Available online at www.sciencedirect.com
SCIENCE DIRECT®
Automatica 40 (2004) 1689–1703
automatica
www.elsevier.com/locate/automatica

Subspace identification by data orthogonalization and model decoupling☆

Alessandro Chiuso, Giorgio Picci*

Department of Information Engineering, University of Padova, via Gradenigo 6/a, 35131 Padova, Italy

Received 7 October 2002; received in revised form 27 April 2004; accepted 18 May 2004

Abstract

It has been observed that identification of state-space models with inputs may lead to unreliable results in certain experimental conditions even when the input signal excites well within the bandwidth of the system. This may be due to ill-conditioning of the identification problem, which occurs when the state space and the future input space are nearly parallel.

We have in particular shown in the companion papers (Automatica 40(4) (2004) 575; Automatica 40(4) (2004) 677) that, under these circumstances, subspace methods operating on input–output data may be ill-conditioned, quite independently of the particular algorithm which is used. In this paper, we indicate that the cause of ill-conditioning can sometimes be cured by using orthogonalized data and by recasting the model into a certain natural block-decoupled form consisting of a "deterministic" and a "stochastic" subsystem. The natural subspace algorithm for the identification of the deterministic subsystem is then a weighted version of the PI-MOESP method of Verhaegen and Dewilde (Int. J. Control 56 (1993) 1187–1211). The analysis shows that, under certain conditions, methods based on the block-decoupled parametrization and orthogonal decomposition of the input–output data, perform better than traditional joint-model-based methods in the circumstance of nearly parallel regressors.

Keywords: Subspace identification; Exogenous inputs; Stochastic realization; Statistical analysis; Stochastic state-space identification

1. Introduction

As observed in (Chiuso & Picci, 1999; Kawauchi, Chiuso, Katayama, & Picci, 1999) the standard subspace methods for identification with inputs (e.g. the N4SID and MOESP-type methods) may lead to unreliable results in certain experimental conditions. As discussed in a previous paper (Chiuso & Picci, 2004d), this behavior can be explained in terms of ill-conditioning of the underlying multiple regression problem which occurs when the future input space and the state space of the system are nearly parallel (i.e. some canonical angle is near zero). It is well-known in regression analysis that the ill-conditioning due to "almost parallel" regressors can be cured by orthogonalizing the data in an appropriate way. In this paper, it is also argued that ill-conditioning can be cured (at least in certain cases) by reformulating the problem using orthogonalized data. Scope of this paper is to demonstrate that the use of a preliminary orthogonal decomposition of the data, together with appropriate subspace algorithms adapted to this decomposition, in the spirit of Verhaegen and Dewilde (1993), Picci and Katayama (1996a), may, in certain circumstances, lead to more robust and accurate estimates. As discussed in Picci and Katayama (1996a), the orthogonal decomposition of the data induces a "canonical" decomposition of the model into a "stochastic" and a "deterministic" subsystem. It turns out that the PI-MOESP method introduced by Verhaegen and Dewilde (1993) in the framework of instrumental variable identification, is in fact a (MOESP-type) subspace identification method of the deterministic subsystem corresponding to the orthogonal decomposition philosophy mentioned above.

The conditioning analysis of Chiuso and Picci (2004d) considers linear models describing jointly the dynamics of

☆ This work has been supported by the national project *Identification* and adaptive control of industrial systems funded by MIUR. Part of this work has been presented at the 2003 IFAC Symposium on System Identification (SYSID) in Rotterdem. This paper was recommended for publication in revised form by Associate Editor Brett Ninness under the direction of Editor Torsten Söderström.

* Corresponding author. Tel.: +39-049-827-7705; fax: +39-049-827-7699.
E-mail address: picci@dei.unipd.it (G. Picci).

0005-1098/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2004.05.008

the deterministic and stochastic components of the output process. These models allow for common poles in the deterministic transfer function and in the shaping filter which models the stochastic error/disturbance process. In practice, however modeling a common dynamics in the two subsystems is seldom needed and in the frequent case where the true deterministic and stochastic output components are dynamically decoupled, the "joint" models turn out to be badly overparametrized. Models with decoupled deterministic and stochastic dynamics admit instead a natural block-diagonal canonical form which involves less parameters and fits in a natural way the preliminary data orthogonalization mentioned before. In this spirit, we shall discuss the numerical conditioning of subspace identification of decoupled models (assuming disjoint dynamics) and compare with that of joint models.

Since ill-conditioning depends on the input signal, in order to assess and compare the performance of competing identification methods, some standard "worst case" input classes should be defined, and the performance of candidate methods should be compared relative to this class of input signals. In this paper, we propose a definition of worst-case input signals which may be called *probing inputs*. As anticipated in Chiuso and Picci (2000), the probing inputs can be defined and designed, in such a way as to lead to the largest state-to-input correlation coefficients, and hence to the worst possible conditioning of the identification problem, for a fixed input power/bandwidth. Numerical experiments are included demonstrating how these input signals may lead to a substantial deterioration of performance in the algorithms using joint model parametrization with respect to those using decoupled models.

Concerning this and previous papers dealing with ill-conditioning of subspace methods, one general remark is in order: we do not want to give the reader the impression that the possible loss of accuracy which may be incurred in presence of ill-conditioning is due to the use of subspace algorithms. Quite the contrary, the Cramèr Rao bounds in the simulations show that also (theoretically) optimal methods, like prediction error methods, will behave, under the same conditions, approximately the same as subspace methods. A point which should be made is that subspace methods are based on well understood system theoretic principles (stochastic realization) and hence are amenable of rather in-depth analysis. Consequently for subspace methods these phenomena can be thoroughly analyzed and, in certain cases, an appropriate cure may be found. Much less of this analysis seems to be possible for prediction error methods.

The structure of the paper is as follows:

- In Section 2, we review the basic background of subspace identification and discuss the orthogonal decomposition of linear stochastic models into a block-diagonal structure where the stochastic and deterministic subsystems are completely decoupled.
- In Section 3, we discuss the finite-interval stochastic realization problem of the deterministic subsystem and describe a finite-time subspace identification algorithm which is a weighted version of the PI-MOESP algorithm of Verhaegen and Dewilde (1993). We also do some elementary numerical conditioning analysis of subspace identification for the deterministic subsystem.
- In Section 4 a class of input signals, called *probing inputs* is introduced which lead to the largest state-input correlation coefficients and hence enhance the possible ill-conditioning of the identification problem.
- Section 6 contains a discussion of the simulation results and some conclusions.

The comparison of the two methods should ultimately be based on asymptotic error variance formulas for the estimates. These formulas are derived in a companion paper (Chiuso & Picci, 2004b) and we shall defer final comments on this issue to tis paper.

2. State-space models for subspace identification

Assume that the observed input–output data of the unknown system, which we want to identify

$$\{u_{t_0},\dots,u_t,\dots\},\ \{y_{t_0},\dots,y_t,\dots\},\quad u_t\in\mathbb{R}^p,\ \ y_t\in\mathbb{R}^m, \tag{2.1}$$

are sample paths of a pair of zero-mean second-order stationary random processes $\mathbf{y}=\{\mathbf{y}(t)\}$, $\mathbf{u}=\{\mathbf{u}(t)\}$ having a rational spectral density or, equivalently, assume that (2.1) is generated by a linear stochastic system of the form

$$\begin{aligned}\mathbf{x}(t+1)&=A\mathbf{x}(t)+B\mathbf{u}(t)+G\mathbf{w}(t),\\ \mathbf{y}(t)&=C\mathbf{x}(t)+D\mathbf{u}(t)+J\mathbf{w}(t),\end{aligned}\qquad t\geqslant t_0, \tag{2.2}$$

where A,B,G,C,D,J are constant matrices, $\{\mathbf{x}(t)\}$ is the state process of dimensions n, and $\{\mathbf{w}(t)\}$ is a normalized white noise process. We shall always make the assumption that *there is no feedback from* $\mathbf{y}$ *to* $\mathbf{u}$. This implies that the processes $\{\mathbf{u}(t)\}$ and $\{\mathbf{w}(t)\}$ are completely uncorrelated. See e.g. Caines and Chan (1976), Gevers and Anderson (1981), Picci and Katayama (1996a) for a discussion of this concept.

System (2.2) is called a *stationary stochastic realization* of the output process $\mathbf{y}$ with input $\mathbf{u}$. There are always infinitely many such linear representations of $\mathbf{y}$, which are equivalent up to (conditional) second-order statistics. A realization which is unique up to change of basis, is the so-called "innovation representation"

$$\begin{aligned}\mathbf{x}(t+1)&=A\mathbf{x}(t)+B\mathbf{u}(t)+K\mathbf{e}(t),\\ \mathbf{y}(t)&=C\mathbf{x}(t)+D\mathbf{u}(t)+\mathbf{e}(t),\end{aligned} \tag{2.3}$$

where the white noise $\{\mathbf{e}(t)\}$ has the meaning of (stationary) one step prediction error of $\{\mathbf{y}(t)\}$, given the infinite past history of $\{\mathbf{y}(t)\}$ $\{\mathbf{u}(t)\}$ up to time $t-1$.

For obvious reasons, in identification one wants to use realizations which are (*stochastically*) *minimal*, in the sense

that the state dimension, n, is the smallest possible. This implies in particular (but is not equivalent to) that the triplet $\{C, A, [B\ G]\}$ is minimal in the usual system-theoretic sense.

In order to discuss our approach to the problem, we need to introduce some background concepts and notations. Since this paper is in a sense a continuation of the articles (Chiuso & Picci, 2004c, d), we shall just quickly review the notations in the next subsection and refer the reader to Chiuso and Picci (2004c, d) for a thorough discussion of the background material.

2.1. Notations

As before, boldface symbols will denote random quantities. For $-\infty \leqslant t_0 \leqslant t \leqslant T \leqslant +\infty$ define the Hilbert spaces of scalar zero-mean square integrable random variables

$$\mathscr{U}_{[t_0,t)} := \overline{\text{span}}\{\mathbf{u}_k(s);\ k = 1, \ldots, p,\ t_0 \leqslant s < t\},$$

$$\mathscr{Y}_{[t_0,t)} := \overline{\text{span}}\{\mathbf{y}_k(s);\ k = 1, \ldots, m,\ t_0 \leqslant s < t\},$$

where the bar denotes closure in mean square, i.e. in the metric defined by the inner product $\langle \xi, \eta \rangle := E\{\xi, \eta\}$, the operator E denoting mathematical expectation. We shall let $\mathscr{P}_{[t_0,t)} := \mathscr{U}_{[t_0,t)} \vee \mathscr{Y}_{[t_0,t)}$ denote the joint past space of the input and output processes at time t (the $\vee$ denotes closed vector sum). Similarly, let $\mathscr{U}_{[t,T]}, \mathscr{Y}_{[t,T]}$ be the respective future spaces up to time T

$$\mathscr{U}_{[t,T]} := \overline{\text{span}}\{\mathbf{u}_k(s);\ k = 1, \ldots, p,\ t \leqslant s \leqslant T\},$$

$$\mathscr{Y}_{[t,T]} := \overline{\text{span}}\{\mathbf{y}_k(s);\ k = 1, \ldots, m,\ t \leqslant s \leqslant T\}.$$

By convention the past spaces do not include the present. When $t_0 = -\infty$ we shall use the shorthands $\mathscr{U}_t^-, \mathscr{Y}_t^-$ for $\mathscr{U}_{[-\infty,t)}, \mathscr{Y}_{[-\infty,t)}$, the closed vector sum $\mathscr{U}_t^- \vee \mathscr{Y}_t^-$ being denoted by $\mathscr{P}_t^-$ (the infinite joint past at time t). These are the Hilbert spaces of random variables spanned by the infinite past of $\mathbf{u}$ and $\mathbf{y}$ up to time t.

Subspaces spanned by random variables at just one time instant (e.g. $\mathscr{U}_{[t,t]}, \mathscr{Y}_{[t,t]}$, etc) are simply denoted $\mathscr{U}_t, \mathscr{Y}_t$, etc. while for the spaces generated by the whole time history of $\mathbf{u}$ and $\mathbf{y}$ we shall use the symbols $\mathscr{U}$, $\mathscr{Y}$, respectively.

All through this paper we shall assume that the input process is "sufficiently rich", in the sense that $\mathscr{U}_{[t_0,T]}$ admits the direct sum decomposition

$$\mathscr{U}_{[t_0,T]} = \mathscr{U}_{[t_0,t)} + \mathscr{U}_{[t,T]}, \quad t_0 \leqslant t < T, \tag{2.4}$$

the $+$ sign denoting direct sum of subspaces. The symbol $\oplus$ will be reserved for *orthogonal* direct sum. Condition (2.4) can be found expressed in various equivalent forms in the literature, see e.g. Verhaegen and Dewilde (1992, formula (10)). Also various conditions ensuring sufficient richness are known. For example, it is well-known that for a full-rank purely non-deterministic process $\mathbf{u}$ to be sufficiently rich, it is necessary and sufficient that the determinant of the spectral density matrix Φ_u should have no zeros on the unit circle (Hannan & Poskitt, 1988).

We shall use indexed capitals, e.g. X_t, Y_t, U_t, to denote finite "tail" matrices, constructed at each time t from sample sequences of $\mathbf{x}, \mathbf{y}, \mathbf{u}$, by letting

$$Y_t := [y_t\ y_{t+1}\ \ldots\ y_{t+N-1}],$$

$$U_t := [u_t\ u_{t+1}\ \ldots\ u_{t+N-1}],$$

$$X_t := [x_t\ x_{t+1}\ \ldots\ x_{t+N-1}]. \tag{2.5}$$

Symbols like $Y_{[t,T]}$ will denote a Hankel matrix, e.g.

$$Y_{[t,T]} := [Y_t^\top\ \ldots\ Y_T^\top]^\top$$

and $\mathscr{Y}^N_{[t,T]}$ the corresponding (finite-dimensional) rowspace.

We shall assume *second-order ergodicity* of all random processes involved. Introducing the notation

$$E_N X_t Y_t^\top := \frac{1}{N} X_t Y_t^\top = \frac{1}{N} \sum_{k=0}^{N-1} x_{t+k} y_{t+k}^\top$$

second-order ergodicity means that, $E_N Y_{t+\tau} Y_t^\top \to E\mathbf{y}(t+\tau)\mathbf{y}(t)^\top$ for $N \to \infty$. For this reason $E_N X_t Y_t^\top$ is a consistent estimate of the covariance $\Sigma_{\mathbf{x}(t)\mathbf{y}(t)}$ and will also be denoted by the symbol $\hat{\Sigma}_{\mathbf{x}(t)\mathbf{y}(t)}$. In a sense (which can be made precise), "finite expectations" ($E_N\{\cdot\}$) operations on tail sequences like Y_t, X_t, etc. behave, for $N \to \infty$, like ordinary expectations on the corresponding random variable $\mathbf{y}(t), \mathbf{x}(t)$, etc. This we shall sometimes express by saying that $Y_t \to \mathbf{y}(t)$ as $N \to \infty$. In the same spirit, $Y_{[t,T]}$ "tends to" the $m(T-t+1)$-dimensional column random vector $\mathbf{y}_{[t,T]}$, as $N \to \infty$, and one can say that $\mathscr{Y}^N_{[t,T]} \to \mathscr{Y}_{[t,T]}$ for $N \to \infty$. "Approximating" spaces of random variables by vector spaces spanned by the rows of tail matrices is a standard device in subspace identification. Finally, we shall write

$$E_N[X \mid Y] := E_N[XY^{\mathrm{T}}](E_N[YY^{\mathrm{T}}])^{-1}Y$$

for the well-known linear regression formula solving the (deterministic) least-squares problem $\min_{A \in \mathbb{R}^{n \times m}} \|Y - AX\|$. Let $\mathbf{x}$ and $\mathbf{y}$ be random vectors, linear functions of some jointly second-order ergodic, stationary processes, whose shifted sample values form the finite tail sequences X and Y of length N. It may be seen that the limit (in a suitable sense) for $N \to \infty$, of $E_N[X \mid Y]$ is the wide-sense conditional expectation

$$E[\mathbf{x} \mid \mathbf{y}] := E[\mathbf{x}\mathbf{y}^{\mathrm{T}}](E[\mathbf{y}\mathbf{y}^{\mathrm{T}}])^{-1}\mathbf{y}.$$

2.2. A decoupled canonical form

Let $\mathscr{U}^\perp$ be the orthogonal complement of $\mathscr{U}$ in $\mathscr{U} \vee \mathscr{Y}$. The stochastic processes $\mathbf{y}_{\mathrm{d}}$ and $\mathbf{y}_{\mathrm{s}}$, called the *deterministic* and the *stochastic component* of $\mathbf{y}$, defined by the complementary projections

$$\mathbf{y}_{\mathrm{d}}(t) := E[\mathbf{y}(t) \mid \mathscr{U}], \quad \mathbf{y}_{\mathrm{s}}(t) := \mathbf{y}(t) - E[\mathbf{y}(t) | \mathscr{U}] \tag{2.6}$$

are obviously uncorrelated at all times. It follows that $\mathbf{y}$ admits an orthogonal decomposition as the sum of its

deterministic and stochastic components

$$\mathbf{y}(t) = \mathbf{y}_{\mathrm{d}}(t) + \mathbf{y}_{\mathrm{s}}(t) \quad E\mathbf{y}_{\mathrm{s}}(t)\mathbf{y}_{\mathrm{d}}(\tau)^{\top} = 0 \quad \text{for all } t, \tau.$$

It is easy to see that, under absence of feedback, $\mathbf{y}_{\mathrm{d}}$ is actually a *causal* linear functional of the input process, i.e.

$$\mathbf{y}_{\mathrm{d}}(t) = E[\mathbf{y}(t) \,|\, \mathscr{U}_{t+1}^{-}], \quad t \in \mathbb{Z}$$

see Picci and Katayama (1996a), and is hence representable as the output of a causal linear time-invariant filter driven only by the input signal $\mathbf{u}$. Consequently, $\mathbf{y}_{\mathrm{s}}(t)$ is also the "causal estimation error" of $\mathbf{y}(t)$ based on the past and present inputs up to time t, i.e.

$$\mathbf{y}_{\mathrm{s}}(t) := \mathbf{y}(t) - E[\mathbf{y}(t) \,|\, \mathscr{U}_{t+1}^{-}]. \tag{2.7}$$

Since, under absence of feedback, $\mathbf{u}$ and $\mathbf{w}$ are uncorrelated, the deterministic and stochastic components of a process represented by a state-space realization of type (2.2) are represented by (generally non minimal) individual state-space realizations, obtained by setting $\mathbf{u} = 0$ and $\mathbf{w} = 0$ in (2.2), or, equivalently, $\mathbf{e} = 0$ in (2.3).

Let us now restrict to the innovation model (2.3), hereafter assumed to be minimal. Its input–output relation has the familiar form $\mathbf{y} = F(z)\mathbf{u} + G(z)\mathbf{e}$ with "stochastic" and "deterministic" transfer functions $F(z) = D + C(zI - A)^{-1}B$ and $G(z) = I + C(zI - A)^{-1}K$. Note that in these formulas the transfer functions are parametrized by the *same dynamic parameters* A, C and therefore need not be represented minimally. This is essentially the well-known ARMAX parametrization, most often considered in the identification literature. In most practical cases, however, unless there are known "physical" disturbances entering the system from the same input channels as $\mathbf{u}$, the stochastic and deterministic dynamics will be completely different. In this case $F(z), G(z)$ are parametrized by non-minimal realizations and, even if the "true" model (2.3) had cancellations leading to "true" subsystem transfer functions of individual degrees $n_{\mathrm{d}}, n_{\mathrm{s}}$, smaller than the overall dimension n, in the identified model the stochastic and deterministic transfer functions will invariably have the same dimension n of the joint system. In other words, fitting a model which allows for the same deterministic and stochastic dynamics, to generic data, will almost surely lead to identified transfer functions $\hat{F}(z), \hat{G}(z)$ both of maximum degree, with too many parameters, near pole-zero cancellations and higher variance of the estimates. For this reasons we may, in many situations, regard the ARMAX modelling philosophy as being unnatural.

On the other hand, (still assuming absence of feedback) the innovation representation of $\mathbf{y}$ can be obtained by combining in parallel a "deterministic" state-space model for $\mathbf{y}_{\mathrm{d}}$

$$\mathbf{x}_{\mathrm{d}}(t+1) = A_{\mathrm{d}}\mathbf{x}_{\mathrm{d}}(t) + B_{\mathrm{d}}\mathbf{u}(t), \tag{2.8a}$$

$$\mathbf{y}_{\mathrm{d}}(t) = C_{\mathrm{d}}\mathbf{x}_{\mathrm{d}}(t) + D\mathbf{u}(t), \tag{2.8b}$$

with the innovation representation of $\mathbf{y}_{\mathrm{s}}$

$$\mathbf{x}_{\mathrm{s}}(t+1) = A_{\mathrm{s}}\mathbf{x}_{\mathrm{s}}(t) + K_{\mathrm{s}}\mathbf{e}_{\mathrm{s}}(t), \tag{2.9a}$$

$$\mathbf{y}_{\mathrm{s}}(t) = C_{\mathrm{s}}\mathbf{x}_{\mathrm{s}}(t) + \mathbf{e}_{\mathrm{s}}(t), \tag{2.9b}$$

where $\mathbf{e}_{\mathrm{s}}(t)$ is the one-step prediction error of the stochastic component $\mathbf{y}_{\mathrm{s}}$ based on its own past, i.e. the innovation process of $\mathbf{y}_{\mathrm{s}}$. It is shown in Picci and Katayama (1996b) that $\mathbf{e}_{\mathrm{s}}$ actually coincides with the innovation process $\mathbf{e}$ of (2.3). Hence, the innovation model of the process $\mathbf{y}$ with inputs, can be described by a "canonical" block-diagonal realization

$$\begin{bmatrix} \mathbf{x}_{\mathrm{d}}(t+1) \\ \mathbf{x}_{\mathrm{s}}(t+1) \end{bmatrix} = \begin{bmatrix} A_{\mathrm{d}} & 0 \\ 0 & A_{\mathrm{s}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{\mathrm{d}}(t) \\ \mathbf{x}_{\mathrm{s}}(t) \end{bmatrix} + \begin{bmatrix} B_{\mathrm{d}} & 0 \\ 0 & K_{\mathrm{s}} \end{bmatrix} \begin{bmatrix} \mathbf{u}(t) \\ \mathbf{e}(t) \end{bmatrix},$$

$$\mathbf{y}(t) = [C_{\mathrm{d}} \; C_{\mathrm{s}}] \begin{bmatrix} \mathbf{x}_{\mathrm{d}}(t) \\ \mathbf{x}_{\mathrm{s}}(t) \end{bmatrix} + D\mathbf{u}(t) + \mathbf{e}(t), \tag{2.10}$$

where $\mathbf{x}_{\mathrm{d}}(t)$ and $\mathbf{x}_{\mathrm{s}}(t)$, are the *deterministic* and *stochastic* components of the state, mutually uncorrelated at all times. The deterministic and stochastic transfer functions of (2.10), can then be parametrized as $F(z) = D + C_{\mathrm{d}}(zI - A_{\mathrm{d}})^{-1}B_{\mathrm{d}}$ and $G(z) = I + C_{\mathrm{s}}(zI - A_{\mathrm{s}})^{-1}K_{\mathrm{s}}$. In this case the deterministic and stochastic models are parametrized independently and more parsimoniously than in the ARMAX model.[1] In fact, since $n_{\mathrm{d}}(m+p) + mp + 2n_{\mathrm{s}}m < m(n_{\mathrm{d}} + n_{\mathrm{s}}) + (n_{\mathrm{d}} + n_{\mathrm{s}})(p+m) + mp$, canonical forms for the decoupled model have always less free parameters than the jointly parametrized model.

In general it may happen that, even if the realizations of the stochastic and deterministic components of $\mathbf{y}$ are individually minimal, the joint model is not, as there may be a loss of observability due to the presence of common modes in the dynamics of the two subsystems. Hence in general a minimal realization takes the form

$$\begin{bmatrix} \check{\mathbf{x}}_{\mathrm{d}}(t+1) \\ \mathbf{x}_0(t+1) \\ \check{\mathbf{x}}_{\mathrm{s}}(t+1) \end{bmatrix} = \begin{bmatrix} \check{A}_{\mathrm{d}} & 0 & 0 \\ 0 & A_0 & 0 \\ 0 & 0 & \check{A}_{\mathrm{s}} \end{bmatrix} \begin{bmatrix} \check{\mathbf{x}}_{\mathrm{d}}(t) \\ \mathbf{x}_0(t) \\ \check{\mathbf{x}}_{\mathrm{s}}(t) \end{bmatrix} + \begin{bmatrix} \check{B}_{\mathrm{d}} & 0 \\ B_0 & K_0 \\ 0 & \check{K}_{\mathrm{s}} \end{bmatrix} \begin{bmatrix} \mathbf{u}(t) \\ \mathbf{e}(t) \end{bmatrix},$$

$$\mathbf{y}(t) = [\check{C}_{\mathrm{d}} \; C_0 \; \check{C}_{\mathrm{s}}] \begin{bmatrix} \check{\mathbf{x}}_{\mathrm{d}}(t) \\ \mathbf{x}_0(t) \\ \check{\mathbf{x}}_{\mathrm{s}}(t) \end{bmatrix} + D\mathbf{u}(t) + \mathbf{e}(t), \tag{2.11}$$

[1] The so-called *Box–Jenkins* models in PEM identification (Ljung, 1997), seem to serve a similar purpose.

where $\mathbf{x}_0$, of dimension n_0, describes the common dynamics, $\dim \check{\mathbf{x}}_s = n_s - n_0 := \check{n}_s$, and $\dim \check{\mathbf{x}}_d = n_d - n_0 := \check{n}_d$. This brings down the dimension of the model (2.10) from $n_d + n_s = \check{n}_d + 2n_0 + \check{n}_s$, to $\check{n}_d + n_0 + \check{n}_s$ for the minimal description (2.11). The analysis of subspace algorithms using minimal models of this general kind leads to notational complications and will not be taken up in this paper. Henceforth cases where there is common dynamics will be regarded as "very unlikely" and excluded from our analysis, on the basis of the following assumption.

Assumption 1. *The deterministic and stochastic subsystems* (2.9) *and* (2.8) *of the true model have no common dynamics, i.e. the sum of respective dimensions* $n_d + n_s$ *is equal to the dimension* n, *of the* (*minimal*) *joint model* (2.3).

Under Assumption 1 there is a non-singular change of basis bringing (2.3) into a decoupled form of the type (2.10).[2] Hence the deterministic transfer function of the system is (minimally) described by the upper left block of realization (2.10) of the "true" system. So, to estimate the deterministic transfer function $F(z)$, one could in principle estimate just the "deterministic" parameters (A_d, C_d, B_d, D). Dually, to estimate $G(z)$ one could in principle estimate just the "stochastic" parameters (A_s, C_s, K_s, Λ). This simple idea is the rationale of the orthogonal decomposition algorithm of Verhaegen and Dewilde (1993), Picci and Katayama (1996a), Chiuso and Picci (1999), Chiuso and Picci (2001).

Our main concern in this paper will be to compare the performance of subspace algorithms for the identification of decoupled models of type (2.10), with the standard subspace algorithms applied to the identification of the joint model (2.3) which is usually considered in the literature. We shall in fact only compare performances of the identification of the deterministic subsystem. Most of the results of this paper will actually hold *without any assumption of finite dimensionality of the stochastic component* $\mathbf{y}_s$, which could be just any stationary, purely non deterministic process.

Even if we shall not discuss the identification of the stochastic component, may we just mention that in case $\mathbf{y}_s$ is modeled by a finite-dimensional realization, and hence the data are to be described by a linear model of overall dimension n, the block structure of (2.10) gives a more parsimonious parametric description also of the stochastic component of the output process and we should expect better results also in the identification of $\mathbf{y}_s$. We shall however postpone this verification to another occasion.

[2] Just consider any choice of basis in the state space, coherent with the orthogonal direct sum decomposition $\mathbf{X} = \mathbf{X}_d \oplus \mathbf{X}_s$, where $\mathbf{X}_d$ and $\mathbf{X}_s$ are the reachable subspaces for $\mathbf{u}$ and $\mathbf{e}$ from $t = -\infty$.

3. Identification from data on a finite interval

Model (2.10) is a stationary model describing the stationary signals $(\mathbf{y}, \mathbf{u})$ on an infinite time interval. If the data were available from the infinite past, we could in principle compute projections (2.6) exactly and split the identification problem into two distinct subproblems with orthogonal data. In fact, from the two orthogonal components $\{\mathbf{y}_d, \mathbf{u}\}$, and $\mathbf{y}_s$, we could estimate separately the deterministic and the stochastic subsystem in (2.10). But infinite data of course never occur in practice and the problem of recovering in a statistically consistent way, the stationary model (2.10) from *finite* input–output data is not completely trivial. Also the advantages of this procedure in terms of accuracy of the estimates are not apparent and some analysis is needed.

The finite-data algorithm we consider is based on a preliminary decomposition of the finite (random) data into finite-interval "deterministic" and a "stochastic" components:

$$\hat{\mathbf{y}}_d(t) := E[\mathbf{y}(t) \mid \mathscr{U}_{[t_0,T]}], \quad \hat{\mathbf{y}}_s(t) := \mathbf{y}(t) - \hat{\mathbf{y}}_d(t).$$

the "hatted" variables being the best reconstruction of the stationary components $\mathbf{y}_d(t)$ and $\mathbf{y}_s(t)$, based on data available on the *finite* time interval $[t_0, T]$.

It is easy to see that the deterministic component $\hat{\mathbf{y}}_d(t)$ admits the finite-interval realization:

$$\begin{aligned} &\hat{\mathbf{x}}_d(t+1) = A_d \hat{\mathbf{x}}_d(t) + B_d \mathbf{u}(t), \\ &\hat{\mathbf{y}}_d(t) = C_d \hat{\mathbf{x}}_d(t) + D\mathbf{u}(t), \qquad t \in [t_0, T], \\ &\hat{\mathbf{x}}_d(t_0) := E[\mathbf{x}_d(t_0) \mid \mathscr{U}_{[t_0,T]}], \end{aligned} \tag{3.1}$$

which can be obtained, for example, by projecting the transient Kalman filter equations (2.16) of Chiuso and Picci (2004d) onto $\mathscr{U}_{[t_0,T]}$. The model is non-causal, due to the initial condition depending on the future input history. Of course, when $t_0 \to -\infty, A_d^{t-t_0}\mathbf{x}_d(t_0) \to 0$ and we recover the steady-state deterministic model which is instead causal. Let $\hat{\mathscr{X}}_t^d := \operatorname{span}\{\mathbf{x}_d(t)\}$ be the state space of the deterministic realization (3.1). A technical condition which will be needed in the following is the (deterministic) "consistency condition":

$$\hat{\mathscr{X}}_t^d \cap \mathscr{U}_{[t,T]} = \{0\}, \tag{3.2}$$

which is similar (although a bit stronger) than the consistency condition of Jansson and Wahlberg (1998). Compare formula (3.2) in Chiuso and Picci (2004d). Note that (3.2) holds trivially if $t_0 = -\infty$ since in this case $\hat{\mathscr{X}}_t^d \subset \mathscr{U}_t^-$ and the past and future input spaces intersect at zero (the "sufficient richness" condition (2.4)). However if t is close to t_0 it may not be satisfied; in particular it is certainly not satisfied for $t = t_0$.

We would like to interpret (3.1) as a regression equation with unknown parameters (A_d, B_d, C_d, D). However, in analogy to what happens in the case of a joint model, see the discussion in Chiuso and Picci (2004d), we find that $\hat{\mathbf{x}}_d(t)$ is not directly constructible from finite input–output data. Therefore (3.1) is, again, only an "ideal" regression model.

We may either resort to a "stationary" approximation assuming $t-t_0$ is large enough so that the effect of the initial state is negligible or, in analogy to what is done in N4SID-type methods, construct a (deterministic) pseudo-state which satisfies an exact (but more complicated) recursion involving the stationary parameters $A_\mathrm{d}, C_\mathrm{d}$. The stationary approximation leads to biased estimates and we shall follow this second route. Actually, we shall refer to a more reliable procedure which is essentially a weighted version of the PI-MOESP method of Verhaegen and Dewilde (1992, 1993). For up to date information on the influence of weighting matrices, the reader may consult (Bauer, Deistler, & Schetter, 2000).

The method is based on computing the orthogonal projection, $\hat{Y}^\mathrm{c}_{[t,T-1]}$, of the future output data $Y_{[t,T-1]}$ onto the "complementary" data space spanned by the rows of the matrix

$$U^\perp_{[t,T]} := U_{[t_0,t-1]} - E_N[U_{[t_0,t-1]}|U_{[t,T]}].$$

This subspace is called "complementary", since it is the orthogonal complement of $\mathscr{U}^N_{[t,T]}$ in the data space $\mathscr{U}^N_{[t_0,T]}$. The future data up to time $T-1$ (instead that up to time T) are used in order to keep the same future time horizon $v := T-t$ in the construction of the complementary state at time $t+1$.

Introduce a non-singular weighting matrix W_d and consider the (weighted) Singular Value Decomposition

$$W_\mathrm{d}\hat{Y}^\mathrm{c}_{[t,T-1]} = USV^\top = [\hat{U}\ \tilde{U}]\begin{bmatrix}\hat{S} & 0\\ 0 & \tilde{S}\end{bmatrix}\begin{bmatrix}\hat{V}^\top\\ \tilde{V}^\top\end{bmatrix}. \tag{3.3}$$

In this SVD we keep, say, the first n_1 "most significant" singular values. Here $\hat{U}$ is the matrix made with the first n_1 columns of U, $\hat{V}$ the first n_1 rows of V, and $\hat{S}$ the upper-left n_1 by n_1 corner of S. By neglecting the "small" singular values, we obtain an approximate rank factorization

$$\hat{Y}^\mathrm{c}_{[t,T-1]} \simeq \hat{U}\hat{S}\hat{V}^\top = \hat{\Gamma}_\mathrm{d}\hat{X}^\mathrm{c}_t, \tag{3.4}$$

where

$$\hat{\Gamma}_\mathrm{d} = W_\mathrm{d}^{-1}\hat{U}\hat{S}^{1/2},$$
$$\hat{X}^\mathrm{c}_t = \hat{S}^{1/2}\hat{V}^\top = \hat{S}^{-1/2}\hat{U}^\top W_\mathrm{d}\hat{Y}^\mathrm{c}_{[t,T-1]}, \tag{3.5}$$

Consider the tail matrix of the finite-interval deterministic component $\hat{\mathbf{y}}_\mathrm{d}(t)$, i.e. the linear regression, denoted $\hat{Y}_{[t,T-1]}$, of the future output data $Y_{[t,T-1]}$ on the whole input history $U_{[t_0,T]}$. It follows from (3.1) that

$$\begin{aligned}\hat{Y}_{[t,T-1]} &:= E_N[Y_{[t,T-1]}\,|\,U_{[t_0,T]}]\\ &= \Gamma_\mathrm{d}\hat{X}_t + H_\mathrm{d}U_{[t,T-1]} + E^\perp,\end{aligned} \tag{3.6}$$

where Γ_d is the observability matrix of the model (3.1), $\hat{X}_t$ is the $n_\mathrm{d}\times N$ tail matrix of the state $\hat{\mathbf{x}}_\mathrm{d}(t)$, and, letting $v := T-t$, H_d is the lower triangular block $mv\times mv$ Toeplitz matrix

$$H_\mathrm{d} := \begin{bmatrix} D & 0 & \dots & 0 & 0\\ C_\mathrm{d}B_\mathrm{d} & D & \dots & 0 & 0\\ \vdots & & & \ddots & \vdots\\ C_\mathrm{d}A_\mathrm{d}^{v-2}B_\mathrm{d} & C_\mathrm{d}A_\mathrm{d}^{v-3}B_\mathrm{d} & \dots & C_\mathrm{d}B_\mathrm{d} & D\end{bmatrix}, \tag{3.7}$$

of the deterministic subsystem $(A_\mathrm{d}, B_\mathrm{d}, C_\mathrm{d}, D)$. The last term, $E^\perp$, is a truncation error term which goes to zero as $N\to\infty$.

Since as $N\to\infty$, the last term in the expression

$$\begin{aligned}\hat{Y}^\mathrm{c}_{[t,T-1]} &= E_N[\hat{Y}_{[t,T-1]}\,|\,U^\perp_{[t,T]}] = \Gamma_\mathrm{d}E_N[\hat{X}_t\,|\,U^\perp_{[t,T]}]\\ &\quad + E_N[E^\perp\,|\,U^\perp_{[t,T]}]\end{aligned}$$

tends to zero, in force of the consistency condition (3.2), the rank of the projected matrix $E_N[\hat{X}_t\,|\,U^\perp_{[t,T]}]$ is equal to n_d (the true state dimension) for N large enough. Hence, the term $\Gamma_\mathrm{d}\hat{X}_t$ and the "complementary predictor" $\hat{Y}^\mathrm{c}_{[t,T-1]}$, will have, for N large enough, the same rank n_d, and for $N\to\infty$, the same column spaces. If the rank determination step in the factorization (3.4) is statistically consistent (i.e. asymptotically $n_1 = n_\mathrm{d}$) the approximate factorization in (3.4) will in the limit assume a special significance, made precise in the following Proposition.

Proposition 1. *Assume that the rank determination step in factorization (3.4) is statistically consistent (i.e. asymptotically $n_1 = n_\mathrm{d}$). Let $\mathscr{U}^\perp_{[t,T]}$ denote the orthogonal complement of $\mathscr{U}_{[t,T]}$ in $\mathscr{U}_{[t_0,T]}$. Then as $N\to\infty$, the tail matrix $\hat{X}^\mathrm{c}_t$ tends to the deterministic complementary state vector*

$$\hat{\mathbf{x}}^\mathrm{c}_\mathrm{d}(t) := \hat{\mathbf{x}}_\mathrm{d}(t) - E[\hat{\mathbf{x}}_\mathrm{d}(t)\,|\,\mathscr{U}_{[t,T]}] = E[\hat{\mathbf{x}}_\mathrm{d}(t)\,|\,\mathscr{U}^\perp_{[t,T]}], \tag{3.8}$$

which satisfies the state equation

$$\begin{aligned}\hat{\mathbf{x}}^\mathrm{c}_\mathrm{d}(t+1) &= A_\mathrm{d}\hat{\mathbf{x}}^\mathrm{c}_\mathrm{d}(t) + B_\mathrm{d}(t)\bar{\mathbf{v}}(t),\\ \hat{\mathbf{y}}^\mathrm{c}_\mathrm{d}(t) &= C_\mathrm{d}\hat{\mathbf{x}}^\mathrm{c}_\mathrm{d}(t),\end{aligned} \tag{3.9}$$

where $\hat{\mathbf{y}}^\mathrm{c}_\mathrm{d}(t) := \hat{\mathbf{y}}_\mathrm{d}(t) - E[\mathbf{y}(t)|\mathscr{U}_{[t,T]}]$, $\bar{\mathbf{v}}(t) = \mathbf{u}(t) - E[\mathbf{u}(t)\,|\,\mathscr{U}_{[t,T]}]$ is the backward innovation process of $\mathbf{u}(t)$, $B_\mathrm{d}(t) := A_\mathrm{d}K_\mathrm{d}(t) + B_\mathrm{d}$, with $K_\mathrm{d}(t) := E[\hat{\mathbf{x}}_\mathrm{d}(t)\bar{\mathbf{v}}(t)^\top]$ $([\bar{\mathbf{v}}(t)\bar{\mathbf{v}}(t)^\top])^{-1}$ and the initial condition is $\hat{\mathbf{x}}^\mathrm{c}_\mathrm{d}(t_0) = 0$.

Let $\Lambda_\mathrm{d} := E\{\hat{\mathbf{y}}^\mathrm{c}_{\mathrm{d}[t,T-1]}(\hat{\mathbf{y}}^\mathrm{c}_{\mathrm{d}[t,T-1]})^\top\}$ and consider the reduced (full rank) factorization,

$$W_\mathrm{d}\Lambda_\mathrm{d}W_\mathrm{d}^\top = US^2U^\top, \quad S = diag\{\sigma_1,\dots,\sigma_{n_\mathrm{d}}\}, \tag{3.10}$$

where $U^\top U = I$, and σ_{n_d} is the smallest non-zero eigenvalue of $W_\mathrm{d}\Lambda_\mathrm{d}W_\mathrm{d}^\top$. Then, for $N\to\infty$, $\hat{S}\xrightarrow{a.s.}S$ and

$$\hat{\mathbf{x}}^\mathrm{c}_\mathrm{d}(t) = S^{-1/2}U^\top W_\mathrm{d}\hat{\mathbf{y}}^\mathrm{c}_{\mathrm{d}[t,T-1]} = \Gamma_\mathrm{d}^{-L}\hat{\mathbf{y}}^\mathrm{c}_{\mathrm{d}[t,T-1]}, \tag{3.11}$$

where Γ_d is the observability matrix of the complementary realization (3.9) and $\Gamma_\mathrm{d}^{-L} := S^{-1/2}U^\top W_\mathrm{d}$.

Proof. The statement about convergence for $N\to\infty$ and expression (3.11), follow by the assumed consistency of the rank determination step and by second-order ergodicity

of the data. The recursion for $\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t)$ follows by projecting the finite-interval deterministic equation (3.1) onto $\mathscr{U}_{[t,T]}^{\perp}$. In particular note that $\mathscr{U}_{[t+1,T]}^{\perp} = \mathrm{span}\{\bar{\mathbf{v}}(t)\} \oplus \mathscr{U}_{[t,T]}^{\perp}$, see e.g. Lemma 3.2 in Chiuso and Picci (2004a), which yields directly the state equation. The output equation follows by computing $\hat{\mathbf{y}}_{\mathrm{d}}^{\mathrm{c}}(t) = E[\mathbf{y}(t) \,|\, \mathscr{U}_{[t,T]}^{\perp}] = C_{\mathrm{d}}\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t) + DE[\mathbf{u}(t) \,|\, \mathscr{U}_{[t,T]}^{\perp}] = C_{\mathrm{d}}\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t)$. Moreover

$$\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t_0) = E[\hat{\mathbf{x}}_{\mathrm{d}}(t_0) | \mathscr{U}_{[t_0,T]}^{\perp}] = 0$$

since $\hat{\mathbf{x}}_{\mathrm{d}}(t_0) \in \mathscr{U}_{[t_0,T]}$. □

By the consistency condition (3.2), the complementary state covariance matrix $E[\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t)(\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t))^{\top}]$ is non-singular, so the matrices A_{d} and C_{d} are uniquely determined by the complementary state by the formulas

$$A_{\mathrm{d}} = E[\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t+1)(\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t))^{\top}](E[\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t)(\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t))^{\top}])^{-1} \tag{3.12}$$

and

$$\begin{aligned} C_{\mathrm{d}} &= E[\hat{\mathbf{y}}_{\mathrm{d}}^{\mathrm{c}}(t)(\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t))^{\top}](E[\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t)(\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t))^{\top}])^{-1} \\ &= E[\mathbf{y}(t)(\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t))^{\top}](E[\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t)(\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t))^{\top}])^{-1}. \end{aligned} \tag{3.13}$$

From Proposition 1 it follows that the tail matrix $\hat{X}_t^{\mathrm{c}}$ defined in (3.5), is an estimate of the complementary state $\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t)$ which can be used to construct consistent sample-covariance estimates of A_{d} and C_{d} by an obvious sample version of formulas (3.12), (3.13).

Note that the covariance matrix of the complementary state $\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t)$ coincides with the *conditional* covariance of $\hat{\mathbf{x}}_{\mathrm{d}}(t)$ given the future inputs $\mathscr{U}_{[t,T]}$, namely

$$E[\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t)(\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t))^{\top}] = \Sigma_{\hat{\mathbf{x}}_{\mathrm{d}}\hat{\mathbf{x}}_{\mathrm{d}} \,|\, \mathbf{u}^+},$$

$$E[\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t+1)(\hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t))^{\top}] = \Sigma_{\hat{\mathbf{x}}_{\mathrm{d},1}\hat{\mathbf{x}}_{\mathrm{d}} \,|\, \mathbf{u}^+},$$

etc. so that the numerical conditioning of the computation of $(A_{\mathrm{d}}, C_{\mathrm{d}})$ by the weighted PI-MOESP procedure described above, is actually governed by the condition number of the *conditional covariance matrix* $\Sigma_{\hat{\mathbf{x}}_{\mathrm{d}}\hat{\mathbf{x}}_{\mathrm{d}}|\mathbf{u}^+}$, in line with the general analysis of Chiuso and Picci (2004d).

It is now natural to compare the conditioning of the computation of the (A, C) parameters of the decoupled model (2.8) with the conditioning of the joint model identification. Assume that we have chosen a basis in the state space which transforms (2.3) into a block-diagonal structure of type (2.10). We show in the appendix that in such a basis, when all modes of the deterministic system have died out, i.e. $A_{\mathrm{d}}^{t-t_0} \simeq 0$, the conditional covariance matrix of the joint state, $\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}$, becomes block-diagonal i.e.

$$\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+} \to \mathrm{diag}\{\Sigma_{\hat{\mathbf{x}}_{\mathrm{d}}\hat{\mathbf{x}}_{\mathrm{d}}|\mathbf{u}^+}, \Sigma_{\hat{\mathbf{x}}_{\mathrm{s}},\hat{\mathbf{x}}_{\mathrm{s}}}\}, \tag{3.14}$$

where $\Sigma_{\hat{\mathbf{x}}_{\mathrm{s}},\hat{\mathbf{x}}_{\mathrm{s}}} = E\{\hat{\mathbf{x}}_{\mathrm{s}}(t)\hat{\mathbf{x}}_{\mathrm{s}}(t)^{\top}\}$. In this case a comparison can be made rather easily. To simplify matters we shall henceforth make the assumption that $t - t_0$ is chosen large enough, so that the above block diagonal structure holds.

Proposition 2. *For $t - t_0$ large enough in the sense described above, the condition number $\kappa(\Sigma_{\hat{\mathbf{x}}_{\mathrm{d}}\hat{\mathbf{x}}_{\mathrm{d}}|\mathbf{u}^+})$ for the computation of $(A_{\mathrm{d}}, C_{\mathrm{d}})$ can be estimated as*

$$\kappa(\Sigma_{\hat{\mathbf{x}}_{\mathrm{d}}\hat{\mathbf{x}}_{\mathrm{d}}|\mathbf{u}^+}) \leqslant \kappa(\Sigma_{\hat{\mathbf{x}}_{\mathrm{d}}\hat{\mathbf{x}}_{\mathrm{d}}}) \frac{1 - \sigma_{\min}^2(\hat{\Pi}_{\mathrm{d}})}{1 - \sigma_{\max}^2(\hat{\Pi}_{\mathrm{d}})}, \tag{3.15}$$

$\hat{\Pi}_{\mathrm{d}}$ being the normalized cross-covariance

$$\hat{\Pi}_{\mathrm{d}} := L_{\mathbf{u}}^{-1}\Pi_{\mathrm{d}}L_{\mathbf{x}_{\mathrm{d}}}^{-\top},$$

where $\Pi_{\mathrm{d}} := E[\mathbf{u}_t^+\mathbf{x}_{\mathrm{d}}(t)^{\top}]$ and $L_{\mathbf{u}}, L_{\mathbf{x}_{\mathrm{d}}}$ are square roots of the covariance matrices of $\mathbf{u}_t^+$ and of $\mathbf{x}_{\mathrm{d}}(t)$. Bound (3.15) is sharp.

Formula (3.15) is just a particular case of formula (3.4) in Chiuso and Picci (2004d) so we shall refer the reader to that paper for a proof.

Since we have, $\lambda_{\max}(\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}) \geqslant \lambda_{\max}(\Sigma_{\hat{\mathbf{x}}_{\mathrm{d}}\hat{\mathbf{x}}_{\mathrm{d}}|\mathbf{u}^+})$ and $\lambda_{\min}(\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}) \leqslant \lambda_{\min}(\Sigma_{\hat{\mathbf{x}}_{\mathrm{d}}\hat{\mathbf{x}}_{\mathrm{d}}|\mathbf{u}^+})$ and all matrices in (3.14) are symmetric positive semidefinite, so that the singular values coincide with the eigenvalues, it trivially follows that

$$\kappa(\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}) \geqslant \kappa(\Sigma_{\hat{\mathbf{x}}_{\mathrm{d}}\hat{\mathbf{x}}_{\mathrm{d}}|\mathbf{u}^+})$$

which says that the estimation of $(A_{\mathrm{d}}, C_{\mathrm{d}})$ is better conditioned than that of the joint parameters (A, C) (relative to the block diagonalizing basis). In other words, joint estimation is less reliable than separate estimation of the two blocks. This is true especially when the eigenvalues of $\Sigma_{\hat{\mathbf{x}}_{\mathrm{s}},\hat{\mathbf{x}}_{\mathrm{s}}}$ differ by orders of magnitude from those of $\Sigma_{\hat{\mathbf{x}}_{\mathrm{d}}\hat{\mathbf{x}}_{\mathrm{d}}|\mathbf{u}^+}$.

An important aspect which is not captured by the elementary conditioning analysis reported in this section, is the influence of the errors in the "off-diagonal" blocks when a non block-diagonal joint model is used. This will be taken up in the statistical analysis of the companion paper (Chiuso & Picci, 2004b).

We warn the reader that the analysis here is only relative to estimation of the (A, C) parameters. More noticeable differences emerge when comparing the estimates of the (B, D) parameters. This comparison will be also discussed in Chiuso and Picci (2004b).

4. Probing inputs for identification

As shown in Chiuso and Picci (2004d), the conditioning of joint subspace identification methods depends on the canonical correlations between the state space $\hat{\mathscr{X}}_t^{+/-}$, spanned by the state $\hat{\mathbf{x}}(t)$, of the (transient) Kalman filter model (see Eq. (2.16) in the above reference), and the finite future space $\mathscr{U}_{[t,T]}$. In this section, we shall investigate a class of inputs which asymptotically leads to the largest canonical correlation coefficients between these two subspaces. This will indicate a situation where the worst conditioning of the (joint) subspace identification problem occurs, at least for estimating the (A, C) parameters of the model.

To simplify the canonical correlation analysis we shall assume that $t - t_0$ is large, so that the conditional covariance matrix of the joint state, $\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}$, is block-diagonal, see

(3.14), and that $T-t$ is also large, so that the principal angles (and principal directions) between $\hat{\mathscr{X}}_t^{+/-}$ and $\mathscr{U}_{[t,T]}$ are approximately the same as those between $\hat{\mathscr{X}}_t^{+/-}$ and $\mathscr{U}_t^+$ (the infinite future). Under these assumptions we can actually work with stationary models, so hereafter we shall fix the present time to $t=0$ and suppress the time subscript from all symbols.

Let

$$\mathscr{X}_{\mathrm{d}}^{+/-} = E_{\|\mathscr{U}^+}[\mathscr{Y}_{\mathrm{d}}^+ \mid \mathscr{U}^-] \subset \mathscr{U}^-$$

be the n_{d}-dimensional oblique predictor space of the deterministic component $\mathbf{y}_{\mathrm{d}}$ of the process $\mathbf{y}$. Let also $\mathscr{X}_{\mathrm{s}}^{+/-} = \mathrm{span}\{\mathbf{x}_{\mathrm{s}}(0)\}$ be the state space of the "stochastic" component (in innovation form). Under the absence of feedback assumption, the two subspaces $\mathscr{X}_{\mathrm{s}}^{+/-}$ and $\mathscr{X}_{\mathrm{d}}^{+/-}$ are orthogonal and still in force of Assumption 1, the state space $\mathscr{X}^{+/-}$ can be decomposed in orthogonal direct sum as [3]

$$\mathscr{X}^{+/-} = \mathscr{X}_{\mathrm{d}}^{+/-} \oplus \mathscr{X}_{\mathrm{s}}^{+/-}.$$

Since by absence of feedback $\mathscr{X}_{\mathrm{s}}^{+/-}$ is orthogonal to the whole input history $\mathscr{U}$, *the* (*non-zero*) *canonical correlation coefficents between* $\mathscr{X}^{+/-}$ *and* $\mathscr{U}^+$ *are the same as those between* $\mathscr{X}_{\mathrm{d}}^{+/-}$ *and* $\mathscr{U}^+$. In particular for $t-t_0$ and $T-t$ large, the normalized cross covariance $\hat{\Pi} := L_{\mathbf{u}}^{-1}\Pi L_{\mathbf{x}}^{-\top}$ of the future inputs and joint state, where $\Pi := E[\mathbf{u}_t^+\mathbf{x}(t)^\top]$, will be

$$\hat{\Pi} = [\hat{\Pi}_{\mathrm{d}}\ 0]$$

so that in particular the maximal singular values of the two matrices will be the same. For this reason, in the following it will be enough to consider only the deterministic subsystem.

Denote by $\sigma_k(\mathscr{X}_{\mathrm{d}}^{+/-}, \mathscr{U}^+)$ the kth canonical correlation coefficient of $\mathscr{X}_{\mathrm{d}}^{+/-}$ and $\mathscr{U}^+$ and by $\sigma_k(\mathscr{U}^-, \mathscr{U}^+)$ the kth canonical correlation coefficient of $\mathscr{U}^-$ and $\mathscr{U}^+$, the canonical correlation coefficients being ordered in decreasing magnitude. Let $\mathscr{X}_u^{+/-} := E[\mathscr{U}^+ | \mathscr{U}^-]$ be the forward predictor space (i.e. the state space of the innovation model) of the process $\mathbf{u}$. It is well-known that the canonical variables [4] of $\mathscr{U}^-$ for the pair of subspaces $(\mathscr{U}^-, \mathscr{U}^+)$, belong to $\mathscr{X}_u^{+/-}$ (Akaike, 1975; Desai, Pal, & Kikpatrick, 1985; Lindquist & Picci, 1996).

For concreteness in what follows we shall assume that the spectral density of $\mathbf{u}$ is rational of MacMillan degree $2r$. The following lemma, whose proof is immediate, will be instrumental in the analysis.

Lemma 3. *Let* $r \geqslant n_{\mathrm{d}}$. *The following inequalities hold*:

$$\sigma_k(\mathscr{X}_{\mathrm{d}}^{+/-}, \mathscr{U}^+) \leqslant \sigma_k(\mathscr{U}^-, \mathscr{U}^+), \quad k = 1, 2, \ldots . \tag{4.1}$$

Moreover

$$\sigma_k(\mathscr{X}_{\mathrm{d}}^{+/-}, \mathscr{U}^+) = \sigma_k(\mathscr{U}^-, \mathscr{U}^+) \quad k = 1, 2, \ldots, n_{\mathrm{d}} \tag{4.2}$$

if and only if the first n_{d} *canonical variables of* $\mathscr{U}^-$ *for the pair* $(\mathscr{U}^-, \mathscr{U}^+)$, *belong to* $\mathscr{X}_{\mathrm{d}}^{+/-}$ (*and hence span* $\mathscr{X}_{\mathrm{d}}^{+/-}$).

Let instead $r < n_{\mathrm{d}}$. *Then inequalities* (4.1) *hold for* $k = 1, 2, \ldots, r$. *Moreover*

$$\sigma_k(\mathscr{X}_{\mathrm{d}}^{+/-}, \mathscr{U}^+) = \sigma_k(\mathscr{U}^-, \mathscr{U}^+) \quad k = 1, 2, \ldots, r \tag{4.3}$$

if and only if the first r canonical variables of $\mathscr{X}_{\mathrm{d}}^{+/-}$ *for the pair* $(\mathscr{X}_{\mathrm{d}}^{+/-}, \mathscr{U}^+)$, *belong to* $\mathscr{X}_u^{+/-}$ (and hence span $\mathscr{X}_u^{+/-}$).

Consider the case $r \geqslant n_{\mathrm{d}}$. Let $\mathbf{x}_1$ be a subvector of the state vector $\mathbf{x}$ spanning the predictor space $\mathscr{X}_u^{+/-}$. Without loss of generality we may assume that $\mathbf{x}$ is decomposed as

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix},$$

where $\mathrm{span}\{\mathbf{x}_1\} := \mathscr{X} \subset \mathscr{X}_u^{+/-}$ and $\mathbf{x}_2 \perp \mathscr{X}$.

Let

$$\begin{bmatrix} \mathbf{x}_1(t+1) \\ \mathbf{x}_2(t+1) \end{bmatrix} = A_u \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} + K_u \mathbf{e}_u(t), \tag{4.4a}$$

$$\mathbf{u}(t) = C_u \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} + \mathbf{e}_u(t) \tag{4.4b}$$

be the corresponding minimal realization of $\mathbf{u}$ with state space $\mathscr{X}_u^{+/-}$. Expressing the innovation in function of $\mathbf{u}$ and substituting in the state equation we obtain

$$\begin{bmatrix} \mathbf{x}_1(t+1) \\ \mathbf{x}_2(t+1) \end{bmatrix} = \begin{bmatrix} (A_u - K_u C_u)_{11} & (A_u - K_u C_u)_{12} \\ (A_u - K_u C_u)_{21} & (A_u - K_u C_u)_{22} \end{bmatrix} \times \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} + \begin{bmatrix} K_{u1} \\ K_{u2} \end{bmatrix} \mathbf{u}(t). \tag{4.5}$$

Now, for the subvector $\mathbf{x}_1$ to qualify also as a state variable evolving in $\mathscr{X}$ (which then becomes an *oblique* Markovian splitting subspace), it must hold that $(A_u - K_u C_u)_{12} = 0$. If this property holds, it clearly holds (modulo change of basis) for any subvector of the type $\hat{\mathbf{x}}_1 = T\mathbf{x}_1$ with T a non singular matrix, and hence is a property of the subspace $\mathscr{X}$. In this case we shall call $\mathscr{X}$ an *invariant subspace of* $\mathscr{X}_u^{+/-}$.

It is well known (Akaike, 1975; Desai et al., 1985; Lindquist & Picci, 1996) that we can pick a basis $\mathbf{x}$ in $\mathscr{X}_u^{+/-}$ made of random variables which are proportional to the principal directions of $\mathscr{U}^-$ for the pair $(\mathscr{U}^-, \mathscr{U}^+)$. In particular, we may pick a basis of *ordered* principal directions. A basis of this kind (with proper weights) leads to the so-called *stochastically balanced form* of the corresponding realization. If $\mathscr{X}$ is an invariant subspace of $\mathscr{X}_u^{+/-}$ spanned by the first n_1 principal components of $\mathscr{U}^-$ for the

[3] Note that in general, i.e. if Assumption 1 is not satisfied, only $\mathscr{X}^{+/-} \subseteq \mathscr{X}_{\mathrm{d}}^{+/-} \oplus \mathscr{X}_{\mathrm{s}}^{+/-}$ holds true.

[4] Also called *principal directions*.

pair $(\mathscr{U}^-, \mathscr{U}^+)$, we shall say that $\mathscr{X}$ is a *principal invariant subspace of* $\mathscr{X}_u^{+/-}$. In this case the eigenvalues of the upper left diagonal block $\lambda\{(A_u - K_uC_u)_{11}\}$, will be called the first n_1 *principal zeros* of system (4.4). Principal zeros, like principal eigenvalues to be introduced later, remain invariant under principal truncation, i.e. extraction of the subsystem with state vector $\mathbf{x}_1$, defined by the upper-left block entries in (4.4) (Lindquist & Picci, 1996).

The following result then follows readily from the statement of Lemma 3 and provides a geometric solution to our problem.

Proposition 4. *Let $r \geqslant n_\mathrm{d}$. The maximal canonical correlation coefficients (smallest canonical angles) between $\mathscr{X}_\mathrm{d}^{+/-}$ and $\mathscr{U}^+$ are obtained when, and only when, $\mathscr{X}_\mathrm{d}^{+/-}$ is a principal invariant subspace of $\mathscr{X}_u^{+/-}$.*

In the following theorem we shall give conditions on the input process and on the input spectrum to insure that the deterministic state space $\mathscr{X}_\mathrm{d}^{+/-}$ is a principal invariant subspace of $\mathscr{X}_u^{+/-}$.

Theorem 5. *Given an input process of rational spectral density matrix Φ_u of degree $2r$, $r \geqslant n_\mathrm{d}$, the maximal canonical correlation coefficients $\sigma_k(\mathscr{X}_\mathrm{d}^{+/-}, \mathscr{U}^+)$ are obtained when, and only when, there are n_d principal zeros of the (forward) innovation realization of* $\mathbf{u}$ *cancelling all poles of the deterministic transfer function $F(z)$ of the system. Equivalently, the spectral density matrix Φ_u has n_d stable principal zeros which cancel all the poles of $F(z)$.*

Proof. Let (4.4) be a stochastically balanced innovation representation of $\mathbf{u}$ and let $(A_\mathrm{d}, B_\mathrm{d}, C_\mathrm{d}, D)$ be a minimal realization of the deterministic subsystem (with state space $\mathscr{X}_\mathrm{d}^{+/-}$). As we have just seen, $\mathscr{X}_u^{+/-}$ admits a principal invariant subspace if and only if the matrix $A_u - K_uC_u$, has the block structure

$$(A_u - K_uC_u) = \begin{bmatrix} (A_u - K_uC_u)_{11} & 0 \\ (A_u - K_uC_u)_{21} & (A_u - K_uC_u)_{22} \end{bmatrix},$$

where $K_u = [K_{u1}^\top \; K_{u2}^\top]^\top$. Moreover $\mathscr{X}_\mathrm{d}^{+/-}$ is spanned by the first n_d canonical vectors of $\mathscr{X}_u^{+/-}$ if and only if $((A_u - K_uC_u)_{11}, K_{u1})$ is similar to the pair $(A_\mathrm{d}, B_\mathrm{d})$. In other words, there exists a non-singular $T \in \mathbb{R}^{n_\mathrm{d} \times n_\mathrm{d}}$ such that $A_\mathrm{d} = T(A_u - K_uC_u)_{11}T^{-1}$ and $B_\mathrm{d} = TK_{u1}$. In particular, $\mathscr{X}_u^{+/-}$ and $\mathscr{X}_\mathrm{d}^{+/-}$ coincide if and only if $(A_u - K_uC_u, K_u)$ is similar to the pair $(A_\mathrm{d}, B_\mathrm{d})$.

The equivalence of the statement of Proposition 4 with the cancellation of the zero dynamics of the innovation realization (4.4) and the dynamics of the deterministic subsystem can be seen from the state space description of the cascade of (4.4) with the deterministic realization $(A_\mathrm{d}, B_\mathrm{d}, C_\mathrm{d}, D)$, namely

$$\begin{bmatrix} \mathbf{x}_\mathrm{d}(t+1) \\ \mathbf{x}_1(t+1) \\ \mathbf{x}_2(t+1) \end{bmatrix} = \begin{bmatrix} A_\mathrm{d} & B_\mathrm{d}C_{u1} & B_\mathrm{d}C_{u2} \\ 0 & (A_u)_{11} & (A_u)_{12} \\ 0 & (A_u)_{21} & (A_u)_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_\mathrm{d}(t) \\ \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} + \begin{bmatrix} B_\mathrm{d} \\ K_{u1} \\ K_{u2} \end{bmatrix} \mathbf{e}_u(t)$$

$$\mathbf{y}_\mathrm{d}(t) = C_\mathrm{d}\mathbf{x}_\mathrm{d}(t) + D_\mathrm{d}C_u\mathbf{x}(t) + D\mathbf{e}_u(t)$$

from which, subtracting the second state component from the first, and recalling from the previous paragraph that, $\mathscr{X}_\mathrm{d}^{+/-}$ is a principal invariant subspace of $\mathscr{X}_u^{+/-}$ if and only if we can substitute $(A_\mathrm{d}, B_\mathrm{d})$ with $((A_u - K_uC_u)_{11}, K_{u1})$, it follows that

$$\mathbf{x}_\mathrm{d}(t+1) - \mathbf{x}_1(t+1) = (A_u - K_uC_u)_{11}(\mathbf{x}_\mathrm{d}(t) - \mathbf{x}_1(t))$$

so that $\mathbf{x}_\mathrm{d}(t) - \mathbf{x}_1(t) = 0$ for all t, by asymptotic stability of $(A_u - K_uC_u)_{11}$. Hence a minimal basis in the state space of the cascade realization is $\mathbf{x}(t)$ and the dynamics of the overall system reduces to

$$\begin{bmatrix} \mathbf{x}_1(t+1) \\ \mathbf{x}_2(t+1) \end{bmatrix} = \begin{bmatrix} (A_u)_{11} & (A_u)_{12} \\ (A_u)_{21} & (A_u)_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} + \begin{bmatrix} K_{u1} \\ K_{u2} \end{bmatrix} \mathbf{e}_u(t)$$

$$\mathbf{y}_\mathrm{d}(t) = (C_\mathrm{d} + D_\mathrm{d}C_{u1})\mathbf{x}_1(t) + DC_{u2}\mathbf{x}_2(t) + D\mathbf{e}_u(t)$$

whose only eigenvalues are those of the innovation realization (4.4). The dynamics of the deterministic system has been cancelled completely. □

Remark 6. If the (deterministic) system is given and we are to design the spectrum of the "probing" input to get maximum ill-conditioning, it is enough to choose a spectral density of degree $2n_\mathrm{d}$ so that the innovation model of $\mathbf{u}$ has dimension n_d and all of its zeros are (trivially) principal. In addition, we should choose the zero dynamics of the innovation model, i.e. of Φ_u, so as to cancel the dynamics of the deterministic system.

There is then freedom to place the poles of Φ_u. These poles determine the "excitation properties" of the input process (in fact, the conditioning of the Toeplitz matrix $\Sigma_{\mathbf{u}^+\mathbf{u}^+}$). It is possible to show (but we shall not do that here) that, by placing the poles of Φ_u arbitrarily close to the unit circle, one can obtain canonical correlation coefficients $\sigma_k(\mathscr{U}^-, \mathscr{U}^+)$, arbitrarily close to one. Hence we can make $\sigma_k(\mathscr{U}^-, \mathscr{U}^+)$, arbitrarily close to one by choosing the poles of the spectrum and make the $\sigma_k(\mathscr{X}_\mathrm{d}^{+/-}, \mathscr{U}^+)$'s equal to their maximum values $\sigma_k(\mathscr{U}^-, \mathscr{U}^+)$, by choosing the zeros of the spectrum.

We shall now take a quick look to the case $r < n_d$. Let $\mathbf{x}_{d1}$ be a subvector of the state vector $\mathbf{x}_d$, spanning the predictor space $\mathscr{X}_u^{+/-}$. Without loss of generality we may assume that $\mathbf{x}_d$ is decomposed as

$$\mathbf{x}_d = \begin{bmatrix} \mathbf{x}_{d1} \\ \mathbf{x}_{d2} \end{bmatrix},$$

where $\operatorname{span}\{\mathbf{x}_{d1}\} := \mathscr{X}_u^{+/-} \subset \mathscr{X}_d^{+/-}$ and $\mathbf{x}_{d2} \perp \mathscr{X}_u^{+/-}$.

Since $\mathbf{x}_{d1}$ is a state in the predictor space $\mathscr{X}_u^{+/-}$, we can write

$$\begin{bmatrix} \mathbf{x}_{d1}(t+1) \\ \mathbf{x}_{d2}(t+1) \end{bmatrix} = A_d \begin{bmatrix} \mathbf{x}_{d1}(t) \\ \mathbf{x}_{d2}(t) \end{bmatrix} + B_d \mathbf{u}(t), \tag{4.6a}$$

$$\mathbf{u}(t) = H_u \mathbf{x}_{d1}(t) + \mathbf{e}_u(t) \tag{4.6b}$$

for some matrix H_u. Expressing $\mathbf{u}$ in function of the innovation in the state equation we obtain

$$\begin{bmatrix} \mathbf{x}_{d1}(t+1) \\ \mathbf{x}_{d2}(t+1) \end{bmatrix} = \begin{bmatrix} (A_d)_{11} + B_{d1}H_u & (A_d)_{12} \\ (A_d)_{21} + B_{d2}H_u & (A_d)_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{d1}(t) \\ \mathbf{x}_{d2}(t) \end{bmatrix} + \begin{bmatrix} B_{d1} \\ B_{d2} \end{bmatrix} \mathbf{e}_u(t). \tag{4.7}$$

Now, for the subvector $\mathbf{x}_{d1}$ to qualify as a state variable evolving in $\mathscr{X}_u^{+/-}$ (which has to be so since $\mathscr{X}_u^{+/-}$ is a Markovian splitting subspace), it must hold that $(A_d)_{12} = 0$. This property clearly holds (modulo change of basis) for any subvector of the type $\hat{\mathbf{x}}_{d1} = T\mathbf{x}_{d1}$ with T a non-singular matrix, and hence is a property of the subspace spanned by $\mathbf{x}_{d1}$. Any subspace of this kind will be called an *invariant subspace of* $\mathscr{X}_d^{+/-}$. This condition is clearly equivalent to $((A_d)_{11} + B_{d1}H_u, B_{d1}, H_u, I)$ being a minimal realization of the innovation model of $\mathbf{u}$. In other words $\mathscr{X}_u^{+/-}$ is an invariant subspace of $\mathscr{X}_d^{+/-}$ iff there exists a non-singular $T \in \mathbb{R}^{r\times r}$ such that for any minimal realization (C_u, A_u, K_u, I) of the innovation model of $\mathbf{u}$, it holds that $A_u = T((A_d)_{11} + B_{d1}H_u)T^{-1}$ and $K_u = TB_{d1}$. But this is the same as

$$T(A_d)_{11}T^{-1} = A_u - K_uC_u,\ B_{d1} = T^{-1}K_u,\ C_u := H_uT^{-1}.$$

Again, we can (and shall) pick a basis $\mathbf{x}_d$ in $\mathscr{X}_d^{+/-}$ made of *ordered* principal directions for the pair $(\mathscr{X}_d^{+/-}, \mathscr{U}^+)$. If $\mathscr{X}_u^{+/-}$ is an invariant subspace of $\mathscr{X}_d^{+/-}$, then by (4.1) of Lemma 3, it is necessarily spanned by the first r principal components of $\mathscr{X}_d^{+/-}$ and hence it is automatically a *principal invariant subspace of* $\mathscr{X}_d^{+/-}$. In this case the eigenvalues of the upper left diagonal block, $\lambda\{(A_d)_{11}\}$, are the first r *principal eigenvalues* of the deterministic system.

Proposition 7. *Let $r < n_d$. The first r canonical correlation coefficients between $\mathscr{X}_d^{+/-}$ and $\mathscr{U}^+$ are maximal when, and only when, $\mathscr{X}_u^{+/-}$ is an invariant subspace of $\mathscr{X}_d^{+/-}$.*

The following is the analog of Theorem 5 which applies in the situation $r \leqslant n_d$.

Theorem 8. *With an input process of rational spectral density matrix Φ_u of degree $2r$, $r \leqslant n_d$, the first r canonical correlation coefficients $\sigma_k(\mathscr{X}_d^{+/-}, \mathscr{U}^+)$ are maximized when, and only when, the deterministic subsystem of transfer function $F(z)$, admits r eigenvalues*[5] *which are all cancelled by the (stable) zeros of Φ_u.*

It may be worth to remark that "bad" inputs are not necessarily a simple thing as inputs cancelling the poles of the system to be identified. In the next section we shall see some examples which should help to clarify this point.

5. Some experimental results

We shall discuss the results of 100 Monte-Carlo runs made on a simple scalar system described in Table 1 below. The model is in the jointly parametrized form and the identification algorithm is a "robustified" version of the N4SID algorithm, described in Overschee and De Moor (1996). The sample size is $N = 500$.

The simulations are performed with three different input spectra denoted Φ_1, Φ_2, Φ_3. See Fig. 1.

(1) The (stable) zeros of Φ_1 cancel exactly the poles of the deterministic system. However Φ_1 is nearly constant with frequency and $\mathbf{u}_1$ is a nearly white process. In this case the singular values $\sigma_k(\mathscr{X}_d^{+/-}, \mathscr{U}^+) = \sigma_k(\mathscr{U}^-, \mathscr{U}^+)$ are all nearly the same and rather small (since past and future of a white noise form angles of 90°). In particular $\sigma^2_{\max}(\hat{\Pi}_d)$ is small, so that the problem, in spite of the cancellation, is well-conditioned. Input 1 gives the best estimates even if the poles of the deterministic transfer function are cancelled exactly.

(2) The zeros of Φ_2 are far apart from the system poles but the input process is ill-conditioned. In this example $\kappa(\Sigma_{\mathbf{u}^+\mathbf{u}^+}) \simeq 10^5$. Note that the estimate of the system poles does not worsen much with respect to input 1. However the identification of the transfer function is rather poor. This may be attributed to the following fact: input 2 is ill conditioned (in the sense that the covariance matrix $\Sigma_{\mathbf{u}^+\mathbf{u}^+}$ is ill conditioned), and this makes the estimation of (B, D) unreliable; on the other hand, as there is no cancellation in the zero structure, the conditional state covariance $\Sigma_{\mathbf{x}_d\mathbf{x}_d|\mathbf{u}^+}$ is not ill-conditioned and therefore the estimation of (A, C) is reliable.

(3) The (stable) zeros of Φ_3 cancel exactly the system poles and in addition the input process is ill-conditioned. In this example $\kappa(\Sigma_{\mathbf{u}^+\mathbf{u}^+})$ is nearly the same as in example 2. The identification is very poor. Note that in this case also the estimates of the eigenvalues (and of (A, C))

[5] These eigenvalues are then necessarily principal.

Table 1
Poles, zeros and gains for the stochastic, deterministic subsystems and for the three inputs considered

	Poles	Zeros	K
Stoch. system	$-0.1 + j0.6$	0.5	1
	$-0.1 - j0.6$	0.7	
Det. system	$0.75 + j0.55$	$-0.1 + j0.8$	0.2
	$0.75 - j0.55$	$-0.1 - j0.8$	
	0.9	0.5	
Input 1	$0.8 + j0.55$	$0.75 + j0.55$	0.7185
	$0.8 - j0.55$	$0.75 - j0.55$	
Input 2	$-0.815 + j0.5$	$0.35 + j0.95$	1.8495
	$-0.815 - j0.5$	$0.35 - j0.95$	
Input 3	$-0.8 + j0.5$	$0.75 + j0.55$	2.2760
	$-0.8 - j0.5$	$0.75 - j0.55$	

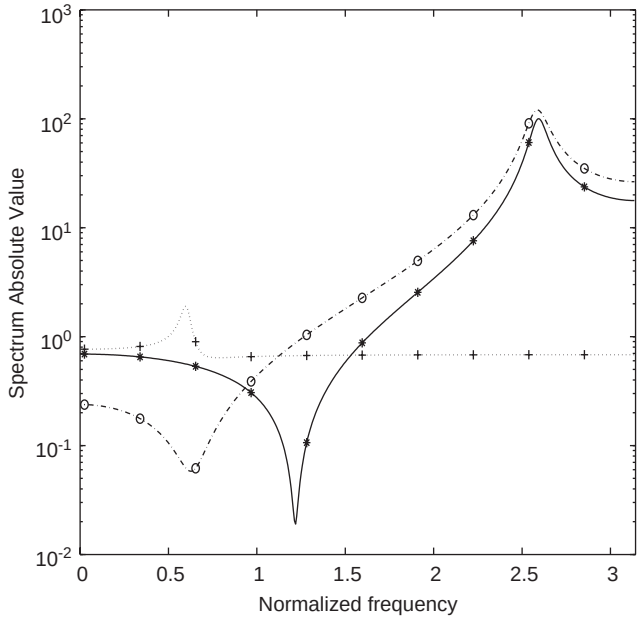


Fig. 1. Input spectrum. Crosses (+): input 1, stars (∗): input 2, circles (o): input 3.

become rather erratic. This fact agrees with the prediction of Theorem 8 that under these conditions $\Sigma_{\mathbf{x}_d\mathbf{x}_d|\mathbf{u}^+}$ is ill conditioned.

In all three experiments, the power (variance) of the deterministic component of the output signal and of the stochastic disturbance are about the same so as to keep the same SNR ratio ($SNR := \sigma^2_{\mathbf{y}_d}/\sigma^2_{\mathbf{y}_s} = 2$).

Comments: The standard deviation of the (deterministic) transfer function estimates corresponding to the three different input processes is shown in Fig. 2. Going from input 1 to 2 to 3, an increase of the standard deviation, roughly of one order of magnitude (in the frequency band of interest), is observed (Fig. 3). The estimated poles (Fig. 4) confirm that the estimates worsen in the order $1 \to 2 \to 3$.

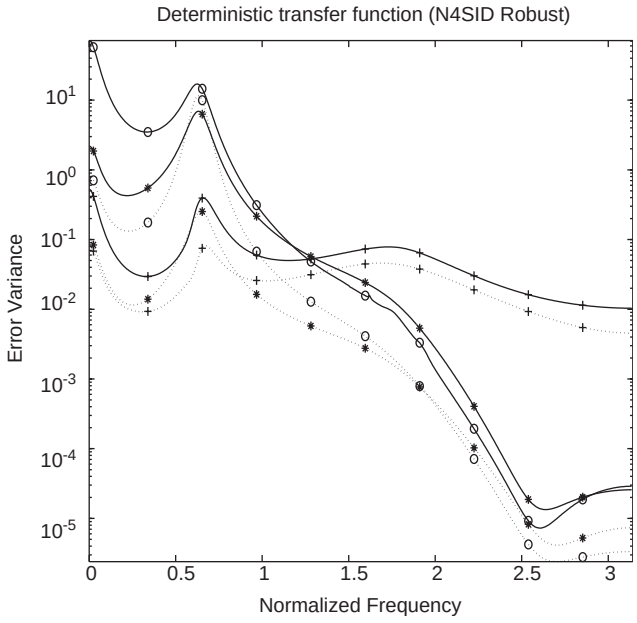


Fig. 2. Joint modeling estimates: Standard deviation of the transfer function estimates (solid) and Cramèr–Rao lower bound (dotted) vs. frequency. Crosses (+): input 1, stars (∗): input 2, circles (o): input 3.

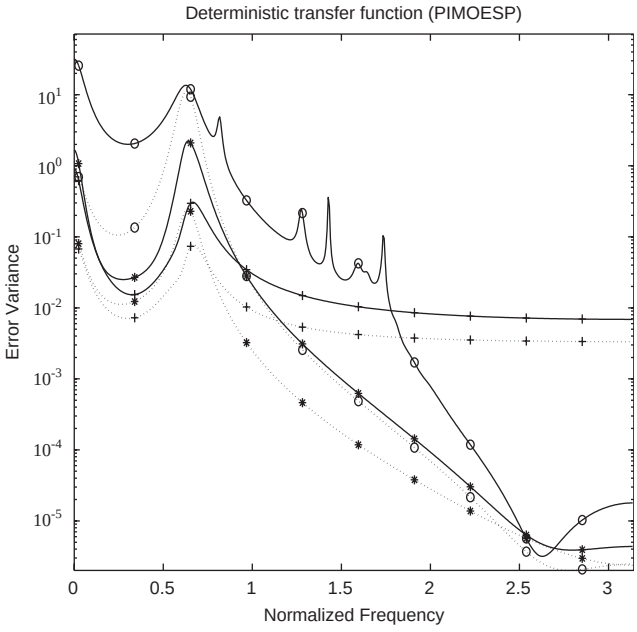


Fig. 3. Orthogonal decomposition (PI-MOESP) estimates: Standard deviation of the transfer function estimates (solid) and Cramèr–Rao lower bound (dotted) vs. frequency. Crosses (+): input 1, stars (∗): input 2, circles (o): input 3.

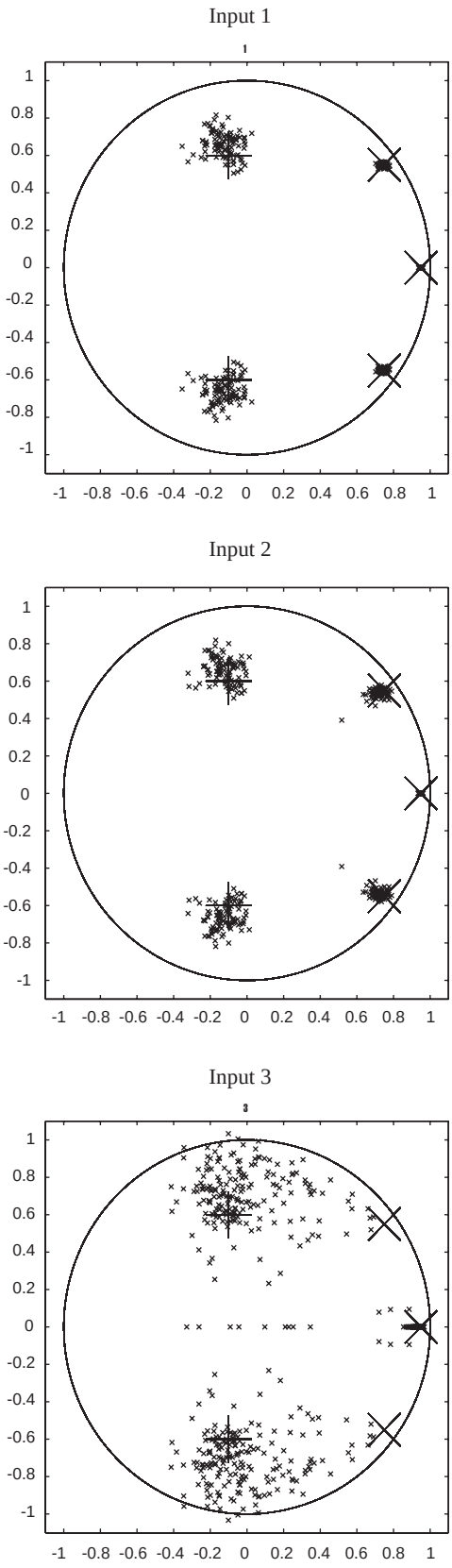


Fig. 4. Estimated poles (N4SID robust).

5.1. Comparison with decoupled identification

The deterministic system poles estimated with the orthogonal decomposition method are shown in Fig. 5. As expected, the accuracy of these estimates is roughly the same as those obtained with the joint parametrization. In particular, the conditioning of the estimation of A and C does not depend on $\Sigma_{\mathbf{u}^+\mathbf{u}^+}$ and hence on the conditioning of the input process. The asymptotic variance formulas which will be derived in the companion paper (Chiuso & Picci, 2004b) confirm these qualitative conclusions.

The accuracy of transfer function estimation however depends on the parametrization. In Figs. 6, and 7, we compare the results of subspace identification of the (deterministic) transfer function a simple scalar system using the joint and the "orthogonal decomposition" methods. The experiment consists of 100 Monte-Carlo runs of a robust N4SID method (joint model) and of the orthogonal decomposition-based algorithm (PI-MOESP), with $t - t_0 = T - t = 10$. The system and input spectra (frequency-domain data) are the same considered in the paper (Chiuso & Picci, 2004c). The input process is a colored ARMA process. The deterministic and stochastic components have completely disjoint dynamics. In (Chiuso & Picci, 2004c), Table 1 reports the details of the simulations and Fig. 1 shows the system and input spectra (frequency-domain data).

It is evident that the orthogonal decomposition-based algorithm performs better. Here, the simulations show that the performance of jointly parametrized methods is much worse than that of decoupled model-methods especially *in the frequency range of the stochastic noise spectrum*, where the off-diagonal error terms, $\tilde{A}_{\mathrm{sd}}, \tilde{B}_{\mathrm{s}}$, (see the companion papers (Chiuso & Picci, 2004b for details) affecting the joint-parametrization methods have a large influence on the transfer function estimate. This is essentially what is predicted by Eqs. (3.30) and (3.32) of Chiuso and Picci (2004b).

The last simulation shows that the difference in performance is quite evident even for white inputs.

6. Conclusions

In this paper we have presented a comparison between two classes of subspace identification methods. The first is based on an orthogonal decomposition of the input-output data, combined with a block-decoupled parametrization of the model while the other methods are subspace methods based on the usual "joint model parametrization". An elementary error analysis has been provided and a comparison of the numerical conditioning of the identification problem has been made. In particular, we have shown that for inputs whose power spectral density has zeros cancelling the poles of the system to be identified, the ill-conditioning may be worsened at will. Expressions for the asymptotic error covariances of the (A, B, C, D) parameters and also of the

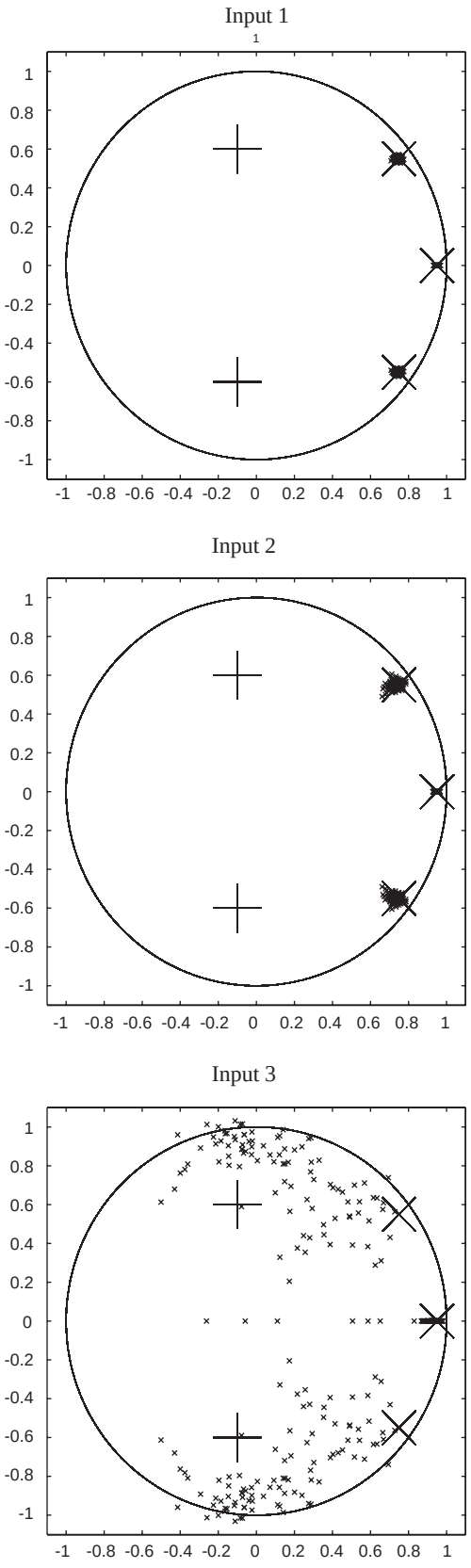


Fig. 5. Estimated poles (PI-MOESP), only deterministic poles shown.

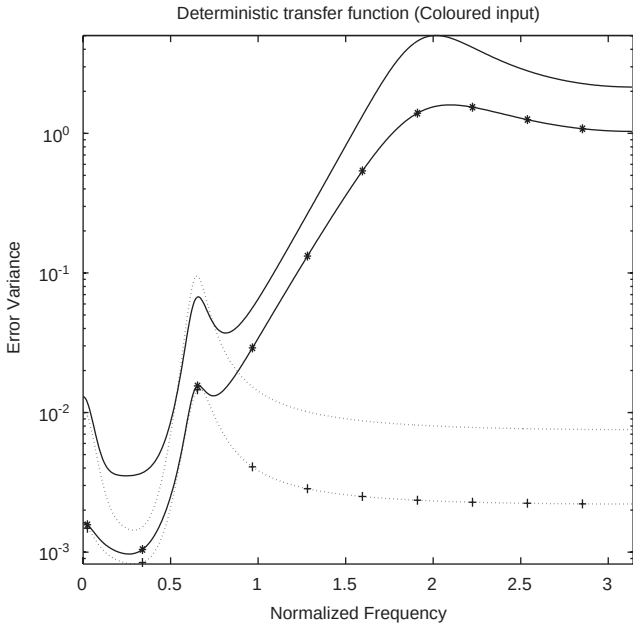


Fig. 6. Error variance (Monte-Carlo estimate) of estimated transfer function for coloured input. Dotted line (…): orthogonal decomposition algorithm (PI-MOESP); dotted line with crosses (....+): Cramer–Rao Bound for block parametrized models. Solid line (—): jointly parametrized Robust N4SID; solid line with stars (—∗): Cramer–Rao bound for joint parametrized models.

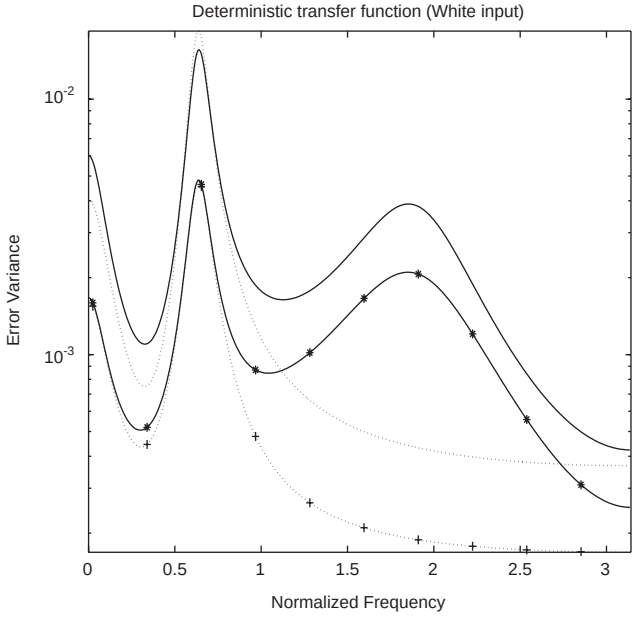


Fig. 7. Error variance (Monte-Carlo estimate) of estimated transfer function for white noise input. Dotted line (....): orthogonal decomposition algorithm (PI-MOESP); Dotted line with crosses (…+): Cramer –Rao Bound for block parametrized models. Solid line (—): jointly parametrized Robust N4SID; solid line with stars (—∗): Cramer–Rao bound for jointly parametrized models.

transfer function estimates, will be given, for both classes of subspace methods, in the companion paper (Chiuso & Picci, 2004b). In certain cases, the formulas show that the

errors for the joint parametrization are larger than for the decoupled parametrization. When models with decoupled deterministic–stochastic dynamics are adequate, there seems to be enough evidence to conclude that subspace estimation based on a block-diagonal model parametrization provides in most cases, better estimates of the system transfer function than the standard "joint" input–output methods.

Appendix

Convergence of conditional state covariances

Consider a "decoupled" basis as in formula (2.10)

$$\mathbf{x}(t) = \begin{bmatrix} \mathbf{x}_{\mathrm{d}}(t) \\ \mathbf{x}_{\mathrm{s}}(t) \end{bmatrix}$$

and let $\hat{\mathbf{x}}(t)$ be the transient Kalman state, defined as in Theorem 1 of Chiuso and Picci (2004d), i.e. $\hat{\mathbf{x}}(t) := E[\mathbf{x}(t) \,|\, \mathscr{P}_{[t_0,t)} \vee \mathscr{U}_{[t,T]}]$ so that $\hat{\mathbf{x}}^{\mathrm{c}}(t) := \hat{\mathbf{x}}(t) - E[\hat{\mathbf{x}}(t)|\mathscr{U}_{[t,T]}]$. Let us denote by $\hat{\hat{\mathbf{x}}}_{\mathrm{d}}(t) := E[\mathbf{x}_{\mathrm{d}}(t) \,|\, \mathscr{P}_{[t_0,t)} \vee \mathscr{U}_{[t,T]}]$ and by $\hat{\hat{\mathbf{x}}}_{\mathrm{s}}(t) := E[\mathbf{x}_{\mathrm{s}}(t)|\mathscr{P}_{[t_0,t)} \vee \mathscr{U}_{[t,T]}]$ the two subvectors of the Kalman filter state $\hat{\mathbf{x}}(t)$. Note that for finite $t - t_0$ in general $\hat{\mathbf{x}}_{\mathrm{d}}(t) \neq \hat{\hat{\mathbf{x}}}_{\mathrm{d}}(t)$. In fact, using the decomposition $\mathbf{x}_{\mathrm{d}}(t) = A_{\mathrm{d}}^{t-t_0}\mathbf{x}_{\mathrm{d}}(t_0) + $"*terms in* $\{\mathscr{U}_{[t_0,t)}\}$" the two projections satisfy

$$\begin{aligned}\mathbf{x}_{\mathrm{d}}(t) - \hat{\hat{\mathbf{x}}}_{\mathrm{d}}(t) = A_{\mathrm{d}}^{t-t_0}(&E[\mathbf{x}_{\mathrm{d}}(t_0)|\mathscr{U}_{[t_0,t)} \vee \mathscr{U}_{[t,T]}] \\ &- E[\mathbf{x}_{\mathrm{d}}(t_0)|\mathscr{P}_{[t_0,t)} \vee \mathscr{U}_{[t,T]}]).\end{aligned}$$

Similarly, if we define $\hat{\hat{\mathbf{x}}}_{\mathrm{d}}^{\mathrm{c}}(t) := \hat{\hat{\mathbf{x}}}_{\mathrm{d}}(t) - E[\hat{\hat{\mathbf{x}}}_{\mathrm{d}}(t)|\mathscr{U}_{[t,T]}]$, we have

$$\hat{\hat{\mathbf{x}}}_{\mathrm{d}}^{\mathrm{c}}(t) - \hat{\mathbf{x}}_{\mathrm{d}}^{\mathrm{c}}(t) = A_{\mathrm{d}}^{t-t_0}(E[\mathbf{x}_{\mathrm{d}}(t_0)|\mathscr{U}_{[t,T]}^{\perp}] - E[\mathbf{x}_{\mathrm{d}}(t_0)|\tilde{\mathscr{P}}_{[t_0,t)}]),$$

where $\tilde{\mathscr{P}}_{[t_0,t)}$ is the orthogonal complement of $\mathscr{U}_{[t,T]}$ in $\mathscr{P}_{[t_0,t)} \vee \mathscr{U}_{[t,T]}$. This last equality guarantees that the covariance matrix of $\hat{\hat{\mathbf{x}}}_{\mathrm{d}}^{\mathrm{c}}(t)$, i.e. the upper left corner of $\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}$ differs from $\Sigma_{\hat{\mathbf{x}}_{\mathrm{d}}\hat{\mathbf{x}}_{\mathrm{d}}|\mathbf{u}^+}$ for terms of the order of $A_{\mathrm{d}}^{t-t_0}$.

Using a similar argument we can also show that $E[\hat{\hat{\mathbf{x}}}_{\mathrm{s}}^{\mathrm{c}}(t)(\hat{\hat{\mathbf{x}}}_{\mathrm{d}}^{\mathrm{c}}(t))^{\top}]$ tends to zero at least as $A_{\mathrm{d}}^{t-t_0}$ for $t - t_0 \to \infty$. In fact

$$\begin{aligned}E[\hat{\hat{\mathbf{x}}}_{\mathrm{d}}^{\mathrm{c}}(t)(\hat{\hat{\mathbf{x}}}_{\mathrm{s}}^{\mathrm{c}}(t))^{\top}] &= E[\hat{\hat{\mathbf{x}}}_{\mathrm{d}}^{\mathrm{c}}(t)\mathbf{x}_{\mathrm{s}}^{\top}(t)] \\ &= A_{\mathrm{d}}^{t-t_0}E[\hat{\hat{\mathbf{x}}}_{\mathrm{d}}^{\mathrm{c}}(t_0)\mathbf{x}_{\mathrm{s}}^{\top}(t)], \qquad (A.1)\end{aligned}$$

where the last equality follows from the fact that the terms which live in $\mathscr{U}_{[t,T]}$ are orthogonal to the stochastic state.

References

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

Bauer, D., Deistler, M., & Scherrer, W. (2000). On the impact of weighting matrices in subspace algorithms. In *Proceedings of IFAC international symposium on system identification*. Santa Barbara.

Caines, P. E., & Chan, C. W. (1976). Estimation, identification and feedback. In R. Mehra, & D. Lainiotis (Eds.), *System identification: Advances and case studies* (pp. 349–405). New York: Academic Press.

Chiuso, A., & Picci, G. (1999). Subspace identification by orthogonal decomposition. In *Proceedings of the 14th IFAC World Congress*, Vol. I. (pp. 241–246).

Chiuso, A., & Picci, G. (2000). Probing inputs for subspace identification. In *Proceedings of the 2000 conference on decision and control*. Sydney, Australia. pp. paper CDC00–INV0201.

Chiuso, A., & Picci, G. (2001). Some algorithmic aspects of subspace identification with inputs. *Applied Mathematics and Computer Sciences*, *11*(1), 55–76.

Chiuso, A., & Picci, G. (2004a). The asymptotic variance of subspace estimates. *Journal of Econometrics*, *118*(1–2), 257–291.

Chiuso, A., & Picci, G. (2004b). Asymptotic variance of subspace methods by data orthogonalization and model decoupling: A comparative analysis. *Automatica*, this issue, doi:10.1016/j.automatica.2004.05.009.

Chiuso, A., & Picci, G. (2004c). Numerical conditioning and asymptotic variance of subspace estimates. *Automatica*, *40*(4), 677–683.

Chiuso, A., & Picci, G. (2004d). On the ill-conditioning of subspace identification with inputs. *Automatica*, *40*(4), 575–589.

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

Gevers, M. R., & Anderson, B. D. O. (1981). Representation of jointly stationary feedback free processes. *International Journal of Control*, *33*, 777–809.

Hannan, E. J., & Poskitt, D. S. (1988). Unit canonical correlations between future and past. *The Annals of Statistics*, *16*, 784–790.

Jansson, M., & Wahlberg, B. (1998). On consistency of subspace methods for system identification. *Automatica*, *34*, 1507–1519.

Kawauchi, H., Chiuso, A., Katayama, T., & Picci, G. (1999). Comparison of two subspace identification methods for combined deterministic -stochastic systems. In *Proceedings of the 31st ISCIE international symposium on stochastic systems theory and its applications*. Yokohama, Japan.

Lindquist, A., & Picci, G. (1996). Canonical correlation analysis, approximate covariance extension and identification of stationary time series. *Automatica*, *32*, 709–733.

Ljung, L. (1997). *System identification; theory for the user*. Englewood Cliffs, NJ: Prentice-Hall.

Overschee, P. Van., & De Moor, B. (1996). *Subspace identification for linear systems*. Dordrecht: Kluwer Academic Publications.

Picci, G., & Katayama, T. (1996a). Stochastic realization with exogenous inputs and "subspace methods" identification. *Signal Processing*, *52*, 145–160.

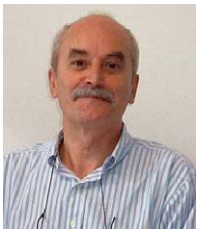
Picci, G., & Katayama, T. (1996b). Stochastic realization with exogenous inputs and "subspace methods" identification. *Signal Processing*, *52*, 145–160.

Verhaegen, M., & Dewilde, P. (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. *International Journal of Control*, *56*, 1187–1210 & 1211–1241.

Verhaegen, M., & Dewilde, P. (1993). Subspace model identification, part 3. Analysis of the ordinary output-error state-space model identification algorithm. *International Journal of Control*, *58*, 555–586.

Alessandro Chiuso received his D.Ing. degree Cum Laude in 1996, and the Ph.D. degree in Systems Engineering in 2000 both from the University of Padova. In 1998/99 he was a Visiting Research Scholar with the Electronic Signal and Systems Research Laboratory (ESSRL) at Washington University, St. Louis. From March 2000 to July 2000 he has been Visiting Post-Doctoral (EU-TMR) fellow with the Division of Optimization and System Theory, Department of Mathematics, KTH, Stockholm, Sweden.

Since March 2001 he is Research Faculty ("Ricercatore") with the Department of Information Engineering, University of Padova. In the summer 2001 he has been visiting researcher with the Department of Computer Science, University of California, Los Angeles. His research interests are mainly in Identification and Estimation Theory, System Theory and Computer Vision.



Giorgio Picci holds a full professorship with the University of Padova, Italy, Department of Information Engineering, since 1980. He graduated (cum laude) from the University of Padova in 1967 and since then has held several long-term visiting appointments with various American and European universities among which Brown University, M.I.T., the University of Kentucky, Arizona State University, the Center for Mathematics and Computer Sciences (C.W.I.) in Amsterdam, the Royal Institute of Technology, Stockholm Sweden, Kyoto University and Washington University in St. Louis, Mo.

He has been contributing to Systems and Control theory mostly in the area of modeling, estimation and identification of stochastic systems and published over 100 papers and edited three books in this area. Since 1992 he has been active also in the field of Dynamic Vision and scene and motion reconstruction from monocular vision.

He has been involved in various joint research projects with industry and state agencies. He is currently general coordinator of the italian national project *New techniques for identification and adaptive control of industrial systems*, funded by MIUR (the Italian ministery for higher education), has been project manager of the italian team for the Commission of the European Communities Network of Excellence *System Identification* (ERNSI) and is currently general project manager of the Commission of European Communities IST project RECSYS, in the fifth Framework Program.

Giorgio Picci is a Fellow of the IEEE, past chairman of the IFAC Technical Committee on Stochastic Systems and a member of the EUCA council.