A Multistage Representation of the Wiener Filter Based on Orthogonal Projections

J. Scott Goldstein, *Senior Member, IEEE,* Irving S. Reed, *Fellow, IEEE,* and Louis L. Scharf, *Fellow, IEEE*

***Abstract*—The Wiener filter is analyzed for stationary complex Gaussian signals from an information-theoretic point of view. A dual-port analysis of the Wiener filter leads to a decomposition based on orthogonal projections and results in a new multistage method for implementing the Wiener filter using a nested chain of scalar Wiener filters. This new representation of the Wiener filter provides the capability to perform an information-theoretic analysis of previous, basis-dependent, reduced-rank Wiener filters. This analysis demonstrates that the recently introduced cross-spectral metric is optimal in the sense that it maximizes mutual information between the observed and desired processes. A new reduced-rank Wiener filter is developed based on this new structure which evolves a basis using successive projections of the desired signal onto orthogonal, lower dimensional subspaces. The performance is evaluated using a comparative computer analysis model and it is demonstrated that the low-complexity multistage reduced-rank Wiener filter is capable of outperforming the more complex eigendecomposition-based methods.**

***Index Terms*—Adaptive filtering, mutual information, orthogonal projections, rank reduction, Wiener filtering.**

I. Introduction

THIS paper is concerned with the discrete-time Wiener filter. Here the desired signal, also termed a reference signal, is assumed to be a scalar process and the observed signal is assumed to be a vector process. By contrast, a scalar Wiener filter is described by a desired signal and an observed signal which are both scalar processes. The so-called matrix Wiener filter, which is not addressed in this paper, is characterized by both a desired signal and an observed signal which are vector processes.

A new approach to Wiener filtering is presented and analyzed in this paper. The process observed by the Wiener filter is first decomposed by a sequence of orthogonal projections. This decomposition has the form of an analysis filterbank, whose output is shown to be a process which is characterized by a tridiagonal covariance matrix. The corresponding error-synthesis filterbank is realized by means of a nested chain of scalar Wiener filters. These Wiener filters can be interpreted as well to be a Gram–Schmidt orthogonalization which results in an error sequence for the successive stages of the decomposed Wiener filter. This new multistage filter structure achieves the identical minimum mean-square error that is obtained by the original multidimensional Wiener filter.

The advantages realized by this new multistage Wiener filter are due to the decomposition being designed from a point of view in which the Wiener filter is treated as a dual-port problem. The multistage decomposition of the Wiener filter in the space spanned by the observed-data covariance matrix utilizes all of the information available to determine a "best" basis representation of the Wiener filter. Since all full-rank decompositions of the space spanned by the observed-data covariance matrix are simply different representations of the same Wiener filter, the term "best" is used here to describe that basis representation which comes the closest to most compactly representing the estimation energy in the lowest rank subspace *without* knowledge of the observed-data covariance matrix inverse. Clearly, if the covariance matrix inverse were known then also the Wiener filter would be known, and the rank-one subspace spanned by the Wiener filter would itself be the optimal basis vector.

Previous decompositions of the space spanned by the observed-data covariance matrix only consider the Wiener filtering problem from the perspective of a single-port problem. In other words, the decompositions considered were based on Gram–Schmidt, Householder, Jacobi, or principal-components analyses of the observed-data covariance matrix (for example, see [1]–[5] and the references contained therein). Treating the Wiener filter as a dual-port problem, however, seems more logical since the true problem at hand is not determining the best representation of the observed data alone, but instead finding the best representation of the useful portion of the observed data in the task of estimating one scalar signal from a vector observed-data process. Here, the projection of the desired signal onto the space spanned by the columns of the observed-data covariance matrix is utilized to determine the basis set. This basis set is generated in a stage-wise manner which maximizes the projected estimation energy in each orthogonal coordinate.

An interesting decomposition which at first appears similar in form to that presented in this paper is developed for constrained adaptive Wiener filters in [6] and [7]. This decomposition also treats the constrained Wiener filter as a single-port problem and does not use the constraint (in place of the desired signal, as detailed in Appendix B) in basis determination; in fact, the technique proposed in [6] and [7] removes the constraint itself from the adaptive portion of

Manuscript received February 22, 1997; revised March 25, 1998. This work was supported in part under a Grant from the Okawa Research Foundation.

J. S. Goldstein was with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2565 USA. He is now with MIT Lincoln Laboratory, Lexington, MA 02173-9108 USA (e-mail: scott@LL.MIT.EDU).

I. S. Reed is with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2565 USA.

L. L. Scharf is with the Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80309-0425 USA.

Publisher Item Identifier S 0018-9448(98)06747-9.

0018–9448/98$10.00 © 1998 IEEE

the processor. In addition, this method decomposes a rank-reduction matrix as opposed to the Wiener filter itself and results in a modular structure which is not a nested recursion. Thus while the modular structure in [6] and [7] is very interesting in its own right, neither the decomposition nor the recursion technique are similar to the new multistage Wiener filter presented here.

Reduced-rank Wiener filtering is concerned with the compression, or reduction in dimensionality, of the observed data prior to Wiener filtering. The purpose of rank reduction is to obtain a minimum mean-square error which is as close as possible to that obtainable if all of the observed data were available to linearly estimate the desired signal. The new multistage Wiener filter structure leads to a natural means to obtain rank reduction.

The performance of such a reduced-rank multistage Wiener filter is compared by computer analysis to the well-known principal components and the lesser known cross-spectral methods of rank reduction. This analysis demonstrates that the new method of rank reduction, using quite simply a truncated version of the above-described nested chain of scalar Wiener filters, is capable of outperforming these previous approaches. Also an information-theoretic analysis of entropy and mutual information is now possible due to this new structure, which provides insight into these results. In particular, it is demonstrated that the cross-spectral method of rank reduction maximizes the mutual information as a function of rank relative to the eigenvector basis. The new reduced-rank multistage Wiener filter does not utilize eigendecomposition or eigenvector-pruning techniques.

Section II provides a brief description of the Wiener filter in terms of the framework to be used in the remainder of this paper. An introduction and analysis of this new representation of the Wiener filter is presented in Section III. A summary of previous reduced-rank Wiener filtering techniques is provided in Section IV, where the reduced-rank multistage Wiener filter is presented and its performance is evaluated via a comparative computer analysis. Concluding remarks are given in Section V.

II. Preliminaries

The classical Wiener filtering problem is depicted in Fig. 1, where there is a desired scalar signal $d_0(k)$, an N-dimensional observed-data vector $\boldsymbol{x}_0(k)$, and an N-dimensional Wiener filter $\boldsymbol{w}_{x_0}$. The error signal is denoted by $\varepsilon_0(k)$. The Wiener filter requires that the signals be modeled as wide-sense-stationary random processes, and the information-theoretic analysis to be considered makes as well the complex Gaussian assumption. Thus in both cases there is no loss in generality to assume that all signals are zero-mean, jointly stationary, complex Gaussian random processes. The covariance matrix of the input vector process $\boldsymbol{x}_0(k)$ is given by

$$\boldsymbol{R}_{x_0} = \boldsymbol{E}[\boldsymbol{x}_0(k)\boldsymbol{x}_0^H(k)] \tag{1}$$

where $\boldsymbol{E}[(\cdot)]$ denotes the expected-value operator and $(\cdot)^H$ is the complex conjugate transpose operator. Similarly, the

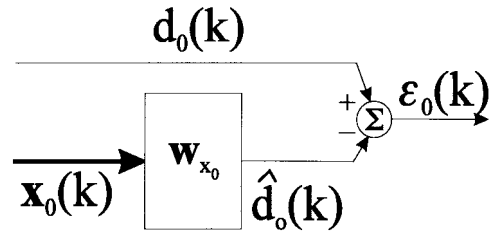

Fig. 1. The classical Wiener filter.

variance of the desired process $d_0(k)$ is

$$\sigma_{d_0}^2 = \boldsymbol{E}[|d_0(k)|^2]. \tag{2}$$

The complex cross-correlation vector between the processes $d_0(k)$ and $\boldsymbol{x}_0(k)$ is given by

$$\boldsymbol{r}_{x_0 d_0} = \boldsymbol{E}[\boldsymbol{x}_0(k) d_0^*(k)] \tag{3}$$

where $(\cdot)^*$ is the complex conjugate operator. The optimal linear filter, which minimizes the mean-square error between the desired signal $d_0(k)$ and its estimate,

$$\hat{d}_0(k) = \boldsymbol{w}_{x_0}^H \boldsymbol{x}_0(k), \tag{4}$$

is the classical Wiener filter

$$\boldsymbol{w}_{x_0} = \boldsymbol{R}_{x_0}^{-1} \boldsymbol{r}_{x_0 d_0} \tag{5}$$

for complex stationary processes. The resulting error is

$$\varepsilon_0(k) = d_0(k) - \hat{d}_0(k). \tag{6}$$

The minimum mean-square error (MMSE) is

$$\begin{aligned} \xi_0 &= \boldsymbol{E}[|\varepsilon_0(k)|^2] \\ &= \sigma_{d_0}^2 - \boldsymbol{r}_{x_0 d_0}^H \boldsymbol{R}_{x_0}^{-1} \boldsymbol{r}_{x_0 d_0} \\ &= \sigma_{d_0}^2 (1 - k_{x_0 d_0}^2) \end{aligned} \tag{7}$$

where the squared canonical correlation $k_{x_0 d_0}^2$ [8]–[11] is

$$k_{x_0 d_0}^2 = \frac{\boldsymbol{r}_{x_0 d_0}^H \boldsymbol{R}_{x_0}^{-1} \boldsymbol{r}_{x_0 d_0}}{\sigma_{d_0}^2}. \tag{8}$$

As will be seen in Section III-D, the squared canonical correlation provides a measure of the information present in the observed vector random process $\boldsymbol{x}_0(k)$ that is used to estimate the scalar random process $d_0(k)$.

Because of the assumed Gaussianity, the self-information or entropy of the signal process $d_0(k)$ is given by (see [12]–[15])

$$H(d_0) = \frac{1}{2} \log(2\pi e \sigma_{d_0}^2) \tag{9}$$

and the entropy of the vector input process $\boldsymbol{x}_0(k)$ is

$$H(\boldsymbol{x}_0) = \frac{1}{2} \log[(2\pi e)^N ||\boldsymbol{R}_{x_0}||] \tag{10}$$

where $||(\cdot)||$ denotes the determinant operator. Next define an augmented vector $\boldsymbol{\kappa}(k)$ by

$$\boldsymbol{\kappa}(k) = \begin{bmatrix} d_0(k) \\ \boldsymbol{x}_0(k) \end{bmatrix}. \tag{11}$$

Then, using (1)–(3) and (11), the covariance matrix associated

with the vector process $\boldsymbol{\kappa}(k)$ is given by

$$\boldsymbol{R}_{\kappa} = \begin{bmatrix} \sigma_{d_0}^2 & \boldsymbol{r}_{x_0 d_0}^H \\ \boldsymbol{r}_{x_0 d_0} & \boldsymbol{R}_{x_0} \end{bmatrix} \tag{12}$$

so that, by (10), the joint entropy of the random processes $d_0(k)$ and $\boldsymbol{x}_0(k)$ is given by

$$H(d_0, \boldsymbol{x}_0) = \frac{1}{2}\log[(2\pi e)^{N+1}||\boldsymbol{R}_{\kappa}||]. \tag{13}$$

Thus by Shannon's chain rule the conditional entropy $H(d_0|\boldsymbol{x}_0)$, or what Shannon called the equivocation of $d_0(k)$ given $\boldsymbol{x}_0(k)$, is given by

$$H(d_0|\boldsymbol{x}_0) = H(d_0, \boldsymbol{x}_0) - H(\boldsymbol{x}_0). \tag{14}$$

Now the mutual information $I(d_0, \boldsymbol{x}_0)$ is the relative entropy between the joint distribution and the product distribution. That is, $I(d_0, \boldsymbol{x}_0)$ represents the reduction in uncertainty of $d_0(k)$ due to the knowledge of $\boldsymbol{x}_0(k)$. This mutual information is given by

$$\begin{aligned} I(d_0, \boldsymbol{x}_0) &= H(d_0) - H(d_0|\boldsymbol{x}_0) \\ &= H(d_0) + H(\boldsymbol{x}_0) - H(d_0, \boldsymbol{x}_0). \end{aligned} \tag{15}$$

By definition the Wiener filter minimizes the mean-square error between the desired process and the filtered observed process. Therefore, the operation of this filter must determine that portion of the observed process which contains the most information about the desired process. Intuitively for Gaussian processes one expects that a minimization of the mean-square error and a maximization of the mutual information are equivalent. This insight is mathematically realized through the multistage representation of the Wiener filter presented next in this paper.

III. The Multistage Wiener Filter

The analysis developed herein emphasizes the standard, unconstrained Wiener filter. It is noted that an identical approach also solves the problem of quadratic minimization with linear constraints [16] and the joint-process estimation problem, both of which can be interpreted as a constrained Wiener filter. The partitioned solution presented in [16] decomposes the constraint in such a manner that the resulting Wiener filter is unconstrained, as is further explored in the example given in Section IV-C and Appendix B. It is further noted that other constraints also may be decomposed similarly [17]. Thus the constrained Wiener filter can be represented as an unconstrained Wiener filter with a prefiltering operation determined by the constraint. It is seen next that the unconstrained Wiener filter can also be represented as a nested chain of constrained Wiener filters.

This new representation of the Wiener filter is achieved by a multistage decomposition. This decomposition forms two subspaces at each stage; one in the direction of the cross-correlation vector at the previous stage and one in the subspace orthogonal to this direction. Then the data orthogonal to the cross-correlation vector is decomposed again in the same manner, stage by stage. This process reduces the dimension of the data vector at each stage. Thus a new coordinate system for Wiener filtering is determined via a pyramid-like structured decomposition which serves as an analysis filterbank. This decomposition decorrelates the observed vector process at lags greater than one, resulting in a tridiagonal covariance matrix associated with the transformed vector process.

A nested chain of scalar Wiener filters form an error-synthesis filterbank, which operates on the output of the analysis filterbank to yield an error process with the same MMSE as the standard multidimensional Wiener filter. It is demonstrated also that the error-synthesis filterbank can be interpreted as an implementation of a Gram–Schmidt orthogonalization process.

A. An Equivalent Wiener Filtering Model

To obtain this new multistage decomposition, note that the preprocessing of the observation data by a full-rank, nonsingular, linear operator prior to Wiener filtering does not modify the MMSE. This fact is demonstrated in Appendix A. Now consider the particular nonsingular operator $\boldsymbol{T}_1$ with the structure

$$\boldsymbol{T}_1 = \begin{bmatrix} \boldsymbol{h}_1^H \\ \boldsymbol{B}_1 \end{bmatrix} \tag{16}$$

where $\boldsymbol{h}_1$ is the normalized cross-correlation vector, a unit vector in the direction of $\boldsymbol{r}_{x_0 d_0}$, given by

$$\boldsymbol{h}_1 = \frac{\boldsymbol{r}_{x_0 d_0}}{\sqrt{\boldsymbol{r}_{x_0 d_0}^H \boldsymbol{r}_{x_0 d_0}}} \tag{17}$$

and $\boldsymbol{B}_1$ is an $(N-1)\times N$ operator which spans the nullspace of $\boldsymbol{r}_{x_0 d_0}$; i.e., $\boldsymbol{B}_1$ is the blocking matrix which annihilates those signal components in the direction of the vector $\boldsymbol{r}_{x_0 d_0}$ [18], [19] such that $\boldsymbol{B}_1 \boldsymbol{h}_1 = \boldsymbol{0}$.

Two fast algorithms for obtaining such a unitary matrix $\boldsymbol{T}_1$ are described in [20] which use either the singular-value decomposition or the QR decomposition. For a $\boldsymbol{T}_1$ which is nonsingular, but not unitary, a new, very efficient, implementation of the blocking matrix $\boldsymbol{B}_1$ is presented in Appendix A.

Let the new transformed data vector, formed by $\boldsymbol{T}_1$ operating on the observed-data vector, be given by

$$\boldsymbol{z}_1(k) = \boldsymbol{T}_1 \boldsymbol{x}_0(k) = \begin{bmatrix} \boldsymbol{h}_1^H \boldsymbol{x}_0(k) \\ \boldsymbol{B}_1 \boldsymbol{x}_0(k) \end{bmatrix} = \begin{bmatrix} d_1(k) \\ \boldsymbol{x}_1(k) \end{bmatrix}. \tag{18}$$

The transform-domain Wiener filter with the preprocessor $\boldsymbol{T}_1$ is shown in Fig. 2. The Wiener filter for the transformed process is computed now to have the form

$$\boldsymbol{w}_{z_1} = \boldsymbol{R}_{z_1}^{-1} \boldsymbol{r}_{z_1 d_0}. \tag{19}$$

Next, the covariance matrix $\boldsymbol{R}_{z_1}$, its inverse $\boldsymbol{R}_{z_1}^{-1}$, and the cross-correlation vector $\boldsymbol{r}_{z_1 d_0}$ are expressed as

$$\boldsymbol{R}_{z_1} = \begin{bmatrix} \sigma_{d_1}^2 & \boldsymbol{r}_{x_1 d_1}^H \\ \boldsymbol{r}_{x_1 d_1} & \boldsymbol{R}_{x_1} \end{bmatrix} \tag{20}$$

$$\boldsymbol{R}_{z_1}^{-1} = \xi_1^{-1} \begin{bmatrix} 1 & -\boldsymbol{r}_{x_1 d_1}^H \boldsymbol{R}_{x_1}^{-1} \\ -\boldsymbol{R}_{x_1}^{-1} \boldsymbol{r}_{x_1 d_1} & \boldsymbol{R}_{x_1}^{-1}(\xi_1 \boldsymbol{I} + \boldsymbol{r}_{x_1 d_1} \boldsymbol{r}_{x_1 d_1}^H \boldsymbol{R}_{x_1}^{-1}) \end{bmatrix} \tag{21}$$

$$\boldsymbol{r}_{z_1 d_0} = \boldsymbol{E}[\boldsymbol{z}_1(k) d_0^*(k)] = \boldsymbol{T}_1 \boldsymbol{r}_{x_0 d_0} = [\delta_1 \;\; 0 \;\; \cdots \;\; 0]^T \tag{22}$$

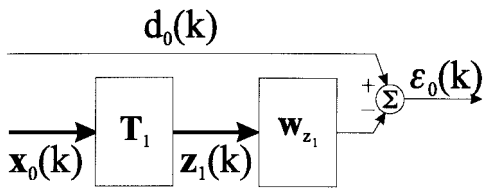

Fig. 2. The transform-domain Wiener filter.

where $(\cdot)^T$ denotes the standard matrix transpose operator and, by (16)–(18), the scalar δ_1 in (22) is obtained as

$$\delta_1 = \boldsymbol{h}_1^H \boldsymbol{r}_{x_0 d_0} = \sqrt{\boldsymbol{r}_{x_0 d_0}^H \boldsymbol{r}_{x_0 d_0}}. \tag{23}$$

The variance of $d_1(k)$ in (18) is calculated to be

$$\sigma_{d_1}^2 = \boldsymbol{h}_1^H \boldsymbol{R}_{x_0} \boldsymbol{h}_1. \tag{24}$$

The covariance matrix $\boldsymbol{R}_{x_1}$ is given by

$$\boldsymbol{R}_{x_1} = \boldsymbol{E}[\boldsymbol{x}_1(k)\boldsymbol{x}_1^H(k)] = \boldsymbol{B}_1 \boldsymbol{R}_{x_0} \boldsymbol{B}_1^H. \tag{25}$$

The cross-correlation vector $\boldsymbol{r}_{x_1 d_1}$ is computed to be

$$\boldsymbol{r}_{x_1 d_1} = \boldsymbol{E}[\boldsymbol{x}_1(k) d_1^*(k)] = \boldsymbol{B}_1 \boldsymbol{R}_{x_0} \boldsymbol{h}_1 \tag{26}$$

and the matrix $\boldsymbol{R}_{z_1}^{-1}$ in (21) is determined by the matrix inversion lemma for partitioned matrices [16]. In terms of the joint-process covariance matrix $\boldsymbol{R}_\kappa$ in (12), the transformations described above in (20)–(26) may be represented by

$$\begin{aligned}\boldsymbol{R}_{\kappa_1} &= \tilde{\boldsymbol{T}}_1 \boldsymbol{R}_\kappa \tilde{\boldsymbol{T}}_1^H \\ &= \tilde{\boldsymbol{T}}_1 \begin{bmatrix} \sigma_{d_0}^2 & \boldsymbol{r}_{x_0 d_0}^H \\ \boldsymbol{r}_{x_0 d_0} & \boldsymbol{R}_{x_0} \end{bmatrix} \tilde{\boldsymbol{T}}_1^H \\ &= \begin{bmatrix} \sigma_{d_0}^2 & \boldsymbol{r}_{z_1 d_0}^H \\ \boldsymbol{r}_{z_1 d_0} & \boldsymbol{R}_{z_1} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{d_0}^2 & [\delta_1^* \quad 0 \quad \cdots \quad 0] \\ \begin{bmatrix} \delta_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} & \begin{bmatrix} \sigma_{d_1}^2 & \boldsymbol{r}_{x_1 d_1}^H \\ \boldsymbol{r}_{x_1 d_1} & \boldsymbol{R}_{x_1} \end{bmatrix} \end{bmatrix}\end{aligned} \tag{27}$$

where

$$\tilde{\boldsymbol{T}}_1 = \begin{bmatrix} 1 & \boldsymbol{0}^T \\ \boldsymbol{0} & \boldsymbol{T}_1 \end{bmatrix}. \tag{28}$$

The structure of the matrix $\boldsymbol{R}_{z_1}$ in (20), its inverse in (21), and the diagrams in Figs. 1 and 2 suggest that a new $(N-1)$-dimensional "weight" vector $\boldsymbol{w}_2$ be defined by

$$\boldsymbol{w}_2 = \boldsymbol{R}_{x_1}^{-1} \boldsymbol{r}_{x_1 d_1} \tag{29}$$

which is the Wiener filter for estimating the scalar $d_1(k)$ from the vector $\boldsymbol{x}_1(k)$. Then a new error $\varepsilon_1(k)$, given by

$$\varepsilon_1(k) = d_1(k) - \boldsymbol{w}_2^H \boldsymbol{x}_1(k), \tag{30}$$

can be defined for the new Wiener filter $\boldsymbol{w}_2$ which is similar in form to the Wiener filter depicted in Fig. 1. The variance of the error $\varepsilon_1(k)$ in (30) is computed readily to be

$$\xi_1 = \boldsymbol{E}[|\varepsilon_1(k)|^2] = \sigma_{d_1}^2 - \boldsymbol{r}_{x_1 d_1}^H \boldsymbol{R}_{x_1}^{-1} \boldsymbol{r}_{x_1 d_1} = \sigma_{d_1}^2 - \boldsymbol{w}_2^H \boldsymbol{r}_{x_1 d_1}. \tag{31}$$

Since ξ_1 is the covariance of the scalar process $\varepsilon_1(k)$, the identical MMSE, ξ_0, is achieved by the filtering diagram in Fig. 3, where the new scalar Wiener filter is defined by

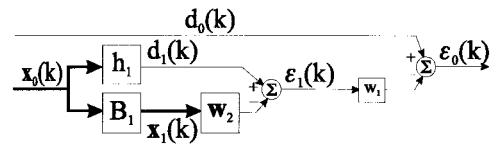

Fig. 3. The first stage of the decomposition.

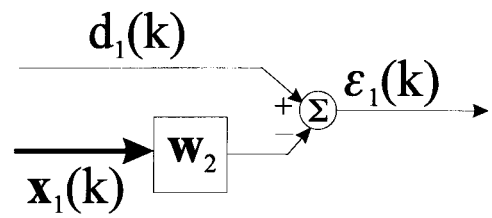

Fig. 4. The first chain in the nested recursion.

$$w_1 = \xi_1^{-1} r_{\varepsilon_1 d_0} = \xi_1^{-1} \delta_1 \tag{32}$$

and where by (18), (22), (23), and (30) the identity of the correlation between the scalar processes $\varepsilon_1(k)$ and $d_0(k)$ with δ_1 is shown by

$$\begin{aligned} r_{\varepsilon_1 d_0} &= \boldsymbol{E}[\varepsilon_1(k) d_0^*(k)] \\ &= (\boldsymbol{h}_1^H - \boldsymbol{w}_2^H \boldsymbol{B}_1) \boldsymbol{r}_{x_0 d_0} = \boldsymbol{h}_1^H \boldsymbol{r}_{x_0 d_0} = \delta_1. \end{aligned} \tag{33}$$

It is evident by (19)–(21) that the filtering diagrams in Figs. 2 and 3 are identical since

$$\boldsymbol{w}_{z_1} = \boldsymbol{R}_{z_1}^{-1} \boldsymbol{r}_{z_1 d_0} = \xi_1^{-1} \delta_1 \begin{bmatrix} 1 \\ -\boldsymbol{R}_{x_1}^{-1} \boldsymbol{r}_{x_1 d_1} \end{bmatrix} = w_1 \begin{bmatrix} 1 \\ -\boldsymbol{w}_2 \end{bmatrix}. \tag{34}$$

Note that the scalar ξ_1 is also the MMSE of the nested lower dimensional Wiener filter with a new scalar signal $d_1(k)$ and a new observed signal vector $\boldsymbol{x}_1(k)$. The first stage of this decomposition partitions the N-dimensional Wiener filter into a scalar Wiener filter and an $(N-1)$-dimensional vector Wiener filter, where the reduced-dimension vector filter spans a space which is orthogonal to the space spanned by the scalar filter. Also note that the nested filtering structure in Fig. 3, which uses $\boldsymbol{w}_2$ to estimate $d_1(k)$ from $\boldsymbol{x}_1(k)$, may be interpreted as a constrained Wiener filter which minimizes the error $\varepsilon_1(k)$ subject to the constraint that the desired signal $d_1(k)$ has the gain and phase provided by the filter $\boldsymbol{h}_1$.

B. A Multistage Representation of the Wiener Filter

The first stage decomposition results in the structure depicted in Fig. 3. The new $(N-1)$-dimensional vector Wiener filter $\boldsymbol{w}_2$ operates on the transformed $(N-1)$-dimensional data vector $\boldsymbol{x}_1(k)$ to estimate the new scalar signal $d_1(k)$, as shown in Fig. 4. This represents a Wiener filter which is identical in form to the original N-dimensional Wiener filter, except that it is one dimension smaller. Thus a recursion of scalar Wiener filters can be derived by following the outline given in Section III-A until the dimension of both the data and the corresponding Wiener filter is reduced to one at level $(N-1)$ in the tree. The error signal at each stage serves as the scalar observed process for the Wiener filter at the next stage. At

each stage i, $1 \le i \le N-1$, the normalized cross-correlation vector $\boldsymbol{h}_i$ is computed in the same manner as (17) to be

$$\boldsymbol{h}_i = \frac{\boldsymbol{r}_{x_{i-1}d_{i-1}}}{\sqrt{\boldsymbol{r}^H_{x_{i-1}d_{i-1}}\boldsymbol{r}_{x_{i-1}d_{i-1}}}} = \frac{\boldsymbol{r}_{x_{i-1}d_{i-1}}}{\boldsymbol{\delta}_i}. \tag{35}$$

The blocking matrix, $\boldsymbol{B}_i = \text{null}\{\boldsymbol{h}_i\}$, may be computed using the method detailed in Appendix A, that presented in [20], or any other method which results in a valid $\boldsymbol{T}_i$. The covariance matrix $\boldsymbol{R}_{x_i}$ is computed corresponding to (25) as follows:

$$\begin{aligned}\boldsymbol{R}_{x_i} &= \boldsymbol{B}_i\boldsymbol{R}_{x_{i-1}}\boldsymbol{B}_i^H \\ &= \left(\prod_{k=i}^{1}\boldsymbol{B}_k\right)\boldsymbol{R}_{x_0}\left(\prod_{k=1}^{i}\boldsymbol{B}_k^H\right), \qquad 1 \le i \le N-1\end{aligned} \tag{36}$$

and the cross-correlation vector $\boldsymbol{r}_{x_i d_i}$ is found recursively in the manner of (26) to be

$$\begin{aligned}\boldsymbol{r}_{x_i d_i} &= \boldsymbol{B}_i\boldsymbol{R}_{x_{i-1}}\boldsymbol{h}_i \\ &= \left(\prod_{k=i}^{1}\boldsymbol{B}_k\right)\boldsymbol{R}_{x_0}\left(\prod_{k=1}^{i-1}\boldsymbol{B}_k^H\right)\boldsymbol{h}_i, \quad 1 \le i \le N-1.\end{aligned} \tag{37}$$

The scalar signal $d_i(k)$ and the $(N-i)$-dimensional observed-data vector $\boldsymbol{x}_i(k)$ at the ith stage are found in accordance with (18) as follows:

$$d_i(k) = \boldsymbol{h}_i^H\boldsymbol{x}_{i-1}(k), \qquad 1 \le i \le N-1 \tag{38}$$

$$\boldsymbol{x}_i(k) = \boldsymbol{B}_i\boldsymbol{x}_{i-1}(k), \qquad 1 \le i \le N-1. \tag{39}$$

The error signals at each stage, in analogy to (30), are given by

$$\varepsilon_i(k) = d_i(k) - w^*_{i+1}\varepsilon_{i+1}(k) \tag{40}$$

where it is notationally convenient to define the scalar output of the last signal blocking matrix in the chain $x_{N-1}(k)$ to be the Nth element of both the sequences $d_i(k)$ and $\varepsilon_i(k)$ as follows:

$$x_{N-1}(k) = d_N(k) = \varepsilon_N(k). \tag{41}$$

The variances associated with the signals $d_i(k)$, $1 \le i \le N-1$, are defined by

$$\sigma^2_{d_i} = \boldsymbol{E}[|d_i(k)|^2] = \boldsymbol{h}_i^H\boldsymbol{R}_{x_{i-1}}\boldsymbol{h}_i \tag{42}$$

where $\sigma^2_{d_N} = R_{x_{N-1}}$. The scalar cross-correlations $r_{\varepsilon_{i+1}d_i}$ are computed in the same manner as (33) to be

$$r_{\varepsilon_{i+1}d_i} = \delta_{i+1} = \sqrt{\boldsymbol{r}^H_{x_i d_i}\boldsymbol{r}_{x_i d_i}} \tag{43}$$

where, using (41), the last term of the recursion in (43) is provided by the identity

$$\delta_N = r_{\varepsilon_N d_{N-1}} = r_{x_{N-1}d_{N-1}}. \tag{44}$$

The scalar Wiener filters w_i are found from the Wiener–Hopf equation to be

$$w_i = \xi_i^{-1}\delta_i \tag{45}$$

where, for $i < N$, the MMSE recursion yields

$$\xi_i = \sigma^2_{d_i} - w^*_{i+1}\delta_{i+1} \tag{46}$$

TABLE I
RECURSION EQUATIONS

FORWARD
$\mathbf{r}_{x_i d_i} = \mathbf{E}\left[x_i(k)d_i^*(k)\right]$
$\delta_{i+1} = \sqrt{\mathbf{r}^H_{x_i d_i}\mathbf{r}_{x_i d_i}}$
$\mathbf{h}_{i+1} = \frac{\mathbf{r}_{x_i d_i}}{\delta_{i+1}}$
$\mathbf{B}_{i+1} = null\left\{\mathbf{h}_{i+1}\right\}$
$d_{i+1}(k) = \mathbf{h}^H_{i+1}\mathbf{x}_i(k)$
$\mathbf{x}_{i+1}(k) = \mathbf{B}_{i+1}\mathbf{x}_i(k)$
$x_{N-1}(k) = d_N(k) = \varepsilon_N(k)$
BACKWARD
$\xi_N = \mathbf{E}\left[\lvert x_{N-1}(k)\rvert^2\right]$
$\delta_N = r_{x_{N-1}d_{N-1}}$
$w_i = \xi_i^{-1}\delta_i$
$\varepsilon_i(k) = d_i(k) - w^*_{i+1}\varepsilon_{i+1}(k)$
$\xi_i = \sigma^2_{d_i} - \xi^{-1}_{i+1}\lvert\delta_{i+1}\rvert^2$

similar to the results of (29), (31), and (32). In accordance with (41), the MMSE of the last stage is given by $\xi_N = \sigma^2_{d_N}$.

The complete series of required recursion relationships are listed in Table I. An example of this decomposition for $N = 4$ is provided in Fig. 5. Note that this new multistage Wiener filter does not require an estimate of the covariance matrix or its inverse when the statistics are unknown since the only requirements are for estimates of the cross-correlation vectors and scalar correlations, which can be calculated directly from the data.

C. Analysis of the Multistage Wiener Filter

This new Wiener filter structure is naturally partitioned into an analysis filterbank and a synthesis filterbank. The analysis filterbank is pyramidal, and the resulting tree structure successively refines the signal $\boldsymbol{x}_i(k)$ in terms of $d_{i+1}(k)$, its component in the direction of the cross-correlation vector, and $\boldsymbol{x}_{i+1}(k)$, its components in the orthogonal subspace. The subspaces formed at level i and level $i+1$ in the tree satisfy the direct-sum relationship

$$S_{x_i} = S_{d_{i+1}} \oplus S_{x_{i+1}} \tag{47}$$

where S_{x_i} denotes the linear subspace spanned by the columns of the covariance matrix $\boldsymbol{R}_{x_i}$ and $(\cdot) \oplus (\cdot)$ represents a direct sum.

The operation of the analysis filterbanks are combined next into one lossless $N \times N$ transfer matrix $\boldsymbol{L}_N$, which is given

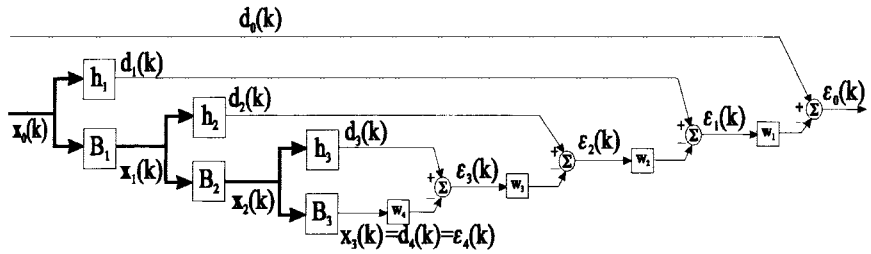

Fig. 5. The nested chain of scalar Wiener filters for $N = 4$.

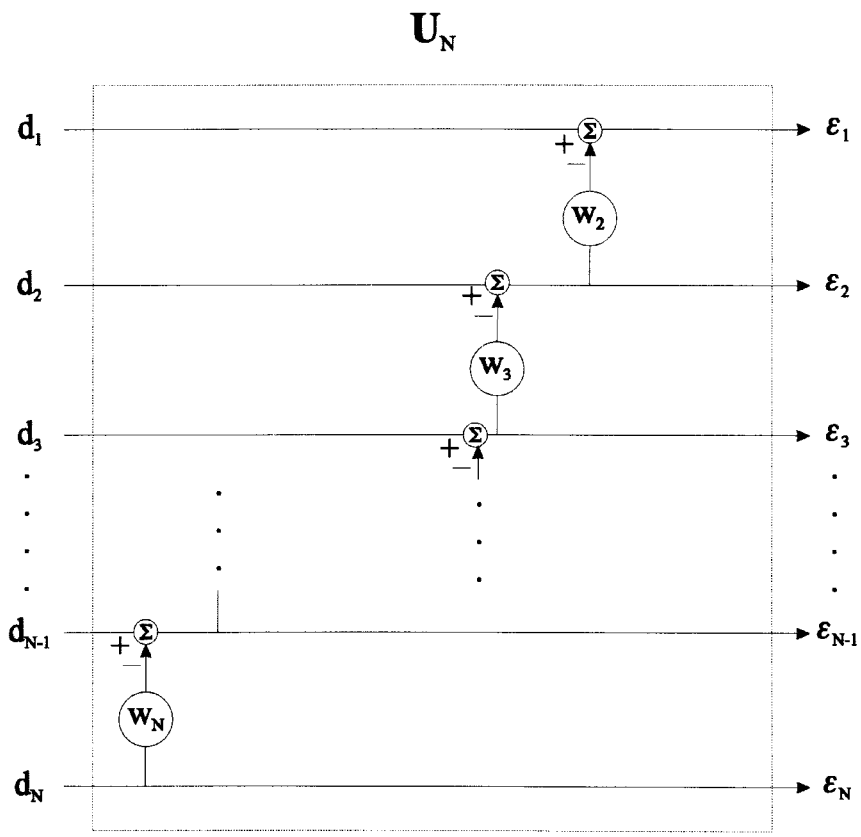

Fig. 6. The Gram–Schmidt interpretation of the synthesis filterbank.

as follows:

$$\boldsymbol{L}_N = \begin{bmatrix} \boldsymbol{h}_1^H \\ \boldsymbol{h}_2^H \boldsymbol{B}_1 \\ \vdots \\ \boldsymbol{h}_{N-1}^H \prod_{i=N-2}^{1} \boldsymbol{B}_i \\ \prod_{i=N-1}^{1} \boldsymbol{B}_i \end{bmatrix} \tag{48}$$

where the dimension of the matrix $\boldsymbol{B}_i$ is $(N-i)\times(N-i+1)$ and the vector $\boldsymbol{h}_i$ is of dimension $N-i+1$. The matrix filter $\boldsymbol{L}_N$ operates on the observed-data vector $\boldsymbol{x}_0(k)$ to produce an $N \times 1$ output vector $\tilde{\boldsymbol{d}}(k)$, defined by

$$\tilde{\boldsymbol{d}}(k) = \boldsymbol{L}_N \boldsymbol{x}_0(k) = [d_1(k) \quad d_2(k) \quad \cdots \quad d_N(k)]^T \tag{49}$$

where $d_N(k) = x_{N-1}(k)$. The definition of the matrix filters $\boldsymbol{h}_i$ and $\boldsymbol{B}_i$ guarantees that $\boldsymbol{L}_N$ decorrelates all lags in the process $\tilde{\boldsymbol{d}}(k)$ greater than one. This means that the covariance matrix

$$\boldsymbol{R}_{\tilde{d}} = \boldsymbol{L}_N \boldsymbol{R}_{x_0} \boldsymbol{L}_N^H \tag{50}$$

is tridiagonal.

The synthesis filterbank is composed of N nested scalar Wiener filters, where the input to each weight is the error signal from the previous stage. This filtering operation may be interpreted also as an implementation of the Gram–Schmidt algorithm, as shown in Fig. 6, which operates on the sequentially correlated random vector $\tilde{\boldsymbol{d}}(k)$, defined in (49), to form the uncorrelated error vector $\tilde{\boldsymbol{\varepsilon}}(k)$ where

$$\tilde{\boldsymbol{\varepsilon}}(k) = [\varepsilon_1(k) \quad \varepsilon_2(k) \quad \cdots \quad \varepsilon_N(k)]^T \tag{51}$$

and the last element in this recursion is given by (41).

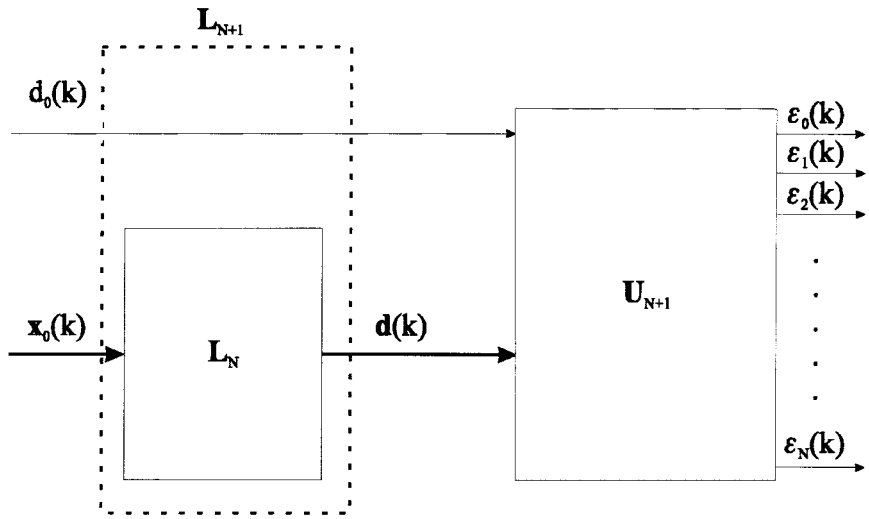

Fig. 7. The combined multiresolution Gram–Schmidt interpretation of the Wiener filter.

The equivalent Gram–Schmidt matrix may be computed by the backward recursion relations in Table I. Explicitly, the $N \times N$ Gram–Schmidt operator $\boldsymbol{U}_N$ is an upper-triangular matrix given by (52) at the bottom of this page. Finally, the $(N+1) \times (N+1)$ matrix $\boldsymbol{U}_{N+1}$ is given by

$$\boldsymbol{U}_{N+1} = \begin{bmatrix} 1 & -\boldsymbol{w}_{z_1}^H \\ \boldsymbol{0} & \boldsymbol{U}_N \end{bmatrix} \tag{53}$$

where the N-dimensional Wiener filter $\boldsymbol{w}_{z_1}$, defined in (19), is now given equivalently by

$$\boldsymbol{w}_{z_1}^H = \begin{bmatrix} w_1^* & -w_1^* w_2^* & \cdots & (-1)^{N+1} \prod_{i=1}^{N} w_i^* \end{bmatrix}. \tag{54}$$

Define the $(N+1)$-dimensional vector $\boldsymbol{d}(k)$ to be

$$\boldsymbol{d}(k) = \begin{bmatrix} d_0(k) \\ \tilde{\boldsymbol{d}}(k) \end{bmatrix}. \tag{55}$$

Thus the complete Gram–Schmidt orthogonalization, which results in the error signal vector

$$\boldsymbol{\varepsilon}(k) = \boldsymbol{U}_{N+1} \boldsymbol{d}(k) = \begin{bmatrix} \varepsilon_0 \\ \tilde{\boldsymbol{\varepsilon}}(k) \end{bmatrix} \tag{56}$$

is realized with N weights, as opposed to the $N(N+1)/2$ weights which are normally required to obtain the $(N+1)$-dimensional error process $\boldsymbol{\varepsilon}(k)$ from an $(N+1)$-dimensional process such as $\boldsymbol{d}(k)$. Fig. 7 depicts the combined multistage decomposition and Gram–Schmidt orthogonalization interpretation of the Wiener filter.

The error covariance matrix $\boldsymbol{R}_{\tilde{\varepsilon}}$, which is associated with the error process $\tilde{\boldsymbol{\varepsilon}}(k)$ in (51), is a diagonal matrix given by

$$\boldsymbol{R}_{\tilde{\varepsilon}} = \operatorname{diag}([\xi_1 \quad \xi_2 \quad \cdots \quad \xi_N]) \tag{57}$$

where the operator $\operatorname{diag}(\cdot)$, with a vector operand, represents a diagonal matrix whose only nonzero elements are along the main diagonal and provided by the corresponding element of the operand. Similarly, the error covariance matrix $\boldsymbol{R}_{\varepsilon}$ associated with the error process $\boldsymbol{\varepsilon}(k)$ in (56) is also a diagonal matrix given by

$$\boldsymbol{R}_{\varepsilon} = \operatorname{diag}([\xi_0 \quad \xi_1 \quad \cdots \quad \xi_N]). \tag{58}$$

The minimum mean-square error of the ith stage ξ_i is explicitly computed in Table I.

One final interpretation of the multistage Wiener filter is realized through an examination of the analysis and error-synthesis filterbanks. Define the $(N+1) \times (N+1)$ transformations

$$\tilde{\boldsymbol{T}}_j = \begin{bmatrix} \boldsymbol{I}_j & \boldsymbol{0} \\ \boldsymbol{0} & \begin{bmatrix} \boldsymbol{h}_j^H \\ \boldsymbol{B}_j \end{bmatrix} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_j & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_j \end{bmatrix} \tag{59}$$

$$\boldsymbol{U}_N = \begin{bmatrix} 1 & -w_2^* & w_2^* w_3^* & \cdots & (-1)^N \prod_{i=2}^{N-1} w_i^* & (-1)^{N+1} \prod_{i=2}^{N} w_i^* \\ 0 & 1 & -w_3^* & \cdots & (-1)^{N-1} \prod_{i=3}^{N-1} w_i^* & (-1)^N \prod_{i=3}^{N} w_i^* \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -w_N^* \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}. \tag{52}$$

where the identity matrix of rank j is denoted $\boldsymbol{I}_j$ and the vectors $\boldsymbol{h}_j$ and $\boldsymbol{B}_j$ are detailed in Table I for $i = 1, 2, \cdots, N-1$. The matrix $\boldsymbol{L}_{N+1}$ is formed by extending the definition of matrix $\boldsymbol{L}_N$ in (48) as follows:

$$\boldsymbol{L}_{N+1} = \begin{bmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{L}_N \end{bmatrix}. \tag{60}$$

Then the $N+1$-dimensional vector $\boldsymbol{d}(k)$ in (55) is given by

$$\boldsymbol{d}(k) = \boldsymbol{L}_{N+1}\boldsymbol{\kappa}(k) = \tilde{\boldsymbol{T}}_{N-1}\tilde{\boldsymbol{T}}_{N-2}\cdots\tilde{\boldsymbol{T}}_1\boldsymbol{\kappa}(k) = \begin{bmatrix} d_0(k) \\ \tilde{\boldsymbol{d}}(k) \end{bmatrix} \tag{61}$$

where the augmented vector $\boldsymbol{\kappa}(k)$ is defined in (11). The tridiagonal covariance matrix $\boldsymbol{R}_d$ is given by

$$\begin{aligned} \boldsymbol{R}_d &= \tilde{\boldsymbol{T}}_{N-1}\tilde{\boldsymbol{T}}_{N-2}\cdots\tilde{\boldsymbol{T}}_1\boldsymbol{R}_\kappa\tilde{\boldsymbol{T}}_1^H\tilde{\boldsymbol{T}}_2^H\cdots\tilde{\boldsymbol{T}}_{N-1}^H \\ &= \boldsymbol{L}_{N+1}\boldsymbol{R}_\kappa\boldsymbol{L}_{N+1}^H \end{aligned} \tag{62}$$

where $\boldsymbol{R}_\kappa$ is defined in (12). Note that the solution in (61) and (62) can be achieved by (59) using the method described in Appendix A or by a slight modification of the Householder tridiagonalization technique [21]–[24]. That is to say, that the decomposition presented here as a component of this new multistage Wiener filter represents a generalization of the unitary Householder tridiagonalization method for arbitrary nonsingular transformations $\boldsymbol{T}_i$ and $\tilde{\boldsymbol{T}}_i$. An example of the tridiagonal decomposition of the matrix $\boldsymbol{R}_\kappa$ is provided by the following progression for $N = 4$ as shown in (63) at the bottom of this page.

Next, define the 2×2 matrices $\boldsymbol{G}_j$ by

$$\boldsymbol{G}_j = \begin{bmatrix} 1 & -w_j^* \\ 0 & 1 \end{bmatrix} \tag{64}$$

for $j = 1, 2, \cdots, N$. Let the $(N+1)\times(N+1)$ matrices $\tilde{\boldsymbol{G}}_j$, for $j = 1, 2, \cdots, N$, be defined as follows:

$$\tilde{\boldsymbol{G}}_j = \begin{bmatrix} \boldsymbol{I}_{j-1} & & \boldsymbol{0} \\ & \boldsymbol{G}_j & \\ \boldsymbol{0} & & \boldsymbol{I}_{N-j} \end{bmatrix}. \tag{65}$$

Then the error-synthesis filterbank can be expressed in the form

$$\boldsymbol{\varepsilon}(k) = \tilde{\boldsymbol{G}}_1\tilde{\boldsymbol{G}}_2\cdots\tilde{\boldsymbol{G}}_N\boldsymbol{d}(k) = \boldsymbol{U}_{N+1}\boldsymbol{d}(k). \tag{66}$$

The error-synthesis filterbank is now seen to provide planar rotations, similar to the Jacobi or Givens transforms [5], [21], [25]. Note that the matrices $\boldsymbol{T}_j$ which form the multistage decomposition may be chosen to be unitary or nonunitary as a design choice and that the matrices $\boldsymbol{G}_j$ which form the error-synthesis filterbank are unimodular.

Therefore, for $N = 4$, the pyramid decomposition and nested scalar Wiener filters are represented in the form depicted in Fig. 8. Here the multistage Wiener filter demonstrates a relationship to planar reflections and rotations. Note that the coarsest estimate of the diagonal covariance matrix $\boldsymbol{R}_\varepsilon$ is given by

$$\hat{\boldsymbol{R}}_\varepsilon = (\tilde{\boldsymbol{G}}_1\tilde{\boldsymbol{G}}_2\boldsymbol{I}\tilde{\boldsymbol{T}}_1)\boldsymbol{R}_\kappa(\tilde{\boldsymbol{G}}_1\tilde{\boldsymbol{G}}_2\boldsymbol{I}\tilde{\boldsymbol{T}}_1)^H \tag{67}$$

where the first row of $\tilde{\boldsymbol{G}}_1\tilde{\boldsymbol{G}}_2\boldsymbol{I}\tilde{\boldsymbol{T}}_1$ is the coarsest estimate of the Wiener filter and finer grain estimates of both $\boldsymbol{R}_\varepsilon$ and the Wiener filter are obtained as additional stages are added.

D. Entropy and Mutual Information

The expressions for entropy and mutual information are now examined in terms of the multistage representation of the Wiener filter. First note from (1), (50), and (57), that matrix $\boldsymbol{R}_{x_0}$ is similar to the matrix $\boldsymbol{R}_{\tilde{\varepsilon}}$. The multiplicative property of determinants, which states that the determinant of a product

$$\begin{aligned} \begin{bmatrix} \sigma_{d_0}^2 & \boldsymbol{r}_{x_0d_0}^H \\ \boldsymbol{r}_{x_0d_0} & \boldsymbol{R}_{x_0} \end{bmatrix} &\overset{\tilde{\boldsymbol{T}}_1}{\Longrightarrow} \begin{bmatrix} \sigma_{d_0}^2 & \delta_1^* & 0 & 0 & 0 \\ \delta_1 & \sigma_{d_1}^2 & r_{x_1d_1}^*(1) & r_{x_1d_1}^*(2) & r_{x_1d_1}^*(3) \\ 0 & r_{x_1d_1}(1) & R_{x_1}(1,1) & R_{x_1}^*(2,1) & R_{x_1}^*(3,1) \\ 0 & r_{x_1d_1}(2) & R_{x_1}(2,1) & R_{x_1}(2,2) & R_{x_1}^*(3,2) \\ 0 & r_{x_1d_1}(3) & R_{x_1}(3,1) & R_{x_1}(3,2) & R_{x_1}(3,3) \end{bmatrix} \\ &\overset{\tilde{\boldsymbol{T}}_2}{\Longrightarrow} \begin{bmatrix} \sigma_{d_0}^2 & \delta_1^* & 0 & 0 & 0 \\ \delta_1 & \sigma_{d_1}^2 & \delta_2^* & 0 & 0 \\ 0 & \delta_2 & \sigma_{d_2}^2 & r_{x_2d_2}^*(1) & r_{x_2d_2}^*(2) \\ 0 & 0 & r_{x_2d_2}(1) & R_{x_2}(1,1) & R_{x_2}^*(2,1) \\ 0 & 0 & r_{x_2d_2}(2) & R_{x_2}(2,1) & R_{x_2}(2,2) \end{bmatrix} \\ &\overset{\tilde{\boldsymbol{T}}_3}{\Longrightarrow} \begin{bmatrix} \sigma_{d_0}^2 & \delta_1^* & 0 & 0 & 0 \\ \delta_1 & \sigma_{d_1}^2 & \delta_2^* & 0 & 0 \\ 0 & \delta_2 & \sigma_{d_2}^2 & \delta_3^* & 0 \\ 0 & 0 & \delta_3 & \sigma_{d_3}^2 & \delta_4^* \\ 0 & 0 & 0 & \delta_4 & \sigma_{d_4}^2 \end{bmatrix}. \end{aligned} \tag{63}$$

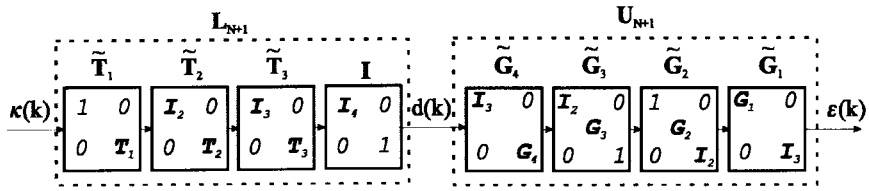

Fig. 8. The plane rotation interpretation of the Wiener filter for $N = 4$.

of two matrices is equal to the product of their determinants, implies

$$\|\boldsymbol{R}_{x_0}\| = \|\boldsymbol{R}_{\tilde{\varepsilon}}\| = \prod_{i=1}^{N} \xi_i. \tag{68}$$

Similarly, from (12), (58), and (62), it is found that

$$\|\boldsymbol{R}_{\kappa}\| = \|\boldsymbol{R}_{\tilde{\varepsilon}}\| = \prod_{i=0}^{N} \xi_i. \tag{69}$$

The entropy of the desired process $H(d_0)$ is unchanged from the expression in (9). The entropy of the observed process, $H(\boldsymbol{x}_0)$ in (10), is now written in the form

$$\begin{aligned} H(\boldsymbol{x}_0) &= \frac{1}{2}\log[(2\pi e)^N \|\boldsymbol{R}_{x_0}\|] \\ &= \frac{1}{2}\log\left[(2\pi e)^N \prod_{i=1}^{N} \xi_i\right] \\ &= \frac{1}{2}\log[(2\pi e)^N] + \frac{1}{2}\sum_{i=1}^{N}\log(\xi_i). \end{aligned} \tag{70}$$

Similarly, the joint entropy, $H(d_0, \boldsymbol{x}_0)$ in (13) becomes

$$\begin{aligned} H(d_0, \boldsymbol{x}_0) &= \frac{1}{2}\log[(2\pi e)^{N+1} \|\boldsymbol{R}_{\kappa}\|] \\ &= \frac{1}{2}\log\left[(2\pi e)(2\pi e)^N \prod_{i=0}^{N} \xi_i\right] \\ &= \frac{1}{2}\log(2\pi e) + \frac{1}{2}\log[(2\pi e)^N] + \frac{1}{2}\log(\xi_0) \\ &\quad + \frac{1}{2}\sum_{i=1}^{N}\log(\xi_i). \end{aligned} \tag{71}$$

Thus the conditional entropy, $H(d_0|\boldsymbol{x}_0)$ in (14), is expressed by

$$\begin{aligned} H(d_0|\boldsymbol{x}_0) &= H(d_0, \boldsymbol{x}_0) - H(\boldsymbol{x}_0) \\ &= \frac{1}{2}\log(2\pi e) + \frac{1}{2}\log(\xi_0). \end{aligned} \tag{72}$$

Finally, the mutual information, $I(d_0, \boldsymbol{x}_0)$ in (15), for a joint Gaussian process is computed to be

$$\begin{aligned} I(d_0, \boldsymbol{x}_0) &= H(d_0) - H(d_0|\boldsymbol{x}_0) \\ &= \frac{1}{2}\log(2\pi e \sigma_{d_0}^2) - \frac{1}{2}\log(2\pi e) - \frac{1}{2}\log(\xi_0) \\ &= \frac{1}{2}\log(\sigma_{d_0}^2) - \frac{1}{2}\log(\xi_0). \end{aligned} \tag{73}$$

From the last expression in (73) the intuitive result, that minimizing the mean-square error and maximizing the mutual information are equivalent, can now be established. Expressing the MMSE in terms of the canonical correlation in (8) results finally in

$$\begin{aligned} I(d_0, \boldsymbol{x}_0) &= \frac{1}{2}\log(\sigma_{d_0}^2) - \frac{1}{2}\log[\sigma_{d_0}^2(1 - k_{x_0 d_0}^2)] \\ &= -\frac{1}{2}\log(1 - k_{x_0 d_0}^2) \end{aligned} \tag{74}$$

from which it is also evident for joint Gaussian processes that maximizing the canonical correlation is equivalent to a maximization of the mutual information.

IV. Reduced-Rank Wiener Filtering

Rank reduction is concerned with finding a low-rank subspace, say of rank $M < N$, of the space spanned by the columns of the rank N observed-data covariance matrix. The rank M covariance matrix whose columns span this subspace has the same dimension as the original covariance matrix. The N-dimensional Wiener filter which is calculated using the rank M covariance matrix estimate is termed a reduced-rank Wiener filter.

An equivalent method of implementing a reduced-rank Wiener filter is obtained by prefiltering the observed data with an $N \times M$ matrix whose columns are the basis vectors of the reduced-rank subspace. The observed data is then compressed to an M-dimensional vector, the corresponding covariance matrix is $M \times M$, and both the transformed cross-correlation vector and the Wiener filter are of dimension M. This method is also termed reduced-rank or reduced-dimension Wiener filtering.

Another rank-reduction technique utilizes rank-shaping and shrinkage methods. The data-adaptive shrinkage method presented in [26] results in a nonlinear filter that uses mode-dependent, nonlinear companders to estimate something akin to the Wiener filter gain. This technique, which is not discussed further in this paper, represents an extension of a linear estimator.

A. Previous Approaches to Rank Reduction

The first approaches to the rank-reduction problem were motivated by the array processing application and were somewhat *ad hoc* [27], [28]. More statistical approaches to this problem were introduced next which were based on the principal-components analysis of the covariance matrix developed originally by Hotelling and Eckart [29], [30].

A new method to achieve a reduction in the number of degrees of freedom, or filter order, in a Wiener filter is introduced in [31] and [32]. This method utilizes a measure, termed the cross-spectral metric, to determine the smallest

number of degrees of freedom to linearly estimate with little loss a scalar random process from a set of N correlated complex random processes. To accomplish this a basis set is selected for the space that is spanned by the columns of the observed-data covariance matrix. What is called the cross-spectral information measure, is used then to rank the relative contributions of these basis vectors in order to minimize the mean-square error. The dimension of the data vector, and therefore the associated Wiener filter, can be reduced optimally to a dimension $M < N$ by discarding the projection of the data upon those basis vectors whose contributions are the smallest. This cross-spectral metric is recognized in [31] to be a vector which has components that are the weighted squared magnitudes of the direction cosines between each basis vector and the cross-correlation vector for the two aforementioned correlated random processes.

The principal-components technique achieves rank reduction to the effective rank of the signal subspace, but is limited in its ability to further reduce the data dimension. On the other hand, the cross-spectral metric provides a capability for reducing the data rank below this dimension without adversely affecting the mean-square error performance. This property represents a form of *data compression*. When the eigenvectors of the observed-data covariance matrix are utilized as a basis, the counter-intuitive result that the cross-spectral metric forms an upper bound on the performance obtained by the principal-components method of rank reduction is well established, e.g., see [31], [32], and Appendix B. Also it is proved that the $\binom{N}{M}$ eigenvectors selected, determined by the ranking provided by the cross-spectral metric, yield the lowest MMSE out of all other possible combinations of M eigenvectors.

It is now demonstrated that the cross-spectral metric is the optimal eigen-based subspace selection rule for Gaussian signals in an information-theoretic sense. The $N \times N$ covariance matrix of the data observed by the Wiener filter is expressed by

$$\boldsymbol{R}_{x_0} = \boldsymbol{E}\boldsymbol{\Lambda}\boldsymbol{E}^H \tag{75}$$

where $\boldsymbol{E}$ is the $N \times N$ matrix of eigenvectors $\{\boldsymbol{E}_i\}_{i=1}^N$ and $\boldsymbol{\Lambda}$ is the diagonal matrix of corresponding eigenvalues $\{\lambda_i\}_{i=1}^N$. The expression for the full-rank MMSE in (7) may be expressed in terms of the eigendecomposition in (75) as follows:

$$\xi_0 = \sigma_{d_0}^2 - \boldsymbol{r}_{x_0 d_0}^H \boldsymbol{E}\boldsymbol{\Lambda}^{-1}\boldsymbol{E}^H \boldsymbol{r}_{x_0 d_0} = \sigma_{d_0}^2 - \sum_{i=1}^{N} \frac{|\boldsymbol{E}_i^H \boldsymbol{r}_{x_0 d_0}|^2}{\lambda_i}. \tag{76}$$

The cross-spectral metric chooses the M eigenvectors which correspond with the largest M values of the term $1/\lambda_i |\boldsymbol{E}_i^H \boldsymbol{r}_{x_0 d_0}|^2$, from which it is evident that the cross-spectral metric is the optimal selection of $\binom{N}{M}$ eigenvectors in an MMSE sense for reduced-rank Wiener filtering and directly minimizes the mean-square error as a function of the rank M, for $1 \le M \le N$. The expression for the canonical correlation in (8) similarly reduces to

$$k_{x_0 d_0}^2 = \frac{1}{\sigma_{d_0}^2} \sum_{i=1}^{N} \frac{|\boldsymbol{E}_i^H \boldsymbol{r}_{x_0 d_0}|^2}{\lambda_i} \tag{77}$$

and it is also seen that the cross-spectral metric maximizes the canonical correlation as a function of the rank M. From the expression for mutual information in (74) it is also concluded that the cross-spectral metric maximizes mutual information as a function of rank. These properties show that the cross-spectral metric provides the capability for optimum compression, defined to be a reduction in the rank of the covariance matrix below the dimension of the signal subspace eigenstructure without incurring a large penalty in MMSE performance. The reduced-rank Wiener filter using the cross-spectral metric is therefore more robust in the sense that an error in underestimating the dimension of the underlying signal eigenstructure does not cause as great of a loss in performance as that which is experienced by the principal-components method. Note that the above claims of optimality are all relative to the eigenvector basis.

B. Rank Reduction Using the Multistage Representation of the Wiener Filter

The development and analysis of a reduced-rank or reduced-dimension Wiener filter which is based on the new multistage Wiener filter utilizes the operator interpretation of this decomposition. The $N \times N$ matrix $\boldsymbol{L}_N$ in (48) provides the complete decomposition of the observed-data vector. A reasonable and straightforward method of rank reduction is accomplished by discarding the last $N - M$ stages of this decomposition. The rank-reducing transformation is then given by the $M \times N$ matrix $\boldsymbol{L}_M$, defined as follows:

$$\boldsymbol{L}_M = \begin{bmatrix} \boldsymbol{h}_1^H \\ \boldsymbol{h}_2^H \boldsymbol{B}_1 \\ \vdots \\ \boldsymbol{h}_M^H \prod_{i=M-1}^{1} \boldsymbol{B}_i \end{bmatrix}. \tag{78}$$

The resulting M-dimensional output vector is processed by a nested chain of M scalar Wiener filters, in a manner which is a direct truncation of the filter structure shown in Fig. 5. Thus rank reduction is accomplished by simply stopping the multistage decomposition at the Mth stage. Note that no eigendecomposition is involved in this process and that a different basis is naturally generated by the multistage decomposition and the nested scalar Wiener filters.

This tridiagonalization of the covariance matrix of a stationary process tends to group most of the significant energy (both autocorrelation terms $\sigma_{d_i}^2$ and cross-correlation terms δ_i) in the first few tridiagonal terms. That is to say, the tridiagonal covariance matrix can be divided into 2×2 blocks along the main diagonal, where the upper-left and lower-right elements of one block are also members of the previous and following blocks that compose the diagonal, respectively. The multistage decomposition then strongly weights the most highly correlated terms and places them in the first few blocks; that is, the information which is most useful for estimating the signal $d_0(k)$ from $\boldsymbol{x}_0(k)$.

A final justification for the reduced-rank multistage Wiener filter is that the covariance matrix of the process $\boldsymbol{x}_i$ tends to

become white as the filter structure increases in stages. This is particularly true if the data is overmodeled, as is often the case in many applications such as sensor array processing. At the point where this covariance matrix is approximately white, the optimal Wiener filter is obtained by truncation of the remaining terms due to the fact that $\boldsymbol{h}_{i+1}$ goes to zero. Thus the optimal solution is obtained as the data becomes white and the decomposition itself whitens the data via successive stages.

It is now emphasized that this new method of reduced-rank Wiener filtering has the capability to outperform the eigendecomposition-based methods. This is due to the fact that a natural basis is generated by the multistage decomposition. Assume for the moment that the Wiener filter is collinear with the cross-correlation vector $\boldsymbol{r}_{x_0 d_0}$. Then the rank-one Wiener filter, which results from the first stage of the decomposition, maximizes the mutual information, and the remaining portion of the tree-structured analysis filterbank contributes nothing to the minimization of the mean-square error. The smallest subspace of the space spanned by the eigenvectors of the covariance matrix $\boldsymbol{L}_N \boldsymbol{R}_{x_0} \boldsymbol{L}_N^H$ which contains this same vector may be composed of many eigenvectors. Thus to obtain the optimal solution under these conditions, any eigendecomposition-based method of rank reduction would require a larger rank than the multistage method. Note that this does not imply that the reduced-rank multistage Wiener filter is optimal; it is, however, a very-low-complexity realization which demonstrates excellent performance as shown next.

C. Example

A computer analysis is now made to demonstrate the performance of the above new reduced-rank Wiener filter and to compare its performance with the principal-components and cross-spectral techniques. The well-known problem of interference mitigation using a narrowband and minimum-variance distortionless response (MVDR) antenna array is considered for this example. The MVDR constrained optimization problem is to minimize the output power of the array subject to a constraint that any signal coming from a particular and predefined direction is passed without distortion. This choice demonstrates the applicability of the new Wiener filter structure to the constrained optimization problem. The generalized sidelobe canceler (GSC) form of the constrained processor, defined, for example in [16]–[19] and Appendix B, is now utilized.

The GSC MVDR array processor has the partitioned structure shown in Fig. 9 and derived in Appendix B. There are $N+1$ sensor elements in the array, and the $(N+1)$-dimensional data vector which serves as the antenna input to the processor is denoted by $\boldsymbol{x}(k)$. The MVDR linear quadratic constraint is implemented by the $(N+1)$-dimensional steering vector $\boldsymbol{s}$ in the upper branch of the GSC. The $N \times (N+1)$ matrix $\boldsymbol{B}$ in the lower branch is a blocking matrix for those signals which impinge the array from the direction specified by $\boldsymbol{s}$. The output of the upper branch provides for the scalar desired signal $d_0(k)$ of the Wiener filter, and the output of the lower branch similarly provides the $N \times 1$ observed-data vector $\boldsymbol{x}_0(k)$. As noted in Section III, the constrained Wiener filter is

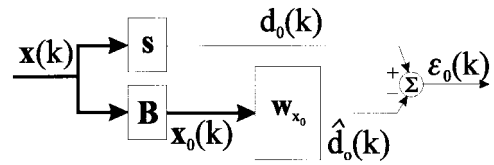

Fig. 9. The GSC MVDR constrained Wiener filter.

TABLE II
SIGNAL GEOMETRY

SIGNAL	DOA	SNR
desired	0°	0 dB
jammer 1	−69°	20 dB
jammer 2	−30°	22 dB
jammer 3	−17°	17 dB
jammer 4	14°	20 dB
jammer 5	34°	19 dB

reduced to an unconstrained Wiener filter by the matrix filters $\boldsymbol{s}$ and $\boldsymbol{B}$. The structure which follows these matrix filters in Fig. 9 is identical in form to the original Wiener filter depicted in Fig. 1. Thus the methods detailed in Sections III and IV can be used directly to implement the equivalent multistage Wiener filter and the reduced-rank Wiener filters. Note, in particular, that the very efficient blocking matrix developed in Appendix A can be employed both to implement the matrix $\boldsymbol{B}$ in Fig. 9 and the matrices $\boldsymbol{B}_i$ at each stage of the multistage Wiener filter.

For the purpose of this analysis, the sensor array is assumed to be linear, equispaced at a half-wavelength, and composed of $N+1=16$ sensor elements. The dimension of the Wiener filter is $N=15$. There are six signals which impinge the array, representing five undesired sources of interference or jammers and one desired signal, as detailed in Table II. All signals are modeled as narrowband, white, complex Gaussian random processes. The theoretical covariance matrix is computed and the optimal Wiener solution for the scenario, described in Table II, is depicted in Fig. 10 by the array power gain as a function of angle of arrival. All jammers are attenuated, and the desired signal receives a unity gain.

Next the MMSE is calculated as a function of rank for the multistage Wiener filter, the cross-spectral metric Wiener filter and the principal-components Wiener filter using the theoretical covariance matrix. As shown in Fig. 11 the principal-components Wiener filter obtains the optimal solution only after the number of significant eigenvectors, retained in the covariance estimate, equals the number of jammers. For this example the number of jammers is five, which corresponds to the effective rank of the signal subspace for that covariance matrix. The principal-components method degrades in performance rapidly if fewer than five eigenvectors are retained. The cross-spectral metric Wiener filter also obtains the same optimal solution with five weights. However, its MMSE performance is much better as a function of rank, achieving values within a fraction of a decibel of the optimum with just three or four weights. The new multistage Wiener filter outperforms both the cross-spectral and the principal-components Wiener filters. This new reduced-rank filtering

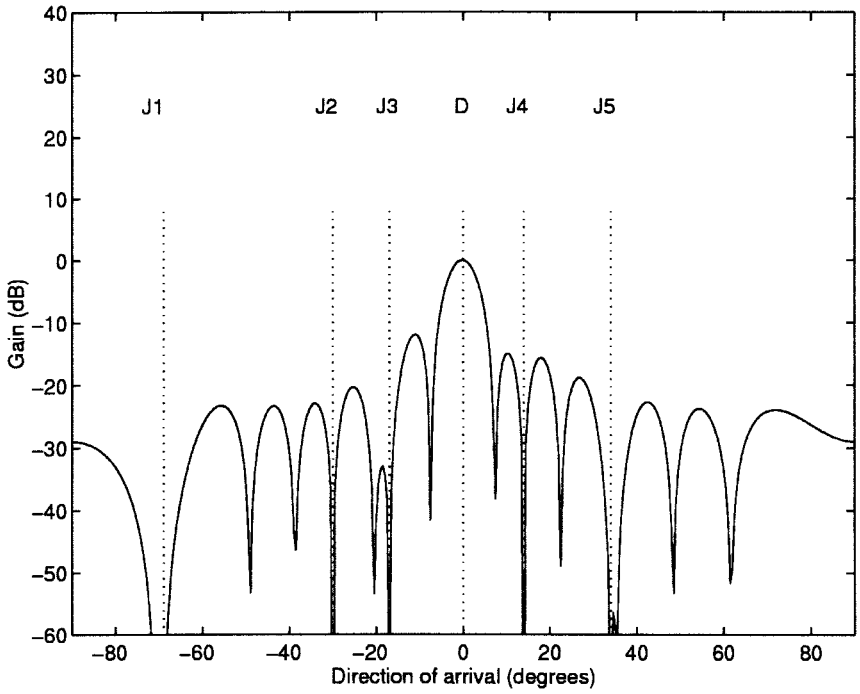

Fig. 10. The gain pattern of the optimal 16-sensor GSC.

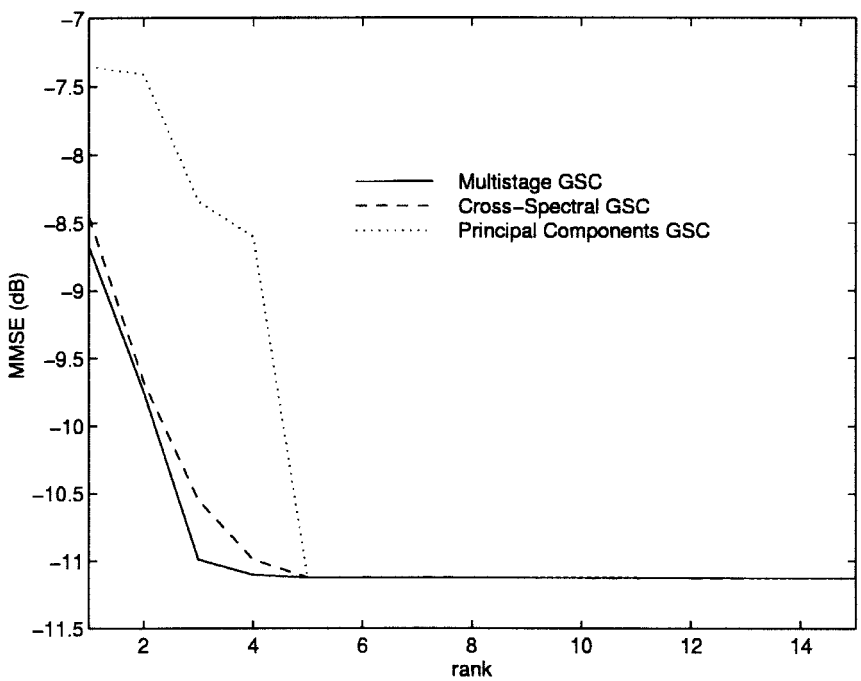

Fig. 11. The MMSE as a function of rank for the three reduced-rank GSC sensor arrays.

structure obtains the optimal solution with four weights and performs better in terms of the MMSE performance for every rank. The computational complexity of the Wiener filter is also greatly reduced for the multistage Wiener filter using the method found in Appendix A.

The mutual information is calculated for these three processors as a function of rank. The mutual information for the reduced-rank processors is given by

$$\hat{I}(d_0, \boldsymbol{x}_0) = \frac{1}{2}\log(\sigma_{d_0}^2) - \frac{1}{2}\log(\hat{\xi}_0) \tag{79}$$

where $\hat{\xi}_0 \geq \xi_0$ is the MMSE of the reduced-rank processor and may take a different value for each rank-reduction technique. The results of this comparison are presented in Fig. 12. The curves for the mutual information depict the same comparative results as noted above for the MMSE. Observe that the mutual information curves are scaled and inverted replicas of the MMSE curves, which is what might be expected since, as established in Section III-D for the assumed Gaussian processes, the maximization of the mutual information and the minimization of the mean-square error are equivalent.

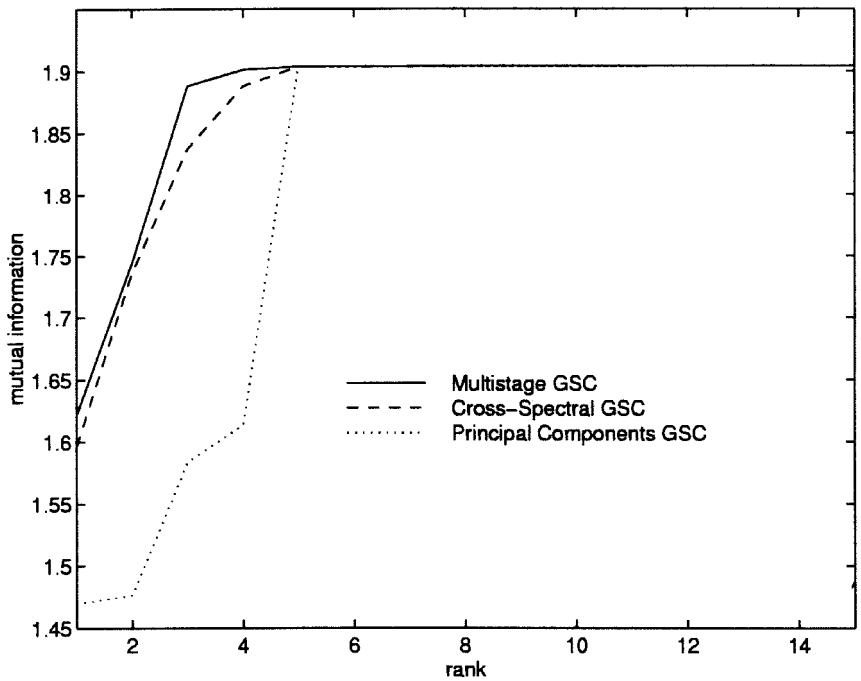

Fig. 12. The mutual information as a function of rank for the three reduced-rank GSC sensor arrays.

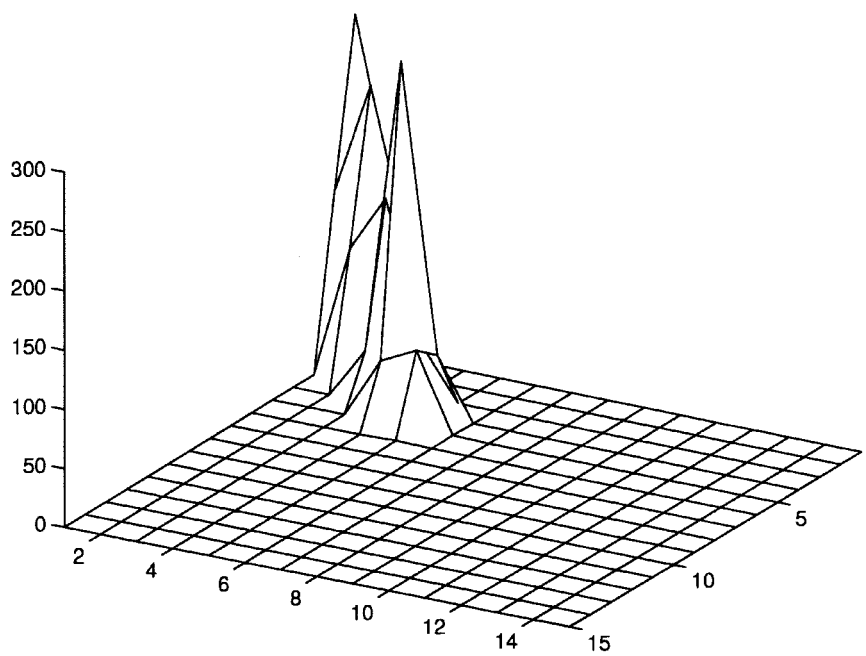

Fig. 13. The magnitude of the elements of the tridiagonal covariance matrix.

Finally, the magnitude of the terms in the theoretical covariance matrix, after being decomposed to a tridiagonal matrix by the multistage decomposition, is depicted in Fig. 13 with a Matlab mesh plot. Here it seen that the number of significant terms in the $N \times N$ covariance matrix $\boldsymbol{R}_{\tilde{d}}$ decay quickly and are not discernible after the sixth row and column entry. The fact that the equivalent cross-spectral energy in the basis generated by the multistage decomposition is compactly represented by this new structure is reflected by the fact that the new reduced-rank processor outperforms the cross-spectral metric using an eigenvector basis.

V. Conclusions

A new Wiener filter structure is derived which uses a multistage decomposition. The properties of this new version of a Wiener filter are analyzed from several points of view. In particular, it is demonstrated that the unconstrained Wiener filter can always be constructed via a chain of constrained Wiener filters. The relationship of the Wiener filter with the properties of entropy, mutual information, and canonical correlation are established. Also previous reduced-rank Wiener filter algorithms are analyzed in terms of these properties. A new

method for performing rank reduction based on the multistage decomposition is derived. This new reduced-rank Wiener filter is shown by example to be capable of outperforming the eigendecomposition-based methods both in terms of MMSE performance and computational complexity.

Appendix A
A Low Complexity Blocking Matrix

By weakening the unitary requirement for $\boldsymbol{T}_i$, an implementation of the blocking matrix $\boldsymbol{B}_i$ is now found which demonstrates a low computational complexity. The MMSE is preserved when the nonunitary matrix $\boldsymbol{T}_i$ is any arbitrary, nonsingular $N \times N$ matrix [19], [33]. To establish this fact, recall that the MMSE for the filter in Fig. 1 is given by (7)

$$\xi_0 = \sigma_{d_0}^2 - \boldsymbol{r}_{x_0 d_0}^H \boldsymbol{R}_{x_0}^{-1} \boldsymbol{r}_{x_0 d_0}. \tag{A.1}$$

If the matrix $\boldsymbol{T}_1$ in Fig. 2 is an arbitrary, nonsingular matrix, then the covariance matrix $\boldsymbol{R}_{z_1}$ is given by

$$\boldsymbol{R}_{z_1} = \boldsymbol{T}_1 \boldsymbol{R}_{x_0} \boldsymbol{T}_1^H, \tag{A.2}$$

and its inverse is given by

$$\boldsymbol{R}_{z_1}^{-1} = \left(\boldsymbol{T}_1^H\right)^{-1} \boldsymbol{R}_{x_0}^{-1} \boldsymbol{T}_1^{-1}. \tag{A.3}$$

Similarly, the cross-correlation vector $\boldsymbol{r}_{z_1 d_0}$ is given by

$$\boldsymbol{r}_{z_1 d_0} = \boldsymbol{T}_1 \boldsymbol{r}_{x_0 d_0}. \tag{A.4}$$

Thus the MMSE is computed to be

$$\begin{aligned}\xi_0 &= \sigma_{d_0}^2 - \boldsymbol{r}_{z_1 d_0}^H \boldsymbol{R}_{z_1}^{-1} \boldsymbol{r}_{z_1 d_0} \\ &= \sigma_{d_0}^2 - \boldsymbol{r}_{x_0 d_0}^H \boldsymbol{T}_1^H ((\boldsymbol{T}_1^H)^{-1} \boldsymbol{R}_{x_0}^{-1} \boldsymbol{T}_1^{-1}) \boldsymbol{T}_1 \boldsymbol{r}_{x_0 d_0} \\ &= \sigma_{d_0}^2 - \boldsymbol{r}_{x_0 d_0}^H \boldsymbol{R}_{x_0}^{-1} \boldsymbol{r}_{x_0 d_0}\end{aligned} \tag{A.5}$$

and it is evident that the MMSE is conserved by any invertible transformation of the observed data.

The special case of interest occurs when the nonsingular matrix $\boldsymbol{T}_1$ is formed so that its first row is given by $\boldsymbol{h}_1^H$ and the remaining $N-1$ rows are linearly independent, composing the matrix $\boldsymbol{B}_1$. The matrix $\boldsymbol{B}_1$ must satisfy $\boldsymbol{B}_1 \boldsymbol{h}_1 = \boldsymbol{0}$. Then this operator $\boldsymbol{T}_1$ is of the same form as detailed in (16), the covariance matrix $\boldsymbol{R}_{z_1}$ and its inverse retain the structure described in (20) and (21), respectively, and the cross-correlation vector $\boldsymbol{r}_{z_1 d_0}$ is again in the form shown in (22). Finally, the structure depicted in Fig. 3 is again obtained, and it yields the identical MMSE as shown above.

The partitioning of the operator $\boldsymbol{T}_i$, given in (16) for $i = 1$, is shown in Fig. 14. It is now desired to find an $(N-i+1) \times (N-i+1)$ diagonal matrix preprocessor $\boldsymbol{H}_i$ at each stage i, with the properties that the $(N-i+1) \times 1$ vector $\boldsymbol{h}_i$ is replaced by a vector $\boldsymbol{a}_i$ (to be defined) and that the $(N-i) \times (N-i+1)$ matrix $\boldsymbol{B}_i$ is replaced by the following multiplier-free difference matrix $\boldsymbol{D}_i$:

$$\boldsymbol{D}_i = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & & & \ddots & & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{bmatrix} \tag{A.6}$$

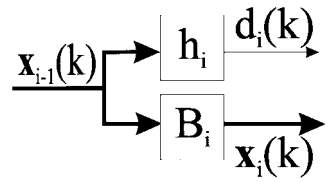

Fig. 14. The matrix filter decomposition of the observed data.

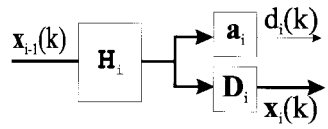

Fig. 15. An equivalent decomposition of the observed data.

as shown in Fig. 15. In order for the operation of the structures in Figs. 14 and 15 to be equivalent in terms of the MMSE, it is clear that the following three conditions must hold:

1) $\boldsymbol{h}_i = \boldsymbol{H}_i^H \boldsymbol{a}_i$.
2) $\boldsymbol{B}_i = \boldsymbol{D}_i \boldsymbol{H}_i$.
3) $\boldsymbol{B}_i \boldsymbol{h}_i = \boldsymbol{D}_i \boldsymbol{H}_i \boldsymbol{H}_i^H \boldsymbol{a}_i = \boldsymbol{0}$.

Let the complex cross-correlation vector $\boldsymbol{h}_i$ be composed of elements θ_j, for $j = 1 \cdots N-i+1$, so that

$$\boldsymbol{h}_i = [\theta_1 \quad \theta_2 \quad \cdots \quad \theta_{N-i+1}]^T. \tag{A.7}$$

Then a solution, which satisfies Conditions 2) and 3), is the $(N-i+1) \times (N-i+1)$ diagonal matrix $\boldsymbol{H}_i$ given by

$$\boldsymbol{H}_i = \operatorname{diag}\left(\left[\frac{1}{\theta_1} \quad \frac{1}{\theta_2} \quad \cdots \quad \frac{1}{\theta_{N-i+1}}\right]\right). \tag{A.8}$$

With (A.8) and Condition 1), the vector $\boldsymbol{a}_i$ must have the form

$$\boldsymbol{a}_i = [|\theta_1|^2 \quad |\theta_2|^2 \quad \cdots \quad |\theta_{N-i+1}|^2]^T. \tag{A.9}$$

Finally, the form of the matrices $\boldsymbol{T}_i$ are now given by

$$\boldsymbol{T}_i = \begin{bmatrix} \boldsymbol{h}_i \\ \boldsymbol{D}_i \boldsymbol{H}_i \end{bmatrix}. \tag{A.10}$$

Therefore, it is observed that the new filtering operation in Fig. 15 may now be implemented with the filtering structure in Fig. 14 where a new signal-blocking matrix $\tilde{\boldsymbol{B}}_i$, defined by

$$\tilde{\boldsymbol{B}}_i = \begin{bmatrix} \frac{1}{\theta_1} & \frac{-1}{\theta_2} & 0 & \cdots & 0 & 0 & 0 \\ 0 & \frac{1}{\theta_2} & \frac{-1}{\theta_3} & \cdots & 0 & 0 & 0 \\ \vdots & & & \ddots & & & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\theta_{N-i-1}} & \frac{-1}{\theta_{N-i}} & 0 \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{\theta_{N-i}} & \frac{-1}{\theta_{N-i+1}} \end{bmatrix}, \tag{A.11}$$

is used to replace the matrix filter $\boldsymbol{B}_i$ in Fig. 14. Fig. 16 depicts this new structure, where the vector $\boldsymbol{h}_i$ is again given by (35) and the matrix $\tilde{\boldsymbol{B}}_i$ in (A.11) is found directly from $\tilde{\boldsymbol{B}}_i = \boldsymbol{D}_i \boldsymbol{H}_i$ using (A.6) and (A.8). All three previously stated conditions for equivalence are thus satisfied.

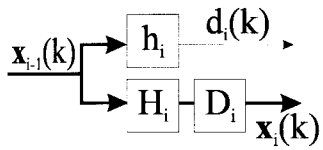

Fig. 16. Final equivalent representation of the decomposition.

APPENDIX B
THE CONSTRAINED WIENER FILTER

Many applications of the Wiener filter utilize a constrained optimization in situations where a desired signal may not be explicitly available. In this Appendix the linearly constrained, minimum-variance, distortionless response (MVDR) Wiener filter is considered to describe the problem of quadratic mininization with linear constraints and rank reduction for constrained Wiener filters.

The MVDR constrained optimization problem is expressed as follows:

$$\min_{\boldsymbol{w}} \boldsymbol{w}^H \boldsymbol{R} \boldsymbol{w} \qquad \text{subject to } \boldsymbol{s}^H \boldsymbol{w} = 1 \tag{B.1}$$

where for convenience in the derivation to follow $\boldsymbol{w}$ and $\boldsymbol{s}$ are defined to be $(N+1)$-dimensional vectors and $\boldsymbol{R}$ is the $(N+1)\times(N+1)$ dimensional covariance matrix associated with the $(N+1)$-dimensional observed-data vector $\boldsymbol{x}(k)$. Here the vector $\boldsymbol{s}$ usually represents a temporal, spatial, or space–time frequency response of interest which is constrained to be undisturbed by the minimization procedure. The solution to the constrained optimization in (B.1) may be found by the method of Lagrange multipliers and is given by

$$\boldsymbol{w} = \frac{\boldsymbol{R}^{-1}\boldsymbol{s}}{\boldsymbol{s}^H \boldsymbol{R}^{-1}\boldsymbol{s}}. \tag{B.2}$$

The resulting MMSE is provided by

$$\xi_0 = \boldsymbol{w}^H \boldsymbol{R} \boldsymbol{w} = \frac{1}{\boldsymbol{s}^H \boldsymbol{R}^{-1}\boldsymbol{s}}. \tag{B.3}$$

Reduced-rank Wiener filtering using the direct-form constrained Wiener filter in (B.2) is possible by performing an eigendecomposition of the matrix $\boldsymbol{R}$

$$\boldsymbol{R} = \boldsymbol{V}\boldsymbol{\Psi}\boldsymbol{V}^H \tag{B.4}$$

where the matrix $\boldsymbol{V}$ is composed of the $N+1$ eigenvectors $\boldsymbol{v}_i$ and the diagonal matrix $\boldsymbol{\Psi}$ is composed of the $N+1$ eigenvalues ψ_i. The principal-components method of rank reduction would retain only those $M+1$ eigenvectors $\boldsymbol{v}_i$ which correspond with the largest $M+1$ eigenvalues ψ_i, where $M < N$.

In a similar manner, the method of rank reduction which utilizes the cross-spectral metric may be implemented with the direct-form constrained Wiener filter. Note that minimizing the mean-square error is equivalent to maximizing the inverse of the mean-square error. In the case of the constrained Wiener filter it may be shown [17] that the inverse of the MMSE is proportional with the output signal-to-interference plus noise ratio (SINR). Hence minimizing the mean-square error as a function of rank is equivalent to maximizing the output SINR as a function of rank, or maximizing the expression

$$\xi_0^{-1} = \boldsymbol{s}^H \boldsymbol{R}^{-1} \boldsymbol{s} = \sum_{i=1}^{N+1} \frac{|\boldsymbol{v}_i^H \boldsymbol{s}|^2}{\psi_i} \tag{B.5}$$

as a function of rank i. Clearly, choosing the $M+1$ eigenvectors $\boldsymbol{v}_i$ corresponding with the largest $M+1$ values of the direct-form cross-spectral metric

$$\frac{|\boldsymbol{v}_i^H \boldsymbol{s}|^2}{\psi_i} \tag{B.6}$$

both maximizes the output SINR and minimizes the mean-square error as a function of rank.

The form of the constrained Wiener filter in (B.2) may be compared with the unconstrained Wiener filter in (5). The cross-correlation vector in (5) is replaced by the term $\beta\boldsymbol{s}$, where the scalar $\beta = (\boldsymbol{s}^H \boldsymbol{R}^{-1}\boldsymbol{s})^{-1}$. The multistage Wiener filter could be derived directly using this relationship, but an interpretation of the constrained Wiener filter known as the GSC [16], [18], [19] provides additional insight.

To derive the GSC form of the constrained Wiener filter it is convenient to redefine the constraint vector $\boldsymbol{s}$ to be

$$\boldsymbol{s} = \frac{\boldsymbol{s}}{\sqrt{\boldsymbol{s}^H \boldsymbol{s}}} \tag{B.7}$$

and define the $(N+1)\times(N+1)$ nonsingular matrix $\boldsymbol{T}$ as follows:

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{s}^H \\ \boldsymbol{B} \end{bmatrix} \tag{B.8}$$

where $\boldsymbol{B}$ is a full row-rank $N\times(N+1)$ matrix with nullspace $\boldsymbol{s}$. The optimization in (B.1) may then be expressed by

$$\min_{\boldsymbol{w}} \boldsymbol{w}^H \boldsymbol{T}^{-1}\boldsymbol{T}\boldsymbol{R}\boldsymbol{T}^H(\boldsymbol{T}^H)^{-1}\boldsymbol{w} \qquad \text{subject to } \boldsymbol{s}^H \boldsymbol{T}^H(\boldsymbol{T}^H)^{-1}\boldsymbol{w} = 1 \tag{B.9}$$

which has the equivalent solution for the $(N+1)$-dimensional Wiener filter in (B.2) expressed in the form

$$\boldsymbol{w} = \frac{\boldsymbol{T}^H(\boldsymbol{T}\boldsymbol{R}\boldsymbol{T}^H)^{-1}\boldsymbol{T}\boldsymbol{s}}{\boldsymbol{s}^H \boldsymbol{T}^H(\boldsymbol{T}\boldsymbol{R}\boldsymbol{T}^H)^{-1}\boldsymbol{T}\boldsymbol{s}}. \tag{B.10}$$

Next note that the product $\boldsymbol{T}\boldsymbol{s}$ results in the compressed $(N+1)$-dimensional vector $\boldsymbol{e}_1$, where $\boldsymbol{e}_1 = [1\ 0\ 0\ \cdots\ 0]^T$. The partitioned, transformed, covariance matrix $\boldsymbol{T}\boldsymbol{R}\boldsymbol{T}^H$ is computed to have the form

$$\boldsymbol{T}\boldsymbol{R}\boldsymbol{T}^H = \begin{bmatrix} \boldsymbol{s}^H \boldsymbol{R}\boldsymbol{s} & \boldsymbol{s}^H \boldsymbol{R}\boldsymbol{B}^H \\ \boldsymbol{B}\boldsymbol{R}\boldsymbol{s} & \boldsymbol{B}\boldsymbol{R}\boldsymbol{B}^H \end{bmatrix} = \begin{bmatrix} \sigma_{d_0}^2 & \boldsymbol{r}_{x_0 d_0}^H \\ \boldsymbol{r}_{x_0 d_0} & \boldsymbol{R}_{x_0} \end{bmatrix} \tag{B.11}$$

which is identical to $\boldsymbol{R}_\kappa$ in (12) with $\boldsymbol{r}_{x_0 d_0}$ being an $N\times 1$ vector and $\boldsymbol{R}_{x_0}$ being an $N\times N$ matrix. The inverse of the covariance matrix $\boldsymbol{T}\boldsymbol{R}\boldsymbol{T}^H$ is computed as follows:

$$(\boldsymbol{T}\boldsymbol{R}\boldsymbol{T}^H)^{-1} = \xi_0^{-1} \begin{bmatrix} 1 & -\boldsymbol{r}_{x_0 d_0}^H \boldsymbol{R}_{x_0}^{-1} \\ -\boldsymbol{R}_{x_0}^{-1}\boldsymbol{r}_{x_0 d_0} & \boldsymbol{R}_{x_0}^{-1}(\xi_0 \boldsymbol{I} + \boldsymbol{r}_{x_0 d_0}\boldsymbol{r}_{x_0 d_0}^H \boldsymbol{R}_{x_0}^{-1}) \end{bmatrix}. \tag{B.12}$$

Finally, using (B.11) and (B.12), the $(N+1)$-dimension constrained Wiener filter in (B.10) is expressed as

$$\boldsymbol{w} = \boldsymbol{T}^H \begin{bmatrix} 1 \\ -\boldsymbol{R}_{x_0}^{-1}\boldsymbol{r}_{x_0 d_0} \end{bmatrix} = \boldsymbol{s} - \boldsymbol{B}^H \boldsymbol{w}_{x_0} \tag{B.13}$$

where $\boldsymbol{w}_{x_0}$ is the N-dimensional unconstrained Wiener filter

$$\boldsymbol{w}_{x_0} = \boldsymbol{R}_{x_0}^{-1}\boldsymbol{r}_{x_0 d_0} \tag{B.14}$$

which yields the optimal estimate of $d_0(k) = \boldsymbol{s}^H \boldsymbol{x}(k)$ from the process $\boldsymbol{x}_0(k) = \boldsymbol{B}\boldsymbol{x}(k)$ as depicted in Fig. 9.

The GSC form of the constrained Wiener filter is now in the standard form of the classical Wiener filter, and the analysis of reduced-rank algorithms may be examined in the context presented in Section IV. The multistage Wiener filter implementation is identical from either the direct-form or the GSC-form derivation; that is, the matrix filters $\boldsymbol{s}$ and $\boldsymbol{B}$ form the first stage of the decomposition and the rest follows directly from the derivation in Section IV. However, the eigendecomposition-based approaches require more attention.

For the eigendecomposition-based techniques, the matrix $\boldsymbol{R}_{x_0}$ is expressed in the form

$$\boldsymbol{R}_{x_0} = \boldsymbol{E}\boldsymbol{\Lambda}\boldsymbol{E}^H \tag{B.15}$$

where $\boldsymbol{E}$ is the $N \times N$ matrix of eigenvectors $\boldsymbol{E}_i$ and $\boldsymbol{\Lambda}$ is the $N \times N$ diagonal matrix of eigenvalues λ_i associated with the matrix $\boldsymbol{R}_{x_0} = \boldsymbol{B}\boldsymbol{R}\boldsymbol{B}^H$. The principal-components method chooses the $M < N$ eigenvectors corresponding with the M largest eigenvalues of $\boldsymbol{R}_{x_0}$ to implement rank reduction. The GSC-form cross-spectral metric approach to reduced-rank Wiener filtering chooses those M eigenvectors which correspond with the largest M values of the cross-spectral metric

$$\frac{|\boldsymbol{E}_i^H \boldsymbol{r}_{x_0 d_0}|^2}{\lambda_i}. \tag{B.16}$$

Thus rank reduction for the GSC-form processor utilizes an $M \times N$ operator to compress the data vector $\boldsymbol{x}_0(k)$ and reduce the dimension of the Wiener filter $\boldsymbol{w}_{x_0}$ from N to M. Note that reducing the dimension of $\boldsymbol{w}_{x_0}$ in (B.14) from N to M is equivalent to reducing the dimension of $\boldsymbol{w}$ in (B.2) and (B.10) from $N+1$ to $M+1$.

The performance of the principal-components and cross-spectral metric Wiener filters for the direct form and the GSC form are not in general the same. The eigenstructure of the GSC-form processor is dependent on the choice of $\boldsymbol{T}$ and the spaces spanned by the two sets of eigenvectors, $\{\boldsymbol{v}_i\}$ and $\{\boldsymbol{E}_i\}$, are of different dimensions. The direct-form cross-spectral metric in (B.6) and the GSC-form cross-spectral metric in (B.16) both minimize the mean-square error and maximize the output SINR as a function of rank, however their performance is not the same as a function of rank. While this may seem counter-intuitive at first, the direct-form cross-spectral metric is the magnitude-squared value of a direction cosine between the eigenvectors of the unconstrained covariance matrix $\boldsymbol{R}$ and the constraint vector $\boldsymbol{s}$, which determines the largest projection of the useful estimation energy. The GSC-form cross-spectral metric is the magnitude-squared value of a direction cosine between the eigenvectors of the constrained covariance matrix $\boldsymbol{R}_{x_0}$ and the induced cross correlation with the constraint vector, $\boldsymbol{r}_{x_0 d_0}$, which also determines the largest projection of the useful estimation energy from a different perspective. Therefore, while the solutions are identical for the full-rank processors, the convergence as a function of rank for the direct-form and GSC-form reduced-rank processors are generally different. The cross-spectral metric is optimal in terms of MMSE or SINR as a function of rank with respect to the eigenvectors of the relevant covariance matrix for both the direct-form and the GSC-form structures.

ACKNOWLEDGMENT

The authors wish to thank Prof. Alfred Hero and the anonymous reviewers for their suggestions which improved the presentation of the material in this paper. In addition, the authors are grateful for insightful discussions with Dr. Allan Steinhardt and Dr. Dennis Healy at DARPA, Dr. John Tague at the Office of Naval Research, Dr. Peter Zulch at the USAF Rome Laboratory, Prof. Robert Gagliardi and Dr. Xiaoli Yu at the University of Southern California, and Dr. Jeff Holder at the Georgia Tech Research Institute.

REFERENCES

[1] L. E. Brennan, J. D. Mallett, and I. S. Reed, "Convergence rate in adaptive arrays," Tech. Rep. TSC-PD-A177-2, Technology Service Corp., July 1977.
[2] L. E. Brennan and I. S. Reed, "Convergence rate in adaptive arrays," Tech. Rep. TSC-PD-A177-4, Technology Service Corp., Jan. 1978.
[3] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays.* New York: Wiley, 1980.
[4] R. T. Compton, *Adaptive Antennas.* Englewood Cliffs, NJ: Prentice-Hall, 1988.
[5] S. Haykin, *Adaptive Filter Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1991.
[6] T. C. Liu and B. D. Van Veen, "A modular structure for implementation of linearly constrained minimum variance beamformers," *IEEE Trans. Signal Processing*, vol. 39, pp. 2343–2346, Oct. 1991.
[7] R. Sharma and B. D. Van Veen, "Large modular structures for adaptive beamformering and the Gram–Schmidt preprocessor," *IEEE Trans. Signal Processing*, vol. 42, pp. 448–451, Feb. 1994.
[8] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, pp. 321–377, 1936.
[9] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis.* New York: Wiley, 1958.
[10] R. J. Muirhead, *Aspects of Multivariate Statistical Theory.* New York: Wiley, 1982.
[11] L. L. Scharf and J. K. Thomas, "Wiener filters in canonical coordinates for transform coding, filtering, and quantizing," *IEEE Trans. Signal Processing*, vol. 46, pp. 647–654, Mar. 1998.
[12] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 350–355 and 623–656, Mar. 1948.
[13] R. M. Fano, *Transmission of Information: A Statistical Theory of Communication.* New York: Wiley, 1961.
[14] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.
[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.
[16] L. L. Scharf, *Statistical Signal Processing.* Reading, MA: Addison-Wesley, 1991.
[17] J. S. Goldstein and I. S. Reed, "Performance measures for optimal constrained beamformers," *IEEE Trans. Antennas Propagat.*, vol. 45, pp. 11–14, Jan. 1997.
[18] S. P. Applebaum and D. J. Chapman, "Adaptive arrays with main beam constraints," *IEEE Trans. Antennas Propagat.*, vol. AP-24, pp. 650–662, Sept. 1976.
[19] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propagat.*, vol. AP-30, pp. 27–34, Jan. 1982.

[20] J. S. Goldstein and I. S. Reed, "Theory of partially adaptive radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, pp. 1309–1325, Oct. 1997.
[21] G. H. Golub and C. F. Van Loan, *Matrix Computations.* Baltimore, MD: The Johns Hopkins Univ. Press, 1990.
[22] G. H. Golub, "Numerical methods for solving linear least squares problems," *Numer. Math.*, vol. 7, pp. 206–216, 1965.
[23] D. S. Watkins, *Fundamentals of Matrix Computations.* New York: Wiley, 1991.
[24] A. O. Steinhardt, "Householder transforms in signal processing," *IEEE ASSP Mag.*, vol. 5, pp. 4–12, 1988.
[25] W. Givens, "Computation of plane unitary rotations transforming a general matrix to triangular form," *J. Soc. Ind. Appl. Math.*, vol. 6, pp. 26–50, 1958.
[26] A. J. Thorpe and L. L. Scharf, "Data adaptive rank-shaping methods for solving least squares problems," *IEEE Trans. Signal Processing*, vol. 43, pp. 1591–1601, July 1995.
[27] D. J. Chapman, "Partial adaptivity for the large array," *IEEE Trans. Antennas Propagat.*, vol. AP-24, pp. 685–696, Sept. 1976.
[28] D. R. Morgan, "Partially adaptive array techniques," *IEEE Trans. Antennas Propagat.*, vol. AP-26, pp. 823–833, Nov. 1978.
[29] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, pp. 417–441 and 498–520, 1933.
[30] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrica*, vol. 1, pp. 211–218, 1936.
[31] J. S. Goldstein and I. S. Reed, "Reduced rank adaptive filtering," *IEEE Trans. Signal Processing*, vol. 45, pp. 492–496, Feb. 1997.
[32] ——, "Subspace selection for partially adaptive sensor array processing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, Apr. 1997.
[33] C. W. Jim, "A comparison of two LMS constrained optimal array structures," *Proc. IEEE*, vol. 65, pp. 1730–1731, Dec. 1977.