Correspondence

Canonical Coordinates are the Right Coordinates for Low-Rank Gauss–Gauss Detection and Estimation

Ali Pezeshki, *Member, IEEE*, Louis L. Scharf, *Fellow, IEEE*, John K. Thomas, *Member, IEEE*, and Barry D. Van Veen, *Fellow, IEEE*

***Abstract*—In this correspondence, our aim is to establish a connection between low-rank detection, low-rank estimation, and canonical coordinates. The key to this connection is the observation that Gauss-Gauss detectors and estimators share canonical coordinates in the case where the underlying model is a signal-plus-noise model. We show that in Gauss–Gauss detection J-divergence is a function of squared canonical correlations, and hence is invariant to nonsingular transformations of the data channels. Further, we show that J-divergence has a special decomposition in canonical coordinates, impelling their use for rank-reduction. Canonical coordinates have been found earlier to be fundamental for low-rank estimation, as they decompose three important performance measures, namely the relative volume of error concentration ellipse, processing gain, and information rate. This correspondence shows that canonical coordinates are also fundamental for low-rank detection, making them more useful in signal processing and communication problems when low-rank modeling is required to achieve computational efficiency or robustness against noise and model uncertainties.**

***Index Terms*—Canonical coordinates, concentration ellipse, detection, estimation, Gaussian random vectors, information rate, J-divergence, mutual information.**

I. Introduction

Reduced-rank detection and estimation find applications in signal processing and communications, where low-rank modeling is desired to achieve computational efficiency or robustness against noise and model uncertainties. Some of the fundamental results in this area are the low-rank detectors of [1] and [2] for Gaussian random vectors, the low-rank detector of [3] for multichannel Gaussian signals, the reduced-rank detector of [4] for direct-sequence code-division multiple-access (DS-CDMA) signals, the reduced-rank Wiener filters of [5]–[11], the reduced-rank maximum-likelihood estimator of [12], and the reduced-rank quantizer of [13].

The choice of a coordinate system for rank-reduction depends on the performance measure to be optimized under the corresponding rank constraint. In [5], [6], and [9]–[11], it has been shown that *canonical coordinates* are optimal for building reduced-rank Wiener filters when the objective is to maximize the rate at which the reduced-rank estimator carries information about the unknown signal of interest.[1] Our aim in this correspondence is to show that canonical coordinates are also optimal for building the low-rank Gauss–Gauss detectors of [1], in the case where the underlying model is a signal-plus-noise model. Our results establish a connection between low-rank detection and low-rank estimation by clarifying that canonical coordinates are the right coordinates for rank reduction in both problems.

As will be seen, four important performance measures in estimation and detection, namely the relative volume of error concentration ellipse, processing gain, information rate, and J-divergence are functions of *squared canonical correlations*, and hence are invariant to nonsingular transformations of the measurement channel and the signal channel. Moreover, these performance measures have special decompositions in canonical coordinates. It turns out that under a rank constraint, the dominant canonical coordinates, i.e., those associated with larger canonical correlations, need to be retained for building detectors and estimators. The *dominant* canonical coordinates are the canonical coordinates with the largest *per-mode signal-to-noise ratios* (SNRs). Therefore, in a signal-plus-noise model, the good coordinates for building low-rank detectors and estimators are the canonical coordinates that carry the signal with large per-mode SNRs. The decompositions and invariances of the performance measures for low-rank estimation (i.e., the relative volume of error concentration ellipse, processing gain, and information rate) in canonical coordinates have been reported earlier in [5], [6], and [9]–[11]. What is original in this correspondence is the decomposition of J-divergence in canonical coordinates, the study of its invariances, and the demonstration that canonical coordinates are also fundamental for constructing the low-rank detectors reported in [1].

II. Problem Setup

Consider the signal-plus-noise model $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where $\mathbf{x} : CN_n[\mathbf{0}, \mathbf{R}_{xx}]$ is the n-dimensional message vector or signal of interest, $\mathbf{n} : CN_n[\mathbf{0}, \mathbf{R}_{nn}]$ is the additive noise vector, uncorrelated with $\mathbf{x}$, and $\mathbf{y} : CN_n[\mathbf{0}, \mathbf{R}_{yy} = \mathbf{R}_{xx} + \mathbf{R}_{nn}]$ is the measurement vector. The notation $CN_n[\mathbf{0}, \mathbf{R}_{xx}]$ denotes the n-variate proper complex normal distribution with mean vector zero and covariance matrix $\mathbf{R}_{xx}$. Throughout, we shall assume that all complex random vectors are proper, in which case $E\mathbf{x}\mathbf{x}^{\mathrm{H}} = \mathbf{R}_{xx}$ and $E\mathbf{x}\mathbf{x}^{\mathrm{T}} = \mathbf{0}$. The estimation problem considered in this correspondence is to estimate $\mathbf{x} : CN_n[\mathbf{0}, \mathbf{R}_{xx}]$, from $\mathbf{y} : CN_n[\mathbf{0}, \mathbf{R}_{yy}]$, and the detection problem is to decide between the hypothesis $H_0 : \mathbf{y} = \mathbf{n}$ and the alternative $H_1 : \mathbf{y} = \mathbf{x} + \mathbf{n}$.

Under H_1, the composite covariance matrix for $\mathbf{z}^{\mathrm{H}} = [\mathbf{x}^{\mathrm{H}}, \mathbf{y}^{\mathrm{H}}]$ may be written as

$$\mathbf{R}_{zz} = E\mathbf{z}\mathbf{z}^{\mathrm{H}} = \begin{bmatrix} \mathbf{R}_{xx} & (\mathbf{R}_{xy} = \mathbf{R}_{xx}) \\ \mathbf{R}_{xy}^{\mathrm{H}} & (\mathbf{R}_{yy} = \mathbf{R}_{xx} + \mathbf{R}_{nn}) \end{bmatrix}. \quad (1)$$

There are several functions of this matrix that will be of interest in this correspondence:

1) SNR matrix

$$\mathbf{R} = \mathbf{R}_{nn}^{-1/2}\,\mathbf{R}_{xx}\,\mathbf{R}_{nn}^{-\mathrm{H}/2} \quad (2)$$

[1]The right coordinates for minimizing mean-squared error under a rank constraint are different. They are *half-canonical coordinates*. The reader is referred to [7]–[9].

Manuscript received May 4, 2005; revised December 8, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Leslie Collins. This work is supported by the DARPA Integrated Sensing and Processing Program under AFOSR contract FA9550-04-1-0371.

A. Pezeshki was with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523 USA. He is now with the Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544 USA (e-mail: pezeshki@princeton.edu).

L. L. Scharf is with the Department of Electrical and Computer Engineering and the Department of Statistics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: scharf@engr.colostate.edu).

J. K. Thomas is with TensorComm Incorporated, Westminster, CO 80234 USA (e-mail: johnthomas@tensorcomm.com).

B. D. Van Veen is with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706 USA (e-mail: vanveen@engr.wisc.edu).

Digital Object Identifier 10.1109/TSP.2006.881249

2) signal-plus-noise-to-noise matrix

$$\mathbf{S} = \mathbf{R}_{nn}^{-1/2}(\mathbf{R}_{xx} + \mathbf{R}_{nn})\mathbf{R}_{nn}^{-\mathrm{H}/2} \tag{3}$$

3) squared coherence matrix

$$\mathbf{C}\mathbf{C}^{\mathrm{H}} = \mathbf{R}_{xx}^{\mathrm{H}/2}(\mathbf{R}_{xx} + \mathbf{R}_{nn})^{-1}\mathbf{R}_{xx}^{1/2}. \tag{4}$$

In these equations, $\mathbf{R}_{nn}^{1/2}$ and $\mathbf{R}_{xx}^{1/2}$ are arbitrary $n \times n$ square-roots of $\mathbf{R}_{nn}$ and $\mathbf{R}_{xx}$. That is, $\mathbf{R}_{nn}^{1/2}$ (or alternatively $\mathbf{R}_{xx}^{1/2}$) satisfies

$$\mathbf{R}_{nn}^{1/2}\mathbf{R}_{nn}^{\mathrm{H}/2} = \mathbf{R}_{nn} \text{ and } \mathbf{R}_{nn}^{-1/2}\mathbf{R}_{nn}\mathbf{R}_{nn}^{-\mathrm{H}/2} = \mathbf{I}. \tag{5}$$

The reason for naming $\mathbf{R}$ the SNR matrix and $\mathbf{S}$ the signal-plus-noise-to-noise ratio matrix is evident from their formulations. The matrix $\mathbf{C}\mathbf{C}^{\mathrm{H}}$ is the square of the *coherence matrix*

$$\begin{aligned}\mathbf{C} &= E(\mathbf{R}_{xx}^{-1/2}\mathbf{x})(\mathbf{R}_{yy}^{-1/2}\mathbf{y})^{\mathrm{H}} = \mathbf{R}_{xx}^{-1/2}\mathbf{R}_{xy}\mathbf{R}_{yy}^{-\mathrm{H}/2}\\ &= \mathbf{R}_{xx}^{\mathrm{H}/2}(\mathbf{R}_{xx} + \mathbf{R}_{nn})^{-\mathrm{H}/2},\end{aligned}$$

which is the cross covariance between the whitened versions of $\mathbf{x}$ and $\mathbf{y}$ [5], [6].

Denoting the eigenvalues of $\mathbf{R}$, $\mathbf{S}$, and $\mathbf{C}\mathbf{C}^{\mathrm{H}}$ as $\gamma_i^2 = \mathrm{ev}_i(\mathbf{R})$, $l_i^2 = \mathrm{ev}_i(\mathbf{S})$, and $k_i^2 = \mathrm{ev}_i(\mathbf{C}\mathbf{C}^{\mathrm{H}})$, it is easy to show the following connections:

$$\gamma_i^2 = l_i^2 - 1 = \frac{k_i^2}{1 - k_i^2} \tag{6}$$

$$l_i^2 = 1 + \gamma_i^2 = \frac{1}{1 - k_i^2} \tag{7}$$

$$k_i^2 = \frac{\gamma_i^2}{\gamma_i^2 + 1} = \frac{l_i^2 - 1}{l_i^2}. \tag{8}$$

We shall call the k_i^2 squared canonical correlations for reasons to be made clear. We call γ_i^2 per-mode SNRs , and the l_i^2 per-mode signal-plus-noise-to-noise ratios, as they are the eigenvalues of the signal-to-noise ratio matrix $\mathbf{R}$ and the signal-plus-noise-to-noise ratio matrix $\mathbf{S}$. When $\mathbf{R}_{xx}$, $\mathbf{R}_{nn}$, and subsequently $\mathbf{R}_{yy}$ are circulant, γ_i^2 and l_i^2 are the SNR and signal-plus-noise-to-noise ratio in the ith discrete Fourier transform (DFT) mode. Throughout, we shall assume that the squared canonical correlations k_i^2 are arranged in descending order, i.e., $k_1^2 \geq \cdots \geq k_n^2 > 0$. Consequently, the per-mode SNRs γ_i^2 and the per-mode signal-plus-noise-to-noise ratios l_i^2 are also descending, by virtue of their dependence on the k_i^2. For reasons to become apparent, we shall take the k_i^2 to be fundamental, and the γ_i^2 and the l_i^2 to be derivative, although this convention is somewhat arbitrary.

III. Canonical Coordinates

The composite covariance matrix $\mathbf{R}_{zz}$ may be taken to block-tridiagonal form as follows [5], [6]:

$$\begin{bmatrix}\mathbf{F}^{\mathrm{H}} & \mathbf{0}\\ \mathbf{0} & \mathbf{G}^{\mathrm{H}}\end{bmatrix}\begin{bmatrix}\mathbf{R}_{xx}^{-1/2} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}_{yy}^{-1/2}\end{bmatrix}\mathbf{R}_{zz}\begin{bmatrix}\mathbf{R}_{xx}^{-\mathrm{H}/2} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}_{yy}^{-\mathrm{H}/2}\end{bmatrix}\begin{bmatrix}\mathbf{F} & \mathbf{0}\\ \mathbf{0} & \mathbf{G}\end{bmatrix} = \begin{bmatrix}\mathbf{I} & \mathbf{K}\\ \mathbf{K} & \mathbf{I}\end{bmatrix}. \tag{9}$$

The trick is to choose $\mathbf{F}$, $\mathbf{K}$, and $\mathbf{G}$ to be the singular value decomposition (SVD) of the coherence matrix $\mathbf{C} = \mathbf{R}_{xx}^{\mathrm{H}/2}(\mathbf{R}_{xx} + \mathbf{R}_{nn})^{-\mathrm{H}/2}$, as follows:

$$\begin{gathered}\mathbf{F}\mathbf{K}\mathbf{G}^{\mathrm{H}} = \mathbf{C} \text{ and } \mathbf{K} = \mathbf{F}^{\mathrm{H}}\mathbf{C}\mathbf{G}\\ \mathbf{F}\mathbf{F}^{\mathrm{H}} = \mathbf{F}^{\mathrm{H}}\mathbf{F} = \mathbf{I} \text{ and } \mathbf{G}\mathbf{G}^{\mathrm{H}} = \mathbf{G}^{\mathrm{H}}\mathbf{G} = \mathbf{I}\\ \mathbf{K} = \mathrm{diag}(k_1, \ldots, k_n) \text{ and } 1 \geq k_1 \geq \cdots \geq k_n > 0.\end{gathered} \tag{10}$$

The transformation

$$\begin{bmatrix}\mathbf{u}\\ \mathbf{v}\end{bmatrix} = \begin{bmatrix}\mathbf{F}^{\mathrm{H}} & \mathbf{0}\\ \mathbf{0} & \mathbf{G}^{\mathrm{H}}\end{bmatrix}\begin{bmatrix}\mathbf{R}_{xx}^{-1/2} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}_{yy}^{-1/2}\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} \tag{11}$$

resolves $\mathbf{z}^{\mathrm{H}}[\mathbf{x}^{\mathrm{H}}, \mathbf{y}^{\mathrm{H}}]$ into their canonical coordinates $\mathbf{w}^{\mathrm{H}} = [\mathbf{u}^{\mathrm{H}}, \mathbf{v}^{\mathrm{H}}]$, with the composite covariance matrix

$$\mathbf{R}_{ww} = E\mathbf{w}\mathbf{w}^{\mathrm{H}} = \begin{bmatrix}(\mathbf{R}_{uu} = \mathbf{I}) & (\mathbf{R}_{uv} = \mathbf{K})\\ (\mathbf{R}_{vu} = \mathbf{K}) & (\mathbf{R}_{vv} = \mathbf{I})\end{bmatrix}. \tag{12}$$

The diagonal matrix $\mathbf{K} = E[\mathbf{u}\mathbf{v}^{\mathrm{H}}] = \mathbf{F}^{\mathrm{H}}\mathbf{C}\mathbf{G}$ is the *canonical correlation matrix* of *canonical correlations* k_i. Correspondingly, $\mathbf{K}^2 = \mathrm{diag}(k_1^2, \ldots, k_n^2)$ is the *squared canonical correlation matrix* of *squared canonical correlations* k_i^2. It is obvious that $\mathbf{C}\mathbf{C}^{\mathrm{H}} = \mathbf{F}\mathbf{K}^2\mathbf{F}^{\mathrm{H}}$ is similar to $\mathbf{K}^2$. The squared canonical correlations k_i^2 are a complete set of invariants to the transformation group $T = \{\mathbf{T} : \det(\mathbf{T}) \neq 0\}$, with group action $\mathbf{T}\mathbf{R}_{zz}\mathbf{T}^{\mathrm{H}}$ and $\mathbf{T} = \mathrm{diag}(\mathbf{T}_1, \mathbf{T}_2)$ block diagonal. Thus, any function of the composite covariance matrix $\mathbf{R}_{zz}$ that is invariant to $\mathbf{T} \in T$ is a function of the k_i^2. We shall find that the relative volume of error concentration ellipse, processing gain, information rate, and J-divergence are four such functions. It is this finding that ties detection to estimation and clarifies the role of canonical coordinates in signal-plus-noise problems.

IV. Estimation in Canonical Coordinates

In [5] and [6], the minimum mean-squared error estimator of $\mathbf{x}$, from $\mathbf{y}$, is written in canonical coordinates as

$$\hat{\mathbf{x}} = \mathbf{R}_{xx}^{1/2}\mathbf{F}\mathbf{K}\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}, \tag{13}$$

with minimum error covariance

$$\mathbf{Q}_{xx} = \mathbf{R}_{xx}^{1/2}\mathbf{F}(\mathbf{I} - \mathbf{K}^2)\mathbf{F}^{\mathrm{H}}\mathbf{R}_{xx}^{\mathrm{H}/2}. \tag{14}$$

As shown in [5] and [6], the volume of the error concentration ellipse, divided by the volume of the prior concentration ellipse (the relative volume) is

$$\begin{aligned}V &= \frac{\det(\mathbf{Q}_{xx})}{\det(\mathbf{R}_{xx})} = \det(\mathbf{I} - \mathbf{K}^2)\\ &= \prod_{i=1}^{n}(1 - k_i^2) = \prod_{i=1}^{n}\left(\frac{1}{1 + \gamma_i^2}\right) = \prod_{i=1}^{n}\left(\frac{1}{l_i^2}\right).\end{aligned} \tag{15}$$

Correspondingly, the processing gain PG and information rate R are $PG = V^{-1}$ and $R = -(1/2)\log V$. These performance measures are invariant to T (or T-invariant), as evidenced by the fact that they are functions of the maximal invariants k_i^2. Moreover, canonical coordinates are the correct coordinates for minimizing the relative volume V, maximizing the processing gain PG, and maximizing the information rate R, under a rank constraint, as V, PG, and R are additive or multiplicative in $(1 - k_i^2)$. It is clear that the *dominant* canonical coordinates, i.e., those associated with larger canonical correlations, minimize V, maximize PG, and maximize R. The dominant canonical correlations correspond to *dominant* per-mode SNRs γ_i^2 and *dominant* per-mode signal-plus-noise-to-noise ratios l_i^2.

The optimal rank-$r \leq n$ estimator $\hat{\mathbf{x}}_r$ of $\mathbf{x}$, from $\mathbf{y}$, in canonical coordinates is

$$\hat{\mathbf{x}}_r = \mathbf{R}_{xx}^{1/2}\mathbf{F}\mathbf{K}(r)\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y} \tag{16}$$

where

$$\mathbf{K}(r) = \begin{bmatrix}\mathbf{K}_r & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}; \mathbf{K}_r = \mathrm{diag}(k_1, \ldots, k_r). \tag{17}$$

This solution is argued in [5] and [6] and proved to be optimal in [9]–[11]. That is, for building $\hat{\mathbf{x}}_r$ the r-dominant canonical coordinates, corresponding to $k_1, \ldots, k_r$, are retained, and the *subdominant* canonical coordinates, corresponding to $k_{r+1}, \ldots, k_n$, are discarded. This result holds with very slight modification for block quantizing of noisy signals, under a bit-rate constraint [13], where canonical coordinates are optimal for building reduced-rank quantizers.

The optimality of $\hat{\mathbf{x}}_r$ is in the sense that it has the smallest relative volume of the error concentration ellipse, the largest processing gain, and the highest rate of information about $\mathbf{x}$, among all rank-r estimators of $\mathbf{x}$ from $\mathbf{y}$. The relative volume of the error concentration ellipse, processing gain, and information rate associated with $\hat{\mathbf{x}}_r$ are the formulas for V, $PG = V^{-1}$, and $R = -(1/2)\log V$, with $\mathbf{K}(r)$ and r replacing $\mathbf{K}$ and n [5], [6].

V. Detection in Canonical Coordinates

In [1], the Neyman–Pearson detector for testing H_0:$\mathbf{y}$:$CN_n[\mathbf{0}, \mathbf{R}_{nn}]$ versus $H_1 : \mathbf{y} : CN_n[\mathbf{0}, \mathbf{R}_{yy} = \mathbf{R}_{xx} + \mathbf{R}_{nn}]$ is written as the log-likelihood ratio

$$l(\mathbf{y}) = (\mathbf{R}_{nn}^{-1/2}\mathbf{y})^{\mathrm{H}}(\mathbf{I} - \mathbf{S}^{-1})(\mathbf{R}_{nn}^{-1/2}\mathbf{y}). \tag{18}$$

Inserting $\mathbf{S} = \mathbf{R}_{nn}^{-1/2}(\mathbf{R}_{xx} + \mathbf{R}_{nn})\mathbf{R}_{nn}^{-\mathrm{H}/2}$, we may write $l(\mathbf{y})$ as

$$l(\mathbf{y}) = (\mathbf{R}_{xx}^{-1/2}\mathbf{y})^{\mathrm{H}}(\mathbf{R}_{xx}^{\mathrm{H}/2}\mathbf{R}_{nn}^{-1}\mathbf{R}_{xx}^{1/2} - \mathbf{R}_{xx}^{\mathrm{H}/2}(\mathbf{R}_{xx} + \mathbf{R}_{nn})^{-1}\mathbf{R}_{xx}^{1/2})(\mathbf{R}_{xx}^{-1/2}\mathbf{y}). \tag{19}$$

Using (4), it is easy to show that

$$\mathbf{R}_{xx}^{\mathrm{H}/2}\mathbf{R}_{nn}^{-1}\mathbf{R}_{xx}^{1/2} = [(\mathbf{C}\mathbf{C}^{\mathrm{H}})^{-1} - \mathbf{I}]^{-1}. \tag{20}$$

Therefore, $l(\mathbf{y})$ may be written as

$$l(\mathbf{y}) = (\mathbf{R}_{xx}^{-1/2}\mathbf{y})^{\mathrm{H}}([(\mathbf{C}\mathbf{C}^{\mathrm{H}})^{-1} - \mathbf{I}]^{-1} - \mathbf{C}\mathbf{C}^{\mathrm{H}})(\mathbf{R}_{xx}^{-1/2}\mathbf{y}). \tag{21}$$

Inserting $\mathbf{C}\mathbf{C}^{\mathrm{H}} = \mathbf{F}\mathbf{K}^2\mathbf{F}^{\mathrm{H}}$, we may rewrite the log-likelihood ratio $l(\mathbf{y})$ as

$$l(\mathbf{y}) = (\mathbf{F}^{\mathrm{H}}\mathbf{R}_{xx}^{-1/2}\mathbf{y})^{\mathrm{H}}([(\mathbf{K}^2)^{-1} - \mathbf{I}]^{-1} - \mathbf{K}^2)(\mathbf{F}^{\mathrm{H}}\mathbf{R}_{xx}^{-1/2}\mathbf{y}). \tag{22}$$

Using the SVD of the coherence matrix $\mathbf{C} = \mathbf{R}_{xx}^{\mathrm{H}/2}\mathbf{R}_{yy}^{-\mathrm{H}/2} = \mathbf{F}\mathbf{K}\mathbf{G}^{\mathrm{H}}$, it is easy to show that $\mathbf{F}^{\mathrm{H}}\mathbf{R}_{xx}^{-1/2} = \mathbf{K}^{-1}\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}$ and write the log-likelihood ratio $l(\mathbf{y})$ as the following quadratic form in $\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}$:

$$l(\mathbf{y}) = (\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y})^{\mathrm{H}}([\mathbf{I} - \mathbf{K}^2]^{-1} - \mathbf{I})(\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}). \tag{23}$$

This is the standard Gauss–Gauss log-likelihood ratio, but in the coordinates $\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}$, which under H_1 are the canonical coordinates of $\mathbf{y}$. This follows by observing that $[\mathbf{I} - \mathbf{K}^2]$ and $\mathbf{I}$ are, respectively, the covariance of $\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}$ under H_0 and H_1. That is

$$E_{H_0}\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}\mathbf{y}^{\mathrm{H}}\mathbf{R}_{yy}^{-\mathrm{H}/2}\mathbf{G} = \mathbf{I} - \mathbf{K}^2 \tag{24}$$

$$E_{H_1}\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}\mathbf{y}^{\mathrm{H}}\mathbf{R}_{yy}^{-\mathrm{H}/2}\mathbf{G} = \mathbf{I}. \tag{25}$$

See the Appendix for the derivations of (24) and (25). Thus, it is the squared canonical correlation matrix that counts. Denoting $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_n]$, the log-likelihood ratio $l(\mathbf{y})$ may be written as

$$\begin{aligned} l(\mathbf{y}) &= \sum_{i=1}^{n} \left|\mathbf{g}_i^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}\right|^2 \left(\frac{k_i^2}{1-k_i^2}\right) \\ &= \sum_{i=1}^{n} \left|\mathbf{g}_i^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}\right|^2 (\gamma_i^2) \\ &= \sum_{i=1}^{n} \left|\mathbf{g}_i^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}\right|^2 (l_i^2 - 1). \end{aligned} \tag{26}$$

It is interesting to note that under H_1 the log-likelihood ratio $l(\mathbf{y})$ is the weighted sum of the magnitude-squared of the canonical coordinates $\mathbf{g}_i^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}$, each of which is weighted by its corresponding per-mode SNR γ_i^2.

The J-divergence between H_1 and H_0 is [14]

$$J = E_{H_1}l(\mathbf{y}) - E_{H_0}l(\mathbf{y}). \tag{27}$$

We note that J-divergence is not the definitive measure of detectability, rather it is a tractable T-invariant measure for rank reduction. The detectors of [1] are low-rank versions of the log-likelihood ratio $l(\mathbf{y})$, which have maximum J-divergence under a rank constraint.

The expected value of $l(\mathbf{y})$ under H_0 is

$$\begin{aligned} E_{H_0}l(\mathbf{y}) &= E_{H_0}(\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y})^{\mathrm{H}}([\mathbf{I}-\mathbf{K}^2]^{-1}-\mathbf{I})(\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}) \\ &= \mathrm{tr}\{([\mathbf{I}-\mathbf{K}^2]^{-1}-\mathbf{I})(E_{H_0}\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}\mathbf{y}^{\mathrm{H}}\mathbf{R}_{yy}^{-\mathrm{H}/2}\mathbf{G})\}. \end{aligned} \tag{28}$$

Using (24), we may simplify this to

$$E_{H_0}l(\mathbf{y}) = \mathrm{tr}\{\mathbf{K}^2\} = \sum_{i=1}^{n} k_i^2. \tag{29}$$

The expected value of $l(\mathbf{y})$ under H_1 is

$$\begin{aligned} E_{H_1}l(\mathbf{y}) &= E_{H_1}(\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y})^{\mathrm{H}}([\mathbf{I}-\mathbf{K}^2]^{-1}-\mathbf{I})(\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}) \\ &= \mathrm{tr}\{\mathbf{I}-\mathbf{K}^2]^{-1}-\mathbf{I})(E_{H_1}\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}\mathbf{y}^{\mathrm{H}}\mathbf{R}_{yy}^{-\mathrm{H}/2}\mathbf{G})\} \end{aligned} \tag{30}$$

which may be simplified to

$$E_{H_1}l(\mathbf{y}) = \mathrm{tr}\{[\mathbf{I} - \mathbf{K}^2]^{-1} - \mathbf{I}\} = \sum_{i=1}^{n} \frac{k_i^2}{1-k_i^2}. \tag{31}$$

Thus, the expected value of the log-likelihood ratio is a T-invariant function of the squared canonical correlations k_i^2, under each hypothesis. Consequently, the J-divergence between H_1 and H_0 is

$$\begin{aligned} J = E_{H_1}l(\mathbf{y}) - E_{H_0}l(\mathbf{y}) &= \mathrm{tr}\{[\mathbf{I} - \mathbf{K}^2]^{-1} - \mathbf{I} - \mathbf{K}^2\} \\ &= \sum_{i=1}^{n} \frac{k_i^2}{1-k_i^2} k_i^2. \end{aligned} \tag{32}$$

This shows that J-divergence is a T-invariant function of the squared canonical correlations k_i^2.

We may write the J-divergence in terms of the γ_i^2 and l_i^2 as

$$J = \sum_{i=1}^{n} \frac{k_i^2}{1-k_i^2} k_i^2 = \sum_{i=1}^{n} \frac{\gamma_i^2}{1+\gamma_i^2} \gamma_i^2 = \sum_{i=1}^{n} \left(l_i - \frac{1}{l_i} \right)^2. \qquad (33)$$

The function $[k_i^2/(1-k_i^2)]k_i^2$ is nonincreasing for $i = 1, \ldots, r$. Consequently, the rank-r detector that maximizes divergence is the detector that uses the coordinates $(\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y})_i$, corresponding to the r-dominant squared canonical correlations $k_1^2, \ldots, k_r^2$. This low-rank detector and its resulting J-divergence are, respectively, given by (26) and (33), with r replacing n in both equations.

Thus, for building low-rank Gauss–Gauss detectors, the dominant canonical coordinates need to be retained. These are the coordinates with the largest per-mode SNRs and the largest per-mode signal-plus-noise-to-noise ratios. This result shows that canonical coordinates are fundamental for low-rank detection, in the case where the underlying model is a signal-plus-noise model, and ties the low-rank detectors of [1] to the maximum information rate low-rank estimators of [5], [6], and [9]–[11].

Remark: Typically, in low-rank estimation and detection, only a few dominant canonical coordinates and correlations are required. A conventional method of canonical coordinate decomposition as in (11), however, does not offer a simple way to compute a small subset of canonical coordinates and correlations. A full SVD for the coherence matrix has to be computed, regardless of the rank reduction. In addition, the conventional method does not allow an easy update of the canonical coordinates and correlations in time to accommodate new data, making it intractable for online applications. Therefore, iterative methods for recursive extraction of canonical coordinates are desired. In [15], simple algorithms called *alternating power methods with deflation* have been reported to recursively compute the canonical coordinates and correlations one-by-one or in groups. These algorithms also allow for an update of the canonical coordinates and correlations in time as new data is collected. Provided that the rank reduction is relatively aggressive and the singular values of the coherence matrix are not close together, the alternating power methods of [15] can be more efficient in computation than the conventional method. The reader is referred to [15] for details.

VI. Conclusion

Evidently, canonical coordinates are the right coordinates for building low-rank Gauss–Gauss detectors and estimators, when the underlying model is a signal-plus-noise model. Four important performance measures for detection and estimation are functions of the squared canonical correlations, and thus are invariant to nonsingular transformations of the measurement and signal channels. These measures are the relative volume of error concentration ellipse, processing gain, information rate, and J-divergence. In addition, these performance measures have additive or multiplicative decompositions in canonical coordinates that make canonical coordinates compelling for rank reduction. Under a rank constraint, it is the dominant canonical coordinates that need to be retained for detection and estimation. The dominant canonical coordinates are also the coordinates with the largest per-mode SNRs.

Appendix
Derivations of (24) and (25)

The covariance of $\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}$ under H_0 is

$$\begin{aligned} &E_{H_0}\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}\mathbf{y}^{\mathrm{H}}\mathbf{R}_{yy}^{-\mathrm{H}/2}\mathbf{G} \\ &= \mathbf{G}^{\mathrm{H}}(\mathbf{R}_{xx}+\mathbf{R}_{nn})^{-1/2}\mathbf{R}_{nn}(\mathbf{R}_{xx}+\mathbf{R}_{nn})^{-\mathrm{H}/2}\mathbf{G} \\ &= \mathbf{G}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}[(\mathbf{C}\mathbf{C}^{\mathrm{H}})^{-1} - \mathbf{I}]\mathbf{C}\mathbf{G}, \end{aligned} \qquad (A.1)$$

where the second equality follows from $(\mathbf{R}_{xx} + \mathbf{R}_{nn})^{-\mathrm{H}/2} = \mathbf{R}_{xx}^{-\mathrm{H}/2}\mathbf{C}$ and (20). Using the SVD $\mathbf{C} = \mathbf{F}\mathbf{K}\mathbf{G}^{\mathrm{H}}$ reduces (A.1) to (24). The covariance of $\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}$ under H_1 is

$$\begin{aligned} &E_{H_1}\mathbf{G}^{\mathrm{H}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}\mathbf{y}^{\mathrm{H}}\mathbf{R}_{yy}^{-\mathrm{H}/2}\mathbf{G} \\ &= \mathbf{G}^{\mathrm{H}}(\mathbf{R}_{xx}+\mathbf{R}_{nn})^{-1/2}(\mathbf{R}_{xx}+\mathbf{R}_{nn})(\mathbf{R}_{xx}+\mathbf{R}_{nn})^{-\mathrm{H}/2}\mathbf{G} \\ &= \mathbf{I}. \end{aligned} \qquad (A.2)$$

References

[1] L. L. Scharf and B. D. Van Veen, “Low rank detectors for Gaussian random vectors,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 11, pp. 1579–1582, Nov. 1987.

[2] J. Kirsteins and D. W. Tufts, “On the probability density of signal-to-noise ratio in an improved adaptive detector,” in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Mar. 1985, pp. 572–575.

[3] P. Strobach, “Low-rank detection of multichannel Gaussian signals using block matrix approximation,” *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 233–242, Jan. 1995.

[4] Y. Song and S. Roy, “Blind adaptive reduced-rank detection for DS-CDMA signals in multipath channels,” *IEEE J. Sel. Areas Commun.*, vol. 17, pp. 1960–1970, Nov. 1999.

[5] 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.

[6] L. L. Scharf and J. K. Thomas, “Wiener filters in canonical coordinates for transform coding, filtering, and quantizing,” *IEEE Trans. Signal Process.*, vol. 46, no. 3, pp. 647–654, Mar. 1998.

[7] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991, pp. 330–331.

[8] ——, “The SVD and reduced rank signal processing,” *Signal Process.*, vol. 25, pp. 113–133, 1991.

[9] 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.

[10] D. R. Brillinger, *Time Series: Data Analysis and Theory*. Philadelphia, PA: SIAM, 2001.

[11] Y. Hua and W. Liu, “Generalized Karhunen-Loeve transform,” *IEEE Signal Process. Lett.*, vol. 5, pp. 141–142, Jun. 1998.

[12] P. Stoica and M. Viberg, “Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions,” *IEEE Trans. Signal Process.*, vol. 44, pp. 3069–3078, Dec. 1996.

[13] 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.

[14] S. Kullback, *Information Theory and Statistics*. New York: Dover, 1968.

[15] A. Pezeshki, L. L. Scharf, M. R. Azimi-Sadjadi, and Y. Hua, “Two-channel constrained least squares problems: Solutions using power methods and connections with canonical coordinates,” *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 121–135, Jan. 2005.