Two-Channel Constrained Least Squares Problems: Solutions Using Power Methods and Connections with Canonical Coordinates

Ali Pezeshki, *Student Member, IEEE*, Louis L. Scharf, *Fellow, IEEE*,
Mahmood R. Azimi-Sadjadi, *Senior Member, IEEE*, and Yingbo Hua, *Fellow, IEEE*

Abstract—The problem of two-channel constrained least squares (CLS) filtering under various sets of constraints is considered, and a general set of solutions is derived. For each set of constraints, the solution is determined by a coupled (asymmetric) generalized eigenvalue problem. This eigenvalue problem establishes a connection between two-channel CLS filtering and transform methods for resolving channel measurements into canonical or half-canonical coordinates. Based on this connection, a unified framework for reduced-rank Wiener filtering is presented. Then, various representations of reduced-rank Wiener filters in canonical and half-canonical coordinates are introduced. An alternating power method is proposed to recursively compute the canonical coordinate and half-canonical coordinate mappings. A deflation process is introduced to extract the mappings associated with the dominant coordinates. The correctness of the alternating power method is demonstrated on a synthesized data set, and conclusions are drawn.

Index Terms—Alternating power method, canonical coordinates, constrained least squares, generalized eigenvalue problem, half-canonical coordinates, reduced-rank Wiener filtering, rank reduction, SVD, two-channel least squares.

I. INTRODUCTION

TWO-CHANNEL problems find numerous applications in signal processing, communication, sonar, radar, and sensor fusion. In filtering and communication, one of the channels (the $\mathbf{x}$-channel) contains the unobserved source variables to be estimated, and the other channel (the $\mathbf{y}$-channel) contains the observed measurement variables. In radar and sonar, the two channels may be the outputs of two subarrays in space or the outputs over two subintervals in time. In sensor fusion, the channels correspond to different sensory measurements of the same process. The problem we pose in this paper is one of estimating a linear function of variables in one channel from a linear combination of variables in the other, under constraints. This is the idea behind constrained least squares (CLS) filtering

and canonical correlation analysis [1]–[5]. By imposing various constraints, we aim to further illuminate the essence of CLS filtering in various important coordinate systems and establish connections between CLS filtering, mutual information, information rate, and linear dependence.

The most general form of two-channel linear least squares is $\min_{\mathrm{W},\mathrm{D}} J$, where J is the quadratic function

$$J = \mathrm{tr}\{E[(\mathrm{W}^T\mathbf{x} - \mathrm{D}^T\mathbf{y})(\mathrm{W}^T\mathbf{x} - \mathrm{D}^T\mathbf{y})^T]\}. \qquad (1)$$

The matrix $\mathrm{W}^T \in \mathrm{R}^{m \times m}$ is a linear map that takes $\mathbf{x} \in \mathrm{R}^m$ into R^m, and $\mathrm{D}^T \in \mathrm{R}^{m \times n}$ is a linear map that takes $\mathbf{y} \in \mathrm{R}^n$ into $\mathrm{R}^m, m \leq n$. This problem, without any constraints on W and D, yields the trivial solution $\mathrm{W} = 0, \mathrm{D} = 0$. However, we show in this paper that under special constraints the optimization problem (1) formulates the canonical coordinate decomposition [1]–[5], half-canonical coordinate decomposition [6], [7], or Programmable Canonical Correlation Analysis (PCCA) coordinate decomposition [8], [9]. The first two coordinate systems are important for optimal reduced-rank Wiener filtering [4]–[7], [10]–[13] and the third for adaptive source separation [8], [9]. Canonical coordinates are also optimal for analyzing information rate and capacity of Gaussian communication channels and for decomposing the linear dependence and mutual information between two data channels [4], [5].

The first goal of this paper is to derive a general set of solutions to the two-channel CLS problem, clarifying in the process the connections between two-channel CLS filtering, various canonical coordinate systems, and reduced-rank Wiener filtering. A unified framework is established for deriving various optimal reduced-rank Wiener filters. Depending on the criterion, either canonical coordinates or half-canonical coordinates are optimal for rank reduction [4]–[7], [10], [13].

The second goal is to develop a simple method for recursively computing the canonical coordinates and half-canonical coordinates required for reduced-rank Wiener filtering. The conventional methods of finding canonical coordinates [4], [5] and half-canonical coordinates [6], [7] for reduced-rank Wiener filters involve computation of square-root-inverses of covariance matrices followed by a singular value decomposition (SVD) [14] of a coherence matrix. For low-rank modeling, only a few dominant coordinates and the corresponding correlations are required. However, the conventional methods do not offer a simple way to recursively compute the canonical or half-canonical coordinates, one by one. In addition, they do

not allow an easy update of the coordinates in time, as new observations for the channels become available. An alternating power method for computing the reduced-rank Wiener filter has been reported in [10]. This method is a generalization of an Iterative Quadratic Minimum Distance (IQMD) algorithm [12]. It may also be viewed as a generalization of power iterations [14]–[17]. In this paper, we present a new alternating power method, with deflation, to recursively compute canonical coordinates, half-canonical coordinates, and reduced-rank Wiener filters, and also to update them in time. The method may be viewed as a *two-step* decomposition of the standard power method [14]–[17] as it solves a coupled (asymmetric) generalized eigenvalue problem through power iterations. The algorithm does not require any matrix square-roots or inverses. One of the algorithms presented in this paper (an alternating *block* power method) has been reported in [9] to solve the PCCA problem. This algorithm may also be generalized to incorporate deflation by blocks.

II. TWO-CHANNEL CLS PROBLEMS AND SOLUTIONS

Consider the two-channel problem of Fig. 1, with two random vectors, $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$, $m \leq n$. If m were greater than n, then we would simply reverse the roles of $\mathbf{x}$ and $\mathbf{y}$. The interpretation of $\mathbf{x}$ and $\mathbf{y}$ depends on the application. In communications, $\mathbf{x}$ is the information-bearing symbol vector, and $\mathbf{y}$ is a filtered version of $\mathbf{x}$, observed in noise. In radar or sonar, $\mathbf{x}$ and $\mathbf{y}$ may be snapshots of subarrays or consecutive vectors of sampled data. In sensor fusion, $\mathbf{x}$ and $\mathbf{y}$ may be different measurements of the same process. Usually, the problem is to design the filters W^T and D^T so that the linear dependence between $\mathbf{x}$ and $\mathbf{y}$ is carried in pairwise dependence between the elements of $\mathbf{u}$ and $\mathbf{v}$ in Fig. 1. When $\mathbf{x}$ and $\mathbf{y}$ are mutually Gaussian, this linear dependence uniquely characterizes mutual information.

Assume that $\mathbf{x}$ and $\mathbf{y}$ have zero means and share the composite covariance matrix

$$R_{zz} = E[\mathbf{z}\mathbf{z}^T] = E\left[\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \begin{pmatrix} \mathbf{x}^T & \mathbf{y}^T \end{pmatrix}\right]$$
$$= \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix}. \quad (2)$$

The two-channel CLS problem is defined as $\min_{W,D} J$, subject to constraints on W and D, where J is the scalar objective function

$$J = E[\|W^T\mathbf{x} - D^T\mathbf{y}\|^2]$$
$$= \text{tr}\{E[(W^T\mathbf{x} - D^T\mathbf{y})(W^T\mathbf{x} - D^T\mathbf{y})^T]\}$$
$$= \text{tr}\{W^T R_{xx} W - W^T R_{xy} D - D^T R_{yx} W + D^T R_{yy} D\}. \quad (3)$$

The matrices $W \in \mathbb{R}^{m \times m}$ and $D \in \mathbb{R}^{n \times m}$ have equal column dimensions, $\text{tr}\{\ \}$ denotes trace of a matrix, and $E[\ ]$ denotes expectation. By completing the square, we may rewrite J as

$$J = \text{tr}\{W^T Q_{xx} W$$
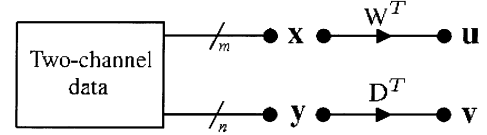$$+ (R_{yy}D - R_{yx}W)^T R_{yy}^{-1}(R_{yy}D - R_{yx}W)\} \quad (4)$$

Fig. 1.   Two-channel problem.

where $Q_{xx} = R_{xx} - R_{xy}R_{yy}^{-1}R_{yx}$ is the Schur complement of R_{xx}. Alternatively

$$J = \text{tr}\{D^T Q_{yy} D$$
$$+ (R_{xx}W - R_{xy}D)^T R_{xx}^{-1}(R_{xx}W - R_{xy}D)\} \quad (5)$$

where $Q_{yy} = R_{yy} - R_{yx}R_{xx}^{-1}R_{xy}$ is the Schur complement of R_{yy}. The virtue of these formulas is that they suggest an alternating sequence of approximations to W and D. For example, for given W in (4), J has minimum $W^T Q_{xx} W$ for $D = R_{yy}^{-1}R_{yx}W$. For given D in (5), J has minimum $D^T R_{yy} D$ for $W = R_{xx}^{-1}R_{xy}D$. This basic idea of iterating between W and D will be a consistent theme of our methods, even when constraints are imposed. Let us now consider the constraints.

A. Case 1: Canonical Coordinates

Referring to Fig. 1, the objective is to whiten $\mathbf{u} = W^T\mathbf{x}$ and $\mathbf{v} = D^T\mathbf{y}$, and diagonally cross-correlate them, while minimizing J. Thus, the constraints are

$$R_{uu} = W^T R_{xx} W = I, \quad R_{vv} = D^T R_{yy} D = I, \quad \text{and}$$
$$R_{uv} = W^T R_{xy} D = \Sigma = \text{diag}[\sigma_1, \sigma_2, \ldots, \sigma_m] \quad (6)$$

where I is the $m \times m$ identity matrix. The diagonal matrix Σ is not known *a priori*. However, it may be assumed, without loss of generality, that the diagonal elements of Σ are arranged in descending order. That is

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m > 0. \quad (7)$$

The coordinates $\mathbf{u}$ and $\mathbf{v}$ are said to be canonical because any function of R_{zz} that is invariant to nonsingular transformations of $\mathbf{x}$ and $\mathbf{y}$ must be a function of Σ [3]. We will have more to say about canonical coordinates as we proceed with our development.

Using the method of Lagrange multipliers, the constrained minimization problem may be written as $\min_{W,D} J_C$, where J_C is the scalar objective function

$$J_C = \text{tr}\{W^T Q_{xx} W + (R_{yy}D - R_{yx}W)^T R_{yy}^{-1}$$
$$\times (R_{yy}D - R_{yx}W)\} + \text{tr}\{(W^T R_{xx} W - I)\Lambda_1\}$$
$$+ \text{tr}\{(D^T R_{yy} D - I)\Lambda_2\} + 2\text{tr}\{(W^T R_{xy} D - \Sigma)\Lambda_3\} \quad (8)$$

or equivalently as

$$J_C = \text{tr}\{D^T Q_{yy} D + (R_{xx}W - R_{xy}D)^T R_{xx}^{-1}$$
$$\times (R_{xx}W - R_{xy}D)\} + \text{tr}\{(W^T R_{xx} W - I)\Lambda_1\}$$
$$+ \text{tr}\{(D^T R_{yy} D - I)\Lambda_2\} + 2\text{tr}\{(W^T R_{xy} D - \Sigma)\Lambda_3\} \quad (9)$$

where Λ_1, Λ_2, and Λ_3 are $m \times m$ Lagrange multipliers. It can be verified easily that the trace constraints in (8) and (9) indeed impose the actual constraints in (6). Taking the derivative of J_C with respect to W and D and setting the results to zero yields the coupled equations

$$R_{xy}D(I - \Lambda_3) = R_{xx}W(I + \Lambda_1) \tag{10}$$
$$R_{yx}W(I - \Lambda_3) = R_{yy}D(I + \Lambda_2). \tag{11}$$

Pre-multiplying (10) by W^T and (11) by D^T yields the coupled equations

$$W^T R_{xy}D(I - \Lambda_3) = W^T R_{xx}W(I + \Lambda_1)$$
$$D^T R_{yx}W(I - \Lambda_3) = D^T R_{yy}D(I + \Lambda_2). \tag{12}$$

At the solution, the Lagrange multipliers force the constraints in (6), and thus, (12) reduces to

$$\Sigma(I - \Lambda_3) = (I + \Lambda_1)$$
$$\Sigma(I - \Lambda_3) = (I + \Lambda_2) \tag{13}$$

which implies that $\Lambda_1 = \Lambda_2$, assuming $I - \Lambda_3$ is nonsingular. Using (13), we may rewrite (10) and (11) as the coupled system

$$R_{xy}D = R_{xx}W\Sigma$$
$$R_{yx}W = R_{yy}D\Sigma. \tag{14}$$

Equivalently, we may combine these two equations as

$$R_{xy}R_{yy}^{-1}R_{yx}W = R_{xx}W\Sigma^2$$
$$R_{yx}R_{xx}^{-1}R_{xy}D = R_{yy}D\Sigma^2. \tag{15}$$

The set of equations in (14) and (15) are key results, for they characterize the solutions for W and D to minimize J under the constraints of (6). These solutions, in turn, produce the coordinates $\mathbf{u} = W^T\mathbf{x}$ and $\mathbf{v} = D^T\mathbf{y}$ and correlations $\Sigma = W^T R_{xy}D$ of $\mathbf{x}$ and $\mathbf{y}$. The equations in (15) are (symmetric) generalized eigenvalue problems for W and D, with Σ^2 the shared eigenvalue matrix. We refer to (15) as a "coupled (symmetric) generalized eigenvalue problem." Correspondingly, (14) may be viewed as a coupled (asymmetric) generalized eigenvalue problem. In Section V, we will give an alternating power method for solving (14). The advantage of solving (14) rather than (15) is that no inverse of R_{xx} or R_{yy} is required. In Section III, we will establish that $\mathbf{u} = W^T\mathbf{x}$ and $\mathbf{v} = D^T\mathbf{y}$ are, indeed, the standard canonical coordinates of $\mathbf{x}$ and $\mathbf{y}$.[1]

B. Case 2: Half-Canonical Coordinates

Referring again to Fig. 1, the objective is now to whiten $\mathbf{v} = D^T\mathbf{y}$ only, and diagonally cross-correlate $\mathbf{u} = W^T\mathbf{x}$ and $\mathbf{v}$, while minimizing J. The constraints in this case are

$$W^T W = I, \quad R_{vv} = D^T R_{yy}D = I, \quad \text{and}$$
$$R_{uv} = W^T R_{xy}D = \Sigma = \text{diag}[\sigma_1, \sigma_2, \ldots, \sigma_m]. \tag{16}$$

---

[1]If, instead of the set of constraints in (6), we had only constrained the *diagonal* elements of $W^T R_{xx}W$ and $D^T R_{yy}D$ to be unity, we could still have obtained the generalized eigenvalue problems in (14) and (15) for W and D. However, solving (14) and (15) under this new set of constraints would not have guaranteed that $W^T R_{xx}W = I$, $D^T R_{yy}D = I$, or $W^T R_{xy}D$ diagonal. Consequently, $\mathbf{u} = W^T\mathbf{x}$ and $\mathbf{v} = D^T\mathbf{y}$ would not have been the canonical coordinates of $\mathbf{x}$ and $\mathbf{y}$.

The diagonal matrix Σ is not known *a priori*. However, similar to Case 1, it is assumed that its diagonal elements satisfy (7). The coordinates $\mathbf{u}$ and $\mathbf{v}$ are said to be half-canonical because any function of R_{zz} that is invariant to nonsingular transformations of $\mathbf{y}$ only must be a function of Σ. Since the matrix W is square and full rank, it follows that $WW^T = I$, as well.

The Lagrange multiplier method for the objective function in (3) and constraints in (16) yields the coupled equations

$$R_{xy}D(I - \Lambda_3) = R_{xx}W + W\Lambda_1 \tag{17}$$
$$R_{yx}W(I - \Lambda_3) = R_{yy}D(I + \Lambda_2) \tag{18}$$

where Λ_1, Λ_2, and Λ_3 are $m \times m$ Lagrange multipliers. Pre-multiplying (17) by W^T and (18) by D^T and enforcing the constraints in (16) yields

$$\Sigma(I - \Lambda_3) = W^T R_{xx}W + \Lambda_1$$
$$\Sigma(I - \Lambda_3) = I + \Lambda_2. \tag{19}$$

Using these solutions for Λ_1 and Λ_2 and assuming that $(I - \Lambda_3)$ is nonsingular, we may rewrite (17) and (18) as the coupled equations

$$R_{xy}D = W\Sigma$$
$$R_{yx}W = R_{yy}D\Sigma \tag{20}$$

or equivalently as

$$R_{xy}R_{yy}^{-1}R_{yx}W = W\Sigma^2$$
$$R_{yx}R_{xy}D = R_{yy}D\Sigma^2. \tag{21}$$

The set of equations in (20) and (21) are key results, for they characterize the solutions for W and D to minimize J under the constraints of (16). Equation (21) is a coupled (symmetric) generalized eigenvalue problem for W and D, with Σ^2 the shared eigenvalue matrix. Correspondingly, (20) is a coupled (asymmetric) generalized eigenvalue problem. In Section III, we will establish that $\mathbf{u}$ and $\mathbf{v}$ are, indeed, half-canonical coordinates of $\mathbf{x}$ and $\mathbf{y}$, and in Section V, we will give an alternating power method for solving (20).

C. Case 3: Programmable Canonical Correlation Analysis

In this case, the objective is to whiten $\mathbf{u} = W^T\mathbf{x}$ and $\mathbf{v} = D^T\mathbf{y}$, while minimizing J [8]. The constraints in this case are

$$W^T R_{xx}W = I, \quad \text{and} \quad D^T R_{yy}D = I. \tag{22}$$

Comparing to Case 1, the constraint on the diagonal cross-covariance is not imposed. The term PCCA [8] suggests that $\mathbf{u} = W^T\mathbf{x}$ and $\mathbf{v} = D^T\mathbf{y}$ are canonical coordinates *programmed* by W and D. We will show that this constrained minimization problem is not well-posed because the solution given in [8] is really the unique solution for Case 1, which happens to be just one of an infinite number of solutions to the PCCA problem.

Contrary to the two-channel CLS problem in Case 1, in PCCA the constraint that $W^T R_{xy}D$ be diagonal is relaxed. Nonetheless, in [8], the coupled (symmetric) generalized eigenvalue problem (15) is solved for W and D. Thus, the solution of [8] for the PCCA coordinates is actually a solution for Case 1, canonical coordinates, and not for the problem posed. We wish to clarify this point by contrasting the two solutions.

In the case of PCCA, the Lagrange multiplier method for (3), with constraints in (22), yields the equations

$$R_{xy}D = R_{xx}W(I + \Lambda_1)$$
$$R_{yx}W = R_{yy}D(I + \Lambda_2) \quad (23)$$

where Λ_1 and Λ_2 are $m \times m$ Lagrange multipliers. At the solution, (22) is satisfied, and thus, we have

$$(I + \Lambda_1) = (I + \Lambda_2)^T = W^T R_{xy} D. \quad (24)$$

Without constraints on $W^T R_{xy} D$, there is no way to determine Λ_1 and Λ_2. Nonetheless, we may proceed with them undetermined to rewrite (23) as

$$R_{xy}D = R_{xx}W(I + \Lambda_1)$$
$$R_{yx}W = R_{yy}D(I + \Lambda_1)^T \quad (25)$$

or equivalently as

$$R_{xy}R_{yy}^{-1}R_{yx}W = R_{xx}W(I + \Lambda_1)(I + \Lambda_1)^T$$
$$R_{yx}R_{xx}^{-1}R_{xy}D = R_{yy}D(I + \Lambda_1)^T(I + \Lambda_1). \quad (26)$$

Since $(I + \Lambda_1)(I + \Lambda_1)^T$ is not necessarily diagonal, neither equation in (26) is a generalized eigenvalue problem. However, noting that $(I + \Lambda_1)(I + \Lambda_1)^T$ and $(I + \Lambda_1)^T(I + \Lambda_1)$ are symmetric and have the same eigenvalues, we may consider the eigenvalue decompositions

$$(I + \Lambda_1)(I + \Lambda_1)^T = AQ^2A^T$$
$$(I + \Lambda_1)^T(I + \Lambda_1) = BQ^2B^T \quad (27)$$

where A and B are orthogonal matrices, i.e., $A^T A = AA^T = I, B^T B = BB^T = I$, and Q^2 is a diagonal eigenvalue matrix. Using (27), we may rewrite (26) as

$$R_{xy}R_{yy}^{-1}R_{yx}(WA) = R_{xx}(WA)Q^2$$
$$R_{yx}R_{xx}^{-1}R_{xy}(DB) = R_{yy}(DB)Q^2. \quad (28)$$

This equation is a coupled (symmetric) generalized eigenvalue problem for WA and DB. The corresponding coupled (asymmetric) generalized eigenvalue problem is

$$R_{xy}(DB) = R_{xx}(WA)Q$$
$$R_{yx}(WA) = R_{yy}(DB)Q. \quad (29)$$

These generalized eigenvalue problems for WA and DB are the same as those for W and D in Case 1, i.e., (14) and (15). Therefore, we may write

$$W_{PCCA}A = W_{CC} \quad \text{and} \quad D_{PCCA}B = D_{CC} \quad (30)$$

where subscript CC stands for Canonical Coordinates (Case 1) and PCCA for Programmable Canonical Correlation Analysis (Case 3). That is, in the PCCA case, solving (29) determines W and D up to unknown (right) orthogonal matrices A and B.

The ambiguity in the solution, however, does not affect the minimum value of (3). The solution for Case 1 ($W = W_{CC}$ and $D = D_{CC}$) would solve the PCCA problem, corresponding to $W^T R_{xx} W = W_{CC}^T R_{xx} W_{CC} = I, D^T R_{yy} D = D_{CC}^T R_{yy} D_{CC} = I$, and $W^T R_{xy} D = W_{CC}^T R_{xy} D_{CC} = \Sigma$, with Σ diagonal. However, so would $W = W_{CC}M$ and $D = D_{CC}M$, where $M \in R^{m \times m}$ is any orthogonal matrix. In this latter case, $W^T R_{xx} W = M^T W_{CC}^T R_{xx} W_{CC} M = I, D^T R_{yy} D = M^T D_{CC}^T R_{yy} D_{CC} M = I$, but $W^T R_{xy} D = M^T W_{CC}^T R_{xy} D_{CC} M = M^T \Sigma M$ is not diagonal. In summary, a nonunique solution to the PCCA problem, as originally posed in [8], is made unique by imposing the additional constraint of Case 1, which we have called the canonical coordinates case.

III. Two-Channel CLS and Various Canonical Coordinate Systems

In Section II, the two-channel CLS problems were given what might have appeared to be arbitrary names. In this section, we legitimize these names by establishing the connections between the two-channel CLS problems in Cases 1 and 2 and various well-established canonical coordinate systems.

A. Canonical Coordinates

Consider the constraints in (6) for the two-channel CLS problem of Case 1. These constraints may be rewritten as

$$F^T F = FF^T = I, G^T G = I, \quad \text{and}$$
$$F^T R_{xx}^{-1/2} R_{xy} R_{yy}^{-T/2} G = \Sigma \quad (31)$$

where $F^T = W^T R_{xx}^{1/2}, G^T = D^T R_{yy}^{1/2}, R_{xx}^{-1/2} R_{xx} R_{xx}^{-T/2} = I$, and $R_{xx}^{1/2} R_{xx}^{T/2} = R_{xx}$. Thus, $F\Sigma G^T = R_{xx}^{-1/2} R_{xy} R_{yy}^{-T/2} = C$ is a thin SVD [14] of the *coherence* matrix $C = R_{xx}^{-1/2} R_{xy} R_{yy}^{-T/2}$ [4]. The thin SVD of a rectangular matrix A is a trimmed-down version of the SVD of A, in which the zero singular values of A and their corresponding singular vectors are discarded in forming the SVD.

Using (31), we may now write

$$\begin{bmatrix} W^T & 0 \\ 0 & D^T \end{bmatrix} \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix} \begin{bmatrix} W & 0 \\ 0 & D \end{bmatrix}$$
$$= \begin{bmatrix} F^T & 0 \\ 0 & G^T \end{bmatrix} \begin{bmatrix} R_{xx}^{-1/2} & 0 \\ 0 & R_{yy}^{-1/2} \end{bmatrix} \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix}$$
$$\times \begin{bmatrix} R_{xx}^{-T/2} & 0 \\ 0 & R_{yy}^{-T/2} \end{bmatrix} \begin{bmatrix} F & 0 \\ 0 & G \end{bmatrix} = \begin{bmatrix} I & \Sigma \\ \Sigma^T & I \end{bmatrix}. \quad (32)$$

Define the *canonical coordinates* of the composite vector $\mathbf{z} = [\mathbf{x}^T \ \mathbf{y}^T]^T$ as [4], [5]

$$\mathbf{w} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} F^T & 0 \\ 0 & G^T \end{bmatrix} \begin{bmatrix} R_{xx}^{-1/2} & 0 \\ 0 & R_{yy}^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \quad (33)$$

and note that they share the composite covariance matrix

$$R_{ww} = E\left[\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} \begin{pmatrix} \mathbf{u}^T & \mathbf{v}^T \end{pmatrix} \right] = \begin{bmatrix} I & \Sigma \\ \Sigma^T & I \end{bmatrix}. \quad (34)$$

Thus, the diagonal matrix $\Sigma = F^T C G = W^T R_{xy} D$ is the *canonical correlation* matrix of canonical correlations $E[u_i v_i] = \sigma_i$. Correspondingly, the matrices

$$W^T = F^T R_{xx}^{-1/2}, \quad \text{and} \quad D^T = G^T R_{yy}^{-1/2} \quad (35)$$

map $\mathbf{x}$ and $\mathbf{y}$ to their respective canonical coordinates $\mathbf{u} = W^T \mathbf{x}$ and $\mathbf{v} = D^T \mathbf{y}$. Thus, we call W^T and D^T, which are solutions to the two-channel CLS problem of Case 1, canonical coordinate maps.

Using (31) and (35), along with the cyclic property of trace, the minimum value of J for Case 1, which is denoted by J_{uu}, may be written as

$$
\begin{aligned}
J_{uu} &= \text{tr}\left\{ F^T R_{xx}^{-1/2} Q_{xx} R_{xx}^{-T/2} F \right\} \\
&\quad + \text{tr}\left\{ I - \Sigma - \Sigma^T + F^T R_{xx}^{-1/2} R_{xy} R_{yy}^{-1} R_{yx} R_{xx}^{-T/2} F \right\} \\
&= \text{tr}\left\{ R_{xx}^{-1/2} Q_{xx} R_{xx}^{-T/2} \right\} + \text{tr}\left\{ (I - \Sigma)(I - \Sigma^T) \right\} \\
&= \text{tr}\left\{ R_{xx}^{-1/2} Q_{xx} R_{xx}^{-T/2} \right\} + \sum_{i=1}^{m} (1 - \sigma_i)^2. \quad (36)
\end{aligned}
$$

It is interesting to compare this index with the mean-squared error (MSE) when estimating $\mathbf{u}$ from $\mathbf{v}$. From (34), the linear minimum MSE estimator of $\mathbf{u}$ from $\mathbf{v}$ is $\hat{\mathbf{u}} = \Sigma \mathbf{v}$, and its corresponding error covariance matrix is $I - \Sigma \Sigma^T$. Then, the MSE of estimating $\mathbf{u}$ from $\mathbf{v}$ is

$$
\begin{aligned}
\text{MSE}_{uu} &= \text{tr}\{ E[(\mathbf{u} - \hat{\mathbf{u}})(\mathbf{u} - \hat{\mathbf{u}})^T] \} = \text{tr}\{ (I - \Sigma \Sigma^T) \} \\
&= \sum_{i=1}^{m} (1 - \sigma_i^2) = \text{tr}\left\{ R_{xx}^{-1/2} Q_{xx} R_{xx}^{-T/2} \right\} \quad (37)
\end{aligned}
$$

which is equal to the first term on the right side of (36). Thus, the connection between J_{uu} and MSE_{uu} is

$$
\begin{aligned}
J_{uu} &= \text{MSE}_{uu} + \text{tr}\{ (I - \Sigma)(I - \Sigma)^T \} \\
&= \text{MSE}_{uu} + \sum_{i=1}^{m} (1 - \sigma_i)^2. \quad (38)
\end{aligned}
$$

We see that J_{uu} is within $\text{tr}\{(I - \Sigma)(I - \Sigma)^T\}$, an invariant for this two-channel problem, of the MSE for estimating $\mathbf{u}$ from $\mathbf{v}$, further clarifying the connection between two-channel CLS filtering and canonical correlation analysis.

B. Half-Canonical Coordinates

For the two-channel CLS problem of Case 2, the constraints in (16) may be written as

$$
\begin{aligned}
U^T U = U U^T = I, V^T V = I, \quad \text{and} \\
U^T R_{xy} R_{yy}^{-T/2} V = \Sigma \quad (39)
\end{aligned}
$$

where $U^T = W^T$, and $V^T = D^T R_{yy}^{1/2}$. Thus, $U \Sigma V^T = R_{xy} R_{yy}^{-T/2}$ is a thin SVD of the *half-coherence* matrix $R_{xy} R_{yy}^{-T/2}$. Therefore, we may write

$$
\begin{aligned}
&\begin{bmatrix} W^T & 0 \\ 0 & D^T \end{bmatrix} \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix} \begin{bmatrix} W & 0 \\ 0 & D \end{bmatrix} \\
&= \begin{bmatrix} U^T & 0 \\ 0 & V^T \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & R_{yy}^{-1/2} \end{bmatrix} \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix} \\
&\quad \times \begin{bmatrix} I & 0 \\ 0 & R_{yy}^{-T/2} \end{bmatrix} \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix} = \begin{bmatrix} W^T R_{xx} W & \Sigma \\ \Sigma^T & I \end{bmatrix}. \quad (40)
\end{aligned}
$$

Define the *half-canonical coordinates* of the composite vector $\mathbf{z} = [\mathbf{x}^T \ \mathbf{y}^T]^T$ as [6], [7]

$$\mathbf{w} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} U^T & 0 \\ 0 & V^T \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & R_{yy}^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \quad (41)$$

and note that they share the composite covariance matrix

$$R_{ww} = E\left[\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} \begin{pmatrix} \mathbf{u}^T & \mathbf{v}^T \end{pmatrix} \right] = \begin{bmatrix} W^T R_{xx} W & \Sigma \\ \Sigma^T & I \end{bmatrix}. \quad (42)$$

The diagonal matrix $\Sigma = U^T R_{xy} R_{yy}^{-T/2} V = W^T R_{xy} D$ is the *half-canonical correlation* matrix of half-canonical correlations $E[u_i v_i] = \sigma_i$. Correspondingly

$$W^T = U^T, \quad \text{and} \quad D^T = V^T R_{yy}^{-1/2} \quad (43)$$

are the matrices that map $\mathbf{x}$ and $\mathbf{y}$ to their respective half-canonical coordinates $\mathbf{u} = W^T \mathbf{x}$ and $\mathbf{v} = D^T \mathbf{y}$. Thus, we call W^T and D^T, which are solutions to the two-channel CLS problem of Case 2, half-canonical coordinate maps.

Using (39) and (43), along with the cyclic property of trace, the minimum value of J for Case 2, which is denoted by J_{xx}, may be written as

$$
\begin{aligned}
J_{xx} &= \text{tr}\{ U^T Q_{xx} U \} + \text{tr}\left\{ I - \Sigma - \Sigma^T + U^T R_{xy} R_{yy}^{-1} R_{yx} U \right\} \\
&= \text{tr}\{ Q_{xx} \} + \text{tr}\{ (I - \Sigma)(I - \Sigma^T) \} \\
&= \text{tr}\{ Q_{xx} \} + \sum_{i=1}^{m} (1 - \sigma_i)^2. \quad (44)
\end{aligned}
$$

The first term on the right side of (44) ($\text{tr}\{Q_{xx}\} = \text{tr}\{R_{xx} - R_{xy} R_{yy}^{-1} R_{yx}\} = \text{MSE}_{xx}$) is the MSE for estimating $\mathbf{x}$ from $\mathbf{y}$, using the Wiener filter $R_{xy} R_{yy}^{-1}$. Therefore, we have

$$J_{xx} = \text{MSE}_{xx} + \text{tr}\{ (I - \Sigma)(I - \Sigma)^T \}. \quad (45)$$

Thus, J_{xx} is within $\text{tr}\{(I - \Sigma)(I - \Sigma)^T\}$, an invariant for this two-channel problem, of the MSE for estimating $\mathbf{x}$ from $\mathbf{y}$, further clarifying the connection between two-channel CLS filtering and half-canonical correlation analysis.

IV. TWO-CHANNEL CLS AND REDUCED-RANK WIENER FILTERING

Our aim now is to establish the connections between two-channel CLS filters and different classes of optimal reduced-rank Wiener filters. Let $\hat{\mathbf{x}} = H(r)\mathbf{y}$ be a rank $r < m$ estimate of $\mathbf{x}$ for some rank-r linear transformation of $\mathbf{y}$. The covariance matrix of the error may be written as

$$
\begin{aligned}
Q_{xx}(r) &= E[(\mathbf{x} - H(r)\mathbf{y})(\mathbf{x} - H(r)\mathbf{y})^T] \\
&= Q_{xx} + [H - H(r)] R_{yy} [H - H(r)]^T \quad (46)
\end{aligned}
$$

where $H = R_{xy} R_{yy}^{-1}$ is the full-rank Wiener filter. The choice of coordinate system for building the optimal rank-r Wiener filter $H(r)$ depends on the measure to be optimized. In [10], common measures for reduced-rank Wiener filtering are reviewed:

$$\text{tr}\left\{ R_{xx}^{-1/2} Q_{xx}(r) R_{xx}^{-T/2} \right\}, \det\{Q_{xx}(r)\}, \quad \text{tr}\{Q_{xx}(r)\}.$$

The first two measures have been shown to be equivalent. In this section, we intend to review the connections between different classes of optimal reduced-rank Wiener filters and

further clarify their connections with canonical coordinates, half-canonical coordinates, and two-channel CLS filters. We demonstrate that canonical coordinates are optimal for rank reduction based on the first two measures, whereas half-canonical coordinates are optimal for rank reduction based on the third measure. Our results reproduce what is known from [10], but our unified method follows the line of argument given in [7] for deriving the reduced-rank Wiener filter in half-canonical coordinates. Finally, we present several equivalent implementations of reduced-rank Wiener filters in canonical and half-canonical coordinates.

A. Reduced-Rank Filtering in Canonical Coordinates

Here, the objective is to find the rank-r filter $\mathrm{H}(r)$ that minimizes the trace of the weighted error covariance matrix $\mathrm{R}_{xx}^{-1/2}\mathrm{Q}_{xx}(r)\mathrm{R}_{xx}^{-T/2}$ [10]. We note that $\mathrm{tr}\{\mathrm{R}_{xx}^{-1/2}\mathrm{Q}_{xx}(r)\mathrm{R}_{xx}^{-T/2}\} = \mathrm{tr}\{\mathrm{F}^T\mathrm{R}_{xx}^{-1/2}\mathrm{Q}_{xx}(r)\mathrm{R}_{xx}^{-T/2}\mathrm{F}\} = \mathrm{tr}\{\mathrm{Q}_{uu}(r)\}$ is the MSE for the rank-r Wiener estimator of $\mathbf{u} = \mathrm{F}^T\mathrm{R}_{xx}^{-1/2}\mathbf{x}$ from $\mathbf{v} = \mathrm{G}^T\mathrm{R}_{yy}^{-1/2}\mathbf{y}$ [5]; therefore, we denote it by $\mathrm{MSE}_{uu}(r)$:

$$
\begin{aligned}
\mathrm{MSE}_{uu}(r) &= \mathrm{tr}\left\{\mathrm{R}_{xx}^{-1/2}\mathrm{Q}_{xx}(r)\mathrm{R}_{xx}^{-T/2}\right\} \\
&= \mathrm{tr}\left\{\mathrm{R}_{xx}^{-1/2}\left(\mathrm{Q}_{xx} + \left[\mathrm{R}_{xy}\mathrm{R}_{yy}^{-1} - \mathrm{H}(r)\right]\mathrm{R}_{yy}\right.\right. \\
&\qquad \left.\left. \cdot \left[\mathrm{R}_{xy}\mathrm{R}_{yy}^{-1} - \mathrm{H}(r)\right]^T\right)\mathrm{R}_{xx}^{-T/2}\right\}.
\end{aligned}
\tag{47}
$$

Using $\mathrm{C} = \mathrm{R}_{xx}^{-1/2}\mathrm{R}_{xy}\mathrm{R}_{yy}^{-T/2}$, we may rewrite $\mathrm{MSE}_{uu}(r)$ as

$$
\begin{aligned}
\mathrm{MSE}_{uu}(r) &= \mathrm{tr}\left\{\mathrm{R}_{xx}^{-1/2}\mathrm{Q}_{xx}\mathrm{R}_{xx}^{-T/2}\right\} \\
&\quad + \mathrm{tr}\left\{\left[\mathrm{C} - \mathrm{R}_{xx}^{-1/2}\mathrm{H}(r)\mathrm{R}_{yy}^{1/2}\right]\left[\mathrm{C} - \mathrm{R}_{xx}^{-1/2}\mathrm{H}(r)\mathrm{R}_{yy}^{1/2}\right]^T\right\}.
\end{aligned}
\tag{48}
$$

The first term on the right side of this equation is fixed. Thus, minimizing $\mathrm{MSE}_{uu}(r)$ is equivalent to minimizing the second term

$$
\epsilon_{uu}^2 = \mathrm{tr}\left\{\left[\mathrm{C} - \mathrm{R}_{xx}^{-1/2}\mathrm{H}(r)\mathrm{R}_{yy}^{1/2}\right]\left[\mathrm{C} - \mathrm{R}_{xx}^{-1/2}\mathrm{H}(r)\mathrm{R}_{yy}^{1/2}\right]^T\right\}
\tag{49}
$$

which is the Frobenius norm of the matrix $\mathrm{C} - \mathrm{R}_{xx}^{-1/2}\mathrm{H}(r)\mathrm{R}_{yy}^{1/2}$. It measures the extra MSE introduced by rank reduction. The optimum choice for the rank-r Wiener filter $\mathrm{H}(r)$ is the rank-r matrix that best approximates the coherence matrix $\mathrm{C} = \mathrm{R}_{xx}^{-1/2}\mathrm{R}_{xy}\mathrm{R}_{yy}^{-T/2} = \mathrm{F}\Sigma\mathrm{G}^T$ by minimizing ϵ_{uu}^2 [18]. Thus, it is given by $\mathrm{R}_{xx}^{-1/2}\mathrm{H}(r)\mathrm{R}_{yy}^{1/2} = \mathrm{F}\Sigma(r)\mathrm{G}^T$, or

$$
\mathrm{H}(r) = \mathrm{R}_{xx}^{1/2}\mathrm{F}\Sigma(r)\mathrm{G}^T\mathrm{R}_{yy}^{-1/2}; \quad \Sigma(r) = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}
\tag{50}
$$

where $\Sigma_r = \mathrm{diag}[\sigma_1, \dots, \sigma_r]$ is the first $r \times r$ block of diagonal Σ. Using (50), the optimal value of $\mathrm{MSE}_{uu}(r)$ is

$$
\begin{aligned}
&\mathrm{MSE}_{uu}(r) \\
&= \mathrm{tr}\left\{\mathrm{R}_{xx}^{-1/2}\mathrm{Q}_{xx}\mathrm{R}_{xx}^{-T/2}\right\} + \mathrm{tr}\{\Sigma\Sigma^T - \Sigma(r)\Sigma(r)^T\} \\
&= \mathrm{MSE}_{uu} + \sum_{i=r+1}^{m}\sigma_i^2.
\end{aligned}
\tag{51}
$$

The first term on the right-hand side of (51) is the minimum MSE for a full-rank estimator of $\mathbf{u}$ from $\mathbf{v}$, and the second term is the extra MSE due to rank reduction.

We now show that the minimization of $\mathrm{MSE}_{uu}(r)$ in (47) is equivalent to minimization of the volume of the concentration ellipse $\{\mathbf{e} \in \mathrm{R}^m : \mathbf{e}^T\mathrm{Q}_{xx}(r)^{-1}\mathbf{e} = 1\}$. This volume is proportional to determinant of the error covariance matrix $\mathrm{Q}_{xx}(r)$ [7]. We may write this determinant as

$$
\begin{aligned}
\det\{\mathrm{Q}_{xx}(r)\} &= \det\left\{\mathrm{Q}_{xx} + \left[\mathrm{R}_{xy}\mathrm{R}_{yy}^{-1} - \mathrm{H}(r)\right]\right. \\
&\qquad \left. \cdot \mathrm{R}_{yy}\left[\mathrm{R}_{xy}\mathrm{R}_{yy}^{-1} - \mathrm{H}(r)\right]^T\right\} \\
&= \det\left\{\mathrm{Q}_{xx}^{1/2}\right\}\det\left\{\mathrm{I} + \left[\mathrm{Q}_{xx}^{-1/2}\mathrm{R}_{xy}\mathrm{R}_{yy}^{-T/2}\right.\right. \\
&\qquad \left. - \mathrm{Q}_{xx}^{-1/2}\mathrm{H}(r)\mathrm{R}_{yy}^{1/2}\right]\left[\mathrm{Q}_{xx}^{-1/2}\mathrm{R}_{xy}\mathrm{R}_{yy}^{-T/2}\right. \\
&\qquad \left.\left. - \mathrm{Q}_{xx}^{-1/2}\mathrm{H}(r)\mathrm{R}_{yy}^{1/2}\right]^T\right\}\det\left\{\mathrm{Q}_{xx}^{T/2}\right\}.
\end{aligned}
\tag{52}
$$

The terms $\det\{\mathrm{Q}_{xx}^{1/2}\}$ and $\det\{\mathrm{Q}_{xx}^{T/2}\}$ do not affect minimization of $\det\{\mathrm{Q}_{xx}(r)\}$ and thus may be dropped. We may then rewrite the middle determinant as

$$
\begin{aligned}
&\det\left\{\mathrm{I} + \left[\mathrm{Q}_{xx}^{-1/2}\mathrm{R}_{xy}\mathrm{R}_{yy}^{-T/2} - \mathrm{Q}_{xx}^{-1/2}\mathrm{H}(r)\mathrm{R}_{yy}^{1/2}\right]\right. \\
&\qquad \left. \times \left[\mathrm{Q}_{xx}^{-1/2}\mathrm{R}_{xy}\mathrm{R}_{yy}^{-T/2} - \mathrm{Q}_{xx}^{-1/2}\mathrm{H}(r)\mathrm{R}_{yy}^{1/2}\right]^T\right\} \\
&= \prod_{i=1}^{m}(1 + \gamma_i^2)
\end{aligned}
\tag{53}
$$

where the γ_i's are the singular values of $\mathrm{Q}_{xx}^{-1/2}\mathrm{R}_{xy}\mathrm{R}_{yy}^{-T/2} - \mathrm{Q}_{xx}^{-1/2}\mathrm{H}(r)\mathrm{R}_{yy}^{1/2}$. Thus, minimization of $\det\{\mathrm{Q}_{xx}(r)\}$ reduces to minimization of the γ_i's. Given a fixed matrix M and a rank-r matrix N of the same dimension as M, the singular values of $\mathrm{M} - \mathrm{N}$ are minimized if N is the rank-r approximation of M that is computed from the SVD of M [13], [19]. Thus, the optimal rank-r Wiener filter $\mathrm{H}(r)$ must satisfy

$$
\mathrm{Q}_{xx}^{-1/2}\mathrm{H}(r)\mathrm{R}_{yy}^{1/2} = \left[\mathrm{Q}_{xx}^{-1/2}\mathrm{R}_{xy}\mathrm{R}_{yy}^{-T/2}\right]_r
\tag{54}
$$

where the matrix $[\mathrm{Q}_{xx}^{-1/2}\mathrm{R}_{xy}\mathrm{R}_{yy}^{-T/2}]_r$ is the rank-r approximation of the matrix $\mathrm{Q}_{xx}^{-1/2}\mathrm{R}_{xy}\mathrm{R}_{yy}^{-T/2}$ computed from the SVD of $\mathrm{Q}_{xx}^{-1/2}\mathrm{R}_{xy}\mathrm{R}_{yy}^{-T/2}$. Plugging in $\mathrm{Q}_{xx} = \mathrm{R}_{xx}^{1/2}(\mathrm{I} - \mathrm{C}\mathrm{C}^T)\mathrm{R}_{xx}^{T/2}$ in the right-hand side of (54) and using the thin SVD in (31) for the coherence matrix $\mathrm{C} = \mathrm{R}_{xx}^{-1/2}\mathrm{R}_{xy}\mathrm{R}_{yy}^{-T/2} = \mathrm{F}\Sigma\mathrm{G}^T$ reduces (54) to

$$
\begin{aligned}
\mathrm{Q}_{xx}^{-1/2}\mathrm{H}(r)\mathrm{R}_{yy}^{1/2} &= [(\mathrm{I} - \mathrm{F}\Sigma\Sigma^T\mathrm{F}^T)^{-1/2}\mathrm{F}\Sigma\mathrm{G}^T]_r \\
&= [(\mathrm{F}(\mathrm{I} - \Sigma\Sigma^T)^{1/2})^{-1}\mathrm{F}\Sigma\mathrm{G}^T]_r \\
&= ([(\mathrm{I} - \Sigma\Sigma^T)^{-1/2}\Sigma\mathrm{G}^T]_r \\
&= (\mathrm{I} - \Sigma\Sigma^T)^{-1/2}\Sigma(r)\mathrm{G}^T.
\end{aligned}
\tag{55}
$$

Premultiplying (55) by $\mathrm{Q}_{xx}^{1/2} = \mathrm{R}_{xx}^{1/2}(\mathrm{I} - \mathrm{F}\Sigma\Sigma^T\mathrm{F}^T)^{1/2}$ and post-multiplying by $\mathrm{R}_{yy}^{-1/2}$ yields

$$
\begin{aligned}
\mathrm{H}(r) &= \mathrm{R}_{xx}^{1/2}(\mathrm{I} - \mathrm{F}\Sigma\Sigma^T\mathrm{F}^T)^{1/2}(\mathrm{I} - \Sigma\Sigma^T)^{-1/2}\Sigma(r)\mathrm{G}^T\mathrm{R}_{yy}^{-1/2} \\
&= \mathrm{R}_{xx}^{1/2}\mathrm{F}(\mathrm{I} - \Sigma\Sigma^T)^{1/2}(\mathrm{I} - \Sigma\Sigma^T)^{-1/2}\Sigma(r)\mathrm{G}^T\mathrm{R}_{yy}^{-1/2} \\
&= \mathrm{R}_{xx}^{1/2}\mathrm{F}\Sigma(r)\mathrm{G}^T\mathrm{R}_{yy}^{-1/2}.
\end{aligned}
\tag{56}
$$

This is the optimal rank-r Wiener filter $H(r)$ that minimizes (52) and is equivalent to the optimal rank-r Wiener filter in (50) that minimizes the measure $MSE_{uu}(r)$ in (47). Therefore, canonical coordinates are the right coordinate system for reduced-rank Wiener filtering with the determinant measure in (52) as well. The corresponding rank-r Wiener estimate of $\mathbf{x}$ from $\mathbf{y}$ is given by $\hat{\mathbf{x}} = H(r)\mathbf{y}$. The optimal value of $\det\{Q_{xx}(r)\}$ in (52) is

$$
\begin{aligned}
\det\{Q_{xx}(r)\} &= \det\{R_{xx}\}\det\{I - F\Sigma(r)\Sigma(r)^T F^T \\
&\quad - F\Sigma(r)\Sigma^T(r)F^T + F\Sigma(r)\Sigma(r)^T F^T\} \\
&= \det\{R_{xx}\}\det\{I - \Sigma(r)\Sigma(r)^T\} \\
&= \det\{R_{xx}\}\prod_{i=1}^{r}\left(1 - \sigma_i^2\right) \\
&= \det\{Q_{xx}\}\frac{1}{\prod_{i=r+1}^{m}(1 - \sigma_i^2)}
\end{aligned}
\tag{57}
$$

which is proportional to the volume of the concentration ellipse for the rank-r estimator $H(r)$.

It can be shown [4] that when the composite vector $\mathbf{z} = [\mathbf{x}^T\ \mathbf{y}^T]^T$ is normally distributed, the rate at which the rank-r estimate $\hat{\mathbf{x}}$ carries information about $\mathbf{x}$ is

$$
\begin{aligned}
R(r) &= \frac{1}{2}\log\det\{R_{xx}\} - \frac{1}{2}\log\det\{Q_{xx}(r)\} \\
&= -\frac{1}{2}\sum_{i=1}^{r}\log\left(1 - \sigma_i^2\right) = R + \frac{1}{2}\sum_{i=r+1}^{m}\log\left(1 - \sigma_i^2\right)
\end{aligned}
\tag{58}
$$

where $R = -\frac{1}{2}\sum_{i=1}^{m}\log(1 - \sigma_i^2)$ is the rate at which $\mathbf{y}$ (or alternatively, the full-rank estimate of $\mathbf{x}$ from $\mathbf{y}$) carries information about $\mathbf{x}$. However, when $\mathbf{z}$ is non-Gaussian, $R(r)$ is not the actual information rate. Nonetheless, the log of the volume of the concentration ellipse in (58) may still be used as a loose measure of information rate. It is seen that the information rate is additively decomposed by the squared canonical correlations σ_i^2.

The standard measure of linear dependence for the composite vector $\mathbf{s} = [\mathbf{x}^T\ \hat{\mathbf{x}}^T]^T$ is the Hadamard ratio inside the inequality [4]

$$
0 \leq \frac{\det\{R_{ss}\}}{\prod_{i=1}^{2m}(R_{ss})_{ii}} \leq 1
\tag{59}
$$

where

$$
\begin{aligned}
R_{ss} &= E\left[\begin{pmatrix}\mathbf{x} \\ \hat{\mathbf{x}}\end{pmatrix}(\mathbf{x}^T\ \hat{\mathbf{x}}^T)\right] = \begin{bmatrix}R_{xx} & R_{x\hat{x}} \\ R_{\hat{x}x} & R_{\hat{x}\hat{x}}\end{bmatrix} \\
&= \begin{bmatrix}I & 0 \\ R_{\hat{x}x}R_{xx}^{-1} & I\end{bmatrix}\begin{bmatrix}R_{xx} & 0 \\ 0 & Q_{xx}(r)R_{xx}^{-1}\end{bmatrix} \\
&\quad \times \begin{bmatrix}I & R_{xx}^{-1}R_{x\hat{x}} \\ 0 & R_{\hat{x}\hat{x}}\end{bmatrix}
\end{aligned}
\tag{60}
$$

is the composite covariance matrix for the composite vector $\mathbf{s}$, and $(R_{ss})_{ii}$'s are the diagonal elements of R_{ss}. The ratio takes the value 0 iff there is linear dependence among elements of $\mathbf{s}$; it takes the value 1 iff elements of $\mathbf{s}$ are mutually uncorrelated.

Using (57) and (60), the Hadamard ratio in (59) may be written as

$$
\begin{aligned}
&\frac{\det\{R_{ss}\}}{\prod_{i=1}^{2m}(R_{ss})_{ii}} \\
&= \frac{\det\{R_{xx}\}}{\prod_{i=1}^{m}(R_{xx})_{ii}}\det\{I - \Sigma(r)\Sigma(r)^T\}\frac{\det\{R_{\hat{x}\hat{x}}\}}{\prod_{i=1}^{m}(R_{\hat{x}\hat{x}})_{ii}}.
\end{aligned}
\tag{61}
$$

The first term on the right-hand side measures the linear dependence among the elements of $\mathbf{x}$, and the third term measures the linear dependence among the elements of $\hat{\mathbf{x}}$. The middle term $L(r) = \det\{I - \Sigma(r)\Sigma(r)^T\}$ measures the linear dependence between the elements of $\mathbf{x}$ and $\hat{\mathbf{x}}$. It is seen that the linear dependence $L(r)$ is proportional to the volume of the concentration ellipse. This linear dependence may also be written as

$$
\begin{aligned}
L(r) &= \det\{I - \Sigma(r)\Sigma(r)^T\} = \prod_{i=1}^{r}\left(1 - \sigma_i^2\right) \\
&= \frac{L}{\prod_{i=r+1}^{m}\left(1 - \sigma_i^2\right)}
\end{aligned}
\tag{62}
$$

where $L = \prod_{i=1}^{m}(1 - \sigma_i^2)$ measures the linear dependence between elements of $\mathbf{x}$ and $\mathbf{y}$. We see that the linear dependence is multiplicatively decomposed by the squared canonical correlations, even for non-Gaussian data.

Various Implementations: Using (35), the rank-r Wiener filter in (50) may be written in terms of the canonical coordinate maps as

$$
H(r) = R_{xx}W\Sigma(r)D^T.
\tag{63}
$$

This rank-r Wiener filter $H(r)$ may also be implemented with projection matrices, as we now show. Let us partition F and G into $F = [F_r\ F_\star]$ and $G = [G_r\ G_\star]$, and define

$$
I(r) = \begin{bmatrix}I_r & 0 \\ 0 & 0\end{bmatrix}
\tag{64}
$$

where I_r is an $r \times r$ identity matrix. In (31), premultiplying $F^T R_{xx}^{-1/2}R_{xy}R_{yy}^{-T/2}G = \Sigma$ by $I(r)$ and post-multiplying it by $G^T R_{yy}^{-1/2}$ yields

$$
\begin{bmatrix}F_r^T \\ 0\end{bmatrix}R_{xx}^{-1/2}R_{xy}R_{yy}^{-1} = \Sigma(r)G^T R_{yy}^{-1/2}.
\tag{65}
$$

Substituting for $\Sigma(r)G^T R_{yy}^{-1/2}$ in (50) yields the implementation

$$
H(r) = R_{xx}^{1/2}F_r F_r^T R_{xx}^{-1/2}R_{xy}R_{yy}^{-1} = R_{xx}^{1/2}P_{F_r}R_{xx}^{-1/2}H
\tag{66}
$$

where $P_{F_r} = F_r F_r^T$ is the orthogonal projection onto the span of F_r, and $H = R_{xy}R_{yy}^{-1}$ is the full-rank Wiener filter. The interpretation here is that the output of the full-rank Wiener filter H is whitened by $R_{xx}^{-1/2}$, projected onto the subspace $\langle F_r \rangle$, and recolored by $R_{xx}^{1/2}$. Alternately, premultiplying $F^T R_{xx}^{-1/2}R_{xy}R_{yy}^{-T/2}G = \Sigma$ in (31) by $R_{xx}^{1/2}F$ and postmultiplying it by $I(r)$ yields

$$
R_{xy}R_{yy}^{-T/2}[G_r\quad 0] = R_{xx}^{1/2}F\Sigma(r).
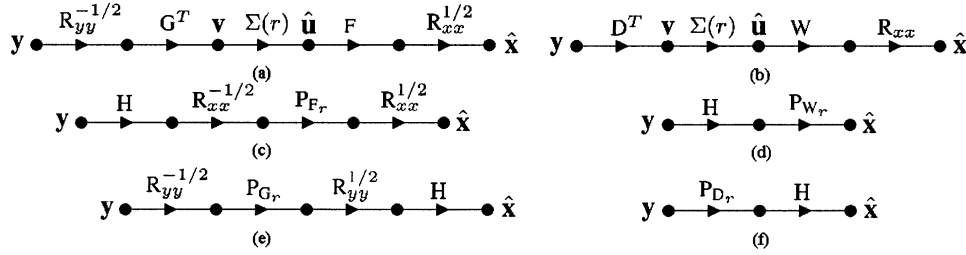\tag{67}
$$

Fig. 2. Equivalent representations of the rank-r Wiener filter in canonical coordinates. (a) $H(r) = R_{xx}^{1/2} F \Sigma(r) G^T R_{yy}^{-1/2}$. (b) $H(r) = R_{xx} W \Sigma(r) D^T$. (c) $H(r) = R_{xx}^{1/2} P_{F_r} R_{xx}^{-1/2} H$. (d) $H(r) = P_{W_r} H$. (e) $H R_{yy}^{1/2} P_{G_r} R_{yy}^{-1/2}$. (f) $H(r) = H P_{D_r}$.

Substituting for $R_{xx}^{1/2} F \Sigma(r)$ in (50) yields the implementation

$$H(r) = R_{xy} R_{yy}^{-T/2} G_r G_r^T R_{yy}^{-1/2} = H R_{yy}^{1/2} P_{G_r} R_{yy}^{-1/2} \quad (68)$$

where $P_{G_r} = G_r G_r^T$ is the orthogonal projection onto the span of G_r. The interpretation here is that the measurement is whitened by $R_{yy}^{-1/2}$, projected onto the subspace $\langle G_r \rangle$, recolored by $R_{yy}^{1/2}$, and then filtered by the full-rank Wiener filter H.

Partitioning W and D into $W = [W_r \ W_\star]$ and $D = [D_r \ D_\star]$ and plugging them into (35) yields

$$R_{xx}^{-T/2} F_r = W_r \quad \text{and} \quad R_{yy}^{-T/2} G_r = D_r. \quad (69)$$

Using (69), the rank-r Wiener filter in (50) and (68) may be implemented as

$$H(r) = R_{xx} W_r W_r^T H = P_{W_r} H, \quad \text{and}$$
$$H(r) = H R_{yy} D_r D_r^T = H P_{D_r} \quad (70)$$

where P_{W_r} and P_{D_r} are the following oblique projection [20], [21] operators:

$$P_{W_r} = R_{xx} W_r W_r^T \quad \text{and} \quad P_{D_r} = R_{yy} D_r D_r^T. \quad (71)$$

The interpretations of these implementations are obvious. Thus, the rank-r Wiener filter $H(r)$ has six equivalent representations, which are depicted in Fig. 2. Reading down the left-hand side, from (a) to (c) to (e), produces various implementations for $H(r)$ in the *orthogonal* coordinates of F_r and G_r, using the *orthogonal* projections P_{F_r} and P_{G_r}. Reading down the right-hand side, from (b) to (d) to (f), produces various implementations for $H(r)$ in the *nonorthogonal* coordinates of W_r and D_r, using the *oblique* projections P_{W_r} and P_{D_r}.

B. Reduced-Rank Filtering in Half-Canonical Coordinates

Here, the objective is to find the rank-r filter $H(r)$ that minimizes the trace of the error covariance matrix $Q_{xx}(r)$ in (46). Thus, the measure to be minimized is

$$MSE_{xx}(r) = \text{tr}\{Q_{xx}(r)\}$$
$$= \text{tr}\{Q_{xx} + [H - H(r)]R_{yy}[H - H(r)]^T\} \quad (72)$$

which we may rewrite as

$$MSE_{xx}(r) = \text{tr}\{Q_{xx}\} + \text{tr}\left\{\left[H R_{yy}^{1/2} - H(r)R_{yy}^{1/2}\right]\right.$$
$$\left. \times \left[H R_{yy}^{1/2} - H(r)R_{yy}^{1/2}\right]^T\right\}. \quad (73)$$

The first term on the right side of this equation is the MSE for estimating x from y with the full-rank Wiener filter $H = R_{xy} R_{yy}^{-1}$ and is fixed. The second term

$$\epsilon_{xx}^2 = \text{tr}\left\{\left[H R_{yy}^{1/2} - H(r)R_{yy}^{1/2}\right]\left[H R_{yy}^{1/2} - H(r)R_{yy}^{1/2}\right]^T\right\} \quad (74)$$

is the Frobenius norm of the matrix $H R_{yy}^{1/2} - H(r)R_{yy}^{1/2}$, which measures the extra variance introduced by rank reduction [6], [7]. The optimum choice for the rank-r Wiener filter $H(r)$ is the rank-r matrix that best approximates the half-coherence matrix $H R_{yy}^{1/2} = R_{xy} R_{yy}^{-T/2} = U \Sigma V^T$, by minimizing ϵ_{xx}^2. Thus, it is given by $H(r) R_{yy}^{1/2} = U \Sigma(r) V^T$ or [6], [7]

$$H(r) = U \Sigma(r) V^T R_{yy}^{-1/2}. \quad (75)$$

Correspondingly, $\hat{x} = H(r)y$ is the rank-r Wiener estimate of x from y. Using (75), the optimal value of $MSE_{xx}(r)$ is

$$MSE_{xx}(r) = \text{tr}\{Q_{xx}\} + \text{tr}\{\Sigma \Sigma^T - \Sigma(r)\Sigma(r)^T\}$$
$$= MSE_{xx} + \sum_{i=r+1}^{m} \sigma_i^2 \quad (76)$$

which is the MSE of the rank-r Wiener estimator $H(r)$. The first term on the right-hand side of (76) is the minimum MSE for the full-rank estimator of x from y, i.e., $\hat{x} = Hy$, and the second term is the extra MSE ϵ_{xx}^2 due to rank reduction.

Various Implementations: Using (43), the rank-r Wiener filter in (75) may be written, in terms of the half-canonical coordinate maps, as

$$H(r) = W \Sigma(r) D^T. \quad (77)$$

Similar to Section IV-A, we may implement $H(r)$ with projection matrices. Let us partition U and V into $U = [U_r \ U_\star]$ and $V = [V_r \ V_\star]$. In (39), premultiplying $U^T R_{xy} R_{yy}^{-T/2} V = \Sigma$ by $I(r)$ and post-multiplying it by $V^T R_{yy}^{-1/2}$ yields

$$\begin{bmatrix} U_r^T \\ 0 \end{bmatrix} R_{xy} R_{yy}^{-1} = \begin{bmatrix} U_r^T \\ 0 \end{bmatrix} H = \Sigma(r) V^T R_{yy}^{-1/2}. \quad (78)$$

Substituting for $\Sigma(r) V^T R_{yy}^{-1/2}$ in (75) yields the implementation

$$H(r) = U_r U_r^T H = P_{U_r} H \quad (79)$$

where $P_{U_r} = U_r U_r^T$ is the orthogonal projection onto the span of U_r. The interpretation is that a full-rank Wiener filter H is
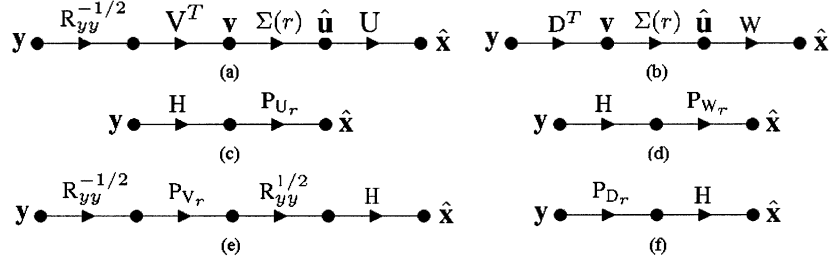
Fig. 3. Equivalent representations of the rank-r Wiener filter in half-canonical coordinates. (a) $\mathrm{H}(r) = \mathrm{U}\Sigma(r)\mathrm{V}^T\mathrm{R}_{yy}^{-1/2}$. (b) $\mathrm{H}(r) = \mathrm{W}\Sigma(r)\mathrm{D}^T$. (c) $\mathrm{H}(r) = \mathrm{P}_{\mathrm{U}_r}\mathrm{H}$. (d) $\mathrm{H}(r) = \mathrm{P}_{\mathrm{W}_r}\mathrm{H}$. (e) $\mathrm{H}(r) = \mathrm{H}\mathrm{R}_{yy}^{1/2}\mathrm{P}_{\mathrm{V}_r}\mathrm{R}_{yy}^{-1/2}$. (f) $\mathrm{H}(r) = \mathrm{H}\mathrm{P}_{\mathrm{D}_r}$.

followed by a projection onto the subspace $\langle \mathrm{U}_r \rangle$. Alternately, premultiplying $\mathrm{U}^T\mathrm{R}_{xy}\mathrm{R}_{yy}^{-T/2}\mathrm{V} = \Sigma$ in (39) by U and post-multiplying it by $\mathrm{I}(r)$ yields

$$\mathrm{R}_{xy}\mathrm{R}_{yy}^{-T/2}[\mathrm{V}_r \quad 0] = \mathrm{U}\Sigma(r). \tag{80}$$

Substituting for $\mathrm{U}\Sigma(r)$ in (75) yields the implementation

$$\mathrm{H}(r) = \mathrm{R}_{xy}\mathrm{R}_{yy}^{-T/2}\mathrm{V}_r\mathrm{V}_r^T\mathrm{R}_{yy}^{-1/2} = \mathrm{H}\mathrm{R}_{yy}^{1/2}\mathrm{P}_{\mathrm{V}_r}\mathrm{R}_{yy}^{-1/2} \tag{81}$$

where $\mathrm{P}_{\mathrm{V}_r} = \mathrm{V}_r\mathrm{V}_r^T$ is the orthogonal projection onto the span of V_r. The interpretation here is that the measurement is whitened with $\mathrm{R}_{yy}^{-1/2}$, projected onto the subspace $\langle \mathrm{V}_r \rangle$, recolored by $\mathrm{R}_{yy}^{1/2}$, and then filtered by the full-rank Wiener filter H.

Partitioning W and D into $\mathrm{W} = [\mathrm{W}_r \ \mathrm{W}_\star]$ and $\mathrm{D} = [\mathrm{D}_r \ \mathrm{D}_\star]$ and plugging them into (43) yields

$$\mathrm{U}_r = \mathrm{W}_r \quad \text{and} \quad \mathrm{R}_{yy}^{-T/2}\mathrm{V}_r = \mathrm{D}_r. \tag{82}$$

Using (82), the rank-r Wiener filter in (79) and (81) may be implemented as

$$\mathrm{H}(r) = \mathrm{W}_r\mathrm{W}_r^T\mathrm{H} = \mathrm{P}_{\mathrm{W}_r}\mathrm{H} \quad \text{and}$$
$$\mathrm{H}(r) = \mathrm{H}\mathrm{R}_{yy}\mathrm{D}_r\mathrm{D}_r^T = \mathrm{H}\mathrm{P}_{\mathrm{D}_r} \tag{83}$$

where $\mathrm{P}_{\mathrm{W}_r}$ and $\mathrm{P}_{\mathrm{D}_r}$ are the following orthogonal and oblique projection operators:

$$\mathrm{P}_{\mathrm{W}_r} = \mathrm{W}_r\mathrm{W}_r^T \quad \text{and} \quad \mathrm{P}_{\mathrm{D}_r} = \mathrm{R}_{yy}\mathrm{D}_r\mathrm{D}_r^T. \tag{84}$$

Thus, similar to Section IV-A, the rank-r Wiener filter $\mathrm{H}(r)$ associated with measure (72) has six equivalent representations. These representations are depicted in Fig. 3. Reading down the left-hand side, from (a) to (c) to (e), produces various implementations for $\mathrm{H}(r)$ in the coordinates of U_r and V_r. Reading down the right-hand side, from (b) to (d) to (f), produces various implementations for $\mathrm{H}(r)$ in the coordinates of W_r and D_r.

V. COMPUTING CANONICAL COORDINATE AND HALF-CANONICAL COORDINATE MAPPINGS

A conventional method of canonical coordinate decomposition does not offer a simple way to compute a small subset of canonical coordinates that are required for low-rank modeling. A full SVD for the coherence matrix

$\mathrm{F}\Sigma\mathrm{G}^T = \mathrm{R}_{xx}^{-1/2}\mathrm{R}_{xy}\mathrm{R}_{yy}^{-T/2} = \mathrm{C}$ has to be computed, regardless of the rank-reduction. There are indeed simple and fast algorithms to compute the principal singular vectors of a matrix, e.g., [14], [22]–[26]. However, the conventional method also requires the computation of the square-root-inverses of the covariance matrices R_{xx} and R_{yy}. In addition, the conventional method does not allow an easy update of the canonical coordinate mappings in time for expanding data samples, making it intractable for online applications. A similar argument may be made about a conventional method of half-canonical coordinate decomposition.

In this section, various alternating power methods are derived to *recursively* compute the canonical and half-canonical coordinate mapping vectors (columns of W and D) one by one or in group. These algorithms also allow for an update of the mapping vectors in time as new samples of the channels are observed. Provided that the rank-reduction is relatively large and the singular values of the coherence matrix are not close together, the alternating power methods can be more efficient in computation than the conventional methods as they require no matrix square-roots. The algorithms presented here are identical in form to the alternating power methods derived in [10] to compute the AB^T factorization of the rank-r Wiener filter $\mathrm{H}(r) = \mathrm{AB}^T$. However, the algorithms in [10] do not yield the canonical and half-canonical coordinate maps W and D, and the corresponding canonical and half-canonical correlation matrix $\Sigma = \mathrm{W}^T\mathrm{R}_{xy}\mathrm{D}$. Therefore, what is original here is the idea that alternating power methods may be used to compute canonical and half-canonical coordinate maps and correlations, making them even more applicable in signal processing problems than they would appear from the work in [10].

A. Computing Canonical Coordinate Mappings

The power method [14]–[17] may be the simplest and oldest method for computing the principal eigenvectors of a matrix. It is also a natural choice for computing the principal subspace of a matrix [26]. Our aim is to adapt it to the coupled generalized (symmetric) eigenvalue problem in (15) or, equivalently, to the coupled (asymmetric) generalized eigenvalue problem in (14).

A standard power method [14]–[17] for computing the first columns of W and D, i.e., $\mathbf{w}_1$ and $\mathbf{d}_1$, associated with the dominant eigenvalue of (15), may be summarized as follows. Let k denote the index of iteration, start with a random choice for

$\mathbf{w}_1(0) \in \mathrm{R}^m$ and $\mathbf{d}_1(0) \in \mathrm{R}^n$, and iterate the following equations on k until convergence:

$$\begin{cases} \bar{\mathbf{w}}_1(k+1) = \left(\mathrm{R}_{xx}^{-1}\mathrm{R}_{xy}\mathrm{R}_{yy}^{-1}\mathrm{R}_{yx}\right)^{k+1}\mathbf{w}_1(k) \\ \mathbf{w}_1(k+1) = \bar{\mathbf{w}}_1(k+1)(\bar{\mathbf{w}}_1(k+1)^T\mathrm{R}_{xx}\bar{\mathbf{w}}_1(k+1))^{-1/2} \\ \bar{\mathbf{d}}_1(k+1) = \left(\mathrm{R}_{yy}^{-1}\mathrm{R}_{yx}\mathrm{R}_{xx}^{-1}\mathrm{R}_{xy}\right)^{k+1}\mathbf{d}_1(k) \\ \mathbf{d}_1(k+1) = \bar{\mathbf{d}}_1(k+1)(\bar{\mathbf{d}}_1(k+1)^T\mathrm{R}_{yy}\bar{\mathbf{d}}_1(k+1))^{-1/2}. \end{cases} \tag{85}$$

The normalization to obtain $\mathbf{w}_1$ from $\bar{\mathbf{w}}_1$ and $\mathbf{d}_1$ from $\bar{\mathbf{d}}_1$ ensures that $\mathbf{w}_1^T\mathrm{R}_{xx}\mathbf{w}_1 = 1$ and $\mathbf{d}_1^T\mathrm{R}_{yy}\mathbf{d}_1 = 1$ for each iteration k.

Alternating Power Method: A simpler algorithm can be developed based on the coupled (asymmetric) generalized eigenvalue problem in (14) to find $\mathbf{w}_1$ and $\mathbf{d}_1$. Let us rewrite (14) as

$$\mathrm{R}_{xx}^{-1}\mathrm{R}_{xy}\mathrm{D} = \mathrm{W}\Sigma \tag{86}$$
$$\mathrm{R}_{yy}^{-1}\mathrm{R}_{yx}\mathrm{W} = \mathrm{D}\Sigma. \tag{87}$$

These equations suggest an alternating sequence of approximations to W and D. Given a random initial guess $\mathbf{d}_1(0)$, the first estimate of $\mathbf{w}_1$ is computed from (86). With this estimate of $\mathbf{w}_1$, (87) is used to compute a new estimate of $\mathbf{d}_1$. This iterative alternation between (86) and (87) continues until convergence. We can summarize the algorithm as follows. Randomly select $\mathbf{d}_1(0) \in \mathrm{R}^n$, start with $k = 0$, and iterate the following equations on k until convergence:

$$\begin{cases} \bar{\mathbf{w}}_1(k+1) = (\mathrm{R}_{xx}^{-1}\mathrm{R}_{xy})\mathbf{d}_1(k) \\ \mathbf{w}_1(k+1) = \bar{\mathbf{w}}_1(k+1)(\bar{\mathbf{w}}_1(k+1)^T\mathrm{R}_{xx}\bar{\mathbf{w}}_1(k+1))^{-1/2} \\ \bar{\mathbf{d}}_1(k+1) = (\mathrm{R}_{yy}^{-1}\mathrm{R}_{yx})\,\mathbf{w}_1(k+1) \\ \mathbf{d}_1(k+1) = \bar{\mathbf{d}}_1(k+1)(\bar{\mathbf{d}}_1(k+1)^T\mathrm{R}_{yy}\bar{\mathbf{d}}_1(k+1))^{-1/2}. \end{cases} \tag{88}$$

This algorithm is a *two-step* decomposition of the standard power method of (85). Therefore, the convergence of it follows from the convergence of the standard power method. Provided that the first eigenvalue of (14) is larger than the second one (i.e., $\sigma_1 > \sigma_2$) and the initial guess of $\mathbf{d}_1$ is not orthogonal to $\mathbf{d}_1$ (i.e., $\mathbf{d}_1(0)^T\mathrm{R}_{yy}\mathbf{d}_1 \neq 0$), the estimates $\mathbf{d}_1(k)$ and $\mathbf{w}_1(k)$ converge to $\mathbf{d}_1$ and $\mathbf{w}_1$. The estimation error at iteration k converges to zero as an exponential function of the ratio of the second eigenvalue to the first one, i.e., as $(\sigma_2/\sigma_1)^k$ [14]–[17].

The algorithm in (88) may be rewritten as

$$\begin{cases} \mathrm{R}_{xx}\bar{\mathbf{w}}_1(k+1) = \mathrm{R}_{xy}\mathbf{d}_1(k) \\ \mathbf{w}_1(k+1) = \bar{\mathbf{w}}_1(k+1)(\bar{\mathbf{w}}_1(k+1)^T\mathrm{R}_{xx}\bar{\mathbf{w}}_1(k+1))^{-1/2} \\ \mathrm{R}_{yy}\bar{\mathbf{d}}_1(k+1) = \mathrm{R}_{yx}\mathbf{w}_1(k+1) \\ \mathbf{d}_1(k+1) = \bar{\mathbf{d}}_1(k+1)(\bar{\mathbf{d}}_1(k+1)^T\mathrm{R}_{yy}\bar{\mathbf{d}}_1(k+1))^{-1/2}. \end{cases} \tag{89}$$

Therefore, at each iteration k, the vectors $\bar{\mathbf{w}}_1(k+1)$ and $\bar{\mathbf{d}}_1(k+1)$ are determined by solving the linear systems of equations $\mathrm{R}_{xx}\bar{\mathbf{w}}_1(k+1) = \mathrm{R}_{xy}\mathbf{d}_1(k)$ and $\mathrm{R}_{yy}\bar{\mathbf{d}}_1(k+1) = \mathrm{R}_{yx}\mathbf{w}_1(k+1)$, respectively. Any standard method for solving a linear system of equations may be used here. For various methods, see [14]. We call the algorithm in (89) an *alternating power method*, in the sense that it solves a coupled generalized eigenvalue problem using alternating iterations. As mentioned earlier, it may be viewed as a two-step decomposition of the standard power iterations for computing the principal eigenvector of a matrix.

The alternating power method reported and analyzed in [10][2] for computing the canonical components of a reduced-rank Wiener filter is a generalization of an Iterative Quadratic Minimum Distance (IQMD) method [12]. It furnishes matrices A and B that decompose the rank-r Wiener filter $\mathrm{H}(r)$ as $\mathrm{H}(r) = \mathrm{AB}^T$. The matrices A and B are related to the canonical coordinate maps W and D as $\mathrm{A} = \mathrm{R}_{xx}\mathrm{W}_r\Sigma_r\mathrm{M}$ and $\mathrm{B} = \mathrm{D}_r\mathrm{M}^{-T}$, where W_r and D_r contain the first r columns of W and D, and M is any $r \times r$ nonsingular matrix. Because of the ambiguous matrix M, the algorithm cannot be used to compute the canonical coordinate maps or their corresponding canonical correlations $\sigma_i = \mathbf{w}_i^T\mathrm{R}_{xy}\mathbf{d}_i$.

Alternating Block Power Method: [9]. If the ratio (σ_2/σ_1) is close to one, then the convergence rate is very slow. One way to address this problem is to combine the block power method of [16] and [17] with the above alternating procedure to solve for several columns of W and D. The idea is to start with $l \leq m$ orthogonal vectors and, after each iteration, use a Gram–Schmidt orthogonalization procedure to guarantee that the constraints $\mathrm{W}_l^T\mathrm{R}_{xx}\mathrm{W}_l = \mathrm{I}_l$ and $\mathrm{D}_l^T\mathrm{R}_{yy}\mathrm{D}_l = \mathrm{I}_l$ are satisfied. This algorithm may be summarized as follows. Initialize $\mathrm{D}_l(0)$ with orthogonal columns, start with $k = 0$, and iterate the following equations on k until convergence:

$$\begin{cases} \text{Solve } \mathrm{R}_{xx}\bar{\mathrm{W}}_l(k+1) = \mathrm{R}_{xy}\mathrm{D}_l(k) \text{ for } \bar{\mathrm{W}}_l(k+1) \\ \bar{\mathrm{W}}_l(k+1) \xrightarrow{\text{GSO}} \mathrm{W}_l(k+1) \text{ such that} \\ \quad \mathrm{W}_l^T(k+1)\mathrm{R}_{xx}\mathrm{W}_l(k+1) = \mathrm{I}_l \\ \text{Solve } \mathrm{R}_{yy}\bar{\mathrm{D}}_l(k+1) = \mathrm{R}_{yx}\mathrm{W}_l(k+1) \text{ for } \bar{\mathrm{D}}_l(k+1) \\ \bar{\mathrm{D}}_l(k+1) \xrightarrow{\text{GSO}} \mathrm{D}_l(k+1) \text{ such that} \\ \quad \mathrm{D}_l^T(k+1)\mathrm{R}_{yy}\mathrm{D}_l(k+1) = \mathrm{I}_l \end{cases} \tag{90}$$

where GSO is a Gram–Schmidt Orthogonalization. The GSO for W_l may be summarized as follows. At each iteration k, do the following for $i = 1, 2, \ldots, l$:

$$\tilde{\mathbf{w}}_i(k+1) = \left[\mathrm{I} - \tilde{\mathrm{W}}_{i-1}(k+1)\tilde{\mathrm{W}}_{i-1}^T(k+1)\mathrm{R}_{xx}\right]\bar{\mathbf{w}}_i(k+1)$$
$$\mathbf{w}_i(k+1) = \tilde{\mathbf{w}}_i(k+1)\left(\tilde{\mathbf{w}}_i^T(k+1)\mathrm{R}_{xx}\tilde{\mathbf{w}}_i(k+1)\right)^{-1/2} \tag{91}$$

where $\tilde{\mathrm{W}}_{i-1}(k+1) = [\tilde{\mathbf{w}}_1(k+1), \ldots, \tilde{\mathbf{w}}_{i-1}(k+1)]$. A similar set of equations may be written for the GSO for D_l. In [9], this algorithm was used to iteratively compute WA and DB of the PCCA.

Alternating Power Method With Deflation: We now extend the previous algorithm by introducing an alternating power method with deflation for computing the canonical coordinate maps. Assume that the first $r < m$ canonical coordinate

---

[2]The alternating power method of [10] is not a simple two-step decomposition of the standard power method, and therefore, its convergence analysis requires special treatment.

mappings (the first r columns of W and D) have already been found. Partition W, D, and Σ into

$$W = [W_r \quad W_\star], \quad D = [D_r \quad D_\star], \quad \text{and}$$
$$\Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & \Sigma_\star \end{bmatrix}. \tag{92}$$

To compute the $(r+1)$th pair of canonical coordinate mappings, deflate the first r canonical coordinate mappings from the left hand sides of (14) to get

$$\left(I - R_{xx}W_r W_r^T\right) R_{xy}D_\star = R_{xx}W_\star \Sigma_\star$$
$$\left(I - R_{yy}D_r D_r^T\right) R_{yx}W_\star = R_{yy}D_\star \Sigma_\star. \tag{93}$$

The $(r+1)$th canonical coordinate mappings $\mathbf{w}_{r+1}$ and $\mathbf{d}_{r+1}$ are the first columns of $W_\star$ and $D_\star$, which are now associated with the dominant eigenvalue of (93). The Appendix makes this claim precise. Thus, these mappings can be computed by iterating between the two equations in (93) with a random initialization. The dominant eigenvalue of (93), i.e., $\sigma_{r+1} = \mathbf{w}_{r+1}R_{xy}\mathbf{d}_{r+1}$, is the $(r+1)$th canonical correlation of $\mathbf{x}$ and $\mathbf{y}$.

The deflation process may also be implemented with projection matrices. Using (71), we may rewrite (93) as

$$(I - P_{W_r})R_{xy}D_\star = R_{xx}W_\star \Sigma_\star$$
$$(I - P_{D_r})R_{yx}W_\star = R_{yy}D_\star \Sigma_\star. \tag{94}$$

Thus, the alternating power method for finding $\mathbf{w}_{r+1}$ and $\mathbf{d}_{r+1}$ may be summarized as follows. Randomly select $\mathbf{d}_{r+1}(0) \in \mathbb{R}^n$, set $k = 0$, and iterate the following equations on k until convergence:

$$\begin{cases} \text{Solve } R_{xx}\bar{\mathbf{w}}_{r+1}(k+1) = (I - P_{W_r}) R_{xy}\mathbf{d}_{r+1}(k) \\ \quad \text{for } \bar{\mathbf{w}}_{r+1}(k+1) \\ \mathbf{w}_{r+1}(k+1) = \bar{\mathbf{w}}_{r+1}(k+1) \\ \quad \times \left(\bar{\mathbf{w}}_{r+1}^T(k+1)R_{xx}\bar{\mathbf{w}}_{r+1}(k+1)\right)^{-1/2} \\ \text{Solve } R_{yy}\bar{\mathbf{d}}_{r+1}(k+1) = (I - P_{D_r}) R_{yx}\mathbf{w}_{r+1}(k+1) \\ \quad \text{for } \bar{\mathbf{d}}_{r+1}(k+1) \\ \mathbf{d}_{r+1}(k+1) = \bar{\mathbf{d}}_{r+1}(k+1) \\ \quad \times \left(\bar{\mathbf{d}}_{r+1}^T(k+1)R_{yy}\bar{\mathbf{d}}_{r+1}(k+1)\right)^{-1/2}. \end{cases} \tag{95}$$

Provided that $\sigma_{r+1} < \sigma_r$, this algorithm yields the $(r+1)$th pair of canonical coordinate mappings at the rate of $(\sigma_{r+1}/\sigma_r)^k$. We note that the alternating block power method in (90) may be used to simultaneously compute several columns of $W_\star$ and $D_\star$ associated with the dominant eigenvalues of the deflated coupled (asymmetric) generalized eigenvalue problem in (94). This produces an alternating block power method with deflation.

Unlike the conventional method of canonical coordinate decomposition, the alternating power method, with deflation, in (95) requires no matrix square-roots. Moreover, all operations are matrix-vector multiplications, where the number of vectors might be much smaller than the number of columns of the matrix. Therefore, provided that the eigenvalues associated with the desired canonical coordinates are not close to each other, the alternating power method, with deflation, in (95) is an efficient algorithm for practical extraction of a few dominant canonical coordinate mappings and canonical correlations.

Order-Recursive Alternating Power Method: In most applications, the number of canonical coordinate pairs to be extracted is not known *a priori*, or it may vary with time. One may run a test of information rate or linear dependence based on (58) and (62) to determine if a prespecified threshold is met. If the threshold is not reached, additional canonical coordinate pairs must be extracted. However, if the threshold is exceeded, computation of the mapping vectors associated with the less significant canonical coordinate pairs may be stopped to reduce the computational load. Thus, the alternating power method will be modified in the next paragraph to allow for changes in the number of columns of W and D to be computed, during the iterations of the algorithm.

The alternating block power method in (90) yields W_r and D_r asymptotically. Thus, (90) and (95) can be run at the same time, with W_r and D_r in (95) being replaced by $W_r(k)$ and $D_r(k)$. At each iteration k, the combination of (90) and (95) can be run successively for $r = 0, 1, \ldots, p-1, p \leq m$ to extract up to p columns of W and D. This algorithm may be summarized as follows. At each iteration k, do the following for $r = 0, 1, \ldots, p-1$:

$$\begin{cases} \mathbf{a}_{r+1}(k+1) = R_{xy}\mathbf{d}_{r+1}(k) \\ \alpha_{r+1}(k+1) = W_r^T(k+1)\mathbf{a}_{r+1}(k+1) \\ \mathbf{b}_{r+1}(k+1) = W_r(k+1)\alpha_{r+1}(k+1) \\ \beta_{r+1}(k+1) = \mathbf{a}_{r+1}(k+1) - R_{xx}\mathbf{b}_{r+1}(k+1) \\ \text{Solve } R_{xx}\bar{\mathbf{w}}_{r+1}(k+1) = \beta_{r+1}(k+1), \text{ for } \bar{\mathbf{w}}_{r+1}(k+1) \\ \mathbf{w}_{r+1}(k+1) = \bar{\mathbf{w}}_{r+1}(k+1) \\ \quad \times \left(\bar{\mathbf{w}}_{r+1}^T(k+1)\beta_{r+1}(k+1)\right)^{-1/2} \\ \mathbf{q}_{r+1}(k+1) = R_{yx}\mathbf{w}_{r+1}(k+1) \\ \gamma_{r+1}(k) = D_r^T(k+1)\mathbf{q}_{r+1}(k+1) \\ \mathbf{s}_{r+1}(k+1) = D_r(k+1)\gamma_{r+1}(k+1) \\ \theta_{r+1}(k+1) = \mathbf{q}_{r+1}(k+1) - R_{yy}\mathbf{s}_{r+1}(k+1) \\ \text{Solve } R_{yy}\bar{\mathbf{d}}_{r+1}(k+1) = \theta_{r+1}(k+1), \text{ for } \bar{\mathbf{d}}_{r+1}(k+1) \\ \mathbf{d}_{r+1}(k+1) = \bar{\mathbf{d}}_{r+1}(k+1) \\ \quad \times \left(\bar{\mathbf{d}}_{r+1}^T(k+1)\theta_{r+1}(k+1)\right)^{-1/2} \\ W_{r+1}(k+1) = [W_r(k+1) \quad \mathbf{w}_{r+1}(k+1)] \\ D_{r+1}(k+1) = [D_r(k+1) \quad \mathbf{d}_{r+1}(k+1)]. \end{cases} \tag{96}$$

Note that the computations that require W_0 and D_0 must be ignored. The value of p may be changed during the iterations of the algorithm to meet the prespecified criterion. The above algorithm only involves scalar-vector and vector-matrix multiplications, and no matrix-matrix multiplication is required.

Online Implementation: For online implementation, the idea is simply to allow the correlation matrices $R_{xx}, R_{yy},$ and R_{xy} to be updated as new data become available during the iteration of the alternating power method. This may be done using the standard rank-one update equation

$$R_{xx}(j) = \delta R_{xx}(j-1) + \mathbf{x}(j)\mathbf{x}^T(j) \tag{97}$$

where $\delta \in (0, 1)$ is a forgetting factor. To prevent R_{xx} from becoming singular at early iterations, $R_{xx}(0)$ may be chosen $R_{xx} = \rho^2 I$, where $|\rho|$ is small. After each rank-one update of the covariance matrices, the alternating power method (any version) may be iterated for one or more iterations. During the iterations, the covariance matrices are kept fixed. When the next

sample pair of the data is observed, covariance matrices are updated again, and the procedure is repeated. The number of times that the equations in an alternating power method are iterated introduces a tradeoff between the accuracy of the algorithm and computational load. This tradeoff is illustrated in Section VI.

B. Computing Half-Canonical Coordinate Mappings

Similar to Section V-A, we may derive alternating power methods for solving the coupled (asymmetric) generalized eigenvalue problem in (20) to find the half-canonical coordinate mappings. However, it is interesting to note that when R_{xx} is set to $R_{xx} = I$, the generalized eigenvalue problem of (14) for canonical coordinate maps reduces to the generalized eigenvalue problem of (20) for half-canonical coordinate maps. Thus, all variations of the alternating power methods introduced for computing the canonical coordinate mappings, may be used to compute the half-canonical coordinate mappings by replacing R_{xx} with $R_{xx} = I$. Note that in the alternating power method with deflation, (95), the oblique projection matrix $P_{W_r} = R_{xx}W_r W_r^T$ is replaced by the orthogonal projection $P_{W_r} = W_r W_r^T$.

VI. SIMULATION RESULTS

This section demonstrates the correctness of the alternating power method for extracting canonical coordinate mappings on a synthesized data set. The data set is constructed from the channel models

$$\mathbf{x} = H_{xx}\eta_x$$
$$\mathbf{y} = H_{yx}\mathbf{x} + H_{yy}\eta_y$$

where $\mathbf{x} \in R^4$, and $\mathbf{y} \in R^5$. The matrices $H_{xx} \in R^{4\times4}, H_{yy} \in R^{5\times5}$, and $H_{yx} \in R^{5\times4}$ are known, and $\eta_x \in R^4$ and $\eta_y \in R^5$ are two independent white Gaussian vectors.

Let $\hat{\mathbf{w}}_i$ and $\hat{\mathbf{d}}_i$ denote the estimates of the ith canonical coordinate mappings $\mathbf{w}_i$ and $\mathbf{d}_i$. We define the normalized error norm of the ith estimated canonical coordinate mapping as

$$e_{\mathbf{w}_i} = \frac{\|\mathbf{w}_i - \hat{\mathbf{w}}_i\|}{\|\mathbf{w}_i\|} \quad \text{and} \quad e_{\mathbf{d}_i} = \frac{\|\mathbf{d}_i - \hat{\mathbf{d}}_i\|}{\|\mathbf{d}_i\|}.$$

We also define the rank-r group errors of the ideal canonical coordinate mappings W_r and D_r as

$$E_{W_r} = \frac{\|W_r - \hat{W}_r\|}{\|W_r\|} \quad \text{and} \quad E_{D_r} = \frac{\|D_r - \hat{D}_r\|}{\|D_r\|}$$

where $\|W\|$ denotes the Frobenius norm $\|W\|^2 = \text{tr}\{W^T W\}$. In these definitions, the vectors $\mathbf{w}_i$ and $\mathbf{d}_i$, their estimates $\hat{\mathbf{w}}_i$ and $\hat{\mathbf{d}}_i$, the matrices W_r and D_r, and their estimates $\hat{W}_r$ and $\hat{D}_r$ are normalized in sign.

Alternating Block Power Method in Batch Mode: The covariance matrices are computed from $N = 500$ samples and kept constant during the iteration of the alternating block power method in (90). Fig. 4(a) and (b) shows the rank-3 group errors associated with W and D when ten independent initializations are used. The errors are very small after the seventh iteration.
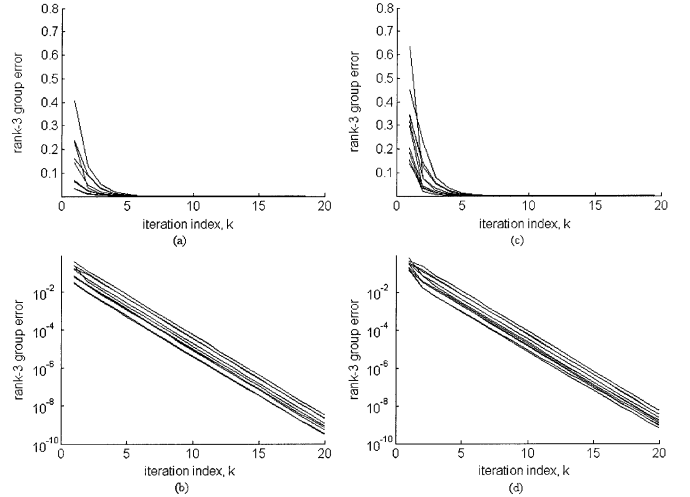

Fig. 4. Rank-3 group errors for the alternating block power method, in batch mode, with ten independent initializations. (a) E_{W_3} linear scale. (b) E_{W_3} logarithmic scale. (c) E_{D_3} linear scale. (d) E_{D_3} logarithmic scale. The results confirm that convergence of the alternating block power method is exponential in iteration number.
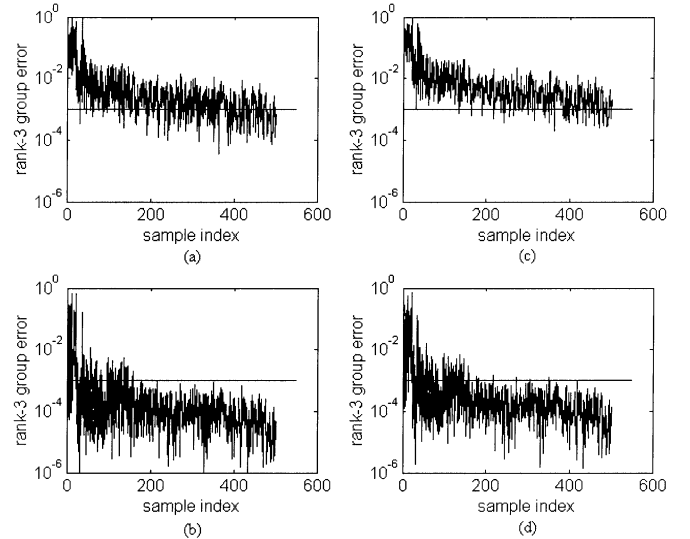

Fig. 5. Rank-3 group errors E_{W_3} and E_{D_3} for the alternating block power method, in online mode, with a variable number of iterations per sample index. (a) E_{W_3}: One iteration. (b) E_{W_3}: Four iterations. (c) E_{D_3}: One iteration. (d) E_{D_3}: Four iterations.

The logarithmic versions of these plots are given in Fig. 4(c) and (d). The exponential convergence of the algorithm is prominent in these figures.

Alternating Block Power Method in Online Mode: For this case, the covariance matrices are updated using the rank-one time updating equation in (97) as a new sample pair $\mathbf{x}(j)$ and $\mathbf{y}(j)$ becomes available. After each time (sample) update, the covariance matrices may be used in one or more iterations of the alternating block power method. The forgetting factor used for updating the covariance matrices is chosen to be $\delta = 0.99$. All other assumptions are as in the previous case. Fig. 5(a)–(d) shows the rank-3 group errors associated with W and D versus the iteration index (for ten independent initializations). In these plots, the number of iterations of the algorithm for each new sample pair is 1 and 4, respectively.
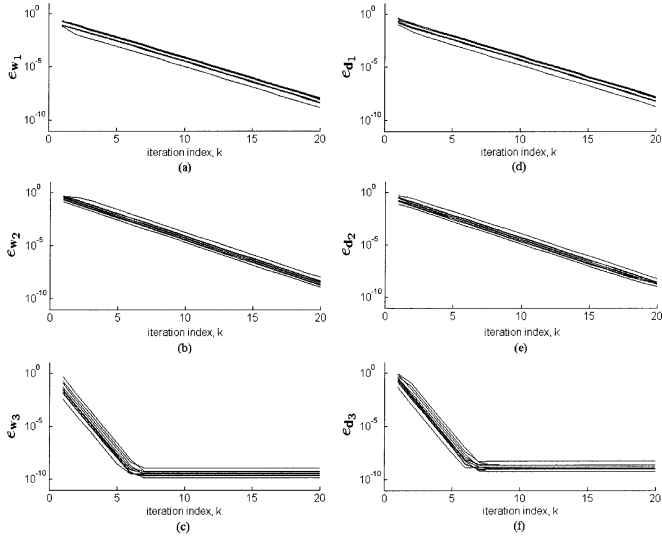
Fig. 6. Normalized error norms of the estimated canonical coordinate mappings for the alternating power method with deflation, in batch mode, in logarithmic scale with ten independent initializations. (a) $e_{\mathbf{w}_1}$. (b) $e_{\mathbf{w}_2}$. (c) $e_{\mathbf{w}_3}$. (d) $e_{\mathbf{d}_1}$. (e) $e_{\mathbf{d}_2}$. (f) $e_{\mathbf{d}_3}$. The results confirm that convergence of the alternating power method with deflation is exponential in iteration number.
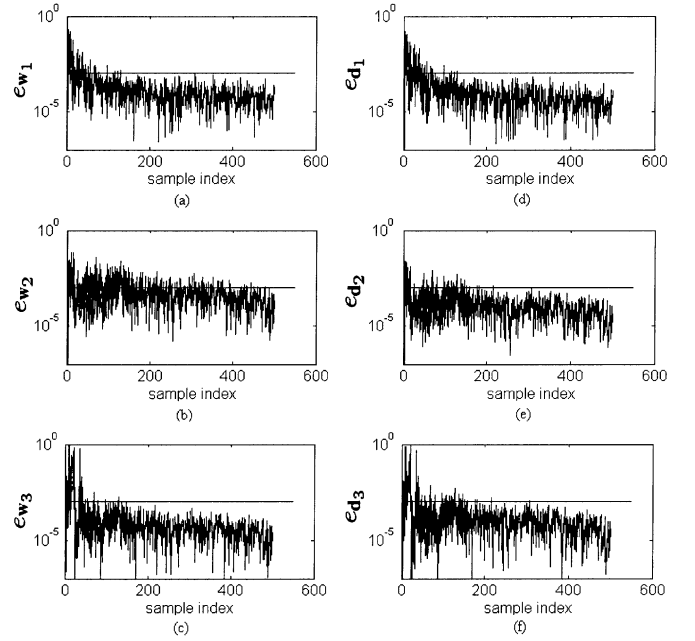

Fig. 7. Normalized error norms of the estimated canonical coordinate mapping for the order recursive alternating power method, in online mode, in logarithmic scale with ten independent initializations. (a) $e_{\mathbf{w}_1}$. (b) $e_{\mathbf{w}_2}$. (c) $e_{\mathbf{w}_3}$. (d) $e_{\mathbf{d}_1}$. (e) $e_{\mathbf{d}_2}$. (f) $e_{\mathbf{d}_3}$.

Alternating Power Method with Deflation in Batch Mode: The covariance matrices are computed in batch mode from the 500 samples of the data channels and kept fixed during the iterations of the alternating power method in (95). We use the algorithm in (95) to compute the first three pair of canonical coordinate mappings. The computation of the ith canonical coordinate mappings $\mathbf{w}_i$ and $\mathbf{d}_i, i \in [1,3]$ is started after the estimates of the $(i-1)$th canonical coordinate mappings have converged. In computing $\mathbf{w}_1$ and $\mathbf{d}_1$, the matrices $\mathrm{P}_{\mathrm{W}_0}$ and $\mathrm{P}_{\mathrm{D}_0}$ are set to zero. Fig. 6(a)–(f) shows the normalized error norms of the estimated canonical coordinate mappings associated with $\mathbf{w}_i$ and $\mathbf{d}_i, i \in [1,3]$ versus iteration number when ten independent initializations are used. The plots are given in logarithmic scale. The straight decaying lines show that the convergence of the algorithm is exponential in iteration number. The constant error levels in Figs. 6(c) and (f), after approximately seven iterations, are due to the numerical precision of MATLAB, and the error propagation caused by deflation.

Order-Recursive Alternating Power Method in Online Mode: Here, the covariance matrices are updated using the rank-one time updating equation in (97) as a new sample pair $\mathbf{x}(j)$ and $\mathbf{y}(j)$ becomes available. Each updated covariance matrix is used for four iterations in the alternating power method in (96) before it is updated again. The forgetting factor used for updating the covariance matrices is $\delta = 0.99$. Fig. 7(a)–(f) shows the normalized error norms of the estimated canonical coordinate mappings $e_{\mathbf{w}_i}$ and $e_{\mathbf{d}_i}, i \in [1,3]$ versus the sample index for ten initializations of the algorithm.

VII. CONCLUSION

A general class of two-channel CLS problems, with various constraints, has been introduced. Depending on the constraints, the two-channel solution decomposes the two data channels into one of three important coordinate systems: canonical coordi-

nates, half-canonical coordinates, or PCCA coordinates. The solution to each two-channel CLS problem is determined from a coupled (asymmetric) generalized eigenvalue problem.

A unified framework for deriving and implementing three different classes of reduced-rank Wiener filters is presented. Each class corresponds to a particular error measure for reduced-rank estimation. Two of the classes, corresponding to whitened MSE and volume of concentration ellipse, are equivalent [10], and canonical coordinates are optimal for reduced-rank Wiener filtering under these error measures. For MSE estimation, half-canonical coordinates are optimal for reduced-rank Wiener filtering [7]. We have derived all of these results in a unified way, using variations on the arguments of [4], [7], and [10].

The alternating power methods presented in this paper are simple methods for recursive computation of the canonical coordinate and half-canonical coordinate mappings. Consequently, they are simple methods for computing reduced-rank Wiener filters, regardless of the coordinate system. They may be used in deflation, block, or block-deflation mode. Moreover, they may be used in batch mode on a fixed data sample or in recursive mode on expanding data samples.

APPENDIX
SOLUTION TO THE DEFLATED COUPLED GENERALIZED EIGENVALUE PROBLEM

From (31) and (35), we have the SVD

$$\mathrm{R}_{xx}^{-1/2}\mathrm{R}_{xy}\mathrm{R}_{yy}^{-T/2} = \mathrm{R}_{xx}^{T/2}\mathrm{W}\Sigma\mathrm{D}^T\mathrm{R}_{yy}^{1/2}. \qquad (\mathrm{A.1})$$

Premultiplying (A.1) by $\mathrm{R}_{xx}^{1/2}$, postmultiplying it by $\mathrm{R}_{yy}^{T/2}$, and using $\Sigma = \mathrm{W}^T\mathrm{R}_{xy}\mathrm{D}$ yields

$$\mathrm{R}_{xy} = \mathrm{R}_{xx}\mathrm{W}\Sigma\mathrm{D}^T\mathrm{R}_{yy} = \mathrm{R}_{xx}\mathrm{W}\mathrm{W}^T\mathrm{R}_{xy}\mathrm{D}\mathrm{D}^T\mathrm{R}_{yy}. \qquad (\mathrm{A.2})$$

Assume that the first $r < m$ columns of W and D and their corresponding σ_i's have already been found. Rewrite (A.2) using (92) as

$$\mathrm{R}_{xy} = \mathrm{R}_{xx}[\mathrm{W}_r \quad \mathrm{W}_\star] \begin{bmatrix} \mathrm{W}_r^T \\ \mathrm{W}_\star^T \end{bmatrix} \mathrm{R}_{xy}[\mathrm{D}_r \quad \mathrm{D}_\star] \begin{bmatrix} \mathrm{D}_r^T \\ \mathrm{D}_\star^T \end{bmatrix} \mathrm{R}_{yy}. \tag{A.3}$$

Postmultiplying (A.3) by $\mathrm{D}_\star$ and recalling that $\mathrm{D}^T \mathrm{R}_{yy} \mathrm{D} = \mathrm{I}$ gives

$$\mathrm{R}_{xy}\mathrm{D}_\star = \mathrm{R}_{xx}[\mathrm{W}_r \quad \mathrm{W}_\star] \begin{bmatrix} \mathrm{W}_r^T \\ \mathrm{W}_\star^T \end{bmatrix} \mathrm{R}_{xy}[\mathrm{D}_r \quad \mathrm{D}_\star] \begin{bmatrix} 0 \\ \mathrm{I}_\star \end{bmatrix}$$
$$= \mathrm{R}_{xx}\left(\mathrm{W}_r\mathrm{W}_r^T \mathrm{R}_{xy}\mathrm{D}_\star + \mathrm{W}_\star\mathrm{W}_\star^T \mathrm{R}_{xy}\mathrm{D}_\star \right) \quad \text{(A.4)}$$

where $\mathrm{I}_\star$ is the $(m - r) \times (m - r)$ identity matrix. Rearranging (A.4) yields

$$\left(\mathrm{I} - \mathrm{R}_{xx}\mathrm{W}_r\mathrm{W}_r^T \right) \mathrm{R}_{xy}\mathrm{D}_\star = \mathrm{R}_{xx}\mathrm{W}_\star\mathrm{W}_\star^T\mathrm{R}_{xy}\mathrm{D}_\star = \mathrm{R}_{xx}\mathrm{W}_\star\Sigma_\star. \tag{A.5}$$

Similarly starting with $\mathrm{C}^T = \mathrm{R}_{yy}^{-1/2}\mathrm{R}_{yx}\mathrm{R}_{xx}^{-T/2}$ and following a similar procedure results in

$$\left(\mathrm{I} - \mathrm{R}_{yy}\mathrm{D}_r\mathrm{D}_r^T \right) \mathrm{R}_{yx}\mathrm{W}_\star = \mathrm{R}_{yy}\mathrm{D}_\star\Sigma_\star. \tag{A.6}$$

Equations (A.5) and (A.6) introduce a coupled generalized eigenvalue problem for $\mathrm{W}_\star, \mathrm{D}_\star$, and $\Sigma_\star$, wherein R_{xy} is deflated by $\mathrm{R}_{xx}\mathrm{W}_r\mathrm{W}_r^T\mathrm{R}_{xy}$ and R_{yx} by $\mathrm{R}_{yy}\mathrm{D}_r\mathrm{D}_r^T\mathrm{R}_{yx}$. Thus, $\mathbf{w}_{r+1}$ and $\mathbf{d}_{r+1}$ are now the generalized eigenvectors associated with the dominant eigenvalue σ_{r+1} of (A.5) and (A.6).

REFERENCES

[1] H. Hotelling, "Relation between two sets of variates," *Biometrica*, vol. 28, pp. 321–377, 1936.

[2] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 1958.

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

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

[5] L. L. Scharf and J. T. Thomas, "Wiener filters in canonical coordinates for transform coding, filtering, and quantizing," *IEEE Trans. Signal Process.*, vol. 46, pp. 647–654, Mar. 1998.

[6] L. L. Scharf, "The SVD and reduced rank signal processing," *Signal Process.*, vol. 25, pp. 113–133, 1991.

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

[8] S. V. Schell and W. A. Gardner, "Programmable canonical correlation analysis: A flexible framework for blind adaptive spatial filtering," *IEEE Trans. Signal Process.*, vol. 43, pp. 2898–2908, Dec. 1995.

[9] M. F. Kahn, W. A. Gardner, and M. A. Mow, "Programmable canonical correlation analyzers with recursion and feedback," in *Conf. Rec. Twenty-Ninth Asilomar Conf. Signals, Syst., Comput.*, vol. 1, 1996, pp. 351–356.

[10] Y. Hua, M. Nikpour, and P. Stoica, "Optimal reduced-rank estimation and filtering," *IEEE Trans. Signal Process.*, vol. 49, pp. 457–469, Mar. 2001.

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

[12] Y. Hua and M. Nikpour, "Computing the reduced-rank Wiener filter by IQMD," *IEEE Signal Process. Lett.*, vol. 6, pp. 240–242, Sep. 1999.

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

[14] G. H. Golub and C. F. V. Loan, *Matrix Computations*, Third ed. Baltimore, MD: John Hopkins Univ. Press, 1996.

[15] D. S. Watkins, *Fundamentals of Matrix Computations*, First ed: Wiley, 1991.

[16] G. Strang, *Linear Algebra and Its Applications*, Second ed. Orlando, FL: Academic, 1980.

[17] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*. London, U.K.: Oxford Univ. Press, 1965.

[18] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrica*, vol. 1, pp. 211–218, 1936.

[19] L. Mirsky, "Symmetric gauge functions and unitary invariant norms," *Quart. J. Math.*, vol. 11, pp. 50–59, 1960.

[20] S. Kayalar and H. L. Weinert, "Oblique projections: Formulas, algorithms, and error bounds," *Math. Contr. Signals Syst.*, vol. 2, pp. 33–45, 1989.

[21] R. T. Behrens and L. L. Scharf, "Signal processing applications of oblique projection operators," *IEEE Trans. Signal Process.*, vol. 42, pp. 1413–1424, Jun. 1994.

[22] G. H. Golub and W. Kahan, "Calculating the singular values and pseduo-inverse of a matrix," in *SIAM J. Numer. Anal.*, vol. 2, 1965, pp. 205–224.

[23] G. H. Golub, F. T. Luk, and M. Overton, "A block Lanczos method for computing the singular values and corresponding singular vectors of a matrix," in *ACM Trans. Math. Soft.*, vol. 7, 1981, pp. 149–169.

[24] T. F. Chan, "An improved algorithm for computing the singular value decomposition," in *ACM Trans. Math. Soft.*, vol. 8, 1982, pp. 72–83.

[25] G. H. Golub, "Tracking a few extreme singular values and vectors in signal processing," *Proc. IEEE*, vol. 78, pp. 1327–1343, Aug. 1990.

[26] D. W. Tufts and C. D. Mellissinos, "Simple, effective computation of principal eigenvectors and their eigenvalues and application to high-resolution estimation of frequencies," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, pp. 1046–1053, Oct. 1986.

Ali Pezeshki (S'95) received the B.S. and M.S. degrees in electrical engineering from the University of Tehran, Tehran, Iran, in 1999 and 2001, respectively.

Since Fall 2001, he has been pursuing the Ph.D. degree in electrical engineering at Colorado State University, Fort Collins. His current research interests lie in the area of statistical signal and array processing.

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

From 1971 to 1982, he served as Professor of electrical engineering and statistics at Colorado State University (CSU), Fort Collins. From 1982 to 1985, he was Professor and Chairman of electrical and computer engineering at the University of Rhode Island, Kingston. From 1985 to 2000, he was Professor of electrical and computer engineering at the University of Colorado, Boulder. In January 2001, he rejoined CSU as Professor of electrical and computer engineering and statistics. He has held several visiting positions in the United States and abroad: Ecole Superieure d'Electricité, Gif-sur-Yvette, France; Ecole Nationale Superieure des Télécommunications Paris, France; EURECOM, Nice, France; the University of La Plata, La Plata, Argentina; Duke University Durham, NC; the University of Wisconsin, Madison; and the University of Tromso, Tromso, Norway. His interests are in statistical signal processing as it applies to adaptive radar, sonar, and wireless communication. His most important contributions to date are to invariance theories for detection and estimation; matched and adaptive subspace detectors and estimators for radar, sonar, and data communication; and canonical decompositions for reduced dimensional filtering and quantizing. His current interests are in rapidly adaptive receiver design for space-time and frequency-time signal processing in the wireless communication channel.

Prof. Scharf was Technical Program Chair for 1980 ICASSP in Denver, CO, Tutorials Chair for ICASSP 2001 in Salt Lake City, UT, and Technical Program Chair for Asilomar 2002, Pacific Grove, CA. He is past chair of the Fellow Committee Chair for Asilomar 2002. He is past chair of the Fellow Committee for Theory and Methods and for Sensor Arrays and Multichannel Signal Processing. He has received numerous awards for his research contributions to statistical signal processing, including an IEEE Distinguished Lectureship, an IEEE Third Millennium Medal, and the Technical Achievement Award from the IEEE Signal Processing Society.

Mahmood R. Azimi-Sadjadi (S'81–M'81–SM'89) received the M.S. and Ph.D. degrees from the Imperial College of Science and Technology, University of London, London, U.K., in 1978 and 1982, respectively, both in electrical engineering with specialization in digital signal/image processing.

He is currently a full professor with the Electrical and Computer Engineering Department, Colorado State University (CSU), Fort Collins, where he is also serving as the director of the Digital Signal/Image Laboratory. His main areas of interest include digital signal and image processing, target detection, classification and tracking using broadband sonar, radar, and IR systems, adaptive filtering and system identification, and neural networks. His research efforts in these areas resulted in over 150 journal and refereed conference publications. He is the co-author of the book *Digital Filtering in One and Two Dimensions* (New York: Plenum, 1989).

Prof. Azimi-Sadjadi is the recipient of the 1999 the ABELL Teaching Award, the 1993 ASEE-Navy Senior Faculty Fellowship Award, the 1991 CSU Dean's Council Award, and the 1984 DOW chemical Outstanding Young Faculty Award. He served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and is currently serving as an Associate Editor of IEEE TRANSACTIONS ON NEURAL NETWORKS.

Yingbo Hua (S'86–M'88–SM'92–F'02) received B.S. degree from Nanjing Institute of Technology, Nanjing, China, in 1982 and the M.S. and Ph.D. degrees from Syracuse University, Syracuse, NY, in 1983 and 1988, respectively.

He was a research fellow at Syracuse and consulted for Syracuse Research and Aeritalia in Italy from 1988 to 1989. He was Lecturer from 1990 to 1992, Senior Lecturer from 1993 to 1995, and Reader from 1996 to 2001 with the University of Melbourne, Parkville, Australia. He served as a visiting professor with Hong Kong University of Science and Technology and consulted for Microsoft Research, Redmond, WA, from 1999 to 2000. Since February 2001, he has been Professor of electrical engineering with the University of California, Riverside. He is an author/coauthor of more than 240 articles in journals, conference proceedings, and books. He is a co-editor of *Signal Processing Advances in Wireless and Mobile Communications* (Englewood Cliffs, NJ: Prentice-Hall, 2001) and *High-Resolution and Robust Signal Processing* (New York: Marcel Dekker, 2003).

Dr. Hua received the Chinese Government Scholarship for Overseas Graduate Study for the period 1983 to 1984 and the Syracuse University Graduate Fellowship for 1985 to 1986. He served as Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS. He has been a member of several IEEE Signal Processing Society's Technical Committees.