A Unifying Discussion of Correlation Analysis for Complex Random Vectors

Peter J. Schreier, *Member, IEEE*

***Abstract*—The assessment of multivariate association between two complex random vectors is considered. A number of correlation coefficients based on three popular correlation analysis techniques, namely canonical correlation analysis, multivariate linear regression, and partial least squares, are reviewed and connected to performance measures in signal processing and communications, such as mean-squared estimation error, mutual information, and signal-to-noise ratio (SNR). For complex data, there are three types of correlation coefficients, which account for rotational, reflectional, and total (i.e., rotational and reflectional) dependencies between two random vectors. These three types are defined and analyzed for different correlation coefficients, and a numerical example is given. It is often required to compare two complex random vectors in a lower-dimensional subspace. For the large class of increasing, Schur-convex correlation coefficients, it is shown that the low-rank approximations of two random vectors maximizing a particular correlation coefficient are determined only by the constraints imposed on the correlation analysis technique. In this context, the correlation spread is defined as a normalized measure of how much of the overall correlation is contained in a low-dimensional subspace.**

***Index Terms*—Canonical correlations, correlation analysis, improper complex random vector, majorization, multivariate linear regression, partial least squares, polarization, widely linear estimator.**

I. Introduction

THE assessment of multivariate association between two random vectors $\mathbf{x}$ and $\mathbf{y}$ is of interest in a multitude of research areas, ranging from geography to psychometrics to communications. A function $\rho(\mathbf{x},\mathbf{y})$ that gives an overall measure of the association between $\mathbf{x}$ and $\mathbf{y}$ is called a *correlation coefficient* if it satisfies the following conditions for all nonzero scalars α and β, provided that $\mathbf{x}$ and $\mathbf{y}$ are not both zero [1], [2]:

$$0 \le \rho(\mathbf{x},\mathbf{y}) \le 1 \tag{1}$$
$$\rho(\mathbf{x},\mathbf{y}) = \rho(\alpha\mathbf{x},\mathbf{y}) = \rho(\mathbf{x},\beta\mathbf{y}) \tag{2}$$
$$\rho(\mathbf{x},\mathbf{y}) = 1 \text{ if } \mathbf{x} = \beta\mathbf{y} \tag{3}$$
$$\rho(\mathbf{x},\mathbf{y}) = 0 \text{ if } \mathbf{x},\mathbf{y} \text{ uncorrelated.} \tag{4}$$

Manuscript received April 2, 2007; revised August 3, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. James Lam. The work of the author was supported by the Australian Research Council (ARC) under Discovery Project Grant DP0664365.

The author is with the School of Electrical Engineering and Computer Science, The University of Newcastle, Callaghan, NSW 2308, Australia (e-mail: peter@peter-schreier.com).

Digital Object Identifier 10.1109/TSP.2007.909054

Many correlation coefficients are symmetric in the sense that $\rho(\mathbf{x},\mathbf{y}) = \rho(\mathbf{y},\mathbf{x})$. However, we shall not generally require this property. Sometimes a correlation coefficient is defined taking values in the interval $[0,c]$ or $[-c,c]$. The conditions (1)–(4) then apply to the normalized absolute value of the coefficient. Conditions (3) and (4) are necessary conditions only. It is usually desirable to have further cases that result in $\rho(\mathbf{x},\mathbf{y}) = 1$, such as $\mathbf{x} = \mathbf{T}\mathbf{y}$, where $\mathbf{T}$ is any full-rank matrix, or $\mathbf{x} = \mathbf{U}\mathbf{y}$, where $\mathbf{U}$ is any unitary matrix.

In this paper, we discuss three popular correlation analysis techniques: canonical correlation analysis (CCA) [3], multivariate linear regression (MLR) [4], [5], and partial least squares (PLS) [6]–[8]. Each of these techniques transforms $\mathbf{x}$ and $\mathbf{y}$ into internal representations $\boldsymbol{\xi}$ and $\boldsymbol{\omega}$, respectively. Different correlation coefficients are then defined as functions of the diagonal correlations $\{k_i\}$ between the internal representations ξ_i and ω_i. CCA, MLR, and PLS are designed to maximize these diagonal correlations. Yet they actually do more than just maximize the $\{k_i\}$. We will demonstrate that CCA, MLR, and PLS produce diagonal correlations $\{k_i\}$ that have maximum spread in the sense of majorization [9]. Because of this, any correlation coefficient that is an increasing, Schur-convex function of the $\{k_i\}$ is maximized. Moreover, the internal coordinate systems of CCA, MLR, and PLS are also the right coordinate systems to determine rank-r approximations of $\mathbf{x}$ and $\mathbf{y}$ such that any increasing, Schur-convex correlation coefficient is maximized, for arbitrary choice of r. This connects our results to closely related problems in optimum reduced-rank estimation and filtering [5], [10].

The key difference between CCA, MLR, and PLS is their invariances. CCA is invariant to linear transformation of $\mathbf{x}$ and $\mathbf{y}$, MLR is invariant to linear transformation of $\mathbf{y}$ but only unitary transformation of $\mathbf{x}$, and PLS is invariant to unitary transformation of $\mathbf{x}$ and $\mathbf{y}$. The correlation analysis technique must be chosen to match the invariances of the problem under consideration. For instance, mutual information between Gaussian $\mathbf{x}$ and $\mathbf{y}$ is invariant under linear transformation of $\mathbf{x}$ and $\mathbf{y}$. Thus, CCA is the right method to deal with mutual information. On the other hand, linear minimum mean-squared estimation error is invariant under linear transformation of the measurement $\mathbf{y}$ but only unitary transformation of the message $\mathbf{x}$. Hence, MLR is the right technique to address linear estimation.

Almost all the work on correlation analysis has been done for *real* data. Yet, in many research areas, *complex* data is routinely used. Complex signals may arise as special cases of bivariate real signals, e.g., as horizontal velocity components in geophysical measurements. Complex signals are also used to describe underlying real wave phenomena. The complex descriptions employed are either the analytic or the complex equivalent baseband representation of a real signal. For instance, in communications, equivalent baseband signals make the detector insensitive to unknown phase in bandpass signals.

Complex data can exhibit three different types of correlation. If x and y are two scalar complex random variables, they are said to be rotationally dependent if $x = ky$ for some constant k. This means that, in the complex plane, sample pairs of x and y rotate in the same direction (clockwise or counterclockwise) by the same angle. They are called reflectionally dependent if $x = \tilde{k}y^*$ for some constant $\tilde{k}$. This means that sample pairs of x and y rotate by the same angle, but in *opposite* directions. *Rotational* and *reflectional correlation* measure the degree of rotational and reflectional dependency, respectively. It is also possible to define a *total correlation* as the combined effect of rotational and reflectional correlations. Rotational and reflectional, but not total, correlations of *scalar* complex random variables are investigated in [11], one of the very few papers on correlation analysis of complex data. Prior work on rotational and reflectional correlations, e.g., [12], often considered real data of double dimension rather than complex data. However, expressing rotational and reflectional correlations in terms of real data of double dimension is significantly more cumbersome and provides less insight.

In this paper, we provide a unifying discussion of correlation analysis for vector-valued complex random data. In Section II, we present a review of CCA, MLR, and PLS. We list six commonly used correlation coefficients that are defined on the basis of CCA, MLR, and PLS, and connect these coefficients with some important performance measures in signal processing and communications: mean-squared estimation error, mutual information, and signal-to-noise ratio (SNR). In Section III, we demonstrate that these correlation coefficients are increasing, Schur-convex functions of the diagonal correlations $\{k_i\}$. This means that they are maximized by CCA, MLR, or PLS, for arbitrary rank r. We also introduce a normalized measure, inspired by the degree of polarization [13], [14], that indicates how much of the overall correlation is contained in a lower-dimensional subspace.

For simplicity, we will assume *real* vectors in Sections II and III before considering *complex* vectors in Sections IV and V. Section IV deals with rotational, reflectional, and total correlations of complex data, and analyzes the properties of the corresponding correlation measures. Finally, Section V presents a numerical example illustrating the usage of the different correlation measures. The Appendix contains some relevant background material on majorization and widely linear transformations.

II. Correlation Measures

Consider two real zero-mean random vectors $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$ with correlation matrices $\mathbf{R}_{xx} = E\mathbf{x}\mathbf{x}^T$ and $\mathbf{R}_{yy} = E\mathbf{y}\mathbf{y}^T$. We shall assume that both $\mathbf{R}_{xx}$ and $\mathbf{R}_{yy}$ are invertible. The cross-correlation properties between $\mathbf{x}$ and $\mathbf{y}$ are described by the cross-correlation matrix $\mathbf{R}_{xy} = E\mathbf{x}\mathbf{y}^T$, but this matrix is generally difficult to interpret. In order to illuminate the underlying cross-correlation structure, many correlation analysis techniques transform $\mathbf{x}$ and $\mathbf{y}$ into p-dimensional internal representations $\boldsymbol{\xi} = \mathbf{A}\mathbf{x}$ and $\boldsymbol{\omega} = \mathbf{B}\mathbf{y}$, with $p = \min(m, n)$, as shown in Fig. 1.

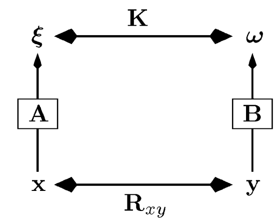

Fig. 1. Principle of CCA, MLR, and PLS.

The full-rank matrices $\mathbf{A} \in \mathbb{R}^{p\times m}$ and $\mathbf{B} \in \mathbb{R}^{p\times n}$ are chosen such that all partial sums over the absolute values of the correlations $k_i = E\xi_i\omega_i$ are maximized

$$\max_{\mathbf{A},\mathbf{B}} \sum_{i=1}^{r} |k_i|, \quad r = 1, \dots, p \tag{5}$$

subject to the following constraints that determine the type of correlation analysis.

- In CCA [3], the internal representations $\boldsymbol{\xi}_{\mathrm{C}}$ and $\boldsymbol{\omega}_{\mathrm{C}}$ are required to be white: $E\boldsymbol{\xi}_{\mathrm{C}}\boldsymbol{\xi}_{\mathrm{C}}^T = \mathbf{A}_{\mathrm{C}}\mathbf{R}_{xx}\mathbf{A}_{\mathrm{C}}^T = \mathbf{I}$ and $E\boldsymbol{\omega}_{\mathrm{C}}\boldsymbol{\omega}_{\mathrm{C}}^T = \mathbf{B}_{\mathrm{C}}\mathbf{R}_{yy}\mathbf{B}_{\mathrm{C}}^T = \mathbf{I}$, where the subscript C refers to CCA.
- In MLR [4], $\mathbf{A}_{\mathrm{M}}$ is restricted to be row-orthogonal (i.e., it has orthogonal rows) and the internal representation $\boldsymbol{\omega}_{\mathrm{M}}$ is required to be white: $\mathbf{A}_{\mathrm{M}}\mathbf{A}_{\mathrm{M}}^T = \mathbf{I}$ and $E\boldsymbol{\omega}_{\mathrm{M}}\boldsymbol{\omega}_{\mathrm{M}}^T = \mathbf{B}_{\mathrm{M}}\mathbf{R}_{yy}\mathbf{B}_{\mathrm{M}}^T = \mathbf{I}$, where the subscript M stands for MLR.
- In PLS [6]–[8], both $\mathbf{A}_{\mathrm{P}}$ and $\mathbf{B}_{\mathrm{P}}$ are restricted to be row-orthogonal: $\mathbf{A}_{\mathrm{P}}\mathbf{A}_{\mathrm{P}}^T = \mathbf{I}$ and $\mathbf{B}_{\mathrm{P}}\mathbf{B}_{\mathrm{P}}^T = \mathbf{I}$, where the subscript P means PLS.

As discussed in Section III, the solution to the maximization problem (5) for CCA, MLR, as well as PLS results in a diagonal cross-correlation matrix between $\boldsymbol{\xi}$ and $\boldsymbol{\omega}$

$$\mathbf{K} = E\boldsymbol{\xi}\boldsymbol{\omega}^T = \operatorname{diag}(k_1, \dots, k_p) \tag{6}$$

with $k_1 \geq k_2 \geq \cdots \geq k_p \geq 0$. Sometimes we need to assume $k_p > 0$. In order to summarize the correlation between $\mathbf{x}$ and $\mathbf{y}$, an overall correlation coefficient ρ can then be defined as a function of the diagonal correlations $\{k_i\}$. This correlation coefficient shares the invariances of the $\{k_i\}$.

CCA and PLS provide a symmetric assessment of correlation, as the roles of $\mathbf{x}$ and $\mathbf{y}$ are interchangeable. MLR, on the other hand, distinguishes between the *message* (or predictor variables) $\mathbf{x}$ and the *measurement* (or criterion variables) $\mathbf{y}$. Because of the maximization (5), all three techniques allow the assessment of correlation in a lower-dimensional subspace of rank $r \leq p = \min(m, n)$. We will use the notation $[\boldsymbol{\xi}]_r$ to denote the vector consisting of the first r elements of $\boldsymbol{\xi}$. For matrices, $[\mathbf{A}]_{r\times m}$ denotes the matrix consisting of the first r rows and m columns of $\mathbf{A}$. Let us now take a more detailed look at CCA, MLR, and PLS.

A. Canonical Correlation Analysis

The most common way of assessing multivariate association is CCA [3]. In CCA, both $\boldsymbol{\xi}_{\mathrm{C}}$ and $\boldsymbol{\omega}_{\mathrm{C}}$ are required to be white, i.e., $\mathbf{A}_\mathrm{C}\mathbf{R}_{xx}\mathbf{A}_\mathrm{C}^T = \mathbf{I}$, $\mathbf{B}_\mathrm{C}\mathbf{R}_{yy}\mathbf{B}_\mathrm{C}^T = \mathbf{I}$. The optimum transformations solving (5) are found as [2], [3]

$$\mathbf{A}_\mathrm{C} = \mathbf{F}_\mathrm{C}^T\mathbf{R}_{xx}^{-1/2} \tag{7}$$
$$\mathbf{B}_\mathrm{C} = \mathbf{G}_\mathrm{C}^T\mathbf{R}_{yy}^{-1/2} \tag{8}$$

where the column-orthogonal matrices $\mathbf{F}_\mathrm{C} \in \mathbb{R}^{m\times p}$ and $\mathbf{G}_\mathrm{C} \in \mathbb{R}^{n\times p}$ are determined by the singular value decomposition (SVD) of the coherence matrix

$$\mathbf{C}_\mathrm{C} = \mathbf{R}_{xx}^{-1/2}\mathbf{R}_{xy}\mathbf{R}_{yy}^{-T/2} = \mathbf{F}_\mathrm{C}\mathbf{K}_\mathrm{C}\mathbf{G}_\mathrm{C}^T. \tag{9}$$

The *canonical correlations* $k_{\mathrm{C},i}$ are normalized such that $0 \le k_{\mathrm{C},i} \le 1$, and *invariant* under nonsingular linear transformation $\mathbf{T}_1 \in \mathbb{R}^{m\times m}$ applied to $\mathbf{x}$ and nonsingular linear transformation $\mathbf{T}_2 \in \mathbb{R}^{n\times n}$ applied to $\mathbf{y}$. The internal representations $\boldsymbol{\xi}_\mathrm{C} = \mathbf{A}_\mathrm{C}\mathbf{x}$ and $\boldsymbol{\omega}_\mathrm{C} = \mathbf{B}_\mathrm{C}\mathbf{y}$ are the *canonical variables*, which are said to be in *canonical coordinates*.

There is a variety of possible correlation coefficients that can be defined based on the first r canonical correlations $\{k_{\mathrm{C},i}\}_{i=1}^r$ for a given rank r. Three particularly compelling coefficients have been proposed for $r = p$ in [3], [15]–[17], and their rank-r generalizations are

$$\rho_{\mathrm{C}_1,r} = \frac{1}{p}\sum_{i=1}^r k_{\mathrm{C},i}^2 \tag{10}$$
$$\rho_{\mathrm{C}_2,r} = 1 - \prod_{i=1}^r \left(1 - k_{\mathrm{C},i}^2\right) \tag{11}$$
$$\rho_{\mathrm{C}_3,r} = \frac{\sum_{i=1}^r \frac{k_{\mathrm{C},i}^2}{1-k_{\mathrm{C},i}^2}}{\sum_{i=1}^r \frac{1}{1-k_{\mathrm{C},i}^2} + (p-r)}. \tag{12}$$

Sometimes we drop the subscript r for convenience. For $r = p$, these coefficients can also be expressed in terms of the original correlation matrices [18]

$$\rho_{\mathrm{C}_1,p} = \frac{1}{p}\mathrm{tr}\left(\mathbf{R}_{xx}^{-1}\mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{xy}^T\right) = \frac{1}{p}\mathrm{tr}\left(\mathbf{C}_\mathrm{C}\mathbf{C}_\mathrm{C}^T\right) \tag{13}$$
$$\rho_{\mathrm{C}_2,p} = 1 - \det\left(\mathbf{I} - \mathbf{R}_{xx}^{-1}\mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{xy}^T\right) = 1 - \det\left(\mathbf{I} - \mathbf{C}_\mathrm{C}\mathbf{C}_\mathrm{C}^T\right) \tag{14}$$
$$\rho_{\mathrm{C}_3,p} = \frac{\mathrm{tr}\left(\mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{xy}^T\left(\mathbf{R}_{xx} - \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{xy}^T\right)^{-1}\right)}{\mathrm{tr}\left(\mathbf{R}_{xx}\left(\mathbf{R}_{xx} - \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{xy}^T\right)^{-1}\right)} = \frac{\mathrm{tr}\left(\mathbf{C}_\mathrm{C}\mathbf{C}_\mathrm{C}^T\left(\mathbf{I} - \mathbf{C}_\mathrm{C}\mathbf{C}_\mathrm{C}^T\right)^{-1}\right)}{\mathrm{tr}\left(\mathbf{I} - \mathbf{C}_\mathrm{C}\mathbf{C}_\mathrm{C}^T\right)^{-1}}. \tag{15}$$

All these coefficients share the invariances of the canonical correlations, i.e., they are invariant under nonsingular linear transformation of $\mathbf{x}$ and $\mathbf{y}$. The first coefficient, ρ_{C_1}, is related to the minimum mean squared error (MMSE) when constructing a rank-r linear estimate $\widehat{\boldsymbol{\xi}}_\mathrm{C}$ of the p-dimensional canonical vector $\boldsymbol{\xi}_\mathrm{C}$ from $\mathbf{y}$. The estimate is $\widehat{\boldsymbol{\xi}}_\mathrm{C} = [\mathbf{K}_\mathrm{C}]_{p\times r}[\mathbf{B}_\mathrm{C}]_{r\times n}\mathbf{y} = [\mathbf{K}_\mathrm{C}]_{p\times r}[\boldsymbol{\omega}_\mathrm{C}]_r$, and the resulting MMSE is [19]

$$E\|\boldsymbol{\xi}_\mathrm{C} - \widehat{\boldsymbol{\xi}}_\mathrm{C}\|^2 = \mathrm{tr}\left(\mathbf{I} - [\mathbf{K}_\mathrm{C}]_{p\times r}[\mathbf{K}_\mathrm{C}]_{p\times r}^T\right) = p - \sum_{i=1}^r k_{\mathrm{C},i}^2 = p\left(1 - \rho_{\mathrm{C}_1}\right). \tag{16}$$

Since CCA is symmetric in $\mathbf{x}$ and $\mathbf{y}$, the same MMSE is obtained when estimating $\boldsymbol{\omega}_\mathrm{C}$ from $\mathbf{x}$.

For $r = p$, the second coefficient, ρ_{C_2}, measures the linear dependence between $\mathbf{x}$ and $\mathbf{y}$. If $\rho_{\mathrm{C}_2} = 0$, then $\mathbf{x}$ and $\mathbf{y}$ are linearly independent. If $\rho_{\mathrm{C}_2} = 1$, then there is at least one canonical coordinate $\xi_{\mathrm{C},i}$ that is perfectly linearly estimable from $\mathbf{y}$. For jointly Gaussian $\mathbf{x}$ and $\mathbf{y}$, ρ_{C_2} also determines the mutual information between $\mathbf{x}$ and $\mathbf{y}$ [19]

$$I(\mathbf{x};\mathbf{y}) = -\frac{1}{2}\log\det\left(\mathbf{I} - [\mathbf{K}_\mathrm{C}]_{p\times r}[\mathbf{K}_\mathrm{C}]_{p\times r}^T\right) = -\frac{1}{2}\sum_{i=1}^r \log\left(1 - k_{\mathrm{C},i}^2\right) \tag{17}$$
$$= -\frac{1}{2}\log\left(1 - \rho_{\mathrm{C}_2}\right). \tag{18}$$

The ith term in the sum (17), $-1/2\log(1 - k_{\mathrm{C},i}^2)$, measures the rate at which the ith canonical coordinate $\omega_{\mathrm{C},i}$ brings information about the ith canonical coordinate $\xi_{\mathrm{C},i}$. For $r < p$, ρ_{C_2} determines the mutual information between rank-r approximations of $\mathbf{x}$ and $\mathbf{y}$.

For $r = p$, an interpretation of ρ_{C_3} is given in [18]. Another interesting interpretation is possible in the signal-plus-uncorrelated-noise case. Let $\mathbf{R}_{xx} = \mathbf{R}_{xy} = \mathbf{S}$ and $\mathbf{R}_{yy} = \mathbf{S} + \mathbf{N}$. It is easy to show that the eigenvalues of the SNR matrix $\mathbf{SN}^{-1}$ are $\{k_{\mathrm{C},i}^2/(1 - k_{\mathrm{C},i}^2)\}$. Hence, they are invariant under nonsingular linear transformation of the signal $\mathbf{x}$, and the numerator of $\rho_{\mathrm{C}_3,p}$ in (15) is $\mathrm{tr}(\mathbf{SN}^{-1})$. Correlation coefficient ρ_{C_3} can, thus, be interpreted as a normalized SNR.

B. *Multivariate Linear Regression (a.k.a. Half-Canonical Correlation Analysis)*

There are many problems where the roles of $\mathbf{x}$ and $\mathbf{y}$ are not interchangeable. For these, the symmetric assessment of correlation by CCA is unsatisfactory. The most obvious example is multivariate linear regression, where a message $\mathbf{x}$ is estimated from a measurement $\mathbf{y}$. The resulting MMSE is invariant under nonsingular transformation of $\mathbf{y}$, but only under *orthogonal* transformation of $\mathbf{x}$. This motivates us to look for transformations $\mathbf{A}_\mathrm{M}$ and $\mathbf{B}_\mathrm{M}$ that maximize the diagonal elements of $\mathbf{K}_\mathrm{M}$ under the constraints that $\mathbf{A}_\mathrm{M}$ be row-orthogonal and only $\boldsymbol{\omega}_\mathrm{M}$, but not $\boldsymbol{\xi}_\mathrm{M}$, be white. Such transformations are found as [2]

$$\mathbf{A}_\mathrm{M} = \mathbf{F}_\mathrm{M}^T \tag{19}$$
$$\mathbf{B}_\mathrm{M} = \mathbf{G}_\mathrm{M}^T\mathbf{R}_{yy}^{-1/2} \tag{20}$$

where the column-orthogonal matrices $\mathbf{F}_\mathrm{M} \in \mathbb{R}^{m\times p}$ and $\mathbf{G}_\mathrm{M} \in \mathbb{R}^{n\times p}$ are determined by the SVD of [5]

$$\mathbf{C}_\mathrm{M} = \mathbf{R}_{xy}\mathbf{R}_{yy}^{-T/2} = \mathbf{F}_\mathrm{M}\mathbf{K}_\mathrm{M}\mathbf{G}_\mathrm{M}^T. \tag{21}$$

The coefficients $k_{\mathrm{M},i}$ are sometimes referred to as *half-canonical correlations* [20], reflecting the fact that they are invariant under nonsingular linear transformation $\mathbf{T}_2 \in \mathbb{R}^{n\times n}$ applied to $\mathbf{y}$, but only *orthogonal* transformation $\mathbf{U}_1 \in \mathbb{R}^{m\times m}$ applied to $\mathbf{x}$. Correspondingly, the internal representations $\boldsymbol{\xi}_{\mathrm{M}} = \mathbf{A}_{\mathrm{M}}\mathbf{x}$ and $\boldsymbol{\omega}_{\mathrm{M}} = \mathbf{B}_{\mathrm{M}}\mathbf{y}$ are said to be in *half-canonical coordinates*.

There are many conceivable correlation coefficients based on $\{k_{\mathrm{M},i}\}_{i=1}^r$. The most common coefficient

$$\rho_{\mathrm{M},r} = \frac{\sum_{i=1}^{r} k_{\mathrm{M},i}^2}{\mathrm{tr}\mathbf{R}_{xx}} \tag{22}$$

has been called the *redundancy index* [21], [22]. For $r = p$, it can also be written as

$$\begin{aligned}\rho_{\mathrm{M},p} &= \frac{\mathrm{tr}\left(\mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{xy}^T\right)}{\mathrm{tr}\mathbf{R}_{xx}}\\ &= \frac{\mathrm{tr}\left(\mathbf{C}_{\mathrm{M}}\mathbf{C}_{\mathrm{M}}^T\right)}{\mathrm{tr}\mathbf{R}_{xx}}.\end{aligned} \tag{23}$$

Coefficient ρ_{M} is related to the MMSE resulting from constructing a rank-r linear estimate $\hat{\mathbf{x}}$ of the message $\mathbf{x}$ from the measurement $\mathbf{y}$ [23]. The estimate is given by $\hat{\mathbf{x}} = [\mathbf{A}_{\mathrm{M}}]_{r\times m}^T[\mathbf{K}_{\mathrm{M}}]_{r\times r}[\mathbf{B}_{\mathrm{M}}]_{r\times n}\mathbf{y} = [\mathbf{A}_{\mathrm{M}}]_{r\times m}^T[\mathbf{K}_{\mathrm{M}}]_{r\times r}[\boldsymbol{\omega}_{\mathrm{M}}]_r$ and the MMSE is

$$E\|\mathbf{x}-\hat{\mathbf{x}}\|^2 = \mathrm{tr}\mathbf{R}_{xx} - \sum_{i=1}^{r} k_{\mathrm{M},i}^2 = \mathrm{tr}\mathbf{R}_{xx}(1-\rho_{\mathrm{M}}). \tag{24}$$

In [21] and [22], ρ_{M} has been expressed and discussed in terms of the (full-)canonical correlations $\{k_{\mathrm{C},i}\}$ and the transformation $\mathbf{A}_{\mathrm{C}}$ in (7), rather than the half-canonical correlations $\{k_{\mathrm{M},i}\}$ as in (22). However, the connection with full-canonical correlations is not very illuminating, as ρ_{M} has the invariances of *half*- rather than *full*-canonical correlations.

A generalization of MLR replaces the orthogonality condition $\mathbf{A}_{\mathrm{M}}\mathbf{A}_{\mathrm{M}}^T = \mathbf{I}$ with $\mathbf{A}_{\mathrm{W}}\mathbf{W}\mathbf{A}_{\mathrm{W}}^T = \mathbf{I}$, where $\mathbf{W} \in \mathbb{R}^{m\times m}$ is a Hermitian, positive definitive weighting matrix [2]. The optimum transformations and corresponding SVD are then given by [2]

$$\mathbf{A}_{\mathrm{W}} = \mathbf{F}_{\mathrm{W}}^T\mathbf{W}^{-1/2} \tag{25}$$

$$\mathbf{B}_{\mathrm{W}} = \mathbf{G}_{\mathrm{W}}^T\mathbf{R}_{yy}^{-1/2} \tag{26}$$

$$\mathbf{C}_{\mathrm{W}} = \mathbf{W}^{-1/2}\mathbf{R}_{xy}\mathbf{R}_{yy}^{-T/2} = \mathbf{F}_{\mathrm{W}}\mathbf{K}_{\mathrm{W}}\mathbf{G}_{\mathrm{W}}^T. \tag{27}$$

The correlation coefficient generalizing (22) to weighted MLR is

$$\rho_{\mathrm{W},r} = \frac{\sum_{i=1}^{r} k_{\mathrm{W},i}^2}{\mathrm{tr}(\mathbf{W}^{-1}\mathbf{R}_{xx})} \tag{28}$$

$$\rho_{\mathrm{W},p} = \frac{\mathrm{tr}\left(\mathbf{W}^{-1}\mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{xy}^T\right)}{\mathrm{tr}(\mathbf{W}^{-1}\mathbf{R}_{xx})}. \tag{29}$$

Therefore, ρ_{W} is related to the weighted MMSE when constructing a rank-r linear estimate $\hat{\mathbf{x}}$ of the message $\mathbf{x}$ from the measurement $\mathbf{y}$. The estimate is given by $\hat{\mathbf{x}} = [\mathbf{A}_{\mathrm{W}}]_{r\times m}^{-1}[\mathbf{K}_{\mathrm{W}}]_{r\times r}[\mathbf{B}_{\mathrm{W}}]_{r\times n}\mathbf{y} = [\mathbf{A}_{\mathrm{W}}]_{r\times m}^{-1}[\mathbf{K}_{\mathrm{W}}]_{r\times r}[\boldsymbol{\omega}_{\mathrm{W}}]_r$, where $[\mathbf{A}_{\mathrm{W}}]_{r\times m}^{-1}$ is a *right* inverse: $[\mathbf{A}_{\mathrm{W}}]_{r\times m}^{-1} = \mathbf{W}^{1/2}[\mathbf{F}_{\mathrm{W}}]_{m\times r}$. The weighted MMSE is

$$\begin{aligned}E\|\mathbf{x}-\hat{\mathbf{x}}\|_{\mathbf{W}^{-1}}^2 &= E(\mathbf{x}-\hat{\mathbf{x}})^T\mathbf{W}^{-1}(\mathbf{x}-\hat{\mathbf{x}})\\ &= \mathrm{tr}(\mathbf{W}^{-1}\mathbf{R}_{xx})(1-\rho_{\mathrm{W}}).\end{aligned} \tag{30}$$

The weighted half-canonical correlations $\{k_{\mathrm{W},i}\}$ and thus ρ_{W} have different invariances than the half-canonical correlations $\{k_{\mathrm{M},i}\}$ and ρ_{M}. It is interesting to note that the full-canonical correlations $\{k_{\mathrm{C},i}\}$ are actually weighted half-canonical correlations $\{k_{\mathrm{W},i}\}$ with the particular weighting matrix $\mathbf{W} = \mathbf{R}_{xx}$. This observation has already been made in [10].

C. Partial Least Squares

Some problems are naturally invariant only to *orthogonal* transformation of both $\mathbf{x}$ and $\mathbf{y}$. We are then interested in maximizing the diagonal elements of $\mathbf{K}_{\mathrm{P}}$ under the constraints $\mathbf{A}_{\mathrm{P}}\mathbf{A}_{\mathrm{P}}^T = \mathbf{I}$ and $\mathbf{B}_{\mathrm{P}}\mathbf{B}_{\mathrm{P}}^T = \mathbf{I}$. This is achieved by [2], [8]

$$\mathbf{A}_{\mathrm{P}} = \mathbf{F}_{\mathrm{P}}^T \tag{31}$$

$$\mathbf{B}_{\mathrm{P}} = \mathbf{G}_{\mathrm{P}}^T \tag{32}$$

where $\mathbf{F}_{\mathrm{P}}$ and $\mathbf{G}_{\mathrm{P}}$ are given by the SVD of

$$\mathbf{C}_{\mathrm{P}} = \mathbf{R}_{xy} = \mathbf{F}_{\mathrm{P}}\mathbf{K}_{\mathrm{P}}\mathbf{G}_{\mathrm{P}}^T. \tag{33}$$

This technique is a variant of PLS. Note, however, that our description of PLS follows [8], which differs from the original PLS algorithm [6], [7]. In the original PLS algorithm, both $E\boldsymbol{\xi}\boldsymbol{\xi}^T$ and $E\boldsymbol{\omega}\boldsymbol{\omega}^T$ are diagonal matrices (albeit not identity matrices), and $\mathbf{K}$ is not generally diagonal. In the version of [8], $\mathbf{K}_{\mathrm{P}}$ is diagonal, but $E\boldsymbol{\xi}_{\mathrm{P}}\boldsymbol{\xi}_{\mathrm{P}}^T$ and $E\boldsymbol{\omega}_{\mathrm{P}}\boldsymbol{\omega}_{\mathrm{P}}^T$ are generally not.

The PLS coefficients $k_{\mathrm{P},i}$ are invariant under *orthogonal* transformation $\mathbf{U}_1 \in \mathbb{R}^{m\times m}$ applied to $\mathbf{x}$ and *orthogonal* transformation $\mathbf{U}_2 \in \mathbb{R}^{n\times n}$ applied to $\mathbf{y}$. The internal representations $\boldsymbol{\xi}_{\mathrm{P}} = \mathbf{A}_{\mathrm{P}}\mathbf{x}$ and $\boldsymbol{\omega}_{\mathrm{P}} = \mathbf{B}_{\mathrm{P}}\mathbf{y}$ are referred to as *latent variables*. Among the possible correlation coefficients

$$\rho_{\mathrm{P},r} = \frac{\sum_{i=1}^{r} k_{\mathrm{P},i}^2}{\sqrt{\mathrm{tr}\left(\mathbf{R}_{xx}\mathbf{R}_{xx}^T\right)\mathrm{tr}\left(\mathbf{R}_{yy}\mathbf{R}_{yy}^T\right)}} \tag{34}$$

is used most frequently. For $r = p$, it can be expressed as [24]

$$\rho_{\mathrm{P},p} = \frac{\mathrm{tr}\left(\mathbf{R}_{xy}\mathbf{R}_{xy}^T\right)}{\sqrt{\mathrm{tr}\left(\mathbf{R}_{xx}\mathbf{R}_{xx}^T\right)\mathrm{tr}\left(\mathbf{R}_{yy}\mathbf{R}_{yy}^T\right)}}. \tag{35}$$

Both PLS and CCA provide a symmetric measure of multivariate association. However, there are key advantages of PLS if it is applied to sample correlation matrices. Since the computation of $\mathbf{C}_{\mathrm{C}}$ in (9) requires the computation of inverses $\mathbf{R}_{xx}^{-1}$ and $\mathbf{R}_{yy}^{-1}$, CCA can become unstable if $\mathbf{R}_{xx}$ or $\mathbf{R}_{yy}$ are close to being singular. That is, the SVD (9) can change significantly after recomputing sample correlation matrices with added samples nearly collinear with previous samples. These problems are rectified in PLS, which is why PLS has been referred to as "robust canonical analysis" [25].

III. Majorization and Correlation Spread

Canonical correlations can be derived in a great number of ways. The most common problem formulation is this: Find two vectors $\mathbf{a}_{C,1} \in \mathbb{R}^m$ and $\mathbf{b}_{C,1} \in \mathbb{R}^n$ such that the absolute value of the scalar correlation coefficient between canonical variables $\xi_{C,1} = \mathbf{a}_{C,1}^T\mathbf{x}$ and $\omega_{C,1} = \mathbf{b}_{C,1}^T\mathbf{y}$ is maximized

$$k_{C,1} = \max_{\mathbf{a}_{C,1},\mathbf{b}_{C,1}} \left| \frac{\mathbf{a}_{C,1}^T\mathbf{R}_{xy}\mathbf{b}_{C,1}}{\sqrt{\mathbf{a}_{C,1}^T\mathbf{R}_{xx}\mathbf{a}_{C,1}}\sqrt{\mathbf{b}_{C,1}^T\mathbf{R}_{yy}\mathbf{b}_{C,1}}} \right|. \quad (36)$$

The next pair of canonical variables $(\xi_{C,2}, \omega_{C,2})$ maximizes the absolute value of the scalar correlation coefficient, $k_{C,2}$, between $\xi_{C,2} = \mathbf{a}_{C,2}^T\mathbf{x}$ and $\omega_{C,2} = \mathbf{b}_{C,2}^T\mathbf{y}$, subject to the constraint that they are to be uncorrelated with the first pair. A total of p correlations is determined in this manner.

In this paper, we have followed [2] by formulating canonical correlations as the solution to the equivalent maximization problem (5). This puts canonical correlations in a majorization context. The constraints imposed by CCA, MLR, or PLS, summarized following (5), determine $\mathbf{A}$ and $\mathbf{B}$ up to multiplication from the left by row-orthogonal matrices $\mathbf{F}^T$ and $\mathbf{G}^T$, respectively. That is, the singular values of $\mathbf{A}\mathbf{R}_{xy}\mathbf{B}^T$ are fixed by these constraints. Maximizing all partial sums (5) requires maximum spread among the diagonal elements of $\mathbf{A}\mathbf{R}_{xy}\mathbf{B}^T$. As outlined in Appendix A, the absolute values of the diagonal elements of an $m \times n$ matrix are *weakly majorized* by its singular values. Maximum spread is thus achieved by making $\mathbf{A}\mathbf{R}_{xy}\mathbf{B}^T$ diagonal, which means that $\mathbf{F}^T$ and $\mathbf{G}^T$ are determined via SVD. This solves the CCA, MLR, and PLS problems.

Using the results from Appendix A, it is not difficult to show that all correlation coefficients presented in this paper are Schur-convex and increasing functions of the diagonal correlations $\{k_i\}$. By maximizing (5) subject to the constraints imposed by the correlation analysis technique, the correlation coefficients are then also maximized for all ranks r.

An interesting question in this context is how much of the overall correlation is captured by r coefficients $\{k_i\}_{i=1}^r$, which can be defined either through CCA, MLR, or PLS. One could, of course, compute the fraction ρ_r/ρ_p for all $1 \le r < p$. Alternatively, we can define the *correlation spread* as

$$\begin{aligned}\sigma^2 &= \frac{1}{p}\frac{\widehat{\mathrm{var}}\left(\{k_i^2\}\right)}{\hat{\mu}^2\left(\{k_i^2\}\right)} \\ &= \frac{p}{p-1}\left(\frac{\sum_{i=1}^p k_i^4}{\left(\sum_{i=1}^p k_i^2\right)^2} - \frac{1}{p}\right) \quad (37)\end{aligned}$$

where $\widehat{\mathrm{var}}$ and $\hat{\mu}$ denote sample variance and sample mean, respectively. The correlation spread provides a *single, normalized measure* of how much of the overall correlation is concentrated in a few coefficients. If there is only one nonzero coefficient k_1, then $\sigma^2 = 1$. If all coefficients are equal, $k_1 = k_2 = \cdots = k_p$, then $\sigma^2 = 0$.

The definition (37) is inspired by the definition of the *degree of polarization* [13], [14] of a random vector $\mathbf{x}$. The degree of polarization measures the spread among the eigenvalues of $\mathbf{R}_{xx}$. A random vector $\mathbf{x}$ is said to be completely polarized if all of its energy is concentrated in one direction, i.e., if there is only one nonzero eigenvalue. On the other hand, $\mathbf{x}$ is unpolarized if its energy is equally distributed among all dimensions, i.e., if all eigenvalues are equal. The correlation spread σ^2 generalizes this idea to the correlation between two random vectors $\mathbf{x}$ and $\mathbf{y}$.

It can sometimes be appropriate to modify the definition (37) by replacing the coefficients k_i^2 with a function of the coefficients, $f(k_i^2)$, as

$$\begin{aligned}\sigma_f^2 &= \frac{1}{p}\frac{\widehat{\mathrm{var}}\left(\{f\left(k_i^2\right)\}\right)}{\hat{\mu}^2\left(\{f\left(k_i^2\right)\}\right)} \\ &= \frac{p}{p-1}\left(\frac{\sum_{i=1}^p f^2\left(k_i^2\right)}{\left[\sum_{i=1}^p f\left(k_i^2\right)\right]^2} - \frac{1}{p}\right). \quad (38)\end{aligned}$$

For instance, if we are interested in the distribution of *information rate* over individual components, then we compute the coefficients $k_{C,i}^2$ through CCA and set $f(k_{C,i}^2) = \log(1 - k_{C,i}^2)$. If $\sigma_{C,f}^2 = 1$, then there is only one canonical coordinate $\omega_{C,1}$ that carries information about $\mathbf{x}$ (or, alternatively, only $\xi_{C,1}$ carries information about $\mathbf{y}$). If $\sigma_{C,f}^2 = 0$, then all canonical coordinates $\{\omega_{C,i}\}_{i=1}^p$ carry an equal amount of information about $\mathbf{x}$ (or, alternatively, all $\{\xi_{C,i}\}_{i=1}^p$ carry an equal amount of information about $\mathbf{y}$).

IV. Correlation Analysis of Complex Vectors

Let us now consider two *complex* zero-mean random vectors $\mathbf{x} \in \mathbb{C}^m$ and $\mathbf{y} \in \mathbb{C}^n$. For complex random vectors, there are three types of correlation: first, the correlation between $\mathbf{x}$ and $\mathbf{y}$, as measured by the standard Hermitian cross-correlation matrix $\mathbf{R}_{xy} = E\mathbf{x}\mathbf{y}^H$, and second, the correlation between $\mathbf{x}$ and the conjugate $\mathbf{y}^*$, as measured by the complementary cross-correlation matrix $\widetilde{\mathbf{R}}_{xy} = E\mathbf{x}\mathbf{y}^T$. If $\widetilde{\mathbf{R}}_{xx} = \mathbf{0}$, $\mathbf{x}$ is called *proper*, otherwise *improper*. If $\mathbf{x}$ is proper and Gaussian, then it is *circular* as $\mathbf{x}$ and $e^{j\alpha}\mathbf{x}$ have the same probability density functions for all real α [26]. If $\widetilde{\mathbf{R}}_{xy} = \mathbf{0}$, $\mathbf{x}$ and $\mathbf{y}$ are said to be *cross-proper*. If $\widetilde{\mathbf{R}}_{xx} = \mathbf{0}$, $\widetilde{\mathbf{R}}_{xy} = \mathbf{0}$, and $\widetilde{\mathbf{R}}_{yy} = \mathbf{0}$, $\mathbf{x}$ and $\mathbf{y}$ are called *jointly proper*.

Finally, the complete second-order statistics are described by the augmented cross-correlation matrix $\overline{\mathbf{R}}_{xy}$ [20], which is the cross-correlation matrix of the augmented vectors $\bar{\mathbf{x}} = [\mathbf{x}^T\ \mathbf{x}^H]^T$ and $\bar{\mathbf{y}} = [\mathbf{y}^T\ \mathbf{y}^H]^T$

$$\overline{\mathbf{R}}_{xy} = E\bar{\mathbf{x}}\bar{\mathbf{y}}^H = \begin{bmatrix} \mathbf{R}_{xy} & \widetilde{\mathbf{R}}_{xy} \\ \widetilde{\mathbf{R}}_{xy}^* & \mathbf{R}_{xy}^* \end{bmatrix}. \quad (39)$$

In order to preserve the structure of $\bar{\mathbf{x}}$, transformations applied to $\bar{\mathbf{x}}$ must be of the form [20]

$$\overline{\mathbf{A}} = \begin{bmatrix} \overline{\mathbf{A}}_1 & \overline{\mathbf{A}}_2 \\ \overline{\mathbf{A}}_2^* & \overline{\mathbf{A}}_1^* \end{bmatrix}. \quad (40)$$

Then, $\bar{\boldsymbol{\xi}}' = \overline{\mathbf{A}}\bar{\mathbf{x}}$ describes the so-called *widely linear* [27] or linear-conjugate linear transformation $\boldsymbol{\xi}' = \overline{\mathbf{A}}_1\mathbf{x} + \overline{\mathbf{A}}_2\mathbf{x}^*$. Matrices that satisfy the block structure (40) are overlined.

A. Rotational, Reflectional, and Total Correlations

Either one of the pairs $(\mathbf{x}, \mathbf{y})$, $(\mathbf{x}, \mathbf{y}^*)$, or $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ can be analyzed by either CCA, MLR, or PLS. The required SVDs and optimum transformations for the nine possible combinations are given in Table I. The expression $\widetilde{\mathbf{C}} = \widetilde{\mathbf{F}}\widetilde{\mathbf{K}}\widetilde{\mathbf{G}}^T$ in this table is

TABLE I
SVDs AND OPTIMUM TRANSFORMATIONS FOR DIFFERENT CORRELATION TYPES AND CORRELATION ANALYSIS TECHNIQUES

correlation type	SVD	CCA	MLR	PLS
$(\mathbf{x},\mathbf{y})$ (rotational)	$\mathbf{C}=\mathbf{FKG}^H$	$\mathbf{C}_\mathrm{C}=\mathbf{R}_{xx}^{-1/2}\mathbf{R}_{xy}\mathbf{R}_{yy}^{-H/2}$ $\mathbf{A}_\mathrm{C}=\mathbf{F}_\mathrm{C}^H\mathbf{R}_{xx}^{-1/2}$ $\mathbf{B}_\mathrm{C}=\mathbf{G}_\mathrm{C}^H\mathbf{R}_{yy}^{-1/2}$	$\mathbf{C}_\mathrm{M}=\mathbf{R}_{xy}\mathbf{R}_{yy}^{-H/2}$ $\mathbf{A}_\mathrm{M}=\mathbf{F}_\mathrm{M}^H$ $\mathbf{B}_\mathrm{M}=\mathbf{G}_\mathrm{M}^H\mathbf{R}_{yy}^{-1/2}$	$\mathbf{C}_\mathrm{P}=\mathbf{R}_{xy}$ $\mathbf{A}_\mathrm{P}=\mathbf{F}_\mathrm{P}^H$ $\mathbf{B}_\mathrm{P}=\mathbf{G}_\mathrm{P}^H$
$(\mathbf{x},\mathbf{y}^*)$ (reflectional)	$\widetilde{\mathbf{C}}=\widetilde{\mathbf{F}}\widetilde{\mathbf{K}}\widetilde{\mathbf{G}}^T$	$\widetilde{\mathbf{C}}_\mathrm{C}=\mathbf{R}_{xx}^{-1/2}\widetilde{\mathbf{R}}_{xy}\mathbf{R}_{yy}^{-T/2}$ $\widetilde{\mathbf{A}}_\mathrm{C}=\widetilde{\mathbf{F}}_\mathrm{C}^H\mathbf{R}_{xx}^{-1/2}$ $\widetilde{\mathbf{B}}_\mathrm{C}=\widetilde{\mathbf{G}}_\mathrm{C}^T\mathbf{R}_{yy}^{-*/2}$	$\widetilde{\mathbf{C}}_\mathrm{M}=\widetilde{\mathbf{R}}_{xy}\mathbf{R}_{yy}^{-T/2}$ $\widetilde{\mathbf{A}}_\mathrm{M}=\widetilde{\mathbf{F}}_\mathrm{M}^H$ $\widetilde{\mathbf{B}}_\mathrm{M}=\widetilde{\mathbf{G}}_\mathrm{M}^T\mathbf{R}_{yy}^{-*/2}$	$\widetilde{\mathbf{C}}_\mathrm{P}=\widetilde{\mathbf{R}}_{xy}$ $\widetilde{\mathbf{A}}_\mathrm{P}=\widetilde{\mathbf{F}}_\mathrm{P}^H$ $\widetilde{\mathbf{B}}_\mathrm{P}=\widetilde{\mathbf{G}}_\mathrm{P}^T$
$(\bar{\mathbf{x}},\bar{\mathbf{y}})$ (total)	$\overline{\mathbf{C}}=\overline{\mathbf{FKG}}^H$	$\overline{\mathbf{C}}_\mathrm{C}=\overline{\mathbf{R}}_{xx}^{-1/2}\overline{\mathbf{R}}_{xy}\overline{\mathbf{R}}_{yy}^{-H/2}$ $\overline{\mathbf{A}}_\mathrm{C}=\overline{\mathbf{F}}_\mathrm{C}^H\overline{\mathbf{R}}_{xx}^{-1/2}$ $\overline{\mathbf{B}}_\mathrm{C}=\overline{\mathbf{G}}_\mathrm{C}^H\overline{\mathbf{R}}_{yy}^{-1/2}$	$\overline{\mathbf{C}}_\mathrm{M}=\overline{\mathbf{R}}_{xy}\overline{\mathbf{R}}_{yy}^{-H/2}$ $\overline{\mathbf{A}}_\mathrm{M}=\overline{\mathbf{F}}_\mathrm{M}^H$ $\overline{\mathbf{B}}_\mathrm{M}=\overline{\mathbf{G}}_\mathrm{M}^H\overline{\mathbf{R}}_{yy}^{-1/2}$	$\overline{\mathbf{C}}_\mathrm{P}=\overline{\mathbf{R}}_{xy}$ $\overline{\mathbf{A}}_\mathrm{P}=\overline{\mathbf{F}}_\mathrm{P}^H$ $\overline{\mathbf{B}}_\mathrm{P}=\overline{\mathbf{G}}_\mathrm{P}^H$

the usual SVD but with column-unitary $\widetilde{\mathbf{G}}^*\in\mathbb{C}^{n\times p}$, so that the regular transpose rather than the Hermitian transpose is taken to match the structure of $\widetilde{\mathbf{C}}$. As discussed in Appendix B, all factors in the SVD $\overline{\mathbf{C}}=\overline{\mathbf{FKG}}^H$ must satisfy the block structure (40). That is

$$\begin{bmatrix}\overline{\mathbf{C}}_1 & \overline{\mathbf{C}}_2\\ \overline{\mathbf{C}}_2^* & \overline{\mathbf{C}}_1^*\end{bmatrix}=\frac{1}{2}\begin{bmatrix}\overline{\mathbf{F}}_1 & \overline{\mathbf{F}}_2\\ \overline{\mathbf{F}}_2^* & \overline{\mathbf{F}}_1^*\end{bmatrix}\begin{bmatrix}\overline{\mathbf{K}}^{(1)}+\overline{\mathbf{K}}^{(2)} & \overline{\mathbf{K}}^{(1)}-\overline{\mathbf{K}}^{(2)}\\ \overline{\mathbf{K}}^{(1)}-\overline{\mathbf{K}}^{(2)} & \overline{\mathbf{K}}^{(1)}+\overline{\mathbf{K}}^{(2)}\end{bmatrix}\times\begin{bmatrix}\overline{\mathbf{G}}_1^H & \overline{\mathbf{G}}_2^T\\ \overline{\mathbf{G}}_2^H & \overline{\mathbf{G}}_1^T\end{bmatrix} \quad (41)$$

where $\overline{\mathbf{K}}^{(1)}$ and $\overline{\mathbf{K}}^{(2)}$ are two diagonal matrices containing two sets of coefficients $\{\bar{k}_i^{(1)}\}_{i=1}^p$ and $\{\bar{k}_i^{(2)}\}_{i=1}^p$ on their respective diagonals, with ordering $\bar{k}_1^{(1)}\ge\bar{k}_1^{(2)}\ge\bar{k}_2^{(1)}\ge\bar{k}_2^{(2)}\ge\cdots\ge\bar{k}_p^{(1)}\ge\bar{k}_p^{(2)}\ge 0$. The matrices $\overline{\mathbf{F}}\in\mathbb{C}^{2m\times 2p}$ and $\overline{\mathbf{G}}\in\mathbb{C}^{2n\times 2p}$ are both column-unitary.

The decompositions corresponding to $(\mathbf{x},\mathbf{y})$, $(\mathbf{x},\mathbf{y}^*)$, and $(\bar{\mathbf{x}},\bar{\mathbf{y}})$ characterize different types of correlation. If we analyze $(\mathbf{x},\mathbf{y})$, the internal representations are $\boldsymbol{\xi}=\mathbf{Ax}$ and $\boldsymbol{\omega}=\mathbf{By}$. Consider two different pairs $(\mathbf{x}^{[1]},\mathbf{y}^{[1]})$ and $(\mathbf{x}^{[2]},\mathbf{y}^{[2]})$ of samples of $(\mathbf{x},\mathbf{y})$ and their corresponding internal descriptions $(\boldsymbol{\xi}^{[1]},\boldsymbol{\omega}^{[1]})$ and $(\boldsymbol{\xi}^{[2]},\boldsymbol{\omega}^{[2]})$. Perfect correlation[1] $k_i=E\xi_i\omega_i^*=(E|\xi_i|^2E|\omega_i|^2)^{-1/2}$ for some index i means $\xi_i=k_i\omega_i$, and hence the angles of $\xi_i^{[1]}$ and $\xi_i^{[2]}$ equal the angles of $\omega_i^{[1]}$ and $\omega_i^{[2]}$: $\measuredangle\xi_i^{[1]}=\measuredangle\omega_i^{[1]}$ and $\measuredangle\xi_i^{[2]}=\measuredangle\omega_i^{[2]}$. Therefore, the correlation k_i measures the tendency of samples of ξ_i and ω_i to turn clockwise or counterclockwise together. This is why we refer to $\{k_i\}$ as *rotational* correlations [11].

If we deal with $(\mathbf{x},\mathbf{y}^*)$, the internal representations are $\widetilde{\boldsymbol{\xi}}=\widetilde{\mathbf{A}}\mathbf{x}$ and $\widetilde{\boldsymbol{\omega}}^*=\widetilde{\mathbf{B}}\mathbf{y}^*$. Perfect correlation $\tilde{k}_i=E\tilde{\xi}_i\tilde{\omega}_i=(E|\tilde{\xi}_i|^2E|\tilde{\omega}_i|^2)^{-1/2}$ for some index i means $\tilde{\xi}_i=\tilde{k}_i\tilde{\omega}_i^*$. For the internal descriptions computed from two sample pairs $(\mathbf{x}^{[1]},\mathbf{y}^{[1]})$ and $(\mathbf{x}^{[2]},\mathbf{y}^{[2]})$, the angles of $\tilde{\xi}_i^{[1]}$ and $\tilde{\xi}_i^{[2]}$ equal *minus* the angles of $\tilde{\omega}_i^{[1]}$ and $\tilde{\omega}_i^{[2]}$: $\measuredangle\tilde{\xi}_i^{[1]}=-\measuredangle\tilde{\omega}_i^{[1]}$ and $\measuredangle\tilde{\xi}_i^{[2]}=-\measuredangle\tilde{\omega}_i^{[2]}$. Therefore, the correlation $\tilde{k}_i$ measures the tendency of samples of $\tilde{\xi}_i$ and $\tilde{\omega}_i$ to turn in opposite directions—one turning clockwise and the other counterclockwise. Thus, we call $\{\tilde{k}_i\}$ *reflectional* correlations [11].

Finally, if we analyze $(\bar{\mathbf{x}},\bar{\mathbf{y}})$, the internal representations are $\boldsymbol{\xi}'=\overline{\mathbf{A}}_1\mathbf{x}+\overline{\mathbf{A}}_2\mathbf{x}^*$ and $\boldsymbol{\omega}'=\overline{\mathbf{B}}_1\mathbf{y}+\overline{\mathbf{B}}_2\mathbf{y}^*$. We use the prime symbol to clarify that $\boldsymbol{\xi}'$ and $\boldsymbol{\omega}'$ are different than $\boldsymbol{\xi}$ and $\boldsymbol{\omega}$ obtained when analyzing $(\mathbf{x},\mathbf{y})$. Note that $\boldsymbol{\xi}'$ and $\boldsymbol{\omega}'$ are jointly proper if and only if $\bar{k}_i^{(1)}=\bar{k}_i^{(2)}$ for all $i=1,\ldots,p$. The correlation $E\xi_i'\omega_i'^*=(1/2)(\bar{k}_i^{(1)}+\bar{k}_i^{(2)})$ provides an assessment of the *total* correlation, taking into account both rotational and reflectional dependencies. The best possible match between ξ_i' and ω_i' is achieved if $\bar{k}_i^{(1)}=\bar{k}_i^{(2)}=(E|\xi_i'|^2E|\omega_i'|^2)^{-1/2}$, which means that ξ_i' and ω_i' are jointly proper.

B. Correlation Measures

On the basis of the three different correlation types—rotational, reflectional, and total—we can define three different types of correlation coefficients. We now discuss these three versions for ρ_{C_2}, ρ_M, ρ_W, and ρ_P. The extension to other correlation coefficients should then be obvious. For ρ_{C_2} based on CCA

$$\rho_{\mathrm{C}_2,r}=1-\prod_{i=1}^{r}\left(1-k_{\mathrm{C},i}^2\right) \quad (42)$$

$$\tilde{\rho}_{\mathrm{C}_2,r}=1-\prod_{i=1}^{r}\left(1-\tilde{k}_{\mathrm{C},i}^2\right) \quad (43)$$

$$\bar{\rho}_{\mathrm{C}_2,r}=1-\prod_{i=1}^{r}\left[\left(1-\left(\bar{k}_{\mathrm{C},i}^{(1)}\right)^2\right)\left(1-\left(\bar{k}_{\mathrm{C},i}^{(2)}\right)^2\right)\right]^{1/2}. \quad (44)$$

For jointly Gaussian $\mathbf{x}$ and $\mathbf{y}$, ρ_{C_2} determines the mutual information between $\mathbf{x}$ and $\mathbf{y}$, if *only rotational* dependencies are taken into account:

$$I(\mathbf{x};\mathbf{y})=-\log(1-\rho_{\mathrm{C}_2}). \quad (45)$$

Note that $I(\mathbf{x};\mathbf{y})$ for complex $\mathbf{x}$ and $\mathbf{y}$ does not contain the normalization by 1/2, as in the real case (18). If *only reflectional* dependencies are considered, $\tilde{\rho}_{\mathrm{C}_2}$ is used in place of ρ_{C_2}. The

[1]For MLR, perfect correlation is $k_{\mathrm{M},i}=(E|\xi_{\mathrm{M},i}|^2)^{-1/2}$, and for CCA, $k_{\mathrm{C},i}=1$.

total mutual information, allowing for both rotation and reflection, is obtained with $\bar{\rho}_{\mathrm{C}_2}$. Note that if $\bar{k}^{(1)}_{\mathrm{C},i} = \bar{k}^{(2)}_{\mathrm{C},i} = k_{\mathrm{C},i}$ for $i = 1,\ldots,r$, then $\rho_{\mathrm{C}_2,r} = \bar{\rho}_{\mathrm{C}_2,r}$. Similarly, if $\bar{k}^{(1)}_{\mathrm{C},i} = \bar{k}^{(2)}_{\mathrm{C},i} = \tilde{k}_{\mathrm{C},i}$ for $i = 1,\ldots,r$, then $\tilde{\rho}_{\mathrm{C}_2,r} = \bar{\rho}_{\mathrm{C}_2,r}$.

The three versions of ρ_{W}, based on weighted MLR with positive definite augmented weighting matrix $\overline{\mathbf{W}}$, are

$$\rho_{\mathrm{W},r} = \frac{\sum_{i=1}^{r} k^2_{\mathrm{W},i}}{\mathrm{tr}\left(\overline{\mathbf{W}}_1^{-1}\mathbf{R}_{xx}\right)} \tag{46}$$

$$\tilde{\rho}_{\mathrm{W},r} = \frac{\sum_{i=1}^{r} \tilde{k}^2_{\mathrm{W},i}}{\mathrm{tr}\left(\overline{\mathbf{W}}_1^{-1}\mathbf{R}_{xx}\right)} \tag{47}$$

$$\bar{\rho}_{\mathrm{W},r} = \frac{\sum_{i=1}^{r} \left(\bar{k}^{(1)}_{\mathrm{W},i}\right)^2 + \left(\bar{k}^{(2)}_{\mathrm{W},i}\right)^2}{\mathrm{tr}(\overline{\mathbf{W}}^{-1}\overline{\mathbf{R}}_{xx})}. \tag{48}$$

For the computation of $\rho_{\mathrm{W},r}$ in (46) and $\tilde{\rho}_{\mathrm{W},r}$ in (47), the weighting matrix $\overline{\mathbf{W}}_1$ is the positive definite northwest block of $\overline{\mathbf{W}}$. The weighted MMSE allowing *strictly linear* estimates $\hat{\mathbf{x}} = \mathbf{H}\mathbf{y}$ only, is obtained, analogously to (30), as $\mathrm{tr}(\overline{\mathbf{W}}_1^{-1}\mathbf{R}_{xx})(1-\rho_{\mathrm{W}})$. The weighted MMSE allowing *conjugate linear* estimates $\tilde{\hat{\mathbf{x}}} = \tilde{\mathbf{H}}\mathbf{y}^*$ only, is $\mathrm{tr}(\overline{\mathbf{W}}_1^{-1}\mathbf{R}_{xx})(1-\tilde{\rho}_{\mathrm{W}})$. Finally, the weighted MMSE using a *widely linear* estimate $\hat{\mathbf{x}}' = \overline{\mathbf{H}}_1\mathbf{y}+\overline{\mathbf{H}}_2\mathbf{y}^*$, is $\mathrm{tr}(\overline{\mathbf{W}}^{-1}\overline{\mathbf{R}}_{xx})(1-\bar{\rho}_{\mathrm{W}})/2$. Results for the standard MMSE and ρ_{M} can be obtained with $\overline{\mathbf{W}} = \mathbf{I}$. For $\bar{\rho}_{\mathrm{M},r}$, the normalization in (48) is then $\mathrm{tr}(\overline{\mathbf{W}}^{-1}\overline{\mathbf{R}}_{xx}) = 2\mathrm{tr}\mathbf{R}_{xx}$.

The three definitions of ρ_{P}, based on PLS, are

$$\rho_{\mathrm{P},r} = \frac{\sum_{i=1}^{r} k^2_{\mathrm{P},i}}{\sqrt{\mathrm{tr}\left(\mathbf{R}_{xx}\mathbf{R}^H_{xx}\right)\mathrm{tr}\left(\mathbf{R}_{yy}\mathbf{R}^H_{yy}\right)}} \tag{49}$$

$$\tilde{\rho}_{\mathrm{P},r} = \frac{\sum_{i=1}^{r} \tilde{k}^2_{\mathrm{P},i}}{\sqrt{\mathrm{tr}\left(\mathbf{R}_{xx}\mathbf{R}^H_{xx}\right)\mathrm{tr}\left(\mathbf{R}_{yy}\mathbf{R}^H_{yy}\right)}} \tag{50}$$

$$\bar{\rho}_{\mathrm{P},r} = \frac{\sum_{i=1}^{r} \left(\bar{k}^{(1)}_{\mathrm{P},i}\right)^2 + \left(\bar{k}^{(2)}_{\mathrm{P},i}\right)^2}{\sqrt{\mathrm{tr}\left(\overline{\mathbf{R}}_{xx}\overline{\mathbf{R}}^H_{xx}\right)\mathrm{tr}\left(\overline{\mathbf{R}}_{yy}\overline{\mathbf{R}}^H_{yy}\right)}}. \tag{51}$$

The coefficients $\rho_{\mathrm{P},r}$, $\tilde{\rho}_{\mathrm{P},r}$, and $\bar{\rho}_{\mathrm{P},r}$ evaluate the match between $\mathbf{x}$ and $\mathbf{y}$ under unitary, conjugate unitary, and widely unitary transformations, respectively. Note that if $\overline{\mathbf{U}}$ is widely unitary, i.e., it satisfies the block structure (40) and $\overline{\mathbf{U}}\overline{\mathbf{U}}^H = \overline{\mathbf{U}}^H\overline{\mathbf{U}} = \mathbf{I}$, its northwest block $\overline{\mathbf{U}}_1$ is generally *not* unitary. In contrast, if $\overline{\mathbf{A}}$ is an augmented matrix that describes a general widely linear transformation, its northwest block $\overline{\mathbf{A}}_1$ always describes a strictly linear transformation.

C. Majorization and Correlation Spread

While there is no general ordering of rotational and reflectional correlations, i.e., either one may be greater, it is clear that total correlations, which take into account both rotational and reflectional dependencies, must always be greater or equal to rotational or reflectional correlations. Yet this does not mean that total correlation *coefficients* are necessarily greater than rotational or reflectional correlation *coefficients*, as we will illustrate now.

The fact that the set of strictly linear transformations, $\mathbf{H}\mathbf{y}$, and the set of conjugate linear transformations, $\tilde{\mathbf{H}}\mathbf{y}^*$, are each proper subsets of the set of widely linear transformations, $\overline{\mathbf{H}}_1\mathbf{y}+\overline{\mathbf{H}}_2\mathbf{y}^*$, implies the following weak majorization results:

$$\begin{aligned}&[k_1,k_1,k_2,k_2,\ldots,k_p,k_p]^T\\ &\qquad \prec_w \left[\bar{k}^{(1)}_1,\bar{k}^{(2)}_1,\bar{k}^{(1)}_2,\bar{k}^{(2)}_2,\ldots,\bar{k}^{(1)}_p,\bar{k}^{(2)}_p\right]^T\end{aligned} \tag{52}$$

$$\begin{aligned}&[\tilde{k}_1,\tilde{k}_1,\tilde{k}_2,\tilde{k}_2,\ldots,\tilde{k}_p,\tilde{k}_p]^T\\ &\qquad \prec_w \left[\bar{k}^{(1)}_1,\bar{k}^{(2)}_1,\bar{k}^{(1)}_2,\bar{k}^{(2)}_2,\ldots,\bar{k}^{(1)}_p,\bar{k}^{(2)}_p\right]^T.\end{aligned} \tag{53}$$

These relations are true for CCA, MLR, as well as PLS. Recall that all $\rho_{\times,r}$, $\tilde{\rho}_{\times,r}$, and $\bar{\rho}_{\times,r}$ presented in this paper are increasing, Schur-convex functions of $\{k_i\}_{i=1}^r$, $\{\tilde{k}_i\}_{i=1}^r$, and $\{\bar{k}_i\}_{i=1}^r$, respectively. An immediate consequence of (52) and (53) is that rotational and reflectional versions of ρ_{C_1}, ρ_{C_2}, ρ_{C_3}, and ρ_{M} are both smaller than their respective total version. For ρ_{W} and ρ_{P}, this ordering is not generally true. This means their rotational or reflective version can either be greater or less than their respective total version. The reason is that rotational and reflective coefficients on the one hand and the total coefficient on the other hand are normalized by different denominators, as can be observed from (46)–(48) and (49)–(51). As an example, consider a pair of cross-proper, but not individually proper $\mathbf{x}$ and $\mathbf{y}$, i.e., $\tilde{\mathbf{R}}_{xy} = \mathbf{0}$, but $\tilde{\mathbf{R}}_{xx} \neq \mathbf{0}$ and $\tilde{\mathbf{R}}_{yy} \neq \mathbf{0}$. Then

$$\sum_{i=1}^{r}\left(\bar{k}^{(1)}_{\mathrm{P},i}\right)^2 + \left(\bar{k}^{(2)}_{\mathrm{P},i}\right)^2 = 2\sum_{i=1}^{r} k^2_{\mathrm{P},i} \tag{54}$$

but

$$\mathrm{tr}\left(\overline{\mathbf{R}}_{xx}\overline{\mathbf{R}}^H_{xx}\right)\mathrm{tr}\left(\overline{\mathbf{R}}_{yy}\overline{\mathbf{R}}^H_{yy}\right) > 4\mathrm{tr}\left(\mathbf{R}_{xx}\mathbf{R}^H_{xx}\right)\mathrm{tr}\left(\mathbf{R}_{yy}\mathbf{R}^H_{yy}\right) \tag{55}$$

thus, $\bar{\rho}_{\mathrm{P},r} < \rho_{\mathrm{P},r}$.

The *rotational correlation spread* σ^2 and the *reflectional correlation spread* $\tilde{\sigma}^2$ are computed as in (37) using $\{k_i\}_{i=1}^p$ or $\{\tilde{k}_i\}_{i=1}^p$, respectively. Analogously, the modified correlation spreads σ^2_f and $\tilde{\sigma}^2_f$ are computed as in (38). The modified *total correlation spread* is defined as

$$\bar{\sigma}^2_f = \frac{p}{p-1}\left(\frac{\sum_{i=1}^{p}\left[f\left(\left(\bar{k}^{(1)}_i\right)^2\right)+f\left(\left(\bar{k}^{(2)}_i\right)^2\right)\right]^2}{\left[\sum_{i=1}^{p} f\left(\left(\bar{k}^{(1)}_i\right)^2\right)+f\left(\left(\bar{k}^{(2)}_i\right)^2\right)\right]^2} - \frac{1}{p}\right) \tag{56}$$

and $\bar{\sigma}^2$ is obtained with $f(x) = x$. Note that if $\bar{k}^{(1)}_i = \bar{k}^{(2)}_i = k_i$ for all i, then $\sigma^2_f = \bar{\sigma}^2_f$. Correspondingly, if $\bar{k}^{(1)}_i = \bar{k}^{(2)}_i = \tilde{k}_i$ for all i, then $\tilde{\sigma}^2_f = \bar{\sigma}^2_f$. We emphasize that there is no general ordering of σ^2_f, $\tilde{\sigma}^2_f$, and $\bar{\sigma}^2_f$, not even for $f(x) = x$. In particular, we may have $\sigma^2_f > \bar{\sigma}^2_f$ and/or $\tilde{\sigma}^2_f > \bar{\sigma}^2_f$, as can be seen from the numerical example that follows.

V. A Numerical Example

In order to illustrate the usage of the concepts discussed in this paper, we shall now work out a numerical example. Con-

TABLE II
CORRELATION AND CROSS-CORRELATION MATRICES FOR THE NUMERICAL EXAMPLE

$$\mathbf{R}_{xx} = \begin{bmatrix} 19 & 4j & -7+j \\ -4j & 14 & 4j \\ -7-j & -4j & 16 \end{bmatrix}$$

$$\widetilde{\mathbf{R}}_{xx} = \begin{bmatrix} 5+6j & 2-4j & -7-5j \\ 2-4j & -6-4j & -4+4j \\ -7-5j & -4+4j & 6+6j \end{bmatrix}$$

$$\mathbf{R}_{yy} = \begin{bmatrix} 72 & -11+13j & -18+45j \\ -11-13j & 43 & 11+18j \\ -18-45j & 11-18j & 84 \end{bmatrix}$$

$$\widetilde{\mathbf{R}}_{yy} = \begin{bmatrix} 38+16j & -31-3j & -6-13j \\ -31-3j & 3+16j & 15+36j \\ -6-13j & 15+36j & 26-4j \end{bmatrix}$$

$$\mathbf{R}_{xy} = \begin{bmatrix} -30 & 2-6j & 2-26j \\ -4+j & 8 & -12+8j \\ 18+5j & -11-j & -7+15j \end{bmatrix}$$

$$\widetilde{\mathbf{R}}_{xy} = \begin{bmatrix} -12-10j & 18-4j & -10-2j \\ j & -6j & 4+12j \\ 22+7j & -5-3j & -1-17j \end{bmatrix}$$

TABLE III
FULL-RANK CORRELATION COEFFICIENTS

	rotational ($\rho_{\times,p}$)	reflectional ($\tilde{\rho}_{\times,p}$)	total ($\bar{\rho}_{\times,p}$)
$\rho_{\mathrm{C}_1,p}$	0.498	0.515	0.699
$\rho_{\mathrm{C}_2,p}$	0.950	0.935	0.998
$\rho_{\mathrm{C}_3,p}$	0.760	0.680	0.964
$\rho_{\mathrm{M},p}$	0.602	0.556	0.812
$\rho_{\mathrm{P},p}$	0.597	0.392	0.695

sider a pair of complex random vectors $\mathbf{x}$ and $\mathbf{y}$, $m = n = p = 3$, with correlation and cross-correlation matrices given in Table II. Some full-rank correlation coefficients for this example, rounded to three decimals, are listed in Table III. The first observation we can make is that different correlation coefficients can paint a very different picture of the multivariate association between $\mathbf{x}$ and $\mathbf{y}$. While $\bar{\rho}_{\mathrm{C}_2,p} = 0.998$ indicates an almost perfect match, $\tilde{\rho}_{\mathrm{P},p} = 0.392$ gives a much lower level of association. In almost all cases, the total correlation coefficient $\bar{\rho}_{\times,p}$ is significantly larger than either the rotational or reflectional correlation coefficients $\rho_{\times,p}$ and $\tilde{\rho}_{\times,p}$. We also note that the rotational correlation coefficient can be smaller than the reflectional coefficient $(\rho_{\mathrm{C}_1,p} < \tilde{\rho}_{\mathrm{C}_1,p})$ and vice versa (all other cases). Moreover, this relationship can change with rank reduction, as seen from Table V. While $\rho_{\mathrm{C}_1,p} < \tilde{\rho}_{\mathrm{C}_1,p}$, we have $\rho_{\mathrm{C}_1,1} > \tilde{\rho}_{\mathrm{C}_1,1}$.

Now consider the correlation spreads listed in Table IV (rounded to three decimals), starting with CCA. Since $\sigma^2_{\mathrm{C}} = 0.172$, $\tilde{\sigma}^2_{\mathrm{C}} = 0.115$, and $\bar{\sigma}^2_{\mathrm{C}} = 0.092$ are all small, the reduced-rank coefficients $\rho_{\mathrm{C}_1,r}$, $\tilde{\rho}_{\mathrm{C}_1,r}$, and $\bar{\rho}_{\mathrm{C}_1,r}$ are all significantly smaller than the corresponding full-rank coefficients $\rho_{\mathrm{C}_1,p}$, $\tilde{\rho}_{\mathrm{C}_1,p}$, and $\bar{\rho}_{\mathrm{C}_1,p}$, as is evident from Table V. The biggest drop occurs from $r = 2$ to $r = 1$.

TABLE IV
CORRELATION SPREADS

	rotational ($\sigma^2_{\times}$)	reflectional ($\tilde{\sigma}^2_{\times}$)	total ($\bar{\sigma}^2_{\times}$)
σ^2_{C}	0.172	0.115	0.092
$\sigma^2_{\mathrm{C},f}$	0.408	0.266	0.275
σ^2_{M}	0.418	0.297	0.348
σ^2_{P}	0.696	0.403	0.820

TABLE V
SOME REDUCED-RANK CORRELATION COEFFICIENTS

r	$\rho_{\mathrm{C}_1,r}$	$\tilde{\rho}_{\mathrm{C}_1,r}$	$\bar{\rho}_{\mathrm{C}_1,r}$
$p=3$	0.498	0.515	0.699
2	0.431	0.435	0.604
1	0.299	0.280	0.327
r	$\rho_{\mathrm{C}_2,r}$	$\tilde{\rho}_{\mathrm{C}_2,r}$	$\bar{\rho}_{\mathrm{C}_2,r}$
$p=3$	0.950	0.937	0.998
2	0.937	0.914	0.997
1	0.896	0.839	0.984
r	$\rho_{\mathrm{P},r}$	$\tilde{\rho}_{\mathrm{P},r}$	$\bar{\rho}_{\mathrm{P},r}$
$p=3$	0.597	0.392	0.695
2	0.587	0.376	0.691
1	0.529	0.291	0.650

The PLS correlation spreads $\sigma^2_{\mathrm{P}} = 0.696$, $\tilde{\sigma}^2_{\mathrm{P}} = 0.403$, and $\bar{\sigma}^2_{\mathrm{P}} = 0.820$ are larger than the corresponding CCA correlation spreads. This indicates that, in contrast to CCA, much of the PLS correlation is contained in a low-dimensional subspace. This makes rank reduction more effective with PLS than CCA. For instance, while $\bar{\rho}_{\mathrm{C}_1,1}/\bar{\rho}_{\mathrm{C}_1,p} = 0.468$, we have $\bar{\rho}_{\mathrm{P},1}/\bar{\rho}_{\mathrm{P},p} = 0.935$. Note that the relationship between rotational/reflectional spread and total spread can be different for different correlation analysis techniques. Indeed, in CCA, $\sigma^2_{\mathrm{C}} > \bar{\sigma}^2_{\mathrm{C}}$ and $\tilde{\sigma}^2_{\mathrm{C}} > \bar{\sigma}^2_{\mathrm{C}}$, but in PLS, $\sigma^2_{\mathrm{P}} < \bar{\sigma}^2_{\mathrm{P}}$ and $\tilde{\sigma}^2_{\mathrm{P}} < \bar{\sigma}^2_{\mathrm{P}}$.

Table IV also lists $\sigma^2_{\mathrm{C},f}$ for $f(k^2_{\mathrm{C},i}) = \log(1 - k^2_{\mathrm{C},i})$. This modified spread function characterizes the distribution of information rate over individual components. Since $\sigma^2_{\mathrm{C},f} > \bar{\sigma}^2_{\mathrm{C},f}$, the ratio of reduced-rank to full-rank information rate is greater in the rotational than the total correlation case. For rotational correlations, $\log(1-\rho_{\mathrm{C}_2,1})/\log(1-\rho_{\mathrm{C}_2,p}) = 0.756$, cf. (45), whereas for total correlations, $\log(1-\bar{\rho}_{\mathrm{C}_2,1})/\log(1-\bar{\rho}_{\mathrm{C}_2,p}) = 0.656$.

VI. CONCLUSIONS

There are two fundamental choices that must be made when analyzing multivariate association between two complex random vectors $\mathbf{x}$ and $\mathbf{y}$: the desired invariances of the analysis (linear-in-$\mathbf{x}$/linear-in-$\mathbf{y}$ for CCA, unitary-in-$\mathbf{x}$/linear-in-$\mathbf{y}$ for MLR, or unitary-in-$\mathbf{x}$/unitary-in-$\mathbf{y}$ for PLS), and the type of correlation (rotational, reflectional, or total). In many cases, these choices will be dictated by the problem at hand. For instance, if we are interested in widely linear MMSE estimation, the total correlation coefficient $\bar{\rho}_{\mathrm{M}}$, based on half-canonical correlations, will have to be selected. As another example, consider the transmission of information over a noncoherent channel. Here, the rotational correlation coefficient ρ_{C_2}, based on full-canonical correlations, is the right choice since it characterizes mutual information in the rotationally invariant case.

Once the type of correlation analysis has been chosen, any correlation coefficient that is an increasing, Schur-convex function of the diagonal correlations $\{k_i\}$ will be maximized, for arbitrary rank r. In this context, we have defined the correlation spread, inspired by the degree of polarization, as a single, normalized measure for how much of the overall correlation is contained in a lower-dimensional subspace. The correlation spread gives a good indication of how effective rank reduction is going to be for the desired invariances and type of correlation.

APPENDIX

A. Background on Majorization

A vector $\boldsymbol{\lambda} \in \mathbb{R}^p$ is said to be *majorized* by a vector $\boldsymbol{\mu} \in \mathbb{R}^p$, written as $\boldsymbol{\lambda} \prec \boldsymbol{\mu}$, if

$$\sum_{i=1}^{r} \lambda_{[i]} \le \sum_{i=1}^{r} \mu_{[i]}, \quad r = 1, \ldots, p-1 \tag{57}$$

$$\sum_{i=1}^{p} \lambda_{[i]} = \sum_{i=1}^{p} \mu_{[i]} \tag{58}$$

where $[\cdot]$ is a permutation such that $\lambda_{[1]} \ge \cdots \ge \lambda_{[p]}$. Intuitively, if $\boldsymbol{\lambda} \prec \boldsymbol{\mu}$, then the components of $\boldsymbol{\lambda}$ are "less spread out" or "more equal" than the components of $\boldsymbol{\mu}$. An excellent overview of majorization results is given in [9].

The idea of majorization becomes most powerful when it is combined with the concept of Schur-convexity. Functions that are Schur-convex preserve the preordering of majorization. A real-valued function f defined on a set $D \subset \mathbb{R}^p$ is said to be *Schur-convex on* D if $\boldsymbol{\lambda} \prec \boldsymbol{\mu}$ on D implies that $f(\boldsymbol{\lambda}) \le f(\boldsymbol{\mu})$. There are a number of tests for Schur-convexity. Particularly useful is the following result [9, 3.A.4]. Let $I \in \mathbb{R}$ be an open interval and $f : I^p \to \mathbb{R}$ be continuously differentiable. Necessary and sufficient conditions for f to be Schur-convex on I^p are: f is symmetric in its arguments on I^p, and for all $\boldsymbol{\lambda} \in I^p$,

$$(\lambda_1 - \lambda_2)\left[\frac{\partial}{\partial \lambda_1} f(\boldsymbol{\lambda}) - \frac{\partial}{\partial \lambda_2} f(\boldsymbol{\lambda})\right] \ge 0. \tag{59}$$

A classic result of majorization is that if $\mathbf{H}$ is a $p \times p$ Hermitian matrix with diagonal elements $\mathbf{diag}(\mathbf{H}) = (H_{11}, \ldots, H_{pp})^T$ and eigenvalues $\mathbf{ev}(\mathbf{H}) = (\lambda_1, \ldots, \lambda_p)^T$, then [9, 9.B.1] $\mathbf{diag}(\mathbf{H}) \prec \mathbf{ev}(\mathbf{H})$. A generalization of this result to $m \times n$ complex matrices is possible if we introduce weak majorization.

A vector $\boldsymbol{\lambda} \in \mathbb{R}^p$ is said to be *weakly majorized* by a vector $\boldsymbol{\mu} \in \mathbb{R}^p$, written as $\boldsymbol{\lambda} \prec_w \boldsymbol{\mu}$, if

$$\sum_{i=1}^{r} \lambda_{[i]} \le \sum_{i=1}^{r} \mu_{[i]}, \quad r = 1, \ldots, p. \tag{60}$$

In (60) and in contrast to regular majorization, equality is not required for $r = p$. If $\mathbf{C} \in \mathbb{C}^{m \times n}$ has diagonal elements $\mathbf{diag}(\mathbf{C}) = (C_{11}, \ldots, C_{pp})^T$, $p = \min(m, n)$, and singular values $\mathbf{sv}(\mathbf{C}) = (k_1, \ldots, k_p)^T$, then the absolute values of the diagonal elements are weakly majorized by the singular values, $|\mathbf{diag}(\mathbf{C})| \prec_w \mathbf{sv}(\mathbf{C})$. This result has been proved in [28] for square complex matrices. It extends trivially to nonsquare matrices, as these can be augmented with zeros to make them square.

In order to apply the concept of Schur-convexity to weakly majorized vectors, we need to introduce increasing functions. A function f defined on a set $D \subset \mathbb{R}^p$ is called *increasing* if $\lambda_i \le \mu_i$ for $i = 1, \ldots, p$ implies that $f(\boldsymbol{\lambda}) \le f(\boldsymbol{\mu})$. If f is Schur-convex and increasing on D, then [9, 3.A.8] $\boldsymbol{\lambda} \prec_w \boldsymbol{\mu}$ on D implies that $f(\boldsymbol{\lambda}) \le f(\boldsymbol{\mu})$.

B. Factorization of an Augmented Correlation Matrix

In factorizations of an augmented correlation matrix, all factors need to satisfy the pattern of (40). This can be achieved as follows [20]. Let

$$\mathbf{T} = \frac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{bmatrix} \tag{61}$$

$\check{\mathbf{x}} = [\sqrt{2}\mathrm{Re}\mathbf{x}^T \ \sqrt{2}\mathrm{Im}\mathbf{x}^T]^T$, and $\check{\mathbf{R}}_{xx} = E\check{\mathbf{x}}\check{\mathbf{x}}^T$, so that $\overline{\mathbf{x}} = \mathbf{T}\check{\mathbf{x}}$, $\overline{\mathbf{R}}_{xx} = \mathbf{T}\check{\mathbf{R}}_{xx}\mathbf{T}^H$, and $\check{\mathbf{R}}_{xx} = \mathbf{T}^H\overline{\mathbf{R}}_{xx}\mathbf{T}$.

First consider the factorization $\check{\mathbf{R}}_{xx}^{-1} = \check{\mathbf{R}}_{xx}^{-1/2}\check{\mathbf{R}}_{xx}^{-T/2}$. We obtain the corresponding factorization of the augmented correlation matrix $\overline{\mathbf{R}}_{xx}$ as

$$\begin{aligned} \overline{\mathbf{R}}_{xx}^{-1} &= \left(\mathbf{T}\check{\mathbf{R}}_{xx}^{-1/2}\mathbf{T}^H\right)\left(\mathbf{T}\check{\mathbf{R}}_{xx}^{-T/2}\mathbf{T}^H\right) \\ &= \overline{\mathbf{R}}_{xx}^{-1/2}\overline{\mathbf{R}}_{xx}^{-H/2}. \end{aligned} \tag{62}$$

The SVD of an augmented cross-correlation matrix $\overline{\mathbf{C}}$ is obtained via the usual SVD of the real-valued matrix $\check{\mathbf{C}} = \mathbf{T}^H\overline{\mathbf{C}}\mathbf{T} = \check{\mathbf{F}}\check{\mathbf{K}}\check{\mathbf{G}}^T$. We assume that the singular values on the diagonal of

$$\check{\mathbf{K}} = \begin{bmatrix} \overline{\mathbf{K}}^{(1)} & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{K}}^{(2)} \end{bmatrix} \tag{63}$$

$\overline{\mathbf{K}}^{(1)} = \mathrm{diag}(\bar{k}_1^{(1)}, \ldots, \bar{k}_p^{(1)})$, $\overline{\mathbf{K}}^{(2)} = \mathrm{diag}(\bar{k}_1^{(2)}, \ldots, \bar{k}_p^{(2)})$, are arranged such that $\bar{k}_1^{(1)} \ge \bar{k}_1^{(2)} \ge \bar{k}_2^{(1)} \ge \bar{k}_2^{(2)} \ge \cdots \ge \bar{k}_p^{(1)} \ge \bar{k}_p^{(2)} \ge 0$. Then the SVD of $\overline{\mathbf{C}}$ is $\overline{\mathbf{C}} = \overline{\mathbf{F}}\overline{\mathbf{K}}\overline{\mathbf{G}}^H$ with $\overline{\mathbf{F}} = \mathbf{T}\check{\mathbf{F}}\mathbf{T}^H$, $\overline{\mathbf{G}} = \mathbf{T}\check{\mathbf{G}}\mathbf{T}^H$, and

$$\overline{\mathbf{K}} = \mathbf{T}\check{\mathbf{K}}\mathbf{T}^H = \frac{1}{2}\begin{bmatrix} \overline{\mathbf{K}}^{(1)} + \overline{\mathbf{K}}^{(2)} & \overline{\mathbf{K}}^{(1)} - \overline{\mathbf{K}}^{(2)} \\ \overline{\mathbf{K}}^{(1)} - \overline{\mathbf{K}}^{(2)} & \overline{\mathbf{K}}^{(1)} + \overline{\mathbf{K}}^{(2)} \end{bmatrix}. \tag{64}$$

Hence, $\overline{\mathbf{K}}$ is diagonal if and only if the singular values of $\overline{\mathbf{C}}$ have even multiplicity.

References

[1] A. Renyi, "On measures of dependence," *Acta Mathematica Acad. Sci. Hungary*, vol. 10, no. 3–4, pp. 441–451, Sep. 1959.

[2] J. O. Ramsay, J. ten Berge, and G. P. H. Styan, "Matrix correlation," *Psychometrika*, vol. 49, no. 3, pp. 402–423, Sep. 1984.

[3] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3–4, pp. 321–377, Dec. 1936.

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

[5] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991.

[6] H. Wold, "Path models with latent variables: The NIPALS approach," in *Quantitative Sociology: International Perspectives on Mathematical and Statistical Modeling*, H. M. Blalock, Ed. *et al.* New York: Academic, 1975, pp. 307–357.

[7] H. Wold, "Partial least squares," in *Encyclopedia of the Statistical Sciences*, S. Kotz and N. L. Johnson, Eds. New York: Wiley, 1985, pp. 581–591.

[8] P. D. Sampson, A. P. Streissguth, H. M. Barr, and F. L. Bookstein, "Neurobehavioral effects of prenatal alcohol: Part II. Partial least squares analysis," *Neurotoxicol. Teratol.*, vol. 11, no. 5, pp. 477–491, Sep.–Oct. 1989.

[9] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1979.

[10] Y. Hua, M. Nikpour, and P. Stoica, "Optimal reduced-rank estimation and filtering," *IEEE Trans. Signal Process.*, vol. 49, no. 3, pp. 457–469, Mar. 2001.

[11] B. Hanson, K. Klink, K. Matsuura, S. M. Robeson, and C. J. Willmott, "Vector correlation: Review, exposition, and geographic application," *Ann. Assoc. Amer. Geographers*, vol. 82, no. 1, pp. 103–116, Mar. 1992.

[12] P. E. Jupp and K. V. Mardia, "A general correlation coefficient for directional data and related regression problems," *Biometrika*, vol. 67, no. 1, pp. 163–173, Apr. 1980.

[13] J. C. Samson, "Descriptions of the polarization states of vector processes: Application to ULF magnetic fields," *Geophys. J. R. Astron. Soc.*, vol. 34, no. 4, pp. 403–419, Dec. 1973.

[14] P. Vahimaa and J. Tervo, "Unified measures for optical fields: degree of polarization and effective degree of coherence," *J. Opt. A: Pure Appl. Opt.*, vol. 6, no. 3, pp. S41–S44, 2004.

[15] H. Yanai, "Unification of various techniques of multivariate analysis by means of generalized coefficient of determination (G.C.D.)," *J. Behaviormetrics*, vol. 1, no. 1, pp. 45–54, 1974.

[16] W. W. Rozeboom, "Linear correlation between sets of variables," *Psychometrika*, vol. 30, no. 1, pp. 57–71, Mar. 1965.

[17] P. Coxhead, "Measuring the relationship between two sets of variables," *Brit. J. Math. Stat. Psych.*, vol. 27, pp. 205–212, 1974.

[18] E. M. Cramer and W. A. Nicewander, "Some symmetric, invariant measures of multivariate association," *Psychometrika*, vol. 44, no. 1, pp. 43–54, Mar. 1979.

[19] L. L. Scharf and C. T. Mullis, "Canonical coordinates and the geometry of inference, rate, and capacity," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 824–831, Mar. 2000.

[20] P. J. Schreier and L. L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 714–725, Mar. 2003.

[21] D. Stewart and W. Love, "A general canonical correlation index," *Psych. Bull.*, vol. 70, no. 3, pt. 1, pp. 160–163, Sep. 1968.

[22] T. C. Gleason, "On redundancy in canonical analysis," *Psych. Bull.*, vol. 83, no. 6, pp. 1004–1006, Nov. 1976.

[23] P. J. Schreier and L. L. Scharf, "Canonical coordinates for transform coding of noisy sources," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 235–243, Jan. 2006.

[24] P. Robert and Y. Escoufier, "A unifying tool for linear multivariate statistical methods: The RV-coefficient," *Appl. Stat.*, vol. 25, no. 3, pp. 257–265, 1976.

[25] A. Tishler and S. Lipovetsky, "Modelling and forecasting with robust canonical analysis: Method and application," *Comput. Oper. Res.*, vol. 27, no. 3, pp. 217–232, Mar. 2000.

[26] B. Picinbono, "On circularity," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3473–3482, Dec. 1994.

[27] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 2030–2033, Aug. 1995.

[28] R. C. Thompson, "Singular values, diagonal elements, and convexity," *SIAM J. Appl. Math.*, vol. 32, no. 1, pp. 39–63, Jan. 1977.

Peter J. Schreier (S'03–M'04) was born in Munich, Germany, in 1975. He received the M.S. degree from the University of Notre Dame, IN, in 1999, and the Ph.D. degree from the University of Colorado at Boulder in 2003, both in electrical engineering.

During the fall semester of 1998, he was a visiting research student with the University of Hawaii. From April 1999 to July 2000, he was with the Research Center 3D Image Analysis and Synthesis at Friedrich-Alexander Universität Erlangen-Nürnberg, Germany. In the spring semester of 2004, he was a Postdoctoral Research Associate with Colorado State University, Ft. Collins. Currently, he is a Senior Lecturer with the School of Electrical Engineering and Computer Science, The University of Newcastle, Australia. His research interests lie in the areas of statistical signal processing and communications.