SOME ALGORITHMIC ASPECTS OF SUBSPACE IDENTIFICATION WITH INPUTS

ALESSANDRO CHIUSO GIORGIO PICCI *

Abstract

It has been experimentally verified that most commonly used subspace methods for identification of linear state-space systems with exogenous inputs may, in certain experimental conditions, run into ill-conditioning and lead to ambiguous results. An analysis of the critical situations has lead to propose a new algorithmic structure wich could be used either to test difficult cases or/and to implement a suitable combination of new and old algorithms presented in the literature to help fixing the problem. A MATLAB code is available upon request at `chiuso@dei.unipd.it`.

1 Introduction

It is well-known that the "classical" approach to system identification is based on parameter optimization, i.e. the system parameters are obtained by minimization of a suitable cost function. These methods have been widely used and shown reasonably successful in modeling single-input single-output systems by ARMA or ARMAX models, see the classical textbook [23] for an up-to-date illustration of this approach.

However, when one has to attack general multi-input multi-output models, these methods suffer from various drawbacks. Since, unless one restricts to rather trivial model classes, the dependence of the cost function on the parameters is in general non-linear, iterative techniques are required for minimization. For multivariable sytems these may well turn out to be very time-consuming. Due to existence of local minima and non-convexity, the outcome is in general very sensitive to the choice of the parametrization and of the starting point in the optimization procedure. There is generally no guarantee of global optimality but only of ending close to a local minimum. Furthermore, to attack general multi-input multi-output models by parameter optimization methods the hussle of choosing canonical parametrizations is unavoidable. In fact, the use of canonical parametrizations has been recognized as a critical issue in MIMO identification since the early 1970's [15, 46], and represents a bottleneck in extending from SISO to MIMO identification.

Geometric, or "subspace" or realization-based, methods, rely on the ideas of stochastic realization theory which have been developed (mainly for time series) by many authors, for instance Akaike [1, 2], Faurre [14], Lindquist and Picci [25, 27], Picci [36], Ruckebusch [44, 43].

Subspace methods, roughly speaking, translate the constructions of stochastic realization theory into procedures for model building which work on measured data [28]. They owe the name "subspace" to the fact that the basic objects which are constructed in the algorithms are subspaces generated by the data, and geometric operations such as orthogonal and oblique projections are all what is needed to compute estimates of the parameters.

The appealing features of subspace methods are that there is no need for canonical parametrizations; no iterative nonlinear optimization is required; only simple and numerically robust tools of numerical linear algebra such as QR, SVD, QSVD, are used; finally, since the methods rely on the theoretical background of stochastic realization theory a deeper system-theoretic understanding of the involved procedures is possible.

*Dipartimento di Elettronica e Informatica, Università di Padova, 35131 Padua, Italy; Email: `picci@dei.unipd.it`

The basics of subspace methods may probably be traced back to old work of Hotelling [17], Ho and Kalman [16], Akaike [1, 2], Faurre [14], Aoki [4] and Moonen et al. [32], but probably, the first "true" subspace algorithm is the "stochastic" algorithm of van Overschee and De Moor [47] for the identification of time series. Various subspace algorithms have been introduced for identification of systems with exogenous inputs, some of the basic references being [32, 33, 50, 49, 52, 48, 54]. Even though these methods have been around for a while, it is fair to say that for subspace methods *with inputs* there are still a number of questions which are not completely understood.

1. One of these questons is numerical ill-conditioning which has been experimentally verified in a number of situations [9, 18]. One should understand when ill-conditioning may occur and how to cure the problem. Recently [13] it has been argued that using orthogonal decomposition and block-parametrized models, together with the orthogonal decomposition algorithm of [38, 9] may be a possible solution to the problem of ill-conditioning. Simulation results and comparison with the N4SID algorithm are discussed in [9, 12, 13].

2. As it is well-known, the dynamics of the input signal is crucial for the outcome of an identification experiment. It is important to have bounds on the performance of an algorithm as a "function" of the input characteristic (bandwith, persistence of excitation, etc..). In particular, for comparing results of simulations, a specification of "probing inputs" for the validation of identification algorithms [12] is needed. By "probing inputs" we mean inputs which are tailored to reveal the main limitations of the algorithms.

3. Subspace identification in the presence of feedback has been addressed by some authors [51, 53], but the problem seems to be very far from being completely understood.

4. The characterization of the accuracy of the estimates is still a partially open problem. Steps toward solving this problem have been made in [42, 5, 7, 35, 34] where results on asymptotic normality of the estimates are obtained and procedures to compute asymptotic variance have been suggested.

On the thoretical side, one should remark that stochastic realization theory with exogenous inputs has not been fully developed until very recently [11]. While the algorithm of van Overschee and De Moor for time series [47] follows exactly the ideal steps suggested by stochastic realization theory, so far it has not been possible to implement the ideal realization procedure in identification with exogenous inputs. In particular, it has been pointed out that there are no known procedures for constructing a basis in the state space of a stochastic system with inputs (to be precise in a finite-time oblique Markovian splitting subspace), directly from finite-time input output data [11]. By "directly" we mean only by means of operations on the data which do not involve preliminary estimation of some system parameters.

In all algorithms existing in the literature ad hoc tricks are used and an approximate version of the state is involved. This can be shown to deteriorate the "ideal" numerical conditioning of the problm [13] and is believed to be the reason why the state-of-the-art in subspace methods may be considered satisfactory only for time-series identification.

Due to page limitations, we cannot give here more details on these aspects and shall have to refer the reader to the literature. The main purpose of this paper is to give a brief guided tour to the algorithms for subspace identification with inputs existing in the literature and to suggest some variations which help in dealing with the possible ill-conditioning of the identification problem. The algorithm may optionally use alternative approaches to those in the literature. A MATLAB software package has been developed as a part of the doctoral thesis of A. Chiuso, [11].

2 Notations

There is a "true" stochastic system (which we assume in innovation form)

$$\left\{ \begin{array}{rcl} \mathbf{x}(t+1) & = & A\mathbf{x}(t) + B\mathbf{u}(t) + K\mathbf{e}(t) \\ \mathbf{y}(t) & = & C\mathbf{x}(t) + D\mathbf{u}(t) + \mathbf{e}(t) \end{array} \right. \tag{1}$$

generating the observed data $\{y(t)\}$ (m-dimensional), $\{u(t)\}$ (p-dimensional). Let $\{x(t)\}$, $\{e(t)\}$ be the sample paths of the corresponding state (n-dimensional) and innovation processes. Suppose (ideally) that we have observations on some (hopefully very long) time interval $[0, N]$, of one sample path $\{y(t)\}$, $\{u(t)\}$, $\{x(t)\}$, of the processes $\{\mathbf{y}(t)\}$, $\{\mathbf{u}(t)\}$, $\{\mathbf{x}(t)\}$. Since these processes generate the data, it is obvious that the finite "tail" matrices, Y_t, U_t, X_t, constructed at each time t from the observed samples by the recipe

$$\begin{array}{lcl} Y_t & := & [\, y(t) \quad y(t+1) \quad \dots \quad y(t+N-1)\,] \\ U_t & := & [\, u(t) \quad u(t+1) \quad \dots \quad u(t+N-1)\,] \\ X_t & := & [\, x(t) \quad x(t+1) \quad \dots \quad x(t+N-1)\,] \end{array}$$

also satisfy equation (1), i.e.:

$$\begin{cases} X_{t+1} = AX_t + BU_t + KE_t \\ Y_t = CX_t + DU_t + E_t \end{cases} \tag{2}$$

where $E_t := [\, e(t) \quad e(t+1) \quad \dots \quad e(t+N-1)\,]$ is the innovation tail. This equation can be interpreted as a regression model. It is straightforward to see that, if the tail matrices X_{t+1}, X_t, U_t, Y_t, are given, then one can solve (2) for the unknown parameters (A, B, C, D), say by least squares. Hence in the ideal situation, when we have available an input, output, and a corresponding *state* sequence at two successive time instants t and $t+1$, the identification of the parameters (A, B, C, D) of the system (1) is a rather trivial matter. In practice, X_{t+1}, X_t are of course not available and will have to be estimated from the input-output data. This is the crucial step of most susbspace identification algorithms.

In the ideal case when infinitely long sample trajectories are available ($N \to \infty$), E_t is orthogonal to the past data, namely $E_t \perp (X_s, U_s)$ for all $s \leq t$ (this is only approximately true for N large but finite). Because of orthogonality of the innovation term, the estimates computed by solving the regression equation, coincide (for $N \to \infty$) with the true parameters (consistency).

We shall use standard notations; in particular use the symbols

$$Y_{t_1|t_2} := \begin{bmatrix} Y_{t_1} \\ Y_{t_1+1} \\ \vdots \\ Y_{t_2} \end{bmatrix} = \begin{bmatrix} y(t_1) & y(t_1+1) & \dots & y(t_1+N-1) \\ y(t_1+1) & y(t_1+2) & \dots & y(t_1+N) \\ \vdots & \dots & \ddots & \vdots \\ y(t_2) & y(t_2+2) & \dots & y(t_2+N-1) \end{bmatrix},$$

$$U_{t_1|t_2} := \begin{bmatrix} U_{t_1} \\ U_{t_1+1} \\ \vdots \\ U_{t_2} \end{bmatrix} = \begin{bmatrix} u(t_1) & u(t_1+1) & \dots & u(t_1+N-1) \\ u(t_1+1) & u(t_1+2) & \dots & u(t_1+N) \\ \vdots & \dots & \ddots & \vdots \\ u(t_2) & u(t_2+2) & \dots & u(t_2+N-1) \end{bmatrix}$$

and denote

$$P_{t_1|t_2} = \begin{bmatrix} U_{t_1|t_2} \\ Y_{t_1|t_2} \end{bmatrix}$$

the joint input-output history between instants t_1 and t_2.

Given a $k_1 \times N$ matrix B and a $k_2 \times N$ matrix A, we will, with slight abuse of notation, denote the orthogonal projection

$$E\,[B \mid A] = B \mid A := BA^T(AA^T)^\dagger A,$$

meaning the $k_1 \times N$ matrix whose rows are the orthogonal projection of the rows of B onto the row span of A. Moreover, let

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$$

and let

$$\text{row-span}\{A_1\} \cap \text{row-span}\{A_2\} = \{0\}$$

For notational convenience we will denote

$$E\,[B \mid A_1 \vee A_2] := E\left[B \mid \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}\right]$$

the orthogonal projection of the rows of B onto the row space of $\begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$. This orthogonal projection may be uniquely decomposed as

$$E\left[B \mid A_1 \vee A_2\right] = E_{\| A_2}\left[B \mid A_1\right] + E_{\| A_1}\left[B \mid A_2\right]$$

which are respectively the oblique projection of the rows of B onto row-span$\{A_1\}$ along row-span$\{A_2\}$ and viceversa. It is immediate to obtain expressions for these oblique projections:

$$E_{\| A_1}\left[B \mid A_2\right] = B\left(A_2 \mid A_1^{\perp}\right)\left[\left(A_2 \mid A_1^{\perp}\right)\left(A_2 \mid A_1^{\perp}\right)^T\right]^{-1} A_2$$

where

$$A_2 \mid A_1^{\perp} := A_2 - A_2|A_1 = A_2 - A_2 A_1^T (A_1 A_1^T)^{\dagger} A_1$$

and similarly for the other.

Define the extended observability matrix

$$\Gamma_k := \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{k-1} \end{bmatrix}$$

the reversed controllability matrices

$$\mathcal{C}_k^d := [\, A^{k-1}B \quad \dots \quad AB \quad B \,] \quad \mathcal{C}_k^s := [\, A^{k-1}K \quad \dots \quad AK \quad K \,],$$

and the Toeplitz matrices

$$H_k^s := \begin{bmatrix} I & 0 & \dots & 0 \\ CK & I & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{k-2}K & CA^{k-3}K & \dots & I \end{bmatrix}, \quad H_k^d := \begin{bmatrix} D & 0 & \dots & 0 \\ CB & D & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{k-2}B & CA^{k-3}B & \dots & D \end{bmatrix}.$$

It follows from straightforward manipulations that we can write:

$$\left\{ \begin{array}{lllllll} X_k & = & A^k X_0 & + & \mathcal{C}_k^d U_{0|k-1} & + & \mathcal{C}_k^s E_{0|k-1} \\ Y_{k|2k-1} & = & \Gamma_k X_k & + & H_k^d U_{k|2k-1} & + & H_k^s E_{k|2k-1} \end{array} \right. \tag{3}$$

These relations are the starting point for most subspace identification methods.

3 The orthogonal decomposition approach

Identification in the presence of exogenous inputs can be done, in principle, following two different approaches, which essentially correspond to different choices of “model structures”. On one hand, one could use stochastic realizations of $\mathbf{y}$ driven by $\mathbf{u}$ of the general form

$$\left\{ \begin{array}{rcl} \mathbf{x}(t+1) & = & A\mathbf{x}(t) + B\mathbf{u}(t) + K\mathbf{e}(t) \\ \mathbf{y}(t) & = & C\mathbf{x}(t) + D\mathbf{u}(t) + \mathbf{e}(t) \end{array} \right. . \tag{4}$$

Identification procedures based on this model will be referred to as “joint identification”. On the other hand one could instead consider models in block diagonal form such as

$$\left\{ \begin{array}{rcl} \begin{bmatrix} \mathbf{x}_d(t+1) \\ \mathbf{x}_s(t+1) \end{bmatrix} & = & \begin{pmatrix} A_d & 0 \\ 0 & A_s \end{pmatrix} \begin{bmatrix} \mathbf{x}_d(t) \\ \mathbf{x}_s(t) \end{bmatrix} + \begin{pmatrix} B_d \\ 0 \end{pmatrix} \mathbf{u}(t) + \begin{pmatrix} 0 \\ K_s \end{pmatrix} \mathbf{e}_s(t) \\ \mathbf{y}(t) & = & \begin{pmatrix} C_d & C_s \end{pmatrix} \begin{bmatrix} \mathbf{x}_d(t) \\ \mathbf{x}_s(t) \end{bmatrix} (t) + D_d \mathbf{u}(t) + \mathbf{e}_s(t). \end{array} \right. \tag{5}$$

which is based on the preliminary decomposition of the state and output processes into the component lying in the input space (the " deterministic component") and its orthogonal complement (the "stochastic component"), see [38, 9, 11]. For identification based on models of this structure, we shall talk about a "orthogonal decoposition" approach. We warn the reader that models of the form (5) may turn out to be non minimal, due to lack of observability, which may occur when the "deterministic" and "stochastic" components share some common dynamics. The most general situation is the one in which the deterministic and stochastic subsystems may share some (and possibly all) "dynamics". In such a situation a minimal realization would have a block diagonal structure formed by three blocks corresponding to deterministic, shared and stochastic dynamics of the form:

$$
\left\{
\begin{array}{rcl}
\begin{bmatrix} \mathbf{x}_d(t+1) \\ \mathbf{x}_{ds}(t+1) \\ \mathbf{x}_s(t+1) \end{bmatrix} & = & \begin{pmatrix} A_d & 0 & 0 \\ 0 & A_{ds} & 0 \\ 0 & 0 & A_s \end{pmatrix} \begin{bmatrix} \mathbf{x}_d(t) \\ \mathbf{x}_{ds}(t) \\ \mathbf{x}_s(t) \end{bmatrix} + \begin{pmatrix} B_d \\ B_{ds} \\ 0 \end{pmatrix} \mathbf{u}(t) + \begin{pmatrix} 0 \\ K_{ds} \\ K_s \end{pmatrix} \mathbf{e}_s(t) \\
\mathbf{y}(t) & = & \begin{pmatrix} C_d & C_{ds} & C_s \end{pmatrix} \begin{bmatrix} \mathbf{x}_d(t) \\ \mathbf{x}_{ds}(t) \\ \mathbf{x}_s(t) \end{bmatrix} (t) + D_d \mathbf{u}(t) + \mathbf{e}_s(t).
\end{array}
\right. \tag{6}
$$

Naturally, the presence of common dynamics is to be regarded as a " non generic" situation, unless there some a priori information on the way the noise enters the system.

In section 7 we shall present simulations comparing the results of subspace algorithms with the Cramèr-Rao lower bounds. It will become apparent that an orthogonal decomposition approach is to be preferred when the dynamics of the deterministic and stochastic parts are disjoint as the Cramèr-Rao bounds are lower for this kind of approach.

This is essentially due to the fact that in this situation more "structure" is used and less parameters (as compared to a joint approach) are to be estimated. Using a joint model in this case leads to worse results. In fact, the identified model will present some near cancellations of poles and zeros in the deterministic and in the stochastic transfer functions. That might be a further source of ill-conditioning.

On the other hand, when it is known that the two subsystems share the same dynamics we are in the opposite situation and the joint approach does better. If only part of the dynamics is shared then things become of course harder to evaluate.

Several "subspace" algorithms have been presented in the literature which could be adapted to both approaches. However, the differences are not just due to the choice between "joint" or "orthogonal decomposition" approaches.

A subspace algorithms can organized into four main steps:

1. Estimation of the state (or of the extended observability matrix), which includes order estimation;
2. Estimation of the matrices (A, C) (or (A_d, C_d) for the orthogonal decomposition case);
3. Estimation of the noise model, i.e. the "Kalman gain" K and the variance of the innovation Λ for the joint apporach or the entire stochastic realization $(A_s, C_s, K_s, \Lambda_s)$ for the orthogonal decomposition approach;
4. Estimation of the input matrices (B, D) (or (B_d, D_d)).

The four steps have been enumerated in the order they are usually performed, as any of them requires (or may require) the estimates obtained in the previous steps but does not require estimates to be obtained in the next steps.

We have defined two main functions, `joint.m` and `ort_dec.m` which respectively implement the joint and orthogonal decomposition aproaches. They are structured in such a way that the user has the freedom to choose independently (to a certain extent) how steps 1), 2), and 3) are performed among the most common choices considered in the literature. Essentialy, in the current implementation step 4) is fixed . In the joint case K and Λ are obtained solving a certain Riccati equation which amounts to computing the steady state Kalman gain (when such a solution exists [28]). In the orthogonal decomposition approach the algorithm implemented for the estimation of the stochastic component is the "stochastic" algorithm of Van Overschee and De Moor [47]. As a matter of fact, this algorithm, which is the only theoretically sound "stochastic" subspace approach, has recently been shown to be asymptotically efficient [6]. We warn the reader that the identification of the stochastic component in the orthogonal decomposition approach requires a somewhat delicate prefiltering algorithm. For further details on the orthogonal decomposition approach one may consult [38, 9, 11].

In the next sections we shall give a brief overview the main procedures of the algorithm. For reasons of space we shall not be able to enter into much detail. The theoretical analysis on which some of the procedures are based will be found in the forthcoming pubblications [13].

The syntax is the following

```
function [Ad,Bd,Cd,Dd,As,Ks,Cs,Lambda] =
                   ort_dec(y,u,ns,nd,ks,kd,BD,T,delay,type,Aest).
```

```
function [A,B,C,D,K,Lambda]=joint(y,u,nn,k,BD,T,delay,type,Aest);
```

where `y` and `u` are respectively output and input data, `ns,nd,ks,kd,k` are indexes related to orders, and `BD,T,delay,type,Aest` are related to the user choices in steps 1), 2), 3) as discussed above.

4 Estimation of the Extended Observability matrix

In this section we shall talk about estimation of the extended observability matrix rather than estimation of the state vector, the reason being that, as we have already pointed out, there are non known recepies to construct directly an oblique Markovian splitting subspace from mesured data [1]. A more precise analysis would require the introduction of a sort of "conditional"model (given future inputs); details will be found in a forthcoming pubblication.

In the code we have implemented three standard approaches for the estimation of the observability matrix which are called the "orthogonal projection", the "oblique projection" and the "canonical variate analysis". They correspond, as it has been pointed out in [49], to different choices of weighting matrices. Infact, the extended observability matrix is determined via SVD of the following matrix:

$$W_1 E_{\|U_{k|2k-1}} \left[Y_{k|2k-1} \mid P_{0|k-1}\right] W_2 = W_1 \Gamma_k E_{\|U_{k|2k-1}} \left[X_k \mid P_{0|k-1}\right] W_2 = USV; \tag{7}$$

The matrix $P_{0|k-1}$ is either $Y_{0|k-1} \vee U_{0|k-1}$ for the combined deterministic-stochastic identification or $P_{0|k-1} = U_{0|k-1}$ for deterministic identification, i.e. for the identification of the deterministic component in the orthogonal decomposition algorithm.

In an ideal situation, when data are generated by an n-dimensional, "true" linear time invariant system, and N goes to infinity, the matrix S has generically n singular values different from zero. We say "generically" since there might be pathological situations in which $\hat{Y}_{k|2k-1}$ looses rank [30]; nevertheless the set of systems for which asymptotically S looses rank is non-generic [7]. We will discuss this point in the following. In fact, even though this matrix looses rank on a set of "measure zero", there are open neighborhoods in the set of parameters that makes the n-th singular value arbitrarily close to zero.

In practical situations, i.e. for finite data, S has full rank and a reduction step has to be performed. This corresponds to order estimation in subspace identification methods and is of primary importance. Note that, if S is partitioned as

$$S = \begin{bmatrix} \hat{S}_n & 0 \\ 0 & \tilde{S}_n \end{bmatrix} \simeq \begin{bmatrix} \hat{S}_n & 0 \\ 0 & 0 \end{bmatrix}$$

and U and V are partitioned accordingly,

$$U = [\, U_n \quad U_n^{\perp} \,] \quad V = [\, V_n \quad V_n^{\perp} \,]$$

[1] We stress again that "directly" here means without preliminary estimation of some of the system parameters, e.g. the Markov parameters

the corresponding estimate of the state space and observablity matrix are:

$$\hat{\Gamma}_k = W_1^{-1} U_n S_n^{1/2} \quad , \quad \hat{\hat{X}}_k = S_n^{1/2} V_n^T W_2^{-1}. \tag{8}$$

Even though a precise theoretical analysis is still lacking, there is some evidence [21, 5, 10, 52] that CVA performs better in a broader range of situations. We now briefly review the aforementioned approaches.

4.1 Oblique Projection

This is the choice of basis which is done in N4SID [48]. It is called *oblique projection* since it corresponds to the weighting matrices $W_1 = W_2 = I$, i.e. to performing SVD of the oblique projection of future outputs along future inputs onto the joint past:

$$USV^T := \hat{Y}_{k|2k-1} = E_{\| U_{k|2k-1}} \left[Y_{k|2k-1} \mid P_{0|k-1} \right] = \Gamma_k \tilde{X}_k \tag{9}$$

4.2 Orthogonal Projection

This factorization is for instance done in the PO-MOESP type of algorithms [52], and is called *orthogonal projection* since it corresponds to projecting the optimal oblique predictor onto the orthogonal complement of $U_{k|2k-1}$ in $P_{0|k-1} \vee U_{k|2k-1}$, i.e.

$$USV^T := E\left[Y_{k|2k-1} \mid \left(P_{0|k-1} \mid U^{\perp}_{k|2k-1} \right) \right] = \Gamma_k E\left[\tilde{X}_k | U^{\perp}_{k|2k-1} \right]. \tag{10}$$

It is apparent that this corresponds to $W_1 = I$ and $W_2 = \Pi_{U^{\perp}_{k|2k-1}}$.

4.3 Canonical Variate Analysis

CVA is a way of choosing basis in the state space which makes use of the concept of Canonical Correlations [22, 19]. The idea is to compute the canonical correlations between joint past $P_{0|k-1}$ and future outputs $Y_{k|2k-1}$, given future inputs $U_{k|2k-1}$. Let us define

$$L_{p|u^\perp} L^T_{p|u^\perp} = \Sigma_{pp|u^\perp} = \frac{1}{N} \left(P_{0|k-1} \mid U^{\perp}_{k|2k-1} \right) \left(P_{0|k-1} \mid U^{\perp}_{k|2k-1} \right)^T$$

and similarly

$$L_{y|u^\perp} L^T_{y|u^\perp} = \Sigma_{yy|u^\perp} = \frac{1}{N} \left(Y_{k|2k-1} \mid U^{\perp}_{k|2k-1} \right) \left(Y_{k|2k-1} \mid U^{\perp}_{k|2k-1} \right)^T,$$

$$\Sigma_{yp|u^\perp} = \frac{1}{N} \left(Y_{k|2k-1} \mid U^{\perp}_{k|2k-1} \right) \left(P_{0|k-1} \mid U^{\perp}_{k|2k-1} \right)^T.$$

The the following decomposition is performed:

$$USV^T := L^{-1}_{y|u^\perp} \Sigma_{yp|u^\perp} L^{-T}_{p|u^\perp}$$

i.e., it corresponds to the factorization (7) with weighting matrices

$$W_1 = L^{-1}_{y|u^\perp} \quad W_2 = \Pi_{U^{\perp}_{k|2k-1}} \tag{11}$$

5 Estimation of A, C

The estimation of the matrices A and C are usually performed in two different ways.

One approach is based on the preliminary construction of an approximated state, say $\tilde{X}_k$ and it conditional shift $\tilde{X}_{k+1}$, (CVA, N4SID "approximated", [21, 48]) or of a "pseudostate" (together with its shifted version) say Z_{k+1}, Z_k, (N4SID, [48]) from which (A, C) are estimated directly solving a linear least square problem, namely:

$$\begin{pmatrix} \tilde{X}_{k+1} \\ Y_k \end{pmatrix} \simeq \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \tilde{X}_k \\ U_k \end{pmatrix} \oplus \begin{pmatrix} K(k)\hat{E}_k \\ \hat{E}_k \end{pmatrix} \tag{12}$$

where the approximate state is computed from the oblique predictor $E_{\|U_{k|2k-1}}\left[Y_{k|2k-1} \mid P_{0|k-1}\right]$ as $\tilde{X}_k := \hat{\Gamma}_k^{-L} E_{\|U_{k|2k-1}}\left[Y_{k|2k-1} \mid P_{0|k-1}\right]$. Similarly, using instead the pseudostate Z_k, the following recursion can be shown to hold [48]

$$\begin{pmatrix} Z_{k+1} \\ Y_k \end{pmatrix} = \begin{pmatrix} A & \mathcal{K}_1 \\ C & \mathcal{K}_2 \end{pmatrix} \begin{pmatrix} Z_k \\ U_{k|2k-1} \end{pmatrix} \oplus \begin{pmatrix} K(k)\hat{E}_k \\ \hat{E}_k \end{pmatrix}. \tag{13}$$

The pseudo-state Z_k is computed starting from the predictor

$$\hat{Y}_{k|2k-1} = E\left[Y_{k|2k-1} \mid P_{0|k-1} \vee U_{k|2k-1}\right]$$

as $Z_k := \hat{\Gamma}_k^{-L} \hat{Y}_{k|2k-1}$.

With these approaches one may also estimate (B, D) directly from (12) or solving an overdetermined linear set of equations from $\mathcal{K}_1$ and $\mathcal{K}_2$ obtained in (13). We shall further comment on this later in section 6.1.

On the other hand one could enforce the shift invariance structure of the observability matrix $\Gamma_k = \begin{bmatrix} C^T & A^T C^T & \dots & \left(A^{k-1}\right)^T C^T \end{bmatrix}^T$.

The matrix C can be taken to be the first p rows of the estimated observability matrix $\hat{\Gamma}_k$. Let us denote by $\widehat{\overrightarrow{\Gamma}}_k$ the estimated observability matrix with the first p rows deleted. It is straightforward to see that the the matrix A should satisfy

$$\hat{\Gamma}_{k-1} A = \widehat{\overrightarrow{\Gamma}}_k.$$

This equation, is not satisfied exactly for finite data when stochastic disturbances are present and hence it has to be solved approximately.

This is usually done in a variety of ways including least squares solution, total least squares, subspace fitting; let us just recall the most common solutions obtained by least squares as

$$\hat{A} = \hat{\Gamma}_{k-1}^{\dagger} \widehat{\overrightarrow{\Gamma}}_k$$

and by total least squares computing the singular value decomposition

$$\begin{bmatrix} \hat{\Gamma}_{k-1} & \widehat{\overrightarrow{\Gamma}}_k \end{bmatrix} = U \begin{pmatrix} S_n & 0 \\ 0 & \tilde{S}_n \end{pmatrix} \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}^T$$

and setting

$$\hat{A} = -V_{12} V_{22}^{\dagger}.$$

For a discussion on these topics see for instance [29] and the references therein.

Our current implementation allows the choice between least squares and total least squares. However, practical experience has shown that there are no big differences between the two approaches; moreover it can be shown that they are asymptotically equivalent [45].

6 Estimation of B and D

It is well known that, once an estimate of A and C has been obtained, the problem of estimating B and D can be formulated as a linear least squares problem. As we have seen in section 5 some approaches for the estimation of (A, C) yield naturally also estimates for (B, D). This is the case for the CVA algorithm of Larimore [21] and for N4SID [48]. In addition a number of different procedures have been proposed in the literature which yields consistent results as the length of data sequences N goes to infinity. However it is not clear which of them gives better results. Our algorithm implements the most common procedures and some variants which seem to give better results in some ill conditioned cases.

As a guideline we may say that the approach proposed by Van Overschee and De Moor with some minor modifications and the "optimally weighted" projection approach (see section 6.6) seem to give the best results.

We shall briefly describe the different approaches.

6.1 N4SID based approach

As we have anticipated in section 5, B and D can be estimated solving an overdetermined set of linear equations, starting from the estimated parameters $\mathcal{K}_1$ and $\mathcal{K}_2$ (see eq. (13)), which are linear functions of B and D once A and C are given. The equations are as follows:

$$\mathcal{K}(B,D) := \begin{pmatrix} \mathcal{K}_1(B,D) \\ \mathcal{K}_2(B,D) \end{pmatrix} = \begin{pmatrix} B - A\Gamma_k^\dagger \begin{pmatrix} D \\ \Gamma_{k-1}B \end{pmatrix} & \Gamma_{k-1}^\dagger H_{k-1}^d - A\Gamma_k^\dagger \begin{pmatrix} 0 \\ H_{k-1}^d \end{pmatrix} \\ D - C\Gamma_k^\dagger \begin{pmatrix} D \\ \Gamma_{k-1}B \end{pmatrix} & -C\Gamma_k^\dagger \begin{pmatrix} 0 \\ H_{k-1}^d \end{pmatrix} \end{pmatrix}. \tag{14}$$

The solution is found solving the weighted problem

$$\min_{B,D} \|(\hat{\mathcal{K}} - \mathcal{K}(B,D))L\|^2$$

which is linear in the elements of B and D. Different choices of L are possible. However the most common are $L = I$, which corresponds to the standard $N4SID$ and L computed from QR factorization of future inputs as $U_{k|2k-1} = LQ$ which corresponds to the so called "robust" algorithm in [49].

6.2 Minimum Prediction Error

A possible solution is computed via the minimization of the prediction error $y(k) - \hat{y}(k)$. Note that the Kalman gain K is needed, which however can be determined without the knowledge of (B, D). Since also the initial condition will be estimated, the gain corresponding to the stationary solution can be used. The one step ahead predictor can be written as:

$$\begin{aligned} \hat{y}(t) &= C(A-KC)^t x(0) + \sum_{i=0}^{t-1} C(A-KC)^{t-1-i}(B-KD)u(i) + \\ &\quad + \sum_{i=0}^{t-1} C(A-KC)^{t-1-i}Ky(i) + Du(t) \end{aligned}$$

Because of linearity in B, D, $x(0)$ the minimization of the cost functional

$$J_{pred}(B,D,x(0)) = \sum_{k=0}^{T} \|y(k) - \hat{y}(k)\|^2 ,$$

can be easily performed.

6.3 Minimum Simulation Error

Another approach has been proposed (see for instance [29]) which is based on the minimization of the "simulation error"

$$y(t) - \hat{y}(t) = y(t) - \left[CA^t x(0) + \sum_{i=0}^{t-1} CA^{t-1-i}Bu(i) + Du(t) \right] \tag{15}$$

which is a linear functional of $x(0)$, $vec(B)$, $vec(D)$. Therefore, minimizing the cost functional

$$J_{sim}(B,D,x(0)) = \sum_{k=0}^{T} \|y(k) - \hat{y}(k)\|^2 ,$$

with respect to B, D, $x(0)$, is just solving a linear least-squares problem.

6.4 Block Minimum Simulation Error

Let us denote $\hat{Y}^d_{0|2k-1} = \Gamma_{2k}\hat{X}^d_o + H^d_{2k}U_{0|2k-1}$ the projection of the outputs $Y_{0|2k-1}$ on the space spanned by the inputs. Assuming A and C known, the i-th column $\hat{Y}^d_{0|2k-1}(i)$ of the matrix $\hat{Y}^d_{0|2k-1}$ is a linear functional of $\hat{X}^d_o(i)$, i.e. the i-th column of the initial state and of vec (B) and vec (D):

$$\hat{Y}^d_{0|2k-1}(i) = \Gamma_{2k}\hat{X}^d_o(i) + H^d_{2k}U_{0|2k-1}(i).$$

Therefore one could consider the following as a cost function for the estimation of B and D:

$$J_{sim_b}(B, D, x_1, x_2, \ldots, x_N) = \sum_{i=1}^{N} \left\| \hat{Y}^d_{0|2k-1}(i) - \Gamma_{2k}x_i - H^d_{2k}U_{0|2k-1}(i) \right\|^2.$$

Again this is linear in $(vec(B), vec(D), x_1, x_2, \ldots, x_N)$, for some choice of the integer N, which will be a tradeoff between speed and accuracy.

6.5 Projection Approach

rewriting the second equation in (3) as

$$Y_{0|k-1} = \Gamma_k X_0 + H^d_k U_{0|k-1} + H^s_k E_{0|k-1}.$$

and projecting this equation onto the space spanned by the inputs $U_{0|k-1}$, one obtains

$$\hat{Y}^d_{0|k-1} := E\left[Y_{0|k-1}|U_{0|k-1}\right] := \Phi U_{0|k-1} = \Gamma_k E\left[X_0|U_{0|k-1}\right] + H^d_k U_{0|k-1}$$

The first term on the right hand side may be removed by multiplying from the left by $\left(\Gamma^\perp_k\right)^T$. For convenience of notation let us denote

$$\Pi_{\Gamma^\perp} := \left(\Gamma^\perp_k\right)^T$$

Since $\Pi_{\Gamma^\perp}\Gamma_k\hat{X}_0 = 0$, in this way we obtain:

$$\Pi_{\Gamma^\perp}\hat{Y}^d_{0|k-1} = \Pi_{\Gamma^\perp}H^d_k U_{0|k-1}$$

Once the matrix $\bar{H}^d_k := \Pi_{\Gamma^\perp}H^d_k = \Pi_{\Gamma^\perp}\Phi$ is available, we obtain

$$K\left[\begin{array}{c} D \\ \hline B \end{array}\right] = \left[\begin{array}{c} \bar{H}_k(:, 1:m) \\ \hline H_k(:, m+1, 2m) \\ \hline \vdots \\ \hline H_k(:, m(k-1):km) \end{array}\right] \tag{16}$$

where

$$K := \left[\begin{array}{c|c} \Pi_{\Gamma^\perp}(:, 1:p) & \Pi_{\Gamma^\perp}(:, p+1:kp)\Gamma_{k-1} \\ \hline \Pi_{\Gamma^\perp}(:, p+1:2p) & \Pi_{\Gamma^\perp}(:, 2p+1:kp)\Gamma_{k-2} \\ \hline \vdots & \vdots \\ \hline \Pi_{\Gamma^\perp}(:, (k-1)p+1:2p) & 0 \end{array}\right]$$

This is implemented by the function `BD_proj.m`.

A similar solution can be obtained orthonormalizing the inputs via LQ factorization [2], i.e. computing

$$LQ = U_{0|k-1}$$

and then solving the weighted problem

$$\min_{B,D} \|(\Pi_{\Gamma^\perp}\Phi - \Pi_{\Gamma^\perp}H^d_k(B, D))L\|^2$$

which is still linear in the elements of B and D. This solution gives much more robust results when some canonical angles between the rows of $U_{0|k-1}$ are small.

[2]This clearly can also be done via SVD decomposition.

6.6 Optimally weighted projection

In this section we shall consider a procedure to estimate the matrices (B, D) which is slightly different from standard procedures proposed in the literature. We shall also see that the procedure proposed in [52] is just a special case. Under this framework it is possible to show that in an ideal situation, i.e. if A and C were known exactly, this approach would guaratee lower variance of the estimated pair $(\hat{B}_d, \hat{D}_d)$.

Let $\hat{Y}_{0|k-1} := E\left[Y_{0|k-1} \mid U_{0|k-1}\right]$ be the projection of outputs onto the input space. Making the dependence on system parameters explicit, we have:

$$\hat{Y}_{0|k-1} = \Gamma_k E\left[X_0 \mid U_{0|k-1}\right] + H_k^d(B,D)U_{0|-1} + H_k^s E\left[E_{0|k-1} \mid U_{0|k-1}\right]. \tag{17}$$

The third term should ideally be zero; in practice, due to finite length effects, it is not. Let us call this perturbation term R_k, i.e.

$$R_k^+ := H_k^s E\left[E_{0|k-1} \mid U_{0|k-1}\right] = H_k^s \hat{R}_{EU} R_{UU}^{-1} U_{0|k-1}$$

with obvious meaning of symbols. Defining $\hat{\Phi}_x$ such that $\hat{\Phi}_x U_{0|k-1} = E\left[E_{0|k-1} \mid U_{0|k-1}\right]$, we can rewrite equation (17) as

$$\hat{\Phi} U_{0|k-1} := \hat{Y}_{0|k-1} = \Gamma_k \Phi_x U_{0|k-1} + H_k^d(B,D)U_{0|k-1} + H_k^s \hat{R}_{EU} \hat{R}_{UU}^{-1} U_{0|k-1} \tag{18}$$

which turns out to be an equation for the coefficients of the following form

$$\hat{\Phi} := \Gamma_k \Phi_x + H_k^d(B,D) + H_k^s \hat{R}_{EU} R_{UU}^{-1} \tag{19}$$

This equation is linear in the parameters (Φ_x, B, D) and can be easily rewritten in the form

$$\mathrm{vec}\left(\hat{\Phi}_y\right) = \left[I_{km} \otimes \Gamma_k\right] \mathrm{vec}\left(\Phi_x\right) + L_d \mathrm{vec}\begin{pmatrix} B \\ D \end{pmatrix} + \left(R_{UU}^{-1} \otimes H_k^s\right) \mathrm{vec}\left(\hat{R}_{EU}\right) \tag{20}$$

for some matrix L_d.

The last term is regarded as a perturbation which covariance matrix may be easily computed. In fact, define

$$R_{UU}(\tau) = \lim_{N\to\infty} \frac{1}{N} U_{\tau|\tau+k-1} U_{0|k-1}^T$$

and

$$R_{EE}(\tau) = \lim_{N\to\infty} \frac{1}{N} E_{\tau|\tau+k-1} E_{0|k-1}^T$$

it follows from standar calculations that

$$\Sigma_{res} = E\left[\mathrm{vec}\left(\hat{R}_{EU}\right) \mathrm{vec}^T\left(\hat{R}_{EU}\right)\right] = \frac{1}{N} \sum_{|\tau| \le T-t} \left(1 - \frac{|\tau|}{N}\right) R_{UU}(\tau) \otimes R_{EE}(\tau)$$

Defining

$$W_0 := \left(R_{UU}^{-1} \otimes H_k^s\right) \Sigma_{res} \left(R_{UU}^{-1} \otimes H_k^s\right)^T$$

and

$$\Theta := W_0^{-1/2} \left[I_{km} \otimes \Gamma_k\right]$$

we can write the estimate of the input matrices as:

$$\mathrm{vec}\begin{pmatrix} \hat{B} \\ \hat{D} \end{pmatrix} = \left(L_d^T W_0^{-1/2} \Theta^{\perp} W_0^{-1/2} L_d\right)^{-1} L_d^T W_0^{-1/2} \Theta^{\perp} W_0^{-1/2} \mathrm{vec}\left(\hat{\Phi}_y\right) \tag{21}$$

i.e., the oblique projection of $\mathrm{vec}\left(\hat{\Phi}_y\right)$ onto the column span of $W^{-1/2} L_d$ along the column span of Θ.

It easy to verify that the projection approach, which was first proposed in Verhaegen [52], is just a particular case of this with a suitably choosen weighting matrix W, different from the optimal W_0 computed above.

7 Simulation results

In this section some simulation results comparing the joint approach with the orthogonal decomposition algorithm are presented. For reason of space we are only able to present one possible choice of the different steps 1), 2) and 3), i.e. the robust N4SID of [49] and the orthogonal decomposition algorithm with 1) corresponding to CVA, 2) using the pseudo-state and 3) with the optimally weighted projection.

Cramer Rao lower bounds for the variance of the estimated transfer function are presented. The Cramer Rao lower bound corresponding to a block parametrization of the form (5) is lower that the CR lower bound for the joint parametrization when the deterministic and stochastic subsystem have completely disjoint dynamics. The opposite happens when the dynamic is fully shared. Even though in our simulations the CR bound has not been reached the plots show how in the case where the dynamics is disjoint the orthogonal decomosition does better, as expected. The interested reader may contact the authors for more informations concernig the eperimental conditions.

The plots show the mean square error of estimatd transfer function (deterministic system and minimum phase spectral factor of stochastic component) versus frequency (ranging from 0 to π) and the corresponding CR lower bound.

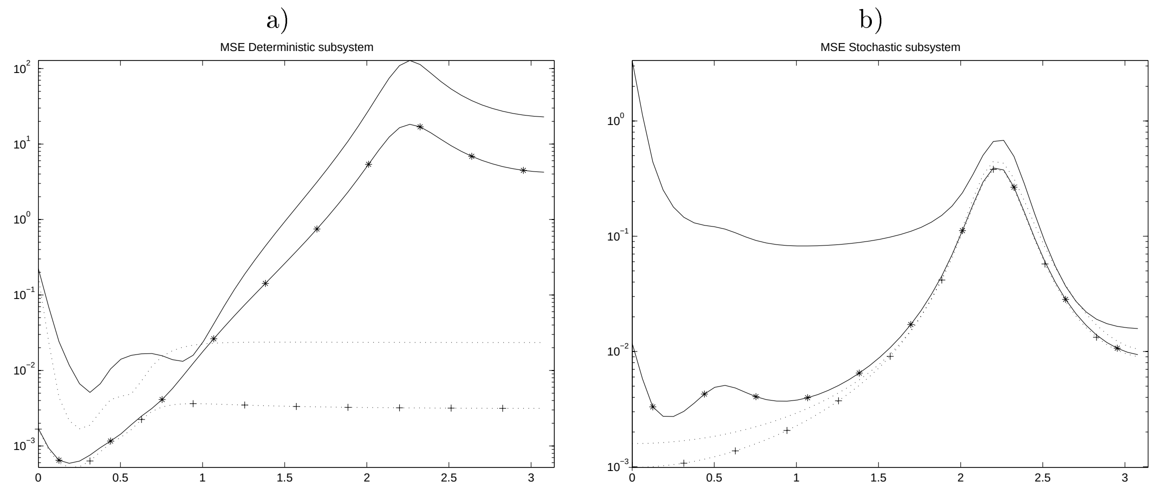

Figure 1: Example 1: the deterministic and stochastic dynamics are completely disjoint. Estimated (Monte Carlo) MSE of transfer functions versus Cramer-Rao lower bound. Left: deterministic subsystem, right stochastic subsystem; dotted : ortogonal decomposition algorithm, solid : joint (N4SID robust). The dotted line with crosses is the CR lower bound for block parametrization, the solid line with stars is the CR lower bound for joint parametrization.

References

[1] H. Akaike, Markovian representation of stochastic processes by canonical variables, *SIAM J. Control*, **13** , pp. 162–173, (1975).

[2] H. Akaike, Stochastic Theory of Minimal Realization. *IEEE Trans. Automat. Contr.*, vol. AC-19, no. 6, pp. 667–674, 1974.

[3] H. Akaike, Canonical Correlation Analysis of Time Series and the Use of an Information Criterion. *System Identification: Advances and Case Studies* (R. Mehra and D. Lainiotis, Eds.), Academic, 1976, pp. 27–96.

a) b)

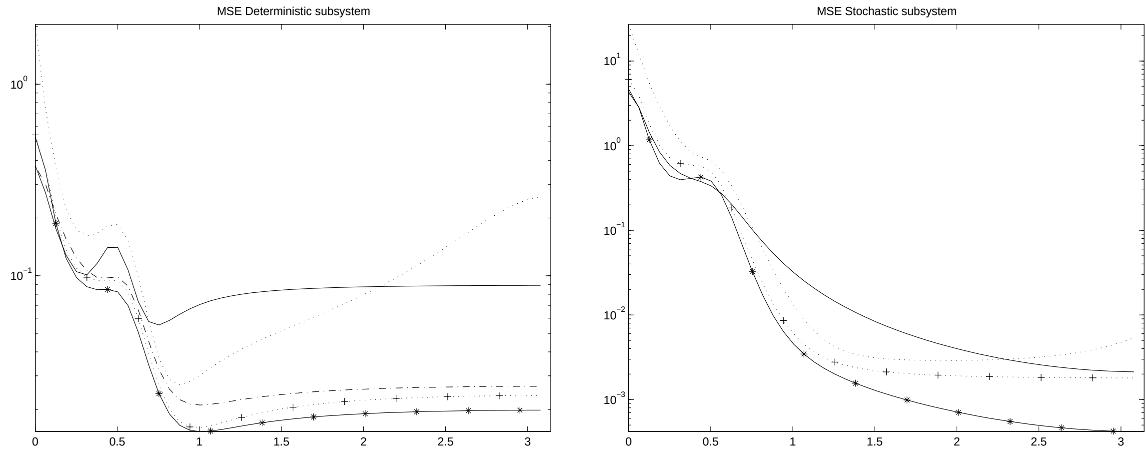

Figure 2: Example 2: the deterministic and stochastic dynamics are in common (a minimal realization has the same order as a minimal realization of both the stochastic and deterministic part.) Estimated (Monte Carlo) MSE of transfer functions versus Cramer-Rao lower bound. Left: deterministic subsystem, right stochastic subsystem; dotted : ortogonal decomposition algorithm, solid : joint (N4SID robust), dashed: other joint algorithm. The dotted line with crosses is the CR lower bound for block parametrization, the solid line with stars is the CR lower bound for joint parametrization.

[4] Aoki, M. (1990). *State Space Modeling of Time Series* (2nd ed.), Springer.

[5] Bauer, D. *Some Asymptotic Theory for the Estimation of Linear Systems Using Maximum Likelihood Methods or Subspace Algorithms.* PhD Thesis, TU Wien 1998.

[6] Bauer, D. (2000) *Asymptotic efficiency of the CCA subspace method in the case of no exogenous inputs.* Submitted for pubblication.

[7] Bauer, D. and Jansson M. Analysis of the asymptotic properties of the moesp type of subspace algorithms. *Automatica*, 2000.

[8] S. Bittanti, M Lovera, Assessing Model Uncertainty in Subspace Identification Methods: a Simulation Study, in *Proceedings MTNS-98*, A. Beghi, L. Finesso, G. Picci eds, Il Poligrafo, Padova 1998, pp. 703–706.

[9] Chiuso, A. and Picci G. Subspace Identification by Orthogonal Decomposition. In *Proc. 14th IFAC World Congress* , Vol. I, 241–246 (1999).

[10] Chiuso, A. and Picci G. Error Analysis of Certain Subspace Methods. *Proc. of SYSID-2000* S. Barbara Calif. (2000).

[11] Chiuso, A. *Geometric Methods for Subspace Identifcation*, Doctoral Thesis, Department of Electronics and Informatics, University of Padova, Italy, Feb. 2000.

[12] Chiuso, A. and Picci G. Probing inputs for subspace Identification. To appear in *Proc. of 39th CDC* , Sydney, Australia, Dec. 2000.

[13] Chiuso, A. and Picci G. On the ill-conditioning of certain subspace identification methods with inputs. In preparation, (2000).

[14] Faurre, P. Stochastic realization algorithms. In R. Mehra and D. Lainotis, editors, *System Identification: Advances and Case Studies.* Academic Press, New York, 1976.

[15] R.P. Guidorzi, Invariants and canonical forms for systems: structural and parameter identification, *Automatica*, 17, (1963), pp.117-133, 1981.

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

[17] Hotelling H. (1936). Relations between two sets of variables. *Biometrica*, **28**, 321-377.

[18] H. Kawauchi, A. Chiuso, T. Katayama, G.Picci. Comparison of Two Subspace Identification Methods fo rCombined Deterministic -Stochastic Systems, *Proc. of The 31st ISCIE International Symposium on Stochastic Systems Theory and its Applications*, Yokohama, Japan, 1999.

[19] T. Katayama and G. Picci (1999). Realization of Stochastic Systems with Exogenous Inputs and Subspace System Identification Methods, *Automatica*, vol 35, No. 10, pp.1635-1652.

[20] T. Katayama, S. Omori and G. Picci: Comparison of some subspace identification methods, in *Proceedings of the 1998 Conference on Decision and Control*, Tampa, Florida, paper n. TA08-6 p. 1850-1852, 1998.

[21] W. E. Larimore, System identification, reduced-order filtering and modeling via canonical variate analysis, *Proc. American Control Conference*, 1983, pp. 445–451.

[22] Larimore, W. E. (1990). Canonical variate analysis in identification, filtering, and adaptive control. In *Proc. 29th IEEE Conf. Decision & Control*, Honolulu, 596–604.

[23] Ljung, L. (1987). *System Identification – Theory for the User*. Prentice-Hall.

[24] A. Lindquist, G. Picci and G. Ruckebusch On minimal splitting subspaces and Markovian representation, *Math. System Theory*, **12**: 271-279, 1979.

[25] A. Lindquist and G. Picci, On the stochastic realization problem *SIAM J. Control and Optimization*, **17**: 365–389, 1979.

[26] A. Lindquist and M. Pavon, On the structure of state space models of discrete-time vector processes, *IEEE Tr. on Automatic Control*, **AC-29**, p.418-432, 1984.

[27] A. Lindquist and G. Picci, A geometric approach to modelling and estimation of linear stochastic systems, *Journal of Mathematical Systems, Estimation and Control*, **1**:241–333, 1991.

[28] A. Lindquist and G. Picci, Canonical correlation analysis approximate covariance extension and identification of stationary time series, *Automatica*, vol. 32, pp. 709-733, 1996.

[29] M. Lovera PhD. *Subspace Metods: Theory and Applications*, Politecnico di Milano, Ph.D. Thesis 1997.

[30] Jansson, M. and Wahlberg, B. (1997). Counterexample to the general concistency of subspace system identification metods. In *Proceedings of SYSID97,* Fukuoka, Japan, 1677-1682.

[31] A. Lindquist and G. Picci, Geometric Methods for State Space Identification, in *Identification, Adaptation, Learning*, (Lectures given at the NATO-ASI School, *From Identifiation to Learning* held in Como, Italy, Aug.1994), Springer Verlag, 1996.

[32] M. Moonen, B. De Moor, L. Vanderberghe and J. Vandewalle, On- and Off-Line Identification of Linear State-Space Models. *Int. J. Control* **49** (1989), pp. 219–232.

[33] M. Moonen and J. Vandewalle, QSVD Approach to On- and Off-Line State-Space Identification. *Int. J. Control* **51** (1990), pp. 1133–1146.

[34] Peternell, K., W. Scherrer and M. Deistler (1996). Statistical analysis of novel subspace identification methods. *Signal Processing*, **52**, 161–177.

[35] Peternell, K. *Identification of linear time-invariant systems via subspace and realization-based methods.* TUWien PhD Thesis, 1995.

[36] G. Picci, Stochastic realization of Gaussian processes *Proc. of the IEEE*, **64** (1976), pp. 112-122.

[37] G. Picci and T. Katayama, Stochastic realization with exogenous inputs and "Subspace Methods" Identification, *Signal Processing*, 1996, **52**, 145–160.

[38] G. Picci and T. Katayama: "A simple "subspace" identification algorithm with exogenous inputs", *Proceedings of the 1996 triennial IFAC Congress*, San Francisco, Ca., paper n. 0916, session 3a-06-5.

[39] G. Picci, Geometric Methods in Stochastic Realization and System Identification *CWI Quarterly* special Issue on System Theory, **9**, pp. 205-240, 1996.

[40] G. Picci, Stochastic Realization and System Identification, in *Statistical Methods in Control and Signal Processing*, T. Katayama and I. Sugimoto eds, M. Dekker, N.Y. 1997.

[41] G. Picci, Oblique Splitting Susbspaces and Stochastic Realization with Inputs in *Operators, Systems and Linear Algebra*, U. Helmke, D. Prätzel-Wolters and E. Zerz eds. pp. 157-174, Teubner, Stuttgart, 1997.

[42] G. Picci, "Statistical properties of certain subspace identification methods", *Proceedings of the SYSID 97, Fukuoka, Japan*, vol**3**, pp. 1093-1099, 1997.

[43] G. Ruckebusch, Répresentations markoviennes de processus gaussiens stationnaires, *C.R.Acad.Sc.Paris* , Series A, **282**, p. 649–651, 1976.

[44] G. Ruckebusch, A state space approach to the stochastic realization problem, *Proc. 1978 IEEE Intern. Symp. Circuits and Systems*, p. 972–977, 1978.

[45] Stoica, P. and Viberg, M. (1995), Weighted LS and TLS approaches yields asympotically equivalent results, *Signal Processing*, 45(2).

[46] R. J. Ober, Balanced Canonical Forms, in *Identification, Adaptation, Learning*, (Lectures given at the NATO-ASI School, *From Identifiation to Learning* held in Como, Italy, Aug.1994), Springer Verlag, 1996.

[47] P. Van Overschee and B. De Moor, Subspace algorithms for the stochastic identification problem, *Automatica* **29** (1993), pp. 649–660.

[48] P. Van Overschee and B. De Moor, N4SID: Subspace algorithms for the identification of combined deterministic– stochastic systems, *Automatica* **30** (1994), pp. 75–93.

[49] Van Overschee, P. and B. De Moor (1996). *Subspace Identification for Linear Systems.* Kluwer Academic Publications.

[50] Verhaegen, M. and P. Dewilde (1992). Subspace model identification, Part 1. The output-error state-space model identification class of algorithms; Part 2. Analysis of the elementary output-error state-space model identification algorithm. *Int. J. Control*, **56**, 1187–1210 & 1211–1241.

[51] M. Verhaegen, Application of a Subspace Model Identification Technique to Identify LTI Systems Operating in Closed-loop, *Automatica* **29** (1993), pp. 1027–1040.

[52] M. Verhaegen, Identification of the deterministic part of MIMO State Space Models given in Innovations form from Input-Output data, *Automatica* **30** (1994), pp. 61–74.

[53] T. Chou, M. Verhaegen, Subspace Algorithms for the Identification of Multivariable Dynamic Errors-in-Variables Models, *Automatica* **33** (1997), pp. 1857–1869.

[54] Viberg, M. (1995). Subspace-based methods for the identification of linear time-invariant systems. *Automatica*, **31**, 1835–1851.