

# A Recursive Algorithm for MIMO Stochastic Model Estimation

Juan C. Agüero\* and Graham C. Goodwin†

School of Electrical Engineering and Computer Science  
The University of Newcastle  
Newcastle, Australia.

\* e-mail: jaguero@ee.newcastle.edu.au

† e-mail: eegcg@ee.newcastle.edu.au

## Abstract

Multivariable system identification is known to be a difficult problem. In part, this is due to the fact that, in general, the likelihood function is non-convex. The most commonly used class of procedures for off-line identification of multivariable systems is the method commonly known as Sub-Space. These methods avoid the non-convexity issue by using a multi-step procedure which includes a singular value decomposition. Unfortunately, it is not easy to develop a recursive form of these Sub-space algorithms due to the singular value decomposition step. Here, we borrow ideas from the Sub-space methodologies to develop a novel recursive algorithm. We assume that the Kronecker invariants for the system are known. We also illustrate the performance of the algorithm via a simple example.

## 1 Introduction

Multiple input - Multiple output (MIMO) system identification is known to be a difficult problem. The main source of difficulty is that the likelihood function is almost always a non-convex functional, and therefore, in general, identification algorithms have to cope with the existence of local minima. In this context, Sub-space based methods have been demonstrated to be a strong alternative to maximum likelihood (ML) methods and related schemes such as prediction error methods (PEM) (See for example [11, 27, 25, 21, 26, 23, 2]). Key advantages of Sub-space methods include:

- (i) These methods are of a non-iterative nature, and under mild conditions, achieve consistent parameter estimates [5, 20, 4].
- (ii) They work directly with general multivariable state space models.

(iii) They employ numerically robust singular value decomposition (SVD) algorithms, and

(iv) They provide an estimate of the order of the system in a natural way.

However, traditional Sub-space methods also suffer from several disadvantages, namely

- (i) The methods do not directly work with closed loop data without significant extra effort [14].
- (ii) Traditional methods do not allow one to use a specific structure in the state space matrices in the model [23].
- (iii) The resulting models are over parameterized. Sub-space methods usually estimate every entry in the matrices of the state space representation, even though less parameters are necessary to represent a system.
- (iv) It is not easy to extend the results to a recursive algorithm, mainly due to the SVD step.

Our goal here is to present a novel recursive algorithm that is inspired by Sub-space procedures. Other recursive Sub-space methods have been described elsewhere [16, 19, 17, 18]. Our procedure is simple and allows us to estimate fewer parameters than is usual in Sub-space methods.

## 2 Review of Sub-Space methods

In Sub-space methods the system is typically modelled via a linear state space description of the following form:

$$x_{t+1} = Ax_t + Bu_t + w'_t \quad (1)$$

$$y_t = Cx_t + Du_t + v_t \quad (2)$$

where  $u_t \in R^{n_u}$ ,  $x_t \in R^n$ , and  $y_t \in R^{n_y}$  are the input, state, and the output of the system respectively, and  $w_t' \in R^n$ , and  $v_t \in R^{n_v}$  are the state and measurement noise. The latter are assumed to be white with correlation matrix given by

$$E \left\{ \begin{bmatrix} w_k' \\ v_k \end{bmatrix} \begin{bmatrix} w_l' \\ v_l \end{bmatrix}^T \right\} = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \delta_{k,l} \quad (3)$$

where  $\delta_{k,l}$  denotes the Kronecker delta. The pair  $[A, C]$  is assumed to be observable.

Under mild conditions (see for example [1, 3, 8]), this model can also be written in innovations form as follows:

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + Kw_t \\ y_t &= Cx_t + Du_t + w_t \end{aligned} \quad (4)$$

Notice that if the state,  $x_t$  were available, then the system could be identified by a straightforward procedure. However, this is rarely the case. Thus, a possible procedure is to estimate the states first and then estimate the state matrices. This is the core ingredient in most Sub-space methods. Specifically, Sub-space methods incorporate several steps:

- calculate an oblique projection, which provides output predictions based on past input ( $\{u_t\}$ ) and output ( $\{y_t\}$ ) sequences.
- use a **SVD** to extract the extended observability matrix and a state sequence for the predictions.
- estimate the matrices  $A$ ,  $B$ ,  $C$ ,  $D$ , and  $K$  using different methods (several variants of the Sub-space method utilize different methodologies to obtain these matrices).

These steps will be briefly explained below. A more detailed description can be found, for example, in a recent survey paper [2].

Traditional Sub-space algorithms (**CVA**, **MOESP**, **N4SID**) can be described via the following two steps [22]:

- (i) Estimate the system order, the extended observability matrix, and the state from output predictions.
- (ii) Estimate the state space matrices using the data and the estimates obtained in the first step.

In Sub-space methods it is common to define the fol-

lowing data matrices:

$$\mathcal{D} = \left[ \begin{array}{c|c} \begin{matrix} u_0 & u_1 & \dots & u_{j-1} \\ u_1 & u_2 & \dots & u_j \\ \vdots & \vdots & \ddots & \vdots \\ u_{i-1} & u_i & \dots & u_{i+j-2} \\ \hline u_i & u_{i+1} & \dots & u_{i+j-1} \\ \vdots & \vdots & \ddots & \vdots \\ u_{2i-1} & u_{2i} & \dots & u_{2i+j-2} \end{matrix} & \begin{matrix} y_0 & y_1 & \dots & y_{j-1} \\ y_1 & y_2 & \dots & y_j \\ \vdots & \vdots & \ddots & \vdots \\ y_{i-1} & y_i & \dots & y_{i+j-2} \\ \hline y_i & y_{i+1} & \dots & y_{i+j-1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{2i-1} & y_{2i} & \dots & y_{2i+j-2} \end{matrix} \end{array} \right] = \quad (5)$$

Note that all these data matrices have a Hankel structure.

It is straightforward to show, using (4), that the output  $y_t$  can be written as

$$\begin{aligned} y_t &= y_t^p + y_t^f + v_t \\ y_t^p &= CA^{d-1}x_{t-d+1} \\ y_t^f &= Du_t + \sum_{k=1}^{d-1} CA^{k-1}Bu_{t-k} \\ v_t &= w_t + \sum_{k=1}^{d-1} CA^{k-1}Kw_{t-k} = F_d(q^{-1})w_t \end{aligned} \quad (6)$$

where  $x_{t-d+1}$  summarizes all of the information in the input and output signals up to time  $t-d$ .  $y_t^p$  is the natural response which describes the effect of the initial conditions, and  $y_t^f$  is the forced response which describes the effect of the input over the interval  $[t-d+1, t-1]$ . We can also see that the sequence  $\{v_t\}$  is not white and its variance increases with the prediction horizon “d”.

These steps in the usual Sub-space methods are then:

## 2.1 First Step: Obtain the extended observability matrix and states

The first step in Sub-space identification is essentially to obtain the extended observability matrix from the data.

Let  $\mathcal{F}_t$  denote the  $\sigma$ -algebra generated by  $\{y_0, y_1, \dots, y_t\}$ , and assume that  $\{u_t\}$  is a known sequence. Also assume that

$$E\{w_t | \mathcal{F}_{t-1}\} = 0, \quad E\{w_t w_t^T | \mathcal{F}_{t-1}\} = \Omega \quad (7)$$

This implies that  $\{w_t\}$  is a martingale difference process of constant (conditional) variance (see for example [3]).

The best estimate of  $y_t$  given the  $\sigma$ -algebra  $\mathcal{F}_{t-d}$  is

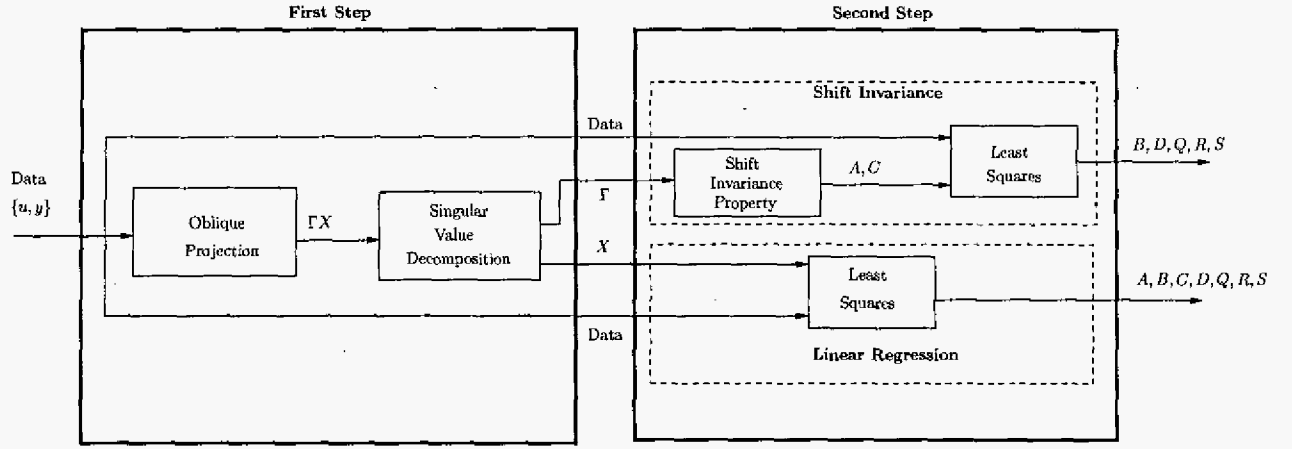


Figure 1: Typical two steps in Sub-space methods

described by the d-step ahead predictor:

$$\begin{aligned}\hat{y}_{t|t-d} &= E\{y_t | \mathcal{F}_{t-d}\} \\ &= CA^{d-1}x_{t-d+1} + Du_t + \sum_{k=1}^{d-1} CA^{k-1}Bu_{t-k}\end{aligned}\quad (8)$$

$$\Gamma = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{i-1} \end{bmatrix} \quad (12)$$

Notice that  $\hat{y}_{t|t-d}$  is, in fact, the sum of the natural and forced response of the system without the influence of noise. Thus, having an estimate of the d-step ahead prediction, it is possible to identify the parameters based on noise free data.

We next show that the matrix  $\hat{Y}_f$  which contains the different d-step ahead predictions can be expressed as a function of the extended observability matrix, namely:

$$\hat{Y}_f = \begin{bmatrix} \hat{y}_{i|i-1} & \hat{y}_{i+1|i} & \cdots & \hat{y}_{i+j-1|i+j-2} \\ \hat{y}_{i+1|i-1} & \hat{y}_{i+2|i} & \cdots & \hat{y}_{i+j|i+j-2} \\ \vdots & \vdots & & \vdots \\ \hat{y}_{2i-1|i-1} & \hat{y}_{2i|i} & \cdots & \hat{y}_{2i+j-2|i+j-2} \end{bmatrix} \quad (9)$$

$$= \Gamma X + \Lambda U_f \quad (10)$$

where we have used equation (5) in equation (10). In equation (10), we have:

$$\Lambda = \begin{bmatrix} D & 0 & 0 & \cdots & \cdots & 0 \\ CB & D & 0 & \cdots & \cdots & 0 \\ CAB & CB & D & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ CA^{i-2}B & CA^{i-3}B & CA^{i-4}B & \cdots & CB & D \end{bmatrix} \quad (11)$$

$$X = [x_i \ x_{i+1} \ x_{i+2} \ \cdots \ x_{i+j-1}] \quad (13)$$

Then we have that:

$$\Gamma X = \hat{Y}_f - \Lambda U_f \quad (14)$$

Equation (14) implies that if we eliminate the influence of the input signal,  $U_f$ , from the d-step ahead predictions,  $\hat{Y}_f$ , we have an estimation of  $\Gamma X$ . Actually this term corresponds to the part of the predicted system outputs generated by the initial state in equation (8). We rewrite this component as

$$\Gamma X = \hat{Y}_f^p = \begin{bmatrix} \hat{y}_{i|i-1}^p & \hat{y}_{i+1|i}^p & \cdots & \hat{y}_{i+j-1|i+j-2}^p \\ \hat{y}_{i+1|i-1}^p & \hat{y}_{i+2|i}^p & \cdots & \hat{y}_{i+j|i+j-2}^p \\ \vdots & \vdots & & \vdots \\ \hat{y}_{2i-1|i-1}^p & \hat{y}_{2i|i}^p & \cdots & \hat{y}_{2i+j-2|i+j-2}^p \end{bmatrix} \quad (15)$$

Note that if the system order is  $n$ , then the extended observability matrix,  $\Gamma$ , will also have rank  $n$ . Thus, we can write  $\Gamma X$ , using an SVD, as follows:

$$\begin{aligned}\Gamma X &= \hat{Y}_f - \Lambda U_f = USV^T \\ &\approx [U_1 \ U_2] \begin{bmatrix} S_1 & 0 \\ 0 & 0 \end{bmatrix} [V_1^T \ V_2^T] = U_1 S_1 V_1^T\end{aligned} \quad (16)$$

We can then associate  $\Gamma$ , and  $X$  with  $U_1$ ,  $S_1$ , and  $V_1^T$

as follows:

$$\begin{aligned}\Gamma &= U_1 S_1^{1/2} \\ X &= S_1^{1/2} V_1^T\end{aligned}\quad (17)$$

A typical variant of this step is to pre and post multiply  $\Gamma X$  by weighting matrices, and then to calculate the SVD. Different weighting matrices lead to different Sub-space methods [22].

Notice that, if we make a similarity transformation,  $x_t = T z_t$ , then the model can also be written as

$$\begin{aligned}z_{t+1} &= A' z_t + B' u_t + K' w_t \\ y_t &= C' z_t + D' u_t + w_t\end{aligned}\quad (18)$$

where

$$\begin{aligned}A' &= T^{-1} A T, B' = T^{-1} B, K' = T^{-1} K, \\ C' &= C T, D' = D\end{aligned}\quad (19)$$

and hence

$$\Gamma' X' = \Gamma X \quad (20)$$

This implies that this procedure yields *one possible realization* of the state space matrices for the system. This representation is dependent on the basis for the states implicit in the singular value decomposition step.

## 2.2 Second Step: Estimate the system parameters

Once the extended observability matrix  $\Gamma$  and the state sequence of the system  $X$  have been extracted as described in 2.1, different approaches can be followed to obtain the state space matrices, as represented in Figure 1. Two procedures that are commonly employed are:

**Using shift invariance properties:** This method basically exploits the structure in the extended observability matrix  $\Gamma$ . The matrix  $C$  corresponds to the first  $n_y$  rows, and the matrix  $A$  is obtained using ordinary least squares (OLS) [24, 26]. Using MATLAB-like notation:

$$\hat{C} = \Gamma(1 : n_y, :) \quad (21)$$

$$\Gamma(1 : n_y(i-1), :) \hat{A} = \Gamma(n_y + 1 : i n_y, :) \quad (22)$$

Then, if the extended observability matrix is consistently estimated from input - output data, the matrix estimates will be also consistent [4].

Having the estimates  $\hat{A}, \hat{C}$ , the matrices  $B$  and  $D$  can also be obtained by solving a least squares problem [12] since

$$y_t = \hat{C}(qI_n - \hat{A})^{-1} B u_t + D u_t \quad (23)$$

Finally, the noise properties can be estimated by using the residuals:

$$K \epsilon_t = x_{t+1} - \hat{A} x_t - \hat{B} u_t \quad (24)$$

$$\epsilon_t = y_t - \hat{C} x_t - \hat{D} u_t \quad (25)$$

**Using a linear regression:** An alternative way to estimate the state space matrices is to first obtain the estimate for the state sequence of the system, and then to solve the least square problem:

$$\begin{bmatrix} \hat{x}_{t+1} \\ y_t \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \hat{x}_t \\ u_t \end{bmatrix} + E \quad (26)$$

Again, the noise properties are also estimated using the residuals.

## 3 Discussion

One of the difficulties inherent in Sub-space identification is that we do not have the "true" model, and the d-step ahead predictions cannot be exactly found. Moreover, we have to obtain  $\hat{y}_{t|t-d}^p$  from the input - output data. This problem is solved by estimating a high order ARX model (see [12] and the references therein) where we assume that the state  $x_{t-d}$  in equation (6) is a function of past inputs and outputs. The number of terms needed in the ARX models is a function of the decay rate of  $\bar{A} = A - KC$ . Indeed, for the system (4) the state is given by

$$x_t = \bar{A}^t x_0 + \sum_{j=0}^{t-1} \bar{A}^j [K y_{t-j-1} + (B - KD) u_{t-j-1}] \quad (27)$$

and we see that  $\|\bar{A}^j s_t\| \leq \|\bar{A}^j\| \times \|s_t\|$ .

It is important to note (see for example [6]) that the estimate obtained by using least squares is the Best Linear Unbiased Estimate (BLUE) if the noise  $\nu_t$  is white. However, this is not the case when the noise is not white as in equation (6). It is also important that the correlation  $r_\nu(\tau)$  of the noise  $\nu_t$  is not zero for  $\tau < d$ , which means that this signal will be more correlated as the prediction horizon  $d$  grows. This fact is important in closed loop identification.

In [13] Sub-space algorithms are described by using a d-step ahead predictor interpretation. It is established that if we estimate the d-step ahead predictions, parameter estimates can be obtained.

In [13] it was demonstrated that when the true system is an ARMAX model ( $A(q^{-1})y_t = B(q^{-1})u_t + C(q^{-1})w_t$ ), then d-step ahead predictions  $\hat{y}_{t|t-d}^p$  and

$\hat{y}_{t|t-d}^f$  satisfy the following equations:

$$\begin{aligned} A(q^{-1})\hat{y}_{t|t-d}^p &= 0 \\ A(q^{-1})\hat{y}_{t|t-d}^f &= B(q^{-1})u_t \end{aligned} \quad (28)$$

where  $q^{-1}\hat{y}_{t|t-d} = \hat{y}_{t-1|t-d}$ . In [13] it was also demonstrated that when the  $d$ -step ahead predictions were determined by using an **ARX** model of any order then these equations are also satisfied.

In the next section we will analyze the case when a particular state space basis (the observability canonical form for multivariable systems) is used. We will obtain equations similar to (28) with the number of parameters corresponding to the canonical form.

To illustrate the key ideas we will use a 2 input, 2 output case. However, the generalization to  $n_u$  input and  $n_y$  outputs contains no additional conceptual issues.

#### 4 Observability Canonical Form

A potential problem with the standard Sub-space methods is that the basis for the state space is set by the **SVD** step. This means that the matrices in the corresponding state space model have no particular structure and hence we need to assume full matrices. This means that  $A, B, C, D, K$  have respectively  $n^2, nn_u, nn_y, n_y n_u, nn_y$  elements. However, it is well known that fewer parameters are actually needed to describe an  $n$ -th order system.

An alternative to the traditional subspace approach would be to introduce *structure* in the state space form to be obtained. Under the assumption that the model (1)-(2) is completely observable, the matrices  $A$  and  $C$  can be written without loss of generality as [6, 10, 7]:

$$A = \left[ \begin{array}{ccc|ccc} 0 & & I_{n_1} & 0 & \dots & 0 \\ * & \dots & * & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & & I_{n_2} \\ * & \dots & * & * & \dots & * \end{array} \right] \quad (29)$$

$$C = \left[ \begin{array}{ccc|ccc} 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{array} \right] \quad (30)$$

where the starred lines represents non zero coefficients. The sum of the *Kronecker invariants* (or *observability indices*)  $\gamma_i = n_i + 1$  gives the system order:

$$\sum_{i=1}^{n_y} \gamma_i = \sum_{i=1}^{n_y} (n_i + 1) = n \quad (31)$$

**Claim 1** The extended observability matrix  $\Gamma$  in the observability canonical form has the following struc-

ture:

$$\Gamma = \left[ \begin{array}{ccc|ccc} 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & & & \vdots & & & \\ a'_1 & a'_2 & \dots & a'_{\gamma_1} & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & 0 & \dots & 1 & \dots & 0 & 0 \\ \vdots & & & \vdots & & & & & & \\ a_1 & a_2 & \dots & a_{\gamma_1+1} & \dots & a_n \\ \vdots & & & \vdots & & & & & & \end{array} \right] \quad (32)$$

**Proof:** We use the form of the matrices as in (29). Then, the extended observability matrix is given by

$$\Gamma = \left[ \begin{array}{c} C \\ CA \\ \vdots \\ CA^{n-1} \\ \vdots \end{array} \right] \quad (33)$$

The result is obtained by straightforward calculation. ■

**Claim 2** The Markov parameters for the system (4) have the following structure:

$$\begin{aligned} CB &= \left[ \begin{array}{ccc} b_{1,1} & \dots & b_{1,n_u} \\ b_{\gamma_1+1,1} & \dots & b_{\gamma_1+1,n_u} \end{array} \right] \\ CAB &= \left[ \begin{array}{ccc} b_{2,1} & \dots & b_{2,n_u} \\ b_{\gamma_1+2,1} & \dots & b_{\gamma_1+2,n_u} \end{array} \right] \\ CA^2B &= \left[ \begin{array}{ccc} b_{3,1} & \dots & b_{3,n_u} \\ b_{\gamma_1+3,1} & \dots & b_{\gamma_1+3,n_u} \end{array} \right] \\ &\vdots \\ CA^k B &= \left[ \begin{array}{ccc} b_{k,1} & \dots & b_{k,n_u} \\ b_{\gamma_1+k,1} & \dots & b_{\gamma_1+k,n_u} \end{array} \right] \\ &\vdots \end{aligned}$$

$$\begin{aligned} CA^{\gamma_1-1}B &= \left[ \begin{array}{ccc} b_{\gamma_1,1} & \dots & b_{\gamma_1,n_u} \\ b_{2\gamma_1,1} & \dots & b_{2\gamma_1,n_u} \end{array} \right] \\ CA^{\gamma_1}B &= \left[ \begin{array}{ccc} * & \dots & * \\ b_{2\gamma_1+1,1} & \dots & b_{2\gamma_1+1,n_u} \end{array} \right] \\ &\vdots \end{aligned}$$

$$CA^{\gamma_1+k}B = \left[ \begin{array}{ccc} * & \dots & * \\ b_{2\gamma_1+k+1,1} & \dots & b_{2\gamma_1+k+1,n_u} \end{array} \right] \quad (34)$$

**Proof:** We take observability matrix in (32) and multiply on the right side by the matrix  $B$ . ■

We next show that the special structures for the matrices  $\Gamma$ , and the Markov parameters found above allow us to find a simple way to estimate the parameters of the matrices  $A$ , and  $B$  in (4).

From (32) we can obtain the following:

$$x_t^1 = \hat{y}_{t|t-1}^{p,1} \quad (35)$$

$$x_t^2 = \hat{y}_{t+1|t-1}^{p,1} \quad (36)$$

$\vdots$

$$x_{t+1}^{\gamma_1} = \hat{y}_{t+\gamma_1-1|t-1}^{p,1} \quad (37)$$

$$\hat{y}_{t+\gamma_1|t-1}^{p,1} = a_1' \hat{y}_{t|t-1}^{p,1} + \dots + a_{\gamma_1}' \hat{y}_{t+\gamma_1-1|t-1}^{p,1} \quad (38)$$

and

$$x_{t+1}^{\gamma_1+1} = \hat{y}_{t|t-1}^{p,2} \quad (39)$$

$$x_t^{\gamma_1+2} = \hat{y}_{t+1|t-1}^{p,2} \quad (40)$$

$\vdots$

$$x_t^n = \hat{y}_{t+\gamma_1+1|t-1}^{p,2} \quad (41)$$

$$\begin{aligned} \hat{y}_{t+\gamma_1|t-1}^{p,2} &= a_1 \hat{y}_{t|t-1}^{p,1} + a_2 \hat{y}_{t+1|t-1}^{p,1} + a_{\gamma_1} \hat{y}_{t+\gamma_1-1|t-1}^{p,1} \\ &\quad + a_{\gamma_1+1} \hat{y}_{t|t-1}^{p,2} + \dots + a_{\gamma_1+2} \hat{y}_{t+1|t-1}^{p,2} \\ &\quad + \dots + a_n \hat{y}_{t+\gamma_1+1|t-1}^{p,2} \end{aligned} \quad (42)$$

where the super indexes 1, 2, ...,  $n$  denote the entries in the vector. We see from (35) to (42) that, if we calculate the  $d$ -step ahead predictions  $\hat{y}_{t|t-d}^p$  from a long regression procedure, then we can also estimate the matrix  $A$  from a linear regression.

Additionally, the matrix  $B$  can be calculated in a straightforward way by a linear regression since the Markov parameters are given by (34). The matrix  $D$  could be also estimated in the same way.

The previous result has important consequences:

1. The first  $n$  predicted outputs due to initial conditions in the matrix  $\hat{Y}_f^p$  corresponds to an estimate of the system state, this is:

$$\begin{bmatrix} \hat{y}_{i|i-1}^p \\ \hat{y}_{i+1|i}^p \\ \vdots \\ \hat{y}_{i+n-1|i+n-2}^p \end{bmatrix} = \hat{x}_{i|i-1}^p \quad (43)$$

2. From the identification point of view this fact is interesting because, instead of estimating  $n^2$  coefficients of the matrix  $A$ , we need only estimate  $\gamma_1 + n$  coefficients. Note the reduction in parameters, consider the case in (35) to (42)  $n = 10$ ,  $\gamma_1 = 5$ ,  $n_u = 2$ ,  $n_y = 2$ . Then, in the typical Sub-space methods we have to estimate  $n^2 + nn_u + 2nn_y + n_y n_u = 164$ . On

the other hand, in our approach we estimate  $\gamma_1 + n + nn_u + nn_y + n_y n_u = 59$  parameters.

3. We have only used the first  $n$  rows of the matrix  $Y_f^p$  to estimate the parameters in the matrix  $A$ . However, it is straightforward to see that it is possible to obtain a set of possible estimates from other rows, and then calculate an average, or weighting average. This is analogous to the pre and post weighting utilized previous to the SVD step in typical Sub-space methods.

4. Once we have estimated the matrix  $A$ , we can estimate  $B$ ,  $D$ ,  $K$  by using linear regressions.

## 5 Recursive Estimation

Due to the use of the observability canonical form our algorithm utilizes only linear regressions. These can be solved by many different methods [9]. This also means that it is straightforward to obtain a recursive form of the algorithm, just using a recursive version of a linear regression. Our procedure can be described as follows:

1. estimate the output predictions  $\hat{Y}_f^p$ , by using a long linear regression.
2. estimate a set of parameters for the matrix  $A$  from different blocks of rows in  $\hat{Y}_f^p$  using a linear regression.
3. estimate the corresponding matrices  $B$ ,  $D$ ,  $K$  via linear regression.

In all of these steps it is possible to use a recursive form for the linear regression.

## 6 Comments on Convergence

We will not give a formal analysis for the convergence of this algorithm. However, heuristically, we can say that, if we use a long enough regression in step (1), then the  $d$ -step ahead predictions will converge to the true  $d$ -step ahead predictions. Then, since the parameter estimates are a simple function of these predictions, they will also converge.

## 7 Examples

Consider the system (4) with the following matrices:

$$A_o = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -0.06 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -0.096 & 0.712 & -1.94 & 2.3 \end{bmatrix} \quad (44)$$

$$B_o = \begin{bmatrix} 0.8 & 0.7 \\ 0.16 & 0.21 \\ 0 & 0 \\ 0.24 & 0.1 \\ 0.24 & 0.13 \\ 0.1824 & 0.129 \end{bmatrix} \quad (45)$$

$$C_o = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \quad (46)$$

$$D_o = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad (47)$$

Note that the “true” matrix  $A_o$  is block diagonal. However, here we will assume that we do not have that information. Indeed, we will estimate two vectors of parameters in the matrix  $A_o$ :  $\theta_o^1 = [-0.06 \ 0.5]$ , and  $\theta_o^2 = [0 \ 0 \ -0.096 \ 0.712 \ -1.94 \ 2.3]$ , and the full matrices  $B$ , and  $K$ .

In this system the inputs and noises are chosen normally distributed with zero mean and variances  $\sigma_u^2 = 5$ , and  $\sigma_w^2 = 0.04$ . We identify the system using  $N = 6000$  data points and we choose the prediction horizon  $i = 50$ . We also compare our results with these obtained from N4SID in MATLAB with the variable “N4Horizon” as  $[i \ i \ i]$ .

For this example we have that the d-step ahead predictions satisfy:

$$\hat{y}_{t|t-1}^{p,1} = x_t^1 \quad (48)$$

$$\hat{y}_{t+1|t-1}^{p,1} = x_t^2 \quad (49)$$

$$\hat{y}_{t+2|t-1}^{p,1} = a_1' \hat{y}_{t|t-1}^{p,1} + a_2' \hat{y}_{t+1|t-1}^{p,1} \quad (50)$$

$$\hat{y}_{t|t-1}^{p,2} = x_t^3 \quad (51)$$

$$\hat{y}_{t+1|t-1}^{p,2} = x_t^4 \quad (52)$$

$$\hat{y}_{t+2|t-1}^{p,2} = x_t^5 \quad (53)$$

$$\hat{y}_{t+3|t-1}^{p,2} = x_t^6 \quad (54)$$

$$\begin{aligned} \hat{y}_{t+4|t-1}^{p,2} = & a_1 \hat{y}_{t|t-1}^{p,1} + a_2 \hat{y}_{t+1|t-1}^{p,1} + a_3 \hat{y}_{t|t-1}^{p,2} \\ & + a_4 \hat{y}_{t+1|t-1}^{p,2} + a_5 \hat{y}_{t+2|t-1}^{p,2} + a_6 \hat{y}_{t+3|t-1}^{p,2} \end{aligned} \quad (55)$$

and we can obtain a recursive estimate for  $A_o$ , using the numerically robust square root algorithm [6, 15].

Fixing the basis of the state space as the observability canonical form, we can obtain a consistent estimate for the matrix  $B_o$  using recursive least squares, and exploiting the structure of the Markov parameters.

We first use off-line OLS to implement the procedure using 6000 data points. Figure 2 captures the frequency response for the  $2 \times 2$  system obtained for N4SID, the new algorithm with that of the true system. We note that the results are qualitatively the same.

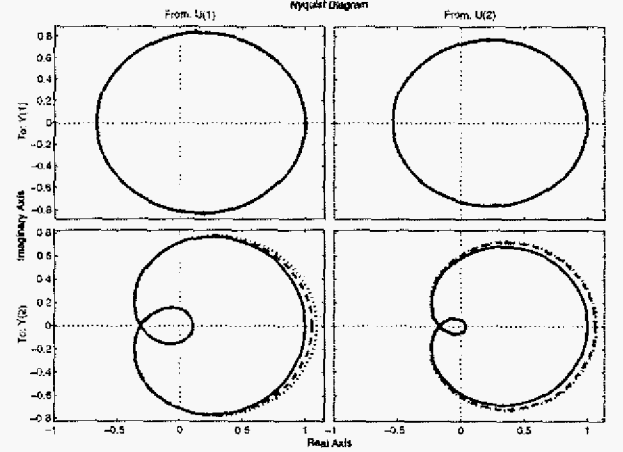


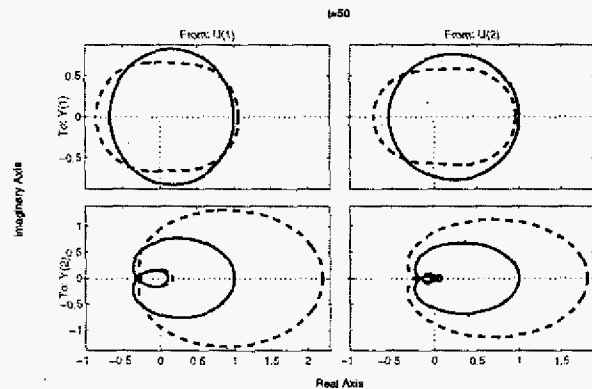
Figure 2: Nyquist plot of the true system (red, continuous line), N4SID (blue, dot line), and the batch form of the algorithm (green, dash line).

We next run the recursive form of the algorithm. Figure 3, 4, and 5 show the resulting frequency response extracted from the estimated parameters after 50, 500, 5000 steps respectively. These plots also show the true frequency response. We see that after 500 steps, the estimated frequency response is very close to the “true” frequency response, whilst after 5000 steps the estimated frequency response is essentially perfect.

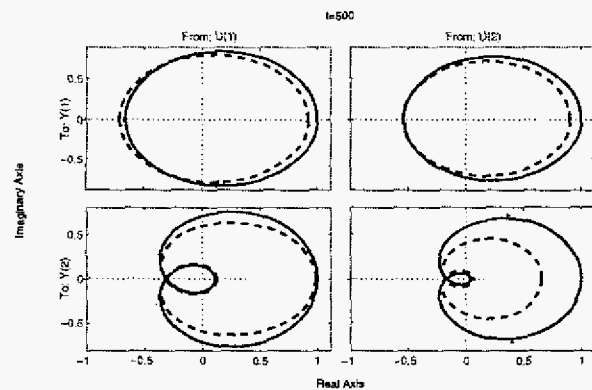
## 8 Conclusions

We have described a simple scheme for recursive estimation of parameters in MIMO stochastic systems. The novel algorithm borrows ideas from Sub-space identification algorithms, specifically the idea of an oblique projection to obtain the d-step ahead predictions. We believe that because of the nature of this algorithm, it will converge globally, which is an important feature with respects to typical recursive algorithms found in the literature. One drawback of this algorithm is the fact that we need to know the Kronecker invariants for the system. However, these parameters can be determined from previous off line identification procedures. It is also possible to run parallel algorithms assuming different Kronecker invariants

parameters, and use the one which performs better in validation. We believe that the same ideas explained in this paper could be used for systems with structures, different from the canonical form, thus allowing structured Sub-space identification methods to be developed.



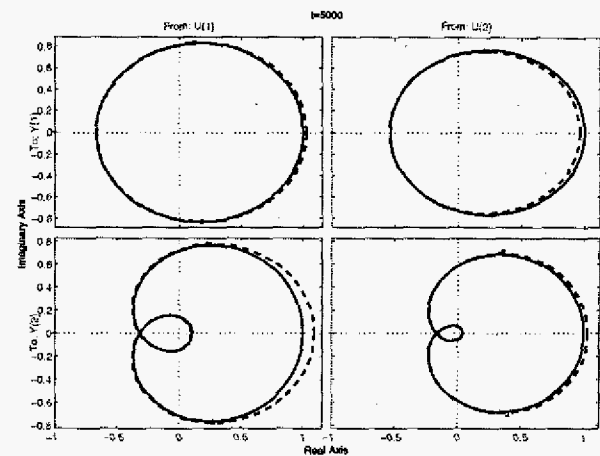
**Figure 3:** Nyquist plot of the true system (red, continuous line), and the recursive form of the algorithm (blue, dash line).



**Figure 4:** Nyquist plot of the true system (red, continuous line), and the recursive form of the algorithm (blue, dash line).

## References

- [1] B. Anderson and J. Moore. *Optimal Filtering*. Prentice Hall, Englewood Cliffs, N.J., 1979.
- [2] D. Bauer. Subspace algorithms. *13th IFAC Symposium on System Identification, Rotterdam, The Netherlands*, pages 1030–1041, 2003.
- [3] P. Caines. *Linear Stochastic Systems*. John Wiley & Sons, 1988.
- [4] C. T. Chou and M. Verhaegen. Subspace algorithms for the identification of multivariable dynamic errors in variables models. *Automatica*, 33(10):1857–1869, 1997.
- [5] M. Deistler, K. Peterzell, and W. Scherrer. Consistency and relative efficiency of subspace methods. *Automatica*, Vol 31(12):1865–1875, 1995.
- [6] G. C. Goodwin and R. Payne. *Dynamic System Identification: Experiment design and data analysis*. Academic Press, 1977.
- [7] G. C. Goodwin and K. S. Sin. *Adaptive Filtering Prediction and Control*. Prentice Hall, 1984.
- [8] R. Johansson, M. Verhaegen, C. T. Chou, and A. Robertsson. Residual models and stochastic realization in state-space identification. *International Journal of Control*, 74(10):988–995, 2001.
- [9] I. T. Jolliffe. *Principal component analysis*. Springer, second edition, 2002.
- [10] T. Kailath. *Linear Systems*. Prentice Hall, USA, 1980.
- [11] W. Larimore. Canonical variate analysis in identification, filtering, and adaptive control. *Proceedings of the 29th conference of decision and control*, pages 596–604, 1990.
- [12] L. Ljung. *System Identification: Theory for the user*. Prentice Hall, 2nd. edition, 1999.
- [13] L. Ljung and T. McKelvey. A least squares interpretation of sub-space methods for system identification. *Proceedings of the 35th conference of decision and control*, 1996.
- [14] L. Ljung and T. McKelvey. Subspace identification from closed loop data. *Signal Processing*, Vol 52:209–215, 1996.
- [15] L. Ljung and T. Söderström. *Theory and Practice of Recursive Identification*. MIT press, 1987.
- [16] M. Lovera, T. Gustafsson, and M. Verhaegen. Recursive subspace identification of linear and non-



**Figure 5:** Nyquist plot of the true system (red, continuous line), and the recursive form of the algorithm (blue, dash line).



- linear wiener state-space models. *Automatica*, Vol. 36(11):1639–1650, 2000.
- [17] G. Mercere, S. Lecoeuche, and C. Vasseur. A new recursive method for subspace identification of noisy systems: Eivpm. *13th IFAC Symposium on System Identification, Rotterdam, The Netherlands*, pages 1637–1642, 2003.
  - [18] A. Ohsumi, Y. Matsuura, and K. Kameyama. Recursive subspace identification for continuous - discrete time stochastic systems. *13th IFAC Symposium on System Identification, Rotterdam, The Netherlands*, pages 911–916, 2003.
  - [19] H. Oku and Kimura H. Recursive 4SID algorithms using gradient type subspace tracking. *Automatica*, 38:1035–1043, 2002.
  - [20] K. Peterzell, W. Scherrer, and M. Deistler. Statistical analysis of novel subspace identification methods. *Signal Processing*, Vol 52:161–177, 1996.
  - [21] P. Van Overschee and B. De Moor. N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30:75–93, 1994.
  - [22] P. Van Overschee and B. De Moor. A unifying theorem for three subspace system identification algorithms. *Automatica*, 31(12):1853–1864, 1995.
  - [23] P. Van Overschee and B. De Moor. *Subspace Identification for Linear Systems: Theory-Implementation-Applications*. Kluwer Academic Publishers, 1996.
  - [24] M. Verhaegen. Identification of the deterministic part of mimo state space models given in innovations form from input-output data. *Automatica*, 30(1):61–74, 1994.
  - [25] M. Viberg. Subspace methods in system identification. *10th IFAC Symposium on System Identification, Copenhagen, Denmark*, 1994.
  - [26] M. Viberg. Subspace base methods for the identification of linear time invariant systems. *Automatica*, 21(12):1835–1851, 1995.
  - [27] M. Viberg, B. Ottersten, B. Wahlberg, and L. Ljung. Performance of subspace state space system identification methods. *12th IFAC World congress, Sydney, Australia*, 7:369–372, 1993.