Minimax MSE Estimation of Deterministic Parameters With Noise Covariance Uncertainties

Yonina C. Eldar, *Member, IEEE*

***Abstract*—In this paper, a minimax mean-squared error (MSE) estimator is developed for estimating an unknown deterministic parameter vector in a linear model, subject to noise covariance uncertainties. The estimator is designed to minimize the worst-case MSE across all norm-bounded parameter vectors, and all noise covariance matrices, in a given region of uncertainty. The minimax estimator is shown to have the same form as the estimator that minimizes the worst-case MSE over all norm-bounded vectors for a least-favorable choice of the noise covariance matrix. An example demonstrating the performance advantage of the minimax MSE approach over the least-squares and weighted least-squares methods is presented.**

***Index Terms*—Covariance uncertainty, minimax estimation, robust estimation.**

I. INTRODUCTION

THE problem of parameter estimation in linear models is pervasive in signal processing and communication applications. It is often common to restrict attention to linear estimators, which simplifies the implementation as well as the mathematical derivations. The simplest design scenario is when the second order statistics of the parameters to be estimated are known and it is desirable to minimize the mean-squared error (MSE). The optimal linear estimator for this problem is straightforward to derive, and is the well-known minimum MSE (MMSE) estimator [1]. However, as discussed in a recent series of papers [2]–[7], if the parameters to be estimated are deterministic, then the MSE depends explicitly on the unknown parameters, and therefore cannot be optimized directly.

We consider the class of estimation problems represented by the linear model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where $\mathbf{y}$ is the data vector, $\mathbf{H}$ is a known linear transformation, $\mathbf{x}$ is the deterministic unknown parameter vector to be estimated, and $\mathbf{w}$ is a noise vector with covariance matrix $\mathbf{C}_w$. Since the expected estimation error depends explicitly on $\mathbf{x}$, often the estimated parameters are chosen to optimize a criterion based on the observed signal $\mathbf{y}$, which no longer depends on the unknown $\mathbf{x}$. For example, the celebrated least-squares estimator , which was first used by Gauss to predict movements of planets [9], seeks the estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ that results in an estimated data vector $\hat{\mathbf{y}} = \mathbf{H}\hat{\mathbf{x}}$ that is closest, in a squared norm sense, to the given data $\mathbf{y}$. However, in an estimation context, the objective typically is to minimize the size of the estimation error $\hat{\mathbf{x}} - \mathbf{x}$, rather than that of the data error $\hat{\mathbf{y}} - \mathbf{y}$. In fact, in many practical scenarios the least-squares estimator is known to result in a large MSE.

To overcome this limitation of the least-squares estimator, it is desirable to develop estimators that optimize a criterion based on the estimation error rather than the data error. To render this strategy practical, the dependence of the estimation error on $\mathbf{x}$ must somehow be eliminated. Several different approaches to estimation in this spirit have been developed, under the assumption that the norm of the unknown vector $\mathbf{x}$ is bounded. One strategy is to minimize the worst-case MSE over all norm-bounded parameter vectors , [3], [6], [7]. An alternative method, developed in [2], is to minimize the worst-case *regret*, i.e., the worst-case difference between the MSE and the best possible MSE achievable with a linear estimator that knows $\mathbf{x}$. Both classes of estimators have been derived under the assumption that the noise covariance matrix $\mathbf{C}_w$ is completely specified. In a wide range of practical applications, this covariance may, unfortunately, be subjected to uncertainties. It is therefore desirable to design a robust estimator whose performance remains relatively insensitive and reasonably good across the region of uncertainty in the noise covariance.

The problem of estimation in the presence of unknown noise covariance has been studied extensively in the statistical literature under the assumption that $\mathbf{H} = \mathbf{I}$, and that an observation $\mathbf{A}$ of the matrix $\mathbf{C}_w$ is given, where $\mathbf{A}$ has a Wishart distribution with parameter $\mathbf{C}_w$. For this setup, different nonlinear estimators have been proposed and analyzed [10]–[14].

In this paper, we treat the case in which $\mathbf{C}_w$ is known to lie in a given *deterministic* uncertainty set. The uncertainty region we consider is defined by first jointly diagonalizing $\mathbf{C}_w$ and $\mathbf{H}\mathbf{H}^*$, and then assuming that the diagonal values of $\mathbf{C}_w$ in this representation lie in some convex constraint set. For example, each of the diagonal values of $\mathbf{C}_w$ may be bounded below and above. As another example, the sum of the diagonal values can be bounded. In practice, the diagonalizing matrix and bounds on the diagonal values can be estimated from training data. The uncertainty interval can then be regarded as a confidence interval around the estimated values. Given an uncertainty set, we seek the linear estimator that minimizes the worst-case MSE over all norm-bounded parameter vectors $\mathbf{x}$, and all noise covariance matrices in the region of uncertainty. Robust estimators of this form have been studied extensively in the context of Wiener estimation, where the signal to be estimated is an infinite-length stationary process with uncertain second order statistics [15]–[21].

In our model, the uncertainty is defined only on the generalized eigenvalues and not on the eigenvectors. Therefore, we refer to this class of uncertainties as "structured uncertainties."

Manuscript received January 16, 2004; revised March 1, 2005. This work was supported in part by the Taub Foundation, by the EU 6th framework programme via the NEWCOM network of excellence, and by the Israel Science Foundation under Grant No. 536/04. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Simon J. Godsill.

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

Digital Object Identifier 10.1109/TSP.2005.861086

In a companion paper [22] we consider the unstructured case in which the deterministic uncertainty set is defined by a norm constraint, so that the true covariance $\mathbf{C}_w$ can be written as $\mathbf{C}_w = \tilde{\mathbf{C}}_w + \Delta$ where $\tilde{\mathbf{C}}_w$ is a given nominal covariance matrix, and Δ is an arbitrary norm-bounded perturbation such that $\mathbf{C}_w$ is nonnegative definite.

Our approach to developing the minimax MSE estimator is to show that it can be derived by first solving a maximin problem, in which we reverse the order of the minimization and the maximization, leading to a simpler optimization problem. We then prove that the resulting maximin MSE estimator is also minimax optimal. As we show, the minimax MSE estimator is matched to a least-favorable noise covariance matrix, which depends directly on the noise covariance uncertainty set.

To illustrate the potential advantage of the minimax MSE approach, we consider an example in which the noise covariance is not known but rather estimated from the data. Simulation results demonstrate that the minimax MSE approach can significantly increase the performance over the conventional least-squares method, even in cases in which the noise covariance matrix and the norm of $\mathbf{x}$ are estimated from the data.

The paper is organized as follows. In Section II, we discuss our problem and introduce (without proof) the minimax MSE estimator. The derivation of the estimator is presented in Section III. Section IV considers some special cases of uncertainty classes. In Section V, we provide an example illustrating the performance of the proposed estimator.

II. Minimax MSE Estimator

In the sequel, we denote vectors in $\mathbb{C}^m$ by boldface lowercase letters and matrices in $\mathbb{C}^{n\times m}$ by boldface uppercase letters. The identity matrix of appropriate dimension is denoted by $\mathbf{I}$, $(\cdot)^*$ denotes the Hermitian conjugate of the corresponding matrix, $(\hat{\cdot})$ denotes an estimated vector or matrix and $\mathrm{diag}(\delta_1,\ldots,\delta_m)$ denotes an $m\times m$ diagonal matrix with diagonal elements δ_i. We use Θ to denote an arbitrary diagonal matrix of appropriate dimensions. The notation $\mathbf{A}\succ 0$ means that the matrix $\mathbf{A}$ is positive definite.

We consider the generic linear estimation problem in which it is desirable to estimate an unknown deterministic vector parameter $\mathbf{x}$ from observations $\mathbf{y}$ which are related through the linear model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}. \tag{1}$$

Here $\mathbf{H}$ is a known $n\times m$ matrix with full rank m, and $\mathbf{w}$ is a zero-mean, length-n random vector with positive definite covariance $\mathbf{C}_w$. The linear transformation $\mathbf{H}$ is assumed to be known completely; on the other hand, we only have partial information about the covariance $\mathbf{C}_w$. We further assume that $\mathbf{x}$ satisfies the norm constraint $\|\mathbf{x}\|\le U$ for some scalar $U>0$, where $\|\mathbf{x}\|^2 = \mathbf{x}^*\mathbf{x}$. In practice, if U is not known, then we can estimate it from the data, as we show in the example in Section V (see also [23]).

Minimax estimators minimizing the worst-case MSE for the model (1) have been recently developed in a series of papers [2], [3], [6], [24]. However, while in previous work the noise covariance $\mathbf{C}_w$ was assumed to be known, here we treat the case in which we only have partial information on $\mathbf{C}_w$.

Given the observations $\mathbf{y}$, we would like to design a linear estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$, so that $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ for some $m\times n$ matrix $\mathbf{G}$. A popular design strategy is to choose $\mathbf{G}$ to minimize the MSE between the estimator $\hat{\mathbf{x}}$ and $\mathbf{x}$, which is given by

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

Unfortunately, as can be seen from (2), the MSE of $\hat{\mathbf{x}}$ depends on the unknown vector $\mathbf{x}$ and the unknown covariance $\mathbf{C}_w$. Therefore, in general, we cannot construct an estimator to directly minimize the MSE. To eliminate the dependence of the design criterion on $\mathbf{x}$, an alternative approach is to minimize the data error $\|\hat{\mathbf{y}}-\mathbf{y}\|^2 = (\hat{\mathbf{y}}-\mathbf{y})^*(\hat{\mathbf{y}}-\mathbf{y})$, where $\hat{\mathbf{y}} = \mathbf{H}\hat{\mathbf{x}}$. Clearly this criterion no longer depends on $\mathbf{x}$, and is straightforward to optimize. The optimal solution is the least-squares (LS) estimator

$$\hat{\mathbf{x}}_{\mathrm{LS}} = (\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*\mathbf{y}. \tag{3}$$

If the covariance of the noise $\mathbf{C}_w$ is known, then we can improve the estimator performance by minimizing a weighted data error $(\hat{\mathbf{y}}-\mathbf{y})^*\mathbf{C}_w^{-1}(\hat{\mathbf{y}}-\mathbf{y})$, resulting in the weighted LS (WLS) estimator

$$\hat{\mathbf{x}}_{\mathrm{WLS}} = \left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}. \tag{4}$$

The WLS estimator has the additional property that it minimizes the variance from all *unbiased* estimators. Note, however, that a smaller MSE may be achieved by allowing for a bias.

To develop an estimator that minimizes an objective directly related to the MSE, it was suggested in [3] to seek the linear estimator that minimizes the worst-case MSE over all possible values of $\mathbf{x}$, assuming that $\mathbf{C}_w$ is known. The minimax estimator of this form was developed in [3], [24] for the case in which $\mathbf{x}^*\mathbf{T}\mathbf{x}\le U^2$, where $\mathbf{T}$ is an arbitrary positive definite weighting matrix; for $\mathbf{T}=\mathbf{I}$ the estimator reduces to

$$\hat{\mathbf{x}}_{\mathrm{MX}}(\mathbf{C}_w) = \frac{U^2}{U^2+\gamma_0}\left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y} \tag{5}$$

where $\gamma_0 = \mathrm{Tr}((\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1})$ is the variance of the WLS estimator. Comparison between (5) and (4) shows that the minimax estimator is a shrinkage of the WLS estimator. The advantage of the minimax estimator $\hat{\mathbf{x}}_{\mathrm{MX}}(\mathbf{C}_w)$ is established in [5] in which it is proven that its MSE is smaller than that of $\hat{\mathbf{x}}_{\mathrm{WLS}}$ for all $\|\mathbf{x}\|\le U$.

Since $\hat{\mathbf{x}}_{\mathrm{MX}}(\mathbf{C}_w)$ depends explicitly on $\mathbf{C}_w$, it cannot be implemented if $\mathbf{C}_w$ is not completely specified. Instead, following the ideas of [3], [6], we may seek the estimator that minimizes the worst-case MSE over all possible choices of $\mathbf{x}$ *and* $\mathbf{C}_w$ that are consistent with our prior information on these unknowns.

To reflect the uncertainty in our knowledge of the true covariance matrix $\mathbf{C}_w$, we consider an uncertainty region which resembles the "band model" widely used in the continuous-time case [16]–[18], [20]. Specifically, since $\mathbf{C}_w\succ 0$, there exists an invertible matrix $\mathbf{Q}$ that jointly diagonalizes $\mathbf{C}_w$ and $\mathbf{H}\mathbf{H}^*$ [Theorem 7.6.4] [25], so that

$$\mathbf{H}\mathbf{H}^* = \mathbf{Q}\tilde{\Sigma}\tilde{\Sigma}^*\mathbf{Q}^* \tag{6}$$

$$\mathbf{C}_w = \mathbf{Q}\tilde{\Delta}\mathbf{Q}^*. \tag{7}$$

Here $\tilde{\Sigma}$ is an $n \times m$ matrix defined by

$$\tilde{\Sigma} = \begin{bmatrix} \Sigma \\ \mathbf{0} \end{bmatrix} \tag{8}$$

where $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_m)$ with $\sigma_i > 0, 1 \le i \le m$, and $\tilde{\Delta}$ is an $n \times n$ diagonal matrix defined by

$$\tilde{\Delta} = \begin{bmatrix} \Delta & \mathbf{0} \\ \mathbf{0} & \Theta \end{bmatrix} \tag{9}$$

where $\Delta = \mathrm{diag}(\delta_1, \ldots, \delta_m)$ with $\delta_i > 0, 1 \le i \le m$, and Θ is an arbitrary diagonal matrix of size $(n-m) \times (n-m)$. In our uncertainty model, we assume that the diagonalizing matrix $\mathbf{Q}$ is known. For example, $\mathbf{Q}$ can be chosen as the matrix that jointly diagonalizes $\mathbf{HH}^*$ and an estimate of $\mathbf{C}_w$. The uncertainty in $\mathbf{C}_w$ is due entirely to the diagonal matrix $\tilde{\Delta}$, which is constrained to an arbitrary convex set $\mathcal{U}$ such that $\Delta \succ 0$ (as we will see, the choice of Θ will not effect the estimator). Thus, the covariance $\mathbf{C}_w$ lies in the set $\mathcal{Q}$ defined by

$$\mathcal{Q} = \{\mathbf{C}_w = \mathbf{Q}\tilde{\Delta}\mathbf{Q}^* \mid \tilde{\Delta} \in \mathcal{U}\}. \tag{10}$$

Note that with $\mathbf{HH}^*$ given by (6), it follows that $\mathbf{H}$ has a decomposition

$$\mathbf{H} = \mathbf{Q}\tilde{\Sigma}\mathbf{V}^* \tag{11}$$

for some $m \times m$ unitary matrix $\mathbf{V}$. We will use this form of $\mathbf{H}$ in the development below.

Given the uncertainty model (10), we seek the linear estimator that minimizes the worst-case MSE over all possible values of $\mathbf{x}$ and $\mathbf{C}_w$. Our estimator is therefore the solution to the problem

$$\min_{\hat{\mathbf{x}}=\mathbf{Gy}} \max_{\|\mathbf{x}\| \le U, \tilde{\Delta} \in \mathcal{U}} E\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\}. \tag{12}$$

Substituting $\mathbf{H}$ of (11) and $\mathbf{C}_w$ of (7) into the expression for the MSE (2), our problem can be written explicitly as

$$\min_{\mathbf{G}} \max_{\|\mathbf{x}\| \le U, \tilde{\Delta} \in \mathcal{U}} \mathcal{M}(\mathbf{G}, \tilde{\Delta}, \mathbf{x}) \tag{13}$$

where

$$\mathcal{M}(\mathbf{G}, \tilde{\Delta}, \mathbf{x}) = \mathrm{Tr}(\mathbf{GQ}\tilde{\Delta}\mathbf{Q}^*\mathbf{G}^*) + \mathbf{x}^*(\mathbf{I} - \mathbf{GQ}\tilde{\Sigma}\mathbf{V}^*)^*(\mathbf{I} - \mathbf{GQ}\tilde{\Sigma}\mathbf{V}^*)\mathbf{x}. \tag{14}$$

Since $\mathcal{M}(\mathbf{G}, \tilde{\Delta}, \mathbf{x})$ is strictly convex in $\mathbf{G}$, there is a unique solution to (13), which is given in the following theorem.

Theorem 1: Let $\mathbf{x}$ denote the unknown deterministic parameter vector in the model $\mathbf{y} = \mathbf{Hx} + \mathbf{w}$, where $\mathbf{H}$ is a known $n \times m$ matrix with rank m, and $\mathbf{w}$ is a zero-mean random vector with positive definite covariance $\mathbf{C}_w$. Let $\mathbf{Q}$ be an invertible matrix that jointly diagonalizes $\mathbf{HH}^*$ and $\mathbf{C}_w$ so that $\mathbf{H} = \mathbf{V}\tilde{\Sigma}\mathbf{Q}^*$ for some unitary matrix $\mathbf{V}$ where $\tilde{\Sigma} = [\Sigma\ \mathbf{0}]^*$ is an $n \times m$ diagonal matrix, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_m)$ with $\sigma_i > 0, 1 \le i \le m$, and $\mathbf{C}_w = \mathbf{Q}\tilde{\Delta}\mathbf{Q}^*$ where $\tilde{\Delta} = \mathrm{diag}(\Delta, \Theta), \Delta = \mathrm{diag}(\delta_1, \ldots, \delta_m)$ with $\delta_i > 0, 1 \le i \le m$, and Θ is an $(n-m) \times (n-m)$ diagonal matrix. Then for any convex set $\mathcal{U}$ such that $\Delta \succ 0$, the solution to the problem

$$\min_{\hat{\mathbf{x}}=\mathbf{Gy}} \max_{\|\mathbf{x}\| \le U, \Delta \in \mathcal{U}} E\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\}$$

is

$$\hat{\mathbf{x}}_{\mathrm{MXC}} = \frac{U^2}{U^2 + \sum_{i=1}^m \hat{\delta}_i/\sigma_i^2} \mathbf{V}\Sigma^{-1}\mathbf{Z}^*\mathbf{Q}^{-1}\mathbf{y}$$

where $\mathbf{Z} = [\mathbf{I}_m\ \mathbf{0}]^*$ and the parameters $\{\hat{\delta}_i, 1 \le i \le m\}$ are the solution to

$$\max_{\Delta \in \mathcal{U}} \sum_{i=1}^m \frac{\delta_i}{\sigma_i^2}.$$

The minimax value is given by

$$\min_{\hat{\mathbf{x}}=\mathbf{Gy}} \max_{\|\mathbf{x}\| \le U, \tilde{\Delta} \in \mathcal{U}} E\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\} = \frac{U^2 \sum_{i=1}^m \hat{\delta}_i/\sigma_i^2}{\sum_{i=1}^m \hat{\delta}_i/\sigma_i^2 + U^2}.$$

The Proof of Theorem 1 is deferred to the next section. Before proving the theorem, we now provide an interpretation of the minimax MSE estimator $\hat{\mathbf{x}}_{\mathrm{MXC}}$.

By direct substitution it can be easily shown that if $\mathbf{C}_w \in \mathcal{Q}$ is any positive definite covariance matrix of the form (10), then the estimator $\hat{\mathbf{x}}_{\mathrm{MXC}}$ can be expressed as a shrinkage of the WLS estimator

$$\hat{\mathbf{x}}_{\mathrm{MXC}} = \alpha \left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1} \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y} \tag{15}$$

with shrinkage factor

$$\alpha = \frac{U^2}{U^2 + \sum_{i=1}^m \hat{\delta}_i/\sigma_i^2}. \tag{16}$$

In particular, choosing $\mathbf{C}_w = \hat{\mathbf{C}}_w$ where

$$\hat{\mathbf{C}}_w = \mathbf{Q} \begin{bmatrix} \hat{\Delta} & \mathbf{0} \\ \mathbf{0} & \Theta \end{bmatrix} \mathbf{Q}^* \tag{17}$$

with $\hat{\Delta} = \mathrm{diag}(\hat{\delta}_1, \ldots, \hat{\delta}_m)$ and Θ is an arbitrary diagonal matrix, $\hat{\mathbf{x}}_{\mathrm{MXC}}$ has the form

$$\hat{\mathbf{x}}_{\mathrm{MXC}} = \frac{U^2}{U^2 + \mathrm{Tr}\left((\mathbf{H}^*\hat{\mathbf{C}}_w^{-1}\mathbf{H})^{-1}\right)} \times \left(\mathbf{H}^*\hat{\mathbf{C}}_w^{-1}\mathbf{H}\right)^{-1}\hat{\mathbf{C}}_w^{-1}\mathbf{H}^*\mathbf{y} = \hat{\mathbf{x}}_{\mathrm{MX}}(\hat{\mathbf{C}}_w) \tag{18}$$

where $\hat{\mathbf{x}}_{\mathrm{MX}}(\hat{\mathbf{C}}_w)$ is given by (5) and is the minimax MSE estimator for fixed $\mathbf{C}_w$. Next, we note that for any $\mathbf{C}_w \in \mathcal{Q}$

$$\sum_{i=1}^m \frac{\delta_i}{\sigma_i^2} = \mathrm{Tr}\left(\left(\mathbf{HC}_w^{-1}\mathbf{H}\right)^{-1}\right)$$

which is the variance of the WLS estimator with weight $\mathbf{C}_w^{-1}$. Therefore, we may view $\hat{\mathbf{C}}_w$ as the covariance matrix in $\mathcal{Q}$ which maximizes the WLS variance. This discussion leads to the following two-stage interpretation of the estimator $\hat{\mathbf{x}}_{\mathrm{MXC}}$:

1) find the worst-case covariance matrix $\hat{\mathbf{C}}_w \in \mathcal{Q}$ which maximizes the variance of the WLS estimator of (4);
2) design a minimax estimator assuming known covariance matrix with $\mathbf{C}_w = \hat{\mathbf{C}}_w$.

As a final comment, if $\mathbf{Q}$ is unitary, then

$$\begin{aligned}(\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^* &= \mathbf{V}(\Sigma\mathbf{Z}^*\mathbf{Q}^*\mathbf{Q}\mathbf{Z}\Sigma)^{-1}\Sigma\mathbf{Z}^*\mathbf{Q}^* \\ &= \mathbf{V}\Sigma^{-1}\mathbf{Z}^*\mathbf{Q}^{-1} \end{aligned} \quad (19)$$

where we used the relations $\mathbf{Q}^*\mathbf{Q} = \mathbf{I}, \mathbf{Z}^*\mathbf{Z} = \mathbf{I}$ and $\mathbf{Q}^* = \mathbf{Q}^{-1}$. Therefore, in this case, the minimax MSE estimator is given by

$$\hat{\mathbf{x}}_{\mathrm{MXC}} = \alpha(\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^* \quad (20)$$

which is just a shrinkage of the LS estimator.

III. Proof of the Theorem

We now prove Theorem 1. The proof is comprised of three parts: We first show that the minimax matrix $\mathbf{G}$ that is the solution to (13) has the form

$$\mathbf{G} = \mathbf{V}\mathbf{D}\mathbf{Z}^*\mathbf{Q}^{-1} \quad (21)$$

for some $m \times m$ matrix $\mathbf{D}$. We then show that $\mathbf{D}$ must be a diagonal matrix. Finally, we prove that $\mathbf{D} = \alpha\Sigma^{-1}$ with α given by (16).

A. Form of $\mathbf{G}$

Suppose that $\mathbf{G}$ is the solution to (13), and let $\mathbf{G}' = \mathbf{G}\mathbf{Q}\mathbf{Z}\mathbf{Z}^*\mathbf{Q}^{-1}$. Then

$$\begin{aligned}&\mathcal{M}(\mathbf{G}', \tilde{\Delta}, \mathbf{x}) \\ &\quad = \mathrm{Tr}(\mathbf{G}\mathbf{Q}\mathbf{Z}\Delta\mathbf{Z}^*\mathbf{Q}^*\mathbf{G}^*) \\ &\qquad + \mathbf{x}^*(\mathbf{I} - \mathbf{G}\mathbf{Q}\tilde{\Sigma}\mathbf{V}^*)^*(\mathbf{I} - \mathbf{G}\mathbf{Q}\tilde{\Sigma}\mathbf{V}^*)\mathbf{x} \\ &\quad = \mathcal{M}(\mathbf{G}, \tilde{\Delta}, \mathbf{x}) + \mathrm{Tr}(\mathbf{G}\mathbf{Q}(\mathbf{Z}\Delta\mathbf{Z}^* - \tilde{\Delta})\mathbf{Q}^*\mathbf{G}^*) \end{aligned} \quad (22)$$

where we used the facts that $\mathbf{Z}^*\tilde{\Delta}\mathbf{Z} = \Delta$ and $\mathbf{Z}\mathbf{Z}^*\tilde{\Sigma} = \tilde{\Sigma}$. Since

$$\mathbf{Z}\Delta\mathbf{Z}^* = \begin{bmatrix}\Delta & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} \quad (23)$$

we have that $\mathbf{Z}\Delta\mathbf{Z}^* \preceq \tilde{\Delta}$ for any $\tilde{\Delta} \in \mathcal{U}$, which implies the inequality $\mathcal{M}(\mathbf{G}', \tilde{\Delta}, \mathbf{x}) \leq \mathcal{M}(\mathbf{G}, \tilde{\Delta}, \mathbf{x})$. Since the optimal $\mathbf{G}$ is unique (by strict convexity), we have $\mathbf{G} = \mathbf{G}'$ or

$$\mathbf{G} = \mathbf{G}\mathbf{Q}\mathbf{Z}\mathbf{Z}^*\mathbf{Q}^{-1} = \mathbf{B}\mathbf{Z}^*\mathbf{Q}^{-1} \quad (24)$$

for some $m \times m$ matrix $\mathbf{B}$. Denoting $\mathbf{D} = \mathbf{V}^*\mathbf{B}, \mathbf{G}$ has the required form (21).

B. Diagonal Solution

Next, we show that the optimal $\mathbf{D}$ is a diagonal matrix. To this end, we first note that with $\mathbf{G}$ given by (21), the MSE can be written as

$$\begin{aligned}\mathcal{M}(\mathbf{D}, \Delta, \mathbf{x}) &= \mathrm{Tr}(\mathbf{D}^*\mathbf{D}\Delta) \\ &\quad + \mathbf{x}^*(\mathbf{I} - \mathbf{V}\mathbf{D}\Sigma\mathbf{V}^*)^*(\mathbf{I} - \mathbf{V}\mathbf{D}\Sigma\mathbf{V}^*)\mathbf{x}. \end{aligned} \quad (25)$$

The minimax problem (13) then becomes

$$\begin{aligned}\min_{\mathbf{D}} \max_{\Delta \in \mathcal{U}} \{&\mathrm{Tr}(\mathbf{D}^*\mathbf{D}\Delta) \\ &+ \max_{\|\mathbf{x}\| \leq U} \mathbf{x}^*(\mathbf{I} - \mathbf{V}\mathbf{D}\Sigma\mathbf{V}^*)^*(\mathbf{I} - \mathbf{V}\mathbf{D}\Sigma\mathbf{V}^*)\mathbf{x}\}. \end{aligned} \quad (26)$$

Denoting $\mathbf{z} = \mathbf{V}^*\mathbf{x}$, the inner maximization problem in (26) is equivalent to

$$\max_{\|\mathbf{z}\| \leq U} \mathbf{z}^*(\mathbf{I} - \mathbf{D}\Sigma)^*(\mathbf{I} - \mathbf{D}\Sigma)\mathbf{z} \quad (27)$$

and our problem can be written as

$$\min_{\mathbf{D}} \max_{\Delta \in \mathcal{U}} \mathcal{G}(\mathbf{D}, \Delta) \quad (28)$$

where

$$\mathcal{G}(\mathbf{D}, \Delta) = \mathrm{Tr}(\mathbf{D}^*\mathbf{D}\Delta) + \max_{\|\mathbf{z}\| \leq U} \mathbf{z}^*(\mathbf{I} - \mathbf{D}\Sigma)^*(\mathbf{I} - \mathbf{D}\Sigma)\mathbf{z}. \quad (29)$$

To show that the optimal $\mathcal{D}$ is diagonal, let $\mathbf{J}$ be any diagonal matrix with diagonal elements equal to ± 1. Then

$$\begin{aligned}&\mathcal{G}(\mathbf{J}\mathbf{D}\mathbf{J}, \Delta) \\ &\quad = \mathrm{Tr}(\mathbf{J}\mathbf{D}^*\mathbf{D}\mathbf{J}\Delta) + \max_{\|\mathbf{z}\| \leq U} \mathbf{z}^*\mathbf{J}(\mathbf{I} - \mathbf{D}\Sigma)^*(\mathbf{I} - \mathbf{D}\Sigma)\mathbf{J}\mathbf{z} \\ &\quad = \mathrm{Tr}(\mathbf{D}^*\mathbf{D}\Delta) + \max_{\|\mathbf{z}'\| \leq U} \mathbf{z}'^*(\mathbf{I} - \mathbf{D}\Sigma)^*(\mathbf{I} - \mathbf{D}\Sigma)\mathbf{z}' \\ &\quad = \mathcal{G}(\mathbf{D}, \Delta) \end{aligned} \quad (30)$$

where we denoted $\mathbf{z}' = \mathbf{J}\mathbf{z}$ and we used the relations $\mathbf{J}^* = \mathbf{J}, \mathbf{J}^2 = \mathbf{I}$ and $\mathbf{J}\mathbf{B} = \mathbf{B}\mathbf{J}$ for any diagonal matrix $\mathbf{B}$. Since $\mathcal{G}(\mathbf{J}\mathbf{D}\mathbf{J}, \Delta) = \mathcal{G}(\mathbf{D}, \Delta)$ for any $\mathbf{D}$, if $\mathbf{D}$ is an optimal matrix then $\mathbf{J}\mathbf{D}\mathbf{J}$ is also an optimal solution; therefore, the optimal $\mathbf{D}$ satisfies $\mathbf{D} = \mathbf{J}\mathbf{D}\mathbf{J}$ for any diagonal $\mathbf{J}$ with diagonal elements ± 1, from which we conclude that $\mathbf{D}$ must be diagonal.

C. Minimax Estimator

Denoting by d_i the diagonal elements of $\mathbf{D}$, the problem (28) is equivalent to

$$\min_{d_i} \max_{\delta_i \in \mathcal{U}, \sum_{i=1}^m |z_i|^2 \leq U^2} \left\{\sum_{i=1}^m d_i^2\delta_i + (1 - d_i\sigma_i)^2|z_i|^2\right\}. \quad (31)$$

In the Appendix we show that the values d_i that are the solution to (31) satisfy $0 \leq d_i \leq 1/\sigma_i$. Letting $s_i = |z_i|^2$, we can therefore write (31) as

$$\min_{d_i \in \mathcal{D}} \max_{\delta_i \in \mathcal{U}, s_i \in \mathcal{S}} \left\{\sum_{i=1}^m d_i^2\delta_i + (1 - d_i\sigma_i)^2 s_i\right\} \quad (32)$$

where

$$\mathcal{D} = \{0 \leq d_i \leq 1/\sigma_i, \quad 1 \leq i \leq m\} \quad (33)$$

and

$$\mathcal{S} = \left\{\sum_{i=1}^m s_i \leq U^2, \quad s_i \geq 0, \quad 1 \leq i \leq m\right\}. \quad (34)$$

Note that $\mathcal{D}$ is a convex and compact set, and $\mathcal{U}$ and $\mathcal{S}$ are convex sets. In addition, the objective in (32) is convex in the minimization arguments d_i and concave (linear) in the maximization arguments s_i and δ_i. Therefore, we can exchange the order of the minimization and maximization [26] so that

$$\min_{d_i\in\mathcal{D}}\max_{\delta_i\in\mathcal{U},s_i\in\mathcal{S}}\left\{\sum_{i=1}^m d_i^2\delta_i+(1-d_i\sigma_i)^2 s_i\right\} = \max_{\delta_i\in\mathcal{U},s_i\in\mathcal{S}}\min_{d_i\in\mathcal{D}}\left\{\sum_{i=1}^m d_i^2\delta_i+(1-d_i\sigma_i)^2 s_i\right\}. \quad (35)$$

Using (35) we can solve the minimax problem of (32) by first solving the maximin problem, which is easier. We then prove that the values d_i that are the solution to the maximin problem are also the solution to the minimax problem, by showing that with that choice of d_i, the minimax value is equal to the optimal value (which is given by the optimal maximin value).

To develop a solution to the maximin problem of (35) we first consider the inner minimization problem

$$\min_{d_i\in\mathcal{D}}\left\{\sum_{i=1}^m d_i^2\delta_i+(1-d_i\sigma_i)^2 s_i\right\}. \quad (36)$$

Differentiating the objective with respect to d_i and equating to 0

$$d_i=\frac{s_i\sigma_i}{s_i\sigma_i^2+\delta_i}. \quad (37)$$

Substituting d_i back into (35), we are left with the double maximization problem

$$\max_{\delta_i\in\mathcal{U}}\max_{s_i\in\mathcal{S}}\sum_{i=1}^m\frac{\delta_i s_i}{\delta_i+s_i\sigma_i^2}. \quad (38)$$

Considering first the maximization with respect to s_i, we form the Lagrangian

$$\mathcal{L}=-\sum_{i=1}^m\frac{\delta_i s_i}{\delta_i+s_i\sigma_i^2}+\mu\sum_{i=1}^m s_i, \quad (39)$$

where $\mu\geq 0$, and from the complementary slackness condition we must have that

$$\mu\left(\sum_{i=1}^m s_i-U^2\right)=0. \quad (40)$$

Differentiating $\mathcal{L}$ with respect to s_i and equating to 0

$$\frac{\delta_i^2}{(\delta_i+s_i\sigma_i^2)^2}=\mu. \quad (41)$$

Since $\delta_i>0$, it follows that $\mu>0$, and

$$s_i=\frac{(1-\sqrt{\mu})\delta_i}{\sqrt{\mu}\sigma_i^2}. \quad (42)$$

To satisfy (40) we must have $\sum_{i=1}^m s_i=U^2$, from which we conclude that

$$\sqrt{\mu}=\frac{\sum_{i=1}^m\delta_i/\sigma_i^2}{U^2+\sum_{i=1}^m\delta_i/\sigma_i^2}, \quad (43)$$

and

$$s_i=\frac{U^2\delta_i/\sigma_i^2}{\sum_{j=1}^m\delta_j/\sigma_j^2}. \quad (44)$$

Substituting (44) back into (38), our problem becomes

$$\max_{\delta_i\in\mathcal{U}}\frac{U^2\sum_{i=1}^m\delta_i/\sigma_i^2}{U^2+\sum_{i=1}^m\delta_i/\sigma_i^2}. \quad (45)$$

Since the function $U^2x/(U^2+x)$ is monotonically increasing in $x>0$, (45) is equivalent to

$$\max_{\delta_i\in\mathcal{U}}\sum_{i=1}^m\frac{\delta_i}{\sigma_i^2}. \quad (46)$$

Let $\hat{\delta}_i$ denote the optimal solution to (46). It then follows that

$$\max_{\tilde{\Delta}\in\mathcal{U},\|\mathbf{x}\|\leq U}\min_{\mathbf{G}}\mathcal{M}(\mathbf{G},\tilde{\Delta},\mathbf{x})=\frac{U^2\sum_{i=1}^m\hat{\delta}_i/\sigma_i^2}{U^2+\sum_{i=1}^m\hat{\delta}_i/\sigma_i^2} \quad (47)$$

and the optimal value of $\mathbf{G}$ is given by (48), where $\mathbf{D}$ is diagonal with diagonal elements d_i that follow from (37) and (44) as

$$d_i=\frac{U^2/\sigma_i}{U^2+\sum_{j=1}^m\hat{\delta}_j/\sigma_j^2}. \quad (48)$$

With d_i given by (48)

$$\mathbf{D}=\frac{U^2}{U^2+\sum_{j=1}^m\hat{\delta}_j/\sigma_j^2}\Sigma^{-1}. \quad (49)$$

From (35) and (47) it follows that

$$\min_{\mathbf{G}}\max_{\tilde{\Delta}\in\mathcal{U},\|\mathbf{x}\|\leq U}\mathcal{M}(\mathbf{G},\tilde{\Delta},\mathbf{x})=\frac{U^2\sum_{i=1}^m\hat{\delta}_i/\sigma_i^2}{U^2+\sum_{i=1}^m\hat{\delta}_i/\sigma_i^2}. \quad (50)$$

Therefore, to show that $\mathbf{D}$ of (49) is also the solution to the minimax problem of (13) it is sufficient to show that with this choice of $\mathbf{D}$,

$$\max_{\tilde{\Delta}\in\mathcal{U},\|\mathbf{x}\|\leq U}\mathcal{M}(\mathbf{D},\tilde{\Delta},\mathbf{x})=\frac{U^2\sum_{i=1}^m\hat{\delta}_i/\sigma_i^2}{U^2+\sum_{i=1}^m\hat{\delta}_i/\sigma_i^2}. \quad (51)$$

Using (32) we have that with $\mathbf{D}$ given by (49)

$$\begin{aligned}\max_{\tilde{\Delta}\in\mathcal{U},\|\mathbf{x}\|\leq U}&\mathcal{M}(\mathbf{D},\tilde{\Delta},\mathbf{x})\\ &=\max_{\delta_i\in\mathcal{U},s_i\in\mathcal{S}}\left\{\frac{U^4\sum_{i=1}^m\delta_i/\sigma_i^2+\left(\sum_{i=1}^m\hat{\delta}_i/\sigma_i^2\right)^2\sum_{i=1}^m s_i}{\left(U^2+\sum_{i=1}^m\hat{\delta}_i/\sigma_i^2\right)^2}\right\}\\ &=\frac{U^2\sum_{i=1}^m\hat{\delta}_i/\sigma_i^2}{U^2+\sum_{i=1}^m\hat{\delta}_i/\sigma_i^2}\end{aligned} \quad (52)$$

where we used that fact that $\max_{s_i\in\mathcal{S}}\sum_{i=1}^m s_i=U^2$, and $\max_{\delta_i\in\mathcal{U}}\sum_{i=1}^m\delta_i/\sigma_i^2=\sum_{i=1}^m\hat{\delta}_i/\sigma_i^2$. Thus, $\mathbf{D}$ given by (49) is also minimax optimal, which completes the proof of the theorem.

IV. Examples of Uncertainty Sets

We now consider some specific examples of covariance uncertainty sets $\mathcal{U}$ and derive the corresponding estimators.

A. Trace and Lower Bound Uncertainty

Suppose that the uncertainty set $\mathcal{U}$ is defined by

$$\mathcal{U}=\left\{\delta_i\,\middle|\,\delta_i\geq l_i,\quad\sum_{i=1}^m\delta_i\leq\alpha\right\} \quad (53)$$

for some $\alpha \geq \sum_{i=1}^{m} l_i \geq 0$. To determine the minimax MSE estimator in this case we need to find the values δ_i that are the solution to

$$\max_{\delta_i \geq l_i, \sum_{i=1}^{m} \delta_i \leq \alpha} \sum_{i=1}^{m} \frac{\delta_i}{\sigma_i^2}. \tag{54}$$

Let $\sigma_1 \geq \cdots \geq \sigma_m$. Then, for any choice of $\delta_i \in \mathcal{U}$,

$$\begin{aligned}\sum_{i=1}^{m} \frac{\delta_i}{\sigma_i^2} &= \beta + \sum_{i=1}^{m} \frac{(\delta_i - l_i)}{\sigma_i^2} \\ &\leq \beta + \frac{1}{\sigma_m^2} \sum_{i=1}^{m} (\delta_i - l_i) \\ &\leq \beta + \frac{\alpha}{\sigma_m^2} - \frac{1}{\sigma_m^2} \sum_{i=1}^{m} l_i \end{aligned} \tag{55}$$

where $\beta = \sum_{i=1}^{m} (l_i/\sigma_i^2)$. We have equality in (55) if $\delta_i = \hat{\delta}_i$, with

$$\hat{\delta}_i = \begin{cases} \alpha - \sum_{i=1}^{m-1} l_i, & i = m; \\ l_i, & i \neq m. \end{cases} \tag{56}$$

The least-favorable matrix in this case is obtained when the noise power is all directed to the generalized eigenvector corresponding to the smallest power of the transformation.

B. Trace and Band Uncertainty

We now consider the case in which we also have an upper bound on δ_i. Specifically, we assume that

$$\mathcal{U} = \left\{ \delta_i \,\middle|\, l_i \leq \delta_i \leq u_i, \sum_{i=1}^{m} \delta_i \leq \alpha \right\} \tag{57}$$

where $l_i < u_i$. Determining the minimax MSE estimator in this case requires solving the problem

$$\max_{l_i \leq \delta_i \leq u_i, \sum_{i=1}^{m} \delta_i \leq \alpha} \sum_{i=1}^{m} \frac{\delta_i}{\sigma_i^2}. \tag{58}$$

To find the optimal δ_i we form the Lagrangian

$$\mathcal{L} = -\sum_{i=1}^{m} \frac{\delta_i}{\sigma_i^2} + \rho \left(\sum_{i=1}^{m} \delta_i - \alpha \right) + \sum_{i=1}^{m} \mu_i (\delta_i - u_i) + \sum_{i=1}^{m} \zeta_i (l_i - \delta_i) \tag{59}$$

where $\rho, \mu_i, \zeta_i \geq 0$ and satisfy the complementary slackness conditions

$$\rho \left(\sum_{i=1}^{m} \delta_i - \alpha \right) = 0 \tag{60}$$

$$\mu_i (\delta_i - u_i) = 0, \quad 1 \leq i \leq m \tag{61}$$

$$\zeta_i (\delta_i - l_i) = 0, \quad 1 \leq i \leq m. \tag{62}$$

Differentiating $\mathcal{L}$ with respect to δ_i and equating to 0,

$$-\frac{1}{\sigma_i^2} + \rho + \mu_i - \zeta_i = 0, \quad 1 \leq i \leq m. \tag{63}$$

Suppose that for some index $1 \leq j \leq m, \mu_j > 0$. Then from (61), $\delta_j = u_j$, and from (62), $\zeta_j = 0$. Since $\mu_j > 0$, from (63) we then have that

$$\rho < \frac{1}{\sigma_j^2} \leq \frac{1}{\sigma_i^2}, \quad j \leq i \leq m. \tag{64}$$

Similarly, if for some index $1 \leq j \leq m, \zeta_j > 0$, then from (61), (62) and (63)

$$\rho > \frac{1}{\sigma_j^2} \geq \frac{1}{\sigma_i^2}, \quad 1 \leq i \leq j. \tag{65}$$

Finally, if $\mu_j = \zeta_j = 0, 1 \leq j \leq m$, then $\rho = 1/\sigma_j^2$, and there is no further constraint on $l_j \leq \delta_j \leq u_j$.

Clearly, to maximize $\sum_{i=1}^{m} \delta_i/\sigma_i^2$ we need to choose the values of $\delta_i, 1 \leq i \leq m$ to be as large as possible. If $\sum_{i=1}^{m} u_i \leq \alpha$, then the optimal choice of $\delta_i, 1 \leq i \leq m$ is $\delta_i = u_i$. If $\sum_{i=1}^{m} u_i > \alpha$, then we cannot have $\delta_i = u_i, 1 \leq i \leq m$. In this case, from our analysis above it follows that $\delta_i = l_i, 1 \leq i \leq k$ for some value k and $\delta_i = u_i, k + r_k + 1 \leq i \leq m$, where r_k is the multiplicity of σ_{k+1}^2. The remaining values $\delta_i, k+1 \leq i \leq k + r_k$ are chosen such that $\sum_{i=i}^{m} \delta_i = \alpha$ and such that $l_i \leq \delta_i \leq u_i$. Since we seek to maximize $\sum_{i=1}^{m} \delta_i/\sigma_i^2$, we would like k to be as small as possible. For a specific choice of k, let

$$a = \alpha - \sum_{i=1}^{k} l_i - \sum_{i=k+r_k+1}^{m} u_i. \tag{66}$$

Then, we need to choose $l_i \leq \delta_i \leq u_i$ such that $\sum_{i=k+1}^{k+r_k} \delta_i = a$. Since $\sigma_i = \sigma_{k+1}, k+1 \leq i \leq k + r_k$, the choice of $\delta_i, k+1 \leq i \leq k + r_k$ does not effect the objective function. Therefore, any choice of δ_i satisfying $l_i \leq \delta_i \leq u_i$ and $\sum_{i=k+1}^{k+r_k} \delta_i = a$ is optimal.

V. Example

We now consider an example illustrating the performance of the minimax MSE estimator.

Consider the problem of estimating an acoustical impulse response using a known input signal. The desired impulse response, of length $m = 100$, is simulated using the image method [27],[1] and shown in Fig. 1. A speech signal taken from the TIMIT database [28] is used to excite the filter, and the filtered output is contaminated by a white Gaussian noise vector with covariance $\mathbf{C}_w = \sigma^2 \mathbf{I}$ for some $\sigma^2 > 0$.

Denoting by $\mathbf{x}$ the filter taps to be estimated, and defining $\mathbf{H}$ as a 200×100 Toeplitz convolution matrix built from the known excitation signal, our problem reduces to that of estimating $\mathbf{x}$ in the linear model $\mathbf{y} = \mathbf{Hx} + \mathbf{w}$.

To estimate the unknown filter $\mathbf{x}$ we first estimate the covariance matrix $\mathbf{C}_w$ using $N = 250$ realizations of the noise $\mathbf{w}$. We then find a matrix $\mathbf{Q}$ that jointly diagonalizes $\mathbf{HH}^*$ and the estimate of $\mathbf{C}_w$; we denote the diagonal elements of $\mathbf{HH}^*$ and the estimate of $\mathbf{C}_w$ in this representation by σ_i and $\tilde{\delta}_i$, respectively.

[1]This method constructs the impulse response between two points in a small rectangular room using images of the original source. Each image is positioned in a symmetric manner with respect to the six faces of the room, and the received signal considers wall absorptions.

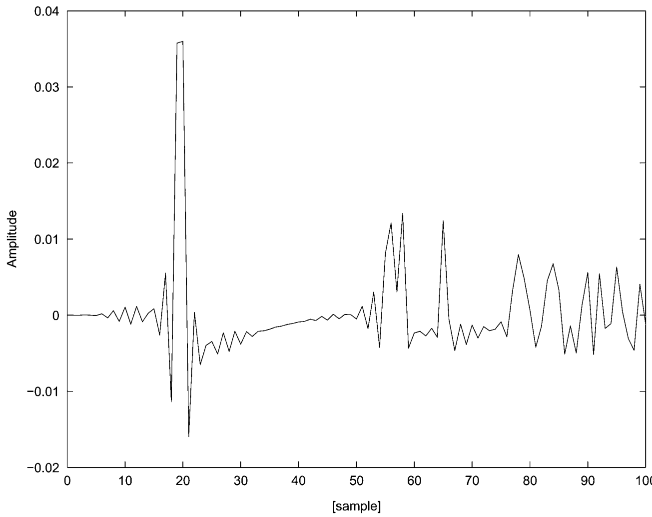

Fig. 1. Impulse response used in the simulation.

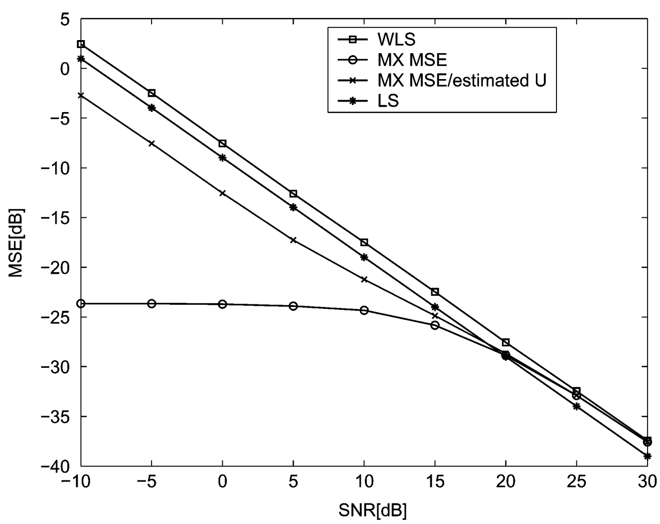

Fig. 2. Mean-squared error as a function of the SNR using the least-squares (LS), weighted LS (WLS), and minimax MSE estimators, where the minimax MSE estimator is implemented with $U = \|\mathbf{x}\|$ and $U = \|\hat{\mathbf{x}}_{\rm LS}\|$.

We next choose an uncertainty set Δ to reflect the uncertainty in our estimate $\tilde{\delta}_i$. In this example, we choose the band uncertainty model of Section IV-B with $l_i = (1-r)\tilde{\delta}_i, u_i = (1+r)\tilde{\delta}_i$ and $\alpha = \sum_i \tilde{\delta}_i$ for some $0 \leq r \leq 1$. In the simulations below, $r = 0.4$.

To estimate the vector $\mathbf{x}$ from the observations $\mathbf{y}$ we consider 3 different estimators. The LS estimator, a WLS estimator with covariance chosen to be equal to the least-favorable covariance matrix, and the minimax MSE estimator $\hat{\mathbf{x}}_{\rm MXC}$ of Theorem 1 (see also Section IV-B). In implementing the minimax MSE estimator, we consider two choices of the norm bound U. In the first choice, the norm of $\mathbf{x}$ is assumed to be known, so that $U = \|\mathbf{x}\|$. In the second choice, U is estimated as $U = \|\hat{\mathbf{x}}_{\rm LS}\|$, where $\hat{\mathbf{x}}_{\rm LS}$ is the LS estimator.

In Fig. 2 we plot the MSE of the LS, WLS and the minimax MSE estimator with $U = \|\mathbf{x}\|$ and $U = \|\hat{\mathbf{x}}_{\rm LS}\|$ as a function of the SNR,

$$\mathrm{SNR} = 10\log_{10}\frac{\|\mathbf{Hx}\|^2}{\mathrm{Tr}(\mathbf{C}_w)}. \tag{67}$$

The MSE is obtained by averaging the squared-norm of the estimation error over 1000 realizations.

As can be seen from the figure, by considering an uncertainty interval around our estimate, we can improve the performance over the LS approach. We note however that in the high SNR regime, the LS estimate is superior to the other methods, as it makes no use of the poorly estimated $\mathbf{C}_w$.

APPENDIX

In this Appendix, we show that the values d_i that are the solution to (31) satisfy $0 \leq d_i \leq 1/\sigma_i$.

To this end, let

$$\mathcal{T}(\mathbf{d}) = \sum_{i=1}^m d_i^2\delta_i + (1-d_i\sigma_i)^2|z_i|^2 \tag{68}$$

where $\mathbf{d} = (d_1,\ldots,d_m)$. Suppose that for some value of $k, d_k < 0$, and let $\mathbf{d}_k = (d_1,\ldots,d_{k-1},-d_k,d_{k+1},\ldots d_m)$. Since $d_k^2 = (-d_k)^2$ and $(1-d_k\sigma_k)^2 > (1+d_k\sigma_k)^2, \mathcal{T}(\mathbf{d}) > \mathcal{T}(\mathbf{d}_k)$, and

$$\max_{\delta_i\in\mathcal{U},\sum_{i=1}^m|z_i|^2\leq U^2}\mathcal{T}(\mathbf{d}) > \max_{\delta_i\in\mathcal{U},\sum_{i=1}^m|z_i|^2\leq U^2}\mathcal{T}(\mathbf{d}_k) \tag{69}$$

so that $\mathbf{d}_k$ leads to a lower objective value then $\mathcal{T}(\mathbf{d})$. Therefore, at the minimal value $d_k \geq 0$. Similarly, suppose that $d_k > 1/\sigma_k$ for some k and let $\mathbf{d}_k = (d_1,\ldots,d_{k-1},\tilde{d}_k,d_{k+1},\ldots d_m)$, where

$$\tilde{d}_k = \begin{cases} 2/\sigma_k - d_k, & d_k \leq 2/\sigma_k \\ 0, & d_k > 2/\sigma_k. \end{cases} \tag{70}$$

Then

$$(1-\tilde{d}_k\sigma_k)^2 + \tilde{d}_k^2 = \begin{cases} (d_k\sigma_k - 1)^2 + (2/\sigma_k - d_k)^2, & d_k \leq 2/\sigma_k \\ 1, & d_k > 2/\sigma_k. \end{cases} \tag{71}$$

Now, if $1/\sigma_k < d_k \leq 2/\sigma_k$, then

$$(2/\sigma_k - d_k)^2 < 1/\sigma_k^2 < d_k^2. \tag{72}$$

Therefore, in this case $\mathcal{T}(\mathbf{d}) > \mathcal{T}(\mathbf{d}_k)$. If, on the other hand, $d_k > 2/\sigma_k$, then $(1-d_k\sigma_k)^2 > 1$, so that again, $\mathcal{T}(\mathbf{d}) > \mathcal{T}(\mathbf{d}_k)$.

ACKNOWLEDGMENT

The author would like to thank T. Dvorkind for his help with the simulations and the anonymous reviewers for their constructive comments, which greatly improved the presentation.

REFERENCES

[1] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.
[2] 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.
[3] ——, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 168–181, Jan. 2005.
[4] Y. C. Eldar, "Robust estimation in linear models with a random model matrix," *IEEE Trans. Signal Process.*, to be published.
[5] Z. Ben-Haim and Y. C. Eldar, "Maximum set estimators with bounded estimation error," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3172–3182, Aug. 2005.

[6] M. S. Pinsker, "Optimal filtering of square-integrable signals in Gaussian noise," *Problems Inform. Trans.*, vol. 16, pp. 120–133, 1980.

[7] J. Pilz, "Minimax linear regression estimation with symmetric parameter restrictions," *J. Stat. Planning Inference*, vol. 13, pp. 297–318, 1986.

[8] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[9] K. F. Gauss, *Theory of Motion of Heavenly Bodies Moving About the Sun in Conic Sections*. New York: Dover, 1963.

[10] L. J. Gleser, "Minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix," *Ann. Statist.*, vol. 14, no. 4, pp. 1625–1633, 1986.

[11] J. Berger and L. R. Haff, "A class of minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix," *Statist. Decisions*, vol. 1, no. 2, pp. 105–129, 1982/83.

[12] L. J. Gleser, "Minimax estimation of a normal mean vector when the covariance matrix is unknown," *Ann. Statist.*, vol. 7, no. 4, pp. 838–846, 1979.

[13] J. Berger, M. E. Bock, L. D. Brown, G. Casella, and L. J. Gleser, "Minimax estimation of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix," *Ann. Statist.*, vol. 5, no. 4, pp. 763–771, 1977.

[14] P. E. Lin and H. L. Tsai, "Generalized bayes minimax estimators of the multivariate normal mean with unknown covariance matrix," *Ann. Statist.*, vol. 1, pp. 142–145, 1973.

[15] L. Breiman, "A note on minimax filtering," *Annals Probability*, vol. 1, pp. 175–179, 1973.

[16] S. A. Kassam and T. L. Lim, "Robust Wiener filters," *J. Franklin Inst.*, vol. 304, pp. 171–185, Oct./Nov. 1977.

[17] 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.

[18] V. P. Kuznetsov, "Stable detection when the signal and spectrum of normal noise are inaccurately known," *Telecommun. Radio Eng.*, vol. 30/31, pp. 58–64, 1976. (English transl.).

[19] H. V. Poor, "On robust Wiener filtering," *IEEE Trans. Autom. Control*, vol. AC-25, no. 3, pp. 531–536, Jun. 1980.

[20] 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.

[21] J. Franke, "Minimax-robust prediction of discrete time series," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 68, pp. 337–364, 1985.

[22] A. Beck, A. Ben-Tal, and Y. C. Eldar, "Robust mean-squared error estimation of multiple signals in linear systems affected by model and noise uncertainties," *Math Prog.*, to be published.

[23] Z. Ben-Haim and Y. C. Eldar, "Minimax estimators dominating the least-squares estimator," in *Proc. Int. Conf. Acoustics. Speech, Signal Processing (ICASSP 2005)*, vol. 4, Mar. 18–23, 2005, pp. iv53–iv56.

[24] Y. C. Eldar, "Comparing between estimation approaches: Admissible and dominating linear estimators," *IEEE Trans. Signal Process.*, to be published.

[25] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[26] M. Sion, "On general minimax theorems," *Pac. J. Math.*, vol. 8, pp. 171–176, 1958.

[27] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[28] The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus, Oct. 1991. CD-RDM NIST Speech Disc 1-1.1, National Institute of Standards and Technology.

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

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

Dr. Eldar is a member of the IEEE Signal Processing Theory and Methods technical committee and an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. She was in the program for outstanding students at TAU from 1992 to 1996. In 1998, she held the Rosenblith Fellowship for study in Electrical Engineering at MIT, and in 2000, she held an IBM Research Fellowship. She is currently a Horev Fellow of the Leaders in Science and Technology program at The Technion and an Alon Fellow. In 2004, she was awarded the Wolf Foundation Krill Prize for Excellence in Scientific Research, and in 2005, the Andre and Bella Meyer Lectureship.