Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Identification of Bilinear Systems Using an Iterative Deterministic-Stochastic Subspace Approach

P. Lopes dos Santos[1], J. A. Ramos[2], J. L. Martins de Carvalho[1]
[1] Faculdade de Engenharia da Universidade do Porto
Departamento de Engenharia Electrotécnica e de Computadores
Rua Dr. Roberto Frias
4200 465 Porto, Portugal

[2] Department of Electrical & Computer Engineering
Purdue University at Indianapolis
723 West Michigan Street, SL 164-E
Indianapolis, IN 46202-5132

Abstract— **In this paper we introduce a new identification algorithm for MIMO bilinear systems driven by white noise inputs. The new algorithm is based on a convergent sequence of linear deterministic-stochastic state space approximations, thus considered a Picard based method. The key to the algorithm is the fact that the bilinear terms behave like white noise processes. Using a linear Kalman filter, the bilinear terms can be estimated and combined with the system inputs at each iteration, leading to a linear system which can be identified with a linear-deterministic subspace algorithm such as MOESP, N4SID, or CVA. Furthermore, the model parameters obtained with the new algorithm converge to those of a bilinear model. Finally, the dimensions of the data matrices are comparable to those of a linear subspace algorithm, thus avoiding the curse of dimensionality.**

I. Introduction

Subspace identificatiom theory has opened new avenues of research for a large class of systems in the state space domain. Among the linear algorithms, the purely deterministic, purely stochastic, and combined deterministic-stochastic approaches have already reached a considerable level of maturity [9], [17], [20], [21]. These algorithms have already been applied in many areas of applications (see [20] and references therein). Despite the power of the linear subspace algorithms, there are instances where the input-output data is more likely to follow a nonlinear model. Such is the case in rainfall-runoff models [14], biological models [16], and distillation column models [2], among many others. It is well known in system theory that a bilinear system can approximate a nonlinear system fairly well, via a finite sum of the Volterra series expansion between the inputs and outputs of the system [15], [5], [6], [7]. However, for more general nonlinear systems, a finite sum of Volterra kernels may no longer hold. Thus, bilinear subspace algorithms can only solve a limited class of nonlinear system identification problems. Despite this limitation, they still provide a higher degree of approximation to nonlinear models than traditional linear models. Furthermore, from a system theory point of view, bilinear models behave similarly to linear models, which is not the case with other nonlinear models [7]. Finally, a full understanding of the bilinear model is important before attempting to model more general nonlinear systems.

In recent years, there has been a considerable effort in developing algorithms for bilinear system identification [10], [11], [12], [13], [1], [2], [3], [4], [6], [14]. The bilinear subspace algorithm in [2] has the tendency to require input-output data matrices whose row dimensions grow exponentially. Since the column dimension must be much larger than the row dimension, the curse of dimensionality will limit the validity of some approaches due to high memory demands. For such reason, most algorithms are tested in the literature with small order models ($n \leq 2$). Nevertheless, the algorithm of [3], [4], [13] introduces a way of reducing the data matrices to a certain extent. Furthermore, only the algorithm of [3] claims to handle non white noise inputs, however, it produces biased system parameter estimates.

In this paper we introduce a new bilinear system identification algorithm based on a convergent sequence of linear deterministic-stochastic state space approximations. The algorithm is based on a Picard method introduced in [7]. The proposed algorithm works with white noise inputs, which is key to proving that the bilinear term behaves like a white noise process. Using a linear Kalman filter, the bilinear term can be estimated and used as inputs in each iteration. Furthermore, the algorithm converges to a bilinear model and the dimensions of the data matrices are comparable to those of a linear subspace algorithms [17], [20]. In section 2 we show that the bilinear system driven by white noise inputs can be modeled as a linear deterministic-stochastic system, where the bilinear terms are white noise sources. In section 3 we propose an iterative method for estimating the states of a bilinear system. A convergence analysis of the proposed method is presented in section 4. In section 5 we propose

0-7803-9568-9/05/$20.00 ©2005 IEEE

an iterative bilinear subspace identification algorithm. The performance of the algorithm is tested with simulated data in section 6. Finally, conclusions are drawn in section 7.

II. Modeling the Bilinear System as a Linear System

Consider a bilinear state space model described by

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + Eu(t) \otimes x(t) + q(t) \quad (1) \\ y(t) &= Cx(t) + Du(t) + Fu(t) \otimes x(t) + r(t), \quad (2) \end{aligned}$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, and $y(t) \in \mathbb{R}^\ell$ are, respectively, the state, input, and output vectors at time t. The unknown parameter matrices have dimensions $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{\ell\times n}$, $D \in \mathbb{R}^{\ell\times m}$, $E = \left[\begin{array}{ccc} E_1 & \cdots & E_m \end{array} \right]$, and $F = \left[\begin{array}{ccc} F_1 & \cdots & F_m \end{array} \right]$, where $E_k \in \mathbb{R}^{n\times n}$ and $F_k \in \mathbb{R}^{\ell\times n}$, for $k = 1, \ldots, m$. The input, $u(t)$, is a zero mean white noise vector with covariance matrix $R_u(0)$, and the process and measurement noise vectors, $q(t) \in \mathbb{R}^n$ and $r(t) \in \mathbb{R}^\ell$, respectively, are white noise processes, uncorrelated with $u(t)$, and having a joint covariance matrix given by

$$\mathbb{E}\left\{ \left[\begin{array}{c} q(t) \\ r(t) \end{array} \right] \left[\begin{array}{cc} q^T(s) & r^T(s) \end{array} \right] \right\} = \left[\begin{array}{cc} Q & S \\ S^T & R \end{array} \right] \delta_{ts},$$

where $Q \in \mathbb{R}^{n\times n}$, $S \in \mathbb{R}^{n\times \ell}$, and $R \in \mathbb{R}^{\ell\times\ell}$. Other properties of bilinear models such as $(1)-(2)$ can be found in [15], [5], [6], [2], [12].

In order to develop our bilinear system identification approach, we need to characterize the bilinear portion of the state equation, $z(t) = u(t) \otimes x(t)$, as a random process. Let us calculate its mean and autocorrelation function. Taking the expected value of $z(t)$ we obtain

$$\begin{aligned} \mathbb{E}\{z(t)\} &= \mathbb{E}\{u(t) \otimes x(t)\} \\ &= \mathbb{E}\{u(t)\} \otimes \mathbb{E}\{x(t)\} = 0_{nm\times 1}, \end{aligned}$$

since $u(t)$ is white noise and $x(t)$ depends on $u(t-\tau)$, for $\tau = 1, 2, \ldots, \infty$. The autocovariance function $R_z(\tau) = \mathbb{E}\{z(t)z^T(t-\tau)\}$ can be shown to be

$$R_z(\tau) = \left\{ \begin{array}{ll} \mathbb{E}\{u(t)u^T(t)\} \otimes \mathbb{E}\{x(t)x^T(t)\}, & \tau = 0 \\ 0_{nm\times nm} & \tau \neq 0 \end{array} \right. .$$

To prove this we need to use properties of Kronecker products [22]. For $\tau > 0$, we obtain

$$\begin{aligned} R_z(\tau) &= \mathbb{E}\{u(t)u^T(t-\tau) \otimes x(t)x^T(t-\tau)\} \\ &= \mathbb{E}\{u(t)\} \otimes \mathbb{E}\{u^T(t-\tau) \otimes x(t)x^T(t-\tau)\} \\ &= 0_{nm\times nm}, \end{aligned}$$

since $x(t)$ and $x(t-\tau)$ do not depend on $u(t)$. Likewise, by symmetry of the autocovariance function, we obtain $R_z(\tau) = 0_{nm\times nm}$ when $\tau < 0$. Furthermore, if $P(t) = \mathbb{E}\{x(t)x^T(t)\}$ is a time invariant $n \times n$ covariance matrix, then the bilinear term, $z(t)$, is also a white noise process.

Let us now study under what conditions this can occur. Multiplying $x(t+1)$ by $x^T(t+1)$ and taking expectations on both sides, we obtain

$$\begin{aligned} & \mathbb{E}\{x(t+1)x^T(t+1)\} \\ = \; & A\mathbb{E}\{x(t)x^T(t)\}A^T + B\mathbb{E}\{u(t)u^T(t)\}B^T \\ & + E\mathbb{E}\{x(t)\} \otimes \mathbb{E}\{u(t)u^T(t)\}B^T \\ & + B\mathbb{E}\{u(t)u^T(t)\} \otimes \mathbb{E}\{x^T(t)\}E^T \\ & + E\mathbb{E}\{u(t)u^T(t)\} \otimes \mathbb{E}\{x(t)x^T(t)\}E^T \\ & + \mathbb{E}\{q(t)q^T(t)\}. \end{aligned}$$

Now taking the expected value of $\mathbb{E}\{x(t)\}$, we obtain

$$\begin{aligned} \mathbb{E}\{x(t+1)\} &= A\mathbb{E}\{x(t)\} + E\mathbb{E}\{z(t)\} + \mathbb{E}\{q(t)\} \\ &= A\mathbb{E}\{x(t)\}, \end{aligned}$$

from which we can conclude that $\mathbb{E}\{x(t)\}$ can be stabilized if all the eigenvalues of A are inside the unit circle. Under these conditions, $\mathbb{E}\{x(t)\} = 0_{n\times 1}$. On the other hand, if there exists an eigenvalue of A outside the unit circle, then $\mathbb{E}\{x(t)x^T(t)\}$ is no longer stable and, consequently, $z(t)$ violates the stationarity conditions that are necessary to identify the system. Henceforth, in the remainder of this paper we will consider systems whose A matrix has all its eigenvalues inside the unit circle. Thus, if $\mathbb{E}\{x(t)\} = 0_{n\times 1}$, then

$$\begin{aligned} & \mathbb{E}\{x(t+1)x^T(t+1)\} \\ = \; & A\mathbb{E}\{x(t)x^T(t)\}A^T + B\mathbb{E}\{u(t)u^T(t)\}B^T \\ & + E\mathbb{E}\{u(t)u^T(t) \otimes x(t)x^T(t)\}E^T + \mathbb{E}\{q(t)q^T(t)\}. \end{aligned}$$

This equation is stabilized if $|\lambda_{max}(M)| < 1$ holds true [22], where $\lambda_{max}(M)$ denotes the maximum eigenvalue of the matrix $M \in \mathbb{R}^{n^2\times n^2}$, given by

$$M = A \otimes A + \sum_{k=1}^{m}\sum_{w=1}^{m}(E_k \otimes E_w)\{R_u(0)\}_{kw},$$

with $\{R_u(0)\}_{kw}$ being the $(kw)^{th}$ entry of $R_u(0)$. If this condition is satisfied, then $P = \mathbb{E}\{x(t)x^T(t)\}$ will converge to a solution of

$$\begin{aligned} P &= APA^T + E(R_u(0) \otimes P)E^T \\ &\quad + BR_u(0)B^T + Q. \end{aligned} \quad (3)$$

Therefore, we conclude that $z(t)$ is a white noise process with zero mean and autocovariance function equal to

$$R_z(\tau) = \left\{ \begin{array}{ll} R_u(0) \otimes P, & \tau = 0 \\ 0_{nm\times nm} & \tau \neq 0 \end{array} \right. . \quad (4)$$

In order to carry the analysis one step further, we need to determine whether $z(t)$ and $u(t)$ are correlated or not. Toward this end we calculate the cross covariance function $R_{zu}(\tau)$

$$\begin{aligned} R_{zu}(\tau) &= \mathbb{E}\{u(t) \otimes x(t)u^T(t-\tau)\} \\ &= \mathbb{E}\{u(t)u^T(t-\tau)\} \otimes \mathbb{E}\{x(t)\} \\ &= 0_{mn\times m}. \end{aligned}$$

Thus, implying that $z(t)$ and $u(t)$ are uncorrelated. This will allow us to model the nonlinear term, $z(t)$, as white process noise and describe the bilinear system by a linear state space model such as

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + \bar{q}(t) \quad (5) \\ y(t) &= Cx(t) + Du(t) + \bar{r}(t), \quad (6) \end{aligned}$$

where the noise terms $\bar{q}(t) = Ez(t) + q(t)$ and $\bar{r}(t) = Fz(t) + r(t)$ are white noise processes uncorrelated with $u(t)$, with

$$\mathbb{E}\left\{ \begin{bmatrix} \bar{q}(t) \\ \bar{r}(t) \end{bmatrix} \begin{bmatrix} \bar{q}^T(s) & \bar{r}^T(s) \end{bmatrix} \right\} = \begin{bmatrix} \bar{Q} & \bar{S} \\ \bar{S}^T & \bar{R} \end{bmatrix} \delta_{ts},$$

where $\bar{Q} = ER_z(0)E^T + Q$, $\bar{R} = FR_z(0)F^T + R$, and $\bar{S} = ER_z(0)F^T + S$. The new linear system $(5) - (6)$ can be identified by any of the existing linear deterministic-stochastic subspace algorithms from the literature [17], [20], [21].

III. An Iterative State Estimation Approach

Suppose the identified linear model $(5) - (6)$ yield state sequence matrices $\{\hat{X}_i, \hat{X}_{i+1}\}$ and parameter matrices $\{A, B, C, D\}$. Then we can calculate the estimated noise matrices $\{\hat{\bar{q}}_{i|i}, \hat{\bar{r}}_{i|i}\}$ from

$$\begin{bmatrix} \hat{\bar{q}}_{i|i} \\ \hat{\bar{r}}_{i|i} \end{bmatrix} = \begin{bmatrix} \hat{X}_{i+1} \\ Y_{i|i} \end{bmatrix} - \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \hat{X}_i \\ U_{i|i} \end{bmatrix},$$

where

$$\begin{bmatrix} \hat{\bar{q}}_{i|i} \\ \hat{\bar{r}}_{i|i} \end{bmatrix} = \begin{bmatrix} E \\ F \end{bmatrix} U_{i|i} \odot \hat{X}_i + \begin{bmatrix} \hat{q}_{i|i} \\ \hat{r}_{i|i} \end{bmatrix}.$$

In the above, $\odot$ is the Khatri-Rao product of matrices [8], i and j are, respectively, the number of block rows and columns of the Hankel data matrices in most standard subspace algorithms [17], [20], [21], and

$$\begin{aligned} \hat{q}_{i|i} &= \begin{bmatrix} \hat{q}(i) & \hat{q}(i+1) & \dots & \hat{q}(i+j-1) \end{bmatrix} \\ \hat{r}_{i|i} &= \begin{bmatrix} \hat{r}(i) & \hat{r}(i+1) & \dots & \hat{r}(i+j-1) \end{bmatrix} \\ \hat{X}_i &= \begin{bmatrix} \hat{x}(i) & \hat{x}(i+1) & \cdots & \hat{x}(i+j-1) \end{bmatrix} \\ U_{i|i} &= \begin{bmatrix} u(i) & u(i+1) & \dots & u(i+j-1) \end{bmatrix} \\ Y_{i|i} &= \begin{bmatrix} y(i) & y(i+1) & \dots & y(i+j-1) \end{bmatrix}. \end{aligned}$$

At first sight one would think that the estimated noise matrices can be used to estimate E and F from

$$\begin{bmatrix} E \\ F \end{bmatrix} = \begin{bmatrix} \hat{\bar{q}}_{i|i} \\ \hat{\bar{r}}_{i|i} \end{bmatrix} \left(U_{i|i} \odot \hat{X}_i \right)^{\dagger}.$$

However, E cannot be estimated this way due to the following argument. Although the errors from $\hat{\bar{q}}_{i|i}$ are orthogonal to $\hat{X}_i$, they are contained in W_p^+, defined as

$$W_p^+ = \begin{bmatrix} U_p^+ \\ Y_p^+ \end{bmatrix} = \begin{bmatrix} U_{0|0} \\ \vdots \\ U_{i|i} \\ \hline Y_{0|0} \\ \vdots \\ Y_{i|i} \end{bmatrix},$$

which in turn is orthogonal to $EU_{i|i} \odot \hat{X}_i$. However, this bilinear term is not contained in the errors $\hat{\bar{q}}_{i|i}$. Thus, it would be impossible to compute E from knowledge of $\hat{\bar{q}}_{i|i}$.

Let us now take another approach. Knowing that the bilinear term $z(t)$ is a white noise process, uncorrelated with $u(t)$, we can re-write the bilinear model as

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + \bar{q}_0(t) \quad (7) \\ y(t) &= Cx(t) + Du(t) + \bar{r}_0(t), \quad (8) \end{aligned}$$

where the process and measurement noises are given by

$$\begin{aligned} \bar{q}_0(t) &= Eu(t) \otimes x(t) + q(t) \\ \bar{r}_0(t) &= Fu(t) \otimes x(t) + r(t), \end{aligned}$$

with joint covariance matrix

$$\mathbb{E}\left\{ \begin{bmatrix} \bar{q}(t) \\ \bar{r}(t) \end{bmatrix} \begin{bmatrix} \bar{q}^T(s) & \bar{r}^T(s) \end{bmatrix} \right\} = \begin{bmatrix} \bar{Q} & \bar{S} \\ \bar{S}^T & \bar{R} \end{bmatrix} \delta_{ts}.$$

As previously shown, the linear model $(7) - (8)$ can be identified by any existing linear deterministic-stochastic subspace algorithms [17], [20], [21]. Let us now consider the stationary Kalman filter of $(7) - (8)$, i.e.,

$$\begin{aligned} \hat{x}_0(t+1) &= A\hat{x}_0(t) + Bu(t) + K_0(y(t) \\ &\quad - C\hat{x}_0(t) - Du(t)) \quad (9) \end{aligned}$$

where

$$\begin{aligned} K_0 &= S_0 R_0^{-1} \\ S_0 &= G - AP_0C^T - BR_u(0)D^T \\ R_0 &= \Lambda(0) - CP_0C^T - DR_u(0)D^T \\ R_u(0) &= \mathbb{E}\{u(t)u^T(t)\} \\ \Lambda(0) &= \mathbb{E}\{y(t)y^T(t)\} \\ G &= \mathbb{E}\{x(t+1)y^T(t)\} \\ P_0 &= AP_0A^T + BR_u(0)B^T + S_0R_0^{-1}S_0^T. \end{aligned}$$

We know from Kalman filter theory that the innovation term, $(y(t) - C\hat{x}_0(t) - Du(t))$, is uncorrelated with $\hat{x}_0$ and $u(t)$ [20]. That is,

$$\begin{aligned} \mathbb{E}\{\hat{x}_0(t)(y(t) - C\hat{x}_0(t) - Du(t))^T\} &= 0_{n\times\ell} \quad (10) \\ \mathbb{E}\{u(t)(y(t) - C\hat{x}_0(t) - Du(t))^T\} &= 0_{m\times\ell}. \quad (11) \end{aligned}$$

If we now write the right hand term in the expectation as

$$\begin{aligned} y(t) - C\hat{x}_0(t) - Du(t) &= Cx(t) + Du(t) \\ &\quad + r(t) - C\hat{x}_0(t) - Du(t) \\ &= C(x(t) - \hat{x}_0(t)) + r(t) \\ &= C\tilde{x}_0(t) + r(t). \end{aligned}$$

Then the first orthogonality condition can be re-written as

$$\begin{aligned} &\mathbb{E}\{\hat{x}_0(t)(y(t) - C\hat{x}_0(t) - Du(t))^T\} \\ &= \mathbb{E}\{\hat{x}_0(t)\tilde{x}_0^T(t)C^T + \hat{x}_0(t)r^T(t)\} \\ &= \mathbb{E}\{\hat{x}_0(t)\tilde{x}_0^T(t)\}C^T + \mathbb{E}\{\hat{x}_0(t)r^T(t)\}. \end{aligned}$$

Since $r(t)$ is uncorrelated with $\hat{x}_0(t)$, we have $\mathbb{E}\{\hat{x}_0(t)r^T(t)\} = 0_{n\times\ell}$. Consequently, $\hat{x}_0(t)$ is also

uncorrelated with $\tilde{x}_0(t)$.

Given that $u(t)$ is a white noise process and that the system is stable, the signal $\hat{z}_0(t) = u(t) \otimes \hat{x}_0(t) \in \mathbb{R}^{nm}$ has a mean value equal to

$$\begin{aligned} \mathbb{E}\{\hat{z}_0(t)\} &= \mathbb{E}\{u(t)\} \otimes \mathbb{E}\{\hat{x}_0(t)\} \\ &= 0_{mn\times 1}. \end{aligned}$$

Furthermore, the autocovariance function, $R_{\hat{z}}(\tau) = \mathbb{E}\{\hat{z}_0(t)\hat{z}_0^T(t-\tau)\}$ can be written as

$$\begin{aligned} R_{\hat{z}}(\tau) &= \left\{ \begin{array}{l} \mathbb{E}\{u(t)u^T(t)\} \otimes \mathbb{E}\{\hat{x}_0(t)\hat{x}_0^T(t)\},\ \tau = 0 \\ 0_{nm\times nm},\ \tau \neq 0 \end{array} \right. \\ &= \left\{ \begin{array}{ll} R_u(0) \otimes P_0, & \tau = 0 \\ 0_{nm\times nm}, & \tau \neq 0 \end{array} \right. , \end{aligned}$$

where, for $\tau > 0$,

$$\begin{aligned} R_{\hat{z}}(\tau) &= \mathbb{E}\{u(t)u^T(t-\tau) \otimes \hat{x}_0(t)\hat{x}_0^T(t-\tau)\} \\ &= \mathbb{E}\{u(t)\} \otimes \mathbb{E}\{u^T(t-\tau) \otimes \hat{x}_0(t)\hat{x}_0^T(t-\tau)\} \\ &= 0_{nm\times nm}, \end{aligned}$$

since $\hat{x}_0(t)$ and $\hat{x}_0(t-\tau)$ do not depend on $u(t)$. By symmetry, when $\tau < 0$, the autocovariance function also vanishes. We then conclude that $\hat{z}_0(t)$ is a white noise process.

We can now use $\hat{z}_0(t)$ as an estimate of $z(t)$. Let us now write the bilinear system as follows

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + Ez(t) + q(t) + E\hat{z}_0(t) \\ &\quad - E\hat{z}_0(t) \\ &= Ax(t) + Bu(t) + E\hat{z}_0(t) + \bar{q}_1(t), \end{aligned} \tag{12}$$

where $\bar{q}_1(t) = E\tilde{z}_0(t) + q(t)$ and $\tilde{z}_0(t) = z(t) - \hat{z}_0(t) = u(t) \otimes (x(t) - \hat{x}_0(t))$. If we now operate on the output equation in a similar fashion, we obtain

$$\begin{aligned} y(t) &= Cx(t) + Du(t) + Fz(t) + r(t) + F\hat{z}_0(t) \\ &\quad - F\hat{z}_0(t) \\ &= Cx(t) + Du(t) + F\hat{z}_0(t) + \bar{r}_1(t), \end{aligned} \tag{13}$$

where $\bar{r}_1(t) = F\tilde{z}_0(t) + r(t)$. It is now easy to show that $\tilde{z}_0(t)$ is also a white noise process, uncorrelated with $u(t)$. On the other hand, since $\tilde{x}_0(t)$ is uncorrelated with $\hat{x}_0(t)$, then $\tilde{z}_0(t)$ is also uncorrelated with $\hat{z}_0(t)$. We can now describe the bilinear model as

$$x(t+1) = Ax(t) + \bar{B}\bar{u}_1(t) + \bar{q}_1(t) \tag{14}$$

$$y(t) = Cx(t) + \bar{D}\bar{u}_1(t) + \bar{r}_1(t), \tag{15}$$

where $\bar{B} = \left[\begin{array}{cc} B & E \end{array} \right]$, $\bar{D} = \left[\begin{array}{cc} D & F \end{array} \right]$, $\bar{q}_1(t)$ and $\bar{r}_1(t)$ are white noise processes, uncorrelated with $\bar{u}_1(t)$, defined as

$$\bar{u}_1(t) = \left[\begin{array}{c} u(t) \\ \hat{z}_0(t) \end{array} \right].$$

If we knew the state estimates, $\hat{x}_0$, we could then identify the bilinear system using any of the existing deterministic-stochastic subspace algorithms. Since we do not know $\{A, B, C, D, K_0\}$, we cannot find $\hat{x}_0(t)$. However, through the identification of $(7) - (8)$, we could consistently identify $\{A, B, C, D, K_0\}$, and from these we can obtain estimates close to those of $\hat{x}_0(t)$. Then we can use these state estimates to generate $\hat{z}_0(t)$, which can be used to identify $(15) - (16)$, thus obtaining all the parameters of the bilinear model.

A standard result in system identification theory is that the larger the noise in the data, the smaller the precision of the estimates of the model parameters. This implies that if we merge the bilinear terms, $\{Ez(t), Fz(t)\}$ of $(7) - (8)$ and $\{E\tilde{z}_0(t), F\tilde{z}_0(t)\}$ of $(15) - (16)$, with the process and measurement noises, respectively, we cannot expect the parameter estimates to be the most precise. Since the covariance of $\tilde{x}_0$ is likely to be smaller than that of $x(t)$, the noise term $\bar{q}_1(t)$ is smaller than $\bar{q}_0(t)$ and, consequently, the estimates obtained from model $(15) - (16)$ are better than those from model $(7) - (8)$. Furthermore, we could obtain better estimates if, in a subsequent stage, we describe the system as

$$x(t+1) = Ax(t) + \bar{B}\bar{u}_2(t) + \bar{q}_2(t) \tag{16}$$

$$y(t) = Cx(t) + \bar{D}\bar{u}_2(t) + \bar{r}_2(t), \tag{17}$$

where $\bar{q}_2(t) = E\tilde{z}_1 + q(t)$, $\bar{r}_2(t) = F\tilde{z}_1 + r(t)$, $\bar{q}_2(t)$ and $\bar{r}_2(t)$ are white noise processes, uncorrelated with $\bar{u}_2(t)$, and

$$\begin{aligned} \bar{u}_2(t) &= \left[\begin{array}{c} u(t) \\ \hat{z}_1(t) \end{array} \right] \\ \tilde{x}_1(t) &= x(t) - \hat{x}_1(t) \\ \tilde{z}_1(t) &= u(t) \otimes \tilde{x}_1(t). \end{aligned}$$

Here, $\hat{x}_1(t)$ is the state estimate obtained from a stationary Kalman filter of $(15) - (16)$. That is,

$$\begin{aligned} \hat{x}_1(t+1) &= A\hat{x}_1(t) + \bar{B}\bar{u}_1(t) \\ &\quad + K_1(y(t) - C\hat{x}_1(t) - \bar{D}\bar{u}_1(t)), \end{aligned} \tag{18}$$

where

$$\begin{aligned} K_1 &= S_1R_1^{-1} \\ S_1 &= G - AP_1C^T - \bar{B}R_{\bar{u}_1}(0)\bar{D}^T \\ R_1 &= \Lambda(0) - CP_1C^T - \bar{D}R_{\bar{u}_1}(0)\bar{D}^T \\ R_{\bar{u}_1}(0) &= \mathbb{E}\{\bar{u}_1(t)\bar{u}_1^T(t)\}, \end{aligned}$$

and $P_1 = \mathbb{E}\{\hat{x}_1(t)\hat{x}_1^T(t)\}$ is the solution of the Ricatti equation

$$P_1 = AP_1A^T + \bar{B}R_{\bar{u}_1}(0)\bar{B}^T + S_1R_1^{-1}S_1^T.$$

The above estimates will lead to a model with a smaller noise variance since $\bar{q}_2(t)$ is smaller than $\bar{q}_1(t)$. The identification of $(17) - (18)$ will then allow us to obtain better estimates of the model parameters. If we continue iterating this way, we obtain an iterative method of the Picard type, which is known to converge under mild conditions [7].

IV. Convergence Analysis

During the p^{th} iteration, the system to be identified is given by

$$\begin{aligned} x(t+1) &= Ax(t) + \bar{B}\bar{u}_p(t) + \bar{q}_p(t) \quad (19)\\ y(t) &= Cx(t) + \bar{D}\bar{u}_p(t) + \bar{r}_p(t), \quad (20)\end{aligned}$$

where $\bar{q}_p(t) = E\tilde{z}_p + q(t)$ and $\bar{r}_p(t) = F\tilde{z}_p + r(t)$ are white noise processes, uncorrelated with $\bar{u}_p(t)$, defined as

$$\bar{u}_p(t) = \begin{bmatrix} u(t) \\ \hat{z}_{p-1}(t) \end{bmatrix},$$

$\tilde{x}_{p-1}(t) = x(t) - \hat{x}_{p-1}(t)$, and $\tilde{z}_{p-1}(t) = u(t) \otimes \tilde{x}_{p-1}(t)$. Here, $\hat{x}_{p-1}(t)$ is the state estimate obtained from the stationary Kalman filter at the $(p-1)^{th}$ iteration. That is,

$$\begin{aligned}\hat{x}_{p-1}(t+1) &= A\hat{x}_{p-1}(t) + \bar{B}\bar{u}_{p-1}(t)\\ &\quad + K_{p-1}(y(t) - C\hat{x}_{p-1}(t) - \bar{D}\bar{u}_{p-1}(t)). \quad (21)\end{aligned}$$

In order for the process to converge, the term $\hat{\tilde{x}}_p(t) = \hat{x}_p(t) - \hat{x}_{p-1}(t)$ must converge to zero as $p \to \infty$. From (23) we can write

$$\begin{aligned}\hat{\tilde{x}}_p(t+1) &= A\hat{\tilde{x}}_p(t) + \bar{B}(\bar{u}_p(t) - \bar{u}_{p-1}(t))\\ &\quad + (K_p - K_{p-1})y(t) - K_pC\hat{x}_p(t)\\ &\quad + K_{p-1}C\hat{x}_{p-1}(t)\\ &\quad - K_p\bar{D}\bar{u}_p(t) + K_{p-1}\bar{D}\bar{u}_{p-1}(t).\end{aligned}$$

If we now substitute $\hat{x}_p(t) = \hat{x}_{p-1}(t) + \hat{\tilde{x}}_p(t)$ in the equation above, we obtain

$$\begin{aligned}\hat{\tilde{x}}_p(t+1) &= (A - K_pC)\hat{\tilde{x}}_p(t) + \bar{B}(\bar{u}_p(t) - \bar{u}_{p-1}(t))\\ &\quad + (K_p - K_{p-1})(y(t) - C\hat{x}_p(t))\\ &\quad - K_p\bar{D}\bar{u}_p(t) + K_{p-1}\bar{D}\bar{u}_{p-1}(t). \quad (22)\end{aligned}$$

Simplifying $\bar{B}(\bar{u}_p(t) - \bar{u}_{p-1}(t))$ as

$$\begin{aligned}\bar{B}(\bar{u}_p(t) - \bar{u}_{p-1}(t)) &= \begin{bmatrix} B & E \end{bmatrix}\left(\begin{bmatrix} u(t) \\ u(t) \otimes \hat{x}_{p-1}(t) \end{bmatrix}\right.\\ &\quad \left. - \begin{bmatrix} u(t) \\ u(t) \otimes \hat{x}_{p-2}(t) \end{bmatrix}\right)\\ &= Eu(t) \otimes \hat{\tilde{x}}_{p-1}(t).\end{aligned}$$

Then we have

$$\begin{aligned}\bar{u}_p(t) &= \begin{bmatrix} u(t) \\ u(t) \otimes \hat{x}_{p-1}(t) \end{bmatrix}\\ &= \begin{bmatrix} u(t) \\ u(t) \otimes (\hat{x}_{p-2}(t) + \hat{\tilde{x}}_{p-1}(t)) \end{bmatrix}\\ &= \begin{bmatrix} u(t) \\ u(t) \otimes \hat{\tilde{x}}_{p-1}(t) \end{bmatrix} + \begin{bmatrix} 0_{m\times 1} \\ u(t) \otimes \hat{x}_{p-2}(t) \end{bmatrix}\\ &= \bar{u}_{p-1}(t) + \begin{bmatrix} 0_{m\times 1} \\ u(t) \otimes \hat{x}_{p-2}(t) \end{bmatrix},\end{aligned}$$

and we can now re-write (24) as

$$\begin{aligned}\hat{\tilde{x}}_p(t+1) &= (A - K_pC)\hat{\tilde{x}}_p(t)\\ &\quad + (E - K_pF)u(t) \otimes \hat{\tilde{x}}_{p-1}(t)\\ &\quad + (K_p - K_{p-1})(y(t) - C\hat{x}_{p-1}(t)\\ &\quad - \bar{D}\bar{u}_{p-1}(t)). \quad (23)\end{aligned}$$

The covariance $\hat{\tilde{P}}_p = \mathbb{E}\{\hat{\tilde{x}}_p(t)\hat{\tilde{x}}_p^T(t)\}$ is the solution of the equation

$$\begin{aligned}\hat{\tilde{P}}_p &= (A - K_pC)\hat{\tilde{P}}_p(A - K_pC)^T\\ &\quad + (E - K_pF)(R_u(0) \otimes \hat{\tilde{P}}_{p-1})(E - K_pF)^T\\ &\quad + (K_p - K_{p-1})R_{p-1}(K_p - K_{p-1})^T, \quad (24)\end{aligned}$$

where

$$\begin{aligned}R_p &= \mathbb{E}\{(y(t) - C\hat{x}_p(t) - \bar{D}\bar{u}_p(t))(y(t)\\ &\quad - C\hat{x}_p(t) - \bar{D}\bar{u}_p(t))^T\}.\end{aligned}$$

If $\lim_{p\to\infty}(K_p - K_{p-1}) = 0_{n\times\ell}$, (26) has a solution since all eigenvalues of $A - K_pC$ are inside the unit circle. Therefore, when $p \to \infty$, the covariance of $\hat{\tilde{x}}_p(t)$ converges to the solution of

$$\begin{aligned}\hat{\tilde{P}} &= (A - KC)\hat{\tilde{P}}(A - KC)^T\\ &\quad + (E - KF)(R_u \otimes \hat{\tilde{P}})(E - KF)^T. \quad (25)\end{aligned}$$

The solution to (27) exists and is equal to $\hat{\tilde{P}} = 0_{n\times n}$ when $|\lambda_{max}(M)| < 1$ holds true [22], where $\lambda_{max}(M)$ denotes the maximum eigenvalue of the matrix $M \in \mathbb{R}^{n^2\times n^2}$, given by

$$\begin{aligned}M &= (A - KC) \otimes (A - KC)\\ &+ \sum_{k=1}^{m}\sum_{w=1}^{m}((E_k - KF_k) \otimes (E_w - KF_w))\{R_u(0)\}_{kw}.\end{aligned}$$

This implies that when $\lim_{p\to\infty}\hat{x}_p(t) = \hat{x}(t)$, where $\hat{x}(t)$ is the state vector of the bilinear Kalman filter [2], i.e.,

$$\begin{aligned}\hat{x}(t+1) &= A\hat{x}(t) + Bu(t) + Eu(t) \otimes \hat{x}(t)\\ &\quad + K(y(t) - C\hat{x}(t) - Du(t) - Fu(t) \otimes \hat{x}(t)) \quad (26)\end{aligned}$$

with

$$\begin{aligned}K &= \mathcal{S}\mathcal{R}^{-1}\\ \mathcal{S} &= G - A\hat{P}C^T - BR_u(0)D^T - E(R_u(0) \otimes \hat{P})F^T\\ \mathcal{R} &= \Lambda(0) - C\hat{P}C^T - DR_u(0)D^T - F(R_u(0) \otimes \hat{P})F^T\end{aligned}$$

and $\hat{P} = \mathbb{E}\{\hat{x}(t)\hat{x}^T(t)\}$ is the solution of the Ricatti equation

$$\begin{aligned}\hat{P} &= A\hat{P}A^T + BR_u(0)B^T + E(R_u(0) \otimes \hat{P})E^T\\ &\quad + \mathcal{S}\mathcal{R}^{-1}\mathcal{S}^T. \quad (27)\end{aligned}$$

To conclude the convergence proof we need to show that $\lim_{p\to\infty}(K_p - K_{p-1}) = 0_{n\times\ell}$. Along these lines, we consider the equation for $\tilde{x}_p(t)$, i.e.,

$$\begin{aligned}\tilde{x}_p(t+1) &= A\tilde{x}_p(t) + Eu(t) \otimes \tilde{x}_{p-1}(t) + q(t)\\ &\quad - K_pC\tilde{x}_p(t) - K_pFu(t) \otimes \tilde{x}_{p-1}(t)\\ &\quad - K_pr(t).\end{aligned}$$

If $\tilde{x}_p(t)$ is a stationary signal with covariance $\tilde{P}_p = \mathbb{E}\{\tilde{x}_p(t)\tilde{x}_p^T(t)\}$, then $\tilde{P}_p$ is the solution to the Ricatti equa-

tion

$$\begin{aligned}\tilde{P}_p &= A\tilde{P}_pA^T + E(R_u(0)\otimes\tilde{P}_{p-1})E^T + Q \\ &\quad -A\tilde{P}_pC^TK_p^T - E(R_u(0)\otimes\tilde{P}_{p-1})F^TK_p^T \\ &\quad -SK_p^T - K_pC\tilde{P}_pA^T \\ &\quad -K_pF(R_u(0)\otimes\tilde{P}_{p-1})E^T \\ &\quad +K_pC\tilde{P}_pC^TK_p^T + K_pF(R_u(0)\otimes\tilde{P}_{p-1})F^TK_p^T \\ &\quad -K_pS^T + K_pRK_p^T \\ &= A\tilde{P}_pA^T + E(R_u(0)\otimes\tilde{P}_{p-1})E^T + Q \\ &\quad -\tilde{G}_p\tilde{\Lambda}_p^{-1}\tilde{G}_p^T \\ &\quad + \left[K_p\tilde{\Lambda}_p - \tilde{G}_p\right]\tilde{\Lambda}_p^{-1}\left[K_p\tilde{\Lambda}_p - \tilde{G}_p\right]^T, \end{aligned} \tag{28}$$

where

$$\begin{aligned}\tilde{G}_p &= A\tilde{P}_pC^T + S + E(R_u(0)\otimes\tilde{P}_{p-1})F^T \\ \tilde{\Lambda}_p &= R + C\tilde{P}_pC^T + F(R_u(0)\otimes\tilde{P}_{p-1})F^T.\end{aligned}$$

From this last equation we see that $\tilde{P}_p$ is minimized when $K_p\tilde{\Lambda}_p = \tilde{G}_p$ or

$$K_p = \tilde{G}_p\tilde{\Lambda}_p^{-1}.$$

This last expression provides an alternative computation for the gain of the Kalman filter. Upon substituting for K_p in (30) we obtain the state error covariance equation at the p^{th} iteration. That is,

$$\tilde{P}_p = A\tilde{P}_pA^T + E(R_u(0)\otimes\tilde{P}_{p-1})E^T + Q - \tilde{G}_p\tilde{\Lambda}_p^{-1}\tilde{G}_p^T.$$

Since $\bar{Q}_p = E(R_u(0)\otimes\tilde{P}_{p-1})E^T + Q$, we can re-write $\tilde{P}_p$ as a function of $\bar{Q}_p$, i.e.,

$$\tilde{P}_p = A\tilde{P}_pA^T + \bar{Q}_p - \tilde{G}_p\tilde{\Lambda}_p^{-1}\tilde{G}_p^T,$$

and because the covariance of $\bar{Q}_p$ decreases with p, the covariance $\tilde{P}_p$ must decrease also. On the other hand, given that $\bar{Q}_p$ is bounded below by Q, $\tilde{P}_p$ can never be zero, unless $Q = 0_{n\times n}$, which occurs in the deterministic case. Furthermore, $\tilde{P}_p$ converges to the solution of

$$\tilde{P} = A\tilde{P}A^T + E(R_u(0)\otimes\tilde{P})E^T + Q - \tilde{G}\tilde{\Lambda}^{-1}\tilde{G}^T,$$

where

$$\begin{aligned}\tilde{G} &= A\tilde{P}C^T + S + E(R_u(0)\otimes\tilde{P})F^T \\ \tilde{\Lambda} &= R + C\tilde{P}C^T + F(R_u(0)\otimes\tilde{P})F^T.\end{aligned}$$

Therefore, $\lim_{p\to\infty}K_p = K = \tilde{G}\tilde{\Lambda}^{-1}\tilde{G}^T$ and $\lim_{p\to\infty}(K_p - K_{p-1}) = 0_{n\times\ell}$.

The above analysis proves that $\hat{x}_p(t)$ converges towards $\hat{x}(t)$, whose covariance matrix $\hat{P}$ is a solution to (29). We should point out that (29) is the covariance equation of the Kalman filter states obtained in [2]. We conclude that the sequence of linear Kalman filters with states $\hat{x}_p(t)$, converges to the Kalman filter of the bilinear system. Furthermore, a system identification algorithm using the above state sequences iteratively, converges to a Kalman filter model for the bilinear system.

V. Iterative Bilinear System Identification Algorithm

The algorithm can be outlined as follows:

Step 0: Initialization. Set $p = 0$. Model the bilinear system by

$$\begin{aligned}x(t+1) &= Ax(t) + Bu(t) + \bar{q}_0(t) \\ y(t) &= Cx(t) + Du(t) + \bar{r}_0(t)\end{aligned}$$

and identify the parameters $\{A_0, B_0, C_0, D_0, K_0\}$ of the "forward innovations model" that describes the above system using a deterministic-stochastic subspace identification algorithm such as MOESP, N4SID, CVA, etc. [17], [20], [21]. Let $\bar{u}_0(t) = u(t)$, $\bar{B}_0 = B_0$, and $\bar{D}_0 = D_0$.

Step 1: Estimate the state sequence $\hat{x}_p(t)$ using the Kalman filter

$$\begin{aligned}\hat{x}_p(t+1) &= A_p\hat{x}_p(t) + \bar{B}_p\bar{u}_p(t) \\ &\quad +K_p\left[y(t) - C_p\hat{x}_p(t) - \bar{D}_p\bar{u}_p(t)\right],\end{aligned}$$

initialized with $\hat{x}_p(0) = 0_{n\times 1}$.

Step 2: Set $p = p+1$ and compute the new input sequence

$$\bar{u}_p(t) = \begin{bmatrix} u(t) \\ u(t)\otimes\hat{x}_{p-1}(t)\end{bmatrix},$$

and identify the parameters $\{A_p, \bar{B}_p, C_p, \bar{D}_p, K_p\}$ of the "forward innovations model" for the system

$$\begin{aligned}x(t+1) &= Ax(t) + \bar{B}\bar{u}_p(t) + \bar{q}_p(t) \\ y(t) &= Cx(t) + \bar{D}\bar{u}_p(t) + \bar{r}_p(t),\end{aligned}$$

where $\bar{B} = \begin{bmatrix} B & E\end{bmatrix}$ and $\bar{D} = \begin{bmatrix} D & F\end{bmatrix}$.

Step 3: If convergence, then go to Step 4, else go to Step 1.

Step 4: Extract identified parameters: $\hat{A} = A_p$, $\hat{B} = \bar{B}_p(:, 1:m)$, $\hat{C} = C_p$, $\hat{D} = \bar{D}_p(:, 1:\ell)$, $\hat{E} = \bar{B}_p(:, m+1:m(n+1))$, $\hat{F} = \bar{D}_p(:, \ell+1:\ell(n+1))$, and $\hat{K} = K_p$.

VI. Numerical Results

In order to test the algorithm we simulated a 3^{rd} order bilinear system with parameters

$$\begin{aligned}A &= \begin{bmatrix} 0.95 & 0.15 & 0.00 \\ -0.50 & 0.70 & 0.00 \\ 0.00 & -0.25 & 0.85\end{bmatrix}, \\ B &= \begin{bmatrix} 0.30 \\ -0.50 \\ 0.00\end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 & 2\end{bmatrix}, \\ D &= [1], \quad E = \begin{bmatrix} 0.20 & 0.00 & 0.00 \\ 0.00 & 0.25 & 0.00 \\ 0.00 & 0.00 & 0.30\end{bmatrix}, \\ F &= 0_{1\times 3}, \quad \begin{bmatrix} Q & S \\ S^T & R\end{bmatrix} = \sigma^2\left[\begin{array}{c|c} I_3 & 0_{3\times 1} \\ \hline 0_{1\times 3} & 1\end{array}\right].\end{aligned}$$

The input $u(t)$ is a white noise binary signal with $R_u(0) = 1$. We simulated three input/output pairs $\{u(t), y(t)\}$, for $t = 1, 2, \ldots, 2000$. From these data sets we used the first 1000 points for identification and the remaining points for model validation. For each simulation we used a different values of σ, namely, $\sigma = 0.1$, $\sigma = 0.2$, and $\sigma = 1.0$. The corresponding signal to noise ratios were, respectively, 17 db, 13.5 db, and 8.5 db. For each simulation, we calculated the prediction error in percentages given by [20]

$$\epsilon_p = 100\sqrt{\frac{\sum_{t=1}^{N} \hat{e}_p^2(t)}{\sum_{t=1}^{N} \hat{y}_p^2(t)}},$$

where $\hat{y}_p(t)$ is the output predicted by the Kalman filter and $\hat{e}_p(t) = y(t) - \hat{y}_p(t)$ is the prediction error. In Table 1 we compare ϵ_p values obtained from the true system (ϵ_{pt}) with the estimated models ϵ_{pe}, for both identification and validation data. The superscripts i and v denote, respectively, identification and validation data.

Table 1. ϵ_p values for identification and validation experiments.

SNR(db)	σ	ϵ_{pt}^i	ϵ_{pe}^i	ϵ_{pt}^v	ϵ_{pe}^v
17.0	0.1	13.5	12.6	12.7	13.0
13.5	0.2	22.5	22.2	22.0	22.3
8.5	1.0	34.8	34.5	42.4	42.0

Table 1 show that despite the high noise levels (low SNR values), there are no significant differences between the prediction errors of both the true and estimated models.

VII. Conclusions

In this paper we proposed a new subspace identification algorithm for MIMO bilinear systems driven by white noise inputs. The algorithm is based on a Picard iterative process, which overcomes the typical curse of dimensionality common to direct bilinear subspace algorithms. We showed the conditions under which the Picard process converges to the true bilinear state space model. We tested the algorithm under extreme simulated noise conditions (low and high noise conditions) and the algorithm performed well. In all cases the estimated and true models predicted the system output with the same accuracy. The case of general inputs will be investigated in future work.

References

[1] Favoreel, W., De Moor, B., and P. Vanoverschee, "Subspace identification of bilinear systems subject to white inputs," *IEEE Trans. Automat. Contr.*, **AC-44**, No. 6, pp. 1157-1169, 1999.

[2] Favoreel, W., "Subspace methods for identification and control linear and bilinear systems," Ph.D. Thesis, Department of Electrical Engineering, Katholieke Universiteit Leuven, Leuven, Belgium, 1999.

[3] Chen, H. and J. Maciejowski, "Subspace identification of combined deterministic-stochastic bilinear systems," Proceedings of the IFAC SYSID 2000, Santa Barbara, USA, pp. , 2000.

[4] Chen, H. and J. Maciejowski, "An improved subspace identification method for bilinear systems," Proceedings of the 39^{th} IEEE Conference on Decision and Control, Sidney, Australia, 2000.

[5] Isidori, A., "Direct construction of minimal bilinear realizations from nonlinear input-output maps," *IEEE Trans. Automat. Contr.*, **AC-18**, No. 6, pp. 626-631, 1973.

[6] Desai, U. B., "Realization of bilinear stochastic systems," *IEEE Trans. Automat. Contr.*, **AC-31**, No. 2, pp. 189-192, 1986.

[7] Hsu, C. S., Desai, U. B., and C. A. Crawley, "Realization and approximation of discrete bilinear systems," In: Applied Digital Control, S. G. Tzafestas (ed.), Elsevier Science Publishers, North-Holland, pp. 171-187, 1985..

[8] Khatri, C. G. and Rao, C.R., "Solutions to some functional equations and their applications to characterization of probability distributions," Sankhya: The Indian J. Stat., series A, 30, pp. 167-180, 1968.

[9] Moonen, M., De Moor, B., Vandenberghe, L., and J. Vandewalle, "On- and off-line identification of linear state space models," *Int. Journal of Control*, Vol. 49, No. 1, pp. 219-232, 1989.

[10] Verdult, V. and M. Verhaegen, "Bilinear state space systems for nonlinear dynamics," *Theory of Biosciences*, vol. 119, pp. 1-9, 2000.

[11] Verdult, V. and M. Verhaegen, "Subspace based identification of MIMO bilinear systems, Proc. of the ECC, Karlruhe, Germany, 1999.

[12] Verdult, V., Verhaegen, M., Chou, C. T., and M. Lovera, "Efficient and systematic identification of MIMO bilinear state space models, Proc. of the 37^{th} IEEE Conference on Decision and Control, Tampa, Florida, 1998.

[13] Verdult, V. and M. Verhaegen, "A kernel method for subspace identification of multivariable bilinear systems, Proc. of the 42^{nd} IEEE Conf. on Decision and Control, Maui, Hawaii, USA, December 2003.

[14] Ramos, J. A., Moonen, M., and D. Mallant, "Modeling the rainfall-runoff process with a bilinear state-space model," Proc. 1991 European Conference, Ghent, Belgium, November 1991.

[15] Bruni, C. Dipillo, G., and G. Koch, "Bilinear systems: An appealing class of nearly linear systems in theory and applications," *IEEE Trans. Automat. Contr.*, **AC-19**, No. 8, pp. 334-348, 1974.

[16] Mohler, R. and W. J. Kolodziej, "An overview of bilinear system theory and applications," *IEEE Trans. Man and Cybernetics*, vol. 10, no. 10, pp. 683-688, 1980.

[17] Verhaegen, M., "Identification of the deterministic part of MIMO state space models given in innovations form from input-output data," *Automatica*, vol. 30, no. 1, pp. 61-74, 1994.

[18] Van Overschee, P. and B. De Moor, "Subspace algorithms for the stochastic identification problem," *Automatica*, Vol. 29, No. 3, pp. 649-660, 1993.

[19] Van Overschee, P. and B. De Moor, "*Subspace identification for linear systems*," Kluwer Academic Publishers, Boston, MA, 1996.

[20] Van Overschee, P. and B. De Moor, "N4SID: Subspace algorithms for the identification of combined deterministic and stochastic systems," *Automatica*, vol. 30, no. 1, pp. 75-93, 1994.

[21] Larimore, W. E., "The optimality of canonical variate identification by example, filtering, and adaptive control," Proc. SYSID'94, Vol. 2, pp. 151-156, 4-6 July, Copenhagen, Denmark, 1994.

[22] Magnus, J. R. and H. Neudecker, "*Matrix differential calculus with applications in statistics and economics,*" Wiley Publisher, New York, 1999.