Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Process Control 14 (2004) 685–697
JOURNAL OF PROCESS CONTROL
www.elsevier.com/locate/jprocont

Identification of block-oriented nonlinear systems using orthonormal bases

Juan C. Gómez [a,*], Enrique Baeyens [b]

[a] Laboratory for System Dynamics and Signal Processing, FCEIA, Universidad Nacional de Rosario, Riobamba 245 Bis, 2000 Rosario, Argentina
[b] Department of Systems Engineering and Automatic Control, ETSII, Universidad de Valladolid, Paseo del Cauce s/n, 47011 Valladolid, Spain

Received 28 May 2002; received in revised form 11 March 2003; accepted 22 September 2003

Abstract

In this paper, new noniterative algorithms for the identification of (multivariable) *block-oriented* nonlinear models consisting of the interconnection of linear time invariant systems and static nonlinearities are presented. The proposed algorithms are numerically robust, since they are based only on least squares estimation and singular value decomposition. Two different block-oriented nonlinear models are considered in this paper, viz., the Hammerstein model, and the Wiener model. For the Hammerstein model, the proposed algorithm provides consistent estimates even in the presence of colored 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 basis functions for the representation of the linear and nonlinear parts of the models. The performance of the proposed identification algorithms is illustrated through simulation examples of two benchmark problems drawn from the process control literature, viz., a binary distillation column and a pH neutralization process.

Keywords: Block-oriented models; Nonlinear identification; Singular value decomposition; Process control

1. 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 behavior 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 [1,2], which consist of the interconnection of linear time invariant (LTI) systems and static (memoryless) nonlinearities. Within this class, three 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 [3] for a review on identification of Hammerstein models),
- the *Wiener* model, in which the order of the linear and the nonlinear blocks in the cascade connection is reversed (see for instance [4–6] for different methods for the identification of Wiener models), and
- the *feedback block-oriented (FBO) model*, which consists of a static nonlinearity in the feedback path around a LTI system. An identification algorithm for the estimation of this type of model is presented in [7], where it is used to represent the phenomenon of output multiplicity [1] which appears in some distillation processes, and that can not be modelled by Hammerstein or Wiener models [2].

These model structures have been successfully used to represent nonlinear systems in a number of practical applications in the areas of chemical processes [2,3,8,9], biological processes [10], signal processing [11], communications, and control [12].

* Corresponding author. Tel.: +54-341-4808543; fax: +54-341-4892654.
E-mail address: jcgomez@fceia.unr.edu.ar (J.C. Gómez).

[1] *Output multiplicity* is the situation in which more than one steady-state output value y_{ss} corresponds to the same steady-state input value u_{ss}.

0959-1524/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jprocont.2003.09.010

In particular, in the area of process control, two of the problems that appear frequently are the control of distillation processes and pH processes. These have been recognized as challenging problems due to the highly nonlinear and time-varying characteristics of the processes. Hammerstein and Wiener models have been successfully used to represent distillation processes and pH neutralization processes. From an identification point of view, pH processes have often been considered in the literature as having a Wiener structure (see for instance [8,13]). In this structure, the linear block represents the mixing dynamics of the reagent streams in the stirring tank reactor (CSTR), while the static nonlinearity represents the nonlinear *titration curve* which gives the pH of the effluent solution as a function of the chemical components. On the other hand, distillation processes have been modelled using both Hammerstein and Wiener models (see for instance [2,3,9,14]).

Several techniques have been proposed in the literature for the identification of Hammerstein and Wiener models. The reader is referred to [3,15–22], and the references therein, for identification of Hammerstein models; and to [4–6,23], and the references therein, for identification of Wiener models. Although there are well established results concerning the stability of FBO models [24], less attention has been paid to the problem of identification of these model structures (see for instance [2]).

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 [15] for the identification of a Hammerstein model. In this algorithm, an appropriate parametrization 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 [20] showed that the original algorithm could be divergent in some particular cases. A second approach, based on correlation techniques, is introduced in [16–18]. 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 [25]. The algorithm in [25] is based on least squares estimation (LSE) and singular value decomposition (SVD), however it only applies to the single-input/single-output (SISO) case due to the particular parametrization (the ARX model in Eq. (1) in [25]) used in that paper. Furthermore, only polynomial representations of the static nonlinearities can be handled by the algorithm, and consistency of the estimates can only be guaranteed for the case of the disturbances being white noise, or in the noise free case. Due also to the particular parametrization used in [25], an extension of the results to the MIMO (multiple input, multiple output) setting does not seem to be straightforward. Inspired by the work in [25], Gómez and Baeyens [26] proposed a noniterative algorithm for the identification of Hammerstein models, which, in contrast to [25], applies also to multivariable systems, allows a more general representation (using basis functions) for the static nonlinearity, and where the consistency of the estimates is guaranteed even in the presence of colored output noise. As in [25], 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 inputs. [2] Key on the derivations of the results in [26] is the use of basis functions for the representation of the linear and the nonlinear blocks in the Hammerstein model. The preliminary results in the conference paper [26] are summarized here, since they are used as a paradigm for the derivation of similar results for the Wiener model.

The use of rational orthonormal bases for the representation of LTI systems has come as a natural answer for the issue of how to incorporate a priori information on the system dynamics in the identification of *black-box* model structures for the systems. Choosing the poles of the bases close to the (approximately known) system poles the accuracy of the estimate can be considerably improved (see [27] for a detailed review on 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 [27–32], 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 colored noise, under the assumption that the actual system belongs to the model class (i.e., there is no undermodelling).

In this paper, basis functions are used to represent both the linear and the nonlinear parts of Hammerstein and Wiener models. For the Hammerstein model, this parametrization results in a linear regressor form, so that least squares techniques can be used to estimate an oversized parameter matrix. Then, by recurring to SVD and rank reduction, optimal estimates of the parameter matrices characterizing the linear and nonlinear parts can be obtained. For the Wiener model, the parametrization

[2] 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 observed from the output.

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. A similar result can be obtained for the FBO model (see the conference paper [33]).

In comparison with other works, the proposed algorithms have the following advantages

- They apply to *multivariable* Hammerstein and Wiener models.
- A more general representation of the static nonlinearity (not limited to polynomial representations) can be handled by the algorithm.
- No special assumptions on the inputs, other than the standard persistency of excitation conditions, are required.
- For the case of the Hammerstein model, the algorithm provides consistent estimates even in the presence of colored noise. On the other hand, the consistency of the estimates for the Wiener model can only be guaranteed in the noise free case.

Even though the proposed identification algorithms are not application specific, an important area for their application is, as already mentioned, in the chemical industry, where Hammerstein and Wiener model structures have been successfully used to represent nonlinear processes. In particular, the algorithms proposed here are suitable for the identification of multivariable, stable, continuous (as opposed to batch) processes. Since an input–output model is obtained, no observability assumptions are required to the processes.

To illustrate the performance of the proposed identification algorithms, two benchmark processes drawn from the process control literature are considered in this paper. Namely, the algorithms are used to estimate block-oriented nonlinear models for a pH neutralization process (taken from [34,35]), and for a binary distillation column considered in [36].

The rest of the paper is organized as follows. In Section 2, the multivariable Hammerstein model is introduced, the identification problem is formulated, and the optimal identification algorithm is derived. The same is done in Section 3 for multivariable Wiener models. Simulation examples illustrating the performance of the algorithms on two benchmark problems are presented in Section 4, and finally, some concluding remarks are provided in Section 5.

2. Hammerstein model identification

2.1. 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) [3] $G(q) \in H_2^{m\times n}(\mathbb{T})$. It is assumed that the measured output y_k contains an unknown additive noise component v_k.

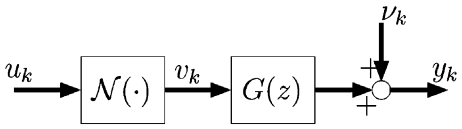

Fig. 1. Multivariable Hammerstein model.

The input–output relationship is then given by

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

where $y_k \in \mathbb{R}^m$, $u_k \in \mathbb{R}^n$, and $v_k \in \mathbb{R}^m$, 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 $g_i(\cdot) : \mathbb{R}^n \to \mathbb{R}^n$ $(i = 1, \ldots, r)$ are known (nonlinear) basis functions, and $a_i \in \mathbb{R}^{n\times n}$ $(i = 1, \ldots, r)$ are unknown matrix parameters. Typically, the nonlinear functions $g_i(\cdot)$ are polynomials (as for instance in [3]) that allows the representation of smooth nonlinearities, [4] but they can also be radial basis functions (RBF) or basis functions generated by translations and dilations of a *mother* function (e.g., wavelets). It is not the intention of this paper to give a complete overview of nonlinear approximation using basis functions, and the reader is referred to the companion papers [37,38], where a unified overview of nonlinear black-box modelling using basis functions, as well as the mathematical foundations behind these modelling approaches are presented.

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

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

where $b_\ell \in \mathbb{R}^{m\times n}$ are unknown matrix parameters, and $\{\mathscr{B}_\ell(q)\}_{\ell=0}^{\infty}$ are rational orthonormal bases [5] on $H_2(\mathbb{T})$.

[3] 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 $H_2^{m\times n}(\mathbb{T})$ will be referred as the space of all stable, causal, discrete-time, $(m \times n)$ transfer matrices.

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

[5] The bases are orthonormal in the sense that

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

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

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

The identification problem is to estimate the unknown parameter matrices a_i $(i = 1, \ldots, r)$ and b_ℓ $(\ell = 0, \ldots, p-1)$ 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 subsection, an optimal identification algorithm is presented, which is based only in LSE and SVD. Under weak conditions the algorithm delivers unbiased estimates of the parameter matrices.

2.2. Identification algorithm

Substituting Eqs. (2) and (3) in (1), the input–output relationship can be written as

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

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

It is clear from Eq. (5) that the parametrization (2) and (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 [25,39], these two sets of parameters will be called *equivalent*. To obtain a one-to-one parametrization, i.e. for the system to be identifiable, additional constrains 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), for instance assuming that $\|a_i\|_2 = 1$ (or $\|b_\ell\|_2 = 1$). A similar methodology was employed in [25] for a scalar Hammerstein–Wiener model. Under this assumption the parametrization (2) and (3) is unique.

Defining now

$$\theta \triangleq [b_0 a_1, \ldots, b_0 a_r, \ldots, b_{p-1}a_1, \ldots, b_{p-1}a_r]^{\mathrm{T}}, \tag{6}$$

$$\phi_k \triangleq [\mathscr{B}_0(q)g_1^{\mathrm{T}}(u_k), \ldots, \mathscr{B}_0(q)g_r^{\mathrm{T}}(u_k), \ldots,$$
$$\vdots \tag{7}$$
$$\mathscr{B}_{p-1}(q)g_1^{\mathrm{T}}(u_k), \ldots, \mathscr{B}_{p-1}(q)g_r^{\mathrm{T}}(u_k)]^{\mathrm{T}},$$

Eq. (5) can be written as

$$y_k = \theta^{\mathrm{T}}\phi_k + v_k, \tag{8}$$

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

$$Y_N \triangleq [y_1, y_2, \ldots, y_N]^{\mathrm{T}}, \tag{9}$$

$$V_N \triangleq [v_1, v_2, \ldots, v_N]^{\mathrm{T}}, \tag{10}$$

$$\Phi_N \triangleq [\phi_1, \phi_2, \ldots, \phi_N], \tag{11}$$

the following equation can be written

$$Y_N = \Phi_N^{\mathrm{T}}\theta + V_N. \tag{12}$$

It is well known [40] that the estimate $\hat{\theta}$ that minimizes a quadratic criterion on the prediction errors $\epsilon_N = Y_N - \Phi_N^{\mathrm{T}}\theta$ (that is the least squares estimate) is given by

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

provided the indicated inverse exists [6] [41].

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

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

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

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

$$\Theta_{ab} \triangleq \begin{bmatrix} a_1^{\mathrm{T}}b_0^{\mathrm{T}} & a_1^{\mathrm{T}}b_1^{\mathrm{T}} & \cdots & a_1^{\mathrm{T}}b_{p-1}^{\mathrm{T}} \\ a_2^{\mathrm{T}}b_0^{\mathrm{T}} & a_2^{\mathrm{T}}b_1^{\mathrm{T}} & \cdots & a_2^{\mathrm{T}}b_{p-1}^{\mathrm{T}} \\ \vdots & \vdots & \cdots & \vdots \\ a_r^{\mathrm{T}}b_0^{\mathrm{T}} & a_r^{\mathrm{T}}b_1^{\mathrm{T}} & \cdots & a_r^{\mathrm{T}}b_{p-1}^{\mathrm{T}} \end{bmatrix} = ab^{\mathrm{T}} \tag{15}$$

with the following definitions for the matrices a and b,

$$a \triangleq [a_1, a_2, \ldots, a_r]^{\mathrm{T}}, \tag{16}$$

$$b \triangleq [b_0^{\mathrm{T}}, b_1^{\mathrm{T}}, \ldots, b_{p-1}^{\mathrm{T}}]^{\mathrm{T}}. \tag{17}$$

An estimate $\widehat{\Theta}_{ab}$ 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 $\widehat{\Theta}_{ab}$. It is clear that the closest, in the 2-norm [7] sense, estimates $\hat{a}$ and $\hat{b}$ are those that solve the following optimization problem

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

The solution to this optimization problem is provided by the SVD [42] of the matrix $\widehat{\Theta}_{ab}$. The result is summarized in the following theorem.

[6] 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

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

[7] 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

$$\|A\|_2 = \mathrm{Sup}_{w\neq 0}\frac{\|Aw\|_2}{\|w\|_2}.$$

Theorem 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*

$$\widehat{\Theta}_{ab} = U_k \Sigma_k V_k^{\mathrm{T}} = \sum_{i=1}^{k} \sigma_i u_i v_i^{\mathrm{T}}, \tag{19}$$

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

$$\widehat{\Theta}_{ab} = U \Sigma V^{\mathrm{T}}, \tag{20}$$

respectively. [8] *Then, the matrices $\hat{a} \in \mathbb{R}^{nr\times n}$ and $\hat{b} \in \mathbb{R}^{mp\times n}$ that minimize the norm $\|\widehat{\Theta}_{ab} - ab^{\mathrm{T}}\|_2^2$, are given by*

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

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

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

and the approximation error is given by

$$\|\widehat{\Theta}_{ab} - \hat{a}\hat{b}^{\mathrm{T}}\|_2^2 = \sigma_{n+1}^2. \tag{23}$$

Proof. See Appendix A. □

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 $\widehat{\Theta}_{ab}$ such that

$$\hat{\theta} = \mathsf{blockvec}(\widehat{\Theta}_{ab}). \tag{24}$$

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

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

$$\hat{a} = U_1, \tag{25}$$

$$\hat{b} = V_1\Sigma_1, \tag{26}$$

respectively.

[8] In Eq. (20), the matrix $\Sigma \in \mathbb{R}^{nr\times mp}$ is given by

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

or

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

Remark 1. It is important to note that the algorithm intrinsically delivers estimates that satisfy the uniqueness condition $\|a_i\|_2 = 1$, since matrix U_1 in the SVD of $\widehat{\Theta}_{ab}$ is a unitary matrix.

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 Eqs. (1)–(8)). Therefore, the observed data have actually been generated by

$$y_k = \theta_0^{\mathrm{T}}\phi_k + v_k^0 \tag{27}$$

for some sequence $\{v_k^0\}$, 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 [40] 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 $N \to \infty$, under the assumption on persistency of excitation of the regressors (as expressed in Footnote 6). Moreover, the consistency of the estimate $\hat{\theta}$ holds even in the presence of colored noise.

The convergence of the estimate $\hat{\theta}$ 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

$$\begin{aligned} \|\hat{a}\hat{b}^{\mathrm{T}} - ab^{\mathrm{T}}\|_2^2 &= \|\hat{a}\hat{b}^{\mathrm{T}} - \widehat{\Theta}_{ab} + \widehat{\Theta}_{ab} - \Theta_{ab}\|_2^2, \\ &\leqslant \|\hat{a}\hat{b}^{\mathrm{T}} - \widehat{\Theta}_{ab}\|_2^2 + \|\widehat{\Theta}_{ab} - \Theta_{ab}\|_2^2, \\ &= \sigma_{n+1}^2 + \|\widehat{\Theta}_{ab} - \Theta_{ab}\|_2^2, \end{aligned} \tag{28}$$

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

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

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. 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 (as expressed in Footnote* 6*), $\hat{a} \xrightarrow{a.s.} a$, and $\hat{b} \xrightarrow{a.s.} b$ as N tends to infinity. The result holds even in the presence of colored noise.*

Remark 2. It is interesting to note that the structure (1)–(3) of the identified Hammerstein model can be interpreted as an equivalent LTI model whose inputs are the actual inputs u_k filtered by the nonlinear basis functions $g_i(\cdot)$ used to represent the static nonlinearity in the Hammerstein model. This interpretation of the

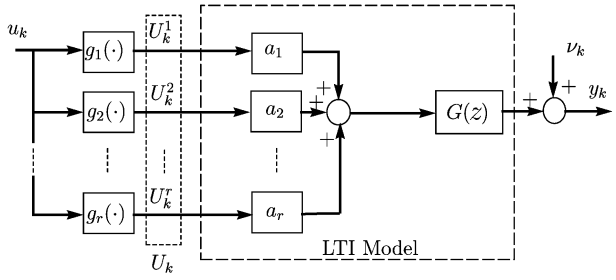

Fig. 2. Equivalent LTI model.

Hammerstein model is illustrated in the block-diagram of Fig. 2, where for simplicity the SISO case has been represented. The equivalent LTI model can be used, for instance, in a standard (i.e., based on a linear model) model predictive control (MPC) scheme [43]. In this way, a nonlinear model predictive control (NMPC) scheme based on a Hammerstein model can be developed with the same stability and numerical properties of a standard MPC scheme. In particular, if a quadratic performance index, and linear constraints are considered, the optimization algorithm in the NMPC scheme based on the Hammerstein model retains the convexity characteristic of linear MPC.

Remark 3. For control purposes, usually a state-space realization of the identified model is required. A state-space realization of the identified Hammerstein model can be constructed resorting to existing results on minimal state-space realizations for systems represented using orthonormal bases (see for instance Chapter 3 in [27]).

3. Wiener model identification

3.1. Problem formulation

A (multivariable) Wiener model is schematically depicted in Fig. 3. The model consists of the cascade of a LTI system with transfer function (matrix) $G(z) \in H_2^{m\times n}(\mathbb{T})$, followed by a zero-memory nonlinear element with input–output characteristic given by $\mathscr{N}(\cdot)$. In this case, $y_k \in \mathbb{R}^m$, $u_k \in \mathbb{R}^n$, and $v_k \in \mathbb{R}^m$, represent the system output, input, and process noise vectors at time k, respectively.

As in the case of the Hammerstein model, it will be assumed that the transfer function matrix $G(z)$ of the LTI subsystem is represented as an orthonormal basis expansion of the form (3). On the other hand, and as it is common in most existing identification algorithms for Wiener models [2,5,8], the nonlinear function $\mathscr{N}(\cdot): \mathbb{R}^m \to \mathbb{R}^m$ will be assumed to be invertible, [9,10] and that its inverse $\mathscr{N}^{-1}(\cdot)$ can be described as

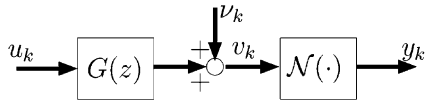

Fig. 3. Multivariable Wiener model.

$$\mathscr{N}^{-1}(y_k) = \sum_{i=1}^{r} a_i g_i(y_k), \tag{29}$$

where now $g_i(\cdot): \mathbb{R}^m \to \mathbb{R}^m$ $(i = 1,\ldots,r)$ are known (nonlinear) basis functions, and $a_i \in \mathbb{R}^{m\times m}$ $(i = 1,\ldots,r)$ 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 $(i = 2,\ldots,r)$ and b_ℓ $(\ell = 0,\ldots,p-1)$ 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.

3.2. Identification algorithm

The intermediate variable v_k in Fig. 3, can be written as

$$v_k = G(q)u_k + v_k, \tag{30}$$

and also as

$$v_k = \mathscr{N}^{-1}(y_k). \tag{31}$$

Equating the right-hand sides of the above two equations, and considering the parameterizations (3) and (29) of the linear and the nonlinear subsystems, respectively, the following equation is obtained

$$g_1(y_k) + \sum_{i=2}^{r} a_i g_i(y_k) = \sum_{\ell=0}^{p-1} b_\ell \mathscr{B}_\ell(q)u_k + v_k, \tag{32}$$

or equivalently

$$g_1(y_k) = -\sum_{i=2}^{r} a_i g_i(y_k) + \sum_{\ell=0}^{p-1} b_\ell \mathscr{B}_\ell(q)u_k + v_k, \tag{33}$$

which is a linear regression. Defining

$$\theta \triangleq [a_2, a_3, \ldots, a_r, b_0, b_1, \ldots, b_{p-1}]^{\mathrm{T}}, \tag{34}$$

$$\phi_k \triangleq [-g_2^{\mathrm{T}}(y_k), -g_3^{\mathrm{T}}(y_k), \ldots, -g_r^{\mathrm{T}}(y_k), \mathscr{B}_0(q)u_k^{\mathrm{T}}, \mathscr{B}_1(q)u_k^{\mathrm{T}}, \ldots, \mathscr{B}_{p-1}(q)u_k^{\mathrm{T}}]^{\mathrm{T}}, \tag{35}$$

Eq. (33) can be written as

[9] As pointed out in [2], this condition exclude the possibility of representing processes where the phenomenon of input multiplicity is present (see next footnote).

[10] *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}.

$$g_1(y_k) = \theta^{\mathrm{T}}\phi_k + v_k. \tag{36}$$

Now, an estimate $\hat{\theta}$ of θ can be computed by minimizing a quadratic criterion on the prediction errors $\epsilon_k = g_1(y_k) - \theta^{\mathrm{T}}\phi_k$ (i.e., the least squares estimate). It is well known [40] that this estimate is given by [11]

$$\hat{\theta} = (\Phi_N \Phi_N^{\mathrm{T}})^{-1}\Phi_N Y_N = \Phi_N^{\dagger} Y_N, \tag{37}$$

where the following definitions have been made

$$Y_N \triangleq [g_1(y_1), g_1(y_2), \ldots, g_1(y_N)]^{\mathrm{T}}, \tag{38}$$

$$\Phi_N \triangleq [\phi_1, \phi_2, \ldots, \phi_N]. \tag{39}$$

Now, estimates of the parameters a_i $(i = 2, \ldots, r)$ and b_ℓ $(\ell = 0, \ldots, p-1)$ can be computed by partitioning the estimate $\hat{\theta}$ in (37), according to the definition of θ in (34).

The nonlinear identification algorithm can then be summarized as follows.

Algorithm 2

Step 1: Compute the least squares estimate of θ as in (37).

Step 2: Compute estimates of the matrices a and b by partitioning the estimate $\hat{\theta}$ obtained in Step 1, according to the definition of θ in (34).

Remark 4. In this case, the consistency of the estimate $\hat{\theta}$ in (37), 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 $\{v_k\}$ at the same instant, even if the disturbance is a white noise process [40].

Remark 5. Comments similar to the ones in Remarks 2 and 3 also hold for the identified Wiener model.

4. Simulation examples

To illustrate the performance of the proposed identification algorithms, two simulation examples drawn from the process control literature, are provided in this section.

Example 1 (*pH neutralization process*). In this example, Algorithm 2 is used to identify a Wiener model based on 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. 4, and corresponds to a bench-scale plant at the University of California, Santa Barbara (see [34,35,44]).

[11] Provided the indicated inverse exists (see Footnote 6).

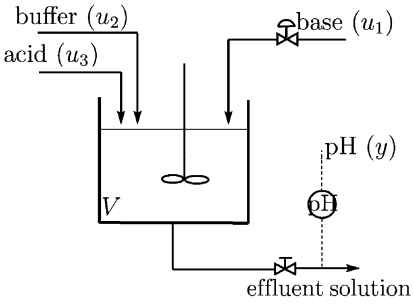

Fig. 4. Schematic representation of the pH neutralization process.

The inputs to the system are the (volumetric) base flow rate (u_1), the buffer flow rate (u_2), and the acid flow rate (u_3), while the output (y) is the pH of the effluent solution. The acid flow rate (u_3), as well as the volume (V) of the tank are assumed to be constant. Usually, the objective is to control the pH of the effluent solution by manipulating the base flow rate u_1, despite the variations of the unmeasured buffer flow rate u_2.

A simulation model, based on first principles, was presented in [35] introducing two *reaction invariants* (one related to a charge balance and the other to a balance on the carbonate ion) for each inlet/outlet stream. The reaction invariants will be denoted here as (W_{a1}, W_{b1}), (W_{a2}, W_{b2}), (W_{a3}, W_{b3}) and (W_a, W_b), for the base stream, the buffer stream, the acid stream, and the effluent solution, respectively. The model is highly nonlinear due to the implicit output equation, known as *titration curve* (Eq. (46)). The dynamic model for the reaction invariants of the effluent solution (W_a, W_b), in state-space form, is given by [35,44]:

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

$$h(x, y) = 0, \tag{41}$$

where

$$x \triangleq [x_1, x_2]^{\mathrm{T}} = [W_a, W_b]^{\mathrm{T}}, \tag{42}$$

$$f(x) = \left[\frac{u_3}{V}(W_{a3} - x_1), \frac{u_3}{V}(W_{b3} - x_2)\right]^{\mathrm{T}}, \tag{43}$$

$$g(x) = \left[\frac{1}{V}(W_{a1} - x_1), \frac{1}{V}(W_{b1} - x_2)\right]^{\mathrm{T}}, \tag{44}$$

$$p(x) = \left[\frac{1}{V}(W_{a2} - x_1), \frac{1}{V}(W_{b2} - x_2)\right]^{\mathrm{T}}, \tag{45}$$

$$h(x, y) = x_1 + 10^{y-14} - 10^{-y} + x_2 \frac{1 + 2 \times 10^{y-pK_2}}{1 + 10^{pK_1 - y} + 10^{y-pK_2}}. \tag{46}$$

The parameters pK_1 and pK_2 in (46) are the first and second disassociation constants of the weak acid H_2CO_3.

The nominal operating conditions of the system are given in [34,35,44], and they are reproduced in Table 1 for the sake of completeness.

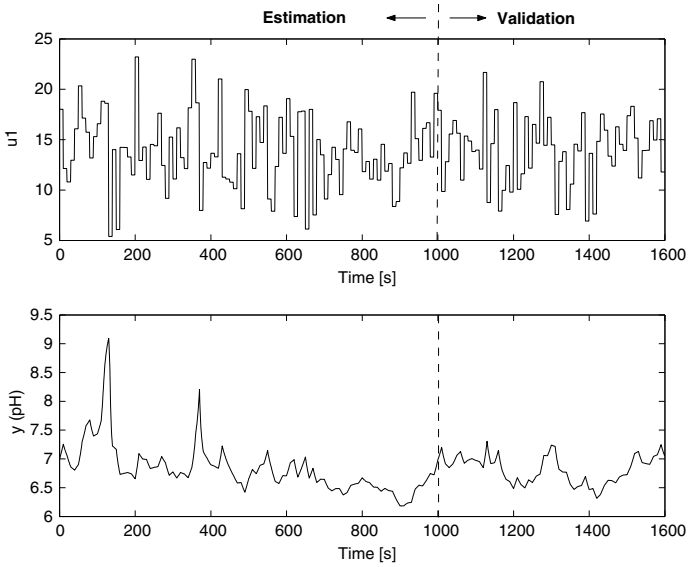

Fig. 5. Estimation (first 1000 points) and validation (remaining 600 points) input–output data.

Table 1
Nominal operating conditions of the pH neutralization process

$\bar{u}_3 = 16.60$ ml/s	$\bar{u}_2 = 0.55$ ml/s
$\bar{u}_1 = 15.55$ ml/s	$V = 2900$ ml
$W_{a1} = -3.05 \times 10^{-3}$ M	$W_{a2} = -3 \times 10^{-2}$ M
$W_{a3} = 3 \times 10^{-3}$ M	$W_a = -4.32 \times 10^{-4}$ M
$W_{b1} = 5 \times 10^{-5}$ M	$W_{b2} = 3 \times 10^{-2}$ M
$W_{b3} = 0$ M	$W_b = 5.28 \times 10^{-4}$ M
$pK_1 = 6.35$	$pK_2 = 10.25$
$\bar{y} = 7.0$	

For the purposes of identification, the model (40) and (41) was excited with band-limited white noise around the nominal value of the base flow rate, keeping the buffer flow rate and the acid flow rate constant in their nominal values. Changes in the base flow rate were produced every 10 s, with a maximum amplitude of ±70% of the nominal value. In order to simulate the more realistic situation of having measurement noise, the output of the system was corrupted with additive Gaussian white noise with zero mean and standard deviation $\sigma = 0.001$. A set of 1600 data points was collected from the simulation with a sampling time of 1 s. The first 1000 data were used for the estimation of the Wiener model, while the following 600 data were used for validation purposes. The estimation and validation input–output data are represented in Fig. 5.

The linear subsystem in the Wiener model was represented using the rational orthonormal bases with fixed poles (OBFP) studied in [27,28], that have the more common FIR, Laguerre [29,31], and Kautz [30,31] bases as special cases. The bases are defined as

$$\mathscr{B}_\ell(q) = \left(\frac{\sqrt{1-|\xi_\ell|^2}}{q-\xi_\ell}\right)\prod_{i=0}^{\ell-1}\left(\frac{1-\overline{\xi}_i q}{q-\xi_i}\right), \quad \ell \geqslant 1, \tag{47}$$

$$\mathscr{B}_0(q) = \frac{\sqrt{1-|\xi_0|^2}}{q-\xi_0}, \tag{48}$$

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 $(\xi_0, \xi_1, \ldots, \xi_{p-1})$, close 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 [27].

In order to determine the model order of the linear subsystem, as well as initial guesses for the location of the poles of the bases, the same input–output data were used to identify a linear model of the process using a subspace method. Namely, the CVA [12] algorithm [45] as implemented by Ljung in the n4sid routine in version 5 of the System Identification Toolbox for use with Matlab (hereafter referred as SIT) [46] was used for the identification of the linear model. The default option ('Prediction') for the 'focus' property of the n4sid routine was chosen. This option means that the models are determined by minimizing the prediction errors, and it corresponds to the optimal weighting from a statistical variance point of view. To determine the optimal model order, the default option 'best' was chosen for the input

[12] CVA: canonical variate analysis.

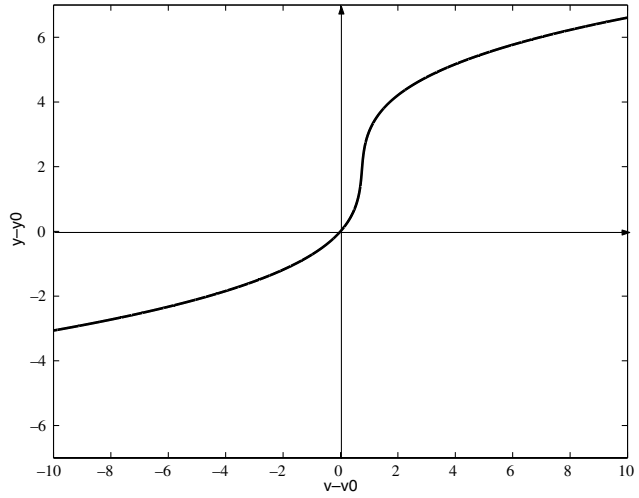

Fig. 6. Estimated static nonlinearity in the Wiener model.

argument 'order' of the routine. The algorithm computes and plots the singular values of the Hankel matrices of the impulse response for orders in the range 1:10, and the optimal order, in the sense that the singular values for higher orders are comparatively smaller, is calculated. Following this procedure, the resulting optimal model order was $n = 3$. The eigenvalues of the transition matrix of the identified linear model were used as initial guesses for the location of the poles of the OBFP in the nonlinear identification algorithm. The final choice for the location of the poles was {0.9897, 0.9784, 0.9479}.

On the other hand, a third order polynomial was used to represent the nonlinear static block of the Wiener model. This choice for the order of the polynomial was a compromise between model complexity and variance error (bias-variance trade-off), and the decision was taken based on the values of Akaike's Information theoretic Criterion (AIC) [13] for polynomials with orders in the range 2:5.

The estimated transfer function of the linear block in the Wiener model is

$$\widehat{G}(z) = \frac{0.0049z^2 - 0.0094z + 0.0045}{z^3 - 2.9160z^2 + 2.8339z - 0.9179}, \tag{49}$$

while the estimated static nonlinearity is represented in Fig. 6. Note that the amplitude of the output signal in Fig. 5 is large enough to cover the dynamic range of the static nonlinearity.

The true and estimated output corresponding to the identified Wiener model are represented in the left plot of Fig. 7, while the corresponding to the identified linear model are represented in the right plot of the same figure.

For the purposes of comparison between the identified Wiener and linear models, three different performance criteria have been computed. Namely, the mean square error (MSE), [14] the best fit, [15] and the variance accounted for (VAF) criterion. [16] The computed values of these criteria, as well as the total number of estimated parameters for the identified Wiener and linear models are shown in Table 2. Clearly, the Wiener model outperforms the linear model for each of the three considered criteria.

Example 2 (*Binary distillation column*). In this example, Algorithm 1 is used to estimate a Hammerstein model based on simulation data of a binary distillation column which separates a mixture of methanol and ethanol. The column was originally modelled by Weischedel and McAvoy in [47] based on component mass balances and steady-state energy balances. The resulting white-box model consists of a set of coupled nonlinear differential algebraic equations. In [47], the objective was to study the feasibility of control decoupling for this particular column. The column has 27 trays, a reboiler on the bottom tray and a condenser on the overhead stream. The column is fed at the 14th tray (counted from the bottom) with a 50–50% mixture of methanol and ethanol. A schematic representation of the process is depicted in Fig. 8.

This process represents a benchmark problem that has been studied by a number of researchers [36,47]. The inputs to the system are the feed flow rate (u_1), the feed composition (u_2), the vapor flow rate (u_3), and the reflux

[13] Akaike's Information theoretic Criterion is defined as:

$$\mathrm{AIC} = \log(V) + \frac{2d}{N},$$

where V is the loss function, d the number of estimated parameters, and N the number of validation data [40].

[14] The MSE is defined as

$$\mathrm{MSE} = \frac{1}{N}\sum_{k=1}^{N}\left(y_k - \hat{y}_k\right)^2,$$

where y_k denotes the real output, $\hat{y}_k$ denotes the output of the model, and N is the number of validation data [40].

[15] The best fit is defined as

$$\mathrm{FIT} = \left(1 - \frac{\|Y - Y_v\|}{\|Y - y_{\mathrm{mean}}\|}\right) \times 100,$$

where Y is a vector containing the output of the model when simulated with the validation input data, Y_v is a vector with the validation output data, and y_{mean} is the mean value of the output y.

[16] The VAF is defined as

$$\mathrm{VAF} = \max\left\{1 - \frac{\mathsf{Var}\{y - \hat{y}\}}{\mathsf{Var}\{y\}}, 0\right\} \times 100\%,$$

where $y = \{y_k\}_{k=1}^{N}$ denotes the real output sequence, $\hat{y} = \{\hat{y}_k\}_{k=1}^{N}$ denotes the model output sequence, and $\mathsf{Var}\{\cdot\}$ denotes the variance of a quasi-stationary signal.

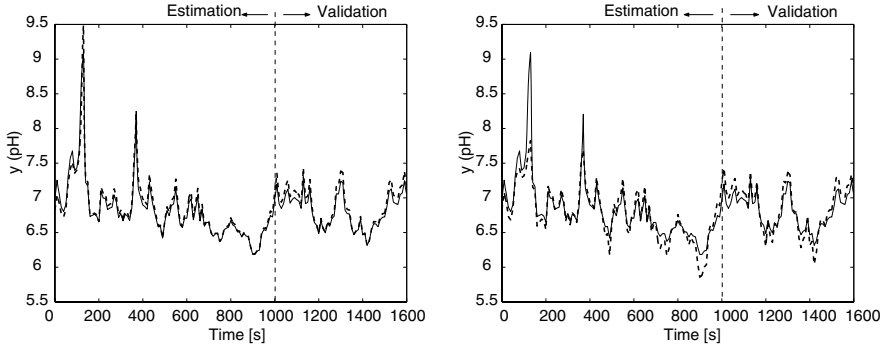

Fig. 7. Left plot: true (—) and estimated (- - -) output for the identified Wiener model. Right plot: true (—) and estimated (- - -) output for the identified linear model.

Table 2
Performance criteria for the identified Wiener and linear models

	# Parameters	MSE	FIT	VAF (%)
Wiener	5	0.0063	78.1098	96.4950
Linear	17	0.0280	53.7798	78.7740

Table 3
Nominal operating conditions of the distillation column

$\bar{u}_1 = 0.025$ m^3/s	$\bar{u}_2 = 0.5$
$\bar{u}_3 = 0.033$ m^3/s	$\bar{u}_4 = 1.75$
$\bar{y}_1 = 0.0125$ m^3/s	$\bar{y}_2 = 0.85$
$\bar{y}_3 = 0.0125$ m^3/s	$\bar{y}_4 = 0.15$

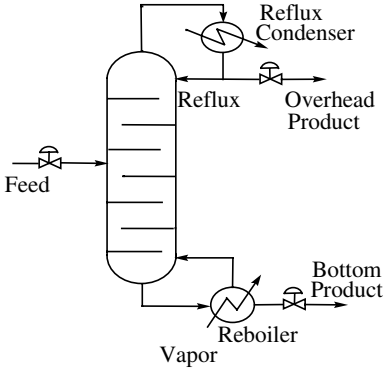

Fig. 8. Schematic representation of the distillation column.

ratio (u_4). Typically, the manipulated variables are the reflux ratio (u_4) and the vapor flow rate (u_3), while the other inputs are kept constant and their variations around the nominal values are considered as disturbances. This configuration corresponds to the so-called L-V control scheme [48]. The outputs of the system are the overhead flow rate (y_1), the overhead methanol composition (y_2), the bottom flow rate (y_3), and the bottom methanol composition (y_4). The outputs of interest (i.e., the controlled variables) are the overhead and the bottom methanol compositions (y_2 and y_4, respectively), and typically the control objective is to achieve a given percentage of methanol composition in the overhead stream and a given percentage of ethanol composition in the bottom stream. This is known as dual composition control problem. In the formulation considered here, the vapor flow rate is assumed to be constant and only the reflux ratio is considered as the manipulated variable. Similar assumptions has been considered in [2], where the authors estimate Hammerstein, Wiener and FBO models for a binary distillation column with similar characteristics.

The nominal operating conditions of the column considered in this example are given in Table 3.

For the purposes of identification, the white-box model of the column was excited with band-limited white noise around the nominal value of the reflux ratio (u_4), while all the other inputs were kept constant in their nominal values. Changes in the reflux ratio were produced every 200 s, with a maximum amplitude of ±100% of the nominal value. A set of 8000 data point were collected from the simulation with a sampling time of 1 s. The first 4000 data were used for the estimation of a Hammerstein model of the column, while the remaining 4000 data were used for validation purposes. The estimation and validation data are represented in Fig. 9.

The linear subsystem in the Hammerstein model was represented using the same rational OBFP (47) and (48), as in Example 1, while the static nonlinearity was represented by a polynomial. A procedure similar to the one employed in the previous example was used to determine the optimal model order of the linear subsystem, initial guesses for the location of the poles of the basis functions, and the optimal order of the polynomial representing the static nonlinearity.

As a result of the identification process a sixth order model was estimated as the linear part of the Hammerstein model, with eigenvalues (the chosen poles of the basis functions) at {0.9988, 0.9958, 0.9935, 0.9846, 0.9835, 0.98}. On the other hand, a third order poly-

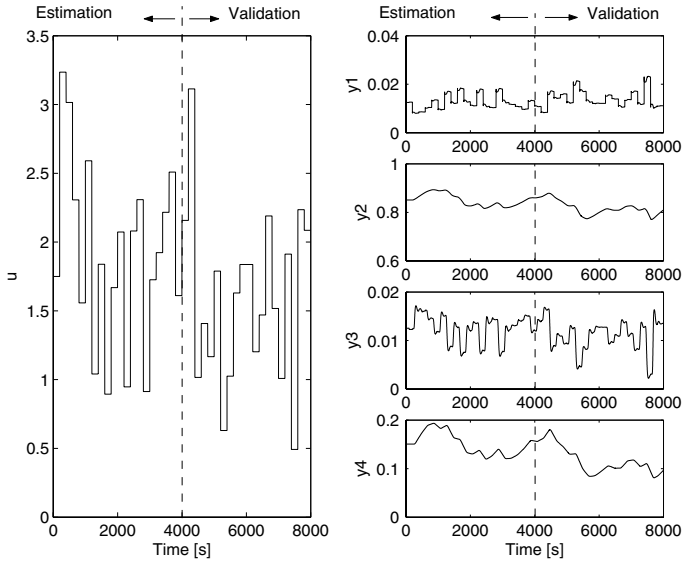

Fig. 9. Left plot: estimation (first 4000 points) and validation (remaining 4000 points) input data. Right plot: estimation (first 4000 points) and validation (remaining 4000 points) output data.

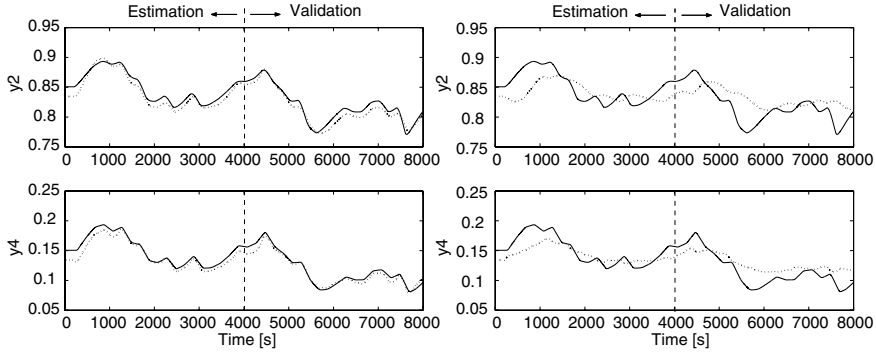

Fig. 10. Left plot: true (—) and estimated (···) outputs (y_2 and y_4) for the identified Hammerstein model. Right plot: true (—) and estimated (···) outputs (y_2 and y_4) for the identified linear model.

nomial was used to represent the static nonlinearity. The true (solid line) and estimated (dotted line) outputs (y_2 and y_4) for the identified Hammerstein model are represented in the left plots of Fig. 10, where a good agreement between them can be observed. The estimated nonlinear characteristic is represented in Fig. 11. Note that the amplitude of the input signal in Fig. 9 is large enough to cover the dynamic range of the static nonlinearity.

For the purposes of comparison, the same input–output data were used to identify a linear model of the process using the CVA subspace algorithm [45] as implemented by Ljung in the SIT [46]. A sixth order linear model was estimated with eigenvalues at $\{0.9940 \pm 0.0355i, 0.9969 \pm 0.0037i, 0.9892, 0.9990\}$. The true (solid line) and estimated (dotted line) outputs (y_2 and y_4) for the identified linear model are represented in

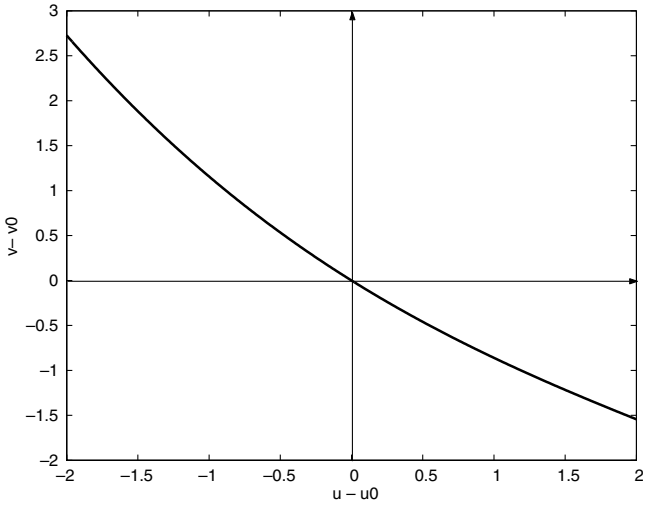

Fig. 11. Estimated nonlinear characteristic of the Hammerstein model.

Table 4
Performance criteria for the identified Hammerstein and linear models

	y_1	y_2	y_3	y_4
MSE (linear)	0.0091×10^{-3}	0.6329×10^{-3}	0.0133×10^{-3}	0.3287×10^{-3}
MSE (Hammerstein)	0.0064×10^{-3}	0.0642×10^{-3}	0.0011×10^{-3}	0.0525×10^{-3}
VAF (linear)	14.3277	36.0207	0	65.0629
VAF (Hammerstein)	40.2700	97.1433	89.4490	96.9223
FIT (linear)	7.4366	19.8203	−20.4021	40.6572
FIT (Hammerstein)	22.5716	74.4696	65.6285	76.2728

the right plots of Fig. 10. To compare the predictive capabilities of the identified Hammerstein and linear models of the column, the same three criteria as in the previous example were computed. Namely, the MSE, the VAF and the best fit criteria were computed, and the results are shown in Table 4. Clearly, the Hammerstein model outperforms the linear model for each of the three considered criteria.

5. Concluding remarks

In this paper, noniterative algorithms 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 colored noise. For the case of the Wiener model, consistency of the estimates can be guaranteed only in the noise free case. The key issue in the derivation of the results is the representation of the linear and nonlinear parts of the system using basis functions which allows to put the system in linear regressor form. In addition, the use of rational orthonormal bases for the representation of the linear subsystem allows a priori information one can have about the dominant dynamics, to be incorporated in the identification process, to improve the estimation accuracy. The suitability of the proposed methods for their use in the identification of pH neutralization processes and distillation columns has been illustrated through simulation of two benchmark processes.

Appendix A

Proof of Theorem 1. Let the SVD of the matrix $\widehat{\Theta}_{ab} \in \mathbb{R}^{nr\times mp}$ be given by

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

where k is the rank of $\widehat{\Theta}_{ab}$. Appealing to Theorem 2.5.2 in [42], 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

$$\Theta = \Theta_n \triangleq \sum_{i=1}^{n} \sigma_i u_i v_i^{\mathrm{T}}, \tag{A.2}$$

and the approximation error is given by

$$\|\widehat{\Theta}_{ab} - \Theta_n\|_2^2 = \sigma_{n+1}^2. \tag{A.3}$$

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

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

what concludes the proof, by equating

$$\hat{a} = U_1, \tag{A.4}$$

$$\hat{b} = V_1\Sigma_1. \qquad \square \tag{A.5}$$

References

[1] R. Haber, H. Unbehauen, Structure identification of nonlinear dynamic systems—A survey on input/output approaches, Automatica 26 (4) (1990) 651–677.
[2] R. Pearson, M. Pottmann, Gray-box identification of block-oriented nonlinear models, Journal of Process Control 10 (2000) 301–315.
[3] E. Eskinat, S. Johnson, W. Luyben, Use of Hammerstein models in identification of nonlinear systems, AIChE Journal 37 (2) (1991) 255–268.
[4] W. Greblicki, Nonparametric identification of Wiener systems by orthogonal series, IEEE Transactions on Automatic Control 39 (10) (1994) 2077–2086.
[5] T. Wigren, Recursive prediction error identification using the nonlinear Wiener model, Automatica 29 (4) (1993) 1011–1025.
[6] T. Wigren, Convergence analysis of recursive identification algorithms based on the nonlinear Wiener model, IEEE Transactions on Automatic Control 39 (11) (1994) 2191–2206.
[7] M. Pottmann, R. Pearson, Block-oriented NARMAX models with output multiplicities, AIChE Journal 44 (1998) 131–140.
[8] A. Kalafatis, N. Arifin, L. Wang, W. Cluett, A new approach to the identification of pH processes based on the Wiener model, Chemical Engineering Science 50 (23) (1995) 3693–3701.
[9] C. Chou, H. Bloemen, V. Verdult, T. van den Boom, T. Backx, M. Verhaegen, Nonlinear identification of high purity distillation columns, in: Proceedings of the IFAC Symposium on System Identification SYSID 2000, Santa Barbara, CA, USA, June 2000, pp. 415–420.
[10] M. Korenberg, Identification of biological cascades of linear and static nonlinear systems, in: Proceedings of the 16th Midwest Symposium on Circuit Theory, 1978, pp. 2.1–2.9.
[11] J. Stapleton, S. Bass, Adaptive noise cancellation for a class of nonlinear dynamic reference channels, IEEE Transactions on Circuits and Systems CAS-32 (1985) 143–150.

[12] K. Fruzzetti, A. Palazoglu, K. McDonald, Nonlinear model predictive control using Hammerstein models, Journal of Process Control 7 (1) (1997) 31–41.
[13] S. Norquay, A. Palazoglu, J. Romagnoli, Model predictive control based on Wiener models, Chemical Engineering Science 53 (1) (1998) 75–84.
[14] H. Bloemen, C. Chou, T. van den Boom, V. Verdult, M. Verhaegen, T. Backx, Wiener model identification and predictive control for dual composition control of a distillation column, Journal of Process Control 11 (2001) 601–620.
[15] K. Narendra, P. Gallman, An iterative method for the identification of nonlinear systems using a Hammerstein model, IEEE Transactions on Automatic Control AC-11 (1966) 546–550.
[16] S. Billings, S. Fakhouri, Identification of nonlinear systems using correlation analysis, Proceedings IEE 125 (1978) 691.
[17] S. Billings, S. Fakhouri, Identification of a class of nonlinear systems using correlation analysis, Proceedings IEE, Corres. 125 (1978) 1307.
[18] S. Billings, Identification of nonlinear systems—A survey, Proceedings of IEE, Part D 127 (1980) 272–285.
[19] S. Billings, S. Fakhouri, Identification of systems containing linear dynamic and static nonlinear elements, Automatica 18 (1) (1982) 15–26.
[20] P. Stoica, On the convergence of an iterative algorithm used for Hammerstein system identification, IEEE Transactions on Automatic Control AC-26 (4) (1981) 967–969.
[21] M. Boutayeb, M. Darouach, Recursive identification method for MISO Wiener–Hammerstein model, IEEE Transactions on Automatic Control 40 (2) (1995) 287–291.
[22] W. Greblicki, M. Pawlak, Nonparametric identification of Hammerstein systems, IEEE Transactions on Information Theory 35 (2) (1989) 409–418.
[23] S. Billings, S. Fakhouri, Identification of nonlinear systems using the Wiener model, Electronics Letters 13 (1978) 502–504.
[24] H. Khalil, Nonlinear Systems, Prentice Hall, Englewood Cliffs, NJ, 1996.
[25] E. Bai, An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems, Automatica 34 (3) (1998) 333–338.
[26] J.C. Gómez, E. Baeyens, Identification of multivariable Hammerstein systems using rational orthonormal bases, in: Proceedings of the 39th IEEE Conference on Decision and Control, Sydney, Australia, December 2000, pp. 2849–2854.
[27] J.C. Gómez, Analysis of dynamic system identification using rational orthonormal bases, Ph.D. Thesis, The University of Newcastle, Australia, PS file downloadable from <http://www.fceia.unr.edu.ar/~jcgomez>.
[28] B. Ninness, F. Gustafsson, A unifying construction of orthonormal bases for system identification, IEEE Transactions on Automatic Control AC-42 (4) (1997) 515–521.
[29] B. Wahlberg, System identification using Laguerre models, IEEE Transactions on Automatic Control AC-36 (5) (1991) 551–562.
[30] B. Wahlberg, System identification using Kautz models, IEEE Transactions on Automatic Control AC-39 (6) (1994) 1276–1282.
[31] B. Wahlberg, P. Mäkilä, On approximation of stable linear dynamical systems using Laguerre and Kautz functions, Automatica 32 (5) (1996) 693–708.
[32] P. Van den Hof, P. Heuberger, J. Bokor, System identification with generalized orthonormal basis functions, Automatica 31 (12) (1995) 1821–1834.
[33] J.C. Gómez, E. Baeyens, Identification of nonlinear systems using orthonormal bases, in: Proceedings of the IASTED International Conference on Intelligent Systems and Control, Clearwater, FL, USA, November 2001, pp. 126–131.
[34] M. Henson, D. Seborg, Nonlinear adaptive control of a pH neutralization process, in: Proceedings of the American Control Conference, 1992, pp. 2586–2590.
[35] M. Henson, D. Seborg, Adaptive nonlinear control of a pH neutralization process, IEEE Transactions on Control Systems Technology 2 (3) (1994) 169–182.
[36] F. Doyle III, E. Gatzke, R. Parker, Process Control Modules—A Software Laboratory for Control Design, in: International Series in the Physical and Chemical Engineering Sciences, Prentice Hall, Englewood Cliffs, NJ, 2000.
[37] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P. Glorennec, H. Hjalmarsson, A. Juditsky, Nonlinear black-box modeling in system identification: a unified approach, Automatica 31 (12) (1995) 1691–1724.
[38] A. Juditsky, H. Hjalmarsson, A. Benveniste, B. Delyon, L. Ljung, J. Sjöberg, Q. Zhang, Nonlinear black-box modeling in system identification: mathematical foundations, Automatica 31 (12) (1995) 1725–1750.
[39] S. Rangan, G. Wolodkin, K. Poolla, New results for Hammerstein system identification, in: Proceedings of the 34th IEEE Conference on Decision and Control, New Orleans, USA, December 1995, pp. 697–702.
[40] L. Ljung, System Identification: Theory for the User, second ed., Prentice-Hall, Inc., Englewood Cliffs, NJ, 1999.
[41] T. Södeström, P. Stoica, System Identification, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1989.
[42] G. Golub, C. Van Loan, Matrix Computations, second ed., The Johns Hopkins University Press, Baltimore and London, 1989.
[43] J. Maciejowski, Predictive Control with Constraints, Prentice Hall, 2002.
[44] Nonlinear process control M. Henson, D. Seborg (Eds.), Feedback Linearizing Control, Prentice Hall PTR, NJ, 1997 (Chapter 4).
[45] W. Larimore, Canonical variate analysis in identification, filtering, adaptive control, in: Proceedings of the 29th Conference on Decision and Control, Honolulu, Hawaii, December 1990, IEEE, pp. 596–604.
[46] L. Ljung, System Identification Toolbox, User's Guide, ver. 5, The MathWorks, Inc., Natick, MA, 2000.
[47] K. Weischedel, T. McAvoy, Feasibility of decoupling in conventionally controlled distillation columns, Industrial and Engineering Chemical Fundamentals 19 (4) (1980) 379–384.
[48] W. Luyben, B. Tyréus, M. Luyben, Plantwide Process Control, McGraw Hill, New York, 1999.