UNBIASED BILINEAR SUBSPACE SYSTEM IDENTIFICATION METHODS

Huixin Chen[†]**, Jan Maciejowski** [*] **and Chris Cox**[†]

[†]School of Computing, Engineering and Technology
University of Sunderland
Sunderland SR6 0DD, England
fax: +44 191 5152781
email: huixin.chen@sunderland.ac.uk; chris.cox@sunderland.ac.uk

[*] Department of Engineering
University of Cambridge
Cambridge CB2 1PZ, England
fax: +44 1223 332662
email: jmm@eng.cam.ac.uk

Keywords: System Identification, Identification of Non-linear Systems, Subspace Methods, Bilinear Systems.

Abstract

Several subspace algorithms for the identification of bilinear systems have been proposed recently. A key practical problem with all of these is the very large size of the data-based matrices which must be constructed in order to 'linearise' the problem and allow parameter estimation essentially by regression. Another shortcoming of currently known subspace algorithms for bilinear systems is that the results are biased for most input signals. This paper focuses on the cause of this bias. A conceptual algorithm which can achieve unbiased estimation under less restrictive assumptions on the system and input signals is presented. It is pointed out that one combination of an existing algorithm and particular conditions on the input is an instance of this conceptual algorithm. Also, the conceptual algorithm may shed light on the trade-off between accuracy and computational complexity which has been noted in our earlier work.

1 Introduction

Recently, subspace methods have been developed for the estimation of bilinear systems. Shortcomings of the methods proposed to date are that they typically give biased estimates for most input signals, and that their computational complexity is extremely high, impeding their practical application.

Favoreel *et al* [6] proposed a 'bilinear N4SID' algorithm which gave unbiased results only if the measured input signal was white. Favoreel and De Moor [7] suggested an alternative algorithm for general input signals. Verdult and Verhaegen [12] pointed out that this algorithm gives biased results, and proposed an alternative algorithm, which involved a nonlinear optimization step. Chen and Maciejowski [2, 3, 4] proposed algorithms for the deterministic and combined deterministic-stochastic cases which give asymptotically unbiased estimates with general inputs, and for which the rate of reduction of bias can be estimated. The computational complexity of these algorithms was also significantly lower than the earlier ones, both because the matrix dimensions were smaller, and because convergence to correct estimates (with sample size) appears to be much faster.

In this paper, we propose a class of unbiased 'four-block' algorithms which forms a basis for the analysis and comparison of various subspace methods for bilinear systems. This class can be viewed as a conceptual algorithm. Various concrete implementations of this conceptual algorithm can be obtained by making different choices of subspaces for the decomposition of the input-output data. By using the framework we present in this paper, we can explain why existing subpace methods can achieve unbiased results, under certain assumptions about the system and the input signals. We hope that the viewpoint adopted here will help to clarify the relations between the various 'bilinear subspace' methods proposed so far, and assist further developments.

The outline of the paper is as follows. Extensive notations are introduced in section 2, and these are followed by some analysis in section 3. Readers familiar with earlier papers on 'bilinear subspace' methods can skip directly to section 4, which introduces the conceptual algorithm referred to above. Some conclusions are made in section 5.

2 Notation

The use of much specialised notation seems to be unavoidable in the current context. Mostly we follow the notation used in [8, 2].

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 [9, 11]: $F \odot G \hat{=} [f_1 \otimes g_1, f_2 \otimes g_2, \dots, f_p \otimes g_p]$ $+$, $\oplus$ and $\cap$ denote the sum, the direct sum and the intersection

[*]Corresponding author

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 Hermitian as $\cdot^{*}$.

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 independent 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.

Our objective is to estimate the matrices A, B, C, D, N (up to state-coordinate transformations) and possibly Q, R, S.

For arbitrary t we define

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

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

$$\begin{aligned} X_p &\hat{=} [x_0\ x_1\ ...\ x_{j-1}] \in \mathbf{R}^{n\times j} \\ X_c &\hat{=} [x_k\ x_{k+1}\ ...\ x_{k+j-1}] \in \mathbf{R}^{n\times j} \\ X_f &\hat{=} [x_{2k}\ x_{2k+1}\ ...\ x_{2k+j-1}] \in \mathbf{R}^{n\times j} \\ X_r &\hat{=} [x_{3k}\ x_{3k+1}\ ...\ x_{3k+j-1}] \in \mathbf{R}^{n\times j} \end{aligned}$$

where k is the *row block size*. The suffices p, c, f and r are supposed to be mnemonic, representing 'past', 'current', '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, W_p, W_f, W_r, V_t, V_p, V_f, V_r, similarly. 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 $\tilde{N} = 4k + j - 1$ in the case of $l \geq n$ and $\tilde{N} = 4k+j-1$ in the case of $l < n$. Let $d_i = \Sigma_{p=1}^{i}(m+1)^{p-1}l$ and $e_i = \Sigma_{p=1}^{i}(m+1)^{p-1}m$.

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

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

Letting:

$$\begin{aligned} X^p &\hat{=} X_{k-1|0},\ X^c \hat{=} X_{2k-1|k} \\ X^f &\hat{=} X_{3k-1|2k},\ X^r \hat{=} X_{4k-1|3k} \\ U^p &\hat{=} U_{k-1|0},\ U^c \hat{=} U_{2k-1|k} \\ U^f &\hat{=} U_{3k-1|2k},\ U^r \hat{=} U_{4k-1|3k} \end{aligned}$$

Y^p, Y^c, Y^f, Y^r, W^p,W^c, W^f, W^r, V^p, V^c, V^f, V^r, U^{+p}, U^{+c}, U^{+f} and U^{+r} can be defined similarly.

Remark 1. We denote by $\mathcal{U}_p$ the space spanned by all the rows of the matrix U_p. That is, $\mathcal{U}_p := \text{span}\{\alpha^* U_p, \quad \alpha \in \mathbf{R}^m\}$ $\mathcal{U}_c, \mathcal{U}_f, \mathcal{U}_r, \mathcal{Y}_p, \mathcal{Y}_c, \mathcal{Y}_f, \mathcal{Y}_r, \mathcal{U}^p, \mathcal{Y}^p, \mathcal{U}^f$ and $\mathcal{Y}^f$ *etc* are defined similarly.

3 Analysis

Lemma 1 *The system (1) can be rewritten in the following matrix equation form:*

$$\begin{aligned} X_{t+1} &= AX_t + NU_t \odot X_t + BU_t + W_t \\ Y_t &= CX_t + DU_t + V_t \end{aligned} \tag{2}$$

Lemma 2 *For $i \geq 0$, and the block size k, we have*

$$X_{k-1+i|i} = \begin{pmatrix} X_i \\ U^+_{k-1+i|i} \odot X_i \end{pmatrix}$$

Lemma 3 *For F, G, H, J of compatible dimensions, $F \in \mathbf{R}^{k\times l}$, $G \in \mathbf{R}^{l\times m}$, $H \in \mathbf{R}^{p\times l}$, $J \in \mathbf{R}^{l\times m}$:*

$$\begin{aligned} (FG \otimes HJ) &= (F \otimes H)(G \otimes J) \\ (FG \odot HJ) &= (F \otimes H)(G \odot J) \end{aligned}$$

Lemma 4 (Input-Output Equation) *For the combined deterministic-stochastic system (1) and $i \geq 0$, we have the following Input-Output Equation*

$$\begin{aligned} X_{k+i+1} &= \triangle^X_k X_{k+i|i-1} + \triangle^U_k U_{k+i|i-1} \\ &\quad + \triangle^W_k W_{k-1+i|i} \\ Y_{k+i|i} &= \mathcal{L}^X_k X_{k+i|i-1} + \mathcal{L}^U_k U_{k+i|i-1} \\ &\quad + \mathcal{L}^W_k W_{k+i|i-1} + \mathcal{L}^V_k V_{k+i|i-1} \end{aligned}$$

where

$$\begin{aligned} \triangle^X_t &\hat{=} [A\triangle^X_{t-1}, N_1\triangle^X_{t-1}, \ldots, N_m\triangle^X_{t-1}] \\ \triangle^X_1 &\hat{=} [A, N_1, \ldots, N_m] \end{aligned}$$

$$
\begin{array}{rcl}
\triangle_t^U & \hat{=} & [B, A\triangle_{t-1}^U, N_1\triangle_{t-1}^U, \ldots, N_m\triangle_{t-1}^U] \\
\triangle_1^U & \hat{=} & B \\
\triangle_t^W & \hat{=} & [I_{n\times n}, A\triangle_{t-1}^W, N_1\triangle_{t-1}^W, \ldots, N_m\triangle_{t-1}^W] \\
\triangle_1^W & \hat{=} & I_{n\times n} \\
\mathcal{L}_t^X & \hat{=} & \begin{bmatrix} C\triangle_{t-1}^X & 0 & \ldots & 0 \\ \mathcal{L}_{t-1}^X & 0 & \ldots & 0 \\ 0 & \mathcal{L}_{t-1}^X & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & \mathcal{L}_{t-1}^X \end{bmatrix}
\end{array}
$$

$$
\begin{array}{rcl}
\mathcal{L}_t^U & \hat{=} & \begin{bmatrix} D & C\triangle_{t-1}^U & 0 & \ldots & 0 \\ 0 & \mathcal{L}_{t-1}^U & 0 & \ldots & 0 \\ 0 & 0 & \mathcal{L}_{t-1}^U & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \vdots & \mathcal{L}_{t-1}^U \end{bmatrix} \\
\mathcal{L}_t^W & \hat{=} & \begin{bmatrix} 0 & C\triangle_{t-1}^W & 0 & \ldots & 0 \\ 0 & \mathcal{L}_{t-1}^W & 0 & \ldots & 0 \\ 0 & 0 & \mathcal{L}_{t-1}^W & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \vdots & \mathcal{L}_{t-1}^W \end{bmatrix} \\
\mathcal{L}_t^V & \hat{=} & \begin{bmatrix} I_{l\times l} & 0 & \ldots & 0 \\ 0 & \mathcal{L}_{t-1}^V & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mathcal{L}_{t-1}^V \end{bmatrix}
\end{array}
$$

with

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

Remark 2. For proofs of Lemmas 1–4 see [1].

Lemma 5 *For the combined deterministic-stochastic system (1) and $i \geq 0$, we have the following Block Form Input-Output Equation*

$$
\begin{array}{rcl}
X_{k+i+1} & = & \triangle_{k,1}^X X_{k+i} + \triangle_{k,2}^X (U_{k-1+i|i}^+ \odot X_{k+i}) \\
& & + \triangle_k^U U_{k+i|i-1} + \triangle_k^W W_{k-1+i|i} \\
Y_{k+i|i} & = & \mathcal{L}_{k,1}^X X_{k+i} + \mathcal{L}_{k,2}^X (U_{k-1+i|i}^+ \odot X_{k+i}) \\
& & + \mathcal{L}_k^U U_{k+i|i-1} + \mathcal{L}_k^W W_{k+i|i-1} + \mathcal{L}_k^V V_{k+i|i-1}
\end{array}
$$

where $\triangle_{k,1}^X$ and $\triangle_{k,2}^X$ are the first n columns and last $((m+1)^k-1)n$ columns of the matrix $\triangle_k^X$; $\mathcal{L}_{k,1}^X$ and $\mathcal{L}_{k,2}^X$ are the first n columns and last $((m+1)^k-1)n$ columns of the matrix $\mathcal{L}_k^X$ respectively.

Proof: From the definitions of $\triangle_{k,1}^X, \triangle_{k,2}^X, \mathcal{L}_{k,1}^X$ and $\mathcal{L}_{k,2}^X$, it is easy to get Lemma 5 from Lemma 4.

Lemma 6 *The system (1) can be written as the following block form equations:*

$$
\begin{array}{rcl}
Y^c & = & \mathcal{L}_{k,1}^X X_c + \mathcal{L}_{k,2}^X (U^{+c} \odot X_c) \\
& & + \mathcal{L}_k^U U^c + \mathcal{L}_k^W W^c + \mathcal{L}_k^V V^c \qquad (3) \\
Y^f & = & \mathcal{L}_{k,1}^X X_f + \mathcal{L}_{k,2}^X (U^{+f} \odot X_f) \\
& & + \mathcal{L}_k^U U^f + \mathcal{L}_k^W W^f + \mathcal{L}_k^V V^f \qquad (4) \\
Y^r & = & \mathcal{L}_{k,1}^X X_r + \mathcal{L}_{k,2}^X (U^{+r} \odot X_f) \\
& & + \mathcal{L}_k^U U^r + \mathcal{L}_k^W W^r + \mathcal{L}_k^V V^r \qquad (5) \\
X_f & = & \triangle_{k,1}^X X_c + \triangle_{k,2}^X (U^{+c} \odot X_c) + \triangle_k^W W^c \\
X_r & = & \triangle_{k,1}^X X_f + \triangle_{k,2}^X (U^{+f} \odot X_f) + \triangle_k^W W^f
\end{array}
$$

4 A Conceptual Algorithm

The bias in bilinear subspace methods is due to the Khatri-Rao product term in the block form equations. How to handle this bilinear term is the key to solving the biased estimation problem. If the system input is white, it is a comparatively easy problem to handle, since the Khatri-Rao product term will disappear after projection onto a suitable data-based subspace. Then the identification problem becomes essentially the same as in the linear case, and can be dealt with in the same way. There are many algorithms which give unbiased results in the linear case — see [6] and references therein. Since the requirement of a white input is a very tight restriction, we consider general inputs in the rest of this paper. The main idea to solve the biased problem is to 'linearise' the system equation, in a sense, and try to find a subspace generated by data sets which contains the block system state. Then bilinear systems can be estimated in a similar way as linear systems.

In the following $\Pi_A B$ denotes the orthogonal projection of the rows of matrix B onto the space spanned by the rows of matrix A.

Theorem 1 (Four Block Form Equation) *Suppose that there exists a block size k, a constant (ie independent of data) matrix $H \in \mathbf{R}^{n\times s}$ and a mapping*

$$
\begin{array}{rcl}
f(.) & : & \underbrace{\mathbf{R}^{m\times j} \times \mathbf{R}^{m\times j} \times \ldots \times \mathbf{R}^{m\times j}}_{2k+1} \\
& & \times \underbrace{\mathbf{R}^{l\times j} \times \mathbf{R}^{l\times j} \times \ldots \times \mathbf{R}^{l\times j}}_{2k+1} \\
& \longrightarrow & \mathbf{R}^{s\times j}
\end{array}
$$

such that

$$X_{k+i+1} = Hf(U_{k+i}, ..., U_{i+1}, Y_{k+i}, ..., Y_{i+1}, W_{k+i}, V_{k+i}) \quad (6)$$

Let

$$Z_{k+i} = f(U_{k+i}, ..., U_{i+1}, Y_{k+i}, ..., Y_{i+1}, W_{k+i}, V_{k+i}) \quad (7)$$

where the mapping $f(.)$ is constructed in such a way that the rows of Z_{k+i} are the rows of

$U_{k+i},...,U_{i+1},Y_{k+i},...,Y_{i+1},W_{k+i},V_{k+i}$ and their combinations of Khatri-Rao products. Then the system (1) can be written in the following 'four block' form:

$$\begin{aligned} Y^c &= \mathcal{L}^X_{k,1}X_c + \mathcal{T}^u_k\tilde{U}^{c,u,z} \\ &\quad +\mathcal{L}^W_k W^c + \mathcal{L}^V_k V^c &(8)\\ Y^f &= \mathcal{L}^X_{k,1}X_f + \mathcal{T}^u_k\tilde{U}^{f,u,z} \\ &\quad +\mathcal{L}^W_k W^f + \mathcal{L}^V_k V^f &(9)\\ Y^r &= \mathcal{L}^X_{k,1}X_r + \mathcal{T}^u_k\tilde{U}^{r,u,z} \\ &\quad +\mathcal{L}^W_k W^r + \mathcal{L}^V_k V^r &(10)\\ X_f &= \triangle^X_{k,1}X_c + \mathcal{G}^u_k\tilde{U}^{c,u,z} + \triangle^W_k W^c &(11)\\ X_r &= \triangle^X_{k,1}X_f + \mathcal{G}^u_k\tilde{U}^{f,u,z} + \triangle^W_k W^f &(12)\end{aligned}$$

where $\mathcal{T}^u_k$ and $\mathcal{G}^u_k$ are system-dependent constant matrices and

$$\tilde{U}^{c,u,z} = \begin{pmatrix} U^c \\ U^{+c}\odot Z_{k-1}\end{pmatrix},\ \tilde{U}^{f,u,z} = \begin{pmatrix} U^f \\ U^{+f}\odot Z_{2k-1}\end{pmatrix}$$
$$\tilde{U}^{r,u,z} = \begin{pmatrix} U^r \\ U^{+r}\odot Z_{3k-1}\end{pmatrix}$$

Proof outline:

From (3) and Lemma 3, we have

$$\begin{aligned} Y^c &= \mathcal{L}^X_{k,1}X_c + \mathcal{L}^X_{k,2}(U^{+c}\odot X_c) \\ &\quad +\mathcal{L}^U_k U^c + \mathcal{L}^W_k W^c + \mathcal{L}^V_k V^c \\ &= \mathcal{L}^X_{k,1}X_c + \mathcal{L}^X_{k,2}(U^{+c}\odot HZ_{k-1}) \\ &\quad +\mathcal{L}^U_k U^c + \mathcal{L}^W_k W^c + \mathcal{L}^V_k V^c \\ &= \mathcal{L}^X_{k,1}X_c + (\mathcal{L}^X_{k,2}\otimes H)(U^{+c}\odot Z_{k-1}) \\ &\quad \mathcal{L}^U_k U^c + \mathcal{L}^W_k W^c + \mathcal{L}^V_k V^c \\ &= \mathcal{L}^X_{k,1}X_c + \mathcal{T}^u_k\tilde{U}^{c,u,z} \\ &\quad +\mathcal{L}^W_k W^c + \mathcal{L}^V_k V^c \end{aligned}$$

where

$$\mathcal{T}^u_k = [\mathcal{L}^U_k, (\mathcal{L}^X_{k,2}\otimes H)]$$

which proves (8). Equations (9) - (12) can be obtained using similar manipulations.

In this and the following theorem 'p' (past), 'c' (current), 'f' (future) and 'r' (remote future) data blocks are used. Hence we refer to the algorithm which follows as a 'four-block' algorithm — note that Z_{k-1} and $\hat{Z}_{k-1}$ contain 'past' data, and that $\tilde{N} = 4k+j-1$. Let

$$\hat{U}^{c,u,z} = \begin{pmatrix} U^c \\ U^{+c}\odot \hat{Z}_{k-1}\end{pmatrix},\ \hat{U}^{f,u,z} = \begin{pmatrix} U^f \\ U^{+f}\odot \hat{Z}_{2k-1}\end{pmatrix}$$
$$\hat{U}^{r,u,z} = \begin{pmatrix} U^r \\ U^{+r}\odot \hat{Z}_{3k-1}\end{pmatrix}$$

where

$$\hat{Z}_{k+i} = \Pi_{\mathcal{F}}Z_{k+i} \qquad (13)$$

and $\mathcal{F}$ is the subspace spanned by the rows of $f(U_{k+i},...,U_{i+1},Y_{k+i},...,Y_{i+1},0,0)$.

Theorem 2 *Suppose that the linear part of the system (1) is observable and*

$$\begin{pmatrix} Y^c \\ \hat{U}^{c,u,z} \\ \hat{U}^{f,u,z} \\ \hat{U}^{r,u,z}\end{pmatrix} \qquad (14)$$

is a full row rank matrix. Denote $\tilde{\mathcal{S}} = \mathcal{Y}^c + \hat{\mathcal{U}}^{c,u,z} + \hat{\mathcal{U}}^{f,u,z} + \hat{\mathcal{U}}^{r,u,z}$ and $\tilde{\mathcal{R}} = \Pi_{\tilde{\mathcal{S}}}\mathcal{Y}^f + \hat{\mathcal{U}}^{f,u,z}$, then

$$\Pi_{\tilde{\mathcal{R}}^\perp}\Pi_{\tilde{\mathcal{S}}}\mathcal{Y}^r = \mathcal{T}^u_k\Pi_{\tilde{\mathcal{R}}^\perp}\hat{U}^{r,u,z} \qquad (15)$$

Proof: (15) is obtained using manipulations as in [1].

Algorithm:

1. Decompose Y^r into $\mathcal{O}_kX_r$ and $\mathcal{T}^u_k\hat{U}^{r,u,z}$ using orthogonal projection: from (15) of Theorem 2, estimate $\mathcal{T}^u_k$ as

$$\hat{\mathcal{T}}^u_k = (\Pi_{\tilde{\mathcal{R}}^\perp}\Pi_{\tilde{\mathcal{S}}}\mathcal{Y}^r)(\Pi_{\tilde{\mathcal{R}}^\perp}\hat{U}^{r,u,z})^\dagger \qquad (16)$$

2. Obtain the SVD decomposition and partition as

$$\begin{aligned}&\left[\Pi_{\tilde{\mathcal{S}}}Y_{4k-1|3k}\ \Pi_{\tilde{\mathcal{S}}}Y_{4k|3k+1}\right] - \hat{\mathcal{T}}^u_k\left[\hat{U}^{u,z}_{4k-1|3k}\ \hat{U}^{u,z}_{4k|3k+1}\right] \\ &=: \Gamma\Sigma\Omega^* = \begin{bmatrix}\Gamma_1 & \Gamma_2\end{bmatrix}\begin{bmatrix}\Sigma_1 & 0\\ 0 & \Sigma_2\end{bmatrix}\begin{bmatrix}\Omega_1^*\\ \Omega_2^*\end{bmatrix}\end{aligned}$$

Since we expect

$$\Gamma\Sigma\Omega^* = \Gamma_1\Sigma_1\Omega_1^* = \mathcal{L}^X_{k,1}\left[X_{4k-1}\quad X_{4k}\right]$$

from (8–10), ($\mathrm{rank}(\Sigma_1) = n$ and $\mathrm{rank}(\Sigma_2) = 0$), form the estimates $\hat{\mathcal{L}}^X_{k,1} = \Gamma_1\Sigma_1^{1/2}$ and $[\hat{X}_{4k-1}\quad \hat{X}_{4k}] = \Sigma_1^{1/2}\Omega_1^*$, retaining only $\hat{n}$ significant singular values in Σ_1. ($\hat{\mathcal{L}}^X_{k,1}$ is not needed later.)

3. Estimate the parameters A, B, C, D, N on the basis of equation (2), by solving

$$\begin{bmatrix}\hat{X}_{4k}\\ Y_{4k-1}\end{bmatrix} = \begin{bmatrix}A & N & B\\ C & 0 & D\end{bmatrix}\begin{bmatrix}\hat{X}_{4k-1}\\ U_{4k-1}\odot\hat{X}_{4k-1}\\ U_{4k-1}\end{bmatrix} \qquad (17)$$

in a least-squares sense.

4. Estimate the covariance matrix by calculating

$$\begin{aligned}\begin{bmatrix}\epsilon_w\\ \epsilon_v\end{bmatrix} &= \begin{bmatrix}\hat{X}_{4k}\\ Y_{4k-1}\end{bmatrix} \\ &\quad -\begin{bmatrix}\hat{A} & \hat{N} & \hat{B}\\ \hat{C} & 0 & \hat{D}\end{bmatrix}\begin{bmatrix}\hat{X}_{4k-1}\\ U_{4k-1}\odot\hat{X}_{4k-1}\\ U_{4k-1}\end{bmatrix}\end{aligned}$$

then estimating Q, S, R from the sample covariance of $[\epsilon_w^T, \epsilon_v^T]^T$.

Remark 3. Theorem 1 allows the states X_c, X_f and X_r of the bilinear system (2) to be related linearly to each other, provided that (6) holds. This is the basis of parameter estimation by regression as in (17). Also, since w and v are assumed to be white sequences, and independent of u, $H\hat{Z}_{k+i}$ is a consistent estimate of X_{k+i+1} (as $j \to \infty$) if condition (6) holds. Hence the estimates obtained from (17) are asymptotically unbiased if (6) holds.

Remark 4. The algorithm presented in [7] will be unbiased if the condition:

$$\mathcal{X}^f \subset \begin{pmatrix} \mathcal{U}^f \\ \mathcal{U}^f \odot \begin{pmatrix} \mathcal{U}^c \\ \mathcal{Y}^c \end{pmatrix} \end{pmatrix} \tag{18}$$

is met.

It is easy to show that the condition (18) is equivalent to (19)

$$X_{k+i+1} = H \begin{pmatrix} U_{k+i|i+1} \\ Y_{k+i|i+1} \end{pmatrix} \tag{19}$$

which is a special case of condition (6) of Theorem 1. This explains why the algorithm presented in [7] can give unbiased results in certain cases.

Remark 5. If $l \geq n$ holds, from [4], we can also deduce that condition (6) of Theorem 1 holds: We have

$$\begin{aligned} X_{k+i+1} &= AX_{k+i} + N(U_{k+i} \odot X_{k+i}) + BU_{k+i} + W_{k+i} \\ &= A\left[C^\dagger(Y_{k+i} - DU_{k+i} - V_{k+i})\right] \\ &\quad +N\left(U_{k+i} \odot \left[C^\dagger(Y_{k+i} - DU_{k+i} - V_{k+i})\right]\right) \\ &\quad +BU_{k+i} + W_{k+i} \\ &= \begin{pmatrix} (AC^\dagger)^T \\ (B - AC^\dagger D)^T \\ (-AC^\dagger)^T \\ (N \otimes C^\dagger)^T \\ -(N \otimes C^\dagger D)^T \\ -(N \otimes C^\dagger)^T \\ I \end{pmatrix}^T \begin{pmatrix} Y_{k+i} \\ U_{k+i} \\ V_{k+i} \\ U_{k+i} \odot Y_{k+i} \\ U_{k+i} \odot U_{k+i} \\ U_{k+i} \odot V_{k+i} \\ W_{k+i} \end{pmatrix} \\ &= Hf(U_{k+i}, Y_{k+i}, W_{k+i}, V_{k+i}) \end{aligned}$$

where

$$H = \begin{pmatrix} (AC^\dagger)^T \\ (B - AC^\dagger D)^T \\ (-AC^\dagger)^T \\ (N \otimes C^\dagger)^T \\ -(N \otimes C^\dagger D)^T \\ -(N \otimes C^\dagger)^T \\ I \end{pmatrix}^T$$

$$Z_{k+i} = f(U_{k+i}, Y_{k+i}, W_{k+i}, V_{k+i}) = \begin{pmatrix} Y_{k+i} \\ U_{k+i} \\ V_{k+i} \\ U_{k+i} \odot Y_{k+i} \\ U_{k+i} \odot U_{k+i} \\ U_{k+i} \odot V_{k+i} \\ W_{k+i} \end{pmatrix}$$

$$\hat{Z}_{k+i} = \Pi_{\mathcal{F}} Z_{k+i} = \begin{pmatrix} Y_{k+i} \\ U_{k+i} \\ U_{k+i} \odot Y_{k+i} \\ U_{k+i} \odot U_{k+i} \end{pmatrix}$$

Thus the unbiased algorithm (for the case $l \geq n$) presented in [4] is a special case of our new conceptual algorithm.

Remark 6.

Similarly, we can deduce the approximate unbiasedness of our algorithm given in [4], if condition (6) of Theorem 1 holds only approximately. In this case (8)–(12) hold only approximately; in order that the approximation error should reduce with the block size k, some kind of stability condition must be imposed. In previous work a rather severe sufficient condition has been imposed for this purpose. It is not currently known how far this condition can be relaxed.

5 Conclusion

Apart from the problem of bias, very high computational cost is also a major problem of currently-proposed subspace methods for bilinear system identification. In our new conceptual algorithm, the key dimensions of the matrices involved depend on the mapping $f(.)$. In particular, the smaller the dimension of the row space spanned by the image of the mapping, the lower the computional cost of the algorithm. The practical implication of this is that one can first try some relatively low dimensional subspace (hence low computational cost) to estimate the system matrices, then try a larger-dimensional subspace to see whether better results are obtained. The various proposals put forward in [1, 4, 7, 8] differ essentially in proposing different mappings $f(.)$, or equivalently, different subspaces onto which projections are made. (The proposal in [12] was different in nature, using projection onto the same subspace as in [8] followed by nonlinear processing to remove bias.)

In this paper we have attempted to clarify some aspects of the differences between various proposals for 'bilinear subspace' algorithms, by formulating a conceptual algorithm which includes most of the existing proposals. The main idea is to find a subspace constructed from input-output data, which contains the system state block and allows the system to be 'linearised' so that an unbiased algorithm can be obtained. Both the resulting bias and the computational complexity depend on the choice of a suitable data-based subspace, on the system behaviour, and on the nature of the input signal. By reference to this conceptual algorithm we have explained why some of the existing proposals give unbiased results in certain circumstances.

We hope that this work will be useful in the development of further algorithms with reduced bias and/or complexity, and in the investigation of any inherent trade-off between these.

6 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., *New Subspace Identification Method for Bilinear Systems* Cambridge University Engineering Department Technical Report CUED/F-INFENG/TR.357, 1999, (Submitted to *Automatica*).

[2] Chen H. and Maciejowski J.M., Subspace Identification of Deterministic Bilinear Systems, *ACC 2000, Chicago, Illinois, USA*, June, 2000

[3] Chen H. and Maciejowski J.M., Subspace Identification of Combined Deterministic-Stochastic Bilinear Systems, *IFAC Symp. on System Identification, SYSID 2000, Santa Barbara, California, USA*, June, 2000

[4] Chen H. and Maciejowski J.M., An Improved Subspace method for Bilinear Systems *Accepted for CDC 2000, Sydney, Australia*, December, 2000

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

[6] Favoreel W, De Moor B. and Van Overschee P., *Subspace identification of Balanced deterministic bilinear systems subject to white inputs*, ECC, 1997

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

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

[9] 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.

[10] 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

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

[12] Verdult V., and Verhaegen M., *Subspace-based Identification of MIMO Bilinear Systems*, ECC 1999, 31.August-3.September, 1999, Karlsruhe, Germany