2}^r a_i g_i(y_k) + \sum_{\ell=0}^{p-1} b_\ell \mathcal{B}_\ell(q)u_k + \nu_k, \quad (29)$$

⁸As pointed out in (Pearson and Pottmann, 2000), this rules out the use of the proposed identification algorithm for processes in which the phenomenon of input multiplicity is present (see next footnote).

⁹*Input Multiplicity* is the situation in which more than one steady-state input value u_{ss} corresponds to the same steady-state output value y_{ss} .

which is a linear regression. Defining

$$\theta \triangleq [a_2, a_3, \dots, a_r, b_0, b_1, \dots, b_{p-1}]^T, \quad (30)$$

$$\phi_k \triangleq [-g_2^T(y_k), -g_3^T(y_k), \dots, -g_r^T(y_k), \mathcal{B}_0(q)u_k^T, \mathcal{B}_1(q)u_k^T, \dots, \mathcal{B}_{p-1}(q)u_k^T]^T, \quad (31)$$

equation (29) can be written as

$$g_1(y_k) = \theta^T \phi_k + \nu_k. \quad (32)$$

Now, an estimate $\hat{\theta}$ of θ can be computed by minimizing a quadratic criterion on the prediction errors $\epsilon_k = g_1(y_k) - \theta^T \phi_k$ (i.e., the least squares estimate). It is well known (Ljung, 1999) that this estimate is given by¹⁰

$$\hat{\theta} = (\Phi_N^T \Phi_N)^{-1} \Phi_N^T Y_N = \Phi_N^\dagger Y_N, \quad (33)$$

where the following definitions have been made

$$Y_N \triangleq [g_1(y_1), g_1(y_2), \dots, g_1(y_N)]^T, \quad (34)$$

$$V_N \triangleq [\nu_1, \nu_2, \dots, \nu_N]^T, \quad (35)$$

$$\Phi_N \triangleq [\phi_1, \phi_2, \dots, \phi_N]. \quad (36)$$

Now, estimates of the parameters a_i , ($i = 2, \dots, r$), and b_ℓ , ($\ell = 0, \dots, p-1$), can be computed by partitioning the estimate $\hat{\theta}$ in (33), according to the definition of θ in (30).

In this case, the consistency of the estimate $\hat{\theta}$ in (33), can only be guaranteed in the noise free case, since in the presence of noise the regressors $\{\phi_k\}$ at time k will be correlated with the disturbances $\{\nu_k\}$ at the same instant, even if the disturbance is a white noise process (Ljung, 1999).

IV. SIMULATION EXAMPLES

The performance of the proposed identification algorithms is illustrated through two simulation examples.

Example 1 (Hammerstein model)

The nonlinear *true* system consists of a third order linear discrete system with transfer function

$$G(z) = \frac{z^2 + 0.7z - 1.5}{z^3 + 0.9z^2 + 0.15z + 0.002}, \quad (37)$$

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

$$\mathcal{N}(u_k) = 0.8585u_k + 0.0149u_k^2 - 0.5113u_k^3 - 0.0263u_k^4. \quad (38)$$

The nonlinear characteristic is shown as curve A (solid line) in Fig. 3. The system was excited with the input

$$\begin{aligned} u_k = & \sin(0.0005\pi k) + 0.5 \sin(0.0015\pi k) \\ & + 0.3 \sin(0.0025\pi k) \\ & + 0.1 \sin(0.0035\pi k) + \gamma_k, \end{aligned} \quad (39)$$

¹⁰Provided the indicated inverse exists.

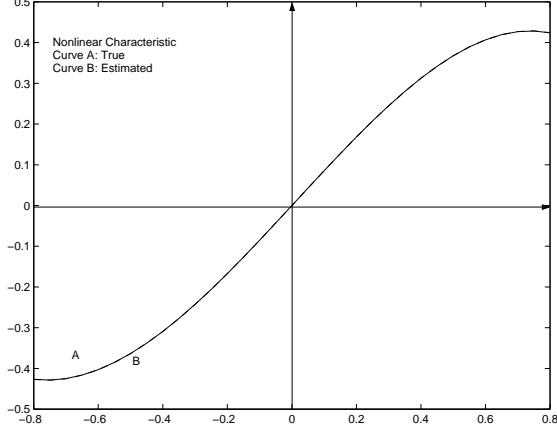


Figure 3: True (Curve A: solid line) and Estimated (Curve B: dashed line) nonlinear characteristic (indistinguishable one from the other).

where γ_k is a zero mean white Gaussian process with variance 10^{-6} , 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 (Ninness and Gustafsson, 1997), (Gómez, 1998), that have the more common FIR, Laguerre (Wahlberg, 1991), and Kautz (Wahlberg, 1994) bases as special cases. The bases are defined as

$$\mathcal{B}_0(q) = \frac{\sqrt{1 - |\xi_0|^2}}{q - \xi_0},$$

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

and they allow prior knowledge about an arbitrary number of system modes to be incorporated in the identification process.

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 (37))

$$\hat{G}(z) = \frac{1.0034z^2 + 0.6941z - 1.4967}{z^3 + 0.91z^2 + 0.149z + 0.0014}.$$

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 (38))

$$\hat{\mathcal{N}}(u_k) = 0.8591u_k - 0.0147u_k^2 - 0.5110u_k^3 - 0.0263u_k^4.$$

The estimated nonlinear characteristic is represented as curve B (dashed line) in Fig. 3. Finally, the measured (solid line) and estimated (dashed line) outputs

are represented in Fig. 4, where a good agreement between them can be observed (they are almost indistinguishable one from the other).

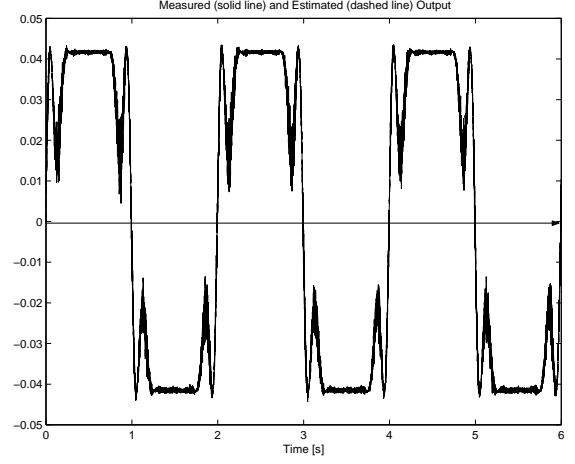


Figure 4: Measured (solid line) and Estimated (dashed line) Outputs.

Example 2 (pH Neutralization Process)

In this example, a Wiener model is identified based on the simulation data of a pH neutralization process. The process consists of an acid (HNO_3) stream, a base (NaOH) stream, and a buffer (NaHCO_3) stream that are mixed in a constant volume (V) stirring tank. The process is schematically depicted in Fig. 5, and corresponds to a bench-scale plant at the University of California, Santa Barbara (see (Henson and Seborg, 1992), (Henson and Seborg, 1994), (Henson and Seborg, 1997)).

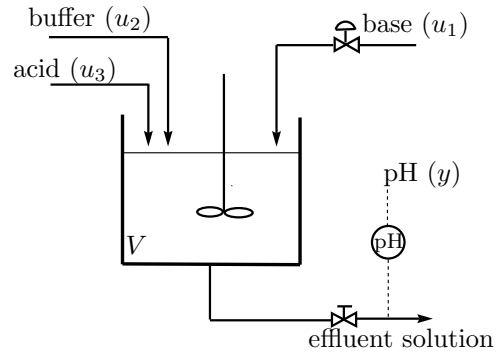


Figure 5: Schematic representation of the pH neutralization process.

The inputs to the system are the base flow rate (u_1) and the buffer flow rate (u_2), while the output (y) is the pH of the effluent solution in the tank. The acid flow rate (u_3), as well as the volume (V) of the tank are assumed to be constant. A simulation model, based on first principles, is derived in (Henson and Seborg, 1992) introducing two *reaction invariants* for each inlet stream ((W_{a1}, W_{b1}) for the base stream, (W_{a2}, W_{b2})

for the buffer stream, (W_{a3}, W_{b3}) for the acid stream, and (W_a, W_b) for the effluent solution). The dynamic model for the reaction invariants of the effluent solution (W_a, W_b) , in state-space form, is given by:

$$\dot{x} = f(x) + g(x)u_1 + p(x)u_2, \quad (40)$$

$$h(x, y) = 0, \quad (41)$$

where

$$\begin{aligned} x &\triangleq [x_1, x_2]^T = [W_a, W_b]^T, \\ f(x) &= \left[\frac{u_3}{V}(W_{a3} - x_1), \frac{u_3}{V}(W_{b3} - x_2) \right]^T, \\ g(x) &= \left[\frac{1}{V}(W_{a1} - x_1), \frac{1}{V}(W_{b1} - x_2) \right]^T, \\ p(x) &= \left[\frac{1}{V}(W_{a2} - x_1), \frac{1}{V}(W_{b2} - x_2) \right]^T, \\ h(x, y) &= x_1 + 10^{y-14} - 10^{-y} \\ &\quad + x_2 \frac{1 + 2 \times 10^{y-pK_2}}{1 + 10^{pK_1-y} + 10^{y-pK_2}}. \end{aligned}$$

The nominal operating conditions of the system are given in (Henson and Seborg, 1992), (Henson and Seborg, 1994), (Henson and Seborg, 1997).

For the purposes of identification, the model was excited with band limited white noise around the nominal values of the base and buffer flow rates. The first six hundred data were used for the estimation of the model, while the following five hundred data were used for validation purposes. The linear subsystem was represented using the same rational Orthonormal Bases with Fixed Poles (OBFP) as in the previous example, with poles at $\{0.978, 0.9897, 0.9897, 0.99, 0.9784\}$, while a third order polynomial was used to represent the nonlinear part of the model. The true and estimated output (estimation-validation data) are represented in the top plot of Fig. 6, where a good agreement between them can be observed. The estimated nonlinear characteristic is represented in the bottom plot of Fig. 6.

V. CONCLUDING REMARKS

In this paper, noniterative methods for the identification of multivariable Hammerstein and Wiener systems have been presented. The proposed algorithms are numerically robust, since they rely only on LSE and SVD. For the case of the Hammerstein model, the algorithm provides consistent estimates under weak assumptions on the persistency of excitation of the inputs, even in the presence of coloured noise. For the case of the Wiener model, consistency of the estimates can be guaranteed only for the noise free case. The key issue in the derivation of the results is the representation of the linear part of the system using orthonormal basis functions which allows to put the system in linear regressor form. In addition, the use of rational orthonormal bases allows *a priori* information one can

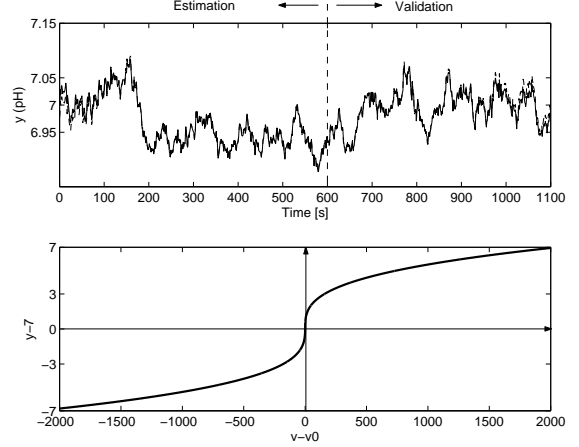


Figure 6: Top plot: True (solid line) and Estimated (dashed line) Output (Estimation-Validation Data). Bottom plot: Estimated nonlinear characteristic.

have about the dominant dynamics of the system, to be incorporated in the identification process, to improve the estimation accuracy.

REFERENCES

- Bai, E.-W., “An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems”, *Automatica*, **34**(3), 333–338 (1998).
- Billings, S., “Identification of nonlinear systems - A survey”, *Proc. of IEE, Part D*, **127**, 272–285 (1980).
- Billings, S. and S. Fakhouri, “Identification of systems containing linear dynamic and static nonlinear elements”, *Automatica*, **18**(1), 15–26 (1982).
- Chou, C., H. Bloemen, V. Verdult, T. van den Boom, T. Backx, and M. Verhaegen. “Nonlinear identification of high purity distillation columns”. In *Proc. of the IFAC Symposium on System Identification SYSID 2000*, 415–420, Santa Barbara, CA (2000).
- Eskinat, E., S. Johnson, and W. Luyben, “Use of Hammerstein models in identification of nonlinear systems”, *AIChE Journal*, **37**(2), 255–268 (1991).
- Fruzzetti, K., A. Palazoglu, and K. McDonald, “Nonlinear model predictive control using Hammerstein models”, *Journal of Process Control*, **7**(1), 31–41 (1997).
- Golub, G. and C. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 2nd edition (1989).
- Gómez, J. C. *Analysis of Dynamic System Identification using Rational Orthonormal Bases*. PhD thesis, The University of Newcastle, Australia, PS file at <http://www.fceia.unr.edu.ar/~jcgomez/> (1998).
- Gómez, J. C. and E. Baeyens. “Identification of multivariable hammerstein systems using rational orthonormal bases”. In *Proc. of the 39th. IEEE CDC*, 2849–2854, Sydney, Australia (2000).
- Greblicki, W., “Nonparametric identification of Wiener systems by orthogonal series”, *IEEE Trans. on Autom. Contr.*, **39**(10), 2077–2086 (1994).

Greblicki, W. and M. Pawlak, "Nonparametric identification of Hammerstein systems", *IEEE Trans. on Information Theory*, **35**(2), 409–418 (1989).

Hagenblad, A. and L. Ljung. "Maximum likelihood estimation of Wiener models". In *Proc. of the 39th IEEE CDC*, 712–713, Sydney, Australia (2000).

Henson, M. and D. Seborg. "Nonlinear adaptive control of a pH neutralization process". In *Proc. of the ACC*, 2586–2590 (1992).

Henson, M. and D. Seborg, "Adaptive nonlinear control of a pH neutralization process", *IEEE Trans. on Control Systems Technology*, **2**(3), 169–182 (1994).

Henson, M. and D. Seborg, editors. *Nonlinear Process Control*, chapter 4: "Feedback Linearizing Control". Prentice Hall PTR, N.J. (1997).

Juditsky, A., H. Hjalmarsson, A. Benveniste, B. Delyon, L. Ljung, J. Sjöberg, and Q. Zhang, "Nonlinear black-box modeling in system identification: Mathematical foundations", *Automatica*, **31**(12), 1725–1750 (1995).

Kalafatis, A., N. Arifin, L. Wang, and W. Cluett, "A new approach to the identification of pH processes based on the Wiener model", *Chemical Engineering Science*, **50**(23), 3693–3701 (1995).

Korenberg, M. "Identification of biological cascades of linear and static nonlinear systems". In *Proc. of the 16th. Midwest Symposium on Circuit Theory*, 2.1–2.9 (1978).

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

Narendra, K. and P. Gallman, "An iterative method for the identification of nonlinear systems using a Hammerstein model", *IEEE Trans. on Autom. Contr.*, **AC-11**, 546–550 (1966).

Ninness, B. and F. Gustafsson, "A unifying construction of orthonormal bases for system identification", *IEEE Trans. on Autom. Contr.*, **AC-42**(4), 515–521 (1997).

Pearson, R. and M. Pottmann, "Gray-box identification of block-oriented nonlinear models", *Journal of Process Control*, **10**, 301–315 (2000).

Sjöberg, J., Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P. Glorennec, H. Hjalmarsson, and A. Juditsky, "Nonlinear black-box modeling in system identification: a unified approach", *Automatica*, **31**(12), 1691–1724 (1995).

Söderström, T. and P. Stoica. *System Identification*. Prentice Hall, Inc., New Jersey (1989).

Van den Hof, P., P. Heuberger, and J. Bokor, "System identification with generalized orthonormal basis functions", *Automatica*, **31**(12), 1821–1834 (1995).

Wahlberg, B., "System identification using Laguerre models", *IEEE Trans. on Autom. Contr.*, **AC-36**(5), 551–562 (1991).

Wahlberg, B., "System identification using Kautz models", *IEEE Trans. on Autom. Contr.*, **AC-39**(6), 1276–1282 (1994).

Wigren, T., "Recursive prediction error identification using the nonlinear Wiener model", *Automatica*, **29**(4), 1011–1025 (1993).

Wigren, T., "Convergence analysis of recursive identification algorithms based on the nonlinear Wiener model", *IEEE Trans. on Autom. Contr.*, **39**(11), 2191–2206 (1994).

APPENDIX

Proof of Theorem 1 Let the Singular Value Decomposition of the matrix $\hat{\Theta}_{ab} \in \mathbb{R}^{nr \times mp}$ be given by

$$\hat{\Theta}_{ab} = \sum_{i=1}^k \sigma_i u_i v_i^T, \quad (42)$$

where k is the rank of $\hat{\Theta}_{ab}$. Appealing to Theorem 2.5.2 in (Golub and Van Loan, 1989), the rank- n matrix $\Theta \in \mathbb{R}^{nr \times mp}$ ($n < k$) which is closest, in the 2-norm sense, to $\hat{\Theta}_{ab}$ is given by

$$\Theta = \Theta_n \triangleq \sum_{i=1}^n \sigma_i u_i v_i^T, \quad (43)$$

and the approximation error is given by

$$\|\hat{\Theta}_{ab} - \Theta_n\|_2^2 = \sigma_{n+1}^2. \quad (44)$$

Considering now the partition of the *economy-size* SVD of $\hat{\Theta}_{ab}$ in (22), it is clear that

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

what concludes the proof, by equating $\hat{a} = U_1$ and $\hat{b} = V_1 \Sigma_1$. ■

Proof of Theorem 2 The convergence of the estimate $\hat{\theta}$ in (13) implies that $\hat{\Theta}_{ab} \rightarrow \Theta_{ab}$ with probability one as N tends to infinity (denoted $\hat{\Theta}_{ab} \xrightarrow{\text{a.s.}} \Theta_{ab}$). Noting now that

$$\begin{aligned} \|\hat{a}\hat{b}^T - ab^T\|_2^2 &= \|\hat{a}\hat{b}^T - \hat{\Theta}_{ab} + \hat{\Theta}_{ab} - \Theta_{ab}\|_2^2, \\ &\leq \|\hat{a}\hat{b}^T - \hat{\Theta}_{ab}\|_2^2 + \|\hat{\Theta}_{ab} - \Theta_{ab}\|_2^2, \\ &= \sigma_{n+1}^2 + \|\hat{\Theta}_{ab} - \Theta_{ab}\|_2^2, \end{aligned} \quad (45)$$

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

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

as N tends to infinity. Now, from the uniqueness of the decomposition ab^T , it can be concluded that $\hat{a} \xrightarrow{\text{a.s.}} a$, and $\hat{b} \xrightarrow{\text{a.s.}} b$ as N tends to infinity, what concludes the proof. ■