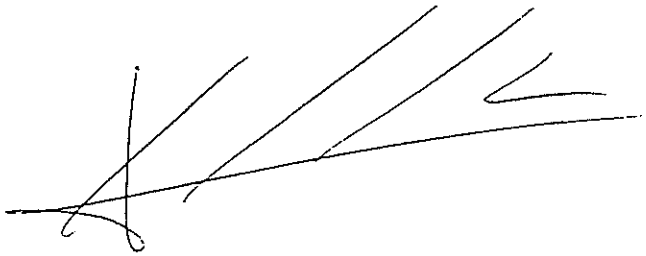
Alberto Isidori (Ed.)

95-20

Trends in Control

A European Perspective

With 101 Figures



Bart De Moor
Leuven, August 1995

Springer

Table of Contents

Numerical Algorithms for Subspace State Space System Identification *

Bart De Moor[1] *and Peter Van Overschee*[2]

[1] Associate Professor at the Department of Electrical Engineering of the Katholieke Universiteit Leuven, Kardinaal Mercierlaan 94, B-3001 Leuven, Belgium, tel: 32/ 16/ 321715, fax:32/ 16/ 321986, email: bart.demoor@esat.kuleuven.ac.be, and Senior Research Associate of the National Fund for Scientific Research (NFWO).
[2] Senior Research Assistant of the National Fund for Scientific Research (NFWO).

Abstract. We present the basic notions on subspace identification algorithms for linear systems. These methods first estimate a state sequence directly from input-output data, through an orthogonal or oblique projection of the row space of certain input-output block Hankel matrices into the row spaces of others. The state estimation procedure is then followed by a least squares problem which delivers the state space model. These algorithms can be elegantly implemented using well-known numerical linear algebra algorithms such as the LQ- and singular value decomposition.

1 Introduction

While at first sight, the class of linear time-invariant systems with lumped parameters, seems to be rather restricted, it turns out that the input-output behavior of many real-life industrial processes, for most practical purposes (such as simulation, prediction, monitoring or control system design), can be approximated very well by a linear time-invariant system. The problem of linear system identification, which is the problem of obtaining (approximate) linear system representations from measured data, is certainly not new. It has been studied now for more than 20 years in the mathematical engineering literature. Excellent books, such as [AE 71] [AW 84], [BJ 76], [Eyk 74], [Lju 87], [Nor 86], [SS 89] and several others, reveal the maturity of the field, which has culminated in the eighties into the development of so-called *prediction-error-methods* (PEM). These have been analyzed in much detail and applied with great success to many practical problems. Excellent software toolboxes, in which these algorithms (together with the subspace algorithms described in this work) have been implemented, such as the Matlab System Identification Toolbox [Lju 91c],

* The research reported on here was supported by Belgian Federal Government Interuniversity Attraction Poles IUAP-17 (Modeling and Control of Dynamic Systems), IUAP-50 (Automation in Design and Production), Flemish Government Concerted Action GOA-MIPS (Model-based Information Processing Systems) and the European Commission Human Capital and Mobility SIMONET (System Identification and Modeling Network).

or the Xmath Interactive System Identification Module [VODMAKB 94], are nowadays commercially available.

Yet, it can not be denied that with these (by now) classical PEM approaches, there are still some problems that are difficult to deal with in practical situations. It so turns out that a new breed of algorithms, called *subspace algorithms*, which have been developed the last couple of years, provide a meaningful alternative for PEM, with respect to the following problems:

Parametrization problems: In PEM the transfer matrix models are parametrized with a minimal number of parameters. This parametrization issue has triggered lots of research (e.g. [GW 74] [Gui 75] [Gui 81] [HD 88] [Kai 80] [Lue 67] [VOL 82]) or for so-called *overlapping* parametrizations (see e.g. [GeWe 82] [HD 88] and the references in these papers and books). However, determining the right parametrization (especially in the case of multi-output systems) starting from observed measured input-output data is certainly not trivial in practice!

Numerical robustness: Even when a parametrization has been fixed, the resulting model structure can be ill-conditioned, meaning that the identified parameters are extremely sensitive to perturbations. In subspace methods, the models are full state space models ('fully parametrized'), typically within a certain state space basis with good properties, such as e.g. the balanced realization, which guarantees insensitivity with respect to perturbations. The only variable to be decided upon is the number of states (the 'order' of the system), a choice which is guided by the singular values of certain matrices and/or trial-and-error based. This automatically leads to more user-friendly algorithmic implementations.

Algorithms: For almost all parametrizations, the identification problem translates into a *nonlinear numerical optimization problem*, with all its unpleasant side effects such as for instance convergence problems and local minima. Subspace algorithms are typically faster (when implemented correctly!) than Prediction Error Methods, because they are basically *non-iterative*, exploiting basic algorithms from numerical linear algebra, such as the LQ-decomposition (transpose of the QR-decomposition) and the singular value decomposition.

Nonzero initial conditions: In parametrized input-output models, the presence of a nonzero initial state requires additional terms in the parametrization. In subspace methods, the initial state is estimated (implicitly or explicitly). More specifically, subspace methods will also model those modes of the system that are observable but not necessarily controllable, including *unstable* ones.

This paper should be considered as a *first introduction* to subspace methods for system identification. We do not aim here for completeness nor complete mathematical rigor nor exhaustive comparison with other methods and algorithms that have appeared in the literature. Nor will we attempt to provide a complete list of variations on the themes proposed in this paper. We simply want to elaborate

on the main insights that have been influential in developing subspace methods for linear system identification during the last couple of years.
This paper is organized as follows: In Section 2, we briefly describe the state space model that will be identified from input-output data. In Section 3, we comment on one of the main ideas behind subspace methods, which is to recover (implicitly or explicitly), the state directly from input-output data. It will be explained how, first of all, this is possible by using orthogonal or oblique projections using block Hankel matrices with data, and second, how first obtaining the state sequence *linearizes* the problem, and allows to obtain the state space model from a simple least squares solution. In Section 4, we introduce (past and future) block Hankel matrices with input-output data and also define the notations to be used throughout. In Section 5, we treat the geometrical operations to be used such as orthogonal and oblique projections, involving the data matrices. In Section 6, we explore a relationship between a certain orthogonal projection of 'future outputs' into the 'past inputs and outputs' and 'future inputs', and the state sequence as it would be generated by a bank of Kalman filters if the state space model and its initial conditions would be completely known. In Section 7, we explore the relationship between a certain oblique projection of 'future outputs', along 'future inputs', into 'past inputs and outputs' and the state sequence as it would be generated by a bank of Kalman filters, which is the same as the one from Section 6, but with different initial conditions. Here we show how the oblique projection contains important information about the model to be identified, such as its order and its observability matrix. In Section 8, we derive how the state space model can be found by solving a least squares problem using the states obtained in Section 6. In Section 9, we derive another algorithm which is more elegant in a certain way, but biased. It uses the state sequence obtained in Section 7. A robustified implementation and the role of the LQ-decomposition is discussed in Section 10. Conclusions and references to extensions and related work are summarized in Section 11.

2 State space models and problem formulation

We will restrict ourselves to discrete time, linear, time-invariant, state space models, which are described as[3]:

$$x_{k+1} = Ax_k + Bu_k + w_k \,, \tag{1}$$

$$y_k = Cx_k + Du_k + v_k \,, \tag{2}$$

with

$$\mathbf{E}[\begin{pmatrix} w_p \\ v_p \end{pmatrix} \begin{pmatrix} w_q^t & v_q^t \end{pmatrix}] = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \delta_{pq} \quad \geq 0 \,. \tag{3}$$

The vectors $u_k \in \mathbb{R}^{m\times 1}$ and $y_k \in \mathbb{R}^{l\times 1}$ are the measurements at time instant k of respectively the m inputs and l outputs of the process. The vector x_k is the state vector of the process at discrete time instant k, $v_k \in \mathbb{R}^{l\times 1}$ and $w_k \in \mathbb{R}^{n\times 1}$

[3] E denotes the expected value operator and δ_{pq} the Kronecker delta.

are unobserved vector signals. v_k is called the measurement noise and w_k is called the process noise. It is assumed that they are zero mean, stationary white noise vector sequences. $A \in \mathbb{R}^{n\times n}$ is the system matrix, $B \in \mathbb{R}^{n\times m}$ is the input matrix, $C \in \mathbb{R}^{l\times n}$ is the output matrix while $D \in \mathbb{R}^{l\times m}$ is the direct feedthrough matrix. 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. As explained in e.g. [Fau 76], the noise model is not unique: there is a set of 'equivalent' noise models that can be described using linear matrix inequalities or equivalently, a pair of algebraic Riccati equations (backward and forward)(see [Cai 88] [Fau 76] [VODM 93a] for details). In addition, we require the following orthogonality conditions to be true:

$$\mathbf{E}[\begin{pmatrix} x_k \\ u_k \end{pmatrix} \begin{pmatrix} w_k^T & v_k^T \end{pmatrix}] = 0 \,. \tag{4}$$

We are now ready to state the main problem treated here:

> Given a large number of measurements of the input $u_k \in \mathbb{R}^m$ and the output $y_k \in \mathbb{R}^l$ generated by the unknown system (1)-(2)-(3). Determine, using only the inputs u_k and outputs y_k, the order n of the unknown system, the system matrices $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}, C \in \mathbb{R}^{l\times n}, D \in \mathbb{R}^{l\times m}$ up to within a similarity transformation and the matrices $Q \in \mathbb{R}^{n\times n}, S \in \mathbb{R}^{n\times l}, R \in \mathbb{R}^{l\times l}$ (or equivalent realizations in the sense of Faurre).

There are many questions that pop up. For instance: What is meant by a large number of measurements? (We'll show that we need an infinite number for consistency). Which noise model is being identified, as there are an infinite number of 'equivalent' ones? etc ....

3 The main idea: Re-discovering the state

One of the important conceptual ideas behind subspace algorithms is to re-introduce the concept of *the state of a dynamical system* within the system identification context. In contrast to 'classical' identification algorithms, subspace algorithms first estimate/calculate the state (sequence) (implicitly or explicitly), while next the (state space) model is determined. This difference between PEM methods and subspace methods is illustrated in Figure 1.
Why would one bother to first obtain the (Kalman filter) state sequence, directly from input-output data and only after that the state space model? The answer is threefold:

- When the state sequence x_k is available and since the inputs u_k and outputs y_k are available as well, it can be seen from the state space model (1)-(2) and the orthogonality conditions (4) that the state space matrices A, B, C and D could be obtained by solving a least squares problem. The covariance matrices Q, R and S could then be estimated as the covariance matrices

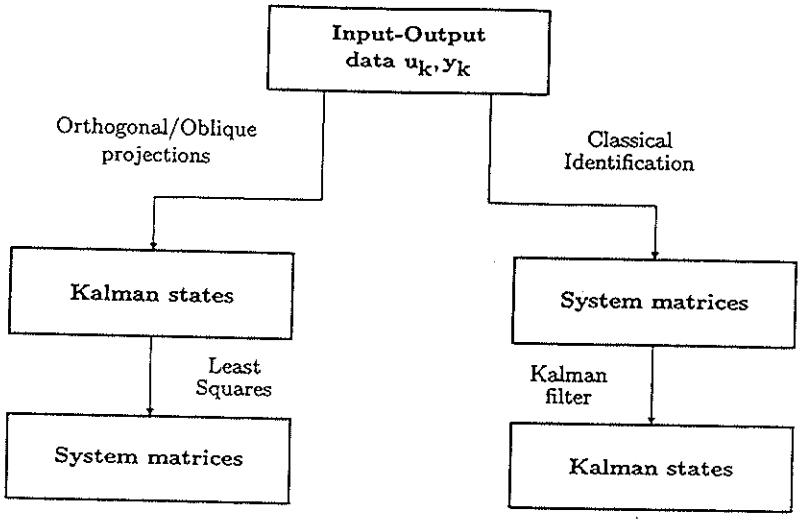


Fig. 1. System identification aims at constructing state space models from input-output data. The left hand side shows the subspace method approach: first the (Kalman filter) states are estimated directly from input-output data, then the system matrices can be obtained. The right hand side is the classical approach: first obtain the system matrices, then estimate the states.

of the least squares residuals. Therefore, it seems worthwhile to first try to obtain the state from input and output measurements, because in doing so, the identification problem is *linearized* in the sense that it is reduced to a simple least squares problem!

- Remains the problem of how to obtain the state sequence, which will be the main subject of Sections 6 and 7 below. The main conclusion will be that a certain Kalman filter state sequence can be calculated directly from input-output data, by doing one or two orthogonal or oblique projections with block Hankel matrices constructed from input-output data.
- It turns out that the state sequence can be calculated using some well-known tools from numerical linear algebra, such as the LQ-decomposition and the singular value decomposition, which ultimately leads to a numerically extremely efficient implementation.

4 The linear algebra tools: Block Hankel matrices with data

Block Hankel matrices with inputs and outputs are the basic linear algebra objects for subspace identification algorithms. Input block Hankel matrices are defined as:

$$U_{0|i-1} \stackrel{\text{def}}{=} \begin{pmatrix} u_0 & u_1 & u_2 & \dots & u_{j-1} \\ u_1 & u_2 & u_3 & \dots & u_j \\ \dots & \dots & \dots & \dots & \dots \\ u_{i-1} & u_i & u_{i+1} & \dots & u_{i+j-2} \end{pmatrix} \in \mathbb{R}^{mi \times j} .$$

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[4]. Most of the times, we will assume that $j \to \infty$ (which means in practice that j should be 'large'). Summarizing, we require that $n < i << j \to \infty$. We will also use the following input matrices:

$$U_p = U_{0|i-1} \ , \ U_f = U_{i|2i-1} \ , \ U_p^+ = U_{0|i} \ , \ U_f^- = U_{i+1|2i-1} \ .$$

Here, the subscript 'p' refers to 'past', 'f' refers to future and the superscripts '+' and '-' refer to adding or removing one more block row. Similar definitions hold for the block Hankel matrices with the output vectors, which will be denoted by Y_p, Y_f, Y_p^+ and Y_f^-. 'Double' block Hankel matrices with inputs and outputs, are defined as:

$$W_{0|i-1} = \begin{pmatrix} U_p \\ Y_p \end{pmatrix} = W_p \ , \ W_{0|i} = \begin{pmatrix} U_p^+ \\ Y_p^+ \end{pmatrix} = W_p^+ \ .$$

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

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

We assume the pair $\{A, C\}$ to be observable, which implies (see for instance [Kai 80]) that the rank of Γ_i is equal to n. The reversed extended controllability matrix Δ_i^d (where the subscript i denotes the number of block columns) is defined as:

$$\Delta_i^d \stackrel{\text{def}}{=} \begin{pmatrix} A^{i-1}B & A^{i-2}B & \dots & AB & B \end{pmatrix} \in \mathbb{R}^{n \times mi} .$$

[4] Theoretically, i should be at least as large as the largest observability index of the observability matrix.

Furthermore, we require the pair $\{A, [B\ ,\ Q^{1/2}]\}$ to be controllable. This implies that all modes are excited by either the external input u_k or the process noise w_k. The lower block triangular Toeplitz matrix H_i^d is defined as:

$$H_i^d \stackrel{\text{def}}{=} \begin{pmatrix} D & 0 & 0 & \dots & 0 \\ CB & D & 0 & \dots & 0 \\ CAB & CB & D & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ CA^{i-2}B & CA^{i-3}B & CA^{i-4}B & \dots & D \end{pmatrix} \in \mathbb{R}^{li \times mi} .$$

The system (1)-(2) is split in a deterministic and a stochastic subsystem, by splitting the state (x_k) and output (y_k) in a deterministic $(\bullet^d)$ and stochastic $(\bullet^s)$ component:

$$x_k = x_k^d + x_k^s\ ,\ y_k = y_k^d + y_k^s\ .$$

The deterministic state (x_k^d) and output (y_k^d) follow from the deterministic subsystem, which describes the influence of the deterministic input (u_k) on the deterministic output while the stochastic state (x_k^s) and output (y_k^s) follow from the stochastic subsystem, which describes the influence of the noise sequences (w_k and v_k) on the stochastic output:

$$\begin{aligned} x_{k+1}^d &= Ax_k^d + Bu_k\ , & x_{k+1}^s &= Ax_k^s + w_k\ , \\ y_k^d &= Cx_k^d + Du_k\ , & y_k^s &= Cx_k^s + v_k\ . \end{aligned} \tag{5}$$

The controllable modes of $\{A, B\}$ can be either stable or unstable. The controllable modes of $\{A, Q^{1/2}\}$ are assumed to be stable. The deterministic inputs (u_k) and states (x_k^d) and the stochastic states (x_k^s) and outputs (y_k^s) are assumed to be quasi-stationary (as defined in [Lju 87, page 27]). Note that even though the deterministic subsystem can have unstable modes, the excitation (u_k) has to be chosen in such a way that the deterministic states and output are finite for all time. Also note that since the systems $\{A, B\}$ and $\{A, Q^{1/2}\}$ are not assumed to be controllable (only the concatenation of the deterministic and stochastic subsystem as a whole should be controllable), the deterministic and stochastic subsystem may have common as well as completely decoupled input-output dynamics. The state sequence is defined as:

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

The *deterministic* state sequence X_i^d and *stochastic* state sequence X_i^s are defined as:

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

In a similar way, the past and future deterministic and stochastic state sequences are defined as $X_p^d = X_0^d$, $X_f^d = X_i^d$, , $X_p^s = X_0^s$, $X_f^s = X_i^s$. These are not the key state sequences for combined system identification. The introduction of the

more important Kalman filter state sequence is postponed to Section 6. We also define the following covariance and cross covariance matrices (which exist due to the quasi stationarity of u_k and x_k^d, see above) [5]:

$$\begin{aligned} R^{uu} &\stackrel{\text{def}}{=} \Phi_{[U_{0|2i-1}, U_{0|2i-1}]} = \begin{pmatrix} \Phi_{[U_p,U_p]} & \Phi_{[U_p,U_f]} \\ \Phi_{[U_f,U_p]} & \Phi_{[U_f,U_f]} \end{pmatrix} = \begin{pmatrix} R_p^{uu} & R_{pf}^{uu} \\ (R_{pf}^{uu})^T & R_f^{uu} \end{pmatrix} , \\ S^{xu} &\stackrel{\text{def}}{=} \Phi_{[X_0^d, U_{0|2i-1}]} = \begin{pmatrix} \Phi_{[X_p^d,U_p]} & \Phi_{[X_p^d,U_f]} \end{pmatrix} = \begin{pmatrix} S_p^{xu} & S_f^{xu} \end{pmatrix} , \\ \Sigma^d &\stackrel{\text{def}}{=} \Phi_{[X_p^d, X_p^d]} . \end{aligned}$$

For the stochastic subsystem (5) we use the following notations. Defining the output covariance matrices as

$$\Lambda_i \stackrel{\text{def}}{=} \mathbf{E}[y_{k+i} y_k^T] ,$$

we find for

$$\begin{aligned} \Lambda_0 &= \mathbf{E}[y_k . y_k^T] = \mathbf{E}[(Cx_k^s + v_k).(Cx_k^s + v_k)^T] \\ &= C\mathbf{E}[x_k^s (x_k^s)^T] C^T + \mathbf{E}[v_k v_k^T] = C\Sigma^s C^T + R , \end{aligned}$$

with an obvious definition for Σ^s. Defining

$$\begin{aligned} G &\stackrel{\text{def}}{=} \mathbf{E}[x_{k+1}^s . y_k^T] = \mathbf{E}[(Ax_k^s + w_k).(Cx_k^s + v_k)^T] \\ &= A\mathbf{E}[x_k^s (x_k^s)^T] C^T + \mathbf{E}[w_k v_k^T] = A\Sigma^s C^T + S , \end{aligned}$$

we get (for $i = 1, 2, \dots$)

$$\Lambda_i = CA^{i-1}G \text{ and } \Lambda_{-i} = G^T (A^{i-1})^T C^T .$$

The reversed extended stochastic controllability matrix Δ_i^c (where the subscript i denotes the number of block columns and the superscript "c" stands for "covariance") is defined as

$$\Delta_i^c \stackrel{\text{def}}{=} \begin{pmatrix} A^{i-1}G & A^{i-2}G & \dots & AG & G \end{pmatrix} \in \mathbb{R}^{n \times li} .$$

[5] In subspace identification we typically assume that there are long time series of data available ($j \to \infty$), and that the data is ergodic. Due to ergodicity and the infinite number of data at our disposition, we can replace the expectation operator $\mathbf{E}$ (average over an infinite number of experiments) with the operator $\mathbf{E_j}$ applied to the sum of variables (average over one, infinitely long, experiment). For instance for the correlation $\mathbf{E}(ae^T)$ between two random variables a and e, on which we have j observations, a_k and e_k, $k = 1, \dots, j$, we get $\mathbf{E}[ae^T] = \lim_{j\to\infty} [\frac{1}{j} \sum_{i=1}^{j} a_i e_i^T] = \mathbf{E_j}[\sum_{i=1}^{j} a_i e_i^T]$ with an obvious definition of $\mathbf{E_j}$: $\mathbf{E_j}[\bullet] \stackrel{\text{def}}{=} \lim_{j\to\infty} \frac{1}{j}[\bullet]$. We define the covariance $\Phi_{[A,B]}$ between two matrices $A \in \mathbb{R}^{p \times j}$ and $B \in \mathbb{R}^{q \times j}$ as $\Phi_{[A,B]} \stackrel{\text{def}}{=} \mathbf{E_j}[AB^T]$.

The block Toeplitz matrix L_i is constructed from the output covariance matrices as:

$$L_i \stackrel{\text{def}}{=} \begin{pmatrix} \Lambda_0 & \Lambda_{-1} & \Lambda_{-2} & \dots & \Lambda_{1-i} \\ \Lambda_1 & \Lambda_0 & \Lambda_{-1} & \dots & \Lambda_{2-i} \\ \Lambda_2 & \Lambda_1 & \Lambda_0 & \dots & \Lambda_{3-i} \\ \dots & \dots & \dots & \dots & \dots \\ \Lambda_{i-1} & \Lambda_{i-2} & \Lambda_{i-3} & \dots & \Lambda_0 \end{pmatrix} \in \mathbb{R}^{li \times li}. \quad (6)$$

5 The geometric tools: Orthogonal and Oblique Projections

In what follows, we will use the matrices $A \in \mathbb{R}^{p \times j}$, $B \in \mathbb{R}^{q \times j}$ and $C \in \mathbb{R}^{r \times j}$ as dummy variables, meaning that they are only 'local' variables in this Section, not to be confused with the system matrices A, B and C we have defined before. The elements of a row of one of the given matrices can be considered as the coordinates of a vector in the j-dimensional ambient space. The rows of each matrix A, B, C thus define a basis for a vector space in this ambient space. In Subsection 5.1 and 5.2 we define orthogonal and oblique projections, in which these row spaces are involved. It should be noted that these geometric operations can be easily implemented using an LQ decomposition, which is the subject of Section 10.

5.1 Orthogonal projections

Π_B denotes the operator that projects the row space of a matrix into the row space of the matrix $B \in \mathbb{R}^{q \times j}$ (where $\bullet^\dagger$ denotes the Moore-Penrose pseudo-inverse [6]) so that $\Pi_B \stackrel{\text{def}}{=} B^T.(BB^T)^\dagger.B$. A/B is shorthand for the projection of the row space of the matrix $A \in \mathbb{R}^{p \times j}$ into the row space of the matrix B so that $A/B \stackrel{\text{def}}{=} A.\Pi_B = AB^T.(BB^T)^\dagger.B$. The projection operator can be interpreted in the ambient j-dimensional space as indicated in Figure 2. Note that in the notation A/B the matrix B is printed bold face, which indicates that the result of the operation A/B lies in the row space of B.

$\Pi_{B^\perp}$ is the geometric operator that projects the row space of a matrix into the orthogonal complement of the row space of the matrix B, for which we have $A/B^\perp \stackrel{\text{def}}{=} A\Pi_{B^\perp}$, where $\Pi_{B^\perp} = I_j - \Pi_B$. Here $B^\perp$ is a matrix, the row space of which is the orthogonal complement to the row space of B. Once again these projections can be interpreted in the j-dimensional space as indicated in Figure 2. The combination of the projections Π_B and $\Pi_{B^\perp}$ decomposes a matrix A into two matrices for which the row spaces are orthogonal to each other as $A = A\Pi_B + A\Pi_{B^\perp}$. The projections also decompose the matrix A as a linear combination of the rows of B and those of $B^\perp$. With $L_B.B \stackrel{\text{def}}{=} A\Pi_B$ and $L_{B^\perp}.B^\perp \stackrel{\text{def}}{=} A\Pi_{B^\perp}$ we find $A = L_B.B + L_{B^\perp}.B^\perp$, which is indeed a decomposition of A as a linear combination of the rows of B and those of $B^\perp$.

[6] This generalized inverse could be replaced by less 'restricted' generalized inverses, but this will not be pursued here.

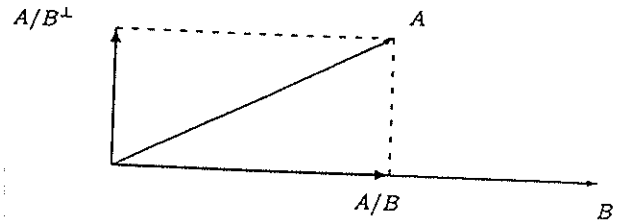


Fig. 2. Interpretation of the orthogonal projection in the j-dimensional space ($j = 2$ in this case). A/B is formed by projecting the row space of A orthogonally into the row space of B. $A/B^\perp$ on the other hand is formed by projecting the row space of A into the orthogonal complement of the row space of B.

5.2 Oblique projections

A matrix A can also be decomposed as a linear combination of two matrices B and C with *non-intersecting row spaces* and another matrix of which the row space is the orthogonal complement of the row spaces of B and C. This is illustrated in Figure 3. The matrix A is decomposed as a linear combination of the rows of B and C and of the rows of a third matrix orthogonal to B and C. This can be written as $A = L_B.B + L_C.C + L_{B^\perp,C^\perp}\begin{pmatrix} B \\ C \end{pmatrix}^\perp$. The row space of the matrix $L_C.C$ is defined as the oblique projection of the row space of A along the row space of B in the row space of C as $A/_B C = L_C.C$. The name *oblique* refers to the non-orthogonal projection direction. 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 C. Mathematically, the orthogonal projection of the row space of A into the joint row space of B and C can be stated as:

$$A/\begin{pmatrix} C \\ B \end{pmatrix} = A\begin{pmatrix} C^T & B^T \end{pmatrix}.\begin{pmatrix} CC^T & CB^T \\ BC^T & BB^T \end{pmatrix}^\dagger.\begin{pmatrix} C \\ B \end{pmatrix}.$$

The oblique projection of the row space of $A \in \mathbb{R}^{p \times j}$ along the row space of $B \in \mathbb{R}^{q \times j}$ into the row space of $C \in \mathbb{R}^{r \times j}$ is then obtained as

$$A/_B C \stackrel{\text{def}}{=} A\begin{pmatrix} C^T & B^T \end{pmatrix}.\left[\begin{pmatrix} CC^T & CB^T \\ BC^T & BB^T \end{pmatrix}^\dagger\right]_{\text{first } r \text{ columns}}.C.$$

Another expression for the oblique projection is given by

$$A/_B C = [A/B^\perp].[C/B^\perp]^\dagger.C. \quad (7)$$

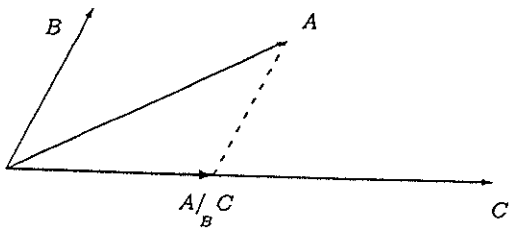


Fig. 3. Interpretation of the oblique projection in the j-dimensional space ($j = 2$ in this case). The oblique projection is formed by projecting the row space of A along the row space of B into the row space of C.

6 Kalman filter states and orthogonal projections

In this section, we first show that the Kalman filter state estimates can be obtained from certain linear combinations of the rows of the input-output block Hankel matrices. Next we show how this state sequence also shows up in the orthogonal projection of the row space of the future outputs into the row spaces of the past outputs and past and future inputs.

6.1 Explicit expressions for the Kalman filter state estimates

In the derivation of the subspace identification algorithms for combined deterministic - stochastic system identification, the Kalman filter plays a crucial role. In this Subsection, we introduce a closed form equation for the Kalman filter state estimate for the combined system. We also introduce a bank of Kalman filters generating a sequence of state estimates. We indicate the Kalman filter state estimate by a hat: $\hat{x}_k$. For the notations, we refer to Section 4.

Theorem 1. Kalman filter and explicit expressions for the state estimate

Let the following be given: An initial state estimate: $\hat{x}_0 = \overline{x}_0$, an initial estimate of the matrix: $P_0 = \mathrm{E}[\overline{x}_0.\overline{x}_0^T]$, the input and output measurements $u_0, y_0, \ldots, u_{k-1}, y_{k-1}$; Then the state estimate $\hat{x}_k$ from the Kalman filter, defined by the following recursive formulas:

$$\hat{x}_k = A\hat{x}_{k-1} + Bu_k + K_{k-1}(y_{k-1} - C\hat{x}_{k-1} - Du_k)\,, \tag{8}$$

$$K_{k-1} = (G - AP_{k-1}C^T)(\Lambda_0 - CP_{k-1}C^T)^{-1}\,, \tag{9}$$

$$P_k = AP_{k-1}A^T + (G - AP_{k-1}C^T) \times (\Lambda_0 - CP_{k-1}C^T)^{-1}(G - AP_{k-1}C^T)^T\,, \tag{10}$$

can be explicitly written as:

$$\hat{x}_k = \left(A^k - \Omega_k\Gamma_k \middle| \Delta_k^d - \Omega_k H_k^d \middle| \Omega_k\right) \left(\begin{array}{c} \overline{x}_0 \\ \hline u_0 \\ \cdots \\ u_{k-1} \\ \hline y_0 \\ \cdots \\ y_{k-1} \end{array}\right)\,, \tag{11}$$

where:

$$\Omega_k \stackrel{def}{=} (\Delta_k^c - A^k P_0 \Gamma_k^T)(L_k - \Gamma_k P_0 \Gamma_k^T)^{-1}\,. \tag{12}$$

The explicit solution of the matrix P_k is equal to:

$$P_k = A^k P_0 (A^T)^k + (\Delta_k^c - A^k P_0 \Gamma_k^T)(L_k - \Gamma_k P_0 \Gamma_k^T)^{-1}(\Delta_k^c - A^k P_0 \Gamma_k^T)^T\,. \tag{13}$$

A proof can be found in [VODM 94a].
The covariance matrix of the state error $\widetilde{P}_k$ is given by:

$$\widetilde{P}_k = \mathrm{E}[\,(x_k - \hat{x}_k).(x_k - \hat{x}_k)^T\,] = \Sigma^s - P_k\,.$$

From this we conclude that the covariance of the initial error on the state estimate is given by $\widetilde{P}_0 = \Sigma^s - P_0$. This indirectly implies that P_0 should be negative definite (or smaller than Σ^s at least).

> The significance of Theorem 1 is that it indicates how the Kalman filter state estimate $\hat{x}_k$ can be written as a linear combination of the past inputs and output measurements $u_0, y_0, \ldots, u_{k-1}, y_{k-1}$ and of the initial state estimate $\overline{x}_0$ if the system matrices A, B, C, D, Q, S and R were known, *which is not the case if only input-output data u_k and y_k are available.*

We now define the state sequence

$$\begin{aligned} \widehat{X}_i &= (\hat{x}_i\ \hat{x}_{i+1}\ \ldots\ \hat{x}_{i+j-1}) \\ &= \left(A^k - \Omega_k\Gamma_k \middle| \Delta_k^d - \Omega_k H_k^d \middle| \Omega_k\right) \left(\begin{array}{c} \overline{X}_0 \\ \hline U_p \\ \hline Y_p \end{array}\right) \\ &= \left(A^k - \Omega_k\Gamma_k \middle| (\Delta_k^d - \Omega_k H_k^d\ \Omega_k)\right) \left(\begin{array}{c} \overline{X}_0 \\ W_p \end{array}\right)\,. \end{aligned} \tag{14}$$

with $\overline{X}_0$ the sequence of initial states. This state sequence is generated by a bank of Kalman filters, working in parallel on each of the columns of the block Hankel matrix of past inputs and outputs, which is illustrated in Figure 4. The Kalman filters run in a *vertical* direction (over the columns). The state estimate is obtained from *partial* input-output information. Each vector in the sequence

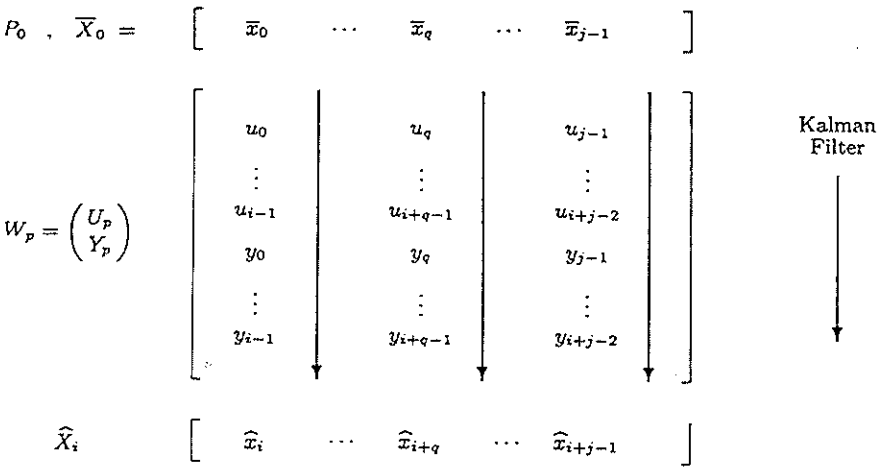


Fig. 4. Interpretation of the sequence $\widehat{X}_i$ as a sequence of Kalman filter state estimates based upon i measurements of u_k and y_k. When the system matrices A, B, C, D, Q, R, S would be known, the state $\widehat{x}_{i+q}$ could be determined from a Kalman filter as follows: start the filter at time q, with an initial state estimate $\overline{x}_q$ and initial error covariance matrix $\Sigma^s - P_0$. Now iterate the Kalman filter over i time steps (the vertical arrow down). The Kalman filter will then return a state estimate $\widehat{x}_{i+q}$. This procedure could be repeated for each of the j columns, and thus we speak about a *bank* of Kalman filters. The major observation in subspace algorithms is that the system matrices A, B, C, D, Q, R, S do not have to be known to determine the state sequence $\widehat{X}_i$. It can be determined directly from input-output data. It is very important to realize that there is NO direct relation between the column vectors in $\widehat{X}_i$ (unless $i \to \infty$). The Kalman filter interpretation only 'works' in the vertical direction, for each column separately, and NOT in the horizontal direction, unless $i \to \infty$!

$\widehat{X}_i$ depends on its corresponding initial state in the sequence $\overline{X}_0$ and the initial covariance matrix P_0 . In what follows we will encounter different Kalman filter sequences (in the sense of different initial states $\overline{X}_0$ and initial matrices P_0). Therefore we will denote the Kalman filter state sequence with initial state $\overline{X}_0$ and initial covariance matrix P_0 by $\widehat{X}_{i[\overline{X}_0, P_0]}$.

6.2 Kalman filter states and orthogonal projections

In this Subsection, we introduce the projection of the future outputs onto the past and future inputs and the past outputs. Through this projection, the Kalman filter states can be determined directly from the data. In the following Theorem, we show that there is a nice mathematical expression for the orthogonal projection of the row space of the future outputs Y_f onto the row space

generated by the past and future inputs and the past outputs. This orthogonal projection is illustrated in Figure 5. Before we proceed, we need the following

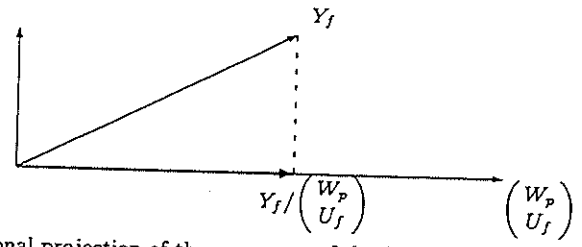


Fig. 5. The orthogonal projection of the row space of the future outputs into the row spaces of past outputs and past and future inputs.

technical definition [Lju 87]:

Definition 2. Persistency of excitation
The input sequence $u_k \in \mathbb{R}^m$ is persistently exciting of order $2i$ if the input covariance matrix $R^{uu} \stackrel{\text{def}}{=} \frac{1}{j}[U_{0|2i-1} U_{0|2i-1}^T]$ is of full row rank, which is $2mi$.

It is tedious though straightforward to prove the following Theorem which extracts Kalman filter state sequences directly from the input-output data.

Theorem 3. Kalman filter and orthogonal projection
Under the assumptions that:

1. *The deterministic input u_k is uncorrelated with the process noise w_k and the measurement noise v_k (see equation (4));*
2. *The input u_k is persistently exciting of order $2i$ (Definition 2) ;*
3. *The number of available data is large: $j \to \infty$;*
4. *The process noise w_k and the measurement noise v_k are not identically zero ;*

Then:

$$\mathcal{Z}_i \stackrel{def}{=} Y_f / \begin{pmatrix} W_p \\ U_f \end{pmatrix} = \Gamma_i \widehat{X}_i + H_i^d U_f \,, \tag{15}$$

with:

$$\widehat{X}_i \stackrel{def}{=} \widehat{X}_{i[\overline{X}_0, P_0]} \,, \tag{16}$$

$$\overline{X}_0 = S^{xu} . (R^{uu})^{-1} . \begin{pmatrix} U_p \\ U_f \end{pmatrix} \,, \tag{17}$$

$$P_0 = -[\Sigma^d - S^{xu} . (R^{uu})^{-1} . (S^{xu})^T] \,. \tag{18}$$

A proof of this Theorem can be found in [VODM 94a].

The importance of Theorem 3 is that it reveals one way in which the Kalman filter state sequence $\widehat{X}_i$ (14) obtained from Theorem 1 relates directly to given input-output data. The projection matrix $\mathcal{Z}_i$ can be computed from the given data, without knowing the system matrices. $\mathcal{Z}_i$ could be considered as an optimal (least squares) prediction of the future output data $Y_{i|2i-1}$, given the past input and output data $U_{0|i-1}$ and $Y_{0|i-1}$ *and* the future input data $U_{i|2i-1}$ (see also Section 7.2).

Examining the formulas for the initial state sequence $\overline{X}_0$ (17) and the initial matrix P_0 (18) a little closer, leads to the following interesting formulas:

$$\overline{X}_0 = X_p^d / \begin{pmatrix} U_p \\ U_f \end{pmatrix}, \tag{19}$$

$$P_0 = -\Phi \left[X_p^d / \begin{pmatrix} U_p \\ U_f \end{pmatrix}^\perp, X_p^d / \begin{pmatrix} U_p \\ U_f \end{pmatrix}^\perp \right]. \tag{20}$$

First note that the "real" initial state would be $X_p^d + X_p^s$. The stochastic part X_p^s is impossible to estimate and is thus set to zero. This explains why X_0^s does not appear in (19). The deterministic part X_p^d of the *real* initial state however enters (19). Note that the row space of $\overline{X}_0$ is a subspace of the combined row spaces of W_p and U_f (since $\mathcal{Z}_i$ is constructed by a projection on the combined row space). From (19) we can see that $\overline{X}_0$ is the best estimate of X_p^d lying in the row space of past and future *inputs*.
From formula (20) we find that the covariance $\widetilde{P}_0$ of the *error* on the initial state estimate is given by:

$$\begin{aligned} \widetilde{P}_0 &= \Sigma^s - P_0 \\ &= \Sigma^s + \Phi \left[X_p^d / \begin{pmatrix} U_p \\ U_f \end{pmatrix}^\perp, X_p^d / \begin{pmatrix} U_p \\ U_f \end{pmatrix}^\perp \right] \\ &= \Phi_{[(X_p^s + X_p^d) - \overline{X}_0, (X_p^s + X_p^d) - \overline{X}_0]}, \end{aligned}$$

which is positive definite. The last equation indicates that the error on the initial state is given by the variance of the part of the *real* initial state $X_p^d + X_p^s$ that does not lie in the combined row space of U_p and U_f. Both equations (17) and (18) can thus be explained intuitively. The initial state estimate is the best estimate of the *real* initial state, lying in the combined row space of U_p and U_f. The initial state error covariance is the covariance of the difference between the *real* initial state and the estimated initial state $\overline{X}_0$. Finally note that when the inputs u_k are white noise the initial state $\overline{X}_0 = 0$, since in this case, there is no correlation between the *real* initial state $X_p^d + X_p^s$ and the inputs U_p and U_f.

7 Kalman filter states and oblique projections

In this Section, we show how an oblique projection with input-output block Hankel matrices forms one of the key observations in subspace system identification algorithms. The oblique projection which is of central interest here, is the projection of the row space of the future outputs Y_f, along the future input row space U_f into the past inputs and outputs row space W_p. This projection is illustrated in Figure 6.

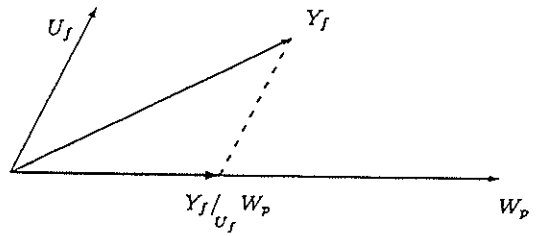


Fig. 6. Interpretation of the oblique projection in the j-dimensional space. The oblique projection is formed by projecting the row space of Y_f along the row space of U_f onto the row space of W_p.

7.1 The main theorem

The next Theorem allows to calculate a Kalman filter state sequence together with the column space of the extended observability matrix Γ_i directly from the input-output data, without any knowledge about the system matrices. The system matrices can then afterwards be extracted from the state sequence $\widehat{X}_i$ or from Γ_i. An overview of the general combined identification procedure is presented in Figure 7.

In what follows, $W_1 \in \mathbb{R}^{li \times li}$ and $W_2 \in \mathbb{R}^{j \times j}$ are user-defined weighting matrices, to which we will come back later.

Theorem 4. Kalman filter states and oblique projections
Under the assumptions that:

1. *The deterministic input u_k is uncorrelated with the process noise w_k and measurement noise v_k (see (4)) ;*
2. *The input u_k is persistently exciting of order $2i$ (Definition 2) ;*
3. *The number of available data is large, so that $j \to \infty$;*
4. *The process noise w_k and the measurement noise v_k are not identically zero ;*
5. *The user-defined weighting matrices $W_1 \in \mathbb{R}^{li \times li}$ and $W_2 \in \mathbb{R}^{j \times j}$ are such that W_1 is of full rank and W_2 obeys (where W_p is the block Hankel matrix containing the past inputs and outputs): rank (W_p) = rank $(W_p W_2)$;*

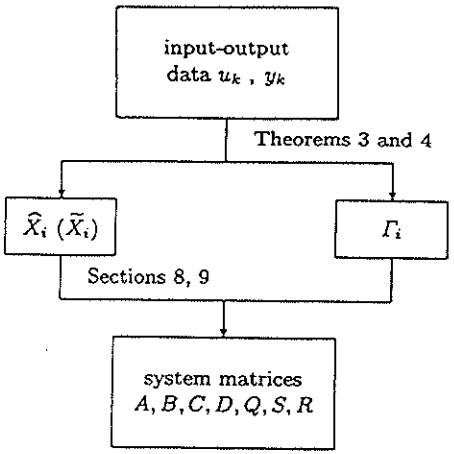


Fig. 7. An overview of the combined deterministic-stochastic subspace identification procedure. Through the Theorems 3 and 4 the state sequence $\widehat{X}_i$ ($\widetilde{X}_i$) and the extended observability matrix Γ_i are determined. The system matrices are then extracted using one of the algorithms described in Sections 8 and 9.

and with $\mathcal{O}_i$ defined as the oblique projection:

$$\mathcal{O}_i \stackrel{def}{=} Y_f /_{U_f} W_p \,, \tag{21}$$

and the singular value decomposition:

$$W_1 \mathcal{O}_i W_2 = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = U_1 S_1 V_1^T \,, \tag{22}$$

we have:

1. *The matrix $\mathcal{O}_i$ is equal to the product of the extended observability matrix and a matrix $\widetilde{X}_i$ with Kalman filter state estimates:*

$$\mathcal{O}_i = \Gamma_i . \widetilde{X}_i \,, \tag{23}$$

with:

$$\begin{aligned} \widetilde{X}_i &\stackrel{def}{=} \widehat{X}_{i[\widehat{X}_0, P_0]} \,, \\ \widehat{X}_0 &= X_p^d /_{U_f} U_p \,, \\ P_0 &= -[\Sigma^d - S^{xu} . (R^{uu})^{-1} . (S^{xu})^T] \,. \end{aligned} \tag{24, 25}$$

2. *The order of the system (1)-(2) is equal to the number of singular values in equation (22) different from zero ;*

3. *The extended observability matrix Γ_i can be chosen as:*

$$\Gamma_i = W_1^{-1} U_1 S_1^{1/2} . T \,. \tag{26}$$

where T is an arbitrary similarity transformation.

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

$$\widetilde{X}_i W_2 = T^{-1} . S_1^{1/2} V_1^T \,. \tag{27}$$

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

$$\widetilde{X}_i = \Gamma_i^{\dagger} . \mathcal{O}_i \,. \tag{28}$$

up to within an arbitrary similarity transformation T.

The proof of this Theorem can be found [VODM 94a].

This Theorem can be summarized algebraically as follows:

$$\begin{aligned} &\text{rank } (Y_f /_{U_f} W_p) = n \,, \\ &\text{range } (Y_f /_{U_f} W_p) = \text{range } (\Gamma_i) \,, \\ &\text{range } ((Y_f /_{U_f} W_p)^T) = \text{range } (\bar{X}_i^T) \,. \end{aligned}$$

Obviously, the singular value decomposition of the matrix $Y_f /_{U_f} W_p$ delivers these three results at once. It should be noted that the state sequence $\widetilde{X}_i$ recovered through this Theorem differs from the state sequence $\widehat{X}_i$ that was introduced in Theorem 3. The two sequences are different due to their different initial state $\widehat{X}_0$. For Theorem 3, we have $\overline{X}_0 = X_p^d / \begin{pmatrix} U_p \\ U_f \end{pmatrix}$ and for Theorem 4 $\widehat{X}_0 = X_p^d /_{U_f} U_p$. This difference in initial states will play a crucial role in the derivation of the algorithms. Finally note that even though their initial state sequence is different, both sequences $\widehat{X}_i$ and $\widetilde{X}_i$ are generated by a bank of Kalman filters.

7.2 Some intuition

In this Subsection we present some intuition behind the different projections of Theorem 3 and 4. More intuitive and rigorous explanations can be found in [VODM 93d]. The goal of an identification procedure is to find a model of which the input/output behavior approximates that of the system under consideration. This goal is classically solved by minimizing a "prediction error criterion" which expresses the "prediction performance" of the model on the given data

set. The minimizing solution is designated as the optimal model (see for instance [Lju 87]). In the framework of subspace identification, the identification goal is attained by solving two subsequent problems:

1. Optimal Prediction: As stated above we want to find a model that will predict the behavior of the process sufficiently accurately. This can be formulated as: predict the future outputs (Y_f) as accurately as possible, using all the information that can be obtained from the past (W_p), and using the knowledge of the inputs that will be presented to the system in the future (U_f).
Inspired by the linearity of the system, we propose to combine the past (W_p) and the future inputs (U_f) linearly to predict the future outputs (Y_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:

$$\min_{\substack{L_p \in \mathbb{R}^{li \times (m+l)i} \\ L_u \in \mathbb{R}^{li \times mi}}} \| Y_f - \left(L_p \; L_u \right) \begin{pmatrix} W_p \\ U_f \end{pmatrix} \|_F^2 . \tag{29}$$

Obviously, this least squares problem leads to the orthogonal projection $\mathcal{Z}_i$ of Theorem 3 and hence provides exactly the desired linear combination of W_p and U_f as $\mathcal{Z}_i = \left(L_p \; L_u \right) \begin{pmatrix} W_p \\ U_f \end{pmatrix}$. The optimal linear combination of the past (W_p) to predict the future is the term $L_p W_p$, which is exactly equal to the oblique projection of Theorem 4 , i.e. $\mathcal{O}_i = L_p . W_p$.

2. Low order model: Apart from the fact that the the model should predict the future, we also want its order to be as small as possible. We thus need to reduce the rank of $\mathcal{O}_i$. Since the rows of $\mathcal{O}_i$ 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:

$$\min_{\mathcal{R} \in \mathbb{R}^{li \times j}} \| W_1 \left[\mathcal{O}_i - \mathcal{R} \right] W_2 \|_F^2 \text{ subject to } \operatorname{rank}(\mathcal{R}) = n , \tag{30}$$

The user-defined weighting matrices W_1 and W_2 determine which part of the "information" of $\mathcal{O}_i$ is important to retain. A more rigorous interpretation of the weighting matrices is presented in [VODM 93d] [VODM 94d]. It is easy to show [VODM 93d] that $\mathcal{R}$ is determined through the singular value decomposition (22) as $\mathcal{R} = W_1^{-1} U_1 S_1 V_1^T W_2^{\dagger}$. Note that $\mathcal{R} = \mathcal{O}_i$ when all the assumptions of Theorem 4 are satisfied exactly (which is however never the case with real-life data). However, when $j \neq \infty$ or when the data generating system is not linear, the singular values of $W_1 \mathcal{O}_i W_2$ (22) are all different from zero. In that case, the row space of $\mathcal{O}_i$ is of dimension li, and the order has to be chosen equal to the number of "dominant" singular values. The complexity reduction step is then truly a reduction of the dimension of the row space of $\mathcal{O}_i$, and the weights W_1 and W_2 play an important role in determining which part of the original row space of $\mathcal{O}_i$ is retained.

7.3 Different weights, different algorithms

One might wonder about the effect of chosing the weights W_1 and W_2 in the singular value decomposition of Theorem 4. Without going into details here, it suffices to say that for certain appropriate choices of W_1 and W_2, one recovers some subspace algorithms that were recently proposed and more or less studied in the literature. Without going into detail, we present in the following table the acronyms and refererences of these algorithms and the weights to be plugged in into Theorem 4 to recover them

Acronym	Full name and Main ref.	W_1	W_2
N4SID	Numerical Algorithm for Subspace State Space System Identification [VODM 94a] [VOWL 93]	I_{li}	I_j
MOESP	Multivariable Output-Error State sPace [Ver 94]	I_{li}	$\Pi_{U_f^\perp}$
CVA	Canonical Variate Analysis [Lar 90]	$(Y_f/U_f^\perp)^T . (Y_f/U_f^\perp)$	$\Pi_{U_f^\perp}$

Our main contribution here is that we have unified these different approaches and put them on an equal footing in the sense that, in one way or another (i.e. implicitly or explicitly) they all recover the state sequence and the observability matrix. We refer to [VODM 93d] for proofs and details.

8 From the states to the state space model

What have we done so far? In Section 6 we have found a relation between the Kalman filter state sequence $\widehat{X}_i$ (14) and the orthogonal projection of the row space of Y_f (future outputs) into the row spaces of past and future inputs and past outputs. In this Section, we discuss an identification algorithm that uses the Kalman filter states as obtained from these orthogonal projections.

8.1 Shifting the state sequence

By the shifted state sequence $\widehat{X}_i$ we mean the state sequence (14) shifted to the right, i.e. starting with $\hat{x}_{i+1}$ and ending with $\hat{x}_{i+j}$. It turns out that this shifted sequence can be obtained from another orthogonal projection of shifted future outputs into shifted past inputs and outputs and future inputs. Why would we try to obtain this shifted state sequence? Because it will give us an elegant way to obtain the state space model A, B, C, D and an estimate of the noise covariance matrices Q, R and S, by just solving a linear least squares problem! This can seen from (1)-(2) and (4).
From Theorem 4, we find:

- The order of the system from inspection of the singular values of equations (22).
- The extended observability matrix Γ_i from equation (26) and the matrix Γ_{i-1} as $\underline{\Gamma_i}$ [7].

The following side result of Theorem 3 can easily be proven. By shifting the separation line between "past" and "future" one block row downwards, we obtain the matrices W_p^+, U_f^- and Y_f^- (notations, see Section 4). A similar orthogonal projection as in Theorem 3 can now be performed with these matrices. This leads to the sequence Z_{i+1} and the Kalman filter states $\widehat{X}_{i+1}$:

$$Z_{i+1} = Y_f^- / \begin{pmatrix} W_p^+ \\ U_f^- \end{pmatrix} = \Gamma_{i-1}\widehat{X}_{i+1} + H_{i-1}^d U_f^- \,, \tag{31}$$

$$\widehat{X}_{i+1} = \widehat{X}_{i+1[\overline{X}_0, P_0]} \,, \tag{32}$$

with $\overline{X}_0$ and P_0 given by equations (17)-(18). There is a very important observation to be made here:

> Corresponding columns of $\widehat{X}_i$ (16) and of $\widehat{X}_{i+1}$ (32) are consecutive state estimates of the same Kalman filters at two consecutive time instants, in the sense that the associated Kalman filters have the same initial state estimate $\overline{X}_0$ and the same initial error covariance P_0.

This statement is not trivial at all and is proven in [VODM 94a]! It implies that we can write with (8):

$$\widehat{X}_{i+1} = A\widehat{X}_i + BU_{i|i} + K_i(Y_{i|i} - C\widehat{X}_i - DU_{i|i}) \,. \tag{33}$$

It is also true that:

$$Y_{i|i} = C\widehat{X}_i + DU_{i|i} + (Y_{i|i} - C\widehat{X}_i - DU_{i|i}) \,. \tag{34}$$

It is now easy to prove (see [VODM 94a]) that the row space of $Y_{i|i} - C\widehat{X}_i - DU_{i|i}$ is orthogonal to the row spaces of W_p, U_f and $\widehat{X}_i$. This result can also be intuitively proven by noticing that the innovations $(Y_{i|i} - C\widehat{X}_i - DU_{i|i})$ of a Kalman filter are uncorrelated with the states $\widehat{X}_i$, the past inputs and outputs W_p and the future inputs U_f. We thus have:

$$\mathbf{E_j}(Y_{i|i} - C\widehat{X}_i - DU_{i|i}).\begin{pmatrix} W_p \\ U_f \\ \widehat{X}_i \end{pmatrix}^T = 0 \,.$$

[7] A 'bar' under a symbol means that the matrix $\underline{\Gamma_i}$ is obtained by omitting the last block row of Γ_i.

This implies that (33)-(34) can be written as:

$$\begin{pmatrix} \widehat{X}_{i+1} \\ Y_{i|i} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \widehat{X}_i \\ U_{i|i} \end{pmatrix} + \begin{pmatrix} \rho_w \\ \rho_v \end{pmatrix} \,, \tag{35}$$

with obvious definitions for ρ_w and ρ_v as residual matrices, the row spaces of which are orthogonal to the row space of W_p, U_f and $\widehat{X}_i$. If we would now be able to compute the state sequences $\widehat{X}_i$ and $\widehat{X}_{i+1}$ from the input-output data, we could solve equation (35) in a least squares sense for the system matrices A, B, C, D. The matrices ρ_w and ρ_v would then contain the least squares residuals. Unfortunately, it is not possible to determine the state sequences $\widehat{X}_i$ and $\widehat{X}_{i+1}$ directly from the input-output data[8]. However, from (15) and (31) we can determine $\widehat{X}_i$ and $\widehat{X}_{i+1}$ as:

$$\widehat{X}_i = \Gamma_i^\dagger \,.\, [Z_i - H_i^d.U_f] \,, \tag{36}$$

$$\widehat{X}_{i+1} = \Gamma_{i-1}^\dagger \,.\, [Z_{i+1} - H_{i-1}^d.U_f^-] \,. \tag{37}$$

In these formulas the only unknowns on the right hand side are H_i^d and H_{i-1}^d, since Z_i and Z_{i+1} can be determined as a projection of the input-output block Hankel matrices and Γ_i and Γ_{i-1} are determined through Theorem 4. Substitution of (36) and (37) into (35) leads to:

$$\begin{pmatrix} \Gamma_{i-1}^\dagger.Z_{i+1} \\ \hline Y_{i|i} \end{pmatrix} = \underbrace{\begin{pmatrix} A \\ \hline C \end{pmatrix}}_{\text{term 1}} .\Gamma_i^\dagger.Z_i + \underbrace{\mathcal{K}}_{\text{term 2}} .U_f + \underbrace{\begin{pmatrix} \rho_w \\ \hline \rho_v \end{pmatrix}}_{\text{term 3}} \,, \tag{38}$$

where we have defined:

$$\mathcal{K} \stackrel{\text{def}}{=} \begin{pmatrix} (B|\Gamma_{i-1}^\dagger.H_{i-1}^d) - A.\Gamma_i^\dagger.H_i^d \\ (D|0) - C.\Gamma_i^\dagger.H_i^d \end{pmatrix} \,. \tag{39}$$

8.2 Determining the state space matrices

Observe that the matrices B and D appear linearly in the matrix $\mathcal{K}$. We can now solve equation (38) in a least squares sense for A,C and $\mathcal{K}$. Since the row spaces of ρ_w and ρ_v have been shown to be orthogonal to the row spaces of Z_i and U_f and since the least squares solution computes residuals that are orthogonal to the regressors, the least squares solution will compute asymptotically unbiased estimates of the system matrices (see [VODM 94a] for details). From (38) we find in this way (term by term):

term 1. The system matrices A and C.

[8] For all clarity, Theorem 4 determines the state sequence $\widetilde{X}_i$ directly from the input-output data. This sequence is different from $\widehat{X}_i$, in a sense that the initial conditions of the Kalman filter are different (see the discussion at the end of Subsection 7.1). We will use the sequence $\widetilde{X}_i$ in the second algorithm.

term 2. The matrix $\mathcal{K}$ from which B and D can be computed. This leads to a least squares problem for B and D, details of which can be found in [VODM 94a] [VO PhD]. Actually, one can show that

> **The estimates of A, B, C and D are consistent as $j \to \infty$, provided that the system that generated the data belongs to the model class described by (1)-(2)-(4).**

term 3. The covariances Q, S and R can be estimated from the residuals ρ_w and ρ_v as:

$$\begin{pmatrix} Q_i & S_i \\ S_i^T & R_i \end{pmatrix} \simeq \mathbf{E_j}[\begin{pmatrix} \rho_w \\ \rho_v \end{pmatrix} . (\rho_w^T \ \rho_v^T)] .$$

This result is biased for finite i as indicated by the subscript i in the estimates Q_i, R_i and S_i. The reason is the fact that the bank of Kalman filters, which was discussed in Figure 4 is not in steady state for finite i (meaning that the Riccati difference equation (10) has not converged yet). As i grows larger, the approximation error grows smaller. For infinite i and j the stochastic system is determined asymptotically unbiased. However, as is obvious by construction, the matrix $\begin{pmatrix} Q_i & S_i \\ S_i^T & R_i \end{pmatrix}$ is guaranteed to be positive definite, which implies the following

> **Fact: The noise model is guaranteed to be positive real, but its covariance matrix estimates Q_i, S_i and R_i are biased for finite i.**

The steps of this first algorithm are summarized in Figure 8.

9 A biased subspace identification algorithm

The algorithm we have just presented, computes unbiased estimates of the system matrices A, B, C and D and biased ones for Q, R and S (unless $i \to \infty$). However, this algorithm is quite complicated (especially the extraction of B and D). One could wonder if there does not exist an algorithm in which the estimation of B and D does NOT require the solution of an additional least squares problem. In turns out that there is such an approach, which uses the the Kalman filter state sequence $\widetilde{X}_i$ (24) found in Section 7 from the oblique projection of the row space of the future outputs, along that of the future inputs, into the row space of past inputs and outputs. However, this algorithm calculates asymptotically biased solutions (see also [VODM 94a]). From a similar reasoning as in Theorem 4, we find that:

$$\mathcal{O}_{i+1} = Y_f^- /_{U_f^-} W_p^+ = \Gamma_{i-1} . \widetilde{X}_{i+1} ,$$

and we now have $\widetilde{X}_i$ and $\widetilde{X}_{i+1}$. The problem however is that this new Kalman filter sequence $\widetilde{X}_{i+1}$ has a different initial state as the sequence $\widetilde{X}_i$ and hence

Algorithm using the states:

1. Calculate the oblique and orthogonal projections:

$$\mathcal{O}_i = Y_f /_{U_f} W_p ,$$

$$\mathcal{Z}_i = Y_f / \begin{pmatrix} W_p \\ U_f \end{pmatrix} ,$$

$$\mathcal{Z}_{i+1} = Y_f^- / \begin{pmatrix} W_p^+ \\ U_f^- \end{pmatrix} .$$

2. Calculate the SVD of the weighted oblique projection $W_1 \mathcal{O}_i W_2 = USV^T$. where W_1 and W_2 are user-defined weighting matrices.
3. Determine the order by inspecting the singular values in S and partition the SVD accordingly to obtain U_1 and S_1.
4. Determine Γ_i and Γ_{i-1} as $\Gamma_i = W_1^{-1} U_1 S_1^{1/2}$ and $\Gamma_{i-1} = \underline{\Gamma_i}$.
5. Solve the set of linear equations for A, C and $\mathcal{K}$ (using least squares) which also gives residuals ρ_w and ρ_v:

$$\begin{pmatrix} \Gamma_{i-1}^{\dagger} . \mathcal{Z}_{i+1} \\ \hline Y_{i|i} \end{pmatrix} = \begin{pmatrix} A \\ \hline C \end{pmatrix} . \Gamma_i^{\dagger} . \mathcal{Z}_i + \mathcal{K} . U_f + \begin{pmatrix} \rho_w \\ \hline \rho_v \end{pmatrix} .$$

6. Determine B and D from $\mathcal{K}, A, C, \Gamma_i, \Gamma_{i-1}$ as described in Subsection 8.2.
7. Determine Q_i, S_i and R_i from the residuals as:

$$\begin{pmatrix} Q_i & S_i \\ S_i^T & R_i \end{pmatrix} = \mathbf{E_j}[\begin{pmatrix} \rho_w \\ \rho_v \end{pmatrix} . (\rho_w^T \ \rho_v^T)] .$$

Fig. 8. A schematic overview of the first combined deterministic-stochastic identification algorithm. See Section 10 for implementation issues.

is not just a 'shifted' version of $\widetilde{X}_i$. Indeed, we have as an initial state for $\widetilde{X}_i$, $X_p^d /_{U_f} U_p$, while the initial state for $\widetilde{X}_{i+1}$ is $X_p^d /_{U_f^-} U_p^+$.

So, we can not write a formula similar to (35) with $\widehat{X}_i$ and $\widehat{X}_{i+1}$ replaced by $\widetilde{X}_i$ and $\widetilde{X}_{i+1}$!

It can be proven however (see [VODM 94a]) that the difference between $\widehat{X}_i$ and $\widetilde{X}_i$ is zero when at least one of the following conditions is satisfied:

1. $i \to \infty$. The difference between $\widehat{X}_i$ and $\widetilde{X}_i$ goes to zero at the same rate as the Riccati difference equation of the Kalman filter converges. This is intuitively clear, since by the time the Kalman filter is in steady state, the effect of the initial conditions has died out. In [VODM 94a] a rigorous proof

is given.

2. The system is purely deterministic.
3. The deterministic input u_k is white noise. In this case the deterministic state X_p^d is uncorrelated with U_p and U_f. This implies that the initial state sequences of $\widehat{X}_i, \widehat{X}_{i+1}, \overline{X}_i$ and $\overline{X}_{i+1}$ are all equal to zero (and are thus equal to each other).

When at least one of these conditions is satisfied, we can replace $\widehat{X}_i$ and $\widehat{X}_{i+1}$ in (35) with $\widetilde{X}_i$ and $\widetilde{X}_{i+1}$, and solve for A, B, C and D in a least squares sense, which will then give an asymptotically unbiased estimate. The details of this simple algorithm are summarized in Figure 9. If none of the above conditions is satisfied, this second algorithm will return biased estimates of the system matrices (since the states are replaced by approximations). In [VODM 94a] an expression for this bias is derived. Even though the algorithm is simpler than the first one we have presented, it turns out that it should be used with care in practice. For many practical measurements the input signal u_k is anything but white noise (steps, impulses, ...). This implies that the algorithm is biased, and this bias can be significant. Our experience with lots of practical data, is that, if computational complexity is no real problem, the first algorithm is to be preferred over the second one, but that the robust version that we will treat next, beats both of them.

10 Engineering a robust identification algorithm

The reader should be aware of the fact that the outlines of the algorithms presented so far only form the central backbone of a more sophisticated implementation. Details can be found in [VODM 94a]. In this Section, we will however not discuss these important alterations. Instead we will show in some detail how the oblique and orthogonal projections can be implemented in a numerically robust manner using the LQ-decomposition of the input-output block Hankel matrices.

10.1 An LQ-decomposition

The common factor in the implementation of all the subspace algorithms is the LQ-decomposition (see [GVL 89]) of the block Hankel matrix formed of the input and output measurements[9]:

$$\mathcal{H} \stackrel{\text{def}}{=} \frac{1}{\sqrt{j}} \underbrace{\begin{pmatrix} U_{0|2i-1} \\ Y_{0|2i-1} \end{pmatrix}}_{\in \mathbb{R}^{2(m+l)i \times j}} = L.Q^T ,$$

[9] The scalar $1/\sqrt{j}$ is used to be conform with the definition of $\mathbf{E_j}$.

with $Q \in \mathbb{R}^{j \times 2(m+l)i}$ orthonormal ($Q^TQ = I_{2(m+l)i}$) and $L \in \mathbb{R}^{2(m+l)i \times 2(m+l)i}$ lower triangular [10]. The LQ-decomposition is partitioned as follows:

$$\begin{pmatrix} U_{0|2i-1} \\ Y_{0|2i-1} \end{pmatrix} = \begin{matrix} mi \\ mi \\ li \\ li \end{matrix} \overset{j}{\begin{pmatrix} U_p \\ U_f \\ Y_p \\ Y_f \end{pmatrix}} = \begin{matrix} m(i+1) \\ m(i-1) \\ l(i+1) \\ l(i-1) \end{matrix} \overset{j}{\begin{pmatrix} U_p^+ \\ U_f^- \\ Y_p^+ \\ Y_f^- \end{pmatrix}} = \begin{matrix} mi \\ m \\ m(i-1) \\ li \\ l \\ l(i-1) \end{matrix} \overset{j}{\begin{pmatrix} U_{0|i-1} \\ U_{i|i} \\ U_{i+1|2i-1} \\ Y_{0|i-1} \\ Y_{i|i} \\ Y_{i+1|2i-1} \end{pmatrix}}$$

$$= L\,Q^T = \begin{matrix} mi \\ m \\ m(i-1) \\ li \\ l \\ l(i-1) \end{matrix} \overset{\begin{matrix} mi & m & m(i-1) & li & l & l(i-1) \end{matrix}}{\begin{pmatrix} L_{11} & 0 & 0 & 0 & 0 & 0 \\ L_{21} & L_{22} & 0 & 0 & 0 & 0 \\ L_{31} & L_{32} & L_{33} & 0 & 0 & 0 \\ L_{41} & L_{42} & L_{43} & L_{44} & 0 & 0 \\ L_{51} & L_{52} & L_{53} & L_{54} & L_{55} & 0 \\ L_{61} & L_{62} & L_{63} & L_{64} & L_{65} & L_{66} \end{pmatrix}} \overset{j}{\begin{pmatrix} Q_1^T \\ Q_2^T \\ Q_3^T \\ Q_4^T \\ Q_5^T \\ Q_6^T \end{pmatrix}} .$$

We will use the shorthand Matlab notation $L_{[4:6],[1:3]}$ for the submatrix of L consisting of *block* rows 4 to 6 and *block* columns 1 to 3, and the shorthand $Q_{3:4}^T$ in a similar way. For instance[11]:

$$L_{[4:5],[4:6]} = \begin{pmatrix} L_{44} & 0 & 0 \\ L_{54} & L_{55} & L_{56} \end{pmatrix} \text{, and also } L_{[1,4],[1,3]} = \begin{pmatrix} L_{11} & 0 \\ L_{41} & L_{43} \end{pmatrix} .$$

10.2 Expressions for the geometric operations

In this Subsection we give expressions for the geometric operations introduced in Section 5 in terms of the LQ-decomposition of the previous Subsection. Again, in this Section, A and B are *dummy* matrices (not to be confused with the model matrices).

Orthogonal projections Orthogonal projections can be easily expressed in function of the LQ-decomposition. We first treat the general case A/B, where A and B consist of any non-overlapping selection of rows of $\mathcal{H}$, which implies they can be expressed as linear combinations of the matrix Q^T as:

$$\begin{aligned} A &= L_A Q^T , \\ B &= L_B Q^T , \end{aligned} \tag{40}$$

[10] Again, there might be some confusion in the notation here as Q is also one of the covariance matrices in the noise model. However, it will always be clear from the context which Q is actually meant. We stick to the notation Q here mainly because of a deeply rooted historical tradition in numerical analysis to call the numerical result of a Gram-Schmidt orthogonalization procedure applied to the rows of a matrix the LQ-decomposition.

[11] By convention, this Matlab subscripting, has priority over transposition $L_{[1,4],[1,3]}^T = \left[L_{[1,4],[1,3]}\right]^T$, $Q_{3:4}^T = [Q_{3:4}]^T$.

for certain matrices L_A and L_B (consisting of a non-overlapping selection of rows of L). We thus get:

$$\begin{aligned} A/B &= \Phi_{[A,B]}.\Phi^{\dagger}_{[B,B]}.B = [L_A Q^T Q L_B^T].[L_B Q^T Q L_B^T]^{\dagger}.L_B Q^T \\ &= L_A L_B^T.[L_B L_B^T]^{\dagger}.L_B Q^T \ . \end{aligned}$$

The important observation to be made here is *that the factor Q does not really play a role in these derivations.* One only needs to do certain operations (products and inverses) with certain submatrices of the L-factor. In many cases, this can be even further simplified. Consider for instance the expression for $\mathcal{Z}_i$ (see Equation (15)): Identifying this with (40), we have:

$$L_A = L_{[5:6],[1:6]} \ , \ \ L_B = L_{[1:4],[1:6]} \ ,$$

which, when L_B is of full row rank, simplifies to

$$\begin{aligned} \mathcal{Z}_i &= L_{[5:6],[1:6]} L^T_{[1:4],[1:6]} [L_{[1:4],[1:6]} L^T_{[1:4],[1:6]}]^{-1} L_{[1:4],[1:6]} Q^T \\ &= L_{[5:6],[1:4]} L^T_{[1:4],[1:4]} [L_{[1:4],[1:4]} L^T_{[1:4],[1:4]}]^{-1} L_{[1:4],[1:4]} Q^T_{1:4} \\ &= L_{[5:6],[1:4]} \cdot \underbrace{L^T_{[1:4],[1:4]} L^{-t}_{[1:4],[1:4]}}_{=I} \cdot \underbrace{L^{-1}_{[1:4],[1:4]} L_{[1:4],[1:4]}}_{=I} \cdot Q^T_{1:4} \\ &= L_{[5:6],[1:4]} Q^T_{1:4} \ , \end{aligned}$$

which is a very simple expression for $\mathcal{Z}_i$ indeed! Similarly, we find

$$\mathcal{Z}_{i+1} = L_{[6],[1:5]} \ Q^T_{1:4} \ ,$$

which again, is very simple.

Oblique projections The oblique projection can also be written in function of the LQ decomposition. For instance with A, B and C being matrices the rows of which are a non-overlapping selection of the rows of $\mathcal{H}$ [12]:

$$A = L_A Q^T \ , \ B = L_B Q^T \ , \ C = R_C Q^T \ ,$$

where L_A, L_B and L_C are certain non-overlapping rows of the L-factor of $\mathcal{H}$, we find that

$$\begin{aligned} A/B^{\perp} &= L_A [I - L_B^T [L_B L_B^T]^{\dagger} L_B].Q^T \ , \\ C/B^{\perp} &= L_C [I - L_B^T [L_B L_B^T]^{\dagger} L_B].Q^T \ , \end{aligned}$$

and with the orthogonal projection operator being idempotent ($\Pi_{B^{\perp}}.\Pi_{B^{\perp}} = \Pi_{B^{\perp}}$), we find through formula (7) for the oblique projection:

$$\begin{aligned} A/_B C &= A/B^{\perp}.\left[C/B^{\perp}\right]^{\dagger}.C \\ &= L_A [I - L_B^T [L_B L_B^T]^{\dagger} L_B] L_C^T \times \left[L_C [I - L_B^T [L_B L_B^T]^{\dagger} L_B] L_C^T\right]^{\dagger} \ . \ L_C Q \ . \end{aligned}$$

[12] Again, A, B and C are *dummy* matrices of this section, not to be confused with the model matrices

For instance with:

$$A = Y_f = L_{[5:6],[1:6]} Q^T \ , B = U_f = L_{[2:3],[1:6]} Q^T \ , C = W_p = L_{[1,4],[1:6]} Q^T \ ,$$

we could compute the oblique projection $Y_f/_{U_f} W_p$ using the above formula. Once again, observe that the matrix Q does not really 'participate' in these expressions. Even though this would lead to a valid expression for the oblique projection, there is a better way to calculate this quantity by first projecting Y_f into the row space of the past outputs and the past and future inputs, and then separating the effect of the future inputs U_f out of this projection (see also Section 5). This can be done as follows: With L_{U_p}, L_{U_f} and L_{Y_p} defined as:

$$\left(\underbrace{L_{U_p}}_{\in \mathbb{R}^{li \times mi}} \ \underbrace{L_{U_f}}_{\in \mathbb{R}^{li \times mi}} \ \underbrace{L_{Y_p}}_{\in \mathbb{R}^{li \times li}} \right) \stackrel{\text{def}}{=} L_{[5:6],[1:4]} L^{\dagger}_{[1:4],[1:4]} \ ,$$

we get for the oblique projection:

$$Y_f/_{U_p} W_p = \left[L_{U_p}.L_{[1:1],[1:4]} + L_{Y_p}.L_{[4:4],[1:4]}\right] Q^T_{1:4} \ .$$

This computation is significantly faster than the previous one, since (when $L_{[1:4],[1:4]}$ is of full rank) the matrices L_{U_p}, L_{U_f} and L_{Y_p} can be computed using back-substitution (since $L_{[1:4],[1:4]}$ is a lower triangular matrix).

10.3 An implementation of the robust identification algorithm

To illustrate the simplicity of using the LQ-decomposition when implementing subspace identification algorithms, for completeness we present here one version of the subspace algorithms which has proven its usefulness on practical applications, without going into the rationales for the different steps. In this algorithm, we need an expression for the following projected oblique projection:

$$\mathcal{O}_i \Pi_{U_f^{\perp}} = \left[L_{U_p}.L_{[1:1],[1:4]} + L_{Y_p}.L_{[4:4],[1:4]}\right] Q^T_{1:4}$$

$$\times Q \left[I_{2(m+l)i} - \begin{pmatrix} L^T_{[2:3],[1:3]} \\ 0 \end{pmatrix} [L_{[2:3],[1:3]} L^T_{[2:3],[1:3]}]^{\dagger} \begin{pmatrix} L_{[2:3],[1:3]} & 0 \end{pmatrix} \right] Q^T \ .$$

With $\Pi = I_{2mi} - L^T_{[2:3],[1:3]} [L_{[2:3],[1:3]} L^T_{[2:3],[1:3]}]^{-1} L_{[2:3],[1:3]}$, this leads to:

$$\mathcal{O}_i \Pi_{U_f^{\perp}} = \left((L_{U_p} L_{[1:1],[1:3]} + L_{Y_p} L_{[4:4],[1:3]}).\Pi | L_{Y_p} L_{44} \right) Q^T_{1:4} \ .$$

The other steps of the implementation are straightforward (except maybe for the step where B and D are determined). The overall implementation is illustrated in Figure 10.

Let us summarize that, by using the LQ-decomposition of the data matrix with input-output data, we achieve a drastic reduction in computational complexity. The Q factor, which has a dimension of $j \times 2(l+m)i$, never needs to be computed! It cancels out in all the expressions we need to obtain the state space model. Only the L-factor is required, the dimension of which is $2(m+l)i \times 2(m+l)i$, which is substantially smaller that Q. In brief, the LQ-decomposition, in which only L is computed, acts as an important data compression step. All quantities of interest can be calculated from L only.

11 Conclusions

In this paper, we have tackled the problem of multivariable system identification for multiple input multiple output, linear, combined deterministic-stochastic systems. Such models often provide good engineering models for real industrial plants, especially for design of model-based controllers. By combining insights, concepts and algorithms from system theory, (numerical) linear algebra and geometry, we have developed a new breed of system identification techniques, called subspace algorithms, that do not suffer from some shortcomings of 'classical' (prediction-error-method) identification approaches.

There are many extensions, special cases or properties that cannot be discussed in this paper due to space (and time) constraints. We restrict ourselves here to a brief literature survey: Some historical developments relating to the roots of subspace algorithms are surveyed in [DMVO 94], including realization theory [DMK 74a] [DKM 74b] [HK 66] [Kun 78] [Moo 81] [MR 76] [Sil 71] [ZM 74], deterministic system identification [Gop 69] [Bud 71] [LS 77] [DMo 88] [Wil 86] [DMV 87], stochastic realization theory [Aka 74] [Aka 75] [Aka 76] [Aok 87] [AK 90] [Cai 88] [DP 84] [DKP 85] [Fau 76] [VODM 93a]. Extensions are described for descriptor systems in [MDMRT 92], continuous time systems [MDMV 91], identification problems with known noise structure [MV 90], periodic systems [He 93] [Verr 94]) and frequency domain versions [McK 94b]. Further properties and interpretation can be found in [MDM 92] [JW 94] [VODM 93d] [MR 93] [OV 94] [VODM 94d] [LP 93] [LP 94]. Results on the statistical analysis can be found in [Vib 94] [DPS 94] [Lar 94] [VOWL 91] and in the context of array signal processing (which is much related to the system identification problem) [Ott 89] [OVSN 93] [OVK 92] [SROK 92] [Vib 89] [VO 91]. Software issues are treated in [AMKMVO 93] [AKVODMB 93] [VODMBAK 94] [VODMAKB 94], [Lju 91c]. Successful applications of subspace identifications methods are reported in vibrational analysis [AML 94], rapid thermal processing of VLSI silicium wafers [Gyu 93] [Cho 93] [CK 93], modelling of industrial processes [FVOMHL 94]

[VODM 93c] [ZVODML 94] and direction-of-arrival estimation for broad band sources [VP 93]. Relations to other identification methods, such as the Eigensystem Realization Algorithm (ERA, [Jua 85], [Jua 94]), Q-Markov covers [AnSk 88] [LSS 92] [Kin 88] [Ske 94], Observability Range Extraction [LS 91] [Liu 92] [LJM 94], etc... still need to be explored. Further research to speed up the calculations by exploitation of the (block) Hankel matrix structure using the notion of displacement rank, is reported in [CXK 94a] [CXK 94b] [CK 95] [Cho 93]. Another interesting idea is to apply subspace algorithms as a first initial guess to start up a nonlinear optimization problem, such as in [Chu 94] [Lju 91b] [Lju 91c].

Finally, there is one problem (which at first sight has to do with elegance but might also have more profound implications) which has bothered us for a long time. The algorithms we have presented here are *asymmetric* in two respects: One needs to make an a priori distinction between inputs and outputs, even if one wants to determine the states only from inputs and outputs (Indeed in the oblique projection it is Y_f that is projected along U_f onto W_p.) The second asymmetry has to do with 'time' and 'causality'. Indeed, it are the 'future' outputs Y_f that are projected onto the past inputs/outputs and the future inputs. Both of these asymmetries do not exist in the pure deterministic and stochastic cases (at least not in the first step where the state is estimated from inputs and outputs, without knowing the model). At present it remains an open problem whether such a double 'symmetric' approach, which in spirit would come very close to Willems' work [Wil 86], is possible.

12 Acknowledgments

We would like to thank Professors Gene Golub, Thomas Kailath, Robert Kosut and Stephen Boyd, all from Stanford University (California) for the many occasions at which we were their guest. That a real company is something completely different from our protected universitarian environment, was taught to us convincingly by Henk Aling and Alexandra Schmidt from Integrated Systems Incorporated, Santa Clara, California. The Swedish gang, consisting of Professor Lennart Ljung from Linköping University, professors Anders Lindquist, Björn Ottersten and Bo Wahlberg (Royal Institute of Technology, Stockholm) and professor Mats Viberg (Chalmers University, Göteborg) have shared with us lots of ideas and suggestions, many of which still need to be worked out. Professors Jan Maciejowski (Cambridge University, UK), Jan Willems en Jan Van Schuppen (Rijksuniversiteit Groningen) and Michel Gevers (UCL, Belgium) have contributed a lot in creating a stimulating European Research Network on System Identification.

References

[AML 94] Abrahamsson T., McKelvey T., Ljung L. *A study of some approaches to vibration data analysis.* Proc. of SYSID '94, Vol. 3, 4-6 July, Copenhagen, Denmark, pp.289-294, 1994.

[Aka 74] Akaike H. *Stochastic theory of minimal realization.* IEEE Transactions on Automatic Control, 19, pp. 667-674, 1974.

[Aka 75] Akaike H. *Markovian representation of stochastic processes by canonical variables.* Siam J. Control, Vol. 13, no.1, pp. 162-173, 1975.

[Aka 76] Akaike H. *Canonical correlation analysis of time series and the use of an information criterion.* System Identification: Advances and case studies, R.K. Mehra and D.G. Lainiotis eds., New York, Academic Press, pp.27-96, 1976.

[AMKMVO 93] Aling H., Milletti U., Kosut R.L., Mesaros M.P., Van Overschee P., De Moor B. *An interactive system identification module for Xmath.* Proc. of the American Control Conference, June 2-4, San Francisco, USA, pp. 3071-3075, 1993.

[AKVODMB 93] Aling H., Kosut R., Van Overschee P., De Moor B., Boyd S. *Xmath Interactive System Identification Module, Part 1.* Integrated Systems Inc., Santa Clara, USA, 1993.

[AnSk 88] Anderson B.D.O., Skelton R.E. *The generation of all q-Markov covers.* IEEE Trans. on Circuits and Systems, 35, pp.375-384, 1988.

[Aok 87] Aoki M. *State space modeling of time series.* Springer Verlag, Berlin, 1987.

[AK 90] Arun K.S., Kung S.Y. *Balanced approximation of stochastic systems.* SIAM J. Matrix Analysis and Applications, 11, pp. 42-68, 1990.

[AE 71] Åström K., Eykhoff P. *System identification - A survey.* Automatica, Vol. 7, pp. 123-167, 1971.

[AW 84] Åström K., Wittenmark B. *Computer Controlled Systems: Theory and Design.* Prentice Hall, 1984.

[BJ 76] Box G.E., Jenkins G.M. *Time series analysis, forecasting and control.* Revised edition, Holden-Day series in time series analysis and digital processing, Holden-Day, Oackland, 1976.

[Bud 71] Budin M. *Minimal realization of discrete linear systems from input-output observations.* IEEE Transactions on Automatic Control, Vol. AC-16, no. 5, pp. 395-401, 1971.

[Cai 88] Caines P. *Linear Stochastic Systems.* Wiley Series in Probability and Mathematical Statistics, 1988.

[Cho 93] Cho Y.M. *Fast subspace based system identification: Theory and practice.* Ph.D. Thesis, Information Systems Lab, Stanford University, CA, USA, 1993.

[CK 93] Cho. Y.M., Kailath T. *Model Identification in Rapid Thermal Processing Systems.* IEEE Trans. on Semicond. Manuf., Vol. 6, no. 3, 1993.

[CXK 94a] Cho Y.M., Xu G., Kailath T. *Fast Recursive Identification of State Space Models via Exploitation of Displacement Structure.* Automatica, Special Issue on Statistical Signal Processing and Control, Vol. 30, no. 1, pp. 45-59, 1994.

[CXK 94b] Cho Y.M., Xu G., Kailath T. *Fast Identification of State Space Models via Exploitation of Displacement Structure.* IEEE Transactions on Automatic Control, Vol. AC-39, no. 10, 1994.

[CK 95] Cho Y.M., Kailath T. *Fast Subspace-Based System Identification : An Instrumental Variable Approach.* To appear in Automatica.

[Chu 94] Chun T.C. *Geometry of linear systems and identification.* PhD Thesis, Cambridge University, England, March 1994, 173 pp.

[DPS 94] Deistler M., Peternell K., Scherrer W. *Consistency and Relative Efficiency of Subspace Methods.* Proc. of SYSID '94, Vol. 2, 4-6 July, Copenhagen, Denmark, pp.157-163, 1994.

[DMV 87] De Moor B., Vandewalle J. *A geometrical strategy for the identification of state space models of linear multivariable systems with singular value decomposition.* Proc. of the 3rd International Symposium on Applications of Multivariable System Techniques, April 13-15, Plymouth, UK, pp. 59-69, 1987.

[DMo 88] De Moor B. *Mathematical concepts and techniques for modeling of static and dynamic systems.* PhD thesis, Department of Electrical Engineering, Katholieke Universiteit Leuven, Belgium, 1988.

[DMo 94] De Moor B. *Numerical algorithms for state space subspace system identification.* Academia Analecta, Klasse der Wetenschappen, Koninklijke Akademie voor Wetenschappen, jaargang 55, no. 5, België, 1994.

[DMVO 94] De Moor B., Van Overschee P., *Graphical User Interface Software for System Identification*, Award winning paper of the *Siemens Award 1994*. ESAT-SISTA Report 94-06I, Department of Electrical Engineering, Katholieke Universiteit Leuven, 1994.

[DP 84] Desai U.B., Pal D. *A transformation approach to stochastic model reduction.* IEEE Transactions on Automatic Control, Vol. AC-29, no. 12, 1984.

[DKP 85] Desai U.B., Kirkpatrick R.D., Pal D. *A realization approach to stochastic model reduction.* International Journal of Control, Vol. 42, no. 4, pp. 821-838, 1985.

[DMK 74a] Dickinson B., Morf M., Kailath T. *A minimal realization algorithm for matrix sequences.* IEEE Transactions on Automatic Control, Vol. AC-19, no. 1, pp. 31-38, 1974.

[DKM 74b] Dickinson B., Kailath T., Morf M. *Canonical Matrix fraction and state space descriptions for deterministic and stochastic linear systems.* IEEE Transactions on Automatic Control, Vol. AC-19, pp. 656-667, 1974.

[Eyk 74] Eykhoff P. *System identification.* Wiley, London, 1974.

[FVOMHL 94] Falkus H., Van Overschee P., Murad G., Hakvoort H., Ludlage J. *Advanced Identification and Robust Control of a Glass Tube Production Process.* Third Philips Conference on Applications of Systems and Control Theory (PACT), November 9-10, Doorwerth, The Netherlands, 1994.

[Fau 76] Faurre P. *Stochastic realization algorithms.* System Identification: Advances and case studies (Eds.) Mehra R., Lainiotis D., Academic Press, 1976.

[GeWe 82] Gevers M., Wertz V. *On the problem of structure selection for the identification of stationary stochastic processes.* 6th IFAC Symposium on Identification and System Parameter Estimation, Eds. G. Bekey and G. Saridis, Washington D.C., 1982, pp.387-392.

[GW 74] Glover K., Willems J. *Parametrizations of linear dynamical systems: canonical forms and identifiability.* IEEE Transactions on Automatic Control, Vol. AC-19, pp. 640-645, 1974.

[GVL 89] Golub G., Van Loan C. *Matrix computations.*, second edition, Johns Hopkins University Press, Baltimore, Maryland, 1989.

[Gop 69] Gopinath B. *On the identification of linear time-invariant systems from input-output data.* The Bell System Technical Journal, Vol. 48, no. 5, pp. 1101-1113, 1969.

[Gui 75] Guidorzi R. *Canonical structures in the identification of multivariable systems.* Automatica, 11, pp. 361-374, 1975.

[Gui 81] Guidorzi R. *Invariants and canonical forms for systems structural and parametric identification.* Automatica, 17, pp. 177-133, 1981.

[Gyu 93] Gyugyi Paul. *Model-Based Control Applied to Rapid Thermal Processing.* Ph.D. Thesis Stanford University, CA, USA, 1993.

[HD 88] Hannan E.J., Deistler M. *The statistical theory of linear systems.* Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, New York,

1988.

[He 93] Hench J.J. *Identification of linear periodic state space models.* Inst. of Information Theory and Automation, Praag, 1993, to be published in the International Journal of Control.

[HK 66] Ho B.L., Kalman R.E. *Efficient construction of linear state variable models from input/output functions.* Regelungstechnik, **14**, pp. 545-548,1966.

[JW 94] Jansson M., Wahlberg B. *4SID Linear Regression.* Proc. 33rd IEEE Conference on Decision and Control, Lake Buena Vista, Florida, USA, 14-16 Dec., pp. 2858-2863, 1994.

[Jua 85] Juang J.N., Pappa R.S. *An eigensystem realization algorithm for model parameter identification and model reduction.* Journal of Guidance, Control and Dynamics, Vol. 8., no.5, Sept-Oct 1985, pp.620-627.

[Jua 94] Juang J.N. *Applied System Identification.* Prentice Hall, Englewood Cliffs, NJ, 1994.

[Kai 80] Kailath T. *Linear Systems.* Prentice Hall, Englewood Cliffs, New Jersey, 1980.

[Kin 88] King A.M., Desai U.B., Skelton R.E. *A generalized approach to Q-Markov Covariance Equivalent Realizations for Discrete Systems.* Automatica, 24, pp.507-515, 1988.

[Kun 78] Kung S.Y. *A new identification method and model reduction algorithm via singular value decomposition.* 12th Asilomar Conf. on Circuits, Systems and Comp., pp. 705-714, Asilomar, CA, 1978.

[Lar 83] Larimore W.E. *System identification, reduced order filtering and modeling via canonical variate analysis.* Proc. of the American Control Conference, San Francisco, USA, 1983.

[Lar 90] Larimore W.E. *Canonical variate analysis in identification, filtering and adaptive control.* Proc. 29th Conference on Decision and Control, Hawai, USA, pp. 596-604, 1990.

[Lar 94] Larimore W.E. *The Optimality of Canonical Variate Identification by Example,* Proc. of SYSID '94, Vol. **2**, 4-6 July, Copenhagen, Denmark, pp.151-156, 1994.

[LP 93] Lindquist A., Picci G. *On "Subspace Methods" Identification.* Proc. of MTNS, Regensburg, Germany, August 2-6, Vol. **2**, pp. 315-320, 1993.

[LP 94] Lindquist A., Picci G. *On "Subspace Methods" Identification and Stochastic Model reduction.* Proc. of SYSID '94, Vol. **2**, 4-6 July, Copenhagen, Denmark, pp. 397-404, 1994.

[LS 91] Liu K., Skelton R.E., *Identification and Control of NASA's ACES Structure.* Proceedings American Control Conference, Boston, MA, USA, pp. 3000-3006, 1991.

[Liu 92] Liu K., *Identification of Multi-Input and Multi-Output Systems by Observability Range Space Extraction.* Proc. 31st Conference on Decision and Control, Tucson, Arizona, USA, pp. 915-920, 1992.

[LSS 92] Liu K., Skelton R.E., Sharkey J.P., *Modeling Hubble Space Telescope Flight Data by Q-Markov Cover Identification.* Proc. of the American Control Conference, pp. 1961-1965, 1992.

[LJM 94] Liu K., Jacques R.N., Miller D.W. *Frequency Domain Structural System Identification by Observability Range Space Extraction.* Proc. of the American Control Conference, Baltimore, Maryland, Vol. 1, pp. 107-111, 1994.

[LS 77] Liu R., Suen L.C. *Minimal dimension realization and identifiability of input-output sequences.* IEEE Transactions on Automatic Control, Vol. AC-**22**, pp. 227-232, 1977.

[Lju 87] Ljung L. *System identification - Theory for the User.* Prentice Hall, Englewood Cliffs, NJ. 1987.

[Lju 91b] Ljung. L. *A Simple Start-Up Procedure for Canonical Form State Space Identification, Based on Subspace Approximation.* 30th IEEE Conference on Decision and Control, Brighton, UK, pp. 1333-1336, 1991.

[Lju 91c] Ljung. L. *System Identification Toolbox For Use with Matlab.* The Mathworks Inc., Mass., U.S.A., 1991.

[Lue 67] Luenberger D.C. *Canonical forms for linear multivariable systems.* IEEE Transactions on Automatic Control, Vol. AC-**12**:290, 1967.

[McK 93] McKelvey T., *System identification using overparametrized state space models.* Dept. of EE, Linköping University, Sweden, 1993, Report LiTH-ISY-R-1454, 41 pp.

[McK 94a] McKelvey T., *On State-Space Models in System Identification,* Thesis no. 447, Department of Electrical Engineering, Linköping university, Sweden, 1994.

[McK 94b] McKelvey T. *An Efficient Frequency Domain State-Space Identification Algorithm.* Proc. 33rd IEEE Conference on Decision and Control, Lake Buena Vista, Florida, USA, 14-16 Dec., pp. 3359-3364, 1994.

[McK 94c] McKelvey T. *SSID - A MATLAB Toolbox for Multivariable State-Space Model Identification.* Dept. of EE, Linköping University, Linköping Sweden, 1994.

[MDMVV 89] Moonen M., De Moor B., Vandenberghe L., Vandewalle J. *On and off-line identification of linear state space models.* International Journal of Control, Vol. **49**, no. 1, pp. 219-232, 1989.

[MV 90] Moonen M., Vandewalle J. *A QSVD approach to on- and off-line state space identification.* International Journal of Control, Vol. **51**, no. 5, pp. 1133-1146, 1990.

[MDMV 91] Moonen M., De Moor B., Vandewalle J. *SVD-based subspace methods for multivariable continuous time system identification.* Identification of continuous-time systems, G.P. Rao, N.K. Sinha (Eds.), Kluwer Academic Publications, pp. 473-488, 1991.

[MDM 92] Moonen M., De Moor B. *Comments on 'State-space model identification with data correlation'.* International Journal of Control, Vol. **55**, no. 1, pp. 257-259, 1992.

[MDMRT 92] Moonen M., De Moor B., Ramos J., Tan S. *A subspace identification algorithm for descriptor systems.* Systems & Control Letters, Vol. **19**, pp. 47-52, 1992.

[MR 93] Moonen M, Ramos J. *A subspace algorithm for balanced state space system identification.* IEEE Transactions on Automatic Control, Vol. **38**, pp. 1727-1729, 1993

[Moo 81] Moore B.C. *Principal component analysis in linear systems: Controllability, Observability and Model Reduction.* IEEE Transactions on Automatic Control, Vol. AC-**26**, no. 1, pp 17-32, 1981.

[MR 76] Mullis C.T., Roberts R.A. *Synthesis of minimum round-off noise fixed point digital filters.* IEEE Transactions on Circuits and Systems, Vol. CAS-**23**, pp. 555-562, 1976.

[Nor 86] Norton J.P. *An introduction to identification.* Academic Press, London, 1986.

[Ott 89] Ottersten B. *Parametric subspace fitting methods for array signal processing.* Ph.D. Thesis, Information Systems Laboratory, Department of Electrical Engineering, Stanford University, CA. USA, 1989.

[OVK 92] Ottersten B., Viberg M., Kailath T., *Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data*, IEEE Tr. on SP, Vol. 40, no. 3, pp. 590-600, 1992.

[OVSN 93] Ottersten B., Viberg M., Stoica P., Nehorai A., *Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing*, in Radar array processing, S.Haykin, J.Litva, T.J.Sherpherd (Eds.), Springer Verlag, pp. 99-151, 1993.

[OV 94] Ottersten B., Viberg M., *A Subspace Based Instrumental Variable Method for State Space System Identification*. Proc. of SYSID '94, Vol. 2, 4-6 July, Copenhagen, Denmark, pp.139-144, 1994.

[Pal 82] Pal D., *Balanced stochastic realization and model reduction*. Master's thesis, Washington State University, Electrical Engineering, 1982.

[Ske 94] Skelton R.E., Grigoriadis Karolus, Zhu Guoming. *An algorithm for iterative identification and control design using an impro ved Q-Markov cover*. SYSID '94, Copenhagen, pp.183-188.

[SS 89] Söderström T., Stoica P. *System Identification*. Prentice Hall International Series in Systems and Control Engineering, Prentice Hall, New York, 1989.

[Sil 71] Silverman L. *Realization of linear dynamical systems*. IEEE Transactions on Automatic Control, Vol. AC-16, pp. 554-567, 1971.

[SROK 92] Swindlehurst A., Roy R., Ottersten B., Kailath T. *System identification via weighted subspace fitting*. Proc. of the American Control Conference, pp. 2158-2163, 1992.

[VDS 93] Van Der Veen A., Deprettere E.F., Swindlehurst A.L. *Subspace-Based Signal Analysis Using Singular Value Decompositions*. Proceedings of the IEEE, Vol. 81, no. 9, pp. 1277-1308, 1993.

[VOL 82] van Overbeek A.J.M., Ljung L. *On-line structure selection for multivariable state space models*. Automatica, 18, no. 5, pp. 529-543, 1982.

[VODM 93a] Van Overschee P., De Moor B. *Subspace algorithms for the stochastic identification problem*. Automatica, Vol. 29, no. 3, 1993, pp. 649-660.

[VODM 93c] Van Overschee P., De Moor B. *Subspace identification of a glass tube manufacturing process*. Proc. of the second European Control Conference, June 28-July 1, Groningen, The Netherlands, pp. 2338-2343, 1993.

[VODM 93d] Van Overschee P., De Moor B., *A Unifying Theorem for three Subspace System Identification Algorithms*. ESAT-SISTA Report 93-50I, Department of Electrical Engineering, Katholieke Universiteit Leuven, 1993. Accepted for *Automatica*.

[VODM 94a] Van Overschee P., De Moor B. *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, 1994.

[VODMBAK 94] Van Overschee P., De Moor B., Boyd S., Aling H., Kosut R. *A fully interactive system identification module for Xmath (ISID)*, Proc. of SYSID '94, Vol. 4, Copenhagen, Denmark, pp. 1, 1994.

[VODM 94d] Van Overschee P., De Moor P. *About the choice of State Space Bases in Combined Deterministic-Stochastic Subspace Identification*, ESAT-SISTA Report 94-24I, Department of Electrical Engineering, Katholieke Universiteit Leuven, 1994. Accepted for publication in *Automatica*.

[VODMAKB 94] Van Overschee P., De Moor B., Aling H., Kosut R., Boyd S. *Xmath Interactive System Identification Module, Part 2*. Integrated Systems Inc., Santa Clara, CA, USA, 1994.

[VO PhD] Van Overschee P., *Subspace identification: Theory - Implementation - Application*. PhD Thesis, Department of Electrical Engineering, Katholieke Universiteit Leuven, Belgium, February 1995.

[VP 93] Van Poucke F., Moonen M. *Direction finding of multiple wide-band emitters using state space modelling*. ESAT-SISTA TR 1993-30, Department of Electrical Engineering, Katholieke Universiteit Leuven, Belgium. To be published in IEEE Trans. on Signal Processing.

[Ver 91] Verhaegen M. *A novel non-iterative MIMO state space model identification technique*. Proc. 9th IFAC/IFORS Symp. on Identification and System Parameter Estimation, Budapest, Hungary, pp. 1453-1458, 1991.

[VD 92] Verhaegen M., Dewilde P. *Subspace model identification, Part I: The output-error state space model identification class of algorithms*. Int. J. Control, Vol. 56, 1187-1210, 1992.; Part II: Analysis of the elementary output-error state space model identification algorithm, Vol. 56, no.5, pp.1211-1241, ; Part III: Analysis of the ordinary output error state space model identification algorithm, Vol. 58, no.3, pp.555-586, 1993.

[Ver 94] Verhaegen M. *Identification of the deterministic part of MIMO state space models given in innovations form from input-output data*, Automatica (Special Issue on Statistical Signal Processing and Control), Vol 30, no. 1, pp. 61-74, 1994

[Verr 94] Verriest E.I., Kullstam J.A. *Realization of discrete-time periodic systems from input-output data*. Int. Report, Georgia Tech Lorraine, 1994, submitted for publication.

[Vib 89] Viberg M. *Subspace fitting concepts in sensor array processing*. Ph.D. thesis, Department of Electrical Engineering, Linköping University, Sweden, 1989.

[VO 91] Viberg M., Ottersten B. *Sensor array processing based on subspace fitting*. IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-39, 1991.

[VOWL 91] Viberg M., Ottersten B., Wahlberg B., Ljung L. *A Statistical Perspective on State-Space Modeling Using Subspace Methods*. Proc. of the 30th IEEE Conference on Decision and Control, Brighton, England, 11-13 Dec, Vol 2, pp 1337-1342, 1991.

[VOWL 93] Viberg M., Ottersten B., Wahlberg B., Ljung L., *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, 1993.

[Vib 94] Viberg M. *Subspace Methods in System Identification*. Proc. of SYSID '94, Vol. 1, 4-6 July, Copenhagen, Denmark, pp. 1-12, 1994.

[Wil 86] Willems J. *From time series to linear systems*. Automatica, Part I: Finite Dimensional Linear Time Invariant Systems, Vol. 22, no. 5, pp. 561-580, 1986, Part II: Exact Modelling, Vol. 22, no. 6, pp. 675-694, 1986, Part III: Approximate modelling, Vol. 23, no. 1, pp. 87-115, 1987.

[ZM 74] Zeiger H., McEwen A. *Approximate linear realizations of given dimension via Ho's algorithm*. IEEE Transactions on Automatic Control, 19, pp. 153. 1974.

[ZVODML 94] Zhu Y., Van Overschee P., De Moor B., Ljung L. *Comparison of three classes of identification methods*. Proc. of SYSID '94, Vol. 1, 4-6 July, Copenhagen, Denmark, pp. 175-180, 1994.

Algorithm (simpler to estimate B and D)

1. Calculate the oblique projections:

$$\mathcal{O}_i = Y_f /_{U_f} W_p \ ,$$

$$\mathcal{O}_{i+1} = Y_f^- /_{U_f^-} W_p^+ \ .$$

2. Calculate the SVD of the weighted oblique projection:

$$W_1 \mathcal{O}_i W_2 = USV^T \ ,$$

where W_1 and W_2 are user-defined weighting matrices.

3. Determine the order by inspecting the singular values in S and partition the SVD accordingly to obtain U_1 and S_1.
4. Determine Γ_i and Γ_{i-1} as:

$$\Gamma_i = W_1^{-1} U_1 S_1^{1/2} \quad , \quad \Gamma_{i-1} = \underline{\Gamma_i} \ .$$

5. Determine the state sequences:

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

$$\widetilde{X}_{i+1} = \Gamma_{i-1}^{\dagger} . \mathcal{O}_{i+1} \ .$$

6. Solve the set of linear equations for A, B, C and D in a linear least squares sense (which also delivers the residuals ρ_w and ρ_v):

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

7. Determine Q_i, S_i and R_i from the residuals as:

$$\begin{pmatrix} Q_i & S_i \\ S_i^T & R_i \end{pmatrix} = \mathbf{E_j}[\begin{pmatrix} \rho_w \\ \rho_v \end{pmatrix} . \begin{pmatrix} \rho_w^T & \rho_v^T \end{pmatrix}] \ .$$

Fig. 9. A schematic overview of the second combined deterministic-stochastic identification algorithm. This algorithm computes asymptotically biased solutions, unless at least one of the three conditions discussed above is satisfied. See Section 10 for implementation issues. Note the 'symmetry' in this algorithm in the determination of the matrices A, B, C and D (Compare to the first algorithm) which gives a certain elegance to this algorithm, although in general it provides biased estimates.

Robust implementation of N4SID

1. Compute $L_{U_p} \in \mathbb{R}^{li \times mi}$ and $L_{Y_p} \in \mathbb{R}^{li \times li}$ from:

$$\begin{pmatrix} L_{U_p} & L_{U_f} & L_{Y_p} \end{pmatrix} = L_{[5:6],[1:4]} L_{[1:4],[1:4]}^{\dagger} \ ,$$

where the pseudo inverse can be substituted by a back-substitution when $L_{[1:4],[1:4]}$ is of full rank.

2. With:

$$\Pi = I_{2mi} - L_{[2:3],[1:3]}^T [L_{[2:3],[1:3]} L_{[2:3],[1:3]}^T]^{-1} L_{[2:3],[1:3]} \ .$$

Compute the SVD of:

$$\left((L_{U_p} L_{[1:1],[1:3]} + L_{Y_p} L_{[4:4],[1:3]}) . \Pi | L_{Y_p} L_{44} \right) \ .$$

3. Determine the order by inspection of the singular values in S and partition the SVD accordingly to obtain U_1 and S_1.
4. Determine $\Gamma_i = U_1 S_1^{1/2}$ and $\Gamma_{i-1} = \underline{\Gamma_i}$. Compute A and C as the first n columns of:

$$\begin{pmatrix} \Gamma_{i-1}^{\dagger} L_{[6:6],[1:5]} \\ L_{[5:5],[1:5]} \end{pmatrix} \begin{pmatrix} \Gamma_i^{\dagger} L_{[5:6],[1:5]} \\ L_{[2:3],[1:5]} \end{pmatrix}^{\dagger} \ .$$

Recompute Γ_i and Γ_{i-1} from A and C.

5. Solve for B and D from the least squares problem discussed and referred to in Subsection 8.2.
6. Form the matrices $\mathcal{P}$ and $\mathcal{Q}$:

$$\mathcal{P} = \begin{pmatrix} \Gamma_{i-1}^{\dagger} L_{[6:6],[1:5]} \\ \hline L_{[5:5],[1:5]} \end{pmatrix} - \begin{pmatrix} A \\ \hline C \end{pmatrix} \Gamma_i^{\dagger} L_{[5:6],[1:5]} \ ,$$

$$\mathcal{Q} = L_{[2:3],[1:5]} \ .$$

7. Determine the matrix $\mathcal{K}$ (39) from A, B, C, D, Γ_i and determine estimates of the covariance matrices Q, S and R as:

$$\begin{pmatrix} Q_i & S_i \\ S_i^T & R_i \end{pmatrix} = (\mathcal{P} - \mathcal{K}\mathcal{Q})(\mathcal{P} - \mathcal{K}\mathcal{Q})^T \ .$$

Fig. 10. Implementation of a robust deterministic-stochastic identification algorithm. For details the reader is referred to [VODM 94a] [VO PhD]. Notice that the 'long' dimension j has disappeared, which implies that the factor Q in the LQ-decomposition of the block Hankel matrix with input-output data is never needed explicitly and hence should not be computed.