Available online at www.sciencedirect.com
SCIENCE DIRECT®
Automatica 41 (2005) 377–391
automatica
www.elsevier.com/locate/automatica

Consistency analysis of some closed-loop subspace identification methods☆

Alessandro Chiuso, Giorgio Picci*

Department of Information Engineering, University of Padova, 35131 Padova, Italy

Received 1 February 2004; received in revised form 18 June 2004; accepted 26 October 2004
Available online 23 December 2004

Abstract

We study statistical consistency of two recently proposed subspace identification algorithms for closed-loop systems. These algorithms may be seen as implementations of an abstract state-space construction procedure described by the authors in previous work on stochastic realization of closed-loop systems. A detailed error analysis is undertaken which shows that both algorithms are biased due to an unavoidable mishandling of initial conditions which occurs in closed-loop identification. Instability of the open loop system may also be a cause of trouble.

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

1. Introduction

It seems fair to say that current state-of-the-art subspace identification methods provide reliable results only when applied to plants operating in open loop. However, feedback is present in a variety of practical situations (even though often one cannot directly recognize physical controllers which "close the loop") and there is a need of reliable identification methods and algorithms which could be used with multivariable systems in the presence of feedback. Various attempts to extend existing subspace identification algorithms to work in the presence of feedback have been made in the last decades. Among early references, we quote Van der Klauw, Verhaegen, and Van den Bosch (1991), Verhaegen (1993), Ljung and McKelvey (1996), Chou and Verhaegen (1999), and Van Overschee and De Moor (1997), while more recent work is presented in Qin and Ljung (2003), Jansson (2003). Yet, as discussed in Chiuso and Picci (2002), Chiuso and Picci (2003) there are fundamental issues related to stochastic realization theory in the presence of feedback which remain unclear. In particular stochastic realization with feedback is still not fully understood when unstable open-loop transfer functions are involved, which of course is a very interesting situation in the applications. On the other hand, even some of the best recently proposed methods seem to run occasionally into troubles with unstable open-loop transfer functions.

Unfortunately, even when restricting to stable open loop plants (a rather stringent restriction to be sure), the existing algorithms turn out to provide biased estimates. We shall argue that this is so mainly because one has to neglect the effect of initial conditions. This is in contrast to the classical open-loop subspace methods (N4SID, MOESP, CCA) from the literature, which instead provide consistent estimates by taking properly into account the "transient" effects due to initial conditions. Of course the usual way to reduce the bias due to neglecting initial conditions is to regress on enough past data (i.e. to keep the past data horizon of the algorithm suitably large). Provided the zeros of the system are not too

☆ This work has been supported in part by the *RECSYS* project of the European Community and 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-2003) in Rotterdam. This paper was recommended for publication in revised form by Guest Editors Torsten Söderström, Paul van der Hof, Bo Wahlberg and Siep Weiland.

* Corresponding author.
E-mail addresses: chiuso@dei.unipd.it (A. Chiuso), picci@dei.unipd.it (G. Picci).

0005-1098/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2004.10.015

close to the unit circle, the bias error can generally be made negligible. However, as we shall argue in this paper, when feedback is present, there may be situations in which, even when the past horizon of the algorithm is chosen very large, the bias is still unacceptable.

In this paper we shall analyze the bias in subspace identification with feedback. We shall in particular point out that mishandling of the initial conditions is an intrinsic difficulty related to feedback, which pops up whenever we restrict to identification of the system in the forward loop only. The difficulty is visible from the structure of the (transient) Kalman filter-like representation of the output process, which in general does involve also the dynamics of the input process $\mathbf{u}$ which we *don't want to model*. This phenomenon can be circumvented in the feedback-free case, see Chiuso and Picci (2004b), but seems to be very hard to bypass when there is feedback. We believe that a correct handling of the initial condition should be a main step towards a satisfactory theory of subspace identification with feedback.

This paper has been partly inspired and motivated by strong connections existing between the recent subspace algorithms of Qin and Ljung (2003) and Jansson (2003), and some theoretical work that we have been carrying through in the last years, dealing with stochastic state-space construction in the presence of feedback. This work is preliminarily exposed in the papers Chiuso and Picci (2002, 2003). We shall demonstrate that these new algorithms, which we regard as a significant step forward in subspace identification of feedback systems, can be interpreted as possible numerical implementations of some stochastic realization constructions described in Chiuso and Picci (2002, 2003). Consistency analysis of the algorithms, and an explicit computation of the bias are then possible by using the framework of stochastic systems and stochastic realization theory.

The structure of the paper is as follows:

Section 2 states the problem and sets up basic notations; Section 3 recalls the basic notions of stochastic realization with inputs. An alternative procedure to construct the state space and to do subspace identification is also proposed.

In Section 4 the problem of modeling $\mathbf{y}$ (given $\mathbf{u}$) on a finite interval by Kalman filter like representations is discussed. It is shown that transient "open-loop" models are in general of the same dimension as a minimal model for the joint process $[\mathbf{y}^\top\ \mathbf{u}^\top]^\top$. In general these models are not parameterized by the stationary parameters (i.e. the "A, B, K" matrices will be time varying, see (4.8)).

Section 5 discusses the finite data formulation of the ideal state construction introduced in Section 3 implemented in the algorithm by Qin and Ljung; also the effect of bias is analyzed.

Section 6 does the same for the algorithm based on the "whitening filter". In Section 7 the effect of (open loop) unstable plants is studied. We show that the unstable dynamics may amplify errors due to approximation of expectations with finite sums, making one of the analyzed procedures quite prone to errors.

In Section 8 some simulations are presented and Section 9 draws some conclusions.

2. Statement of the problem

Let $\{\mathbf{y}(t)\}, \{\mathbf{u}(t)\}$ be jointly (weakly) stationary zero-mean second-order ergodic[1] stochastic processes of dimension m and p, respectively, which are representable as the output and input signals of a linear stochastic system in innovation form

$$\begin{aligned}\mathbf{x}(t+1) &= A\mathbf{x}(t) + B\mathbf{u}(t) + K\mathbf{e}(t),\\ \mathbf{y}(t) &= C\mathbf{x}(t) + D\mathbf{u}(t) + \mathbf{e}(t),\end{aligned} \qquad t \geqslant t_0. \tag{2.1}$$

Without loss of generality we shall assume that the dimension n of the state vector $\mathbf{x}(t)$ is as small as possible, i.e. representation (2.1) is minimal.

In general there may be feedback from $\{\mathbf{y}(t)\}$ to $\{\mathbf{u}(t)\}$ (Granger, 1963; Caines & Chan, 1976; Gevers & Anderson, 1982; Gevers & Anderson, 1981). We shall assume that $D = 0$, i.e. there is no direct feedthrough from $\mathbf{u}$ to $\mathbf{y}$. Under this assumption the feedback (if any) could be quite general. In Section 4, to carry on the analysis, we shall restrict to linear time-invariant finite-dimensional feedback channels; however this assumption is not needed for the derivation of the algorithms.

Assuming D equal to zero serves to guarantee well-posedness and identifiability of $F(z) = C(zI - A)^{-1}B + D$ with an arbitrary feedback interconnection. Obviously this is the same as imposing that the transfer function $F(z)$, of the forward loop is strictly causal, i.e. $F(\infty) = 0$. In case of a linear feedback channel this is equivalent to a certain block-triangular normalization at infinity of the spectral factor of the joint spectrum, which can always be assumed to hold. We refer the reader to the papers Caines and Chan (1976), Gevers and Anderson (1982, 1981) for a thorough discussion of this and related questions.

The white noise process $\mathbf{e}$, the innovation of $\mathbf{y}$ given the joint past of $\mathbf{y}$, $\mathbf{u}$, is defined as the one step ahead prediction error of $\mathbf{y}(t)$ given the joint (strict) past of $\mathbf{u}$ and $\mathbf{y}$ up to time t. We shall not make any assumption on the correlation structure of $\mathbf{u}$ and $\mathbf{e}$; it is well known that $\mathbf{e}(t)$ is uncorrelated with the whole history [2] of $\mathbf{u}$ if and only if there is no feedback from $\mathbf{y}$ to $\mathbf{u}$.

We shall analyze the asymptotic behavior of certain subspace algorithms which estimate a parametrization (A, B, C) of the open-loop transfer function $F(z)$, starting from input–output sample data $\{y_s, u_s\}$, $s \in [t_0, T + N]$,

[1] Second-order ergodicity guarantees that sample second-order moments converge (a.s.) to the true variances/covariances as the sample size tends to infinity. Various sufficient conditions for this property to hold are found in the literature. It can for example be guaranteed by certain "low level" assumptions, say on the innovation process of the rational model considered in the paper. Conditions adapted to the setting we consider in this paper (but which we shall not report here) can be found in the book of Hannan and Deistler (1988), see Theorem 4.1.1.

[2] I.e. for $t \in (-\infty, \infty)$.

generated by system (2.1), without a priori excluding the presence of a possible feedback channel from **y** to **u**. No model for the feedback channel is sought. This is in contrast with the so-called joint input–output identification, where a full model of the joint input–output process is estimated. Joint input–output identification is well-established but needs an assumption of linearity (and time-invariance) of the feedback channel and requires to work with more complex (joint) models. Also, obtaining the parameters (A, B, C) of the direct transfer function from the joint identified model requires in general to solve a nontrivial model reduction problem.

Consistency is a well-known key asymptotic property of estimators (i.e. estimation algorithms), according to which the "true model parameters" [3] should be obtained as a limit of the parameter estimates when the data sequence becomes "infinitely long". In this paper, rather than working with notationally cumbersome finite Hankel matrices and then taking limits for $N \to \infty$ as is often done in the subspace identification literature, we shall work entirely in a stochastic setting. For the subspace estimators of linear state-space models are generally expressible as certain simple functions of the sample cross covariances of the state and of the input/output processes and under the assumed second-order ergodicity the transition between the two settings essentially amounts to formally identify infinitely long sequences of observed data with random variables.[4] This permits to disregard random fluctuations due to finite sample length altogether (e.g. the error in approximating expectations with finite time averages, etc.).

Still, in order to deal with realistic algorithms which can only regress on a finite amount of data, we shall keep *finite past and future horizons* (the "*i*" parameter of Overschee & De Moor (1994) or the *p* and *f* parameters in most subsequent subspace literature) in subspace identification. This setting we shall describe as using data from a *finite observation interval* later on. In fact, in this paper finite (and generally fixed) past and future horizons will hold even when the sample size N is let going to ∞ for the purpose of asymptotic analysis. Because of this intrinsic limitation, the effect of initial conditions has to be taken into account and will play an important role as we shall see.

[3] Of course defined modulo coordinate changes.

[4] This can be made precise by taking advantage of a natural isomorphism which can be established between the (L^2 space of) random variables of a second-order ergodic process and a certain Hilbert space of semi-infinite sequences linearly generated by a (semi-infinite) observed sample path of the same process, which is described e.g. in Lindquist and Picci(1996a,b). Using this isomorphism, taking limits for $N \to \infty$ of quadratic functions of a finite sample of the observed processes $\mathbf{y}, \mathbf{u}$, coincides with taking expectation of the products of random quantities which correspond to the "abstract" random quantities which are being "sampled" in the measurement experiment.

2.1. Notations

In this paper t_0 and T will denote fixed initial and terminal times; for $-\infty \leqslant t_0 \leqslant t \leqslant T \leqslant +\infty$ we define the Hilbert spaces of scalar zero-mean 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. These are the *past spaces* at time t of the processes **u** and **y**. Similarly, let $\mathscr{U}_{[t,T]}, \mathscr{Y}_{[t,T]}$ be the future input and output spaces up to time T

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

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

The *joint* future, $\mathscr{Z}_{[t,T]}$ and joint past $\mathscr{Z}_{[t_0,t)}$ spaces are defined as $\mathscr{U}_{[t,T]} \vee \mathscr{Y}_{[t,T]}$ and $\mathscr{U}_{[t_0,t)} \vee \mathscr{Y}_{[t_0,t)}$ respectively, the $\vee$ denoting closed vector sum. 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 $\mathscr{U}_{[-\infty,t)}$, $\mathscr{Y}_{[-\infty,t)}$, and $\mathscr{Z}_t^- := \mathscr{U}_t^- \vee \mathscr{Y}_t^-$. 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 **u** and **y** we shall use the symbols $\mathscr{U}$, $\mathscr{Y}$, respectively. In particular, the ambient Hilbert space for all future computations will be the (closed) vector sum $\mathscr{Z} := \mathscr{U} \vee \mathscr{Y}$.

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

$$\mathscr{U}_{[t_0,T]} = \mathscr{U}_{[t_0,t)} + \mathscr{U}_{[t,T]}, \quad t_0 \leqslant t < T \tag{2.2}$$

the $+$ sign denoting direct sum of subspaces. The symbol $\oplus$ will be reserved for *orthogonal* direct sum. Various conditions ensuring sufficient richness are known. For example, it is well known that for a full-rank purely nondeterministic (p.n.d.) process **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).

Given two zero mean random vectors **a** and **b** we shall use the notation $\Sigma_{\mathbf{ab}} := E\{\mathbf{a}\mathbf{b}^\top\}$ to denote the covariance matrix. The symbol $E[\zeta \mid \mathscr{X}]$ denotes the vector of orthogonal projections (conditional expectations in the Gaussian case) of the components of $\zeta \in \mathscr{Z}$ onto the subspace $\mathscr{X}$. In particular, if the subspace has a basis given by the components of a vector **x**, we have the well-known representation

$$E[\zeta \mid \mathscr{X}] := \Sigma_{\zeta\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x}.$$

Let the subspaces $\mathscr{A}$ and $\mathscr{B}$ of $\mathscr{Z}$ be in direct sum, i.e. $\mathscr{A} \cap \mathscr{B} = \{0\}$, then the orthogonal projection of any element

$\zeta \in \mathscr{Z}$ onto the direct sum $\mathscr{A}+\mathscr{B}$ can be written uniquely as a sum of elements of $\mathscr{A}$ and $\mathscr{B}$, namely

$$E\{\zeta \mid \mathscr{A}+\mathscr{B}\} = E_{\|\mathscr{A}}\{\zeta \mid \mathscr{B}\} + E_{\|\mathscr{B}}\{\zeta \mid \mathscr{A}\},$$

where $E_{\|\mathscr{A}}\{\zeta \mid \mathscr{B}\}$ is called the *oblique projection of* ζ *onto* $\mathscr{B}$ *along* $\mathscr{A}$ and $E_{\|\mathscr{B}}\{\zeta \mid \mathscr{A}\}$ is called the *oblique projection of* ζ *onto* $\mathscr{A}$ *along* $\mathscr{B}$. If the subspaces $\mathscr{A}$ and $\mathscr{B}$ are finite-dimensional, with bases the random vectors $\mathbf{a}$ and $\mathbf{b}$, respectively, then the oblique projection of ζ onto $\mathscr{B}$ along $\mathscr{A}$ can be computed by the formula

$$E_{\|\mathscr{A}}\{\zeta \mid \mathscr{B}\} = [\Sigma_{\zeta\mathbf{b}}\ \Sigma_{\zeta\mathbf{a}}]\begin{bmatrix}\Sigma_{\mathbf{bb}} & \Sigma_{\mathbf{ba}}\\ \Sigma_{\mathbf{ab}} & \Sigma_{\mathbf{aa}}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{b}\\ 0\end{bmatrix}.$$

The notation $\mathscr{A} \perp \mathscr{B} \mid \mathscr{C}$ means that the two subspaces $\mathscr{A}$ and $\mathscr{B}$ are *conditionally orthogonal given a third subspace* $\mathscr{C}$, namely, for any $\alpha \in \mathscr{A}$ and $\beta \in \mathscr{B}$

$$\langle \alpha - E\{\alpha \mid \mathscr{C}\}, \beta - E\{\beta \mid \mathscr{C}\}\rangle = 0.$$

If $\mathscr{C} = \{0\}$, conditional orthogonality reduces to the usual orthogonality $\mathscr{A} \perp \mathscr{B}$.

3. Stochastic realization with feedback

Stochastic realization with inputs is discussed in Picci (1997), Picci and Katayama (1996), and Chiuso and Picci (2002, 2003). In this section we shall just recall the basic notions needed in the paper. A key concept is that of an *Oblique Markovian Splitting Subspace*, introduced in Picci (1997), and Chiuso and Picci (2002), which is the basic coordinate-free object equivalent to state space recursions as (2.1). We shall show that an oblique Markovian splitting subspace can be constructed by an oblique projection once a certain space generated by the future innovations has been constructed. We shall first operate in an ideal infinite-data setup, i.e. assuming that the random variables $\mathbf{u}(t)$ and $\mathbf{y}(t)$, are available for all t.

Definition 1. The subspace $\mathscr{X}_t$ is *Oblique Markovian Splitting*, if

$$\begin{aligned} &E[\mathscr{X}_{t+1} \vee \mathscr{Y}_t \mid \mathscr{X}_{t+1}^- \vee \mathscr{U}_t \vee \mathscr{Z}_t^-]\\ &\quad = E[\mathscr{X}_{t+1} \vee \mathscr{Y}_t \mid \mathscr{X}_t + \mathscr{U}_t], \end{aligned} \tag{3.1}$$

which is equivalent to the *conditional orthogonality* property:

$$(\mathscr{X}_{t+1} \vee \mathscr{Y}_t) \perp (\mathscr{X}_t^- \vee \mathscr{Z}_t^-) \mid (\mathscr{X}_t + \mathscr{U}_t). \tag{3.2}$$

If $\mathscr{U}_t \cap \mathscr{X}_t = \{0\}$ then (3.2) implies $E_{\|\mathscr{U}_t}[\mathscr{X}_{t+1} \mid \mathscr{X}_t^- \vee \mathscr{U}_t^-] = E_{\|\mathscr{U}_t}[\mathscr{X}_{t+1} \mid \mathscr{X}_t]$, which is a Markov property, *conditional also on* $\mathscr{U}_t$. An oblique Markovian splitting subspace is *causal* if $\mathscr{X}_t \subseteq \mathscr{Z}_t^-$ and *purely non deterministic* (p.n.d.) if the family $\{\mathscr{X}_t^-\}$ has the p.n.d. property $\bigcap_{t<0}(\mathscr{X}_t^- \vee \mathscr{Z}_t^-) = \{0\}$.

The Oblique Markovian splitting property is precisely what is needed for the space $\mathscr{X}$ to qualify as a state space for a stochastic model described by standard state space equations. The following result is taken from Chiuso and Picci (2002).

Theorem 3.1. *Let $\mathscr{X}_t$ be a p.n.d. oblique Markovian splitting subspace for $(\mathscr{Y}, \mathscr{U})$; then there exists a stationary family of orthogonal subspaces $\mathscr{W}_t$ such that*

$$\mathscr{X}_{t+1} \subseteq (\mathscr{X}_t + \mathscr{U}_t) \oplus \mathscr{W}_t, \tag{3.3}$$

$$\mathscr{Y}_t \subseteq (\mathscr{X}_t + \mathscr{U}_t) \oplus \mathscr{W}_t. \tag{3.4}$$

Conversely, if these inclusions hold $\mathscr{X}_t$ is an oblique Markovian splitting subspace. Let $\mathscr{X}_t$ be finite dimensional with basis vector $\mathbf{x}(t)$ and $\mathbf{w}(t)$ be a basis for $\mathscr{W}_t$; $\mathbf{w}(t)$ is called a generating process of $\mathbf{y}(t)$ given $\mathbf{u}(t)$. Then $\mathbf{x}(t)$ satisfies

$$\begin{aligned} \mathbf{x}(t+1) &= A\mathbf{x}(t) + B\mathbf{u}(t) + G\mathbf{w}(t),\\ \mathbf{y}(t) &= C\mathbf{x}(t) + D\mathbf{u}(t) + J\mathbf{w}(t), \end{aligned} \qquad t \geqslant t_0, \tag{3.5}$$

for suitable constant matrices (A, B, C, D, G, J).

An oblique Markovian splitting subspace of minimal dimension n (equal to the dimension of the state space of the data-generating model (2.1)), is called a minimal oblique Markovian splitting subspace. Oblique Markovian splitting subspaces are highly nonunique; in fact, even imposing minimality, there are in general infinitely many such subspaces. However, in identification we are usually interested in the special one which is attached to the (forward) innovation model of $\mathbf{y}$ given $\mathbf{u}$, i.e. the one for which $\mathbf{w}(t)$ is the one step ahead prediction error of $\mathbf{y}(t)$ given $\mathscr{Z}_t^-$, i.e.

$$\mathbf{w}(t) = \mathbf{e}(t) := \mathbf{y}(t) - E[\mathbf{y}(t) \mid \mathscr{Z}_t^-]. \tag{3.6}$$

The space spanned by $\mathbf{e}(t)$ will be denoted as $\mathscr{E}_t$. A central question in subspace identification is how to construct the state space of the innovation model from the random variables of the processes $\mathbf{y}$, $\mathbf{u}$, possibly in the presence of feedback. The following construction gives an answer to the above question in the context of infinite data. Introduce

$$\mathscr{G}_t := \mathscr{E}_t \vee \mathscr{U}_t$$

and accordingly define $\mathscr{G}_{[t,T]}$, and $\mathscr{G}_t^+$. We may think of $\mathscr{G}_t$ as being an "extended input" space and $\mathscr{G}_{[t,T]}$ and $\mathscr{G}_t^+$ as being, respectively, the finite and infinite extended future space.

Theorem 3.2. *Assume that $\mathscr{G}_{[t,T]} \cap \mathscr{Z}_t^- = \{0\}$. Then, for all $k \geqslant n$,*

$$\begin{aligned} &E_{\|\mathscr{G}_{[t,t+k-1]}}[\mathscr{Y}_{[t,t+k-1]} \mid \mathscr{Z}_t^-]\\ &\quad = E_{\|\mathscr{G}_{[t,t+k]}}[\mathscr{Y}_{[t,t+k]} \mid \mathscr{Z}_t^-] \end{aligned} \tag{3.7}$$

holds, where n is the dimension of the model (2.1) *and*

$$\mathscr{X}_t^{+/-} := E_{\|\mathscr{G}_{[t,T]}}[\mathscr{Y}_{[t,T]} \mid \mathscr{Z}_t^-], \quad T \geqslant t+n-1 \tag{3.8}$$

is a minimal oblique Markovian splitting subspace, called the (*oblique*) *predictor space of* $\mathbf{y}$ *which serves as a state space of the innovation model* (2.1).

Proof. We just need to verify that $\mathscr{X}_t^{+/-}$ defined in (3.8) satisfies conditions (3.3) and (3.4) where $\mathscr{W}_t = \mathscr{E}_t$. Note that since $\mathscr{Y}_t \subseteq \mathscr{X}_t^- \oplus \mathscr{E}_t$ we have $E_{\|\mathscr{G}_{[t,T]}}[\mathscr{Y}_t \mid \mathscr{Z}_t^-] = E_{\|\mathscr{E}_t}[\mathscr{Y}_t \mid \mathscr{Z}_t^-] = E[\mathscr{Y}_t \mid \mathscr{Z}_t^-]$ which by construction is contained in $\mathscr{X}_t^{+/-}$. Therefore $\mathscr{Y}_t \subseteq \mathscr{X}_t^{+/-} \oplus \mathscr{E}_t$, which is (3.4). As far as (3.3), note that for $(T > t + n - 1)$

$$\begin{aligned}\mathscr{X}_{t+1}^{+/-} &= E_{\|\mathscr{G}_{[t+1,T]}}[\mathscr{Y}_{[t+1,T]} \mid \mathscr{Z}_{t+1}^-] \\ &= E_{\|\mathscr{G}_{[t+1,T]}}[\mathscr{Y}_{[t+1,T]} \mid \mathscr{Z}_t^- + \mathscr{G}_t] \\ &\subseteq E_{\|\mathscr{G}_{[t,T]}}[\mathscr{Y}_{[t+1,T]} \mid \mathscr{Z}_t^-] + \mathscr{G}_t \\ &\subseteq E_{\|\mathscr{G}_{[t,T]}}[\mathscr{Y}_{[t,T]} \mid \mathscr{Z}_t^-] + \mathscr{G}_t \\ &= (\mathscr{X}_t^{+/-} + \mathscr{U}_t) \oplus \mathscr{E}_t\end{aligned}$$

which is condition (3.3). □

From (3.3) and (3.4) finding a representation of form (2.1) is just a matter of choosing a basis $\mathbf{x}(t)$ in $\mathscr{X}_t^{+/-}$.

The theorem above gives a precise recipe for constructing the oblique predictor space $\mathscr{X}_t^{+/-}$ in the stationary *infinite data* case. It can be implemented by the following conceptual algorithm:

(a) Choose a suitably large[5] k (or equivalently a suitably large $T := t + k$) and construct the future innovations $\mathbf{e}(t+i) := \mathbf{y}(t+i) - E[\mathbf{y}(t+i) \mid \mathscr{Z}_{t+i}^-], i = 0, 1, \ldots, k$.
(b) Form the extended future space $\mathscr{G}_{[t,t+k]}$.
(c) Compute the oblique projection (3.8) to get $\mathscr{X}_t^{+/-}$.
(d) Repeat the construction shifting time to $t + 1$ to form $\mathscr{X}_{t+1}^{+/-}$.
(e) Choose bases consistently in $\mathscr{X}_t^{+/-}$ and $\mathscr{X}_{t+1}^{+/-}$ and solve a linear regression to estimate A, B, C by least squares.

This conceptual procedure is implemented by the following algorithm, which has been introduced in Qin and Ljung (2003). The main steps of this algorithm are as follows (further details will be given in Section 5).

Algorithm 1 (*Qin and Ljung, 2003*).

(a) Choose (the "past" and "future" horizons) $t - t_0$ and $T - t$ and compute the *transient* innovations (which depend on the initial time t_0)

$$\hat{\mathbf{e}}(t + i) := \mathbf{y}(t + i) - E[\mathbf{y}(t + i) \mid \mathscr{Z}_{[t_0,t+i)}]$$

for $i = 0, \ldots, T - t$.

(b) Using the transient innovations form the subspace $\hat{\mathscr{G}}_{[t,T)}$, a "transient" version of $\mathscr{G}_{[t,T)}$.
(c) Approximate the oblique projection (3.8) by computing $E_{\|\hat{\mathscr{G}}_{[t,T)}}[\mathbf{y}_{[t,T)} \mid \mathscr{Z}_{[t_0,t)}]$ and $E_{\|\hat{\mathscr{G}}_{[t+1,T]}}[\mathbf{y}_{[t+1,T]} \mid \mathscr{Z}_{[t_0,t+1)}]$ at time $t + 1$.
(d) These projections should theoretically be n-dimensional, but will be of full rank for real data. Using standard procedures based on truncated SVD, find coherent bases $\hat{\mathbf{x}}(t)$ and $\hat{\mathbf{x}}(t + 1)$ for the two subspaces.
(e) Using $\hat{\mathbf{x}}(t)$ and $\hat{\mathbf{x}}(t + 1)$ solve a linear regression to get the parameters (A, B, C).

Unfortunately, there are two pitfalls in this procedure.

(a) In Theorem 3.2 we made the assumption that $\mathscr{G}_{[t,T]} \cap \mathscr{Z}_t^- = \{0\}$. It can be shown that this condition indeed holds true for any finite future horizon k, provided $F(z)$ is nonanticipative, i.e. it has no poles at infinity. However, if the transfer function $F(z)$ has unstable poles, the condition fails asymptotically, i.e. $\mathscr{G}_t^+ \cap \mathscr{Z}_t^- \neq \{0\}$ as shown in Chiuso and Picci (2002).
This fact, which may seem marginal at a first glance, has a significant impact if the unstable dynamics is "fast", i.e. there are unstable eigenvalues $\lambda_i(A)$ which are far from the unit circle ($|\lambda_i(A)| \gg 1$). In this case the spaces $\mathscr{G}_{[t,T]}$ and the finite past $\mathscr{Z}_{[t_0,t)}$ may become close (or nearly parallel, in the sense of subspace angles) also for k finite and possibly small. It is well known that the oblique projection along subspaces which are nearly parallel may be ill-conditioned and hence the estimation based on oblique projection may not be well-behaved. Something similar has been observed and studied in Chiuso and Picci(2004a,b) for subspace identification in open loop.
(b) The second problem has to do with the fact that in practice one can only regress on a finite amount of data and some finite past $\mathscr{Z}_{[t_0,t)}$ must be used in place of $\mathscr{Z}_t^-$ to approximate the construction of $\mathscr{X}_t^{+/-}$. This implies, as we shall see later, that the estimates will be biased for finite $t - t_0$. One straightforward solution could be to take $t - t_0$ very large i.e. $t_0 \to -\infty$. However, as it is well known in open loop identification, this solution does not provide consistent estimators when there are zeros (of the stochastic subsystem) on the unit circle. Even if this is not the case, the amount of bias will anyway depend on how close to the unit circle are the zeros.

There is an alternative procedure to construct the oblique predictor space $\mathscr{X}_t^{+/-}$ which does not suffer from the possible ill-conditioning mentioned in (a) above, occurring when the open loop system is unstable. It is based on the observation that $\mathscr{X}_t^{+/-}$ is also the state space of the inverse system generating the innovations $\mathbf{e}(t)$ from the joint process $\mathbf{y}$ and $\mathbf{u}$ (the "whitening filter") which is well known to be asymp-

[5] "Suitably large" means larger than the true system order n, which we shall assume is known. Of course in practice n must be estimated from the data. However, we shall not discuss the order estimation problem in this paper.

totically stable under mild conditions on the zeros of the system. The following proposition states this well-known fact in geometric terms.

Proposition 3.3. *Let $\mathscr{E}_t$ be the space generated by the stationary innovation* $\mathbf{e}(t)$, *defined in* (3.6), *and $\mathscr{X}_t^{+/-}$ the oblique predictor space. Then the following holds*:

$$\mathscr{X}_{t+1}^{+/-} \subseteq \mathscr{X}_t^{+/-} + \mathscr{U}_t + \mathscr{Y}_t, \tag{3.9}$$

$$\mathscr{Y}_t \subseteq \mathscr{X}_t^{+/-} \oplus \mathscr{E}_t, \tag{3.10}$$

$$\mathscr{E}_t \subseteq \mathscr{X}_t^{+/-} + \mathscr{Y}_t. \tag{3.11}$$

In fact $\mathscr{X}_t^{+/-}$ is also a minimal oblique Markovian splitting subspace for $\mathbf{e}$.

Proof. The proof is just based on rearranging the inclusions in Theorem 3.1 and recalling that by definition of splitting, $\mathbf{e}(t) := \mathbf{y}(t) - E[\mathbf{y}(t) \mid \mathscr{Z}_t^-] = \mathbf{y}(t) - E[\mathbf{y}(t) \mid \mathscr{X}_t^{+/-}]$. □

Remark 3.1. Eqs. (3.9)–(3.11) show that $\mathscr{X}_t^{+/-}$ is in fact the state space of the "inverse" system producing the innovation $\mathbf{e}(t)$ from past input and output measurements $\{\mathbf{y}(s), \mathbf{u}(s), s \leqslant t\}$. It follows that $\mathscr{X}_t^{+/-}$ is the oblique predictor space of $\mathscr{E}_t^+$ given $\mathscr{Z}_t^+$, i.e.

$$\mathscr{X}_t^{+/-} = E_{\parallel \mathscr{Z}_t^+}[\mathscr{E}_t^+ \mid \mathscr{Z}_t^-]. \tag{3.12}$$

This oblique projection of course make sense if and only if $\mathscr{Z}_t^+ \cap \mathscr{Z}_t^- = \{0\}$, which is guaranteed if the spectrum of the joint process is bounded away from zero (Hannan & Poskitt, 1988). For finite-dimensional models this in particular requires that $\bar{A} := A - KC$ be strictly stable, i.e. should have no zeros on the unit circle.[6]

One advantage of this construction is that one needs not pre-compute the future innovation space $\mathscr{E}_t^+$ to obtain $\mathscr{X}_t^{+/-}$.

Theorem 3.4. *Assume that the joint process satisfies*

$$\mathscr{Z}_t^+ \cap \mathscr{Z}_t^- = \{0\}$$

then the space $\mathscr{X}_t^{+/-}$ is generated by the oblique projections $E_{\parallel \mathscr{Z}_{[t,t+k)}}[\mathscr{Y}_{t+k} \mid \mathscr{Z}_t^-]$, for $k = 0, 1, \ldots, \infty$, i.e.

$$\mathscr{X}_t^{+/-} = \bigvee_{k=0}^{\infty} E_{\parallel \mathscr{Z}_{[t,t+k)}}[\mathscr{Y}_{t+k} \mid \mathscr{Z}_t^-]. \tag{3.13}$$

In the finite-dimensional case (*i.e. $F(z)$ and $G(z)$ rational*), *the closed vector sum can be stopped at any $k \geqslant n$ where n is the system order, i.e. the dimension of $\mathscr{X}_t^{+/-}$, in which case it is only required that $\mathscr{Z}_{[t,t+k)} \cap \mathscr{Z}_t^- = \{0\}$.*

[6] Recall that strict stability of the predictor is always required for prediction error methods, and it is also postulated in Jansson (2003).

Proof. From (3.10) and iterating (3.9), the output space at time $t+k$, $k \geqslant 0$ satisfies the following inclusion:

$$\mathscr{Y}_{t+k} \subseteq (\mathscr{X}_t^{+/-} + \mathscr{Z}_{[t,t+k)}) \oplus \mathscr{E}_{t+k}.$$

Furthermore, minimality of $\mathscr{X}_t^{+/-}$ (in fact, observability of the inverse system) ensures that indeed

$$\begin{aligned}\mathscr{X}_t^{+/-} &= \bigvee_{k>0} \overline{\operatorname{span}}\{E_{\parallel \mathscr{Z}_t^+}[\mathbf{e}(t+k) \mid \mathscr{Z}_t^-]\} \\ &= \bigvee_{k>0} \overline{\operatorname{span}}\{E_{\parallel \mathscr{Z}_t^+}[E[\mathbf{y}(t+k) \mid \mathscr{Z}_{t+k}^-] \mid \mathscr{Z}_t^-]\} \\ &= \bigvee_{k>0} \overline{\operatorname{span}}\{E_{\parallel \mathscr{Z}_{[t,t+k)}}[\mathbf{y}(t+k) \mid \mathscr{Z}_t^-]\},\end{aligned}$$

where the second equality stems from the fact that $\mathbf{e}(t+k) = \mathbf{y}(t+k) - E[\mathbf{y}(t+k) \mid \mathscr{Z}_{t+k}^-]$ and $E_{\parallel \mathscr{Z}_t^+}[\mathbf{y}(t+k) \mid \mathscr{Z}_t^-] = 0$.

In the finite-dimensional case, using a minimal realization

$$E_{\parallel \mathscr{Z}_{[t,t+k)}}[\mathbf{y}(t+k) \mid \mathscr{Z}_t^-] = CA^{k-1}\mathbf{x}(t)$$

holds. Therefore, by Cayley–Hamilton theorem, the sum can be stopped at the system order $k = n$. □

We would like to stress that Eq. (3.13) just involves computing oblique projections of future outputs ($\mathbf{y}(t+k)$) along the future input and output space ($\mathscr{Z}_{[t,t+k)}$) onto the past data ($\mathscr{Z}_t^-$). This yields an alternative procedure (based on infinite past data) to estimate the system matrices (A, B, C):

(a) Compute the oblique projections

$$E_{\parallel \mathscr{Z}_{[t,t+k)}}[\mathscr{Y}_{t+k} \mid \mathscr{Z}_t^-], \quad k = 0, \ldots, K$$

and find $\mathscr{X}_t^{+/-}$ as a "best" n-dimensional[7] approximation of the space generated by these oblique predictors. Fix a suitable basis in $\mathscr{X}_t^{+/-}$.

(b) Repeat the same procedure shifting time to $t+1$, to get a (coherent) basis in $\mathscr{X}_{t+1}^{+/-}$.

(c) Solve by standard least squares for the system matrices (A, B, C).

The only practical drawback of this procedure is that the infinite past is not available and one has to work with its truncated version $\mathscr{Z}_{[t_0,t)}$. Again the conceptual procedure can be approximately implemented by the following algorithm:

Algorithm 2 (*"Whitening Filter" Algorithm*).

(a) Choose "past" and "future" horizons $t - t_0$ and $T - t$ and compute the oblique predictors

$$\hat{\mathbf{y}}(t+i \mid t) := E_{\parallel \mathscr{Z}_{[t,t+i)}}[\mathbf{y}(t+i) \mid \mathscr{Z}_{[t_0,t)}]$$

[7] Here the system order n is also assumed to be known. Of course any consistent order estimation procedure used in subspace identification would serve to the purpose. Order estimation is performed in most subspace identification algorithms by a (weighted) SVD truncation step which we shall not discuss in this paper.

for $i = 0, \dots, T-1$ and

$$\hat{\mathbf{y}}(t+i \mid t+1) := E_{\|\mathscr{Z}_{[t+1,t+i)}}[\mathbf{y}(t+i) \mid \mathscr{Z}_{[t_0,t+1)}]$$

for $i = 1, \dots, T$.

(b) The projections $\{\hat{\mathbf{y}}(t+i \mid t),\ i = 0, \dots, T-1\}$ and $\{\hat{\mathbf{y}}(t+i \mid t+1),\ i = 1, \dots, T\}$ generate the state spaces at time t and $t+1$, respectively (see (3.13)). With real data the two families of generators will be of full rank and one should use standard procedures based on SVD truncation to find coherent bases $\hat{\mathbf{x}}(t)$ and $\hat{\mathbf{x}}(t+1)$ of dimension n.

(c) Once $\hat{\mathbf{x}}(t)$ and $\hat{\mathbf{x}}(t+1)$ are computed, one could solve a linear regression (see (4.23) below) to estimate A, B, C.

The use of finite past data will also introduce a bias which we shall analyze in detail in Section 6.

The algorithm of the paper (Jansson, 2003) implements a slight variation of this procedure. The oblique projection step is replaced by a preliminary estimation of the Markov parameters of the system by a "long" autoregression and by an orthogonal projection after the effect of future inputs is removed. In the experiments reported in this paper the implementation introduced by Jansson and the oblique projection procedure proposed above show a very similar behavior, which in both cases is quite close to that of standard prediction error methods. However, the geometric flavor of the procedure based on the oblique projection is more in the spirit of subspace identification and geometric stochastic realization and in the following we shall mostly deal with the "geometric" version.

4. Joint vs. input–output modeling

To better understand how transients could be handled, in this section we shall discuss state-space modeling of the joint process $[\mathbf{y}^\top \mathbf{u}^\top]^\top$ based on data from a finite interval $[t_0, T]$. Stationary (infinite-interval) input–output models with feedback have been thoroughly analyzed in the literature, see Gevers and Anderson (1981, 1982). In this section we shall instead discuss nonstationary joint state space representations; in particular, innovation representations which turn out to be of the Kalman filter type.

Let (2.1) be a stationary innovation representation of $\mathbf{y}$ given $\mathbf{u}$ and similarly let

$$\begin{aligned} \mathbf{s}(t+1) &= F\mathbf{s}(t) + G\mathbf{y}(t) + L\mathbf{v}(t), \\ \mathbf{u}(t) &= H\mathbf{s}(t) + J\mathbf{y}(t) + \mathbf{v}(t), \end{aligned} \qquad t \geqslant t_0, \tag{4.1}$$

be a stationary innovation representation of $\mathbf{u}$ given $\mathbf{y}$ with the innovation process of $\mathbf{u}$ based on the joint past as the white noise input

$$\mathbf{v}(t) := \mathbf{u}(t) - E[\mathbf{u}(t) \mid \mathscr{Z}_t^- \vee \mathscr{Y}_t].$$

For future reference we define $\bar{F} := F - LH$.

Under mild assumptions $\mathbf{q}(t) := [\mathbf{x}^\top(t)\ \mathbf{s}^\top(t)]^\top$ is a minimal state vector for the joint innovation model

$$\begin{aligned} \mathbf{q}(t+1) &= \mathscr{A}\mathbf{q}(t) + \mathscr{K}\mathbf{n}(t), \\ \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{u}(t) \end{bmatrix} &= \mathscr{C}\mathbf{q}(t) + \mathbf{n}(t), \end{aligned} \tag{4.2}$$

where

$$\mathscr{A} := \begin{bmatrix} A + BJC & BH \\ GC & F \end{bmatrix},$$

$$\mathscr{K} := \begin{bmatrix} K + BJ & B \\ G & L \end{bmatrix}, \quad \mathscr{C} := \begin{bmatrix} C & 0 \\ JC & H \end{bmatrix}$$

and

$$\mathbf{n}(t) := \begin{bmatrix} I & 0 \\ J & I \end{bmatrix} \begin{bmatrix} \mathbf{e}(t) \\ \mathbf{v}(t) \end{bmatrix}$$

is the joint stationary innovation process. In practice one has only a finite amount of data and the (stationary) innovations $\mathbf{e}(t)$ and $\mathbf{v}(t)$ will have to be replaced by the transient ones

$$\hat{\mathbf{e}}(t) := \mathbf{y}(t) - E[\mathbf{y}(t) \mid \mathscr{Z}_{[t_0,t)}] \tag{4.3}$$

and

$$\hat{\mathbf{v}}(t) := \mathbf{u}(t) - E[\mathbf{u}(t) \mid \mathscr{Z}_{[t_0,t)} \vee \mathscr{Y}_t]. \tag{4.4}$$

We use the notation $\hat{\mathbf{e}}(t)$ to remind that this random variable is a function of the initial time instant t_0 and should not be confused with the stationary innovation $\mathbf{e}(t)$. For $t - t_0 \to \infty$, $\hat{\mathbf{e}}(t)$ tends to $\mathbf{e}(t)$ (in mean square).

A finite-interval state-space innovation model for $\mathbf{y}$ and $\mathbf{u}$ is obtained by setting up a (transient) Kalman filter based on the stationary model (4.2), which updates the projections $\hat{\mathbf{x}}(t) := E[\mathbf{x}(t) \mid \mathscr{Z}_{[t_0,t)}]$ and $\hat{\mathbf{s}}(t) := E[\mathbf{s}(t) \mid \mathscr{Z}_{[t_0,t)}]$ based on the finite observation interval $[t_0, T]$. The projections form the state of the transient model and are updated recursively according to the decomposition $\mathscr{Z}_{[t_0,t+1)} = (\mathscr{Z}_{[t_0,t)} \oplus \hat{\mathscr{E}}_t) \oplus \hat{\mathscr{V}}_t$ where

$$\hat{\mathscr{E}}_t := \operatorname{span}\{\hat{\mathbf{e}}(t)\}, \quad \hat{\mathscr{V}}_t := \operatorname{span}\{\hat{\mathbf{v}}(t)\}. \tag{4.5}$$

Letting $\hat{\mathbf{q}}(t) := [\hat{\mathbf{x}}^\top(t)\ \hat{\mathbf{s}}^\top(t)]^\top$ and

$$\mathscr{K}(t) = \begin{bmatrix} K(t) + B(t)J(t) & B(t) \\ G(t) & L(t) \end{bmatrix} \tag{4.6}$$

leads to a representation of the following form:

$$\begin{aligned} \hat{\mathbf{q}}(t+1) &= \mathscr{A}\hat{\mathbf{q}}(t) + \mathscr{K}(t) \begin{bmatrix} \hat{\mathbf{e}}(t) \\ \hat{\mathbf{v}}(t) \end{bmatrix}, \\ \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{u}(t) \end{bmatrix} &= \mathscr{C}\hat{\mathbf{q}}(t) + \begin{bmatrix} I & 0 \\ J(t) & I \end{bmatrix} \begin{bmatrix} \hat{\mathbf{e}}(t) \\ \hat{\mathbf{v}}(t) \end{bmatrix}, \end{aligned} \tag{4.7}$$

associated to zero initial conditions $\hat{\mathbf{q}}(t_0) = 0$. The matrices $K(t)$, $B(t)$, $G(t)$, $J(t)$, $L(t)$ are computed using standard Riccati equations involving the joint model parameters in (4.2). The upper left block of the transient Kalman gain

has been written in a form which resembles the stationary structure $K+BJ$ for ease of comparison. It is clear that even though we are interested in updating only the projection of $\mathbf{x}$, the recursion for $\hat{\mathbf{x}}$ involves $\hat{\mathbf{s}}$ and $\hat{\mathbf{v}}$, which in turn require an explicit model of $\mathbf{u}$. This is the reason why it is not possible to describe the evolution of $\mathbf{y}$ given $\mathbf{u}$ on a finite interval, without the current state of $\mathbf{u}$ itself taking part in this evolution.

Using the last equation in (4.7) to express $\hat{\mathbf{v}}(t)$ as a function if $\mathbf{u}(t), \hat{\mathbf{x}}(t), \hat{\mathbf{s}}(t)$ and $\hat{\mathbf{e}}(t)$, and substituting back into the state equation in (4.7) we obtain an exact transient state-space representation of $\mathbf{y}$ given $\mathbf{u}$,

$$\hat{\mathbf{q}}(t+1) = A_J(t)\hat{\mathbf{q}}(t) + K_u(t)\mathbf{u}(t) + K_e(t)\hat{\mathbf{e}}(t),$$

$$\mathbf{y}(t) = [\,C \quad 0\,]\,\hat{\mathbf{q}}(t) + \hat{\mathbf{e}}(t) \tag{4.8}$$

also started at zero initial condition. The various time-varying parameters in this model are defined as follows. Letting $K_u(t) := [B(t)^\top\ L(t)^\top]^\top$ to be the second block-column of $\mathscr{K}(t)$ in (4.6) we have

$$\begin{aligned} A_J(t) &:= \mathscr{A} - K_u(t)\,[\,JC \quad H\,] \\ &= \begin{bmatrix} A + \tilde{B}(t)JC & \tilde{B}(t)H \\ GC - L(t)JC & F - L(t)H \end{bmatrix}, \end{aligned} \tag{4.9}$$

where $\tilde{B}(t) := B - B(t)$. The gain matrix $K_e(t)$ is instead obtained by subtracting from the first block-column of $\mathscr{K}(t)$ the term $K_u(t)J(t)$, so that

$$K_e(t) = \begin{bmatrix} K(t) \\ G(t) - L(t)J(t) \end{bmatrix}. \tag{4.10}$$

Note that this representation is of the same dimension of the joint model (4.2).

From standard Kalman filter theory, the two block components of the"Kalman gain" $K_u(t)$ tend, in the limit for $t - t_0 \to \infty$, to the constant gain matrices B and L, so that the matrix $\tilde{B}(t)$ converges to zero as $t - t_0 \to \infty$. It is well known (Anderson & Moore, 1979), that the rate of convergence is governed by the (stable) zeros of the joint spectrum of the stationary model, i.e. by the eigenvalues of

$$\mathscr{A} - \mathscr{K}\begin{bmatrix} I & 0 \\ -J & I \end{bmatrix}\mathscr{C} = \begin{bmatrix} \bar{A} & 0 \\ 0 & \bar{F} \end{bmatrix}.$$

In fact, it can be seen that asymptotically with $t - t_0 \to \infty$, $\tilde{B}(t)$ is of the same order (i.e. tends to zero at the same rate) of the product of the matrices $\bar{A}^{t-t_0}$ and $\bar{F}^{t-t_0}$. In particular we have convergence to zero in a finte number of steps of $\tilde{B}(t)$ if at least one of the two matrices $\bar{A}$ and $\bar{F}$ is *nilpotent*. Nilpotency of $\bar{F}$, and/or $\bar{A}$ means of course that the (conditional) models of $\mathbf{u}$ given $\mathbf{y}$, and/or of $\mathbf{y}$ given $\mathbf{u}$, are of the ARX type. If $\bar{A}$ is nilpotent, both $B(t)$ and $K(t)$ converge to the stationary values in at most n steps. In any case when $\tilde{B}(t) = 0$ and $K(t) = K$, the model (4.8) reduces[8] to

$$\begin{aligned} \begin{bmatrix} \hat{\mathbf{x}}(t+1) \\ \hat{\mathbf{s}}(t+1) \end{bmatrix} &= \begin{bmatrix} A & 0 \\ GC - L(t)JC & F - L(t)H \end{bmatrix}\begin{bmatrix} \hat{\mathbf{x}}(t) \\ \hat{\mathbf{s}}(t) \end{bmatrix} \\ &\quad + \begin{bmatrix} B \\ L(t) \end{bmatrix}\mathbf{u}(t) + \begin{bmatrix} K \\ G(t) - L(t)J(t) \end{bmatrix}\hat{\mathbf{e}}(t), \end{aligned}$$

$$\mathbf{y}(t) = [\,C \quad 0\,]\begin{bmatrix} \hat{\mathbf{x}}(t) \\ \hat{\mathbf{s}}(t) \end{bmatrix} + \hat{\mathbf{e}}(t),$$

$$\hat{\mathbf{x}}(t_0) = 0 \quad \hat{\mathbf{s}}(t_0) = 0, \tag{4.11}$$

which is obviously not observable as the state $\hat{\mathbf{s}}(t)$ is not coupled to the measurement of $\mathbf{y}$ and can be dropped yielding back the original n-dimensional stationary model (2.1). Hence independence of the one-step predictor of $\mathbf{y}$ from the past dynamics of $\mathbf{u}$, which is well known to hold in the stationary setting[9] holds in general only asymptotically (or at most after a finite number of steps in the nilpotent case).

Remark 4.2. Note that in the absence of feedback the coupling of $\hat{\mathbf{x}}(t+1)$ with the state of the input process can be avoided using a "conditional" Kalman filter given the whole input history (see Overschee & De Moor, 1994; Chiuso & Picci, 2004b, formula (2.12)). This can be done because in the absence of feedback the dependence of $\mathbf{x}(t)$ on future inputs $\mathscr{U}_{[t,T]}$ conditional on $\mathscr{Y}_{[t_0,t)} \vee \mathscr{U}_{[t_0,t)}$ is all contained in the initial condition $\hat{\mathbf{x}}(t_0) := E[\mathbf{x}(t_0) \mid \mathscr{U}_{[t_0,T]}]$ (see Chiuso & Picci, 2004b, Theorem 1 & formula (2.16)). Unfortunately this is no longer true when feedback is present.

The transient *whitening filter realization* of $\hat{\mathbf{e}}$ is just the inverse system of (4.7). Defining $\tilde{G}(t) := G - G(t)$, $\tilde{J}(t) := J - J(t)$ and also the state transition matrix for the whitening filter representation of $\mathbf{y}$ given $\mathbf{u}$

$$\bar{A}_J(t) := \begin{bmatrix} A - K(t)C + \tilde{B}(t)JC & \tilde{B}(t)H \\ \tilde{G}(t)C - L(t)\tilde{J}(t)C & F - L(t)H \end{bmatrix} \tag{4.12}$$

one gets,

$$\begin{aligned} \hat{\mathbf{q}}(t+1) &= \bar{A}_J(t)\hat{\mathbf{q}}(t) + K_u(t)\mathbf{u}(t) + K_e(t)\mathbf{y}(t) \\ \hat{\mathbf{e}}(t) &= -[\,C \quad 0\,]\,\hat{\mathbf{q}}(t) + \mathbf{y}(t) \end{aligned} \tag{4.13}$$

again with zero initial condition. All the models above have time-varying parameters and is not clear how they could be used for identification. Remarkably, the deviation from the stationary model can be lumped into an additive error term, as explained in the following proposition.

Proposition 4.1. *Define*

$$\tilde{\mathbf{x}}(t) := E[\mathbf{x}(t) \mid \mathscr{Z}_{[t_0,t+1)}] - \hat{\mathbf{x}}(t). \tag{4.14}$$

[8] We remind the reader that $L(t)$ and $J(t)$ converge in a finite number of steps only if $\bar{F}$ is also nilpotent.

[9] And is actually the key condition on which PEM methods with feedback are based.

Then the following representation holds:

$$\hat{\mathbf{x}}(t+1) = A\hat{\mathbf{x}}(t) + B\mathbf{u}(t) + K\hat{\mathbf{e}}(t) + \bar{A}\tilde{\mathbf{x}}(t), \tag{4.15}$$

$$\mathbf{y}(t) = C\hat{\mathbf{x}}(t) + \hat{\mathbf{e}}(t) \tag{4.16}$$

which, letting $\hat{K}(t) := K(t) - \tilde{B}(t)J(t)$, *can also be written as*

$$\hat{\mathbf{x}}(t+1) = A\hat{\mathbf{x}}(t) + B\mathbf{u}(t) + \hat{K}(t)\hat{\mathbf{e}}(t) - \tilde{B}(t)\hat{\mathbf{v}}(t), \tag{4.17}$$

$$\mathbf{y}(t) = C\hat{\mathbf{x}}(t) + \hat{\mathbf{e}}(t), \tag{4.18}$$

moreover we have the identity

$$\bar{A}\tilde{\mathbf{x}}(t) = -(\tilde{B}(t)J(t) + \tilde{K}(t))\hat{\mathbf{e}}(t) - \tilde{B}(t)\hat{\mathbf{v}}(t). \tag{4.19}$$

Proof. It follows from (4.14) that $E[\mathbf{x}(t) \mid \mathscr{Z}_{[t_0,t+1)}] = \hat{\mathbf{x}}(t) + \tilde{\mathbf{x}}(t)$; therefore $E[\mathbf{e}(t) \mid \mathscr{Z}_{[t_0,t+1)}] = \mathbf{y}(t) - CE[\mathbf{x}(t) \mid \mathscr{Z}_{[t_0,t+1)}] = \hat{\mathbf{e}}(t) - C\tilde{\mathbf{x}}(t)$; (4.15) hence follows by projecting $\mathbf{x}(t+1)$ onto $\mathscr{Z}_{[t_0,t+1)}$. Representation (4.17) follows instead by rearranging terms in (4.7); it suffices to write $H\hat{\mathbf{s}}(t) = \mathbf{u}(t) - JC\hat{\mathbf{x}}(t) - J(t)\hat{\mathbf{e}}(t) - \hat{\mathbf{v}}(t)$ from the output equation and substitute into the state update for the $\mathbf{x}$ component. Finally, letting $\tilde{K}(t) := K - K(t)$, it is immediate to recognize from (4.15) and (4.17) that $\tilde{\mathbf{x}}(t)$ satisfies (4.19). □

These representations will be used to compute various bias terms in the next section. In the rest of this section we shall derive some explicit formulas describing how fast the error $\tilde{\mathbf{x}}(t)$ and the transient gains $\tilde{B}(t)$ and $\tilde{K}(t)$ tend to zero with $t - t_0$. The proofs may be found in the appendix.

Lemma 4.2. *The projection error* $\tilde{\mathbf{x}}(t)$ *can be expressed as*

$$\tilde{\mathbf{x}}(t) = \bar{A}^{t-t_0}(E[\mathbf{x}(t_0) \mid \hat{\mathscr{E}}_t] + E[\mathbf{x}(t_0) \mid \hat{\mathscr{V}}_t]). \tag{4.20}$$

Letting $\Lambda_e(t) := \mathrm{Var}\{\hat{\mathbf{e}}(t)\}$, $\Lambda_v(t) := \mathrm{Var}\{\hat{\mathbf{v}}(t)\}$, *the two last terms can in turn be written as*:

$$E[\mathbf{x}(t_0) \mid \hat{\mathscr{E}}_t] = \mathrm{Var}\{\mathbf{x}(t_0) \mid \mathscr{Z}_{[t_0,t)}\}(\bar{A}^\top)^{t-t_0}C^\top \times \Lambda_e^{-1}(t)\hat{\mathbf{e}}(t) \tag{4.21}$$

and

$$E[\mathbf{x}(t_0) \mid \hat{\mathscr{V}}_t] = \mathrm{Cov}\{\mathbf{x}(t_0), \mathbf{s}(t_0) \mid \mathscr{Z}_{[t_0,t)} \vee \mathscr{Y}_t\} \times (\bar{F}^\top)^{t-t_0}H^\top\Lambda_v^{-1}(t)\hat{\mathbf{v}}(t). \tag{4.22}$$

Remark 4.3. These last two equations, together with (4.19), show that $\tilde{\mathbf{x}}(t)$, similarly to what happens to $\tilde{B}(t)$, converges to zero at a rate determined by both $\bar{A}$ and $\bar{F}$, while $K(t) \to K$ at a rate basically depending only upon $\bar{A}$. As we shall see later, one important consequence of this fact is that some bias term will disappear when subsystem (4.1) generating the feedback signal is of the ARX type.

There is one last technical result of the same kind which will be needed in the next sections. Observe that, defining $\bar{A}(t) := A - \hat{K}(t)C$, the whitening filter of $\mathbf{y}$ given $\mathbf{u}$ can be written in a "perturbed" form similar to (4.17),

$$\hat{\mathbf{x}}(t+1) = \bar{A}(t)\hat{\mathbf{x}}(t) + B\mathbf{u}(t) + \hat{K}(t)\mathbf{y}(t) - \tilde{B}(t)\hat{\mathbf{v}}(t),$$

$$\hat{\mathbf{e}}(t) = -C\hat{\mathbf{x}}(t) + \mathbf{y}(t). \tag{4.23}$$

Lemma 4.3. *Let* $\hat{\mathbf{v}}(t)$ *be the transient innovation process of the input* $\mathbf{u}$. *The following formulas hold*:

$$E_{\|\hat{\mathscr{G}}_{[t,T]}}[\hat{\mathbf{v}}(t+k) \mid \mathscr{Z}_{[t_0,t)}] = [-JC \quad -H] \prod_{h=t}^{t+k-1} A_J(h) \begin{bmatrix} \hat{\mathbf{x}}(t) \\ \hat{\mathbf{s}}(t) \end{bmatrix} \tag{4.24}$$

for $0 \leqslant k \leqslant T - t$ *and*

$$E_{\|\mathscr{Z}_{[t,t+i)}}[\hat{\mathbf{v}}(t+k) \mid \mathscr{Z}_{[t_0,t)}] = [-\tilde{J}(t+k)C \quad -H] \prod_{h=t}^{t+k-1} \bar{A}_J(h) \begin{bmatrix} \hat{\mathbf{x}}(t) \\ \hat{\mathbf{s}}(t) \end{bmatrix} \tag{4.25}$$

for $0 \leqslant k < i$.

The proof will also be postponed to the appendix.

5. Analysis of the innovation estimation algorithm by Qin and Ljung

In this section we shall do some error analysis of the algorithm proposed by Qin and Ljung (2003) which is named by the authors "innovation estimation" as it is based on a preliminary step in which the future innovations are estimated from the observed data. Instead of using data Hankel matrices we shall work with stochastic column vectors, as this allows to simplify notations. We shall also disregard most computational details. Introduce the notations

$$\mathbf{y}_{[t,T]} := \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{y}(t+1) \\ \vdots \\ \mathbf{y}(T) \end{bmatrix}, \quad \mathbf{u}_{[t,T]} := \begin{bmatrix} \mathbf{u}(t) \\ \mathbf{u}(t+1) \\ \vdots \\ \mathbf{u}(T) \end{bmatrix},$$

$$\hat{\mathbf{e}}_{[t,T]} := \begin{bmatrix} \hat{\mathbf{e}}(t) \\ \hat{\mathbf{e}}(t+1) \\ \vdots \\ \hat{\mathbf{e}}(T) \end{bmatrix}, \quad \hat{\mathbf{v}}_{[t,T]} := \begin{bmatrix} \hat{\mathbf{v}}(t) \\ \hat{\mathbf{v}}(t+1) \\ \vdots \\ \hat{\mathbf{v}}(T) \end{bmatrix}.$$

As pointed out in Section 3 the algorithm can be seen as an implementation with finite data of the state-space construction formula of Theorem 3.2. The starting point is to construct the future transient innovations $\hat{\mathbf{e}}(t+i)$, $i = 0, \ldots, k$ using formula (4.3). One then forms the "extended future space" at time t, using data from time t_0 on, as

$$\hat{\mathscr{G}}_{[t,t+k]} := \mathrm{span}\{\hat{\mathbf{e}}(t+i), \mathbf{u}(t+i), i = 0, \ldots, k\} \tag{5.1}$$

and decomposes the vector $\mathbf{y}_{[t,T]}$ as a sum of two pieces belonging to $\hat{\mathscr{G}}_{[t,T]}$ and $\mathscr{Z}_{[t_0,t]}$

$$\mathbf{y}_{[t,T]} = E_{\|\hat{\mathscr{G}}_{[t,T]}}[\mathbf{y}_{[t,T]} \mid \mathscr{Z}_{[t_0,t]}] + E_{\|\mathscr{Z}_{[t_0,t]}}[\mathbf{y}_{[t,T]} \mid \hat{\mathscr{G}}_{[t,T]}]. \tag{5.2}$$

Let Γ be the observability matrix of the pair (A, C). In Qin and Ljung (2003) the first term on the right-hand side of (5.2) is used to implement the following approximation:

$$E_{\|\hat{\mathscr{G}}_{[t,T]}}[\mathbf{y}_{[t,T]} \mid \mathscr{Z}_{[t_0,t)}] \simeq \Gamma E_{\|\hat{\mathscr{G}}_{[t,T]}}[\mathbf{x}(t)|\mathscr{Z}_{[t_0,t)}] \simeq \Gamma E[\mathbf{x}(t) \mid \mathscr{Z}_{[t_0,t)}] = \Gamma\hat{\mathbf{x}}(t) \tag{5.3}$$

from which an estimate of the observability matrix Γ and hence of (A, C), can be obtained by standard methods.

Next, from the regression equation obtained after computing the second term

$$E_{\|\mathscr{Z}_{[t_0,t)}}[\mathbf{y}_{[t,T]} \mid \hat{\mathscr{G}}_{[t,T]}] \simeq H_{\mathbf{u}}\mathbf{u}_{[t,T]} + H_{\mathbf{e}}\hat{\mathbf{e}}_{[t,T]} \tag{5.4}$$

one could (approximately) estimate the block Toeplitz matrices $H_{\mathbf{u}}, H_{\mathbf{e}}$, containing the Markov parameters of $F(z)$ and of the noise model.

Alternatively, one could use regression (4.17), Proposition 4.1, to estimate $A, B, C, \hat{K}(t)$. Note that, unlike $\hat{\mathbf{v}}(t)$, $\hat{\mathbf{e}}(t)$ is correlated with the preceding terms in the regression and cannot be treated as "additive noise". Therefore one needs to include $\hat{\mathbf{e}}(t)$ as a regressor to estimate A and B, and hence estimate also $\hat{K}(t)$.

Remark 5.4. Note that this last step could also be implemented in a different way. For instance, after A, C have been computed from the estimated observability matrix, then $B, \hat{K}(t)$ could be estimated by linear regression from the Toeplitz matrices $H_{\mathbf{u}}, H_{\mathbf{e}}$ of the Markov parameters of the system obtained from regression (5.4). Incidentally, the estimated Toeplitz matrices will then, by construction, be lower triangular, i.e. "parsimonious" in the sense of Qin and Ljung (2003). Since $H_{\mathbf{u}}, H_{\mathbf{e}}$ are linear in B and $\hat{K}(t)$, using the estimates of A, C, the estimation of $B, \hat{K}(t)$ could be done by solving a linear regression. However in this paper we shall follow the strategy outlined above for it is also the one followed by the second algorithm which will be analyzed in the next section.

Unfortunately, as we shall show below, unless $\bar{A}$ and/or $\bar{F}$ are nilpotent, the algorithm yields biased estimates for finite $t - t_0$. In fact, from Eq. (4.8), one can see that the oblique projection $E_{\|\hat{\mathscr{G}}_{[t,T]}}[\mathbf{y}_{[t,T]} \mid \mathscr{Z}_{[t_0,t)}]$ is contained in the joint ($\mathbf{y}$ and $\mathbf{u}$) predictor space and not necessarily in the space spanned by $\hat{\mathbf{x}}(t)$. This fact is formalized in the following lemma, whose proof is deferred to the appendix.

Lemma 5.1. *Let* $v := T - t$ *and*

$$H_{\mathbf{v}}(t,T) := \begin{bmatrix} 0 & \dots & 0 \\ C\tilde{B}(t) & \dots & 0 \\ \vdots & \ddots & \vdots \\ CA^{v-2}\tilde{B}(t) & \dots & C\tilde{B}(T-1) \end{bmatrix}.$$

Then the oblique projection $E_{\|\hat{\mathscr{G}}_{[t,T]}}[\mathbf{y}_{[t,T]} \mid \mathscr{Z}_{[t_0,t]}]$ *can be split into two parts*:

$$E_{\|\hat{\mathscr{G}}_{[t,T]}}[\mathbf{y}_{[t,T]} \mid \mathscr{Z}_{[t_0,t]}] = \Gamma\hat{\mathbf{x}}(t) - H_{\mathbf{v}}(t,T)E_{\|\hat{\mathscr{G}}_{[t,T]}}[\hat{\mathbf{v}}_{[t,T-1]} \mid \mathscr{Z}_{[t_0,t)}] \tag{5.5}$$

the second of which originates bias.

Remark 5.5. From Lemma 4.3 and in particular Eq. (4.24) it is easy to see that the second term on the right-hand side belongs to $\text{span}\{\hat{\mathbf{x}}(t), \hat{\mathbf{s}}(t)\}$. Notice also that if $t - t_0$ is large enough, $\tilde{B}(t+i) = 0$ if at least one of the matrices $\bar{A}$, $\bar{F}$ is nilpotent. In this case $H_{\mathbf{v}}(t,T) = 0$ and the bias disappears.

6. Analysis of the "whitening filter"-based algorithm and relation to Jansson's method

In Section 3 we have introduced an algorithm based on the whitening filter to construct the state space and identify the system matrices (A, B, C).

The first step of the algorithm is to compute the oblique projections $E_{\|\mathscr{Z}_{[t,t+i)}}[\mathbf{y}(t+i) \mid \mathscr{Z}_{[t_0,t)}]$ and thereby construct a basis $\hat{\mathbf{x}}(t)$ for the state space. Unfortunately, also in this case a bias term is present since initial conditions are not properly handled. The bias is evaluated in the following lemma.

Lemma 6.1. *Let* $\Phi(t,s) := \prod_{k=t}^{s-1}\bar{A}(k)$ *where* $\Phi(t,t) = I$ *and*

$$\bar{\Gamma}(t,T) := [\,C^\top \quad \Phi^\top(t,t+1)C^\top \quad \dots \quad \Phi^\top(t,T-1)C^\top\,]^\top$$

be the "time varying" observability matrix associated to model (4.23). *Define also*

$$\bar{H}_{\mathbf{v}}(t,T) := \begin{bmatrix} 0 & \dots & 0 \\ C\tilde{B}(t) & \dots & 0 \\ \vdots & \ddots & \vdots \\ C\Phi(t,T-2)\tilde{B}(t) & \dots & C\tilde{B}(T-1) \end{bmatrix}.$$

Then the following equality holds:

$$\begin{bmatrix} E[\mathbf{y}_t \mid \mathscr{Z}_{[t_0,t)}] \\ E_{\|\mathscr{Z}_t}[\mathbf{y}_{t+1} \mid \mathscr{Z}_{[t_0,t)}] \\ \vdots \\ E_{\|\mathscr{Z}_{[t,T-1]}}[\mathbf{y}_T \mid \mathscr{Z}_{[t_0,t)}] \end{bmatrix} = \bar{\Gamma}(t,T)\hat{\mathbf{x}}(t) - \bar{H}_{\mathbf{v}}(t,T)E_{\|\mathscr{Z}_{[t,T-1]}}[\hat{\mathbf{v}}_{[t,T-1]} \mid \mathscr{Z}_{[t_0,t)}]. \tag{6.1}$$

The proof is deferred to the appendix.

Remark 6.6. From Lemma 4.3 and in particular Eq. (4.25), it is easy to see that the second term on the right-hand side belongs to span$\{\hat{\mathbf{x}}(t), \hat{\mathbf{s}}(t)\}$. Again, if $t - t_0$ is large enough and $\bar{F}$ is nilpotent, $\bar{B}(t)$ is zero and the second term on the right-hand side vanishes. Nevertheless, unless $\bar{A}$ is also nilpotent, $\bar{\Gamma}(t, T)$ is still time varying. This implies that the state space can be estimated without bias but there may be problems in constructing a basis at time $t + 1$, $\hat{\mathbf{x}}(t + 1)$, coherent with the one chosen at time t, $\hat{\mathbf{x}}(t)$. Equivalently the standard "shift-invariance" method applied to the estimated time-varying $\bar{\Gamma}(t, T)$ may lead to errors.

As we have mentioned earlier, an algorithm based on similar ideas has been proposed in Jansson (2003), where the oblique projection step is replaced by a preliminary estimation of the Markov parameters (which should ideally remove the contribution due to terms in $\mathscr{Z}_{[t,t+i)}$), followed by an orthogonal projection. Whether this approach is to be preferred to our implementation is still an open question, even though the simulations with simple systems that we report in this paper does not show remarkable differences between the two approaches; both the implementations have performances very close to standard PEM in the examples considered. A thorough analysis of these aspects will however be undertaken in future work.

7. Stable vs. unstable plants

The case of unstable plants deserves a particular attention since it brings into play the ill-conditioning of oblique projections. As it has already been noticed, when the A matrix has eigenvalues strictly outside of the unit circle, the spaces $\mathscr{Z}_t^-$ and $\mathscr{G}_t^+$ intersect, so that one may expect that with moderately large past and future horizons, $\mathscr{Z}_{[t_0,t)}$ and $\mathscr{G}_{[t,T]}$ will be almost collinear. The more so, the larger the $|\lambda_i(A)|$'s.

How serious is this phenomenon? In order to simplify the analysis we shall for a moment assume that $\hat{\mathscr{G}}_{[t,T]} \simeq \mathscr{G}_{[t,T]}$, which is reasonable for $t - t_0$ large enough. The following lemma helps understanding why near parallelism occurs.[10]

Lemma 7.1. *Assume A has eigenvalues outside of the closed unit disc. Let $\mathbf{x}_a(t) = T_a\mathbf{x}(t)$ be a basis for the "unstable manifold" of the state space*[11] *and let A_a, B_a, K_a be the matrices A, B, K restricted to that subspace of the state space. Then $\mathbf{x}_a(t)$ can be written in the following forward and backward forms*:

$$\mathbf{x}_a(t) = T_a\bar{A}^{t-t_0}\mathbf{x}(t_0) + \sum_{i=0}^{t-t_0-1} T_a\bar{A}^{t-t_0-i-1}(B\mathbf{u}(t_0+i) + K\mathbf{y}(t_0+i)), \tag{7.1}$$

$$\mathbf{x}_a(t) = A_a^{-(T-t+1)}\mathbf{x}_a(T+1) - \sum_{i=0}^{T-t} A_a^{-(T-t-i+1)}(B_a\mathbf{u}(T-i) + K_a\mathbf{e}(T-i)). \tag{7.2}$$

Eq. (7.1) shows that for $t-t_0$ large, $\mathbf{x}_a(t)$ "nearly" belongs to the past space $\mathscr{Z}_{[t_0,t)}$ since the eigenvalues of $\bar{A}$ in the innovation model are stable, and the first term dies out for $t - t_0$ large. But at the same time, for $T - t$ large, $\mathbf{x}_a(t)$ "nearly" belongs also to the extended future space $\mathscr{G}_{[t,T]}$. This is so since the first component in (7.2) dies out as $T - t \to \infty$, and the second one lies in $\mathscr{G}_{[t,T]}$. It is clear that near parallelism shows up rather quickly (as $t - t_0$ and $T - t$ grow) if the eigenvalues of A_a and $\bar{A}$ are far from the unit circle.

Now, it is well known that for near parallel subspaces the oblique projection may amplify errors in the data. In fact it is known (and easy to show) that the norm of the oblique projection is given by

$$\|E_{\|\hat{\mathscr{G}}_{[t,T]}}[\cdot \mid \mathscr{Z}_{[t_0,t)}]\| = \frac{1}{\sqrt{1-\cos^2(\theta_{\min}(\hat{\mathscr{G}}_{[t,T]}, \mathscr{Z}_{[t_0,t)}))}},$$

where $\theta_{\min}(\hat{\mathscr{G}}_{[t,T]}, \mathscr{Z}_{[t_0,t)})$ is the smallest canonical angle between $\hat{\mathscr{G}}_{[t,T]}$ and $\mathscr{Z}_{[t_0,t)}$.

The situation may become dramatic when working with real data. Finite-sample perturbations due to, e.g. approximations of inner products by finite sums, may get amplified by oblique projections and lead to a large variance of the estimates. Experimental evidence shows that indeed the variance of the estimates using the algorithm of Qin and Ljung with unstable plants is considerably larger as compared to the other approaches.

There is another aspect which has to do with instability and may be attributed to scaling problems. As $T-t$ becomes large, matrix coefficients involving powers of A (e.g. the observability matrix) may show unstable components which will grow very large while the stable components will instead become small and eventually die out. This fact is also seen when using the SVD in the truncation step, as some singular values grow fast when enlarging $T - t$ while some other decrease, making it difficult to perform order selection. The situation becomes serious if the system order is large and hence one is forced to make $T - t$ large. These scaling problems may become an issue working with finite-precision arithmetic. For the reasons explained above, the approach of Qin and Ljung (2003) is not to be recommended with unstable plants. In any case, particular care must be exercised in the choice of the "future horizon" $T - t$.

[10] Actually one should work with the transient future space $\hat{\mathscr{G}}_{[t,T]}$, but for simplcity of presentation we approximate it with the stationary version. A similar argument, but with more complicated notations holds for $\hat{\mathscr{G}}_{[t,T]}$.

[11] The subscript a stands for "acausal".

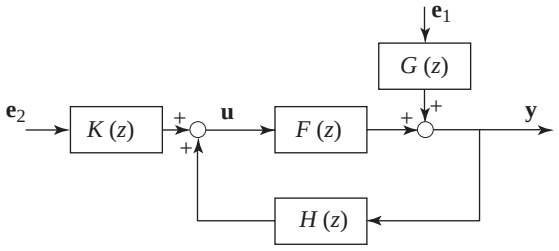

Fig. 1. Joint model of the signals **y** and **u**.

Fortunately, these issues do not affect the whitening filter approach and the state construction described in Theorem 3.4. Oblique projections (3.13) play of course a central role also here, but the space along which the projection if performed is now different. The spaces $\mathscr{Z}_{[t_0,t)}$ and $\mathscr{Z}_{[t,T)}$ may in some unfortunate circumstances become close (depending on the zeros of the joint spectrum getting close to the unit circle) but generally do not intersect in the limit, as $\mathscr{Z}_{[t_0,t)}$ and $\mathscr{G}_{[t,T]}$ instead do.

Table 1
Transfer functions for the examples considered

#	$F(z)$	$H(z)$	$G(z)$	$K(z)$
Ex. 1	$\frac{0.3}{z-0.7}$	-1	$\frac{z+0.5}{z}$	1
Ex. 2	$\frac{2.5}{z-3}$	-1	$\frac{z+0.5}{z}$	1
Ex. 3	$\frac{2.5}{z-3}$	-1	$\frac{z+0.999}{z}$	$\frac{0.2(z+0.999)}{z-0.99}$
Ex. 4	$\frac{2.5}{z-3}$	-1	$\frac{z+0.999}{z}$	1

8. Experiments

The simulation setup is as in Fig. 1 where $\mathbf{e}_1$ and $\mathbf{e}_2$ are uncorrelated, zero mean and unit variance white gaussian noises. The four examples are as in Table 1. The first is a stable plant, controlled by a proportional controller with white reference signal and moving average observation noise. The second is an unstable plant controlled by a proportional controller with white reference signal and moving average observation noise. The third and the fourth examples use the same plant and controller as the second but different driving

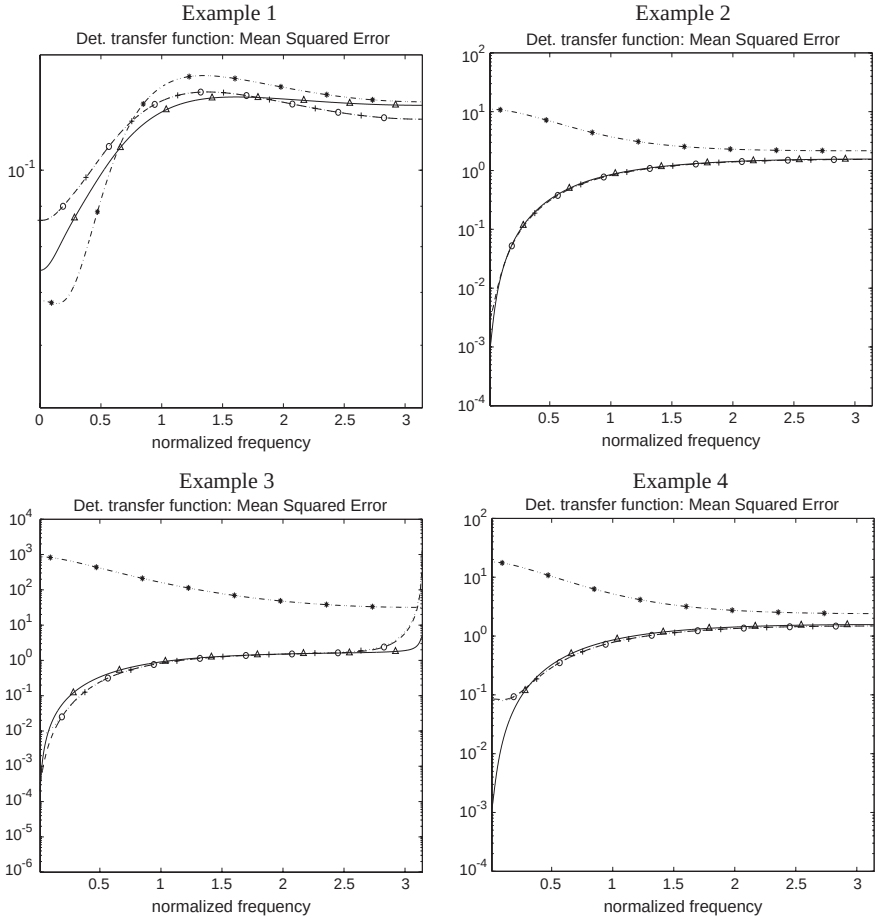

Fig. 2. Mean squared error (rms) of estimated transfer function $\hat{F}(e^{j\omega})$ vs. normalized frequency ($\omega \in [0, \pi]$). Crosses (+): Whitening filter method, stars (∗): Qin–Ljung "innovation estimation" method, circles (○): Jansson method, triangles (△): Matlab 6.5 PEM.

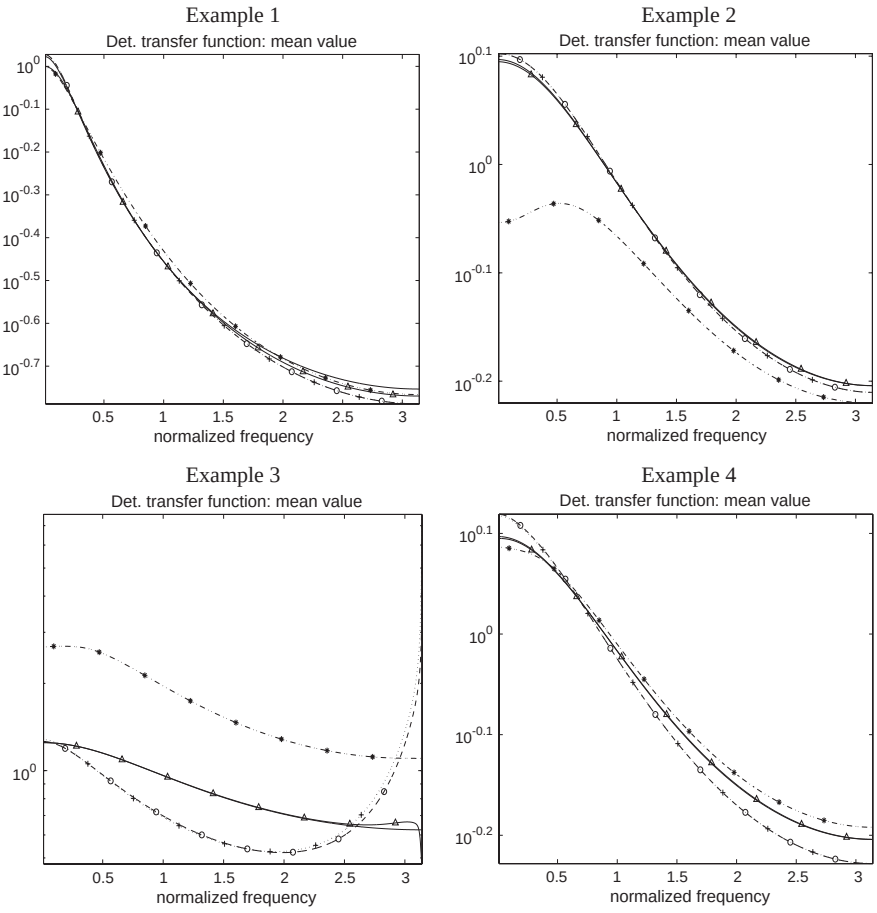

Fig. 3. Magnitude of $E[\hat{F}(e^{j\omega})]$ (Monte Carlo estimate) vs. normalized frequency ($\omega \in [0, \pi]$) Solid: True $|F(e^{j\omega})|$, crosses (+): Whitening filter method, stars (∗): Qin–Ljung "innovation estimation" method, circles (○): Jansson method, triangles (△): Matlab 6.5 PEM.

noises. Example 3 has colored reference signal and observation noise, where the zeros are placed close to the unit circle while in Example 4 the reference is white. Note that with a proportional controller a white reference signal implies that the eigenvalues of $\bar{F}$ are in zero so that the matrix is nilpotent. In all experiments the past and future horizons have been chosen to be 10 and 100 Monte-Carlo runs have been repeated with 1000 data points each. Fig. 2 and Fig. 3

The behavior of these simple examples is quite in accordance with the predictions one could have made, based on the analysis of this paper. In Example 1, where $F(z)$ is stable, the reference signal is white and the eigenvalues of $\bar{A}$ are far from the unit circle, the bias is negligible. The algorithm of Qin and Ljung behaves similarly to the one based on the whitening filter. In Example 2, where $F(z)$ is unstable, $\bar{F} = 0$ and $\lambda_i(\bar{A})$ far from the unit circle, bias is negligible for both Qin–Ljung and "whitening-filter". As far as variance, as predicted by our analysis, Qin–Ljung performs worst for the reasons explained in Section 7. In Example 3, where the closed loop zeros are very close to the unit circle, the bias becomes large for all algorithms while the mean squared error is much larger for the Qin–Ljung algorithm, in line with Example 2 and the discussion in Section 7. In Example 4 we have again $\bar{F} = 0$. The bias is considerably reduced and the mean squared error is significantly larger for the Qin–Ljung approach, for the same reasons as above.

In all the examples the algorithm based on the whitening filter is basically indistinguishable from the two step procedures proposed by Jansson and PEM has a very similar behavior.

9. Discussion and conclusions

In this paper we have discussed some theoretical issues related to subspace identification in the presence of feedback. We have shown that neglecting the effect of initial condition results in bias in the estimates which can be significant in case of (closed loop) zeros close to the unit circle.

We have analyzed two recent subspace algorithms, (Qin & Ljung, 2003; Jansson, 2003), and proposed a slight variation on the last one. According to the general results of this paper, all of these algorithms are generally biased. We have argued that identification of high order systems with unstable dynamics may become unfeasible due to numerical problems with the algorithm of Qin and Ljung (2003).

Different considerations hold for the other approach proposed in Jansson (2003) which seems not to suffer the same drawbacks.

The above conclusions are supported by experimental evidence. Simulations reported in the last section show that with unstable plants and $T-t$ large the "innovation estimation" approach fails, while the other methods provide fairly reliable estimates.

Although we have not been able to suggest procedures leading to unbiased algorithms, we hope nevertheless that the ideas and the system-theoretic background of this paper will be of some help to gain more understanding and to foster progress in this area.

Appendix A. Proofs

Proof of Lemma 4.2. Since $\mathscr{Z}_{[t_0,t+1)}=(\mathscr{Z}_{[t_0,t)}\oplus\hat{\mathscr{E}}_t)\oplus\hat{\mathscr{V}}_t$,

$$\begin{aligned}\tilde{\mathbf{x}}(t) &:= E[\mathbf{x}(t)\mid\mathscr{Z}_{[t_0,t+1)}]-E[\mathbf{x}(t)\mid\mathscr{Z}_{[t_0,t)}]\\ &= E[\mathbf{x}(t)\mid\hat{\mathscr{E}}_t]+E[\mathbf{x}(t)\mid\hat{\mathscr{V}}_t].\end{aligned}$$

From (4.2)

$$\mathbf{x}(t)=\bar{A}^{t-t_0}\mathbf{x}(t_0)+\text{“ terms in } \mathscr{Z}_{[t_0,t)}\text{”}$$

and since $\hat{\mathscr{V}}_t\perp\mathscr{Z}_{[t_0,t)}$, $\hat{\mathscr{E}}_t\perp\mathscr{Z}_{[t_0,t)}$

$$\tilde{\mathbf{x}}(t)=\bar{A}^{t_0-t}[E[\mathbf{x}(t_0)\mid\hat{\mathscr{E}}_t]+E[\mathbf{x}(t_0)\mid\hat{\mathscr{V}}_t]]$$

which is (4.20). Again, from (4.2) we get

$$\begin{aligned}\hat{\mathbf{e}}(t) &= \mathbf{y}(t)-CE[\mathbf{x}(t)|\mathscr{Z}_{[t_0,t)}]\\ &= \mathbf{y}(t)-C\mathbf{x}(t)+CE[\mathbf{x}(t)|\mathscr{Z}^{\perp}_{[t_0,t)}]\\ &= \mathbf{e}(t)+C\bar{A}^{t-t_0}E[\mathbf{x}(t_0)|\mathscr{Z}^{\perp}_{[t_0,t)}].\end{aligned}\tag{A.1}$$

From the well-known formula

$$E[\mathbf{x}(t_0)\mid\hat{\mathscr{E}}_t]=\text{Cov}\{\mathbf{x}(t_0),\hat{\mathbf{e}}(t)\}[\text{Var}\{\hat{\mathbf{e}}(t)\}]^{-1}\hat{\mathbf{e}}(t)$$

(4.21) follows noting that, from (A.1),

$$\text{Cov}\{\mathbf{x}(t_0),\hat{\mathbf{e}}(t)\}=\text{Var}\{\mathbf{x}(t_0)\mid\mathscr{Z}_{[t_0,t)}\}C^{\top}(\bar{A}^{\top})^{t-t_0}.$$

Observe that, using (A.1), $\Lambda_e(t):=\text{Var}\{\hat{\mathbf{e}}(t)\}$ can be given the following expression:

$$\Lambda_e(t)=\Lambda_e+C\bar{A}^{t-t_0}\text{Var}\{\mathbf{x}(t_0)\mid\mathscr{Z}_{[t_0,t)}\}C^{\top}(\bar{A}^{\top})^{t-t_0},$$

where $\Lambda_e:=\text{Var}\{\mathbf{e}(t)\}$ Eq. (4.22) follows from a similar argument involving $\hat{\mathbf{v}}(t)$. □

Proof of Lemma 4.3. It follows from (4.7) that

$$\hat{\mathbf{v}}(t)=\mathbf{u}(t)-\tilde{J}(t)C\hat{\mathbf{x}}(t)-H\hat{\mathbf{s}}(t)-J(t)\mathbf{y}(t)$$

and from (4.13),

$$\hat{\mathbf{q}}(t+k)=\prod_{h=t}^{t+k-1}\bar{A}_J(h)\hat{\mathbf{q}}(t)+\text{“ terms in } \mathscr{Z}_{[t,t+k)}\text{”}.$$

Combining the two last equations (4.25) follows, i.e.

$$\begin{aligned}&E_{\|\mathscr{Z}_{[t,t+i)}}[\hat{\mathbf{v}}(t+k)\mid\mathscr{Z}_{[t_0,t)}]\\ &\quad=[-\tilde{J}(t+k)C\quad -H]\prod_{h=t}^{t+k-1}\bar{A}_J(h)\begin{bmatrix}\hat{\mathbf{x}}(t)\\ \hat{\mathbf{s}}(t)\end{bmatrix}\end{aligned}$$

for $0\leqslant k<i$. Similarly (4.25) follows from (4.7) and (4.8). □

Proof of Lemma 5.1. From (4.17) it follows that

$$\mathbf{y}_{[t,T]}=\Gamma\hat{\mathbf{x}}(t)-H_{\mathbf{v}}\hat{\mathbf{v}}_{[t,T-1]}+\text{“ terms in } \hat{\mathscr{G}}_{[t,T]}\text{”}$$

so that (5.5) follows. □

Proof of Lemma 6.1. Denoting by $\bar{H}_{\mathbf{v},i+1}$ the i-th block row of $\bar{H}_{\mathbf{v}}$ the output can be written

$$\begin{aligned}E[\mathbf{y}(t+i)|\mathscr{Z}_{[t_0,t+i)}] &= C\Phi(t,t+i)\hat{\mathbf{x}}(t)-\bar{H}_{\mathbf{v},i+1}\hat{\mathbf{v}}_{[t,T-1]}\\ &\quad+\text{“ terms in } \mathscr{Z}_{[t,t+i)}\text{”}.\end{aligned}$$

Recalling that, from (4.25),

$$\begin{aligned}&E_{\|\mathscr{Z}_{[t,t+k)}}[\hat{\mathbf{v}}(t+k)\mid\mathscr{Z}_{[t_0,t)}]\\ &\quad=E_{\|\mathscr{Z}_{[t,t+i+1)}}[\hat{\mathbf{v}}(t+k)\mid\mathscr{Z}_{[t_0,t)}]\end{aligned}$$

for $0\leqslant k<i$, (6.1) immediately follows. □

References

Anderson, B. D. O., & Moore, J. B. (1979). *Optimal filtering*. Englewood cliffs, NJ: Prentice-Hall.

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

Chiuso, A., & Picci, G. (2002). *Geometry of oblique splitting, minimality and hankel operators. Lecture notes in control and Information Sciences*, Vol. 286 (pp. 85–124). Berlin: Springer.

Chiuso, A., & Picci, G. (2003). Subspace identification random processes with feedback. In *Proceedings of the IFAC international symposium on system identification (SYSID)*, Rotterdam.

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

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

Chou, C. T., & Verhaegen, M. (1999). Closed-loop identification using canonical correlation analysis. In *Proceedings of the European control conference*, Karlsruhe (pp. CD-ROM, F162).

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

Gevers, M. R., & Anderson, B. D. O. (1982). On jointly stationary feedback-free stochastic processes. *IEEE Transactions on Automatic Control*, *27*, 431–436.

Granger, C. W. J. (1963). Economic processes involving feedback. *Information and Control*, *6*, 28–48.

Hannan, E. J., & Deistler, M. (1988). *The statistical theory of linear systems*. New York: Wiley.

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

Jansson, M. (2003). Subspace identification and arx modelling. In *Proceedings of SYSID 2003*, Rotterdam.

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

Lindquist, A., & Picci, G. (1996b). Geometric methods for state-space identification. In: S. Bittanti, & G. Picci (Eds.), *Identification, adaptation, learning* (pp. 1–69). Springer Verlag.

Ljung, L., & McKelvey, T. (1996). Subspace identification from closed loop data. *Signal Processing*, *52*(2), 209–216.

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

Picci, G. (1997). Oblique splitting susbspaces and stochastic realization with inputs. In: D. Prätzel-Wolters, U. Helmke, & E. Zerz (Eds.), *Operators, systems and linear algebra* (pp. 157–174). Stuttgart: Teubner.

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

Qin, S. J., & Ljung, L. (2003). Closed-loop subspace identification with innovation estimation. In *Proceedings of SYSID 2003*. Rotterdam.

Van der Klauw, A. C., Verhaegen, M., & Van den Bosch, P. P. J. (1991). State space identification of closed loop systems. In *Proceedings of the 30th IEEE conference on decision & control*, Brighton, UK (pp. 1327–1332).

Van Overschee, P., & De Moor, B. (1997). Closed loop subspace systems identification. In *Proceedings of 36th IEEE conference on decision & control*, San Diego, CA (pp. 1848–1853).

Verhaegen, M. (1993). Application of a subspace model identification technique to identify lti systems operating in closed-loop. *Automatica*, *29*, 1027–1040.

Giorgio Picci 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 minestery 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.

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 Laboaratory (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. Alessandro Chiuso is a member of the IEEE Control System Society Conference Editorial Board. His research interests are mainly in Identification and Estimation Theory, System Theory and Computer Vision.