Empirical Canonical Correlation Analysis in Subspaces

Ali Pezeshki [1], Louis L. Scharf [1,2], Mahmood R. Azimi-Sadjadi [1], and Magnus Lundberg [1]
[1]Department of Electrical and Computer Engineering, [2]Department of Statistics
Colorado State University, Fort Collins, Colorado 80523–1373
{ali,scharf,azimi,mlg}@engr.colostate.edu

***Abstract*— This paper addresses canonical correlation analysis of two-channel data, when channel covariances are estimated from a limited number of samples, and are not necessarily full-rank. We show that empirical canonical correlations measure the cosines of the principal angles between the *row spaces* of the data matrices for the two channels. When the number of samples is smaller than the sum of the ranks of the two data matrices, some of the empirical canonical correlations become one, regardless of the two-channel model that generates the samples. In such cases, the empirical canonical correlations may not be used as estimates of correlation between random variables.**

I. Introduction

The developments and results reported to date for canonical correlation analysis [1],[2] of two-channel data are based on the assumption that either the theoretical covariance and cross-covariance matrices of the channels are known, or enough independent copies of the channels are available to obtain full-rank estimates of the covariance matrices. Little attention has been devoted to the algebraic limits to canonical correlation analysis. That is, just how poor can sample support become before sample canonical correlations cease to carry any information about the true canonical correlations? This is one of the particular questions we address in this paper.

More generally, we study the empirical canonical correlation analysis of two-channel data. The term empirical implies that the canonical correlation analysis is based on covariance matrices that are estimated from a limited number of copies (samples) of the two-channel data, and are not necessarily full-rank. The question to be addressed in this paper is whether or not the canonical correlations and coordinates obtained from sample data have the same algebraic and geometric properties as the underlying theoretical ones. For example, when do empirical canonical correlations estimate theoretical canonical correlations between two data channels?

We shall clarify how the number of samples drawn from two-channel data and the ranks of the data matrices (column-wise collections of the vector samples drawn from the two data channels) affect the algebraic and geometric properties of empirical canonical correlations and canonical coordinates. We demonstrate that empirical canonical correlations measure the cosines of the principal angles between the row spaces of the two data matrices in Euclidean space, and hence are experimental surrogates for the true canonical correlations, which measure the principal cosines between subspaces of the Hilbert space second-order random variables. This result has been reported in [1] for the case where sample covariance matrices are full-rank.

We establish that when the number of vector samples drawn from each vector channel is smaller than the sum of the ranks of the two data matrices, some of the empirical canonical correlations become one, regardless of the two-channel model that generates the samples. In such cases, the empirical canonical correlations should be interpreted solely as principal cosines between two linear subspaces of a Euclidean space, and not as estimates of canonical correlations or principal cosines between subspaces in the Hilbert space of second-order random variables. However, when the number of samples is greater than the sum of the ranks of the two data matrices, it may be possible for the empirical canonical correlations to estimate canonical correlations and principal cosines between random variables. This implies that the sum of the ranks of the two data matrices determines the minimum number of data samples (sample support) required to estimate the theoretical canonical correlations.

A more comprehensive account of the findings in this paper have been reported in [3]. Some of the findings, concerning the effect of sample support on empirical canonical correlations, have been reported in [4], without rigorous proof and analysis.

This work was supported by the Office of Naval Research under contracts N00014-02-1-0006 and N00014-04-1-0084.

II. A Review of Canonical Coordinates

Consider the composite data vector $\mathbf{z} = [\mathbf{x}^T \ \mathbf{y}^T]^T \in \mathbb{R}^{m+n}$, consisting of two random vectors $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$, $m \leq n$. Assume that $\mathbf{x}$ and $\mathbf{y}$ have zero means and share the nonsingular composite covariance matrix

$$\mathbf{R}_{zz} = E[\mathbf{z}\,\mathbf{z}^T] = \begin{bmatrix} \mathbf{R}_{xx} & \mathbf{R}_{xy} \\ \mathbf{R}_{yx} & \mathbf{R}_{yy} \end{bmatrix}. \quad (1)$$

We may think of the elements of the cross-covariance matrix $\mathbf{R}_{xy}$ as inner products in the Hilbert space of second-order random variables. That is $[\mathbf{R}_{xy}]_{ij} = E[x_i y_j]$ is the inner product between random variables x_i and y_j, in the Hilbert space.

The *coherence matrix* of $\mathbf{x}$ and $\mathbf{y}$ is defined as [2]

$$\mathbf{C} = E[(\mathbf{R}_{xx}^{-1/2}\mathbf{x})(\mathbf{R}_{yy}^{-1/2}\mathbf{y})^T] = \mathbf{R}_{xx}^{-1/2}\mathbf{R}_{xy}\mathbf{R}_{yy}^{-T/2}, \quad (2)$$

where $\mathbf{R}_{xx}^{1/2} \in \mathbb{R}^{m\times m}$ is a square root of $\mathbf{R}_{xx}$. That is, $\mathbf{R}_{xx}^{1/2}\mathbf{R}_{xx}^{T/2} = \mathbf{R}_{xx}$ and $\mathbf{R}_{xx}^{-1/2}\mathbf{R}_{xx}\mathbf{R}_{xx}^{-T/2} = \mathbf{I}$. A singular value decomposition (SVD) for the coherence matrix $\mathbf{C}$ may be written as

$$\mathbf{C} = \mathbf{R}_{xx}^{-1/2}\mathbf{R}_{xy}\mathbf{R}_{yy}^{-T/2} = \mathbf{F}\boldsymbol{\Sigma}\mathbf{G}^T \tag{3}$$

where $\mathbf{F} \in \mathbb{R}^{m\times m}$ and $\mathbf{G} \in \mathbb{R}^{n\times n}$ are orthogonal matrices. The matrix $\boldsymbol{\Sigma} = [\ \boldsymbol{\Sigma}(m) \quad 0\ ] \in \mathbb{R}^{m\times n}$ is a diagonal singular value matrix, with $\boldsymbol{\Sigma}(m) = \text{diag}[\sigma_1, \sigma_2, \ldots, \sigma_m]$; $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_m > 0$.

The *canonical coordinates* of the composite vector $\mathbf{z} = [\ \mathbf{x}^T \quad \mathbf{y}^T\ ]^T$ are defined as [2]

$$\mathbf{w} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} \mathbf{F}^T\mathbf{R}_{xx}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^T\mathbf{R}_{yy}^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \tag{4}$$

and they share the composite covariance matrix

$$\mathbf{R}_{ww} = E[\mathbf{w}\mathbf{w}^T] = \begin{bmatrix} \mathbf{I} & \boldsymbol{\Sigma} \\ \boldsymbol{\Sigma}^T & \mathbf{I} \end{bmatrix}. \tag{5}$$

Here the elements of $\mathbf{u} \in \mathbb{R}^m$ are canonical coordinates of $\mathbf{x}$ and the elements of $\mathbf{v} \in \mathbb{R}^n$ are canonical coordinates of $\mathbf{y}$. The diagonal matrix

$$\boldsymbol{\Sigma} = \mathbf{F}^T\mathbf{C}\mathbf{G} = \mathbf{F}^T\mathbf{R}_{xx}^{-1/2}\mathbf{R}_{xy}\mathbf{R}_{yy}^{-T/2}\mathbf{G} \tag{6}$$

is the *canonical correlation matrix* of *canonical correlations* σ_i, where $1 \geq \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m > 0$. The canonical correlations σ_i are the singular values of the coherence matrix $\mathbf{C}$. They are also the correlations between pairs of canonical coordinates (u_i, v_i), i.e. $E[u_iv_j] = \sigma_i\delta_{ij}$; $i \in [1, m]$, $j \in [1, n]$, with δ_{ij} the Kronecker delta.

The ith canonical correlation $\sigma_i = E[u_iv_i]$ measures the cosine of the angle between u_i, the ith canonical coordinate of $\mathbf{x}$, and v_i, the ith canonical coordinate of $\mathbf{y}$. This angle is the ith principal angle between the $m-$ and $n-$dimensional subspaces of the Hilbert space of second-order random variables that are spanned by the elements of $\mathbf{x}$ and $\mathbf{y}$. Therefore, the canonical correlations σ_i measure the cosines of the principal angles between two linear subspaces in the Hilbert space of second order random variables.

III. Empirical Canonical Correlation Analysis in Subspaces

Consider the sample data matrices

$$\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_M] \quad \text{and} \quad \mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_M] \tag{7}$$

consisting of M vector samples drawn from the two-channel vector process $\mathbf{z} = [\mathbf{x}^T \ \ \mathbf{y}^T]^T$. Without loss of generality, we assume that the columns of the sample data matrices $\mathbf{X}$ and $\mathbf{Y}$ are centered, i.e. $\sum_{i=1}^M \mathbf{x}_i = \mathbf{0}$ and $\sum_{i=1}^M \mathbf{y}_i = \mathbf{0}$. Then, we may write the composite sample covariance matrix of $\mathbf{z} = [\mathbf{x}^T \ \ \mathbf{y}^T]^T$ as

$$\mathbf{S}_{zz} = \mathbf{Z}\mathbf{Z}^T = \begin{bmatrix} \mathbf{X}\mathbf{X}^T & \mathbf{X}\mathbf{Y}^T \\ \mathbf{Y}\mathbf{X}^T & \mathbf{Y}\mathbf{Y}^T \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{xx} & \mathbf{S}_{xy} \\ \mathbf{S}_{yx} & \mathbf{S}_{yy} \end{bmatrix} \tag{8}$$

where $\mathbf{S}_{xx} = \mathbf{X}\mathbf{X}^T$, $\mathbf{S}_{yy} = \mathbf{Y}\mathbf{Y}^T$, and $\mathbf{S}_{xy} = \mathbf{X}\mathbf{Y}^T$ are sample auto-covariance and cross-covariance matrices of $\mathbf{x}$ and $\mathbf{y}$, computed from the sample data matrices $\mathbf{X}$ and $\mathbf{Y}$. Note that the normalization by M in the sample covariance matrices is ignored, as it does not affect the discussion.

We think of the elements of $\mathbf{S}_{xx}$, $\mathbf{S}_{yy}$, and $\mathbf{S}_{xy}$ as inner products in Euclidean space. For example, the ijth element of $\mathbf{S}_{xy}$, i.e. $[\mathbf{S}_{xy}]_{ij} = [\mathbf{X}\mathbf{Y}^T]_{ij} = \mathbf{X}(i,:)\mathbf{Y}(j,:)^T$ is the inner product between the ith row of $\mathbf{X}$, i.e. $\mathbf{X}(i,:)$, and the jth row of $\mathbf{Y}$, i.e. $\mathbf{Y}(j,:)$. In analogy with the Hilbert space case in Section II, where $[\mathbf{R}_{xy}]_{ij} = E[x_iy_j]$ is the inner product between random varibles x_i and y_j, we may think of the ith row of $\mathbf{X}$, i.e. $\mathbf{X}(i,:)$, as an experimental *surrogate* for x_i, and $\mathbf{Y}(j,:)$ as an experimental surrogate for y_j. Correspondingly, the inner product $[\mathbf{X}\mathbf{Y}^T]_{ij} = \mathbf{X}(i,:)\mathbf{Y}(j,:)^T$ in the Euclidean space of $M-$vectors may be viewed as a stand-in for the inner product $[\mathbf{R}_{xy}]_{ij} = E[x_iy_j]$ in the Hilbert space of second-order random variables.

Assuming that Rank$\{\mathbf{X}\} = p$ and Rank$\{\mathbf{Y}\} = q$, we may write the SVD's of the sample data matrices $\mathbf{X}$ and $\mathbf{Y}$ as,

$$\begin{aligned} \mathbf{X} = \mathbf{U}_x\boldsymbol{\Sigma}_x\mathbf{V}_x^T \quad &\text{and} \quad \mathbf{U}_x^T\mathbf{X}\mathbf{V}_x = \boldsymbol{\Sigma}_x, \\ \mathbf{Y} = \mathbf{U}_y\boldsymbol{\Sigma}_y\mathbf{V}_y^T \quad &\text{and} \quad \mathbf{U}_y^T\mathbf{Y}\mathbf{V}_y = \boldsymbol{\Sigma}_y, \end{aligned} \tag{9}$$

where $\mathbf{U}_x \in \mathbb{R}^{m\times m}$, $\mathbf{V}_x \in \mathbb{R}^{M\times M}$, $\mathbf{U}_y \in \mathbb{R}^{n\times n}$, and $\mathbf{V}_y \in \mathbb{R}^{M\times M}$ are orthogonal matrices and $\boldsymbol{\Sigma}_x \in \mathbb{R}^{m\times M}$ and $\boldsymbol{\Sigma}_y \in \mathbb{R}^{n\times M}$ are diagonal:

$$\boldsymbol{\Sigma}_x = \begin{bmatrix} \boldsymbol{\Sigma}_x(p) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_y = \begin{bmatrix} \boldsymbol{\Sigma}_y(q) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{10}$$

The diagonal matrices $\boldsymbol{\Sigma}_x(p) \in \mathbb{R}^{p\times p}$ and $\boldsymbol{\Sigma}_y(q) \in \mathbb{R}^{q\times q}$ contain the nonzero singular values of $\mathbf{X}$ and $\mathbf{Y}$.

In analogy to Section II, we define the *sample coherence matrix* $\mathbf{C}$ for $\mathbf{X}$ and $\mathbf{Y}$ as[1]

$$\mathbf{C} = \mathbf{S}_{xx}^{-1/2}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-T/2} = (\mathbf{X}\mathbf{X}^T)^{-1/2}\mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T)^{-T/2} \tag{11}$$

and write an SVD for it of the form

$$\mathbf{C} = \mathbf{F}\boldsymbol{\Sigma}\mathbf{G}^T \quad \text{and} \quad \mathbf{F}^T\mathbf{C}\mathbf{G} = \boldsymbol{\Sigma}, \tag{12}$$

where $\mathbf{F} \in \mathbb{R}^{m\times m}$ and $\mathbf{G} \in \mathbb{R}^{n\times n}$ are orthogonal matrices and $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times n}$ is diagonal. Note that since Rank$\{\mathbf{X}\} = p$ and Rank$\{\mathbf{Y}\} = q$, the sample coherence matrix $\mathbf{C}$ is of rank $r = \min(p, q)$. Thus, the singular value matrix $\boldsymbol{\Sigma}$ has only r nonzero elements $\sigma_i > 0$, $i \in [1, r]$. The diagonal matrix $\boldsymbol{\Sigma}$ is the *empirical canonical correlation matrix* of the *empirical canonical correlations* σ_i, $i \in [1, r]$.

Correspondingly, we define the composite canonical coordinate matrix $\mathbf{W}$ consisting of the canonical coordinate matrices $\mathbf{U}$ and $\mathbf{V}$ as

$$\mathbf{W} = \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix} = \begin{bmatrix} \mathbf{F}^T(\mathbf{X}\mathbf{X}^T)^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^T(\mathbf{Y}\mathbf{Y}^T)^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \tag{13}$$

[1]Note that when $\mathbf{X}$ and $\mathbf{Y}$ are not full-rank, $(\mathbf{X}\mathbf{X}^T)^{-1/2}$ and $(\mathbf{Y}\mathbf{Y}^T)^{-1/2}$ denote Moore-Penrose pseudo inverses of $(\mathbf{X}\mathbf{X}^T)^{1/2}$ and $(\mathbf{Y}\mathbf{Y}^T)^{1/2}$.

Then, the composite sample covariance matrix for $\mathbf{W}$ may be written as

$$\mathbf{S}_{ww} = \mathbf{W}\mathbf{W}^T = \begin{bmatrix} \mathbf{S}_{uu} & \mathbf{S}_{uv} \\ \mathbf{S}_{vu} & \mathbf{S}_{vv} \end{bmatrix} = \begin{bmatrix} \mathbf{F}^T \mathbf{P}_{\mathbf{U}_x(:,1:p)} \mathbf{F} & \mathbf{\Sigma} \\ \mathbf{\Sigma}^T & \mathbf{G}^T \mathbf{P}_{\mathbf{U}_y(:,1:q)} \mathbf{G} \end{bmatrix} \tag{14}$$

where $\mathbf{P}_{\mathbf{U}_x(:,1:p)} = \mathbf{U}_x(:,1:p)\mathbf{U}_x(:,1:p)^T$ and $\mathbf{P}_{\mathbf{U}_y(:,1:q)} = \mathbf{U}_y(:,1:q)\mathbf{U}_y(:,1:q)^T$ are orthogonal projection matrices onto the $p-$ and $q-$dimensional subspaces spanned by the first p and q columns of $\mathbf{U}_x$ and $\mathbf{U}_y$, respectively.

Equation (14) shows that the empirical canonical correlation $\sigma_i = [\mathbf{U}\mathbf{V}^T]_{ii}$ is the experimental surrogate for the ith canonical correlation $\sigma_i = E[u_i v_i]$, where the standard inner product $[\mathbf{S}_{uv}]_{ij} = [\mathbf{U}\mathbf{V}^T]_{ij} = \mathbf{U}(i,:)\mathbf{V}(j,:)^T$ stands for the inner product $[\mathbf{R}_{uv}]_{ij} = E[u_i v_j]$ in the Hilbert space of second-order random variables. Correspondingly, the rows of the matrices $\mathbf{U}$ and $\mathbf{V}$ in (13) are surrogates for the canonical coordinates $\mathbf{u}$ and $\mathbf{v}$, respectively. However, contrasting to the Hilbert space case in Section II, the sample auto-covariance matrices for $\mathbf{U}$ and $\mathbf{V}$ in (14) are not necessarily identity. Rather they are orthogonal projection matrices given by $\mathbf{U}\mathbf{U}^T = \mathbf{F}^T\mathbf{P}_{\mathbf{U}_x(:,1:p)}\mathbf{F}$ and $\mathbf{V}\mathbf{V}^T = \mathbf{G}^T\mathbf{P}_{\mathbf{U}_y(:,1:q)}\mathbf{G}$. Only when $p = m$ and $q = n$, which yields full-rank sample covariance matrices for $\mathbf{X}$ and $\mathbf{Y}$, are the sample covariance matrices for $\mathbf{U}$ and $\mathbf{V}$ identity, as $\mathbf{P}_{\mathbf{U}_x(:,1:m)} = \mathbf{U}_x\mathbf{U}_x^T = \mathbf{I}(m)$ and $\mathbf{P}_{\mathbf{U}_y(:,1:n)} = \mathbf{U}_y\mathbf{U}_y^T = \mathbf{I}(n)$.

Inserting the SVD's $\mathbf{X} = \mathbf{U}_x\mathbf{\Sigma}_x\mathbf{V}_x^T$ and $\mathbf{Y} = \mathbf{U}_y\mathbf{\Sigma}_y\mathbf{V}_y^T$ in (11), and using (12) yields

$$\begin{bmatrix} \mathbf{V}_x(:,1:p)^T\mathbf{V}_y(:,1:q) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \mathbf{U}_x^T\mathbf{F}\begin{bmatrix} \mathbf{\Sigma}(r) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{G}^T\mathbf{U}_y \tag{15}$$

where $\mathbf{\Sigma}(r) = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$ is the diagonal matrix that carries the nonzero singular values of $\mathbf{C}$, and $\mathbf{V}_x(:,1:p) \in \mathbb{R}^{M\times p}$ and $\mathbf{V}_y(:,1:q) \in \mathbb{R}^{M\times q}$ are column-wise matrices that carry the first p columns of $\mathbf{V}_x$ and the first q columns of $\mathbf{V}_y$.

The column-wise matrices $\mathbf{V}_x(:,1:p) \in \mathbb{R}^{M\times p}$ and $\mathbf{V}_y(:,1:q) \in \mathbb{R}^{M\times q}$ are orthonormal bases for $p-$ and $q-$dimensional subspaces of $\mathbb{R}^M$. Therefore, singular values of the matrix $\mathbf{V}_x(:,1:p)^T\mathbf{V}_y(:,1:q)$ measure the cosines of principal angles between Col-Span$\{\mathbf{V}_x(:,1:p)\}$ and Col-Span$\{\mathbf{V}_y(:,1:q)\}$. Since $\mathbf{F}$, $\mathbf{G}$, $\mathbf{U}_x$, and $\mathbf{U}_y$ are orthogonal matrices, the rank-r matrices $\mathbf{V}_x(:,1:p)^T\mathbf{V}_y(:,1:q)$ and $\mathbf{\Sigma}(r)$ in (15) are similar. That is, the diagonal elements of $\mathbf{\Sigma}(r)$ are the singular values of $\mathbf{V}_x(:,1:p)^T\mathbf{V}_y(:,1:q)$ and hence σ_i measures the cosine of the ith principal angle between Col-Span$\{\mathbf{V}_x(:,1:p)\}$ and Col-Span$\{\mathbf{V}_y(:,1:q)\}$. However, the SVD's in (9) show that Col-Span$\{\mathbf{V}_x(:,1:p)\}$ and Col-Span$\{\mathbf{V}_y(:,1:q)\}$ are the same as Row-Span$\{\mathbf{X}\}$ and Row-Span$\{\mathbf{Y}\}$, respectively. Thus, the empirical canonical correlations σ_i are cosines of the principal angles between Row-Span$\{\mathbf{X}\}$ and Row-Span$\{\mathbf{Y}\}$. Therefore, it is the *row spaces* of $\mathbf{X}$ and $\mathbf{Y}$ that determine the sample canonical correlations. This fits intuition, as in the sample data case the ith rows of $\mathbf{X}$ and $\mathbf{Y}$ are experimental surrogates for x_i and y_i, where the Euclidean space inner product $[\mathbf{S}_{xy}]_{ij} = [\mathbf{X}\mathbf{Y}^T]_{ij} = \mathbf{X}(i,:)\mathbf{Y}(j,:)^T$ stands for the Hilbert space inner product $[\mathbf{R}_{xy}]_{ij} = E[x_i y_j]$.

So far, we have established that the empirical canonical correlations measure the cosines of principal angles in Euclidean space, where row vectors are surrogates for random variables. The question is then, can these principal angles in Euclidean space estimate the principal angles in the Hilbert space of second-order random variables. In what follows we shall discuss how the algebraic and geometric properties of the empirical canonical correlations, carried by the surrogate rows of $\mathbf{U}$ and $\mathbf{V}$, may be affected by M, the number of samples, and p and q, the dimensions of the row spaces $\mathbf{X}$ and $\mathbf{Y}$. For this purpose, we consider two different cases; namely *sample-poor* and *sample-rich*.

A. *Case 1: Sample-Poor* $(M < p + q)$

In this case, the number of data samples M drawn from each channel is smaller than the sum of the dimensions of the row spaces of $\mathbf{X}$ and $\mathbf{Y}$, i.e. $M < p + q$. The row-space of $\mathbf{X} \in \mathbb{R}^{m\times M}$ is a $p-$dimensional subspace of $\mathbb{R}^M$, and the row space of $\mathbf{Y} \in \mathbb{R}^{n\times M}$ is a $q-$dimensional subspace of $\mathbb{R}^M$. Therefore, when $p + q > M$, Row-Span$\{\mathbf{X}\}$ and Row-Span$\{\mathbf{Y}\}$ have to share a subspace of $\mathbb{R}^M$ of dimension at least $d = (p + q) - M$. Since these two subspaces overlap in at least d dimensions, the cosines of at least d principal angles, or equivalently d sample canonical correlations will be equal to one, regardless of the two-channel vector process that generates the samples. This result has also been reported in [4], but no rigorous proof or analysis for the geometric properties of sample canonical correlations is given.

Equation (16) shows that the $d-$dimensional subspace of $\mathbb{R}^M$ that is shared by Row-Span$\{\mathbf{X}\}$ and Row-Span$\{\mathbf{Y}\}$ is spanned by the first d surrogate rows of $\mathbf{U}$, or alternatively by the first d surrogate rows of $\mathbf{V}$, as illustrated in Figure 1. In this figure, the planes show the row spaces of $\mathbf{X}$ and $\mathbf{Y}$ and the intersection line shows the $d-$dimensional subspace shared between Row-Span$\{\mathbf{X}\}$ and Row-Span$\{\mathbf{Y}\}$.

Therefore, in this case, the sample canonical correlation matrix $\mathbf{\Sigma}$ may be expressed as

$$\mathbf{\Sigma} = \mathbf{U}\mathbf{V}^T = \begin{bmatrix} \mathbf{I}(d) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_\star & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{16}$$

where $\mathbf{\Sigma}_\star = \mathrm{diag}(\sigma_{d+1}, \ldots, \sigma_r)$ with $1 \geq \sigma_{d+1} \geq \cdots \geq \sigma_r > 0$ and $r = \min(p,q)$. This shows that even when the samples are drawn from a two-channel process in which the elements of $\mathbf{x}$ and $\mathbf{y}$ are mutually uncorrelated, some empirical canonical correlations become one as a result of poor sampling. Therefore, in a sample poor case the empirical canonical correlations between the data matrices $\mathbf{X}$ and $\mathbf{Y}$ definitely do not estimate the canonical correlations between

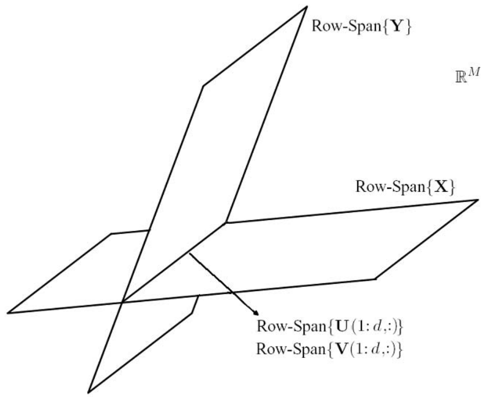

Fig. 1. Geometry of empirical canonical correlations in a sample-poor case.

the random vectors $\mathbf{x}$ and $\mathbf{y}$. Equivalently, the principal angles between the Euclidean spaces spanned by rows of $\mathbf{X}$ and $\mathbf{Y}$ *do not estimate* the principal angles between the Hilbert spaces spanned by the random variables in $\mathbf{x}$ and $\mathbf{y}$. In other words, in a sample poor case the empirical canonical correlations are *defective* and may not be used in any inference based on estimates of theoretical canonical correlations.

B. Case 2: Sample-Rich $(M \geq p+q)$

In this case the number of samples drawn from each channel is greater than or equal to the sum of the dimensions of the row spaces of $\mathbf{X}$ and $\mathbf{Y}$, i.e. $M \geq p+q$. Note that being sample-rich does not guarantee that the sample covariance matrices $\mathbf{S}_{xx} = \mathbf{X}\mathbf{X}^T$, $\mathbf{S}_{xy} = \mathbf{X}\mathbf{Y}^T$, and $\mathbf{S}_{yy} = \mathbf{Y}\mathbf{Y}^T$ are nonsingular, as $p = \text{Rank}\{\mathbf{X}\}$ and $q = \text{Rank}\{\mathbf{Y}\}$ may be smaller than m and n.

In this case, the row spaces of $\mathbf{X}$ and $\mathbf{Y}$ are still $p-$ and $q-$dimensional subspaces of $\mathbb{R}^M$. However, as $p+q \leq M$, Row-Span$\{\mathbf{X}\}$ and Row-Span$\{\mathbf{Y}\}$ do not necessarily share a subspace of $\mathbb{R}^M$. Therefore, in this case, the first principal cosine between Row-Span$\{\mathbf{X}\}$ and Row-Span$\{\mathbf{Y}\}$, or equivalently the first sample canonical correlation σ_1, is not necessarily equal to one. This implies that, when $M \geq p+q$ it is possible to use sample canonical correlations or principal cosines between $\mathbf{X}$ and $\mathbf{Y}$ to estimate the canonical correlations or principal cosines between the random variables in $\mathbf{x}$ and $\mathbf{y}$. However, these sample canonical correlations may provide poor estimates of the theoretical ones, as we typically need a large number of samples to estimate the theoretical covariance matrices. Thus, all we say here is that contrasting to the sample poor case, in a sample rich case the sample canonical correlations or principal cosines in Euclidean space are no longer defective for estimating canonical correlations or principal cosines in the corresponding Hilbert space of second-order random variables. In other words, when the number of samples drawn from each channel exceeds the sum of the ranks of the two data matrices, it is possible to use the sample canonical correlations as estimates of the theoretical ones.

IV. Conclusions

In this paper, we have studied the canonical correlation analysis of two-channel data, when the channel covariances are estimated from a limited number of samples. Depending on the number of samples drawn from each channel, and the ranks of the sample data matrices, two different cases emerge: the sample-poor case, in which the number of data samples is smaller than the sum of the ranks of the data matrices, and a sample-rich case, in which the number of data samples is greater than the sum of the ranks of the data matrices. In either case, it is the rows of the sample data matrices that determine the empirical canonical correlations, and that the empirical canonical correlations measure the cosines of the principal angles between the row spaces of the two data matrices [3].

We have established that in a sample-poor case some of the empirical canonical correlations or principal cosines are always one, regardless of the two-channel model that generates the data samples [3]. This result has also been reported in [4], without a rigorous analysis. Therefore, the empirical canonical correlations are defective and may not be used as estimates of canonical correlations between random variables. Geometrically, this means that principal angles between linear subspaces of Euclidean space can not be used as estimates of principal angles between corresponding linear subspaces of the Hilbert space of second-order random variables. In a sample-rich case, however, the empirical canonical correlations do estimate the canonical correlations and principal cosines between the random variables that generate the samples.

References

[1] M. L. Eaton, *Multivariate Statistics: A Vector Space Approach.* New York: Wiley, 1983, ch. 10.

[2] L. L. Scharf and C. T. Mullis, "Canonical coordinates and the geometry of inference, rate and capacity," *IEEE Trans. Signal Processing*, vol. 48, pp. 824–831, Mar. 2000.

[3] A. Pezeshki, L. L. Scharf, and M. R. Azimi-Sadjadi, "Empirical canonical coordinate decompositions in subspaces for two-channel linear and nonlinear maps," submitted to *IEEE Trans. Signal Processing*, Sept. 2004.

[4] M. Kuss and T. Graepel, "The geometry of kernel canonical correlation analysis," Max Planck Institute for Biological Cybernetics, Technical Report 108, May 2003.