SUBSPACE IDENTIFICATION METHODS

Katrien De Cock, Bart De Moor, K.U.Leuven, Department of Electrical Engineering (ESAT – SCD), Kasteelpark Arenberg 10, B–3001 Leuven, Belgium, tel: +32-16-321709, fax: +32-16-321970, http://www.esat.kuleuven.ac.be/sista-cosic-docarch/, {decock,demoor}@esat.kuleuven.ac.be

Keywords: Systems, Discrete time systems, System identification, Subspace identification, Linear systems, Kalman filter, Numerical linear algebra, Orthogonal projection, Oblique projection, State space models, Time invariant systems, Multivariable models, Deterministic models, Stochastic models, Row space, State sequence, Singular value decomposition, QR factorization, LQ factorization, Least squares, Block Hankel matrix, Principal angles, Software, Positive realness, Stability, Bilinear systems, N4SID, MOESP, CVA

Contents

Glossary

CACSD: Computer Aided Control System Design
CVA: Canonical Variate Analysis
Deterministic systems: Systems of which the input and output signals are known exactly. There is no process nor observation noise.
Discrete time systems: Systems described in the discrete time domain
Eigenvalue: An eigenvalue of a matrix A is a root z of the characteristic equation: $\det(zI - A) = 0$
MIMO: Multiple-input-multiple-output
MOESP: Multivariable Output-Error State sPace

N4SID: Numerical algorithms for Subspace State Space System IDentification
Noise: A disturbing or driving signal that is not measured or observed
Row space: The vector space that contains all linear combinations of the rows of a matrix
SISO: Single-input-single-output
State space: The vector space of states of a linear system
State vector: Vector whose elements are the state variables of a dynamical system
Stochastic systems: Systems of which only the output signal is observed. The process and observation noise are assumed to be white sequences.
Subspace: A subset of vectors of an ambient vector space that is closed under vector addition and multiplication by a scalar
SVD: Singular value decomposition
System identification: The discipline of making mathematical models of systems, starting from experimental data
Time invariant systems: Dynamical systems whose properties are time invariant. The parameters of the model of a time-invariant system are constants.

Summary

This paper gives a short introduction to and survey of subspace identification algorithms. Deterministic, stochastic and combined deterministic-stochastic subspace identification algorithms are treated. These methods estimate state sequences directly from the given data, either explicitly or implicitly, through an orthogonal or oblique projection of the row spaces of certain block Hankel matrices of data into the row spaces of other block Hankel matrices, followed by a singular value decomposition (SVD) to determine the order, the observability matrix and /or the state sequence. The extraction of the state space model is then achieved through the solution of a least squares problem. Each of these steps can be elegantly implemented using well-known numerical linear algebra algorithms such as the singular value decomposition and the QR decomposition.

1. Introduction

This Section contains a description of the central ideas of this paper. First, in Section 1.1, we describe state space models, *which is the type of models that is delivered by subspace identification algorithms. In Section 1.2 we explain how subspace identification algorithms work.*

1.1. State space models

Models in this paper are lumped, discrete time, linear, time-invariant, state space models. From the number of epithets used, this might seem like a highly restricted class of models (especially the fact they are linear), but, surprisingly enough, many industrial processes can be described very accurately by this type of models, especially locally in the neighborhood of a working point. Moreover, there is a large number of control system design tools available to build controllers for such systems and models.

Mathematically, these models are described by the following set of difference equations:

$$\left\{ \begin{array}{rcl} x_{k+1} & = & Ax_k + Bu_k + w_k \,, \\ y_k & = & Cx_k + Du_k + v_k \,, \end{array} \right. \tag{1}$$

with[1]

$$\mathbf{E}[\begin{pmatrix} w_p \\ v_p \end{pmatrix} \begin{pmatrix} w_q^T & v_q^T \end{pmatrix}] = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \delta_{pq} \geq 0 \,. \tag{2}$$

In this model, we have

– **vectors:** The vectors $u_k \in \mathbb{R}^m$ and $y_k \in \mathbb{R}^l$ are the observations at time instant k of respectively the m inputs and l outputs of the process. The vector $x_k \in \mathbb{R}^n$ is the state vector of the process at discrete time instant k and contains the numerical values of n states. $v_k \in \mathbb{R}^l$ and $w_k \in \mathbb{R}^n$ are unobserved vector signals, usually called the measurement, respectively process noise. It is assumed that they are zero mean, stationary, white noise vector sequences. (The Kronecker delta in (2) means $\delta_{pq} = 0$ if $p \neq q$, and $\delta_{pq} = 1$ if $p = q$.) The effect of the process w_k is different from that of v_k: w_k as an input will have a dynamic effect on the state x_k and output y_k, while v_k only affects the output y_k directly and therefore is called a measurement noise.

– **matrices:** $A \in \mathbb{R}^{n \times n}$ is called the (dynamical) system matrix. It describes the dynamics of the system (as characterized by its eigenvalues). $B \in \mathbb{R}^{n \times m}$ is the input matrix, which represents the linear transformation by which the deterministic inputs influence the next state. $C \in \mathbb{R}^{l \times n}$ is the output matrix, which describes how the internal state is transferred to the outside world in the observations y_k. The term with the matrix $D \in \mathbb{R}^{l \times m}$ is called the direct feedthrough term. The matrices $Q \in \mathbb{R}^{n \times n}$, $S \in \mathbb{R}^{n \times l}$ and $R \in \mathbb{R}^{l \times l}$ are the covariance matrices of the noise sequences w_k and v_k. The block matrix in (2) is assumed to be positive definite, as is indicated by the inequality sign. The matrix pair $\{A, C\}$ is assumed to be observable, which implies that all *modes* in the system can be observed in the output y_k and can thus be identified. The matrix pair $\{A, [\, B \;\; Q^{1/2} \,]\}$ is assumed to be controllable, which in its turn implies that all *modes* of the system can be excited by either the deterministic input u_k and/or the stochastic input w_k.

A graphical representation of the system can be found in Figure 1.
We are now ready to state the main mathematical problem of this paper.

> Given s consecutive input and output observations $u_0, \ldots, u_{s-1}$, and $y_0, \ldots, y_{s-1}$. Find an appropriate order n and the system matrices A, B, C, D, Q, R, S.

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

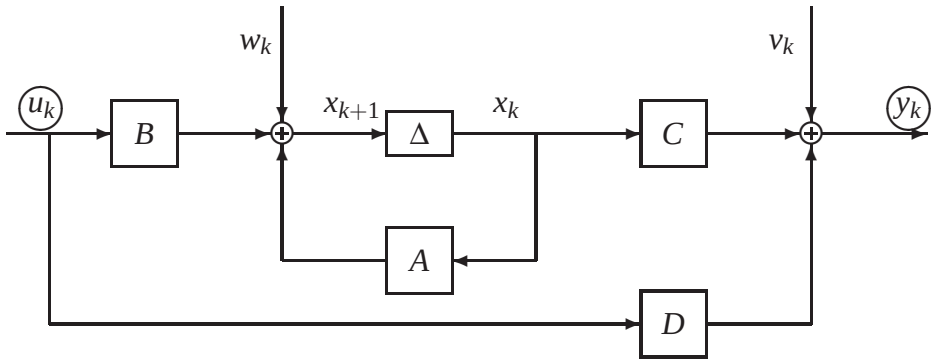

Figure 1: The (circled) vector signals u_k and y_k are available (observed) while v_k, w_k are unknown disturbances. The symbol Δ represents a delay. Note the inherent feedback via the matrix A (which represents the dynamics). Sensor or actuator dynamics are completely contained in A too. It is assumed that u_k is available without measurement noise.

1.2. The basic idea behind subspace identification algorithms

The goal of this Section is to provide a verbal description of the main principles on which subspace identification algorithms are based. The mathematical derivations will be elaborated on in the next sections.

Subspace identification algorithms are based on concepts from system theory, (numerical) linear algebra and statistics. The main concepts in subspace identification algorithms are

1. The *state sequence of the dynamical system* is determined first, directly from input/output observations, without knowing the model. That this is possible for the model class (1) is one of the main contributions of subspace algorithms, as compared to "classical" approaches that are based on an input-output framework. The difference is illustrated in Figure 2. So an important achievement of the research in subspace identification was to demonstrate how the Kalman filter states can be obtained directly from input-output data using linear algebra tools (QR and singular value decomposition) without knowing the mathematical model. An important consequence is that, once these states are known, the identification problem becomes a linear least squares problem in the unknown system matrices, and the process and measurement noise covariance matrices follow from the least squares residuals, as is easy to see from Equations (1):

$$\underbrace{\begin{pmatrix} x_{i+1} & x_{i+2} & \cdots & x_{i+j} \\ y_i & y_{i+1} & \cdots & y_{i+j-1} \end{pmatrix}}_{\text{known}} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \underbrace{\begin{pmatrix} x_i & x_{i+1} & \cdots & x_{i+j-1} \\ u_i & u_{i+1} & \cdots & u_{i+j-1} \end{pmatrix}}_{\text{known}} + \begin{pmatrix} w_i & w_{i+1} & \cdots & w_{i+j-1} \\ v_i & v_{i+1} & \cdots & v_{i+j-1} \end{pmatrix}. \tag{3}$$

   The meaning of the parameters i and j will become clear henceforth.
   Even though the state sequence can be determined explicitly, in most variants and implementations, this is not done explicitly but rather implicitly. Said in other words, the set of linear equations above can be solved 'implicitly' as will become clear below, without an explicit calculation of the state sequence itself. Of course, when needed, the state sequence can be computed explicitly.

   The two main steps that are taken in subspace algorithms are the following.

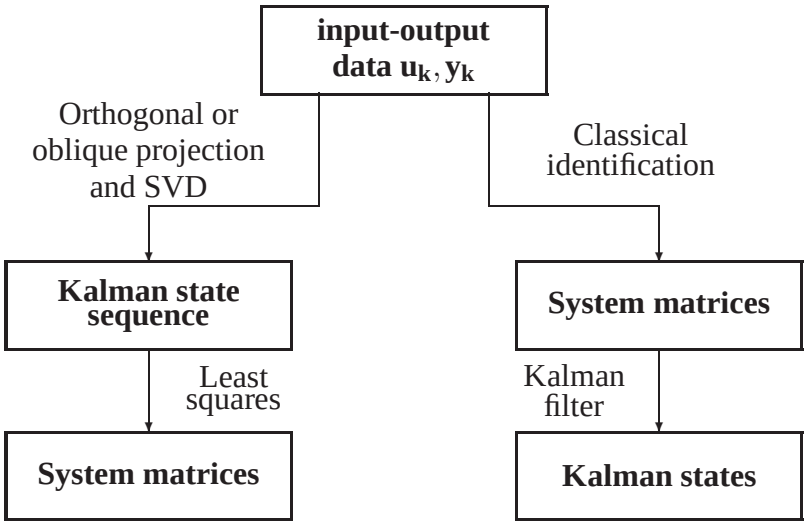

Figure 2: Subspace identification aims at constructing state space models from input-output data. The left hand side shows the subspace identification approach: first the (Kalman filter) states are estimated directly (either implicitly or explicitly) from input-output data, then the system matrices can be obtained. The right hand side is the classical approach: first obtain the system matrices, then estimate the states.

(a) Determine the model order n and a state sequence $\hat{x}_i, \hat{x}_{i+1}, \ldots, \hat{x}_{i+j}$ (estimates are denoted by a $\hat{\cdot}$). They are typically found by first projecting row spaces of data block Hankel matrices, and then applying a singular value decomposition (see Sections 4, 5, 6).

(b) Solve a least squares problem to obtain the state space matrices:

$$\begin{pmatrix} \widehat{A} & \widehat{B} \\ \widehat{C} & \widehat{D} \end{pmatrix} = \min_{A,B,C,D} \left\| \begin{pmatrix} \hat{x}_{i+1} & \hat{x}_{i+2} & \cdots & \hat{x}_{i+j} \\ y_i & y_{i+1} & \cdots & y_{i+j-1} \end{pmatrix} - \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \hat{x}_i & \hat{x}_{i+1} & \cdots & \hat{x}_{i+j-1} \\ u_i & u_{i+1} & \cdots & u_{i+j-1} \end{pmatrix} \right\|_F^2, \tag{4}$$

where $\|\cdot\|_F$ denotes the Frobenius-norm of a matrix. The estimates of the noise covariance matrices follow from

$$\begin{pmatrix} \widehat{Q} & \widehat{S} \\ \widehat{S}^T & \widehat{R} \end{pmatrix} = \frac{1}{j} \begin{pmatrix} \rho_{w_i} & \rho_{w_{i+1}} & \cdots & \rho_{w_{i+j-1}} \\ \rho_{v_i} & \rho_{v_{i+1}} & \cdots & \rho_{v_{i+j-1}} \end{pmatrix} \begin{pmatrix} \rho_{w_i} & \rho_{w_{i+1}} & \cdots & \rho_{w_{i+j-1}} \\ \rho_{v_i} & \rho_{v_{i+1}} & \cdots & \rho_{v_{i+j-1}} \end{pmatrix}^T, \tag{5}$$

where $\rho_{w_k} = \hat{x}_{k+1} - \widehat{A}\hat{x}_k - \widehat{B}u_k$ and $\rho_{v_k} = y_k - \widehat{C}\hat{x}_k - \widehat{D}u_k$ $(k = i, \ldots, i+j-1)$ are the least squares residuals.

2. Subspace system identification algorithms make full use of the well developed body of *concepts and algorithms from numerical linear algebra*. Numerical robustness is guaranteed because of the well-understood algorithms, such as the QR-decomposition, the singular value decomposition and its generalizations. Therefore, they are very well suited for large data sets ($s \to \infty$) and large scale systems (m, l, n large). Moreover, subspace algorithms are not iterative. Hence, there are no *convergence* problems. When carefully implemented, they are computationally very efficient, especially for large datasets (implementation details are however not contained in this survey).

3. The conceptual straightforwardness of subspace identification algorithms translates into *user-friendly software implementations*. To give only one example: since there is no explicit need

for parameterizations in the geometric framework of subspace identification, the user is not confronted with highly technical and theoretical issues such as canonical parameterizations. The number of user choices is greatly reduced when using subspace algorithms because we use full state space models and the only parameter to be specified by the user, is the order of the system, which can be determined by inspection of certain singular values.

2. Notation

In this section, we set some notation. In Section 2.1, we introduce the notation for the data block Hankel matrices and in Section 2.2 for the system related matrices.

2.1. Block Hankel matrices and state sequences

Block Hankel matrices with output and/or input data play an important role in subspace identification algorithms. These matrices can be easily constructed from the given input-output data. Input block Hankel matrices are defined as

$$U_{0|2i-1} \overset{\text{def}}{=} \left(\begin{array}{ccccc} u_0 & u_1 & u_2 & \cdots & u_{j-1} \\ u_1 & u_2 & u_3 & \cdots & u_j \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ u_{i-1} & u_i & u_{i+1} & \cdots & u_{i+j-2} \\ \hline u_i & u_{i+1} & u_{i+2} & \cdots & u_{i+j-1} \\ u_{i+1} & u_{i+2} & u_{i+3} & \cdots & u_{i+j} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ u_{2i-1} & u_{2i} & u_{2i+1} & \cdots & u_{2i+j-2} \end{array}\right) = \left(\frac{U_{0|i-1}}{U_{i|2i-1}}\right) = \left(\frac{U_p}{U_f}\right) \tag{6}$$

$$= \left(\begin{array}{ccccc} u_0 & u_1 & u_2 & \cdots & u_{j-1} \\ u_1 & u_2 & u_3 & \cdots & u_j \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ u_{i-1} & u_i & u_{i+1} & \cdots & u_{i+j-2} \\ u_i & u_{i+1} & u_{i+2} & \cdots & u_{i+j-1} \\ \hline u_{i+1} & u_{i+2} & u_{i+3} & \cdots & u_{i+j} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ u_{2i-1} & u_{2i} & u_{2i+1} & \cdots & u_{2i+j-2} \end{array}\right) = \left(\frac{U_{0|i}}{U_{i+1|2i-1}}\right) = \left(\frac{U_p^+}{U_f^-}\right) \tag{7}$$

where:

- The number of block rows (i) is a user-defined index which is large enough, i.e. it should at least be larger than the maximum order of the system one wants to identify. Note that, since each block row contains m (number of inputs) rows, the matrix $U_{0|2i-1}$ consists of $2mi$ rows.
- The number of columns (j) is typically equal to $s-2i+1$, which implies that all s available data samples are used. In any case, j should be larger than $2i-1$. Throughout the paper, for

statistical reasons, we will often assume that $j, s \to \infty$. For deterministic (noiseless) models, i.e. where $v_k \equiv 0$ and $w_k \equiv 0$, this will however not be needed.

- The subscripts of $U_{0|2i-1}, U_{0|i-1}, U_{0|i}, U_{i|2i-1}, etc...$ denote the subscript of the first and last element of the first column in the block Hankel matrix. The subscript "p" stands for "past" and the subscript "f" for "future". The matrices U_p (the past inputs) and U_f (the future inputs) are defined by splitting $U_{0|2i-1}$ in two equal parts of i block rows. The matrices U_p^+ and U_f^- on the other hand are defined by shifting the border between past and future one block row down[2]. They are defined as $U_p^+ = U_{0|i}$ and $U_f^- = U_{i+1|2i-1}$.

The output block Hankel matrices $Y_{0|2i-1}, Y_p, Y_f, Y_p^+, Y_f^-$ are defined in a similar way.
State sequences play an important role in the derivation and interpretation of subspace identification algorithms. The state sequence X_i is defined as:

$$X_i \stackrel{\text{def}}{=} \begin{pmatrix} x_i & x_{i+1} & \dots & x_{i+j-2} & x_{i+j-1} \end{pmatrix} \in \mathbb{R}^{n \times j}, \tag{8}$$

where the subscript i denotes the subscript of the first element of the state sequence.

2.2. Model matrices

Subspace identification algorithms make extensive use of the observability and of its structure. The extended ($i > n$) observability matrix Γ_i (where the subscript i denotes the number of block rows) is defined as:

$$\Gamma_i \stackrel{\text{def}}{=} \begin{pmatrix} C \\ CA \\ CA^2 \\ \dots \\ CA^{i-1} \end{pmatrix} \in \mathbb{R}^{li \times n}. \tag{9}$$

We assume the pair $\{A, C\}$ to be observable, which implies that the rank of Γ_i is equal to n.

3. Geometric Tools

In Sections 3.1 through 3.2 we introduce the main geometric tools used to reveal some system characteristics. They are described from a linear algebra point of view, independently of the subspace identification framework we will be using in the next sections.

In the following sections we assume that the matrices $A \in \mathbb{R}^{p \times j}, B \in \mathbb{R}^{q \times j}$ and $C \in \mathbb{R}^{r \times j}$ are given (they are dummy matrices in this section). We also assume that $j \geq \max(p, q, r)$, which will always be the case in the identification algorithms.

[2]The superscript "$+$" stands for "add one block row" while the superscript "$-$" stands for "delete one block row".

3.1. Orthogonal projections

The orthogonal projection of the row space of A into the row space of B is denoted by A/B and its matrix representation is

$$A/B \stackrel{\text{def}}{=} AB^T(BB^T)^\dagger B\,, \tag{10}$$

where $\bullet^\dagger$ denotes the Moore-Penrose pseudo-inverse of the matrix $\bullet$. $A/B^\perp$ is the projection of the row space of A into $B^\perp$, the orthogonal complement of the row space of B, for which we have $A/B^\perp = A - A/B = A(I_j - B(BB^T)^\dagger B)$. The projections Π_B and $\Pi_{B^\perp}$ decompose a matrix A into two matrices, the row spaces of which are orthogonal:

$$A = A\Pi_B + A\Pi_{B^\perp}\,. \tag{11}$$

The matrix representations of these projections can be easily computed via the LQ decomposition of $\begin{pmatrix} B \\ A \end{pmatrix}$, which is the numerical matrix version of the Gram–Schmidt orthogonalization procedure. Let A and B be matrices of full row rank and let the LQ decomposition of $\begin{pmatrix} B \\ A \end{pmatrix}$ be denoted by

$$\begin{pmatrix} B \\ A \end{pmatrix} = LQ^T = \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} Q_1^T \\ Q_2^T \end{pmatrix}\,, \tag{12}$$

where $L \in \mathbb{R}^{(p+q)\times(p+q)}$ is lower triangular, with $L_{11} \in \mathbb{R}^{q\times q}$, $L_{21} \in \mathbb{R}^{p\times q}$, $L_{22} \in \mathbb{R}^{p\times p}$ and $Q \in \mathbb{R}^{j\times(p+q)}$ is orthogonal, i.e. $Q^TQ = \begin{pmatrix} Q_1^T \\ Q_2^T \end{pmatrix} \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} = \begin{pmatrix} I_q & 0 \\ 0 & I_p \end{pmatrix}$. Then, the matrix representations of the orthogonal projections can be written as

$$A/B = L_{21}Q_1^T\,, \tag{13}$$
$$A/B^\perp = L_{22}Q_2^T\,. \tag{14}$$

3.2. Oblique projections

Instead of decomposing the rows of A as in (11) as a linear combination of the rows of two orthogonal matrices (Π_B and $\Pi_{B^\perp}$), they can also be decomposed as a linear combination of the rows of two non-orthogonal matrices B and C and of the orthogonal complement of B and C. This can be written as $A = L_BB + L_CC + L_{B^\perp,C^\perp}\begin{pmatrix} B \\ C \end{pmatrix}^\perp$. The matrix L_CC is defined[3] as the oblique projection of the row space of A along the row space of B into the row space of C:

$$A/_B C \stackrel{\text{def}}{=} L_CC\,. \tag{15}$$

[3] Note that L_B and L_C are only unique when B and C are of full row rank and when the intersection of the row spaces of B and C is $\{0\}$, said in other words, $\text{rank}(\begin{pmatrix} B \\ C \end{pmatrix}) = \text{rank}(B) + \text{rank}(C) = q + r$.

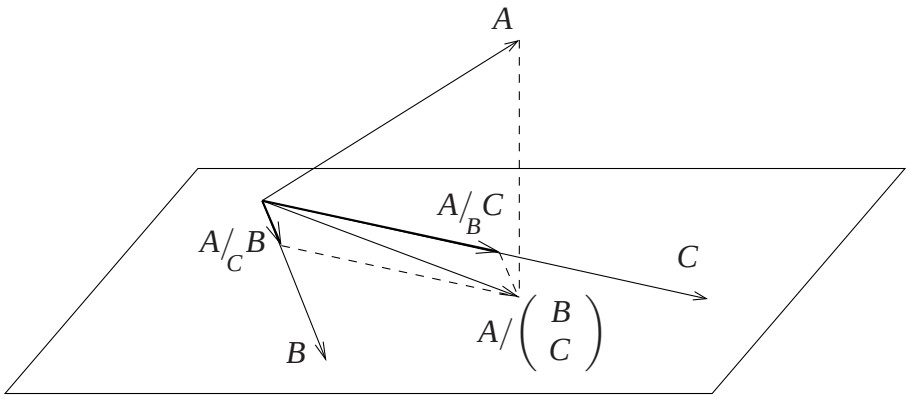

Figure 3: Interpretation of the oblique projection in the j-dimensional space ($j = 3$ in this case).

The oblique projection can also be interpreted through the following recipe: project the row space of A orthogonally into the joint row space of B and C and decompose the result along the row space of B and C. This is illustrated in Figure 3 for $j = 3$ and $p = q = r = 1$, where $A/\begin{pmatrix} B \\ C \end{pmatrix}$ denotes the orthogonal projection of the row space of A into the joint row space of B and C, $A/_B C$ is the oblique projection of A along B into C and $A/_C B$ is the oblique projection of A along C into B.

Let the LQ decomposition of $\begin{pmatrix} B \\ C \\ A \end{pmatrix}$ be given by $\begin{pmatrix} B \\ C \\ A \end{pmatrix} = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} Q_1^T \\ Q_2^T \\ Q_3^T \end{pmatrix}$. Then, the matrix representation of the orthogonal projection of the row space of A into the joint row space of B and C is equal to (see previous section):

$$A/\begin{pmatrix} B \\ C \end{pmatrix} = \begin{pmatrix} L_{31} & L_{32} \end{pmatrix} \begin{pmatrix} Q_1^T \\ Q_2^T \end{pmatrix}. \tag{16}$$

Obviously, the orthogonal projection of A into $\begin{pmatrix} B \\ C \end{pmatrix}$ can also be written as a linear combination of the rows of B and C:

$$A/\begin{pmatrix} B \\ C \end{pmatrix} = L_B B + L_C C = \begin{pmatrix} L_B & L_C \end{pmatrix} \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} Q_1^T \\ Q_2^T \end{pmatrix}. \tag{17}$$

Equating (16) and (17) leads to

$$\begin{pmatrix} L_B & L_C \end{pmatrix} \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix} = \begin{pmatrix} L_{31} & L_{32} \end{pmatrix}. \tag{18}$$

The oblique projection of the row space of A along the row space of B into the row space of C can thus be computed as

$$A/_B C = L_C C = L_{32} L_{22}^{-1} \begin{pmatrix} L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} Q_1^T \\ Q_2^T \end{pmatrix}. \tag{19}$$

Note that when $B = 0$ or when the row space of B is orthogonal to the row space of C ($BC^T = 0$) the oblique projection reduces to an orthogonal projection, in which case $A/_B C = A/C$.

4. Deterministic subspace identification

In this section, we treat subspace identification of purely time-invariant deterministic systems, with no measurement nor process noise ($v_k \equiv 0$ and $w_k \equiv 0$ in Figure 1).
As explained in Section 1.2, we first determine a state sequence (Section 4.1) and then solve a least squares problem to find A, B, C, D (Section 4.2).

4.1. Calculation of a state sequence

The state sequence of a deterministic system can be found by computing the intersection of the past input and output and the future input and output spaces. This can be seen as follows. Consider w_k and v_k in (1) to be identically 0, and derive the following matrix input-output equations:

$$Y_{0|i-1} = \Gamma_i X_i + H_i U_{0|i-1} \,, \tag{20}$$
$$Y_{i|2i-1} = \Gamma_i X_{2i} + H_i U_{i|2i-1} \,, \tag{21}$$

in which H_i is an $li \times mi$ lower block Triangular Toeplitz matrix with the so-called Markov parameters of the system:

$$H_i = \begin{pmatrix} D & 0 & 0 & \dots & 0 \\ CB & D & 0 & \dots & 0 \\ CAB & CB & D & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ CA^{i-2}B & CA^{i-3}B & \dots & \dots & D \end{pmatrix}.$$

From this we find that

$$\begin{pmatrix} Y_{0|i-1} \\ U_{0|i-1} \end{pmatrix} = \begin{pmatrix} \Gamma_i & H_i \\ 0 & I_{mi} \end{pmatrix} \begin{pmatrix} X_i \\ U_{0|i-1} \end{pmatrix} \,, \tag{22}$$

from which we get

$$\text{rank} \begin{pmatrix} Y_{0|i-1} \\ U_{0|i-1} \end{pmatrix} = \text{rank} \begin{pmatrix} X_i \\ U_{0|i-1} \end{pmatrix} .$$

Hence,

$$\text{rank} \begin{pmatrix} Y_{0|i-1} \\ U_{0|i-1} \end{pmatrix} = mi + n$$

provided that $U_{0|i-1}$ is of full row rank (we assume throughout that $j \gg mi$, that there is no intersection between the row spaces of X_i and that of $U_{0|i-1}$ and that the state sequence is of full row rank as well ('full state space excited')). These are experimental conditions that are generically satisfied and that can be considered as 'persistancy-of-excitation' requirements for subspace algorithms to work.
A similar derivation under similar conditions can be done for

$$\begin{aligned}\text{rank}\begin{pmatrix} Y_{i|2i-1} \\ U_{i|2i-1} \end{pmatrix} &= mi+n\,,\\ \text{rank}\begin{pmatrix} Y_{0|2i-1} \\ U_{0|2i-1} \end{pmatrix} &= 2mi+n\,.\end{aligned}$$

We can also relate X_{2i} to X_i as

$$X_{2i} = A^i X_i + \Delta_i^r U_{0|i-1}\,, \tag{23}$$

in which $\Delta_i^r = (\, A^{i-1}B\, A^{i-2}B\, \ldots\, AB\, B\,)$ is a reversed extended controllability matrix. Assuming that the model is observable and that $i \geq n$, we find from (21) that

$$X_{2i} = \begin{pmatrix} -\Gamma_i^\dagger H_i & \Gamma_i^\dagger \end{pmatrix} \begin{pmatrix} U_{i|2i-1} \\ Y_{i|2i-1} \end{pmatrix}\,,$$

which implies that the row space of X_{2i} is contained within the row space of $\begin{pmatrix} U_f \\ Y_f \end{pmatrix}$. But similarly, from (23) and (20) we find that

$$X_{2i} = A^i(\Gamma_i^\dagger Y_{0|i-1} - \Gamma_i^\dagger H_i U_{0|i-1}) + \Delta_i^r U_{0|i-1} = \begin{pmatrix} \Delta_i^r - A^i\Gamma_i^\dagger H_i & A^i\Gamma_i^\dagger \end{pmatrix} \begin{pmatrix} U_{0|i-1} \\ Y_{0|i-1} \end{pmatrix}\,,$$

which implies that the row space of X_{2i} is equally contained within the row space of $\begin{pmatrix} U_p \\ Y_p \end{pmatrix}$. Let's now apply Grassmann's dimension theorem (under the generic assumptions on persistency of excitation)

$$\begin{aligned}\dim\left(\text{row space}\begin{pmatrix} U_p \\ Y_p \end{pmatrix} \cap \text{row space}\begin{pmatrix} U_f \\ Y_f \end{pmatrix}\right) &= \text{rank}\begin{pmatrix} U_p \\ Y_p \end{pmatrix} + \text{rank}\begin{pmatrix} U_f \\ Y_f \end{pmatrix} - \text{rank}\begin{pmatrix} U_p \\ Y_p \\ U_f \\ Y_f \end{pmatrix} &(24)\\ &= (mi+n)+(mi+n)-(2mi+n) = n\,. &(25)\end{aligned}$$

Indeed, above we have shown that any basis for the intersection between 'past' and 'future' represents a valid state sequence X_i. The state sequence X_{i+1} can be obtained analogously. Different ways to compute the intersection have been proposed. A first way, is by making use of a singular value decomposition of a concatenated Hankel matrix $\begin{pmatrix} U_{0|2i-1} \\ Y_{0|2i-1} \end{pmatrix}$. This allows to estimate the model

order n and to calculate the linear combination of the rows of $\begin{pmatrix} U_p \\ Y_p \end{pmatrix}$, or equivalently of $\begin{pmatrix} U_f \\ Y_f \end{pmatrix}$, that generate the intersection. A second way is by taking as a basis for the intersection the principal directions between the row space of the past inputs and outputs and the row space of the future inputs and outputs. A non-empty intersection between two subspaces is characterized by a number of principal angles equal to zero, and the principal directions corresponding to these zero angles form a basis for the row space of the intersection.

4.2. Computing the system matrices

As soon as the order of the model and the state sequences X_i and X_{i+1} are known, the state space matrices A,B,C,D can be solved from

$$\underbrace{\begin{pmatrix} X_{i+1} \\ Y_{i|i} \end{pmatrix}}_{\text{known}} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \underbrace{\begin{pmatrix} X_i \\ U_{i|i} \end{pmatrix}}_{\text{known}}, \tag{26}$$

where $U_{i|i}, Y_{i|i}$ are block Hankel matrices with only one block row of inputs respectively outputs, namely $U_{i|i} = \begin{pmatrix} u_i & u_{i+1} & \cdots & u_{i+j-1} \end{pmatrix}$ and similarly for $Y_{i|i}$. This set of equations can be solved. As there is no noise, it is consistent.

5. Stochastic subspace identification

In this section, we treat subspace identification of linear time-invariant stochastic systems with no external input ($u_k \equiv 0$). The stochastic identification problem thus consists of estimating the stochastic system matrices A,C,Q,S,R from given output data only. We show how this can be done using geometric operations. In Section 5.1 we show how a state sequence can be found and in Section 5.2 the system matrices are computed.

5.1. Calculation of a state sequence

The state sequence of a stochastic model can be obtained in two steps: first, the future output space is projected orthogonally into the past output space and next, a singular value decomposition is carried out.

1. **Orthogonal projection:** As explained in Section 3.1, we will use the LQ decomposition to compute the orthogonal projection. Let $Y_{0|2i-1}$ be the $2li \times j$ output block Hankel matrix. Then, we partition the LQ decomposition of $Y_{0|2i-1}$ as follows

$$\begin{matrix} li \\ l \\ l(i-1) \end{matrix} \begin{pmatrix} Y_{0|i-1} \\ Y_{i|i} \\ Y_{i+1|2i-1} \end{pmatrix} = \begin{matrix} \begin{matrix} li & l & l(i-1) \end{matrix} \\ \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \end{matrix} \begin{matrix} j \\ \begin{pmatrix} Q_1^T \\ Q_2^T \\ Q_3^T \end{pmatrix} \end{matrix}. \tag{27}$$

We will need two projections. The orthogonal projection Y_f/Y_p of the future output space into the past output space, which is denoted by O_i, and the orthogonal projection Y_f^-/Y_p^+ of Y_f^- into Y_p^+, denoted by O_{i-1} (see Section 2.1 for the definitions of Y_p, Y_f, Y_p^+ and Y_f^-). Applying (13) leads to

$$\begin{aligned} O_i &= Y_f/Y_p &&= \begin{pmatrix} L_{21} \\ L_{31} \end{pmatrix} Q_1^T \\ O_{i-1} &= Y_f^-/Y_p^+ &&= \begin{pmatrix} L_{31} & L_{32} \end{pmatrix} \begin{pmatrix} Q_1^T \\ Q_2^T \end{pmatrix} . \end{aligned} \tag{28}$$

It can be shown that the matrix O_i is equal to the product of the extended observability matrix and a matrix $\widehat{X}_i$, which contains certain Kalman filter states (the interpretation is given in Figure 4):

$$O_i = \Gamma_i \widehat{X}_i \, , \tag{29}$$

where Γ_i is the $li \times n$ observability matrix (see (9)) and $\widehat{X}_i = \begin{pmatrix} \hat{x}_i^{[0]} & \hat{x}_i^{[1]} & \cdots & \hat{x}_i^{[j-1]} \end{pmatrix}$. Similarly, O_{i-1} is equal to

$$O_{i-1} = \Gamma_{i-1} \widehat{X}_{i+1} \, , \tag{30}$$

where $\widehat{X}_{i+1} = \begin{pmatrix} \hat{x}_{i+1}^{[0]} & \hat{x}_{i+1}^{[1]} & \cdots & \hat{x}_{i+1}^{[j-1]} \end{pmatrix}$.

2. **Singular value decomposition:** The singular value decomposition of O_i allows us to find the order of the model (the rank of O_i), and the matrices Γ_i and $\widehat{X}_i$.
Let the singular value decomposition of $\begin{pmatrix} L_{21} \\ L_{31} \end{pmatrix}$ be equal to

$$\begin{pmatrix} L_{21} \\ L_{31} \end{pmatrix} = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = U_1 S_1 V_1^T \, , \tag{31}$$

where $U_1 \in \mathbb{R}^{li \times n}$, $S_1 \in \mathbb{R}^{n \times n}$, and $V_1 \in \mathbb{R}^{li \times n}$. Then, we can choose $\Gamma_i = U_1 S_1^{1/2}$ and $\widehat{X}_i = S_1^{1/2} V_1^T Q_1^T$. This state sequence is generated by a bank of non-steady state Kalman filters working in parallel on each of the columns of the block Hankel matrix of past outputs Y_p. Figure 4 illustrates this interpretation. The j Kalman filters run in a *vertical* direction (over the columns). It should be noted that each of these j Kalman filters only uses *partial* output information. The qth Kalman filter ($q = 0, \ldots, j-1$)

$$\hat{x}_{k+1}^{[q]} = (A - K_k C) \hat{x}_k^{[q]} + K_k y_{k+q} \, , \tag{32}$$

runs over the data in the qth column of Y_p, for $k = 0, 1, \ldots, i-1$.

The "shifted" state sequence $\widehat{X}_{i+1}$, on the other hand, can be obtained as

$$\widehat{X}_{i+1} = \left(\underline{\Gamma_i}\right)^{\dagger} O_{i-1} \, , \tag{33}$$

where $\underline{\Gamma_i} = \Gamma_{i-1}$ denotes the matrix Γ_i without the last l rows, which is also equal to $\underline{U_1} S_1^{1/2}$.

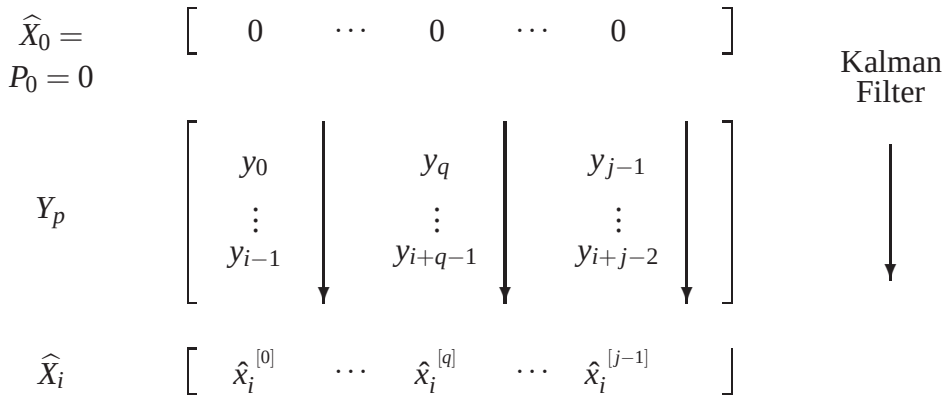

Figure 4: Interpretation of the sequence $\widehat{X}_i$ as a sequence of non-steady state Kalman filter state estimates based upon i observations of y_k. When the system matrices A,C,Q,R,S were known, the state $\hat{x}_i^{[q]}$ could be determined from a non-steady state Kalman filter as follows: Start the filter at time q, with an initial state estimate 0. Now iterate the non-steady state Kalman filter over i time steps (the vertical arrow down). The Kalman filter will then return a state estimate $\hat{x}_i^{[q]}$. This procedure could be repeated for each of the j columns, and thus we speak about a *bank* of non-steady state Kalman filters. The major observation in subspace algorithms is that the system matrices A,C,Q,R,S do not have to be known to determine the state sequence $\widehat{X}_i$. It can be determined directly from output data through geometric manipulations.

5.2. Computing the system matrices

At this moment, we have calculated $\widehat{X}_i$ and $\widehat{X}_{i+1}$, using geometrical and numerical operations on output data only. We can now form the following set of equations:

$$\underbrace{\begin{pmatrix} \widehat{X}_{i+1} \\ Y_{i|i} \end{pmatrix}}_{\text{known}} = \begin{pmatrix} A \\ C \end{pmatrix} \underbrace{\left(\widehat{X}_i \right)}_{\text{known}} + \underbrace{\begin{pmatrix} \rho_w \\ \rho_v \end{pmatrix}}_{\text{residuals}}, \tag{34}$$

where $Y_{i|i}$ is a block Hankel matrix with only one block row of outputs. This set of equations can be easily solved for A,C. Since the Kalman filter residuals ρ_w,ρ_v (the innovations) are uncorrelated with $\widehat{X}_i$, solving this set of equations in a least squares sense (since the least squares residuals are orthogonal and thus uncorrelated with the regressors $\widehat{X}_i$) results in an asymptotically (as $j \to \infty$) unbiased estimate $\widehat{A},\widehat{C}$ of A,C as $\begin{pmatrix} \widehat{A} \\ \widehat{C} \end{pmatrix} = \begin{pmatrix} \widehat{X}_{i+1} \\ Y_{i|i} \end{pmatrix} \widehat{X}_i^{\dagger}$. An estimate $\widehat{Q}_i$, $\widehat{S}_i$, $\widehat{R}_i$ of the noise covariance matrices Q,S and R can be obtained from the residuals: $\begin{pmatrix} \widehat{Q}_i & \widehat{S}_i \\ \widehat{S}_i^T & \widehat{R}_i \end{pmatrix} = \frac{1}{j} \begin{pmatrix} \rho_w \\ \rho_v \end{pmatrix} \begin{pmatrix} \rho_w^T & \rho_v^T \end{pmatrix}$, where the subscript i indicates that the estimated covariances are biased, with however an exponentially decreasing bias as $i \to \infty$.
By making the following substitutions:

$$\widehat{X}_i = \Gamma_i^{\dagger} O_i = S_1^{1/2} V_1^T Q_1^T , \tag{35}$$

$$\widehat{X}_{i+1} = \Gamma_{i-1}^{\dagger} O_{i-1} = \left(\underline{\Gamma_i}\right)^{\dagger} O_{i-1} = \left(\underline{U_1} S_1^{1/2}\right)^{\dagger} \begin{pmatrix} L_{31} & L_{32} \end{pmatrix} \begin{pmatrix} Q_1^T \\ Q_2^T \end{pmatrix} , \tag{36}$$

$$Y_{i|i} \;=\; \begin{pmatrix} L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} Q_1^T \\ Q_2^T \end{pmatrix}, \tag{37}$$

the least squares solution reduces to

$$\begin{pmatrix} \widehat{A} \\ \widehat{C} \end{pmatrix} = \begin{pmatrix} \left(\underline{U_1} S_1^{1/2}\right)^{\dagger} L_{31} \\ L_{21} \end{pmatrix} V_1 S_1^{-1/2}, \tag{38}$$

and the noise covariances are equal to

$$\begin{pmatrix} \widehat{Q}_i & \widehat{S}_i \\ \widehat{S}_i^T & \widehat{R}_i \end{pmatrix} = \frac{1}{j} \begin{pmatrix} \left(\underline{U_1} S_1^{1/2}\right)^{\dagger} L_{31} & \left(\underline{U_1} S_1^{1/2}\right)^{\dagger} L_{32} \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} I - V_1 V_1^T & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} L_{31}^T \left(S_1^{1/2} (\underline{U_1})^T\right)^{\dagger} & L_{21}^T \\ L_{32}^T \left(S_1^{1/2} (\underline{U_1})^T\right)^{\dagger} & L_{22}^T \end{pmatrix}. \tag{39}$$

Note that the Q-matrices of the LQ factorization cancel out of the least-squares solution and the noise covariances. This implies that in the first step, the Q-matrix should never be calculated explicitly. Since typically $j \gg 2mi$, this reduces the computational complexity and memory requirements significantly.

6. Combined deterministic–stochastic subspace identification algorithm

In this section, we give one variant of subspace algorithms for the identification of A, B, C, D, Q, R, S. Other variants can be found in the literature. The algorithm works in two main steps. First, the row space of a Kalman filter state sequence is obtained directly from the input-output data, without any knowledge of the system matrices. This is explained in Section 6.1. In the second step, which is given in Section 6.2, the system matrices are extracted from the state sequence via a least squares problem.

6.1. Calculation of a state sequence

The state sequence of a combined deterministic–stochastic model can again be obtained from input-output data in two steps. First, the future output row space is projected along the future input row space into the joint row space of past input and past output. A singular value decomposition is carried out to obtain the model order, the observability matrix and a state sequence, which has a very precise and specific interpretation.

1. **Oblique projection:** We will use the LQ decomposition to compute the oblique projection $Y_f /_{U_f} \begin{pmatrix} U_p \\ Y_p \end{pmatrix}$. Let $U_{0|2i-1}$ be the $2mi \times j$ and $Y_{0|2i-1}$ the $2li \times j$ block Hankel matrices of the

input and output observations. Then, we partition the LQ decomposition of $\begin{pmatrix} U \\ Y \end{pmatrix}$ as follows

$$\begin{pmatrix} U_{0|i-1} \\ U_{i|i} \\ U_{i+1|2i-1} \\ Y_{0|i-1} \\ Y_{i|i} \\ Y_{i+1|2i-1} \end{pmatrix} = \begin{pmatrix} L_{11} & 0 & 0 & 0 & 0 & 0 \\ L_{21} & L_{22} & 0 & 0 & 0 & 0 \\ L_{31} & L_{32} & L_{33} & 0 & 0 & 0 \\ L_{41} & L_{42} & L_{43} & L_{44} & 0 & 0 \\ L_{51} & L_{52} & L_{53} & L_{54} & L_{55} & 0 \\ L_{61} & L_{62} & L_{63} & L_{64} & L_{65} & L_{66} \end{pmatrix} \begin{pmatrix} Q_1^T \\ Q_2^T \\ Q_3^T \\ Q_4^T \\ Q_5^T \\ Q_6^T \end{pmatrix}. \tag{40}$$

The matrix representation of the oblique projection $Y_f /_{U_f} \begin{pmatrix} U_p \\ Y_p \end{pmatrix}$ of the future output row space along the future input row space into the joint space of past input and past output is denoted by O_i. Analogously to the derivation in Section 3.2, the oblique projection can be obtained as

$$O_i = Y_f /_{U_f} \begin{pmatrix} U_p \\ Y_p \end{pmatrix} = L_{U_p} L_{11} Q_1^T + L_{Y_p} \begin{pmatrix} L_{41} & L_{42} & L_{43} & L_{44} \end{pmatrix} \begin{pmatrix} Q_1^T \\ Q_2^T \\ Q_3^T \\ Q_4^T \end{pmatrix}, \tag{41}$$

where

$$\begin{pmatrix} L_{U_p} & L_{U_f} & L_{Y_p} \end{pmatrix} \left(\begin{array}{c|c|c} L_{11} & 0 & 0 & 0 \\ \hline L_{21} & L_{22} & 0 & 0 \\ L_{31} & L_{32} & L_{33} & 0 \\ \hline L_{41} & L_{42} & L_{43} & L_{44} \end{array}\right) = \left(\begin{array}{c|c|c} L_{51} & L_{52} & L_{53} & L_{54} \\ L_{61} & L_{62} & L_{63} & L_{64} \end{array}\right), \tag{42}$$

from which L_{U_p}, L_{U_f} and L_{Y_p} can be calculated. The oblique projection $Y_f^- /_{U_f^-} \begin{pmatrix} U_p^+ \\ Y_p^+ \end{pmatrix}$, denoted by O_{i-1}, on the other hand, is equal to

$$O_{i-1} = L_{U_p^+} \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} Q_1^T \\ Q_2^T \end{pmatrix} + L_{Y_p^+} \begin{pmatrix} L_{41} & L_{42} & L_{43} & L_{44} & 0 \\ L_{51} & L_{52} & L_{53} & L_{54} & L_{55} \end{pmatrix} \begin{pmatrix} Q_1^T \\ Q_2^T \\ Q_3^T \\ Q_4^T \\ Q_5^T \end{pmatrix}, \tag{43}$$

where

$$\begin{pmatrix} L_{U_p^+} & L_{U_f^-} & L_{Y_p^+} \end{pmatrix} \left(\begin{array}{cc|c|cc} L_{11} & 0 & 0 & 0 & 0 \\ L_{21} & L_{22} & 0 & 0 & 0 \\ \hline L_{31} & L_{32} & L_{33} & 0 & 0 \\ \hline L_{41} & L_{42} & L_{43} & L_{44} & 0 \\ L_{51} & L_{52} & L_{53} & L_{54} & L_{55} \end{array}\right) = \left(\begin{array}{cc|c|cc} L_{61} & L_{62} & L_{63} & L_{64} & L_{65} \end{array}\right). \tag{44}$$

Under the assumptions that:

(a) the process noise w_k and measurement noise v_k are uncorrelated with the input u_k,

(b) the input u_k is persistently exciting of order $2i$, i.e. the input block Hankel matrix $U_{0|2i-1}$ is of full row rank,.

(c) the sample size goes to infinity: $j \to \infty$,

(d) the process noise w_k and the measurement noise v_k are not identically zero,

one can show that the oblique projection O_i is equal to the product of the extended observability matrix Γ_i and a sequence of Kalman filter states, obtained from a bank of non-steady state Kalman filters, in essence the same as in Figure 4:

$$O_i = \Gamma_i \widetilde{X}_i \,. \tag{45}$$

Similarly, the oblique projection O_{i-1} is equal to

$$O_{i-1} = \Gamma_{i-1} \widetilde{X}_{i+1} \,. \tag{46}$$

2. **Singular value decomposition:** Let the singular value decomposition of $L_{U_p} \begin{pmatrix} L_{11} & 0 & 0 & 0 \end{pmatrix} + L_{Y_p} \begin{pmatrix} L_{41} & L_{42} & L_{43} & L_{44} \end{pmatrix}$ be equal to

$$L_{U_p} \begin{pmatrix} L_{11} & 0 & 0 & 0 \end{pmatrix} + L_{Y_p} \begin{pmatrix} L_{41} & L_{42} & L_{43} & L_{44} \end{pmatrix} = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} \tag{47}$$

$$= U_1 S_1 V_1^T \,, \tag{48}$$

Then, the order of the system (1) is equal to the number of singular values in equation (47) different from zero. The extended observability matrix Γ_i can be taken to be

$$\Gamma_i = U_1 S_1^{1/2} \,, \tag{49}$$

and the state sequence $\widetilde{X}_i$ is equal to

$$\widetilde{X}_i = \Gamma_i^{\dagger} O_i = S_1^{1/2} V_1^T \begin{pmatrix} Q_1^T \\ Q_2^T \\ Q_3^T \\ Q_4^T \end{pmatrix} \,. \tag{50}$$

The "shifted" state sequence $\widetilde{X}_{i+1}$, on the other hand, can be obtained as

$$\widetilde{X}_{i+1} = \left(\underline{\Gamma_i}\right)^{\dagger} O_{i-1} \,, \tag{51}$$

where $\underline{\Gamma_i} = \Gamma_{i-1}$ denotes the matrix Γ_i without the last l rows.

There is an important observation to be made. Corresponding columns of $\widetilde{X}_i$ and of $\widetilde{X}_{i+1}$ are state estimates of X_i and X_{i+1} respectively, obtained from the same Kalman filters at two consecutive time instants, but with different initial conditions. This is in contrast to the stochastic identification algorithm, where the initial states are equal to 0 (see Figure 4).

6.2. Computing the system matrices

From Section 6.1, we find:

- The order of the system from inspection of the singular values of equation (47).
- The extended observability matrix Γ_i from equation (49) and the matrix Γ_{i-1} as $\underline{\Gamma_i}$, where $\underline{\Gamma_i}$ denotes the matrix Γ_i without the last l rows.
- The state sequences $\widetilde{X}_i$ and $\widetilde{X}_{i+1}$.

The state space matrices A, B, C and D can now be found by solving a set of over-determined equations in a least squares sense:

$$\begin{pmatrix} \widetilde{X}_{i+1} \\ Y_{i|i} \end{pmatrix} = \begin{pmatrix} \widehat{A} & \widehat{B} \\ \widehat{C} & \widehat{D} \end{pmatrix} \begin{pmatrix} \widetilde{X}_i \\ U_{i|i} \end{pmatrix} + \begin{pmatrix} \rho_w \\ \rho_v \end{pmatrix} , \tag{52}$$

where ρ_w and ρ_v are residual matrices. The estimates of the covariances of the process and measurement noise are obtained from the residuals ρ_w and ρ_v of equation (52) as:

$$\begin{pmatrix} \widehat{Q}_i & \widehat{S}_i \\ \widehat{S}_i^T & \widehat{R}_i \end{pmatrix} = \frac{1}{j} \begin{pmatrix} \rho_w \\ \rho_v \end{pmatrix} \begin{pmatrix} \rho_w^T & \rho_v^T \end{pmatrix} , \tag{53}$$

where i again indicates that the estimated covariances are biased, with an exponentially decreasing bias as $i \to \infty$. As in the stochastic identification algorithm, the Q-matrices of the LQ factorization cancel out in the least-squares solution and the computation of the noise covariances. This implies that the Q-matrix of the LQ factorization should never be calculated explicitly. Note however, that corresponding columns of $\widetilde{X}_i$ and of $\widetilde{X}_{i+1}$ are state estimates of X_i and X_{i+1} respectively, obtained with different initial conditions. As a consequence, the set of equations (52) is not theoretically consistent, which means that the estimates of the system matrices are slightly biased. It can however be proven that the estimates of A, B, C and D are unbiased if at least one of the following conditions is satisfied:

- $i \to \infty$,
- the system is purely deterministic, i.e. $v_k = w_k = 0, \ \forall k$,
- the deterministic input u_k is white noise.

If none of the above conditions is satisfied, one obtains biased estimates. However, there exist more involved algorithms that provide consistent estimates of A, B, C and D, even if none of the above conditions is satisfied, for which we refer to the literature.

6.3. Variations

Several variants on the algorithm that was explained above, exist. First, we note that the oblique projection O_i can be weighted left and right by user defined weighting matrices $W_1 \in \mathbb{R}^{li \times li}$ and $W_2 \in \mathbb{R}^{j \times j}$ respectively, which should satisfy the following conditions: W_1 should be of full rank and the rank of $\begin{pmatrix} U_p \\ Y_p \end{pmatrix} W_2$ should be equal to the rank of $\begin{pmatrix} U_p \\ Y_p \end{pmatrix}$. Furthermore, one can distinguish between two classes of subspace identification algorithms. The first class uses the state estimates $\widetilde{X}_i$ (the right singular vectors of $W_1 O_i W_2$) to find the system matrices. The algorithm in Section 6.2 belongs to this class. The second class of algorithms uses the extended observability matrix Γ_i (the left singular vectors of $W_1 O_i W_2$) to first determine estimates of A and C and subsequently of B, D and Q, S, R.
It can be shown that three subspace algorithms that have been described in the literature (N4SID, MOESP and CVA[4]) all start from $W_1 O_i W_2$ with for each of the algorithms a specific choice of weighting matrices W_1 and W_2. The results are summarized in Table 1. From this table is clear that the algorithm described above is the N4SID algorithm ($W_1 = I_{li}$ and $W_2 = I_j$).

Acronym	W_1	W_2
N4SID	I_{li}	I_j
CVA	$\left(\lim_{j\to\infty} \frac{1}{j}[(Y_f/U_f^\perp)(Y_f/U_f^\perp)^T]\right)^{-1/2}$	$\Pi_{U_f^\perp}$
MOESP	I_{li}	$\Pi_{U_f^\perp}$

Table 1: In this table we give interpretations of different existing subspace identification algorithms in a unifying framework. All these algorithms first calculate an oblique projection O_i, followed by an SVD of the weighted matrix $W_1 O_i W_2$. The first two algorithms, N4SID and CVA, use the state estimates $\widetilde{X}_i$ (the right singular vectors) to find the system matrices, while MOESP is based on the extended observability matrix Γ_i (the left singular vectors). The matrix $U_f^\perp$ in the weights of CVA and MOESP represents the orthogonal complement of the row space of U_f.

7. Comments and perspectives

In this Section, we briefly comment on the relation with other identification methods for linear systems, we elaborate on some important open problems and briefly discuss several extensions.

As we have shown in Figure 2, so-called classical identification methods first determine a model (and if needed then proceed via a Kalman filter to estimate a state sequence). A good introduction to these methods (such as least squares methods, instrumental variables, prediction error methods (PEM), etc...) can be found in this Encyclopedia under Identification of linear Systems in Time Domain. Obviously, subspace identification algorithms are just one (important) group of methods for identifying linear systems. But many users of system identification prefer to start from linear input-output models, parametrized by numerator and denominator polynomials and then use maximum likelihood or instrumental variables based techniques. The at first sight apparant advantage of having an input-output parametrization however often turns out to be a disadvantage, as the theory of parametrizations of multivariable systems is certainly not easy nor straightforward and therefore complicates the required

[4]The acronym **N4SID** stands for "**N**umerical algorithms for **S**ubspace **S**tate **S**pace **S**ystem **ID**entification", **MOESP** for "**M**ultivariable **O**utput-**E**rror **S**tate s**P**ace" and **CVA** is the acronym of "**C**anonical **V**ariate **A**nalysis".

optimization algorithms (e.g. there is not one single parametrization for a multiple-output system). In many implementations of PEM-identification, a model obtained by subspace identification typically serves as a good intitial guess (PEMs require a nonlinear conconvex optimization problem to be solved, for which a good intial guess if required).
Another often mentioned disadvantage of subspace methods is the fact that it does not optimize a certain cost function. The reason for this is that, contrary to input-output models (transfer matrices), we can not (as of this moment) formulate a likelihood function for the identification of the state space model, that also leads to an amenable optimization problem. So, in a certain sense, subspace identification algorithms provide (often surprisingly good) 'approximations' of the linear model, but there is still a lot of ongoing research on how the identified model relates to a maximum likelihood formulation of the problem. In particular, it is also not straightforward at all to derive expressions for the error covariances on the estimates, nor the quantify exactly in what sense the obtained state sequence is an approximation to the 'real' (theoretical) Kalman filter state sequence, if some of the assumptions we made are not satisfied and/or the block dimensions i and/or j are not infinite (which they never are in practice). Yet, it is our experience that subspace algorithms often tend to give very good linear models for industrial data sets. By now, in the literature, many succesful implementations and cases have been reported in mechanical engineering (modal and vibrational analysis of mechanical structures such as cars, bridges (civil engineering), airplane wings (flutter analysis), missiles (ESA's Ariane), etc...), process industries (chemical, steel, paper and pulp,....), data assimilation methods (in which large systems of PDEs are discretized and reconciliated with observations using large scale Kalman filters and subspace identification methods are used in an 'error correction' mode), dynamic texture (reduction of sequences of images that are highly correlated in both space (within one image) and time (over several images)).

Since the introduction of subspace identification algorithms, the basic ideas have been extended to other system classes, such as closed-loop systems, linear parameter-varying state-space systems, bilinear systems, continuous-time systems, descriptor systems, periodic systems. We refer the reader to the bibliography for more information. Furthermore, efforts have been made to fine-tune the algorithms as presented in this paper. For example, several algorithms have been proposed to ensure stability of the identified model. For stochastic models, the positive-realness property should hold, which is not guaranteed by the raw subspace algorithms for certain data sets. Also for this problem extensions have been made. More information can be found in the bibliography.

8. Software

The described basis algorithm and variants have been incorporated in commercial software standards for system identification:

- the System Identification Toolbox in Matlab, developed by Prof. L. Ljung (Linköping, Sweden): `http://www.mathworks.com/products/sysid/`
- the system identification package ADAPTx of Adaptics, Inc, developed by dr. W. E. Larimore: `http://www.adaptics.com/`
- the ISID-module in Xmath, developed by dr. P. Van Overschee and Prof. B. De Moor and in license sold to ISI Inc. (now Wind River), USA: `http://www.windriver.com`
- the software packages RaPID and INCA of IPCOS International: `http://www.ipcos.be`

- the package MACEC, developed at the department of Civil Engineering of the K.U.Leuven in Belgium: http://www.kuleuven.ac.be/bwm/macec/
- products of LMS International: http://www.lms-international.com

Also public domain software, as SLICOT (http://www.win.tue.nl/niconet/NIC2/slicot.html), the SMI toolbox of the Control Laboratory at the T.U.Delft (http://lcewww.et.tudelft.nl/~verdult/smi/), the Cambridge University System Identification Toolbox (http://www-control.eng.cam.ac.uk/jmm/cuedsid/) and the website of the authors (http://www.esat.kuleuven.ac.be/sista-cosic-docarch/) contain subspace identification algorithms.

Acknowledgements

The authors are grateful to dr. Peter Van Overschee.
Our research is supported by grants from several funding agencies and sources: **Research Council KUL**: GOA-Mefisto 666, several PhD/postdoc & fellow grants; **Flemish Government**: - FWO: PhD/postdoc grants, projects, G.0240.99 (multilinear algebra), G.0407.02 (support vector machines), G.0197.02 (power islands), G.0141.03 (Identification and cryptography), G.0491.03 (control for intensive care glycemia), G.0120.03 (QIT), research communities (ICCoS, ANMMM); - AWI: Bil. Int. Collaboration Hungary/ Poland; - IWT: PhD Grants, Soft4s (softsensors); **Belgian Federal Government**: DWTC (IUAP IV-02 (1996-2001) and IUAP V-22 (2002-2006), PODO-II (CP/40: TMS and Sustainibility); **EU**: CAGE; ERNSI; Eureka 2063-IMPACT; Eureka 2419-FliTE; **Contract Research/agreements**: Data4s, Electrabel, Elia, LMS, IPCOS, VIB;

Bibliography

Akaike H. (1974). Stochastic theory of minimal realization. *IEEE Transactions on Automatic Control* **19**, 667–674.

— (1975). Markovian representation of stochastic processes by canonical variables. *SIAM Journal on Control* **13**(1), 162–173.

Aoki M. (1987). *State Space Modeling of Time Series*. Berlin, Germany: Springer Verlag.

Bauer D. (2001). Order estimation for subspace methods. *Automatica* **37**, 1561–1573. [In this paper the question of estimating the order in the context of subspace methods is addressed. Three different approaches are presented and the asymptotic properties thereof derived. Two of these methods are based on the information contained in the estimated singular values, while the third method is based on the estimated innovation variance].

Bauer D., Deistler M., Scherrer W. (1999). Consistency and asymptotic normality of some subspace algorithms for systems without observed inputs. *Automatica* **35**, 1243–1254. [The main result presented here states asymptotic normality of subspace estimates. In addition, a consistency result for the system matrix estimates is given. An algorithm to compute the asymptotic variances of the estimates is presented].

Bauer D., Ljung L. (2002). Some facts about the choice of the weighting matrices in Larimore type of subspace algorithms. *Automatica* **38**(5). [In this paper the effect of some weighting matrices on the asymptotic variance of the estimates of linear discrete time state space systems estimated using subspace methods is investigated. The analysis deals with systems with white or without observed inputs and refers to the Larimore type of subspace procedures. The main result expresses the asymptotic variance of the system matrix estimates in canonical form as a function of some of the user choices, clarifying the question on how to choose them optimally. It is shown, that the CCA weighting scheme leads to optimal accuracy].

Chiuso A., Picci G. (2001). Some algorithmic aspects of subspace identification with inputs. *International Journal of Applied Mathematics and Computer Science* **11**(1), 55–75. [In this paper a new structure for subspace identification algorithms is proposed to help fixing problems when certain experimental conditions cause ill-conditioning].

Cho Y., Xu G., Kailath T. (1994a). Fast identification of state space models via exploitation of displacement structure. *IEEE Transactions on Automatic Control* **AC-39**(10). [The major costs in the identification of state-space models still remain because of the need for the singular value (or sometimes QR) decomposition. It turns out that proper exploitation, using results from the theory of displacement structure, of the Toeplitz-like nature of several matrices arising in the procedure reduces the computational effort].

— (1994b). Fast recursive identification of state space models via exploitation of displacement structure. *Automatica* **30**(1), 45–59. [In many on-line identification scenarios with slowly time-varying systems, it is desirable to update the model as time goes on with the minimal computational burden. In this paper, the results of the batch processing algorithm are extended to allow updating of the identified state space model with few flops].

Chou C.T., Verhaegen M. (1997). Subspace algorithms for the identification of multivariable dynamic errors-in-variables models. *Automatica* **33**(10), 1857–1869. [The problem of identifying multivariable finite dimensional linear time-invariant systems from noisy input/output measurements is considered. Apart from the fact that both the measured input and output are corrupted by additive white noise, the output may also be contaminated by a term which is caused by a white input process noise; furthermore, all these noise processes are allowed to be correlated with each other].

Chui N.L.C., Maciejowski J.M. (1996). Realization of stable models with subspace methods. *Automatica* **32**(11), 1587–1595. [In this paper the authors present algorithms to find stable approximants to a least-squares problem, which are then applied to subspace methods to ensure stability of the identified model].

Dahlén A., Lindquist A., Marí J. (1998). Experimental evidence showing that stochastic subspace identification methods may fail. *Systems & Control Letters* **34**, 303–312. [It is known that certain popular stochastic subspace identification methods may fail for theoretical reasons related to positive realness. In this paper, the authors describe how to generate data for which the methods do not find a model].

De Moor B., Van Overschee P. (1994). Graphical user interface software for system identification. Tech. Rep. Report 94–06I, ESAT-SISTA, Department of Electrical Engineering, Katholieke Universiteit Leuven, Leuven, Belgium. [Award winning paper of the *Siemens Award 1994*].

De Moor B., Van Overschee P., Favoreel W. (1999). Algorithms for subspace state-space system identification: An overview. In B.N. Datta, ed., *Applied and Computational Control, Signals, and Circuits*, vol. 1, chap. 6, pp. 247–311, Birkhauser,. [This chapter presents an overview of the field of subspace

identification and compares it with the traditional prediction error methods. The authors present several comparisons between prediction error methods and subspace methods, including comparisons of accuracy and computational effort].

Desai U.B., Kirkpatrick R.D., Pal D. (1985). A realization approach to stochastic model reduction. *International Journal of Control* **42**(2), 821–838.

Desai U.B., Pal D. (1984). A transformation approach to stochastic model reduction. *IEEE Transactions on Automatic Control* **AC-29**(12), 1097–1100.

Faure P. (1976). Stochastic realization algorithms. In R. Mehra, D. Lainiotis, eds., *System Identification: Advances and Case Studies*, Academic Press.

Favoreel W., De Moor B., Van Overschee P. (1999). Subspace identification of bilinear systems subject to white inputs. *IEEE Transactions on Automatic Control* **44**(6), 1157–1165. [The class of existing linear subspace identification techniques is generalized to subspace identification algorithms for bilinear systems].

— (2000). Subspace state space system identification for industrial processes. *Journal of Process Control* **10**, 149–155. [A general overview of subspace system identification methods is given. A comparison between subspace identification and prediction error methods is made on the basis of computational complexity and precision of the methods by applying them on 10 industrial data sets].

Ho B.L., Kalman R.E. (1966). Effective construction of linear state-variable models from input/output functions. *Regelungstechnik* **14**(12), 545–548. [].

Jansson M., Wahlberg B. (1996). A linear regression approach to state-space subspace system identification. *Signal Processing* **52**(2), 103–129. [In this paper, one shows that state-space subspace system identification (4SID) can be viewed as a linear regression multistep-ahead prediction error method with certain rank constraints].

— (1998). On consistency of subspace methods for system identification. *Automatica* **34**(12), 1507–1519. [In this paper, the consistency of a large class of methods for estimating the extended observability matrix is analyzed. Persistence of excitation conditions on the input signal are given which guarantee consistent estimates for systems with only measurement noise. For systems with process noise, it is shown that a persistence of excitation condition on the input is not sufficient. More precisely, an example for which the subspace methods fail to give a consistent estimate of the transfer function is given. This failure occurs even if the input is persistently exciting of any order. It is also shown that this problem can be eliminated if stronger conditions on the input signal are imposed].

Juricek B.C., Seborg D.E., Larimore W.E. (2001). Identification of the Tennessee Eastman challenge process with subspace methods. *Control Engineering Practice* **9**(12), 1337–1351. [The Tennessee Eastman challenge process is a realistic simulation of a chemical process that has been widely used in process control studies. In this case study, several identification methods are examined and used to develop MIMO models that contain seven inputs and ten outputs. For a variety of reasons, the only successful models are the state-space models produced by two popular subspace algorithms, N4SID and canonical variate analysis (CVA). The CVA model is the most accurate].

Kalman R.E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the American Society of Mechanical Engineers, Journal of Basic Engineering* **83**(1), 35–45.

— (1963). Mathematical description of linear dynamical systems. *SIAM Journal on Control* **1**, 152–192.

— (1981). Realization of covariance sequences. In *Proceedings of the Toeplitz Memorial Conference*, Tel Aviv, Israel.

Kung S.Y. (1978). A new identification method and model reduction algorithm via singular value decomposition. In *Proceedings of the 12th Asilomar Conference on Circuits, Systems and Computations*, pp. 705–714.

Larimore W.E. (1996). Statistical optimality and canonical variate analysis system identification. *Signal Processing* **52**(2), 131–144. [The Kullback information is developed as the natural measure of the error in model approximation for general model selection methods including the selection of model state order in large as well as small samples. It also plays a central role in developing statistical decision procedures for the optimal selection of model order as well as structure based on the observed data. The optimality of the canonical variate analysis (CVA) method is demonstrated for both an open- and closed-loop multivariable system with stochastic disturbances].

Lindquist A., Picci G. (1996). Canonical correlation analysis, approximate covariance extension, and identification of stationary time series. *Automatica* **32**, 709–733. [In this paper the authors analyze a class of state space identification algorithms for time-series, based on canonical correlation analysis, in the light of recent results on stochastic systems theory. In this paper the statistical problem of stochastic modeling from estimated covariances is phrased in the geometric language of stochastic realization theory].

Lovera M., Gustafsson T., Verhaegen M. (2000). Recursive subspace identification of linear and non-linear Wiener state-space models. *Automatica* **36**, 1639–1650. [The problem of MIMO recursive identification is analyzed within the framework of subspace model identification and the use of recent signal processing algorithms for the recursive update of the singular value decomposition is proposed].

McKelvey T., Akçay H., Ljung L. (1996). Subspace-based identification of infinite-dimensional multivariable systems from frequency-response data. *Automatica* **32**(6), 885–902. [An identification algorithm which identifies low complexity models of infinite-dimensional systems from equidistant frequency-response data is presented. The new algorithm is a combination of the Fourier transform technique with subspace techniques].

Moonen M., De Moor B., Vandenberghe L., Vandewalle J. (1989). On- and off-line identification of linear state-space models. *International Journal of Control* **49**(1), 219–232.

Peternell K., Scherrer W., Deistler M. (1996). Statistical analysis of novel subspace identification methods. *Signal processing* **52**, 161–177. [In this paper four subspace algorithms which are based on an initial estimate of the state are considered. For the algorithms considered a consistency result is proved. In a simulation study the relative (statistical) efficiency of these algorithms in relation to the maximum likelihood algorithm is investigated].

Picci G., Katayama T. (1996). Stochastic realization with exogenous inputs and 'subspace-methods' identification. *Signal Processing* **52**(2), 145–160. [In this paper the stochastic realization of stationary processes with exogenous inputs in the absence of feedback is studied, and its application to identification is briefly discussed].

Schaper C.D., Larimore W.E., Sborg D.E., Mellichamp D.A. (1994). Identification of chemical processes using canonical variate analysis. *Computers & Chemical Engineering* **18**(1), 55–69. [A method of identification of linear input-output models using canonical variate analysis (CVA) is developed for application to chemical processes].

Silverman L. (1971). Realization of linear dynamical systems. *IEEE Transactions on Automatic Control* **AC-16**, 554–567.

Stoica P., Jansson M. (2000). MIMO system identification: state-space and subspace approximations versus transfer function and instrumental variables. *IEEE Transactions on Signal Processing* **48**(11), 3087–3099. [A simulation study, in which the performances of the subspace and the transfer function approaches are compared, shows that the latter can provide more accurate models than the former at a lower computational cost].

Van Gestel T., Suykens J., Van Dooren P., De Moor B. (2001). Identification of stable models in subspace identification by using regularization. *IEEE Transactions on Automatic Control* **46**(9), 1416–1420. [This paper shows how one can impose stability to the model that is identified with a subspace algorithm. The method proposed is based on regularization].

Van Overschee P., De Moor B. (1993). Subspace algorithms for the stochastic identification problem. *Automatica* **29**, 649–660. [In this paper a subspace algorithm is derived to consistently identify stochastic state space models from given output data].

— (1994). N4SID – Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica* **30**(1), 75–94. [In this paper two subspace algorithms for the identification of mixed deterministic-stochastic systems are derived].

— (1995). A unifying theorem for three subspace system identification algorithms. *Automatica* **31**(12), 1853–1864. [The authors indicate and explore similarities between three different subspace algorithms for the identification of combined deterministic-stochastic systems. It is shown that all three algorithms are special cases of one unifying scheme].

— (1996). *Subspace Identification for linear systems: Theory – Implementation – Applications*. Dordrecht, The Netherlands: Kluwer Academic Publishers. [In this book the theory of subspace identification algorithms is presented in detail].

Verdult V., Verhaegen M. (2001). Identification of multivariable bilinear state space systems based on subspace techniques and separable least squares optimization. *International Journal of Control* **74**(18), 1824–1836. [The identification of discrete-time bilinear state space systems with multiple inputs and multiple outputs is discussed. The subspace algorithm is modified such that it reduces the dimension of the matrices involved].

— (2002). Subspace identification of multivariable linear parameter-varying systems. *Automatica* **38**(5). [A subspace identification method is discussed that deals with multivariable linear parameter-varying state-space systems with affine parameter dependence].

Verhaegen M. (1993). Application of a subspace model identification technique to identify LTI systems operating in closed-loop. *Automatica* **29**(4), 1027–1040. [In this paper the identification of linear time-invariant (LTI) systems operating in a closed-loop with an LTI compensator is reformulated to an open-loop multi-input-multi-output (MIMO) (state space model) identification problem, followed by a model reduction step. The open-loop identification problem is solved by the MOESP (MIMO output-error state space model) identification technique].

— (1994). Identification of the deterministic part of MIMO state space models given in innovations form from input-output data. *Automatica* **30**(1), 61–74. [In this paper two algorithms to identify a linear, time-invariant, finite dimensional state space model from input-output data are described. The system to be identified is assumed to be excited by a measurable input and an unknown process noise and the measurements are disturbed by unknown measurement noise. Both noise sequences are discrete zero-mean white noise].

— (1996). A subspace model identification solution to the identification of mixed causal, anti-causal LTI systems. *SIAM Journal on Matrix Analysis and Applications* **17**(2), 332–347. [This paper describes the modification of the family of MOESP subspace algorithms when identifying mixed causal and anti-causal systems].

Verhaegen M., Dewilde P. (1992a). Subspace model identification part 1. The output-error state-space model identification class of algorithms. *International journal of control* **56**(5), 1187–1210. [In this paper two subspace algorithms are presented to realize a finite dimensional, linear time-invariant state-space model from input-output data. Both schemes are versions of the MIMO Output-Error State Space model identification (MOESP) approach].

— (1992b). Subspace model identification part 2. Analysis of the elementary output-error state-space model identification algorithm. *International journal of control* **56**(5), 1211–1241. [The elementary MOESP algorithm is analyzed in this paper. It is shown that the MOESP1 implementation yields asymptotically unbiased estimates. Furthermore, the model reduction capabilities of the elementary MOESP schemes are analyzed when the observations are error-free].

— (1993). Subspace model identification part 3. Analysis of the ordinary output-error state-space model identification algorithm. *International journal of control* **56**(3), 555–586. [The ordinary MOESP algorithm is analyzed and extended in this paper. The extension of the ordinary MOESP scheme with instrumental variables increases the applicability of this scheme].

Verhaegen M., Westwick D. (1996). Identifying MIMO Hammerstein systems in the context of subspace model identification methods. *International Journal of Control* **63**(2), 331–349. [In this paper, the extension of the MOESP family of subspace model identification schemes to the Hammerstein-type of nonlinear system is outlined].

Verhaegen M., Yu X. (1995). A class of subspace model identification algorithms to identify periodically and arbitrarily time-varying systems. *Automatica* **31**(2), 201–216. [In this paper subspace model identification algorithms that allow the identification of a linear, time-varying state space model from an ensemble set of input-output measurements are presented].

Viberg M. (1995). Subspace-based methods for the identification of linear time-invariant systems. *Automatica* **31**(12), 1835–1851. [An overview of existing subspace-based techniques for system identification is given. The methods are grouped into the classes of realization-based and direct techniques. Similarities between different algorithms are pointed out, and their applicability is commented upon].

Viberg M., Wahlberg B., Ottersten B. (1997). Analysis of state space system identification methods based on instrumental variables and subspace fitting. *Automatica* **33**(9), 1603–1616. [This paper gives a statistical investigation of subspace-based system identification techniques. Explicit expressions for the asymptotic estimation error variances of the corresponding pole estimates are given].

Westwick D., Verhaegen M. (1996). Identifying MIMO Wiener systems using subspace model identification methods. *Signal Processing* **52**, 235–258. [In this paper is shown that the MOESP class of subspace identification schemes can be extended to identify Wiener systems, a series connection of a linear dynamic system followed by a static nonlinearity].

Zeiger H., McEwen A. (1974). Approximate linear realization of given dimension via Ho's algorithm. *IEEE Transactions on Automatic Control* **19**, 153.