CANONICAL CORRELATION ANALYSIS OF LINEAR STOCHASTIC SYSTEMS WITH INPUTS

A. CHIUSO and G. PICCI

REPORT No. 45, 2002/2003, spring

ISSN 1103-467X
ISRN IML-R- -45-02/03- -SE+spring

INSTITUT MITTAG-LEFFLER
THE ROYAL SWEDISH ACADEMY OF SCIENCES

CANONICAL CORRELATION ANALYSIS OF LINEAR STOCHASTIC SYSTEMS WITH INPUTS

ALESSANDRO CHIUSO * and GIORGIO PICCI†

Feb. 2003

Abstract : In this paper we show that the canonical correlations betwee the state space and the future input space of a (minimal) stochastic system with inputs, are maximized (equivalently the canonical angles are minimized) when the zeros of the spectral density of the input process cancel exactly the poles of the deterministic component of the system. This result is relevant in the analysis of subspace identification with inputs.

Keywords: Canonical Correlations; Linear stochastic systems with inputs; Exogenous inputs; Stochastic realization; Subspace identification.

1 Introduction

Let $\mathbf{y} = \{\mathbf{y}(t)\}$, $\mathbf{u} = \{\mathbf{u}(t)\}$ be m and p-dimensional zero-mean second-order stationary random processes admitting a representation by a linear stochastic

*Dipartimento di Ingegneria dell'Informazione, University of Padova, 35131 Padua, Italy; Email: chiuso@dei.unipd.it

†Dipartimento di Ingegneria dell'Informazione, University of Padova and ISB-CNR, 35131 Padua, Italy; Email: picci@dei.unipd.it This work has been supported in part through the national project *Identification and adaptive control of industrial systems* funded by the Italian ministery for higher education (MIUR).

system of the form

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

where A, B, G, C, D, J are constant matrices, $\{\mathbf{x}(t)\}$ is the state process of dimensions n, and $\{\mathbf{w}(t)\}$ is a normalized white noise process. We shall assume that *there is no feedback from* $\mathbf{y}$ *to* $\mathbf{u}$, see e.g. [3, 12, 17] for a discussion of this concept. This implies that the processes $\{\mathbf{u}(t)\}$ and $\{\mathbf{w}(t)\}$ are completely uncorrelated.

The system (1.1) is then called a *stationary stochastic realization of the output process* $\mathbf{y}$ *with input* $\mathbf{u}$ [17, 6, 16]. There are always infinitely many such linear representations of $\mathbf{y}$, which are equivalent up to (conditional) second-order statistics. A realization which is unique up to change of basis, is the so-called "innovation representation"

$$\begin{cases} \mathbf{x}(t+1) &= A\mathbf{x}(t) + B\mathbf{u}(t) + K\mathbf{e}(t) \\ \mathbf{y}(t) &= C\mathbf{x}(t) + D\mathbf{u}(t) + \mathbf{e}(t) \end{cases} \tag{1.2}$$

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

For obvious reasons, it will be convenient to assume throughout that the realization (1.2) is *(stochastically) minimal*, in the sense that the state dimension, n, is the smallest possible. This implies in particular (but is not equivalent to) that the triplet $\{C, A, [B\ K]\}$ is minimal in the usual system-theoretic sense.

In this paper we shall give for granted that the reader is familiar with canonical correlation analysis and with the concept of *principal (or canonical) angles* between subspaces spanned by zero-mean random variables. See e.g. [2, 1]. We shall be interested in computing certain (canonical) angles between the state space and the future input space of the system (1.2). The motivation for this problem comes from *subspace identification*, since as discussed in [5, 9, 4], these angles play an essential role in the conditioning analysis of the subspace identification problem, and are related to the performance (e.g. the asymptotic variance of the estimates) of subspace identification methods.

1.1 Notations

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

$$\begin{aligned} \mathcal{U}_{[t_0,\,t)} &:= \overline{\operatorname{span}}\,\{\mathbf{u}_k(s);\ k=1,\dots,p,\ t_0 \le s < t\,\} \\ \mathcal{Y}_{[t_0,\,t)} &:= \overline{\operatorname{span}}\,\{\mathbf{y}_k(s);\ k=1,\dots,m,\ t_0 \le s < t\,\} \end{aligned}$$

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

$$\langle\, \boldsymbol{\xi}, \boldsymbol{\eta}\,\rangle := \boldsymbol{E}\{\boldsymbol{\xi}, \boldsymbol{\eta}\} \tag{1.3}$$

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

$$\begin{aligned} \mathcal{U}_{[t,\,T]} &:= \overline{\operatorname{span}}\,\{\mathbf{u}_k(s);\ k=1,\dots,p,\ t \le s \le T\,\} \\ \mathcal{Y}_{[t,\,T]} &:= \overline{\operatorname{span}}\,\{\mathbf{y}_k(s);\ k=1,\dots,m,\ t \le s \le T\,\} \end{aligned}$$

By convention the past spaces do not include the present. When $t_0 = -\infty$ we shall use the shorthands $\mathcal{U}_t^-$, $\mathcal{Y}_t^-$ to denote the Hilbert spaces $\mathcal{U}_{[-\infty,\,t)}$, $\mathcal{Y}_{[-\infty,\,t)}$. of random variables spanned by the infinite past of $\mathbf{u}$ and $\mathbf{y}$ up to time t.

Subspaces spanned by random variables at just one time instant (e.g. $\mathcal{U}_{[t,\,t]}$, $\mathcal{Y}_{[t,\,t]}$, etc) are denoted simply as $\mathcal{U}_t$, $\mathcal{Y}_t$, etc. while for the spaces generated by the whole time history of $\mathbf{u}$ and $\mathbf{y}$ we shall use the symbols $\mathcal{U}$, $\mathcal{Y}$. We shall take $\mathcal{H} := \mathcal{U} \vee \mathcal{Y}$ (the $\vee$ denoting closed vector sum) as the *ambient space*, of all random quantities considered hereafter.

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

$$\mathcal{U}_{[t_0,\,T]} = \mathcal{U}_{[t_0,\,t)} + \mathcal{U}_{[t,\,T]}, \qquad t_0 \le t < T \tag{1.4}$$

the + sign denoting direct sum of subspaces. The symbol $\oplus$ will be reserved for *orthogonal* direct sum. Condition (1.4) can be found expressed in various equivalent forms in the literature, see e.g. [19, formula (10)]. Also various conditions ensuring sufficient richness are known. For example, it is well-known that for a full-rank purely non deterministic process $\mathbf{u}$ to be sufficiently rich, it is necessary and sufficient that the determinant of the spectral density matrix Φ_u should have no zeros on the unit circle [13].

The symbol $E\,[\cdot|\mathcal{A}]$ will denote (wide sense) conditional expectation, i.e. orthogonal projection onto the subspace $\mathcal{A} \subseteq \mathcal{H}$, orthogonality being with respect to the inner product (1.3). To streamline notation we shall sometimes also use the symbol $E^{\mathcal{A}}[\cdot]$; in particular, $E^{\mathcal{A}}\cdot_{|\mathcal{B}}$ will denote the restriction of the orthogonal projection to the subspace $\mathcal{B}$.

2 A decoupled canonical form

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

$$\mathbf{y}_d(t) := E[\mathbf{y}(t) \mid \mathcal{U}] \qquad \mathbf{y}_s(t) := \mathbf{y}(t) - E[\mathbf{y}(t) \mid \mathcal{U}] = E[\mathbf{y}(t) \mid \mathcal{U}^{\perp}] \tag{2.5}$$

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

$$\mathbf{y}(t) = \mathbf{y}_d(t) + \mathbf{y}_s(t) \qquad E\mathbf{y}_s(t)\mathbf{y}_d(\tau)' = 0 \quad \text{for all } t, \tau.$$

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

$$\mathbf{y}_d(t) = E[\mathbf{y}(t) \mid \mathcal{U}_{t+1}^{-}] \qquad t \in \mathbb{Z}$$

see [17], and is hence representable as the output of a causal linear time-invariant system driven only by the input signal $\mathbf{u}$. Consequently, $\mathbf{y}_s(t)$ is also the "causal estimation error" of $\mathbf{y}(t)$ based on the past and present inputs up to time t, i.e.

$$\mathbf{y}_s(t) := \mathbf{y}(t) - E[\mathbf{y}(t) \mid \mathcal{U}_{t+1}^{-}] \tag{2.6}$$

Since, under absence of feedback, $\mathbf{u}$ and $\mathbf{w}$ are uncorrelated, the deterministic and stochastic components of a process represented by a state space realization of the type (1.1) are represented by (generally non minimal) individual state space realizations, obtained by setting $\mathbf{u} = 0$ and $\mathbf{w} = 0$ in (1.1), or, equivalently, $\mathbf{e} = 0$ in (1.2).

The input-output relation of the innovation model (1.2) has the familiar form $\mathbf{y} = F(z)\mathbf{u} + G(z)\mathbf{e}$ with "stochastic" and "deterministic" transfer functions $F(z) = D + C(zI - A)^{-1}B$ and $G(z) = I + C(zI - A)^{-1}K$. Note that in these formulas the transfer functions are parametrized by the *same dynamic parameters* A, C and therefore need not be represented minimally. This is essentially the well-known ARMAX parametrization, most often considered in the identification literature. In most practical cases, however, unless there are known "physical" disturbances entering the system from the same input channels as $\mathbf{u}$, the stochastic and deterministic dynamics will be different. In this case $F(z)$, $G(z)$ are parametrized by non-minimal realizations and cancellations leading to subsystem transfer functions of individual degrees n_d, n_s, smaller than the overall dimension n in general occur.

The innovation representation of $\mathbf{y}$ can be split into the parallel a "deterministic" state-space model for $\mathbf{y}_d$

$$\mathbf{x}_d(t+1) = A_d\mathbf{x}_d(t) + B_d\mathbf{u}(t) \tag{2.7a}$$

$$\mathbf{y}_d(t) = C_d\mathbf{x}_d(t) + D\mathbf{u}(t). \tag{2.7b}$$

and the innovation representation of $\mathbf{y}_s$

$$\mathbf{x}_s(t+1) = A_s\mathbf{x}_s(t) + K_s\mathbf{e}_s(t) \tag{2.8a}$$

$$\mathbf{y}_s(t) = C_s\mathbf{x}_s(t) + \mathbf{e}_s(t) \tag{2.8b}$$

where $\mathbf{e}_s(t)$ is the one-step prediction error of the stochastic component $\mathbf{y}_s$ based on its own past, i.e. the innovation process of $\mathbf{y}_s$. Hence, the process $\mathbf{y}$ has a "canonical" block-diagonal innovation realization

$$\begin{bmatrix} \mathbf{x}_d(t+1) \\ \mathbf{x}_s(t+1) \end{bmatrix} = \begin{bmatrix} A_d & 0 \\ 0 & A_s \end{bmatrix} \begin{bmatrix} \mathbf{x}_d(t) \\ \mathbf{x}_s(t) \end{bmatrix} + \begin{bmatrix} B_d \\ 0 \end{bmatrix} \mathbf{u}(t) + \begin{bmatrix} 0 \\ K_s \end{bmatrix} \mathbf{e}_s(t) \tag{2.9a}$$

$$\mathbf{y}(t) = \begin{bmatrix} C_d & C_s \end{bmatrix} \begin{bmatrix} \mathbf{x}_d(t) \\ \mathbf{x}_s(t) \end{bmatrix} + D\mathbf{u}(t) + \mathbf{e}_s(t) \tag{2.9b}$$

where $\mathbf{x}_d(t)$ and $\mathbf{x}_s(t)$, are the *deterministic* and *stochastic* components of the state, mutually uncorrelated at all times. There is a nonsingular change of basis bringing (1.2) into a decoupled form of the type (2.9)[1]. In the input-output relation of (2.9), the transfer functions can then be parametrized as $F(z) = D + C_d(zI - A_d)^{-1}B_d$ and $G(z) = I + C_s(zI - A_s)^{-1}K_s$.

In general it may happen that, even if the realizations of the stochastic and deterministic components of $\mathbf{y}$ are individually minimal, the joint model is not, as there may be a loss of observability due to the presence of common

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

modes in the dynamics of the two subsystems. Hence in general a minimal realization takes the form

$$\begin{bmatrix} \check{\mathbf{x}}_d(t+1) \\ \mathbf{x}_0(t+1) \\ \check{\mathbf{x}}_s(t+1) \end{bmatrix} = \begin{bmatrix} \check{A}_d & 0 & 0 \\ 0 & A_0 & 0 \\ 0 & 0 & \check{A}_s \end{bmatrix} \begin{bmatrix} \check{\mathbf{x}}_d(t) \\ \mathbf{x}_0(t) \\ \check{\mathbf{x}}_s(t) \end{bmatrix} + \begin{bmatrix} \check{B}_d \\ B_0 \\ 0 \end{bmatrix} \mathbf{u}(t) + \begin{bmatrix} 0 \\ K_0 \\ \check{K}_s \end{bmatrix} \mathbf{e}_s(t) \tag{2.10}$$

$$\mathbf{y}(t) = \begin{bmatrix} \check{C}_d & C_0 & \check{C}_s \end{bmatrix} \begin{bmatrix} \check{\mathbf{x}}_d(t) \\ \mathbf{x}_0(t) \\ \check{\mathbf{x}}_s(t) \end{bmatrix} + D\mathbf{u}(t) + \mathbf{e}_s(t) \tag{2.11}$$

where $\mathbf{x}_0$, of dimension n_0, describes the common dynamics, $\dim \check{\mathbf{x}}_s = n_s - n_0 := \check{n}_s$, and $\dim \check{\mathbf{x}}_d = n_d - n_0 := \check{n}_d$. This brings down the dimension of the model (2.9) from $n_d + n_s = \check{n}_d + 2n_0 + \check{n}_s$, to $\check{n}_d + n_0 + \check{n}_s$ for the minimal description (2.10). In order to simplify the analysis, below we shall assume that there is no common dynamics between the "stochastic" and "deterministic" part.

Assumption 1 *The deterministic and stochastic subsystems (2.8) and (2.7) of the true model have no common dynamics, i.e. the sum of respective dimensions $n_d + n_s$ is equal to the dimension n, of the (minimal) joint model (1.2).*

This assumption is not necessary and can be avoided at the price of notational complications.

3 Canonical correlation analysis

In this section we shall investigate how the input signal affects the canonical correlation coefficients between the state space and the future input space.

From a system identification viewpoint, this analysis will point out which systems and which inputs lead to the worst conditioning of the subspace identification problem [7, 8]. To be precise, in system identification one is actually interested in studying the canonical correlation between the state and the *finite* future space $\mathcal{U}_{[t,T]}$, $T < \infty$. However here we shall make the simplifying assumption that $T - t$ is "very large", so that the angles between the finite future $\mathcal{U}_{[t,T]}$ and the state will not differ too much from those between the *infinite* future $\mathcal{U}_t^+$ and the state $\mathcal{X}_t^{+/-} := \text{span}\{\mathbf{x}(t)\}$ of the innovation model (1.2) .

We shall henceforth fix the present time to be $t = 0$, and hereafter suppress the present time subscript from all symbols.

Let

$$\mathcal{X}_d^{+/-} = E_{\|\mathcal{U}^+}\left[\mathcal{Y}_d^+ \mid \mathcal{U}^-\right] \subset \mathcal{U}^-$$

denote the oblique predictor space of the subspace spanned by the deterministic component $\mathbf{y}_d$ of the process $\mathbf{y}$ [17]. This is the state space of the model (2.7) of dimension n_d. Let $\mathbf{x}_d$ be a basis in $\mathcal{X}_d^{+/-}$, i.e. the deterministic component of the state (at time zero).

Let also $\mathcal{X}_s^{+/-} = \text{span}\{\mathbf{x}_s(0)\}$ be the state space off the "stochastic" component (in innovation form). Under the absence of feedback assumption the two subspaces $\mathcal{X}_s^{+/-}$ and $\mathcal{X}_d^{+/-}$ are orthogonal and moreover, under Assumption 1 the state space $\mathcal{X}^{+/-}$ can be decomposed in the orthogonal direct sum [2]

$$\mathcal{X}^{+/-} = \mathcal{X}_d^{+/-} \oplus \mathcal{X}_s^{+/-}.$$

Since by absence of feedback $\mathcal{X}_s^{+/-}$ is orthogonal to the whole input history $\mathcal{U}$, the (non zero) canonical correlation coefficients between $\mathcal{U}^+$ and $\mathcal{X}^{+/-}$ are the same as those between $\mathcal{U}^+$ and $\mathcal{X}_d^{+/-}$. For the computations of the canonical correlations we shall hence only need to consider the deterministic subsystem.

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 $\mathcal{U}^+$, the canonical correlation coefficients being ordered in decreasing magnitude. Let $\mathcal{X}_u^{+/-} := E\left[\mathcal{U}^+ \mid \mathcal{U}^-\right]$ be the forward predictor space (i.e. the state space of the innovation model) of the process $\mathbf{u}$. It is well-known that the canonical variables[3] of $\mathcal{U}^-$ for the pair of subspaces $(\mathcal{U}^-,\mathcal{U}^+)$, belong to $\mathcal{X}_u^{+/-}$ [1, 11, 15].

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

Lemma 3.1 *Let $r \geq n_d$. The following inequalities hold:*

$$\sigma_k(\mathcal{X}_d^{+/-},\mathcal{U}^+) \leq \sigma_k(\mathcal{U}^-,\mathcal{U}^+) \quad , \quad k = 1, 2, \ldots. \tag{3.1}$$

[2]Note that when Assumption 1 is not satisfied, we merely have the inclusion $\mathcal{X}^{+/-} \subseteq \mathcal{X}_d^{+/-} \oplus \mathcal{X}_s^{+/-}$.

[3]Also called *principal directions*.

Moreover

$$\sigma_k(\mathcal{X}_d^{+/-},\mathcal{U}^+) = \sigma_k(\mathcal{U}^-,\mathcal{U}^+) \quad k = 1, 2, \dots, n_d \tag{3.2}$$

if and only if the first n_d *canonical variables of* $\mathcal{U}^-$ *for the pair* $(\mathcal{U}^-,\mathcal{U}^+)$*, belong to* $\mathcal{X}_d^{+/-}$ *(and hence span* $\mathcal{X}_d^{+/-}$*).*

Let instead $r < n_d$*. Then the inequalities (3.1) hold for* $k = 1, 2, \dots, r$*. Moreover*

$$\sigma_k(\mathcal{X}_d^{+/-},\mathcal{U}^+) = \sigma_k(\mathcal{U}^-,\mathcal{U}^+) \quad k = 1, 2, \dots, r \tag{3.3}$$

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

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

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

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

Let

$$\begin{bmatrix} \mathbf{x}_1(t+1) \\ \mathbf{x}_2(t+1) \end{bmatrix} = A_u \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} + K_u \mathbf{e}_u(t) \tag{3.4a}$$

$$\mathbf{u}(t) = C_u \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} + \mathbf{e}_u(t) \tag{3.4b}$$

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

$$\begin{bmatrix} \mathbf{x}_1(t+1) \\ \mathbf{x}_2(t+1) \end{bmatrix} = \begin{bmatrix} (A_u - K_u C_u)_{11} & (A_u - K_u C_u)_{12} \\ (A_u - K_u C_u)_{21} & (A_u - K_u C_u)_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} + \begin{bmatrix} K_{u1} \\ K_{u2} \end{bmatrix} \mathbf{u}(t) \tag{3.5}$$

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

It is well known [1, 11, 15] that we can pick a basis $\mathbf{x}$ in $\mathcal{X}_u^{+/-}$ made of random variables which are proportional to the principal directions of $\mathcal{U}^-$ for the pair ($\mathcal{U}^-$,$\mathcal{U}^+$). In particular, we may pick a basis of *ordered* principal directions. A basis of this kind (with proper weights) leads to the so-called *stochastically balanced form* of the corresponding realization. If $\mathcal{X}$ is an invariant subspace of $\mathcal{X}_u^{+/-}$ spanned by the first n_1 principal components of $\mathcal{U}^-$ for the pair ($\mathcal{U}^-$,$\mathcal{U}^+$), we shall say that $\mathcal{X}$ is a *principal invariant subspace of* $\mathcal{X}_u^{+/-}$. In this case the eigenvalues of the upper left diagonal block $\lambda\{(A_u - K_u C_u)_{11}\}$, will be called the first n_1 *principal zeros* of the system (3.4). Principal zeros, like principal eigenvalues to be introduced later, remain invariant under principal truncation, i.e. extraction of the subsystem with state vector $\mathbf{x}_1$, defined by the upper-left block entries in (3.4) [15].

The following result then follows readily from the statement of Lemma 3.1 and provides a geometric solution to our problem.

Proposition 3.1 *Let $r \geq n_d$. The maximal canonical correlation coefficients (smallest canonical angles) between $\mathcal{X}_d^{+/-}$ and $\mathcal{U}^+$ are obtained when, and only when, $\mathcal{X}_d^{+/-}$ is a principal invariant subspace of $\mathcal{X}_u^{+/-}$.*

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

Theorem 3.1 *Given an input process of rational spectral density matrix Φ_u of degree $2r$, $r \geq n_d$, the maximal canonical correlation coefficients $\sigma_k(\mathcal{X}_d^{+/-}$,$\mathcal{U}^+)$ are obtained when, and only when there are n_d principal zeros of the (forward) innovation realization of* $\mathbf{u}$ *cancelling all the poles of the deterministic transfer function $F(z)$ of the system. Equivalently, the spectral density matrix Φ_u has n_d stable principal zeros which cancel all the poles of $F(z)$.*

Proof. Let (3.4) be a stochastically balanced innovation representation of $\mathbf{u}$ and let (A_d, B_d, C_d, D) be a minimal realization of the deterministic subsystem (with state space $\mathcal{X}_d^{+/-}$). As we have just seen, $\mathcal{X}_u^{+/-}$ admits a principal invariant subspace if and only if the matrix $A_u - K_u C_u$, has the block structure

$$(A_u - K_u C_u) = \begin{bmatrix} (A_u - K_u C_u)_{11} & 0 \\ (A_u - K_u C_u)_{21} & (A_u - K_u C_u)_{22} \end{bmatrix}, \qquad K_u = \begin{bmatrix} K_{u1} \\ K_{u2} \end{bmatrix}$$

moreover $\mathcal{X}_d^{+/-}$ is spanned by the first n_d canonical vectors of $\mathcal{X}_u^{+/-}$ if and only if $((A_u - K_u C_u)_{11},\ K_{u1})$ is similar to the pair (A_d, B_d). In other words, there exists a non singular $T \in \mathbb{R}^{n_d \times n_d}$ such that $A_d = T(A_u - K_u C_u)_{11} T^{-1}$ and $B_d = T K_{u1}$. In particular, $\mathcal{X}_u^{+/-}$ and $\mathcal{X}_d^{+/-}$ coincide if and only if $(A_u - K_u C_u, K_u)$ is similar to the pair (A_d, B_d).

The equivalence of the statement of Proposition 3.1 with the cancellation of the zero dynamics of the innovation realization (3.4) and the dynamics of the deterministic subsystem can be seen from the state space description of the cascade of (3.4) with the deterministic realization (A_d, B_d, C_d, D), namely

$$\begin{aligned} \begin{bmatrix} \mathbf{x}_d(t+1) \\ \mathbf{x}_1(t+1) \\ \mathbf{x}_2(t+1) \end{bmatrix} &= \begin{bmatrix} A_d & B_d C_{u1} & B_d C_{u2} \\ 0 & (A_u)_{11} & (A_u)_{12} \\ 0 & (A_u)_{21} & (A_u)_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_d(t) \\ \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} + \begin{bmatrix} B_d \\ K_{u1} \\ K_{u2} \end{bmatrix} \mathbf{e}_u(t) \\ \mathbf{y}_d(t) &= C_d \mathbf{x}_d(t) + D_d C_u \mathbf{x}(t) + D \mathbf{e}_u(t) \end{aligned}$$

from which, subtracting the second state component from the first, and recalling from the previous paragraph that, $\mathcal{X}_d^{+/-}$ is a principal invariant subspace of $\mathcal{X}_u^{+/-}$ if and only if we can substitute (A_d, B_d) with $((A_u - K_u C_u)_{11},\ K_{u1})$, it follows that

$$\mathbf{x}_d(t+1) - \mathbf{x}_1(t+1) = (A_u - K_u C_u)_{11} (\mathbf{x}_d(t) - \mathbf{x}_1(t))$$

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

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

whose only eigenvalues are those of the innovation realization (3.4). The dynamics of the deterministic system has been cancelled completely.
□

Remark 3.1 If the (deterministic) system is given and we are to design the spectrum of the "probing" input to get maximum ill-conditioning, it is enough to choose a spectral density of degree $2n_d$ so that the innovation model of $\mathbf{u}$ has dimension n_d and all of its zeros are (trivially) principal. In addition, we should choose the zero dynamics of the innovation model, i.e. of Φ_u, so as to cancel the dynamics of the deterministic system.

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

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

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

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

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

$$\begin{bmatrix} \mathbf{x}_{d1}(t+1) \\ \mathbf{x}_{d2}(t+1) \end{bmatrix} = A_d \begin{bmatrix} \mathbf{x}_{d1}(t) \\ \mathbf{x}_{d2}(t) \end{bmatrix} + B_d\mathbf{u}(t) \tag{3.6a}$$

$$\mathbf{u}(t) = H_u\mathbf{x}_{d1}(t) + \mathbf{e}_u(t) \tag{3.6b}$$

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

$$\begin{bmatrix} \mathbf{x}_{d1}(t+1) \\ \mathbf{x}_{d2}(t+1) \end{bmatrix} = \begin{bmatrix} (A_d)_{11} + B_{d1}H_u & (A_d)_{12} \\ (A_d)_{21} + B_{d2}H_u & (A_d)_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{d1}(t) \\ \mathbf{x}_{d2}(t) \end{bmatrix} + \begin{bmatrix} B_{d1} \\ B_{d2} \end{bmatrix} \mathbf{e}_u(t) \tag{3.7}$$

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

$\mathcal{X}_d^{+/-}$ iff there exists a non singular $T \in \mathbb{R}^{r \times r}$ such that for any minimal realization (C_u, A_u, K_u, I) of the innovation model of $\mathbf{u}$, it holds that $A_u = T((A_d)_{11} + B_{d1}H_u)T^{-1}$ and $K_u = TB_{d1}$. But this is the same as

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

Again, we can (and shall) pick a basis $\mathbf{x}_d$ in $\mathcal{X}_d^{+/-}$ made of random variables made of *ordered* principal directions for the pair $(\mathcal{X}_d^{+/-}, \mathcal{U}^+)$. If $\mathcal{X}_u^{+/-}$ is an invariant subspace of $\mathcal{X}_d^{+/-}$, then by (3.1) of Lemma 3.1, it is necessarily spanned by the first r principal components of $\mathcal{X}_d^{+/-}$ and hence it is automatically a *principal invariant subspace of* $\mathcal{X}_d^{+/-}$. In this case the eigenvalues of the upper left diagonal block, $\lambda\{(A_d)_{11}\}$, are the first r *principal eigenvalues* of the deterministic system.

Proposition 3.2 *Let $r < n_d$. The first r canonical correlation coefficients between $\mathcal{X}_d^{+/-}$ and $\mathcal{U}^+$ are maximal when, and only when, $\mathcal{X}_u^{+/-}$ is an invariant subspace of $\mathcal{X}_d^{+/-}$.*

The analogue of Theorem 3.1 in case $r < n_d$, is as follows.

Theorem 3.2 *With an input process of rational spectral density matrix Φ_u of degree $2r$, $r < n_d$, the first r canonical correlation coefficients $\sigma_k(\mathcal{X}_d^{+/-}, \mathcal{U}^+)$ are maximized when, and only when the deterministic subsystem of transfer function $F(z)$, admits r eigenvalues[4] which are all cancelled by the zeros of the forward innovation realization of* $\mathbf{u}$.

4 Matrix representation and computation of the canonical angles

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

$$\mathcal{H}_d := E^{\mathcal{X}_d^{+/-}}{}_{|\mathcal{U}^+} \tag{4.8}$$

where the subscript "|" means "restricted to". It is well-known that the canonical correlation coefficients between future inputs $\mathcal{U}^+$ and the deterministic state space $\mathcal{X}_d^{+/-}$ can be computed as the singular values of the Hankel operator $\mathcal{H}_d$.

[4]These eigenvalues are then necessarily principal.

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

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

where the last operator in the formula is the adjoint of the *observability operator* $\mathcal{O} := E^{\mathcal{U}^+}{}_{|\mathcal{X}_u^{+/-}}$, of the state space $\mathcal{X}_u^{+/-}$ for the process $\mathbf{u}$ [14, 15]. Introducing the projection operator

$$\mathcal{P} := E^{\mathcal{X}_d^{+/-}}{}_{|\mathcal{X}_u^{+/-}}, \qquad \mathcal{P}^* = E^{\mathcal{X}_u^{+/-}}{}_{|\mathcal{X}_d^{+/-}}$$

it is evident that $\mathcal{H}_d$ admits the factorization $\mathcal{H}_d = \mathcal{P}\mathcal{O}^*$.

Since the squared canonical correlation coefficients $\{\sigma_1^2, \sigma_2^2, \dots, \sigma_{n_d}^2\}$ of $\mathcal{U}^+$ and $\mathcal{X}_d^{+/-}$ are the eigenvalues of $\mathcal{H}_d^*\mathcal{H}_d$, i.e.,

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

using the factorization introduced above we obtain

$$\{\sigma_1^2, \sigma_2^2, \dots, \sigma_{n_d}^2\} = \lambda\{\mathcal{O}\mathcal{P}^*\mathcal{P}\mathcal{O}^*\} = \lambda\{\mathcal{O}^*\mathcal{O}\mathcal{P}^*\mathcal{P}\} \tag{4.9}$$

The following lemma provides an explicit matrix representation of the operator $\mathcal{O}^*\mathcal{O}\mathcal{P}^*\mathcal{P}$.

Lemma 4.1 *Let* $\mathbf{x}$ *be a basis in the forward predictor space* $\mathcal{X}_u^{+/-}$ *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 $P_{uu} := \mathrm{Var}\,\{\mathbf{x}\}$. *Then*

$$\mathcal{P}^*\mathcal{P}a'\mathbf{x} = a'P_{ud}P_{dd}^{-1}P_{du}P_{uu}^{-1}\mathbf{x} \qquad a \in \mathbb{R}^r \tag{4.10}$$

and

$$\mathcal{O}^*\mathcal{O}a'\mathbf{x} = a'P_{uu}\bar{P}_{uu}\mathbf{x} \qquad a \in \mathbb{R}^r \tag{4.11}$$

where $\bar{P}_{uu}$ *is the covariance of the dual (backward) basis in the backward predictor space* $\mathcal{U}^{-/+}$ *([14]). Hence the (squared) canonical correlation coefficients of* $\mathcal{U}^+$ *and* $\mathcal{X}_d^{+/-}$ *are given by*

$$\{\sigma_1^2, \sigma_2^2, \dots, \sigma_{n_d}^2\} = \lambda\{\mathcal{O}^*\mathcal{O}\mathcal{P}^*\mathcal{P}\} = \lambda\{P_{ud}P_{dd}^{-1}P_{du}\bar{P}_{uu}\} \tag{4.12}$$

Proof. Equation (4.10) follows from

$$E^{\mathcal{X}_u^{+/-}} E^{\mathcal{X}_d^{+/-}} a^T \mathbf{x} = E^{\mathcal{X}_u^{+/-}} a^T P_{ud} P_{dd}^{-1} \mathbf{x}_d = a^T P_{ud} P_{dd}^{-1} P_{du} P_{uu}^{-1} \mathbf{x}$$

To prove the second expression, recall that the *dual basis*, $\bar{\mathbf{x}}$, of $\mathbf{x}$ is a basis such that $E\,\bar{\mathbf{x}}\mathbf{x}' = I_r$ so that $\bar{\mathbf{x}} = P_{uu}^{-1}\mathbf{x}$. Recall also from stochastic realization theory [14, 18] that the projection of $\bar{\mathbf{x}}$ onto the future $\mathcal{U}^+$ gives exactly the natural basis, $\bar{\mathbf{x}}_+$, of "backward predictors" in $\mathcal{U}^{-/+}$ (i.e. the canonical vectors of $\mathcal{U}^+$)

$$E^{\mathcal{U}^+}\mathbf{x} = E^{\mathcal{U}^+} P_{uu}\bar{\mathbf{x}} = P_{uu}\bar{\mathbf{x}}_+$$

and that

$$E^{\mathcal{X}_u^{+/-}} \bar{\mathbf{x}}_+ = E^{\mathcal{X}_u^{+/-}} \bar{P}_{uu}\mathbf{x}_+ = \bar{P}_{uu}\mathbf{x}$$

which together imply (4.11).

Since the canonical correlation coefficients are the eigenvalues of the matrix representation of (4.9), (4.12) follows.
□

Remark.
A (stochastically balanced) basis $\mathbf{x}$ in $\mathcal{X}_u^{+/-}$ is made of random variables which are proportional to the principal directions of $\mathcal{U}^-$ [11, 15]. If the first n_d 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 $\operatorname{span}\{\mathbf{x}_-\} = \mathcal{X}_d^{+/-}$ and $\mathbf{x}_\perp \perp \mathcal{X}_d^{+/-}$.

Now, when $\mathcal{X}_u^{+/-} \supset \mathcal{X}_d^{+/-}$ we can choose $\mathbf{x}_d = \mathbf{x}_-$ whereby $P_{ud} = [P_- 0] = P'_{du}$ where $P_- := \operatorname{Var}\mathbf{x}_-$. In this situation $P_{uu} = \operatorname{diag}\{P_-,\ P_\perp\}$ and it can be checked that also $\bar{P}_{uu}$ has a block diagonal structure, namely $\bar{P}_{uu} = \operatorname{diag}\{\bar{P}_+,\ \bar{P}_\perp\}$. Hence

$$P_{ud}P_{dd}^{-1}P_{du}\bar{P}_{uu} = [P_- 0]P_-^{-1}\begin{bmatrix} P_- \\ 0 \end{bmatrix}\begin{bmatrix} \bar{P}_+ & 0 \\ 0 & \bar{P}_\perp \end{bmatrix} = P_-\bar{P}_+$$

and, by a well-known result in stochastic system theory, [10, 11, 18], formula (4.12) reduces to

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

where $\sigma_{u,i}^2$ are the first n_d canonical correlation coefficients of past and future inputs. In other words, $\sigma_i^2 = \sigma_{u,i}^2$, $i = 1, \ldots, n_d$. This agrees with the results of the previous section stating that the maximal canonical correlation coefficients (smallest canonical angles) between $\mathcal{X}_d^{+/-}$ and $\mathcal{U}^+$ are obtained when $\mathcal{X}_u^{+/-} \supset \mathcal{X}_d^{+/-}$.

References

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

[2] T.W. Anderson, *An introduction to multivariate statistical analysis*, John Wiley and Sons, 1962.

[3] P.E. Caines and C.W. Chan, *Estimation, identification and feedback*, System Identification: Advances and Case Studies (R. Mehra and D. Lainiotis, eds.), Academic, 1976, pp. 349–405.

[4] A. Chiuso and G. Picci, *Probing inputs for subspace identification*, Proceedings of the 2000 Conference on Decision and Control (Sydney, Australia), Dec. 2000, pp. paper CDC00–INV0201.

[5] ———, *On the ill-conditioning of subspace identification with inputs*, Tech. Report TRITA/MATH-01-OS5, Department of Mathematics, Royal Institute of Technology, Stockholm, Sweden, 2001, submitted for pubblication.

[6] ———, *Geometry of oblique splitting, minimality and hankel operators*, Lect. Notes in Control and information Sciences, no. 286, pp. 85–124, Springer, 2002.

[7] ———, *Subspace identification by data orthogonalization and model decoupling*, submitted to Automatica (2003).

[8] ———, *Subspace identification by orthogonalization and model decoupling*, Proc. of the IFAC Int. Symposium on System Identification (SYSID) (Rotterdam), August 2003.

[9] ———, *The asymptotic variance of subspace estimates*, Journal of Econometrics **118** (2004), pp. 257–291.

[10] U.B. Desai and D. Pal, *A realization approach to stochastic model reduction*, IEEE Transactions Automatic Control **29** (1984), 1097–1100.

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

[12] M.R. Gevers and B.D.O. Anderson, *Representation of jointly stationary feedback free processes*, Intern. Journal of Control **33** (1981), 777–809.

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

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

[15] ______, *Canonical correlation analysis, approximate covariance extension and identification of stationary time series*, Automatica **32** (1996), 709–733.

[16] G. Picci, *Stochastic realization and system identification*, Statistical Methods in Control and Signal Processing (T. Katayama and I. Sugimoto, eds.), M. Dekker, N.Y., 1997, pp. 205–240.

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

[18] G. Picci and S. Pinzoni, *Acausal models and balanced realizations of stationary processes*, Linear Algebra and its Applications **205-206** (1994), 957–1003.

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