Canonical Correlations Between Input and Output Processes of Linear Stochastic Models

Katrien De Cock and Bart De Moor[1]
K.U.Leuven
Dept. of Electrical Engineering (ESAT–SCD)
Kasteelpark Arenberg 10
B–3001 Leuven (Heverlee), Belgium
Tel. +32/(0)16.32.17.09, Fax +32/(0)16.32.19.70
http://www.esat.kuleuven.ac.be/sista-cosic-docarch
decock@esat.kuleuven.ac.be, demoor@esat.kuleuven.ac.be

Abstract

In this paper, we obtain expressions for the principal angles between the row spaces of input and output data block Hankel matrices of a linear stochastic model in terms of the model parameters. The canonical correlations of the corresponding processes are equal to the limiting values of the cosines of the principal angles. From these parametric expressions, the relations between the different sets of canonical correlations can be easily deduced.

1 Introduction

Canonical correlation analysis (CCA) is a well developed tool in statistical analysis that is used for measuring the linear relationship between two sets of random variables. It was developed by H. Hotelling [10]. Although a wide variety of applications exists in econometrics, biometrics, chemometrics, statistics, meteorology, etc., the technique has only got introduced quite recently in the communities of signal processing, system theory and identification and neural networks [4, 14, 20]. In a classic paper by Gel'fand and Yaglom [9], CCA is extended to stochastic processes and related to the notion of mutual information, a concept from information theory that is closely related to CCA and that was introduced by Shannon [18] in 1948. A slightly different interpretation in terms of channel capacity and information rate is given in [17]. Another area where CCA is applied, is stochastic realization and identification of dynamical models [1, 3, 5, 11, 12, 15, 16, 21, 22]. The order of the model and a state sequence can be derived from the canonical correlations and the canonical variates of the past and the future output data.

[1]Katrien De Cock is a research assistant at the K.U.Leuven. Dr. Bart De Moor is a full professor at the K.U.Leuven. Our research is supported by grants from several funding agencies and sources: Research Council KUL: Concerted Research Action GOA-Mefisto 666 (Mathematical Engineering), IDO (IOTA Oncology, Genetic networks), several PhD, postdoc & fellow grants; Flemish Government: Fund for Scientific Research Flanders (several PhD and postdoc grants, projects G.0256.97 (subspace), G.0115.01 (bio-i and microarrays), G.0240.99 (multilinear algebra), G.0197.02 (power islands), G.0407.02 (support vector machines), research communities ICCoS, ANMMM), AWI (Bil. Int. Collaboration Hungary/ Poland), IWT (Soft4s (softsensors), STWW-Genprom (gene promotor prediction), GBOU-McKnow (Knowledge management algorithms), Eureka-Impact (MPC-control), Eureka-FLiTE (flutter modeling), several PhD grants); Belgian Federal Government: DWTC (IUAP IV-02 (1996-2001) and IUAP V-22 (2002-2006): Dynamical Systems and Control: Computation, Identification & Modelling), Program Sustainable Development PODO-II (CP/40: Sustainibility effects of Traffic Management Systems); Direct contract research: Verhaert, Electrabel, Elia, Data4s, IPCOS;

In this paper we will work with the geometric interpretation of canonical correlation analysis, as is usually done in the subspace identification literature, see e.g. [22]. The canonical correlations and the canonical variates are respectively equal to the cosines of the principal angles between and the principal vectors in two linear subspaces. These subspaces are the row spaces of block Hankel matrices obtained by stacking the measured input and output sequences. In this way it is straightforward and computationally efficient to compute an approximation of the canonical correlations between two measured processes, of which in practice, only a finite amount of data is available. Meanwhile, we are able to give expressions for the real canonical correlations, viz. the asymptotic values for infinite data.
The paper is organized as follows. In Section 2 we describe the models we will work with. The principal angles between two subspaces are defined in Section 3. In Section 4 we discuss the principal angles and canonical correlations between the past and future input and output spaces, respectively processes, of a linear stochastic model.

2 Model class

We describe in Section 2.1 the state space representation of the model class that we will work with throughout the paper. We also give the assumptions on the different stochastic processes involved. We define the controllability and observability matrices and Gramians of the model and the inverse model in Section 2.2. In Section 2.3 we introduce the past and future input and output block Hankel matrices.

2.1 State space representation

The forward innovation representation of a given stationary stochastic process $\{y(k)\}_{k\in\mathbb{Z}}$ with m components, i.e. $y(k) \in \mathbb{R}^m \ \forall k$, is the following

$$\left\{ \begin{array}{rcl} x(k+1) &=& Ax(k) + Ku(k) \ , \\ y(k) &=& Cx(k) + u(k) \ . \end{array} \right. \tag{2.1}$$

The process $\{x(k)\}_{k\in\mathbb{Z}} \in \mathbb{R}^n$ is the state process associated to this model, where n is the model order, and $A \in \mathbb{R}^{n\times n}, C \in \mathbb{R}^{m\times n}$ are the system and output matrices, respectively. The matrix $K \in \mathbb{R}^{n\times m}$ is the Kalman gain.
We will denote the model (2.1) by the threesome (A, K, C).
Its Markov parameters are denoted by the matrices $H(k), \ k \geq 0$:

$$\left\{ \begin{array}{ll} H(0) & = I_m \\ H(k) & = CA^{k-1}K \quad \text{for } k > 0 \end{array} \right. \tag{2.2}$$

The model has the following properties. The input process $\{u(k)\}_{k\in\mathbb{Z}}$, i.e. the innovation process of the stochastic process $\{y(k)\}_{k\in\mathbb{Z}}$, is an m-component zero-mean, stationary, white stochastic process with full rank covariance matrix $S_u \in \mathbb{R}^{m\times m}$. Its autocovariance function $R_u(\tau) = E\left\{u(k+\tau)u(k)^T\right\}$ is thus equal to[2] $R_u(\tau) = S_u\delta(\tau)$. The state process $\{x(k)\}_{k\in\mathbb{Z}}$ is a zero-mean,

[2] $\delta(\tau)$ is the Kronecker delta: $\delta(0) = 1$ and $\delta(\tau) = 0 \ \forall \tau \neq 0$.

stationary and ergodic stochastic process with covariance matrix $E\{x(k)x(k)^T\} = \Sigma \in \mathbb{R}^{n\times n}$, which satisfies the Lyapunov equation

$$\Sigma = A\Sigma A^T + KS_uK^T \ . \tag{2.3}$$

Furthermore, the state $x(k)$ is independent of the present and all future inputs. Consequently, $E\{u(k+\tau)^T x(k)\} = 0$ for $\tau \geq 0$ and $E\{u(k+\tau)^T y(k)\} = 0$ for $\tau > 0$. The system in (2.1) is stable and strictly minimum phase. This means that all the poles and zeros of the model are less than one in modulus. The inverse model is then also stable and minimum phase. Its state space description is readily derived from (2.1):

$$\begin{cases} x(k+1) &= (A-KC)x(k) + Ky(k) \ , \\ u(k) &= -Cx(k) + y(k) \ . \end{cases}$$

The state space matrices of the inverse model are denoted by (A_z, K_z, C_z):

$$(A_z, K_z, C_z) = (A - KC, K, -C) \ . \tag{2.4}$$

The Markov parameters of the inverse model are denoted by $H_z(k)$ and they are equal to

$$\begin{cases} H_z(0) &= I_m \ , \\ H_z(k) &= -C(A-KC)^{k-1}K \quad \text{for } k > 0 \ . \end{cases} \tag{2.5}$$

2.2 The controllability and observability matrices and Gramians

The controllability matrix $\mathcal{C}_i$ of the forward innovation model (2.1) is defined as

$$\mathcal{C}_i = \begin{pmatrix} K & AK & A^2K & \cdots & A^{i-1}K \end{pmatrix} \ ,$$

and its observability matrix Γ_i is

$$\Gamma_i = \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{i-1} \end{pmatrix} \ . \tag{2.6}$$

The controllability Gramian P of the forward innovation model (2.1) is defined as the solution of the controllability Lyapunov equation

$$P = APA^T + KK^T \ , \tag{2.7}$$

while the observability Gramian Q follows from the observability Lyapunov equation

$$Q = A^TQA + C^TC \ . \tag{2.8}$$

Since the model is stable and minimal, the matrices P and Q are the unique and positive definite solutions of the respective equations. The explicit solution for P is of the form

$$P = \sum_{k=0}^{\infty} A^k KK^T (A^k)^T = \mathcal{C}_\infty \mathcal{C}_\infty^T \ ,$$

where $\mathcal{C}_\infty$ is the infinite controllability matrix. Similarly, the observability Gramian Q can be obtained as

$$Q = \sum_{k=0}^{\infty} (A^k)^T C^T C A^k = \Gamma_\infty^T \Gamma_\infty \ ,$$

where Γ_∞ is the infinite observability matrix of the model. We will also need the observability matrix of the inverse model, denoted by Γ_{z_i}:

$$\Gamma_{z_i} = \begin{pmatrix} -C \\ -C(A-KC) \\ -C(A-KC)^2 \\ \vdots \\ -C(A-KC)^{i-1} \end{pmatrix} ,$$

where the subscript i in Γ_{z_i} denotes the number of block rows.
The observability Gramian of the inverse model is denoted by Q_z and it is equal to

$$Q_z = \Gamma_{z_\infty}^T \Gamma_{z_\infty} \ . \tag{2.9}$$

It is the solution of the observability Lyapunov equation for the inverse model

$$Q_z = (A-KC)^T Q_z (A-KC) + C^T C \ . \tag{2.10}$$

2.3 Data block Hankel matrices

We define the input and output block Hankel matrices U and Y. These matrices play an important role in the computation of the canonical correlations. The output block Hankel matrix Y is defined as

$$Y = \frac{1}{\sqrt{j}} \left(\begin{array}{ccccc} y(0) & y(1) & y(2) & \cdots & y(j-1) \\ y(1) & y(2) & y(3) & \cdots & y(j) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ y(i-1) & y(i) & y(i+1) & \cdots & y(i+j-2) \\ \hline y(i) & y(i+1) & y(i+2) & \cdots & y(i+j-1) \\ y(i+1) & y(i+2) & y(i+3) & \cdots & y(i+j) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ y(2i-1) & y(2i) & y(2i+1) & \cdots & y(2i+j-2) \end{array} \right) \tag{2.11}$$

$$= Y_{0|2i-1} = \left(\frac{Y_{0|i-1}}{Y_{i|2i-1}} \right) = \left(\frac{Y_p}{Y_f} \right) \in \mathbb{R}^{2mi \times j} \ , \tag{2.12}$$

where

- The number of columns (j) is typically equal to $K - 2i + 1$, where K is the total number of data samples, which implies that all given data samples are used. For statistical reasons we will assume that $j, K \to \infty$ throughout this paper.

- The subscripts of $Y_{0|2i-1}, Y_{0|i-1}, Y_{i+1|2i-1}$ denote the subscript of the first and last element of the first column in the block Hankel matrix. The subscript 'p' stands for 'past' and the subscript 'f' for 'future'.

The input block Hankel matrices $U_{0|2i-1}, U_p, U_f$ are defined in a similar way.
We will also need the state sequence matrix, which is defined as

$$X_i = \frac{1}{\sqrt{j}} \begin{pmatrix} x(i) & x(i+1) & \cdots & x(i+j-1) \end{pmatrix} , \tag{2.13}$$

where the subscript i denotes the subscript of the first element of the state sequence. Analogously to the past inputs and outputs, we denote the past state sequence by X_p and the future state sequence by X_f: $X_p = X_0 \in \mathbb{R}^{n \times j}$ and $X_f = X_i \in \mathbb{R}^{n \times j}$. The state space equations (2.1) can now be formulated in terms of data block Hankel matrices as follows

$$X_f = A^i X_p + \Delta_i U_p , \tag{2.14}$$

$$Y_p = \Gamma_i X_p + H_i U_p , \tag{2.15}$$

$$Y_f = \Gamma_i X_f + H_i U_f , \tag{2.16}$$

where $\Delta_i \in \mathbb{R}^{n \times mi}$ is the reversed controllability matrix:

$$\Delta_i = \begin{pmatrix} A^{i-1}K & A^{i-2}K & \cdots & AK & K \end{pmatrix} ,$$

the matrix $\Gamma_i \in \mathbb{R}^{mi \times n}$ is the observability matrix of the model (see (2.6)) and the matrix $H_i \in \mathbb{R}^{mi \times mi}$ is a block lower triangular and block Toeplitz matrix with the Markov parameters of the model (the impulse response sequence) as its elements:

$$H_i = \begin{pmatrix} I_m & 0 & 0 & \cdots & 0 \\ CK & I_m & 0 & \cdots & 0 \\ CAK & CK & I_m & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{i-2}K & CA^{i-3}K & CA^{i-4}K & \cdots & I_m \end{pmatrix} . \tag{2.17}$$

From (2.15) or (2.16) it immediately follows that the observability matrix of the inverse model (see (2.18)) is equal to

$$\Gamma_{z_i} = -H_i^{-1}\Gamma_i . \tag{2.18}$$

Note that the input covariance matrix $\lim_{j\to\infty} U_p U_p^T = \lim_{j\to\infty} U_f U_f^T$, which will be denoted by Q_{u_i} is a block diagonal matrix with diagonal blocks all equal to S_u. By using the state sequence matrices, we can write the state covariance matrix Σ as $\Sigma = \lim_{j\to\infty} X_p X_p^T = \lim_{j\to\infty} X_f X_f^T$. The fact that the states are uncorrelated with the present and future inputs and that the output is uncorrelated with the future inputs, translates to

$$\begin{cases} \lim_{j\to\infty} X_p U^T & = & 0 \\ \lim_{j\to\infty} X_f U_f^T & = & 0 \end{cases} \quad \text{and} \quad \lim_{j\to\infty} Y_p U_f^T = 0 , \tag{2.19}$$

respectively.

3 Principal angles between and principal directions in subspaces

The concept of principal angles between subspaces of linear vector spaces is due to Jordan [13] in the 19th century. In the area of systems and control, the principal angles between and the principal directions in two subspaces are used in subspace identification methods [22] and also in model updating [7] and damage location [8]. In the latter two applications, one starts from a finite element model and measurements of a certain mechanical structure and one tries to find the subset of parameters of the model that should be adapted to explain the measurements, which is done by computing the principal angles between a certain measurement space and the parameterized space. In that way, damage to the structure can be located. The subspace-based fault detection algorithm of Basseville et al. [2], on the other hand, is based on linear dynamical models, the type of models that we deal with. Changes in the eigenmodes of the observed system are determined by monitoring the difference between the column spaces of the observability matrix of the nominal linear dynamical model and the observability matrix of the model that can be identified from the measurements. The difference between the column spaces can be quantified by the principal angles between the subspaces.

The principal angles between and principal directions in two subspaces S_1 and S_2 are defined as follows.

Definition 3.1. The principal angles between and principal directions in two subspaces *Let S_1 and S_2 be subspaces of dimension p and q, respectively, where $p \leq q$. Then, the p principal angles between S_1 and S_2, denoted by $\theta_1, \ldots, \theta_p$, and the corresponding principal directions $u_i \in S_1$ and $v_i \in S_2$ are recursively defined as*

$$\cos\theta_1 = \max_{u \in S_1} \max_{v \in S_2} |u^T v| = u_1^T v_1$$
$$\cos\theta_k = \max_{u \in S_1} \max_{v \in S_2} |u^T v| = u_k^T v_k \quad (k = 2, \ldots, p)$$

subject to $\|u\| = \|v\| = 1$, and for $k > 1$: $u^T u_i = 0$ and $v^T v_i = 0$, where $i = 1, \ldots, k-1$.

Let $A \in \mathbb{R}^{p \times n}$ be of rank r_a and $B \in \mathbb{R}^{q \times n}$ of rank r_b, where $r_a < r_b$. Then, the ordered set of r_a principal angles between the row spaces of A and B is denoted by

$$(\theta_1, \theta_2, \ldots, \theta_{r_a}) = [A \sphericalangle B] \ .$$

Assume that the matrices A and B are of full row rank and that $p \leq q$. Then, the squared cosines of the principal angles between their row spaces can be computed as the eigenvalues of $(AA^T)^{-1}AB^T(BB^T)^{-1}BA^T$:

$$\cos^2 [A \sphericalangle B] = \lambda\left((AA^T)^{-1}AB^T(BB^T)^{-1}BA^T\right) \ . \qquad (3.20)$$

Since we will have to compute the principal angles between subspaces in $\mathbb{R}^n$, where n is a large number, e.g. 10 000, it is useful to have an efficient algorithm. We present here an algorithm that is based on the LQ decomposition, which is first defined.

Definition 3.2. The LQ factorization of a matrix
The LQ factorization of a real $m \times n$ matrix A is given by

$$A = LQ^T \ ,$$

where $Q \in \mathbb{R}^{n \times n}$ is orthogonal and $L \in \mathbb{R}^{m \times n}$ is lower triangular.

Note that the LQ decomposition of a matrix A boils down to the QR decomposition of A^T, which is the numerical version of the Gram-Schmidt orthogonalization (see e.g. [19]).
It can be shown (see e.g. [6]) that the principal angles between two full row rank matrices $A \in \mathbb{R}^{p \times n}$ and $B \in \mathbb{R}^{q \times n}$, where $p \leq q$ and $p + q \leq n$, can be computed as follows.

1. Compute the triangular part of the LQ factorization of the matrix $\begin{pmatrix} A \\ B \end{pmatrix}$. The triangular part is denoted by
$$\begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix} \in \mathbb{R}^{(p+q) \times (p+q)} \ ,$$
where $L_{11} \in \mathbb{R}^{p \times p}$, $L_{21} \in \mathbb{R}^{q \times p}$ and $L_{22} \in \mathbb{R}^{q \times q}$.

2. Compute the triangular part of the LQ factorization of $\begin{pmatrix} L_{21} & L_{22} \end{pmatrix}$:
$$\begin{pmatrix} L_{21} & L_{22} \end{pmatrix} = \begin{pmatrix} S & 0 \end{pmatrix} T \ .$$
The resulting lower triangular matrix $S \in \mathbb{R}^{q \times q}$ is non-singular.

3. The cosines of the principal angles between row(A) and row(B) are the singular values of the matrix $S^{-1}L_{21}$.

The above described computational scheme leads to a very simple Matlab program, which is given in Table 1.

```
function cosines = cosines_lq(A,B)

p = size(A,1);
q = size(B,1);
L = triu(qr([A;B]'));
L = L(1:p+q,p+1:p+q);
S = triu(qr(L));
S = S(1:q,:);
L = L(1:p,:);
cosines = svd(L/S);
```

Table 1: The Matlab program `cosines_lq.m` for the computation of the principal angles between the row spaces of the matrices A and B.

4 Principal angles and canonical correlations of input and output

In this section we compute the principal angles between the past and future input and output spaces and the canonical correlations of the corresponding processes. In Section 4.1 we first describe the future and the past of a stochastic process. We show how the canonical correlations of the processes will be computed and indicate how they are related. In Section 4.2, we derive expressions for the principal angles between different combinations of past and future input and output spaces, i.e. the row spaces of the $mi \times j$ data block Hankel matrices, where we assume $j \to \infty$. The expressions are in terms of the system matrices (A, K, C) and the input covariance matrix S_u. As we will see, the principal angles converge for $i \to \infty$. The cosines of the limiting principal angles are the canonical correlations of the corresponding processes. In Section 4.4 the relations between the different canonical correlations are derived.

4.1 Introduction

4.1.1 Past and future input and output processes of a linear model

Let $\{u(k)\}_{k\in\mathbb{Z}}$ and $\{y(k)\}_{k\in\mathbb{Z}}$ denote the input and output process of a linear stochastic model in forward innovation form. We assume that the processes are zero-mean, stationary and ergodic. The past output process is defined as

$$y_p = \{y(k)\ (k < 0)\}\ , \tag{4.21}$$

and the future output process is

$$y_f = \{y(k)\ (k \geq 0)\}\ . \tag{4.22}$$

Analogous definitions hold for the past and the future input process, u_p and u_f, respectively.

4.1.2 The canonical correlations of the past and future input and output processes

The canonical correlations of the past and future input and output processes are defined as the canonical correlations of the corresponding random variables $\mathcal{U}_{-1|-\infty}$, $\mathcal{U}_{0|\infty}$, $\mathcal{Y}_{-1|-\infty}$ and $\mathcal{Y}_{0|\infty}$, where

$$\mathcal{Y}_{-1|-\infty} = \begin{pmatrix} y(-1) \\ y(-2) \\ \vdots \end{pmatrix} \quad \text{and} \quad \mathcal{Y}_{0|\infty} = \begin{pmatrix} y(0) \\ y(1) \\ \vdots \end{pmatrix}\ ,$$

and analogously for $\mathcal{U}_{-1|-\infty}$ and $\mathcal{U}_{0|\infty}$. For example, the canonical correlations of the past and future of the output process are equal to

$$\mathrm{cc}(y_p, y_f) = \mathrm{cc}(\mathcal{Y}_{-1|-\infty}, \mathcal{Y}_{0|\infty})\ . \tag{4.23}$$

Due to the stationarity and ergodicity of the processes, the canonical correlations are equal to the cosines of the principal angles between the row spaces of the doubly infinite block Hankel matrices (see (2.11) and (2.12)):

$$\mathrm{cc}(y_p, y_f) = \lim_{j\to\infty} \cos\left[Y_{-1|-\infty} \sphericalangle Y_{0|\infty}\right]\ .$$

We can already treat three trivial cases:

1. Due to the independence of the past and future input processes, the canonical correlations of u_p and u_f are all equal to 0:

$$\boxed{\mathrm{cc}(u_p, u_f) = 0, 0, \ldots}$$

2. The future input is also independent of the past output process. Consequently, their canonical correlations are all equal to 0:

$$\boxed{\mathrm{cc}(y_p, u_f) = 0, 0, \ldots}$$

3. The output at a certain time step k is a linear combination of the present input and all past inputs:

$$y(k) = \sum_{i=1}^{\infty} CA^{i-1}Ku(k-i) + u(k) = \sum_{i=0}^{\infty} H(i)u(k-i) \ ,$$

where $H(i)$ is the ith Markov parameter of the linear model (see (2.2)). Consequently, all random variables in $\mathcal{Y}_{-1|-\infty}$ can be obtained as linear combinations of the random variables in $\mathcal{U}_{-1|-\infty}$. Otherwise formulated, the row space of $Y_{-1|-\infty}$ is contained in the row space of $U_{-1|-\infty}$:

$$\mathrm{row}(Y_{-1|-\infty}) \subseteq \mathrm{row}(U_{-1|-\infty}) \ . \tag{4.24}$$

The canonical correlations of the past input and past output processes are consequently all equal to 1:

$$\boxed{\mathrm{cc}(u_p, y_p) = 1, 1, \ldots} \tag{4.25}$$

Moreover, by applying the same reasoning to the inverse model, we obtain $u(k)$ as a linear combination of the present and past outputs:

$$u(k) = -\sum_{i=1}^{\infty} C(A-KC)^{i-1}Ky(k-i) + y(k) = \sum_{i=0}^{\infty} H_z(i)y(k-i) \ ,$$

where $H_z(i)$ are the Markov parameters of the inverse model (see (2.5)). This leads to

$$\mathrm{row}(U_{-\infty|-1}) \subseteq \mathrm{row}(Y_{-\infty|-1}) \ . \tag{4.26}$$

It follows from (4.24) and (4.26) that the past input and the past output process span the same space:

$$\mathrm{row}(Y_{-\infty|-1}) = \mathrm{row}(U_{-\infty|-1}) \ . \tag{4.27}$$

The canonical correlations of the other combinations of past and future input and output processes can be obtained as the following limits:

$$\mathrm{cc}(u_f, y_f) = \lim_{i\to\infty} \mathrm{cc}(\mathcal{U}_{0|i-1}, \mathcal{Y}_{0|i-1}) \ , \tag{4.28a}$$

$$\mathrm{cc}(u_p, y_f) = \lim_{i\to\infty} \mathrm{cc}(\mathcal{U}_{-i|-1}, \mathcal{Y}_{0|i-1}) \ , \tag{4.28b}$$

$$\mathrm{cc}(y_p, y_f) = \lim_{i\to\infty} \mathrm{cc}(\mathcal{Y}_{-i|-1}, \mathcal{Y}_{0|i-1}) \ , \tag{4.28c}$$

where

$$\mathcal{Y}_{-i|-1} = \begin{pmatrix} y(-i) \\ y(-i+1) \\ \vdots \\ y(-1) \end{pmatrix} \text{ and } \mathcal{Y}_{0|i-1} = \begin{pmatrix} y(0) \\ y(1) \\ \vdots \\ y(i-1) \end{pmatrix} ,$$

and analogously for the input random variables. The parameter i describes how far we go back into the past ($k = -i$) and forward into the future ($k = i-1$), where the present is at $k = 0$. Since the processes are stationary, we can as well take the present at time instant $k = i$, the past from $k = 0$ to $k = i-1$ and the future from $k = i$ to $k = 2i-1$. This is only a convention that allows us to estimate the canonical correlations from measured data sequences. The canonical correlations of the past and future output processes, e.g., can then be computed as

$$\text{cc}(y_p, y_f) = \lim_{i\to\infty} \text{cc}(\mathcal{Y}_{0|i-1}, \mathcal{Y}_{i|2i-1}) .$$

Due to the stationarity and ergodicity of the processes, the canonical correlations of $\mathcal{Y}_{0|i-1}$ and $\mathcal{Y}_{i|2i-1}$ can be obtained as the cosines of the principal angles between the row spaces of the $mi \times j$ data block Hankel matrices $Y_{0|i-1}$ and $Y_{i|2i-1}$, provided $j \to \infty$. These block Hankel matrices are equal to the past and future output block Hankel matrices Y_p and Y_f, which are defined in (2.12). Consequently, we can compute the canonical correlations of the combinations of past and future processes in (4.28a–4.28c) as follows:

$$\text{cc}(u_f, y_f) = \lim_{i\to\infty} \lim_{j\to\infty} \cos\left[U_f \sphericalangle Y_f\right] , \tag{4.29a}$$

$$\text{cc}(u_p, y_f) = \lim_{i\to\infty} \lim_{j\to\infty} \cos\left[U_p \sphericalangle Y_f\right] , \tag{4.29b}$$

$$\text{cc}(y_p, y_f) = \lim_{i\to\infty} \lim_{j\to\infty} \cos\left[Y_p \sphericalangle Y_f\right] . \tag{4.29c}$$

This explains why we denote the first i block rows of the output block Hankel matrix by Y_p and the following i block rows by Y_f (see (2.12)), and similarly for U_p and U_f.

4.1.3 Overview of the relations between the canonical correlations

From Equation (4.27) we can already deduce that the canonical correlations of the past input and future output are equal to the canonical correlations of the past and future output:

$$\text{cc}(u_p, y_f) = \text{cc}(y_p, y_f) .$$

The parametric expressions that we derive in Section 4.2, will also reveal that the canonical correlations of the future input and future output are related to the canonical correlations of the past and future output in the following way:

$$\text{cc}^2(u_f, y_f) = 1 - \text{cc}^2(y_p, y_f) .$$

The corresponding principal angles are complementary:

$$\left[U_f \sphericalangle Y_f\right] = \frac{\pi}{2} - \left[Y_p \sphericalangle Y_f\right] \quad \text{for } i, j \to \infty .$$

An overview of the relations of the canonical correlations of the different combinations of processes is given in Table 2. The canonical correlations of the past and the future output are denoted by ρ_k in this table.

	u_p	u_f	y_p	y_f
u_p	1	0	1	ρ_k
u_f	0	1	0	$\sqrt{1-\rho_k^2}$
y_p	1	0	1	ρ_k
y_f	ρ_k	$\sqrt{1-\rho_k^2}$	ρ_k	1

Table 2: Overview of the relations between the different sets of canonical correlations.

4.2 The principal angles between the input and output spaces

Based on the state space equations (2.14)–(2.16), the properties in (2.19) and (3.20) the following expressions are derived for the principal angles between the past and future input and output spaces. From these expressions, the canonical correlations of the corresponding processes are deduced in Section 4.3. We only give the results. The computations can be found in [6].

The principal angles between row(U_f) and row(Y_f)

The squared cosines of the largest n principal angles between row(U_f) and row(Y_f) for $j \to \infty$ and finite i are the eigenvalues of $(I_n + \mathcal{G}_{z_i}\Sigma)^{-1}$, where $\mathcal{G}_{z_i}$ is equal to

$$\begin{aligned} \mathcal{G}_{z_i} &= \Gamma_{z_i}^T Q_{u_i}^{-1} \Gamma_{z_i} \\ &= \sum_{k=0}^{i-1} (A-KC)^{k^T} C^T S_u^{-1} C (A-KC)^k \ . \end{aligned} \tag{4.30}$$

The other $mi - n$ principal angles are equal to zero:

$$\boxed{\lim_{j\to\infty} \cos^2 [U_f \sphericalangle Y_f] = \lambda\left((I_n + \mathcal{G}_{z_i}\Sigma)^{-1}\right), \underbrace{1,\ldots,1}_{mi-n} \ .} \tag{4.31}$$

Remark 4.1. $\mathcal{G}_{z_i}$ as the solution of a Lyapunov equation

If the state space matrices (A, K, C) and the input covariance matrix S_u are known, then the matrix $\mathcal{G}_{z_i}$ can be computed by making the sum in (4.30). However, $\mathcal{G}_{z_i}$ is also the solution of the following Lyapunov equation:

$$\mathcal{G}_{z_i} = (A-KC)^T \mathcal{G}_{z_i} (A-KC) + C^T S_u^{-1} C - (A-KC)^{i^T} C^T S_u^{-1} C (A-KC)^i \ . \tag{4.32}$$

○

The principal angles between row(U_p) and row(Y_f)

The squared cosines of the smallest n principal angles between row(U_p) and row(Y_f) for $j \to \infty$ and finite i are the eigenvalues of $\mathcal{D}_i(\mathcal{G}_{z_i}^{-1} + \Sigma)^{-1}$, where

$$\mathcal{D}_i = \Delta_i Q_{u_i} \Delta_i^T = \sum_{k=0}^{i-1} A^k K S_u K^T A^{k^T} \ .$$

The other $mi-n$ principal angles are equal to $\frac{\pi}{2}$:

$$\boxed{\lim_{j\to\infty} \cos^2 [U_p \sphericalangle Y_f] = \lambda\left(\mathcal{D}_i(\mathcal{G}_{z_i}^{-1}+\Sigma)^{-1}\right), \underbrace{0,\ldots,0}_{mi-n}\ .} \tag{4.33}$$

The principal angles between row(Y_p) and row(Y_f)
The squared cosines of the smallest n principal angles between row(Y_p) and row(Y_f) for $j \to \infty$ can be computed as the eigenvalues of

$$\Big(-A^i\Sigma R_i^T + \mathcal{D}_i - R_iT_iR_i^T + A^i\Sigma\mathcal{G}_{z_i}(\Sigma A^{i^T} + T_iR_i^T) \\ + \left(-R_i - A^i\Sigma\mathcal{G}_{z_i}T_i\mathcal{G}_{z_i} + R_iT_i\mathcal{G}_{z_i}\right)\Sigma A^{i^T}\Big)\left(\mathcal{G}_{z_i}^{-1}+\Sigma\right)^{-1}\ , \tag{4.34}$$

where

$$T_i = (\Sigma^{-1} + \mathcal{G}_{z_i})^{-1}\ ,$$
$$R_i = \Delta_i\Gamma_{z_i} = -\sum_{k=0}^{i-1} A^{i-1-k}KC(A-KC)^k\ .$$

The other $mi-n$ angles are equal to $\frac{\pi}{2}$.

4.3 The canonical correlations of the input and output processes

The canonical correlations of u_f and y_f
The smallest n canonical correlations of u_f and y_f are the square roots of the eigenvalues of $(I_n + \mathcal{G}_z\Sigma)^{-1}$, where Σ is the state covariance matrix, which can be found by solving the Lyapunov equation $\Sigma = A\Sigma A^T + KS_uK^T$, and $\mathcal{G}_z = \lim_{i\to\infty}\mathcal{G}_{z_i}$ is the solution of the Lyapunov equation

$$\mathcal{G}_z = (A-KC)^T\mathcal{G}_z(A-KC) + C^TS_u^{-1}C\ . \tag{4.35}$$

The other canonical correlations are equal to 1.

$$\boxed{\mathrm{cc}^2(u_f, y_f) = \lambda\left((I_n+\mathcal{G}_z\Sigma)^{-1}\right), 1, 1, 1, \ldots}$$

The canonical correlations of y_p and y_f / u_p and y_f
The largest n canonical correlations of the past and the future output (and also of the past input and future output) are the square roots of the eigenvalues of $\Sigma(\mathcal{G}_z^{-1}+\Sigma)^{-1}$. The other canonical correlations are equal to 0.

$$\boxed{\mathrm{cc}^2(y_p, y_f) = \lambda\left(\Sigma(\mathcal{G}_z^{-1}+\Sigma)^{-1}\right), 0, 0, 0, \ldots}$$

4.4 Relation of the canonical correlations between the different processes

The canonical correlations of the different pairs of processes (or the principal angles between the pairs of subspaces) are closely related, as we have already indicated in Table 2. Here, we show that the canonical correlations of the future input and output are complementary[3] to the canonical correlations of the past and future output (or past input and future output). The relation is straightforwardly proven by means of the matrices given in Section 4.3.

Property 4.1. Complementarity of $\mathrm{cc}(\boldsymbol{u_f}, \boldsymbol{y_f})$ and $\mathrm{cc}(\boldsymbol{y_p}, \boldsymbol{y_f})$
The canonical correlations of u_f and y_f are complementary to the canonical correlations of y_p and y_f (u_p and y_f).

Proof.
The smallest n squared canonical correlations of u_f and y_f are the eigenvalues of $(I_n + \mathcal{G}_z \Sigma)^{-1} = I_n - (\mathcal{G}_z^{-1} + \Sigma)^{-1}\Sigma$ and the other canonical correlations are equal to 1. The eigenvalues of $I_n - (\mathcal{G}_z^{-1} + \Sigma)^{-1}\Sigma$ are equal to the eigenvalues of $I_n - \Sigma(\mathcal{G}_z^{-1} + \Sigma)^{-1}$. Since the eigenvalues of $\Sigma(\mathcal{G}_z^{-1} + \Sigma)^{-1}$ are the largest n squared canonical correlations of y_p and y_f (u_p and y_f) and the other canonical correlations are equal to 0, we have proven that the canonical correlations of u_f and y_f are complementary to the canonical correlations of y_p and y_f (u_p and y_f). □

Remark 4.2. Simplifications for single-input single-output (SISO) models
For SISO models, the expressions for the canonical correlations can be simplified. By comparing (4.35) with (2.10), we see that for SISO models, the matrix $\mathcal{G}_z$ is equal to $\frac{1}{\sigma^2} Q_z$, where σ^2 is the variance of the input process and Q_z is the observability Gramian of the inverse model. Similarly, a comparison of (2.3) and (2.7) shows that the state covariance matrix Σ of a SISO model is equal to $\sigma^2 P$, where P is the controllability Gramian of the model. We consequently obtain the following expressions for the canonical correlations of the input and output processes of a SISO model:

$$\begin{aligned} \mathrm{cc}^2(u_f, y_f) &= \lambda\left((I_n + Q_z P)^{-1}\right), 1, 1, 1, \ldots \\ \mathrm{cc}^2(y_p, y_f) &= \lambda\left(P(Q_z^{-1} + P)^{-1}\right), 0, 0, 0, \ldots \end{aligned}$$

○

5 Conclusions

In this paper we have given expressions for the canonical correlations of the different past and future input and output processes of a linear stochastic model, in terms of the model parameters.

References

[1] H. Akaike, "Stochastic theory of minimal realization", *IEEE Transactions on Automatic Control* 19, 667–674, 1974.

[2] M. Basseville, M. Abdelghani, and A. Benveniste, "Subspace-based fault detection algorithms for vibration monitoring", *Automatica* 36, 101–109, 2000.

[3] D. Bauer, "Order estimation for subspace methods", *Automatica* 37, 1561–1573, 2001.

[3] Two canonical correlations ρ_1 and ρ_2 are complementary if $\rho_1^2 = 1 - \rho_2^2$. For the corresponding principal angles, θ_1 and θ_2 holds: $\theta_1 = \frac{\pi}{2} - \theta_2$.

[4] M. Borga, *Learning Multidimensional Signal Processing*, PhD thesis, Linköping University, Linköping, Sweden, 1998. Available on `http:// people.imt.liu.se/~magnus`.

[5] P. E. Caines, *Linear Stochastic Systems*, Wiley, New York, 1988.

[6] K. De Cock, *Principal Angles in System Theory, Information Theory and Signal Processing*, PhD thesis, Faculty of Applied Sciences, K.U.Leuven, Leuven, Belgium, 2002. Available on `ftp.esat.kuleuven.ac.be/pub/sista/decock/reports/` as file `phd.ps.gz`.

[7] M. I. Friswell, J. E. Mottershead, and H. Ahmadian, “Combining subset selection and parameter constraints in model updating”, *Journal of Vibration and Acoustics* 120, 854–859, 1998.

[8] M. I. Friswell, J. E. T. Penny, and S. D. Garvey, “Parameter subset selection in damage location”, *Inverse Problems in Engineering* 5, 189–215, 1997. Available on `http://www.swan.ac.uk/mecheng/staff/ mfriswell/PDF_Files/` as file `J33.pdf`.

[9] I. M. Gel’fand and A. M. Yaglom, “Calculation of the amount of information about a random function contained in another such function”, *American Mathematical Society Translations, Series (2)*, 12, 199–236, 1959.

[10] H. Hotelling, “Relations between two sets of variates”, *Biometrika* 28, 321–372, 1936.

[11] N. P. Jewell and P. Bloomfield, “Canonical correlations of past and future for time series: definitions and theory”, *The Annals of Statistics* 11, no. 3, 837–847, 1983.

[12] N. P. Jewell, P. Bloomfield, and F. C. Bartmann, “Canonical correlations of past and future for time series: bounds and computation”, *The Annals of Statistics* 11, no. 3, 848–855, 1983.

[13] C. Jordan, “Essai sur la géométrie à n dimensions”, *Bulletin de la Société Mathématique* 3, 103–174, 1875.

[14] J. Kay, “Feature discovery under contextual supervision using mutual information”, In *Proceedings of the 1992 International Joint Conference on Neural Networks* volume 4, pages 79–84, Baltimore, 1992.

[15] W. E. Larimore, “Statistical optimality and canonical variate analysis system identification”, *Signal Processing* 52, no. 2, 131–144, July 1996.

[16] M. Pavon, “Canonical correlations of past inputs and future outputs for linear stochastic systems”, *Systems & Control Letters* 4, no. 4, 209–215, June 1984.

[17] L. L. Scharf and C. T. Mullis, “Canonical coordinates and the geometry of inference, rate, and capacity”, *IEEE Transactions on Signal Processing* 48, no. 3, 824–831, March 2000. Available on `http://schof.Colorado.EDU/~scharf/scharf.html`.

[18] C. E. Shannon, “A mathematical theory of communication”, *Bell System Technical Journal* 27, 379–423 and 623–656, July and October 1948. Available on `http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html`.

[19] L. N. Trefethen and D. Bau, *Numerical Linear Algebra*, SIAM, Philadelphia, 1997.

[20] T. Van Gestel, J. A. K. Suykens, J. De Brabanter, B. De Moor, and J. Vandewalle, "Kernel canonical correlation analysis and least squares support vector machines", In *Proceedings of the International Conference on Artificial Neural Networks (ICANN 2001)*, Vienna, Austria, August 2001. Available on `ftp://ftp.esat.kuleuven. ac.be/pub/SISTA/suykens/reports` as file `lssvm_01_24.ps.gz`.

[21] P. Van Overschee and B. De Moor, "Subspace algorithms for the stochastic identification problem", *Automatica* 29, 649–660, 1993. Available on `ftp://ftp.esat.kuleuven.ac.be/pub/SISTA/vanoverschee/reports` as file `stoch_auto2.ps.Z`.

[22] P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems: Theory – Implementation – Applications*, Kluwer Academic Publishers, Boston, 1996.