Wiener Filters in Canonical Coordinates for Transform Coding, Filtering, and Quantizing

Louis L. Scharf, *Fellow, IEEE*, and John K. Thomas

Abstract— Canonical correlations are used to decompose the Wiener filter into a whitening transform coder, a canonical filter, and a coloring transform decoder. The outputs of the whitening transform coder are called canonical coordinates; these are the coordinates that are reduced in rank and quantized in our finite-precision version of the Gauss–Markov theorem. Canonical correlations are, in fact, cosines of the canonical angles between a source vector and a measurement vector. They produce new formulas for error covariance, spectral flatness, and entropy.

Index Terms—Adaptive filtering, canonical coordinates, canonical correlations, quantizing, transform coding, Wiener filters.

I. INTRODUCTION

CANONICAL correlations were introduced by Hotelling [1], [2] and further developed by Anderson [3]. They are now a standard topic in texts on multivariate analysis [4], [5]. Canonical correlations are closely related to *coherency spectra*, and these spectra have engaged the interest of acousticians and others for decades. In this paper, we take a fresh look at canonical correlations, in a filtering context, and discover that they provide a natural decomposition of the Wiener filter. In this decomposition, the singular value decomposition (SVD) of a *coherence matrix* plays a central role: The right singular vectors are used in a whitening transform coder to produce *canonical coordinates* of the measurement vector; the diagonal singular value matrix is used as a *canonical Wiener filter* to estimate the canonical source coordinates from the canonical measurement coordinates; and the left singular vectors are used in a coloring transform decoder to reconstruct the estimate of the source. The canonical source coordinates and the canonical measurement coordinates are white, but their cross correlation is the diagonal singular value matrix of the SVD, which is also called the *canonical correlation matrix*.

The Wiener filter is reduced in rank by purging subdominant canonical measurement coordinates that have small *squared-canonical correlation* with the canonical source coordinates. Quantizing is done by independently quantizing the canonical measurement coordinates to produce a quantized Wiener filter or a *quantized Gauss–Markov theorem*.

The abstract motivation for studying canonical correlations is that they provide a minimal description of the correlation between a source vector and a measurement vector. Canonical correlations are also cosines of canonical angles; therefore, some very illuminating geometrical insights are gained from a study of Wiener filters in canonical coordinates. The concrete motivation for studying canonical correlations is that they are the variables that determine how a Wiener filter can be reduced in rank and quantized for a finite-precision implementation.

Canonical correlations decompose formulas for error covariance, spectral flatness, and entropy, and they produce geometrical interpretations of all three. These decompositions show that canonical correlations play the role of direction cosines between random vectors, lending new insights into old formulas. All of these finite-dimensional results generalize to cyclic time series and to wide-sense stationary time series. Finally, experimental training data may be used in place of second-order information to produce formulas for adaptive Wiener filters in *adaptive* canonical coordinates.

II. PRELIMINARY OBSERVATIONS

Let us begin our discussion of canonical coordinates by revisiting an old problem in linear prediction. The zero-mean random vector $\mathbf{x} = [x(1) \quad x(2) \quad \cdots \quad x(m)]^\mathrm{T} = [x(1) \quad \mathbf{x}^\mathrm{T}(1)]$ has covariance matrix

$$\mathbf{R}_{xx} = E\begin{bmatrix} x(1) \\ \mathbf{x}(1) \end{bmatrix}[x(1) \quad \mathbf{x}^\mathrm{T}(1)] = \begin{bmatrix} r_{xx}(1) & \mathbf{r}_{xx}^\mathrm{T}(1) \\ \mathbf{r}_{xx}(1) & \mathbf{R}_{xx}(1) \end{bmatrix}. \quad (1)$$

The determinant of $\mathbf{R}_{xx}$ may be written as

$$\det[\mathbf{R}_{xx}] = q_{xx}(1)\det[\mathbf{R}_{xx}(1)] \quad (2)$$

where $q_{xx}(1)$ is the error variance for estimating the scalar $x(1)$ from the vector $\mathbf{x}(1)$. This error variance may be written as

$$q_{xx}(1) = r_{xx}(1)[1 - k_{xx}^2(1)] \quad (3)$$

$$k_{xx}^2(1) = \frac{\mathbf{r}_{xx}^\mathrm{T}(1)\mathbf{R}_{xx}^{-1}(1)\mathbf{r}_{xx}(1)}{r_{xx}(1)}. \quad (4)$$

We call $k_{xx}^2(1)$ the *squared coherence* between the scalar $x(1)$ and the vector $\mathbf{x}(1)$ because it may be written as the product

$$k_{xx}^2(1) = \mathbf{k}_{xx}(1)\mathbf{k}_{xx}^\mathrm{T}(1) \quad (5)$$

$$\mathbf{k}_{xx}(1) = r_{xx}^{-1/2}(1)\mathbf{r}_{xx}^\mathrm{T}(1)\mathbf{R}_{xx(1)}^{-\mathrm{T}/2}. \quad (6)$$

L. L. Scharf is with the Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80309-0425 USA (e-mail: scharf@boulder.colorado.edu).

J. K. Thomas is with the Data Fusion Corporation, Westminster, CO 80021 USA (e-mail: thomasjk@datafusion.com).

The vector $\mathbf{k}_{xx}(1)$ is the coherence between $x(1)$ and $\mathbf{x}(1)$, or the cross correlation between the white random scalar $r_{xx}^{-1/2}(1)x(1)$ and the white random vector $\mathbf{R}_{xx(1)}^{-1/2}\mathbf{x}(1)$:

$$\mathbf{k}_{xx}(1) = E[r_{xx}^{-1/2}(1)x(1)\mathbf{x}^{\mathrm{T}}(1)\mathbf{R}_{xx(1)}^{-\mathrm{T}/2}]$$
$$= r_{xx}^{-1/2}(1)\mathbf{r}_{xx}^{\mathrm{T}}(1)\mathbf{R}_{xx(1)}^{-\mathrm{T}/2}. \tag{7}$$

This basic idea may be iterated to write $\det[\mathbf{R}_{xx}]$ as

$$\det[\mathbf{R}_{xx}] = \prod_{i=1}^{m} q_{xx}(i)$$
$$= \prod_{i=1}^{m} r_{xx}(i)[1 - k_{xx}^2(i)] \tag{8}$$
$$k_{xx}^2(i) = \frac{\mathbf{r}_{xx}^{\mathrm{T}}(i)\mathbf{R}_{xx}^{-1}(i)\mathbf{r}_{xx}(i)}{r_{xx}(i)} \tag{9}$$

where $k_{xx}^2(i)$ is the squared coherence between the scalar $x(i)$ and the vector $\mathbf{x}(i) = [x(i+1) \cdots x(m)]^{\mathrm{T}}$. This formula for $\det[\mathbf{R}_{xx}]$ is the Gram determinant, with each prediction error variance written in terms of squared coherence. It provides a fine-grained resolution of *entropy* and *spectral flatness*:

$$H_{xx} = \frac{m}{2} \log(2\pi e) + \frac{1}{2} \log \det[\mathbf{R}_{xx}]$$
$$= \frac{m}{2} \log(2\pi e) + \frac{1}{2} \sum_{i=1}^{m} \log\left[r_{xx}(i)[1 - k_{xx}^2(i)]\right] \tag{10}$$
$$SF_{xx} = \frac{\det[\mathbf{R}_{xx}]}{\prod_{i=1}^{m} [\mathbf{R}_{xx}]_{ii}} = \prod_{i=1}^{m} [1 - k_{xx}^2(i)]. \tag{11}$$

Therefore, entropy is near its maximum, and spectral flatness is near 1 when the squared coherences between $x(i)$ and $\mathbf{x}(i)$ are near zero for all i.

The sequence of Wiener filters that underlies this decomposition of $\det[\mathbf{R}_{xx}]$ is

$$\mathbf{w}(i) = \mathbf{r}_{xx}^{\mathrm{T}}(i)\mathbf{R}_{xx}^{-1}(i) = r_{xx}^{1/2}(i)\mathbf{k}_{xx}(i)\mathbf{R}_{xx}^{-1/2}(i) \tag{12}$$

which is a decomposition of the filter into a whitener $\mathbf{R}_{xx}^{-1/2}(i)$, a coherence filter $\mathbf{k}_{xx}(i)$, and a colorer $r_{xx}^{1/2}(i)$. This idea is fundamental.

III. CANONICAL CORRELATIONS IN A FILTERING CONTEXT

The context for our further development of canonical correlations is illustrated in Fig. 1. The $m \times 1$ source vector $\mathbf{x}$ and the $n \times 1$ measurement vector $\mathbf{y}$ are generated by Mother Nature. Father Nature views only the measurement vector $\mathbf{y}$, and from it, he must estimate Mother Nature's source vector $\mathbf{x}$. This problem is meaningful because the zero-mean random vectors $\mathbf{x}$ and $\mathbf{y}$ share the covariance matrix $\mathbf{R}_{zz}$:

$$\mathbf{R}_{zz} = E\begin{bmatrix}\mathbf{x}\\\mathbf{y}\end{bmatrix}[\mathbf{x}^{\mathrm{T}} \quad \mathbf{y}^{\mathrm{T}}] = \begin{bmatrix}\mathbf{R}_{xx} & \mathbf{R}_{xy}\\\mathbf{R}_{xy}^{\mathrm{T}} & \mathbf{R}_{yy}\end{bmatrix}. \tag{13}$$
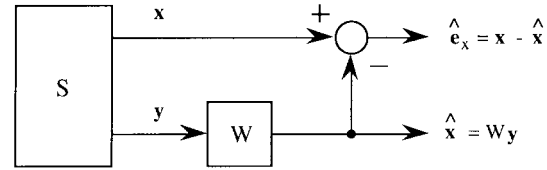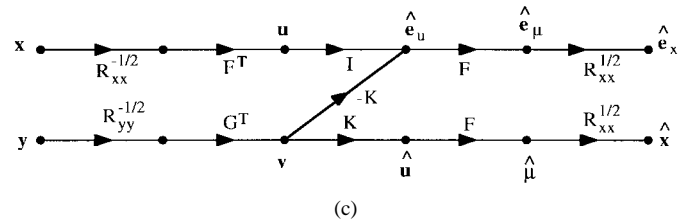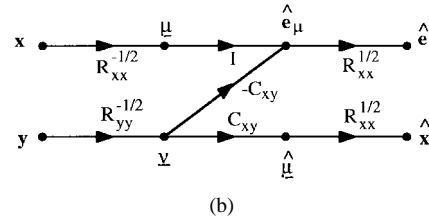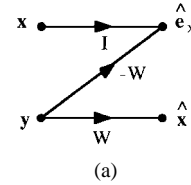

Fig. 1.   Filtering problem.

Fig. 2.   Wiener filter in various coordinate systems.

A. Standard Coordinates

The linear MMSE estimator of $\mathbf{x}$ from $\mathbf{y}$ is $\hat{\mathbf{x}} = \mathbf{W}\mathbf{y}$, and the corresponding (orthogonal) error is $\hat{\mathbf{e}}_x = \mathbf{x} - \hat{\mathbf{x}}$. In standard coordinates, the Wiener filter $\mathbf{W}$ and the error covariance matrix $\mathbf{Q}_{xx}$ are

$$\mathbf{W} = \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1} \tag{14}$$
$$\mathbf{Q}_{xx} = E[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^{\mathrm{T}}] = \mathbf{R}_{xx} - \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{xy}^{\mathrm{T}}. \tag{15}$$

We shall call Fig. 2(a) the Wiener filter in standard coordinates.

The linear transformation

$$\begin{bmatrix}\hat{\mathbf{e}}_x\\\mathbf{y}\end{bmatrix} = \begin{bmatrix}\mathbf{I} & -\mathbf{W}\\\mathbf{0} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{x}\\\mathbf{y}\end{bmatrix} \tag{16}$$

resolves the source vector $\mathbf{x}$ and the measurement vector $\mathbf{y}$ into orthogonal vectors $\hat{\mathbf{e}}_x$ and $\mathbf{y}$, with respective covariances $\mathbf{Q}_{xx}$ and $\mathbf{R}_{yy}$

$$\begin{bmatrix}\mathbf{Q}_{xx} & \mathbf{0}\\\mathbf{0} & \mathbf{R}_{yy}\end{bmatrix} = \begin{bmatrix}\mathbf{I} & -\mathbf{W}\\\mathbf{0} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{R}_{xx} & \mathbf{R}_{xy}\\\mathbf{R}_{xy}^{\mathrm{T}} & \mathbf{R}_{yy}\end{bmatrix}\begin{bmatrix}\mathbf{I} & \mathbf{0}\\-\mathbf{W}^{\mathrm{T}} & \mathbf{I}\end{bmatrix}. \tag{17}$$

This is one of the Schur decompositions of $\mathbf{R}_{zz}$. From this formula, it follows that $\det[\mathbf{R}_{zz}]$ may be written as

$$\det[\mathbf{R}_{zz}] = \det[\mathbf{Q}_{xx}]\det[\mathbf{R}_{yy}] \tag{18}$$
$$\det[\mathbf{Q}_{xx}] = \det[\mathbf{R}_{xx} - \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{xy}^{\mathrm{T}}]. \tag{19}$$

B. Coherence Coordinates

The *coherence matrix* measures the cross-correlation between the *white* vectors $\mathbf{R}_{xx}^{-1/2}\mathbf{x}$ and $\mathbf{R}_{yy}^{-1/2}\mathbf{y}$:

$$\mathbf{C}_{xy} = E[\mathbf{R}_{xx}^{-1/2}\mathbf{x}(\mathbf{R}_{yy}^{-1/2}\mathbf{y})^{\mathrm{T}}] = \mathbf{R}_{xx}^{-1/2}\mathbf{R}_{xy}\mathbf{R}_{yy}^{-\mathrm{T}/2}. \quad (20)$$

Using coherence, we can refine the Wiener filter $\mathbf{W}$ and its corresponding error covariance matrix $\mathbf{Q}_{xx}$ as

$$\mathbf{W} = \mathbf{R}_{xx}^{1/2}\mathbf{C}_{xy}\mathbf{R}_{yy}^{-1/2} \quad (21)$$

$$\mathbf{Q}_{xx} = \mathbf{R}_{xx}^{1/2}(I - \mathbf{C}_{xy}\mathbf{C}_{xy}^{\mathrm{T}})\mathbf{R}_{xx}^{\mathrm{T}/2}. \quad (22)$$

We shall call the matrix $\mathbf{C}_{xy}\mathbf{C}_{xy}^{\mathrm{T}}$ the *squared coherence matrix.*

The corresponding Wiener filter, in coherence coordinates, is illustrated in Fig. 2(b). It resolves the source vector $\mathbf{x}$ and the measurement vector $\mathbf{y}$ into the error vector $\hat{\mathbf{e}}_x$ and the estimate $\hat{\mathbf{x}}$ in three stages. The first stage whitens both $\mathbf{x}$ and $\mathbf{y}$ to produce the coherence coordinates $\underline{\mu}$ and $\underline{\nu}$, the second stage filters $\underline{\nu}$ with the coherence filter $\overline{\mathbf{C}}_{xy}$ to produce the estimator error $\hat{\mathbf{e}}_\mu$ and the estimator $\hat{\underline{\mu}}$, and the third stage colors these to produce $\hat{\mathbf{e}}_x$ and $\hat{\mathbf{x}}$. We shall call this the Wiener filter in coherence coordinates.

The refined linear transformation from $(\mathbf{x}, \mathbf{y})$ to $(\hat{\mathbf{e}}_x, \mathbf{y})$ is

$$\begin{bmatrix} \hat{\mathbf{e}}_x \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{xx}^{1/2} & 0 \\ 0 & \mathbf{R}_{yy}^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{C}_{xy} \\ 0 & \mathbf{I} \end{bmatrix}$$
$$\cdot \begin{bmatrix} \mathbf{R}_{xx}^{-1/2} & 0 \\ 0 & \mathbf{R}_{yy}^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}. \quad (23)$$

The corresponding refinement for the covariance matrix for $\hat{\mathbf{e}}_x$ and $\mathbf{y}$ is

$$\begin{bmatrix} \mathbf{Q}_{xx} & 0 \\ 0 & \mathbf{R}_{yy} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{xx}^{1/2} & 0 \\ 0 & \mathbf{R}_{yy}^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{I} - \mathbf{C}_{xy}\mathbf{C}_{xy}^{\mathrm{T}} & 0 \\ 0 & \mathbf{I} \end{bmatrix}$$
$$\cdot \begin{bmatrix} \mathbf{R}_{xx}^{\mathrm{T}/2} & 0 \\ 0 & \mathbf{R}_{yy}^{\mathrm{T}/2} \end{bmatrix}. \quad (24)$$

The diagonal structure of this covariance matrix shows that the estimator error $\hat{\mathbf{e}}_\mu$ and the measurement $\underline{\nu}$, in coherence coordinates, are also uncorrelated, providing an orthogonal decomposition of the coherence coordinate $\underline{\mu}$ into the estimator $\hat{\underline{\mu}}$ and the error $\hat{\mathbf{e}}_\mu$. It also shows that the covariance matrix for the error in coherence coordinates is $\mathbf{Q}_{\mu\mu} = \mathbf{I} - \mathbf{C}_{xy}\mathbf{C}_{xy}^{\mathrm{T}}$. The formula for $\det[\mathbf{R}_{zz}]$ is now

$$\det[\mathbf{R}_{zz}] = \det[\mathbf{Q}_{xx}]\det[\mathbf{R}_{yy}] \quad (25)$$

$$\det[\mathbf{Q}_{xx}] = \det[\mathbf{R}_{xx}]\det[\mathbf{I} - \mathbf{C}_{xy}\mathbf{C}_{xy}^{\mathrm{T}}]. \quad (26)$$

C. Canonical Coordinates

We achieve one more level of refinement by replacing the coherence matrix $\mathbf{C}_{xy}$ by its SVD:

$$\mathbf{C}_{xy} = \mathbf{F}\mathbf{K}\mathbf{G}^{\mathrm{T}}; \quad \mathbf{K} = \mathbf{F}^{\mathrm{T}}\mathbf{C}_{xy}\mathbf{G}$$

$$\mathbf{F}\mathbf{F}^{\mathrm{T}} = \mathbf{I}_{m \times m}\mathbf{G}\mathbf{G}^{\mathrm{T}} = \mathbf{I}_{n \times n} \quad (27)$$

$$\mathbf{K} = [\mathbf{K}(m)\, 0]; \qquad m < n$$

$$= \begin{bmatrix} \mathbf{K}(n) \\ 0 \end{bmatrix}; \qquad m > n \quad (28)$$

$$\mathbf{K}(m) = \mathrm{diag}[k(1)\, k(2)\, \cdots\, k(m)]. \quad (29)$$

We shall call the orthogonal matrices $\mathbf{F}^{\mathrm{T}}$ and $\mathbf{G}^{\mathrm{T}}$ *transform coders*, the matrix $\mathbf{K}$ the *canonical correlation matrix*, and the matrix $\mathbf{K}\mathbf{K}^{\mathrm{T}}$ the *squared canonical correlation matrix*. The canonical correlation matrix $\mathbf{K}$ is the cross correlation between the white vector $\mathbf{F}^{\mathrm{T}}\mathbf{R}_{xx}^{-1/2}\mathbf{x}$ and the white vector $\mathbf{G}^{\mathrm{T}}\mathbf{R}_{yy}^{-1/2}\mathbf{y}$:

$$\mathbf{K} = E[\mathbf{F}^{\mathrm{T}}\mathbf{R}_{xx}^{-1/2}\mathbf{x}\mathbf{y}^{\mathrm{T}}\mathbf{R}_{yy}^{-\mathrm{T}/2}\mathbf{G}] = \mathbf{F}^{\mathrm{T}}\mathbf{C}_{xy}\mathbf{G}. \quad (30)$$

The Wiener filter $\mathbf{W}$ and error covariance matrix $\mathbf{Q}_{xx}$ in these canonical coordinates are

$$\mathbf{W} = \mathbf{R}_{xx}^{1/2}\mathbf{F}\mathbf{K}\mathbf{G}^{\mathrm{T}}\mathbf{R}_{yy}^{-1/2} \quad (31)$$

$$\mathbf{Q}_{xx} = \mathbf{R}_{xx}^{1/2}\mathbf{F}(\mathbf{I} - \mathbf{K}\mathbf{K}^{\mathrm{T}})\mathbf{F}^{\mathrm{T}}\mathbf{R}_{xx}^{\mathrm{T}/2}. \quad (32)$$

The corresponding Wiener filter, in canonical coordinates, is illustrated in Fig. 2(c). It resolves the source vector $\mathbf{x}$ and the measurement vector $\mathbf{y}$ into the error vector $\hat{\mathbf{e}}_x$ and the estimator $\hat{\mathbf{x}}$ in five stages. The first stage whitens both $\mathbf{x}$ and $\mathbf{y}$ to produce the coherence coordinates $\underline{\mu}$ and $\underline{\nu}$, the second stage transforms the coherence coordinates $\underline{\mu}$ and $\underline{\nu}$ into the canonical coordinates $\mathbf{u}$ and $\mathbf{v}$, the third stage filters $\mathbf{v}$ with the canonical filter $\mathbf{K}$ to produce the estimator $\hat{\mathbf{u}}$ and the estimator error $\hat{\mathbf{e}}_u$, the fourth stage transforms $\hat{\mathbf{u}}$ and $\hat{\mathbf{e}}_u$ into the coherence coordinates $\hat{\underline{\mu}}$ and $\hat{\mathbf{e}}_\mu$, and the fifth stage colors these to produce $\hat{\mathbf{x}}$ and $\hat{\mathbf{e}}_x$. We shall call this the Wiener filter in canonical coordinates.

The refined linear transformation from $(\mathbf{x}, \mathbf{y})$ to $(\hat{\mathbf{e}}_x, \mathbf{y})$ is

$$\begin{bmatrix} \hat{\mathbf{e}}_x \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{xx}^{1/2} & 0 \\ 0 & \mathbf{R}_{yy}^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{F} & 0 \\ 0 & \mathbf{G} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{K} \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{F}^{\mathrm{T}} & 0 \\ 0 & \mathbf{G}^{\mathrm{T}} \end{bmatrix}$$
$$\cdot \begin{bmatrix} \mathbf{R}_{xx}^{-1/2} & 0 \\ 0 & \mathbf{R}_{yy}^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}. \quad (33)$$

The corresponding refinement of the covariance matrix for $\hat{\mathbf{e}}_x$ and $\mathbf{y}$ is

$$\begin{bmatrix} \mathbf{Q}_{xx} & 0 \\ 0 & \mathbf{R}_{yy} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{xx}^{1/2} & 0 \\ 0 & \mathbf{R}_{yy}^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{F} & 0 \\ 0 & \mathbf{G} \end{bmatrix} \begin{bmatrix} \mathbf{I} - \mathbf{K}\mathbf{K}^{\mathrm{T}} & 0 \\ 0 & \mathbf{I} \end{bmatrix}$$
$$\cdot \begin{bmatrix} \mathbf{F}^{\mathrm{T}} & 0 \\ 0 & \mathbf{G}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{R}_{xx}^{\mathrm{T}/2} & 0 \\ 0 & \mathbf{R}_{yy}^{\mathrm{T}/2} \end{bmatrix}. \quad (34)$$

The diagonal structure of this covariance matrix shows that the estimator error $\hat{\mathbf{e}}_u$ and the measurement $\mathbf{v}$ are also uncorrelated, meaning that the estimator $\hat{\mathbf{u}}$ and the error $\hat{\mathbf{e}}_u$ orthogonally decompose the canonical coordinate $\mathbf{u}$. It also shows that the covariance matrix for the error in canonical coordinates is $\mathbf{Q}_{uu} = \mathbf{I} - \mathbf{K}\mathbf{K}^{\mathrm{T}}$. The formula for $\det[\mathbf{R}_{zz}]$ is now

$$\det[\mathbf{R}_{zz}] = \det[\mathbf{R}_{xx}]\det[\mathbf{I} - \mathbf{K}\mathbf{K}^{\mathrm{T}}]\det[\mathbf{R}_{yy}] \quad (35)$$

$$\det[\mathbf{Q}_{xx}] = \det[\mathbf{R}_{xx}]\det[\mathbf{I} - \mathbf{K}\mathbf{K}^{\mathrm{T}}]$$
$$= \det[\mathbf{R}_{xx}]\prod_{i=1}^{\min(m, n)}[1 - k^2(i)]. \quad (36)$$

This formula shows that the squared canonical correlations $k^2(i)$ are objects of fundamental importance for filtering. We pursue this point in Section IV.

IV. Filtering Formulas in Canonical Coordinates

We summarize as follows. The Wiener filter in *canonical coordinates* replaces the source and measurement vectors in *standard* coordinates with source and measurement vectors in *canonical* coordinates. In these coordinates, the source and measurement are white but diagonally cross correlated according to the canonical correlation matrix $\mathbf{K}$. This canonical correlation matrix is also the Wiener filter for estimating the canonical source coordinates from the canonical measurement coordinates. The error covariance matrix associated with Wiener filtering in these canonical coordinates is just $\mathbf{Q}_{uu} = \mathbf{I} - \mathbf{K}\mathbf{K}^{\mathrm{T}}$.

Recall that the canonical correlations are defined as

$$\mathbf{K} = E[\mathbf{F}^{\mathrm{T}}\mathbf{R}_{xx}^{-1/2}\mathbf{x}(\mathbf{G}^{\mathrm{T}}\mathbf{R}_{yy}^{-1/2}\mathbf{y})^{\mathrm{T}}] = \mathbf{F}^{\mathrm{T}}\mathbf{C}_{xy}\mathbf{G} \quad (37)$$

$$k(i) = [\mathbf{K}]_{ii}; \qquad i = 1, 2, \cdots, \min(m, n) \quad (38)$$

so that each canonical correlation $k(i)$ measures the cosine of the angle between two unit variance random variables: one drawn from the canonical source coordinates and one drawn from the canonical measurement coordinates. For this reason, we call the squared canonical correlations $k^2(i)$ direction cosines. By making the canonical variables diagonally correlated, we have uncoupled the measurement of one direction cosine from the measurement of another.

A. Linear Dependence

We think of the Hadamard ratio $\det[\mathbf{R}_{zz}]/\Pi(\mathbf{R}_{zz})_{ii}$ as a measure of linear dependence of the variables $\mathbf{x}$ and $\mathbf{y}$. Using the results of (8) and (35), we may write the Hadamard ratio as the product

$$\frac{\det[\mathbf{R}_{zz}]}{\Pi(\mathbf{R}_{zz})_{ii}}$$

$$= \frac{\det[\mathbf{R}_{xx}]}{\Pi(\mathbf{R}_{xx})_{ii}} \prod_{i=1}^{\min(m, n)} [1 - k^2(i)] \frac{\det[\mathbf{R}_{yy}]}{\Pi(\mathbf{R}_{yy})_{ii}}$$

$$= \prod_{i=1}^{m}[1 - k_{xx}^2(i)] \prod_{i=1}^{\min(m, n)} [1 - k^2(i)] \prod_{i=1}^{n}[1 - k_{yy}^2(i)].$$

$$(39)$$

This formula tells us that what matters is the *intra*dependence within $\mathbf{x}$ as measured by its direction cosines, the *intra*dependence within $\mathbf{y}$ as measured by its direction cosines, and the *inter*dependence between $\mathbf{x}$ and $\mathbf{y}$ as measured by the direction cosines between $\mathbf{x}$ and $\mathbf{y}$. These latter direction cosines are measured in canonical coordinates, much as principal angles between subspaces are measured in something akin to canonical coordinates. They are scale invariant.

B. Relative Filtering Errors

The prior error covariance for the message vector $\mathbf{x}$ is $\mathbf{R}_{xx}$, and the posterior error covariance for the error $\hat{\mathbf{e}}_x = \mathbf{x} - \hat{\mathbf{x}}$ is $\mathbf{Q}_{xx}$. The volumes of the concentration ellipses associated with these covariances are proportional to $\det[\mathbf{R}_{xx}]$ and $\det[\mathbf{Q}_{xx}]$.

The relative volumes depend only on the direction cosines $k^2(i)$:

$$\frac{\det[\mathbf{Q}_{xx}]}{\det[\mathbf{R}_{xx}]} = \prod_{i=1}^{\min(m, n)} [1 - k^2(i)]. \quad (40)$$

C. Entropy and Rate

The *entropy* of the random vector $\mathbf{z}$ is

$$H_{zz} = \frac{m + n}{2} \log(2\pi e) + \frac{1}{2} \log \det[\mathbf{R}_{zz}]$$

$$= H_{xx} + \frac{1}{2} \sum_{i=1}^{\min(m, n)} \log[1 - k^2(i)] + H_{yy}$$

$$H_{xx} = \frac{m}{2} \log(2\pi e) + \frac{1}{2} \log \det[\mathbf{R}_{xx}]$$

$$H_{yy} = \frac{n}{2} \log(2\pi e) + \frac{1}{2} \log \det[\mathbf{R}_{yy}]. \quad (41)$$

Normally, we write this entropy as the conditional entropy of $\mathbf{x}$ given $\mathbf{y}$, plus the entropy of $\mathbf{y}$. The conditional entropy, or *equivocation,* is therefore

$$H_{x|y} = H_{xx} + \frac{1}{2} \sum_{i=1}^{\min(m, n)} \ln[1 - k^2(i)] \quad (42)$$

and the direction cosines determine how $\mathbf{y}$ brings information about $\mathbf{x}$ to reduce its entropy from its prior value of H_{xx}. The second term on the right-hand side of this equation is the negative of *rate* in canonical coordinates. Thus, the rate at which $\mathbf{y}$ brings information about $\mathbf{x}$ is determined by the direction cosines or squared canonical correlations between the source and the measurement.

V. Rank Reduction for Transform Coding, Filtering, and Quantizing

The Wiener filter in canonical coordinates is a filterbank idea. That is, the measurement is decomposed into canonical coordinates that bring information about the canonical coordinates of the source. It is also a spread-spectrum idea because the canonical coordinates are white. The question of rank reduction and bit allocation for finite-precision Wiener filtering or, equivalently, for source coding from noisy measurements is clarified in canonical coordinates. The problem is to quantize the canonical coordinates $[\mathbf{v}]_i$ so that the trace of the error covariance matrix $\mathbf{Q}_{xx}$ is minimized. The error covariance matrix and its trace are

$$\mathbf{Q}_{xx} = \mathbf{R}_{xx}^{1/2}\mathbf{F}(\mathbf{I} - \mathbf{K}\mathbf{K}^{\mathrm{T}})\mathbf{F}^{\mathrm{T}}\mathbf{R}_{xx}^{\mathrm{T}/2} \quad (43)$$

$$\mathrm{tr}[\mathbf{Q}_{xx}] = \sum_{i=1}^{\min(m, n)} [1 - k^2(i)]\xi^2(i) \quad (44)$$

where the $\xi^2(i)$ are the energies of the "impulse responses" for the coloring (or synthesizing) transform decoder:

$$\xi^2(i) = \mathbf{s}_i^{\mathrm{T}}\mathbf{s}_i; \quad \mathbf{s}_i = (\mathbf{R}_{xx}^{1/2}\mathbf{F})_i. \quad (45)$$

If the canonical measurement coordinates $[\mathbf{v}]_i$ that are weakly correlated with the canonical source coordinates are purged

and the remaining are uniformly quantized with b_i bits, then the resulting error covariance matrix for estimating the source vector $\mathbf{x}$ from the reduced-rank and quantized canonical measurement vector $\mathbf{v}$ is

$$
\begin{aligned}
\text{tr}[\mathbf{Q}_{xx}(r)] = & \sum_{i=1}^{r} [1 - k^2(i)]\xi^2(i) + \sum_{i=r+1}^{\min(m,n)} \xi^2(i) \\
& + \sum_{i=1}^{r} k^2(i)\xi^2(i)2^{-2b_i} \\
= & \sum_{i=1}^{\min(m,n)} [1 - k^2(i)]\xi^2(i) + \sum_{i=r+1}^{\min(m,n)} k^2(i)\xi^2(i) \\
& + \sum_{i=1}^{r} k^2(i)\xi^2(i)2^{-2b_i}.
\end{aligned}
\tag{46}
$$

In this latter form, we observe that $\text{tr}[\mathbf{Q}_{xx}(r)]$ consists of three terms: the infinite-precision filtering error, the bias-squared introduced by rank reduction, and the variance introduced by quantizing. The trick is to properly balance the second and third. To this end, we will consider the rate-distortion problem

$$
\min D = \min \frac{1}{m} \text{tr}[\mathbf{Q}_{xx}(r)] \text{ under constraint } \sum_{i=1}^{r} b_i = mR.
\tag{47}
$$

Using the standard procedure for minimizing with constraint (see, for example, [9] and [10]), we obtain the solution

$$
b_i = \frac{1}{2} \log_2 \frac{k^2(i)\xi^2(i)}{\nu}
\tag{48}
$$

$$
r = \max \arg_i \left[k^2(i)\xi^2(i) > \nu\right]
\tag{49}
$$

$$
\begin{aligned}
R &= \frac{1}{m} \sum_{i=1}^{r} \frac{1}{2} \log_2 \frac{k^2(i)\xi^2(i)}{\nu} \\
&= \frac{1}{m} \sum_{i=1}^{\min(m,n)} \max \left[\frac{1}{2} \log_2 \frac{k^2(i)\xi^2(i)}{\nu}, 0\right]
\end{aligned}
\tag{50}
$$

$$
\begin{aligned}
D &= \frac{1}{m} \sum_{i=1}^{r} [1 - k^2(i)]\xi^2(i) + \frac{1}{m} \sum_{i=r+1}^{\min(m,n)} \xi^2(i) + \frac{1}{m} \sum_{i=1}^{r} \nu \\
&= \frac{1}{m} \sum_{i=1}^{\min(m,n)} \min\{[1 - k^2(i)]\xi^2(i) + \nu, \xi^2(i)\}.
\end{aligned}
\tag{51}
$$

These formulas generalize the formulas of [9] by providing a solution to the problem of uniformly quantizing the Wiener filter or quantizing the Gauss–Markov theorem. They may be interpreted as follows.

If the bit rate R is specified, then the slicing level ν is adjusted to achieve it. The slicing level determines the bit allocation b_i, the rank r, and the minimum achievable distortion D. Conversely, if the distortion D is specified, ν is adjusted to achieve it. This determines b_i, r, and the minimum rate R. These formulas are illustrated in Fig. 3 for the idealized case where the $\xi^2(i)$ are unity. The components of distortion illustrate the tradeoff between bias and variance.
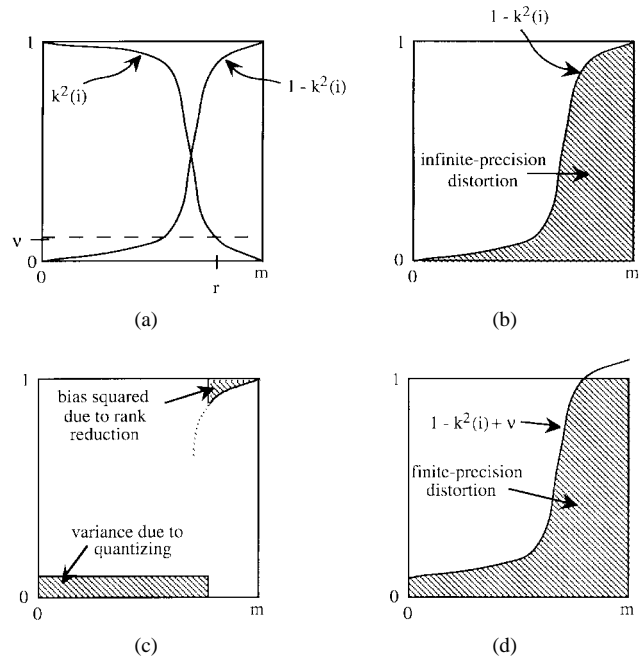

Fig. 3. Components of distortion. (a) Squared canonical correlation. (b) Infinite-precision distortion. (c) Extra components of distortion due to rank reduction and quantizing. (d) Finite-precision distortion.

VI. CANONICAL TIME SERIES

If $\mathbf{x}$ and $\mathbf{y}$ are jointly stationary random vectors whose dimensions increase without bound (that is, they are stationary time series), then all of the correlation matrices in these formulas are infinite Toeplitz matrices with Fourier representations

$$
\mathbf{R}_{xx} \leftrightarrow S_{xx}(e^{j\theta}), \ \mathbf{R}_{xy} \leftrightarrow S_{xy}(e^{j\theta}), \ \mathbf{R}_{yy} \leftrightarrow S_{yy}(e^{j\theta}).
\tag{52}
$$

Furthermore, if the time series are not perfectly predictable (that is, the power spectra S_{xx} and S_{yy} satisfy the Szëgo conditions), then $S_{xx}(e^{j\theta})$ and $S_{yy}(e^{j\theta})$ may be spectrally factored as

$$
\begin{aligned}
S_{xx}(e^{j\theta}) &= \frac{1}{A(e^{j\theta})A(e^{-j\theta})} \\
S_{yy}(e^{j\theta}) &= \frac{1}{B(e^{j\theta})B(e^{-j\theta})}
\end{aligned}
\tag{53}
$$

where the filters $A(z)$ and $B(z)$ are minimum phase, meaning that $A(z)$, $B(z)$, $1/A(z)$, and $1/B(z)$ are causal and stable filters. Then, the various square roots in the filtering fomulas have the Fourier representations

$$
\mathbf{R}_{xx}^{1/2} \leftrightarrow \frac{1}{A(e^{j\theta})}; \ \mathbf{R}_{xx}^{T/2} \leftrightarrow \frac{1}{A(e^{-j\theta})}
\tag{54a}
$$

$$
\mathbf{R}_{xx}^{-1/2} \leftrightarrow A(e^{j\theta}); \ \mathbf{R}_{xx}^{-T/2} \leftrightarrow A(e^{-j\theta})
\tag{54b}
$$

$$
\mathbf{R}_{yy}^{-1/2} \leftrightarrow B(e^{j\theta}); \ \mathbf{R}_{yy}^{-T/2} \leftrightarrow B(e^{-j\theta}).
\tag{54c}
$$

The SVD representation for $\mathbf{C}_{xy}$ becomes a Fourier representation; therefore

$$
\begin{aligned}
K(e^{j\theta}) &\leftrightarrow \mathbf{F}\mathbf{K}\mathbf{G}^{T} = \mathbf{C}_{xy} = \mathbf{R}_{xx}^{-1/2}\mathbf{R}_{xy}\mathbf{R}_{yy}^{-T/2} \\
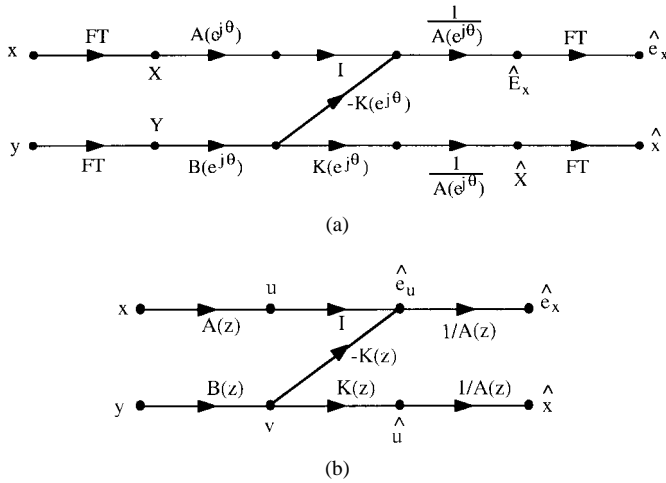&\leftrightarrow A(e^{j\theta})S_{xy}(e^{j\theta})B(e^{-j\theta})
\end{aligned}
\tag{55}
$$

Fig. 4. Canonical Wiener filter in (a) frequency and (b) time domains.

where

$K(e^{j\theta})$ *coherence spectrum*;

$\mathbf{K}$ spectral mask $\mathrm{diag}[K(e^{j\theta})]$;

$\mathbf{F}$ and $\mathbf{G}$ Fourier matrices.

That is, the coherence spectrum is the Fourier representation for the coherence matrix, and the *squared coherence spectrum* is the representation for the squared coherence matrix:

$$\mathbf{C}_{xy} \leftrightarrow K(e^{j\theta}) = A(e^{j\theta})S_{xy}(e^{j\theta})B(e^{-j\theta}) \quad (56)$$

$$\mathbf{C}_{xy}\mathbf{C}_{xy}^{\mathrm{T}} \leftrightarrow |K(e^{j\theta})|^2 = \frac{|S_{xy}(e^{j\theta})|^2}{S_{xx}(e^{j\theta})S_{yy}(e^{j\theta})}. \quad (57)$$

With these results, we summarize the filtering fomulas in canonical time series coordinates as

$$\mathbf{W} \leftrightarrow W(e^{j\theta}) = \frac{1}{A(e^{j\theta})}K(e^{j\theta})B(e^{j\theta}) \quad (58)$$

$$\mathbf{Q}_{xx} \leftrightarrow Q_{xx}(e^{j\theta}) = S_{xx}(e^{j\theta})[1 - |K(e^{j\theta})|^2]. \quad (59)$$

This equation for $W(e^{j\theta})$ shows it to be a product of spectral masks, as illustrated in Fig. 4(a). Alternatively, $W(e^{j\theta})$ is the frequency response of the Wiener filter $W(z)$, and $Q_{xx}(e^{j\theta})$ is its error spectrum. This filter is illustrated in Fig. 4(b). If the filter is constrained to be causal, then the coherence filter $K(z)$ is replaced by the causal filter

$$K_+(z) = A(z)[S_{xy}(z)B(z^{-1})]_+. \quad (60)$$

A. Error Variance, Spectral Flatness, and Entropy

The Toeplitz matrix $\mathbf{Q}_{xx}$ has the error variance σ_x^2 on its diagonal. Therefore, this variance is

$$\sigma_x^2 = \lim_{m \to \infty} \left(\prod_{i=1}^{m}[\mathbf{Q}_{xx}]_{ii}\right)^{1/m}$$

$$= \int_{-\pi}^{\pi} Q_{xx}(e^{j\theta})\frac{d\theta}{2\pi}$$

$$= \int_{-\pi}^{\pi} S_{xx}(e^{j\theta})[1 - |K(e^{j\theta})|^2]\frac{d\theta}{2\pi}. \quad (61)$$

This formula shows the error spectrum to be the product of the source spectrum and an error spectrum, where the latter is determined by the squared coherence spectrum.

The spectral flatness of the error spectrum is

$$SF_{ee} = \frac{\lim_{m \to \infty}(\det[\mathbf{Q}_{xx}])^{1/m}}{\lim_{m \to \infty}\left(\prod_{i=1}^{m}[\mathbf{Q}_{xx}]_{ii}\right)^{1/m}}$$

$$= \frac{\exp\left\{\int_{-\pi}^{\pi} \log Q_{xx}(e^{j\theta})\frac{d\theta}{2\pi}\right\}}{\int_{-\pi}^{\pi} Q_{xx}(e^{j\theta})\frac{d\theta}{2\pi}} \quad (62)$$

which is the ratio of prediction error variance to prior variance.

The equivocation of $\mathbf{x}$ given $\mathbf{y}$ is

$$H_{x|y} = \lim_{m \to \infty}[\ln(\det[\mathbf{R}_{xx}])^{1/m} + \ln(\det[\mathbf{I} - \mathbf{K}\mathbf{K}^{\mathrm{T}}])^{1/m}]$$

$$= \int_{-\pi}^{\pi} \log S_{xx}(e^{j\theta})\frac{d\theta}{2\pi} + \int_{-\pi}^{\pi} \log[1 - |K(e^{j\theta})|^2]\frac{d\theta}{2\pi}. \quad (63)$$

The negative of the second term is the rate at which $\mathbf{y}$ brings information about $\mathbf{x}$, and it is determined by the squared coherence spectrum.

B. Quantizing

It is a straightforward exercise to reduce rank and quantize according to

$$b(e^{j\theta}) = \frac{1}{2}\log_2 \frac{|K(e^{j\theta})|^2 S_{xx}(e^{j\theta})}{\nu} \; : \; \text{b/Hz} \quad (64)$$

$$r = \frac{|\Omega|}{2\pi}$$

$$\Omega = \{\theta: |K(e^{j\theta})|^2 S_{xx}(e^{j\theta}) > \nu\} \quad (65)$$

$$R = \int_{-\pi}^{\pi} \max\left[\frac{1}{2}\log_2 \frac{|K(e^{j\theta})|^2 S_{xx}(e^{j\theta})}{\nu}, 0\right]\frac{d\theta}{2\pi} \quad (66)$$

$$D = \int_{-\pi}^{\pi} \min\{[1 - |K(e^{j\theta})|^2]S_{xx}(e^{j\theta}) + \nu$$

$$S_{xx}(e^{j\theta})\}\frac{d\theta}{2\pi}. \quad (67)$$

These formulas are the asymptotic versions of the finite-dimensional formulas. Their interpretation is the same as illustrated in Fig. 3, with index i replaced by frequency θ.

VII. ADAPTIVE CANONICAL COORDINATES

The story of adaptive canonical correlations may be told as follows. Assume that the covariance matrix $\mathbf{R}_{zz}$ is unknown. In place of it, we have M independent snapshots of $\mathbf{z}$ assembled into the matrix $\mathbf{Z}$

$$\mathbf{Z} = [\mathbf{z}_1 \;\; \mathbf{z}_2 \; \cdots \; \mathbf{z}_M] = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_M \\ \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}. \quad (68)$$

Then, $\hat{\mathbf{R}}_{zz}$ is a crude estimate of the covariance of $\mathbf{z}$:

$$\hat{\mathbf{R}}_{zz} = \mathbf{ZZ}^{\mathrm{T}} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} [\mathbf{X}^{\mathrm{T}} \quad \mathbf{Y}^{\mathrm{T}}] = \begin{bmatrix} \mathbf{XX}^{\mathrm{T}} & \mathbf{XY}^{\mathrm{T}} \\ \mathbf{YX}^{\mathrm{T}} & \mathbf{YY}^{\mathrm{T}} \end{bmatrix}$$

$$= \begin{bmatrix} \hat{\mathbf{R}}_{xx} & \hat{\mathbf{R}}_{xy} \\ \hat{\mathbf{R}}_{yx}^{\mathrm{T}} & \hat{\mathbf{R}}_{yy} \end{bmatrix}. \tag{69}$$

From the estimated covariance matrix $\hat{\mathbf{R}}_{zz}$, we may determine the adaptive Wiener filter $\hat{\mathbf{W}}$ that minimizes

$$\mathrm{tr}\, \hat{\mathbf{Q}}_{xx} = \mathrm{tr} \sum_{i=1}^{M} (\mathbf{x}_i - \mathbf{W}\mathbf{y}_i)(\mathbf{x}_i - \mathbf{W}\mathbf{y}_i)^{\mathrm{T}} \tag{70}$$

which is the sum of squared errors between the experimental source vectors $\mathbf{x}_i$ and the experimental estimates $\mathbf{W}\mathbf{y}_i$. The solution is

$$\hat{\mathbf{W}} = \hat{\mathbf{R}}_{xy}\hat{\mathbf{R}}_{yy}^{-1} \tag{71}$$

$$\min \hat{\mathbf{Q}}_{xx} = \hat{\mathbf{R}}_{xx} - \hat{\mathbf{R}}_{xy}\hat{\mathbf{R}}_{yy}^{-1}\hat{\mathbf{R}}_{yx}. \tag{72}$$

No explicit identification or equalization of the channel model takes place. Of course, $\hat{\mathbf{G}} = (\hat{\mathbf{R}}_{xy})^{\mathrm{T}}\hat{\mathbf{R}}_{xx}^{-1}$ is what we would mean by the *channel model* for generating $\mathbf{y} = \mathbf{Gx} + \mathbf{n}$, but there is no need to explicitly estimate it. If the filter is to be reduced in rank and quantized, then $\hat{\mathbf{W}}$ will be resolved into adaptive canonical coordinates according to

$$\hat{\mathbf{W}} = \hat{\mathbf{R}}_{xx}^{1/2}\hat{\mathbf{F}}\hat{\mathbf{K}}\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{R}}_{yy}^{-1/2} \tag{73}$$

$$\hat{\mathbf{Q}}_{xx} = \hat{\mathbf{R}}_{xx}^{1/2}\hat{\mathbf{F}}(\mathbf{I} - \hat{\mathbf{K}}\hat{\mathbf{K}}^{\mathrm{T}})\hat{\mathbf{F}}^{\mathrm{T}}\hat{\mathbf{R}}_{xx}^{\mathrm{T}/2} \tag{74}$$

$$\hat{\mathbf{F}}\hat{\mathbf{K}}\hat{\mathbf{G}}^{\mathrm{T}} = \hat{\mathbf{C}}_{xy} = \hat{\mathbf{R}}_{xx}^{-1/2}\hat{\mathbf{R}}_{xy}\hat{\mathbf{R}}_{yy}^{-\mathrm{T}/2}. \tag{75}$$

Therefore, the estimated covariance matrix $\hat{\mathbf{R}}_{zz}$, which is constructed from snapshots of the *measurement* and the *source*, is used to determine the estimated coherence matrix $\hat{\mathbf{C}}_{xy}$, which is SVD'd to produce the adaptive Wiener filter in canonical coordinates. In these coordinates, the adaptive canonical filter $\hat{\mathbf{K}}$ may be reduced in rank, and the canonical vector $\mathbf{v}$ may be quantized.

These results solve the problem of designing least sum of squares adaptive filters for transform coding, filtering, and quantizing. They may also be used to experimentally test for linear independence by testing the experimental direction cosines $\hat{k}^2(i)$.

VIII. CONCLUSIONS

Canonical correlations produce insightful forms for old, familiar filtering formulas. They illustrate that canonical measurement coordinates are the natural coordinates to use when solving reduced-rank and quantized filtering problems. Squared canonical correlations determine which canonical measurement coordinates bring enough information to be retained and quantized. Typically, the canonical measurement coordinates would be computed at the transmitter using the whitener and orthogonal transform coder, reduced in dimension by purging those with small squared canonical correlation, and then quantized with variable bit allocations for transmission or storage. At the receiver, these quantized variables would be decoded with an orthogonal transform decoder and colored to produce the estimated source vector.

This might be called a noise-rejecting quantizer [12] because the quantizer is designed to filter out noise and produce a low-variance quantized estimate of the source vector. The results clarify the low-rank filters studied in [8] and [11]. This scheme can be made data adaptive by constructing the estimated correlation matrix for the source and measurement vectors from experimental data and proceeding as if the estimated covariance matrix were the true one. This procedure solves the least squares problem of minimizing the sum of squared errors between the filtered experimental measurements and the experimental source vectors, generalizing the rank reduction and bit allocation problems studied in [9] to the case where experimental data, and not second-order information, is given.

REFERENCES

[1] H. Hotelling, "Analysis of a complex pair of statistical variables into principle components," *J. Educ. Psychol.*, vol. 24, pp. 417–441, 498–520, 1933.
[2] ——, "Relations between two sets of variates," *Biometrika*, vol. 28, pp. 321–377, 1936.
[3] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis.* New York: Wiley, 1958.
[4] M. L. Eaton, *Multivariate Statistics: A Vector Space Approach.* New York: Wiley, 1983.
[5] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis.* London, U.K.: Academic, 1979, 1995.
[6] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.
[7] ——, "Coding theorem for a discrete source with a fidelity criterion," in *IRE Nat. Conv. Rec.*, 1959, pt. 4, pp. 142–163.
[8] L. L. Scharf, *Statistical Signal Processing.* Reading, MA: Addison-Wesley, 1991, pp. 330–331.
[9] L. L. Scharf and D. W. Tufts, "Rank reduction for modeling stationary signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 350–355, Mar. 1987.
[10] A. Segall, "Bit allocations and encoding for vector sources," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 162–169, Mar. 1976.
[11] J. S. Goldstein and I. S. Reed, "Reduced-rank adaptive filtering," in *Proc. ICASSP*, Atlanta, GA, 1996.
[12] S. D. Voran and L. L. Scharf, "Memoryless scalar quantizers for noisy environments," Nat. Sci. Foundation Cent. Optoelectron. Comput. Syst., Univ. Colorado, Boulder, Tech. Rep., 1989.

Louis L. Scharf (F'86) received the Ph.D. degree in electrical engineering in 1969 from the University of Washington, Seattle.

From 1969 to 1971, he was a Member of the Technical Staff at Honeywell's Marine Systems Center, Seattle. He served as Professor of Electrical Engineering and Statistics at Colorado State University, Fort Collins, from 1971 to 1981. From 1982 to 1985, he was Professor and Chair of Electrical and Computer Engineering at the University of Rhode Island, Kingston. He is currently Professor of Electrical and Computer Engineering at the University of Colorado, Boulder, where he teaches and conducts research in signal processing. In 1974, he was a Visiting Associate Professor at Duke University, Durham, NC. In 1977, he was a Member of the Technical Staff at the CNRS Laboratoire des Signaux et Systemes, Gif-sur-Yvette, France, and Professeur Associe at the University of South Paris, Orsay, France. In 1981, he was a Visiting Professor at Ecole Nationale Superiere des Telecommunications, Paris, France, and at the University of La Plata, Buenos Aires, Argentina. He was a Visiting Professor at Institut Eurecom, Sophia-Antipolis, France, in 1992.

Prof. Scharf is a past member of the ASSP ADCOM. He has served on the Editorial Board of *Signal Processing* and is a past Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was Technical Program Chairman of the IEEE International Conference on Acoustics, Speech, and Signal Processing in 1980. He currently serves on the Asilomar Conference Board. In 1994, he served as a Distinguished Lecturer for the IEEE Signal Processing Society, and in 1995, he received the Society's Technical Achievement Award.

John K. Thomas received the B.S.E.E. (Hons.) degree in 1985 and the M.S.E.E. degree in 1987, both from the University of Arkansas, Fayetteville, and the Ph.D. degree in 1996 in electrical engineering from the University of Colorado, Boulder, with specialization in statistical signal processing.

In 1986, he was a Consultant to the Jet Propulsion Laboratory (JPL), Pasadena, CA, working on the Shuttle Imaging Radar (SIR-B) radargrammetry experiment. From 1987 to 1992, he worked at Vexcel Corporation, Boulder, where he was involved in development of image/signal processing algorithms for radar signal analysis and reconstruction. He is a Co-founder and Vice President of Data Fusion Corporation, where he is involved in the development of technology for applications in communications, surveillance, target tracking, and recognition.

Dr. Thomas is a member of Eta Kappa Nu and Tau Beta Pi.