A New Subspace Identification Method for Bilinear Systems

Huixin Chen and **Jan Maciejowski** *

Department of Engineering
University of Cambridge
Cambridge CB2 1PZ U.K.
Tel. +44 1223 332732
Fax. +44 1223 332662
Email: hc240@eng.cam.ac.uk, jmm@eng.cam.ac.uk
4 October, 1999

Keywords:

System Identification, Bilinear System, Subspace Method, Kalman Filter, State Sequence Approach.

Abstract

In this paper, a subspace method for the identification of bilinear systems is developed. A two-block subspace method is developed in the deterministic system case and a three-block one is set up for the combined deterministic-stochastic system. An extended non-steady state bilinear Kalman filter is also derived. The input signal to the system does not have to be white, which is a major advantage over an existing subspace method for bilinear systems. Simulation results also show that the new algorithm converges much more rapidly (with sample size) than the existing method, and hence is more effective with small sample sizes. The faster convergence is presumably due to the insensitivity of the algorithm to the sample-spectrum of the input signal. These advantages are achieved by a different arrangement of the input-output equations into 'blocks', and projections onto different spaces than the ones used in the existing method. A further advantage of our algorithm is that the dimensions of the matrices involved are significantly smaller, so that the computational complexity is lower.

*Corresponding author

1 Introduction

Bilinear systems are attractive models for many dynamical processes, because they allow a significantly larger class of behaviours than linear systems, yet retain a rich theory which is closely related to the familiar theory of linear systems [14, 8]. They exhibit phenomena encountered in many engineering systems, such as amplitude-dependent time constants. Many practical system models are bilinear, and more general nonlinear systems can often be well approximated by bilinear models [16].

Most studies of the identification problem of bilinear systems have assumed an input-output formulation. Standard methods such as recursive least squares, extended least squares, recursive auxiliary variable and recursive prediction error algorithms, have been applied to identifying bilinear systems. Simulation studies have been undertaken [13], and some statistical results (strong consistency and parameter estimate convergence rates) are also available [7].

In this paper, we consider the identification of MIMO bilinear systems in state-space form. There are many advantages of using state-space models, particularly in the multivariable case [3]. In recent years 'subspace' methods have been developed which have proved to be extremely effective for the identification of linear systems [4, 17, 18, 20]. In [9, 10, 12] extension of such methods were given for bilinear systems, but the algorithm presented there was effective only if the measured input signal to the system being identified is white. To our knowledge this was the first extension of the subspace approach to bilinear systems. In [11] another subspace algorithm for bilinear systems was presented by the same authors, which apparently does not require a white input signal. However this algorithm is known to give biased results, and it must therefore be questioned whether it can really be considered to be an effective algorithm for the case of non-white inputs.

In this paper an alternative subspace algorithm for identifying bilinear systems is proposed. It does not require the measured input to be white, and the matrices which need to be constructed and operated upon are much smaller than those which appear in [11, 12]. Simulations show that it works well when the input signal is not white; they also show that if the input signal is white, then good results are obtained with much smaller sample sizes than are required for the algorithm of [11, 12]. The theoretical and simulation results are shown in [1] and [2] for deterministic system and combined deterministic-stochastic system respectively.

The paper is organised as follows. Section 2 introduces a considerable, but apparently unavoidable, amount of notation. Section 3 develops new subspce system identification methodologies for bilinear systems, which consists of a 'two-block' one for the deterministic bilinear system and a 'three-block' one for the combined deterministic-stochastic system. In the new algorithms, 'two-block' and 'three-block' form of the data equations was established and some relations between the spaces spanned by the rows of the data matrices which appear in these equations. Section 4, contains a extended blinear Kalman filter which is established under the condition that the system input does not have to be white noise sequence. Section 5 contains some examples which demonstrate that the algorithm does indeed work with non-white input signals, and that it requires much smaller sample sizes to obtain comparable results, than the algorithm presented in [9, 11], when the input *is* white.

All proofs are included in the appendix. Some references to earlier (non-subspace) work on identification of bilinear systems are also given.

2 Notations

The use of much specialised notation seems to be unavoidable in the current context. Mostly we follow the notation used in [11], but we introduce all the notation here for completeness.

We use $\otimes$ to denote the Kronecker product and $\odot$ the Khatri-Rao product of two matrices with $F \in \mathbf{R}^{t\times p}$ and $G \in \mathbf{R}^{u\times p}$ defined in [15, 19]:

$$F \odot G \triangleq [f_1 \otimes g_1, f_2 \otimes g_2, \ldots, f_p \otimes g_p]$$

$+, \oplus, \cap$ and $\setminus$ denote the sum, the direct sum, the intersection and set minus of two vector spaces, $\cdot^{\perp}$ denotes the orthogonal complement of a subspace with respect to the predefined ambient space, the Moore-Penrose inverse is written as $\cdot^{\dagger}$, and the transpose as $\cdot^{T}$.

In this paper we consider combined deterministic-stochastic time-invariant bilinear system of the form:

$$\begin{aligned} x_{t+1} &= Ax_t + Nu_t \otimes x_t + Bu_t + w_t \\ y_t &= Cx_t + Du_t + v_t \end{aligned} \tag{1}$$

where $x_t \in \mathbf{R}^n, y_t \in \mathbf{R}^l, u_t \in \mathbf{R}^m$, and $N = [N_1\ N_2\ ...N_m] \in \mathbf{R}^{n\times nm}$, $N_i \in \mathbf{R}^{n\times n}\ \ (i = 1, ..., m)$.

The input u_t is assumed to be indepedent of the measurement noise v_t and the process noise w_t. The covariance matrix of w_t and v_t is

$$\mathbf{E}\left[\begin{pmatrix} w_p \\ v_p \end{pmatrix}\begin{pmatrix} w_q \\ v_q \end{pmatrix}^T\right] = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix}\delta_{pq} \geq 0$$

We assume that the sample size is $\tilde{N}$, namely that input-output data $\{u(t), y(t) : t = 0, 1, \ldots, \tilde{N}\}$ are available. For arbitrary t we define

$$X_t \triangleq [x_t\ x_{t+1}\ ...\ x_{t+j-1}] \in \mathbf{R}^{n\times j}$$

but for the special cases $t = 0$ and $t = k$ we define, with some abuse of notation,

$$\begin{aligned} X_p &\triangleq [x_0\ x_1\ ...\ x_{j-1}] \in \mathbf{R}^{n\times j} \\ X_f &\triangleq [x_k\ x_{k+1}\ ...\ x_{k+j-1}] \in \mathbf{R}^{n\times j} \\ X_r &\triangleq [x_{2k}\ x_{2k+1}\ ...\ x_{2k+j-1}] \in \mathbf{R}^{n\times j} \end{aligned}$$

where k is the *row block size*. The suffices p, f and r are supposed to be mnemonic, representing 'past', 'future' and 'remote future' respectively.

We define U_t, U_p, U_f, U_r, Y_t, Y_p, Y_f, Y_r, W_t, U_p, W_f, W_r, V_t, V_p, V_f, V_r, similarly:

$$\begin{aligned}
U_i &\triangleq [u_i \; u_{i+1} \; ... \; u_{i+j-1}] \in \mathbf{R}^{m\times j} \\
U_p &\triangleq [u_0 \; u_1 \; ... \; u_{j-1}] \in \mathbf{R}^{m\times j} \\
U_f &\triangleq [u_k \; u_{k+1} \; ... \; u_{k+j-1}] \in \mathbf{R}^{m\times j} \\
U_r &\triangleq [u_{2k} \; u_{2k+1} \; ... \; u_{2k+j-1}] \in \mathbf{R}^{m\times j} \\
Y_i &\triangleq [y_i \; y_{i+1} \; ... \; y_{i+j-1}] \in \mathbf{R}^{l\times j} \\
Y_p &\triangleq [y_0 \; y_1 \; ... \; y_{j-1}] \in \mathbf{R}^{l\times j} \\
Y_f &\triangleq [y_k \; y_{k+1} \; ... \; y_{k+j-1}] \in \mathbf{R}^{l\times j} \\
Y_r &\triangleq [y_{2k} \; y_{2k+1} \; ... \; y_{2k+j-1}] \in \mathbf{R}^{l\times j} \\
W_i &\triangleq [w_i \; w_{i+1} \; ... \; w_{i+j-1}] \in \mathbf{R}^{n\times j} \\
W_p &\triangleq [w_0 \; w_1 \; ... \; w_{j-1}] \in \mathbf{R}^{n\times j} \\
W_f &\triangleq [w_k \; w_{k+1} \; ... \; w_{k+j-1}] \in \mathbf{R}^{n\times j} \\
W_r &\triangleq [w_{2k} \; w_{2k+1} \; ... \; w_{2k+j-1}] \in \mathbf{R}^{n\times j} \\
V_i &\triangleq [v_i \; v_{i+1} \; ... \; v_{i+j-1}] \in \mathbf{R}^{l\times j} \\
V_p &\triangleq [v_0 \; v_1 \; ... \; v_{j-1}] \in \mathbf{R}^{l\times j} \\
V_f &\triangleq [v_k \; v_{k+1} \; ... \; v_{k+j-1}] \in \mathbf{R}^{l\times j} \\
V_r &\triangleq [v_{2k} \; v_{2k+1} \; ... \; v_{2k+j-1}] \in \mathbf{R}^{l\times j}
\end{aligned}$$

These matrices will later be used to construct larger matrices with a 'generalised block-Hankel' structure. In order to use all the available data in these, the number of columns j is such that $\tilde{N} = 3k + j - 1$ for the three-block combined deterministic-stochastic case and $\tilde{N} = 2k + j - 1$ for the two-block deterministic case.

For arbitrary q and $i \geq q + 2$, we define

$$\begin{aligned}
X_{q|q} &\triangleq \begin{pmatrix} X_q \\ U_q \odot X_q \end{pmatrix} \in \mathbf{R}^{(m+1)n\times j} \\
X_{i-1|q} &\triangleq \begin{pmatrix} X_{i-2|q} \\ U_{i-1} \odot X_{i-2|q} \end{pmatrix} \in \mathbf{R}^{(m+1)^{i-q} n\times j} \\
Y_{q|q} &\triangleq Y_q \\
Y_{i-1|q} &\triangleq \begin{pmatrix} Y_{i-1} \\ Y_{i-2|q} \\ U_{i-1} \odot Y_{i-2|q} \end{pmatrix} \in \mathbf{R}^{d_{i-q}\times j}
\end{aligned}$$

$$
\begin{aligned}
U_{q|q} &\triangleq U_q \\
U_{i-1|q} &\triangleq \begin{pmatrix} U_{i-1} \\ U_{i-2|q} \\ U_{i-1} \odot U_{i-2|q} \end{pmatrix} \in \mathbf{R}^{e_{i-q} \times j} \\
W_{q|q} &\triangleq W_q \\
W_{i-1|q} &\triangleq \begin{pmatrix} W_{i-1} \\ W_{i-2|q} \\ U_{i-1} \odot W_{i-2|q} \end{pmatrix} \in \mathbf{R}^{f_{i-q} \times j} \\
V_{q|q} &\triangleq V_q \\
V_{i-1|q} &\triangleq \begin{pmatrix} V_{i-1} \\ V_{i-2|q} \\ U_{i-1} \odot V_{i-2|q} \end{pmatrix} \in \mathbf{R}^{d_{i-q} \times j} \\
U^{+}_{q|q} &\triangleq U_q \\
U^{+}_{i-1|q} &\triangleq \begin{pmatrix} U^{+}_{i-2|q} \\ U_{i-1} \\ U_{i-1} \odot U^{+}_{i-2|q} \end{pmatrix} \in \mathbf{R}^{((m+1)^{i-q}-1) \times j}
\end{aligned}
$$

$$
\begin{aligned}
U^{++}_{q|q} &\triangleq \begin{pmatrix} U_q(1,:) \odot U_q \\ U_q(2,:) \odot U_q(2:m,:) \\ U_q(3,:) \odot U_q(3:m,:) \\ \vdots \\ U_q(m,:) \odot U_q(m,:) \end{pmatrix} \in \mathbf{R}^{\frac{m(m+1)}{2} \times j} \\
U^{++}_{i-1|q} &\triangleq \begin{pmatrix} U^{++}_{i-2|q} \\ U_{i-1} \odot U^{++}_{i-2|q} \end{pmatrix} \in \mathbf{R}^{\frac{m}{2}(m+1)^{i-q} \times j} \\
U^{y}_{i-1|q} &\triangleq U^{+}_{i-1|q} \odot Y_q \\
U^{u,y}_{i-1|q} &\triangleq \begin{pmatrix} U_{i-1|q} \\ U^{++}_{i-1|q} \\ U^{y}_{i-1|q} \end{pmatrix}
\end{aligned}
$$

where $U_q(i,:)$ and $U_q(i:m,:)$ denote the submatrix of U_q, which contains ith row and from the ith row to the mth row of the matrix U_q respectively, $i = 1, 2, ..., m$.

Remark 1. The meaning of $U^{+}_{i-1|q}$ is different from that in [9]. $U^{++}_{i-1|q}$ is newly introduced here.

$$
\begin{aligned}
X^p &\triangleq X_{k-1|0} \in \mathbf{R}^{(m+1)^k n \times j} \\
X^f &\triangleq X_{2k-1|k} \in \mathbf{R}^{(m+1)^k n \times j} \\
X^r &\triangleq X_{3k-1|2k} \in \mathbf{R}^{(m+1)^k n \times j}
\end{aligned}
$$

$$\begin{aligned}
U^p &\triangleq U_{k-1|0} \in \mathbf{R}^{e_k \times j} \\
U^f &\triangleq U_{2k-1|k} \in \mathbf{R}^{e_k \times j} \\
U^r &\triangleq U_{3k-1|2k} \in \mathbf{R}^{e_k \times j} \\
Y^p &\triangleq Y_{k-1|0} \in \mathbf{R}^{d_k \times j} \\
Y^f &\triangleq Y_{2k-1|k} \in \mathbf{R}^{d_k \times j} \\
Y^r &\triangleq Y_{3k-1|2k} \in \mathbf{R}^{d_k \times j} \\
W^p &\triangleq W_{k-1|0} \in \mathbf{R}^{(m+1)^k n \times j} \\
W^f &\triangleq W_{2k-1|k} \in \mathbf{R}^{(m+1)^k n \times j} \\
W^r &\triangleq W_{3k-1|2k} \in \mathbf{R}^{(m+1)^k n \times j} \\
V^p &\triangleq V_{k-1|0} \in \mathbf{R}^{d_k \times j} \\
V^f &\triangleq V_{2k-1|k} \in \mathbf{R}^{d_k \times j} \\
V^r &\triangleq V_{3k-1|2k} \in \mathbf{R}^{d_k \times j} \\
U^{+p} &\triangleq U^{+}_{k-1|0} \in \mathbf{R}^{[(m+1)^k - 1] \times j} \\
U^{+f} &\triangleq U^{+}_{2k-1|k} \in \mathbf{R}^{[(m+1)^k - 1] \times j} \\
U^{+r} &\triangleq U^{+}_{3k-1|2k} \in \mathbf{R}^{[(m+1)^k - 1] \times j} \\
U^{++p} &\triangleq U^{++}_{k-1|0} \in \mathbf{R}^{\frac{m}{2}(m+1)^k \times j} \\
U^{++f} &\triangleq U^{++}_{2k-1|k} \in \mathbf{R}^{\frac{m}{2}(m+1)^k \times j} \\
U^{++r} &\triangleq U^{++}_{3k-1|2k} \in \mathbf{R}^{\frac{m}{2}(m+1)^k \times j}
\end{aligned}$$

$$\begin{aligned}
U^{p,y} &\triangleq U^{+p} \odot Y_p \in \mathbf{R}^{l[(m+1)^k - 1] \times j} \\
U^{f,y} &\triangleq U^{+f} \odot Y_f \in \mathbf{R}^{l[(m+1)^k - 1] \times j} \\
U^{r,y} &\triangleq U^{+r} \odot Y_r \in \mathbf{R}^{l[(m+1)^k - 1] \times j} \\
U^{p,u,y} &\triangleq \begin{pmatrix} U^p \\ U^{++p} \\ U^{p,y} \end{pmatrix} \in \mathbf{R}^{f_k \times j} \\
U^{f,u,y} &\triangleq \begin{pmatrix} U^f \\ U^{++f} \\ U^{f,y} \end{pmatrix} \in \mathbf{R}^{f_k \times j} \\
U^{r,u,y} &\triangleq \begin{pmatrix} U^r \\ U^{++r} \\ U^{r,y} \end{pmatrix} \in \mathbf{R}^{f_k \times j}
\end{aligned}$$

where $d_i = \Sigma_{p=1}^{i}(m+1)^{p-1}l$, $e_i = \Sigma_{p=1}^{i}(m+1)^{p-1}m$ and $f_k = e_k + \frac{m}{2}(m+1)^k + l[(m+1)^k - 1]$. We denote by $\mathcal{U}_p$ the space spanned by all the rows of the matrix U_p. That is,

$$\mathcal{U}_p := \text{span}\{\alpha^* U_p, \quad \alpha \in \mathbf{R}^{km}\}$$

$\mathcal{U}_f$, $\mathcal{Y}_p$, $\mathcal{Y}_f$, $\mathcal{U}^p$, $\mathcal{Y}^p$, $\mathcal{U}^f$, $\mathcal{Y}^f$, $\mathcal{U}^{p,u,y}$, $\mathcal{U}^{f,u,y}$ *etc* are defined similarly. Finally, $\mathcal{U}^{+p} \odot \mathcal{X}_p$, $\mathcal{U}^{+f} \odot \mathcal{X}_f$ and $\mathcal{U}^{+r} \odot \mathcal{X}_r$ can be defined as the space spanned by all the rows of the matrix $U^{+p} \odot X_p, U^{+f} \odot X_f$ and $U^{+r} \odot X_r$ and so on.

3 New subspace identification algorithms

In this section, first some system description and analysis is given as a priminary knowledge for the latter two subsections. Secondly, a 'two-block' algorithm for deterministic bilinear system is presented. In the last subsection, a 'three-block' algorithms for combined deterministic-stochastic bilinear system is proposed.

3.1 System description and analysis

Every bilinear system of the form (1) can be considered as the sum of two subsystems, one only containing deterministic variables (index d) and the other containing all stochastic variables (index s). The state and output of (1) are then simply the sum of states and the outputs of both subsystems respectively:

$$\begin{aligned} x_t &= x_t^d + x_t^s \\ y_t &= y_t^d + y_t^s \end{aligned}$$

where the two subsystems are decribed as follows respectively.

$$\begin{aligned} x_{t+1}^d &= Ax_t^d + Nu_t \otimes x_t^d + Bu_t \\ y_t^d &= Cx_t^d + Du_t \end{aligned} \tag{2}$$

and

$$\begin{aligned} x_{t+1}^s &= Ax_t^s + Nu_t \otimes x_t^s + w_t \\ y_t^s &= Cx_t^s + v_t \end{aligned} \tag{3}$$

with the state covariance matrix ubder the assumption that alll the relevant processes are second order statistical stationary processes.

$$\begin{aligned} \Sigma^s &\triangleq \mathbf{E}[x_k^s (x_k^s)^T] \\ G^s &\triangleq \mathbf{E}[x_{k+1}^s (y_k^s)^T] \\ \Lambda_0^s &\triangleq \mathbf{E}[y_k^s (y_k^s)^T] \end{aligned} \tag{4}$$

The three lemmas are given in this subsection as the priminary knowledge for the future uses in this section. For simplicity, some of the lemmas are given in the form of three block and they are hold for two block form obviously.

Lemma 1 *For F, G, H, J of compatible dimensions, $F \in \mathbf{R}^{k\times l}$, $G \in \mathbf{R}^{l\times m}$, $H \in \mathbf{R}^{p\times l}$, $J \in \mathbf{R}^{l\times m}$:*

$$(FG \otimes HJ) = (F \otimes H)(G \otimes J) \tag{5}$$

$$(FG \odot HJ) = (F \otimes H)(G \odot J) \tag{6}$$

Proof: see appendix.

Lemma 1 states some properties of bilinear algebra operator, which will be used in our paper later.

Lemma 2 *For any integer $i \geq 1$ and q*

$$\mathcal{U}^{+}_{i+q|q} \odot \mathcal{U}_q \setminus \mathcal{U}^{++}_{i+q|q} = \mathcal{U}^{+}_{i+q|q+1} \odot \mathcal{U}_q$$

Lemma 2 states the row space relationship between the $\mathcal{U}^{+}_{i+q|q} \odot \mathcal{U}_q$ and $\mathcal{U}^{++}_{i+q|q}$, which will be used in our paper later.

Lemma 3

$$X^p = \begin{pmatrix} X_p \\ U^{+p} \odot X_p \end{pmatrix} \tag{7}$$

$$X^f = \begin{pmatrix} X_f \\ U^{+f} \odot X_f \end{pmatrix} \tag{8}$$

$$X^r = \begin{pmatrix} X_r \\ U^{+r} \odot X_r \end{pmatrix} \tag{9}$$

Proof: see appendix.

Lemma 3 states that $\mathcal{X}^p, \mathcal{X}^f$ and $\mathcal{X}^r$ can be represented as $\mathcal{X}_p + \mathcal{U}^{+p} \odot \mathcal{X}_p, \mathcal{X}_f + \mathcal{U}^{+f} \odot \mathcal{X}_f$ and $\mathcal{X}_r + \mathcal{U}^{+r} \odot \mathcal{X}_r$ respectively.

Lemma 4 *Let $F \in \mathbf{R}^{m\times n}$, and F^- is a generlized inverse matrix of F. Let G be an arbitrary generlized matrix of F, G can be represented as the following form:*

$$G = F^- + H - F^- FHFF^-$$

where $H \in \mathbf{R}^{n\times m}$ is an arbitrary matrix

Proof: see [19]

Lemma 5 *From (1) we have, modulo a state coordinate transformation,*

$$X_p = C^\dagger (Y_p - DU_p - V_p) \tag{10}$$

$$X_f = C^\dagger (Y_f - DU_f - V_f) \tag{11}$$

$$X_r = C^\dagger (Y_r - DU_r - V_r) \tag{12}$$

Proof: see appendix.

Lemma 5 shows that $\mathcal{X}_p, \mathcal{X}_f$ and $\mathcal{X}_r$ are contained in the space of $\mathcal{Y}_p + \mathcal{U}_p + \mathcal{V}_p, \mathcal{Y}_f + \mathcal{U}_f + \mathcal{V}_f$ and $\mathcal{Y}_r + \mathcal{U}_r + \mathcal{V}_r$ respectively.

Remark 2 In the case of $l = n$, the presentations of X_p, X_f and X_r are unique if the system is observable. If $l < n$, this holds for any right inverse of C. Different choices of right inverse correspond to different choices of state coordinates. Note that the spaces $\mathcal{X}_p$, $\mathcal{X}_f$ and $\mathcal{X}_r$ do not depend on this choice from **Lemma 5**. We will assume that the Moore-Penrose pseudo-inverse of C is used.

3.2 Deterministic system identification

In this subsection, a so-called two-block subspace method for the identification of deterministic bilinear systems is developed. In [9], an input-output equation relevant to the output, input and state block data matrix is set up in the case of deterministic case as follows:

Lemma 6 (Input-Output Equation) *The system (2) can be written in the following 'input-output equation' form:*

$$Y^p_{k-1|0} = \overline{\Gamma}_k X^p_{k-1|0} + H_k U^p_{k-1|0} \tag{13}$$
$$X_f = \overline{A}_k X^p_{k-1|0} + \triangle^U_k U^p_{k-1|0} \tag{14}$$

where

$$\overline{\Gamma}_i \triangleq \begin{pmatrix} C\overline{A}_{i-1} & 0 & \dots & 0 \\ \overline{\Gamma}_{i-1} & 0 & \dots & 0 \\ 0 & \overline{\Gamma}_{i-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \overline{\Gamma}_{i-1} \end{pmatrix} \in \mathbf{R}^{d_i \times (m+1)^i n} \quad (i > 1)$$

$$\overline{\Gamma}_1 \triangleq (C \quad 0_{l\times(m+1)n})$$

$$H_i \triangleq \begin{pmatrix} D & C\triangle^U_{i-1} & 0 & \dots & 0 \\ 0 & H_{i-1} & 0 & \dots & 0 \\ 0 & 0 & H_{i-1} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & H_{i-1} \end{pmatrix} \in \mathbf{R}^{d_i \times e_i} \quad (i > 1)$$

$$H_1 \triangleq D,$$

where

$$\begin{aligned} \triangle_n^U &\triangleq \begin{bmatrix} B & A\triangle_{n-1}^U & N_1\triangle_{n-1}^U \dots, & N_m\triangle_{n-1}^U \end{bmatrix} \\ \triangle_1^U &\triangleq B \\ \overline{A}_i &\triangleq (A\overline{A}_{i-1} \quad N_1\overline{A}_{i-1} \quad \dots \quad N_m\overline{A}_{i-1}), \\ \overline{A}_0 &\triangleq I_{n\times n} \end{aligned}$$

Proof: see [9].

The data equation and state equation linking the past and future data equation can be written in the following way:

Theorem 1 (Input-Output Equation in Two Block Form) *The system (2) can be written in the following 'two block' form:*

$$Y^p = \mathcal{C}_k X_p + \mathcal{D}_k U^{p,u,y} \tag{15}$$

$$Y^f = \mathcal{C}_k X_f + \mathcal{D}_k U^{f,u,y} \tag{16}$$

$$X_f = \mathcal{A}_k X_p + \mathcal{B}_k U^{p,u,y} \tag{17}$$

Proof: see appendix.

¿From **Theorem 1**, we deduct that all the block data matrix have a linear relationship with the state block matrix X_p and X_f. Also X_f is linear to X_p. This is achieved by putting all the bilinear factors of the system into the newly created data matrices $U^{p,u,y}$ and $U^{f,u,y}$. In such a way, the bilinear system can be transfered into an similar to linear system representation and the restriction that system input has to be white can be removed.

It is well known that the quality of a model obtained from an identification experiment depends highly on the degree of excitation of the input signal. 'Informative experiments' for linear system are standard, for example. For bilinear system, we need some alternative conditions including the input, output and their Khatri-Rao product.

Theorem 2 *If*

$$\begin{pmatrix} Y^p \\ U^{p,u,y} \\ U^{f,u,y} \end{pmatrix} \tag{18}$$

is a full row rank matrix, then

$$\begin{aligned} \mathcal{X}_f &\subset \mathcal{Y}^p + \mathcal{U}^{p,u,y} \\ \mathcal{X}_f + \mathcal{U}^{f,u,y} &= \mathcal{X}_f \oplus \mathcal{U}^{f,u,y} \\ &\subset (\mathcal{Y}^p + \mathcal{U}^{p,u,y}) \oplus \mathcal{U}^{f,u,y} \end{aligned} \tag{19}$$

Proof: see appendix.

Theorem 2. states that the space $\mathcal{X}_f$ are contained by $\mathcal{Y}^p + \mathcal{U}^{p,u,y}$, has empty intersection with $\mathcal{U}^{f,u,y}$. Therefore $\mathcal{Y}^f$ in equation (16) has a unique decomposition into $\mathcal{C}_k\mathcal{X}_f$ and $\mathcal{D}_k U^{f,u,y}$. Consequently, $\mathcal{D}_k$ can be determined by projecting both sides of equation (16) on to a suitable subspace. Once $\mathcal{D}_k$ has been determined, it is possible to estimate state sequences $\hat{X}_k$ and $\hat{X}_{k-1}$, and hence estimate the system parameters.

The main procedures of the two-block subspace identification algorithm are given as follows:

1. Decompose Y^f into $\mathcal{C}_k X_f$ and $\mathcal{D}_k U^{f,u,y}$ using orthogonal projection: from (16) and (19) it follows that

$$\Pi_{\Omega^\perp} Y^f = \mathcal{D}_k \Pi_{\Omega^\perp} U^{f,u,y} \tag{20}$$

where $\Omega = \mathcal{Y}^p + \mathcal{U}^{p,u,y}$. Determine $\mathcal{D}_k \in \mathbf{R}^{d_k \times (e_k + \frac{m}{2}(m+1)^k + l[(m+1)^k - 1])}$ from

$$\mathcal{D}_k = \left(\Pi_{\Omega^\perp} Y^f\right)\left(\Pi_{\Omega^\perp} U^{f,u,y}\right)^\dagger \tag{21}$$

2. Obtain the SVD decomposition and partition, retaining singular values and selecting a model order.

$$\begin{aligned} &\left[Y_{2k-2|k-1} \quad Y_{2k-1|k}\right] - \mathcal{D}_k \left[U^{u,y}_{2k-2|k-1} \quad U^{u,y}_{2k-1|k}\right] \\ &\qquad =: \Gamma\Sigma\Omega^T = \begin{bmatrix} \Gamma_1 & \Gamma_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} \Omega_1^T \\ \Omega_2^T \end{bmatrix} \end{aligned} \tag{22}$$

Since we expect

$$\Gamma\Sigma\Omega^T = \Gamma_1\Sigma_1\Omega_1^T = \mathcal{C}_k \left[X_{k-1} \quad X_k\right] \tag{23}$$

from (15) and (16) ($rank(\Sigma_1) = n$ and $rank(\Sigma_2) = 0$), form the estimates $\hat{\mathcal{C}}_k = \Gamma_1\Sigma_1^{1/2}$ and $[\hat{X}_{k-1} \quad \hat{X}_k] = \Sigma_1^{1/2}\Omega_1^T$, retaining only significant singular values in Σ_1. ($\hat{\mathcal{C}}_k$ is not needed later.)

3. Estimate the parameters A, B, C, D, N by solving

$$\begin{bmatrix} \hat{X}_k \\ Y_{k-1} \end{bmatrix} = \begin{bmatrix} A & N & B \\ C & 0 & D \end{bmatrix} \begin{bmatrix} \hat{X}_{k-1} \\ U_{k-1} \odot \hat{X}_{k-1} \\ U_{k-1} \end{bmatrix} \tag{24}$$

in a least-squares sense.

3.3 Combined deterministic-stochastic system identification

In this section, a 'three-block' subspace method for the identification of stochastic bilinear systems is developed. As shown in the previous section, a two-block configuration, splitting the data into 'past' ($\cdot^p$) and 'future' ($\cdot^f$) blocks, is adequate to identify the parameters in the deterministic case. However, when it comes to the combined deterministic-stochastic case, a two-block configuration cannot determine the parameter matrices of a system since there exists some stochastical noises effects. In the remaining part of this paper, we proposed a three-block configuration to estimate system parameter matrices in stochastical environment.

Lemma 7 (Input-Output Equation) *For the combined deterministic-stochastic system (1), we have the following Input-Output Equation*

$$\begin{aligned} X_k &= \triangle_k^X X_{k-1|0} + \triangle_k^U U_{k-1|0} + \triangle_k^W W_{k-1|0} &\quad (25)\\ Y_{k-1|0} &= \mathcal{L}_k^X X_{k-1|0} + \mathcal{L}_k^U U_{k-1|0} + \mathcal{L}_k^W W_{k-1|0} + \mathcal{L}_k^V V_{k-1|0} &\quad (26)\end{aligned}$$

where

$$\begin{aligned} \triangle_n^X &\triangleq [A\triangle_{n-1}^X, N_1\triangle_{n-1}^X, \dots, N_m\triangle_{n-1}^X] \\ \triangle_1^X &\triangleq [A, N_1, \dots, N_m] \\ \triangle_n^U &\triangleq [B, A\triangle_{n-1}^U, N_1\triangle_{n-1}^U, \dots, N_m\triangle_{n-1}^U] \\ \triangle_1^U &\triangleq B \\ \triangle_n^W &\triangleq [I_{n\times n}, A\triangle_{n-1}^W, N_1\triangle_{n-1}^W, \dots, N_m\triangle_{n-1}^W] \\ \triangle_1^W &\triangleq I_{n\times n} \end{aligned}$$

$$\begin{aligned} \mathcal{L}_k^X &\triangleq \begin{bmatrix} C\triangle_{k-1}^X & 0 \\ \mathcal{L}_{k-1}^X & 0 \\ 0 & \mathcal{L}_{k-1}^X \end{bmatrix} \\ \mathcal{L}_k^U &\triangleq \begin{bmatrix} D & C\triangle_{k-1}^U & 0 \\ 0 & \mathcal{L}_{k-1}^U & 0 \\ 0 & 0 & \mathcal{L}_{k-1}^U \end{bmatrix} \\ \mathcal{L}_k^W &\triangleq \begin{bmatrix} 0 & C\triangle_{k-1}^W & 0 \\ 0 & \mathcal{L}_{k-1}^W & 0 \\ 0 & 0 & \mathcal{L}_{k-1}^W \end{bmatrix} \\ \mathcal{L}_k^V &\triangleq \begin{bmatrix} I_{l\times l} & 0 & 0 \\ 0 & \mathcal{L}_{k-1}^V & 0 \\ 0 & 0 & \mathcal{L}_{k-1}^V \end{bmatrix} \end{aligned}$$

with

$$\mathcal{L}_1^X \triangleq [C, 0_{l\times m}], \quad \mathcal{L}_1^U \triangleq D, \quad \mathcal{L}_1^U \triangleq 0_{l\times n}, \quad \mathcal{L}_1^V \triangleq I_{l\times l}$$

Proof: see appendix.

¿From **Lemma 7**, the data equations and state equations linking the state sequences can be written in the following three-block form.

Theorem 3 (Three Block Form Equation) *The system (1) can be written in the following 'three block' form:*

$$\begin{aligned} Y^p &= \mathcal{O}_k X_p + \mathcal{T}_k^u U^{p,u,y} \\ &\quad + \mathcal{T}_k^v U^{+p} \odot V_p + \mathcal{L}_k^W W^p + \mathcal{L}_k^V V^p \qquad (27) \\ Y^f &= \mathcal{O}_k X_f + \mathcal{T}_k^u U^{f,u,y} \\ &\quad + \mathcal{T}_k^v U^{+f} \odot V_f + \mathcal{L}_k^W W^f + \mathcal{L}_k^V V^f \qquad (28) \\ Y^r &= \mathcal{O}_k X_r + \mathcal{T}_k^u U^{r,u,y} \\ &\quad + \mathcal{T}_k^v U^{+r} \odot V_r + \mathcal{L}_k^W W^r + \mathcal{L}_k^V V^r \qquad (29) \end{aligned}$$

$$\begin{aligned} X_f &= \mathcal{F}_k X_p + \mathcal{G}_k^u U^{p,u,y} + \mathcal{G}_k^v U^{+p} \odot V_p + \triangle_k^W W^p \qquad (30) \\ X_r &= \mathcal{F}_k X_f + \mathcal{G}_k^u U^{f,u,y} + \mathcal{G}_k^v U^{+f} \odot V_f + \triangle_k^W W^f \qquad (31) \end{aligned}$$

Proof: see appendix.

We make an assumption of ergodicity and stationarity of the variables. Then all the covariances used in this paper can be estimated by replacing ensemble means by time means. Therefore, if an infinite amount of data is available, the estimated value operator $\mathbf{E}$ is equivalent to the operator $\mathbf{E}_j$ which is defined as follows:

$$\mathbf{E}_j[\sum_{k=0}^{j} \bullet] \triangleq \frac{1}{j}[\sum_{k=0}^{j} \bullet]$$

and

$$\lim_{j\to\infty} \mathbf{E}_j[\sum_{k=0}^{j} \bullet] = \mathbf{E}[\bullet]$$

The orthogonal projection operator Π is defined as [9] as follows:

$$\Pi_B A \triangleq \mathbf{E}[AB^T]\mathbf{E}^{-1}[BB^T]B$$

Due to the presence of noise, a stronger condition than (18) is needed for experiments to be informative in the combined deterministic-stochastic case.

Theorem 4 *If*

$$\begin{pmatrix} Y^p \\ U^{p,u,y} \\ U^{f,u,y} \\ U^{r,u,y} \end{pmatrix} \qquad (32)$$

is a full row rank matrix, then

$$\begin{aligned}\Pi_{\mathcal{S}}\mathcal{X}_r &\subset \Pi_{\mathcal{S}}\mathcal{Y}^f + \mathcal{U}^{f,u,y} \\ \Pi_{\mathcal{S}}\mathcal{X}_r &+ \mathcal{U}^{r,u,y} = \Pi_{\mathcal{S}}\mathcal{X}_r \oplus \mathcal{U}^{r,u,y} \\ &\subset (\Pi_{\mathcal{S}}\mathcal{Y}^f + \mathcal{U}^{f,u,y}) \oplus \mathcal{U}^{r,u,y}\end{aligned}$$

where $\mathcal{S} := \mathcal{Y}^p + \mathcal{U}^{p,u,y} + \mathcal{U}^{f,u,y} + \mathcal{U}^{r,u,y}$*. Denote* $\mathcal{R} = \Pi_{\mathcal{S}}\mathcal{Y}^f + \mathcal{U}^{f,u,y}$*, then,*

$$\Pi_{\mathcal{R}^\perp}\Pi_{\mathcal{S}}\mathcal{Y}^r = \mathcal{T}_k^u \Pi_{\mathcal{R}^\perp} U^{r,u,y} \tag{33}$$

Proof: see appendix.

In this three-block algorithm, the projection space will be $\mathcal{S} := \mathcal{Y}^p + \mathcal{U}^{p,u,y} + \mathcal{U}^{f,u,y} + \mathcal{U}^{r,u,y}$. Due to the uncorrelated property of past, future and remote future noises, we can remove all the noise effects by projecting on the subspace $\mathcal{S}$ and the 'past' block can be regarded as an auxiliary variable block reminiscent of our instrumental variable. In such a way, the parameter matrix $\mathcal{T}_k^u$ can be determined. The following summarises the algorithm for the combined deterministic-stochastic case,

1. Decompose Y^r into $\mathcal{O}_k X_r$ and $\mathcal{T}_k^u U^{r,u,y}$ using orthgonal projection: from (33) of **Theorem 4**, it follows that

$$\mathcal{T}_k^u = (\Pi_{\mathcal{R}^\perp}\Pi_{\mathcal{S}}\mathcal{Y}^r)(\Pi_{\mathcal{R}^\perp} U^{r,u,y})^\dagger \tag{34}$$

2. Obtain the SVD decompositon and partition, remaining singular values and selecting a model order.

$$\begin{aligned}&\left[\Pi_{\mathcal{S}} Y_{2k-1|k}\ \Pi_{\mathcal{S}_1} Y_{2k|k+1}\right] - \mathcal{T}_k^u \left[U^{u,y}_{2k-1|k}\ U^{u,y}_{2k|k+1}\right] \\ &\qquad =: \Gamma\Sigma\Omega^T = \begin{bmatrix}\Gamma_1 & \Gamma_2\end{bmatrix}\begin{bmatrix}\Sigma_1 & 0 \\ 0 & \Sigma_2\end{bmatrix}\begin{bmatrix}\Omega_1^T \\ \Omega_2^T\end{bmatrix}\end{aligned}$$

Since we expect

$$\Gamma\Sigma\Omega^T = \Gamma_1\Sigma_1\Omega_1^T = \mathcal{O}_k \left[X_{k-1} \quad X_k\right]$$

from (27-29) ($rank(\Sigma_1) = n$ and $rank(\Sigma_2) = 0$), form the estimates $\hat{\mathcal{O}}_k = \Gamma_1\Sigma_1^{1/2}$ and $[\hat{X}_{k-1} \quad \hat{X}_k] = \Sigma_1^{1/2}\Omega_1^T$, retaining only significant singular values in Σ_1. ($\hat{\mathcal{O}}_k$ is not needed later.)

3. Estimate the parameters A, B, C, D, N by solving

$$\begin{bmatrix}\hat{X}_k \\ Y_{k-1}\end{bmatrix} = \begin{bmatrix}A & N & B \\ C & 0 & D\end{bmatrix}\begin{bmatrix}\hat{X}_{k-1} \\ U_{k-1} \odot \hat{X}_{k-1} \\ U_{k-1}\end{bmatrix} \tag{35}$$

in a least-squares sense.

Remark 3 The make of retained singular values, $\hat{n}$, will be the order of the estimated model. In this paper, we do not discuss how $\hat{n}$ should be determined.

Remark 4 Other estimates could be obtained by using other right-inverses in steps 1 and 3, and another factorisation in step 2. In [11] it is suggested that constrained least-squares could be used in step 3, because of the known structure of the solution. Our initial simulation experience is that this does not have much effect on the resulting estimates.

Remark 5 Our new algorithm has considerably lower computational complexity than the one given in [11]. The major computational load is involved in finding the right-inverse in (34). The matrix involved here has dimensions $f_k \times j$, where $f_k = e_k + (m/2)(m+1)^k + l[(m+1)^k - 1]$. For example, with $k = 2$, $n = m = l = 2$, we have $f_k = 33$. In [11], equation (9), the matrix whose right-inverse has to be found has dimensions $(d_k + 2e_k + e_k d_k + e_k^2) \times j$. The row dimension increases exponentially quickly with k. For example, with $k = 2$, $n = m = l = 2$ this row dimension is 152. Furthermore, since our algorithm seems to require much smaller values of j for comparable performance, the column dimension is also much smaller for our algorithm in practice.

Estimate the covariance matrix by calculating

$$\begin{aligned}
\begin{bmatrix} \epsilon_w \\ \epsilon_v \end{bmatrix} &= \begin{bmatrix} \hat{X}_{k+1} \\ Y_k \end{bmatrix} \\
&\quad - \begin{bmatrix} A & N & B \\ C & 0 & D \end{bmatrix} \begin{bmatrix} \hat{X}_k \\ U_k \odot \hat{X}_k \\ U_k \end{bmatrix} \\
\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} &= \begin{bmatrix} \Sigma^w & \Sigma^{wv} \\ \Sigma^{vw} & \Sigma^v \end{bmatrix} \\
&= \mathbf{E} \left[\begin{pmatrix} \epsilon_w \\ \epsilon_v \end{pmatrix} \begin{pmatrix} \epsilon_w \\ \epsilon_v \end{pmatrix}^T \right]
\end{aligned}$$

4 Bilinear Kalman filter

In this section, we derive the equation of the bilinear Kalman filter of the system (1). The derivation is done in a similaar way as done for the linear Kalman filter as well as in [10]. The difference betweem the linear and bilinear Kalman filter is an addition term depending on the system matrix N and unlike in [10], we have no assumption here that the system input u_t is a white noise.

Theorem 5 (Bilinear kalman filter) *1. The non-steady state bilinear Kalman filter state esti-*

mates of the system (1) can be found by solving the following recursive formulas:

$$\hat{x}_{k+1} = A\hat{x}_k + Nu_k \otimes \hat{x}_k + Bu_k + K_k(y_k - C\hat{x}_k - Du_k) \tag{36}$$

where

$$K_k = (G^s - (\sum_{p=1}^{m} \mu_p N_p)P_k C^T - AP_k C^T)(\Lambda_0 - CP_k C^T)^{-1} \tag{37}$$

$$\begin{aligned} P_k &= AP_{k-1}A^T + \sum_{p=1}^{m}\sum_{q=1}^{m} \mu_{p,q} N_p P_{k-1} N_q^T \\ &+AP_{k-1}(\sum_{p=1}^{m} \mu_p N_p^T) + (\sum_{p=1}^{m} \mu_p N_p)P_{k-1}A^T \\ &+ \left[G^s - (\sum_{p=1}^{m} \mu_p N_p)P_k C^T - AP_k C^T\right](\Lambda_0^s - CP_k C^T)^{-1} \\ &\quad \left[G^s - (\sum_{p=1}^{m} \mu_p N_p)P_k C^T - AP_k C^T\right]^T \end{aligned} \tag{38}$$

where $\mu = Eu_t = [\mu_1, \ldots, \mu_m]$ and $\mu_{i,j} = Eu_{t,i}u_{t,j}$ under the assumption that input u_t is a stationary input.

Proof: see appendix.

Theorem 6 *If the condition*

$$\lambda(I - A - \sum_{i=1}^{m} \mu_i N_i) < 1 \tag{39}$$

then set $\mathbf{E}X = \mathbf{E}x_k$

$$\mathbf{E}X = (I - A - \sum_{i=1}^{m} \mu_i N_i)^{-1} B\mu. \tag{40}$$

Proof: see appendix.

Theorem 7 *The algebriac equation*

$$\begin{aligned} P &= APA^T + AP(\sum_{i=1}^{m} \mu_i N_i^T) + A\mathbf{E}X\mu^T B^T \\ &+(\sum_{i=1}^{m} \mu_i N_i)PA^T + \sum_{p=1}^{m}\sum_{q=1}^{m} \mu_{p,q} N_p P N_q^T + \sum_{i=1}^{m} N_i \mathbf{E}X \Lambda_{i.}^T B^T \\ &+B\mu \mathbf{E}X^T A^T + B\sum_{i=1}^{m} \Lambda_{i.} \mathbf{E}X^T N_i^T + Q \end{aligned} \tag{41}$$

where $\Lambda_{i.} = \mathbf{E}u_k u_{k,i}.$ *has a solution* $P \geq 0$ *if and only if*

$$\lambda \left[I - A \otimes A - (\sum_{i=1}^{m} \mu_i N_i) \otimes A - A \otimes (\sum_{i=1}^{m} \mu_i N_i) - \sum_{p=1}^{m} \sum_{q=1}^{m} \mu_{p,q} N_p \otimes N_q \right] < 1 \tag{42}$$

Proof: see appendix.

5 Examples

This section is organised as two subsections. The first section, the examples and comparisons with the existing bilinear subspace algorithm are given in the deterministic system case by using 'two-block' algorithm. In the last subsection, the examples and comparisons with the existing algorithm are given in the case of combined deterministic-stochastic bilinear system.

5.1 Deterministic bilinear system case

In this subsection, two second-order bilinear systems introduced in [9, 12] are used to see how the new algorithm works, and how it compares with existing algorithms.

Example 5.1.1 The system matrices are

$$A = \begin{pmatrix} 0 & 0.5 \\ -0.5 & 0 \end{pmatrix}, B = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, C = \begin{pmatrix} 1 & 1 \end{pmatrix},$$
$$D = 2, \quad N_1 = \begin{pmatrix} 0.4 \\ 0 \end{pmatrix} \quad N_2 = \begin{pmatrix} 0 \\ 0.3 \end{pmatrix}$$

Table 1 shows the eigenvalues of the estimated A and N in various cases. The row labelled 'N4SID' gives the results obtained in [9], with a white input and $k = 3$, $j = 8191$. 'Case I' is for a white input, with uniform distribution in the interval [0,0.01] and $k = 2$. 'Case II' is for a white input with normal distribution $N(0, 0.1^2)$ and $k = 6$. 'Case III' is for a coloured input u with mean 0, standard deviation 3.3e-05 and $r_q = Eu_k u_{k+q} = 0.5^q, \quad q = 0, 1, 2, ...$ and $k = 2$. 'Case IV' is for a white input with exponential distribution with parameter $\lambda = 0.04$ and $k = 2$. In all the cases I–IV the number of columns is only $j = 597$, compared with $j = 8191$ for the N4SID case. It is seen that the eigenvalues of the true and estimated matrices are very close to each other. Table 2 shows how the eigenvalues of the estimated A and N depend on j, in each case with $k = 2$.

Example 5.1.2 The system matrices are

$$A = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.3 \end{pmatrix}, B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, C = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$
$$D = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, N_1 = \begin{pmatrix} 0.6 & 0 \\ 0 & 0.4 \end{pmatrix}, N_2 = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.5 \end{pmatrix}$$

	A	N
True	$\pm 0.5i$	0.4, 0.3
N4SID	$-0.0027 \pm 0.4975i$	0.4011, 0.3055
Case I	$0.0000 \pm 0.5001i$	0.3994, 0.2952
Case II	$0.0000 \pm 0.4993i$	0.4019 0.3064
Case III	$0.0000 \pm 0.5000i$	0.3953 0.3085
Case IV	$0.0006 \pm 0.5011i$	0.3934, 0.3007

Table 1: Example 5.1.1: Results with different inputs and algorithms

	A	N
True	$\pm 0.5i$	0.4, 0.3
j=97	$0.0000 \pm 0.5003i$	0.4092, 0.2789
j=297	$0.0000 \pm 0.5002i$	0.4043, 0.2914
j=597	$0.0000 \pm 0.5001i$	0.3994, 0.2952

Table 2: Example 5.1.1: Effect of sample size

Table 3 shows the results obtained in [12] with $j = 4095$ and $k = 2$, but *with stochastic inputs*, and the results obtained by our 'two-block' algorithm with $j = 597$ and $k = 2$ in the deterministic case. A fairer comparison is available in [1, 2]. In both cases the input signal was white, with a uniform two-dimensional distribution.

	True	N4SID	Two block
A	0.5, 0.3	0.5001, 0.2979	0.5000, 0.3000
N_1	0.6, 0.4	0.5994, 0.4020	0.6000, 0.4000
N_1	0.2, 0.5	0.1914, 0.5016	0.2000, 0.5000

Table 3: Example 5.1.2: Results with different inputs and algorithms

5.2 Combined deterministic-stochastic case

In this section, two simple second order bilinear systems introduced in [9, 11] are given to see how the new algorithm works and some comparison of the simulation results are also presented.

Example 5.2.1 The system matrices are

$$A = \begin{pmatrix} 0 & 0.5 \\ -0.5 & 0 \end{pmatrix}, \ B = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \ C = \begin{pmatrix} 1 & 1 \end{pmatrix},$$

$$D = 2, \quad N_1 = \begin{pmatrix} 0.4 \\ 0 \end{pmatrix}, \quad N_2 = \begin{pmatrix} 0 \\ 0.3 \end{pmatrix}$$

and the noise covariance matrices

$$Q = \begin{pmatrix} 0.16 & 0 \\ 0 & 0.04 \end{pmatrix}, \ R = 0.09, \ S = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

In [9], the input is a white noise series and $k = 3, j = 8191$. In both of the Table 4 and Table 5, system input of case I is a uniform distribution with mean value equals zero and variance equals to 1. In the case II, we adjust the system noise as follows:

$$Q = \begin{pmatrix} 0.0016 & 0 \\ 0 & 0.0004 \end{pmatrix}, \ R = 0.0009, \ S = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

to increase the noise ratio between the input signal and system noise, the simulation results are also given in the Table 4. It is shown that the greater the ratio, the better convergence results will be achieved. For the case of III and IV we introduce a colored noise input signal to have a test, u is a colored noise series with the mean 0, standard deviation 1.1664 and $r_q = Eu_k u_{k+q} = 0.5^q, \quad q = 0, 1, 2, ...$ and the system noise is taken as the same as the case I and II respectively, the simulation results are also shown in Table 4.

It is noticed that the parameter $j = 595$ is applied in all the simulation both in Table 4 and Table 5, which is quite small compare to $j = 8191$ and we take $k = 2$ all in our simulations. The

	A	N
Original	$\pm 0.5i$	0.4, 0.3
N4SID	$-0.0027 \pm 0.4975i$	0.4011, 0.3055
Case I	$-0.0076 \pm 0.4960i$	0.3838, 0.2829
Case II	$0.0000 \pm 0.5000i$	0.4005, 0.3030
Case III	$0.0044 \pm 0.4847i$	0.4048, 0.2688
Case IV	$0.0089 \pm 0.4945i$	0.3906, 0.3149

Table 4: Example 5.2.1: Results with different inputs, noise ratios and algorithms

simulation results with different sample number in the case of I and II is given in Table 5 to show the relationship between the preciseness with the increase of the sample number.

	A	N
Original	$\pm 0.5i$	0.4, 0.3
j=95 (I)	$-0.0103 \pm 0.4725i$	0.3872, 0.2791
j=95 (II)	$0.0066 \pm 0.4956i$	0.3749, 0.2997
j=295 (I)	$0.0323 \pm 0.4789i$	0.4267, 0.3030
j=295 (II)	$0.0084 \pm 0.4965i$	0.3997, 0.3012
j=595 (I)	$0.0076 \pm 0.4960i$	0.3838, 0.2829
j=595 (II)	$0.0000 \pm 0.5000i$	0.4005, 0.3030

Table 5: Example 5.2.1: Effect of sample size

Example 5.2.2 The system matrices and the noise covariance matrices are

$$\begin{aligned}
A &= \begin{pmatrix} 0.5 & 0 \\ 0 & 0.3 \end{pmatrix}, B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, C = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \\
D &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, N_1 = \begin{pmatrix} 0.6 & 0 \\ 0 & 0.4 \end{pmatrix}, N_2 = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.5 \end{pmatrix} \\
Q &= \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix}, \; R = \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix} \\
S &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}
\end{aligned}$$

The input we selected here two-dimensional uniform distribution notation as case V and VI and colored noise input u with $Eu_iu_{i+q} = 0.9^iI_2$ as case VII and case VIII with $j = 595, k = 2$. The difference between case V and case VI, case VII and case VIII is that, the ordinary LS method is used to solve (35) in case V, VII, while the constrained LS method is used in case VI and VIII. It should be pointed that in the N4SID simulation case ($j = 4095, k = 2$) and the comparison of the simulation results are given in Table 6.

The identified model is validated by comparing the eigenvalues of A and N of the model and the original system. It is shown that the eigenvalues of the system matrices of the model and the original system are very close to each other. If the parameter j approches to infinite, they would be identical. It is noticed that the parameter j ¿From the Table 4 and Table 6, it is shown that the algorithm presented here has a quicker convergence convergence rate and smaller sample number requirement compared to the N4SID algorithm. As shown in [9, 11], the noise covariance matrix in here is also not very accurate since the k-parameter here is too small.

	A		N_1		N_2	
Original	0.5,	0.3	0.6,	0.4	0.2,	0.5
N4SID	0.5001		0.4020		0.1914	
(j=4095)	0.2979		0.5994		0.5016	
case V	0.4998		0.5998		0.5000	
(j=595)	0.3002		0.4000		0.2001	
case VI	0.5004		0.5998		0.4999	
(j=595)	0.2990		03997		0.1997	
case VII	0.4992		0.6028		0.5070	
(j=595)	0.2968		0.4007		0.2019	
case VIII	0.4992		0.6035		0.5080	
(j=595)	0.2973		0.3990		0.2027	

Table 6: Example 5.2.2: Comparisons with different algorithms, LS and constrained LS

6 Conclusion

A new subspace algorithm for the identification of bilinear systems and bilinear Kalman filter have been developed. Its main advantage is that the system input does not have to be white. It also has lower computational complexity than the previously proposed algorithm, because the dimensions of the matrices involved in it are much smaller. Its wider applicability has been demonstrated by two examples, which also show that even when coloured inputs are available the new algorithm converges to correct estimates relatively quickly. The presumed reason for this is that, since the algorithm does not depend on whiteness of the input, it is insensitive to the large errors in the sample spectrum which are inevitable with small sample sizes.

7 Acknowledgement

The work reported in this paper was supported by the UK Engineering and Physical Science Research Council under Research Grant GR/M08332, and by the European Research Network on System Identification (ERNSI) under TMR contract ERB FMRX CT98 0206.

References

[1] Chen H. and Maciejowski J.M., Subspace identification of deterministic bilinear systems, *Submitted to ACC 2000, Chicago, Illinois, USA*, June 2000.

[2] Chen H. and Maciejowski J.M., Subspace identification of combined deterministic-stochastic bilinear systems, *Submitted to IFAC Symp. on System Identification, SYSID 2000, Santa Barbara*, June 2000.

[3] Chou C.T. and Maciejowski J.M., System Identification Using Balanced Parameterizations, *IEEE Trans. Auto. Contr.*, Vol. 42, No. 7. July, 1997 pp.956-974.

[4] Chui N.L.C. and Maciejowski J. M., *Sbspace Identification – a Markov Parameter Approach* Technical Report CUED/F-INFENG/TR.337, 1998 Cambridge University

[5] Chui N.L.C. and Maciejowski J. M., *Criteria for Informative Experiments with Subspace Identification* Technical Report CUED/F-INFENG/TR.338, 1998 Cambridge University

[6] Chui N.L.C. and Maciejowski J. M., *A Geometric Interpretation for Subspace Mehtods* Technical Report CUED/F-INFENG/TR.339, 1998 Cambridge University

[7] Chen H., Zinober A.S.I. and Ruan R., Strong Consistency and Convergence Rate of Parameter Identification for Bilinear Systems, *Int.J.Control*, 1996, vol. 63, No.5 pp. 907-919

[8] Desai U. B., *Realization of Bilinear Stochastic Systems* IEEE Trans on AC, Vol. AC-31, No.2, Feb, 1986

[9] Favoreel W, De Moor B. and Van Overschee P., *Subspace identification of bilinear systems subject to white inputs*, ESAT-SISTA/TR 1996-53I, Katholieke Universiteit Leuven, 1996.

[10] Favoreel W, De Moor B. and Van Overschee P., Subspace identification of Balanced deterministic bilinear systems subject to white inputs, *Proc. European Ccontrol Conf, Brussels*, 1997.

[11] Favoreel W and De Moor B, Subspace identification of bilinear systems, *Proc. MTNS, Padova*, 1998. pp. 787-790.

[12] Favoreel W, De Moor B. and Van Overschee P., Subspace identification of bilinear systems subject to white inputs, *IEEE Trans. Auto. Contr.*, Vol 44., No. 6, June 1999. pp. 1157–1165.

[13] Fnaiech F. and Ljung L., Recursive Identification of Bilinear System, *Int.J. Control*, vol. 45, pp. 453-470., 1987.

[14] Isidori A., *Nonlinear Control Systems* Springer-Verlag, 1989

[15] 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.

[16] Krener A.J., Bilinear and nonlinear realizations of input-output maps. *SIAM Journal on Control*, vol. 13, pp.827-834., 1975.

[17] Larimore,W.E, System identification, reduced-order filtering and modeling via canonical variate analysis, *Proc. American Contr.Conf., San Francisco*, (1983).

[18] Van Overschee P and De Moor B, N4SID: subspace algorithms of combined deterministic and stochastic systems, *Automatica*, vol.30, No. 1, pp. 75-93, 1994

[19] Suda N., Kodama S. and Ikeda M., *Matrix Theory in Automatical Control,* Japanese Automatical Control Association, 1973.

[20] Verhaegen,M, and Dewilde,P.M, Subspace model identification, Parts 1 and 2, *Int. J. Contr.*, **56**, 1187-1241, (1992).

8 Appendix

The proof of Lemma 1:

The proof of (5) is in [19].

Now we prove (6) in Lemma 1. From (5), the definition of Kronecker product and the Khatri-Rao product, we can derive that

$$\begin{aligned}(FG \odot HJ) &= [FG_1 \otimes HJ_1, FG_2 \otimes HJ_2, \ldots, FG_l \otimes HJ_l] \\ &= (F \otimes H)\,[G_1 \otimes J_1, G_2 \otimes J_2, \ldots, F_l \otimes J_l] \\ &= (F \otimes H)(G \odot J)\end{aligned}$$

This proves the Lemma 1. ∎

The proof of Lemma 2:

We prove the Lemma 2 by induction.

¿From the definition of $U^{+}_{q|q}$ and $U^{++}_{q|q}$, we know that $\mathcal{U}^{+}_{q|q} \odot \mathcal{U}_q = \mathcal{U}^{++}_{q|q}$. First we prove that the Lemma 2 holds for $i = 1$.

$$\mathcal{U}^{+}_{1+q|q} \odot \mathcal{U}_q = \begin{pmatrix} \mathcal{U}^{+}_{q|q} \odot \mathcal{U}_q \\ \mathcal{U}_{q+1} \odot \mathcal{U}_q \\ \mathcal{U}_{q+1} \odot \mathcal{U}_q \odot \mathcal{U}_q \end{pmatrix}$$

and

$$\mathcal{U}^{++}_{1+q|q} = \begin{pmatrix} \mathcal{U}^{++}_{q|q} \\ \mathcal{U}_{q+1} \odot \mathcal{U}^{++}_{q|q} \end{pmatrix}$$

We can derive that

$$\begin{aligned} \mathcal{U}^{+}_{1+q|q} \odot \mathcal{U}_q \setminus \mathcal{U}^{++}_{1+q|q} &= \begin{pmatrix} \mathcal{U}_{q+1} \odot \mathcal{U}_q \end{pmatrix} \\ &= \mathcal{U}^{+}_{q+1|q+1} \odot \mathcal{U}_q \end{aligned}$$

We suppose that the Lemma 2 holds for $i = n$. Then for $i = n + 1$,

$$\mathcal{U}^{+}_{n+1+q|q} \odot \mathcal{U}_q = \begin{pmatrix} \mathcal{U}^{+}_{n+q|q} \odot \mathcal{U}_q \\ \mathcal{U}_{n+q+1} \odot \mathcal{U}_q \\ \mathcal{U}_{n+q+1} \odot \mathcal{U}^{+}_{n+q|q} \odot \mathcal{U}_q \end{pmatrix}$$

and

$$\mathcal{U}^{++}_{n+1+q|q} = \begin{pmatrix} \mathcal{U}^{++}_{n+q|q} \\ \mathcal{U}_{n+q+1} \odot \mathcal{U}^{++}_{n+q|q} \end{pmatrix}$$

We can derive that

$$\begin{aligned} \mathcal{U}^{+}_{n+q+1|q} \odot \mathcal{U}_q \setminus \mathcal{U}^{++}_{n+1+q|q} &= \begin{pmatrix} \mathcal{U}^{+}_{n+q|q+1} \odot \mathcal{U}_q \\ \mathcal{U}_{n+q+1} \odot \mathcal{U}_q \\ \mathcal{U}_{n+q+1} \odot \mathcal{U}^{+}_{n+q|q+1} \odot \mathcal{U}_q \end{pmatrix} \\ &= \mathcal{U}^{+}_{n+q+1|q+1} \odot \mathcal{U}_q \end{aligned}$$

This proves the Lemma 2. ∎

The proof of Lemma 3

We prove the Lemma 3 by induction.

First we prove the Lemma 3 holds for $i = 1$:

$$\begin{aligned} X_{1|0} &= \begin{pmatrix} X_{0|0} \\ U_1 \odot X_{0|0} \end{pmatrix} \\ &= \begin{pmatrix} X_0 \\ U_0 \odot X_0 \\ U_1 \odot X_0 \\ U_1 \odot U_0 \odot X_0 \end{pmatrix} \\ &= \begin{pmatrix} X_0 \\ U^{+}_{1|0} \odot X_0 \end{pmatrix} \end{aligned}$$

We suppose that the Lemma 3 holds for $i = n$. Then, for $i = n+1$,

$$
\begin{aligned}
X_{n+1|0} &= \begin{pmatrix} X_{n|0} \\ U_{n+1} \odot X_{n|0} \end{pmatrix} \\
&= \begin{pmatrix} X_0 \\ U^{+}_{n|0} \odot X_0 \\ U_{n+1} \odot X_0 \\ U_{n+1} \odot U^{+}_{n|0} \odot X_0 \end{pmatrix} \\
&= \begin{pmatrix} X_0 \\ U^{+}_{n+1|0} \odot X_0 \end{pmatrix}
\end{aligned}
$$

This proves the Lemma 3. ∎

The proof of Lemma 5:

From the notation given in the paper and system (1), we have the block form equation as follows:

$$
\begin{aligned}
Y_p &= CX_p + DU_p + V_p && (43) \\
Y_f &= CX_f + DU_f + V_f && (44) \\
Y_r &= CX_r + DU_r + V_r && (45)
\end{aligned}
$$

¿From (43), we have $CX_p = Y_p - DU_p - V_p$. Therefore there exists a matrix C^{-}, which is a generilized inverse matrix of C, such that $X_p = C^{-}(Y_p - DU_p - V_p)$. This states that $\mathcal{X}_p$ is definitly contained in the row space spanned by $\mathcal{Y}_p + \mathcal{U}_p + \mathcal{V}_p$. As we mentioned in remark 2, we can select the Moore-Penrose pseudo-inverse of C as C^{-} here for instance, similarily for equations (44) and (45), the Lemma 5 is then proved. ∎

The proof of Theorem 1: ¿From Lemma 6., we know that

$$
\begin{aligned}
Y^p &= \overline{\Gamma}_k X^p + H_k U^p \\
&= \overline{\Gamma}_k \begin{pmatrix} X_p \\ U^{+p} \odot X_p \end{pmatrix} + H_k U^p && (46)
\end{aligned}
$$

Let $\overline{\Gamma}_k = \left[\mathcal{C}_k, \overline{\Gamma}_{k,2}\right]$, where $\mathcal{C}_k$ is the first n columns of the matrix $\overline{\Gamma}_k$ and $\overline{\Gamma}_{k,2}$ is the last $\left[(m+1)^k - 1\right] n$ columns of $\overline{\Gamma}_k$. In suach a way, the equation (46) can be written as follows

$$
\begin{aligned}
Y^p &= \mathcal{C}_k X_p + \overline{\Gamma}_{k,2} U^{+p} \odot X_p + H_k U^p \\
&= \mathcal{C}_k X_p + \overline{\Gamma}_{k,2} U^{+p} \odot (C^{\dagger}(Y_p - DU_p) + H_k U^p \\
&= \mathcal{C}_k X_p + (\overline{\Gamma}_{k,2} \otimes C^{\dagger})(U^{+p} \odot Y_p) - (\overline{\Gamma}_{k,2} \otimes C^{\dagger} D)(U^{+p} \odot U_p) + H_k U^p && (47)
\end{aligned}
$$

It is noticed that the last three term of the equation (47) can be regarded as the linear combination of the row vecters spanned by the row space of the matrix $\mathcal{U}^{+p} \odot \mathcal{Y}_p, \mathcal{U}^{+p} \odot \mathcal{U}_p$ and $\mathcal{U}^{+p}$. We divide the item $(\mathcal{U}^{+p} \odot \mathcal{U}_p)$ into two parts *i.e.* $(\mathcal{U}^{++p} \oplus [(\mathcal{U}^{+p} \odot \mathcal{U}_p \setminus \mathcal{U}^{++p})]$ and according to Lemma 2, we know that $[(\mathcal{U}^{+p} \odot \mathcal{U}_p \setminus \mathcal{U}^{++p})] = \mathcal{U}^{+}_{k|1} \subset \mathcal{U}^p$. From the above analysis, there exists a matrix $\mathcal{D}_k$ such that $\mathcal{D}_k U^{p,u,y} = (\overline{\Gamma}_{k,2} \otimes C^{\dagger})(U^{+p} \odot Y_p) - (\overline{\Gamma}_{k,2} \otimes C^{\dagger} D)(U^{++p}) + H_k U^p$, so the equation (15) holds. (16) and (17) of Theorem 1 can be proved similarily. ∎

The proof of Theorem 2:

¿From (14) and Lemma 5, we know that $\mathcal{X}_f \subset \mathcal{X}_p + \mathcal{U}^{p,u,y} \subset \mathcal{Y}^p + \mathcal{U}^{p,u,y}$
Since the condition (18) holds, $\mathcal{X}_f + \mathcal{U}^{f,u,y} \subset \mathcal{X}_f \oplus \mathcal{U}^{f,u,y} \subset (\mathcal{Y}^p + \mathcal{U}^{p,u,y}) \oplus \mathcal{U}^{f,u,y}$. ∎

The proof of Lemma 7:

The proof is by induction.
For $n = 1$, both (25) and (26) are true from the structure of the system, deinitions and notations.

We make an assumption that (25), (26) hold for $n = k$. Suppose that $n = k + 1$. Then

$$\begin{aligned}
X_{k+1} &= AX_k + NU_k \odot X_k + BU_k + W_k \\
&= A(\triangle^X_{k-1} X_{k-1|0} + \triangle^U_{k-1} U_{k-1|0} + \triangle^W_{k-1} W_{k-1|0}) \\
&\quad + NU_K \odot (\triangle^X_{k-1} X_{k-1|0} + \triangle^U_{k-1} U_{k-1|0} + \triangle^W_{k-1} W_{k-1|0}) + BU_k + W_k \\
&= (A\triangle^X_{k-1} X_{k-1|0} + \left[N_1 \triangle^X_{k-1}, N_2 \triangle^X_{k-1}, \dots, N_l \triangle^X_{k-1}\right] U_K \odot X_{k-1|0}) \\
&\quad + (A\triangle^U_{k-1} U_{k-1|0} + BU_k + \left[N_1 \triangle^U_{k-1}, N_2 \triangle^U_{k-1}, \dots, N_l \triangle^U_{k-1}\right] U_k \odot U_{k-1|0}) \\
&\quad + (W_k + A\triangle^W_{k-1} W_{k-1|0} + \left[N_1 \triangle^U_{k-1}, N_2 \triangle^U_{k-1}, \dots, N_l \triangle^U_{k-1}\right] U_k \odot W_{k-1|0}) \\
&= \triangle^X_k X_{k|0} + \triangle^U_k U_{k|0} + \triangle^W_k W_{k|0}
\end{aligned}$$

so equation (25) holds for $n = k + 1$.

$$Y_{k+1|0} = \begin{pmatrix} Y_{k+1} \\ Y_{k|0} \\ U_{K+1} \odot Y_{k|0} \end{pmatrix}$$

$$\begin{aligned}
Y_{k+1} &= CX_{k+1} + DU_{k+1} + V_{k+1} \\
&= C(\triangle^X_k X_{k|0} + \triangle^U_k U_{k|0} + \triangle^W_k W_{k|0}) + DU_{k+1} + V_{k+1} \\
&= C\triangle^X_k X_{k|0} + C\triangle^U_k U_{k|0} + C\triangle^W_k W_{k|0} + DU_{k+1} + V_{k+1}
\end{aligned}$$

$$Y_{k|0} = \mathcal{L}^X_k X_{k|0} + \mathcal{L}^U_k U_{k|0} + \mathcal{L}^W_k W_{k|0} + \mathcal{L}^V_k V_{k|0}$$

$$\begin{aligned}
U_{k+1} \odot Y_{k|0} &= U_{k+1} \odot (\mathcal{L}^X_k X_{k|0} + \mathcal{L}^U_k U_{k|0} + \mathcal{L}^W_k W_{k|0} + \mathcal{L}^V_k V_{k|0}) \\
&= U_{k+1} \odot \mathcal{L}^X_k X_{k|0} + U_{k+1} \odot \mathcal{L}^U_k U_{k|0} + U_{k+1} \odot \mathcal{L}^W_k W_{k|0} + U_{k+1} \odot \mathcal{L}^V_k V_{k|0})
\end{aligned}$$

so equation (26) holds for $n = k + 1$. This proves Lemma 7. ∎

The proof of Theorem 3:

¿From equation (26) of Lemma 5, we know that

$$\begin{aligned} Y^p &= \mathcal{L}_k^X X^p + \mathcal{L}_k^U U^p + \mathcal{L}_k^W W^p + \mathcal{L}_k^V V^p \\ &= \mathcal{L}_k^X \begin{pmatrix} X_p \\ U^{+p} \odot X_p \end{pmatrix} + \mathcal{L}_k^U U^p + \mathcal{L}_k^W W^p + \mathcal{L}_k^V V^p \end{aligned} \tag{48}$$

Let $\mathcal{L}_k^X = \left[\mathcal{O}_k, \mathcal{L}_{k,2}^X\right]$, where $\mathcal{O}_k$ is the first n columns of the matrix $\mathcal{L}_k^X$ and $\mathcal{L}_{k,2}^X$ is the last $(m+1)^k - 1$ columns of $\mathcal{L}_k^X$. Then equation (48) can be written as:

$$\begin{aligned} Y^p &= \mathcal{O}_k X_p + \mathcal{L}_{k,2}^X U^{+p} \odot X_p + \mathcal{L}_k^U U^p + \mathcal{L}_k^W W^p + \mathcal{L}_k^V V^p \\ &= \mathcal{O}_k X_p + \mathcal{L}_{k,2}^X U^{+p} \odot \left[C^\dagger (Y_p - DU_p - V_p)\right] + \mathcal{L}_k^U U^p + \mathcal{L}_k^W W^p + \mathcal{L}_k^V V^p \\ &= \mathcal{O}_k X_p + (\mathcal{L}_{k,2}^X \otimes C^\dagger)(U^{+p} \odot Y_p) - (\mathcal{L}_{k,2}^X \otimes C^\dagger D)(U^{+p} \odot U_p) + \mathcal{L}_{k,2}^X C^\dagger (U^{+p} \odot V_p) \\ &\quad + \mathcal{L}_k^U U^p + \mathcal{L}_k^W W^p + \mathcal{L}_k^V V^p \end{aligned} \tag{49}$$

The sum of the second, third term and fifth term of the equation (49) is a linear combination of row vectors in the span of the row spaces of the matrices $\mathcal{U}^{+p} \odot \mathcal{Y}_p, \mathcal{U}^{+p} \odot \mathcal{U}_p$ and $\mathcal{U}^{+p}$. According to Lemma 2, $(\mathcal{U}^{+p} \odot \mathcal{U}_p)$ can be decomposed into two subspace *i.e.* $\mathcal{U}^{++p} \oplus \left(\mathcal{U}_{k|1}^{+} \odot \mathcal{U}_p\right)$ and the latter subspace is contained in $\mathcal{U}^p$. From above analysis, there exists a matrix $\mathcal{T}_k^u$ such that $\mathcal{T}_k^u U^{p,u,y} = (\mathcal{L}_{k,2}^X \otimes C^\dagger)(U^{+p} \odot Y_p) - (\mathcal{L}_{k,2}^X \otimes C^\dagger D)(U^{++p}) + \mathcal{L}_k^U U^p$, and let $\mathcal{T}_k^v = \mathcal{L}_{k,2}^X C^\dagger$, so the equation (27) of Theorem 3 holds. Equations (28), (29, (30) and (31) of Theorem 3 can be proved similarily. ∎

The proof of Theorem 4:

¿From (31), (28) and Lemma 2, we have

$$\Pi_{\mathcal{S}} \mathcal{X}_r \subset \Pi_{\mathcal{S}} \mathcal{X}_f + \mathcal{U}^{f,u,y} \subset \Pi_{\mathcal{S}} \mathcal{Y}^f + \mathcal{U}^{f,u,y}$$

and from (30) and (27), we get

$$\begin{aligned} \Pi_{\mathcal{S}} \mathcal{Y}^f &\subset \Pi_{\mathcal{S}} \mathcal{X}_f + \mathcal{U}^{f,u,y} \subset \Pi_{\mathcal{S}} \mathcal{X}_p + \mathcal{U}^{p,u,y} + \mathcal{U}^{f,u,y} \\ &\subset \Pi_{\mathcal{S}} \mathcal{Y}^p + \mathcal{U}^{p,u,y} + \mathcal{U}^{f,u,y} \subset \mathcal{Y}^p + \mathcal{U}^{p,u,y} + \mathcal{U}^{f,u,y} \end{aligned}$$

so

$$\begin{aligned} \Pi_{\mathcal{S}} \mathcal{X}_r + \mathcal{U}^{r,u,y} &= \Pi_{\mathcal{S}} \mathcal{X}_r \oplus \mathcal{U}^{r,u,y} \\ &\subset (\Pi_{\mathcal{S}} \mathcal{Y}^f + \mathcal{U}^{f,u,y}) \oplus \mathcal{U}^{r,u,y} \end{aligned}$$

Since

$$\Pi_{\mathcal{S}} X_r \subset \mathcal{R}$$

and from (29), we know that (34) holds due to the last three terms in (29) are orthogonal to $\mathcal{S}$.

The proof of Bilinear Kalman Filter

The aim of the Kalman filter is to find a matrix K_k such that the trace of the error covariance matrix $\tilde{P}_k$ is minimized: $\min_{K_k} tr\tilde{P}_k$. Let us first define the estimate state error $\tilde{x}_k$, the error covariance matrix $\tilde{P}_k$ and the estimate covariance matrix P_k.

$$\begin{aligned}
\tilde{x}_k &\triangleq x_k - \hat{x}_k \\
\tilde{P}_k &\triangleq \mathbf{E}[\tilde{x}_k \tilde{x}_k^T] \\
P_k &\triangleq \mathbf{E}[\hat{x}_k \hat{x}_k^T]
\end{aligned}$$

Let us now look for an expression of $tr\tilde{P}_k$. From the filter equation (36) and the state space description we find that

$$\tilde{x}_{k+1} = A\tilde{x}_k + Nu_k \otimes \tilde{x}_k - K_k C\tilde{x}_k - K_k v_k + w_k$$

$tr\tilde{P}_k$ then becomes :

$$\begin{aligned}
tr(\tilde{P}_{k+1}) &= tr\mathbf{E}[\tilde{x}_{k+1}\tilde{x}_{k+1}^T] \\
&= tr\mathbf{E}[(x_{k+1} - \hat{x}_{k+1})(x_{k+1} - \hat{x}_{k+1})^T] \\
&= tr\mathbf{E}[(A\tilde{x}_k + Nu_k \otimes \tilde{x}_k - K_k C\tilde{x}_k - K_k v_k + w_k) \\
&\qquad (A\tilde{x}_k + Nu_k \otimes \tilde{x}_k - K_k C\tilde{x}_k - K_k v_k + w_k)^T] \\
&= tr\{A\tilde{P}_k A^T + A\tilde{P}_k(\sum_{p=1}^{m} \mu_p N_p^T) - A\tilde{P}_k C^T K_k^T + (\sum_{p=1}^{m} \mu_p N_p)\tilde{P}_k A^T \\
&\quad + (\sum_{p=1}^{m}\sum_{q=1}^{m} \Lambda_{p,q} N_p \tilde{P}_k N_q^T) - (\sum_{p=1}^{m} \mu_p N_p)\tilde{P}_k C^T K_k^T - K_k C\tilde{P}_k A^T \\
&\quad - K_k C\tilde{P}_k(\sum_{p=1}^{m} \mu_p N_p) + K_k C\tilde{P}_k C^T K_k^T + K_k R K_k^T - K_k S^T - S K_k^T + Q\} \\
&= tr\{[K_k(R + C\tilde{P}_k C^T) - A\tilde{P}_k C^T - (\sum_{p=1}^{m} \mu_p N_p)\tilde{P}_k C^T - S](R + C\tilde{P}_k C^T)^{-1} \\
&\quad [K_k(R + C\tilde{P}_k C^T) - A\tilde{P}_k C^T - (\sum_{p=1}^{m} \mu_p N_p)\tilde{P}_k C^T - S]^T\} \\
&\quad + tr\{A\tilde{P}_k A^T + A\tilde{P}_k(\sum_{p=1}^{m} \mu_p N_p^T) + (\sum_{p=1}^{m} \mu_p N_p)\tilde{P}_k A^T + (\sum_{p=1}^{m}\sum_{q=1}^{m} \Lambda_{p,q} N_p \tilde{P}_k N_q^T) \\
&\quad + Q\} \\
&\quad - tr\{A\tilde{P}_k C^T(R + C\tilde{P}_k C^T)^{-1} C\tilde{P}_k A^T + A\tilde{P}_k C^T(R + C\tilde{P}_k C^T)^{-1} C^T \tilde{P}_k(\sum_{p=1}^{m} \mu_p N_p^T) \\
&\quad + A\tilde{P}_k C^T(R + C\tilde{P}_k C^T)^{-1} S^T + (\sum_{p=1}^{m} \mu_p N_p)\tilde{P}_k C^T(R + C\tilde{P}_k C^T)^{-1} C\tilde{P}_k A^T
\end{aligned}$$

$$
+(\sum_{p=1}^{m}\mu_p N_p)\tilde{P}_k C^T(R+C\tilde{P}_k C^T)^{-1}C^T\tilde{P}_k(\sum_{p=1}^{m}\mu_p N_p^T)
$$
$$
+(\sum_{p=1}^{m}\mu_p N_p)\tilde{P}_k C^T(R+C\tilde{P}_k C^T)^{-1}S^T+S(R+C\tilde{P}_k C^T)^{-1}C\tilde{P}_k A^T
$$
$$
+S(R+C\tilde{P}_k C^T)^{-1}C^T\tilde{P}_k(\sum_{p=1}^{m}\mu_p N_p^T)+S(R+C\tilde{P}_k C^T)^{-1}S^T\}
$$

The last equation consists of two terms of which the second one is independent of K_k. Therefore our criterion will be minimized when this first term becomes zero *i.e.*

$$
K_k=(A\tilde{P}_k C^T-(\sum_{p=1}^{m}\mu_p N_p)\tilde{P}_k C^T-S)(R+C\tilde{P}_k C^T)^{-1}
$$

Using the definitions of Λ_0^s, G^s, P_k and the fact that $\tilde{P}_k=\Sigma^s-P_k$ we finally find that

$$
K_k=(G^s-(\sum_{p=1}^{m}\mu_p N_p)P_k C^T-AP_k C^T)(\Lambda_0^s-CP_k C^T)^{-1}, \tag{50}
$$

$$
\begin{aligned}
P_k \quad = \quad & AP_{k-1}A^T+(\sum_{p=1}^{m}\sum_{q=1}^{m}\mu_{p.q}N_p P_{k-1}N_q^T \\
& +AP_{k-1}(\sum_{p=1}^{m}\mu_p N_p^T)+(\sum_{p=1}^{m}\mu_p N_p)P_{k-1}A^T \\
& +(G^s-(\sum_{p=1}^{m}\mu_p N_p)P_k C^T-AP_k C^T)(\Lambda_0^s-CP_k C^T)^{-1} \\
& \qquad (G^s-(\sum_{p=1}^{m}\mu_p N_p)P_k C^T-AP_k C^T)^T
\end{aligned} \tag{51}
$$

∎

The proof of Theorem 6.

Since both u_k and w_k are stationary process, we can make an assumption that x_k is a second order statistics stationary process as well. We take expection on both sides of the first equation of (1) and from the condition (39), we have

$$
\mathbf{E}X=(I-A-\sum_{i=1}^{m}\mu_i N_i)^{-1}B\mu.
$$

where $\mathbf{E}X=\mathbf{E}x_k$. ∎

The proof of Theorem 7.

¿From equation (41), we have

$$\left[I - A \otimes A - (\sum_{i=1}^{m} \mu_i N_i) \otimes A - A \otimes (\sum_{i=1}^{m} \mu_i N_i) - \sum_{p=1}^{m} \sum_{q=1}^{m} \Lambda_{p,q} N_p \otimes N_q \right] vec(P)$$
$$= vec(A\mathbf{E}X\mu^T B^T + \sum_{i=1}^{m} N_i \mathbf{E}X\Lambda_{i.}^T B^T + B\mu\mathbf{E}X^T A^T + B\sum_{i=1}^{m} \Lambda_{i.}\mathbf{E}X^T N_i^T + Q)$$

Since the condition (42) meets, we can complete the proof of theorem 7. ∎