

Robust Competitive Estimation With Signal and Noise Covariance Uncertainties

Yonina C. Eldar, *Member, IEEE*

Abstract—Robust estimation of a random vector in a linear model in the presence of model uncertainties has been studied in several recent works. While previous methods considered the case in which the uncertainty is in the signal covariance, and possibly the model matrix, but the noise covariance is assumed to be completely specified, here we extend the results to the case where the noise statistics may also be subjected to uncertainties. We propose several different approaches to robust estimation, which differ in their assumptions on the given statistics. In the first method, we assume that the model matrix and both the signal and the noise covariance matrices are uncertain, and develop a minimax mean-squared error (MSE) estimator that minimizes the worst case MSE in the region of uncertainty. The second strategy assumes that the model matrix is given and tries to uniformly approach the performance of the linear minimum MSE estimator that knows the signal and noise covariances by minimizing a worst case regret measure. The regret is defined as the difference or ratio between the MSE attainable using a linear estimator, ignorant of the signal and noise covariances, and the minimum MSE possible when the statistics are known. As we show, earlier solutions follow directly from our more general results. However, the approach taken here in developing the robust estimators is considerably simpler than previous methods.

Index Terms—Covariance uncertainty, linear estimation, minimax mean-squared error (MSE), minimax regret, robust estimation.

I. INTRODUCTION

ROBUST methods for dealing with model uncertainties have been applied to a variety of problems in communications, signal processing, and statistics (see, e.g., [1], [2] and references therein). It is well known that in many cases the performance of signal processing methods that are designed for a nominal model may deteriorate considerably if the actual model deviates from the one assumed.

Here, we develop methods for robust linear estimation in the presence of model uncertainties. Specifically, we consider a finite-dimensional analogue of the classical Wiener filtering problem [3], [4], in which the goal is to estimate a random vector \mathbf{x} that is observed through a linear transformation \mathbf{H} and corrupted by additive noise \mathbf{w} , where the signal and noise covariance matrices, as well as the model matrix \mathbf{H} , may not be known precisely.

When \mathbf{H} and the signal and noise covariance matrices are completely specified, the estimator minimizing the mean-squared error (MSE) is the minimum MSE (MMSE) estimator [5]. It is well known that the MMSE estimator is sensitive to the exact knowledge of the signal and noise statistics, as well as the model matrix. However, in many practical scenarios, the actual covariances and the model matrix may not be specified exactly. In such cases, it is desirable to design a robust estimator whose performance is reasonably good across all possible covariances and model matrices, in the region of uncertainty.

The minimax approach, initiated by Huber [6], [7], is the most common method for handling uncertainties. In this strategy, an estimator is designed to minimize the worst case MSE over a given uncertainty class [2], [8]–[12]. However, as demonstrated in the context of concrete examples in [13]–[15] and in Section VI, the minimax approach may lead to poor performance in situations other than the worst case.

To partially compensate for the conservative character of the minimax MSE method, a new competitive approach to robust estimation was developed in [13], [14]. In this strategy, we seek a linear estimator whose performance is uniformly close, in the region of uncertainty, to that of the optimal estimator that knows the exact model. Two competitive design criteria were proposed: In the first, developed in [13], a linear estimator was designed to minimize the worst case *difference regret*, which is the worst case difference between the MSE of a linear estimator, ignorant of the exact model, and the MSE of the optimal linear estimator based on complete knowledge of the model. The second approach [14] is based on minimizing the worst case *ratio regret*, which is the worst case ratio of the estimator's MSE and the optimal MSE. The rationale behind these strategies is that the resulting methods perform uniformly close to the linear optimal estimator across the uncertainty region, and since the minimax criterion is applied to the difference or ratio of MSEs, rather than the total MSE, it is not as pessimistic as the ordinary minimax approach. Similar competitive methods have also been applied in [15] for the case where the unknown desired vector is *deterministic* rather than stochastic. As noted in earlier work [13]–[15], the concept of competitive minimax methods is by no means new, and has been used extensively in a variety of other problem areas, such as universal source coding [16] and hypothesis testing [17], among others.

In this paper, we treat the same finite-dimensional model as that considered in [13], [14], however, in contrast to [13], [14] in which the noise covariance was completely specified, here we assume that the noise statistics may also be subjected to uncertainties. We develop a general framework for robust linear estimation under this model, which generalizes all of the pre-

Manuscript received January 30, 2004; revised December 25, 2005. This work was supported in part by the Israel Science Foundation under Grant 536/04.

The author is with the Technion–Israel Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: yonina@ee.technion.ac.il).

Communicated by A. Kavčić, Associate Editor for Detection and Estimation. Color version of Fig. 4 is available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2006.881749

vious results of [13], [14] to the case in which the noise covariance may not be known exactly. Although the problem we deal with here is more general than that discussed in [13], [14], the approach we take to developing the robust estimators is much simpler than previous methods.

Following the popular minimax approach, in Section III we consider the case in which \mathbf{H} and the signal and noise covariances are subjected to uncertainty, and seek the linear estimator that minimizes the worst case MSE over the uncertainty region. In our development, we treat both structured and unstructured uncertainty models. In the structured case, the minimax MSE estimator can be derived in closed form. The minimax MSE estimator in the unstructured setting is shown to be a solution to a semidefinite programming (SDP) problem [18]–[20].

In Sections IV and V, we develop the competitive minimax regret estimators. For analytical tractability, we restrict our attention to the case in which \mathbf{H} is completely specified, and the signal and noise covariance matrices obey a structured uncertainty model. In Section IV, we derive a closed-form solution for the minimax difference regret estimator, and show that it can be interpreted as an MMSE estimator corresponding to a specific choice of the unknown covariance matrices that depend on the uncertainty region. Besides adding further insight into the difference regret solution, this interpretation offers a general method for estimating the unknown covariance matrices, which may be useful in other contexts. In Section V, we develop the minimax ratio regret estimator, and present the optimal estimator under this criterion in two ways: The first representation is as a solution to a second-order cone program (SOCP), which is a convex optimization problem that can be solved very efficiently, e.g., using interior point methods [20]–[22]. The second is as a simpler SOCP together with a line search algorithm. Using this alternative description, we develop an explicit expression for the minimax ratio regret estimator for the case in which the signal covariance is known exactly.

The purpose of this paper is to extend the previous results of [13], [14] to the case in which the noise covariance is also subject to uncertainties, and to introduce a simpler method of proof, which, as we show, can also be used to derive the minimax MSE estimator for the structured case. Therefore, our main focus is on the theoretical derivations. In Section VI, we briefly discuss guidelines for choosing an estimator in a specific problem. A more detailed performance analysis can be found in [13], [14].

In Section II, we provide an overview of our problem before proceeding to the detailed development.

II. PROBLEM FORMULATION AND UNCERTAINTY MODELS

In the sequel, we denote vectors in \mathbb{C}^m by bold-face lower case letters and matrices in $\mathbb{C}^{n \times m}$ by bold-face upper case letters. The matrix \mathbf{I} is the identity matrix of appropriate dimension, $(\cdot)^*$ and $(\cdot)^\dagger$ are the Hermitian conjugate and the pseudo-inverse, respectively, of the corresponding matrix, and $\hat{(\cdot)}$ denotes an estimated vector or matrix. We use the notation $\text{diag}_m(\sigma_1, \dots, \sigma_m)$ for an $n \times m$ ($n \geq m$) diagonal matrix with diagonal elements σ_i ; $\text{diag}(\sigma_1, \dots, \sigma_m) = \text{diag}_m(\sigma_1, \dots, \sigma_m)$

denotes a square, size m diagonal matrix with diagonal elements σ_i . The covariance matrix of \mathbf{x} is written as \mathbf{C}_x . Finally, $\mathbf{A} \geq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive semidefinite.

We consider the generic linear estimation problem of estimating the random vector \mathbf{x} in the model

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

where \mathbf{H} is an $n \times m$ 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 covariance \mathbf{C}_w , uncorrelated with \mathbf{x} . Our objective is to design a linear estimator $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ of \mathbf{x} to minimize the MSE, which is given by

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

If \mathbf{C}_x and \mathbf{C}_w are known, then the linear estimator minimizing (2) is the MMSE (Wiener) estimator [5]

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

where we assume that $\mathbf{C}_y = \mathbf{H}\mathbf{C}_x\mathbf{H}^* + \mathbf{C}_w$ is invertible.

If \mathbf{C}_x , \mathbf{C}_w , or \mathbf{H} are not completely specified, then we cannot implement the MMSE estimator (3). To reflect our incomplete knowledge of \mathbf{C}_x , \mathbf{C}_w , and \mathbf{H} , we consider two different models of uncertainty which resemble the ‘‘band model’’ widely used in the continuous-time case [2], [9], [23], [24]. Depending on the optimality criteria, a particular model may be mathematically more convenient. As we detail further below, in the first model, we impose a particular structure on the eigenvectors of the matrices and therefore we refer to this case as structured uncertainty. In contrast, the second model does not assume any structure, and is consequently referred to as the unstructured uncertainty model.

Structured Uncertainty: In this model, we assume that \mathbf{C}_x and $\mathbf{H}^*\mathbf{H}$ have the same $m \times m$ eigenvector matrix \mathbf{V} , and that \mathbf{C}_w and $\mathbf{H}\mathbf{H}^*$ have the same $n \times n$ eigenvector matrix \mathbf{U} . Thus, \mathbf{H} , \mathbf{C}_x , and \mathbf{C}_w are constrained to have the form

$$\begin{aligned} \mathbf{H} &= \mathbf{U}\Sigma\mathbf{V}^* \\ \mathbf{C}_x &= \mathbf{V}\Delta\mathbf{V}^* \\ \mathbf{C}_w &= \mathbf{U}\Theta\mathbf{U}^* \end{aligned} \quad (4)$$

where $\Sigma = \text{diag}_n(\sigma_1, \dots, \sigma_m)$, $\Delta = \text{diag}(\delta_1, \dots, \delta_m)$, $\Theta = \text{diag}(\theta_1, \dots, \theta_n)$, and the diagonal values satisfy

$$\begin{aligned} \ell_i &\leq \sigma_i \leq \mu_i, & 1 \leq i \leq m \\ l_i &\leq \delta_i \leq u_i, & 1 \leq i \leq m \\ L_i &\leq \theta_i \leq U_i, & 1 \leq i \leq n \end{aligned} \quad (5)$$

where the bounds $\{\ell_i, \mu_i, l_i, u_i, L_i, U_i\}$ are known, $\ell_i, L_i > 0$, and $l_i \geq 0$. For brevity, we will denote by σ and δ the length- m vectors of components σ_i and δ_i , respectively, with $1 \leq i \leq m$, and by θ the length- n vector of components θ_i , $1 \leq i \leq n$. Note

that in deriving the minimax regret estimator we assume that \mathbf{H} is known so that $\ell_i = \mu_i$, $1 \leq i \leq m$.

The assumption (4) is made for analytical tractability. If \mathbf{x} and \mathbf{w} are stationary random vectors and \mathbf{H} represents convolution of \mathbf{x} with some filter, then \mathbf{H} , \mathbf{C}_x , and \mathbf{C}_w will be Toeplitz matrices and are therefore approximately diagonalized by a Fourier transform matrix [25], so that in this general case (4) is approximately satisfied. As we discuss further in Section IV, in many cases, our general approach can still be used even if (4) does not hold.

The model (4) is reasonable when the covariance matrices are estimated from the data. Specifically considering, for example, the eigenvalues of \mathbf{C}_x , and denoting $\zeta_i = (u_i + l_i)/2$, $\epsilon_i = (u_i - l_i)/2$ for $1 \leq i \leq m$, the conditions (5) can equivalently be expressed as

$$\delta_i = \zeta_i + e_i, \quad e_i^2 \leq \epsilon_i^2, \quad 1 \leq i \leq m \quad (6)$$

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 ζ_i . The same interpretation may be given to the eigenvalues of \mathbf{C}_w and the singular values of \mathbf{H} .

Unstructured Uncertainty: In this model

$$\begin{aligned} \mathbf{H} &= \tilde{\mathbf{H}} + \delta\mathbf{H}, & \|\delta\mathbf{H}\| &\leq \rho_h \\ \mathbf{C}_x &= \tilde{\mathbf{C}}_x + \delta\mathbf{C}_x, & \|\delta\mathbf{C}_x\| &\leq \rho_x \\ \mathbf{C}_w &= \tilde{\mathbf{C}}_w + \delta\mathbf{C}_w, & \|\delta\mathbf{C}_w\| &\leq \rho_w \end{aligned} \quad (7)$$

where $\tilde{\mathbf{H}}$, $\tilde{\mathbf{C}}_x$, and $\tilde{\mathbf{C}}_w$ are known, $\|\cdot\|$ denotes the spectral norm [26], i.e., the largest singular value of the corresponding matrix, and ρ_x and ρ_w are chosen such that $\tilde{\mathbf{C}}_x + \delta\mathbf{C}_x \geq 0$ for all $\|\delta\mathbf{C}_x\| \leq \rho_x$ and $\tilde{\mathbf{C}}_w + \delta\mathbf{C}_w \geq 0$ for all $\|\delta\mathbf{C}_w\| \leq \rho_w$. Here, the singular vectors of \mathbf{H} , \mathbf{C}_x , and \mathbf{C}_w are not constrained. As a consequence, we can no longer restrict each of the corresponding singular values but rather we bound the largest singular value, or equivalently, the spectral norm.

In Section III, we develop the linear minimax MSE estimator that minimizes the worst case MSE in the region of uncertainty, for both the structured and the unstructured models. The unstructured minimax MSE estimator was developed in [13] for the case in which \mathbf{C}_w is completely specified; the minimax MSE approach for the structured model was not previously treated. In Sections IV and V, we develop a competitive estimation approach in which we seek linear estimators that minimize a worst case regret. In this case, for analytical tractability, we consider only the structured uncertainty model and further assume that \mathbf{H} is known.

III. MINIMAX MSE ESTIMATOR

A. Structured Uncertainty

We begin by developing the minimax MSE estimator for the structured uncertainty model. Thus, we consider the problem

$$\begin{aligned} \min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\mathbf{H}, \mathbf{C}_x, \mathbf{C}_w \in \mathcal{S}} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) \\ = \min_{\mathbf{G}} \max_{\mathbf{H}, \mathbf{C}_x, \mathbf{C}_w \in \mathcal{S}} \mathcal{Q}(\mathbf{H}, \mathbf{C}_x, \mathbf{C}_w, \mathbf{G}) \end{aligned} \quad (8)$$

where from (2) the MSE is given by

$$\begin{aligned} \mathcal{Q}(\mathbf{H}, \mathbf{C}_x, \mathbf{C}_w, \mathbf{G}) \\ = \text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \text{Tr}(\mathbf{C}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})) \end{aligned} \quad (9)$$

and \mathcal{S} is the set of matrices defined by (4) and (5).

The minimax MSE estimator under this model is given by the following theorem.

Theorem 1: Let \mathbf{x} denote the unknown parameters in the model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where \mathbf{x} is a zero-mean random vector with covariance \mathbf{C}_x and \mathbf{w} is a zero-mean random vector with covariance \mathbf{C}_w , uncorrelated with \mathbf{x} . Let \mathcal{S} be the set of matrices \mathbf{H} , \mathbf{C}_x , \mathbf{C}_w such that $\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^*$, $\mathbf{C}_x = \mathbf{V}\Delta\mathbf{V}^*$, and $\mathbf{C}_w = \mathbf{U}\Theta\mathbf{U}^*$ where \mathbf{U} and \mathbf{V} are unitary matrices, $\Sigma = \text{diag}_n(\sigma_1, \dots, \sigma_m)$ with $0 < \ell_i \leq \sigma_i \leq \mu_i$, $\Delta = \text{diag}(\delta_1, \dots, \delta_m)$ with $l_i \leq \delta_i \leq u_i$, and $\Theta = \text{diag}(\theta_1, \dots, \theta_n)$ with $0 < L_i \leq \theta_i \leq U_i$. Then the solution to

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\mathbf{H}, \mathbf{C}_x, \mathbf{C}_w \in \mathcal{S}} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) \quad (10)$$

is

$$\hat{\mathbf{x}} = \mathbf{V}\mathbf{D}\mathbf{Z}^*\mathbf{U}^*\mathbf{y}$$

where \mathbf{Z} is the $n \times m$ matrix defined by $\mathbf{Z} = \text{diag}_n(1, \dots, 1)$ and $\mathbf{D} = \text{diag}(d_1, \dots, d_m)$ with

$$d_i = \begin{cases} \frac{2}{\ell_i + \mu_i}, & \mu_i - \ell_i \geq \frac{2U_i}{u_i \ell_i} \\ \frac{u_i \ell_i}{U_i + u_i \ell_i^2}, & \mu_i - \ell_i < \frac{2U_i}{u_i \ell_i}. \end{cases} \quad (11)$$

Proof: The proof of Theorem 1 is comprised of three parts. First, we show that the optimal \mathbf{G} minimizing the worst case MSE has the form $\mathbf{G} = \mathbf{V}\mathbf{D}\mathbf{Z}^*\mathbf{U}^*$ for some $m \times m$ matrix \mathbf{D} . We then prove that \mathbf{D} must be a diagonal matrix. Finally, we establish that the diagonal elements d_i of \mathbf{D} are given by (11).

Lemma 1: Let \mathbf{G} be the solution to (10). Then $\mathbf{G} = \mathbf{V}\mathbf{D}\mathbf{Z}^*\mathbf{U}^*$ for some $m \times m$ matrix \mathbf{D} .

Proof: We first note that the MSE $\mathcal{Q}(\mathbf{H}, \mathbf{C}_x, \mathbf{C}_w, \mathbf{G})$ of (9) 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}

$$\begin{aligned} \text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) &= \text{Tr}(\mathbf{G}\mathbf{U}\Theta\mathbf{U}^*\mathbf{G}^*) \geq \text{Tr}(\mathbf{G}\mathbf{U}\mathbf{Z}\Theta\mathbf{Z}^*\mathbf{U}^*\mathbf{G}^*) \\ &= \text{Tr}(\mathbf{G}\mathbf{P}\mathbf{C}_w\mathbf{P}^*) \end{aligned} \quad (12)$$

where

$$\mathbf{P} = \mathbf{U}\mathbf{Z}\mathbf{Z}^*\mathbf{U}^* \quad (13)$$

is the orthogonal projection onto the range space of \mathbf{H} . In addition, $\mathbf{G}\mathbf{H} = \mathbf{G}\mathbf{P}\mathbf{H}$ since $\mathbf{P}\mathbf{H} = \mathbf{H}$. Therefore,

$$\mathcal{Q}(\mathbf{H}, \mathbf{C}_x, \mathbf{C}_w, \mathbf{G}\mathbf{P}) \leq \mathcal{Q}(\mathbf{H}, \mathbf{C}_x, \mathbf{C}_w, \mathbf{G}).$$

Since $\mathcal{Q}(\mathbf{H}, \mathbf{C}_x, \mathbf{C}_w, \mathbf{G})$ is strictly convex in \mathbf{G} (because $\mathbf{C}_w > 0$ in the uncertainty region), the minimizing \mathbf{G} is unique, so that the optimal solution must satisfy

$$\mathbf{G} = \mathbf{G}\mathbf{P} = \mathbf{G}\mathbf{U}\mathbf{Z}\mathbf{Z}^*\mathbf{U}^* = \mathbf{B}\mathbf{Z}^*\mathbf{U}^* \quad (14)$$

for some $m \times m$ matrix \mathbf{B} . Denoting $\mathbf{B} = \mathbf{V}\mathbf{D}$, completes the proof. \square

Substituting $\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^*$, $\mathbf{C}_x = \mathbf{V}\Delta\mathbf{V}^*$, $\mathbf{C}_w = \mathbf{U}\Theta\mathbf{U}^*$, and \mathbf{G} of Lemma 1 into (9), we can express the MSE as

$$\begin{aligned} \mathcal{Q}(\mathbf{H}, \mathbf{C}_x, \mathbf{C}_w, \mathbf{G}) &= \text{Tr}(\mathbf{D}\mathbf{Z}^*\Theta\mathbf{Z}\mathbf{D}^*) + \text{Tr}(\Delta(\mathbf{I} - \mathbf{D}\mathbf{Z}^*\Sigma)^*(\mathbf{I} - \mathbf{D}\mathbf{Z}^*\Sigma)) \\ &\triangleq \mathcal{Q}(\Sigma, \Delta, \Theta, \mathbf{D}). \end{aligned} \quad (15)$$

Our problem then is to find \mathbf{D} that minimizes

$$\mathcal{G}(\mathbf{D}) = \max_{\sigma, \delta, \theta \in \mathcal{M}} \mathcal{Q}(\Sigma, \Delta, \Theta, \mathbf{D}) \quad (16)$$

where

$$\mathcal{M} = \{\sigma, \delta, \theta : \ell_i \leq \sigma_i \leq \mu_i, l_i \leq \delta_i \leq u_i, L_i \leq \theta_i \leq U_i\}. \quad (17)$$

Lemma 2: The matrix \mathbf{D} that minimizes $\mathcal{G}(\mathbf{D})$ of (16) is diagonal.

Proof: Since $\mathcal{Q}(\Sigma, \Delta, \Theta, \mathbf{D})$ is strictly convex in \mathbf{D} (because $\mathbf{Z}^*\Theta\mathbf{Z} > 0$ in the uncertainty region), so is $\mathcal{G}(\mathbf{D})$, and consequently, $\mathcal{G}(\mathbf{D})$ has a unique global minimizer. Now, let \mathbf{J} be any diagonal matrix with diagonal elements equal to ± 1 . Then using the facts that $\mathbf{J}^2 = \mathbf{I}$, for any diagonal matrix \mathbf{M} we have $\mathbf{J}\mathbf{M}\mathbf{J} = \mathbf{M}$, and $\mathbf{Z}^*\Sigma$ and $\mathbf{Z}^*\Theta\mathbf{Z}$ are diagonal matrices, we can immediately show that $\mathcal{Q}(\Sigma, \Delta, \Theta, \mathbf{J}\mathbf{D}\mathbf{J}) = \mathcal{Q}(\Sigma, \Delta, \Theta, \mathbf{D})$. Since $\mathcal{G}(\mathbf{D})$ has a unique minimizer, we conclude that the minimizing \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 diagonal. \square

Denoting by d_i the diagonal elements of \mathbf{D} , we can express $\mathcal{G}(\mathbf{D})$ as

$$\begin{aligned} \mathcal{G}(\mathbf{D}) &= \max_{\sigma, \delta, \theta \in \mathcal{M}} \sum_{i=1}^m (\theta_i d_i^2 + \delta_i (1 - d_i \sigma_i)^2) \\ &= \sum_{i=1}^m \left(\max_{L_i \leq \theta_i \leq U_i} \theta_i d_i^2 + \max_{l_i \leq \delta_i \leq u_i} \delta_i \max_{\ell_i \leq \sigma_i \leq \mu_i} (1 - d_i \sigma_i)^2 \right) \\ &= \sum_{i=1}^m \left(U_i d_i^2 + u_i \max_{\ell_i \leq \sigma_i \leq \mu_i} (1 - d_i \sigma_i)^2 \right). \end{aligned} \quad (18)$$

Therefore, minimizing $\mathcal{G}(\mathbf{D})$ reduces to solving m independent problems of the form

$$\min_d \left\{ U d^2 + u \max_{\ell \leq \sigma \leq \mu} (1 - d\sigma)^2 \right\} \quad (19)$$

where for brevity we omitted the index i .

Since $(1 - d\sigma)^2$ is convex in σ , the maximum of this function over a closed interval is obtained at one of the boundaries. Thus,

$$\max_{\ell \leq \sigma \leq \mu} (1 - d\sigma)^2 = \max((1 - d\ell)^2, (1 - d\mu)^2) \quad (20)$$

and (19) becomes

$$\min_d \{ U d^2 + u \max((1 - d\ell)^2, (1 - d\mu)^2) \} \quad (21)$$

which can equivalently be expressed as

$$\min_{t, d} \{ t : f_\ell(d) \leq t, f_\mu(d) \leq t \} \quad (22)$$

where we defined

$$f_x(d) = U d^2 + u(1 - dx)^2, \quad x = \ell, \mu. \quad (23)$$

To develop a solution to (22) we note that $f_x(d)$ is a quadratic function in d that obtains a minimum at

$$d_x = \frac{ux}{U + ux^2}. \quad (24)$$

Denoting

$$d_l = \min(d_\ell, d_\mu), \quad d_u = \max(d_\ell, d_\mu) \quad (25)$$

it then follows that both $f_\ell(d)$ and $f_\mu(d)$ are monotonically decreasing for $d < d_l$ and monotonically increasing for $d > d_u$ so that the optimal value of d , denoted d_0 , must satisfy $d_0 \in \mathcal{I}$ where $\mathcal{I} = [d_l, d_u]$. The exact value of d_0 will depend on the intersection points between $f_\ell(d)$ and $f_\mu(d)$. It is easy to see that $f_\ell(d)$ and $f_\mu(d)$ intersect at exactly two points: $d = 0$ and $d = d_e$ where

$$d_e = \frac{2}{\mu + \ell}. \quad (26)$$

For any value of the parameters u, U, ℓ and μ

$$d_e = \frac{2}{\ell + \mu} \geq \frac{1}{\mu} \geq \frac{u\mu}{U + u\mu^2} = d_\mu. \quad (27)$$

Therefore, there are three possibilities:

- 1) $d_\mu \leq d_e \leq d_\ell$,
- 2) $d_\mu \leq d_\ell \leq d_e$,
- 3) $d_\ell \leq d_\mu \leq d_e$.

In Fig. 1, we illustrate schematically the functions $f_\ell(d)$ and $f_\mu(d)$ for each of the three options. For the specific choices of $f_\ell(d)$ and $f_\mu(d)$ drawn in the figure, it can be seen that if $d_e \in \mathcal{I}$, as in Fig. 1(a), then the optimal value of d is $d_0 = d_e$. On the other hand, if $d_e \notin \mathcal{I}$, as in Fig. 1(b) and (c), then $d_0 = d_\ell$. The following lemma shows that these conclusions hold true for all parameter values.

Lemma 3: Let d_0 denote the optimal value of d in the problem

$$\min_{t, d} \{ t : f_\ell(d) \leq t, f_\mu(d) \leq t \}$$

where $f_\ell(d)$ and $f_\mu(d)$ are defined by (23). Then

$$d = \begin{cases} d_e, & d_e \in \mathcal{I} \\ d_\ell, & d_e \notin \mathcal{I} \end{cases} \quad (28)$$

where $\mathcal{I} = [d_l, d_u]$ with $d_l = \min(d_\ell, d_\mu)$, $d_u = \max(d_\ell, d_\mu)$, d_ℓ , and d_μ are given by (24), and $d_e = 2/(\mu + \ell)$.

Proof: See Appendix A. \square

To complete the proof of the theorem we need to prove that $d_e \in \mathcal{I}$ if and only if $\mu - \ell \geq 2U/(u\ell)$. Since $d_\mu \leq d_e$, it is sufficient to prove that $d_e \leq d_\ell$ if and only if $\mu - \ell \geq 2U/(u\ell)$. Now, if $\mu \geq 2U/(u\ell) + \ell$, then

$$d_e = \frac{2}{\mu + \ell} \leq \frac{u\ell}{U + u\ell^2} = d_\ell. \quad (29)$$

Similarly, if $\mu < 2U/(u\ell) + \ell$, then $d_e > u\ell/(U + u\ell^2) = d_\ell$, completing the proof of the theorem. \square

1) *Interpretation of the Minimax MSE Estimator:* We now provide some insight into the minimax MSE estimator of Theorem 1. Let $\Sigma = \text{diag}_n(\sigma_1, \dots, \sigma_m)$ with $\sigma_i = (1/2)(\mu_i + \ell_i)$,

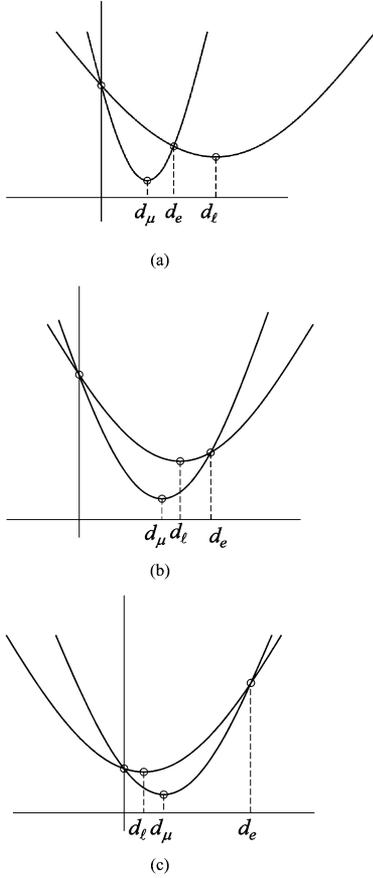


Fig. 1. Illustration of the functions $f_l(d)$ and $f_\mu(d)$ of (23) for different choices of d_μ, d_l and d_e . (a) $d_\mu \leq d_e \leq d_l$; (b) $d_\mu \leq d_l \leq d_e$; (c) $d_l \leq d_\mu \leq d_e$.

and let $\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^*$. Then, in the case $\mu_i - \ell_i \geq 2U_i/(u_i\ell_i)$ for all i , the minimax MSE estimator can be written as

$$\hat{\mathbf{x}} = \mathbf{V}\Sigma^\dagger\mathbf{U}^*\mathbf{y} = \mathbf{H}^\dagger\mathbf{y} \quad (30)$$

where $(\cdot)^\dagger$ denotes the pseudoinverse of the corresponding matrix. The estimator of (30) is just a least-squares estimator matched to the average singular values $\sigma_i = (1/2)(\mu_i + \ell_i)$. Defining the output signal-to-noise ratio (SNR) as $\gamma_i = u_i\ell_i/U_i$, it follows that for high enough SNR with respect to the reciprocal of the uncertainty region in \mathbf{H} , $\epsilon_i = 2/(\mu_i - \ell_i)$, the minimax MSE approach reduces to least-squares estimation matched to the average \mathbf{H} .

If $\mu_i - \ell_i < 2U_i/(u_i\ell_i)$ for all i , or equivalently, $\gamma_i < \epsilon_i$, then the minimax MSE solution is an MMSE estimator matched to $\sigma_i = \ell_i$, $\delta_i = u_i$, and $\theta_i = U_i$. To see this, we note that under the model (4), the estimator of (3) can be written as

$$\hat{\mathbf{x}} = \mathbf{V}\Delta\Sigma^*(\Sigma\Delta\Sigma^* + \Theta)^{-1}\mathbf{U}^*\mathbf{y} = \mathbf{V}\mathbf{Q}\mathbf{Z}^*\mathbf{U}^*\mathbf{y} \quad (31)$$

where $\mathbf{Q} = \text{diag}(q_1, \dots, q_m)$ with

$$q_i = \frac{\delta_i\sigma_i}{\theta_i + \delta_i\sigma_i^2}. \quad (32)$$

The result then follows from comparing (31) and (32) with the minimax MSE estimator of Theorem 1.

In particular, in the special case in which \mathbf{H} is known, $\ell_i = \mu_i$ and $\hat{\mathbf{x}}$ is an MMSE estimator matched to the worst case eigen-

values of the signal and noise covariance matrices: $\delta_i = u_i$ and $\theta_i = U_i$.

More generally, at each coordinate, the minimax MSE approach reduces to inverse filtering if the coordinate SNR is high with respect to the coordinate uncertainty in \mathbf{H} , or MMSE filtering matched to the worst case signal and noise values and the best case channel, at low SNR.

Note, that the values of $U_i, m+1 \leq i \leq n$ do not affect the solution of Theorem 1. This result is to be expected from the structure of the estimation problem. Indeed, under the model (4)

$$\tilde{\mathbf{y}} = \Sigma\tilde{\mathbf{x}} + \tilde{\mathbf{w}}, \quad (33)$$

where $\tilde{\mathbf{y}} = \mathbf{U}^*\mathbf{y}$, $\tilde{\mathbf{x}} = \mathbf{V}^*\mathbf{x}$, and $\tilde{\mathbf{w}} = \mathbf{U}^*\mathbf{w}$. Since the covariance of $\tilde{\mathbf{w}}$ is

$$\mathbf{C}_{\tilde{\mathbf{w}}} = \mathbf{U}^*\mathbf{C}_w\mathbf{U} = \Theta, \quad (34)$$

the components of $\tilde{\mathbf{w}}$ are uncorrelated. If $\tilde{\mathbf{G}}\tilde{\mathbf{y}}$ is an optimal estimate of $\tilde{\mathbf{x}}$ from $\tilde{\mathbf{y}}$, then $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ with $\mathbf{G} = \mathbf{V}\tilde{\mathbf{G}}\mathbf{U}^*$ is an optimal estimate of \mathbf{x} from \mathbf{y} . Now, for $m+1 \leq i \leq n$, we have from (33) that $\tilde{\mathbf{y}}_i = \tilde{\mathbf{w}}_i$, where z_i denotes the i th component of the vector \mathbf{z} . Since the parameter vector $\tilde{\mathbf{x}}$ is not present in $\tilde{\mathbf{y}}_i, m+1 \leq i \leq n$, and the components of $\tilde{\mathbf{w}}$ are uncorrelated, these observations are not useful in estimating $\tilde{\mathbf{x}}$, so that $\tilde{\mathbf{G}}$, and consequently, \mathbf{G} , will not be affected by the eigenvalues of \mathbf{C}_w corresponding to indices larger than m .

As a final comment, since the minimax MSE problem of (10) is not concave in the maximization argument \mathbf{H} , it is not a convex-concave problem, and therefore in general the solution does not satisfy a saddle-point property. Nonetheless, it is interesting to note that when $\mu_i - \ell_i < 2U_i/(u_i\ell_i)$ for all i , the minimax MSE estimator is also the solution to the max-min problem

$$\max_{\mathbf{H}, \mathbf{C}_x, \mathbf{C}_w} \min_{\mathbf{x} \in \mathcal{S}} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2). \quad (35)$$

Thus, in the regime, the minimax estimate is also a saddle-point solution.

To see this, recall that in the region considered, the minimax MSE method is an MMSE estimator matched to $\sigma_i = \ell_i$, $\delta_i = u_i$, and $\theta_i = U_i$ for $1 \leq i \leq m$. Now, the inner minimization problem in (35) is that of minimizing the MSE when $\mathbf{C}_x, \mathbf{C}_w$, and \mathbf{H} are known. The solution therefore is the MMSE estimator given by (3) and the resulting optimal MSE is [13]

$$\begin{aligned} \min_{\hat{\mathbf{x}} \in \mathcal{G}\mathbf{y}} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) \\ = \text{Tr}((\mathbf{I} - \mathbf{C}_x\mathbf{H}^*(\mathbf{H}\mathbf{C}_x\mathbf{H}^* + \mathbf{C}_w)^{-1}\mathbf{H})\mathbf{C}_x). \end{aligned} \quad (36)$$

Substituting $\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^*$, $\mathbf{C}_x = \mathbf{V}\Delta\mathbf{V}^*$, and $\mathbf{C}_w = \mathbf{U}\Theta\mathbf{U}^*$ into (36), the problem (35) becomes

$$\begin{aligned} \max_{\sigma_i, \delta_i, \theta_i \in \mathcal{M}} \text{Tr}((\mathbf{I} - \Delta\Sigma(\Sigma\Delta\Sigma^* + \Theta)^{-1}\Sigma)\Delta) \\ = \sum_{i=1}^m \max_{\sigma_i, \delta_i, \theta_i} \frac{\delta_i\theta_i}{\theta_i + \delta_i\sigma_i^2}. \end{aligned} \quad (37)$$

Since the objective in (37) is monotonically decreasing in each of the variables σ_i and monotonically increasing in each of the δ_i and θ_i , the maximum is obtained at $\sigma_i = \ell_i, \delta_i = u_i$, and

$\theta_i = U_i$ for $1 \leq i \leq m$. Therefore, the max-min solution of (35) is an MMSE estimator matched to these values.

B. Unstructured Uncertainty

We now treat the minimax MSE problem for the unstructured uncertainty model

$$\begin{aligned} & \min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\mathcal{S}} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) \\ & = \min_{\mathbf{G}} \max_{\mathcal{S}} \left\{ \mathcal{L}(\delta\mathbf{C}_x, \delta\mathbf{H}) + \text{Tr}(\mathbf{G}(\tilde{\mathbf{C}}_w + \delta\mathbf{C}_w)\mathbf{G}^*) \right\} \end{aligned} \quad (38)$$

where the maximization is over perturbation matrices satisfying $\delta\mathbf{H}, \delta\mathbf{C}_x, \delta\mathbf{C}_w \in \mathcal{S}$ with

$$\mathcal{S} = \left\{ \delta\mathbf{H}, \delta\mathbf{C}_x, \delta\mathbf{C}_w : \|\delta\mathbf{H}\| \leq \rho_h, \|\delta\mathbf{C}_x\| \leq \rho_x, \|\delta\mathbf{C}_w\| \leq \rho_w \right\} \quad (39)$$

and we defined

$$\begin{aligned} & \mathcal{L}(\delta\mathbf{C}_x, \delta\mathbf{H}) \\ & = \text{Tr}((\tilde{\mathbf{C}}_x + \delta\mathbf{C}_x)(\mathbf{I} - \mathbf{G}(\tilde{\mathbf{H}} + \delta\mathbf{H}))(\mathbf{I} - \mathbf{G}(\tilde{\mathbf{H}} + \delta\mathbf{H}))^*). \end{aligned} \quad (40)$$

To develop a solution to (38) we rely on the following lemma, the proof of which can be found, e.g., in [13].

Lemma 4: Let \mathbf{W}, \mathbf{T} and \mathbf{M} be nonnegative definite matrices with $\mathbf{W} \leq \mathbf{T}$. Then $\text{Tr}(\mathbf{M}\mathbf{W}) \leq \text{Tr}(\mathbf{M}\mathbf{T})$.

From Lemma 4

$$\max_{\|\delta\mathbf{C}_w\| \leq \rho_w} \text{Tr}(\mathbf{G}(\tilde{\mathbf{C}}_w + \delta\mathbf{C}_w)\mathbf{G}^*) = \text{Tr}(\mathbf{G}(\tilde{\mathbf{C}}_w + \rho_w\mathbf{I})\mathbf{G}^*). \quad (41)$$

Substituting (41) into (38), our problem becomes

$$\min_{\mathbf{G}} \left\{ \text{Tr}(\mathbf{G}(\mathbf{C}_w + \rho_w\mathbf{I})\mathbf{G}^*) + \max_{\delta\mathbf{H}, \delta\mathbf{C}_x \in \mathcal{S}} \mathcal{L}(\delta\mathbf{C}_x, \delta\mathbf{H}) \right\}. \quad (42)$$

The problem of (42) can be viewed as linear minimax MSE estimation in which the noise covariance is known and is given by $\mathbf{C}_w + \rho_w\mathbf{I}$, and \mathbf{H} and \mathbf{C}_x are subjected to uncertainty. This minimax problem was considered in [13, Sec. 4], in which it was shown that (42) can be formulated as a convex SDP [18]–[20], which is the problem of minimizing a linear functional subject to linear matrix inequalities. The main advantage of the SDP formulation is that it readily lends itself to efficient computational methods [19], [20], which are guaranteed to converge to the global optimum in polynomial time. Specifically, based on the results of [13], (42) is equivalent to the SDP

$$\min_{t, \mathbf{G}, \lambda, \mathbf{X}, \mathbf{Y}} t \quad (43)$$

subject to

$$\begin{aligned} & \text{Tr}((\tilde{\mathbf{C}}_x + \rho_x\mathbf{I})\mathbf{X}) + \text{Tr}(\mathbf{Y}) \leq t \\ & \begin{bmatrix} \mathbf{Y} & \mathbf{G}(\mathbf{C}_w + \rho_w\mathbf{I})^{1/2} \\ (\mathbf{C}_w + \rho_w\mathbf{I})^{1/2}\mathbf{G}^* & \mathbf{I} \end{bmatrix} \geq 0 \end{aligned}$$

$$\begin{bmatrix} \mathbf{X} - \lambda\mathbf{I} & (\mathbf{I} - \mathbf{G}\tilde{\mathbf{H}})^* & \mathbf{0} \\ \mathbf{I} - \mathbf{G}\tilde{\mathbf{H}} & \mathbf{I} & -\rho_h\mathbf{G} \\ \mathbf{0} & -\rho_h\mathbf{G}^* & \lambda\mathbf{I} \end{bmatrix} \geq 0. \quad (44)$$

Note that each of the matrices in (44) is indeed linear in the unknowns.

In the special case in which $\mathbf{H} = \tilde{\mathbf{H}}$ is known so that $\rho_h = 0$, using Lemma 4 it follows immediately that the solution to the inner maximization in (42) is $\mathbf{C}_x = \tilde{\mathbf{C}}_x + \rho_x\mathbf{I}$, and the minimax MSE method reduces to an MMSE estimator matched to $\mathbf{C}_x = \tilde{\mathbf{C}}_x + \rho_x\mathbf{I}$ and $\mathbf{C}_w = \tilde{\mathbf{C}}_w + \rho_w\mathbf{I}$.

IV. MINIMAX DIFFERENCE REGRET ESTIMATOR

To partially compensate for the conservative character of the minimax MSE approach, in the next two sections we develop competitive strategies in which we seek linear estimators with MSE performance uniformly close to that of the MMSE estimator for all possible values of \mathbf{C}_x and \mathbf{C}_w satisfying (4), where we assume that \mathbf{H} is completely specified. Thus, instead of choosing a linear method to minimize the worst case MSE, we now seek the linear estimator that minimizes the worst case regret. In this section, we treat the difference regret, and in Section V the ratio regret.

The difference regret $\mathcal{D}(\mathbf{C}_x, \mathbf{C}_w, \mathbf{G})$ is defined as the difference between the MSE of $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ and the smallest possible MSE attainable with an estimator of the form $\hat{\mathbf{x}} = \mathbf{G}(\mathbf{C}_x, \mathbf{C}_w)\mathbf{y}$ when the covariances \mathbf{C}_x and \mathbf{C}_w are known, which we denote by MSE^0 . If \mathbf{C}_x and \mathbf{C}_w are specified, then the MMSE estimator is given by (3) and the resulting optimal MSE is [13]

$$\text{MSE}^0 = \text{Tr}((\mathbf{I} - \mathbf{C}_x\mathbf{H}^*(\mathbf{H}\mathbf{C}_x\mathbf{H}^* + \mathbf{C}_w)^{-1}\mathbf{H})\mathbf{C}_x). \quad (45)$$

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

$$\min_{\mathbf{G}} \max_{\delta, \theta \in \mathcal{M}} \mathcal{D}(\mathbf{C}_x, \mathbf{C}_w, \mathbf{G}) \quad (46)$$

where \mathbf{C}_x and \mathbf{C}_w have eigendecompositions of the form (4),

$$\mathcal{M} = \{\delta, \theta : l_i \leq \delta_i \leq u_i, L_i \leq \theta_i \leq U_i\} \quad (47)$$

and

$$\begin{aligned} & \mathcal{D}(\mathbf{C}_x, \mathbf{C}_w, \mathbf{G}) \\ & = E(\|\mathbf{G}\mathbf{y} - \mathbf{x}\|^2) - \text{MSE}^0 \\ & = \text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \text{Tr}(\mathbf{C}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})) \\ & \quad - \text{Tr}((\mathbf{I} - \mathbf{C}_x\mathbf{H}^*(\mathbf{H}\mathbf{C}_x\mathbf{H}^* + \mathbf{C}_w)^{-1}\mathbf{H})\mathbf{C}_x). \end{aligned} \quad (48)$$

The minimax difference regret estimator is given by 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 , and \mathbf{w} is a zero-mean random vector with covariance \mathbf{C}_w , uncorrelated with \mathbf{x} . Let $\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^*$ where \mathbf{U} and \mathbf{V} are unitary matrices and $\Sigma = \text{diag}_n(\sigma_1, \dots, \sigma_m)$ with $\sigma_i > 0$, and let

$\mathbf{C}_x = \mathbf{V}\Delta\mathbf{V}^*$, $\mathbf{C}_w = \mathbf{U}\Theta\mathbf{U}^*$ where $\Delta = \text{diag}(\delta_1, \dots, \delta_m)$, $\Theta = \text{diag}(\theta_1, \dots, \theta_n)$, and

$$\mathcal{M} = \{\delta, \theta : l_i \leq \delta_i \leq u_i, 0 < L_i \leq \theta_i \leq U_i\}.$$

Then the solution to the problem

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\delta, \theta \in \mathcal{M}} \left\{ E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) - \min_{\hat{\mathbf{x}}=\mathbf{G}(\mathbf{C}_x, \mathbf{C}_w)\mathbf{y}} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) \right\} \quad (49)$$

is

$$\hat{\mathbf{x}} = \mathbf{V}\mathbf{D}\Sigma^\dagger\mathbf{U}^*\mathbf{y}$$

where $\mathbf{D} = \text{diag}(d_1, \dots, d_m)$ with

$$d_i = \frac{1}{\sqrt{U_i + l_i\sigma_i^2} + \sqrt{L_i + u_i\sigma_i^2}} \cdot \left(\frac{l_i\sigma_i^2}{\sqrt{U_i + l_i\sigma_i^2}} + \frac{u_i\sigma_i^2}{\sqrt{L_i + u_i\sigma_i^2}} \right). \quad (50)$$

Proof: As in the proof of Theorem 1, we first show that the \mathbf{G} minimizing the worst case regret has the form

$$\mathbf{G} = \mathbf{V}\mathbf{D}\Sigma^\dagger\mathbf{U}^* \quad (51)$$

for some $m \times m$ matrix \mathbf{D} . We then prove that \mathbf{D} must be a diagonal matrix with diagonal elements d_i given by (50).

To establish (51) we note that the regret $\mathcal{D}(\mathbf{C}_x, \mathbf{C}_w, \mathbf{G})$ of (48) depends on \mathbf{G} only through $\mathbf{G}\mathbf{H}$ and $\text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*)$. Therefore, using the same arguments as in the proof of Lemma 1

$$\mathcal{D}(\mathbf{C}_x, \mathbf{C}_w, \mathbf{G}\mathbf{P}) \leq \mathcal{D}(\mathbf{C}_x, \mathbf{C}_w, \mathbf{G})$$

where $\mathbf{P} = \mathbf{H}(\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*$ is an orthogonal projection onto the range of \mathbf{H} . Since the regret is strictly convex in \mathbf{G} (because $\mathbf{C}_w > 0$ in the uncertainty region), the optimal solution is unique and therefore must satisfy

$$\mathbf{G} = \mathbf{G}\mathbf{P} = \mathbf{G}\mathbf{H}(\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^* = \mathbf{B}\mathbf{V}\Sigma^\dagger\mathbf{U}^* \quad (52)$$

for some $m \times m$ matrix \mathbf{B} . Denoting $\mathbf{B} = \mathbf{V}\mathbf{D}\mathbf{V}^*$, (52) reduces to (51).

Substituting $\mathbf{C}_x = \mathbf{V}\Delta\mathbf{V}^*$, $\mathbf{C}_w = \mathbf{U}\Theta\mathbf{U}^*$ and \mathbf{G} of (51) into (48), we can express $\mathcal{D}(\mathbf{C}_x, \mathbf{C}_w, \mathbf{G})$ as

$$\begin{aligned} \mathcal{D} &= \text{Tr}(\mathbf{D}^*\mathbf{D}\Sigma^\dagger\Theta(\Sigma^\dagger)^*) + \text{Tr}(\Delta(\mathbf{I} - \mathbf{D})^*(\mathbf{I} - \mathbf{D})) \\ &\quad - \text{Tr}\left(\left(\mathbf{I} - \Delta\Sigma(\Sigma\Delta\Sigma^* + \Theta)^{-1}\Sigma\right)\Delta\right) \\ &\triangleq \mathcal{L}(\mathbf{D}, \Delta, \Theta). \end{aligned} \quad (53)$$

Our problem then reduces to finding \mathbf{D} that minimizes

$$\mathcal{G}(\mathbf{D}) = \max_{\delta, \theta \in \mathcal{M}} \mathcal{L}(\mathbf{D}, \Delta, \Theta). \quad (54)$$

Since $\mathcal{L}(\mathbf{D}, \Delta, \Theta)$ is strictly convex in \mathbf{D} , $\mathcal{G}(\mathbf{D})$ has a unique global minimum. Now, as in the proof of Lemma 2 we can show that $\mathcal{L}(\mathbf{J}\mathbf{D}\mathbf{J}, \Delta, \Theta) = \mathcal{L}(\mathbf{D}, \Delta, \Theta)$ for any diagonal matrix \mathbf{J} with diagonal elements equal to ± 1 , from which we conclude

that the optimal \mathbf{D} satisfies $\mathbf{D} = \mathbf{J}\mathbf{D}\mathbf{J}$ for any such \mathbf{J} , which in turn implies that \mathbf{D} must be a diagonal matrix.

Denoting by d_i the diagonal elements of \mathbf{D} , we can express $\mathcal{G}(\mathbf{D})$ as

$$\mathcal{G}(\mathbf{D}) = \sum_{i=1}^m \max_{\delta, \theta \in \mathcal{M}} \left\{ \frac{\theta_i d_i^2}{\sigma_i^2} + \delta_i(1 - d_i)^2 - \frac{\delta_i \theta_i}{\theta_i + \delta_i \sigma_i^2} \right\}. \quad (55)$$

Thus, we need to solve m independent problems of the form

$$\min_d \mathcal{T}(d) \quad (56)$$

where

$$\mathcal{T}(d) = \max_{l \leq \delta \leq u, L \leq \theta \leq U} \left\{ \frac{\theta d^2}{\sigma^2} + \delta(1 - d)^2 - \frac{\delta \theta}{\theta + \delta \sigma^2} \right\} \quad (57)$$

and for brevity we omitted the index i . To develop an explicit expression for $\mathcal{T}(d)$ we note that the function

$$h(x) = ax - \frac{bx}{c + dx} \quad (58)$$

with $b, c, d \geq 0$ is convex¹ in $x \geq 0$. It follows that for fixed d and σ

$$f(\theta, \delta) = \frac{\theta d^2}{\sigma^2} + \delta(1 - d)^2 - \frac{\delta \theta}{\theta + \delta \sigma^2} \quad (59)$$

is convex in δ for fixed θ , and convex in θ for fixed δ . Consequently, the maximum of $f(\theta, \delta)$ over a closed interval of δ or θ is obtained at one of the boundaries. Thus,

$$\begin{aligned} \mathcal{T}(d) &= \max_{L \leq \theta \leq U} \max_{l \leq \delta \leq u} f(\theta, \delta) \\ &= \max_{L \leq \theta \leq U} \max(f(\theta, l), f(\theta, u)) \\ &= \max(f(L, l), f(L, u), f(U, l), f(U, u)). \end{aligned} \quad (60)$$

Note that $f(L, l)$ is a function of d , and can be expressed as

$$f_{Ll}(d) = \left(\frac{\sqrt{L + l\sigma^2}}{\sigma} d - \frac{l\sigma}{\sqrt{L + l\sigma^2}} \right)^2. \quad (61)$$

Defining in a similar way the functions $f_{Lu}(d)$, $f_{Ul}(d)$, and $f_{Uu}(d)$, the problem (56) can be formulated as

$$\min_{t, d} \{t : f_{Ll}(d) \leq t, f_{Lu}(d) \leq t, f_{Ul}(d) \leq t, f_{Uu}(d) \leq t\}. \quad (62)$$

We now rely on the following lemma.

Lemma 5: The optimal value d_0 of (62) is given by $d_0 = d_e$ where

$$d_e = \frac{1}{\sqrt{U + l\sigma^2} + \sqrt{L + u\sigma^2}} \left(\frac{l\sigma^2}{\sqrt{U + l\sigma^2}} + \frac{u\sigma^2}{\sqrt{L + u\sigma^2}} \right). \quad (63)$$

The proof of Lemma 5 is given in Appendix C. Below, we provide some intuition into the optimality of d_e .

¹Indeed, the second derivative is given by $2bd/(c + dx)^3$ which is nonnegative for $x \geq 0$.

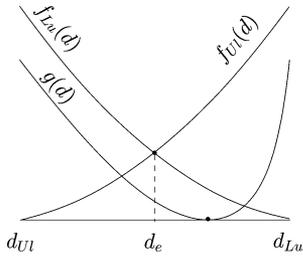


Fig. 2. Illustration of the optimality of d_e when $g(d) \leq \max(f_{Lu}(d), f_{Ul}(d))$ for all $d \in \mathcal{I}$.

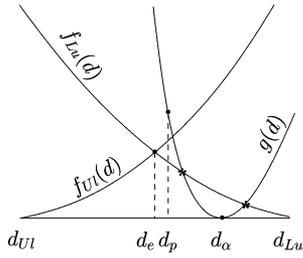


Fig. 3. Illustration of the fact that if $g(d_p) > \max(f_{Lu}(d_p), f_{Ul}(d_p))$ for some $d_p < d_\alpha$, then $g(d)$ intersects $f_{Lu}(d)$ at two points.

We first note that the optimal solution d_0 must satisfy $d_{Ul} \leq d_0 \leq d_{Lu}$, where

$$d_{Ul} = \frac{l\sigma^2}{U + l\sigma^2}, \quad d_{Lu} = \frac{u\sigma^2}{L + u\sigma^2} \quad (64)$$

are the zeros of f_{Ul} and f_{Lu} , respectively. This follows from the fact that the functions $f_{Ll}(d)$, $f_{Lu}(d)$, $f_{Ul}(d)$, $f_{Uu}(d)$ are monotonically increasing for $d > d_{Lu}$ and monotonically decreasing for $d < d_{Ul}$. Now, the point d_e is the unique intersection point in the interval $\mathcal{I} = [d_{Ul}, d_{Lu}]$ of the functions $f_{Lu}(d)$ and $f_{Ul}(d)$. To show that this intersection is optimal it suffices to prove that

$$g(d) \leq \max(f_{Lu}(d), f_{Ul}(d)), \quad d \in \mathcal{I} \quad (65)$$

where $g(d)$ is equal to $f_{Uu}(d)$ or $f_{Ll}(d)$, as illustrated in Fig. 2.

In Appendix B, we show that $g(d)$ intersects $f_{Lu}(d)$ and $f_{Ul}(d)$ at exactly one point in \mathcal{I} . Let d_α denote the value at which $g(d_\alpha) = 0$. It is easy to see that $d_\alpha \in \mathcal{I}$. Fig. 3 illustrates the fact that if $g(d_p) > \max(f_{Lu}(d_p), f_{Ul}(d_p))$ for some $d_p < d_\alpha$, then $g(d)$ intersects $f_{Lu}(d)$ at two points in \mathcal{I} which is a contradiction to the fact that there is only one intersection point. Similarly, it can be shown that if $d_p > d_\alpha$, then $g(d)$ intersects $f_{Ul}(d)$ at two points in \mathcal{I} , so that (65) must be true. \square

Before discussing an interpretation of the minimax regret method of Theorem 2 we note that this estimator does not satisfy a saddle-point property. This follows from the fact that the optimal objective value of the max-min problem which results from interchanging the order of the minimization and the maximization in (49), is 0. Indeed, the solution to the inner minimization is the MMSE estimator of (3) which results in a zero objective value.

A. MMSE Interpretation

The following corollary of Theorem 2 shows that the difference regret strategy can be interpreted as an MMSE estimator.

Corollary 1: Let \mathbf{x} denote the unknown parameters in the model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$. Then, under the assumptions of Theorem 2, the solution to the problem

$$\min_{\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}} \max_{\delta, \theta \in \mathcal{M}} \left\{ E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) - \min_{\hat{\mathbf{x}} = \mathbf{G}(\mathbf{C}_x, \mathbf{C}_w)\mathbf{y}} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) \right\}$$

is an MMSE estimator matched to the covariance matrices $\mathbf{C}_x = \mathbf{V}\mathbf{X}\mathbf{V}^*$, $\mathbf{C}_w = \mathbf{U}\mathbf{W}\mathbf{U}^*$, where $\mathbf{X} = \text{diag}(x_1, \dots, x_m)$, $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$ with

$$\begin{aligned} x_i &= \alpha_i l_i + (1 - \alpha_i) u_i, & 1 \leq i \leq m \\ w_i &= \alpha_i L_i + (1 - \alpha_i) U_i, & 1 \leq i \leq m \end{aligned} \quad (66)$$

where

$$\alpha_i = \frac{\sqrt{L_i + u_i \sigma_i^2}}{\sqrt{L_i + u_i \sigma_i^2} + \sqrt{U_i + l_i \sigma_i^2}} \quad (67)$$

and $w_i > 0$, $m + 1 \leq i \leq n$ are arbitrary.

Proof: From (31) and (32) it follows that the MMSE estimate of \mathbf{x} with $\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^*$ and $\mathbf{C}_x, \mathbf{C}_w$ given by Corollary 1 is $\hat{\mathbf{x}} = \mathbf{V}\mathbf{Q}\Sigma^\dagger\mathbf{U}^*\mathbf{y}$, where $\mathbf{Q} = \text{diag}(q_1, \dots, q_m)$ with

$$\begin{aligned} q_i &= \frac{x_i \sigma_i^2}{x_i \sigma_i^2 + w_i} \\ &= \frac{l_i \sigma_i^2 \sqrt{L_i + u_i \sigma_i^2} + u_i \sigma_i^2 \sqrt{U_i + l_i \sigma_i^2}}{(U_i + l_i \sigma_i^2) \sqrt{L_i + u_i \sigma_i^2} + (L_i + u_i \sigma_i^2) \sqrt{U_i + l_i \sigma_i^2}} \\ &= d_i \end{aligned} \quad (68)$$

where d_i is given by (50). \square

Note that if $l_i = u_i$, so that the i th eigenvalue of the true covariance of \mathbf{C}_x is equal to u_i then, as we expect, $x_i = u_i$. Similarly, if $L_i = U_i$, so that the i th eigenvalue of the true covariance of \mathbf{C}_w is equal to U_i , then $w_i = U_i$. When the output SNR is high so that $l_i \sigma_i^2 \gg U_i$

$$x_i \approx \frac{l_i \sqrt{u_i \sigma_i^2} + u_i \sqrt{l_i \sigma_i^2}}{\sqrt{u_i \sigma_i^2} + \sqrt{l_i \sigma_i^2}} = \sqrt{u_i l_i}, \quad (69)$$

and is approximately the geometric average of the bounds. Similarly, when $L_i \gg u_i \sigma_i^2$, then

$$w_i \approx \sqrt{U_i L_i}. \quad (70)$$

Since the minimax difference regret estimator minimizes the regret for the covariance matrices given by Corollary 1, we may view these matrices as the ‘‘least favorable’’ in the regret sense. These covariance matrices can also be viewed as estimators of the true, unknown covariance. Specifically, considering for example \mathbf{C}_x , each of the unknown eigenvalues of \mathbf{C}_x is estimated as a weighted combination of the bounds u_i and l_i , where the

weights depend explicitly on the uncertainty level of the corresponding eigenvalue of the signal and noise covariances, and on the matching singular value of \mathbf{H} .

The interpretation of Corollary 1 can also be used to implement the estimator of Theorem 2 in the case in which the commuting assumptions of (4) do not hold. Specifically, we can choose an estimator of the form

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

where \mathbf{X} and \mathbf{W} are given by Corollary 1, and \mathbf{V} and \mathbf{U} are the eigenvector matrices of \mathbf{C}_x and \mathbf{C}_w , respectively. This estimator can be implemented whether or not (4) is satisfied.

B. Difference Regret Estimator for Known \mathbf{C}_x

We now consider the difference regret estimator in the special case in which \mathbf{C}_x is completely specified. The scenario in which \mathbf{C}_w is known has been previously treated in [13], and also follows immediately from Theorem 2.

Suppose that $u_i = l_i = \delta_i$, $1 \leq i \leq m$. From Theorem 2

$$d_i = \frac{\delta_i \sigma_i^2}{\sqrt{L_i + \delta_i \sigma_i^2} \sqrt{U_i + \delta_i \sigma_i^2}}. \quad (72)$$

Using Corollary 1, the difference regret approach is equivalent to an MMSE estimator matched to

$$\mathbf{C}_w = \mathbf{U} \text{diag}(w_1, \dots, w_n) \mathbf{U}^*$$

where

$$w_i = \sqrt{L_i + \delta_i \sigma_i^2} \sqrt{U_i + \delta_i \sigma_i^2} - \delta_i \sigma_i^2, \quad 1 \leq i \leq m \quad (73)$$

and $w_i > 0$, $m + 1 \leq i \leq n$ are arbitrary. Since $\sqrt{ab} \leq (a + b)/2$, we have that $w_i \leq (U_i + L_i)/2$. In contrast, the minimax MSE estimator of Theorem 1 in this case is matched to an MMSE estimator with noise covariance eigenvalues $U_i > w_i$.

In the high- and low-SNR regimes, we have the approximations

$$w_i = \begin{cases} \frac{1}{2}(U_i + L_i), & U_i \ll \delta_i \sigma_i^2 \\ \sqrt{L_i U_i}, & L_i \gg \delta_i \sigma_i^2. \end{cases} \quad (74)$$

Thus, in these regions, w_i is an average of the boundary values of the eigenvalues of \mathbf{C}_w , where for high SNR, w_i is equal to the arithmetic average, and for low SNR, w_i is equal to the geometric average.

V. MINIMAX RATIO REGRET ESTIMATOR

We now treat the problem of minimizing the worst case ratio regret, which was suggested in [14] as an alternative to the minimax difference regret approach.

The ratio regret $\mathcal{R}(\mathbf{C}_x, \mathbf{C}_w, \mathbf{G})$ 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 $\hat{\mathbf{x}} = \mathbf{G}(\mathbf{C}_x, \mathbf{C}_w) \mathbf{y}$

when the covariance matrices \mathbf{C}_x and \mathbf{C}_w are known. Thus, our problem is

$$\min_{\mathbf{G}} \max_{\delta, \theta \in \mathcal{M}} \mathcal{R}(\mathbf{C}_x, \mathbf{C}_w, \mathbf{G}) \quad (75)$$

where

$$\begin{aligned}\mathcal{R}(\mathbf{C}_x, \mathbf{C}_w, \mathbf{G}) &= \frac{E(\|\mathbf{G} \mathbf{y} - \mathbf{x}\|^2)}{\text{MSE}^0} \\ &= \frac{\text{Tr}(\mathbf{G} \mathbf{C}_w \mathbf{G}^*) + \text{Tr}(\mathbf{C}_x (\mathbf{I} - \mathbf{G} \mathbf{H})^* (\mathbf{I} - \mathbf{G} \mathbf{H}))}{\text{Tr}((\mathbf{I} - \mathbf{C}_x \mathbf{H}^* (\mathbf{H} \mathbf{C}_x \mathbf{H}^* + \mathbf{C}_w)^{-1} \mathbf{H}) \mathbf{C}_x)}.\end{aligned}\quad (76)$$

To ensure that $\text{MSE}^0 \neq 0$ in the uncertainty region, we assume that $l_i \neq 0$ for at least on value of i .

In Theorem 3 below, we show that the problem of (75) can be formulated as an SOCP. Note, that SOCPs can be solved very efficiently in polynomial time using standard software packages, for example, the Self-Dual-Minimization (SeDuMi) package [21], [22] for Matlab.

Theorem 3: Consider the setting of Theorem 2. Then the solution to the problem

$$\min_{\hat{\mathbf{x}} = \mathbf{G} \mathbf{y}} \max_{\delta, \theta \in \mathcal{M}} \left\{ \frac{E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)}{\min_{\hat{\mathbf{x}} = \mathbf{G}(\mathbf{C}_x, \mathbf{C}_w) \mathbf{y}} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)} \right\}$$

is

$$\hat{\mathbf{x}} = \mathbf{V} \mathbf{D} \Sigma^\dagger \mathbf{U}^* \mathbf{y}$$

where $\mathbf{D} = \text{diag}(d_1, \dots, d_m)$ and $\{d_i\}$ are the solutions to the second-order cone program

$$\min_{\gamma, t_i, d_i} \gamma \quad (77)$$

subject to

$$\begin{aligned}f_{Ll}(d_i, \gamma) &\leq t_i, & 1 \leq i \leq m \\ f_{Lu}(d_i, \gamma) &\leq t_i, & 1 \leq i \leq m \\ f_{Ul}(d_i, \gamma) &\leq t_i, & 1 \leq i \leq m \\ f_{Uu}(d_i, \gamma) &\leq t_i, & 1 \leq i \leq m \\ \sum_{i=1}^m t_i &\leq 0\end{aligned}\quad (78)$$

with

$$\begin{aligned}f_{xy}(d_i, \gamma) &= \left(\frac{\sqrt{x + y \sigma_i^2}}{\sigma_i} d_i - \frac{y \sigma_i}{\sqrt{x + y \sigma_i^2}} \right)^2 \\ &\quad - (\gamma - 1) \frac{xy}{x + y \sigma_i^2}.\end{aligned}\quad (79)$$

Proof: We begin by noting that any problem of the form

$$\min_y \max_{x \in \mathcal{X}} \frac{f(x, y)}{g(x, y)} \quad (80)$$

is equivalent to

$$\min_{x, y} \{ \gamma : f(x, y) - \gamma g(x, y) \leq 0, \forall x \in \mathcal{X} \} \quad (81)$$

which, in turn, can be written as

$$\min_{x,y} \left\{ \gamma : \max_{x \in \mathcal{X}} \{f(x,y) - \gamma g(x,y)\} \leq 0 \right\}. \quad (82)$$

Therefore, (75) can be expressed as

$$(\Gamma) : \min_{\gamma, \mathbf{G}} \left\{ \gamma : \max_{\delta, \theta \in \mathcal{M}} \mathcal{L}(\gamma, \mathbf{G}, \mathbf{C}_x, \mathbf{C}_w) \leq 0 \right\} \quad (83)$$

where we defined

$$\begin{aligned} \mathcal{L}(\gamma, \mathbf{G}, \mathbf{C}_x, \mathbf{C}_w) &= \text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \text{Tr}(\mathbf{C}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})) \\ &\quad - \gamma \text{Tr}(\mathbf{I} - \mathbf{C}_x\mathbf{H}^*(\mathbf{H}\mathbf{C}_x\mathbf{H}^* + \mathbf{C}_w)^{-1}\mathbf{H})\mathbf{C}_x). \end{aligned} \quad (84)$$

Next, we note that the optimal \mathbf{G} has the form $\mathbf{G} = \mathbf{V}\mathbf{D}\Sigma^\dagger\mathbf{U}^*$, for some $\mathbf{D} = \text{diag}(d_1, \dots, d_m)$. This follows using a similar argument to that used in the proof of Theorem 2. The problem of (83) then reduces to finding \mathbf{D} that minimizes γ subject to $\mathcal{G}(\mathbf{D}) \leq 0$ where

$$\begin{aligned} \mathcal{G}(\mathbf{D}) &= \max_{\delta, \theta \in \mathcal{M}} \sum_{i=1}^m \left(\frac{\theta_i d_i^2}{\sigma_i^2} + \delta_i(1 - d_i)^2 - \gamma \frac{\delta_i \theta_i}{\theta_i + \sigma_i^2 \delta_i} \right) \\ &= \sum_{i=1}^m \mathcal{T}(d_i) \end{aligned} \quad (85)$$

with

$$\mathcal{T}(d_i) = \max_{\delta, \theta \in \mathcal{M}} \left\{ \frac{\theta_i d_i^2}{\sigma_i^2} + \delta_i(1 - d_i)^2 - \gamma \frac{\delta_i \theta_i}{\theta_i + \delta_i \sigma_i^2} \right\}. \quad (86)$$

Thus, our problem can be formulated as

$$\min_{\gamma, t_i, d_i} \left\{ \gamma : \mathcal{T}(d_i) \leq t_i, 1 \leq i \leq m, \sum_{i=1}^m t_i \leq 0 \right\}. \quad (87)$$

Now, as in the proof of Theorem 2

$$h(\theta, \delta) = \frac{\theta d^2}{\sigma^2} + \delta(1 - d)^2 - \gamma \frac{\delta \theta}{\theta + \delta \sigma^2} \quad (88)$$

is convex in δ for fixed $\theta, \gamma \geq 0$, and is convex in θ for fixed $\delta, \gamma \geq 0$, where for brevity we omitted the index i . Consequently

$$\mathcal{T}(d) = \max(h(L, l), h(L, u), h(U, l), h(U, u)). \quad (89)$$

The proof then follows from noting that $h(L, l) = f_{Ll}(d_i, \gamma)$. \square

A. Alternative Derivation of the Minimax Ratio Estimator

In Theorem 3, we showed that the minimax ratio regret estimator can be found by solving an SOCP in $2m + 1$ unknowns. In this subsection, we provide further insight into the estimator, by developing an alternative formulation.

Specifically, the minimax ratio estimator, that is the solution to (83), can be determined by first solving the problem

$$(\Phi) : \min_{t, \mathbf{G}} \left\{ t : \max_{\delta, \theta \in \mathcal{M}} \mathcal{L}(\gamma, \mathbf{G}, \mathbf{C}_x, \mathbf{C}_w) \leq t \right\} \quad (90)$$

where $\gamma \geq 1$ is fixed. Let $\hat{t}(\gamma)$ denote the optimal value of t in the problem (Φ) of (90), and let $\hat{\gamma}$ be the unique value of

$\gamma \geq 1$ such that $\hat{t}(\gamma) = 0$ (as we show below in Proposition 1, such a γ always exists, and is unique). Then, denoting by $\hat{\mathbf{G}}$ the optimal value of (Φ) with $\gamma = \hat{\gamma}$, we now show that $\hat{\mathbf{G}}$ and $\hat{\gamma}$ are the solutions to (Γ) of (83): 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 exist 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 *unique* value for which $\hat{t}(\gamma) = 0$. Therefore, to solve (Γ) we may first solve the simpler problem (Φ) , and then find $\hat{\gamma}$ by a line search algorithm using, for example, bisection. Specifically, we may start by choosing $\gamma = 1$. For each choice of γ we compute $\hat{t}(\gamma)$. If $\hat{t}(\gamma) > 0$, then we increase γ , and if $\hat{t}(\gamma) < 0$, then we decrease γ , continuing until $\hat{t}(\gamma) = 0$. Due to the continuity and monotonicity properties of $\hat{t}(\gamma)$, established in Proposition 1 below, the algorithm is guaranteed to converge.

Proposition 1: Let $\hat{t}(\gamma)$ denote the optimal value of t in the problem (Φ) of (90). Then

- 1) $\hat{t}(\gamma)$ is continuous in γ ;
- 2) $\hat{t}(\gamma)$ is strictly decreasing in γ ;
- 3) there is a unique value of γ for which $\hat{t}(\gamma) = 0$.

Proof: See Appendix D. \square

From Proposition 1, we conclude that instead of solving (Γ) of (83), we may solve (Φ) of (90), which in some cases may provide more insight. From the proof of Theorem 2 it follows that the solution to (Φ) is given by $\mathbf{G} = \mathbf{V}\mathbf{D}\Sigma^\dagger\mathbf{U}^*$, where \mathbf{D} is a diagonal matrix with diagonal elements d_i that solve the problem

$$\min_{t_i, d_i} \left\{ t_i : f_{Ll}(d_i) \leq t_i, f_{Lu}(d_i) \leq t_i, \right. \\ \left. f_{Ul}(d_i) \leq t_i, f_{Uu}(d_i) \leq t_i \right\} \quad (91)$$

where $f_{xy}(d_i)$ is given by (79). The optimal value $\hat{t}(\gamma)$ is then $\hat{t}(\gamma) = \sum_{i=1}^m t_i$.

To illustrate the possible advantage of this approach, in the next subsection we treat the case in which \mathbf{C}_x is known completely and \mathbf{C}_w is subject to uncertainty, and show that this method leads to new insight into the optimal solution. The minimax ratio estimator for the case in which \mathbf{C}_w is known and \mathbf{C}_x is uncertain is developed in [14], and can also be found by following similar steps as those in Section V-B.

B. Alternative Derivation for Known Signal Covariance

Suppose that \mathbf{C}_x is known, so that $u_i = l_i = \delta_i, 1 \leq i \leq m$. In this case the problem of (91) reduces to

$$\min_{t, d} \{ t : f_L(d) \leq t, f_U(t) \leq t \} \quad (92)$$

where for brevity we omitted the index i , and

$$f_x(d) = \left(\frac{\sqrt{x + \delta \sigma^2}}{\sigma} d - \frac{\delta \sigma}{\sqrt{x + \delta \sigma^2}} \right)^2 - (\gamma - 1) \frac{\delta x}{x + \delta \sigma^2} \quad (93)$$

with $x = L, U$. Using the same arguments as those used in the proof of Lemma 5 it follows that the optimal d , denoted d_0 , satisfies

$$d_U \leq d_0 \leq d_L \quad (94)$$

where

$$d_L = \frac{\delta\sigma^2}{L + \delta\sigma^2}, \quad d_U = \frac{\delta\sigma^2}{U + \delta\sigma^2}. \quad (95)$$

In addition, if there is a unique point $d_e \in \mathcal{I} = [d_L, d_U]$ such that $f_U(d_e) = f_L(d_e)$, then d_e is optimal.

If $\delta = 0$, then from (94) and (95) we have that $d_0 = 0$. Similarly, if $L = U$, then from (94) and (95), $d_0 = d_L = d_U$. Next assume that $\delta > 0$ and $L \neq U$. Then $f_U(d) = f_L(d)$ at $d = d_{\pm}$ where

$$d_{\pm} = \pm\sqrt{\gamma} \frac{\delta\sigma^2}{\sqrt{U + \delta\sigma^2}\sqrt{L + \delta\sigma^2}}. \quad (96)$$

Since $d_- < 0$, $d_- \notin \mathcal{I}$. From the fact that $\sqrt{\gamma} \geq 1$, we can show that $d_+ \geq d_U$. Thus, $d_+ \in \mathcal{I}$ as long as $\sqrt{\gamma}$ is such that

$$d_+ = \sqrt{\gamma} \frac{\delta\sigma^2}{\sqrt{U + \delta\sigma^2}\sqrt{L + \delta\sigma^2}} \leq d_L \quad (97)$$

or $\gamma \leq \gamma_0$ where

$$\gamma_0 = \frac{U + \delta\sigma^2}{L + \delta\sigma^2}. \quad (98)$$

If $\gamma > \gamma_0$, then $d_0 = d_L$. To see this, we first note that if $f_L(d) > f_U(d)$ for all $d \in \mathcal{I}$, then the optimal value of t is

$$t_0 = \min_{d \in \mathcal{I}} \max(f_L(d), f_U(d)) = \min_{d \in \mathcal{I}} f_L(d) = f_L(d_L). \quad (99)$$

Furthermore, if $f_L(d) > f_U(d)$ for some $d \in \mathcal{I}$, then $f_L(d) > f_U(d)$ for all $d \in \mathcal{I}$. This follows from the fact that $f_L(d)$ and $f_U(d)$ are continuous, and $f_L(d) \neq f_U(d)$ for all $d \notin \mathcal{I}$. Thus, to prove that $t_0 = f_L(d_L)$ it is sufficient to show that $f_L(d) > f_U(d)$ for some $d \in \mathcal{I}$. But this follows immediately from the fact that $f_U(d_U) < f_L(d_L) < f_L(d_U)$.

We conclude that the optimal value of d_i is given by

$$d_i(\gamma) = \begin{cases} \sqrt{\gamma} \frac{\delta_i\sigma_i^2}{\sqrt{U_i + \delta_i\sigma_i^2}\sqrt{L_i + \delta_i\sigma_i^2}}, & \gamma \leq \gamma_i^0 \\ \frac{\delta_i\sigma_i^2}{L_i + \delta_i\sigma_i^2}, & \gamma > \gamma_i^0 \end{cases} \quad (100)$$

where

$$\gamma_i^0 = \frac{U_i + \delta_i\sigma_i^2}{L_i + \delta_i\sigma_i^2}, \quad (101)$$

and the optimal value of $t(\gamma)$ is $\hat{t}(\gamma) = \sum_{i=1}^m t_i(\gamma)$, with $t_i(\gamma) = f_L(d_i(\gamma))$ we have (102) at the bottom of the

page. Finally, the minimax ratio regret estimator is equal to $\hat{\mathbf{x}} = \mathbf{V}\mathbf{D}\Sigma^\dagger\mathbf{U}^*\mathbf{y}$ where \mathbf{D} is a diagonal matrix with diagonal elements d_i that are given by (100), with γ chosen as the unique value for which $\hat{t}(\gamma) = \sum_{i=1}^m t_i(\gamma) = 0$. Since $\hat{t}(\gamma)$ is continuous and strictly decreasing in γ , we can find γ by a simple line search.

Note, that the minimax difference regret estimator of (72) has the same form as the minimax ratio regret estimator, where d_i is given by (100) with $\gamma = 1$.

From (31) and (32) it follows that we can interpret the minimax ratio regret estimator as an MMSE estimate of \mathbf{x} with least favorable covariance $\mathbf{C}_w = \mathbf{U}\text{diag}(w_1, \dots, w_n)\mathbf{U}^*$ where

$$w_i = \begin{cases} \frac{\sqrt{U_i + \delta_i\sigma_i^2}\sqrt{L_i + \delta_i\sigma_i^2}}{\sqrt{\gamma}} - \delta_i\sigma_i^2, & \gamma \leq \gamma_i^0, \\ L_i, & \gamma \geq \gamma_i^0, \end{cases} \quad 1 \leq i \leq m \quad (103)$$

and $w_i > 0$, $m+1 \leq i \leq n$ are arbitrary.

In Section IV-B it was shown that the minimax difference regret estimator is an MMSE estimator matched to a covariance matrix with eigenvalues

$$\tilde{w}_i = \sqrt{L_i + \delta_i\sigma_i^2}\sqrt{U_i + \delta_i\sigma_i^2} - \delta_i\sigma_i^2. \quad (104)$$

Since the optimal value of γ is greater than 1 (unless there is no uncertainty), $w_i < \tilde{w}_i$, $1 \leq i \leq m$, 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 estimator.

VI. PERFORMANCE ANALYSIS

In previous sections, we proposed and developed three estimators for the linear model (1), for the case in which \mathbf{H} is known and \mathbf{C}_x and \mathbf{C}_w are subject to uncertainties: minimax MSE, minimax difference regret, and minimax ratio regret. The performance of these methods clearly depends on the value of the unknown covariances. For example, if the true eigenvalues of \mathbf{C}_x and \mathbf{C}_w are equal to the worst case values, i.e., $\delta_i = u_i$ and $\theta_i = U_i$, then the minimax MSE approach will have the best performance, since it is designed to minimize the MSE for this choice of parameters. However, for other values of these unknowns, the minimax different regret or ratio regret estimators may perform better. One possible way of assessing the performance of the three estimators is to compute the MSE of each of the methods for the best possible choice of parameters, the worst possible choice, and the nominal (average) parameters. If, for example, the MSE of one of the estimators is considerably lower than the MSE of the other estimators for the best possible choice of parameters, and only slightly larger than the MSE of the other strategies for the worst choice, then we may prefer using this estimator. Thus, examining the performance of the proposed methods for the worst case, best case, and nom-

$$t_i(\gamma) = \begin{cases} \delta_i^2\sigma_i^2 \left(\frac{\sqrt{\gamma}}{\sqrt{U_i + \delta_i\sigma_i^2}} - \frac{1}{\sqrt{L_i + \delta_i\sigma_i^2}} \right)^2 + (1 - \gamma) \frac{\delta_i L_i}{L_i + \sigma_i^2 \delta_i}, & \gamma \leq \gamma_i^0 \\ (1 - \gamma) \frac{\delta_i L_i}{L_i + \sigma_i^2 \delta_i}, & \gamma > \gamma_i^0. \end{cases} \quad (102)$$

inal case can provide insight into their behavior, and can suggest guidelines for choosing which estimator to use for a specific problem. We now demonstrate these ideas in the context of a concrete example.

Consider the estimation problem in which

$$\mathbf{y} = \mathbf{x} + \mathbf{w} \quad (105)$$

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

$$E(x_i x_j) = \rho^{|j-i|} \quad (106)$$

for some parameter ρ , and \mathbf{w} is a zero-mean random vector uncorrelated with \mathbf{x} with covariance $\sigma^2 \mathbf{I}$. We assume that we know the model (105) and the eigenvectors of the true covariance matrices \mathbf{C}_x^0 and \mathbf{C}_w^0 . The eigenvalues of \mathbf{C}_x^0 are assumed to lie in an uncertainty interval $l_i \leq \delta_i \leq u_i$, where $l_i = (1 - \alpha)x_i$ and $u_i = (1 + \alpha)x_i$. Here x_i denotes the i th eigenvalue of the true covariance matrix \mathbf{C}_x^0 , and $0 < \alpha < 1$ is a parameter that defines the size of the uncertainty set. Similarly, the eigenvalues of \mathbf{C}_w^0 are assumed to lie in an uncertainty interval $L_i \leq \theta_i \leq U_i$, where $L_i = (1 - \beta)w_i$ and $U_i = (1 + \beta)w_i$ with w_i the i th eigenvalue of \mathbf{C}_w^0 , and $0 < \beta < 1$. In practice, the eigenvectors of \mathbf{C}_x^0 and \mathbf{C}_w^0 as well as the uncertainty sets may not be given. In this case, we may estimate these parameters from the data, as described in [13], [14].

For any linear estimator $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ of \mathbf{x} , we can find the worst choice of $\mathbf{C}_x, \mathbf{C}_w$, denoted $\mathbf{C}_x^{\text{WC}}, \mathbf{C}_w^{\text{WC}}$, that maximize the MSE, and the best choice of $\mathbf{C}_x, \mathbf{C}_w$, denoted $\mathbf{C}_x^{\text{BC}}, \mathbf{C}_w^{\text{BC}}$, that minimize the MSE. Using Lemma 4 we have that

$$\begin{aligned} \mathbf{C}_x^{\text{WC}} &= \arg \max_{l_i \leq \delta_i \leq u_i} \mathcal{L}(\Delta, \Theta) = \mathbf{V}\mathbf{X}_u\mathbf{V}^* \\ \mathbf{C}_w^{\text{WC}} &= \arg \max_{L_i \leq \theta_i \leq U_i} \mathcal{L}(\Delta, \Theta) = \mathbf{U}\mathbf{W}_U\mathbf{U}^* \end{aligned} \quad (107)$$

where

$$\begin{aligned} \mathcal{L}(\Delta, \Theta) &= \text{Tr}(\mathbf{G}\mathbf{U}\Theta\mathbf{U}^*\mathbf{G}^*) \\ &\quad + \text{Tr}(\mathbf{V}\Lambda\mathbf{V}^*(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})) \end{aligned} \quad (108)$$

\mathbf{V}, \mathbf{U} are the known eigenvector matrices of $\mathbf{C}_x^0, \mathbf{C}_w^0$ and $\mathbf{X}_u = \text{diag}(u_1, \dots, u_n)$, $\mathbf{W}_U = \text{diag}(U_1, \dots, U_n)$. Similarly

$$\begin{aligned} \mathbf{C}_x^{\text{BC}} &= \arg \min_{l_i \leq \delta_i \leq u_i} \mathcal{L}(\Delta, \Theta) = \mathbf{V}\mathbf{X}_l\mathbf{V}^* \\ \mathbf{C}_w^{\text{BC}} &= \arg \min_{L_i \leq \theta_i \leq U_i} \mathcal{L}(\Delta, \Theta) = \mathbf{U}\mathbf{W}_L\mathbf{U}^* \end{aligned} \quad (109)$$

where $\mathbf{X}_l = \text{diag}(l_1, \dots, l_n)$, $\mathbf{W}_L = \text{diag}(L_1, \dots, L_n)$. Note, that $\mathbf{C}_x^{\text{WC}}, \mathbf{C}_w^{\text{WC}}, \mathbf{C}_x^{\text{BC}}$, and \mathbf{C}_w^{BC} do not depend on the choice of \mathbf{G} .

In Fig. 4, we plot the MSE of the minimax MSE (MX), minimax difference regret (DRG), and minimax ratio regret (RRG) estimators as a function of the SNR defined by $-10 \log \sigma^2$ for $\rho = 0.9$, $n = 10$, $\alpha = 0.9$, and $\beta = 0.2$. The MSE of each of the estimators, which is given by $\text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \text{Tr}(\mathbf{C}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H}))$, is plotted for three different choices of \mathbf{C}_x and \mathbf{C}_w : the worst case $\mathbf{C}_x = \mathbf{C}_x^{\text{WC}}, \mathbf{C}_w = \mathbf{C}_w^{\text{WC}}$, the best case $\mathbf{C}_x = \mathbf{C}_x^{\text{BC}}, \mathbf{C}_w = \mathbf{C}_w^{\text{BC}}$, and the nominal (true) value $\mathbf{C}_x = \mathbf{C}_x^0, \mathbf{C}_w = \mathbf{C}_w^0$. As can be seen from the figure, in the worst case,

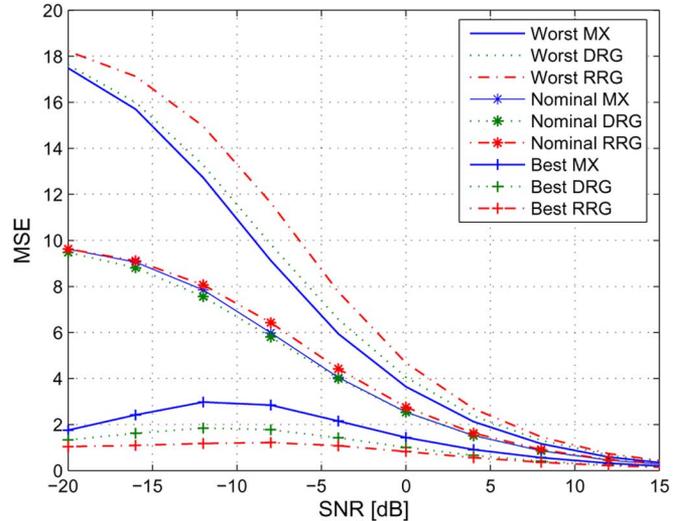


Fig. 4. MSE in estimating \mathbf{x} as a function of SNR using the minimax MSE (MX), the minimax difference regret (DRG), and the minimax ratio regret (RRG) estimators, for three different choices of \mathbf{C}_x and \mathbf{C}_w : $\mathbf{C}_x^{\text{WC}}, \mathbf{C}_w^{\text{WC}}$ (worst), $\mathbf{C}_x^0, \mathbf{C}_w^0$ (nominal), and $\mathbf{C}_x^{\text{BC}}, \mathbf{C}_w^{\text{BC}}$ (best).

the minimax MSE estimator has the best performance. This result is of course expected, since this method is designed to minimize the worst MSE. On the other hand, in the best case, the performance of the minimax MSE estimator deteriorates considerably. For $\mathbf{C}_x = \mathbf{C}_x^0, \mathbf{C}_w = \mathbf{C}_w^0$, the behavior of the three estimators is very similar. In this example, we may prefer using the difference regret approach over the minimax MSE method, since the loss in performance of the minimax MSE estimator in the best case is much more significant than the loss in performance of the difference regret estimator in the worst case. For the nominal value of \mathbf{C}_x , the estimators behave roughly the same, with a slight advantage to the difference regret strategy.

VII. CONCLUSION

We treated the problem of estimating a random vector \mathbf{x} in the linear model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where the covariance matrix \mathbf{C}_x of \mathbf{x} , the covariance matrix \mathbf{C}_w of \mathbf{w} and possibly also the model matrix \mathbf{H} , are subjected to uncertainties. The main contribution of this paper is to extend earlier results of [13], [14] to the setting in which the noise covariance \mathbf{C}_w may not be known precisely, and to introduce a simpler method of proof.

We developed the minimax MSE estimators for the case in which $\mathbf{C}_x, \mathbf{C}_w$ and \mathbf{H} are all subject to structured or unstructured uncertainties. We then treated the setting in which \mathbf{H} is known, and developed two competitive minimax strategies for the structured uncertainty model: the difference regret and the ratio regret estimators. As shown, earlier results of [13], [14] follow directly from the more general case we treat here. However, the method of proof we present is considerably simpler than that used in [13], [14]. Furthermore, using this new approach, we are able to gain more insight into the optimal estimators for cases treated in [13], [14]. For example, the results of Theorem 1 can be used to obtain a closed-form expression for the minimax MSE approach in the case in which \mathbf{C}_w is known, and \mathbf{C}_x and \mathbf{H} are subject to uncertainties.

In our derivation of the competitive minimax methods, we assumed a specific model on the eigenvectors of \mathbf{C}_x , \mathbf{C}_w , $\mathbf{H}^* \mathbf{H}$, and $\mathbf{H} \mathbf{H}^*$. If the true eigenvectors do not obey this model, then we can still apply our general results by exploiting the fact that the competitive strategies can be viewed as MMSE estimators with a specific choice of covariance matrices. Thus, we may implement the competitive estimators in this case using the form (3), where we replace the eigenvalues of \mathbf{C}_x and \mathbf{C}_w by their estimated values, as given, for example, by Corollary 1.

APPENDIX A PROOF OF LEMMA 3

To prove the lemma we first note that if $\ell = \mu$, then $\mathcal{I} = \{d_\ell\}$ and $d_0 = d_\ell$.

Next, assume that $\ell < \mu$. We first show that if $d_e \in \mathcal{I}$, then $d_0 = d_e$. Note, that if $d_e \in \mathcal{I}$, then from (27) it follows that we must have $d_\mu < d_\ell$. Therefore, for $d \in \mathcal{I}$, $f_\mu(d)$ is monotonically increasing and $f_\ell(d)$ is monotonically decreasing. Denoting $t_e = t(d_e)$ and noting that $t_e = f_\ell(d_e) = f_\mu(d_e)$, we conclude that for $d_e < d \leq d_\ell$, $f_\mu(d) > t_e$, and for $d_\mu \leq d < d_e$, $f_\ell(d) > t_e$ so that $t(d) > t_e$ for any $d \in \mathcal{I}$ such that $d \neq d_e$. It follows that if $d_e \in \mathcal{I}$, then $d_0 = d_e$.

Finally, we show that if $d_e \notin \mathcal{I}$, then $d = d_\ell$. We first note that if $f_\ell(d) > f_\mu(d)$ for all $d \in \mathcal{I}$, then the optimal value of t is

$$t_0 = \min_{d \in \mathcal{I}} \max(f_\ell(d), f_\mu(d)) = \min_{d \in \mathcal{I}} f_\ell(d) = f_\ell(d_\ell) \quad (110)$$

and $d_0 = d_\ell$. Furthermore, if $f_\ell(d) > f_\mu(d)$ for some $d \in \mathcal{I}$, then $f_\ell(d) > f_\mu(d)$ for all $d \in \mathcal{I}$. This follows from the fact that $f_\ell(d)$ and $f_\mu(d)$ are continuous, and since $d_e \notin \mathcal{I}$, $f_\ell(d)$, and $f_\mu(d)$ cannot intersect anywhere in \mathcal{I} . Thus, to prove that $d_0 = d_\ell$ it is sufficient to show that $f_\ell(d) > f_\mu(d)$ for some $d \in \mathcal{I}$. But this follows immediately from the inequality $f_\ell(d_\mu) > f_\ell(d_\ell) > f_\mu(d_\mu)$, since

$$f_\ell(d_\ell) = \frac{uU}{U + u\ell^2} > \frac{uU}{U + u\mu^2} = f_\mu(d_\mu). \quad (111)$$

APPENDIX B

INTERSECTION POINTS OF THE FUNCTIONS (61)

In this appendix, we show that $f_{U\ell}(d)$ and $f_{Lu}(d)$ intersect at exactly one point d_e in the interval $\mathcal{I} = [d_{U\ell}, d_{Lu}]$, where $f_{U\ell}(d)$ and $f_{Lu}(d)$ are defined by (61), and $d_{U\ell}$ and d_{Lu} are given by (64). We also show that each of the functions $f_{Uu}(d)$ and $f_{L\ell}(d)$ intersect $f_{Lu}(d)$ and $f_{U\ell}(d)$ at exactly one point in \mathcal{I} .

If $f_{Lu}(d) = f_{U\ell}(d)$, then

$$\frac{u\sigma}{\sqrt{L + u\sigma^2}} - \frac{\sqrt{L + u\sigma^2}}{\sigma} d = \frac{\sqrt{U + l\sigma^2}}{\sigma} d - \frac{l\sigma}{\sqrt{U + l\sigma^2}} \quad (112)$$

or

$$\frac{u\sigma}{\sqrt{L + u\sigma^2}} - \frac{\sqrt{L + u\sigma^2}}{\sigma} d = -\frac{\sqrt{U + l\sigma^2}}{\sigma} d + \frac{l\sigma}{\sqrt{U + l\sigma^2}}. \quad (113)$$

Denoting by d_1 and d_2 the solutions to (112) and (113), respectively, we have that

$$d_1 = \frac{1}{\sqrt{U + l\sigma^2} + \sqrt{L + u\sigma^2}} \left(\frac{l\sigma^2}{\sqrt{U + l\sigma^2}} + \frac{u\sigma^2}{\sqrt{L + u\sigma^2}} \right) \quad (114)$$

and, if $\sqrt{U + l\sigma^2} \neq \sqrt{L + u\sigma^2}$, then

$$d_2 = \frac{1}{\sqrt{U + l\sigma^2} - \sqrt{L + u\sigma^2}} \left(\frac{l\sigma^2}{\sqrt{U + l\sigma^2}} - \frac{u\sigma^2}{\sqrt{L + u\sigma^2}} \right). \quad (115)$$

If $\sqrt{U + l\sigma^2} = \sqrt{L + u\sigma^2}$, then there is no solution to (113), since we are assuming that $L \neq U$ or $l \neq u$. To show that $d_2 \notin \mathcal{I}$, suppose first that $\sqrt{U + l\sigma^2} > \sqrt{L + u\sigma^2}$. In this case

$$\frac{l\sigma^2}{\sqrt{U + l\sigma^2}} < \frac{u\sigma^2}{\sqrt{L + u\sigma^2}}, \quad (116)$$

which implies that $d_2 < 0$, so that $d_2 \notin \mathcal{I}$. If, on the other hand, $\sqrt{U + l\sigma^2} < \sqrt{L + u\sigma^2}$, then

$$d_2 - d_{Lu} = \frac{\sqrt{U + l\sigma^2}}{\sqrt{U + l\sigma^2} - \sqrt{L + u\sigma^2}} \left(\frac{l\sigma^2}{U + l\sigma^2} - \frac{u\sigma^2}{L + u\sigma^2} \right) \geq 0 \quad (117)$$

so that $d_2 \notin \mathcal{I}$. Similarly

$$\begin{aligned} d_1 - d_{U\ell} &= \frac{1}{\sqrt{U + l\sigma^2} + \sqrt{L + u\sigma^2}} \\ &\cdot \left(\frac{-l\sigma^2\sqrt{L + u\sigma^2}}{U + l\sigma^2} + \frac{u\sigma^2}{\sqrt{L + u\sigma^2}} \right) \\ &= \frac{1}{\sqrt{U + l\sigma^2} + \sqrt{L + u\sigma^2}} \\ &\cdot \left(\frac{u\sigma^2(U + l\sigma^2) - l\sigma^2(L + u\sigma^2)}{(U + l\sigma^2)\sqrt{L + u\sigma^2}} \right) \\ &\geq 0 \end{aligned} \quad (118)$$

and

$$\begin{aligned} d_1 - d_{Lu} &= \frac{1}{\sqrt{U + l\sigma^2} + \sqrt{L + u\sigma^2}} \\ &\cdot \left(\frac{-u\sigma^2\sqrt{U + l\sigma^2}}{L + u\sigma^2} + \frac{l\sigma^2}{\sqrt{U + l\sigma^2}} \right) \\ &= \frac{1}{\sqrt{L + u\sigma^2} + \sqrt{U + l\sigma^2}} \\ &\cdot \left(\frac{l\sigma^2(L + u\sigma^2) - u\sigma^2(U + l\sigma^2)}{(L + u\sigma^2)\sqrt{U + l\sigma^2}} \right) \\ &\leq 0 \end{aligned} \quad (119)$$

so that $d_1 \in \mathcal{I}$. Thus, the unique intersection point in \mathcal{I} of the functions $f_{Lu}(d)$ and $f_{U\ell}(d)$ is $d_e = d_1$.

We now show that $f_{Uu}(d)$ intersects $f_{Lu}(d)$ and $f_{U\ell}(d)$ at exactly one point in \mathcal{I} .

Suppose that $f_{Uu}(d) = f_{Lu}(d)$. Then

$$\frac{u\sigma}{\sqrt{L + u\sigma^2}} - \frac{\sqrt{L + u\sigma^2}}{\sigma} d = \frac{\sqrt{U + u\sigma^2}}{\sigma} d - \frac{u\sigma}{\sqrt{U + u\sigma^2}} \quad (120)$$

or

$$\frac{u\sigma}{\sqrt{L+u\sigma^2}} - \frac{\sqrt{L+u\sigma^2}}{\sigma} dl = -\frac{\sqrt{U+u\sigma^2}}{\sigma} d + \frac{u\sigma}{\sqrt{U+u\sigma^2}}. \quad (121)$$

Denoting by d_1 and d_2 the solutions to (120) and (121), respectively, we have that

$$d_1 = \frac{u\sigma^2}{\sqrt{L+u\sigma^2}\sqrt{U+u\sigma^2}} \quad (122)$$

and

$$d_2 = -\frac{u\sigma^2}{\sqrt{L+u\sigma^2}\sqrt{U+u\sigma^2}}. \quad (123)$$

Since $d_2 < 0$, clearly $d_2 \notin \mathcal{I}$. On the other hand, using the fact that $1/\sqrt{U+u\sigma^2} \leq 1/\sqrt{L+u\sigma^2}$ and that

$$\frac{u\sigma^2}{\sqrt{L+u\sigma^2}\sqrt{U+u\sigma^2}} \geq \frac{u\sigma^2}{U+u\sigma^2} \geq \frac{l\sigma^2}{U+l\sigma^2} \quad (124)$$

we have $d_1 \in \mathcal{I}$. Thus, $f_{Lu}(d) = f_{Uu}(d)$ for exactly one value $d = d_1$ in \mathcal{I} .

Next, suppose that $f_{Uu}(d) = f_{Ul}(d)$. Then

$$\frac{u\sigma}{\sqrt{U+u\sigma^2}} - \frac{\sqrt{U+u\sigma^2}}{\sigma} d = \frac{\sqrt{U+l\sigma^2}}{\sigma} d - \frac{l\sigma}{\sqrt{U+l\sigma^2}} \quad (125)$$

or

$$\frac{u\sigma}{\sqrt{U+u\sigma^2}} - \frac{\sqrt{U+u\sigma^2}}{\sigma} d = -\frac{\sqrt{U+l\sigma^2}}{\sigma} d + \frac{l\sigma}{\sqrt{U+l\sigma^2}}. \quad (126)$$

Denoting by d_1 and d_2 the solutions to (125) and (126), respectively,

$$d_1 = \frac{1}{\sqrt{U+l\sigma^2} + \sqrt{U+u\sigma^2}} l \cdot \left(\frac{u\sigma^2}{\sqrt{U+u\sigma^2}} + \frac{l\sigma^2}{\sqrt{U+l\sigma^2}} \right) \quad (127)$$

$$= \frac{1}{\sqrt{U+l\sigma^2} + \sqrt{U+u\sigma^2}} \cdot \left(\frac{u\sigma^2\sqrt{U+l\sigma^2} + l\sigma^2\sqrt{U+u\sigma^2}}{\sqrt{U+l\sigma^2}\sqrt{U+u\sigma^2}} \right) \quad (128)$$

$$= 1 - \frac{U}{\sqrt{U+l\sigma^2}\sqrt{U+u\sigma^2}} \quad (129)$$

where (129) follows from adding and subtracting the term $U(\sqrt{U+l\sigma^2} + \sqrt{U+u\sigma^2})$ to the numerator of (128). Similarly, we can show that

$$d_2 = 1 + \frac{U}{\sqrt{U+l\sigma^2}\sqrt{U+u\sigma^2}}. \quad (130)$$

Since $d_2 > 1$, clearly $d_2 \notin \mathcal{I}$. Using the fact that

$$\frac{U}{U+u\sigma^2} \leq \frac{U}{\sqrt{U+l\sigma^2}\sqrt{U+u\sigma^2}} \leq \frac{U}{U+l\sigma^2}, \quad (131)$$

$d_1 \in \mathcal{I}$, and $f_{Uu}(d) = f_{Ul}(d)$ for one value $d = d_1 \in \mathcal{I}$.

Since $f_{Li}(d)$ is equal to $f_{Uu}(d)$, where we replace U and u by L and l , respectively, and $f_{Lu}(d)$ is equal to $f_{Ul}(d)$ where

we replace U and l by L and u , respectively, we have from (122) and (123) that $f_{Li}(d) = f_{Ul}(d)$ at the points

$$d_1 = \frac{l\sigma^2}{\sqrt{L+l\sigma^2}\sqrt{U+l\sigma^2}} \in \mathcal{I} \quad (132)$$

and

$$d_2 = -\frac{l\sigma^2}{\sqrt{L+l\sigma^2}\sqrt{U+l\sigma^2}} \notin \mathcal{I}. \quad (133)$$

Similarly, from (129) and (130), $f_{Li}(d) = f_{Lu}(d)$ at the points

$$d_1 = 1 - \frac{L}{\sqrt{L+l\sigma^2}\sqrt{L+u\sigma^2}} \in \mathcal{I} \quad (134)$$

and

$$d_2 = 1 + \frac{L}{\sqrt{L+l\sigma^2}\sqrt{L+u\sigma^2}} \notin \mathcal{I}. \quad (135)$$

APPENDIX C

PROOF OF LEMMA 5

To prove the lemma, we first show that $d_{Ul} \leq d_0 \leq d_{Lu}$.

When $d > u\sigma^2/(L+u\sigma^2) = d_{Lu}$, we have that $g_{Li}(d), g_{Lu}(d), g_{Ul}(d), g_{Uu}(d) > 0$, where

$$g_{xy}(d) = \frac{\sqrt{L+l\sigma^2}}{\sigma} d - \frac{l\sigma}{\sqrt{L+l\sigma^2}}. \quad (136)$$

Therefore, the functions $f_{Li}(d)$, $f_{Lu}(d)$, $f_{Ul}(d)$, $f_{Uu}(d)$ are monotonically increasing, so that $d_0 \leq d_{Lu}$. When $d < l\sigma^2/(U+l\sigma^2) = d_{Ul}$, we have that

$$g_{Li}(d), g_{Lu}(d), g_{Ul}(d), g_{Uu}(d) < 0$$

and $f_{Li}(d)$, $f_{Lu}(d)$, $f_{Ul}(d)$, $f_{Uu}(d)$ are monotonically decreasing. Therefore, $d_0 \geq d_{Ul}$.

If $U = L$ and $u = l$, then it follows that $d_0 = d_{Lu} = d_{Ul}$.

Next, suppose that either $U \neq L$ or $u \neq l$, and let \mathcal{I} denote the interval $\mathcal{I} = [d_{Ul}, d_{Lu}]$ containing d_0 . Then, we now show that $d_0 = d_e$, where d_e is the unique intersection point in \mathcal{I} of the functions $f_{Lu}(d)$ and $f_{Ul}(d)$, as proved in Appendix B. To this end, it suffices to establish that

$$g(d) \leq \max(f_{Lu}(d), f_{Ul}(d)), \quad d \in \mathcal{I} \quad (137)$$

where $g(d)$ is equal to $f_{Uu}(d)$ or $f_{Li}(d)$. If (137) holds, then

$$\begin{aligned} d_0 &= \arg \min_{d \in \mathcal{I}} \max(f_{Lu}(d), f_{Uu}(d), f_{Li}(d), f_{Ul}(d)) \\ &= \arg \min_{d \in \mathcal{I}} \max(f_{Lu}(d), f_{Ul}(d)). \end{aligned} \quad (138)$$

Since f_{Lu} is monotonically decreasing in $d \in \mathcal{I}$, and f_{Ul} is monotonically increasing in $d \in \mathcal{I}$

$$\begin{aligned} f_{Lu}(d) &> f_{Ul}(d) > t_e, & d_{Ul} \leq d < d_e \\ f_{Ul}(d) &> f_{Lu}(d) > t_e, & d_e < d \leq d_{Lu} \end{aligned} \quad (139)$$

where $t_e = t(d_e)$ and $t(d) = \max(f_{Lu}(d), f_{Ul}(d))$. Therefore, $t(d) > t_e$ for any $d \in \mathcal{I}$ such that $d \neq d_e$, and d_e is optimal.

Thus, to complete the proof of the theorem it remains to show that (137) holds. In Appendix B, we showed that $g(d)$ intersects $f_{Lu}(d)$ and $f_{Ul}(d)$ at exactly one point in \mathcal{I} . Now, let d_α denote the value at which $g(d)$ is minimized, i.e., the value for which $g(d_\alpha) = 0$. Then

$$d_\alpha = \frac{u\sigma^2}{U + u\sigma^2} \quad \text{or} \quad d_\alpha = \frac{l\sigma^2}{L + l\sigma^2} \quad (140)$$

so that $d_\alpha \in \mathcal{I}$. Since

$$\begin{aligned} 0 &= g(d_\alpha) \leq f_{Lu}(d_\alpha) \\ 0 &= f_{Lu}(d_{Lu}) \leq g(d_{Lu}) \end{aligned} \quad (141)$$

and $g(d)$ and $f_{Lu}(d)$ are continuous, it follows that $g(d)$ and $f_{Lu}(d)$ intersect at some point in the interval $[d_\alpha, d_{Lu}]$. Similarly, $g(d)$ and $f_{Ul}(d)$ intersect at a point in $[d_{Ul}, d_\alpha]$. Now, suppose that

$$g(d_p) > \max(f_{Lu}(d_p), f_{Ul}(d_p)) \quad (142)$$

for some $d_p \in \mathcal{I}$. Clearly, $d_p \neq d_\alpha$ since $g(d_\alpha) = 0$, and $f_{Lu}(d), f_{Ul}(d) \geq 0$ for all d . Since $g(d_\alpha) \leq f_{Ul}(d_\alpha)$ and $g(d_\alpha) \leq f_{Lu}(d_\alpha)$, (142) implies that $g(d)$ intersects both $f_{Lu}(d)$ and $f_{Ul}(d)$ in the interval $[\min(d_p, d_\alpha), \max(d_p, d_\alpha)]$. If $d_p < d_\alpha$ then, as illustrated in Fig. 3, this implies that $g(d)$ intersects $f_{Lu}(d)$ at a point in $[d_\alpha, d_{Lu}]$ and at a point in $[d_p, d_\alpha]$, which contradicts the fact that there is only one intersection point in \mathcal{I} . Similarly, if $d_p > d_\alpha$, then (142) implies that $g(d)$ intersects $f_{Ul}(d)$ at a point in $[d_{Ul}, d_\alpha]$ and at a point in $[d_\alpha, d_p]$, which contradicts the fact that there is only one intersection point in \mathcal{I} . Therefore, we conclude that $g(d) < \max(f_{Lu}(d_p), f_{Ul}(d_p))$ so that d_e is optimal, which completes the proof of the lemma.

APPENDIX D PROOF OF PROPOSITION 1

To prove that $\hat{t}(\gamma)$ is continuous in γ , we rely on the following lemma.²

Lemma 6: Let

$$g(\alpha) = \min_{\mathbf{x} \in \mathcal{C}} f(\mathbf{x}, \alpha), \quad \alpha \in \mathcal{A} \quad (143)$$

where $f(\mathbf{x}, \alpha)$ is a continuous function in the variables \mathbf{x} and the parameters α , which are defined over a set \mathcal{A} , and \mathcal{C} is a compact set. Then $g(\alpha)$ is continuous.

Proof: To prove Lemma 6, we show that if $\{\alpha_k\}$ is a sequence converging to a point α_* in \mathcal{A} , then the sequence $\{g(\alpha_k)\}$ converges to $g(\alpha_*)$. To this end, it is sufficient to show that every convergent subsequence of $\{g(\alpha_k)\}$ converges to $g(\alpha_*)$.

Let $\{g(\alpha_{kl})\}$ be a convergent subsequence of $\{g(\alpha_k)\}$, and define

$$\mathbf{x}_{kl} = \arg \min_{\mathbf{x} \in \mathcal{C}} f(\mathbf{x}, \alpha_{kl}). \quad (144)$$

Since \mathcal{C} is compact, there is a subsequence of \mathbf{x}_{kl} , which we denote by $\{\mathbf{x}_{klm}\}$, that converges to a point \mathbf{x}_* in \mathcal{C} . Now, for every $\mathbf{x} \in \mathcal{C}$

$$f(\mathbf{x}, \alpha_{klm}) \geq f(\mathbf{x}_{klm}, \alpha_{klm}) = g(\alpha_{klm}). \quad (145)$$

Since $f(\mathbf{x}, \alpha)$ is continuous, $f(\mathbf{x}, \alpha_{klm}) \rightarrow f(\mathbf{x}, \alpha_*)$ and $f(\mathbf{x}_{klm}, \alpha_{klm}) \rightarrow f(\mathbf{x}_*, \alpha_*)$ as $m \rightarrow \infty$, where we used the fact that because $\alpha_k \rightarrow \alpha_*$, every subsequence $\{\alpha_{klm}\}$ of $\{\alpha_k\}$ converges to α_* . Taking the limit of (145) as $m \rightarrow \infty$ we then have

$$f(\mathbf{x}, \alpha_*) \geq f(\mathbf{x}_*, \alpha_*) = \lim_{m \rightarrow \infty} g(\alpha_{klm}), \quad \forall \mathbf{x} \in \mathcal{C} \quad (146)$$

from which we conclude that

$$\mathbf{x}_* = \arg \min_{\mathbf{x} \in \mathcal{C}} f(\mathbf{x}, \alpha_*) \quad (147)$$

and

$$\lim_{m \rightarrow \infty} g(\alpha_{klm}) = g(\alpha_*) \quad (148)$$

completing the proof of the lemma. \square

To apply Lemma 6 to the problem (Φ) , we first note that following the proof of Theorem 2, the optimal \mathbf{G} has the form $\mathbf{G} = \mathbf{V}\mathbf{D}\Sigma^\dagger\mathbf{U}^*$, where \mathbf{D} is a diagonal matrix with diagonal elements d_i that are the solution to

$$\min_{t, d_i} \left\{ t : \max_{l_i \leq \delta_i \leq u_i, L_i \leq \theta_i \leq U_i} f(d_i, \delta_i, \theta_i, \gamma) \leq t \right\} \quad (149)$$

where

$$\begin{aligned} f(d_i, \delta_i, \theta_i, \gamma) \\ = \sum_{i=1}^m \left(\frac{\theta_i d_i^2}{\sigma_i^2} + \delta_i (1 - d_i)^2 - \gamma \frac{\delta_i \theta_i}{\theta_i + \sigma_i^2 \delta_i} \right). \end{aligned} \quad (150)$$

We can therefore express $\hat{t}(\gamma)$ as

$$\hat{t}(\gamma) = \min_{d_i} g(d_i, \gamma) \quad (151)$$

with

$$g(d_i, \gamma) = \max_{l_i \leq \delta_i \leq u_i, L_i \leq \theta_i \leq U_i} f(d_i, \delta_i, \theta_i, \gamma). \quad (152)$$

Since the set $l_i \leq \delta_i \leq u_i, L_i \leq \theta_i \leq U_i$ is closed and bounded, it is compact, and from Lemma 6, together with the fact that $f(d_i, \delta_i, \theta_i, \gamma)$ is continuous, we conclude that $g(d_i, \gamma)$ is continuous. It then follows from (151) and Lemma 6, that if the minimization over d_i can be confined to a compact set, then $\hat{t}(\gamma)$ is also continuous.

We now show that the optimal values of d_i satisfy $0 \leq d_i \leq 1$, so that the minimization in (151) can be confined to a compact set. Let $d_i < 0$ be arbitrary, and let $z_i = -d_i > 0$. Since $(1 - d_i)^2 > (1 - z_i)^2 = (1 + d_i)^2$, $f(d_i, \delta_i, \theta_i, \gamma) \geq f(z_i, \delta_i, \theta_i, \gamma)$, and consequently, $g(d_i, \gamma) \geq g(z_i, \gamma)$. Therefore, the optimal value of d_i can always be chosen to be nonnegative. Next, let

²This proof is due to A. Beck.

$d_i > 1$ be arbitrary, and let $z_i = 2 - d_i < 1$. Since $(1 - d_i)^2 = (1 - z_i)^2$ and $d_i^2 > z_i^2$, $g(d_i, \gamma) > g(z_i, \gamma)$, and the optimal value of d_i satisfies $d_i \leq 1$. We conclude that $\hat{t}(\gamma)$ can be expressed as

$$\hat{t}(\gamma) = \min_{0 \leq d_i \leq 1} g(d_i, \gamma). \quad (153)$$

Since $g(d_i, \gamma)$ is continuous, it follows from Lemma 6 that $\hat{t}(\gamma)$ is also continuous, completing the proof of the first part of the proposition.

The second part of the proposition follows from the fact that by our assumption

$$\text{Tr} \left(\left(\mathbf{I} - \mathbf{C}_x \mathbf{H}^* (\mathbf{H} \mathbf{C}_x \mathbf{H}^* + \mathbf{C}_w)^{-1} \mathbf{H} \right) \mathbf{C}_x \right) > 0.$$

Finally, to prove the last part of the proposition, we note that since $\hat{t}(\gamma)$ is strictly decreasing in γ , there is at most one value of γ such that $\hat{t}(\gamma) = 0$. To show that such a value always exists it is sufficient to show that $\hat{t}(\gamma_1) \geq 0$ and $\hat{t}(\gamma_2) < 0$ for some γ_1 and γ_2 , since $\hat{t}(\gamma)$ is continuous. Now, $\hat{t}(\gamma) \rightarrow -\infty$ as $\gamma \rightarrow \infty$ and $\hat{t}(1) \geq 0$. This is because $f(d_i, \delta_i, \theta_i, 1) \geq 0$, for any choice of d_i, δ_i and θ_i . Indeed, differentiating $f(d_i, \delta_i, \theta_i, 1)$ with respect to d_i and equating to 0, the minimal value of $f(d_i, \delta_i, \theta_i, 1)$ is obtained at

$$d_i^0 = \frac{\delta_i \sigma_i^2}{\delta_i \sigma_i^2 + \theta_i} \quad (154)$$

and $f(d_i^0, \delta_i, \theta_i, 1) = 0$. It follows that for any choice of d_i, δ_i and θ_i , $f(d_i, \delta_i, \theta_i, 1) \geq f(d_i^0, \delta_i, \theta_i, 1) \geq 0$, which implies that $\hat{t}(1) \geq 0$. Therefore, there is a unique γ such that $\hat{t}(\gamma) = 0$, completing the proof of the proposition.

ACKNOWLEDGMENT

The author wishes to thank the anonymous reviewers for their helpful comments which greatly improved the presentation.

REFERENCES

- [1] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proc. IEEE*, vol. 73, no. 3, pp. 433–481, Mar. 1985.
- [2] K. S. Vastola and H. V. Poor, "Robust Wiener-Kolmogorov theory," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 2, pp. 316–327, Mar. 1984.
- [3] N. Wiener, *The Extrapolation, Interpolation and Smoothing of Stationary Time Series*. New York: Wiley, 1949.
- [4] A. Kolmogorov, "Interpolation and extrapolation," *Bull. Acad. Sci., USSR, ser. Math.*, vol. 5, pp. 3–14, 1941.

- [5] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.
- [6] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Statist.*, vol. 35, pp. 73–101, 1964.
- [7] ———, *Robust Statistics*. New York: Wiley, 1981.
- [8] L. Breiman, "A note on minimax filtering," *Ann. Probab.*, vol. 1, pp. 175–179, 1973.
- [9] S. A. Kassam and T. L. Lim, "Robust Wiener filters," *J. Franklin Inst.*, vol. 304, pp. 171–185, Oct./Nov. 1977.
- [10] H. V. Poor, "On robust Wiener filtering," *IEEE Trans. Automat. Control*, vol. AC-25, no. 3, pp. 521–526, Jun. 1980.
- [11] J. Franke, "Minimax-robust prediction of discrete time series," *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, vol. 68, pp. 337–364, 1985.
- [12] 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.
- [13] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation of random parameters," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1931–1946, Jul. 2004.
- [14] ———, "Minimax MSE—Ratio estimation with signal covariance uncertainties," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1335–1347, Apr. 2005.
- [15] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Linear minimax regret estimation of deterministic parameters with bounded data uncertainties," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2177–2188, Aug. 2004.
- [16] L. D. Davisson, "Universal noiseless coding," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 6, pp. 783–795, Nov. 1973.
- [17] M. Feder and N. Merhav, "Universal composite hypothesis testing: A competitive minimax approach," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1504–1517, Jun. 2002.
- [18] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*, ser. MPS-SIAM Series on Optimization. Philadelphia, PA: SIAM, 2001.
- [19] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 40–95, Mar. 1996.
- [20] Y. Nesterov and A. Nemirovski, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA: SIAM, 1994.
- [21] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimiz. Methods and Software*, vol. 11–12, pp. 625–653, 1999.
- [22] D. Peaucelle, D. Henrion, and Y. Labit, Users Guide for SeDuMi Interface 1.03 [Online]. Available: <http://www.laas.fr/~peaucell/SeDuMiInt.html>
- [23] S. A. Kassam, "Robust hypothesis testing for bounded classes of probability densities," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 2, pp. 242–247, Mar. 1981.
- [24] V. P. Kuznetsov, "Stable detection when the signal and spectrum of normal noise are inaccurately known," Transl.: in English: *Telecomm. Radio Eng.*, vol. 30/31, pp. 58–64, 1976.
- [25] R. Gray, Toeplitz and circulant Matrices: A review Inform. Syst. Lab., Stanford Univ., Stanford, CA, 1977, Tech. Rep. 6504-1.
- [26] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.