Bilinear System Identification Using Subspace Method

*Wesam M. Jasim**

Date of acceptance 14/6/2007

Abstract

In this paper, a subspace identification method for bilinear systems is used . Wherein a " three-block " and " four-block " subspace algorithms are used. In this algorithms the input signal to the system does not have to be white . Simulation of these algorithms shows that the " four-block " gives fast convergence and the dimensions of the matrices involved are significantly smaller so that the computational complexity is lower as a comparison with " three-block " algorithm .

Introduction

Bilinear systems are attractive models for many dynamical processes, because they allow a significantly larger class of behaviors than linear systems, yet retain a rich theory which is closely related to the familiar theory of linear systems . 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 .

Most studies of 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 and some statistical results (strong consistency and parameter estimate convergence rates) are also available [(1)].

Favoreel et al proposed a " bilinear N4SID " algorithm which gave unbiased results only if the measured input signal was white [(2)]. Favoreel and De Moor suggested an alternative algorithm for general input signals [(3)]. Verdult and Verhaegen pointed out that this algorithm gives biased results, and proposed an alternative algorithm, which involved a nonlinear optimization step [(4)]. Chen and Maciejowski 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 [(5-6)].

In this paper, A comparison between the " three-block " and " four-block " subspace algorithms is shown in two examples.

A " three-block " algorithm can remove the effective of unmeasured noise sources and obtain accurate estimates . In the linear system case, its linked with the systems Markov parameters. Unfortunately, the realization theory for bilinear systems is more complicated than for linear systems.

In " four-block " subspace method

*Al-anbar university, College of computer, Information system

for bilinear systems, the data matrix arrangement and the observation matrix equation are used to linearise the system equation in the block matrix form using linear and bilinear algebra . This allows bilinear models to be obtained from row and column spaces of certain matrices, calculated from the input-output data by means of some bilinear-algebraic operations [6].

Notations

The use of much specialized notation seems to be unavoidable in the current context . Mostly we follow the notation used in [5-7], 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 $F \in R^{t\times p}$ and $G \in R^{u\times p}$ as:

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

$+$, $\oplus$ and $\cap$ denote the sum, the direct sum and the intersection of two vector spaces, $\perp$ denotes the orthogonal complement of a subspace with respect to the predefined ambient space, the Moore-penrose inverse is written as ^{T}, and the Hermitian as * .

In this paper we consider the bilinear system of the form:

$$x_{t+1} = Ax_t + Nu_t \otimes x_t + Bu_t + w_t$$
$$y_t = Cx_t + Du_t + v_t$$
…(1)

where,

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

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:

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

we assume that the sample size is Ñ, namely that input-output data {u(t),y(t):t=0,1,…., Ñ} are available . For arbitrary t we define

$$x_t \triangleq [x_t\, x_{t+1} \ldots\ldots x_{t+j-1}] \in R^{nxj}$$

but for the special cases t=0, t=k, t=2k and t=3k we define with some abuse of notation,

$$X_p \triangleq [x_0 x_1 \ldots x_{j-1}] \in R^{n\times j}$$
$$X_c \triangleq [x_k x_{k+1} \ldots x_{k+j-1}] \in R^{n\times j}$$
$$X_f \triangleq [x_{2k} x_{2k+1} \ldots x_{2k+j-1}] \in R^{n\times j}$$
$$X_r \triangleq [x_{3k} x_{3k+1} \ldots x_{3k+j-1}] \in R^{n\times j}$$

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. This division of the state history into four (overlapping) segments is the reason for 'four-block' method. The 'remote future' segment is not necessary in 'three-block' method.

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 'generalized block-Hankel' structure. In order to use all the available data in these, the number of columns j is such that Ñ=3k+j-1 and let

$$d_i = \sum_{p=1}^{i} (m+1)^{p-1} l,$$
$$e_i = \sum_{p=1}^{i} (m+1)^{p-1} m \quad \text{and}$$
$$f_k = e_k + \frac{m}{2}(m+1)^k + l\left[(m+1)^k - 1\right] + e_k^2.$$

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

$$X_{q/q} \triangleq \begin{pmatrix} X_q \\ U_q \odot X_q \end{pmatrix} \in R^{(m+1)n\times j}$$
$$X_{i-1/q} \triangleq \begin{pmatrix} X_{i-2/q} \\ U_{i-1} \odot X_{i-2/q} \end{pmatrix} \in R^{(m+1)^{i-q} n\times j}$$

$Y_{q/q} \underline{\underline{\Delta}} Y_q$

$$Y_{i-1/q} \underline{\underline{\Delta}} \begin{pmatrix} Y_{i-1} \\ Y_{i-2/q} \\ U_{i-1} \odot Y_{i-2/q} \end{pmatrix} \in R^{d_{i-q} \times j}$$

$U^{+}_{q/q} \underline{\underline{\Delta}} U_q$

$$U^{+}_{i-1/q} \underline{\underline{\Delta}} \begin{pmatrix} U^{+}_{i-2} \\ U_{i-1} \\ U_{i-1} \odot U^{+}_{i-2/q} \end{pmatrix} \in R^{\left((m+1)^{i-q}-1\right) \times j}$$

$$U^{++}_{q/q} \underline{\underline{\Delta}} \begin{pmatrix} U_{q,1} \odot U_q \\ U_{q,2} \odot U_q(2:m,:) \\ U_{q,3} \odot U_q(3:m,:) \\ \vdots \\ U_{q,m} \odot U_{q,m} \end{pmatrix} \in R^{\frac{m(m+1)}{2} \times j}$$

$$U^{++}_{i-1/q} \underline{\underline{\Delta}} \begin{pmatrix} U^{++}_{i-2/q} \\ U_{i-1} \odot U^{++}_{i-2/q} \end{pmatrix} \in R^{\frac{m}{2}(m+1)^{i-q} \times j}$$

$U^{y}_{i-1/q} \underline{\underline{\Delta}} U^{+}_{i-1/q} \odot Y_q$

$$U_{i+k-1/k+q} \underline{\underline{\Delta}} \begin{pmatrix} U_{i+k-1/k+q} \\ U^{++}_{i+k-1/k+q} \\ U^{y}_{i+k-1/k+q} \\ U^{+}_{i+k-1/k+q} \odot U_{i-1/q} \end{pmatrix}$$

$U_{i-1/q}, W_{i-1/q}$ and $V_{i-1/q}$ can be defined similarly [(6)].

Remark 1. The meaning of $U^{++}_{i-1/q}$ is:

$X^{c} \underline{\underline{\Delta}} X_{2k-1/k}$, $X^{f} \underline{\underline{\Delta}} X_{3k-1/2k}$, $X^{r} \underline{\underline{\Delta}} X_{4k-1/3k}$

$U^{p} \underline{\underline{\Delta}} U_{k-1/0}$, $U^{c} \underline{\underline{\Delta}} U_{2k-1/k}$, $U^{f} \underline{\underline{\Delta}} U_{3k-1/2k}$

$U^{p,y} \underline{\underline{\Delta}} U^{+p} \odot Y_p, U^{c,y} \underline{\underline{\Delta}} U^{+c} \odot Y_c$

$U^{f,y} \underline{\underline{\Delta}} U^{+f} \odot Y_f, U^{r,y} \underline{\underline{\Delta}} U^{+r} \odot Y_r$,

$$U^{c,u,y} \underline{\underline{\Delta}} \begin{pmatrix} U^{c} \\ U^{++c} \\ U^{c,y} \\ U^{+c} \odot U^{p} \end{pmatrix}, \quad U^{f,u,y} \underline{\underline{\Delta}} \begin{pmatrix} U^{f} \\ U^{++f} \\ U^{f,y} \\ U^{+f} \odot U^{c} \end{pmatrix}$$

$U^{r}, Y^{p}, Y^{c}, Y^{f}, Y^{r}, W^{c}, W^{f}, W^{r}, V^{c}, V^{f}, V^{r}, U^{+c}$, $U^{+f}, U^{+r}, U^{++c}, U^{++f}, U^{++r}$ and $U^{r,u,y}$ can be defined similarly . Finally, we denote by u_p the space spanned by all the rows of the matrix U_p . That is,

$$u_p := span\{\alpha * U_p, \alpha \in R^{km}\}$$

$u_c, u_f, u_r, y_p, y_c, y_f, y_r, u^{p}, y^{p}, u^{f}, y^{f}, u^{p,u,y}$, $u^{f,u,y}$ etc are defined similarly.

Analysis

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

$$X_{t+1} = AX_t + NU_t \mathrm{O} X_t + BU_t + W_t$$
$$Y_t = CX_t + DU_t + \overset{0}{V_t}$$

...(2)

Lemma 2. For $j \geq 0$, and the block size k, we have

$$X_{k-1+j/j} = \begin{pmatrix} X_j \\ U^{+}_{k-1+j/j} \odot X_j \end{pmatrix}$$

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

$$(FG \otimes HJ) = (F \otimes H)(G \otimes J)$$
$$(FG \odot HJ) = (F \otimes H)(G \odot J)$$

Lemma 4. (Input-Output Equation). For the system (1) and $j \geq 0$, we have the following Input-Output Equation:

$$X_{k+j} = \Delta^{X}_{k} X_{k-1+j/j} + \Delta^{U}_{k} U_{k-1+j/j} + \Delta^{W}_{k} W_{k-1+j/j}$$

$$Y_{k-1+j/j} = L^{X}_{k} X_{k-1+j/j} + L^{U}_{k} U_{k-1+j/j} + L^{W}_{k} W_{k-1+j/j} + L^{V}_{k} V_{k-1+j/j}$$

where

$\Delta^{X}_{n} \underline{\underline{\Delta}} \left[A\Delta^{X}_{n-1}, N_1 \Delta^{X}_{n-1}, \cdots, N_m \Delta^{X}_{n-1}\right]$

$\Delta^{X}_{1} \underline{\underline{\Delta}} [A, N_1, \cdots, N_m]$

$\Delta^{U}_{n} \underline{\underline{\Delta}} \left[B, A\Delta^{U}_{n-1}, N_1 \Delta^{U}_{n-1}, \cdots, N_m \Delta^{U}_{n-1}\right]$

$\Delta^{U}_{1} \underline{\underline{\Delta}} B$

$\Delta^{W}_{n} \underline{\underline{\Delta}} \left[I_{n \times n,} A\Delta^{W}_{n-1}, N_1 \Delta^{W}_{n-1}, \cdots, N_m \Delta^{W}_{n-1}\right]$

$\Delta^{W}_{1} \underline{\underline{\Delta}} I_{n \times n}$

$$L_k^X \triangleq \begin{bmatrix} C\Delta_{k-1}^X & 0 \\ L_{k-1}^X & 0 \\ 0 & L_{k-1}^X \end{bmatrix},$$

$$L_k^U \triangleq \begin{bmatrix} D & C\Delta_{k-1}^U & 0 \\ 0 & L_{k-1}^U & 0 \\ 0 & 0 & L_{k-1}^U \end{bmatrix}$$

$$L_k^W \triangleq \begin{bmatrix} 0 & C\Delta_{k-1}^W & 0 \\ 0 & L_{k-1}^W & 0 \\ 0 & 0 & L_{k-1}^W \end{bmatrix}$$

$$L_k^V \triangleq \begin{bmatrix} I_{l\times l} & 0 & 0 \\ 0 & L_{k-1}^V & 0 \\ 0 & 0 & L_{k-1}^V \end{bmatrix}$$

with

$$L_1^X \triangleq [C, 0_{l\times m}],\ L_1^U \triangleq D,\ L_1^W \triangleq 0_{l\times n},\ L_1^V \triangleq I_{l\times l}$$

Lemma 5. For system (1) , if

$$\lambda = \max_{j=0,\cdots,N} \left| eig\left(A + \sum_{i=1}^{n} u_{j,i} N_i \right) \right| \langle 1, \quad \ldots(3)$$

then

$$X_c = \mathrm{E}(Y_c - DU_c - V_c) + (I - \mathrm{E}C)\Delta_n^U U^p + o(\lambda^k)$$
$$X_f = \mathrm{E}(Y_f - DU_f - V_f) + (I - \mathrm{E}C)\Delta_n^U U^c + o(\lambda^k)$$
$$X_r = \mathrm{E}(Y_r - DU_r - V_r) + (I - \mathrm{E}C)\Delta_n^U U^f + o(\lambda^k)$$

where $o(\lambda^k)$ is used to denote a matrix M, such that $\|M\|_1 = o(\lambda^k)$.

Remark 2. This holds for any matrix *E* of compatible dimensions. In particular, it holds for $E = C^T$, where $CC^T = I$, and if $l \geq n$, then $I - C^T C = 0$ and the expression become exact. In the sequel, we will assume the Moore-Penrose pseudo-inverse is used.

Theorem 1. The system (1) can be written in the following form if the condition (3) holds:

$$Y^c = O_k X_c + T_k^u U^{c,u,y} + T_k^v U^{+c} \odot V_c + L_k^W W^c + L_k^V V^c + o(\lambda^k) \quad \ldots(4)$$

$$Y^f = O_k X_f + T_k^u U^{f,u,y} + T_k^v U^{+f} \odot V_f + L_k^W W^f + L_k^V V^f + o(\lambda^k) \quad \ldots(5)$$

$$Y^r = O_k X_r + T_k^u U^{r,u,y} + T_k^v U^{+r} \odot V_r + L_k^W W^r + L_k^V V^r + o(\lambda^k) \quad \ldots(6)$$

$$X_f = F_k X_c + g_k^u U^{c,u,y} + g_k^v U^{+c} \odot V_c + \Delta_k^W W^c + o(\lambda^k)$$

$$X_r = F_k X_f + g_k^u U^{f,u,y} + g_k^v U^{+f} \odot V_f + \Delta_k^W W^f + o(\lambda^k)$$

where $O_k, T_k^u, T_k^v, F_k, g_k^u$ and g_k^v are system-dependent constant matrices.

Theorem 2. If the linear part of the system (1) is observable and

$$\begin{pmatrix} Y^c \\ U^{c,u,y} \\ U^{f,u,y} \\ U^{r,u,y} \end{pmatrix} \quad \ldots(7)$$

is a full row rank matrix, denoting

$S := y^c + u^{c,u,y} + u^{f,u,y} + u^{r,u,y}$

and $R = \Pi_S y_f + u^{f,u,y}$, then

$$\Pi_R \perp \Pi_S y^r = T_k^u \Pi_R \perp U^{r,u,y} + o(\lambda^k)$$
….(8)

where Π is the orthogonal projection operator [6].

Algorithm

Step 1. Decompose Y^r into $O_k X_r$ and $T_k^u U^{r,y,u}$ using orthogonal projection .

Step 2. Computing the constant matrix via pseudo-inverse .

Step 3. Constructing matrices for SVD decomposition .

Step 4. Performing SVD decomposition and selecting model order.

Step 5. Determining the system matrices using constrained least squares .

Examples

Tow second-order bilinear systems introduced in [5] are used to see how the two algorithms work, and how it compare between them.

The two examples work in the same conditions , equal ' j ' , ' k ' ,and input signals.

Example 1. The system matrices are

$$A=\begin{bmatrix}0 & 0.5\\ -0.5 & 0\end{bmatrix}, \quad B=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix},$$

$$C=\begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix}, \quad D=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix},$$

$$N_1=\begin{bmatrix}0.3 & 0\\ 0 & 0.3\end{bmatrix}, \; N_2=\begin{bmatrix}0.2 & 0\\ 0 & 0.4\end{bmatrix}$$

Table (1) shows the eigen values of the estimated A and N using the two algorithms .

Table (1): results for different algorithms.

	True	'Three-block'	'Four-block'
A	±0.5i	0.0010±0.4920i	±0.5001i
N_1	0.3 , 0.3	0.2520 , 0.2895	0.3021 , 0.3015
N_2	0.2 ,0.4	0.3193 , 0.4545	0.1998 , 0.4009

Example 2. The system matrices are

$$A=\begin{bmatrix}0.5 & 0\\ 0 & 0.5\end{bmatrix}, \quad B=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix},$$

$$C=\begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix},$$

$$D=\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}, \quad N_1=\begin{bmatrix}0.6 & 0\\ 0 & 0.4\end{bmatrix},$$

$$N_2=\begin{bmatrix}0.2 & 0\\ 0 & 0.5\end{bmatrix}$$

Table (2) shows the eigen values of the estimated A and N using the two algorithms.

Table (2): results for different algorithms.

	True	'Three-block'	'Four-block'
A	0.5 , 0.3	0.5046 , 0.3033	0.5008 , 0.3011
N_1	0.6 , 0.4	0.5518 , 0.2864	0.6409 , 0.3972
N_2	0.5 ,0.2	0.4783 , 0.1046	0.5055 , 0.2309

Conclusions

A two different subspace algorithms for identifying the bilinear systems has been used. Its major advantage is that the system input does not have to be white.

From the results above we show that the ' Four-block ' algorithm gives accurate estimation in a comparison with the ' Three-block ' algorithm, its convergence is faster and its need small sample size.

References

1. Chen, H.; and Maciejowski, J.;1999. A new subspace identification method for bilinear systems. CUED/FINFENG/ TR .357.
2. Favoreel, W.; De Moor, B.; and Van Overschee, P.; 1997. Subspace identification of balanced deterministic bilinear systems subject to white inputs . ECC.
3. Favoreel, W.; and De Moor, B.; 1998. Subspace identification of bilinear systems . Proc., MTNS, pp:787-790.
4. Verdult, V.; and Verhaegen, M.; 1999. Subspace-based Identification of MIMO bilinear systems. Karlsruhe, Germany , ECC , 31. August-3.September.
5. Chen, H.; and Maciejowski, J.; 2000 . Subspace identification of deterministic bilinear systems. ACC 2000, Chicago , Illinois, USA.
6. Chen, H.; and Maciejowski, J.; 2000 . Subspace identification of combined deterministic-stochastic bilinear systems . Proc., IFAC Symposium on System Identification SYSID, Santa Barbara.
7. Favoreel, W.; De Moor, B.; and Van Overschee , P.; 1999 . Subspace identification of bilinear systems subject to white inputs . IEEE Trans., On AC44(6), pp: 1157-1165.

تشخيص الأنظمة الخطية الثنائية بإستخدام طريقة مصفوفات التباعد

وسام محمد جاسم*

*جامعة الأنبار – كلية الحاسوب – قسم نظم المعلومات

الخلاصة:

تم تطبيق طريقة التشخيص المسماة subspace لتشخيص الأنظمة bilinear . بإستخدام خوارزميات ذات الثلاث صفوف وأخرى ذات أربع صفوف . وفي هذه الخوارزميات لم يكن من الضروري أن تكون إشارة الدخول إلى النظام إشارة بيضاء . وأظهرت نتائج محاكاة هذه الخوارزميات أن الخوارزمية ذات الأربع صفوف أعطت تقارب سريع وتعقيدات أبعاد المصفوفات أصغر وكذلك تعقيدات الحسابات اقل بالمقارنة مع الخوارزمية ذات الثلاث صفوف .