PROBING INPUTS FOR SUBSPACE IDENTIFICATION

ALESSANDRO CHIUSO [1] and GIORGIO PICCI[2]

Abstract

There is experimental evidence that the standard subspace methods (e.g. the N4SID method) perform poorly in certain conditions where the past signals (past inputs and past outputs) and future input spaces are nearly parallel. Based on an elementary numerical conditioning analysis, this paper describes a class of (system-dependent) input signals (called *probing inputs*) which lead to the worst possible conditioning of the identification problem. Numerical results are included demonstrating how these input signals may lead to a substantial deterioration of performance of the algorithms in some experimental conditions.

Keywords: Subspace Identification; Exogenous inputs; Numerical Conditioning; State-space identification

1 Background and Notations

Assume, ideally, that we have observations on some long time interval $[0, T]$ of one sample path $\{y(t)\}$, $\{u(t)\}$, $\{x(t)\}$, of the zero mean stationary processes $\{\mathbf{y}(t)\}$, $\{\mathbf{u}(t)\}$, $\{\mathbf{x}(t)\}$, which satisfy the dynamical equations of a linear stochastic system

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

where $\mathbf{u}$ is an exogenous input process. The system can be taken, without loss of generality, in "innovation form", in particular one can assume that $\{\mathbf{e}(t)\}$ is a white noise process uncorrelated with the past history of all other variables. All through this paper we shall assume that there is no feedback from $\mathbf{u}$ to $\mathbf{y}$. In this case $\mathbf{e}$ is in fact uncorrelated with the full history of the input process (see e.g. [9]).

Since the processes generating the data satisfy the the dynamical relation (1.1), it is obvious that the "tail" matrices (of length N) at time t, U_t, Y_t, X_t, defined as

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

also satisfy equation (1.1), i.e.:

$$\begin{cases} X_{t+1} &= AX_t + BU_t + KE_t \\ Y_t &= CX_t + DU_t + E_t \end{cases} \tag{1.2}$$

where, for $N \to \infty$, $E_t \perp (X_s, U_s)$ for all $s \le t$, orthogonality of two semi-infinite sequences, $\xi, \eta \in \mathbb{R}^{\mathbb{Z}_+}$ being understood with respect to the scalar product

$$\langle\, \xi,\, \eta \,\rangle := \lim_{N\to\infty} \frac{1}{N+1} \sum_{k=0}^{N} \xi(k)\eta(k)$$

It is straightforward to see that knowledge of X_{t+1}, X_t, U_t and Y_t permits to solve uniquely (1.2) for the unknown parameters (A, B, C, D), say by least squares. Hence in this ideal situation the identification (of the deterministic part) of the system (1.1) is rather trivial. In practice however we only have finite input-output data, say $2k$ tail matrices $\{U_t, Y_t\}_{t=0,\dots,2k-1}$ and the state needs to be constructed from the available data. This construction can be seen as the central conceptual step in subspace identification.

2 Conditioning of N4SID

In this section we shall present an elementary numerical analysis of the N4SID algorithm of [10]. A similar analysis can be made for the so-called MOESP algorithm of [12] and also for the CCA algorithm [7]. Since a unified treatment of these procedures would require a lengthy preliminary technical detour, they will not be discussed in this paper. A general approach will be found in the forthcoming publication [2].

We shall use the notations of [10]. The first object which is computed in the N4SID algorithm is the predictor matrix based on joint input-output data

$$\begin{array}{rcl} Z_k &:=& E[Y_{k|2k-1} \mid Y_{0|k-1} \vee U_{0|2k-1}] \\ &=& [\,L_1^k \quad L_2^k \quad L_3^k\,] \begin{pmatrix} U_{0|k-1} \\ U_{k|2k-1} \\ Y_{0|k-1} \end{pmatrix}. \end{array}$$

From this an estimate, $\hat{\Gamma}_k$, of the observability matrix Γ_k can be obtained by an oblique projection along

[1]Dipartimento di Elettronica e Informatica, University of Padova, 35131 Padua, Italy; Email: chiuso@dei.unipd.it

[2]Dipartimento di Elettronica e Informatica, University of Padova and LADSEB-CNR, 35131 Padua, Italy; Email: picci@dei.unipd.it

[†]This work was supported by the TMR network *System Identification* and by the national project *Identification and control of industrial systems* funded by MURST.

$\mathcal{U}_{k|2k-1}$ followed by an SVD step*. Once $\hat{\Gamma}_k$ is computed, the second crucial step of the procedure is to solve the overdetermined linear system

$$\left[\begin{array}{c} \hat{\Gamma}_{k-1}^{-L} Z_{k+1} \\ Y_k \end{array}\right] = \left[\begin{array}{c} A \\ C \end{array}\right] \hat{\Gamma}_k^{-L} Z_k + \\ + \left[\begin{array}{c} \mathcal{K}_1 \\ \mathcal{K}_2 \end{array}\right] U_{k|2k-1} + W^{\perp} \qquad (2.3)$$

This relation is interpreted as a multiple regression problem for the unknown parameters $(A\,, C)$ and $(\mathcal{K}_1,\, \mathcal{K}_2)$, (see eq. (43) in [10]). We recall, for further reference, that

$$\Gamma_k^{-L} Z_k = \hat{X}_k + \Gamma_k^{-L} H_k U_{k|2k-1} \qquad (2.4)$$

(see equation (15) in [10]), where $\hat{X}_k$ is the $n \times N$ state tail matrix and H_k is a lower triangular block Toeplitz matrix made with the first k Markov parameters $(D,\, CB,\, CAB, ..., CA^{k-2}B)$, of the "deterministic" subsystem (A, B, C, D).

In an ideal situation $(N \to \infty)$ the term $W^{\perp}$ is orthogonal to the row span of $U_{k|2k-1}$ and to the row span of $\Gamma_k^{-L} Z_k$. In practice, due to the fact that only a finite amount of data is available, this is only approximately true. Assuming that orthogonality is satisfied with good approximation, it follows, see e.g. [6], that the Least-Squares parameter estimates of $(A,\, C)$ and $(\mathcal{K}_1,\, \mathcal{K}_2)$ obtained from (2.3), are the oblique projections of the LHS onto the rowspace of $\hat{\Gamma}_k^{-L} Z_k$ along the rowspace of $U_{k|2k-1}$ and, respectively, onto $U_{k|2k-1}$ along the rowspace of $\hat{\Gamma}_k^{-L} Z_k$.

Apparently, a crucial issue in the computation of the solutions is how "parallel" are the rowspaces of $U_{k|2k-1}$ and $\hat{\Gamma}_k^{-L} Z_k$. Here "nearly parallel" subspaces means subspaces for which some canonical angles are nearly zero (see [3], p. 584). For nearly parallel row spaces one expects that the computation of the parameters $(A,\, C)$ and $(\mathcal{K}_1,\, \mathcal{K}_2)$ of the regression will be ill-conditioned, and the parameter estimates could be affected by large errors.

Actually, it turns out that the regression problem and, in particular, the computation of $(\mathcal{K}_1\,, \mathcal{K}_2)$ will be ill-conditioned also if the canonical correlations between past and future inputs is close to one (see Proposition 2.2 below). This possible cause of ill-conditioning has to do with wide variations in the amplitude of the input spectrum and with frequency bands where the spectrum is nearly zero causing "insufficient excitation". We shall not consider this aspect in this paper. Some simulations in [1] show that the consequences of this kind of ill-conditioning can be quite dramatic. In this paper we shall concentrate instead on the situations where there could be "small" canonical angles between the state space of the (deterministic) system to be identified, and the future input space. Since sample fluctuations (i.e. finite data length) play a secondary role in this analysis, we shall henceforth assume that the sample size $N \to \infty$ and adopt the stochastic setup (the corrispondence between the stochastic and the tail matrices settings is described in detail in [8]). All tail matrices will henceforth be substituted by their corresponding stochastic quantities, denoted by lower case boldface symbols. For simplicity, the pk-dimensional stochastic vector corresponding to $U_{k|2k-1}$ will be denoted by $\mathbf{u}_k^+$ and that corresponding to $Y_{k|2k-1}$ by $\mathbf{y}_k^+$. The state vector of the model at time k, $\hat{\mathbf{x}}(k)$, evolves according to a transient Kalman filter

$$\begin{array}{rcl} \hat{\mathbf{x}}(t+1) & = & A\hat{\mathbf{x}}(t) + B\mathbf{u}(t) + K(t)\hat{\mathbf{e}}(t) \quad t \in [0, 2k-1] \\ \hat{\mathbf{x}}(0) & = & E\left[\mathbf{x}(0) \mid \mathcal{U}_{0|2k-1}\right] \end{array}$$

In (2.4) the quantity $\mathbf{x}_u(k) := \Gamma_k^{-L} H_k \mathbf{u}_k^+$ is a component of the state vector at time k made of future inputs, which contributes to the predictor variables. Hence, denoting $\hat{\bar{\mathbf{x}}}(k) := \hat{\mathbf{x}}(k) + \Gamma_k^{-L} H_k \mathbf{u}_k^+ = \hat{\mathbf{x}}(k) + \mathbf{x}_u(k)$ we may rewrite (2.3) as

$$\left[\begin{array}{c} \hat{\bar{\mathbf{x}}}(k+1) \\ \mathbf{y}(k) \end{array}\right] = \left[\begin{array}{c} A \\ C \end{array}\right] \hat{\bar{\mathbf{x}}}(k) + \left[\begin{array}{c} \mathcal{K}_1 \\ \mathcal{K}_2 \end{array}\right] \mathbf{u}_k^+ + \mathbf{w}_k^{\perp} \quad (2.5)$$

where $\mathbf{w}_k^{\perp}$ is orthogonal (i.e. uncorrelated) to the space spanned by the random variables $\hat{\bar{\mathbf{x}}}(k)$, $\mathbf{u}_k^+$. We shall assume that the subspaces spanned by (the components of) $\hat{\bar{\mathbf{x}}}(k)$ and $\mathbf{u}_k^+$ have only the zero random variable in common (the generic consistency condition of [4]). In this situation we can use the oblique projection Lemma of [6] to express the parameter estimates as solutions of the Wiener-Hopf type equations

$$\left[\begin{array}{c} A \\ C \end{array}\right] \Sigma_{\bar{\mathbf{x}}\bar{\mathbf{x}}|\mathbf{u}^+} = \left[\begin{array}{c} \Sigma_{\bar{\mathbf{x}}_1\bar{\mathbf{x}}|\mathbf{u}^+} \\ \Sigma_{\bar{\mathbf{y}}\bar{\mathbf{x}}|\mathbf{u}^+} \end{array}\right], \qquad (2.6)$$

$$\left[\begin{array}{c} \mathcal{K}_1 \\ \mathcal{K}_2 \end{array}\right] \Sigma_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}} = \left[\begin{array}{c} \Sigma_{\bar{\mathbf{x}}_1\mathbf{u}^+|\bar{\mathbf{x}}} \\ \Sigma_{\mathbf{y}\mathbf{u}^+|\bar{\mathbf{x}}} \end{array}\right], \qquad (2.7)$$

involving various *conditional* covariance matrices. The expressions appearing on the left hand side are

$$\begin{array}{l} \Sigma_{\bar{\mathbf{x}}\bar{\mathbf{x}}|\mathbf{u}^+} = E\{\left[\hat{\bar{\mathbf{x}}}(k) - E\left(\hat{\bar{\mathbf{x}}}(k) \mid \mathbf{u}_k^+\right)\right]\left[\hat{\bar{\mathbf{x}}}(k) - E\left(\hat{\bar{\mathbf{x}}}(k) \mid \mathbf{u}_k^+\right)\right]'\} \\ \Sigma_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}} = E\{\left[\mathbf{u}_k^+ - E(\mathbf{u}_k^+ \mid \hat{\bar{\mathbf{x}}}(k))\right]\left[\mathbf{u}_k^+ - E(\mathbf{u}_k^+ \mid \hat{\bar{\mathbf{x}}}(k))\right]'\} \end{array} \qquad (2.8)$$

the formulas for $\Sigma_{\bar{\mathbf{x}}_1\bar{\mathbf{x}}|\mathbf{u}^+}$, etc. are similar, involving $\bar{\mathbf{x}}_1 \equiv \hat{\bar{\mathbf{x}}}(k+1)$ and $\mathbf{y} \equiv \mathbf{y}(k)$. Note that the equations (2.6) (2.7) express how the solutions of (2.5) depend on the data and have nothing to do with the algorithm which is used to compute them. It follows that the conditioning of the problems (2.6) and (2.7) is determined by the singular values of the symmetric matrices $\Sigma_{\bar{\mathbf{x}}\bar{\mathbf{x}}|\mathbf{u}^+}$ and $\Sigma_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}}$. Note the obvious equality $\Sigma_{\bar{\mathbf{x}}\bar{\mathbf{x}}|\mathbf{u}^+} = E\{\left[\hat{\mathbf{x}}(k) - E\left(\hat{\mathbf{x}}(k) \mid \mathbf{u}_k^+\right)\right]\left[\hat{\mathbf{x}}(k) - E\left(\hat{\mathbf{x}}(k) \mid \mathbf{u}_k^+\right)\right]'\} := \Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}$.

*This oblique projection may be the first source of ill-conditioning. [12] argues that an orthogonal projection onto the space $\mathcal{U}_{k|2k-1}^{\perp}$ is to be recommended instead.

Introduce the covariance matrices $\Sigma_{\hat{\mathbf{x}}} = E\left[\hat{\mathbf{x}}(k)\,\hat{\mathbf{x}}(k)'\right]$, $\bar{\Pi} := E\left[\mathbf{u}_k^+\,\bar{\hat{\mathbf{x}}}(k)'\right]$, $\Pi := E\left[\mathbf{u}_k^+\,\hat{\mathbf{x}}(k)'\right]$, and let Λ_u be the covariance matrix of $\mathbf{u}_k^+$. We shall need also the Cholesky factors L_u of Λ_u, and $L_{\hat{\mathbf{x}}}$ of $\Sigma_{\hat{\mathbf{x}}}$, so that $L_u L_u' = \Lambda_u \quad , \quad L_{\hat{\mathbf{x}}} L_{\hat{\mathbf{x}}}' = \Sigma_{\hat{\mathbf{x}}}$

Proposition 2.1 *The condition number $\kappa\left(\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}\right)$ of $\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}$ can be bounded by*

$$\kappa\left(\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}\right) \leq \kappa\left(\Sigma_{\hat{\mathbf{x}}}\right) \frac{1-\sigma_{min}^2(\hat{\Pi})}{1-\sigma_{max}^2(\hat{\Pi})} \tag{2.9}$$

where $\hat{\Pi}$ is the normalized cross-covariance

$$\hat{\Pi} := L_u^{-1} \Pi L_{\hat{\mathbf{x}}}^{-T} \tag{2.10}$$

Similarly, consider the factorization

$$\Sigma_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}} = L_u \left(I - \hat{\bar{\Pi}}\hat{\bar{\Pi}}'\right) L_u', \tag{2.11}$$

where $\hat{\bar{\Pi}}$ is the normalized cross-covariance $\hat{\bar{\Pi}} := L_u^{-1}\bar{\Pi}L_{\bar{x}}^{-T}$ Then the condition number $\kappa\left(\Sigma_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}}\right)$ can be estimated by

$$\kappa\left(\Sigma_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}}\right) \leq \kappa\left(\Lambda_u\right) \frac{1}{1-\sigma_{max}^2(\hat{\bar{\Pi}})} \tag{2.12}$$

These bounds are sharp.

The conditioning of the matrices (2.8) depends on the maximum canonical correlation coefficient between the subspaces generated by $\mathbf{u}_k^+$ and $\hat{\bar{\mathbf{x}}}(k)$ and on the maximum canonical correlation coefficient between the subspaces generated by $\mathbf{u}_k^+$ and $\hat{\mathbf{x}}_d(k)$. As the maximum canonical correlation coefficient approaches one, these matrices tend to be singular, making the problem ill-conditioned. More useful bounds are obtained if we assume that the data are available from the infinite past. In this case one uses the predictor based on the the infinite past

$$\mathbf{z}_k = E\left[\mathbf{y}_k^+ | \mathcal{U}_k^- \vee \mathcal{Y}_k^- \vee \mathcal{U}_{k|2k-1}\right] = \Gamma_k \mathbf{x}(k) + H_k \mathbf{u}_k^+$$

where the state vector $\mathbf{x}(k)$ is just the state of a stationary innovation model (1.1). When dealing with the infinite past case (which we shall do from now on), we shall eliminate the hats (= prediction based on a finite history) on the relevant symbols.

Recall [9] that there exist an orthogonal change of basis on the state space under which (1.1) can be written in a block diagonal form with state $\mathbf{x}(k) = [\mathbf{x}_d(k)' \;\; \mathbf{x}_s(k)']'$, where $\mathbf{x}_d(k)$ and $\mathbf{x}_s(k)$ are the *deterministic* and *stochastic* components of the state, mutually uncorrelated at all times. Let $\Sigma_d := E\{\mathbf{x}_d(k)\,(\mathbf{x}_d(k))'\}$, $\tilde{\mathbf{x}}_d(k) = \mathbf{x}_d(k) - E\{\mathbf{x}_d(k) \mid \mathbf{u}_k^+\}$, $\tilde{\Sigma}_d := E\{\tilde{\mathbf{x}}_d(k)\,(\tilde{\mathbf{x}}_d(k))')\}$, $\Pi_d := E\left[\mathbf{u}_k^+\,\mathbf{x}_d(k)'\right]$ and introduce the Cholesky factor L_{x_d} of Σ_d

Proposition 2.2 *Assume there is no common dynamics to the deterministic and stochastic components of the state process. Then there exist an orthogonal change of basis on the state space under which $\Pi = [\Pi_d \; 0]$ and*

$$\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+} = \tilde{\Sigma}_d \oplus \hat{\Sigma}_s \tag{2.13}$$

(direct sum) where

$$\tilde{\Sigma}_d = L_{x_d}\left(I - \hat{\Pi}_d'\hat{\Pi}_d\right) L_{\mathbf{x}_d}' \tag{2.14}$$

$\hat{\Pi}_d$ being the normalized cross-covariance $\hat{\Pi}_d := L_u^{-1}\Pi_d L_{\mathbf{x}_d}^{-T}$. The condition number $\kappa\left(\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}\right)$ in this case can be estimated as

$$\kappa\left(\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{u}^+}\right) \leq \kappa\left(\Sigma_{\mathbf{x}}\right) \frac{1}{1-\sigma_{max}^2(\hat{\Pi}_d)} \tag{2.15}$$

This bound is sharp.

Now the singular values of the normalized covariance matrices $\hat{\bar{\Pi}}$, $\hat{\Pi}_d$ are the cosines of the canonical angles between the spaces spanned by $\mathbf{u}_k^+$ and $\bar{\mathbf{x}}(k)$, and by $\mathbf{u}_k^+$ and $\mathbf{x}_d(k)$ respectively. From (2.12) it can also be seen that if $\kappa\left(\Lambda_{\mathbf{u}}\right)$ is large (which happens when the amplitude of the spectrum of $\mathbf{u}$ varies widely and if the deterministic state at time k is highly correlated with future inputs, the maximum and minimum singular values of $\Sigma_{\mathbf{u}^+\mathbf{u}^+|\bar{\mathbf{x}}}$ may differ widely. The situation is relatively safe if $\mathbf{u}$ is a white noise process, in which case $\kappa(\Lambda_u) = 1$. Formula (2.13) reduces to the direct sum of the two covariance matrices of the deterministic and stochastic components of the state. This means that, if an appropriate basis is chosen and the system order is estimated correctly, the computation of $(A,\, C)$ should not provide surprises. However, even if $\mathbf{u}$ is nearly white, it may appear from (2.12) that (2.7) could still be ill-conditioned.

3 Probing Inputs for Identification

In this section we shall investigate the class of inputs which lead to the highest correlation coefficient $\sigma_{\max}(\hat{\Pi}_d)$. From what we have seen before, if $\sigma_{\max}(\hat{\Pi}_d)$ is close to one, then in most cases $\sigma_{\max}(\hat{\bar{\Pi}})$ will also be near one. To simplify the canonical correlation analysis between the spaces spanned by $\mathbf{x}_d(k)$ and $\mathbf{u}_k^+$, we shall assume that $\text{span}\{\mathbf{u}_k^+\} \simeq \mathcal{U}_k^+$ (infinite future).

The following lemma is instrumental in the analysis. Here $\text{span}\{\mathbf{x}_d(0)\} = \mathcal{X}_d^{+/-} \subset \mathcal{U}^-$, is the oblique predictor space of the deterministic component $\mathbf{y}_d$ of the process $\mathbf{y}$ [**?**].

Lemma 3.1 *Denote by $\sigma_k(\mathcal{X}_d^{+/-}, \mathcal{U}^+)$ the $k-th$ canonical correlation coefficient of $\mathcal{X}_d^{+/-}$ and $\mathcal{U}^+$ and*

by $\sigma_k(\mathcal{U}^-,\mathcal{U}^+)$ the $k-th$ canonical correlation coefficient of $\mathcal{U}^-$ and $\mathcal{U}^+$. Then the following inequalities hold:

$$\sigma_k(\mathcal{X}_d^{+/-},\mathcal{U}^+) \leq \sigma_k(\mathcal{U}^-,\mathcal{U}^+) \quad , \quad k=1,2,\dots. \tag{3.16}$$

Moreover

$$\sigma_k(\mathcal{X}_d^{+/-},\mathcal{U}^+) = \sigma_k(\mathcal{U}^-,\mathcal{U}^+) \quad k=1,2,\dots,n \tag{3.17}$$

if and only if the first n principal directions of $\mathcal{U}^-$ are contained in $\mathcal{X}_d^{+/-}$.

Assume the spectral density of $\mathbf{u}$ is rational and let $\mathcal{U}^{+/-}$ be the ν-dimensional forward predictor space of the process $\mathbf{u}$. A (stochastically balanced) basis $\mathbf{x}$ in $\mathcal{U}^{+/-}$ is made of random variables which are proportional to the principal directions [8]. If the first n principal directions of $\mathcal{U}^-$ are contained in $\mathcal{X}_d^{+/-}$, then $\mathbf{x}$ can be decomposed as

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_- \\ \mathbf{x}_\perp \end{bmatrix}$$

where $\text{span}\{\mathbf{x}_-\} = \mathcal{X}_d^{+/-}$ and $\mathbf{x}_\perp \perp \mathcal{X}_d^{+/-}$.

Let $\mathcal{H}_d$ be the Hankel operator

$$\mathcal{H}_d := E^{\mathcal{X}_d^{+/-}}_{|\mathcal{U}^+} \tag{3.18}$$

The canonical correlation coefficients between future inputs $\mathcal{U}^+$ and the deterministic state space $\mathcal{X}_d^{+/-}$ can be computed as singular values of a matrix representation of the Hankel operator (3.18).

Since $\mathcal{X}_d^{+/-} \subseteq \mathcal{U}^-$ the Hankel operator can be factorized as follows:

$$\mathcal{H}_d = E^{\mathcal{X}_d^{+/-}} E^{\mathcal{U}^-}_{|\mathcal{U}^+} = E^{\mathcal{X}_d^{+/-}} E^{\mathcal{U}^{+/-}}_{|\mathcal{U}^+}$$

The operator $\mathcal{O} := E^{\mathcal{U}^+}_{|\mathcal{U}^{+/-}}$ is the usual observability operator of the process $\mathbf{u}$. Let us define the projection operator

$$\mathcal{P} := E^{\mathcal{X}_d^{+/-}}_{|\mathcal{U}^{+/-}}$$

It is easy to show that the adjoint of $\mathcal{P}$ is given by

$$\mathcal{P}^* = E^{\mathcal{U}^{+/-}}_{|\mathcal{X}_d^{+/-}}$$

Therefore $\mathcal{H}_d$ admits the factorization $\mathcal{H}_d = \mathcal{P}\mathcal{O}^*$ and the squared canonical correlation coefficients between future inputs $\mathcal{U}^+$ and the deterministic state space $\mathcal{X}_d^{+/-}$ are the eigenvalues of the "squared" Hankel operator $\mathcal{H}_d^*\mathcal{H}_d$, i.e.,

$$\mathcal{H}_d^*\mathcal{H}_d\xi_i = \sigma_i^2\xi_i, \quad i=1,\dots,n$$

Using the factorization introduced above we obtain

$$\{\sigma_1^2,\sigma_2^2,\dots,\sigma_n^2\} = \lambda\{\mathcal{O}\mathcal{P}^*\mathcal{P}\mathcal{O}^*\} = \lambda\{\mathcal{O}^*\mathcal{O}\mathcal{P}^*\mathcal{P}\} \tag{3.19}$$

The following Lemma gives an explicit matrix representation of the Hankel operator.

Lemma 3.2 *Let $\mathbf{x}$ be a basis in the forward predictor space $\mathcal{U}^{+/-}$ of the process $\mathbf{u}$ and $\mathbf{x}_d$ be a basis in the oblique Markovian splitting subspace $\mathcal{X}_d^{+/-}$. Let P be the covariance matrix*

$$P = E\begin{bmatrix} \mathbf{x}_d \\ \mathbf{x} \end{bmatrix}[\,\mathbf{x}_d' \quad \mathbf{x}'\,] = \begin{pmatrix} P_{dd} & P_{du} \\ P_{ud} & P_{uu} \end{pmatrix}$$

where $P_{uu} := \text{Var}\{\mathbf{x}\}$. Then

$$\mathcal{P}^*\mathcal{P}a'\mathbf{x} = a'P_{ud}P_{dd}^{-1}P_{du}P_{uu}^{-1}\mathbf{x} \tag{3.20}$$

and

$$\mathcal{O}^*\mathcal{O}a'\mathbf{x} = a'P_{uu}\bar{P}_{uu}\mathbf{x} \tag{3.21}$$

where $\bar{P}_{uu}$ is the covariance of the dual (backward) basis in the backward predictor space $\mathcal{U}^{-/+}$ ([5]).

From the Lemma above we obtain the expression of the canonical correlation coefficients as eigenvalues of the matrix representation of (3.19) as follows

$$\{\sigma_1^2,\sigma_2^2,\dots,\sigma_n^2\} = \lambda\{\mathcal{O}^*\mathcal{O}\mathcal{P}^*\mathcal{P}\} = \lambda\{P_{ud}P_{dd}^{-1}P_{du}\bar{P}_{uu}\} \tag{3.22}$$

Remark. Note that, if $\mathcal{U}^{+/-} \supset \mathcal{X}_d^{+/-}$ we can choose $\mathbf{x}_d = \mathbf{x}_-$ whereby $P_{ud} = [P_-0] = P_{du}'$ where $P_- := \text{Var}\,\mathbf{x}_-$. In this situation formula (3.22) reduces to

$$\lambda\{\mathcal{O}^*\mathcal{O}\mathcal{P}^*\mathcal{P}\} = \lambda\{P_-\bar{P}_+\} = \{\sigma_{u,1}^2,\sigma_{u,2}^2,\dots,\sigma_{u,n}^2\}$$

where $\sigma_{u,i}^2$ are the canonical correlation coefficients between past and future inputs. In other words, we have proven that the maximal canonical correlation coefficients (smallest canonical angles) between $\mathcal{X}_d^{+/-}$ and $\mathcal{U}^+$ are obtained when $\mathcal{U}^{+/-} \supset \mathcal{X}_d^{+/-}$.

In the following we shall give spectral conditions to insure that the deterministic state space is contained in the forward predictor space of the input process $\mathbf{u}$.

Proposition 3.1 *Let (A_u,K_u,C_u,I) be a minimal innovation representation of $\mathbf{u}$ and let (A_d,B_d,C_d,D_d) be a minimal realization of the deterministic subsystem associated to the oblique Markovian splitting subspace $\mathcal{X}_d^{+/-}$. We have $\mathcal{U}^{+/-} \supset \mathcal{X}_d^{+/-}$ if and only if there is an n-dimensional invariant subspace $\mathcal{Z} \subset \mathbb{R}^\nu$ for $A_u - K_uC_u$, such that the restricted pair $((A_u - K_uC_u)_{|\mathcal{Z}},(K_u)_{|\mathcal{Z}})$ is similar to the pair (A_d,B_d). In other words there exists a non singular T such that $A_d = T(A_u - K_uC_u)_{|\mathcal{Z}}T^{-1}$ and $T(K_u)_{|\mathcal{Z}} = B_d$. In particular, $\mathcal{U}^{+/-}$ and $\mathcal{X}_d^{+/-}$ coincide if and only if $(A_u - K_uC_u, K_u)$ is similar to the pair (A_d,B_d).*

Remark. When $\mathcal{U}^{+/-} \supset \mathcal{X}_d^{+/-}$ the poles of the deterministic subsystem are cancelled by the stable zeros of the spectrum of $\mathbf{u}$.

Proposition 3.2 *For a scalar input process of given rational spectral density matrix Φ_u and state dimension ν having canonical correlation coefficients, $\sigma_k(\mathcal{U}^-,\mathcal{U}^+)$, $k = 1,\ldots,\nu$, the maximal canonical correlation coefficients $\sigma_k(\mathcal{X}_d^{+/-},\mathcal{U}^+)$ are obtained when and only when there are n zeros of the spectral density matrix Φ_u which cancel the poles of the deterministic transfer function of the system.*

Remark. In a sense this proposition describes the worst n-dimensional deterministic subsystem to identify with a given input process. Conversely, if the system is given and we have freedom to choose the input spectrum, once its zero-structure is fixed, we still have freedom to place the poles of Φ_u. It is easy to see that in this case, by placing the poles arbitrarily close to the unit circle, it is possible to move the maximal canonical correlation coefficients $\sigma_k(\mathcal{U}^-,\mathcal{U}^+)$, arbitrarily close to one. Hence in this setting we can make $\sigma_{\max}(\mathcal{X}_d^{+/-},\mathcal{U}^+)$ arbitrarily close to one.

4 Experimental Results

We shall report the results of simulations made on a simple scalar system described in Table 1. below. The identification algorithm used in the simulations is a "robustified" version of the N4SID algorithm described in [11].

	Poles	Zeros	K
Stoch. System	$-0.1 + j0.6$ $-0.1 - j0.6$	0.5 0.7	1
Det. System	$0.75 + j0.55$ $0.75 - j0.55$ 0.9	$-0.1 + j0.8$ $-0.1 - j0.8$ 0.5	0.2
Input 1	$0.8 + j0.55$ $0.8 - j0.55$	$0.35 + j0.95$ $0.35 - j0.95$	0.77
Input 2	$-0.815 + j0.5$ $-0.815 - j0.5$	$0.75 + j0.55$ $0.75 - j0.55$	1.35
Input 3	$-0.8 + j0.5$ $-0.8 - j0.5$	$0.75 + j0.55$ $0.75 - j0.55$	3

Table 1: Poles, Zeros and Gains for the stochastic, deterministic subsystems and for the 3 inputs considered.

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

1. The (stable) zeros of Φ_1 cancel exactly the poles of the deterministic system. However Φ_1 is nearly constant with frequency and $\mathbf{u}_1$ is a nearly white process. In this case the singular values $\sigma_k(\mathcal{X}_d^{+/-},\mathcal{U}^+) = \sigma_k(\mathcal{U}^-,\mathcal{U}^+)$ are all nearly the same and rather small (since past and future of a white noise form angles of 90 degrees). Then $\sigma^2_{max}(\hat{\Pi}_d)$ is small so that the problem, in spite of the cancellation, is well-conditioned.
2. The zeros of Φ_2 are far apart from the system poles but the input process is ill-conditioned. In this example $\kappa(\Lambda_u) \simeq 10^5$ and the identification is rather poor.
3. The (stable) zeros of Φ_3 cancel exactly the system poles and in addition the input process is ill-conditioned. In this example $\kappa(\Lambda_u)$ is nearly the same as in example 2. The identification is very poor.

In all three experiments the input r.m.s. power and the stochastic disturbance spectrum are the same so as to keep the same SNR ratio.

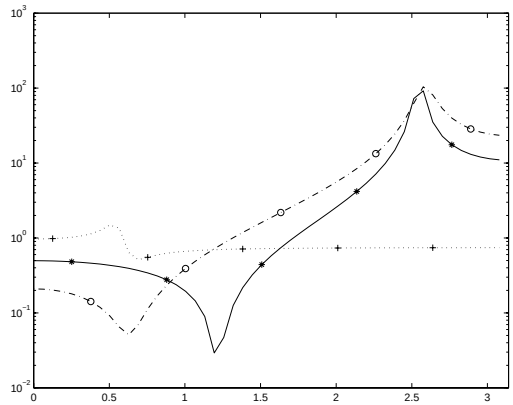

Figure 1: Input spectrum. Crosses (+): Input 1, stars ($*$): Input 2, circles (o): Input 3.

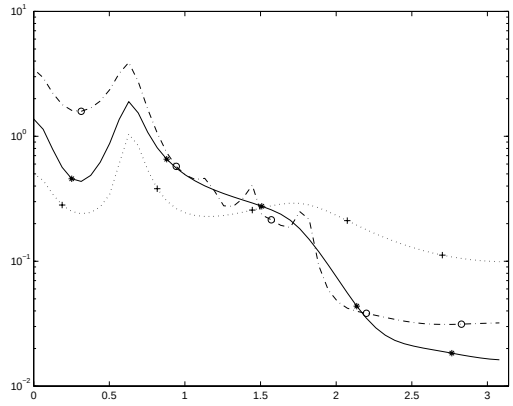

Figure 2: Standard deviation vs. frequency. Crosses (+): Input 1, stars ($*$): Input 2, circles (o): Input 3.

5 Conclusions

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

References

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

[2] Chiuso, A. and Picci G. (2000). On the ill-conditioning of certain subspace identification methods with inputs, *submitted for publication.*

[3] Golub, G. H. and C. R. Van Loan (1989). *Matrix Computations* (2nd ed.). The Johns Hopkins Univ. Press.

[4] Jansson M. and Wahlberg B. Counterexample to general consistency of subspace system identification methods. In *Proceedings of SYSID97*, Fukuoka, Japan,(1997) 1677–1682.

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

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

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

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

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

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

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

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

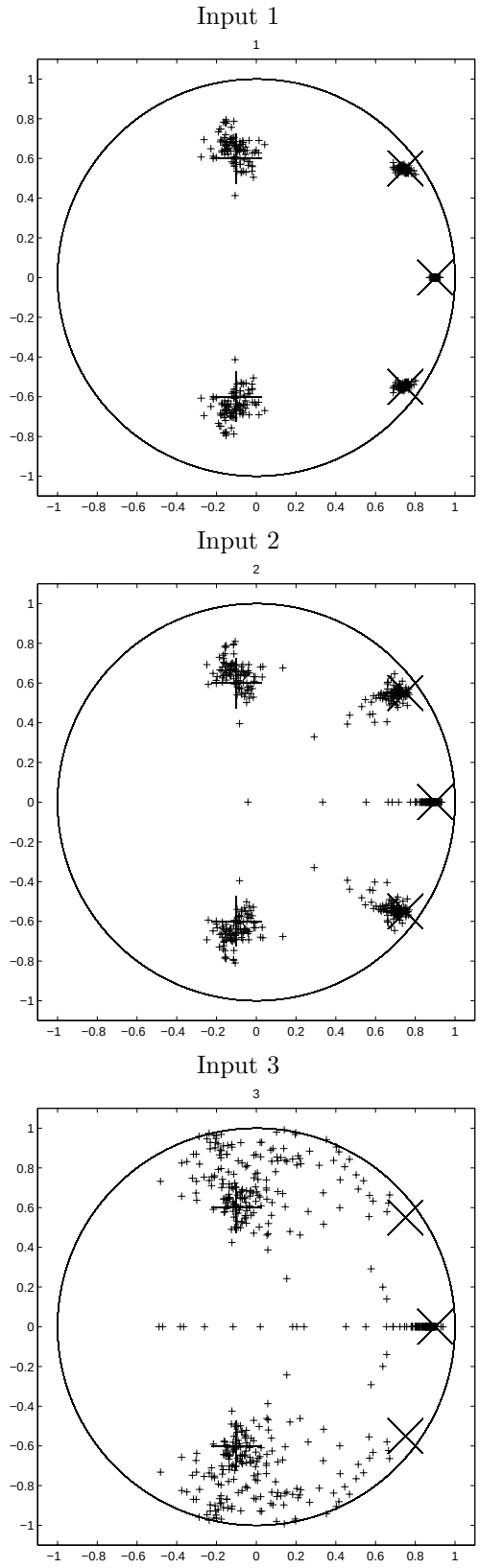

Figure 3: Estimated poles (N4SID robust).