Abstract

This paper is motivated by sensing and wireless communication, where data compression or dimension reduction may be used to reduce the required communication bandwidth. High-dimensional measurements are converted into low-dimensional representations through linear compression. Our aim is to compress a noisy sensor measurement, allowing for the fact that the compressed measurement will then be transmitted over a noisy channel. We give the closed-form expression for the optimal compression matrix that minimizes the trace or determinant of the error covariance matrix. We show that the solutions share a common architecture consisting of a canonical coordinate transformation, scaling by coefficients which account for canonical correlations and channel noise variance, followed by a coordinate transformation into the sub-dominant invariant subspace of the channel noise. Furthermore, we explore the design problem with respect to more general criteria and provide a unified factorization for the corresponding optimal compression matrix. A necessary condition is obtained for the optimal compression.

Index Terms

compressive sensing, canonical coordinates, dimension reduction, optimal linear compression, signal-plus-noise model, reduced-rank filtering, precoding and equalizing.

I. INTRODUCTION

When communication efficiency or model reduction is desired, one can precompress observations to lower-dimensional measurements before transmitting them around a network or to a fusion center. Such data compression or dimension reduction reduces the communication burden, but increases mean-squared error and reduces information...
rate. In this paper, we are interested in designing the linear compression matrix that minimizes the mean-squared error or maximizes the information rate at the optimal compression ratio, under a power constraint.

![Diagram](attachment:linear_compression.png)

Fig. 1. Linear compression of a noisy measurement $x$ with channel noise $v$.

Let $\theta \in \mathbb{R}^p$ be a signal of interest and $x \in \mathbb{R}^n$ a noisy measurement of $\theta$. The diagram of Fig. 1 frames the problem of interest in this paper. The noisy measurement $x$ is to be compressed with the linear transformation $W \in \mathbb{R}^{m \times n} (m < n)$ and then transmitted through a noisy channel. The channel transforms the measurement by a channel matrix $D \in \mathbb{R}^{m \times m}$ and adds noise $v \in \mathbb{R}^m$ to produce a measurement $z \in \mathbb{R}^m$. In distributed sensor networks, each sensor makes a noisy measurement of a state. The sensor measurement is then to be compressed into a lower-dimensional measurement, for low bandwidth transmission over a noisy channel. The channel could be electromagnetic, acoustic, or magnetic (as in storage). The compression is to be designed so that subsequent estimation of the state from the noisy channel measurement is as accurate as it can be at the given power budget for the transmission. The important element of our model, and the model of Schizas, et al. [1], is that noise enters at the input and the output of the compressor, making the results applicable in radar, sonar, storage, and imaging.

Our goal is to design the compressor $W$ so that the noisy and compressed measurement $z$ may be processed for an estimator of the signal $\theta$ whose error covariance has minimum trace or minimum determinant. The minimum trace solution minimizes mean-squared error of the estimate and the minimum determinant solution minimizes volume of the error concentration ellipsoid. In the Gaussian case, it maximizes differential information rate. The mean-squared error result is due to Schizas, et al. [1], but the determinant is new.

We are not the first to consider this problem and its variants. In fact as we will show, this paper is an extension of the original work of Schizas, et al. [1], which in turn generalizes the work of [2]-[6] and [7]-[8]. The innovation of this paper is this. First we replace the mean-squared error criterion of [1] with the determinant criterion. The determinant of the error covariance matrix is proportional to the volume of the error covariance matrix, or equivalently, the volume of the concentration ellipse for the state estimator. For some applications, volume is more illuminating than the trace of the error covariance matrix, although admittedly neither completely characterizes error covariance. This statement requires no assumption of multivariate normality. But in the multivariate normal case, the determinant of the error covariance matrix is proportional to the infinitesimal rate at which the compressed measurement, transmitted through the noisy channel, brings information about the state. This seems to us to be an argument in favor of the determinant. This is not to discredit the trace, which is mean-squared error, but to credit determinant, which is volume or information rate, as an equally credible scalar measure of estimator performance. Moreover, when one or more elements of the state is relatively more important than others, a weighted trace would
be used as the minimizing principle. If only a few out of many state variables are to be estimated, with the others treated as nuisance parameters, then the weighted trace would be used. Our results for weighted trace minimization treat these cases. Second we show that our designs and those of Schizas, et al. [1] may be cast as scaled and rotated canonical coordinate designs. This finding is important, for it generalizes the theory of canonical coordinates to a much more general class of problems than the class for which they were originally designed [13] and subsequently applied [3]-[6]. The maximum information rate designs of this paper require a different proof technique than the proof technique of Schizas, et al. [1]. The main difference in proof arises in the nonlinearity of the determinant function and the complexity of its derivative.

Let us place our work in the context of prior art, by again making reference to Fig. 1. The problem addressed by Schizas et al. [1] is to design the compression matrix $W$ so that the measurement $z$ may be filtered to produce a minimum mean-squared error estimate of the signal $\theta$. We generalize this problem to the maximization of information rate and show that canonical coordinates are central to both criteria. The solutions of [1] and this paper generalize the work on minimum mean-squared error and maximum information rate designs for reduced-rank filtering [2]-[6] and for precoders and equalizers (e.g., [7], [8], [10], [11], and many others). The literature on reduced-rank filtering assumes that the channel matrix $D$ is identity and the channel noise $v$ is zero. In this case, the compression matrix $W$ to minimize mean-squared error was originally found in [2] and the design to maximize information rate was originally found in [5]. These findings were then exploited for a variety of filtering and compression purposes in [3], [4], and [6]. The result of [10] is an instance of pre-coding and equalizing.

So we may summarize by saying that the theory of canonical coordinates treats the problem of compression when there is noise at the input to the compressor and the theory of scaled and rotated canonical coordinates developed in this paper treats the problem of compression when there is noise at the input and the output of the compressor. Noise at the output brings an important element of design to the compression problem, for it forces a constraint on the power out of the compressor $W$, a constraint that leads to rather complicated reasoning about Lagrangians and the KKT conditions for optimality, as for example in the prior work of [1], [7], and [8].

The rest of this paper is organized as follows. In Section II, we briefly introduce the problem of interest. In Section III, in the channel-noise-free case, the compression matrix returns half canonical coordinates for trace minimization, and full canonical coordinates for determinant minimization. In Section IV, when the compressed measurement is transmitted over a noisy channel, the compression matrix for trace or determinant minimization returns a scaled and rotated canonical coordinate design. Moreover, the scaling matrix, which accounts for canonical correlations and channel noise variance, has a mercury/waterfilling interpretation. In Section V, we extend the trace and determinant criteria to differentiable functions of the error covariance and establish a unified factorization for the optimal compression. Section VI concludes the paper.

**Notation.** The set of length $m$ real vectors is denoted by $\mathbb{R}^m$ and the set of $m \times n$ real matrices is denoted $\mathbb{R}^{m \times n}$. Bold upper case letters denote matrices, boldface lower case letters denote column vectors, and italics denote scalars. The scalar $x_i$ denotes the $i$th element of vector $x$, and $x_{ij}$ denotes the element of $X$ at row $i$ and column $j$. The $n \times n$ identity matrix is denoted $I_n$, and the $m \times n$ matrix with elements all 0 is denoted by $0_{m \times n}$. 

December 27, 2013 DRAFT
The transpose, inverse, pseudo inverse, trace and determinant of a matrix are denoted by \((\cdot)^T\), \((\cdot)^{-1}\), \((\cdot)^+\), tr\((\cdot)\) and det\((\cdot)\), respectively.

A covariance matrix is denoted by bold upper case \(Q\) with specified subscripts: \(Q_{zz}\) denotes the covariance matrix of a random vector \(z\); \(Q_{z_1z_2}\) is the cross-covariance matrix between \(z_1\) and \(z_2\); and \(Q_{z_1|z_2}\) is the error covariance for the linear minimum mean-squared error estimator (LMMSE estimator) of \(z_1\) from \(z_2\). Moreover, \(Q_{z_1|z_1}^{1/2}\) is defined from the Cholesky decomposition \(Q_{z_1z_1} = Q_{z_1|z_1}^{1/2}Q_{z_1z_1}^{T/2}\).

II. Problem Statement

Suppose that \(\theta \in \mathbb{R}^p\) is a random signal of interest, and \(x \in \mathbb{R}^n\) is a noisy measurement of \(\theta\). The measurement model used to describe the underlying correlation between the signal \(\theta\) and the measurement \(x\) can take any form. Notice that the measurement model is not necessarily linear, but it is sufficient to assume it so. That is, only second-order modeling is assumed, and once second-order correlations are specified, there is always an equivalent linear model that reproduces these second-order specifications. Consider the linear compression model, as depicted in Fig. 1,

\[
z = DWx + v. \tag{1}
\]

Here \(W \in \mathbb{R}^{m \times n}\) is the compression matrix, and \(Wx\) is the signal to be transmitted over a noisy channel with a full-rank channel matrix \(D \in \mathbb{R}^{m \times m}\) and random noise \(v \in \mathbb{R}^m\). It is assumed that the channel noise \(v\) has mean 0, and is independent of \(\theta\) and \(x\). Our objective is to design the compression matrix \(W\) such that the compressed measurement \(z\) is optimal according to a specified performance metric.

In this paper, we use linear estimation which is optimal in the multivariate normal case. In particular, given a measurement \(z\), the linear minimum mean squared error (LMMSE) estimator of \(\theta\) is

\[
\hat{\theta}_z = \mu_\theta + Q_{\theta z}Q_{zz}^+(z - \mu_z),
\]

where \(\mu_\theta, \mu_z\) are the means of \(\theta\) and \(z\), respectively, and \(Q_{zz}^+\) is the pseudo inverse of \(Q_{zz}\). The error covariance matrix of \(\hat{\theta}_z\), denoted by \(Q_{ee}\), is

\[
Q_{ee} = \mathbb{E}[(\theta - \hat{\theta}_z)(\theta - \hat{\theta}_z)^T] = Q_{\theta \theta} - Q_{\theta z}Q_{zz}^+Q_{z \theta}.
\]

Under model (1), \(Q_{ee}\) is a function of the compression matrix \(W\); that is

\[
Q_{ee} = Q_{\theta \theta} - Q_{\theta x}W^TD^T \\
(DWQ_{xx}W^TD^T + Q_{vv})^{-1}DWQ_{x \theta} \tag{2}
\]

In this paper, we assume that the covariance matrices \(Q_{\theta \theta}, Q_{xx}, Q_{vv}\) and the cross-covariance matrix \(Q_{\theta x}\) are known. In practice, the covariance matrices are determined from the physics of a problem or estimated from a two-channel experiment that generates realizations of \((\theta, x)\). Only the second order moments are required, not the exact distribution of the random signals.
The performance of the compression matrix is determined by evaluating functions of the resulting error covariance $Q_{ee}$. In the literature, the most prominent functions are the determinant criterion, $\det(Q_{ee})$, the average-variance criterion, $(\text{tr}(Q_{ee}^{-1}))^{-1}$, the smallest-eigenvalue criterion, $\lambda_{\min}(Q_{ee})$, and the trace criterion, $\text{tr}(Q_{ee})$. See [14] for more detailed review and discussion. All these criteria provide a reasonable measure of "largeness" of the error covariance matrix $Q_{ee}$. Consequently, the optimal compression matrix $W$ can be obtained by solving an optimization problem using one of the aforementioned criteria. In Sections III and IV, we will focus on two classical criteria: $\text{tr}(Q_{ee})$ and $\det(Q_{ee})$. The first measure $\text{tr}(Q_{ee})$ is the mean squared error of $\hat{\theta}$. This criterion has been studied by Schizas, et al. [1]. The second measure $\det(Q_{ee})$ is the volume of the error concentration ellipsoid. When $z$ and $\theta$ are jointly Gaussian distributed, minimizing $\det(Q_{ee})$ is equivalent to maximizing the mutual information between $z$ and $\theta$, or the differential information rate at which measurement $z$ brings information about $\theta$ [12]. For simplicity, let us refer to the problems where we try to minimize $\text{tr}(Q_{ee})$ and $\det(Q_{ee})$ as the min-trace and min-det problems, respectively. In Section V, we will explore more general criteria, a class of differentiable functions of the error covariance matrix.

III. CHANNEL-NOISE-FREE COMPRESSION DESIGN

In this section, we study a special case of (1) in which the compressed measurement can be transmitted perfectly, i.e., $z = Wx$. In particular, the error covariance matrix is

$$Q_{ee} = Q_{\theta\theta} - Q_{\theta x} W^T (W Q_{xx} W^T)^{-1} W Q_{x\theta}.$$  \hfill (3)

The solutions of the min-trace and min-det problems can be obtained by directly applying the results on optimal reduced-rank filtering [2]-[6].

First, we will discuss a notion of canonical coordinates. The basic idea is to transfer $(\theta, x)$ to canonical coordinates $(\tilde{\theta}, \tilde{x})$ which have a diagonal cross-covariance matrix. For the min-trace problem, we consider the Singular Value Decomposition (SVD) of the half coherence matrix ([2], [5], [6])

$$Q_{\theta x} Q_{xx}^{-T/2} = F K G^T,$$  \hfill (4)

where $K \in \mathbb{R}^{p \times n}$ is a diagonal matrix with diagonal elements $k_1 \geq \ldots \geq k_{\min\{n,p\}} \geq 0$, and $F \in \mathbb{R}^{p \times p}$ and $G \in \mathbb{R}^{n \times n}$ are orthogonal matrices. The vectors $\tilde{\theta} = F^T \theta$ and $\tilde{x} = G^T Q_{xx}^{-1/2} x$ are the half canonical coordinates for $\theta$ and $x$, respectively. Note that the cross-covariance matrix between $\tilde{\theta}$ and $\tilde{x}$ is the diagonal matrix $K$ given in (4).

For the min-det problem, the choice of the canonical coordinates is different. In this case, we consider an SVD of the coherence matrix ([3]-[6], [13])

$$Q_{\theta\theta}^{-1/2} Q_{\theta x} Q_{xx}^{-T/2} = F K G^T,$$  \hfill (5)

where $K \in \mathbb{R}^{p \times n}$ is a diagonal matrix with diagonal elements $k_1 \geq \ldots \geq k_{\min\{n,p\}} \geq 0$, and $F \in \mathbb{R}^{p \times p}$ and $G \in \mathbb{R}^{n \times n}$ are orthogonal matrices. Now, the vectors $\tilde{\theta} = F^T Q_{\theta\theta}^{-1/2} \theta$ and $\tilde{x} = G^T Q_{xx}^{-1/2} x$ are the full canonical
coordinates of $\theta$ and $x$ respectively. Note that for the simplicity of our notation, we choose to re-use the variables $F$, $K$, and $G$ for both SVDs.

The optimal compression matrix is given in Proposition 1, which is a re-statement of the results of [2] and [5].

**Proposition 1:**

For the min-trace and min-det problems, the optimal compression matrix $W^*_0 \in \mathbb{R}^{m \times n}$ can be written as

$$W^*_0 = G_m^T Q_{xx}^{-1/2}$$

where $G_m$ consists of the first $m$ columns of $G$. The matrix $G$ is defined in (4) for the min-trace problem and in (5) for the min-det problem. Moreover, for any $m \times m$ nonsingular matrix $T$, $TW^*_0$ is also an optimal compression matrix.

Proposition 1 figures prominently in our derivation of scaled and rotated canonical coordinates for optimum compression with channel noise. It is also worth mentioning that $W^*_0 x$ returns the first $m$ canonical coordinates in $\tilde{x}$. Let $W^*_{tr,0}$ and $W^*_{det,0}$ denote the optimal compression matrices given in Proposition 1. Straightforward calculation yields that, using the compression matrix $W^*_{tr,0}$, the minimum MSE of $\hat{\theta}_z$ [2] is

$$\text{tr}(Q_{ee}(W^*_{tr,0})) = \text{tr}(Q_{\theta\theta|x}) + \sum_{i=m+1}^{\min\{n,p\}} k_i^2,$$

where $Q_{\theta\theta|x}$ is the error covariance for the LMMSE of $\theta$ given $x$, and $\sum_{i=m+1}^{\min\{n,p\}} k_i^2$ is the minimum increase of the MSE. In addition, using $W^*_{det,0}$, the resulting minimum volume of the error concentration ellipsoid ([3], [4], [6]) is

$$\det(Q_{ee}(W^*_{det,0})) = \det Q_{\theta\theta|x} \prod_{i=m+1}^{\min\{n,p\}} \frac{1}{1-k_i^2}.$$  

Note that, in the min-det problem, the diagonal elements of $K$, i.e., $k_1, \ldots, k_{\min\{n,p\}}$, are the full canonical correlations that measure cosines of principle angles between $\theta$ and $x$ [6]. In general, the $k_i$’s take values between 0 and 1, but in (7), we assume the $k_i$’s are strictly less than 1. It is easy to see that $\det(Q_{ee}(W^*_{det,0})) \geq \det(Q_{\theta\theta|x})$, which shows that compression indeed discards some information about $\theta$ by compressing $x$ to a lower-dimensional measurement.

**IV. COMPRESSION DESIGN WITH SENSOR NOISE AND CHANNEL NOISE**

Now we extend the results in Section III by considering the linear compression of the noisy measurement to be transmitted over a noisy channel. We assume the channel noise $v$ has mean zero and covariance matrix $Q_{vv}$, and $v$ is independent of $\theta$ and $x$.

A significant feature of the design for noisy transmission is the need for a power constraint on the compression matrix, for otherwise the design problem is not well-defined. In this paper, we restrict the compression matrix $W$ subject to $\text{tr}(WQ_{xx}W^T) \leq P$ for a given power budget $P$.

Define $Q_{\omega\omega} = D^{-1}Q_{vv}(D^{-1})^T$ with the eigendecomposition $Q_{\omega\omega} = U_\omega \Sigma_\omega U_\omega^T$, where $U_\omega$ is an $m \times m$ orthogonal matrix and $\Sigma_\omega \in \mathbb{R}^{m \times m}$ is a diagonal matrix with diagonal elements $0 < \sigma_{\omega,11}^2 \leq \ldots \leq \sigma_{\omega,mm}^2$. 

December 27, 2013 DRAFT
A. Min-Trace Compression with Channel Noise

Under the power constraint, Schizas, et al. [1] have derived the optimal compression matrix to minimize \( \text{tr}(Q_{ee}) \). In Theorem 1, we re-state their result as a scaled and rotated canonical coordinate design.

**Theorem 1:** An optimal compression matrix \( W^*_\text{tr} \) minimizing \( \text{tr}(Q_{ee}) \) is given by [1]

\[
W^*_\text{tr} = U_\omega \Sigma^*_\text{tr} G^T Q^{-1/2}_{xx} \tag{8}
\]

Here the matrices \( G \) and \( K \) are given in (4), \( \Sigma^*_\text{tr} \) is an \( m \times n \) diagonal matrix with diagonal elements

\[
\sigma_{ii} = \begin{cases} 
\sqrt{k_i} \sigma_{\omega,ii} \sqrt{1 - \sigma^2_{\omega,ii}} & i = 1, \ldots, \kappa \\
0 & i = \kappa + 1, \ldots, m,
\end{cases}
\tag{9}
\]

with \( \kappa \) the maximum integer between 1 and \( \text{rank}(K) \) such that \( \sigma_{ii} > 0 \) for \( i = 1, \ldots, \kappa \), and

\[
\mu = \left( P + \sum_{i=1}^{\kappa} \sigma^2_{\omega,ii} \right)^{-1} \sum_{i=1}^{\kappa} \sigma_{\omega,ii} k_i \right)^2.
\]

It can be seen that \( W^*_\text{tr} \) factors into whitening \( Q^{-1/2}_{xx} \), canonical coordinate transformation \( G^T \), scaling \( \Sigma^*_\text{tr} \) and rotation \( U_\omega \) into the sub-dominant invariant subspace of \( Q_{\omega\omega} \).

B. Min-Det Compression with Channel Noise

The optimal compression matrix \( W \) to minimize \( \det(Q_{ee}) \) under the power constraint solves the optimization problem,

\[
W^*_{\text{det}} = \arg \min_{W \in \mathbb{R}^{m \times n}} \det(Q_{ee}) \text{ subject to } \text{tr}(WQ_{xx}W^T) \leq P. \tag{10}
\]

The matrix \( W \) has \( mn \) degrees of freedom. But let’s restrict \( W \) to a subset of \( \mathbb{R}^{m \times n} \), over which the local minimizer of \( \det(Q_{ee}) \) can be expressed explicitly. For a given \( n \times n \) orthogonal matrix \( V \), define \( \Omega_V = \{ U_\omega \Pi_m \Sigma \Pi_n^T V^T Q^{-1/2}_{xx} : \text{where } \Pi_m \in \mathbb{R}^{m \times m}, \Pi_n \in \mathbb{R}^{n \times n} \text{ are permutation matrices, and } \Sigma \in \mathbb{R}^{m \times n} \text{ is diagonal with } \sum_{i=1}^{m} \sigma^2_{ii} \leq P \} \). For fixed matrices \( U_\omega \) and \( Q^{-1/2}_{xx} \), the set \( \Omega_V \) is a subset of the constrained space of problem (10). Therefore, the local minimizer of \( \det(Q_{ee}) \) over \( W \in \Omega_V \) generally gives a suboptimal solution for problem (10). However, in Lemma 1, we show that, for a suitable choice of \( V \), the suboptimal solution on \( \Omega_V \) is a global optimal solution for problem (10).

**Lemma 1:** Suppose that \( G \) is the orthogonal matrix given in (5). Then,

\[
\min_{W \in \Omega_G} \det(Q_{ee}) = \det(Q_{ee}(W^*_{\text{det}})).
\]

The proof is given in Appendix A. Following a similar proof, we can show that Lemma 1 holds for the min-trace problem as well, with \( G \) given in (4).
From Lemma 1, it can be seen that the local minimizer over $\Omega_G$ is also a global minimizer of (10). For any $W \in \Omega_G$, we have

$$
det(Q_{ee}(W)) = det(Q_{00|x}) \times$$

$$
det \left( I_n + \Pi_n^T \Gamma \Pi_n (I_n + \Sigma_n^T \Pi_n^T \Sigma_n^{-1} \Pi_m \Sigma)^{-1} \right).$$

(11)

Here $\Gamma = \mathcal{K}^T (I_p - \mathcal{K} \mathcal{K}^T)^{-1} \mathcal{K}$, with $\mathcal{K}$ given in (5), is an $n \times n$ positive semi-definite diagonal matrix with diagonal elements $\gamma_i^2 = k_i^2/(1 - k_i^2)$ for $i = 1, \ldots, \min\{n, p\}$ and 0 otherwise. We require $0 \leq k_i < 1$ for all $i$, and consequently, $\gamma_1^2, \ldots, \gamma_n^2$ is a finite decreasing sequence. The permutation matrices $\Pi_n$ and $\Pi_m$ reorder the diagonal elements of $\Gamma$ and $\Sigma_n^{-1}$, respectively. In fact, for any $W \in \Omega_G$, $\Pi_n$ reorders the canonical coordinates $G^T Q_{ex}^{-1/2} x$ and determines which $m$ coordinates will be transmitted, and the permutation matrix $\Pi_m$ reorders the selected coordinates and determines which subchannel the coordinates will be transmitted over. The optimal compression matrix can be obtained by minimizing $det(Q_{ee}(W))$ with respect to the permutation matrices $\Pi_m$, $\Pi_n$ and the diagonal matrix $\Sigma$. The computational complexity of this optimization has been greatly reduced since there are just $2m + n$ degrees of freedom in the permutation matrices $\Pi_m$, $\Pi_n$ and the diagonal matrix $\Sigma$. We give in Theorem 2 the closed-form expression for the optimal compression matrix $W_{det}^*$.

**Theorem 2:** Suppose the matrix $Q_{\omega \omega}$ has distinct eigenvalues, i.e., $0 < \sigma_{\omega,11}^2 < \ldots < \sigma_{\omega,mm}^2$. Then, the optimal compression matrix $W_{det}^*$ solving problem (10) is

$$W_{det}^* = U_{\omega} \Sigma_{det}^* G^T Q_{xx}^{-1/2}$$

(12)

Here $G$ is given in (5) where the matrix $K$ contains singular values $0 \leq k_i < 1$ for all $i$; $\Sigma_{det}^* \in \mathbb{R}^{m \times n}$ is a diagonal matrix with diagonal elements $\sigma_{11}^*, \ldots, \sigma_{mm}^*$ such that

$$\sigma_{ii}^* = \begin{cases} 
\frac{1}{2} \sigma_{\omega,ii}^2 \left( -2 - \gamma_i^2 + \sqrt{\gamma_i^4 + 4 \gamma_i^2 - \omega_{ii}^2} \right) & \text{for } i = 1, \ldots, \kappa \\
0 & \text{for } i = \kappa + 1, \ldots, m
\end{cases}$$

(13)

where $\kappa$ is the maximum integer between 1 and $m$ such that $\sigma_{ii}^* > 0$ or equivalently $\sigma_{\omega,ii}^2/k_i^2 < 1/\mu$ for $i = 1, \ldots, \kappa$. The value of $\mu$ is nonnegative and uniquely solves $\sum_{i=1}^m \sigma_{ii}^2 = P$. Moreover, the diagonal element of $\Sigma_{det}^*$ are decreasingly ordered, i.e., $\sigma_{11}^* \geq \ldots \geq \sigma_{mm}^* \geq 0$.

The proof of Theorem 2 is given in Appendix B. Notice that the min-det design of this paper requires a different proof technique than the proof technique of [1]. More specifically, in our proof, we first show that the determinant minimizing solution can be restricted to a class of special structured matrices with reduced degrees of freedom. This result follows the standard Lagrangian technique and is similar to the proof of [1]. The second part shows the optimal permutation matrix is in fact the identity matrix. For the min-trace problem, this result is easy to show by the properties of the trace function. However, for the min-det problem, due to the nonlinearity and complexity of the determinant function, we use a completely different proof technique, with more details given in Appendix B.

Given that $\gamma_1^2 \geq \ldots \geq \gamma_n^2$ and $\sigma_{\omega,11}^2 \leq \ldots \leq \sigma_{\omega,mm}^2$, the optimal permutation matrices $\Pi_m$ and $\Pi_n$ are both identity matrices. This indicates that the canonical coordinates of the measurement with higher canonical correlation
with the canonical coordinates of $\mathbf{\theta}$ are transmitted over the subchannels with lower noise. The decreasingly ordered sequence of scalings $\sigma_{11}^* \geq \ldots \geq \sigma_{mm}^* \geq 0$ shows that the subchannels with higher correlation and lower noise are assigned higher power.

Simple calculation shows the minimum determinant of the error covariance to be

$$
\det(Q_{ee}(W_{det}^*)) = \det Q_{\bar{\theta}\bar{\theta}|x} \prod_{i=\kappa+1}^{\min\{n,p\}} \frac{1}{1 - k_i^2} \times \prod_{i=1}^{\kappa} \left(1 + \frac{2}{\sqrt{1 + 4(\gamma_i^2 \sigma_{ii}^2 \omega_{ii})^{-1} - 1}}\right).
$$

(14)

The first term on the right hand side is the minimum volume of the error concentration ellipsoid with no dimension reduction; the second term scales this volume according to canonical correlations of discarded canonical coordinates; the third term scales the volume by a term that depends on the channel noise variance, the power $P$, and the full canonical correlations. The integer $\kappa$ is the number of subchannels assigned with positive power. In fact, $\kappa/n$ is the optimal compression ratio for a given power $P$.

In Theorem 2, it is assumed that all eigenvalues of $Q_{\omega\omega}$ are distinct. Notice that, if some eigenvalues have multiplicity greater than 1, one can perturb $Q_{\omega\omega}$ by $\delta Q$ such that the matrix $\bar{Q}_{\omega\omega} = Q_{\omega\omega} + \delta Q$ has distinct eigenvalues. Moreover, we can restrict $\delta Q$ such that the eigenspace of $\bar{Q}_{\omega\omega}$ is fixed. Because the optimal entries in (13) are continuous functions of $\sigma_{11}^2, \ldots, \sigma_{mm}^2$, we can obtain the optimal compression matrix by letting $\delta Q$ go to zero.

It is worth mentioning that, for a sufficiently large $P$, we have $1/\mu > \sigma_{ii}^2/k_i^2$ (or equivalently $\sigma_{ii}^2 > 0$) for $i = 1, \ldots, m$. Consequently, the solution given in Theorem 2 is also an optimal compression for the channel-noise-free case. We simply let the nonsingular matrix $T$ in Proposition 1 be $T = U_\omega \text{diag}(\sigma_{11}^*, \ldots, \sigma_{mm}^*)$. When $P$ goes to infinity, the third part in (14) goes to 1, and the minimum determinant of $Q_{ee} \text{diag}(\sigma_{11}^*, \ldots, \sigma_{mm}^*)$ converges to the channel-noise-free case. On the other hand, the diagonal elements of $\Sigma_{det}^*$ go to infinity. Therefore, we can see that the optimization problem is ill-posed without a (finite) power constraint.

Finally, we comment on the canonical correlations. Under our current framework, all canonical correlations, $k_i$, are less than 1. In fact, the factorization in (12) still holds when $k_i = 1$ with a different scaling matrix $\Sigma_{det}^*$. In the sensor-noise-free case, suppose that $\mathbf{x} = \mathbf{H}\mathbf{\theta}$ and the matrix $\mathbf{H} \in \mathbb{R}^{n \times p}$ has rank $p$. The full canonical correlations between $\mathbf{\theta}$ and $\mathbf{x}$ are all 1. In this case, the compressor $\mathbf{W}$ operates on $\mathbf{H}\mathbf{\theta}$ directly and the design of $\mathbf{W}$ becomes a precoder design problem [7], [8], [10]. The optimal scaling matrix $\Sigma_{det}^*$ has diagonal elements

$$
\sigma_{ii}^* = \begin{cases} 
\sqrt{\frac{1}{\mu} - \sigma_{ii}^2} & \sigma_{ii}^2 < 1/\mu \\
0 & \sigma_{ii}^2 \geq 1/\mu
\end{cases}
$$

(15)

where the value of $\mu$ is determined by the power constraint $\sum_{i=1}^{m} \sigma_{ii}^2 = P$.

C. A Circulant Model for Fourier Coefficients

Consider a system where Fourier coefficients $\mathbf{\theta}$ are to be measured, as in imaging or interferometry. Suppose that $\mathbf{\theta}$ and $\mathbf{x}$ have circulant covariance and cross-covariance matrices with the Discrete Fourier Transform (DFT)
representations

\[ Q_{\theta\theta} = V_n S_{\theta\theta} V_n^H; \]
\[ Q_{xx} = V_n S_{xx} V_n^H; \]
\[ Q_{\theta x} = V_n S_{\theta x} V_n^H. \]

Here \( S_{\theta\theta}, S_{\theta x} \) and \( S_{xx} \) are diagonal matrices with diagonal elements \( s_{\theta\theta,ii}, s_{\theta x,ii}, s_{xx,ii} \), respectively and \( V_n \in \mathbb{C}^{n \times n} \) is the DFT matrix. The coherence matrix \( Q_{\theta\theta}^{-1/2} Q_{\theta x} Q_{xx}^{-H/2} \) is also a circulant matrix with DFT representation

\[ Q_{\theta\theta}^{-1/2} Q_{\theta x} Q_{xx}^{-H/2} = V_n S_{\theta\theta}^{-1/2} S_{\theta x} S_{xx}^{-H/2} V_n^H. \]

Notice that \( S_{\theta\theta}^{-1/2} S_{\theta x} S_{xx}^{-H/2} \) is a diagonal matrix with diagonal elements \( s_{\theta x,ii}/\sqrt{s_{\theta\theta,ii}s_{xx,ii}} \), which are the canonical correlations between \( \theta \) and \( x \). The squared canonical correlations are \( k_i^2 = s_{\theta x,ii}^2/(s_{\theta\theta,ii}s_{xx,ii}) \). WLOG we assume the diagonal elements are decreasingly ordered. Based on the results in Theorem 2, the optimal compression that minimizing the determinant of the error covariance is

\[ W_{det}^* = U \Sigma_{det}^* V_n^H Q_{xx}^{-1/2}. \]

The canonical coordinates of \( x \), i.e., \( V_n^H Q_{xx}^{-1/2} x \), are the whitened \( x \), rotated by the DFT matrix \( V_n^H \). The diagonal scaling matrix \( \Sigma_{det}^* \) has diagonal elements \( \sigma_{ii}^* \) given in (13) with \( \gamma_i^2 = k_i^2/(1 - k_i^2) \). In summary, with the original measurement \( x \), the error covariance for the LMMSE has determinant

\[ \prod_{i=1}^n s_{\theta\theta,ii}(1 - k_i^2); k_i^2 = \frac{s_{\theta x,ii}^2}{s_{\theta\theta,ii}s_{xx,ii}}. \]

When \( x \) is compressed to the \( \kappa \)-dimensional space which is then transmitted over the noisy channel (1), the minimum determinant of the error covariance increases to

\[ \prod_{i=1}^n (s_{\theta\theta,ii}(1 - k_i^2)) \times \prod_{i=\kappa+1}^n \frac{1}{1 - k_i^2} \times \prod_{i=1}^\kappa \left( 1 + \frac{2}{\frac{4(1-k_i^2)}{k_i^2}\sigma_{\omega,ii}^2 - 1} \right). \]

Here \( \sigma_{\omega,ii}^2 \) is the noise variance in frequency band \( i \).

D. A Mercury/Waterfilling Explanation for Min-Det Optimal Compressor

First, consider a sensor-noise-free case, \( x = H\theta \), in which case \( Q_{\theta x} = Q_{\theta\theta} H^T \) and \( Q_{xx} = HQ_{\theta\theta} H^T \). The optimal compressor has been discussed in Section IV-B, with the factorization in (12) and \( \Sigma_{det}^* \) given in (15). The scaling matrix \( \Sigma_{det}^* \) distributes the power among all the \( m \) subchannels according to a waterfilling policy [12], with a graphical display given in Fig. 2. There are \( m \) vessels, each of which represents a subchannel. The goal is to pour water of total volume \( P \) into these vessels. Here, each vessel has its own solid base with height \( \sigma_{\omega,ii}^2 \).
Recall that $\sigma_{\omega,11}^2, \ldots, \sigma_{\omega,mm}^2$ are the eigenvalues of $Q_{\omega\omega}$, and can be viewed as the variances of the channel noise in the channel coordinates since the optimal compression rotates the rescaled canonical coordinates by $U_\omega$, the eigenvectors of $Q_{\omega\omega}$. The desired water level $1/\mu$ is determined by the power constraint, or equivalently, the total volume of water equals $P$. The optimal compressor pours water into each vessel until the water level reaches $1/\mu$. As a result, the water height in each vessel gives the power assigned to the corresponding subchannel. Note that less power will be allocated to noisier subchannels, and no power will be assigned to subchannels with noise variance larger than $1/\mu$.

![Waterfilling diagram](image)

Fig. 2. Waterfilling without sensor noise. The total volume of water is $P$, and the water height over the solid base on the $i$th vessel gives the power for the $i$th subchannel.

In general, $x$ is a noisy measurement of $\theta$ and the full canonical correlations are strictly less than 1. Therefore, the optimal power allocation policy needs to be adjusted according to the canonical correlations. As a consequence of Theorem 2, the solution can be interpreted as a mercury/waterfilling policy, which is a three-step procedure that has been introduced in [16]:

1) For the $i$th vessel, fill in the solid base with height $\sigma_{\omega,ii}^2/k_i^2$.
2) Compute $\mu$ from the power constraint. For the vessels with base height less than $1/\mu$, fill in mercury in the vessel until the height reaches the maximum of $\sigma_{\omega,ii}^2/k_i^2$ and
   $$\frac{1}{\mu} - \frac{1}{2} \sigma_{\omega,ii}^2 \left( -2 - \gamma_i^2 + \sqrt{\gamma_i^4 + 4 \frac{\gamma_i^2}{\sigma_{\omega,ii}^2/\mu}} \right).$$
3) Pour water into all vessels until the height of water in each vessel reaches $1/\mu$.

In this mercury/waterfilling policy, $1/\mu$ is the parameter in the formula for water volume $\sigma_{ii}^2$ that minimizes $\det(Q_{xx})$ under the constraint that the total volume of water is $P$. Given the value of $\mu$, the determinant of the error covariance is minimized when the value of $\sigma_{ii}^2$ equals the height of water in the corresponding vessel.

The height of the solid base, $\sigma_{\omega,ii}^2/k_i^2$, is the variance of the channel noise in the $i$th vessel divided by the $i$th squared canonical correlation. A higher solid base means a less informative channel with high channel noise and weak correlation of $x$ with $\theta$. For any vessel with base height exceeding $1/\mu$, neither mercury nor water will be added, or equivalently, no power will be assigned to the corresponding subchannel.

While the base height determines whether water will be added, the mercury stage regulates the water level for each vessel. Without adding mercury, the optimal power allocation will have variable solid-plus-water levels among different vessels. The mercury is added to balance the sensor noise contained in $x$ and the channel noise added in
transmission. Recall that no mercury is added in the special case when \( x = \theta \). The water height in each vessel is the optimal power assigned to the corresponding subchannel. As demonstrated in Theorem 2, the water height for each vessel is decreasingly ordered.

![Fig. 3. A mercury/waterfilling policy. The total volume of water is \( P \), and the water height over mercury on the \( i \)th vessel gives \( \sigma_{ii}^2 \).](image)

E. Scaled and Rotated Canonical Coordinate Design

\[
W \xrightarrow{Q_{xx}^{-1/2}} [G^T \Sigma^* U_{\omega}] \xrightarrow{D} y \\
\xrightarrow{z}
\]

Fig. 4. Scaled canonical coordinate transformation for compressing a noisy measurement with transmission over a noisy channel.

Theorems 1 and 2 suggest a common architecture for compression, which specializes to all previous designs for reduced-rank filtering and for reduced rank precoding and equalizing. The optimal compressor can be factored into four component matrices. As shown in Fig. 4, the first matrix \( Q_{xx}^{-1/2} \) whitens the noisy measurement \( x \). The second matrix \( G^T \) transforms the whitened measurement into a canonical coordinate system. For the min-det problem, the full canonical coordinates, \( G^T Q_{xx}^{-1/2} x \), are uncorrelated and have unit variance. The third matrix \( \Sigma^* \in \mathbb{R}^{m \times n} \) is diagonal. The role of \( \Sigma^* \) is to extract the first \( m \) full canonical coordinates and distribute power across the canonical channels. The \( i \)th canonical coordinate is scaled to have power \( \sigma_{ii}^2 \). For the min-det problem, when \( \gamma_i^2 = 0 \) (i.e., \( k_i = 0 \)), the corresponding scaling is \( \sigma_{ii} = 0 \), which means those canonical coordinates uncorrelated or weakly correlated with \( \theta \) will be eliminated. In general, the diagonal elements of \( \Sigma^* \) have a mercury/waterfilling interpretation. The matrix \( U_{\omega} \) rotates the compressed canonical coordinates into the sub-dominant invariant subspace of the matrix \( Q_{\omega \omega} \).

The difference between the trace and determinant designs is in the canonical coordinates and in the values of scaling constants in the diagonal scaling matrix. But the universe architecture of Fig. 4 remains unchanged.

V. A Unified Framework For Optimal Compression

In the previous sections, our interest has centered on optimal compression under two commonly used criteria: trace and determinant. Next, we consider the problem of designing a compression matrix to minimize a general
where the following condition:

\[ W_\phi^* = \arg\min_{W \in \mathbb{R}^{m \times n}} \varphi(Q_{ee}) \text{ s.t. } \text{tr}(WQ_{xx}W^T) \leq P. \tag{16} \]

Here \( \varphi \) is a differentiable function on the space of \( p \times p \) positive definite matrices.

Denote the first derivative of \( \varphi \) by \( \varphi' \). Then \( \varphi' \) is a mapping from \( \mathbb{R}^{p \times p} \) to \( \mathbb{R}^{p \times p} \), with

\[ (\varphi'(Q_{ee}))_{ij} = \lim_{t \to 0} \frac{\varphi(Q_{ee} + tJ_{ij}) - \varphi(Q_{ee})}{t}, \]

where \( J_{ij} \) is the \( p \times p \) single-entry matrix with 1 at \( (i, j) \) and 0 elsewhere.

We first establish a unified factorization of \( W_\phi^* \) in the following theorem.

**Theorem 3:** Suppose that the diagonal matrix \( \Sigma_\omega \) has distinct diagonal elements. Then for any minimizer \( W_\phi^* \) of (16), there exists an \( m \times m \) permutation matrix \( \Pi_m^* \), an \( m \times n \) diagonal matrix \( \Sigma^* \), and an \( n \times n \) orthogonal matrix \( V^* \) such that

\[ W_\phi^* = U_\omega \Pi_m^* \Sigma^* V^{*T} Q_{xx}^{-1/2}. \tag{17} \]

The proof of Theorem 3 is given in Appendix C. Theorem 3 suggests that the optimal compression matrix can be expressed as a sequence of operations, including whitening \((Q_{xx}^{-1/2})\), coordinate system transformation \((V^T)\), scaling \((\Sigma)\), re-ordering \((\Pi_m)\) and rotation to the invariant subspace of the channel noise \((U_\omega)\). The factorization structure in Theorem 3 holds for the min-trace and min-det problems studied in Section IV, where \( V^* \) is given by the full and half canonical coordinate systems in (4) and (5), respectively. Moreover, by choosing \( U_\omega \) as the subdominant invariant space of \( Q_{\omega\omega} \), the optimal permutation matrix \( \Pi_m^* \) is the identity matrix. As a consequence, in both cases, the coordinates with higher correlation are transmitted over the subchannels with lower noise.

In general, searching for a global minimizer of problem (16) is challenging, even with the specified structure in (17). The following proposition provides a necessary condition for the optimal compression matrix \( W_\phi^* \), which further restricts the factorization structure in (17).

**Proposition 2:** Let \( W_\phi^* \) be a solution to problem (16) with \( W_\phi^* = U_\omega \Pi_m^* \Sigma^* V^{*T} Q_{xx}^{-1/2} \). Then, it must satisfy the following condition:

\[ [V^{*T}L^T \varphi'(Q_{ee})LV^*, (I_n + \Delta^*)^{-1}] = 0_{n \times n}, \tag{18} \]

where \( \Delta^* = \Sigma^* \Pi_m^T \Sigma_{\omega}^{-1} \Pi_m^* \Sigma^* \in \mathbb{R}^{n \times n}, Q_{ee}^* \) is the error covariance of \( \theta \) corresponding to the compression \( W_\phi^* \), \( \varphi' \) is the matrix derivative function of the criterion \( \varphi \), \( L = Q_{\theta x} Q_{xx}^{-1/2} \in \mathbb{R}^{p \times n} \) and \([A, B] = AB - BA\).

The proof is given in Appendix D. The main idea is that the directional derivative of the criterion function \( \varphi \) at \( V^* \) on the manifold of orthogonal matrices is zero, since the power \( \text{tr}(WQ_{xx}W^T) \) is invariant to the orthogonal matrix \( V^* \).

We must point out that Proposition 2 does not yield a closed-form expression for \( W_\phi^* \). In fact, the compression matrix satisfying condition (18) is not necessarily the solution to problem (16). Moreover, the solution for condition (18) is generally intractable. The primary purpose for Proposition 2, as a necessary condition for any optimal compression, is to further specify the factorization structure in (17), with the hope of reducing the computation cost.
of searching for the optimal compression. For example, in many scenarios, condition (18) confines the choice of $V^*$ to a reduced subset of the set of $n \times n$ orthogonal matrices. Suppose that the nonnegative diagonal matrix $\Delta^*$ has $r$ ($r \leq n$) strictly positive and distinct diagonal elements and $n - r$ zeros. Moreover, we say an orthogonal matrix $V$ diagonalizes a non-negative matrix $A$ if the product matrix $V^TAV$ is diagonal. Then, condition (18) implies that the matrix $V_r^*$ diagonalizes $L^T\varphi'(Q_{ee})L$, where $V_r^*$ is a $n \times r$ matrix consisting of the first $r$ columns of $V^*$. Moreover, one can show that the value of $\varphi(Q_{ee})$ is invariant to the choice of the last $n - r$ columns of $V^*$. A choice is to choose the last $n - r$ columns of $V^*$ to diagonalize $L^T\varphi'(Q_{ee})L$. In summary, the optimal orthogonal matrix $V^*$ diagonalizes the matrix $L^T\varphi'(Q_{ee})L$. For some problems, this fact could lead to a deterministic choice of the matrix $V^*$, without knowing the scaling matrix $\Lambda^*$ and the permutation matrix $\Pi_m^*$, as shown in the next example.

**Example 1: Linear Criteria.** Consider a linear criterion $\varphi$, i.e.,

$$\varphi(A + B) = \varphi(A) + \varphi(B);$$

$$\varphi(\alpha A) = \alpha \varphi(A),$$

for any $A, B \in \mathbb{R}^{p \times p}$ and $\alpha \in \mathbb{R}^1$. It can be shown that the matrix derivative $\varphi'(Q_{ee})$, denoted by $M$, is a constant matrix, independent of the compression matrix. The condition (18) shows that the optimal orthogonal matrix $V$ must diagonalize the nonnegative matrix $L^TML$, meaning the columns of $V$ are the eigenvectors of $L^TML$. Consider the weighted trace criterion where $\varphi(Q_{ee}) = \text{tr}(BQ_{ee}B^T)$ for some matrix $B$. The derivative $\varphi'(Q_{ee})$ is $M = BB^T$. The optimal orthogonal matrix $V^*$ then is given by the eigenvectors of the nonnegative matrix $LBB^TL^T$, or equivalently the right singular vectors of $LB$. The order of the columns of $V^*$ can be arbitrary, and the optimal permutation matrix is determined correspondingly. As a special case, in the min-trace problem where $B = I$, the matrix $V^*$ is given by the right singular vectors of $L$, as stated in Theorem 1.

In Example 1, we consider the class of linear criteria. We have shown that condition (18) indeed gives an explicit form for the optimal orthogonal matrix $V^*$, which can be used to solve the optimization problem (16) analytically. However, this is a special case in which the choice of $V^*$ does not depend on the scaling matrix $\Sigma^*$. For more general criteria, the practical implementation of condition (18) could be limited due to the large size of the solution set. As pointed out by one anonymous reviewer, an interesting problem is how far a solution to condition (18) is away from the optimal compression. This is an open question that requires further investigation and it is beyond the scope of this paper.

**VI. Conclusion**

In this paper we have considered the problem of compressing a noisy measurement for transmission over a noisy channel, introduced in [1]. This problem generalizes the problem of reduced rank filtering and the problem of reduced rank precoder and equalizer design [2]-[11], producing those designs as special cases. We have shown that designs for minimizing trace or determinant of an error covariance matrix share a common architecture. In this architecture, a noisy sensor measurement is first transformed into a system of canonical coordinates. These
coordinates are then scaled and rotated into the sub-dominant subspace of the channel noise. The difference between the two designs resides in the definition of canonical coordinates and in the determination of the scaling constants. A generalization to differentiable functions of error covariance leads to a factorization theorem that supports practical design for general criteria.

APPENDIX A

PROOF FOR LEMMA 1

Let $A$ and $B$ be two $n \times n$ positive semi-definite matrices with eigenvalues $\mu_1 \geq \ldots \mu_n \geq 0$ and $\lambda_1 \geq \ldots \lambda_n \geq 0$. Then,

$$\det(I_n + AB) \geq \sum_{i=1}^{n} \lambda_i \mu_{n-i+1}. \quad (A.1)$$

The proof uses the same technique as the proof of Lemma 3 in [17] and is omitted here.

Since the determinant is a differentiable function, by Theorem 3, it suffices to consider the compression matrix $W$ with factorization

$$W = U_\omega \Pi_n \Sigma \Pi_n^T V^T Q_{xx}^{-1/2} \quad (A.2)$$

where $V$ is an orthogonal matrix, and $\Pi_n$ and $\Pi_n^T$ are permutation matrices. Define $\Lambda = (I_n + \Sigma \Pi_n^T \Sigma^{-1} \Pi_n) \Sigma^{-1}$. Then, $\det(Q_{xx})$ can be expressed as

$$\det(Q_{xx}) = \det(Q_{\theta \theta}) \det(I_p - CC^T) \times \det(I + \Pi_n \Lambda \Pi_n^T V^T C^T (I_p - CC^T)^{-1} CV)$$

Given the SVD $C = F K G^T$ in (5), the matrix $V^T C^T (I_p - CC^T)^{-1} CV = V^T G K^T (I_p - K K^T)^{-1} K G^T V$ has eigenvalues $\gamma_1^2 \geq \ldots \geq \gamma_n^2 \geq 0$ where $\gamma_i^2 = k_i^2 / (1 - k_i^2)$ for $i = 1, \ldots, \min\{n, p\}$ and 0 otherwise. Moreover, the matrix $\Pi_n \Lambda \Pi_n^T$ has the same eigenvalues as $\Lambda$. Let $W^0 = U_\omega \Pi_n \Sigma (\Pi_n^0)^T G^T Q_{xx}^{-1/2}$ where $\Pi_n^0$ is an $n \times n$ permutation matrix such that the diagonal elements of the diagonal matrix $\Pi_n \Lambda \Pi_n^T$ are increasingly ordered. Then, $W^0 \in \Omega_G$ and

$$\det(Q_{xx}(W)) \geq \det(Q_{\theta \theta}) \det(I_p - CC^T) \prod_{i=1}^{n} \gamma_i^2 \lambda_{(i)}$$

$$= \det(Q_{xx}(W^0)). \quad (A.3)$$

where the inequality is a direct consequence of (A.1). Minimizing both sides of (A.3) over $W$ with factorization (A.2) and $W^0 \in \Omega_G$, respectively, we have

$$\det(W^*_{det}) \geq \min_{W \in \Omega_G} \det(Q_{xx}(W)). \quad (A.4)$$

The proof is therefore completed by the fact that $\det(W^*_{det}) \leq \min_{W \in \Omega_G} \det(Q_{xx}(W))$ by definition.
APPENDIX B

PROOF FOR THEOREM 2

Given Lemma 1, we can restrict $W = U_\omega \Pi_m \Sigma \Pi_n^T G^T Q^{-1/2}_{xx}$ where $\Pi_n$ and $\Pi_m$ are permutation matrices, and $\Sigma \in \mathbb{R}^{m \times n}$ is a diagonal matrix with diagonal elements $\sigma_{11}, \ldots, \sigma_{mm}$. Let $\pi_m(i)$ be the index of the entry equal to unity in the $i$th column of $\Pi_m$, and $\pi_n(i)$ be the index of the unity entry in the $i$th column of $\Pi_n$. Then

$$\det(Q_{ee}(\sigma_{11}, \ldots, \sigma_{mm}, \pi_m, \pi_n)) = \prod_{i=1}^{m} \left(1 + \frac{\gamma_{\pi_m(i)}^2}{1 + \lambda_{\pi_m(i)} \sigma_{ii}^2}\right) \prod_{j=m+1}^{n} \log(1 + \gamma_{\pi_n(j)}^2) \quad (B.1)$$

where $\lambda_i = \sigma_{\omega,ii}^{-2}$ for $i = 1, \ldots, m$ with $\lambda_1 \geq \ldots \geq \lambda_m$. We try to minimize $\det(Q_{ee})$ over all possible permutations and $(\sigma_{11}, \ldots, \sigma_{mm})^T$ subject to $\sum_{i=1}^{m} \sigma_{ii}^2 \leq P$.

First of all, we show that the minimum of $\det(Q_{ee})$ can be achieved when $\pi_m(i) = i$ for $i = 1, \ldots, m$ and $\pi_n(j) = j$ for $j = 1, \ldots, n$, or equivalently, when the optimal permutation matrices $\Pi_m = I_m$ and $\Pi_n = I_n$. For a given permutation, define

$$f(\pi_n, \pi_m) = \min_{\sum_{i=1}^{m} \sigma_{ii}^2 \leq P} \det(Q_{ee}).$$

The optimal permutation $(\pi^*_n, \pi^*_m)$ satisfies

$$f(\pi^*_n, \pi^*_m) \leq f(\pi_n, \pi_m) \quad (B.2)$$

for all other permutations $(\pi_n, \pi_m)$. It is easy to see that for any $i = 1, \ldots, m$, one must have $\gamma_{\pi_m(i)}^2 \geq \max\{\gamma_{\pi_m(m+1)}^2, \ldots, \gamma_{\pi_m(n)}^2\}$. Moreover, since the orders of $\{\pi_n(j)\}_{j=m+1}^{n}$ do not affect the value of (B.1), we can set WLOG $\pi^*_n(j) = j$ for $j = m + 1, \ldots, n$. For $i = 1, \ldots, m$, it can be seen that $\pi_n(i)$ and $\pi_m(i)$ appear in (B.1) pairwise. Therefore, we can set WLOG that $\pi^*_n(i) = i$ for $i = 1, \ldots, \kappa$ and $f(\pi_n) := f(\pi_n, \pi^*_m)$. Then the objective is to search for the optimal permutation $\pi^*_n(i)$ that minimizes

$$\prod_{i=1}^{m} \left(1 + \frac{\gamma_{\pi_n(i)}^2}{1 + \lambda_i \sigma_{ii}^2}\right). \quad (B.3)$$

Let’s start from the simple case with $m = 2$. When $\lambda_1 = \lambda_2$ or $\gamma_1^2 = \gamma_2^2$, (B.3) is permutation invariant. When $\lambda_1 > \lambda_2$ and $\gamma_1^2 > \gamma_2^2$, only two permutations are possible, $\pi^*_n(i) = i$ or $\pi^*_n(i) = 2 - i$ for $i = 1, 2$. To minimize (B.3), consider the following functions

$$h_1(x) = (1 + \frac{\gamma_1^2}{1 + \lambda_1 Px}) \times (1 + \frac{\gamma_2^2}{1 + \lambda_2 P(1 - x)});$$

$$h_2(x) = (1 + \frac{\gamma_2^2}{1 + \lambda_2 Px}) \times (1 + \frac{\gamma_1^2}{1 + \lambda_1 P(1 - x)}),$$

with $Px = \sigma_1^2$ and $P(1 - x) = \sigma_2^2$, in which case $\sigma_1^2 + \sigma_2^2 = P$. Then, $f(\pi_n^1) < f(\pi_n^2)$ is equivalent to

$$\min_{x \in [0, 1]} h_1(x) < \min_{x \in [0, 1]} h_2(x). \quad (B.4)$$

Straightforward calculation gives that

$$h_2(x) - h_1(x) = \frac{P(\gamma_1^2 - \gamma_2^2)((\lambda_1 + \lambda_2)x - \lambda_2)}{(1 + \lambda_1 Px)(1 + \lambda_2 P(1 - x))}.$$
Therefore $h_2(x) > h_1(x)$ for any $x \in (\frac{\lambda_2}{\lambda_1 + \lambda_2}, 1]$. For $x \in [0, \frac{\lambda_2}{\lambda_1 + \lambda_2}]$, one can show that $h_2(x) > h_1(\frac{\lambda_2}{\lambda_1}(1 - x))$ with $\frac{\lambda_2}{\lambda_1}(1 - x) \in [\frac{\lambda_2}{\lambda_1 + \lambda_2}, 1]$. Therefore (B.4) holds and $\pi_n^*(i) = i$ for $i = 1, 2$. By checking the first and second derivative of $h_1(x)$, the minimum of $h_1$ is attained at $x \in [1/2, 1]$. This directly yields that $Px \geq P(1 - x)$ or equivalently, $\sigma_{11}^2 \geq \sigma_{22}^2$, meaning that allocation of power decreases with increasing channel index.

For the general cases where $m \geq 2$, define

$$
(\sigma_{11}^n, \ldots, \sigma_{mm}^n) := \arg \min_{\sum m, \sigma_{ii}^n \leq P} \det(Q_{ee}(\sigma_{11}, \ldots, \sigma_{mm}, \pi_n, \pi_n^m)).
$$

Suppose that there exist $1 \leq i < j \leq m$ with $\gamma_{\pi_n(i)}^2 < \gamma_{\pi_n(j)}^2$. Let $\hat{\pi}_n$ be a new permutation with $\hat{\pi}_n(i) = \pi_n(j)$, $\hat{\pi}_n(j) = \pi_n(i)$, and $\hat{\pi}_n(k) = \pi_n(k)$ for $k \neq i, j$. Define $(\tilde{\sigma}_{ii}, \tilde{\sigma}_{jj})$ to be

$$
\arg \min_{\sigma_{ii}^n + \sigma_{jj}^n \leq \sigma_{ii}^n + \sigma_{jj}^n} \left( 1 + \frac{\gamma_{\pi_n(i)}^2}{1 + \lambda_i \sigma_{ii}^n} \right) \left( 1 + \frac{\gamma_{\pi_n(j)}^2}{1 + \lambda_j \sigma_{jj}^n} \right).
$$

(B.5)

Given the result for the $m = 2$ case, it is straightforward to see that

$$
\det(Q_{ee}(\sigma_{11}^n, \ldots, \tilde{\sigma}_{ii}, \ldots, \tilde{\sigma}_{jj}, \ldots, \sigma_{mm}^n, \pi_n, \pi_m)) < \det(Q_{ee}(\sigma_{11}^n, \ldots, \sigma_{mm}^n, \pi_n, \pi_m))
$$

Therefore, the permutation $\pi_n$ cannot be the optimal permutation. Among all the permutations $\gamma_{\pi_n(1)}^2 \geq \ldots \geq \gamma_{\pi_n(m)}^2$, we can choose WLOG that $\pi_n^*(i) = i$ for $i = 1, \ldots, m$.

Next our problem focuses on finding the sequence

$$
\{\sigma_{ii}^n\}_{i=1}^m = \arg \min_{\sum_{i=1}^m \sigma_{ii}^n \leq P} \sum_{i=1}^m \log \left( 1 + \frac{\gamma_i^2}{1 + \lambda_i \sigma_{ii}^n} \right).
$$

(B.6)

The log operator is implemented to simplify calculation. Note that the objective function in (B.6) is a strictly convex function, therefore (B.6) is a convex optimization problem with unique minimizer. Moreover, the function is strictly decreasing in $\sigma_{ii}^2$. Hence the minimum is attained at $\sum_{i=1}^m \sigma_{ii}^2 = P$. The Lagrangian is

$$
L(\sigma_{11}, \ldots, \sigma_{mm}; \mu) = \sum_{i=1}^m \log \left( 1 + \frac{\gamma_i^2}{1 + \lambda_i \sigma_{ii}^2} \right) + \mu(\sum_{i=1}^m \sigma_{ii}^2 - P)
$$

By setting the first derivative of the Lagrangian with respect to $\sigma_{11}, \ldots, \sigma_{mm}$, and $\mu$ to zero, the necessary conditions for any minimizer of problem (B.6) are

$$
\begin{align*}
- \left( 1 + \frac{\gamma_i^2}{1 + \lambda_i \sigma_{ii}^2} \right)^{-1} \frac{2 \gamma_i^2 \lambda_i \sigma_{ii}^2}{(1 + \lambda_i \sigma_{ii}^2)^2} + 2 \mu \sigma_{ii}^2 &= 0 \\
\sum_{i=1}^m \sigma_{ii}^2 &= P
\end{align*}
$$

(B.7)

(B.8)

Equation (B.7) yields either $\sigma_{ii} = 0$ or

$$
\sigma_{ii} = \sqrt{\frac{1}{2 \lambda_i} \left( \sqrt{\gamma_i^4 + 4 \lambda_i \gamma_i^2 \mu^{-1} - 2} - \gamma_i^2 \right)}
$$

(B.9)
The solution in (B.9) provides a feasible solution for $\sigma_{ii}$ only when $\mu \leq \lambda_i k_i^2$ where $k_i^2 = \gamma_i^2 (1 + \gamma_i^2)^{-1}$ is the squared canonical correlation. The value of $\mu$ is solved by Equation (B.8).

Next we investigate the possible minimizers by checking the second derivative of $L$, which is

$$\frac{\partial^2 L(\sigma_{11}, \ldots, \sigma_{mm}; \mu)}{\partial \sigma_{ii}^2} = 2\mu - \frac{2\gamma_i^2 \lambda_i}{(1 + \gamma_i^2 + \lambda_i \sigma_{ii}^2)(1 + \lambda_i \sigma_{ii}^2)} + \frac{4\gamma_i^2 \lambda_i^2 \sigma_{ii}^2}{(1 + \gamma_i^2 + \lambda_i \sigma_{ii}^2)^2}$$

Substituting (B.9), the second derivative

$$\frac{\partial^2 L(\sigma_{11}, \ldots, \sigma_{mm}; \mu)}{\partial \sigma_{ii}^2} = 4\mu \lambda_i \sigma_{ii} \left( \frac{1}{1 + \gamma_i^2 + \lambda_i \sigma_{ii}^2} + \frac{1}{1 + \lambda_i \sigma_{ii}^2} \right)$$

is strictly positive when $\mu < \lambda_i k_i^2$. When $\sigma_{ii} = 0$,

$$\frac{\partial^2 L(\sigma_{11}, \ldots, \sigma_{mm}; \mu)}{\partial \sigma_{ii}^2} \bigg|_{\sigma_{ii} = 0} = 2\mu - \frac{2\lambda_i \gamma_i^2}{1 + \gamma_i^2}$$

which is negative when $\mu \leq \lambda_i k_i^2$ and positive when $\mu > \lambda_i k_i^2$.

Let $\kappa$ be the maximum integer between 1 and $m$ such that $\mu < \lambda_\kappa k_\kappa^2$ (or equivalently $\sigma_{ii} > 0$ for $i = 1, \ldots, \kappa$) and $\mu \geq \lambda_{\kappa+1} k_{\kappa+1}^2$ ($\sigma_{ii} = 0$ for $i = \kappa+1, \ldots, m$), where the value of $\mu$ is determined by the power constraint (B.8). Then, the Hessian matrix at

$$\sigma_{ii}^* = \begin{cases} \sqrt{\frac{1}{2\lambda_i} \left( \sqrt{\gamma_i^4 + 4\lambda_i \gamma_i^2 / \mu} - 2 - \gamma_i^2 \right)}, & \text{for } i = 1, \ldots, \kappa \\ 0, & \text{for } i = \kappa+1, \ldots, m \end{cases}$$

is strictly positive and (B.11) is the minimizer. As a summary, the optimal compression matrix minimizing $\det(Q_{ee})$ is

$$W_{\text{det}}^* = U_\omega \Sigma_{\text{det}}^* G^T Q_{xx}^{-1/2}$$

where $\Sigma_{\text{det}}^* \in \mathbb{R}^{m \times n}$ is a diagonal matrix with diagonal elements given in (B.11).

**Remark:** Our proof in Theorem 2 is parallel to the min-trace problem in [1]. In the min-trace problem, the Lagrange multiplier $\mu$ has an explicit solution. Therefore, the minimum trace over all possible scaling matrices can be obtained for each permutation matrix. The optimal permutation matrix is then derived by finding the global minimum trace over all possible permutations. Unlike the min-trace problem, our major challenge is the nonlinearity of the determinant function. More specifically, the Lagrange multiplier $\mu$ cannot be expressed explicitly in terms of $\lambda_i$ and $\gamma_i$, as given in (B.8) and (B.9). It is rather difficult to compare the minimum determinant across different permutations. To circumvent this problem, we propose a two-step proof. First, we show that the identity matrix is
the global optimal permutation matrix, given any scaling matrix. Second, we derive the optimal scaling matrix by solving the KKT conditions.

**APPENDIX C**

**PROOF FOR THEOREM 3**

Since $U_\omega$ and $Q_{xx}^{-1/2}$ are invertible, one can uniquely define a matrix $\Phi \in \mathbb{R}^{m \times n}$ such that

$$W = U_\omega \Phi Q_{xx}^{-1/2}. \quad \text{(C.1)}$$

The power constraint is equivalent to $\text{tr}(\Phi \Phi^T) \leq P$ since $\text{tr}(W Q_{xx} W^T) = \text{tr}(\Phi \Phi^T)$ for any pair of $(W, \Phi)$ satisfying (C.1).

We define an alternative optimization w.r.t. $\Phi$ as

$$\Phi^* = \arg \min_{\Phi} \varphi(Q_{ee})$$

subject to $\Phi \in \mathbb{R}^{m \times n}, \text{tr}(\Phi \Phi^T) \leq P. \quad \text{(C.2)}$

The Lagrangian is

$$L(\Phi; \mu) = \varphi(Q_{ee}) + \mu(\text{tr}(\Phi \Phi^T) - P),$$

and the necessary condition for any optimizer $\Phi$ is

$$\frac{\partial}{\partial \Phi} L(\Phi; \mu) = \frac{\partial \varphi(Q_{ee})}{\partial \Phi} + 2\mu \Phi = 0$$

Notice that the error covariance $Q_{ee}$ simplifies to

$$Q_{ee} = Q_{\theta \theta} - L \Phi^T (\Phi \Phi^T + \Sigma_\omega)^{-1} \Phi L^T, \quad \text{(C.3)}$$

where $L = Q_{\theta x} Q_{xx}^{-T/2}$. Using the matrix inversion lemma $(I + \Phi^T \Sigma_\omega^{-1} \Phi)^{-1} = I - \Phi^T (\Phi \Phi^T + \Sigma_\omega)^{-1} \Phi$, the error covariance can be rewritten as

$$Q_{ee} = Q_{\theta \theta|x} + L(I + \Phi^T \Sigma_\omega^{-1} \Phi)^{-1} L^T, \quad \text{(C.4)}$$

where $Q_{\theta \theta|x} = Q_{\theta \theta} - LL^T$ is a constant matrix with respect to $\Phi$.

Simple matrix calculation yields

$$\frac{\partial \varphi(Q_{ee})}{\partial \Phi} = -\Sigma_\omega^{-1} \Phi (I + \Phi^T \Sigma_\omega^{-1} \Phi)^{-1} L^T \times \left(\varphi'(Q_{ee})^T + \varphi'(Q_{ee})\right) L(I + \Phi^T \Sigma_\omega^{-1} \Phi)^{-1}$$

Left multiply $\frac{\partial}{\partial \Phi} L(\Phi; \mu)$ by $\Sigma_\omega$ and right multiply by $\Phi$. Then

$$2\mu \Sigma_\omega \Phi \Phi^T = \Phi (I + \Phi^T \Sigma_\omega^{-1} \Phi)^{-1} L^T \times \left(\varphi'(Q_{ee})^T + \varphi'(Q_{ee})\right) L(I + \Phi^T \Sigma_\omega^{-1} \Phi)^{-1} \Phi^T$$

Since the RHS is a symmetric matrix, $\Phi \Phi^T \Sigma_\omega = \Sigma_\omega \Phi \Phi^T$. When the diagonal matrix $\Sigma_\omega$ has distinct diagonal elements, it can be seen that the symmetric matrix $\Phi \Phi^T$ must be a diagonal matrix.
Given the SVD $\Phi = U \Sigma V^T$ where $U, V$ are orthogonal matrices, and $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal. Then by Lemma 6.8 of [14], there exists an $m \times m$ permutation matrix $\Pi_m$ such that $\Phi \Phi^T = \Pi_m \Sigma \Sigma^T \Pi_m^T$. Therefore, $\Phi = \Pi_m \Sigma V^T$. Plugging $\Phi$ in (C.1), it can be seen that $W^*$ can be factorized as $W^* = U_\omega \Pi_m^* \Sigma^* V^{*T} Q_{\pi \omega}^{-1/2}$.

**APPENDIX D**

**PROOF FOR LEMMA 2**

Suppose that $W^* = U_\omega \Pi_m^* \Sigma^* V^{*T} Q_{\pi \omega}^{-1/2}$ is a solution for problem (C.2). Define a function $f(t)$ for $t \in \mathbb{R}^1$ with

$$f(t) = \phi(Q_{\theta \theta}|_t + L(I + V^* e^{tX} \Delta^* e^{-tX} V^{*T})^{-1} L^T),$$

where $\Delta^* = \Sigma^T \Pi_m^T \Sigma_\omega \Pi_m^* \Sigma^*$, and $X \in \mathbb{R}^{n \times n}$ is an anti-symmetric matrix. The value of $f(t)$ is $\phi(Q_{ee})$ corresponding a compression $W = U_\omega \Pi_m^* \Sigma^* e^{-tX} V^{*T} Q_{\pi \omega}^{-1/2}$. Since $W^*$ is a solution for problem (C.2), we have $f(t) \geq f(0)$ for any $t \in \mathbb{R}$. Therefore, the necessary condition for $V^*$ is $\frac{\partial f}{\partial t}|_{t=0} = 0$ for any anti-symmetric matrix $X$. Let $X = [V^{*T} L^T \phi'(Q_{ee}) L V^*] (I_n + \Delta^*)^{-1}$, one can show that the necessary condition yields $\text{tr}(X X^T) = 0$. Therefore $X = 0$, i.e., $V^{*T} L^T \phi'(Q_{ee}) L V^*$ and $(I_n + \Delta^*)^{-1}$ commute.

**REFERENCES**


Yuan Wang received the B.S. (Statistics) degree from Special Class for Gifted Young, University of Science and Technology of China, Hefei, China, in 2008. She received her Ph.D degree in Statistics from Colorado State University in 2013.

She is currently a postdoctoral research fellow with University of Texas MD Anderson Cancer Center. Her research interest focuses on information fusion, dimension reduction, and object-oriented data analysis including tree-structure data, imaging data, and functional data.

Haonan Wang received his Ph.D. degree in statistics from the University of North Carolina at Chapel Hill in 2003. Currently, he is an Associate Professor of Statistics at Colorado State University, Fort Collins, CO. His research interests are in object oriented data analysis, statistical analysis on tree-structured objects, functional dynamic modeling of neuron activities and spatial statistics.

Louis L. Scharf (S’67-M’69-SM’77-F’86-LF’07) received the Ph.D. degree from the University of Washington, Seattle. From 1971 to 1982, he served as Professor of Electrical Engineering and Statistics with Colorado State University (CSU), Ft. Collins. From 1982 to 1985, he was Professor and Chairman of Electrical and Computer Engineering, University of Rhode Island, Kingston. From 1985 to 2000, he was Professor of Electrical and Computer Engineering, 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 here and abroad, including the Ecole Superieure d’Electricite, Gif-sur-Yvette, France; Ecole Nationale Superieure des Telecommunications, Paris, France; EURECOM, Nice, Italy; the University of La Plata, La Plata, Argentina; Duke University, Durham, NC; the University of Wisconsin, Madison; and the University of Troms, Troms, 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 radar/sonar and wireless communication channels.

Prof. Scharf was Technical Program Chair for the 1980 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Denver, CO; Tutorials Chair for ICASSP 2001, Salt Lake City, UT; and Technical Program Chair for the Asilomar Conference on Signals, Systems, and Computers 2002. He is past-Chair of the Fellow Committee for the IEEE Signal Processing Society and serves on the Technical Committee for Sensor Arrays and Multichannel Signal Processing. He has received numerous awards for his research contributions to statistical signal processing, including a College Research Award, an IEEE Distinguished Lectureship, an IEEE Third Millennium Medal, and the Technical Achievement and Society Awards from the IEEE Signal Processing Society.