A Unifying Theorem for three Subspace System Identification Algorithms[1]

Peter Van Overschee[2] Bart De Moor[3]
ESAT - Katholieke Universiteit Leuven

Kardinaal Mercierlaan 94
3001 Leuven (Heverlee), Belgium
tel 32/16/22 09 31, fax 32/16/22 18 55
peter.vanoverschee@esat.kuleuven.ac.be
bart.demoor@esat.kuleuven.ac.be

Abstract

The aim of this paper is to indicate the striking similarities between three different subspace algorithms for the identification of combined deterministic-stochastic systems. The similarities between these algorithms have been obscured, due to different notations and backgrounds. It is shown that all three algorithms are special cases of one unifying Theorem. The comparison also indicates that the three algorithms use exactly the same subspace to determine the order and the extended observability matrix, but the weighting of the space is different in the three cases.

1. Introduction

In the literature, a number of algorithms to identify multi-input multi-output (MIMO) combined deterministic-stochastic systems have been published. Contrary to "classical" algorithms [3, 4], these subspace algorithms do not suffer from the problems caused by a-priori parametrizations and non-linear optimizations. They identify MIMO systems in a very simple and elegant way. In this paper, we will indicate and explore some striking similarities between 3 different subspace algorithms for the identification of combined deterministic-stochastic systems. This comparison is done through the introduction of a unifying Theorem, of which all three published algorithms are a special case. We believe that this observation will contribute considerably to a further understanding of subspace algorithms for system identification.

We consider three different algorithms in this paper. The first one is the algorithm due to Larimore [2]. It is based on statistical arguments and makes extensive use of principal angles and directions. The method is often referred to as the "**C**anonical **V**ariate **A**nalysis" (**CVA**). The second algorithm we will consider is the **MOESP** algorithm by Verhaegen [8]. **MOESP** stands for "**M**ultivariable **O**utput-**E**rror **S**tate s**P**ace". The third algorithm is the **N4SID** algorithm by Van Overschee & De Moor [7], which is also treated from a different point of view in Viberg, Ottersten, Wahlberg & Ljung [9]. **N4SID** stands for "**N**umerical algorithms for **S**ubspace **S**tate **S**pace **S**ystem **ID**entification" and should be read as a Californian license plate : *enforce it*. The last algorithms [7, 8, 9] are based on geometrical and linear algebra concepts.

The identification problem is the following : let $u_k \in \mathbb{R}^m, y_k \in \mathbb{R}^l$ be the observed input and output generated by the unknown system :

$$x_{k+1} = Ax_k + Bu_k + w_k \quad (1)$$
$$y_k = Cx_k + Du_k + v_k \quad (2)$$

[1] The following text presents research results obtained within the framework of the Belgian programme on interuniversity attraction poles (IUAP-17) initiated by the Belgian State - Prime Minister's Office - Science Policy Programming. The scientific responsibility is assumed by its authors.

[2] Research assistant of the Belgian National Fund for Scientific Research.

[3] Senior research associate of the Belgian National Fund for Scientific Research.

with

$$\mathbf{E}[\begin{pmatrix} w_k \\ v_k \end{pmatrix} \begin{pmatrix} w_l^t & v_l^t \end{pmatrix}] = \begin{pmatrix} Q & S \\ S^t & R \end{pmatrix} \delta_{kl} \geq 0^1 \quad (3)$$

and $A, Q \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times m}, C \in \mathbb{R}^{l \times n}, D \in \mathbb{R}^{l \times m}, S \in \mathbb{R}^{n \times l}$ and $R \in \mathbb{R}^{l \times l}$. $v_k \in \mathbb{R}^l$ and $w_k \in \mathbb{R}^n$ are unobserved, Gaussian distributed, zero mean, white noise vector sequences. $\{A, C\}$ is assumed to be observable, while $\{A, [\, B \;\; Q^{1/2} \,]\}$ is assumed to be controllable.

The main problem is then stated as : Given input and output measurements $u_1, \ldots, u_N$ and $y_1, \ldots, y_N$. where N is large and sometimes (e. g. for statistical analysis reasons) is required to go to infinity ($N \to \infty$). Given the fact that these two sequences (u_k and y_k) are generated by an unknown combined deterministic-stochastic model of the form described by equations (1)-(3). Find A, B, C, D, Q, R, S up to within a similarity transformation.

Subspace algorithms basically consist of two steps (see Figure 1). As a first step, the algorithm computes a certain characteristic subspace from the given input-output data, which coincides with the subspace generated by the columns of the extended observability matrix of the system (Γ_i). The dimension of this subspace is equal to n, the order of the system to be identified. Thus, in a first step, the order and the extended observability matrix of the system are determined, directly from the given input-output data (full lines in Figure 1).

For the second step, the algorithms apply two different strategies as illustrated in Figure 1. For more details see also [6].

Since all algorithms determine in a first step the order and the extended observability matrix, we will focus our attention to that problem in this paper. We introduce a unifying Theorem that allows the determination of the order and the extended observability matrix from input-output data. The basic subspace in this Theorem is obtained by an oblique projection, as was already mentioned in [7, 9]. The only difference between the algorithms lies in the use of different weighting matrices.

This paper is organized as follows : In Section 2 we introduce some notation. Section 3 presents the unifying Theorem, which is the main result of this paper. In Section 4, it is shown how the 3 algorithms fit in this unifying theory. Section 5 summarizes the main results of this paper.

[1] $\mathbf{E}$ denotes the expected value operator and δ_{kl} the Kronecker delta.

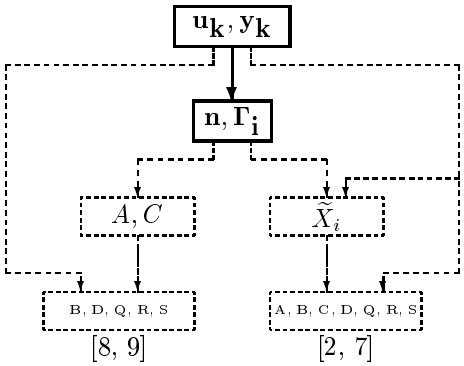

Figure 1: The two steps of a subspace algorithm. The full lines represent the first step : the determination of the order (n) and the extended observability matrix (Γ_i) from the input-output data u_k, y_k. This first step is analyzed in the paper. The dotted lines represent the two possibilities for the second step. The left hand side illustrates the strategy followed in [8, 9], while the right hand side illustrates the strategy of [2, 7].

2. Notation

In this section we introduce the notation used for input and output block Hankel matrices, for projections and for some matrix operations.

Input and output block Hankel matrices are defined as :

$$U_{0|i-1} \stackrel{\text{def}}{=} \begin{pmatrix} u_0 & u_1 & u_2 & \ldots & u_{j-1} \\ u_1 & u_2 & u_3 & \ldots & u_j \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ u_{i-1} & u_i & u_{i+1} & \ldots & u_{i+j-2} \end{pmatrix}$$

$$Y_{0|i-1} \stackrel{\text{def}}{=} \begin{pmatrix} y_0 & y_1 & y_2 & \ldots & y_{j-1} \\ y_1 & y_2 & y_3 & \ldots & y_j \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ y_{i-1} & y_i & y_{i+1} & \ldots & y_{i+j-2} \end{pmatrix}$$

where we assume that $j \to \infty$ throughout the paper. i is a user-defined index which is "large

enough". The subscripts of U and Y denote the subscript of the first and last element of the first column.

We denote the "past" input block Hankel matrix with $U_{0|i-1}$ and the "future" input block Hankel matrix with $U_{i|2i-1}$. A similar notation applies for the past and future output block Hankel matrices. Furthermore, for notational convenience and following [2], we define the matrices p, f and u as : **the past inputs and outputs** $p \stackrel{\text{def}}{=} \begin{pmatrix} U_{0|i-1} \\ Y_{0|i-1} \end{pmatrix}$, **the future outputs** $f \stackrel{\text{def}}{=} \begin{pmatrix} Y_{i|2i-1} \end{pmatrix}$, **the future inputs** $u \stackrel{\text{def}}{=} \begin{pmatrix} U_{i|2i-1} \end{pmatrix}$.

The extended ($i > n$) observability matrix Γ_i (where the subscript i denotes the number of block rows) is defined as :

$$\Gamma_i \stackrel{\text{def}}{=} \begin{pmatrix} C \\ CA \\ \dots \\ CA^{i-1} \end{pmatrix}$$

The Kalman filter state sequence $\widetilde{X}_i$ is defined as in [7] :

$$\widetilde{X}_i \stackrel{\text{def}}{=} \begin{pmatrix} \widetilde{x}_i & \widetilde{x}_{i+1} & \widetilde{x}_{i+2} & \dots & \widetilde{x}_{i+j-1} \end{pmatrix}$$

Each column is the output of a non-steady state Kalman filter built from the matrices of the system (if they were known). The j columns of $\tilde{X}_i$ are thus the outputs of a bank of j Kalman filters in parallel. When the system matrices A, B, C, D, Q, R, S would be known, this sequence could be determined easily by combining the input-output data with the known system matrices. The point is that this state sequence can also be obtained directly from the input-output data, without any knowledge of the system matrices. This observation is at the heart of the approach elaborated on in [7] to which we refer for more details.

Π_A denotes the operator that projects the row space of a matrix onto the row space of A (which is assumed to be of full row rank) : $\Pi_A \stackrel{\text{def}}{=} A^t(AA^t)^{-1}A$. $A^\perp$ denotes the subspace perpendicular to the row space of A. B/A is short hand for the projection of the row space of B onto the row space of A : $B/A \stackrel{\text{def}}{=} B\Pi_A = BA^t(AA^t)^{-1}A$. The Moore-Penrose inverse of a matrix A is denoted by $A^\dagger$ and the $n \times n$ identity matrix by I_n.

3. A unifying framework

In this section we present a unifying framework for the determination of the order and the extended observability matrix of a system, directly from given input-output data. In a first subsection, we will stress the intuition that leads to the Theorem. In a second subsection, we state the Theorem. In section 4 we show that the three aforementioned algorithms are special cases of the unifying Theorem.

3.1. Intuition

The goal of an identification procedure is to find a model that behaves in approximately the same way as the process under consideration. In the framework of subspace identification, we attain this goal by solving two subsequent problems :

Optimal Prediction : The optimal prediction problem can be restated as : predict the future outputs (f) as accurately as possible, using all the information that can be obtained from the past (p), and using the knowledge of the inputs that will be presented to the system in the future (u).
Inspired by the linearity of the system, we propose to combine the past (p) and the future inputs (u) linearly to predict the future outputs (f). We denote the linear combinations respectively with L_p and L_u. The quality of the prediction is measured in the Frobenius norm. Mathematically, the first part of the identification goal thus becomes :

$$\min_{\substack{L_p \in \mathbb{R}^{li \times (m+l)i} \\ L_u \in \mathbb{R}^{li \times mi}}} \| f - \begin{pmatrix} L_p & L_u \end{pmatrix} \begin{pmatrix} p \\ u \end{pmatrix} \|_F^2 \tag{4}$$

The optimal combination of the past (p) to predict the future is thus $L_p p$. Geometrically, the row space of $L_p p$ can be interpreted as the **oblique projection** of the row space of f along the row space of u on the row space of p. This **oblique projection** is denoted with $\mathcal{O} \in \mathbb{R}^{li \times j}$:

$$\mathcal{O} \stackrel{\text{def}}{=} L_p p$$

Complexity Reduction : Apart from the fact that we want to find a model that can pre-

dict the future, we also want the complexity of this model to be as low as possible. We thus need to reduce the complexity of $\mathcal{O}$. Since the rows of $\mathcal{O}$ span an li dimensional subspace in the j dimensional ambient space, we can introduce a complexity reduction by reducing the subspace dimension to n (the order of the system). Intuitively, this implies that we only have to remember n different directions of the past to predict the future. Mathematically, the second step can be formulated :

$$\min_{\mathcal{R} \in \mathbb{R}^{li \times j}} \|W_1 \left[\mathcal{O} - \mathcal{R}\right] W_2\|_F^2 \qquad (5)$$

$$\text{constrained to : } \operatorname{rank}(\mathcal{R}) = n$$

where $W_1 \in \mathbb{R}^{li \times li}$ and $W_2 \in \mathbb{R}^{j \times j}$ are user defined weighting matrices. These weighting matrices determine which part of the "information" of $\mathcal{O}$ is important to retain. Since we do not want to loose any "information" (row rank) due to the weighting, we should make sure that : $\operatorname{rank}(W_1 \mathcal{O} W_2) = \operatorname{rank}(\mathcal{O})$.

3.2. A unifying Theorem

Theorem 1 *Under the assumption that :*

1. *The process noise w_k and the measurement noise v_k are not identically zero.*
2. *The input u_k is uncorrelated with the process noise w_k and the measurement noise v_k.*
3. *The input u_k is persistently exciting of order $2i$ [3, pp 363].*
4. *There is an infinite amount of measurements available : $j \to \infty$.*
5. *W_1 is of full rank and W_2 is such that $rank(p) = rank(pW_2)$*

Then :

$$\mathcal{O} = [\ (f/u^\perp)\ (p/u^\perp)^t\ ][\ (p/u^\perp)\ (p/u^\perp)^t\ ]^{-1} p$$

and with the singular value decomposition

$$W_1 \mathcal{O} W_2 = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^t \\ V_2^t \end{pmatrix} \qquad (6)$$

We have :

1. *The order of the system (1)-(3) is equal to the number of singular values in equation (6) different from zero.*
2. *The optimal reduction $\mathcal{R}$ can be taken equal to[2] :*

$$\mathcal{R} = W_1^{-1} U_1 S_1 V_1^t W_2^\dagger$$

3. *The extended observability matrix Γ_i can be taken equal to :*

$$\Gamma_i = W_1^{-1} U_1 S_1^{1/2}$$

4. *The part of the Kalman state sequence $\widetilde{X}_i$ that lies in the column space of W_2 can be recovered from :*

$$\widetilde{X}_i W_2 = S_1^{1/2} V_1^t \qquad (7)$$

5. *The "full" Kalman state sequence $\widetilde{X}_i$ can be recovered from :*

$$\widetilde{X}_i = \Gamma_i^\dagger \mathcal{O} \qquad (8)$$

The proof of the unifying Theorem and related discussions can be found in [6].

4. Three algorithms and the unifying Theorem

*In this section we show how the three subspace algorithms (***N4SID***, ***MOESP*** and ***CVA***) determine the system order and the matrix Γ_i from the given input-output data as stated in Theorem 1, but with different weighting matrices W_1 and W_2 for each algorithm. The proofs of the Theorems can be found in [6].*

4.1. N4SID

Theorem 2 N4SID

The algorithms of [7, 9] correspond to Theorem 1 with the weights :

$$\begin{aligned} W_1 &= I \\ W_2 &= I \end{aligned}$$

A consequence of $W_2 = I$ is that we do not need (8) to determine the states $\widetilde{X}_i$, but that they can be determined simply from the singular value decomposition (using (7)). In the next subsections,

[2] This is the minimum norm solution of the optimization problem 5. When W_2 is rank deficient there is more than one solution.

it will become clear that for the other two algorithms, the determination of the state sequence $\widetilde{X}_i$ requires the use of formula (8).

In [9] an interpretation of this algorithm is given in an instrumental variable framework. In [7] the connection with non-steady state Kalman filters is elaborated on (see also Section 2).

4.2. MOESP

Theorem 3 MOESP

The algorithm of [8] corresponds to Theorem 1 with the following weights :

$$\begin{aligned} W_1 &= I \\ W_2 &= \Pi_{u^\perp} \end{aligned}$$

Note that, since W_2 is not of full rank, we do not recover the full state from the singular value decomposition. According to formula (7), we only recover the projection of the state :

$$\widetilde{X}_i W_2 = \widetilde{X}_i \Pi_{u^\perp}$$

The state could be determined through formula (8). However, since **MOESP** does not use state sequences, we will not elaborate on this any further.

4.3. CVA

Theorem 4 CVA

The algorithm of [2] corresponds to Theorem 1 with the following weights :

$$\begin{aligned} W_1 &= [(f/u^\perp)(f/u^\perp)^t]^{-1/2} \\ W_2 &= \Pi_{u^\perp} \end{aligned}$$

From Theorem 4 we conclude that applying the weighting matrices of Theorem 4 to the results of Theorem 1 leads to a principal direction analysis between the past inputs and outputs orthogonalized to the future inputs $(p/u^\perp)$ and the future outputs orthogonalized to the future inputs $(f/u^\perp)$. The same principal angles are calculated as in [2].

In [6] it is proven that the principal directions lying in the past $(p/u^\perp)$ correspond to the projected states of Formula (7). This proves formally that the principal directions are no states, but projected states.

The "full" state sequence is given by equation (8) of Theorem 1. First Γ_i is determined from the left singular vectors (and the weights of Theorem 4). Then, we find with (8) :

$$\widetilde{X}_i^c = \Gamma_i^\dagger \mathcal{O}$$

In [6], it is proven that $\widetilde{X}_i^c$ is exactly equal to the "memory" used by Larimore [2].

5. Conclusion

In this paper we have introduced a unifying Theorem that allows the extraction of the system order and the extended observability matrix from given input-output data. The Theorem succeeds in unifying three algorithms described in the literature.

References

[1] Kung S.Y. (1978). *A New Identification and Model Reduction Algorithm via Singular Value Decomposition.* Proceedings of the 12th Asilomar Conference on Circuits, Systems and Computers, Pacific Grove, pp 705-714.

[2] Larimore W.E. (1990). *Canonical Variate Analysis in Identification, Filtering, and Adaptive Control.* Proceedings of the 29th Conference on Decision and Control, Hawaii, pp 596-604.

[3] Ljung L. (1987). *System Identification : theory for the user.* Prentice Hall Information and System Sciences Series.

[4] Söderström T., Stoica P. (1989). *System identification.* Prentice-Hall International Series in Systems and Control Engineering.

[5] Van Overschee P., De Moor B. (1993). *Subspace Algorithms for the Stochastic Identification Problem*, Automatica, Vol. **29**, No. 3, pp 649-660.

[6] Van Overschee P., De Moor B. (1993). *A Unifying Theorem for three Subspace System Identification Algorithms*, ESAT-SISTA report No. 1993-50, Kath. Universiteit Leuven, Dept. E.E., Belgium. Submitted for publication.

[7] Van Overschee P., De Moor B. (1994). **N4SID** *: Subspace Algorithms for the Identification of Combined deterministic-Stochastic Systems.* Automatica (Special Issue on Statistical Signal Processing and Control), Vol **30**, No. 1, pp. 75-93.

[8] Verhaegen M. (1993). *Identification of the deterministic part of MIMO state space models given in innovations form from input-output data.* Internal report of Technical University Delft, Department of Electrical Engineering, Network Theory Section, The Netherlands.

[9] Viberg M., Ottersten B., Wahlberg B., Ljung L. (1993). *Performance of Subspace Based State Space System Identification Methods.* Proc. of the 12th IFAC World Congress, Sydney, Australia, 18-23 July, Vol **7**, pp 369-372.