AAS 14-307

BILINEAR OBSERVER/KALMAN FILTER IDENTIFICATION

Francesco Vicario,[*] Minh Q. Phan,[†] Raimondo Betti[‡] and Richard W. Longman[§]

Bilinear systems are important per se since several phenomena in engineering and other fields are inherently bilinear. Even more interestingly, bilinear systems can approximate more general nonlinear systems, providing a promising approach to handle various nonlinear identification and control problems, such as satellite attitude control. This paper develops and demonstrates via numerical examples a method for discrete-time state-space model identification for bilinear systems in the presence of noise in the process and in the measurements. The formulation relies on a bilinear observer which is proven to have properties similar to the linear Kalman filter under the sole additional assumption of stationary white excitation input, and on a novel approach to system identification based on the estimation of the observer residuals. The latter are used to construct a new, noise-free identification problem, in which the observer is identified and the matrices of the system state-space model are recovered. The resulting method represents the bilinear counterpart of the Observer/Kalman filter Identification (OKID) approach for linear systems, originally developed for the identification of lightly-damped structures and distributed by NASA.

INTRODUCTION

System identification as a research topic has attracted a lot of interest over the last decades and applications have been increasing in many fields including aerospace engineering. Although well-established techniques exist for linear-time-invariant system identification, this is not the case for linear-time-varying systems and for nonlinear systems. Bilinear systems are a specific type of nonlinear systems. Bilinear systems are important per se since several phenomena in engineering (in particular chemical processes), biology, physiology, sociology and other fields (Reference 1) are inherently bilinear. Even more appealing is the fact that bilinear models can approximate more general nonlinear systems (References 2, 3), namely those with input appearing linearly (input-affine dynamic systems). Interest in bilinear systems has recently grown after Carleman linearization was found to be a technique to obtain such an approximation (Reference 4). Bilinear systems can mean systems having multiple variables, and having the property that they are linear in each variable if the remaining variables are held constant. In satellite attitude dynamics, the Euler equations for the satellite angular velocity as a function of applied torque are bilinear equations in this sense. One needs to add to these equations, additional equations that represent the kinematics of rigid body motion, which then give the satellite attitude as a function of time that results from the angular

[*]Doctoral Student, Department of Mechanical Engineering, Columbia University, New York, NY 10027, USA.
[†]Associate Professor, Thayer School of Engineering, Dartmouth College, Hanover, NH 03755, USA.
[‡]Professor, Department of Civil Engineering and Engineering Mechanics, Columbia University, New York, NY 10027, USA.
[§]Professor, Department of Mechanical Engineering and Department of Civil Engineering and Engineering Mechanics, Columbia University, New York, NY 10027, USA.

velocity history. There are various choices for these kinematic equations: one can use 9 direction cosine matrix elements, or 6 direction cosine matrix elements, or 4 parameter representations such as quaternions, Rodriguez parameters, and Euler Rodriguez parameters, or 3 parameter sets of Euler angles. The first few of these choices also produce bilinear equations in this sense, so that the satellite attitude control problem can be thought of as a set of 12 or 9 bilinear equations. The literature on bilinear systems usually makes use of a different form of bilinear equation that only contains nonlinear terms that are products of a state variable with the input. Carleman bilinearization will convert the satellite attitude dynamics bilinear problem into this form of bilinear equations (References 5, 6), so that one can think of system identification of bilinear systems, developing state estimators (observers) and control methods for bilinear systems as an approach to handle the satellite attitude control problem and other nonlinear control problems of interest.

A first significant step towards this objective has been made in References 7, 8, where the concept of interaction matrix was successfully extended to discrete-time bilinear systems and used to develop several bilinear system identification algorithms of deterministic type (i.e. formulated without taking into account the presence of noise in the measured input-output data). Successive developments have led the authors to propose a deterministic bilinear observer (Reference 9), and to interpret the interaction matrices used in References 7,8 as the gains of the fastest possible deterministic bilinear observer. The finding parallels one of the results of the well-known Observer-Kalman filter Identification (OKID) method for linear systems (References 10,11,12). OKID was originally developed for lightly damped linear structures, it is at the core of the software package distributed by NASA with title SOCIT, and over the last twenty years has been proven to be very successful in many other engineering applications (see, for example, References 12, 13, 14, 15). OKID is able to optimally perform the identification process also in the presence of zero-mean white noise in the process and in the measurements (tandard Kalman filter assumptions), in which case the interaction matrix turns out to be the gain of the optimal stochastic observer (Kalman filter) corresponding to the system to be identified and the noise covariance embedded in the data (both of which are unknown before the identification). The original OKID algorithm is referred to as OKID/ERA because after estimating by least-squares from the OKID core equation the Markov parameters of the underlying Kalman filter and recovering from them the Markov parameters of the system, it uses the latter as the input to the Eigenvalue Realization Algorithm (ERA, References 16, 17) to find a realization of the system matrices and the corresponding Kalman gain.

In this paper, we develop a new method for the identification of bilinear discrete-time state-space models from noisy input-output data, based on the OKID approach. First of all, we prove that the bilinear observer proposed in Reference 9 has stochastic properties similar to those of the linear Kalman filter, under the sole additional assumption of dealing with stationary white excitation input. The second main contribution of the paper allows us to get around the lack of the bilinear counterpart of ERA, by taking the following alternative approach. Instead of focusing on the observer Markov parameters estimated in the first step of the method, we prove that the corresponding least-squares residuals are the residuals of the above bilinear observer. We then use such residuals to construct a new identification problem where the dynamic system to be identified is the bilinear observer (instead of the original system). The key advantage of the new identification problem is that no (unknown) noise term is present in its formulation. The identification of the observer can then be accomplished by the deterministic bilinear identification algorithms of References 7, 8. Finally, the system matrices can be easily recovered by the identified bilinear Kalman filter matrices. The outcome of the paper is a bilinear system identification method optimally taking into account the noise corrupting the measured data, which marks a significant improvement over the deterministic methods of Reference 8 and represents the first extension of the well-known OKID approach to nonlinear systems.

PROBLEM STATEMENT

Consider the following n-state, single-input, q-output discrete-time bilinear system in state-space form

$$x(k+1) = Ax(k) + Nx(k)u(k) + Bu(k) + w_p(k) \tag{1a}$$
$$y(k) = Cx(k) + Du(k) + w_m(k) \tag{1b}$$

where $x \in \mathbb{R}^{n\times 1}$ is the state vector, $u \in \mathbb{R}$ is the input, $y \in \mathbb{R}^{q\times 1}$ is the output vector and $A \in \mathbb{R}^{n\times n}$, $N \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times 1}$, $C \in \mathbb{R}^{q\times n}$, $D \in \mathbb{R}^{q\times 1}$ are the matrices governing the dynamics of the system. Both the process (or state) equation, Eq. (1a), and the measurement (or observation) equation, Eq. (1b), are affected by stationary zero-mean, white noises $w_p \in \mathbb{R}^{n\times 1}$ and $w_m \in \mathbb{R}^{q\times 1}$ with covariance matrices $Q \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{q\times q}$, respectively. Both w_p and w_m are assumed to be uncorrelated with u and y as well as, for simplicity, mutually uncorrelated.

A single set of length l of input-output data that starts from some unknown initial state $x(0)$ is given

$$\{u(k)\} = \{u(0),\ \ u(1),\ \ u(2),\ \ ...,\ \ u(l-1)\} \tag{2a}$$
$$\{y(k)\} = \{y(0),\ \ y(1),\ \ y(2),\ \ ...,\ \ y(l-1)\} \tag{2b}$$

The objective is to identify the system of Eq. (1), i.e. the matrices A, N, B, C, D, with the input-output data provided in Eq. (2). The process and measurement noises are unknown, as well as their covariance matrices. The data of Eq. (2) is assumed to be of sufficient length and richness so that the system of Eq. (1) can be correctly identified. For simplicity, we focus on a single-input model in this work. Extension to the multi-input case can be made without conceptual difficulties.

APPROACH OVERVIEW

As mentioned in the introduction, OKID/ERA (References 10, 11, 12, 13) has proven to be a very successful technique for the identification of linear systems, solving the linear counterpart ($N = 0$) of the problem addressed in this paper. OKID/ERA is based on a relationship between the measured input and output which was originally derived via the interaction matrix technique. In the presence of noise in the data, the least-squares solution of the corresponding set of equations is proven to yield the Markov parameters (or unit pulse response) of the optimal linear observer (Kalman filter) for the system to be identified and the noise statistics embedded in the data. From the Markov parameters of the observer, those of the system can be recovered and fed to the Eigensystem Realization Algorithm (ERA, Reference 16) or to the ERA with Data Correlation (ERA/DC, Reference 17) to find a realization of the system (matrices A, B, C, D) and the corresponding Kalman gain. The use of ERA (or ERA/DC) to complete the OKID process is not the only possible choice as proven in Reference 18, and in this work we exploit such finding to overcome the lack of a bilinear version of ERA and develop the first OKID-based identification method for bilinear systems. The method is articulated around the main steps described below.

Input-Output Relationship via Interaction Matrices

In system identification, the measured input and output are the only known signals. No knowledge of the evolution of the state over time is generally available. It is then useful to derive an equation relating the output of the system directly to the input, without the state appearing explicitly. In the same fashion as in the original OKID work (Reference 12), interaction matrices are used to derive such a relation, which can be classified as a bilinear AutoRegressive model with eXogenous input (ARX). The proof for the existence of interaction matrices for bilinear systems is given in Reference 8. Due to its central role in this paper, the corresponding theorem is restated and briefly discussed in the next section.

Estimation of Observer Residuals

The bilinear ARX gives rise to a set of algebraic equations, which represents the core of OKID. The least-squares (LS) solution of the bilinear ARX equations are proven to be related to the optimal bilinear observer corresponding to the system to be identified and to the noise statistics embedded in the data. In particular, the LS residuals of the bilinear OKID core equation are the residuals ϵ of the optimal bilinear observer.

Construction of a Noise-Free Identification Problem

Along the same lines of Reference 18, the observer residuals are used to construct the following new identification problem

$$\hat{x}(k+1) = A\hat{x}(k) + N\hat{x}(k)u(k) + Bu(k) + K'\epsilon(k) + K''\epsilon(k)u(k) \tag{3a}$$

$$\hat{y}(k) = C\hat{x}(k) + Du(k) \tag{3b}$$

Since, from the definition of observer residual we can compute $\hat{y}(k) = y(k) - \epsilon(k)$, Eq. (3) can be looked at as the state-space model of a dynamic system whose inputs, $u(k), \epsilon(k), \epsilon(k)u(k)$, and output, $\hat{y}(k)$, are known. Additionally, notice that no (unknown) noise term is present in Eq. (3). We can then think of identifying the matrices of such system, namely A, N, B, K', K'', C and D, with a deterministic bilinear system identification method.

Observer Identification

The methods developed in Reference 8 are used to solve the noise-free system identification problem in Eq. (3). The reader will notice similarities in the technique used to derive the Input-Output-to-State relationships at the core of the methods in Reference 8 and the way the bilinear ARX model is derived in the present work. Indeed, the proposed bilinear OKID method can be interpreted as the extension of the methods of Reference 8 to bilinear identification problems where the dynamic process and the measurements are corrupted by noise. The crucial difference is that in the latter case the optimal bilinear observer is identified instead of the original dynamical system. Note that the identification of the observer of Eq. (3) solves the original problem of identifying the system of Eq. (1). Due to the peculiarities of the identification problem of Eq. (3), namely different input in the state equation and the observation equation, the deterministic identification algorithms in Reference 8 are briefly reviewed and accordingly modified in a dedicated section.

USEFUL CONCEPTS

Before plunging into the detailed derivation of the bilinear OKID method, a few concepts central to this work are presented below.

Existence of Bilinear Interaction Matrices

The concept of interaction matrix was originally formulated by Minh Q. Phan in the context of linear system identification of lightly-damped large flexible space structures. The dynamics of such structures is described by an infinitely long sequence of Markov parameters, and the interaction matrix provides a mechanism to compress such sequence into a finite sequence that can be easily identified from input-output measurements (References 10, 11). Interaction matrices have found various applications, as summarized in Reference 19. One of them is notably the OKID method for linear system identification, which the present paper extends to the bilinear case.

The existence of interaction matrices for bilinear systems is guaranteed by a theorem proven and thoroughly discussed in Reference 8. Such theorem is invoked at several points in this paper, therefore it is briefly reviewed in this section.

Bilinear Interaction Matrix Theorem. *Given the matrices $A \in \mathbb{R}^{n\times n}$, $N \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{q\times n}$, define the matrices $\bar{A}_M = A - M'C$ and $\bar{N}_M = N - M''C$ and the product*

$$\mathcal{S}_{M,p}(k) = \left(\bar{A}_M + \bar{N}_M u(k-1)\right)\left(\bar{A}_M + \bar{N}_M u(k-2)\right) \dots \left(\bar{A}_M + \bar{N}_M u(k-p)\right)$$

where M', $M'' \in \mathbb{R}^{n\times q}$ are called interaction matrices and $\{u(k-p), u(k-p+1), ..., u(k-1)\}$ is a scalar sequence. If (A,C) is an observable pair, then there exist some interaction matrices M', M'' and a positive scalar γ such that $\mathcal{S}_{M,p}(k)$ converges asymptotically to 0 with p provided that $|u(i)| < \gamma$ for all $i = k-p, k-p+1..., k-1$.

As discussed in details in Reference 8, it is worth noting that there are bilinear systems, referred to as *ideal*, for which there exist M' and M'' such that $\mathcal{S}_{M,p}(k)$ is identically equal to 0 for $p \geq n$ and $\gamma \to \infty$ (i.e. no bound on $|u|$). The class of ideal bilinear systems comprises, among others, all those with $\operatorname{rank} C = n$. For arbitrary (*non-ideal*) bilinear systems, $\mathcal{S}_{M,p}(k)$ cannot be identically equal to 0 but it is guaranteed to converge to 0 asymptotically with p thanks to M', M'' and the finite bound γ on the magnitude of the values that u can take.

Bilinear Observer

At the core of the linear OKID method are steady-state linear state observers, in particular linear-time-invariant observers minimizing $\mathbb{E}\left[e^T(k)e(k)\right]$ for all k after the initial transient, where $e(k) = x(k) - \hat{x}(k)$ is the state estimation error. Such observers are the deadbeat observer (in the absence of noise) and the Kalman filter (in the presence of noise). In the bilinear version of OKID we rely on a bilinear observer defined by a similar optimality criterion. We introduce it below together with its properties of interest. Its structure is the same presented in Reference 9, where the connection between the interaction matrices in the noise-free bilinear identification problem of Reference 8 and the bilinear observer gains is established and numerical methods to design the latter are provided for both the deterministic and stochastic case.

Structure. Define the following bilinear time-invariant observer for the system of Eq. (1)

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + N\hat{x}(k)u(k) + G'(y(k) - \hat{y}(k)) + G''(y(k) - \hat{y}(k))u(k) \tag{4a}$$

$$\hat{y}(k) = C\hat{x}(k) + Du(k) \tag{4b}$$

where $\hat{y}$ is the predicted output based on the estimated state $\hat{x}$. To remark the role of the observer in bilinear OKID, in this work $\hat{y}$ and $\hat{x}$ are often referred to as the observer output and state, respectively. By defining the observer residual as

$$\epsilon(k) = y(k) - \hat{y}(k) \tag{5}$$

the structure of Eq. (4) is equivalent to Eq. (3). Another equivalent form for the above bilinear observer is the following

$$\hat{x}(k+1) = \bar{A}_G\hat{x}(k) + \bar{N}_G\hat{x}(k)u(k) + \bar{B}_G v(k) \tag{6a}$$

$$\hat{y}(k) = C\hat{x}(k) + Du(k) \tag{6b}$$

where

$$\bar{A}_G = A - G'C \qquad \bar{B}_G = \begin{bmatrix} B - G'D & G' & -G''D & G'' \end{bmatrix} \tag{7a}$$

$$\bar{N}_G = N - G''C \qquad v(k) = \begin{bmatrix} u(k) & y^T(k) & u^2(k) & y^T(k)u(k) \end{bmatrix}^T \tag{7b}$$

Stability. By subtracting Eq. (4a) from Eq. (1a), it can be shown that the dynamics of the state estimation error $e(k) = x(k) - \hat{x}(k)$ is governed by

$$e(k+1) = \bar{A}_G e(k) + \bar{N}_G e(k)u(k) + w_p(k) - G'w_m(k) - G''w_m(k)u(k) \tag{8}$$

and the stability of the observer depends on the homogeneous part of Eq. (8)

$$e(k+1) = \bar{A}_G e(k) + \bar{N}_G e(k)u(k) = \left(\bar{A}_G + \bar{N}_G u(k)\right) e(k) \tag{9}$$

Propagating Eq. (9) forward in time produces

$$\begin{aligned} e(1) &= \left(\bar{A}_G + \bar{N}_G u(0)\right) e(0) \\ e(2) &= \left(\bar{A}_G + \bar{N}_G u(1)\right) \left(\bar{A}_G + \bar{N}_G u(0)\right) e(0) \\ &\vdots \\ e(k) &= \mathcal{S}_{G,k}(k) e(0) \end{aligned} \tag{10}$$

where

$$\mathcal{S}_{G,k}(k) = \left(\bar{A}_G + \bar{N}_G u(k-1)\right) \left(\bar{A}_G + \bar{N}_G u(k-2)\right) \dots \left(\bar{A}_G + \bar{N}_G u(0)\right) \tag{11}$$

The bilinear interaction matrix theorem ensures the existence of some observer gains G' and G'' and a bound γ to the input u such that $\mathcal{S}_{G,k}(k)$ converges to 0 with k. Therefore the state estimation error decays to 0 regardless of its initial value, provided that $|u(k)| < \gamma$ at all k.

Optimality. Similarly to the linear Kalman filter, the optimality criterion for the bilinear observer of Eq. (4) or (6) concerns the expected value of the squared norm of the state estimation error $\mathbb{E}\left[e^T(k)e(k)\right]$. The criterion is equivalent to minimizing the trace of the covariance matrix of the state estimation error $\mathbb{E}\left[e(k)e^T(k)\right]$. However, in contrast to the linear case, even after the transient such matrix keeps changing over time because of the dependence of the state estimation error on the input. Under the assumption of stationary white input, the time average P of such covariance matrix can be shown to satisfy the following relation

$$P = \bar{A}_G P \bar{A}_G^T + G' R \left(G'\right)^T + Q + \left(\bar{N}_G P \bar{N}_G^T + G'' R \left(G''\right)^T\right) \beta^2 + \left(\bar{A}_G P \bar{N}_G^T + \bar{N}_G P \bar{A}_G^T + G' R \left(G''\right)^T + G'' R \left(G'\right)^T\right) \mu \tag{12}$$

where $\mu = \mathbb{E}\left[u\right]$ and $\beta^2 = \mathbb{E}\left[u^2\right]$. The optimality criterion corresponds to minimizing P. Note that in the linear case, since the estimation error covariance matrix is constant over time, its minimization at all k corresponds to the minimization of its time average. We impose the first-order conditions

$$\frac{\partial \operatorname{trace} P}{\partial G'} = -2\left(\bar{A}_G P C^T - G' R\right) - 2\left(\bar{N}_G P C^T - G'' R\right)\mu = 0 \tag{13a}$$

$$\frac{\partial \operatorname{trace} P}{\partial G''} = -2\left(\bar{N}_G P C^T - G'' R\right)\beta^2 - 2\left(\bar{A}_G P C^T - G' R\right)\mu = 0 \tag{13b}$$

which lead to the two following optimality conditions

$$\bar{A}_G P C^T - G' R = 0 \tag{14a}$$

$$\bar{N}_G P C^T - G'' R = 0 \tag{14b}$$

To remark their optimality, the gains G' and G'' satisfying Eq. (14) are denoted by K' and K'' in this paper.

Properties of the Residuals. Under the assumption made to derive the optimal gain conditions of Eq. (14), i.e. stationary white excitation input u, the residuals $\epsilon(k)$ of the bilinear observer at steady state ($k \geq p$) can be shown to have the following properties if and only if the observer is optimal ($G' = K', G'' = K''$).

1. $\{\epsilon(k)\}$ is a zero-mean sequence, i.e. $\lim_{l\to+\infty} \frac{1}{l-p}\sum_{k=p}^{l-1} \epsilon(k) = 0$
2. $\{\epsilon(k)\}$ is a white sequence, i.e. $\lim_{l\to+\infty} \frac{1}{l-p-j}\sum_{k=p+j}^{l-1} \epsilon^T(k)\epsilon(k-j) = 0$ for $j > 0$
3. $\epsilon(k)$ is orthogonal to the past outputs and to the current and past inputs, and some products of them, i.e. $\lim_{l\to+\infty} \frac{1}{l-p}\sum_{k=p}^{l-1} \epsilon(k)u(k) = 0$ and $\lim_{l\to+\infty} \frac{1}{l-p}\sum_{k=p}^{l-1} \epsilon^T(k)z_p^T(k) = 0$

where $z_p(k)$ contains past inputs and outputs and is defined more precisely later in the paper, below Eq. (21). Note that such properties are similar to those of the Kalman filter that led to the development of OKID in the linear case. The proofs of the above properties will be presented in a separate paper.

ESTIMATION OF THE OBSERVER RESIDUALS

The proposed bilinear OKID method consists of two main steps, the first of which is the estimation of the optimal observer residuals. The task is accomplished by solving by LS a set of equations arising from an input-output relationship in which the state does not appear explicitly. The derivation in this section is done following the same strategy as in the original OKID paper (Reference 12). Accordingly, the presence of noise in Eq. (1) is temporarily ignored and its effect is discussed when solving the above mentioned LS problem. Additionally, connections with previous work on bilinear discrete-time system identification are made along the derivation to provide a comprehensive framework for the method proposed in this paper. Add and subtract the terms $H'y(k)$ and $H''y(k)u(k)$ to Eq. (1a) where H', $H'' \in \mathbb{R}^{n\times q}$ are interaction matrices, getting

$$\begin{aligned} x(k+1) &= Ax(k) + Nx(k)u(k) + Bu(k) + H'y(k) - H'y(k) + H''y(k)u(k) - H''y(k)u(k) \\ &= Ax(k) + Nx(k)u(k) + Bu(k) + H'y(k) - H'Cx(k) - H'Du(k) \\ &\quad + H''y(k)u(k) - H''Cx(k)u(k) - H''Du^2(k) \\ &= \bar{A}_H x(k) + \bar{N}_H x(k)u(k) + \bar{B}_H v(k) \end{aligned} \tag{15}$$

where

$$\bar{A}_H = A - H'C \qquad \bar{B}_H = \begin{bmatrix} B - H'D & H' & -H''D & H'' \end{bmatrix} \tag{16a}$$

$$\bar{N}_H = N - H''C \qquad v(k) = \begin{bmatrix} u(k) & y^T(k) & u^2(k) & y^T(k)u(k) \end{bmatrix}^T \tag{16b}$$

The interaction matrices convert the bilinear model A, B, N of Eq. (1a) into the equivalent bilinear model of Eq. (15) with matrices $\bar{A}_H$, $\bar{N}_H$, $\bar{B}_H$. The observation equation, Eq. (1b), does not change. The freedom introduced by H' and H'' will be used to impose conditions to express the state at the current time step k solely in terms of past input and output data. Equation (15) has the same form as Eq. (6), anticipating the connection between interaction matrices and observer gains. Propagating Eq. (15) one step forward, we get

$$\begin{aligned} x(k+2) &= \left(\bar{A}_H + \bar{N}_H u(k+1)\right) x(k+1) + \bar{B}_H v(k+1) \\ &= \left(\bar{A}_H + \bar{N}_H u(k+1)\right) \left(\left(\bar{A}_H + \bar{N}_H u(k)\right) x(k) + \bar{B}_H v(k)\right) + \bar{B}_H v(k+1) \\ &= \mathcal{S}_{H,2}(k)x(k) + T_{H,2} z_2(k+2) \end{aligned} \tag{17}$$

where

$$\mathcal{S}_{H,2}(k+2) = \left(\bar{A}_H + \bar{N}_H u(k+1)\right) \left(\bar{A}_H + \bar{N}_H u(k)\right) \tag{18a}$$

$$T_{H,2} = \begin{bmatrix} \bar{A}_H \bar{B}_H & \bar{N}_H \bar{B}_H & \bar{B}_H \end{bmatrix} \qquad z_2(k+2) = \begin{bmatrix} v^T(k) & v^T(k)u(k+1) & v^T(k+1) \end{bmatrix}^T \tag{18b}$$

Propagating Eq. (17) further in time, we obtain the more general expression

$$x(k+p) = \mathcal{S}_{H,p}(k+p)x(k) + T_{H,p} z_p(k+p) \tag{19}$$

which, by shifting the time index backward by p time steps, can be more conveniently written as

$$x(k) = \mathcal{S}_{H,p}(k)x(k-p) + T_{H,p} z_p(k) \tag{20}$$

where

$$\mathcal{S}_{H,p}(k) = \left(\bar{A}_H + \bar{N}_H u(k-1)\right) \left(\bar{A}_H + \bar{N}_H u(k-2)\right) ... \left(\bar{A}_H + \bar{N}_H u(k-p)\right) \tag{21}$$

$T_{H,p}$ contains products of $\bar{A}_H$, $\bar{N}_H$ and $\bar{B}_H$, and the general pattern to construct $z_p(k)$ is (Reference 7):

- $v(k-p), v(k-p+1),$ ... to $v(k-1)$

– $v(k-p)$ multiplied with products of $u(k-p+1)$ to $u(k-1)$ in all possible combinations $(p-1)C1$, $(p-1)C2$, ..., $(p-1)C(p-1)$ of $\{u(k-p+1), u(k-p+2), ..., u(k-1)\}$

– $v(k-p+1)$ multiplied with products of $u(k-p+2)$ to $u(k-1)$ in all possible combinations $(p-2)C1$, $(p-2)C2$, ..., $(p-2)C(p-2)$ of $\{u(k-p+2), u(k-p+3), ..., u(k-1)\}$
$\vdots$

– $v(k-3)$ multiplied with products of $u(k-2)$ to $u(k-1)$ in all possible combinations $2C1$, $2C2$ of $\{u(k-p+3), u(k-p+4), ..., u(k-1)\}$

– $v(k-2)$ multiplied with $1C1$ of $u(k-1)$, which of course is $u(k-1)$

By choosing p sufficiently large, the bilinear interaction matrix theorem ensures that $\mathcal{S}_{H,p}(k)$ vanishes and we can write

$$x(k) = T_{H,p} z_p(k) \tag{22}$$

So far the development is the same as in Reference 8, where Eq. (22) is referred to as an Inoput-Output-to-State Relationship (IOSR) and used as the starting point to develop the Equivalent Linear Model (ELM) and Intersection Subspace (IS) methods for bilinear system identification in the absence of noise.

Putting together Eqs. (22) and (1b), always assuming no noise in the measured data, we realize that

$$x(k) = T_{H,p} z_p(k) \tag{23a}$$
$$y(k) = Cx(k) + Du(k) \tag{23b}$$

are the equations of an observer in non-recursive form, whose state is determined solely by past input-output values. Indeed, the observation that the matrices H' and H'' embedded in $T_{H,p}$ can be interpreted as observer gains is exploited in Reference 9 to devise a technique to design the fastest bilinear observers in the absence of noise. For ideal bilinear systems, such observers are deadbeat in a strict sense, i.e. the observer state is exactly equal to the system state after $p = n$ steps. For arbitrary bilinear systems, the fastest deterministic observer state converges asymptotically to the system state, i.e. one needs to wait a larger number p of steps before the observer state approaches the system state. By plugging Eq. (23a) into Eq. (23b) we obtain the following relationship between the input and output

$$y(k) = CT_{H,p} z_p(k) + Du(k) \tag{24}$$

which is the equation used in Reference 20 to directly identify bilinear input-output maps in the absence of noise. By defining

$$\Phi_p = \begin{bmatrix} CT_{H,p} & D \end{bmatrix} \qquad v_p(k) = \begin{bmatrix} z_p^T(k) & u(k) \end{bmatrix}^T \tag{25}$$

Eq. (24) takes the form of the classic OKID equation

$$y(k) = \Phi_p v_p(k) \tag{26}$$

which can be written at all time steps $p \leq k \leq l-1$ to obtain the following set of equations

$$Y = \Phi_p V_p \tag{27}$$

where

$$Y = \begin{bmatrix} y(p) & y(p+1) & \dots & y(l-1) \end{bmatrix} \tag{28a}$$

$$V_p = \begin{bmatrix} v_p(p) & v_p(p+1) & \dots & v_p(l-1) \end{bmatrix} \tag{28b}$$

The presence of noise in Eq. (1) makes Eq. (27) inconsistent. The inconsistency can be expressed by an error term E

$$Y = \Phi_p V_p + E \tag{29}$$

with

$$E = \begin{bmatrix} \epsilon(p) & \epsilon(p+1) & \dots & \epsilon(l-1) \end{bmatrix} \tag{30}$$

The same symbol ϵ indicating the observer residuals is used for the error terms of Eq. (29). As shown later in this section, the two turns out to correspond.

By having a sufficiently long record, it is possible to find the least-squares (LS) solution to Eq. (29)

$$\tilde{\Phi}_{H,p} = Y V_p^T \left(V_p V_p^T \right)^{-1} = Y V_p^{\dagger} \tag{31}$$

where † denotes the Moore-Penrose pseudoinverse of V_p. Right-multiplying Eq. (29) by V_p^T and replacing Φ_p with its LS estimate $\tilde{\Phi}_{H,p}$, we obtain

$$Y V_p^T = Y V_p^T \left(V_p V_p^T \right)^{-1} V_p V_p^T + E V_p^T = Y V_p^T + E V_p^T \tag{32}$$

which implies that $E V_p^T = 0$, i.e. that

$$\sum_{k=p}^{l-1} \epsilon(k) v_p^T(k) = 0 \tag{33}$$

Assuming l is large and dividing Eq. (33) by $l-p$, we recognize its left-hand side as the time average of each entry of the product ϵv_p^T. In other words, the LS residuals $\epsilon(k)$ are orthogonal to all the entries of $v_p(k)$, i.e. $u(k)$ and all the entries of $z_p(k)$. The latter are products of past input-output data, in accordance with the pattern provided above Eq. (22). This is the same property characterizing the bilinear observer minimizing P, provided u is a stationary white process. Hence, under the stated assumptions, for a long record (large l) and for a sufficiently large p, the residuals corresponding to the LS solution of Eq. (29) are an estimate of the optimal bilinear observer residuals in the presence of noise, and they can be computed by

$$\tilde{E} = Y - \tilde{\Phi}_{H,p} V_p \tag{34}$$

In deriving Eq. (27), the bilinear interaction matrix theorem was invoked to make $\mathcal{S}_{H,p}(k)$ vanish in Eq. (20). In the absence of noise, as considered in Reference 8, the interaction matrices H' and H'' will attempt to place the eigenvalues of $\bar{A}_H$ and $\bar{N}_H$ to make $\mathcal{S}_{H,p}(k)$ decay as fast as possible under the given input history. For ideal bilinear systems, this results in placing the eigenvalues of $\bar{A}_H$ and $\bar{N}_H$ at the origin to obtain a decay of $\mathcal{S}_{H,p}(k)$ in n time steps (deadbeat) under any input history. Such a deadbeat decay corresponds to $\mathbb{E}\left[e^T(k)e(k)\right] = 0$ for all $k \geq p \geq n$. The presence of noise in Eq. (1) makes the latter equality impossible and the interaction matrices attempt to place the eigenvalues of $\bar{A}_H$ and $\bar{N}_H$ so that P is minimized, i.e. $H' = K'$ and $H'' = K''$. The bilinear interaction matrix theorem still guarantees that $\mathcal{S}_{H,p}(k)$ vanishes, possibly for a lower value of γ.

To better clarify the result, we briefly provide an alternative derivation. One could have started from Eq. (6) with the optimal gains $G' = K'$ and $G'' = K''$, and propagated Eq. (6a) forward in time like Eq. (15) to obtain

$$\hat{x}(k) = \mathcal{S}_{K,p}(k)\hat{x}(k-p) + T_{K,p} z_p(k) \tag{35}$$

where $\mathcal{S}_{K,p}(k)$ and $T_{K,p}$ are defined as in Eq. (21), except for $\bar{A}_G$ and $\bar{N}_G$ being repleaced by $\bar{A}_K$ and $\bar{N}_K$. Plugging Eq. (35) into Eq. (6b) and recalling the definition of observer residual, we get

$$y(k) = C\mathcal{S}_{K,p}(k)\hat{x}(k) + CT_{K,p} z_p(k) + \epsilon(k) \tag{36}$$

From the bilinear interaction matrix theorem, we know that there is a bound γ to the input magnitude for which $\mathcal{S}_{K,p}(k)$ converges to 0 for all k, allowing us to write

$$y(k) = \Phi_{K,p} z_p(k) + \epsilon(k) \tag{37}$$

Note that $z_p(k)$ contains past outputs and current and past inputs, hence Eq. (37) is a bilinear ARX model (with noise ϵ in the output). We can rewrite Eq. (36), in matrix form, to obtain Eq. (29). By the above orthogonality argument, the LS solution to Eq. (29) produces LS residuals which are indeed the residuals of the optimal bilinear observer.

OBSERVER IDENTIFICATION

The second main step of the proposed bilinear OKID method consists in the identification of the optimal observer. Once the residuals $\epsilon(k)$ of the optimal bilinear observer have been estimated, they can be used to construct the noise-free identification problem of Eq. (3). Equation (3) describes the dynamics of the optimal bilinear observer, it has no noise terms, its inputs $u(k)$, $\epsilon(k)$ and $\epsilon(k)u(k)$ are known for $k = p, p+1, ..., l-1$ as well as its output $\hat{y}(k)$. Among the matrices to be identified in Eq. (3), there are those of the system of Eq. (1). Identifying the observer of Eq. (3) solves then the problem addressed in this paper. In this step lies the essence of the proposed bilinear OKID method. The identification problem of Eq. (1), affected by (unmeasured) noise, is transformed into the identification problem of Eq. (3), with no noise term in it. Note that the identification of the optimal observer could also be performed in the form of Eq. (6), as discussed in Reference 18 for the linear case.

To identify the noise-free system of Eq. (3), one can use either of the approaches illustrated in Reference 8, i.e. the Equivalent Linear Model (ELM) or the Intersection Subspace (IS) approach. Both of them are based on the following Input-Output-to-State Relationship between the state $\hat{x}(k)$ of the bilinear observer in Eq. (3) and a superstate $z(k)$ made of its input-output data only

$$\hat{x}(k) = Tz(k) \tag{38}$$

where T is a constant matrix. Depending on the specific choice of IOSR or, equivalently, on the specific definition chosen for the superstate $z(k)$, several identification algorithms of ELM and IS type can be devised. In this paper we choose to use the IS method with causal and mixed-anticausal IOSRs, appropriately modified to take into account the extra additive input components $\epsilon(k)$ and $\epsilon(k)u(k)$ in the state equation. For clarity, the method is reviewed below together with the necessary modifications. The latter are not of conceptual nature, therefore their derivation is kept as short as possible. For more details, see Reference 8.

Intersection Subspace (IS) Method

Assume two IOSRs are available for the bilinear system of Eq. (3)

$$\hat{x}(k) = T_a z_a(k) \qquad \hat{x}(k) = T_b z_b(k) \tag{39}$$

and define the following matrices

$$\hat{X} = \begin{bmatrix} \hat{x}(k_i) & \hat{x}(k_i+1) & \dots & \hat{x}(k_f) \end{bmatrix} \tag{40a}$$
$$Z_a = \begin{bmatrix} z_a(k_i) & z_a(k_i+1) & \dots & z_a(k_f) \end{bmatrix} \tag{40b}$$
$$Z_b = \begin{bmatrix} z_b(k_i) & z_b(k_i+1) & \dots & z_b(k_f) \end{bmatrix} \tag{40c}$$

where k_i and k_f are the initial and final time steps for which Eq. (39) holds. Then we can write

$$\hat{X} = T_a Z_a \qquad \hat{X} = T_b Z_b \tag{41}$$

The row space of the observer state sequence $\hat{X}$ is a subspace of the row space of Z_a and also a subspace of the row space of Z_b. The row space of $\hat{X}$ must then lie in the intersection between the row spaces of the two superspaces Z_a and Z_b. The problem of reconstructing the state history $\{\hat{x}(k)\}$ is therefore reduced to finding the intersection of two vector spaces, which can be accomplished via two singular value decompositions (SVD) as follows.

The intersection of Z_a and Z_b is spanned by common row vectors h_i, which can be expressed as linear combinations of the rows of Z_a or of the rows of Z_b

$$h_i = a_i^T Z_a = b_i^T Z_b \tag{42}$$

where a_i and b_i are column vectors with the corresponding linear combination coefficients. Defining

$$Z_{a,b} = \begin{bmatrix} Z_a^T & Z_b^T \end{bmatrix} \qquad c_i = \begin{bmatrix} a_i \\ -b_i \end{bmatrix} \tag{43}$$

we can rewrite Eq. (42) as

$$Z_{a,b} c_i = 0 \tag{44}$$

which shows that the column vectors c_i lie in the null space of $Z_{a,b}$ and therefore can be conveniently found by SVD as the right singular vectors associated with the zero singular values of $Z_{a,b}$. Since all possible pair combinations of basis vectors of the null space of Z_a^T and of the null space of Z_b^T satisfy Eq. (44), the null space of $Z_{a,b}$ generally has dimension $m \geq n$ and a second SVD is necessary to get a basis for the intersection subspace (i.e. n linearly independent h_i vectors). Knowledge of c_i's allows one to compute the corresponding h_i's through Eq. (43) and either of equalities in Eq. (42). Stacking the m common row vectors h_i's in a matrix H, the n basis vectors $\hat{X}_i$'s can be found as the right singular vectors associated with the non-zero singular values of H. A basis for the actual bilinear state space is obtained, which means the state history of the bilinear system is now known.

Once the state history is reconstructed, the identification problem is dramatically simplified and can be solved by the classic least-squares method. From Eq. (3) we can write

$$\begin{bmatrix}\hat{x}(k_i+1) & ... & \hat{x}(k_f)\end{bmatrix} = \begin{bmatrix}A & B & N & K' & K''\end{bmatrix} \begin{bmatrix}\hat{x}(k_i) & ... & \hat{x}(k_f-1) \\ u(k_i) & ... & u(k_f-1) \\ \hat{x}(k_i)u(k_i) & ... & \hat{x}(k_f-1)u(k_f-1) \\ \epsilon(k_i) & ... & \epsilon(k_f-1) \\ \epsilon(k_i)u(k_i) & ... & \epsilon(k_f-1)u(k_f-1)\end{bmatrix} \tag{45a}$$

$$\begin{bmatrix}\hat{y}(k_i) & ... & \hat{y}(k_f)\end{bmatrix} = \begin{bmatrix}C & D\end{bmatrix} \begin{bmatrix}\hat{x}(k_i) & ... & \hat{x}(k_f) \\ u(k_i) & ... & u(k_f)\end{bmatrix} \tag{45b}$$

and recover A, B, N, K', K'' and C, D via pseudo-inversion (Moore-Penrose). Note that the reconstructed state is not necessarily in the original coordinate system, and so will be the identified bilinear observer matrices. As usual with state-space formulation, the change in coordinate system does not affect the identified model validity.

In the examples at the end of the paper, the superspaces Z_a and Z_b are constructed with the causal and mixed-anticausal IOSRs. briefly reviewed below together with the necessary modifications due to the extra additive input components $\epsilon(k)$ and $\epsilon(k)u(k)$ in the state equation of Eq. (3).

Input-Output-to-State Representations

Causal IOSRs

A causal IOSR is a representation in the form of Eq. (38) where the state depends on past input-output data only. In the following equations, the subscript c remarks the *causality* of the representation. By introducing two interaction matrices M_c' and M_c'' and adding and subtracting the terms $M_c'\hat{y}(k)$ and $M_c''\hat{y}(k)u(k)$, Eq. (3a) can be written as

$$\hat{x}(k+1) = \bar{A}_c\hat{x}(k) + \bar{N}_c\hat{x}(k)u(k) + \bar{B}_c v_c(k) \tag{46}$$

where

$$\bar{A}_c = A - M_c'C \qquad \bar{N}_c = N - M_c''C \tag{47a}$$

$$\bar{B}_c = \begin{bmatrix}B - M_c'D & M_c' & -M_c''D & M_c'' & K' & K''\end{bmatrix} \qquad v_c(k) = \begin{bmatrix}u(k) \\ \hat{y}(k) \\ u^2(k) \\ \hat{y}(k)u(k) \\ \epsilon(k) \\ \epsilon(k)u(k)\end{bmatrix} \tag{47b}$$

Propagating Eq. (46) forward in time by $p_c - 1$ steps, we obtain

$$\hat{x}(k+p_c) = \mathcal{S}_{c,p_c}\hat{x}(k) + T_{c,p_c}z_{c,p_c}(k+p_c) \tag{48}$$

where $\mathcal{S}_{c,p_c}$ is defined as in Eq. (21), except for $\bar{A}_M$ and $\bar{N}_M$ being replaced by $\bar{A}_c$ and $\bar{N}_c$, and the causal superstate $z_{c,p_c}(k)$, made of input-output data at steps $k-1, k-2, ..., k-p_c$ only, is defined like $z_p(k)$ before Eq. (22) except for $v(k)$ replaced by $v_c(k)$. By invoking the bilinear interaction matrix theorem and shifting the time index backward by p_c steps, Eq. (48) yields the causal IOSR

$$\hat{x}(k) = T_{c,p_c} z_{c,p_c}(k) \tag{49}$$

Mixed-Anticausal IOSRs

Rewrite Eq. (3) as

$$\hat{x}(k) = A^{-1}\hat{x}(k+1) - A^{-1}N\hat{x}(k)u(k) - A^{-1}Bu(k) - A^{-1}K'\epsilon(k) - A^{-1}K''\epsilon(k)u(k) \tag{50a}$$
$$\hat{y}(k) = C\hat{x}(k) + Du(k) \tag{50b}$$

and introduce the interaction matrices, M_a' and M_a'', where the subscript a stands for *anticausal*. Adding and subtracting the terms $M_a'\hat{y}(k)$ and $M_a''\hat{y}(k)u(k)$, Eq. (50a) can be written as

$$\hat{x}(k) = \bar{A}_a\hat{x}(k+1) + \bar{N}_a\hat{x}(k)u(k) + \bar{B}_a v_a(k+1) \tag{51}$$

where

$$\bar{A}_a = A^{-1} - M_a'C \qquad \bar{N}_a = -A^{-1}N - M_a''C \tag{52a}$$
$$\bar{B}_a = \begin{bmatrix} -A^{-1}B & -M_a'D & M_a' & -M_a''D & M_a'' & A^{-1}K' & A^{-1}K'' \end{bmatrix} \tag{52b}$$
$$v_a(k+1) = \begin{bmatrix} u(k) & u(k+1) & \hat{y}^T(k+1) & u^2(k) & \hat{y}^T(k)u(k) & \epsilon^T(k) & \epsilon^T(k)u(k) \end{bmatrix}^T \tag{52c}$$

Shifting the time index by 1, Eq. (51) becomes

$$\hat{x}(k+1) = \bar{A}_a\hat{x}(k+2) + \bar{N}_a\hat{x}(k+1)u(k+1) + \bar{B}_a v_a(k+2) \tag{53}$$

Work on the right-hand side of Eq. (51) by replacing $\hat{x}(k+1)$ with Eq. (53) and $\hat{x}(k)$ with Eq. (49) to get

$$\begin{aligned}\hat{x}(k) =& \bar{A}_a^2\hat{x}(k+2) + \bar{A}_a\bar{N}_a\hat{x}(k+1)u(k+1) \\ &+ \bar{A}_a\bar{B}_a v_a(k+2) + \bar{N}_a\bar{T}_{p_c,c} z_{p_c,c}(k)u(k) + \bar{B}_a v_a(k+1)\end{aligned} \tag{54}$$

Replace $\hat{x}(k+1)$ in Eq. (54) with Eq. (49) shifted forward by one time step, getting

$$\begin{aligned}\hat{x}(k) =& \bar{A}_a^2\hat{x}(k+2) + \bar{A}_a\bar{N}_a T_{p_c,c} z_{p_c,c}(k+1)u(k+1) \\ &+ \bar{A}_a\bar{B}_a v_a(k+2) + \bar{N}_a T_{p_c,c} z_{p_c,c}(k)u(k) + \bar{B}_a v_a(k+1)\end{aligned} \tag{55}$$

Observe that the only state-dependent term on the right-hand side of Eq. (55) is multiplied by $\bar{A}_a^2$. If $n = 2$, it is sufficient that A^{-1} and C form an observable pair to guarantee the existence of an interaction matrix M_a' such that $\bar{A}_a^2$ vanishes. Note that the interaction matrix M_a'' becomes unnecessary, therefore $\bar{N}_a = N_a$. If $n > 2$, the above derivation can be continued replacing $\hat{x}(k+2)$ with Eq. (53) shifted one step forward in time and taking care of the resulting terms as done above, getting an equation for $\hat{x}(k)$ with the only state-dependent term multiplied by $\bar{A}_a^3$, again guaranteed to be zero if (A^{-1}, C) is an observable pair. Proceeding in that way, we obtain the general mixed-anticausal IOSR

$$\hat{x}(k) = T_{ma,p_c,p_a} z_{ma,p_c,p_a}(k) \tag{56}$$

where the mixed-anticausal superstate is defined as

$$z_{ma,p_c,p_a}(k) = \begin{bmatrix} z_{c,p_c}(k+p_a-1)u(k+p_a-1) \\ \dots \\ z_{c,p_c}(k)u(k) \\ v_a(k+p_a) \\ \dots \\ v_a(k+1) \end{bmatrix} \tag{57}$$

The IOSR of Eq. (56) is *mixed* because it relates the state to both past and future input-output data. The advantage of such an IOSR is that it requires the bilinear interaction matrix theorem only for the existence of M_c' and M_c''. For the anticausal portion only M_a' is required to exist, which is guaranteed by the sole conditions of $\left(A^{-1}, C\right)$ being an observable pair and $p_a \geq n$. When the IS method is applied on arbitrary bilinear systems using the causal and mixed-anticausal IOSRs, the asymptotic approximation introduced by the bilinear interaction matrix theorem is due only to the finite value selected for p_c, instead of both p_c and p_a as when using the causal and anticausal IOSRs (see Reference 8).

EXAMPLES

Numerical examples are provided to demonstrate bilinear OKID and show more details about its implementation. Since the proposed method can be interpreted as the stochastic extension of the deterministic bilinear identification method described in Reference 8, the examples refer to the same bilinear systems utilized in the latter and highlight how the new method can achieve accurate identification even in cases where the direct application of the deterministic algorithms of Reference 8 fail because of the process and measurement noise affecting the data.

In each example, measured input-output data are simulated as follows. First we generate a random input sequence $\{u(k)\}$ of 10,000 samples (from a uniform distribution between -0.5 and 0.5) and two zero-mean gaussian sequences $\{w_p(k)\}$ and $\{w_m(k)\}$, respectively with covariance

$$Q = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \times 10^{-4} \qquad\qquad R = 10^{-4} \tag{58}$$

Said sequences are used to generate the measured output via Eq. (1).

Ideal Bilinear System

The following system is used as a prototype of ideal bilinear system (Reference 8)

$$A = \begin{bmatrix} 0 & 0.5 \\ 0.5 & -0.5 \end{bmatrix} \qquad B = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad N = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix} \qquad C = \begin{bmatrix} 0 & 1 \end{bmatrix} \qquad D = 0 \tag{59}$$

The first step of the bilinear OKID method is the estimation of the residuals of the optimal observer of Eq. (3). This can be done by LS on the set of equations arising from the bilinear ARX model. However notice how the ideal property of the system of Eq. (59) is irrelevant, since the poles of $\bar{A}_H$ and $\bar{N}_H$ are dictated by the statistics of the noise embedded in the input-output data. As a consequence, the noise structure determines how $\mathcal{S}_{L,p}$ converges to 0 with p. Figure 1 shows the error in the estimation of the observer residuals, for both approaches, over a portion of the dataset. As expected, for $p = 2$ the estimates are poor because $\mathcal{S}_{L,2}$ is not negligible. The estimates improve

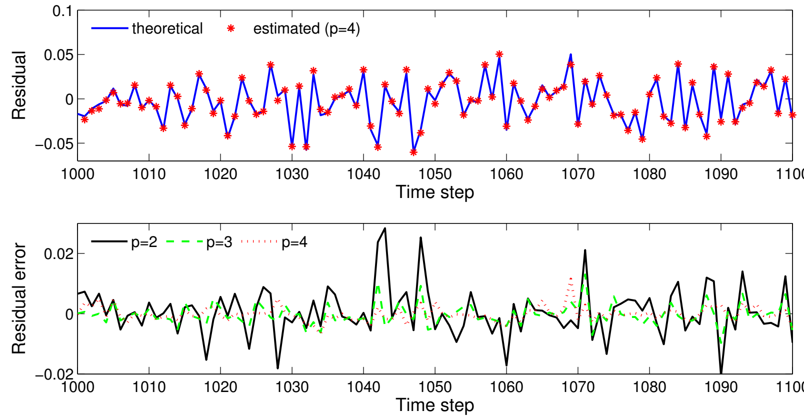
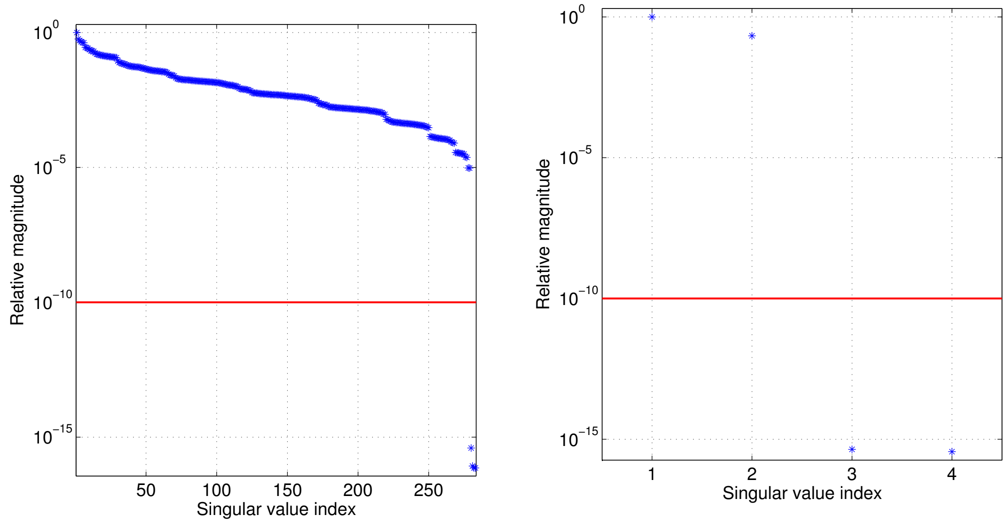


Figure 1: **Estimation of observer residuals for ideal system: comparison for different values of p.**

as p increases. As a last note on Figure 1, it is remarkable how it is possible to accurately estimate a white random process such as the observer residual sequence.

Having now estimated $\epsilon(k)$ for $k \geq p$, say with $p = 4$, we can proceed with the second step of bilinear OKID, i.e. the identification of the observer of Eq. (3), using the IS method with causal and mixed-anticausal IOSRs with $p_c = 4$ and $p_a = 2$. The choice of the values for p, p_c and p_a is consistent with the assumption that the true order of the system is unknown and believed to be $n \leq 4$. When intersecting the two IOSR superspaces to reconstruct the sequence of the observer state $\hat{x}$, two SVDs have to be performed and they reveal the order of the observer of Eq. (3), which is the same as the order of the system of Eq. (1). In order to reduce the SVD computational effort, the identification of the observer can be done on a reduced portion of the record. In this example, we select the last 3000 samples. Figure 2a shows the first SVD, to find vectors spanning the intersection subspace of the two superspaces. At this stage we need to split the zero and non-zero singular values, and the cut is indeed very clear. Figure 2b shows the second SVD, performed to find a basis for the intersection subspace. Here we need to retain the non-zero singular values, and again the difference with respect to the zero singular values is of 10 order of magnitudes, making the selection unquestionable. Since two singular values are retained, the order of the identified model will be $n_{id} = 2$, which is indeed correct. It is worth adding here that applying directly the IS method to the identification of the bilinear system (as prescribed by the deterministic identification method in Reference 8), without passing through the identification of the associated observer, makes the singular values of both SVDs decrease in a continuous fashion, making it impossible to identify the order of the system and reconstruct correctly the state sequence. In other words, the noise embedded in the data can heavily affect the outcome of the identification methods presented in Reference 8, where it is assumed the measured data are noise-free. The bilinear OKID method proposed in this paper represents a significant improvement over Reference 8 for any practical application, where noise cannot be eliminated.

The right singular vectors from the second SVD provide the observer state sequence that allows one to construct the LS problem of Eq. (45) and solve it for the observer matrices. The estimated matrices A, B, N, C, D yield the state-space model for the bilinear system. Its accuracy can be verified by driving both the true and identified models with the same input sequence (without noise),



(a) **Intersection of superspaces Z_a and Z_b (zero singular values are selected).**

(b) **Minimum basis for the intersection subspace (non-zero singular values are selected).**

Figure 2: **Observer identification for ideal system: SVDs of the IS method.**

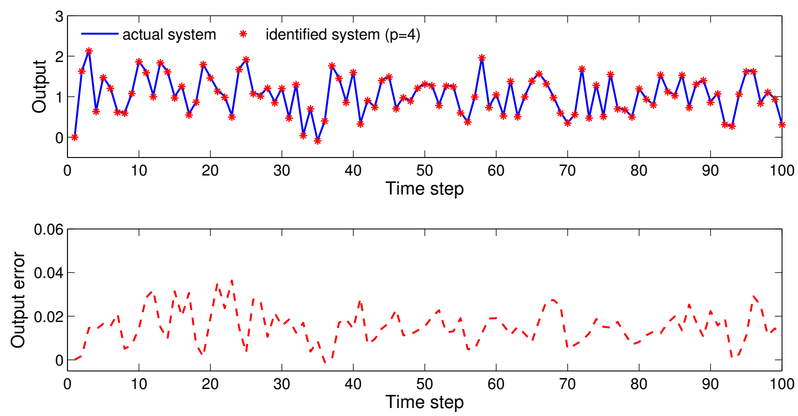


Figure 3: **Comparison of output from actual ideal system and identified models.**

generated independently from the one used for the identification, and comparing the corresponding output, as shown in Figure 3.

Arbitrary Bilinear System

Modifying matrix N in Eq. (59) to

$$N = \begin{bmatrix} 0.3 & 1 \\ -1 & 1 \end{bmatrix} \tag{60}$$

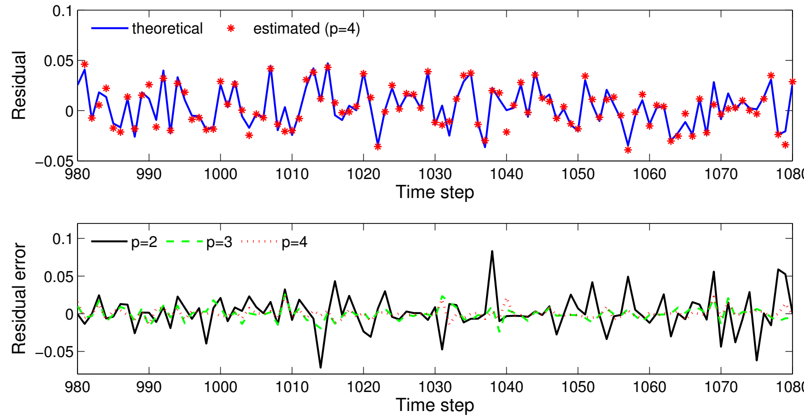


Figure 4: **Estimation of observer residuals for arbitrary system: comparison for different values of p.**

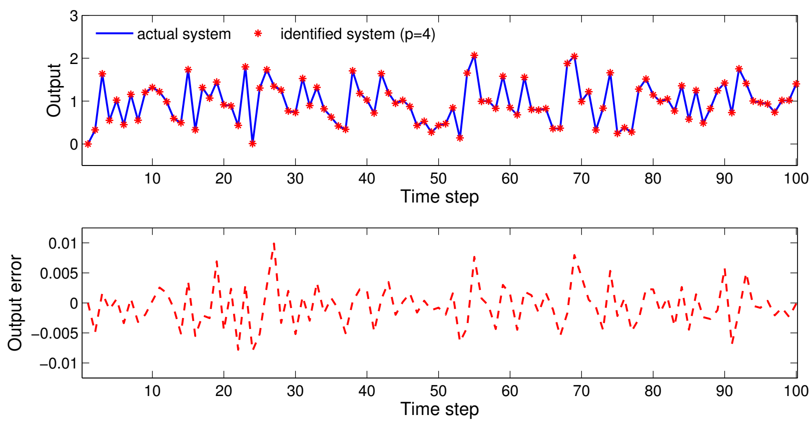


Figure 5: **Comparison of output from actual arbitrary system and identified models.**

is sufficient to lose the above mentioned ideal properties (Reference 8). The resulting system is therefore an example of arbitrary bilinear system. In the deterministic case addressed in Reference 8, the identification process is affected by whether the underlying system is ideal or arbitrary, the former leading to an exact identified model even for minimum values of p ($p = n$) and the latter providing comparable accuracy only for significantly larger values of p. In contrast, when noise is present in the input-output data, any identification method can only estimate an approximate model. Therefore, the difference between ideal and arbitrary bilinear system identification is less significant in the presence of noise. As shown in Figure 4, the estimates of the residuals improve as p is increased.

Using the residuals estimated with $p = 4$, the following bilinear model is identified (matrices rounded to the 4^{th} significant digit) via the IS method with causal and mixed-anticausal IOSRs ($p_c = 4, p_a = 2$)

$$A_{id} = \begin{bmatrix} -0.8525 & -0.1467 \\ 0.3453 & 0.3529 \end{bmatrix} \qquad B_{id} = \begin{bmatrix} -0.02807 \\ -0.04807 \end{bmatrix} \tag{61a}$$

$$N_{id} = \begin{bmatrix} 0.4323 & -0.6828 \\ 1.352 & 0.8648 \end{bmatrix} \qquad C_{id} = \begin{bmatrix} -34.50 & -21.40 \end{bmatrix} \qquad D_{id} = 6.805 \times 10^{-3} \tag{61b}$$

Again, the system order is obtained by singular value plots with a clear difference between zero and non-zero singular values, similar to the ones in Figures 2a and 2b. The accuracy of the identified model can be assessed by comparing the predicted output with the true output when both the identified model and the true system are driven by the same input sequence (without noise), as shown in Figure 5. Note that the LS solution to Eq. (45a) also yields

$$K_{1,id} = -0.01314 \qquad K_{2,id} = -0.008434 \tag{62}$$

Such gains can be used to construct a bilinear observer for the system of Eq. (61). In operation conditions where the input is white, the identification experiment can be performed on line and the proposed bilinear OKID method provides simultaneously both the system and observer mathematical models, which can then be used in similar operation conditions. The key advantage with respect to separate system identification and observer design lies with the difficulties in estimating the process noise covariance Q (necessary, together with R, to design K' and K'' as shown in Reference 9). The difficulties are completely overcome by bilinear OKID, which yields the optimal observer gains without requiring knowledge of Q or R.

CONCLUSIONS

This paper has presented the first extension to nonlinear systems of the well-established OKID approach for linear system identification, under the sole additional assumptions that the excitation input is stationary and white. The result is a new method for bilinear system identification in the presence of process and measurement noise. The method relies on a bilinear steady-state observer, which is proven to have properties similar to the well-known linear Kalman filter. The estimation of the residuals of such bilinear observer allows one to construct a new noise-free identification problem that can be solved via methods previously proposed by the authors. The resulting bilinear OKID method represents indeed a significant improvement over such methods, being able to accurately identify the bilinear system under consideration even in cases where the noise prevents the direct application of the previous methods from identifying even the correct order. In other words, the presented bilinear OKID method can be interpreted as the extension of the methods presented in Reference 8 to the case where the measured input-output data are corrupted by noise. The need for white excitation input parallels the result in Reference 21, which proposed a subspace method for the identification of bilinear systems in the presence of noise. OKID is indeed an alternative approach to subspace methods and has proven to be very successful in linear system identification, in particular in cases where one is interested in the system model as well as in the corresponding optimal observer. Therefore this paper represents a significant step towards the realization of the above mentioned bridge between linear and nonlinear systems in the areas of system identification and controls. Further research will aim to remove the requirements on the excitation input to be used in the identification process, in particular its whiteness, and to reduce the curse of dimensionality as the parameter p is increased. The latter problem has already been successfully addressed in the deterministic case (Reference 22).

REFERENCES

[1] C. Bruni, G. Di Pillo, and G. Koch, "Bilinear Systems: An Appealing Class of "Nearly Linear" Systems in Theory and Applications," *IEEE Transactions on Automatic Control*, Vol. AC-19, No. 4, 1974, pp. 334–348.

[2] J. T.-H. Lo, "Global Bilinearization of Systems with Control Appearing Linearly," *SIAM Journal on Control*, Vol. 13, No. 4, 1975, pp. 879–885.

[3] S. Svoronos, G. Stephanopoulos, and A. Rutherford, "Bilinear Approximation of General Non-Linear Dynamic Systems with Linear Inputs," *International Journal of Control*, Vol. 31, No. 1, 1980, pp. 109–126.

[4] J.-N. Juang, "Continuous-Time Bilinear System Identification," *Nonlinear Dynamcs*, Vol. 39, 2005, pp. 79–94.

[5] J.-N. Juang, "Generalized Bilinear System Identification," *The Journal of the Astronautical Sciences*, Vol. 57, No. 1/2, 2009.

[6] M. Q. Phan, Y. Shi, R. Betti, and R. W. Longman, "Discrete-Time Bilinear Representation of Continuous-Time Bilinear State-Space Models," *Advances in the Astronautical Sciences*, Vol. 143, 2012, pp. 571–589.

[7] M. Q. Phan and H. Celik, "A Superspace Method for Discrete-Time Bilinear Model Identification by Interaction Matrices," *Advances in the Astronautical Sciences*, Vol. 139, 2011, pp. 445–464.

[8] F. Vicario, M. Q. Phan, R. Betti, and R. W. Longman, "Linear State Representations for Identification of Bilinear Discrete-Time Models by Interaction Matrices," *Advances in the Astronautical Sciences*, Vol. 148, 2013, pp. 2039–2058.

[9] M. Q. Phan, F. Vicario, R. W. Longman, and R. Betti, "Observers for Bilinear State-Space Models by Interaction Matrices," *Proceedings of the 16th Yale Workshop on Adaptive and Learning Systems*, New Haven, CT, 2013.

[10] M. Q. Phan, J.-N. Juang, and R. W. Longman, "Identification of Linear Multivariable Systems by Identification of Observers with Assigned Real Eigenvalues," *Journal of the Astronautical Sciences*, Vol. 40, No. 2, 1992, pp. 261–27.

[11] M. Q. Phan, L. G. Horta, J.-N. Juang, and R. W. Longman, "Linear System Identification via an Asymptotically Stable Observer," *Journal of Optimization Theory and Applications*, Vol. 79, No. 1, 1993, pp. 59–86.

[12] J.-N. Juang, M. Q. Phan, L. G. Horta, and R. W. Longman, "Identification of Observer/Kalman Filter Markov Parameters: Theory and Experiments," *Journal of Guidance, Control, and Dynamics*, Vol. 16, No. 2, 1993, pp. 320–329.

[13] M. Q. Phan, L. G. Horta, J.-N. Juang, and R. W. Longman, "Improvement of Observer/Kalman Filter Identification (OKID) by Residual Whitening," *Journal of Vibrations and Acoustics*, Vol. 117, 1995, pp. 232–238.

[14] T. Clarke and X.-D. Sun, "Minimal State-Space Model Realisation of a Nonlinear Helicopter," *Control Theory and Applications, IEE Proceedings*, Vol. 145, No. 4, 1998, pp. 415–422.

[15] A. Tiano, R. Sutton, A. Lozowicki, and W. Naeem, "Observer Kalman Filter Identification of an Autonomous Underwater Vehicle," *Control Engineering Practice*, Vol. 15, 2007, pp. 727–739.

[16] J.-N. Juang and R. S. Pappa, "An Eigensystem Realization Algorithm for Modal Parameter Identification and Model Reduction," *Journal of Guidance, Control, and Dynamics*, Vol. 8, No. 5, 1985, pp. 620–627.

[17] J.-N. Juang, J. E. Cooper, and J. R. Wright, "An Eigensystem Realization Algorithm Using Data Correlations (ERA/DC) for Modal Parameter Identification," *Journal of Control Theory and Advanced Technology*, Vol. 4, No. 1, 1988, pp. 5–14.

[18] F. Vicario, M. Q. Phan, R. W. Longman, and R. Betti, "OKID as a Unified Approach to System Identification," *24th AAS/AIAA Space Flight Mechanics Meeting*, Santa Fe, NM, 2013, AAS14-450.

[19] M. Q. Phan, "Interaction Matrices in System Identification and Control," *Proceedings of the15th Yale Workshop on Adaptive and Learning Systems*, New Haven, CT, 2011.

[20] H. Celik and M. Q. Phan, "Identification of Input-Output Maps for Bilinear Discrete-Time State-Space Models," *Advances in the Astronautical Sciences*, Vol. 142, 2012, pp. 393–412.

[21] W. Favoreel, B. De Moor, and P. Van Overschee, "Subspace identification of bilinear systems subject to white inputs," *Automatic Control, IEEE Transactions on*, Vol. 44, No. 6, 1999, pp. 1157–1165.

[22] F. Vicario, M. Q. Phan, R. W. Longman, and R. Betti, "A Linear-Time-Varying Approach for Exact Identification of Bilinear Discrete-Time Systems by Interaction Matrices," *Advances in the Astronautical Sciences*, Vol. 150, 2014, pp. 1056–1076.