European Journal of Control (2005)11:1–10

European Journal of Control

Subspace Identification of Multivariable Hammerstein and Wiener Models*

Juan C. Gómez[1,**] and Enrique Baeyens[2]

[1]Laboratory for System Dynamics and Signal Processing, FCEIA, Universidad Nac. de Rosario Riobamba 245 Bis, 2000 Rosario, Argentina; [2]Department of Systems Engineering and Automatic Control, ETSII, Universidad de Valladolid Paseo del Cauce s/n, 47011 Valladolid, Spain

In this paper, subspace-based algorithms for the simultaneous identification of the linear and nonlinear parts of multivariable Hammerstein and Wiener models are presented. The proposed algorithms consist basically of two steps. The first one is a standard (linear) subspace algorithm applied to an equivalent linear system whose inputs (respectively outputs) are filtered (by the nonlinear functions describing the static nonlinearities) versions of the original inputs (respectively outputs). The second step consists of a 2-norm minimization problem which is solved via Singular Value Decomposition. Under weak assumptions, consistency of the estimates can be guaranteed. The performance of the proposed identification algorithms is illustrated through simulation examples.

Keywords: Hammerstein and Wiener Models; Nonlinear System Identification; Singular Value Decomposition; Subspace Methods

1. Introduction

Modeling, identification, and control of nonlinear systems have been the subject of many research activities in the last decades. In contrast to linear models that only approximate the system around a given operating point, nonlinear models are able to describe the global behavior of the system over the entire operating range. One of the most frequently studied class of nonlinear models is the one corresponding to the so-called *block-oriented* models, which consist of the interconnection of linear time-invariant (LTI) systems and static (memoryless) nonlinearities. Within this class, two of the most frequently studied models are:

- the *Hammerstein model*, where the static nonlinearity is followed by an LTI system in a cascade connection (see [2–5,9,12,24,25,28] for different identification algorithms for Hammerstein models), and
- the *Wiener model*, in which the order of the linear and nonlinear blocks in the cascade connection is reversed (see for instance [15,33,34] for different identification methods for Wiener models). Wiener models have the capability of approximating, with arbitrary accuracy, any fading memory nonlinear time-invariant system [6].

These models have been successfully used to represent nonlinear systems in a number of practical applications in the areas of chemical processes [9,18,25], biological processes [19], signal processing [27], and control [11].

*A preliminary version of this paper was presented by the authors at the 15th IFAC World Congress, Barcelona, Spain, July 2002.

**E-mail: jcgomez@fceia.unr.edu.ar
Correspondence to: E. Baeyens.
E-mail: enrbae@eis.uva.es

Received 23 July 2003; Accepted 17 December 2004.
Recommended by E. Mosca and D. Clarke.

In recent years, considerable amount of research has been devoted to the development of new identification methods that are able to deliver reliable state-space models of multivariable LTI systems directly from input–output data, and that require a modest computational load without the need of iterative optimization procedures. These techniques have become collectively known as Subspace-based State Space System IDentification (4SID) methods (see [20,29,30], and the references therein). The methods have their origin in state-space realization theory as developed in the 1960/70's, and the main computational tools are QR and singular value decomposition (SVD). Although there is a well-developed theory for subspace methods for LTI systems, it is not so for nonlinear systems. Among some recent contributions in this area, the works by Verhaegen and Westwick on subspace-based identification of MIMO Hammerstein and Wiener models [31,32], and the works by Chen and coauthors [7,8], and by Favoreel and coauthors [10] on subspace identification of bilinear systems, can be mentioned.

In this paper, new subspace-based algorithms for the identification of Hammerstein and Wiener models are presented. The proposed algorithms consist of two basic steps. The first step is a standard (linear) subspace algorithm applied on an equivalent linear system whose inputs (respectively outputs) are filtered (by the nonlinear functions describing the static nonlinearities) versions of the original inputs (respectively outputs), while the second step consists of a 2-norm minimization problem which is solved via SVD.

In contrast to the subspace algorithms for identification of Hammerstein models in Ref. [31], and of Wiener models in Ref. [32], which extend only the MOESP[1] family of linear subspace methods, the subspace algorithms presented in this paper extend any linear subspace method including the MOESP, the N4SID,[2] and the CVA[3] to this class of nonlinear models. In addition, the subspace algorithms presented here are more general than those in Refs [31,32], in the sense that arbitrary basis functions (not restricted to polynomials) can be used for the representation of the nonlinearities. Furthermore, a simulation example is presented for which the algorithm in Ref. [31] fails to identify a Hammerstein model for the system, while the algorithm proposed in this paper successfully identifies the model.

[1]MOESP stands for Multivariable Output Error State sPace [30].
[2]N4SID stands for Numerical algorithms for Subspace State-Space System IDentification [29].
[3]CVA stands for Canonical Variate Analysis [20].

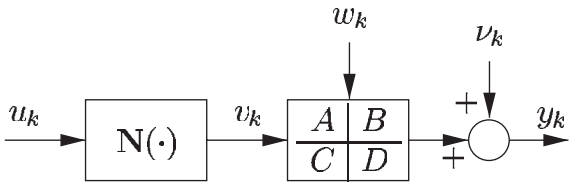

Fig. 1. Multivariable Hammerstein model.

The rest of the paper is organized as follows. In Section 2, the Hammerstein model is introduced, the identification problem is formulated and the subspace algorithm is derived. The same is done in Section 3 for the Wiener model. Finally, simulation examples illustrating the performance of the algorithms are presented in Section 4, and some conclusions are given 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 $\mathbf{N}(\cdot)$ in cascade with LTI system with state-space representation

$$x_{k+1} = Ax_k + Bv_k + \omega_k, \tag{1}$$

$$y_k = Cx_k + Dv_k + \nu_k, \tag{2}$$

where $y_k \in \mathbb{R}^m$, $x_k \in \mathbb{R}^n$, $v_k \in \mathbb{R}^p$, $\omega_k \in \mathbb{R}^n$, and $\nu_k \in \mathbb{R}^m$, are the LTI system output, state, input, process noise and output measurement noise vectors at time k, respectively, and where A, B, C and D are the (unknown) system matrices of appropriate dimensions.

It will be assumed that the nonlinear zero-memory block can be described by a linear combination of basis functions in the form

$$v_k = \mathbf{N}(u_k) = \sum_{i=1}^{r} \alpha_i g_i(u_k), \tag{3}$$

where $g_i(\cdot) : \mathbb{R}^p \to \mathbb{R}^p, (i = 1, \ldots, r)$, are the assumed known basis functions, $\alpha_i \in \mathbb{R}^{p \times p}, (i = 1, \ldots, r)$ are unknown matrix parameters, and where $u_k \in \mathbb{R}^p$ is the Hammerstein model input vector at time k. Typically, the basis functions are polynomials,[4] but they can also be basis functions generated by translations and dilations of a *mother* function (e.g., wavelets, or Radial Basis Functions).

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

The identification problem is to estimate the unknown parameter matrices α_i, $(i = 1, \ldots, r)$, and A, B, C, D, characterizing, respectively, the nonlinear and the linear parts of the system, and the model order n, from an N-point data record $\{u_k, y_k\}_{k=1}^{N}$ of observed input–output measurements.

2.2. Subspace Identification Algorithm

Substituting Eq. (3) into (1) and (2), these last two equations can be written as

$$x_{k+1} = Ax_k + \sum_{i=1}^{r} B\alpha_i g_i(u_k) + \omega_k, \tag{4}$$

$$y_k = Cx_k + \sum_{i=1}^{r} D\alpha_i g_i(u_k) + \nu_k. \tag{5}$$

Notice from Eqs (4) and (5) that the parametrization (1)–(3) is not unique, since any parameter matrices $B\beta, D\beta$ and $\beta^{-1}\alpha_i$, for some nonsingular matrix $\beta \in \mathbb{R}^{p\times p}$, provide the same description (4) and (5). In other words, any identification experiment cannot distinguish between the parameters B, D, α_i and $B\beta, D\beta, \beta^{-1}\alpha_i$, respectively. To obtain a one-to-one parametrization, i.e., for the system to be identifiable, additional constraints must be imposed on the parameters. A technique that can be used to obtain uniqueness is to normalize the parameter matrices α_i, i.e. to assume for instance that $\|[\alpha_1\, \alpha_2 \cdots \alpha_r]^{\mathrm{T}}\|_2 = 1$. A similar methodology was employed in [1] for a scalar Hammerstein–Wiener model. Under this assumption, the parametrization (1)–(3) is unique.

Defining now $\widetilde{B} \triangleq [B\alpha_1, \ldots, B\alpha_r]$, $\widetilde{D} \triangleq [D\alpha_1, \ldots \alpha_r]$, and $U_k \triangleq [g_1(u_k)^{\mathrm{T}}, \ldots, g_r(u_k)^{\mathrm{T}}]^{\mathrm{T}}$, Eqs (4) and (5) can be written as

$$x_{k+1} = Ax_k + \widetilde{B}U_k + \omega_k, \tag{6}$$

$$y_k = Cx_k + \widetilde{D}U_k + \nu_k. \tag{7}$$

Equations (6) and (7) can be interpreted as a state-space realization of an equivalent LTI system whose input U_k is a filtered (by the assumed known basis functions $g_i(\cdot)$) version of the original input u_k. Any available subspace identification algorithm (such as the N4SID algorithm by Van Overschee and de Moor [29], the MOESP algorithm by Verhaegen [30], or the CVA algorithm by Larimore [20,21]) can then be employed to obtain estimates $\widehat{A}, \widehat{\widetilde{B}}, \widehat{C}$, and $\widehat{\widetilde{D}}$ of the system matrices $A, \widetilde{B}, C$, and $\widetilde{D}$, respectively, from input–output data.

Defining $\alpha \triangleq [\alpha_1\, \alpha_2 \cdots \alpha_r]^{\mathrm{T}}$, matrices $\widetilde{B}$ and $\widetilde{D}$ can be written as $\widetilde{B} = B\alpha^{\mathrm{T}}$, and $\widetilde{D} = D\alpha^{\mathrm{T}}$, which can be expressed in a combined form as

$$\left[\widetilde{B}^T \widetilde{D}^T\right]^{\mathrm{T}} \triangleq \Theta_{BD} = \begin{bmatrix} B \\ D \end{bmatrix} \alpha^{\mathrm{T}}. \tag{8}$$

The problem now is how to compute estimates of the parameter matrices B, D and α from an estimate $\widehat{\Theta}_{BD}$ of the matrix Θ_{BD}. It is clear that the closest, in the 2-norm[5] sense, estimates $\widehat{B}, \widehat{D}$ and $\widehat{\alpha}$ are such that

$$\left(\widehat{B}, \widehat{D}, \widehat{\alpha}\right) = \arg\min_{B,D,\alpha} \left\{ \left\| \widehat{\Theta}_{BD} - \begin{bmatrix} B \\ D \end{bmatrix} \alpha^{\mathrm{T}} \right\|_2^2 \right\}. \tag{9}$$

The solution to this optimization problem is provided by the SVD [13] of the matrix $\widehat{\Theta}_{BD}$. The result is summarized in the following theorem.

Theorem 2.1. Let $\widehat{\Theta}_{BD} \in \mathbb{R}^{(n+m)\times rp}$ have rank $s > p$, and let the *economy-size* SVD of $\widehat{\Theta}_{BD}$ be given by

$$\widehat{\Theta}_{BD} = U_s \Sigma_s V_s^T = \sum_{i=1}^{s} \sigma_i u_i v_i^{\mathrm{T}} \tag{10}$$

where Σ_s is a diagonal matrix containing the s nonzero singular values $(\sigma_i, i = 1, \ldots, s)$ of $\widehat{\Theta}_{BD}$ in nonincreasing order, and where the matrices $U_s = [u_1\, u_2 \cdots u_s] \in \mathbb{R}^{(n+m)\times s}$ and $V_s = [v_1\, v_2 \cdots v_s] \in \mathbb{R}^{rp\times s}$ contain only the first s columns of the unitary matrices $U \in \mathbb{R}^{(n+m)\times(n+m)}$ and $V \in \mathbb{R}^{rp\times rp}$ provided by the full SVD of $\widehat{\Theta}_{BD}$,

$$\widehat{\Theta}_{BD} = U\Sigma V^{\mathrm{T}}, \tag{11}$$

respectively. Then, the matrices $\widehat{\alpha} \in \mathbb{R}^{rp\times p}$, $\widehat{B} \in \mathbb{R}^{n\times p}$, and $\widehat{D} \in \mathbb{R}^{m\times p}$ that minimize the norm

$$\left\| \widehat{\Theta}_{BD} - \begin{bmatrix} \widehat{B} \\ \widehat{D} \end{bmatrix} \widehat{\alpha}^{\mathrm{T}} \right\|_2^2,$$

are given by

$$\left(\begin{bmatrix} \widehat{B} \\ \widehat{D} \end{bmatrix}, \widehat{\alpha} \right) = (U_1 \Sigma_1, V_1), \tag{12}$$

where $\Sigma_1 = \mathrm{diag}\,\{\sigma_1, \sigma_2, \ldots, \sigma_p\}$, $U_1 \in \mathbb{R}^{(n+m)\times p}$, and $V_1 \in \mathbb{R}^{rp\times p}$, are given by the following partition of the

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

$$\|A\|_2 = \sup_{w\neq 0} \frac{\|Aw\|_2}{\|w\|_2} = \bar{\sigma}(A)$$

where $\bar{\sigma}(A)$ is the largest singular value of matrix A.

economy-size SVD in (10),

$$\widehat{\Theta}_{BD} = [U_1 \quad U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^{\mathrm{T}} \\ V_2^{\mathrm{T}} \end{bmatrix}, \tag{13}$$

and the approximation error is given by

$$\left\| \widehat{\Theta}_{BD} - \begin{bmatrix} \widehat{B} \\ \widehat{D} \end{bmatrix} \widehat{\alpha}^{\mathrm{T}} \right\|_2^2 = \sigma_{p+1}^2. \tag{14}$$

Proof. The result is a direct application of Theorem 2.5.3 (pp. 72–73) in Ref. [13]. □

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

Algorithm 2.1.
Step 1: Compute estimates $\left(\widehat{A}, \widehat{\widetilde{B}}, \widehat{C}, \widehat{\widetilde{D}}\right)$ of the systems matrices $\left(A, \widetilde{B}, C, \widetilde{D}\right)$ in (6) and (7) using any available subspace algorithm for LTI systems.

Step 2: Based on the estimates $\widehat{\widetilde{B}}$ and $\widehat{\widetilde{D}}$ compute an estimate $\widehat{\Theta}_{BD}$ of the matrix Θ_{BD} defined in (8).

Step 3: Compute the *economy-size* SVD of $\widehat{\Theta}_{BD}$ as in Theorem 2.1, and the partition of this decomposition as in Eq. (13).

Step 4: Compute the estimates of the parameter matrices B, D and α as

$$\begin{bmatrix} \widehat{B} \\ \widehat{D} \end{bmatrix} = U_1 \Sigma_1,$$

and $\widehat{\alpha} = V_1$, respectively, with U_1, V_1 and Σ_1 defined as in Theorem 2.1.

Remark 2.1. It is important to note that the algorithm intrinsically delivers estimates that satisfy the uniqueness condition $\|\alpha\|_2 = 1$, since matrix V_1 in the SVD of $\widehat{\Theta}_{BD}$ is a unitary matrix.

Under some assumptions on the persistency of excitation of the inputs (which depend on the particular subspace identification method used as the first step of the algorithm[6]) the estimates $\left(\widehat{A}, \widehat{\widetilde{B}}, \widehat{C}, \widehat{\widetilde{D}}\right)$ are consistent in the sense that they converge to the *true* values as the number of data points $N \to \infty$. The convergence of the estimates $\widehat{\widetilde{B}}$ and $\widehat{\widetilde{D}}$ implies that of $\widehat{B}, \widehat{D}$ and $\widehat{\alpha}$. The result is summarized in the following theorem.

Theorem 2.2. Let $\widehat{\widetilde{B}}$ and $\widehat{\widetilde{D}}$ be consistent estimates computed using the identification Algorithm 2.1. Then, under the uniqueness condition, the estimates $\widehat{B}, \widehat{D}$, and $\widehat{\alpha}$ provided by Algorithm 2.1 are also consistent,[7] in the sense that $\widehat{B} \xrightarrow{\text{a.s.}} B, \widehat{D} \xrightarrow{\text{a.s.}} D$, and $\widehat{\alpha} \xrightarrow{\text{a.s.}} \alpha$, respectively, as $N \to \infty$.

Proof. See Appendix. □

Remark 2.2. A similar procedure was suggested in Ref. [26]. However, the formulation here is more general, in the sense that any available subspace algorithm can be used as a first step, and proofs of consistency are provided that are not given in Ref. [26].

Remark 2.3. In comparison with the work in Ref. [31], the algorithm presented here is more general in the sense that any family of subspace algorithms (not limited to the MOESP class as in Ref. [31], but also including the N4SID and CVA classes) can be used as a first step of the algorithm, and that a more general representation of the nonlinearity using arbitrary basis functions (not limited to polynomial bases as in Ref. [31]) can be employed.

3. Wiener Model Identification

3.1. Problem Formulation

A (multivariable) Wiener model is schematically depicted in Fig. 2. The model consists of the cascade interconnection of an LTI system followed by a zero-memory nonlinear element with input–output characteristic given by $\mathbf{N}(\cdot)$. The LTI subsystem has a state-space representation of the form

$$x_{k+1} = Ax_k + Bu_k + \omega_k, \tag{15}$$
$$v_k = Cx_k + Du_k + \nu_k, \tag{16}$$

where A, B, C and D, are the system matrices of appropriate dimensions, and where $x_k \in \mathbb{R}^n$, $v_k \in \mathbb{R}^m$,

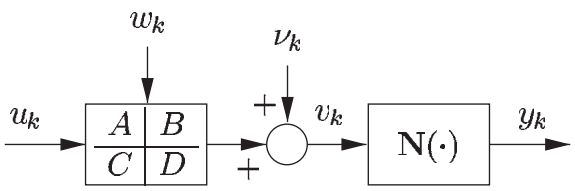

Fig. 2. Multivariable Wiener model.

[6]The reader is referred to Refs [29,30], and [20] for the consistency conditions for the N4SID, MOESP, and CVA algorithms, respectively.

[7]Here, the notation $\widehat{X} \xrightarrow{\text{a.s.}} X$ stands for *almost sure* convergence, or convergence with probability 1.

$u_k \in \mathbb{R}^p$, and $\nu_k \in \mathbb{R}^m$, represent the LTI system state, output, input, and process noise vectors at time k, respectively.

It will be assumed that the nonlinear function $\mathbf{N}(\cdot) : \mathbb{R}^m \to \mathbb{R}^m$ is invertible, and that its inverse $\mathbf{N}^{-1}(\cdot)$ can be described as

$$\mathbf{N}^{-1}(y_k) = \sum_{i=1}^{r} \alpha_i g_i(y_k), \tag{17}$$

where now $g_i(\cdot) : \mathbb{R}^m \to \mathbb{R}^m, (i = 1, \ldots, r)$ are known basis functions, and $\alpha_i \in \mathbb{R}^{m\times m}, (i = 1, \ldots, r)$ are unknown matrix parameters. With this representation for the static nonlinearity, Eq. (16) can be written as

$$\alpha Y_k \triangleq \sum_{i=1}^{r} \alpha_i g_i(y_k) = Cx_k + Du_k + \nu_k, \tag{18}$$

where $\alpha \triangleq [\alpha_1, \ldots, \alpha_r], Y_k \triangleq [g_1^{\mathrm{T}}(y_k), \ldots, g_r^{\mathrm{T}}(y_k)]^{\mathrm{T}}$. The Wiener model can then be described as

$$x_{k+1} = Ax_k + Bu_k + \omega_k, \tag{19}$$

$$Y_k = \widetilde{C}x_k + \widetilde{D}u_k + \widetilde{\nu}_k, \tag{20}$$

with $\widetilde{C} \triangleq \alpha^{\dagger}C, \widetilde{D} \triangleq \alpha^{\dagger}D, \widetilde{\nu}_k \triangleq \alpha^{\dagger}\nu_k$ and where $\alpha^{\dagger}$ stands for the left pseudoinverse of α. As in the case of the Hammerstein model, here also the condition $\|\alpha\|_2 = 1$ must be imposed in order to have uniqueness in the representation (19)–(20).

As in the case of the Hammerstein model, here also any available subspace identification algorithm (such as the N4SID algorithm by Van Overschee and de Moor [29], the MOESP algorithm by Verhaegen [30], or the CVA algorithm by Larimore [20,21]) can be employed to obtain estimates of the system matrices $A, B, \widetilde{C}$, and $\widetilde{D}$ from input–output data.

3.2. Subspace Identification Algorithm

Given estimates of the matrices $A, B, \widetilde{C}$, and $\widetilde{D}$, the problem is how to compute estimates of the matrices C, D and α. Proceeding similarly as in the Hammerstein model, the best (in the mean squares sense) estimates of matrices C, D and α are such that

$$\left(\widehat{C}, \widehat{D}, \widehat{\alpha}^{\dagger}\right) = \underset{C,D,\alpha^{\dagger}}{\arg\min} \left\{ \left\| \left[\widehat{\widetilde{C}}\ \widehat{\widetilde{D}}\right] - \alpha^{\dagger}[C\ D] \right\|_2^2 \right\} \tag{21}$$

The solution to this minimization problem is provided by the SVD of the matrix $[\widehat{\widetilde{C}}\ \widehat{\widetilde{D}}]$. A re-statement, *mutatis mutandi*, of Theorem 2.1 could be made for this case. The Subspace Identification Algorithm for the Wiener model can be summarized as follows.

Algorithm 3.1.

Step 1: Compute estimates $\left(\widehat{A}, \widehat{B}, \widehat{\widetilde{C}}, \widehat{\widetilde{D}}\right)$ of the systems matrices $(A, B, \widetilde{C}, \widetilde{D})$ in (19)–(20) using any available subspace algorithm for LTI systems.

Step 2: Compute the economy-size SVD of $[\widehat{\widetilde{C}}\ \widehat{\widetilde{D}}]$ as

$$[\widehat{\widetilde{C}}\ \widehat{\widetilde{D}}] = U_s \Sigma_s V_s^{\mathrm{T}} = \sum_{i=1}^{s} \sigma_i u_i v_i^{\mathrm{T}} \tag{22}$$

and the partition of this decomposition as

$$[\widehat{\widetilde{C}}\ \widehat{\widetilde{D}}] = [U_1\ U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^{\mathrm{T}} \\ V_2^{\mathrm{T}} \end{bmatrix}, \tag{23}$$

where $\Sigma_1 = \mathrm{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_m\}, U_1 \in \mathbb{R}^{mr\times m}$, and $V_1 \in \mathbb{R}^{(n+p)\times m}$.

Step 3: Compute the estimates of the parameter matrices C, D and α as $[\widehat{C}\ \widehat{D}] = \Sigma_1 V_1^{\mathrm{T}}$, and $\widehat{\alpha} = U_1^{\dagger}$.

Remark 3.1. It is important to note that, as in the case of Algorithm 2.1, Algorithm 3.1 intrinsically delivers estimates that satisfy the uniqueness condition $\|\alpha\|_2 = 1$, since matrix U_1 in the SVD of $[\widehat{\widetilde{C}}\ \widehat{\widetilde{D}}]$ is a unitary matrix. Furthermore, and for the same reason, the pseudoinverse $U_1^{\dagger}$ always exists.

The results on consistency of the estimates presented in Theorem 2.2 for the Hammerstein model can be straightforwardly extended to the Wiener model, and therefore they are omitted here.

Remark 3.2. In comparison with the work in Ref. [32], the algorithm presented here is more general in the sense that any family of subspace algorithms (not limited to the MOESP class as in Ref. [32], but also including the N4SID and CVA classes) can be used as a first step of the algorithm, and that a more general representation of the nonlinearity using arbitrary basis functions (not limited to polynomial bases as in Ref. [32]) can be employed.

4. Simulation Examples

To illustrate the proposed identification schemes, two simulation examples are presented in this section. In Example 4.1, the (luxurious) assumption that the actual system belongs to the model class is considered. The *rationale* for presenting this example is to show that in this case the proposed algorithm is able to accurately identify the system. In addition, Example 4.1 shows that the proposed method works properly in a case where the method in Ref. [31] fails. Example 4.2, on the other hand, deals with the identification of a Wiener model for a real system corresponding to a pH neutralization process.

Example 4.1 (*Hammerstein Model*). The nonlinear *true* system consists of a third order linear discrete system with transfer function

$$G(z) = \frac{z^2 + 0.7z - 1.5}{z^3 + 0.9z^2 + 0.15z + 0.002}, \tag{24}$$

preceded by a static nonlinearity described by a third order polynomial containing only odd terms, of the form

$$\mathbf{N}(u_k) = 0.8593u_k - 0.5115u_k^3. \tag{25}$$

Note the reader that in this case the true system has a Hammerstein structure (i.e., the system belongs to the model class). The nonlinear characteristic is shown in solid line in the left plot of Fig. 3.

For the purposes of identification, the system was excited with zero-mean Gaussian white noise with variance $\sigma_u^2 = \frac{0.8953}{3\times 0.5115} = 0.56$. The corresponding output was corrupted with zero-mean colored noise with spectrum $\Phi_\nu(\omega) = \frac{0.64\times 10^{-8}}{1.2-0.4\cos(\omega)}$.

Remark 4.1. This example corresponds to Example 2. in Ref. [31], where the authors prove that their subspace identification method fails when the variance of the input is of the form $\sigma_u^2 = \frac{\alpha_1}{3\alpha_3}$, with α_1 and α_3 being the coefficients of the linear and cubic terms of the polynomial nonlinearity, respectively. As will be seen in the following, the method proposed in this paper will allow a successful identification of the Hammerstein model.

Algorithm 2.1 was employed to identify a Hammerstein model for the system, using the first ($N = 5000$) points of observed input–output measurements. The remaining 3000 data were used for validation purposes. *Step 1* in Algorithm 2.1 was performed using the CVA algorithm by Larimore [20,21] as implemented by Ljung in the n4sid routine in the *System Identification Toolbox* for use with Matlab (hereafter referred as SIT) [23]. The default option 'Prediction' for the 'focus' property of the n4sid routine was chosen, which means that the models are determined by minimizing the prediction errors, and corresponds to the optimal weighting from a statistical variance point of view [23]. A third-order LTI subsystem was identified. The estimated transfer function was (compare with the *true* transfer function (24))

$$\widehat{G}(z) = \frac{1.0000z^2 + 0.7000z - 1.5000}{1.0000z^3 + 0.9000z^2 + 0.1500z + 0.0002}.$$

On the other hand, a third order polynomial containing only odd terms, was used to represent the nonlinear part of the Hammerstein model. The estimated nonlinear model was (compare with the *true* nonlinearity (25))

$$\widehat{\mathbf{N}}(u_k) = 0.859300u_k - 0.511471u_k^3.$$

The estimated nonlinear characteristic is represented in dashed line in the left plot of Fig. 3. It can be observed that it is indistinguishable from the true nonlinear characteristic.

Finally, the measured (solid line) and estimated (dashed line) outputs are represented in the right plot of Fig. 3 (only the last 100 ms), where a good

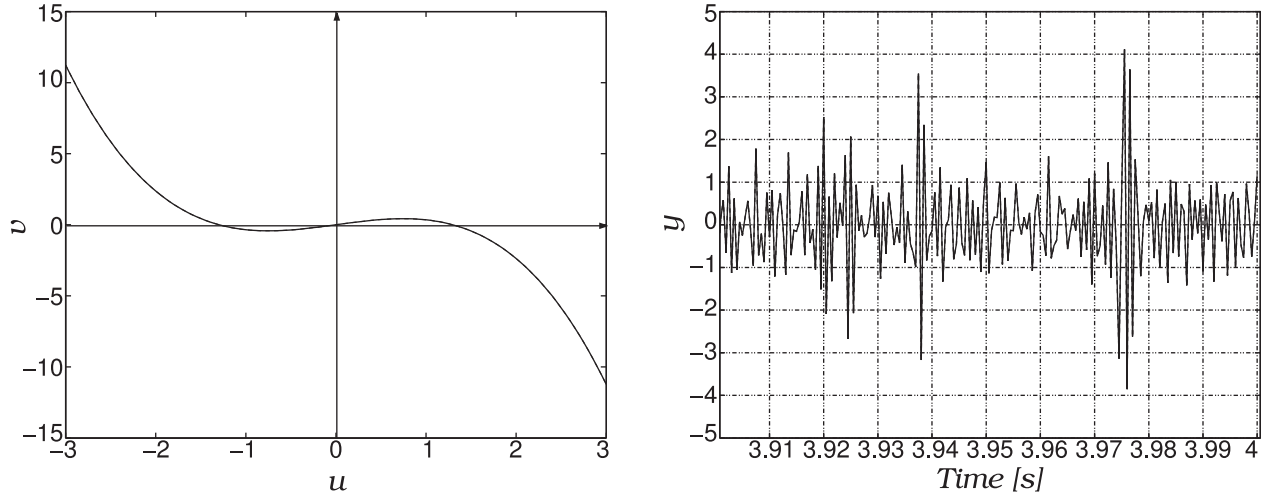

Fig. 3. Left Plot: True (solid line) and estimated (dashed line) nonlinear characteristic (indistinguishable one from the other). Right Plot: Measured (solid line) and estimated (dashed line) outputs (last 100 ms).

agreement between them can be observed. The mean square (Output) error[8] for the validation data was $\mathrm{MSE} = 6.6887 \times 10^{-9}$, while the Best Fit[9] was 99.9942%.

Example 4.2 (*Wiener Model*). In this example, Algorithm 3.1 is used to identify a Wiener model based on the simulation data of a pH neutralization process in a constant volume stirring tank considered in [16,17]), which corresponds to a bench-scale plant at the University of California, Santa Barbara.

The control of pH processes has been recognized as a challenging problem due to the time-varying and nonlinear characteristics of the pH processes. This is particularly true when control has to be achieved in the neutral range (pH between 6 and 8) when only strong acids and strong bases are present. From an identification point of view, pH processes have often been considered in the literature as having a Wiener structure (see for instance [18]). 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.

The pH neutralization process considered here 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.

The inputs to the system are the base (volumetric) 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, despite the variations of the unmeasured buffer flow rate, which can be considered as a disturbance.

A simulation model, based on first principles, was presented in Ref. [16] 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. (32)). The dynamic model for the reaction invariants of the effluent

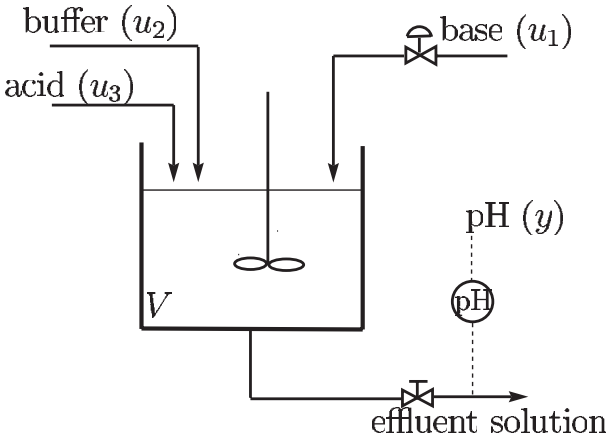

Fig. 4. Schematic representation of the pH neutralization process.

solution (W_a, W_b), in state-space form, is given by Refs [16,17].

$$\dot{x} = f(x) + g(x)u_1 + p(x)u_2, \tag{26}$$

$$h(x, y) = 0, \tag{27}$$

where

$$x \triangleq [x_1, x_2]^{\mathrm{T}} = [W_a, W_b]^{\mathrm{T}}, \tag{28}$$

$$f(x) = \left[\frac{u_3}{V}(W_{a3} - x_1), \frac{u_3}{V}(W_{b3} - x_2)\right]^{\mathrm{T}}, \tag{29}$$

$$g(x) = \left[\frac{1}{V}(W_{a1} - x_1), \frac{1}{V}(W_{b1} - x_2)\right]^{\mathrm{T}}, \tag{30}$$

$$p(x) = \left[\frac{1}{V}(W_{a2} - x_1), \frac{1}{V}(W_{b2} - x_2)\right]^{\mathrm{T}}, \tag{31}$$

$$h(x, y) = x_1 + 10^{y-14} - 10^{-y} + x_2 \frac{1 + 2 \times 10^{y-pK_2}}{1 + 10^{pK_1 - y} + 10^{y-pK_2}}. \tag{32}$$

Here, the parameters pK_1 and pK_2 are the first and second disassociation constants of the weak acid H_2CO_3. The nominal operating conditions of the system are given in Refs [16,17], and they are reproduced in Table 1 for the sake of completeness.

For the purposes of identification, the model (26) and (27) was excited with band limited white noise around the nominal value of the base flow rate (u_1) and the buffer flow rate (u_2), keeping the acid flow rate constant at its nominal value. The output of the

[8]See footnote 10.
[9]See footnote 11.

Table 1. Nominal operating conditions.

$u_3 = 16.60\,\mathrm{ml\,s^{-1}}$	$u_2 = 0.55\,\mathrm{ml\,s^{-1}}$
$u_1 = 15.55\,\mathrm{ml\,s^{-1}}$	$V = 2900\,\mathrm{ml}$
$W_{a1} = -3.05 \times 10^{-3}\,\mathrm{M}$	$W_{a2} = -3 \times 10^{-2}\,\mathrm{M}$
$W_{a3} = 3 \times 10^{-3}\,\mathrm{M}$	$W_a = -4.32 \times 10^{-4}\,\mathrm{M}$
$W_{b1} = 5 \times 10^{-5}\,\mathrm{M}$	$W_{b2} = 3 \times 10^{-2}\,\mathrm{M}$
$W_{b3} = 0\,\mathrm{M}$	$W_b = 5.28 \times 10^{-4}\,\mathrm{M}$
$pK_1 = 6.35$	$pK_2 = 10.25$
$y = 7.0$	

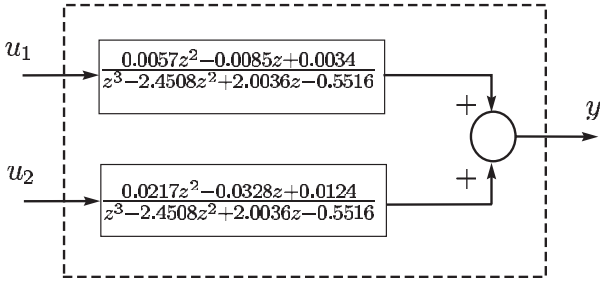

Fig. 5. Identified linear model.

system was corrupted with additive Gaussian white noise with zero mean and standard deviation $\sigma = 0.001$, in order to simulate the more realistic situation of having measurement noise. The first 1000 data were used for the estimation of the Wiener model, while the following 600 data were used for validation purposes.

For the purposes of comparison the same data were used to estimate a linear model of the process. The CVA algorithm option of the n4sid routine in SIT, with the default option 'Prediction' for the 'focus' property, was employed for the first step of the nonlinear identification algorithm, as well as for the identification of the linear model. As already mentioned, this option means that the models are determined by minimizing the prediction errors, and corresponds to the optimal weighting from a statistical variance point of view [23].

The order of the linear model was chosen as $n = 3$, using the default choice 'best' of the order argument in the n4sid routine. The algorithm computes the singular values of the Hankel matrices of the impulse responses for model orders in the range [1 : 10], and the best order is selected, in the sense that for higher orders the singular values are comparatively smaller. To make a fair comparison, the same order ($n = 3$) was chosen for the linear block in the Wiener model.

Polynomial basis functions were used for the representation of the nonlinear block in the Wiener model. In order to determine the number of terms in the polynomial representation, simulations were performed with polynomials with orders in the range [1 : 7], and an 'optimal' order $r = 3$ was selected based on the minimum Mean Square Error.

The estimated linear model is represented in Fig. 5, while the estimated Wiener model is represented in Fig. 6. The estimated nonlinear block in the Wiener model is given by

$$\widehat{\mathbf{N}}^{-1}(y_k) = 0.1595y_k^3 + 0.0225y_k^2 + 0.9869y_k, \tag{33}$$

and it is depicted in the left plot of Fig. 7. The true and estimated output (for both the estimation and the validation data) for the Wiener model are represented in the right plot of Fig. 7, where a good agreement between them can be observed.

For the purposes of comparison between the Wiener and the linear models, three different performance criteria were considered, viz., the mean square error[10] (MSE), the Best Fit[11] (FIT) and the Variance Accounted For[12] (VAF) criterion [22]. The results are summarized in Table 2. It can be seen that, for this example, the prediction capability of the Wiener model is better than that of the linear model, for the three different criteria considered. In particular, an improvement of 12% in the MSE is obtained by resorting to the Wiener model as an alternative to the linear model. It should also be noted that there is no significant increase in the model complexity by choosing a Wiener model instead of a linear model (in this case, only 3 extra parameters need to be estimated).

Remark 4.2. The identified Wiener model has been used in Ref. [14] in a Wiener model predictive control scheme. In that paper, it is shown that the improvement in the prediction capabilities of the Wiener model with respect to the linear model is translated

[10]The MSE is defined as

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

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

[11]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 y.

[12]The Variance Accounted For is defined as

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

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

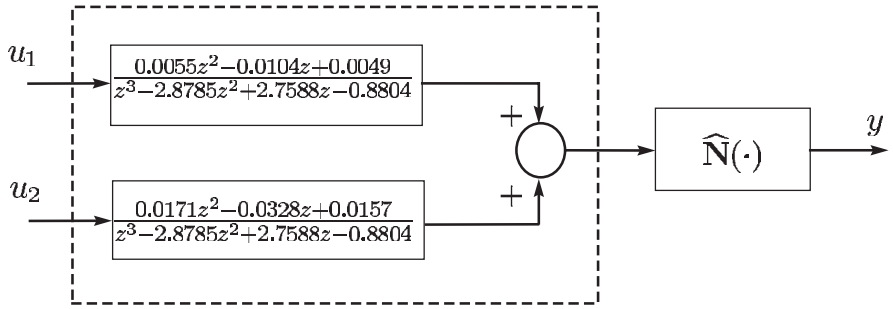

Fig. 6. Identified Wiener model.

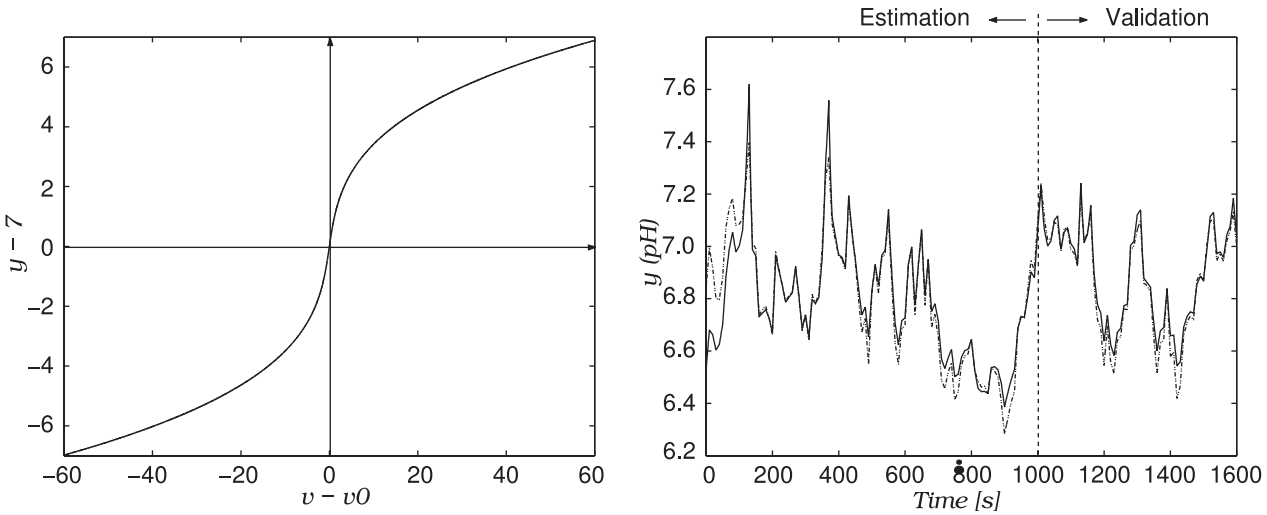

Fig. 7. Left Plot: Estimated nonlinear characteristic in the Wiener model. Right Plot: True (solid-line) and estimated (dotted-line) output of the Wiener model.

Table 2. Performance criteria for Wiener and linear models.

	MSE	FIT	VAF
Wiener	0.0040	69.9768	91.1417
Linear	0.0045	68.3404	90.1153

into a better performance of the model predictive controller.

5. Concluding Remarks

In this paper, new subspace algorithms for the simultaneous identification of the linear and nonlinear parts of Hammerstein and Wiener models have been presented. The algorithms consist of two basic steps. The first one is a standard (linear) subspace algorithm, while the second one is a 2-norm minimization problem solved via SVD. Under weak assumptions, consistency of the estimates is guaranteed. It has been shown that the proposed algorithms deliver nonlinear models which are suitable for their use in standard (i.e., based on linear model) MPC schemes. The performance of the proposed algorithms has been illustrated through simulation examples.

References

1. Bai E. An optimal two-stage identification algorithm for Hammerstein–Wiener nonlinear systems. Automatica 1998; 34(3): 333–338
2. Billings S. Identification of nonlinear systems – A survey. Proc IEE, Part D 1980; 127: 272–285
3. Billings S, Fakhouri S. Identification of a class of nonlinear systems using correlation analysis. Proc IEE, Corresp 1978; 125: 1307
4. Billings S, Fakhouri S. Identification of systems containing linear dynamic and static nonlinear elements. Automatica 1982; 18(1): 15–26
5. Boutayeb M, Darouach M. Recursive identification method for MISO Wiener–Hammerstein model. IEEE Trans Autom Control 1995; 40(2): 287–291
6. Boyd S, Chua L. Fading memory and the problem of approximating nonlinear operators with Volterra series. IEEE Trans Circuits Syst 1985; CAS-32(11): 1150–1161
7. Chen H, Maciejowski J. An improved subspace identification method for bilinear systems. In: Proceedings of the 39th IEEE conference on decision

and control CDC2000, Sydney, Australia, December 2000 pp 1573–1578
8. Chen H, Maciejowski J. Subspace identification method for combined deterministic-stochastic bilinear systems. In: Proceedings of the IFAC symposium on system identification SYSID 2000, Santa Barbara, CA, USA, June 2000
9. Eskinat E, Johnson S, Luyben W. Use of Hammerstein models in identification of nonlinear systems. AIChE J 1991; 37(2): 255–268
10. Favoreel W, de Moor B, Van Overschee P. Subspace identification of bilinear systems subject to white inputs. IEEE Trans Autom Control 1999; 44(6): 1157–1165
11. Fruzzetti K, Palazoglu A, McDonald K. Nonlinear model predictive control using Hammerstein models. J Process Control 1997; 7(1): 31–41
12. Garulli A, Giarrè L, Zappa G. Identification of approximated Hammerstein models in a worst-case setting. IEEE Trans Autom Control 2002; 47(12): 2046–2050
13. Golub G, Van Loan C. Matrix computations. 2nd edn. The Johns Hopkins University Press, Baltimore and London, 1989
14. Gómez JC, Jutan A, Baeyens E. Wiener model identification and predictive control of a pH neutralization process. IEE Proc Control Theory Appl 2004; 151(3): 329–338
15. Greblicki W. Nonparametric identification of Wiener systems by orthogonal series. IEEE Trans Autom Control 1994; 39(10): 2077–2086
16. Henson M, Seborg D. Adaptive nonlinear control of a pH neutralization process. IEEE Trans Control Syst Technol 1994; 2(3): 169–182
17. Henson M, Seborg D. (eds) 'Feedback linearizing control.' Nonlinear Process Control, Prentice Hall PTR, Upper Saddle River, NJ, 1997, chapter 4
18. Kalafatis A, Arifin N, Wang L, Cluett W. A new approach to the identification of pH processes based on the Wiener model. Chem Eng Sci 1995; 50(23): 3693–3701
19. Korenberg M. 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
20. Larimore W. Canonical variate analysis in identification, filtering, and adaptive control. In: Proceedings of the 29th IEEE Conference on Decision and Control, Honolulu, Hawaii, December 1990. pp. 596–604
21. Larimore W. The optimality of canonical variate identification by example. In: Proceedings of 10th IFAC symposium on system identification SYSID'94, vol 2, Copenhagen, Denmark, July 1994, pp. 151–156
22. Ljung L. System Identification: Theory for the User. 2nd edn. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1999
23. Ljung L. System Identification Toolbox, User's Guide, ver. 5. The MathWorks, Inc., Natick, MA, 2000
24. Narendra K, Gallman P. An iterative method for the identification of nonlinear systems using a Hammerstein model. IEEE Trans Autom Control 1966; AC-11: 546–550
25. Pearson R, Pottmann M. Gray-box identification of block-oriented nonlinear models. J Process Control 2000; 10: 301–315
26. Rangan S, Wolodkin G, Poolla K. 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
27. Stapleton J, Bass S. Adaptive noise cancellation for a class of nonlinear dynamic reference channels. IEEE Trans Circuits Syst 1985; CAS-32: 143–150
28. Stoica P. On the convergence of an iterative algorithm used for Hammerstein system identification. IEEE Trans Autom Control 1981; AC-26(4): 967–969
29. Van Overschee P, de Moor B. N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. Automatica 1994; 30(1): 75–93
30. Verhaegen M. Identification of the deterministic part of MIMO state space models given in innovations form from input–ouput data. Automatica 1994; 30(1): 61–74
31. Verhaegen M, Westwick D. Identifying MIMO Hammerstein systems in the context of subspace model identification methods. Int J Control 1996; 63(2): 331–349
32. Westwick D, Verhaegen M. Identifying MIMO Wiener systems using subspace model identification methods. Signal Process, 1996; 52: 235–258
33. Wigren T. Recursive prediction error identification using the nonlinear Wiener model. Automatica 1993; 29(4): 1011–1025
34. Wigren T. Convergence analysis of recursive identification algorithms based on the nonlinear Wiener model. IEEE Trans Autom Control 1994; 39(11): 2191–2206

Appendix

Proof of Theorem 2.2. Let $\widehat{\Theta}_{BD}$ be a consistent estimate of Θ_{BD} defined in (8). Noting now that

$$\begin{aligned}
&\left\| \begin{bmatrix} \widehat{B} \\ \widehat{D} \end{bmatrix} \widehat{\alpha}^{\mathrm{T}} - \begin{bmatrix} B \\ D \end{bmatrix} \alpha^{\mathrm{T}} \right\|_2^2 \\
&\quad = \left\| \begin{bmatrix} \widehat{B} \\ \widehat{D} \end{bmatrix} \widehat{\alpha}^{\mathrm{T}} - \widehat{\Theta}_{BD} + \widehat{\Theta}_{BD} - \begin{bmatrix} B \\ D \end{bmatrix} \alpha^{\mathrm{T}} \right\|_2^2 \\
&\quad \leq \left\| \begin{bmatrix} \widehat{B} \\ \widehat{D} \end{bmatrix} \widehat{\alpha}^{\mathrm{T}} - \widehat{\Theta}_{BD} \right\|_2^2 + \left\| \widehat{\Theta}_{BD} - \begin{bmatrix} B \\ D \end{bmatrix} \alpha^{\mathrm{T}} \right\|_2^2 . \\
&\quad = \sigma_{p+1}^2 + \left\| \widehat{\Theta}_{BD} - \begin{bmatrix} B \\ D \end{bmatrix} \alpha^{\mathrm{T}} \right\|_2^2 , \qquad (34)
\end{aligned}$$

and considering that Θ_{BD} is a rank-p matrix (so that $\sigma_{p+1}^2 = 0$), and that $\widehat{\Theta}_{BD}$ is a consistent estimate of Θ_{BD}, then

$$\left\| \begin{bmatrix} \widehat{B} \\ \widehat{D} \end{bmatrix} \widehat{\alpha}^{\mathrm{T}} - \begin{bmatrix} B \\ D \end{bmatrix} \alpha^{\mathrm{T}} \right\|_2^2 \xrightarrow{\text{a.s.}} 0,$$

as $N \to \infty$ Now, from the uniqueness of the decomposition $\begin{bmatrix} B \\ D \end{bmatrix} \alpha^{\mathrm{T}}$, it can be concluded that $\widehat{B} \xrightarrow{\text{a.s.}} B, \widehat{D} \xrightarrow{\text{a.s.}} D$, and $\widehat{\alpha} \xrightarrow{\text{a.s.}} \alpha$, what ends the proof. □