Available online at www.sciencedirect.com
SCIENCE DIRECT®
Automatica 40 (2004) 677–683
automatica
www.elsevier.com/locate/automatica

Brief paper

Numerical conditioning and asymptotic variance of subspace estimates☆

Alessandro Chiuso*, Giorgio Picci

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

Received 18 November 2003; accepted 21 November 2003

Abstract

New formulas for the asymptotic variance of the parameter estimates in subspace identification, show that the accuracy of the parameter estimates depends on certain indices of 'near collinearity' of the state and future input subspaces of the system to be identified. This complements the numerical conditioning analysis of subspace methods presented in the companion paper (On the ill-conditioning of subspace identification with inputs, Automatica, doi:10.1016/j.automatica.2003.11.009).

Keywords: Asymptotic variance; Subspace identification; Exogenous inputs; Numerical conditioning; Collinearity; Oblique projections; State-space identification

1. Introduction

The asymptotic properties of the parameter estimates in subspace identification with inputs have been studied recently in a number of contributions, which include Bauer and Jansson (2000), Bauer and Ljung (2001), Chiuso and Picci (2004a) and Jansson (2000).

In this paper we shall discuss some new asymptotic variance formulas for the estimated parameters (A,B,C,D) of a stationary linear system with observable exogenous inputs $\mathbf{u}$. The system is assumed in "innovation representation"

$$\begin{cases} \mathbf{x}(t+1) = A\mathbf{x}(t) + B\mathbf{u}(t) + K\mathbf{e}(t), \\ \mathbf{y}(t) = C\mathbf{x}(t) + D\mathbf{u}(t) + \mathbf{e}(t), \end{cases} \tag{1.1}$$

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 $t-1$. The parameter estimates are computed by standard subspace methods (N4SID, MOESP, etc.) taken from the literature. Some of these variance formulas have been introduced earlier in Chiuso and Picci (2004a), but the relation of asymptotic variance to ill-conditioning of the identification problem is discussed here for the first time. It will be shown that the asymptotic variances depend on (the inverse of) certain conditional cross-covariance matrices, $\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}$, and $\Sigma_{\mathbf{u}^+\mathbf{u}^+|\hat{\mathbf{x}}}$ of the state, given the future inputs, and of the future inputs given the current state. After suitable normalization, the singular values of these matrices describe the numerical conditioning of the system parameter estimation problem in a wide variety of subspace identification methods. In Chiuso and Picci (2004b) we have also discussed certain indices which can be extracted from these covariance matrices, which may be used to assess the degree of collinearity of the regressors. Collinearity is defined as a geometric condition of near-parallelism of the regressors which entails ill-conditioning of the subspace identification problem.

Loose observations relating ill-conditioning to the accuracy or "performance" of identification methods have been circulating for a while in the identification community but, to our best knowledge, a precise relation between the two concepts seems to have never been pinpointed in an explicit way. Here the numerical conditioning of the subspace identification problem is related quite explicitly to the asymptotic variance of the estimates.

☆ This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Brett Ninness under the direction of Editor Torsten Söderström. This work has been supported in part by the ERNSI TMR network *System Identification* and by the national project *Identification and adaptive control of industrial systems* funded by MIUR. Part of this work has been done while the first author (A.C.) was a post-doctoral researcher with the Division of Optimization and Systems Theory, KTH, Stockholm, supported by ERNSI.

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

0005-1098/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2003.11.008

1.1. Notations

Boldface symbols will denote random quantities. For $-\infty \leqslant t_0 \leqslant t \leqslant T \leqslant +\infty$ define the Hilbert spaces of zero-mean (square integrable) random variables

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

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

where the bar denotes closure in mean square, i.e. in the metric defined by the inner product

$$\langle \boldsymbol{\xi}, \boldsymbol{\eta} \rangle := E\{\boldsymbol{\xi}, \boldsymbol{\eta}\} \tag{1.2}$$

the operator E denoting mathematical expectation. Similarly, let $\mathscr{U}_{[t,T]}, \mathscr{Y}_{[t,T]}$ be the respective future spaces up to time T

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

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

By convention the past spaces do not include the present. When $t_0 = -\infty$ we shall use the shorthands $\mathscr{U}_t^-, \mathscr{Y}_t^-$ to denote the Hilbert spaces $\mathscr{U}_{[-\infty,\,t)}, \mathscr{Y}_{[-\infty,\,t)}$. of random variables spanned by the infinite past of $\mathbf{u}$ and $\mathbf{y}$ up to time t while for the spaces generated by the whole time history of $\mathbf{u}$ and $\mathbf{y}$ we shall use the symbols $\mathscr{U}$, $\mathscr{Y}$. The *joint past* space of the input and output processes is denoted $\mathscr{P}_{[t_0,t)} := \mathscr{U}_{[t_0,t)} \vee \mathscr{Y}_{[t_0,t)}$, the $\vee$ denoting closed vector sum. The whole of $\mathscr{H} := \mathscr{U} \vee \mathscr{Y}$ will be taken as the *ambient space*, where all random quantities considered hereafter are assumed to live.

All through this paper we shall assume that there is *no feedback from* $\mathbf{y}$ *to* $\mathbf{u}$ which implies that $\mathbf{u}$ and $\mathbf{e}$ are uncorrelated, and that the input process is "sufficiently rich", in the sense that $\mathscr{U}_{[t_0,T]}$ admits the direct sum decomposition

$$\mathscr{U}_{[t_0,T]} = \mathscr{U}_{[t_0,t)} + \mathscr{U}_{[t,T]}, \quad t_0 \leqslant t < T, \tag{1.3}$$

the $+$ sign denoting direct sum of subspaces. The symbol $\oplus$ will be reserved for *orthogonal* direct sum. Condition (1.3) can be found expressed in various equivalent forms in the literature (see e.g. Hannan & Poskitt, 1988; Verhaegen & Dewilde, 1992, formula (10)).

The symbol $E[\,\cdot\,|\mathscr{A}]$ will denote (wide sense) conditional expectation, i.e. orthogonal projection onto the subspace $\mathscr{A} \subseteq \mathscr{H}$, orthogonality being with respect to the inner product (1.2).

Concerning system (1.1), it is well-known that the vector $\mathbf{y}_t^+$, of future stacked outputs from time t to T can be represented by

$$\mathbf{y}_t^+ = \Gamma \mathbf{x}(t) + H_{\text{d}} \mathbf{u}_t^+ + H_{\text{s}} \mathbf{e}_t^+, \tag{1.4}$$

where Γ denotes the observability matrix and by H_{d} and H_{s} the lower triangular block-Toeplitz matrices made with the Markov parameters of the deterministic subsystem (A,B,C,D) and of the stochastic subsystem (A,B,K,I). A bar over the various symbols will denote the same vector or matrix "augmented" so as to correspond to a vector $\bar{\mathbf{y}}_t^+$ with one extra block (namely $\mathbf{y}(T+1)$) appended at the bottom.

Although we shall try to make this paper reasonably self-contained, a thorough understanding of the concepts involved will require the background material exposed in Chiuso and Picci (2004b). We shall have to refer the reader to this paper also for a detailed explanation of some notations which will be used in the following.

2. Conditioning and asymptotic variances of the *A, C* estimates

In Chiuso and Picci (2004b) various subspace algorithms have been recasted into a common framework using ideas from stochastic realization theory, a useful result of this effort being that, at least for the estimation of the (A,C) parameters, the N4SID method, the "Robust" N4SID and PO-MOESP methods (and also CCA) can be dealt with in a unified manner. What distinguishes these methods is essentially the state construction step. Assuming the order estimation step is statistically consistent (i.e. the true order is eventually obtained when the sample size tends to infinity), the identification procedure is defined (asymptotically) by assigning a so-called *complementary state vector* obtained by projecting the transient Kalman filter state $\hat{\mathbf{x}}(t)$ onto the orthogonal complement of the future input space

$$\hat{\mathbf{x}}^{\text{c}}(t) := E[\hat{\mathbf{x}}(t)|\mathscr{U}_{[t,T]}^{\perp}] = (\hat{\mathbf{x}}(t) - E[\hat{\mathbf{x}}(t)|\mathscr{U}_{[t,T)}]) \tag{2.1}$$

whose covariance matrix is

$$\Sigma_{\hat{\mathbf{x}}^{\text{c}}\hat{\mathbf{x}}^{\text{c}}} := E\{\hat{\mathbf{x}}^{\text{c}}(t)\hat{\mathbf{x}}^{\text{c}}(t)^{\top}\} = \Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}. \tag{2.2}$$

Introduce the Cholesky factors, $L_{\hat{\mathbf{x}}}$ of $\Sigma_{\hat{\mathbf{x}}\ \hat{\mathbf{x}}}$ and $L_{\mathbf{u}^+}$ of $\Sigma_{\mathbf{u}^+\mathbf{u}^+}$. Using a well-known formula for the conditional covariances, we have

$$\begin{aligned}\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+} &= L_{\hat{\mathbf{x}}}[I - \Pi_{\hat{\mathbf{x}}\mathbf{u}^+}\Pi_{\hat{\mathbf{x}}\mathbf{u}^+}^{\top}]L_{\hat{\mathbf{x}}}^{\top}\\ \Sigma_{\mathbf{u}^+\mathbf{u}^+|\hat{\mathbf{x}}} &= L_{\mathbf{u}^+}[I - \Pi_{\mathbf{u}^+\hat{\mathbf{x}}}\Pi_{\mathbf{u}^+\hat{\mathbf{x}}}^{\top}]L_{\mathbf{u}^+}^{\top},\end{aligned} \tag{2.3}$$

where $\Pi_{\hat{\mathbf{x}}\mathbf{u}^+}$ is the normalized cross-covariance

$$\Pi_{\hat{\mathbf{x}}\mathbf{u}^+} := L_{\mathbf{x}}^{-1}\Sigma_{\mathbf{x}\mathbf{u}^+}L_{\mathbf{u}^+}^{-\top} = \Pi_{\mathbf{u}^+\hat{\mathbf{x}}}^{\top}$$

whose singular values (bounded by one in magnitude) are the well-known *canonical correlation coefficients* of the present state space, spanned by the random vector $\hat{\mathbf{x}}(t)$, and future input space $\mathscr{U}_{[t,T]}$. The *index of collinearity* of the identification problem is the maximal singular value of $\Pi_{\hat{\mathbf{x}}\mathbf{u}^+}$. Clearly $\sigma_{\text{Max}}(\Pi_{\hat{\mathbf{x}}\mathbf{u}^+}) \simeq 1 \Leftrightarrow \Sigma_{\hat{\mathbf{x}}^{\text{c}}\hat{\mathbf{x}}^{\text{c}}}$ is nearly singular. In fact, if the covariance matrix of the state at time t is normalized to the identity i.e. we choose a basis in the model in such a way that $\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}} = I$, we have exactly

$$\Sigma_{\hat{\mathbf{x}}^{\text{c}}\hat{\mathbf{x}}^{\text{c}}} = I - \Pi_{\hat{\mathbf{x}}\mathbf{u}^+}\Pi_{\hat{\mathbf{x}}\mathbf{u}^+}^{\top}. \tag{2.4}$$

The Gaussian distribution with mean μ and covariance matrix Σ is denoted $\mathscr{N}(\mu, \Sigma)$. If a sequence of random vectors $\{\mathbf{z}_N\}$ converges almost surely to a constant z_0 and is *asymptotically normal*, i.e. $\sqrt{N}(\mathbf{z}_N - z_0) \xrightarrow{\text{d}} \mathscr{N}(0, \Sigma)$, where $\xrightarrow{\text{d}}$ denotes convergence in distribution, one says that

Σ is the *asymptotic variance* of $\{\sqrt{N}\,\mathbf{z}_N\}$. Notation: $\Sigma = \mathrm{AsVar}(\sqrt{N}\,\mathbf{z}_N)$. The *asymptotic covariance* of two, asymptotically jointly Gaussian, sequences is defined in a similar way.

In order to guarantee the existence of the asymptotic variances, we shall assume that in model (1.1) of the true system generating the data, the innovation process $\{\mathbf{e}(t)\}$ is a martingale difference with respect to the σ-algebra $\mathscr{E}_t \vee \mathscr{U}$ generated by the random variables $\{\mathbf{e}(s);\ s < t\}$ and $\{\mathbf{u}(t);\ t \in \mathbb{Z}\}$, more precisely, assume for $j,k \geqslant 0$, that

$$E\{\mathbf{e}(t+k)\,|\,\mathscr{E}_t \vee \mathscr{U}\} = 0, \quad k \geqslant 0, \tag{2.5a}$$

$$E\{\mathbf{e}(t+j)\mathbf{e}(t+k)^\top\,|\,\mathscr{E}_t \vee \mathscr{U}\} = E\{\mathbf{e}(t+j)\mathbf{e}(t+k)^\top\} = \Lambda\delta_{jk} \tag{2.5b}$$

for a positive definite matrix Λ. We shall also need boundedness of the fourth moment of $\{\mathbf{e}(t)\}$ and $\{\mathbf{u}(t)\}$. These "*noise conditions*" are often found in the statistical literature (see e.g. Hannan & Deistler, 1988); they hold, for example, if $\{\mathbf{e}(t)\}$ is a i.i.d. process (strict sense white noise) with finite fourth-order moments, independent of $\mathbf{u}$, or if $\{\mathbf{e}(t)\}$ is Gaussian, independent of $\mathbf{u}$. In the first situation we shall also assume that the observed joint input–output process $[\mathbf{y},\ \mathbf{u}]$ is ergodic. For Gaussian processes, second-order ergodicity suffices since it is the same as ergodicity.

Theorem 1. *Assume that the stationary innovation process, $\{\mathbf{e}(t)\}$, in model (1.1) of the true system generating the data, satisfies the noise conditions described above. Then the vectorized parameter estimates with a sample consisting of N data, $[\mathrm{vec}(\hat{A}_N)^\top \mathrm{vec}(\hat{C}_N)^\top]^\top$, form an asymptotically Gaussian sequence with*

$$\mathrm{AsVar}(\sqrt{N}\mathrm{vec}(\hat{A}_N)) = \bar{F}\left\{\sum_{|\tau|\leqslant \nu+1} \Sigma_{\hat{\mathbf{x}}^c\hat{\mathbf{x}}^c}(\tau) \otimes \Sigma_{\bar{\mathbf{e}}^+\bar{\mathbf{e}}^+}(\tau)\right\}\bar{F}^\top, \tag{2.6}$$

$$\mathrm{AsVar}(\sqrt{N}\mathrm{vec}(\hat{C}_N)) = F\left\{\sum_{|\tau|\leqslant \nu} \Sigma_{\hat{\mathbf{x}}^c\hat{\mathbf{x}}^c}(\tau) \otimes \Sigma_{\mathbf{e}^+\mathbf{e}^+}(\tau)\right\}F^\top, \tag{2.7}$$

$$\mathrm{AsCov}(\sqrt{N}\mathrm{vec}(\hat{A}_N), \sqrt{N}\mathrm{vec}(\hat{C}_N)) = \bar{F}\left\{\sum_{\tau=-\nu-1}^{\tau=\nu} \Sigma_{\hat{\mathbf{x}}^c\hat{\mathbf{x}}^c}(\tau) \otimes \Sigma_{\bar{\mathbf{e}}^+\mathbf{e}^+}(\tau)\right\}F^\top, \tag{2.8}$$

where, setting

$$M := [(K\ \ \Gamma^\dagger) - A(\Gamma^\dagger\ \ 0_{n\times m})],$$
$$R := [(I_m\ \ 0_{m\times m(\nu-1)}) - C\Gamma^\dagger], \tag{2.9}$$

$\Gamma^\dagger \in \mathbb{R}^{n\times m(\nu+1)}$ being a left-inverse [1] of the observability matrix Γ of the system, the matrices $F, \bar{F}$ are defined by

$$F := \Sigma^{-1}_{\hat{\mathbf{x}}^c\hat{\mathbf{x}}^c} \otimes [RH_s], \quad \bar{F} := \Sigma^{-1}_{\hat{\mathbf{x}}^c\hat{\mathbf{x}}^c} \otimes [M\bar{H}_s]. \tag{2.10}$$

Further,

$$\Sigma_{\hat{\mathbf{x}}^c\hat{\mathbf{x}}^c}(\tau) := E\{\hat{\mathbf{x}}^c_\tau(t)\hat{\mathbf{x}}^c(t)^\top\},$$
$$\Sigma_{\mathbf{e}^+\mathbf{e}^+}(\tau) = E\{\mathbf{e}^+_{t+\tau}(\mathbf{e}^+_t)^\top\} \tag{2.11}$$

$\hat{\mathbf{x}}^c_\tau(t)$ being the τ-steps ahead stationary shift of the processes $\hat{\mathbf{x}}^c(t)$, namely

$$\hat{\mathbf{x}}^c_\tau(t) := E[\mathbf{x}(t+\tau)|\mathscr{U}^\perp_{[t+\tau,T+\tau+1]}], \tag{2.12}$$

$\mathscr{U}^\perp_{[t+\tau,T+\tau+1]}$ being the orthogonal complement of $\mathscr{U}_{[t+\tau,T+\tau+1]}$ in $(\mathscr{P}_{[t_0+\tau,t+\tau)} \vee \mathscr{U}_{[t+\tau,T+\tau+1]})$.

Comments on the proof of Theorem 1. The statement is essentially the same as Theorem 4.1 in Chiuso and Picci, 2004a). We have only introduced a slight modification in the definition of the time-updated estimate, $\hat{\mathbf{x}}^c(t+1)$ (and of $\bar{\mathbf{x}}(t+1)$), which we both define using a future horizon of $\nu := T - t$ data points. [2] To get the formulas right, we just need to assume that our reference future interval is $[t, T+1]$ and use the extra data point to this purpose, substituting $T+1$ in place of T wherever needed.

Formulas (2.6)–(2.8) are valid for a variety of estimation methods, including also CCA, provided the complementary state $\hat{\mathbf{x}}^c$ is properly defined.

From (2.6), (2.7) one can see that the inverse of the conditional covariance, $\Sigma^{-1}_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}$, determines the magnitude of the variance of the estimation errors. This is even more visible if we "normalize" the system parameters by fixing an orthonormal basis $\hat{\mathbf{x}}(t)$. In this case, by (2.4), we see that the asymptotic covariances are roughly "proportional" to the inverse of the matrix $I - \Pi_{\hat{\mathbf{x}}\mathbf{u}^+}\Pi^\top_{\hat{\mathbf{x}}\mathbf{u}^+}$. In particular, in the presence of near collinearity of the regressors (see, Chiuso & Picci, 2004b), $\sigma_{\min}(\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}) = \sigma_{\min}(I - \hat{\Pi}\hat{\Pi}^\top) \simeq 0$, and the variance of the estimation errors will explode.

2.1. Asymptotic variance of the N4SID estimator of B, D

There exist a plethora of methods in the literature for estimation of the (B, D) parameters and no unified analysis seems to be possible. We have chosen to analyze one, approximately linear, estimation scheme of (B, D), due to Van Overschee and De Moor (1994). The asymptotic (conditional) variance of this estimate can be written down explicitly and could therefore make a natural term of comparison to study the influence of ill conditioning on the statistical accuracy of estimation of (B, D).

A different estimator which slightly generalizes the "linear regression" estimator of Verhaegen and Dewilde (1992),

[1] See Chiuso and Picci (2004a), formula (4.19).

[2] The reason for introducing this modification is explained in Chiuso and Picci (2004a), Remark 4.1.

which also seems to be one of the most widely used, is analyzed in the paper by Chiuso and Picci (2004a).

We shall now quickly review the estimation of (B,D) as proposed in the paper Van Overschee and De Moor (1994). The first step is to estimate, by ordinary least squares, the matrices $\mathscr{K}_1, \mathscr{K}_2$ from a linear regression of the form

$$\begin{bmatrix} \bar{X}_{t+1} \\ Y_t \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} \bar{X}_t, + \begin{bmatrix} \mathscr{K}_1 \\ \mathscr{K}_2 \end{bmatrix} U_{[t,T+1]} + (\text{Error}),$$

where the finite-sample size "pseudostates" $\bar{X}_t, \bar{X}_{t+1}$ are defined in Appendix A, cf. (A.4). The estimate of the observability matrix, $\hat{\Gamma}$ used to form $\bar{X}_t$ may be taken either as proposed in the classical N4SID procedure or as in the "robust" version recalled in Section 4 of Chiuso and Picci (2004b).

The asymptotic variance of these estimates can be computed explicitly and are given in the following Theorem. The proof is very similar to that of Theorem 1; details can be found in the appendix.

Theorem 2. *Assume that the stationary innovation process,* $\{\mathbf{e}(t)\}$, *in model* (1.1) *of the true system generating the data, satisfies the noise conditions. Then the vectorized parameter estimates* $\text{vec}(\hat{\mathscr{K}}_{1,N})$ *and* $\text{vec}(\mathscr{K}_{2,N})$ *form an asymptotically Gaussian sequence with*

$$\text{AsVar}(\sqrt{N}\text{vec}(\hat{\mathscr{K}}_{1,N})) = \bar{G}\left\{\sum_{|\tau|\leqslant v+1} \Sigma_{\bar{\mathbf{u}}^+\bar{\mathbf{u}}^+|\bar{\mathbf{x}}}(\tau) \otimes \Sigma_{\bar{\mathbf{e}}^+\bar{\mathbf{e}}^+}(\tau)\right\}\bar{G}^\top, \tag{2.13}$$

$$\text{AsVar}(\sqrt{N}\text{vec}(\hat{\mathscr{K}}_{2,N})) = G\left\{\sum_{|\tau|\leqslant v} \Sigma_{\bar{\mathbf{u}}^+\bar{\mathbf{u}}^+|\bar{\mathbf{x}}}(\tau) \otimes \Sigma_{\mathbf{e}^+\mathbf{e}^+}(\tau)\right\}G^\top, \tag{2.14}$$

$$\text{AsCov}(\sqrt{N}\text{vec}(\hat{\mathscr{K}}_{1,N}), \sqrt{N}\text{vec}(\hat{\mathscr{K}}_{2,N})) = \bar{G}\left\{\sum_{\tau=-v-1}^{\tau=v} \Sigma_{\bar{\mathbf{u}}^+\bar{\mathbf{u}}^+|\bar{\mathbf{x}}}(\tau) \otimes \Sigma_{\bar{\mathbf{e}}^+\mathbf{e}^+}(\tau)\right\}G^\top, \tag{2.15}$$

where $\bar{\mathbf{u}}^+$ *stands for the random vector made of* $v+2$ *stacked input values* $\mathbf{u}(s)$ *with* $t \leqslant s \leqslant T+1$, *and*

$$G := \Sigma^{-1}_{\bar{\mathbf{u}}^+\bar{\mathbf{u}}^+\,|\,\bar{\mathbf{x}}} \otimes [RH_s], \quad \bar{G} := \Sigma^{-1}_{\bar{\mathbf{u}}^+\bar{\mathbf{u}}^+|\bar{\mathbf{x}}} \otimes [M\bar{H}_s]$$

R and M being as in (2.9), *and,*

$$\Sigma_{\bar{\mathbf{u}}^+\bar{\mathbf{u}}^+|\bar{\mathbf{x}}}(\tau) := E\{\tilde{\bar{\mathbf{u}}}^+_{t+\tau}(\tilde{\bar{\mathbf{u}}}^+_t)^\top\},$$

$$\Sigma_{\mathbf{e}^+\mathbf{e}^+}(\tau) = E\{\mathbf{e}^+_{t+\tau}(\mathbf{e}^+_t)^\top\}, \tag{2.16}$$

$\tilde{\bar{\mathbf{u}}}^+_{t+\tau}$ *being the* τ*-steps ahead stationary shift of the random vector* $\tilde{\bar{\mathbf{u}}}^+_t := \bar{\mathbf{u}}^+_t - E[\bar{\mathbf{u}}^+_t \mid \bar{\mathbf{x}}(t)]$.

Since $(\mathscr{K}_1, \mathscr{K}_2)$ are known functions of the parameters of the stationary system (Van Overschee & De Moor, 1994, formula (44)), in particular are *linear functions of the parameters* B,D, one may write in vectorized form

$$\text{vec}(\mathscr{K}_1) = L_1(A,C)\text{vec}\begin{pmatrix} B \\ D \end{pmatrix},$$

$$\text{vec}(\mathscr{K}_2) = L_2(A,C)\text{vec}\begin{pmatrix} B \\ D \end{pmatrix}, \tag{2.17}$$

where $L := [L_1(A,C), L_2(A,C)]$ is a known matrix function of A, C. The estimator of (B,D) is based on expressions similar to the above and is normally implemented by "linear regression" *after A and C have been estimated in a preceding step*. Of course, to compute the variance of the estimates in principle one should treat these A and C as sample values of random variables. For the scope of this paper however we shall just consider the estimates and the relative variance expressions which will reported below as *conditional, given the observed value of* $\hat{A}, \hat{C}$. The full asymptotic covariances, taking care of the randomness of A and C can be computed, at the price of some complications (cf. Chiuso & Picci, 2004a; Jansson, 2000).

Denote by Σ_K the joint asymptotic covariance matrix of $(\mathscr{K}_1, \mathscr{K}_2)$. Then by a standard formula the (conditional) variance of the estimates of the B,D parameters follows:

$$\text{AsVar}\left\{\sqrt{N}\text{vec}\begin{pmatrix} \hat{B}_N \\ \hat{D}_N \end{pmatrix}\right\} = (L^\top L)^{-1}L^\top \Sigma_K L (L^\top L)^{-1} \tag{2.18}$$

(the inverses must exist if the parametrization is identifiable).

The asymptotic variance is then seen to be roughly "proportional" to Σ_K, and hence, looking at the expressions (2.13)–(2.15), the important role in the analysis is now played by the smallest singular value of the conditional covariance matrix $\Sigma_{\bar{\mathbf{u}}^+\bar{\mathbf{u}}^+\,|\,\bar{\mathbf{x}}}$. This of course is influenced both by the collinearity of the subspaces generated by the pseudostate $\bar{\mathbf{x}}(t)$ and the future $\bar{\mathbf{u}}^+_t$, *and* by the possible near singularity of the matrix $\Sigma_{\bar{\mathbf{u}}^+\bar{\mathbf{u}}^+}$ which in turn has to do with persistence of excitation and with the canonical correlation structure of the input process. One can see here that the situation could get worse than for the estimates of A and C.

3. Some experimental results and conclusions

To give an idea of possible consequences of collinearity we present some simulations made on a very simple system. The system and input spectra (frequency-domain data) are shown in Fig. 1 below.

The input is a colored ARMA process with roughly the same bandwidth of the deterministic transfer function to be identified. The deterministic transfer function and the stochastic shaping filter have disjoint dynamics. The (power) signal-to-noise ratio is of the order of 10. The data length for this experiment is 500.

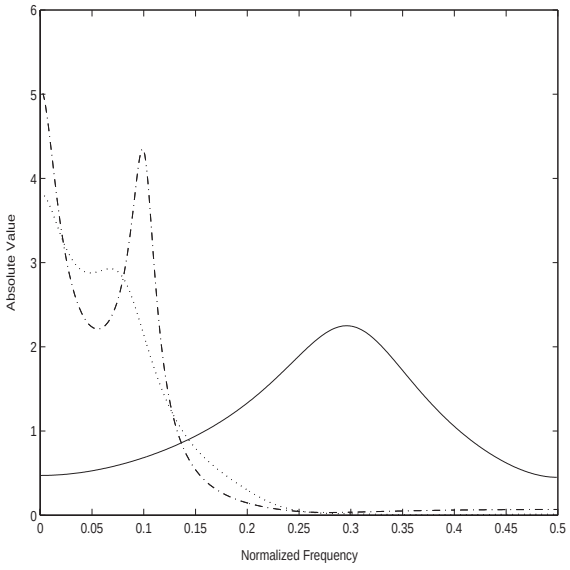

Fig. 1. True model. Solid: square root of stochastic component spectrum; dotted: square root of input spectrum; dashed: absolute value of deterministic transfer function.

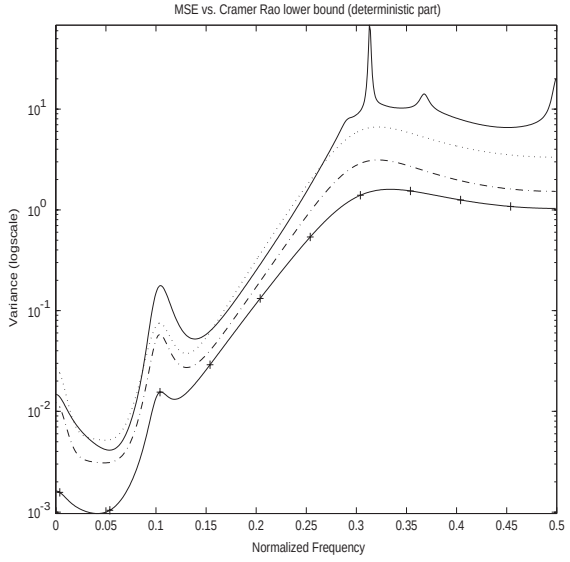

Fig. 2. Asymptotic variance (Monte Carlo estimate) of estimated transfer function versus normalized frequency. Solid : Matlab 5 N4SID (with refinement for B and D), dotted: MOESP, dashed: robust N4SID; solid with crosses: Cramer–Rao bound.

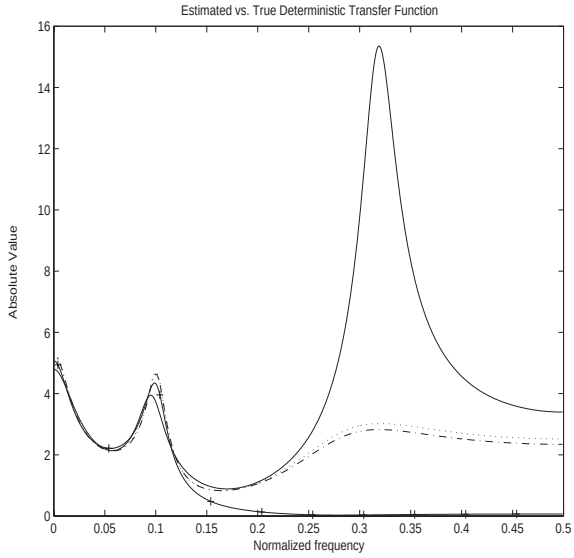

Fig. 3. Typical estimates of the deterministic transfer function versus normalized frequency. Solid : Matlab 5 N4SID (with refinement for B and D), dotted: PO-MOESP, dashed: robust N4SID; solid with crosses: Cramer–Rao bound.

Table 1
Poles, Zeros and Gains for the stochastic, deterministic subsystems and input

	Poles	Zeros	K
Stoch. system	$-0.2 + j0.6$	0.5	1
	$-0.2 - j0.6$	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	$0.7 + j0.4$	$-0.6 + j0.6$	0.1
	$0.7 - j0.4$	$-0.6 - j0.6$	
	0.85	0.7	
	$0.2 + j0.7$	$-0.1 + j0.8$	
	$0.2 - j0.7$	$-0.1 - j0.8$	

In Figs. 2 and 3 we compare the results of subspace identification of the (deterministic) transfer function of a simple third order scalar system. The details of the simulations are reported in Table 1.

The plots shown in Fig. 2 reports the results of the estimated mean squared error using 100 Monte-Carlo runs, compared with Cramèr–Rao lower bound.

- the standard N4SID method (Matlab 5.3 implementation with refinement of the B, D estimates, solid line),
- the MOESP algorithm (dashed),
- the robust N4SID method (dotted).

It is evident that all the algorithms tend to identify a non-existing frequency response with a rather high resonance in the frequency band of the stochastic disturbance input. This is a structural feature of the problem which shows also in the Cramèr–Rao bound. The Monte Carlo estimates of the asymptotic variance remain however quite far from the Cramèr–Rao bound, with a relative standard error of the frequency response estimate (in the frequency band of the stochastic disturbance) of about 100%. It will be shown in

a forthcoming paper (Chiuso & Picci, 2003), how and why this kind of problems can be avoided using an "orthogonal decomposition" based algorithm.

3.1. Conclusions

In this paper explicit expressions have been provided pinpointing the sensitive dependence of the asymptotic variances of the estimates on the index of collinearity $\sigma_{\min}(I - \hat{\Pi}\hat{\Pi}^\top)$. The accuracy of the B, D-parameter estimation has been discussed for the N4SID algorithm but a similar analysis to the one reported in this paper also applies to the "linear regression method" of Verhaegen and Dewilde (1992).

Appendix A.

Proof of Theorem 2. Assuming we have data up to time $T+1$, we now define $Z_{[t,T]} := E_N[Y_{[t,T]} \,|\, P_{[t_0,t)} \vee U_{[t,T+1]}]$ and let $\bar{X}_t := \hat{\Gamma}^\dagger Z_{[t,T]}$ so that by the finite-sample version of (1.4)

$$\bar{X}_t = \hat{\Gamma}^\dagger \Gamma \hat{X}_t + \hat{\Gamma}^\dagger H_d U_{[t,T]} + \hat{\Gamma}^\dagger H_s \tilde{E}_{[t,T]}, \tag{A.1}$$

where $\tilde{E}_{[t,T]} := E_N[E_{[t,T]} \,|\, P_{[t_0,t)} \vee U_{[t,T+1]}]$ and the finite-data approximate Kalman filter state $\hat{X}_t := E_N[X_t \,|\, P_{[t_0,t)} \vee U_{[t,T+1]}]$ satisfies the linear recursion

$$\begin{bmatrix} \hat{X}_{t+1} \\ Y_t \end{bmatrix} = \begin{bmatrix} A \\ \mathbb{C} \end{bmatrix} \hat{X}_t + \begin{bmatrix} B \\ D \end{bmatrix} U_t + \begin{bmatrix} K \\ I \end{bmatrix} \tilde{E}_t + \begin{bmatrix} K(t) \\ I \end{bmatrix} \hat{E}_t. \tag{A.2}$$

Here $\hat{E}_t := Y_t - E_N[Y_t \,|\, P_{[t_0,t)} \vee U_{[t,T+1]}]$ is the finite data innovation tail matrix, and $K(t)$ comes from the usual definition of Kalman gain, $K(t)\hat{E}_t := E_N[Y_t|\hat{E}_t]$.

The updated pseudostate $\bar{X}_{t+1}$, is now defined using $Z_{[t+1,T+1]}$ so that

$$\bar{X}_{t+1} = \hat{\Gamma}^\dagger \Gamma \hat{X}_{t+1} + \hat{\Gamma}^\dagger H_d U_{[t+1,T+1]} + \hat{\Gamma}^\dagger H_s \tilde{E}_{[t+1,T+1]}.$$

Now introduce the change of basis matrix $T_N := \hat{\Gamma}^\dagger \Gamma$ (nonsingular for N large enough) and let

$$A_N := T_N A T_N^{-1}, \quad C_N := C T_N^{-1}, \quad B_N := T_N B \tag{A.3}$$

and substitute (A.2) into (A.1) written for time $t+1$, to obtain

$$\begin{bmatrix} \bar{X}_{t+1} \\ Y_t \end{bmatrix} = \begin{bmatrix} A_N \\ C_N \end{bmatrix} \bar{X}_t, + \begin{bmatrix} \mathscr{K}_{1,N} \\ \mathscr{K}_{2,N} \end{bmatrix} U_{[t,T+1]} + \begin{bmatrix} K_N(t) \\ I \end{bmatrix} \hat{E}_t + \begin{bmatrix} M_N \bar{H}_s \tilde{E}_{[t,T+1]} \\ R_N H_s \tilde{E}_{[t,T]} \end{bmatrix}, \tag{A.4}$$

where $K_N(t) := T_N K(t), K_N := T_N K$ and

$$\mathscr{K}_{1,N} := ([B_N \;\; \hat{\Gamma}^\dagger] - [A_N \hat{\Gamma}^\dagger \;\; 0])\bar{H}_d, \tag{A.5}$$

$$\mathscr{K}_{2,N} := ([D \;\; 0] - C_N \hat{\Gamma}^\dagger) H_d, \tag{A.6}$$

$$M_N := ([K_N \;\; \hat{\Gamma}^\dagger] - [A_N \hat{\Gamma}^\dagger \;\; 0]), \tag{A.7}$$

$$R_N := ([I \;\; 0] - C_N \hat{\Gamma}^\dagger). \tag{A.8}$$

In these expressions $\bar{H}_d, \bar{H}_s$ are the Toeplitz matrices H_d, H_s bordered as defined in Section 1. Since $\hat{\Gamma}$ is assumed to be a consistent estimate of Γ, for $N \to \infty$ $T_N \to I$, and we shall have

$$\mathscr{K}_{1,N} \to K_1 := ([B \;\; \Gamma^\dagger] - [A\Gamma^\dagger \;\; 0])\bar{H}_d, \tag{A.9}$$

$$\mathscr{K}_{2,N} \to K_2 := ([D \;\; 0] - C\Gamma^\dagger) H_d, \tag{A.10}$$

$$M_N \to M := ([K \;\; \Gamma^\dagger] - [A\Gamma^\dagger \;\; 0]), \tag{A.11}$$

$$R_N \to R := ([I \;\; 0] - C\Gamma^\dagger), \tag{A.12}$$

where $(\mathscr{K}_1, \mathscr{K}_2)$ are known functions of the parameters of the stationary system, as defined in Van Overschee and De Moor (1994).

From (A.4), using again the oblique projection Lemma, we can now write down the least-squares estimates of $\mathscr{K}_{1,N}, \mathscr{K}_{2,N}$ as

$$\hat{\mathscr{K}}_{1,N} = E_N[\bar{X}_{t+1} U_{[t,T+1]}^\top \,|\, \bar{X}_t] \hat{\Sigma}^{-1}_{\bar{U}^+\bar{U}^+ \,|\, \bar{X}}, \tag{A.13}$$

$$\hat{\mathscr{K}}_{2,N} = E_N[Y_t U_{[t,T+1]}^\top \,|\, \bar{X}_t] \hat{\Sigma}^{-1}_{\bar{U}^+\bar{U}^+ \,|\, \bar{X}}, \tag{A.14}$$

where the symbol $\bar{U}^+$ designates the bordered tail matrix corresponding to $\bar{\mathbf{u}}^+$, so that $\hat{\Sigma}_{\bar{U}^+\bar{U}^+ \,|\, \bar{X}} := E_N[U_{[t,T+1]} U_{[t,T+1]}^\top \,|\, \bar{X}_t]$. All the conditional (sample) covariances involve projections onto the orthogonal complement of the rowspace of $\bar{X}_t$ in $P_{[t_0,t)} \vee U_{[t,T+1]}$. Letting $\bar{X}_t^\perp$ span this orthogonal complement, we have $E_N[\hat{E}_t \,|\, \bar{X}_t^\perp] = 0$ and

$$E_N(E_N[\tilde{E}_{[t,T+1]} \,|\, \bar{X}_t^\perp] E_N[U_{[t,T+1]} \,|\, \bar{X}_t^\perp]^\top) = E_N(E_{[t,T+1]} E_N[U_{[t,T+1]} \,|\, \bar{X}_t^\perp]^\top) := \hat{\Sigma}_{\bar{E}^+\bar{U}^+ \,|\, \bar{X}}$$

whereby (A.13) and (A.14) can be rewritten as

$$\hat{\mathscr{K}}_{1,N} = \mathscr{K}_{1,N} + M_N \bar{H}_s \hat{\Sigma}_{\bar{E}^+\bar{U}^+ \,|\, \bar{X}} \hat{\Sigma}^{-1}_{\bar{U}^+\bar{U}^+ \,|\, \bar{X}},$$

$$\hat{\mathscr{K}}_{2,N} = \mathscr{K}_{2,N} + R_N H_s \hat{\Sigma}_{E^+\bar{U}^+ \,|\, \bar{X}} \hat{\Sigma}^{-1}_{\bar{U}^+\bar{U}^+ \,|\, \bar{X}}, \tag{A.15}$$

where $\bar{E}^+$ stands for the augmented matrix $E_{[t,T+1]}$. These provide exact expressions for the finite sample estimation errors, from which, following exactly the same arguments in the proof of Theorem 4.1 in Chiuso and Picci (2004a), one gets the asymptotic variance expressions (2.13)–(2.15). The rest is nearly obvious.

References

Bauer, D., & Ljung, L. (2001). Some facts about the choice of the weighting matrices in larimore type of subspace algorithm. *Automatica*, *36*, 763–773.

Bauer, D., & Jansson, M. (2000). Analysis of the asymptotic properties of the moesp type of subspace algorithms. *Automatica*, *36*, 497–509.

Chiuso, A., & Picci, G. (2003). Subspace identification by data orthogonalization and model decoupling. Automatica, submitted.

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

Chiuso, A., & Picci, G. (2004b). On the ill-conditioning of subspace identification with inputs. *Automatica*, doi:10.1016/j.automatica.2003.11.009.

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

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

Jansson, M. (2000). Asymptotic variance analysis of subspace identification methods. In *Proceedings of SYSID2000*, S. Barbara, CA.

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

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 and 1211–1241.

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 Dept. of Information Engineering, University of Padova. In the summer 2001 he has been visiting resarcher with the Dept. of Computer Science, University of California Los Angeles. His research interests are mainly in Identification and Estimation Theory, System Theory and Computer Vision.