

Minimax MSE–Ratio Estimation With Signal Covariance Uncertainties

Yonina C. Eldar, *Member, IEEE*, and Neri Merhav, *Fellow, IEEE*

Abstract—In continuation to an earlier work, we further consider the problem of robust estimation of a random vector (or signal), with an uncertain covariance matrix, that is observed through a known linear transformation and corrupted by additive noise with a known covariance matrix. While, in the earlier work, we developed and proposed a competitive minimax approach of minimizing the worst–case mean-squared error (MSE) *difference regret* criterion, here, we study, in the same spirit, the minimum worst–case MSE *ratio regret* criterion, namely, the worst–case *ratio* (rather than difference) between the MSE attainable using a linear estimator, ignorant of the exact signal covariance, and the minimum MSE (MMSE) attainable by optimum linear estimation with a known signal covariance. We present the optimal linear estimator, under this criterion, in two ways: The first is as a solution to a certain semidefinite programming (SDP) problem, and the second is as an expression that is of closed form up to a single parameter whose value can be found by a simple line search procedure. We then show that the linear minimax ratio regret estimator can also be interpreted as the MMSE estimator that minimizes the MSE for a certain choice of signal covariance that depends on the uncertainty region. We demonstrate that in applications, the proposed minimax MSE ratio regret approach may outperform the well-known minimax MSE approach, the minimax MSE difference regret approach, and the “plug-in” approach, where in the latter, one uses the MMSE estimator with an estimated covariance matrix replacing the true unknown covariance.

Index Terms—Covariance uncertainties, linear estimation, mean squared error, minimax regret, robust estimation, semidefinite programming.

I. INTRODUCTION

THE classical solutions to the problem of optimum linear estimation and filtering in stochastic models, dating back to Wiener [1] and Kolmogorov [2], are well known to be sensitive to the exact knowledge of the second-order statistics of the signal and noise. However, in a wide range of practical applications, these statistics may, unfortunately, be subjected to uncertainties. This fact has been the origin of the need for solid theories and methodologies of designing robust estimators and filters whose performance remains relatively insensitive and reasonably good across the region of uncertainty.

The most common approach to handle such uncertainties has been the minimax strategy, initiated by Huber [3], [4]. According to this method, an estimator is sought to minimize

the worst–case mean-squared error (MSE) over a given uncertainty class of spectral densities [5]–[9]. The idea of using minimax criteria for devising robust schemes that circumvent uncertainties has been applied in quite a few additional problem areas in communications, signal processing, and statistics (see, e.g., [8], [10], [11], and references therein). Recently, minimax ideas have also been applied to several different beamforming problems [12]–[15], [16].

In spite of the widespread use of the minimax approach, its performance sometimes turns out to be disappointingly poor. The reason seems to be rooted in the very definition of the minimax criterion, which is fundamentally pessimistic in nature: Optimizing performance for the worst case might come at the expense of deteriorated performance in all other cases, since in this worst case, the conditions may be so poor that they leave very little or no room for powerful solution strategies.

In light of this fact, to improve the performance of the ordinary minimax MSE approach, we have studied, in an earlier work [17], a modified minimax criterion, which is *competitive* in character. This competitive minimax criterion has been derived in [17] in the context of a simple linear model. Specifically, given an observation vector, resulting from a known linear transformation of the desired random vector (or signal) to be estimated and corrupted by an uncorrelated additive noise vector with a known covariance matrix, we sought in [17] a linear estimator, which is robust to covariance uncertainties of the desired signal, using the following approach: Rather than minimizing the worst–case (total) MSE, we derived the linear estimator that minimizes the worst–case *difference regret*, namely, the worst–case difference between the MSE of a linear estimator, ignorant of the exact signal covariance, and the MSE of the linear optimal estimator based on the exact signal covariance. The rationale was that such an estimator performs uniformly as closely as possible to the linear optimal estimator across the uncertainty region, and since the minimax criterion is applied to the difference of MSEs, rather than the total MSE, it is not as pessimistic as the ordinary minimax approach. The same idea was also applied in [18] for the case where the unknown desired vector is *deterministic* rather than stochastic. As we have pointed out in [17], the competitive minimax approach is by no means new as a general concept. It has been used extensively in a variety of other problem areas, such as, universal source coding [19], hypothesis testing [20], [21], and prediction (see [22] for a survey and references therein).

A possible drawback of the difference regret method is that the value of the regret may not adequately reflect the estimator performance, since even a large regret should be considered insignificant if the value of the optimal MSE is relatively large. On

Manuscript received October 31, 2003; revised April 23, 2004. The work of Y. C. Eldar was supported in part by the Taub Foundation and by a Technion V.P.R. Fund. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Fulvio Gini.

The authors are with the Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: yonina@ee.technion.ac.il; merhav@ee.technion.ac.il).

Digital Object Identifier 10.1109/TSP.2005.843701

the other hand, if the optimal MSE is small, then even a small regret should be considered significant. Therefore, instead of considering the worst-case difference regret, we suggest developing a minimax ratio estimator that minimizes the worst-case *ratio* between the MSE of a linear estimator that does not know the exact signal covariance but knows only that it lies in a specified uncertainty set and the best possible MSE as a function of the unknown covariance. Generally speaking, the rationale is that, as before, such an estimator performs uniformly as close as possible to the linear optimal estimator across the uncertainty region, but where now, the MSE is measured in decibels. This makes sense as one might expect the relative loss in MSE performance to be scale-invariant.

In this paper, we study robust linear estimation for the same linear model under the criterion of minimax *ratio-regret*, rather than the difference-regret. We present the optimal linear estimator, under this new criterion, in two ways: The first representation is as a solution to a certain semidefinite programming (SDP) problem. This is practically meaningful since SDP programs are efficiently executable using standard software packages. The second representation is as an analytical expression, which is of closed form up to a single parameter whose value can be found by a simple line search procedure. We also show that the linear minimax ratio regret estimator (or, for short, the minimax ratio estimator) can be interpreted as the minimum MSE (MMSE) estimator corresponding to a certain choice of signal covariance that depends on the uncertainty region. We demonstrate that in applications, the proposed minimax MSE ratio-regret approach may outperform the ordinary MSE approach, the minimax MSE difference-regret approach, and the “plug-in” approach, where in the latter, one uses the MMSE estimator with an estimated covariance matrix replacing the true unknown covariance.

In our development, we assume explicitly that the linear model matrix is known exactly. In [17], we consider a minimax MSE estimator for the case in which the model matrix is also subjected to uncertainty. However, developing the minimax regret estimator for this case appears to be a much more difficult problem. We note that the problem of estimating a random vector comprised of independent random variables that is observed through an unknown linear transformation is known as the independent component (or factor) analysis problem, or the blind source separation problem, and has received considerable attention in the literature (see, e.g., [23]–[25]). Blind source separation methods attempt to extract the input vector from the observations by exploiting the fact that the inputs are independent and are therefore typically nonlinear methods that involve higher order statistics.

The outline of this paper is as follows. In Section II, we formulate the problem. In Section III, we present the problem as an SDP. In Section IV, the alternative, closed-form solution is derived, and finally, in Section V, performance is demonstrated through several examples.

II. PROBLEM FORMULATION

In the sequel, we denote vectors in \mathbb{C}^m by boldface lowercase letters and matrices in $\mathbb{C}^{n \times m}$ by boldface uppercase letters. The

matrix \mathbf{I} denotes the identity matrix of the appropriate dimension, $(\cdot)^*$ denotes the Hermitian conjugate of the corresponding matrix, and $(\hat{\cdot})$ denotes an estimated vector. The cross-covariance matrix between the random vectors \mathbf{x} and \mathbf{y} is denoted by \mathbf{C}_{xy} , and the covariance matrix of \mathbf{x} is denoted by \mathbf{C}_x .

Consider the problem of estimating the unknown random parameters \mathbf{x} in the linear model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \quad (1)$$

where \mathbf{H} is a known $n \times m$ matrix with rank m , \mathbf{x} is a zero-mean, length- m random vector with covariance matrix \mathbf{C}_x , and \mathbf{w} is a zero-mean, length- n random vector with known positive definite covariance \mathbf{C}_w , uncorrelated with \mathbf{x} . We assume that we only have partial information about the covariance \mathbf{C}_x .

We seek to estimate \mathbf{x} using a linear estimator so that $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ for some $m \times n$ matrix \mathbf{G} . We would like to design an estimator $\hat{\mathbf{x}}$ of \mathbf{x} to minimize the MSE, which is given by

$$\begin{aligned} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) &= \text{Tr}(\mathbf{C}_x) + \text{Tr}(\mathbf{C}_{\hat{\mathbf{x}}}) - 2\text{Tr}(\mathbf{C}_{x\hat{\mathbf{x}}}) \\ &= \text{Tr}(\mathbf{C}_x) + \text{Tr}(\mathbf{G}(\mathbf{H}\mathbf{C}_x\mathbf{H}^* + \mathbf{C}_w)\mathbf{G}^*) - 2\text{Tr}(\mathbf{C}_x\mathbf{H}^*\mathbf{G}^*) \\ &= \text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \text{Tr}(\mathbf{C}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})). \end{aligned} \quad (2)$$

If \mathbf{C}_x is known, then the linear estimator minimizing (2) is the *MMSE estimator* [26]

$$\hat{\mathbf{x}} = \mathbf{C}_x\mathbf{H}^*(\mathbf{H}\mathbf{C}_x\mathbf{H}^* + \mathbf{C}_w)^{-1}\mathbf{y}. \quad (3)$$

An alternative form for $\hat{\mathbf{x}}$, that is sometimes more convenient, can be obtained by applying the matrix inversion lemma [27] to $(\mathbf{H}\mathbf{C}_x\mathbf{H}^* + \mathbf{C}_w)^{-1}$, resulting in

$$\begin{aligned} (\mathbf{H}\mathbf{C}_x\mathbf{H}^* + \mathbf{C}_w)^{-1} &= \mathbf{C}_w^{-1} - \mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}. \end{aligned} \quad (4)$$

Substituting (4) into (3), the MMSE estimator $\hat{\mathbf{x}}$ can be expressed as

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{C}_x(\mathbf{I} - \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1})\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y} \\ &= \mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}. \end{aligned} \quad (5)$$

If \mathbf{C}_x is unknown, then we cannot implement the MMSE estimator (3). In this case, we may choose an estimator to optimize a worst-case performance measure over all covariance matrices in the region of uncertainty. To reflect the uncertainty in our knowledge of the true covariance matrix, we consider an uncertainty model that resembles the “band model” widely used in the continuous-time case [6], [8], [28], [29] and is the same as the model considered in [17]. Specifically, we assume that \mathbf{C}_x and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have the same eigenvector matrix¹ and that each of the non-negative eigenvalues $\delta_i \geq 0$, $1 \leq i \leq m$ of \mathbf{C}_x satisfies

$$l_i \leq \delta_i \leq u_i, \quad 1 \leq i \leq m \quad (6)$$

where l_i and u_i are known.

¹If the eigenvalues of $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ and \mathbf{C}_x have geometric multiplicity 1, then $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ and \mathbf{C}_x have the same eigenvector matrix if and only if they commute [27].

The assumption that \mathbf{C}_x and $\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H}$ have the same eigenvector matrix is made for analytical tractability. In practice, there are many cases where this assumption is satisfied exactly or approximately, for example, if \mathbf{C}_x is proportional to the identity matrix, or if \mathbf{H} and \mathbf{C}_w are proportional to the identity matrix. Another case is when \mathbf{C}_x and $\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H}$ are circulant matrices, in which case, both matrices are diagonalized by a Fourier transform matrix. The circulant model is used in a variety of different signal processing applications, such as image restoration [30] and cyclic convolution filterbanks [31]. If \mathbf{x} is a stationary random vector and \mathbf{H} represents convolution of \mathbf{x} with some filter, then both \mathbf{C}_x and \mathbf{H} will be Toeplitz matrices and are, therefore approximately diagonalized by a Fourier transform matrix so that in this general case, \mathbf{C}_x and $\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H}$ approximately have the same eigenvector matrix [32]. We note that, although in our derivations we assume that \mathbf{C}_x and $\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H}$ have the same eigenvector matrix, our proposed estimator can still be applied if this assumption is violated. Specifically, as we show in Section IV the ratio regret estimator can be viewed as an MMSE estimator matched to a certain choice of signal covariance, which can therefore be implemented, regardless of whether or not the eigenvector matrices are the same. This point is illustrated in the examples in Section V.

In our development, we explicitly assume that the joint eigenvector matrix of \mathbf{C}_x and $\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H}$ is given. In practice, if the eigenvalues of $\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H}$ have geometric multiplicity one, then we choose the eigenvector matrix of \mathbf{C}_x to be equal to the eigenvector matrix of $\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H}$. In the case in which the eigenvector matrix of $\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H}$ is not uniquely specified, e.g., in the case in which $\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H}$ is proportional to \mathbf{I} , as in one of the examples in Section V, we may resolve this ambiguity by estimating the eigenvector matrix of \mathbf{C}_x from the data. Specifically, given an observation vector \mathbf{y} and assuming that the noise covariance is equal to $\mathbf{C}_w = \mathbf{I}$, we may estimate \mathbf{C}_x as $\mathbf{C}_x = (\mathbf{H}^* \mathbf{H})^{-1} \mathbf{H}^* [\hat{\mathbf{C}}_y - \mathbf{C}_w]_+ \mathbf{H} (\mathbf{H}^* \mathbf{H})^{-1}$. Here, $\hat{\mathbf{C}}_y$ is an estimate of \mathbf{C}_y (see Section V), and $[\mathbf{A}]_+$ denotes the matrix in which the negative eigenvalues of \mathbf{A} are replaced by 0. This estimate can be regarded as the analog for finite-length processes of the spectrum estimate based on the spectral subtraction method for infinite-length processes [33], [34]. For general \mathbf{C}_w , a similar estimate can be obtained by first whitening the observation vector.

The model (6) is reasonable when the covariance is estimated from the data. Specifically, denoting by $\zeta_i = (u_i + l_i)/2$, $\epsilon_i = (u_i - l_i)/2$ for $1 \leq i \leq m$, (6) can equivalently be expressed as

$$\delta_i = \zeta_i + e_i, \quad e_i^2 \leq \epsilon_i^2, \quad 1 \leq i \leq m \quad (7)$$

so that each of the eigenvalues of \mathbf{C}_x lies in an interval of length $2\epsilon_i$ around some nominal value ζ_i , which we can think of as an estimate of the i th eigenvalue of \mathbf{C}_x from the data vector \mathbf{y} . The interval specified by ϵ_i may be regarded as a confidence interval around our estimate ζ_i and can be chosen to be proportional to the standard deviation of the estimate ζ_i .

Given $\{\zeta_i\}$, a straightforward approach to estimating \mathbf{x} is to use an MMSE estimate corresponding to the estimated covariance. However, as demonstrated in [17] and in Section V, by taking an uncertainty interval around ζ_i into account, and

seeking a competitive minimax estimator in this interval, we can further improve the estimation performance.

To develop a competitive estimator, we consider a minimax ratio criterion, in which the estimator is obtained by minimizing the worst-case *ratio* between the MSE of a linear estimator that does not know the exact signal covariance but only that it lies in a predefined uncertainty region and the best possible MSE as function of the unknown covariance. In Section III, we show that the minimax ratio estimator can be formulated as an SDP. In Section IV, we use the necessary and sufficient conditions for optimality of an SDP to develop more insight into the minimax ratio estimator. Specifically, we show that the minimax ratio estimator is an MMSE estimator matched to a covariance matrix that depends on a single parameter that can be found using a simple line search algorithm, for example, using the bisection method. In the examples in Section V, we demonstrate that the minimax ratio estimator can improve the performance over the minimax MSE estimator and the minimax regret estimator of [17] for low SNR values.

III. MINIMAX RATIO ESTIMATOR

We seek the linear estimator $\hat{\mathbf{x}}$ that minimizes the worst-case ratio $\mathcal{R}(\mathbf{C}_x, \mathbf{G})$, which is defined as the ratio between the MSE using an estimator $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$, and the smallest possible MSE attainable with an estimator of the form $\hat{\mathbf{x}} = \mathbf{G}(\mathbf{C}_x)\mathbf{y}$, assuming that the true covariance is \mathbf{C}_x , which we denote by $\text{MSE}^o(\mathbf{C}_x)$.

If the true covariance matrix is equal to \mathbf{C}_x , then the MMSE estimator is given by (3), and the resulting optimal MSE is

$$\text{MSE}^o(\mathbf{C}_x) = \text{Tr}(\mathbf{C}_x) - \text{Tr}(\mathbf{C}_x \mathbf{H}^* (\mathbf{H} \mathbf{C}_x \mathbf{H}^* + \mathbf{C}_w)^{-1} \mathbf{H} \mathbf{C}_x). \quad (8)$$

From (4) and (5), we have that $\mathbf{C}_x \mathbf{H}^* (\mathbf{H} \mathbf{C}_x \mathbf{H}^* + \mathbf{C}_w)^{-1} = \mathbf{C}_x (\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H} \mathbf{C}_x + \mathbf{I})^{-1} \mathbf{H}^* \mathbf{C}_w^{-1}$ so that (8) can be written in the equivalent form

$$\begin{aligned} \text{MSE}^o(\mathbf{C}_x) &= \text{Tr}(\mathbf{C}_x (\mathbf{I} - (\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H} \mathbf{C}_x + \mathbf{I})^{-1} \mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H} \mathbf{C}_x)) \\ &= \text{Tr}(\mathbf{C}_x (\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H} \mathbf{C}_x + \mathbf{I})^{-1}) \end{aligned} \quad (9)$$

which will be more convenient for our derivations. Throughout the paper, we assume that $l_i > 0$ for at least one value of i , so that $\text{MSE}^o(\mathbf{C}_x) \neq 0$ for all \mathbf{C}_x in the region of uncertainty.

When the true covariance matrix is unknown, the best possible MSE cannot be attained. For every possible choice of the covariance \mathbf{C}_x in the region of uncertainty, we can define the ratio between the MSE attainable with an arbitrary linear estimator and the best possible MSE attainable $\text{MSE}^o(\mathbf{C}_x)$, assuming that \mathbf{C}_x is the true covariance matrix. We then seek to minimize this worst-case ratio in the region of uncertainty.

Thus, we seek the matrix \mathbf{G} that is the solution to the problem

$$\min_{\mathbf{G}} \max_{l_i \leq \delta_i \leq u_i} \mathcal{R}(\mathbf{C}_x, \mathbf{G}) \quad (10)$$

where

$$\begin{aligned} \mathcal{R}(\mathbf{C}_x, \mathbf{G}) &= \frac{E(\|\mathbf{G}\mathbf{y} - \mathbf{x}\|^2)}{\text{MSE}^o(\mathbf{C}_x)} = \\ &= \frac{\text{Tr}(\mathbf{G} \mathbf{C}_w \mathbf{G}^*) + \text{Tr}(\mathbf{C}_x (\mathbf{I} - \mathbf{G} \mathbf{H})^* (\mathbf{I} - \mathbf{G} \mathbf{H}))}{\text{Tr}(\mathbf{C}_x (\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H} \mathbf{C}_x + \mathbf{I})^{-1})} \end{aligned} \quad (11)$$

which can alternatively be expressed as

$$\text{subject to} \quad \min_{\gamma, \mathbf{G}} \gamma \quad (12)$$

$$\max_{l_i \leq \delta_i \leq u_i} \left\{ \frac{\text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \text{Tr}(\mathbf{C}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H}))}{\text{Tr}(\mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1})} \right\} \leq \gamma. \quad (13)$$

The constraint (13) is equivalent to

$$\begin{aligned} & \text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \text{Tr}(\mathbf{C}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})) \\ & - \gamma \text{Tr}(\mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}) \leq 0, \quad \forall l_i \leq \delta_i \leq u_i \end{aligned} \quad (14)$$

or

$$\max_{l_i \leq \delta_i \leq u_i} \{ \text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \text{Tr}(\mathbf{C}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})) - \gamma \text{Tr}(\mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}) \} \leq 0. \quad (15)$$

Thus, the problem of (10) is equivalent to (12) subject to (15).

We now show that the problem of (12) and (15) can be formulated as a convex *semidefinite programming (SDP) problem* [35]–[37], which is the problem of minimizing a linear functional subject to linear matrix inequalities (LMIs), i.e., matrix inequalities in which the matrices depend *linearly* on the unknowns. (Note that even though the matrices are linear in the unknowns, the inequalities are nonlinear since a positive semidefinite constraint on a matrix reduces to nonlinear constraints on the matrix elements.) The main advantage of the SDP formulation is that it readily lends itself to efficient computational methods. Specifically, by exploiting the many well-known algorithms for solving SDPs [36], [35], e.g., interior point methods, [37], [38] which are guaranteed to converge to the global optimum, the optimal estimator can be computed very efficiently in polynomial time. Using principles of duality theory in vector space optimization, the SDP formulation can also be used to derive optimality conditions.

After a description of the general SDP problem in Section III-A, in Section III.B, we show that our minimax problem can be formulated as an SDP. In Section IV, we use the SDP formulation to develop more insight into the minimax ratio estimator.

A. Semidefinite Programming

A standard SDP is the problem of minimizing

$$\text{subject to} \quad P(\mathbf{x}) = \mathbf{c}^*\mathbf{x} \quad (16)$$

$$\text{where} \quad \mathbf{F}(\mathbf{x}) \geq 0, \quad (17)$$

$$\mathbf{F}(\mathbf{x}) = \mathbf{F}_0 + \sum_{i=1}^m x_i \mathbf{F}_i. \quad (18)$$

Here, $\mathbf{x} \in \mathcal{R}^m$ is the vector to be optimized, x_i denotes the i th component of \mathbf{x} , \mathbf{c} is a given vector in \mathcal{R}^m , and \mathbf{F}_i are given matrices in the space \mathcal{B}_n of $n \times n$ Hermitian matrices.²

²Although typically in the literature, the matrices \mathbf{F}_i are restricted to be real and symmetric, the SDP formulation can be easily extended to include Hermitian matrices \mathbf{F}_i ; see, e.g., [39]. In addition, many of the standard software packages for efficiently solving SDPs, for example, the Self-Dual-Minimization (SeDuMi) package [40], [41], allow for Hermitian matrices.

The constraint (17) is an LMI, in which the unknowns x_i appear linearly. Indeed, any constraint of the form $\mathbf{A}(\mathbf{x}) \geq 0$ where the matrix \mathbf{A} depends linearly on \mathbf{x} can be put in the form of (17).

The problem of (16) and (17) is referred to as the *primal problem*. A vector \mathbf{x} is said to be *primal feasible* if $\mathbf{F}(\mathbf{x}) \geq 0$ and is *strictly primal feasible* if $\mathbf{F}(\mathbf{x}) > 0$. If there exists a strictly feasible point, then the primal problem is said to be strictly feasible. We denote the optimal value of $P(\mathbf{x})$ by \hat{P} .

An SDP is a convex optimization problem and can be solved very efficiently. Furthermore, iterative algorithms for solving SDPs are guaranteed to converge to the global minimum. The SDP formulation can also be used to derive necessary and sufficient conditions for optimality by exploiting principles of duality theory. The essential idea is to formulate a *dual problem* of the form $\max_{\mathbf{Z}} D(\mathbf{Z})$ for some linear functional D whose maximal value \hat{D} serves as a certificate for \hat{P} . That is, for all feasible values of $\mathbf{Z} \in \mathcal{B}_n$, i.e., values of $\mathbf{Z} \in \mathcal{B}_n$ that satisfy a certain set of constraints, and for all feasible values of \mathbf{x} , $D(\mathbf{Z}) \leq P(\mathbf{x})$, so that the dual problem provides a lower bound on the optimal value of the original (primal) problem. If, in addition, we can establish that $\hat{P} = \hat{D}$, then this equality can be used to develop conditions of optimality on \mathbf{x} .

The dual problem associated with the SDP of (16) and (17) is the problem of maximizing

$$D(\mathbf{Z}) = -\text{Tr}(\mathbf{F}_0\mathbf{Z}) \quad (19)$$

subject to

$$\text{Tr}(\mathbf{F}_i\mathbf{Z}) = c_i, \quad 1 \leq i \leq m \quad (20)$$

$$\mathbf{Z} \geq 0 \quad (21)$$

where $\mathbf{Z} \in \mathcal{B}_n$. A matrix $\mathbf{Z} \in \mathcal{B}_n$ is said to be *dual feasible* if it satisfies (20) and (21) and is *strictly dual feasible* if it satisfies (20) and $\mathbf{Z} > 0$. If there exists a strictly feasible point, then the dual problem is said to be strictly feasible.

For any feasible \mathbf{x} and \mathbf{Z} , we have that

$$\begin{aligned} P(\mathbf{x}) - D(\mathbf{Z}) &= \mathbf{c}^*\mathbf{x} + \text{Tr}(\mathbf{F}_0\mathbf{Z}) \\ &= \sum_{i=1}^m x_i \text{Tr}(\mathbf{F}_i\mathbf{Z}) + \text{Tr}(\mathbf{F}_0\mathbf{Z}) = \text{Tr}(\mathbf{F}(\mathbf{x})\mathbf{Z}) \geq 0 \end{aligned} \quad (22)$$

so that as required, $D(\mathbf{Z}) \leq P(\mathbf{x})$. Furthermore, it can be shown that if either the primal or the dual problem are strictly feasible, then $\hat{P} = \hat{D}$, and \mathbf{x} is an optimal primal point if and only if \mathbf{x} is primal feasible, and there exists a dual feasible $\mathbf{Z} \in \mathcal{B}_n$ such that

$$\mathbf{Z}\mathbf{F}(\mathbf{x}) = 0. \quad (23)$$

Equation (23), together with (17), (20), and (21), constitute a set of necessary and sufficient conditions for \mathbf{x} to be an optimal solution to the problem of (16) and (17) when either the primal or the dual are strictly feasible.

B. Semidefinite Programming Formulation of the Estimation Problem

In Theorem 1 below, we show that the problem of (12) and (15) can be formulated as an SDP. The remainder of this section is devoted to the proof of the theorem.

Theorem 1: Let \mathbf{x} denote the unknown parameters in the model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where \mathbf{H} is a known $n \times m$ matrix with rank m , \mathbf{x} is a zero-mean random vector uncorrelated with \mathbf{w} with covariance \mathbf{C}_x , and \mathbf{w} is a zero-mean random vector with covariance \mathbf{C}_w . Let $\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H} = \mathbf{V} \Lambda \mathbf{V}^*$, where \mathbf{V} is a unitary matrix and Λ is an $m \times m$ diagonal matrix with diagonal elements $\lambda_i > 0$, and let $\mathbf{C}_x = \mathbf{V} \Delta \mathbf{V}^*$, where Δ is an $m \times m$ diagonal matrix with diagonal elements $0 \leq l_i \leq \delta_i \leq u_i$. Then, the solution to the problem

$$\min_{\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}} \max_{l_i \leq \delta_i \leq u_i} \left\{ \frac{E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)}{\min_{\hat{\mathbf{x}} = \mathbf{G}(\mathbf{x})\mathbf{y}} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)} \right\}$$

has the form

$$\hat{\mathbf{x}} = \mathbf{V} \mathbf{D} \Lambda^{-1} \mathbf{V}^* \mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{y}$$

where \mathbf{D} is an $m \times m$ diagonal matrix with diagonal elements d_i that are the solution to the semidefinite programming problem

$$\min_{t_i, d_i, \alpha_i, \gamma} \gamma \quad (24)$$

subject to (25), shown at the bottom of the page, where $\zeta_i = (u_i + l_i)/2$, and $\epsilon_i = (u_i - l_i)/2$.

Proof: We prove Theorem 1 in three stages. First, we show that the optimal \mathbf{G} has the form

$$\mathbf{G} = \mathbf{V} \mathbf{D} \Lambda^{-1} \mathbf{V}^* \mathbf{H}^* \mathbf{C}_w^{-1} \quad (26)$$

for some $m \times m$ matrix \mathbf{D} . We then show that \mathbf{D} must be a diagonal matrix. Finally, we develop an expression for the diagonal elements of \mathbf{D} . The first two parts of the proof are similar to the proof of [17, Th. 3].

We begin by showing that the optimal \mathbf{G} has the form given by (26). To this end, note that the ratio $\mathcal{R}(\mathbf{C}_x, \mathbf{G})$ of (11) depends on \mathbf{G} only through $\mathbf{G}\mathbf{H}$ and $\text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*)$. Now, for any choice of \mathbf{G}

$$\begin{aligned} & \text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) \\ &= \text{Tr}(\mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{1/2}\mathbf{G}^*) + \text{Tr}(\mathbf{G}\mathbf{C}_w^{1/2}(\mathbf{I} - \mathbf{P})\mathbf{C}_w^{1/2}\mathbf{G}^*) \\ &\geq \text{Tr}(\mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{1/2}\mathbf{G}^*) \end{aligned} \quad (27)$$

where

$$\mathbf{P} = \mathbf{C}_w^{-1/2} \mathbf{H} (\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^* \mathbf{C}_w^{-1/2} \quad (28)$$

is the orthogonal projection onto the range space of $\mathbf{C}_w^{-1/2} \mathbf{H}$. In addition, $\mathbf{G}\mathbf{H} = \mathbf{G}\mathbf{C}_w^{1/2} \mathbf{P} \mathbf{C}_w^{-1/2} \mathbf{H}$ since $\mathbf{P}\mathbf{C}_w^{-1/2} \mathbf{H} = \mathbf{C}_w^{-1/2} \mathbf{H}$. It follows then that any choice of \mathbf{G} can be replaced by $\tilde{\mathbf{G}} = \mathbf{G}\mathbf{C}_w^{1/2} \mathbf{P} \mathbf{C}_w^{-1/2}$, with

$\text{Tr}(\tilde{\mathbf{G}}\mathbf{C}_w\tilde{\mathbf{G}}^*) \leq \text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*)$ and $\tilde{\mathbf{G}}\mathbf{H} = \mathbf{G}\mathbf{H}$, implying that $\tilde{\mathbf{G}}$ is always at least as good as \mathbf{G} in the sense of reducing the ratio regret. Since $\tilde{\mathbf{G}} = \tilde{\mathbf{G}}\mathbf{C}_w^{1/2} \mathbf{P} \mathbf{C}_w^{-1/2}$, we conclude that when seeking the optimal matrix \mathbf{G} , it is sufficient to confine attention to matrices that satisfy

$$\mathbf{G}\mathbf{C}_w^{1/2} = \mathbf{G}\mathbf{C}_w^{1/2} \mathbf{P}. \quad (29)$$

Substituting (28) into (29), we have

$$\begin{aligned} \mathbf{G} &= \mathbf{G}\mathbf{C}_w^{1/2} \mathbf{P} \mathbf{C}_w^{-1/2} = \mathbf{G}\mathbf{H} (\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^* \mathbf{C}_w^{-1} \\ &= \mathbf{B} (\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^* \mathbf{C}_w^{-1} \end{aligned} \quad (30)$$

for some $m \times m$ matrix \mathbf{B} . Denoting $\mathbf{B} = \mathbf{V} \mathbf{D} \mathbf{V}^*$ and using the fact that $\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H} = \mathbf{V} \Lambda \mathbf{V}^*$, (30) reduces to (26).

We now show that \mathbf{D} must be a diagonal matrix. Since $\mathbf{C}_x = \mathbf{V} \Delta \mathbf{V}^*$, $\mathbf{H} \mathbf{C}_w^{-1} \mathbf{H} = \mathbf{V} \Lambda \mathbf{V}^*$, and \mathbf{G} is given by (30), we have that

$$\begin{aligned} & \text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \text{Tr}(\mathbf{C}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})) \\ & \quad - \gamma \text{Tr}(\mathbf{C}_x(\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H} \mathbf{C}_x + \mathbf{I})^{-1}) \\ &= \text{Tr}(\mathbf{D}^* \mathbf{D} \Lambda^{-1}) + \text{Tr}(\Delta(\mathbf{I} - \mathbf{D})^*(\mathbf{I} - \mathbf{D})) \\ & \quad - \gamma \text{Tr}(\Delta(\Lambda \Delta + \mathbf{I})^{-1}) \triangleq \mathcal{L}(\mathbf{D}, \Delta). \end{aligned} \quad (31)$$

Therefore, the problem of (12) and (15) reduces to finding \mathbf{D} that minimizes γ subject to

$$\mathcal{G}(\mathbf{D}) = \max_{l_i \leq \delta_i \leq u_i} \mathcal{L}(\mathbf{D}, \Delta) \leq 0. \quad (32)$$

Clearly, $\mathcal{L}(\mathbf{D})$ is strictly convex in \mathbf{D} . Therefore, for any $0 < \alpha < 1$

$$\begin{aligned} & \mathcal{G}(\alpha \mathbf{D}_1 + (1 - \alpha) \mathbf{D}_2) \\ &= \max_{l_i \leq \delta_i \leq u_i} \mathcal{L}(\alpha \mathbf{D}_1 + (1 - \alpha) \mathbf{D}_2, \Delta) \\ &< \max_{l_i \leq \delta_i \leq u_i} \{ \alpha \mathcal{L}(\mathbf{D}_1, \Delta) + (1 - \alpha) \mathcal{L}(\mathbf{D}_2, \Delta) \} \\ &\leq \alpha \max_{l_i \leq \delta_i \leq u_i} \mathcal{L}(\mathbf{D}_1, \Delta) + (1 - \alpha) \max_{l_i \leq \delta_i \leq u_i} \mathcal{L}(\mathbf{D}_2, \Delta) \\ &= \alpha \mathcal{G}(\mathbf{D}_1) + (1 - \alpha) \mathcal{G}(\mathbf{D}_2) \end{aligned} \quad (33)$$

so that $\mathcal{G}(\mathbf{D})$ is also strictly convex in \mathbf{D} , and consequently, our problem has a unique global minimum. Let \mathbf{J} be any diagonal matrix with diagonal elements equal to ± 1 . Then, using the relations $\mathbf{J} = \mathbf{J}^*$, $\mathbf{J}^2 = \mathbf{I}$ and the fact that for any diagonal matrix \mathbf{M} , $\mathbf{J} \mathbf{M} \mathbf{J} = \mathbf{M}$, we can show that $\mathcal{L}(\mathbf{D}, \Delta) = \mathcal{L}(\mathbf{J} \mathbf{D} \mathbf{J}, \Delta)$ (see also [17]). Since $\mathcal{G}(\mathbf{D})$ has a unique minimizer, we conclude that the matrix \mathbf{D} that minimizes $\mathcal{G}(\mathbf{D})$ satisfies $\mathbf{D} = \mathbf{J} \mathbf{D} \mathbf{J}$ for any diagonal matrix \mathbf{J} with diagonal elements equal to ± 1 , which in turn implies that \mathbf{D} must be a diagonal matrix.

$$\left[\begin{array}{ccc} \alpha_i & \frac{t_i \lambda_i^2}{2} + \frac{\lambda_i(\gamma-1)}{2} & \lambda_i(d_i - 1) \\ \frac{t_i \lambda_i^2}{2} + \frac{\lambda_i(\gamma-1)}{2} & t_i \lambda_i (\lambda_i \zeta_i + 1) + \lambda_i \zeta_i (\gamma - 1) - \alpha_i \epsilon_i^2 & d_i (1 + \lambda_i \zeta_i) - \lambda_i \zeta_i \\ \lambda_i (d_i - 1) & d_i (1 + \lambda_i \zeta_i) - \lambda_i \zeta_i & 1 \end{array} \right] \geq 0, \quad 1 \leq i \leq m$$

$$\sum_{i=1}^m t_i \leq 0 \quad (25)$$

Denote by d_i , λ_i , and δ_i the diagonal elements of \mathbf{D} , Λ , and Δ , respectively. Then, we can express $\mathcal{G}(\mathbf{D})$ as

$$\begin{aligned} \mathcal{G}(\mathbf{D}) &= \max_{l_i \leq \delta_i \leq u_i} \left\{ \sum_{i=1}^m \left(\frac{d_i^2}{\lambda_i} + \delta_i(1-d_i)^2 - \gamma \frac{\delta_i}{\lambda_i \delta_i + 1} \right) \right\} \\ &= \sum_{i=1}^m \max_{l_i \leq \delta_i \leq u_i} \left\{ \frac{(\lambda_i(d_i-1)\delta_i + d_i)^2 + (1-\gamma)\lambda_i\delta_i}{\lambda_i(\lambda_i\delta_i + 1)} \right\}. \end{aligned} \quad (34)$$

Our problem can now be formulated as

$$\min_{\gamma, t_i, d_i} \gamma \quad (35)$$

subject to

$$\max_{l_i \leq \delta_i \leq u_i} \left\{ \frac{(\lambda_i(d_i-1)\delta_i + d_i)^2 + (1-\gamma)\lambda_i\delta_i}{\lambda_i(\lambda_i\delta_i + 1)} \right\} \leq t_i, \quad 1 \leq i \leq m \quad (36)$$

$$\sum_{i=1}^m t_i \leq 0. \quad (37)$$

Expressing (36) as

$$\frac{(\lambda_i(d_i-1)\delta_i + d_i)^2 + (1-\gamma)\lambda_i\delta_i}{\lambda_i(\lambda_i\delta_i + 1)} \leq t_i \quad \forall \delta_i : l_i \leq \delta_i \leq u_i, \quad 1 \leq i \leq m \quad (38)$$

we develop a solution to our problem by first considering each of the constraints (38), where for brevity, we omit the index i .

Let $\delta = \zeta + e$, where $\zeta = (u+l)/2$. Then, the condition $l \leq \delta \leq u$ is equivalent to the condition $e^2 \leq \epsilon^2$, where $\epsilon = (u-l)/2$, so that (38) can be written as

$$(\lambda(d-1)(\zeta+e) + d)^2 + (1-\gamma)\lambda\delta \leq t\lambda(\lambda(\zeta+e) + 1) \quad \forall e : e^2 \leq \epsilon^2 \quad (39)$$

which in turn is equivalent to the following implication:

$$P(e) \triangleq \epsilon^2 - e^2 \geq 0 \Rightarrow Q(e) \geq 0 \quad (40)$$

where

$$\begin{aligned} Q(e) &= t\lambda(\lambda(\zeta+e) + 1) - (\lambda(d-1)(\zeta+e) + d)^2 - (1-\gamma)\lambda\delta \\ &= -e^2\lambda^2(d-1)^2 \\ &\quad + 2e \left(\frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} + \lambda(1-d)(d(\lambda\zeta+1) - \lambda\zeta) \right) \\ &\quad + t\lambda(\lambda\zeta+1) - (d(\lambda\zeta+1) - \lambda\zeta)^2 + \lambda\zeta(\gamma-1). \end{aligned} \quad (41)$$

We now rely on the following lemma [42, p. 23].

Lemma 1—S-procedure: Let $P(\mathbf{z}) = \mathbf{z}^* \mathbf{A} \mathbf{z} + 2\mathbf{u}^* \mathbf{z} + v$ and $Q(\mathbf{z}) = \mathbf{z}^* \mathbf{B} \mathbf{z} + 2\mathbf{x}^* \mathbf{z} + y$ be two quadratic functions of \mathbf{z} , where \mathbf{A} and \mathbf{B} are symmetric, and there exists a \mathbf{z}_0 satisfying $P(\mathbf{z}_0) > 0$. Then, the implication

$$P(\mathbf{z}) \geq 0 \Rightarrow Q(\mathbf{z}) \geq 0$$

holds true if and only if there exists an $\alpha \geq 0$ such that

$$\begin{bmatrix} \mathbf{B} - \alpha \mathbf{A} & \mathbf{x} - \alpha \mathbf{u} \\ \mathbf{x}^* - \alpha \mathbf{u}^* & y - \alpha v \end{bmatrix} \geq 0.$$

Combining (40) with Lemma 1, it follows immediately that (38) is equivalent to (42), shown at the bottom of the page. Note that if (42) is satisfied, then $\alpha - \lambda^2(d-1)^2 \geq 0$, which implies that $\alpha \geq 0$.

We can express (42) as

$$\mathbf{A} - \mathbf{b}\mathbf{b}^* \geq 0 \quad (43)$$

where

$$\mathbf{A} = \begin{bmatrix} \alpha & \frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} \\ \frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} & t\lambda(\lambda\zeta+1) + \lambda\zeta(\gamma-1) - \alpha\epsilon^2 \end{bmatrix} \quad (44)$$

and

$$\mathbf{b} = \begin{bmatrix} \lambda(d-1) \\ d(1+\lambda\zeta) - \lambda\zeta \end{bmatrix}. \quad (45)$$

To treat the constraint (43), we rely on the following lemma [27, p. 472].

Lemma 2: Let

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B}^* \\ \mathbf{B} & \mathbf{C} \end{bmatrix}$$

be a Hermitian matrix. Then, with $\mathbf{C} > 0$, $\mathbf{M} \geq 0$ if and only if $\Delta_{\mathbf{C}} \geq 0$, where $\Delta_{\mathbf{C}}$ is the Schur complement of \mathbf{C} in \mathbf{M} and is given by $\Delta_{\mathbf{C}} = \mathbf{A} - \mathbf{B}^* \mathbf{C}^{-1} \mathbf{B}$.

The proof of the theorem then follows from applying Lemma 2 to (43). \square

IV. ALTERNATIVE DERIVATION OF THE MINIMAX RATIO ESTIMATOR

In Theorem 1, we showed that the minimax ratio estimator can be formulated as an SDP. In this section, we develop further insight into the minimax ratio estimator, using the SDP optimality conditions. Specifically, we show that the minimax ratio estimator can be expressed in terms of a single parameter, which is a solution to a nonlinear equation, and be found using a simple line search algorithm.

To this end, we first show that the minimax ratio estimator, which is the solution to the problem

$$\begin{aligned} (\Gamma) : \min_{\gamma, \mathbf{G}} \left\{ \gamma : \max_{l_i \leq \delta_i \leq u_i} \{ \text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \text{Tr}(\mathbf{C}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^* \right. \\ \left. (\mathbf{I} - \mathbf{G}\mathbf{H})) - \gamma \text{Tr}(\mathbf{C}_x(\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}) \} \leq 0 \right\} \end{aligned} \quad (46)$$

can be determined by first solving the problem

$$\begin{aligned} (\Phi) : \min_{t, \mathbf{G}} \left\{ t : \max_{l_i \leq \delta_i \leq u_i} \{ \text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \text{Tr}(\mathbf{C}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^* \right. \\ \left. (\mathbf{I} - \mathbf{G}\mathbf{H})) - \gamma \text{Tr}(\mathbf{C}_x(\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}) \} \leq t \right\} \end{aligned} \quad (47)$$

$$\begin{bmatrix} \alpha - \lambda^2(d-1)^2 & \frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} + \lambda(1-d)(d(\lambda\zeta+1) - \lambda\zeta) \\ \frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} + \lambda(1-d)(d(\lambda\zeta+1) - \lambda\zeta) & t\lambda(\lambda\zeta+1) - (d(\lambda\zeta+1) - \lambda\zeta)^2 + \lambda\zeta(\gamma-1) - \alpha\epsilon^2 \end{bmatrix} \geq 0. \quad (42)$$

where $\gamma \geq 1$ is fixed. Specifically, let $\hat{t}(\gamma)$ denote the optimal value of t in problem Φ of (47), and let $\hat{\gamma}$ be the minimal value of $\gamma \geq 1$ such that $\hat{t}(\gamma) = 0$ (as we show below in Proposition 2, such a γ always exists). Then, denoting by $\hat{\mathbf{G}}$ the optimal \mathbf{G} in the problem Φ with $\gamma = \hat{\gamma}$, we now show that $\hat{\mathbf{G}}$ and $\hat{\gamma}$ are the optimal solutions to the problem Γ of (49). Since $\hat{\mathbf{G}}$ and $\hat{\gamma}$ are feasible for Φ with $t = 0$, they are also feasible for Γ . Now, suppose, conversely, that there exists a feasible \mathbf{G} and $\gamma < \hat{\gamma}$ for Γ . It then follows that $\hat{t}(\gamma) \leq 0$, but since $\hat{t}(\gamma)$ is decreasing in γ and $\gamma < \hat{\gamma}$, we have that $\hat{t}(\gamma) \geq \hat{t}(\hat{\gamma}) = 0$, from which we conclude that $\hat{t}(\gamma) = 0$, which is a contradiction since $\hat{\gamma}$ is the minimal value for which $\hat{t}(\gamma) = 0$.

In Proposition 2 below, we show that $\hat{t}(\gamma)$ is continuous and strictly decreasing in γ if $l_i \neq 0$ for at least one value of i , because in this case, $\text{Tr}(\mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}) > 0$ for all $l_i \leq \delta_i \leq u_i$. Since $\hat{t}(1) \geq 0$ and $\hat{t}(\gamma) \rightarrow -\infty$ for $\gamma \rightarrow \infty$ (again, because $\text{Tr}(\mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}) > 0$), we conclude that there is a unique γ such that $\hat{t}(\gamma) = 0$. Therefore, we can find $\hat{\gamma}$ by a simple line search, as we discuss further in the paragraph following Proposition 2.

A. Solution to Problem Φ

Since the minimax ratio can be found using the solution of problem Φ , we first consider this problem. Following the proof of Theorem 1, we have immediately that the optimal \mathbf{G} in problem Φ has the form

$$\mathbf{G} = \mathbf{V}\mathbf{D}\mathbf{\Lambda}^{-1}\mathbf{V}^*\mathbf{H}^*\mathbf{C}_w^{-1} \quad (48)$$

where \mathbf{D} is a diagonal matrix with diagonal elements d_i that are the solution to

$$\min_{t, d_i} t \quad (49)$$

subject to

$$\max_{l_i \leq \delta_i \leq u_i} \left\{ \sum_{i=1}^m \left(\frac{d_i^2}{\lambda_i} + \delta_i(1-d_i)^2 - \gamma \frac{\delta_i}{\lambda_i \delta_i + 1} \right) \right\} \leq t. \quad (50)$$

Noting that

$$\begin{aligned} \max_{l_i \leq \delta_i \leq u_i} \left\{ \sum_{i=1}^m \left(\frac{d_i^2}{\lambda_i} + \delta_i(1-d_i)^2 - \gamma \frac{\delta_i}{\lambda_i \delta_i + 1} \right) \right\} &= \\ &= \sum_{i=1}^m \max_{l_i \leq \delta_i \leq u_i} \left\{ \frac{(\lambda_i(d_i-1)\delta_i + d_i)^2 + (1-\gamma)\lambda_i\delta_i}{\lambda_i(\lambda_i\delta_i + 1)} \right\} \end{aligned} \quad (51)$$

our problem can be formulated as

$$\min_{t_i, d_i} \sum_{i=1}^m t_i \quad (52)$$

subject to

$$\max_{l_i \leq \delta_i \leq u_i} \left\{ \frac{(\lambda_i(d_i-1)\delta_i + d_i)^2 + (1-\gamma)\lambda_i\delta_i}{\lambda_i(\lambda_i\delta_i + 1)} \right\} \leq t_i \quad 1 \leq i \leq m. \quad (53)$$

The constraints (53) are the same as the constraints (36), which, in turn, were shown in the proof of Theorem 1 to be equivalent to the LMIs in (25). Thus, to solve Φ , we need to develop a solution to the problem

$$\min_{t, d, \alpha} t \quad (54)$$

subject to (55), shown at the bottom of the next page, which, using Lemma 2, can be equivalently expressed as

$$\mathbf{A} - \mathbf{b}\mathbf{b}^* \geq 0$$

where \mathbf{A} and \mathbf{b} are defined by (44) and (45), respectively.

The problem of (54) and (55) for the special case in which $\gamma = 1$ was considered in [17], in which it was shown that the solution is given by the smallest value of t such that there exists a triplet (t, d, α) satisfying $\mathbf{A} = \mathbf{b}\mathbf{b}^*$. For $\gamma \geq 1$, the smallest value of t and the corresponding values of d and α such that there exists a triplet (t, d, α) satisfying $\mathbf{A} = \mathbf{b}\mathbf{b}^*$ are given by (56) below. As we show in Proposition 1, this solution is also optimal for the problem (54) and (55), as long as γ is smaller than a threshold.

Proposition 1: The solution of

$$\min_{t, d, \alpha} t$$

subject to (55), is given by

$$\begin{aligned} \hat{d} &= 1 - \frac{\sqrt{\gamma}}{\sqrt{(1+\lambda\zeta)^2 - \lambda^2\epsilon^2}} \\ \hat{\alpha} &= \frac{\gamma\lambda^2}{(1+\lambda\zeta)^2 - \lambda^2\epsilon^2} \\ \hat{t} &= \frac{1}{\lambda} - \frac{2\sqrt{\gamma}}{\lambda\sqrt{(1+\lambda\zeta)^2 - \lambda^2\epsilon^2}} + \frac{\gamma(\lambda^2(\epsilon^2 - \zeta^2) + 1)}{\lambda((1+\lambda\zeta)^2 - \lambda^2\epsilon^2)} \end{aligned} \quad (56)$$

for $1 \leq \gamma \leq \gamma^0$, where

$$\gamma^0 = \frac{1 + \lambda(\zeta + \epsilon)}{1 + \lambda(\zeta - \epsilon)} \quad (57)$$

and by

$$\begin{aligned} \hat{d} &= \frac{\lambda(\zeta - \epsilon)}{1 + \lambda(\zeta - \epsilon)} \\ \hat{\alpha} &= \frac{\lambda(\gamma - 1)}{2\epsilon(1 + \lambda(\zeta - \epsilon))} \\ \hat{t} &= \frac{(\gamma - 1)(\epsilon - \zeta)}{1 + \lambda(\zeta - \epsilon)} \end{aligned} \quad (58)$$

for $\gamma \geq \gamma^0$.

Proof: See the Appendix. \square

$$\begin{bmatrix} \alpha & \frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} & \lambda(d-1) \\ \frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} & t\lambda(\lambda\zeta + 1) + \lambda\zeta(\gamma-1) - \alpha\epsilon^2 & d(1+\lambda\zeta) - \lambda\zeta \\ \lambda(d-1) & d(1+\lambda\zeta) - \lambda\zeta & 1 \end{bmatrix} \geq 0 \quad (55)$$

From Proposition 1, it follows that the solution to the problem Φ of (47) is given by

$$\mathbf{G} = \mathbf{V}\widehat{\mathbf{D}}\mathbf{\Lambda}^{-1}\mathbf{V}^*\mathbf{H}^*\mathbf{C}_w^{-1} \quad (59)$$

where $\widehat{\mathbf{D}}$ is the diagonal matrix with diagonal elements

$$\hat{d}_i = \begin{cases} 1 - \frac{\sqrt{\gamma}}{\sqrt{(1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2}} & 1 \leq \gamma \leq \gamma_i^0 \\ \frac{\lambda_i(\zeta_i - \epsilon_i)}{1 + \lambda_i(\zeta_i - \epsilon_i)}, & \gamma \geq \gamma_i^0 \end{cases} \quad (60)$$

with

$$\gamma_i^0 = \frac{1 + \lambda_i(\zeta_i + \epsilon_i)}{1 + \lambda_i(\zeta_i - \epsilon_i)} \quad (61)$$

and $\hat{t}(\gamma) = \sum_{i=1}^m t_i(\gamma)$, where we have (62), shown at the bottom of the page.

B. Minimax Ratio Estimator

To develop an alternative expression for the minimax ratio estimator, we now relate the solutions of problems Φ and Γ of (47) and (46), respectively. To this end, we first establish the required properties of $\hat{t}(\gamma)$, which were outlined at the beginning of the section.

Proposition 2: Let $\hat{t}(\gamma) = \sum_{i=1}^m t_i(\gamma)$, where $t_i(\gamma)$ is given by (62). Then, we have the following.

- 1) $\hat{t}(\gamma)$ is continuous.
- 2) If $\epsilon_i < \zeta_i$ for some i , then $\hat{t}(\gamma)$ is strictly decreasing in $\gamma \geq 1$, and there is a unique γ such that $\hat{t}(\gamma) = 0$.

Proof: For $\gamma < \gamma_i^0$, $t_i(\gamma)$ is quadratic in γ , and for $\gamma > \gamma_i^0$, $t_i(\gamma)$ is linear in γ . Therefore, in both of these intervals, $t_i(\gamma)$ is continuous. We can also immediately verify that $t_i(\gamma)$ is continuous at $\gamma = \gamma_i^0$.

We now consider the monotonicity property of $t_i(\gamma)$. Suppose first that $\lambda_i^2(\zeta_i^2 - \epsilon_i^2) = 1$. In this case, $t_i(\gamma)$ is linear in γ for $\gamma \leq \gamma_i^0$ and is strictly decreasing. Next, suppose that $\lambda_i^2(\zeta_i^2 - \epsilon_i^2) \neq 1$. Then, for $\gamma \leq \gamma_i^0$, $t_i(\gamma)$ is a quadratic function in $\sqrt{\gamma}$, which we denote by $Q(\gamma)$, with an extremum point at

$$\sqrt{\gamma} = \frac{\sqrt{(1 + \lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2}}{1 - \lambda_i^2(\zeta_i^2 - \epsilon_i^2)} \triangleq \gamma_i^t. \quad (63)$$

If $\lambda_i^2(\zeta_i^2 - \epsilon_i^2) < 1$, then $Q(\gamma)$ has a minimum at γ_i^t . Consequently, $Q(\gamma)$ is decreasing for any $\gamma \leq (\gamma_i^t)^2$. Since, as we now show, $(\gamma_i^t)^2 \geq \gamma_i^0$, $t_i(\gamma)$ is decreasing for $\gamma \leq \gamma_i^0$. To prove that $(\gamma_i^t)^2 \geq \gamma_i^0$, we must have that

$$\frac{(1 + \lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2}{(1 - \lambda_i^2(\zeta_i^2 - \epsilon_i^2))^2} \geq \frac{1 + \lambda_i(\zeta_i + \epsilon_i)}{1 + \lambda_i(\zeta_i - \epsilon_i)} \quad (64)$$

or, equivalently

$$1 + \lambda(\zeta_i - \epsilon_i) \geq 1 - \lambda_i^2(\zeta_i^2 - \epsilon_i^2) \quad (65)$$

which is clearly satisfied. If, on the other hand, $\lambda_i^2(\zeta_i^2 - \epsilon_i^2) > 1$, then $Q(\gamma)$ has a maximum at $\gamma_i^t < 0$, and $t_i(\gamma)$ is strictly decreasing for $1 \leq \gamma \leq \gamma_i^0$.

For $\gamma \geq \gamma_i^0$, $t_i(\gamma)$ is strictly decreasing as long as $\epsilon_i < \zeta_i$. If $\epsilon_i = \zeta_i$, then $t_i(\gamma) = 0$ for all $\gamma \geq \gamma_i^0$.

We conclude that $\hat{t}(\gamma)$ is strictly decreasing in γ if for some i , $\epsilon_i < \zeta_i$. Since $\hat{t}(1) > 0$ (unless $\epsilon_i = 0$, $1 \leq i \leq m$, in which case, there is no uncertainty) and $\hat{t}(\gamma) \leq 0$ for $\gamma \rightarrow -\infty$, it follows that in this case, there is a unique value γ such that $\hat{t}(\gamma) = 0$. \square

From Proposition 2, it follows that the solution to the original problem Γ is also given by (59) and (60), with γ chosen as the minimal value such that $\hat{t}(\gamma) = \sum_{i=1}^m t_i(\gamma) = 0$. Since $\hat{t}(\gamma)$ is continuous and strictly decreasing in γ such that $\hat{t}(\gamma) \neq 0$, we can find the minimal value of γ satisfying $\hat{t}(\gamma) = 0$ by a simple line search. Specifically, we may start by choosing $\gamma = 1$. For each choice of γ , we compute $\sum_{i=1}^m t_i(\gamma)$. If $\sum_{i=1}^m t_i(\gamma) > 0$, then we increase γ , and if $\sum_{i=1}^m t_i(\gamma) < 0$, we decrease γ , continuing until $\sum_{i=1}^m t_i(\gamma) = 0$. Assuming that $\epsilon_i < \zeta_i$ for some i (which is our standing assumption), this is the optimal value of γ . Due to the continuity and monotonicity properties of $\hat{t}(\gamma)$, the algorithm is guaranteed to converge.

We summarize our results in the following theorem.

Theorem 2: Let \mathbf{x} denote the unknown parameters in the model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where \mathbf{H} is a known $n \times m$ matrix with rank m , \mathbf{x} is a zero-mean random vector with covariance \mathbf{C}_x uncorrelated with \mathbf{w} , and \mathbf{w} is a zero-mean random vector with covariance \mathbf{C}_w . Let $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^*$, where \mathbf{V} is a unitary matrix, and $\mathbf{\Lambda}$ is an $m \times m$ diagonal matrix with diagonal elements $\lambda_i > 0$, and let $\mathbf{C}_x = \mathbf{V}\mathbf{\Delta}\mathbf{V}^*$, where $\mathbf{\Delta}$ is an $m \times m$ diagonal matrix with diagonal elements $0 \leq l_i \leq \delta_i \leq u_i$. Then, the solution to the problem

$$\min_{\hat{\mathbf{x}} \in \mathbf{G}\mathbf{y}} \max_{l_i \leq \delta_i \leq u_i} \left\{ \frac{E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)}{\min_{\hat{\mathbf{x}} \in \mathbf{G}(\mathbf{x})\mathbf{y}} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)} \right\}$$

is given by

$$\hat{\mathbf{x}} = \mathbf{V}\mathbf{D}\mathbf{\Lambda}^{-1}\mathbf{V}^*\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}$$

where \mathbf{D} is an $m \times m$ diagonal matrix with diagonal elements

$$\hat{d}_i = \begin{cases} 1 - \frac{\sqrt{\gamma}}{\sqrt{(1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2}}, & 1 \leq \gamma \leq \gamma_i^0 \\ \frac{\lambda_i(\zeta_i - \epsilon_i)}{1 + \lambda_i(\zeta_i - \epsilon_i)}, & \gamma \geq \gamma_i^0 \end{cases} \quad (66)$$

with

$$\gamma_i^0 = \frac{1 + \lambda_i(\zeta_i + \epsilon_i)}{1 + \lambda_i(\zeta_i - \epsilon_i)} \quad (67)$$

and γ is the unique value such that $\hat{t}(\gamma) = \sum_{i=1}^m t_i(\gamma) = 0$, where $t_i(\gamma)$ is given by (62).

$$t_i(\gamma) = \begin{cases} \frac{1}{\lambda_i} - \frac{2\sqrt{\gamma}}{\lambda_i\sqrt{(1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2}} + \frac{\gamma(\lambda_i^2(\epsilon_i^2 - \zeta_i^2) + 1)}{\lambda_i((1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2)}, & 1 \leq \gamma \leq \gamma_i^0 \\ \frac{(\gamma-1)(\epsilon_i - \zeta_i)}{1 + \lambda_i(\zeta_i - \epsilon_i)}, & \gamma \geq \gamma_i^0. \end{cases} \quad (62)$$

Note that the minimax regret estimator has the same form as the minimax ratio estimator, where d_i is given by (66) with $\gamma = 1$.

In the special case in which $\mathbf{x} = x$ is a scalar and $\mathbf{H} = \mathbf{h}$ is a vector, we can immediately show from Theorem 2 that the optimal estimator is

$$\hat{x} = \left(1 - \frac{1}{1 + \mathbf{h}^* \mathbf{C}_w^{-1} \mathbf{h} \sqrt{(\zeta + \epsilon)(\zeta - \epsilon)}} \right) \frac{\mathbf{h}^* \mathbf{y}}{\mathbf{h}^* \mathbf{C}_w^{-1} \mathbf{h}}. \quad (68)$$

C. MMSE Interpretation of the Minimax Ratio Estimator

As we now show, we can interpret the estimator of Theorem 2 as an MMSE estimator matched to a covariance matrix

$$\mathbf{C}_x = \mathbf{V} \mathbf{X} \mathbf{V}^* \quad (69)$$

where \mathbf{X} is a diagonal matrix with diagonal elements

$$x_i = \begin{cases} \frac{1}{\lambda_i} \left(\frac{\sqrt{(1 + \lambda_i \zeta_i)^2 - \lambda_i^2 \epsilon_i^2}}{\sqrt{\gamma}} - 1 \right), & \gamma \leq \gamma_i^0 \\ \zeta_i - \epsilon_i, & \gamma \geq \gamma_i^0. \end{cases} \quad (70)$$

It follows from (70) that if $\gamma \geq \gamma_i^0$, then x_i is equal to the lower bound on the uncertainty region of the i th eigenvalue of \mathbf{C}_x .

From (5), the MMSE estimate of \mathbf{x} with covariance \mathbf{C}_x given by (77) and $\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^*$ is

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{C}_x (\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H} \mathbf{C}_x + \mathbf{I})^{-1} \mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{y} \\ &= \mathbf{V} \mathbf{X} (\mathbf{\Lambda} \mathbf{X} + \mathbf{I})^{-1} \mathbf{V}^* \mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{y}. \end{aligned} \quad (71)$$

Thus, the estimator $\hat{\mathbf{x}}$ of (71) is equivalent to the minimax ratio estimator given by Theorem 2, if $\mathbf{X} (\mathbf{\Lambda} \mathbf{X} + \mathbf{I})^{-1} = \mathbf{D} \mathbf{\Lambda}^{-1}$. Now

$$\frac{x_i}{\lambda_i x_i + 1} = \begin{cases} \frac{1}{\lambda_i} \left(1 - \frac{\sqrt{\gamma}}{\sqrt{(1 + \lambda_i \zeta_i)^2 - \lambda_i^2 \epsilon_i^2}} \right), & 1 \leq \gamma \leq \gamma_i^0 \\ \frac{\zeta_i - \epsilon_i}{1 + \lambda_i (\zeta_i - \epsilon_i)}, & \gamma \geq \gamma_i^0 \end{cases} \quad (72)$$

so that indeed $\mathbf{X} (\mathbf{\Lambda} \mathbf{X} + \mathbf{I})^{-1} = \mathbf{D} \mathbf{\Lambda}^{-1}$.

Since the minimax ratio estimator minimizes the MSE for $\mathbf{C}_x = \mathbf{V} \mathbf{X} \mathbf{V}^*$, we may view the covariance $\mathbf{C}_x = \mathbf{V} \mathbf{X} \mathbf{V}^*$ as the “least-favorable” covariance in the ratio sense.

In [17], it was shown that the minimax regret estimator is an MMSE estimator matched to a covariance matrix with eigenvalues

$$c_i = \frac{1}{\lambda_i} \left(1 - \frac{1}{\sqrt{(1 + \lambda_i \zeta_i)^2 - \lambda_i^2 \epsilon_i^2}} \right). \quad (73)$$

Since the optimal value of γ is greater than 1 (unless there is no uncertainty), $x_i < c_i$, $1 \leq i \leq m$ so that the minimax ratio estimator is matched to a covariance matrix with eigenvalues that are strictly smaller than the eigenvalues of the covariance matrix matched to the minimax difference regret estimator.

V. EXAMPLE OF THE MINIMAX RATIO ESTIMATOR

We now consider examples illustrating the minimax ratio estimator. These examples are the same as those presented in [17] for evaluating the minimax difference regret estimator.

Consider the estimation problem in which

$$\mathbf{y} = \mathbf{x} + \mathbf{w} \quad (74)$$

where \mathbf{x} is a length- n segment of a zero-mean stationary first-order AR process with components x_i so that

$$E(x_i x_j) = \rho^{|j-i|} \quad (75)$$

for some parameter ρ , and \mathbf{w} is a zero-mean random vector uncorrelated with \mathbf{x} with known covariance $\mathbf{C}_w = \sigma^2 \mathbf{I}$. We assume that we know the model (74) and that \mathbf{x} is a segment of a stationary process; however, its covariance \mathbf{C}_x is unknown.

To estimate \mathbf{x} , we may first estimate \mathbf{C}_x from the observations \mathbf{y} . A natural estimate of \mathbf{C}_x is given by

$$\hat{\mathbf{C}}_x = [\hat{\mathbf{C}}_y - \mathbf{C}_w]_+ = [\hat{\mathbf{C}}_y - \sigma^2 \mathbf{I}]_+ \quad (76)$$

where

$$\hat{\mathbf{C}}_y(i, j) = \frac{1}{n} \sum_{k=1}^{n-|j-i|} y_k y_{k+|j-i|} \quad (77)$$

is an estimate of the covariance of \mathbf{y} , and $[\mathbf{A}]_+$ denotes the matrix in which the negative eigenvalues of \mathbf{A} are replaced by 0. Thus, if \mathbf{A} has an eigendecomposition $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{U}^{-1}$, where $\mathbf{\Sigma}$ is a diagonal matrix with diagonal elements σ_i , then $[\mathbf{A}]_+ = \mathbf{U} [\mathbf{\Sigma}]_+ \mathbf{U}^{-1}$, where $[\mathbf{\Sigma}]_+$ is a diagonal matrix with the i th diagonal element equal to $\max(0, \sigma_i)$. The estimate (76) can be regarded as the analog for finite-length processes of the spectrum estimate based on the spectral subtraction method for infinite-length processes [33], [34].

Given $\hat{\mathbf{C}}_x$, we may estimate \mathbf{x} using an MMSE estimate matched to $\hat{\mathbf{C}}_x$, which we refer to as a plug-in estimator. However, as can be seen in Figs. 1 and 2, we can further improve the estimation performance by using the minimax ratio estimator.

To compute the minimax ratio estimator, we choose \mathbf{V} to be equal to the eigenvector matrix of the estimated covariance matrix $\hat{\mathbf{C}}_x$ and $\zeta_i = \sigma_i$, where σ_i are the eigenvalues of $\hat{\mathbf{C}}_x$. We would then like to choose ϵ_i to reflect the uncertainty in our estimate ζ_i . Since computing the standard deviation of ζ_i is difficult, we choose ϵ_i to be proportional to the standard deviation of an estimator $\tilde{\sigma}_x^2$ of the variance σ_x^2 of \mathbf{x} , where

$$\tilde{\sigma}_x^2 = \frac{1}{n} \sum_{i=1}^n y_i^2 - \sigma_w^2. \quad (78)$$

We further assume that \mathbf{x} and \mathbf{w} are uncorrelated Gaussian random vectors. In this case, the variance of $\tilde{\sigma}_x^2$ is given by [17]

$$E \left\{ (\tilde{\sigma}_x^2 - \sigma_x^2)^2 \right\} = \frac{2}{n} \left((\sigma_x^2 + \sigma^2)^2 + \sum_{i=2}^n \mathbf{C}_x^2(1, i) \right). \quad (79)$$

Since σ_x^2 and $\mathbf{C}_x(1, i)$ are unknown, we substitute their estimates $\hat{\mathbf{C}}_x(1, i)$, $1 \leq i \leq m$. Finally, to ensure that $\epsilon_i \leq \zeta_i$, we choose

$$\epsilon_i = \min \left(\zeta_i, A \sqrt{\frac{2}{n} \left((\hat{\mathbf{C}}_x^2(1, 1) + \sigma^2)^2 + \sum_{i=2}^n \hat{\mathbf{C}}_x^2(1, i) \right)} \right) \quad (80)$$

where A is a proportionality factor.

In Fig. 1, we plot the MSE of the minimax ratio estimator averaged over 1000 noise realizations as a function of the SNR defined by $-10 \log \sigma^2$ for $\rho = 0.8$, $n = 10$, and $A = 4$. The performance of the “plug-in” MMSE estimator matched to the

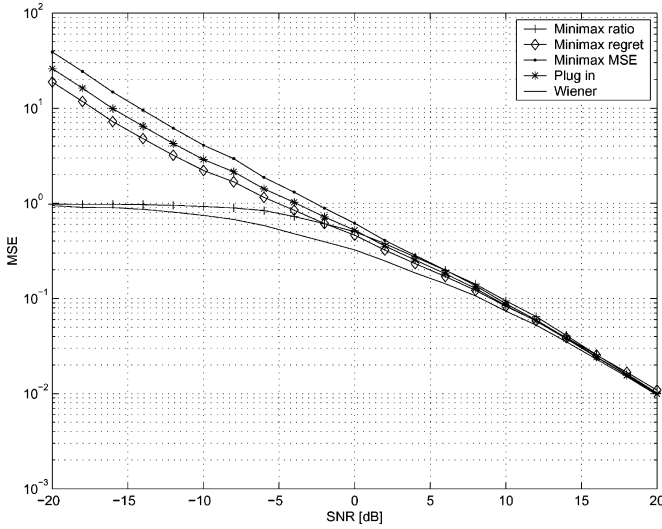


Fig. 1. MSE in estimating \mathbf{x} as a function of SNR using the minimax ratio estimator, the minimax regret estimator, the minimax MSE estimator, and the plug-in MMSE estimator matched to the estimated covariance matrix. The performance of the optimal Wiener estimator is plotted for comparison.

estimated covariance matrix $\hat{\mathbf{C}}_x$, the minimax MSE estimator, and the minimax regret estimator of [17] are plotted for comparison. We also plot the MSE resulting from a Wiener filter matched to the known covariance, which is the optimal MSE attainable when \mathbf{C}_x is known. As can be seen from the figure, the minimax ratio estimator can significantly increase the estimation performance at low to intermediate SNR values, and in this range, the performance of the minimax ratio estimator is close to the optimal performance.

We next consider the case in which the vector \mathbf{x} is filtered with an LTI filter with length-4 impulse response given by

$$h[0] = 1, \quad h[1] = 0.4, \quad h[2] = 0.2, \quad h[3] = 0.1. \quad (81)$$

Note that in this case, the eigenvector matrices of $\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H}$ and \mathbf{C}_x are no longer equal. Nonetheless, we can still implement the minimax ratio estimator by using the results of Section IV-C. Specifically, in Section IV-C, it was shown that the minimax ratio estimator can be formulated as an MMSE estimator matched to a least-favorable covariance. Therefore, to implement the estimator, we may first estimate the eigenvalues and the eigenvector matrix of \mathbf{C}_x from the data and then compute the least-favorable covariance matrix with eigenvalues given by (78).

In Fig. 2, we plot the MSE of the minimax ratio, minimax regret, plug-in, and minimax MSE estimators averaged over 1000 noise realizations as a function of the SNR for $\rho = 0.8$, $n = 10$, and $A = 4$. For comparison, we also plot the bound on the performance given by the MSE of the Wiener estimator. As can be seen, the trends in performance are similar to the previous example.

VI. DISCUSSION—TRADING OFF RATIO REGRET AND DIFFERENCE REGRET

As described in the Introduction, in [17], we have studied a robust estimation problem in the same spirit, where instead

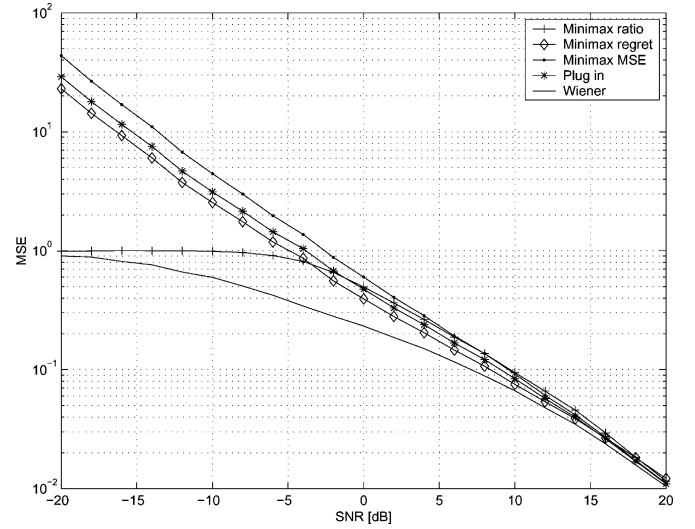


Fig. 2. MSE in estimating \mathbf{x} from a noisy filtered version as a function of SNR using the minimax ratio estimator, the minimax regret estimator, the minimax MSE estimator, and the plug-in MMSE estimator matched to the estimated covariance matrix. The performance of the optimal Wiener estimator is plotted for comparison.

of the ratio regret $\mathcal{R}(\mathbf{C}_x, \mathbf{G})$, the objective function was the difference regret

$$\mathcal{D}(\mathbf{C}_x, \mathbf{G}) = E(\|\mathbf{G}\mathbf{y} - \mathbf{x}\|^2) - \text{MSE}^o(\mathbf{C}_x). \quad (82)$$

Each of the two criteria $\mathcal{R}(\mathbf{C}_x, \mathbf{G})$ and $\mathcal{D}(\mathbf{C}_x, \mathbf{G})$ has its own rationale, as well as its own advantages and disadvantages. Ideally, one would wish, of course, to minimize them both at the same time. However, since their optimum solutions are different, in general, these minimizations are conflicting and, therefore, cannot be achieved simultaneously. If both criteria are of interest, nevertheless, a plausible approach would be to seek the best possible tradeoff between them. More specifically, it is natural to ask what is the achievable region of points in the plane (r, d) for which there exists a linear estimator \mathbf{G} such that for every \mathbf{C}_x in the region of uncertainty

$$\mathcal{R}(\mathbf{C}_x, \mathbf{G}) \leq r \quad \text{and} \quad \mathcal{D}(\mathbf{C}_x, \mathbf{G}) \leq d. \quad (83)$$

In this partition of the plane, between the achievable region and the unachievable region, the most interesting part is, of course, the boundary curve between them, as it characterizes the best possible tradeoff. The boundary curve $r(d)$ corresponds to the solution of the problem

$$r(d) = \min_{\mathbf{G}} \max_{\mathbf{C}_x} \mathcal{R}(\mathbf{C}_x, \mathbf{G}) \quad (84)$$

subject to

$$\max_{\mathbf{C}_x} \mathcal{D}(\mathbf{C}_x, \mathbf{G}) \leq d \quad (85)$$

where the maximizations over \mathbf{C}_x are across the uncertainty region. The techniques that have been used here and in [17] can be applied to solve this minimization problem, thus yielding the spectrum of optimum tradeoffs between the ratio and the difference regrets, but the details are omitted due to space constraints. As a general note, we mention that the two extreme points of

this curve correspond, of course, to the pure difference regret problem of [17] and the pure ratio regret problem considered in this paper. In addition, it is easy to show that the achievability region (and hence also the function $r(d)$) is convex.

APPENDIX

In this Appendix, we prove Proposition 1. We begin by showing that the values given by (56) are optimal for $1 \leq \gamma \leq \gamma^0$. To this end, it is sufficient to show that $\mathbf{F}(\hat{d}, \hat{\alpha}, \hat{t}) \geq 0$, where $\mathbf{F}(d, \alpha, t)$ is the matrix in (55), that there exists a matrix $\mathbf{Z} \geq 0$ such that

$$\begin{aligned} \mathbf{F}(\hat{d}, \hat{\alpha}, \hat{t})\mathbf{Z} &= 0 \\ \text{Tr}(\mathbf{F}_d\mathbf{Z}) &= 0 \\ \text{Tr}(\mathbf{F}_\alpha\mathbf{Z}) &= 0 \\ \text{Tr}(\mathbf{F}_t\mathbf{Z}) &= 1 \end{aligned} \quad (86)$$

where

$$\begin{aligned} \mathbf{F}_d &= \begin{bmatrix} 0 & 0 & \lambda \\ 0 & 0 & 1 + \lambda\zeta \\ \lambda & 1 + \lambda\zeta & 0 \end{bmatrix} \\ \mathbf{F}_\alpha &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\epsilon^2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ \mathbf{F}_t &= \begin{bmatrix} 0 & \frac{\lambda^2}{2} & 0 \\ \frac{\lambda^2}{2} & \lambda(1 + \lambda\zeta) & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned} \quad (87)$$

and that the dual problem is strictly feasible, i.e., there exists a matrix \mathbf{B} satisfying

$$\begin{aligned} \mathbf{B} &> 0 \\ \text{Tr}(\mathbf{F}_d\mathbf{B}) &= 0 \\ \text{Tr}(\mathbf{F}_\alpha\mathbf{B}) &= 0 \\ \text{Tr}(\mathbf{F}_t\mathbf{B}) &= 1. \end{aligned} \quad (88)$$

Let

$$\mathbf{B} = \begin{bmatrix} \frac{b\epsilon^2}{\lambda^2(1 - b\lambda(\lambda\zeta + 1))} & \frac{1}{\lambda^2}(1 - b\lambda(\lambda\zeta + 1)) & 0 \\ \frac{1}{\lambda^2}(1 - b\lambda(\lambda\zeta + 1)) & b & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (89)$$

where

$$b = \frac{1}{\lambda(\lambda\zeta + 1) + \lambda^2(\epsilon - a)} \quad (90)$$

and a is chosen such that $\epsilon^2 > (\epsilon - a)^2$. We can immediately verify that \mathbf{B} satisfies (88). Next, we verify that $\mathbf{F}(\hat{d}, \hat{\alpha}, \hat{t}) \geq 0$. Substituting (56) into the matrix in (55), we have (91), shown at the bottom of the page. From Lemma 2, it follows that $\mathbf{F}(\hat{d}, \hat{\alpha}, \hat{t}) \geq 0$ if and only if $\mathbf{W} \geq 0$, where

$$\mathbf{W} = \begin{bmatrix} \gamma\lambda^2 & \lambda(\gamma(1 + \lambda\zeta) - \beta\sqrt{\gamma}) \\ \lambda(\gamma(1 + \lambda\zeta) - \beta\sqrt{\gamma}) & (\beta - \sqrt{\gamma}(1 + \lambda\zeta))^2 \end{bmatrix} - \begin{bmatrix} \lambda^2\gamma & -\lambda\sqrt{\gamma}(\beta - \sqrt{\gamma}(1 + \lambda\zeta)) \\ -\lambda\sqrt{\gamma}(\beta - \sqrt{\gamma}(1 + \lambda\zeta)) & (\beta - \sqrt{\gamma}(1 + \lambda\zeta))^2 \end{bmatrix} = 0. \quad (92)$$

We must still find a $\mathbf{Z} \geq 0$ satisfying (86). Solving (86) for \mathbf{Z} with t, d, α given by (56) results in (93), shown at the bottom of the page, where, for brevity, we defined

$$\beta \triangleq \sqrt{(1 + \lambda\zeta)^2 - \lambda^2\epsilon^2} \quad (94)$$

and

$$\kappa \triangleq 2\sqrt{\gamma}(1 + \lambda\zeta) - \beta. \quad (95)$$

Since \mathbf{Z} satisfies (86), the values given by (56) are optimal if, in addition, $\mathbf{Z} \geq 0$, which implies that

$$\kappa \geq \beta\gamma \quad (96)$$

or, equivalently

$$\beta\gamma - 2\sqrt{\gamma}(1 + \lambda\zeta) + \beta \leq 0. \quad (97)$$

It is straightforward to show that (97) is satisfied for

$$\frac{(1 + \lambda(\zeta - \epsilon))^2}{\beta^2} \leq \gamma \leq \frac{(1 + \lambda(\zeta + \epsilon))^2}{\beta^2}. \quad (98)$$

Since

$$\begin{aligned} \frac{(1 + \lambda(\zeta - \epsilon))^2}{\beta^2} &= \frac{(1 + \lambda(\zeta - \epsilon))^2}{(1 + \lambda(\zeta + \epsilon))(1 + \lambda(\zeta - \epsilon))} \\ &= \frac{1 + \lambda(\zeta - \epsilon)}{1 + \lambda(\zeta + \epsilon)} \leq 1 \end{aligned} \quad (99)$$

$$\mathbf{F}(\hat{d}, \hat{\alpha}, \hat{t}) = \begin{bmatrix} \gamma\lambda^2 & \lambda(\gamma(1 + \lambda\zeta) - \beta\sqrt{\gamma}) & -\lambda\sqrt{\gamma}\beta \\ \lambda(\gamma(1 + \lambda\zeta) - \beta\sqrt{\gamma}) & (\beta - \sqrt{\gamma}(1 + \lambda\zeta))^2 & \beta^2 - \beta\sqrt{\gamma}(1 + \lambda\zeta) \\ -\lambda\sqrt{\gamma}\beta & \beta^2 - \beta\sqrt{\gamma}(1 + \lambda\zeta) & \beta^2 \end{bmatrix}. \quad (91)$$

$$\mathbf{Z} = \frac{1}{\lambda\beta^2\sqrt{\gamma}} \begin{bmatrix} \epsilon^2\kappa & \frac{1}{\lambda}(\beta^2\sqrt{\gamma} - (1 + \lambda\zeta)\kappa) & \frac{1}{\lambda}(1 + \lambda\zeta)(\kappa - \beta\gamma) \\ \frac{1}{\lambda}(\beta^2\sqrt{\gamma} - (1 + \lambda\zeta)\kappa) & \kappa & \beta\gamma - \kappa \\ \frac{1}{\lambda}(1 + \lambda\zeta)(\kappa - \beta\gamma) & \beta\gamma - \kappa & \kappa - \beta\gamma \end{bmatrix} \quad (93)$$

Equation (96) is satisfied for $1 \leq \gamma \leq \gamma^0$, where γ^0 is given by (57).

We now show that for $1 \leq \gamma \leq \gamma^0$, $\mathbf{Z} \geq 0$. First, suppose that $\kappa = \beta\gamma$, or equivalently, $\gamma = \gamma^0$. In this case, $\mathbf{Z} \geq 0$ if and only if

$$\begin{bmatrix} \epsilon^2 \kappa & \frac{1}{\lambda} (\beta^2 \sqrt{\gamma} - (1 + \lambda\zeta)\kappa) \\ \frac{1}{\lambda} (\beta^2 \sqrt{\gamma} - (1 + \lambda\zeta)\kappa) & \kappa \end{bmatrix} \geq 0 \quad (100)$$

or

$$\lambda^2 \epsilon^2 \kappa^2 - (\beta^2 \sqrt{\gamma} - (1 + \lambda\zeta)\kappa)^2 \geq 0. \quad (101)$$

Using the fact that $\kappa = \beta\gamma$, (101) can be expressed as

$$-\beta - \beta\gamma + 2\sqrt{\gamma}(1 + \lambda\zeta) \geq 0 \quad (102)$$

which is equivalent to (97) and is therefore satisfied for $\gamma = \gamma^0$.

Next, suppose that $\kappa > \beta\gamma$. From Lemma 2, we then have that $\mathbf{Z} \geq 0$ if and only if

$$\begin{aligned} & \begin{bmatrix} \epsilon^2 \kappa & \frac{1}{\lambda} (\beta^2 \sqrt{\gamma} - (1 + \lambda\zeta)\kappa) \\ \frac{1}{\lambda} (\beta^2 \sqrt{\gamma} - (1 + \lambda\zeta)\kappa) & \kappa \end{bmatrix} \\ & + (\beta\gamma - \kappa) \begin{bmatrix} \frac{1}{\lambda^2} (1 + \lambda\zeta)^2 & -\frac{1}{\lambda} (1 + \lambda\zeta) \\ -\frac{1}{\lambda} (1 + \lambda\zeta) & 1 \end{bmatrix} = \\ & = \begin{bmatrix} \epsilon^2 \kappa + \frac{1}{\lambda^2} (\beta\gamma - \kappa)(1 + \lambda\zeta)^2 & \frac{1}{\lambda} (\beta^2 \sqrt{\gamma} - \beta\gamma(1 + \lambda\zeta)) \\ \frac{1}{\lambda} (\beta^2 \sqrt{\gamma} - \beta\gamma(1 + \lambda\zeta)) & \gamma\beta \end{bmatrix} \\ & \geq 0 \end{aligned} \quad (103)$$

which is equivalent to the conditions

$$\epsilon^2 \kappa + \frac{1}{\lambda^2} (\beta\gamma - \kappa)(1 + \lambda\zeta)^2 \geq 0 \quad (104)$$

$$\begin{aligned} & \gamma\beta \left(\epsilon^2 \kappa + \frac{1}{\lambda^2} (\beta\gamma - \kappa)(1 + \lambda\zeta)^2 \right) \\ & - \frac{1}{\lambda^2} (\beta^2 \sqrt{\gamma} - \beta\gamma(1 + \lambda\zeta))^2 \geq 0. \end{aligned} \quad (105)$$

Now

$$\begin{aligned} & \epsilon^2 \kappa + \frac{1}{\lambda^2} (\beta\gamma - \kappa)(1 + \lambda\zeta)^2 \\ & = \frac{\kappa}{\lambda^2} (\lambda^2 \epsilon^2 - (1 + \lambda\zeta)^2) + \frac{\beta\gamma}{\lambda^2} (1 + \lambda\zeta)^2 \\ & = \frac{\beta}{\lambda^2} (\gamma(1 + \lambda\zeta)^2 - \beta\kappa) \\ & = \frac{\beta}{\lambda^2} (\sqrt{\gamma}(1 + \lambda\zeta) - \beta)^2 \geq 0 \end{aligned} \quad (106)$$

so that (96) is satisfied. Finally

$$\begin{aligned} & \gamma\beta \left(\epsilon^2 \kappa + \frac{1}{\lambda^2} (\beta\gamma - \kappa)(1 + \lambda\zeta)^2 \right) \\ & - \frac{1}{\lambda^2} (\beta^2 \sqrt{\gamma} - \beta\gamma(1 + \lambda\zeta))^2 = \\ & = \frac{\gamma\beta^2}{\lambda^2} (\sqrt{\gamma}(1 + \lambda\zeta) - \beta)^2 - \frac{1}{\lambda^2} (\beta^2 \sqrt{\gamma} - \beta\gamma(1 + \lambda\zeta))^2 \\ & = 0 \end{aligned} \quad (107)$$

so that (105) is also satisfied, and $\mathbf{Z} \geq 0$.

We now show that for $\gamma \geq \gamma^0$, the values given by (58) are optimal. For these values, we have that

$$\mathbf{F}(\hat{d}, \hat{\alpha}, \hat{t}) = \frac{1}{1 + \lambda(\zeta - \epsilon)} \begin{bmatrix} \frac{\lambda(\gamma-1)}{2\epsilon} & \frac{\lambda(\gamma-1)}{2} & -\lambda \\ \frac{\lambda(\gamma-1)}{2} & \frac{\lambda\epsilon(\gamma-1)}{2} & -\lambda\epsilon \\ -\lambda & -\lambda\epsilon & 1 + \lambda(\zeta - \epsilon) \end{bmatrix}. \quad (108)$$

From Lemma 2 it then follows that $\mathbf{F}(\hat{d}, \hat{\alpha}, \hat{t}) \geq 0$ if and only if

$$\begin{aligned} & \begin{bmatrix} \frac{\lambda(\gamma-1)}{2\epsilon} & \frac{\lambda(\gamma-1)}{2} \\ \frac{\lambda(\gamma-1)}{2} & \frac{\lambda\epsilon(\gamma-1)}{2} \end{bmatrix} - \frac{\lambda^2}{1 + \lambda(\zeta - \epsilon)} \begin{bmatrix} 1 & \epsilon \\ \epsilon & \epsilon^2 \end{bmatrix} \\ & = \frac{\lambda(\gamma-1)(1 + \lambda(\zeta - \epsilon)) - 2\epsilon\lambda^2}{2\epsilon(1 + \lambda(\zeta - \epsilon))} \begin{bmatrix} 1 & \epsilon \\ \epsilon & \epsilon^2 \end{bmatrix} \geq 0. \end{aligned} \quad (109)$$

Clearly, (109) is satisfied if

$$\lambda(\gamma-1)(1 + \lambda(\zeta - \epsilon)) - 2\epsilon\lambda^2 \geq 0 \quad (110)$$

or

$$\gamma \geq \frac{1 + \lambda(\zeta + \epsilon)}{1 + \lambda(\zeta - \epsilon)} = \gamma^0. \quad (111)$$

We now show that there exists a $\mathbf{Z} \geq 0$ satisfying (86). Solving (86) for \mathbf{Z} with t, d, α given by (58) results in

$$\mathbf{Z} = \frac{1}{\lambda^2(\zeta - \epsilon) + \lambda} \begin{bmatrix} \epsilon^2 & -\epsilon & 0 \\ -\epsilon & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (112)$$

and $\mathbf{Z} \geq 0$. Thus, the values given by (58) are optimal for $\gamma \geq \gamma^0$, completing the proof of the proposition.

REFERENCES

- [1] N. Wiener, *The Extrapolation, Interpolation and Smoothing of Stationary Time Series*. New York: Wiley, 1949.
- [2] A. Kolmogorov, "Interpolation and extrapolation," *Bull. Acad. Sci., USSR, Ser. Math.*, vol. 5, pp. 3–14, 1941.
- [3] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Statist.*, vol. 35, pp. 73–101, 1964.
- [4] —, *Robust Statistics*. New York: Wiley, 1981.
- [5] L. Breiman, "A note on minimax filtering," *Ann. Probab.*, vol. 1, pp. 175–179, 1973.
- [6] S. A. Kassam and T. L. Lim, "Robust Wiener filters," *J. Franklin Inst.*, vol. 304, pp. 171–185, Oct./Nov. 1977.
- [7] H. V. Poor, "On robust Wiener filtering," *IEEE Trans. Automat. Control*, vol. AC-25, pp. 521–526, Jun. 1980.
- [8] K. S. Vastola and H. V. Poor, "Robust Wiener-Kolmogorov theory," *IEEE Trans. Inf. Theory*, vol. IT-30, pp. 316–327, Mar. 1984.
- [9] J. Franke, "Minimax-robust prediction of discrete time series," *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, vol. 68, pp. 337–364, 1985.
- [10] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proc. IEEE*, vol. 73, pp. 433–481, Mar. 1985.
- [11] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 168–181, Jan. 2005.
- [12] Y. C. Eldar and A. Nehorai, "Competitive mean-squared error beamforming," in *Proc. 12th Annu. Workshop Adaptive Sensor Array Processing*, Lincoln Laboratory, Mass. Inst. Technol., Lexington, MA, Mar. 2004.
- [13] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst case performance optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 313–324, Feb. 2003.
- [14] J. Li, P. Stoica, and Z. Wang, "On robust capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, Jul. 2003.

- [15] S. Shahbazpanahi, A. B. Gershman, Z.-Q. Luo, and K. M. Wong, "Robust adaptive beamforming for general-rank signal models," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2257–2269, Sep. 2003.
- [16] Y. C. Eldar and A. Nehorai, "Mean squared error beamforming for signal estimation: A competitive approach," in *Robust Adaptive Beamforming*, J. Li and P. Stoica, Eds. Hoboken, NJ: Wiley, to be published.
- [17] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation in linear models," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1931–1946, Jul. 2004.
- [18] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Linear minimax regret estimation with bounded data uncertainties," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2177–2188, Aug. 2004.
- [19] L. D. Davisson, "Universal noiseless coding," *IEEE Trans. Inf. Theory*, vol. IT-19, pp. 783–795, Nov. 1973.
- [20] M. Feder and N. Merhav, "Universal composite hypothesis testing: A competitive minimax approach," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1504–1517, Jun. 2002.
- [21] E. Levitan and N. Merhav, "A competitive Neyman-Pearson approach to universal hypothesis testing with applications," *IEEE Trans. Inf. Theory*, vol. 48, pp. 2215–2229, Aug. 2002.
- [22] N. Merhav and M. Feder, "Universal prediction," *IEEE Trans. Inf. Theory*, vol. 44, pp. 2124–2147, Oct. 1998.
- [23] A. Hyvarinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, pp. 411–430, 2000.
- [24] H. Attias, "Independent factor analysis," *Neural Comput.*, vol. 11, no. 4, pp. 803–851, 1999.
- [25] P. Comon, "Independent component analysis: A new concept?," *Signal Process.*, vol. 36, pp. 287–314, Apr. 1994.
- [26] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, Third ed. New York: McGraw-Hill, 1991.
- [27] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
- [28] S. A. Kassam, "Robust hypothesis testing for bounded classes of probability densities," *IEEE Trans. Inf. Theory*, vol. IT-27, pp. 242–247, 1980.
- [29] V. P. Kuznetsov, "Stable detection when the signal and spectrum of normal noise are inaccurately known," *Telecomm. Radio Eng.*, vol. 30/31, pp. 58–64, 1976. English translation.
- [30] A. K. Katsaggelos, K. T. Lay, and N. P. Galatsanos, "A general framework for frequency domain multi-channel signal processing," *IEEE Trans. Image Process.*, vol. 2, pp. 417–420, Jul. 1993.
- [31] H. Murakami, "Discrete wavelet transform based on cyclic convolution," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 165–174, Jan. 2004.
- [32] R. Gray, "Toeplitz and Circulant Matrices: A Review," Stanford Univ., Inform. Syst. Lab., Stanford, CA, Tech. Rep. 6504-1, 1977.
- [33] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Signal Process.*, vol. 27, pp. 113–120, 1979.
- [34] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 1979, pp. 208–211.
- [35] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*. Philadelphia, PA: SIAM, 2001, Series on Optimization.
- [36] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 40–95, Mar. 1996.
- [37] Y. Nesterov and A. Nemirovski, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA: SIAM, 1994.
- [38] F. Alizadeh, "Combinatorial Optimization with Interior Point Methods and Semi-Definite Matrices," Ph.D. dissertation, Univ. Minnesota, Minneapolis, MN, 1991.
- [39] M. X. Goemans and D. P. Williamson, "Approximation algorithms for MAX-3-CUT and other problems via complex semidefinite programming," in *ACM Symp. Theory Computing*, 2001, pp. 443–452.
- [40] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11–12, pp. 625–653, 1999.
- [41] D. Peaucelle, D. Henrion, and Y. Labit. Users Guide for SeDuMi Interface 1.03. [Online]. Available: <http://www.laas.fr/peaucell/SeDuMiInt.html>
- [42] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. Philadelphia, PA: SIAM, 1994.



Yonina C. Eldar (S'98–M'02) received the B.Sc. degree in physics in 1995 and the B.Sc. degree in electrical engineering in 1996, both from Tel-Aviv University (TAU), Tel-Aviv, Israel, and the Ph.D. degree in electrical engineering and computer science in 2001 from the Massachusetts Institute of Technology (MIT), Cambridge.

From January 2002 to July 2002, she was a Postdoctoral fellow at the Digital Signal Processing Group at MIT. She is currently a Senior Lecturer with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa, Israel. She is also a Research Affiliate with the Research Laboratory of Electronics at MIT. Her current research interests are in the general areas of signal processing, statistical signal processing, and quantum information theory.

Dr. Eldar was in the program for outstanding students at TAU from 1992 to 1996. In 1998, she held the Rosenblith Fellowship for study in Electrical Engineering at MIT, and in 2000, she held an IBM Research Fellowship. She is currently a Horev Fellow of the Leaders in Science and Technology program at the Technion as well as an Alon Fellow. In 2004, she received the Wolf Foundation Krill Prize. She is a member of the IEEE Signal Processing Theory and Methods technical committee.



Neri Merhav (S'86–M'87–SM'93–F'99) was born in Haifa, Israel, on March 16, 1957. He received the B.Sc., M.Sc., and D.Sc. degrees from the Technion—Israel Institute of Technology, Haifa, in 1982, 1985, and 1988, respectively, all in electrical engineering.

From 1988 to 1990 he was with AT&T Bell Laboratories, Murray Hill, NJ. Since 1990 he has been with the Electrical Engineering Department of the Technion, where he is now the Irving Shepard Professor. From 1994 to 2000, he was also serving

as a consultant to the Hewlett-Packard Laboratories—Israel (HPL-I). His research interests include information theory, statistical communications, and statistical signal processing. He is especially interested in the areas of lossless/lossy source coding and prediction/filtering, relationships between information theory and statistics, detection, estimation, and Shannon Theory, including topics in joint source-channel coding, source/channel simulation, and coding with side information with applications to information hiding and watermarking systems. He is currently on the Editorial Board of *Foundations and Trends in Communications and Information Theory*.

Dr. Merhav was a co-reipient of the 1993 Paper Award of the IEEE Information Theory Society. He also received the 1994 American Technion Society Award for Academic Excellence and the 2002 Technion Henry Taub Prize for Excellence in Research. From 1996 to 1999, he served as an Associate Editor for source coding of the IEEE TRANSACTIONS ON INFORMATION THEORY. He also served as a co-chairman of the Program Committee of the 2001 IEEE International Symposium on Information Theory.