Linear Minimax Regret Estimation of Deterministic Parameters with Bounded Data Uncertainties

Yonina C. Eldar, *Member, IEEE*, Aharon Ben-Tal, and Arkadi Nemirovski

Abstract—We develop a new linear estimator for estimating an unknown parameter vector x in a linear model in the presence of bounded data uncertainties. The estimator is designed to minimize the worst-case *regret* over all bounded data vectors, namely, the worst-case difference between the mean-squared error (MSE) attainable using a linear estimator that does not know the true parameters x and the optimal MSE attained using a linear estimator that knows x. We demonstrate through several examples that the minimax regret estimator can significantly increase the performance over the conventional least-squares estimator, as well as several other least-squares alternatives.

Index Terms—Deterministic parameter estimation, linear estimation, mean squared error bounded data uncertainties estimation, minimax estimation, regret.

I. INTRODUCTION

THE problem of estimating a vector of unknown parameters x from noisy observations $\mathbf{y} = \mathbf{Hx} + \mathbf{w}$, where $\mathbf{H}$ is a known matrix and $\mathbf{w}$ is a noise vector, arises in many different fields in science and engineering and has consequently attracted much attention in the estimation literature.

If the unknown parameter vector x is assumed to be random with known second-order statistics, then the linear estimator minimizing the mean-squared error (MSE) is the well-known Wiener estimator [1], [2]. However, in many problems of practical interest, there is no statistical information available on x so that x is treated as an unknown deterministic parameter vector. In this case, the MSE of an estimator $\hat{\mathbf{x}}$ of x will in general depend explicitly on the unknown parameter vector x and, therefore, cannot be minimized directly.

Since the MSE between $\hat{\mathbf{x}}$ and x generally depends on x, a common approach is to seek linear estimators that minimize some function of the data error $\hat{\mathbf{y}} - \mathbf{y}$, where $\hat{\mathbf{y}} = \mathbf{H}\hat{\mathbf{x}}$ is the estimated data vector. The celebrated least-squares estimator, first studied by Gauss [3], seeks the estimator $\hat{\mathbf{x}}$ of x that minimizes the squared-norm of the data error $\|\hat{\mathbf{y}} - \mathbf{y}\|^2$. It is well known that the least-squares estimate is also the best linear unbiased estimator [4], i.e., it has the smallest variance among all

linear *unbiased* estimators. On the negative side, an unbiased estimator does not necessarily lead to a small MSE. In fact, it is well known that in many cases, the least-squares estimator can result in a large MSE.

Various modifications of the least-squares estimator for the case in which the data model holds, so that $\mathbf{y} = \mathbf{Hx} + \mathbf{w}$ with $\mathbf{H}$ and $\mathbf{y}$ known exactly, and x represents a deterministic parameter vector, have been proposed. Among the alternatives are Tikhonov regularization [5], which is also known in the statistical literature as the ridge estimator [6], the linear shrunken estimator [7], and the covariance shaping least-squares estimator [8]. In general, these linear least-squares alternatives attempt to reduce the MSE in estimating x by allowing for a bias. Each of the estimators above can be shown to be a solution to an optimization problem that involves minimizing some function that depends on the *data error*.

In an estimation context, we typically would like to minimize the *estimation error*, rather than the data error. To this end, we assume that x is known to satisfy a (possibly weighted) norm constraint and then seek a robust estimator whose performance is reasonably good across all possible choices of the parameters x in the region of uncertainty. The most common approach for designing robust estimators is the minimax MSE approach, initiated by Huber [9], [10], in which we seek the estimator that minimizes the worst-case MSE in the region of uncertainty. This approach has been applied to a variety of different estimation problems in which the unknown parameter vector x is assumed to be random, but the statistics of x are not completely specified [11]–[17]. This approach has also been applied to the case in which x is deterministic, under the limiting assumptions that $\mathbf{H} = \mathbf{I}$, and each element $\mathbf{x}_i$ of x is estimated from the corresponding observation $\mathbf{y}_i$ of y [18]. Note that effectively, this is equivalent to assuming that the noise vector has independent elements. The minimax approach, in which the goal is to optimize the worst-case performance, is one of the major techniques for designing robust systems with respect to modeling uncertainties and has been applied to many problems in detection and estimation [19]–[21].

Following the popular minimax approach, we may seek the linear estimator that minimizes the worst-case MSE over all possible values of x that satisfy a weighted norm constraint. The minimax estimator of this form for arbitrary matrices $\mathbf{H}$ and arbitrary noise vectors $\mathbf{w}$ is developed in [22], in which the case of uncertainties in the model matrix $\mathbf{H}$ is also considered.

Although the minimax approach has enjoyed widespread use in the design of robust methods for signal processing and communication [19], [21], its performance is often unsatisfactory. The main limitation of this approach is that it tends to be overly

Y. C. Eldar is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: yonina@ee.technion.ac.il).

A. Ben-Tal and A. Nemirovski are with MINERVA Optimization Center, Department of Industrial Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: morbt@ie.technion.ac.il; nemirovs@ie.technion.ac.il).

conservative since it optimizes the performance for the worst possible choice of unknowns. As we show in the context of concrete examples in Section VI, this can often lead to degraded performance.

To improve the performance of the minimax MSE estimator, we propose, in Section III, a new approach to linear estimation in which we seek a linear estimator whose performance is as close as possible to that of the optimal linear estimator, i.e., the one minimizing the MSE when $\mathbf{x}$ is assumed to be known. Specifically, we seek the estimator that minimizes the worst-case *regret*, which is the difference between the MSE of the linear estimator that does not know $\mathbf{x}$ and the smallest attainable MSE with a linear estimator that knows $\mathbf{x}$. Note that as we show in Section III, since we are restricting ourselves to linear estimators, we cannot achieve zero MSE even in the case in which the parameters $\mathbf{x}$ are known. By considering the *difference* between the MSE and the optimal MSE rather than the MSE directly, we can counterbalance the conservative character of the minimax approach, as is evident in the examples we consider in Section VI.

The minimax regret concept has recently been used to develop a linear estimator for the unknown vector $\mathbf{x}$ in the same linear model considered in this paper, but for the case that $\mathbf{x}$ is random with an unknown covariance matrix [17]. Similar competitive approaches have been used in a variety of other contexts, for example, universal source coding [23], hypothesis testing [24], [25], and prediction [26].

The paper is organized as follows. In Section II, we provide an overview of our problem. In Section III, we develop the form of the minimax regret estimator when the uncertainty region is defined by $\mathbf{x}^*\mathbf{T}\mathbf{x} \leq L^2$ for some positive definite weighting matrix $\mathbf{T}$. For analytical tractability, we restrict our attention to matrices $\mathbf{T}$ such that $\mathbf{T}$ and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have the same eigenvector matrix,[1] where $\mathbf{C}_w$ is the noise covariance matrix. We then specialize the results to the case in which $\mathbf{T} = \mathbf{H}\mathbf{C}_w^{-1}\mathbf{H}$ in Section IV and to the case in which $\mathbf{T} = \mathbf{I}$ in Section V. In these special cases, we show that the minimax regret estimator can be derived as the solution to explicit, simple, and computationally tractable convex optimization problems. Section VI presents several examples illustrating the performance advantage of the minimax regret estimator.

II. PROBLEM FORMULATION AND MAIN RESULTS

We denote vectors in $\mathbb{C}^m$ by boldface lowercase letters and matrices in $\mathbb{C}^{n\times m}$ by boldface uppercase letters. $\mathbf{I}$ denotes the identity matrix of appropriate dimension, $(\cdot)^*$ denotes the Hermitian conjugate of the corresponding matrix, and $\hat{(\cdot)}$ denotes an estimated vector or matrix.

Consider the problem of estimating the unknown deterministic parameter vector $\mathbf{x}$ in the linear model

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

where $\mathbf{H}$ is a known $n \times m$ matrix with full rank m, and $\mathbf{w}$ is a zero-mean random vector with covariance $\mathbf{C}_w$. We assume that $\mathbf{x}$ is known to satisfy the weighted norm constraint $\|\mathbf{x}\|_{\mathbf{T}} \leq L$ for some positive definite covariance $\mathbf{T}$ and scalar $L > 0$, where $\|\mathbf{x}\|_{\mathbf{T}}^2 = \mathbf{x}^*\mathbf{T}\mathbf{x}$.

We estimate $\mathbf{x}$ using a linear estimator so that $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ for some $m \times n$ matrix $\mathbf{G}$. The variance of the estimator $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ is given by

$$V(\hat{\mathbf{x}}) = E\left(\|\hat{\mathbf{x}} - E(\hat{\mathbf{x}})\|^2\right) = \mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) \tag{2}$$

and the bias of the estimator is

$$B(\hat{\mathbf{x}}) = \mathbf{x} - E(\hat{\mathbf{x}}) = (\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{x}. \tag{3}$$

We would like to design an estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ to minimize the MSE, which is given by

$$\begin{aligned} E\left(\|\hat{\mathbf{x}} - \mathbf{x}\|^2\right) &= V(\hat{\mathbf{x}}) + \|B(\hat{\mathbf{x}})\|^2 \\ &= \mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \mathbf{x}^*(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{x}. \end{aligned} \tag{4}$$

Since $B(\hat{\mathbf{x}})$ depends explicitly on the unknown parameter vector $\mathbf{x}$, we cannot choose an estimate to directly minimize the MSE (4).

A common approach is to restrict the estimator $\hat{\mathbf{x}}$ to be unbiased so that $B(\hat{\mathbf{x}}) = \mathbf{0}$ and then seek the estimator of this form that minimizes the variance $V(\hat{\mathbf{x}})$ or the MSE. The resulting estimator is the (weighted) least-squares estimator [4], which is given by

$$\hat{\mathbf{x}} = \left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}. \tag{5}$$

If $\mathbf{w}$ is a zero-mean Gaussian random vector, then the least-squares estimator is also the minimum variance unbiased estimator, i.e., it minimizes the variance from all linear and nonlinear *unbiased* estimators. However, an unbiased estimator does not necessarily lead to a small MSE. Various alternative linear estimators for the model (1) have been proposed, for example, Tikhonov regularization [5], [6], the shrunken estimator [7], and the covariance shaping least-squares estimator [8]. These estimators attempt to reduce the MSE of the least-squares estimator by allowing for a bias. However, each of the estimators above is designed to optimize an objective that depends on the data error and not directly on the MSE.

In this paper, we consider an alternative method for developing optimal linear estimators. Specifically, we develop estimators that minimize the worst-case regret, i.e., the difference between the MSE of an estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ and the best possible MSE attainable using any estimator of the form $\hat{\mathbf{x}} = \mathbf{G}(\mathbf{x})\mathbf{y}$, where $\mathbf{x}$ is assumed to be known, so that $\mathbf{G}$ can depend explicitly on $\mathbf{x}$. As we show in Section III, since we are restricting ourselves to linear estimators of the form $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$, even in the case in which the parameters $\mathbf{x}$ are known, we cannot achieve zero MSE. The best possible MSE is illustrated schematically in Fig. 1. Instead of seeking an estimator to minimize the worst-case MSE, we therefore propose seeking

[1]If the eigenvalues of $\mathbf{T}$ and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have geometric multiplicity 1, then $\mathbf{T}$ and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have the same eigenvector matrix if and only if they commute [27].
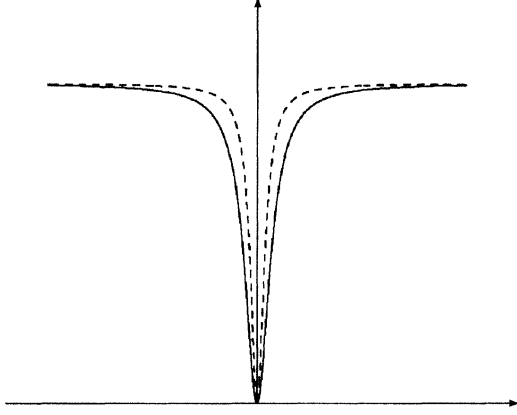
Fig. 1   Line represents the best attainable MSE as a function of $\mathbf{x}$ when $\mathbf{x}$ is known, and the dashed line represents a desirable graph of MSE with small regret as a function of $\mathbf{x}$ using some linear estimator that does not depend on $\mathbf{x}$.

an estimator to minimize the worst-case difference between its MSE and the best possible MSE, as illustrated in Fig. 1.

In Section III, we develop the form of the estimator minimizing the worst-case regret for the case in which $\mathbf{T}$ and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have the same eigenvector matrix (this assumption is made for analytical tractability). We show that the minimax regret estimator can be described by m parameters, which are the solution to a convex optimization problem (Theorem 1). In Section IV, we consider the case in which $\mathbf{T} = \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$. This case may be of interest, for example, when an unknown signal $\mathbf{x}$ is sent through a known channel, and the output signal-to-noise ratio is bounded. As we show, when L is large enough with respect to m, the optimal minimax regret estimator is a shrunken estimator with a specific choice of shrinkage factor. For small values of L, the optimal estimator is given in terms of a single parameter, which is the solution to a nonlinear equation (Theorem 2). In Section V, we consider the case in which $\mathbf{T} = \mathbf{I}$ and show that the minimax regret estimator can be determined by solving m convex optimization problems, each in three unknowns (Theorem 3). The case $\mathbf{T} = \mathbf{I}$ may be of interest when an unknown signal $\mathbf{x}$ is sent through a known channel, and the power of $\mathbf{x}$ is bounded.

In Section VI, we demonstrate by examples that the minimax regret estimator can significantly improve the performance over the traditional least-squares estimator. Furthermore, its performance is often better than that of the minimax estimator that minimizes the worst-case MSE [22] and the Wiener estimator, which results from assuming that $\mathbf{x}$ is a random vector with covariance $L^2\mathbf{I}$.

III. Minimax Regret Estimator

The minimax regret estimator $\hat{\mathbf{x}}$ is designed to minimize the worst-case regret $\mathcal{R}(\mathbf{x}, \mathbf{G})$, which is defined as the difference 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 $\hat{\mathbf{x}} = \mathbf{G}(\mathbf{x})\mathbf{y}$ when the parameters $\mathbf{x}$ are known, which we denote by MSE^o.

To develop an explicit expression for MSE^o, we first determine the estimator $\hat{\mathbf{x}} = \mathbf{G}(\mathbf{x})\mathbf{y}$ that minimizes the MSE when

$\mathbf{x}$ is known. To this end, we differentiate[2] the MSE of (4) with respect to $\mathbf{G}$ and equate to 0, which results in

$$\mathbf{G}(\mathbf{x})\mathbf{C}_w + (\mathbf{G}(\mathbf{x})\mathbf{H} - \mathbf{I})\,\mathbf{x}\mathbf{x}^*\mathbf{H}^* = 0 \qquad (6)$$

so that

$$\mathbf{G}(\mathbf{x}) = \mathbf{x}\mathbf{x}^*\mathbf{H}^*(\mathbf{C}_w + \mathbf{H}\mathbf{x}\mathbf{x}^*\mathbf{H}^*)^{-1}. \qquad (7)$$

Using the Matrix Inversion Lemma [27], we can express $\mathbf{G}$ as

$$\mathbf{G}(\mathbf{x}) = \frac{1}{1 + \mathbf{x}^*\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{x}}\mathbf{x}\mathbf{x}^*\mathbf{H}^*\mathbf{C}_w^{-1}. \qquad (8)$$

Substituting $\mathbf{G}(\mathbf{x})$ back into (4), MSE^0 is given by

$$\mathrm{MSE}^o = \frac{\mathbf{x}^*\mathbf{x}}{1 + \mathbf{x}^*\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{x}}. \qquad (9)$$

Since $\mathbf{x}$ is unknown, we cannot implement the optimal estimator (8). Instead, we seek the estimator $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ that minimizes the worst-case regret $\mathcal{R}(\mathbf{x}, \mathbf{G})$, where

$$\begin{aligned}
\mathcal{R}(\mathbf{x}, \mathbf{G}) &= E\left(\|\mathbf{G}\mathbf{y} - \mathbf{x}\|^2\right) - \mathrm{MSE}^o \\
&= \mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \mathbf{x}^*(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{x} \\
&\quad - \frac{\mathbf{x}^*\mathbf{x}}{1 + \mathbf{x}^*\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{x}}
\end{aligned} \qquad (10)$$

subject to the constraint $\|\mathbf{x}\|_{\mathbf{T}} \leq L$. Thus, we seek the matrix $\mathbf{G}$ that is the solution to the problem

$$\min_{\mathbf{G}} \max_{\mathbf{x}^*\mathbf{T}\mathbf{x} \leq L^2} \mathcal{R}(\mathbf{x}, \mathbf{G}). \qquad (11)$$

For analytical tractability, we restrict our attention to weighting matrices $\mathbf{T}$ such that $\mathbf{T}$ and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have the same eigenvector matrix. Thus, if $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ has an eigendecomposition $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\mathbf{\Sigma}\mathbf{V}^*$, where $\mathbf{V}$ is a unitary matrix and $\mathbf{\Sigma}$ is a diagonal matrix, then $\mathbf{T} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^*$ for some diagonal matrix $\mathbf{\Lambda}$.

Theorem 1 below establishes the general form of the solution to (11) for any $\mathbf{T}$ such that $\mathbf{T}$ and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have the same eigenvector matrix. In Sections IV and V, we use Theorem 1 to develop the solution for the case in which $\mathbf{T} = \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ and $\mathbf{T} = \mathbf{I}$, respectively.

Theorem 1: Let $\mathbf{x}$ denote the unknown deterministic parameter vector in the model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where $\mathbf{H}$ is a known $n \times m$ matrix with rank m, and $\mathbf{w}$ is a zero-mean random vector with covariance $\mathbf{C}_w$. Let $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\mathbf{\Sigma}\mathbf{V}^*$, where $\mathbf{V}$ is a unitary matrix, and $\mathbf{\Sigma}$ is an $m \times m$ diagonal matrix with diagonal elements $\sigma_i > 0$, and let $\mathbf{T} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^*$, where $\mathbf{\Lambda}$ is an $m \times m$

---

[2]We use the following derivatives: For any Hermitian $\mathbf{A}$

$$\frac{\partial \mathrm{Tr}(\mathbf{B}\mathbf{A}\mathbf{B}^*)}{\partial \mathbf{B}} = 2\mathbf{B}\mathbf{A}$$

and

$$\frac{\partial \mathbf{x}^*\mathbf{B}^*\mathbf{B}\mathbf{x}}{\partial \mathbf{B}} = 2\mathbf{B}\mathbf{x}\mathbf{x}^*.$$

diagonal matrix with diagonal elements $\lambda_i > 0$. Then, the solution to the problem

$$\min_{\hat{\mathbf{x}}=\mathbf{Gy}} \max_{\|\mathbf{x}\|_\mathbf{T} \leq L} \left\{ E\left(\|\hat{\mathbf{x}}-\mathbf{x}\|^2\right) - \min_{\hat{\mathbf{x}}=\mathbf{G(x)y}} E\left(\|\hat{\mathbf{x}}-\mathbf{x}\|^2\right) \right\}$$

has the form

$$\hat{\mathbf{x}} = \mathbf{VDV}^* \left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1} \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}$$

where $\mathbf{D}$ is an $m \times m$ diagonal matrix with diagonal elements d_i, which are the solution to the convex optimization problem in (12), shown at the bottom of the page.

Proof: The proof of Theorem 1 is comprised of three parts. We first show that the optimal $\mathbf{G}$ minimizing the worst-case regret has the form

$$\mathbf{G} = \mathbf{VDV}^* \left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1} \mathbf{H}^*\mathbf{C}_w^{-1} \tag{13}$$

for some $m \times m$ matrix $\mathbf{D}$. We then show that $\mathbf{D}$ can be chosen as a diagonal matrix. Finally, we show that the diagonal elements of $\mathbf{D}$, which are denoted d_i, are the solution to (12).

We begin by showing that the optimal $\mathbf{G}$ has the form given by (13). To this end, note that the regret $\mathcal{R}(\mathbf{x},\mathbf{G})$ of (10) depends on $\mathbf{G}$ only through $\mathbf{GH}$ and $\text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*)$. Now, for any choice of $\mathbf{G}$

$$\text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) = \text{Tr}\left(\mathbf{G}\mathbf{C}_w^{\frac{1}{2}}\mathbf{P}\mathbf{C}_w^{\frac{1}{2}}\mathbf{G}^*\right)$$
$$+ \text{Tr}\left(\mathbf{G}\mathbf{C}_w^{\frac{1}{2}}(\mathbf{I}-\mathbf{P})\mathbf{C}_w^{\frac{1}{2}}\mathbf{G}^*\right) \geq \text{Tr}\left(\mathbf{G}\mathbf{C}_w^{\frac{1}{2}}\mathbf{P}\mathbf{C}_w^{\frac{1}{2}}\mathbf{G}^*\right) \tag{14}$$

where

$$\mathbf{P} = \mathbf{C}_w^{-\frac{1}{2}}\mathbf{H}\left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^*\mathbf{C}_w^{-\frac{1}{2}} \tag{15}$$

is the orthogonal projection onto the range space of $\mathbf{C}_w^{-1/2}\mathbf{H}$. Therefore, if $\mathbf{G}$ is chosen such that

$$\mathbf{G}\mathbf{C}_w^{\frac{1}{2}} = \mathbf{G}\mathbf{C}_w^{\frac{1}{2}}\mathbf{P} \tag{16}$$

then $\text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) = \text{Tr}(\mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{1/2}\mathbf{G}^*)$. In addition, $\mathbf{GH} = \mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{-1/2}\mathbf{H}$ since $\mathbf{P}\mathbf{C}_w^{-1/2}\mathbf{H} = \mathbf{C}_w^{-1/2}\mathbf{H}$. Thus, to minimize the worst-case regret, it is sufficient to consider matrices $\mathbf{G}$ that satisfy (16), or equivalently, that satisfy

$$\mathbf{G} = \mathbf{G}\mathbf{C}_w^{\frac{1}{2}}\mathbf{P}\mathbf{C}_w^{-\frac{1}{2}}. \tag{17}$$

Substituting (15) into (17), we have

$$\begin{aligned}\mathbf{G} &= \mathbf{G}\mathbf{C}_w^{\frac{1}{2}}\mathbf{P}\mathbf{C}_w^{-\frac{1}{2}} \\ &= \mathbf{GH}\left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^*\mathbf{C}_w^{-1} \\ &= \mathbf{B}\left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\end{aligned} \tag{18}$$

for some $m \times m$ matrix $\mathbf{B}$. Denoting $\mathbf{B} = \mathbf{VDV}^*$, (18) reduces to (13).

We now show that $\mathbf{D}$ can always be chosen as a diagonal matrix. Since $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\Sigma\mathbf{V}^*$, we can express $\mathcal{R}(\mathbf{x},\mathbf{G})$ as

$$\begin{aligned}\mathcal{R}(\mathbf{x},\mathbf{G}) &= \text{Tr}(\mathbf{D}^*\mathbf{D}\Sigma^{-1}) + \mathbf{x}^*\mathbf{V}(\mathbf{I}-\mathbf{D})^*(\mathbf{I}-\mathbf{D})\mathbf{V}^*\mathbf{x} \\ &\quad - \frac{\mathbf{x}^*\mathbf{x}}{1+\mathbf{x}^*\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{Hx}} \\ &= \text{Tr}(\mathbf{D}^*\mathbf{D}\Sigma^{-1})+\mathbf{z}^*(\mathbf{I}-\mathbf{D})^*(\mathbf{I}-\mathbf{D})\mathbf{z} - \frac{\mathbf{z}^*\mathbf{z}}{1+\mathbf{z}^*\Sigma\mathbf{z}}\end{aligned} \tag{19}$$

where $\mathbf{z} = \mathbf{V}^*\mathbf{x}$. Combining (19) with

$$\mathbf{x}^*\mathbf{Tx} = \mathbf{x}^*\mathbf{V}\Lambda\mathbf{V}^*\mathbf{x} = \mathbf{z}^*\Lambda\mathbf{z} \tag{20}$$

we conclude that (11) reduces to finding $\mathbf{D}$ that minimizes

$$\mathcal{G}(\mathbf{D}) = \max_{\mathbf{z}^*\Lambda\mathbf{z}\leq L^2} \left\{ \text{Tr}(\mathbf{D}^*\mathbf{D}\Sigma^{-1})+\mathbf{z}^*(\mathbf{I}-\mathbf{D})^*(\mathbf{I}-\mathbf{D})\mathbf{z} - \frac{\mathbf{z}^*\mathbf{z}}{1+\mathbf{z}^*\Sigma\mathbf{z}} \right\}. \tag{21}$$

Let $\mathbf{J}$ be a diagonal matrix with diagonal elements equal to ±1. Then

$$\begin{aligned}\mathcal{G}(\mathbf{JDJ}) &= \max_{\mathbf{z}^*\Lambda\mathbf{z}\leq L^2} \left\{ \text{Tr}(\mathbf{D}^*\mathbf{D}\Sigma^{-1})+ \right. \\ &\quad \left. \mathbf{z}^*\mathbf{J}(\mathbf{I}-\mathbf{D})^*\ (\mathbf{I}-\mathbf{D})\mathbf{Jz} - \frac{\mathbf{z}^*\mathbf{z}}{1+\mathbf{z}^*\Sigma\mathbf{z}} \right\} \\ &= \max_{\mathbf{z}'^*\Lambda\mathbf{z}'\leq L^2} \left\{ \text{Tr}(\mathbf{D}^*\mathbf{D}\Sigma^{-1})+ \right. \\ &\quad \left. \mathbf{z}'^*(\mathbf{I}-\mathbf{D})^*\ (\mathbf{I}-\mathbf{D})\mathbf{z}' - \frac{\mathbf{z}'^*\mathbf{z}'}{1+\mathbf{z}'^*\Sigma\mathbf{z}'} \right\} \\ &= \mathcal{G}(\mathbf{D})\end{aligned} \tag{22}$$

where $\mathbf{z}' = \mathbf{Jz}$, and we used the fact that $\mathbf{J}^2 = \mathbf{I}$, and for any diagonal matrix $\mathbf{M}$, $\mathbf{JMJ} = \mathbf{M}$. Therefore, if $\mathbf{D}$ minimizes $\mathcal{G}(\mathbf{D})$, then $\mathbf{JDJ}$ also minimizes $\mathcal{G}(\mathbf{D})$. Now, since the problem of minimizing $\mathcal{G}(\mathbf{D})$ is convex, the set of optimal solutions is also convex [28], which implies that if $\mathbf{JDJ}$ is optimal for any diagonal $\mathbf{J}$ with diagonal elements ±1, then so is $\mathbf{D}' = (1/2^m)\sum_\mathbf{J}\mathbf{JDJ}$, where the summation is over all 2^m diagonal matrices $\mathbf{J}$ with diagonal elements ±1. It is easy to see that $\mathbf{D}'$ has diagonal elements. Therefore, we have shown that there exists an optimal diagonal solution $\mathbf{D}$.

Denoting by d_i the diagonal elements of $\mathbf{D}$ and by $s_i = |z_i|^2$ where z_i are the components of $\mathbf{z}$, we can express $\mathcal{G}(\mathbf{D})$ as

$$\begin{aligned}\mathcal{G}(\mathbf{D}) &= \\ &= \max_{s_i\geq0,\sum_{i=1}^m \lambda_i s_i \leq L^2} \left\{ \sum_{i=1}^m (1-d_i)^2 s_i - \frac{\sum_{i=1}^m s_i}{1+\sum_{i=1}^m \sigma_i s_i} \right\} \\ &\quad + \sum_{i=1}^m \frac{d_i^2}{\sigma_i} = \max_{\mathbf{s}\in\mathcal{I}} \Phi(\mathbf{s}) + \sum_{i=1}^m \frac{d_i^2}{\sigma_i}\end{aligned} \tag{23}$$

$$\min_{\tau,d_i} \left\{ \tau : \begin{array}{c} \sum_{i=1}^m \frac{d_i^2}{\sigma_i} \leq \tau \\ \max_{s_i\geq0,\sum_i \lambda_i s_i = L^2} \left\{ \sum_{i=1}^m (1-d_i)^2 s_i - \frac{\sum_{i=1}^m s_i}{1+\sum_{i=1}^m \sigma_i s_i} \right\} + \sum_{i=1}^m \frac{d_i^2}{\sigma_i} \leq \tau \end{array} \right\}. \tag{12}$$

where

$$\Phi(\mathbf{s}) = \sum_{i=1}^{m}(1-d_i)^2 s_i - \frac{\sum_{i=1}^{m} s_i}{1 + \sum_{i=1}^{m} \sigma_i s_i} \tag{24}$$

and $\mathcal{I}$ is the set of vectors $\mathbf{s} \in \mathcal{R}^m$ with components s_i such that $s_i \geq 0$, and $\sum_{i=1}^{m} \lambda_i s_i \leq L^2$, i.e.,

$$\mathcal{I} \triangleq \left\{ \mathbf{s} \in \mathcal{R}^m \Big| s_i \geq 0, \sum_{i=1}^{m} \lambda_i s_i \leq L^2 \right\}. \tag{25}$$

To complete the proof of Theorem 1, we rely on the following lemma.

Lemma 1: Let

$$\Phi(\mathbf{s}) = \sum_{i=1}^{m}(1-d_i)^2 s_i - \frac{\sum_{i=1}^{m} s_i}{1 + \sum_{i=1}^{m} \sigma_i s_i}$$

for some given σ_i, $1 \leq i \leq m$, and d_i, $1 \leq i \leq m$. If

$$\hat{\mathbf{s}} = \arg\max_{\mathbf{s} \in \mathcal{I}} \Phi(\mathbf{s})$$

where $\mathcal{I}$ is defined by (25), then $\hat{\mathbf{s}} = \mathbf{0}$ or $\sum_{i=1}^{m} \lambda_i \hat{s}_i = L^2$.

Proof: Let $\mathcal{S}$ be the set of vectors $\mathbf{s}$ such that $\mathbf{s} = \mathbf{0}$ or $\sum_{i=1}^{m} \lambda_i s_i = L^2$. To establish the lemma, we need to show that for any $\mathbf{s} \in \mathcal{I}$, $\Phi(\mathbf{s}) \leq \Phi(\mathbf{s}')$ for some $\mathbf{s}' \in \mathcal{S}$.

Fix $\mathbf{s} \in \mathcal{I}$ such that $\mathbf{s} \neq \mathbf{0}$, and let $h(r) = \Phi(r\mathbf{s})$ be defined on the segment $[0, r_*]$, where r_* is the largest value of r for which $r\mathbf{s} \in \mathcal{I}$. Clearly, $r_*\mathbf{s} \in \mathcal{S}$ and $r_* \geq 1$. Since $h(1) = \Phi(\mathbf{s})$, $h(0) = \Phi(\mathbf{0})$, and $h(r_*) = \Phi(\mathbf{s}')$, where $\mathbf{s}' = r_*\mathbf{s} \in \mathcal{S}$, to prove that $\Phi(\mathbf{s}) \leq \Phi(\mathbf{s}')$ for some $\mathbf{s}' \in \mathcal{S}$, it suffices to show that

$$h(1) \leq \max\left(h(0), h(r_*)\right). \tag{26}$$

We now establish (26) by first showing that $h(r)$ is convex. It then follows that $h(r)$ obtains its maximum at one of its end points. Since $h(r)$ is defined on $[0, r_*]$, this implies that $h(r) \leq \max(h(0), h(r_*))$ for any $r \in [0, r_*]$ and, in particular, for $r = 1$, which establishes (26). It remains to show that $h(r)$ is convex. Writing

$$h(r) = \alpha r - \frac{\beta r}{1 + \gamma r} \tag{27}$$

where $\alpha = \sum_{i=1}^{m}(1-d_i)^2 s_i \geq 0$, $\beta = \sum_{i=1}^{m} s_i > 0$ and $\gamma = \sum_{i=1}^{m} \sigma_i s_i > 0$, we can express $h(r)$ as

$$h(r) = \alpha r - \frac{\beta}{\gamma}\frac{1 + \gamma r - 1}{1 + \gamma r} = \alpha r + \frac{\beta}{\gamma}\frac{1}{1 + \gamma r} - \frac{\beta}{\gamma}. \tag{28}$$

Since $1/r$ is convex in r, $h(r)$ is convex. $\square$

From (23) and Lemma 1, it follows that finding $\mathbf{D}$ to minimize $\mathcal{G}(\mathbf{D})$ is equivalent to finding d_i to minimize

$$\max\left(\Phi(\mathbf{0}), \max_{s_i \geq 0, \sum_{i=1}^{m} \lambda_i s_i = L^2} \Phi(\mathbf{s})\right) + \sum_{i=1}^{m}\frac{d_i^2}{\sigma_i}. \tag{29}$$

Since $\Phi(\mathbf{0}) = 0$, this problem can be written as (12), completing the proof of Theorem 1. $\square$

Theorem 1 reduces the problem of minimizing the regret to the simpler optimization problem (12). As we show in Sections IV and V, for certain choices of $\mathbf{T}$, the problem can be further simplified, and in some cases, a closed-form solution for the minimax regret estimator exists. In Section IV, we consider the case in which the weighting $\mathbf{T} = \mathbf{H}^* \mathbf{C}_w \mathbf{H}$, and in Section V, we consider the case in which $\mathbf{T} = \mathbf{I}$. As we show, when L is large enough with respect to m, and $\mathbf{T} = \mathbf{H}^* \mathbf{C}_w \mathbf{H}$, the minimax regret estimator of Theorem 1 reduces to a shrunken estimator with a shrinkage factor that depends only on the bound L. For small values of L, the minimax regret estimator is a function of a single parameter, that is the solution to a nonlinear equation. In the case in which $\mathbf{T} = \mathbf{I}$, the minimax regret estimator depends on three parameters, which can be found by solving m convex optimization problems in three unknowns.

IV. MINIMAX REGRET ESTIMATOR WITH $\mathbf{T} = \mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H}$

We now consider the case in which the weighting $\mathbf{T} = \mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H}$ so that the eigenvalues λ_i of $\mathbf{T}$ are equal to σ_i.

From Theorem 1, the optimal $\mathbf{G}$ that minimizes the regret in this case is given by (13), where the diagonal elements d_i of $\mathbf{D}$ are the solution to the problem (Δ), which is defined in (30), shown at the bottom of the page.

To develop a solution to (Δ), define the set $\mathcal{P}$ as

$$\mathcal{P} \triangleq \left\{ \mathbf{s} \in \mathcal{R}^m \Big| s_i \geq 0, \sum_{i=1}^{m} \sigma_i s_i = L^2 \right\}. \tag{31}$$

Then

$$\max_{\mathbf{s} \in \mathcal{P}} \left\{ \sum_{i=1}^{m}(1-d_i)^2 s_i - \frac{\sum_{i=1}^{m} s_i}{1 + L^2} \right\}$$

is a linear program [29]. From linear programming duality theory, it follows that

$$\max_{\mathbf{s} \in \mathcal{P}} \left\{ \sum_{i=1}^{m}\left((1-d_i)^2 - \frac{1}{1+L^2}\right) s_i \right\} = \min_{y \in \mathcal{D}} L^2 y \tag{32}$$

where $\mathcal{D}$ is the set of scalars y for which

$$y \geq \frac{1}{\sigma_i}\left((1-d_i)^2 - \frac{1}{1+L^2}\right), \quad 1 \leq i \leq m. \tag{33}$$

Thus, the problem (Δ) can be written as

$$(\Gamma): \min_{\tau, d_i, y} \left\{ \tau : \begin{array}{l} \sum_{i=1}^{m}\frac{d_i^2}{\sigma_i} \leq \tau \\ L^2 y + \sum_{i=1}^{m}\frac{d_i^2}{\sigma_i} \leq \tau \\ \frac{1}{\sigma_i}\left((1-d_i)^2 - \frac{1}{1+L^2}\right) \leq y, \quad 1 \leq i \leq m \end{array} \right\}. \tag{34}$$

$$(\Delta): \min_{\tau, d_i} \left\{ \tau : \begin{array}{c} \sum_{i=1}^{m}\frac{d_i^2}{\sigma_i} \leq \tau \\ \max_{s_i \geq 0, \sum_{i=1}^{m} \sigma_i s_i = L^2}\left\{\sum_{i=1}^{m}(1-d_i)^2 s_i - \frac{\sum_{i=1}^{m} s_i}{1+L^2}\right\} + \sum_{i=1}^{m}\frac{d_i^2}{\sigma_i} \leq \tau \end{array} \right\}. \tag{30}$$

Since (Γ) is a convex optimization problem, from Lagrange duality theory [30], it follows that $A = \min \tau$ in the problem (Γ) is equal to the optimal value of the dual problem, namely

$$A = \max_{\alpha, \beta, \gamma_i \geq 0} \min_{\tau, d_i, y} \mathcal{L}(\tau, d_i, y) \qquad (35)$$

where the Lagrangian $\mathcal{L}$ is given by

$$\mathcal{L}(\tau, d_i, y) = \tau + \alpha \left(\sum_{i=1}^{m} \frac{d_i^2}{\sigma_i} - \tau \right) + \beta \left(L^2 y + \sum_{i=1}^{m} \frac{d_i^2}{\sigma_i} - \tau \right)$$
$$+ \sum_{i=1}^{m} \gamma_i \left((1 - d_i)^2 - \frac{1}{1 + L^2} - \sigma_i y \right). \qquad (36)$$

Substituting (36) into (35), we have

$$A = \max_{\alpha, \beta, \gamma_i \geq 0} \left\{ -\frac{1}{1 + L^2} \sum_{i=1}^{m} \gamma_i + \min_{\tau} \{ (1 - \alpha - \beta)\tau \} \right.$$
$$+ \min_{y} \left\{ \left(\beta L^2 - \sum_{i=1}^{m} \sigma_i \gamma_i \right) y \right\}$$
$$\left. + \sum_{i=1}^{m} \min_{d_i} \left\{ \frac{\alpha + \beta}{\sigma_i} d_i^2 + \gamma_i (1 - d_i)^2 \right\} \right\}$$
$$= \max_{\alpha, \beta, \gamma_i \geq 0} \left\{ -\frac{1}{1 + L^2} \sum_{i=1}^{m} \gamma_i + \sum_{i=1}^{m} \frac{\gamma_i}{1 + \sigma_i \gamma_i} : \right.$$
$$\left. \alpha + \beta = 1, \beta L^2 = \sum_{i=1}^{m} \sigma_i \gamma_i \right\}$$
$$= \max_{\gamma_i \geq 0} \left\{ -\frac{1}{1 + L^2} \sum_{i=1}^{m} \gamma_i + \sum_{i=1}^{m} \frac{\gamma_i}{1 + \sigma_i \gamma_i} : \right.$$
$$\left. \sum_{i=1}^{m} \sigma_i \gamma_i \leq L^2 \right\} \qquad (37)$$

where we used the fact that the optimal d_i are given by

$$d_i = \frac{\sigma_i \gamma_i}{1 + \sigma_i \gamma_i}, \quad 1 \leq i \leq m. \qquad (38)$$

The dual problem of (Γ) is therefore the problem

$$\max_{\gamma_i} \left\{ -\frac{1}{1 + L^2} \sum_{i=1}^{m} \gamma_i + \sum_{i=1}^{m} \frac{\gamma_i}{1 + \sigma_i \gamma_i} \right\} \qquad (39)$$

subject to

$$\sum_{i=1}^{m} \sigma_i \gamma_i \leq L^2$$
$$\gamma_i \geq 0, \quad 1 \leq i \leq m. \qquad (40)$$

Once we find the dual optimal values γ_i, the optimal values d_i can be calculated using (38).

Since the problem of (39) subject to (40) is a convex optimization problem, we can find an optimal solution by forming the Lagrangian

$$\mathcal{L} = \frac{1}{1 + L^2} \sum_{i=1}^{m} \gamma_i - \sum_{i=1}^{m} \frac{\gamma_i}{1 + \sigma_i \gamma_i} + \rho \sum_{i=1}^{m} \sigma_i \gamma_i - \sum_{i=1}^{m} \zeta_i \gamma_i \qquad (41)$$

where from the Karush–Kuhn–Tucker conditions [29], we must have that $\rho, \zeta_i \geq 0$. The values γ_i are optimal if and only if they satisfy (41) and there exist $\rho, \zeta_i \geq 0$ such that

$$\frac{\partial \mathcal{L}}{\partial \gamma_i} = \frac{1}{1 + L^2} - \frac{1}{(\gamma_i \sigma_i + 1)^2} + \rho \sigma_i - \zeta_i = 0 \qquad (42)$$

and the complementary slackness conditions are satisfied, namely

$$\rho \left(\sum_{i=1}^{m} \sigma_i \gamma_i - L^2 \right) = 0$$
$$\zeta_i \gamma_i = 0. \qquad (43)$$

Suppose first that $\rho = 0$. If $\gamma_i = 0$, then from (42), $\zeta_i = -L^2/(1 + L^2) < 0$, which contradicts the condition $\zeta_i \geq 0$. Therefore, we must have that $\gamma_i > 0$, which implies from (43) that $\zeta_i = 0$. Substituting $\rho = \zeta_i = 0$ into (42)

$$\frac{1}{1 + L^2} = \frac{1}{(\gamma_i \sigma_i + 1)^2} \qquad (44)$$

or

$$\gamma_i = \frac{1}{\sigma_i}(\sqrt{1 + L^2} - 1), \quad 1 \leq i \leq m. \qquad (45)$$

From (38), we then have

$$d_i = 1 - \frac{1}{\sqrt{1 + L^2}}, \quad 1 \leq i \leq m \qquad (46)$$

so that from Theorem 1, the optimal estimator in this case is

$$\hat{\mathbf{x}} = \left(1 - \frac{1}{\sqrt{1 + L^2}} \right) \left(\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H} \right)^{-1} \mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{y}. \qquad (47)$$

The outlined candidate solution is valid if and only if it satisfies (40), which happens if and only if

$$m(\sqrt{1 + L^2} - 1) \leq L^2 \qquad (48)$$

or, equivalently

$$L^2 \geq (m - 1)^2 - 1. \qquad (49)$$

Now, assume that the inequality opposite to (48) holds true, i.e.,

$$m(\sqrt{1 + L^2} - 1) > L^2. \qquad (50)$$

From the above analysis, in this case, $\rho > 0$, and (43), (40), and (42) become

$$\sum_{i=1}^{m} \sigma_i \gamma_i = L^2$$
$$\zeta_i \gamma_i = 0, \quad 1 \leq i \leq m; \quad \gamma_i \geq 0$$
$$\frac{1}{1 + L^2} - \frac{1}{(1 + \sigma_i \gamma_i)^2} + \sigma_i \rho - \zeta_i = 0. \qquad (51)$$

Let $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_m > 0$. If $\gamma_j = 0$ for some $1 \leq j \leq m$, then $\gamma_i = 0$, $i \leq j$. For suppose that $\gamma_i > 0$ for some $i \leq j$. Then, from (51), $\zeta_i = 0$, and

$$\sigma_i \rho = \frac{1}{(1 + \sigma_i \gamma_i)^2} - \frac{1}{1 + L^2} < 1 - \frac{1}{1 + L^2}. \qquad (52)$$

On the other hand, since $\gamma_j = 0$, $\zeta_j \geq 0$, and from (51)

$$\sigma_j \rho \geq 1 - \frac{1}{1 + L^2} \qquad (53)$$

which contradicts (52) because $\sigma_j \leq \sigma_i$. Thus, we conclude that there exists a $0 \leq k \leq m - 1$ such that $\gamma_i = 0$ for $i \leq k$, and $\gamma_i > 0$ for $i > k$. If $k = 0$, then $\gamma_i > 0$, $1 \leq i \leq m$.

Since $\gamma_i > 0$ for $i > k$, from (51), $\zeta_i = 0$ for $i > k$, and

$$\gamma_i = \frac{1}{\sigma_i} \left(\frac{1}{\sqrt{\rho \sigma_i + \frac{1}{(L^2 + 1)}}} - 1 \right), \quad i > k \qquad (54)$$

where $\rho > 0$, and k should be such that $\gamma_i > 0$ for $i > k$. On the other hand, from (51) combined with the fact that $\gamma_i = 0$ and $\zeta_i \geq 0$ for $i \leq k$, it follows that

$$\frac{1}{\sigma_i} \left(\frac{1}{\sqrt{\rho \sigma_i + \frac{1}{(L^2 + 1)}}} - 1 \right) \leq 0$$

for $i \leq k$. Therefore, for $1 \leq i \leq m$

$$\gamma_i = \gamma_i(\rho) \triangleq \max \left[\frac{1}{\sigma_i} \left(\frac{1}{\sqrt{\rho \sigma_i + \frac{1}{(L^2 + 1)}}} - 1 \right), 0 \right] \qquad (55)$$

where $\rho > 0$ is chosen such that

$$\sum_{i=1}^{m} \sigma_i \gamma_i(\rho) = L^2. \qquad (56)$$

Equivalently, ρ is chosen as a positive root of $\phi(\rho)$, where

$$\phi(\rho) = \sum_{i=1}^{m} \sigma_i \gamma_i(\rho) - L^2. \qquad (57)$$

We can immediately verify that $\phi(\rho)$ is continuous in $\rho \geq 0$. Furthermore, for each i, $\gamma_i(\rho)$ is nonincreasing in ρ, and $\gamma_i(\rho) \to 0$ as $\rho \to \infty$. Therefore, $\phi(\rho)$ has a positive root if and only if $\phi(0) > 0$, which is exactly condition (50). Note also that the function $\phi(\rho)$ is strictly decreasing in $\rho \geq 0$ on the entire segment where $\phi(\rho) > -L^2$. Thus, in the case of (50), $\phi(\rho)$ has exactly one positive root ρ_* (which can be easily found by bisection).

We summarize our results in the following theorem.

Theorem 2: 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 covariance $\mathbf{C}_w$. Let $\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H} = \mathbf{V} \Sigma \mathbf{V}^*$, where $\mathbf{V}$ is a unitary matrix, and Σ is an $m \times m$ diagonal matrix with diagonal elements $\sigma_1 \geq \ldots \geq \sigma_m > 0$. Then, the solution to the problem

$$\min_{\hat{\mathbf{x}} = \mathbf{Gy}} \max_{\|\mathbf{x}\|_{\mathbf{T}} \leq L} \left\{ E \left(\|\hat{\mathbf{x}} - \mathbf{x}\|^2 \right) - \min_{\hat{\mathbf{x}} = \mathbf{G}(\mathbf{x})\mathbf{y}} E \left(\|\hat{\mathbf{x}} - \mathbf{x}\|^2 \right) \right\}$$

with $\mathbf{T} = \mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H}$ is given by

$$\hat{\mathbf{x}} = \begin{cases} \left(1 - \sqrt{\frac{1}{1+L^2}} \right) \left(\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H} \right)^{-1} \mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{y}, & L^2 \geq a \\ \mathbf{V} \mathbf{D} \mathbf{V}^* \left(\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H} \right)^{-1} \mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{y}, & L^2 < a \end{cases}$$

where $a = (m - 1)^2 - 1$, and $\mathbf{D}$ is an $m \times m$ diagonal matrix with diagonal elements d_i that are given by

$$d_i = \begin{cases} 0, & i \leq k(\rho_*) \\ 1 - \sqrt{\rho_* \sigma_i + \frac{1}{(L^2 + 1)}}, & i \geq k(\rho_*) + 1. \end{cases}$$

Here, ρ_* is the unique positive root of $\phi(\rho)$ defined by (57), and

$$k(\rho_*) = \begin{cases} 0, & \gamma_1(\rho_*) > 0 \\ \max \{k : \gamma_k(\rho_*) = 0\}, & \text{otherwise} \end{cases}$$

where $\gamma_i(\rho)$ is defined by (55).

The minimax regret estimator of Theorem 2 for the case in which $L^2 \geq (m - 1)^2 - 1$ is a shrunken estimator proposed by Mayer and Willke [7], which is simply a scaled version of the least-squares estimator, with an optimal choice of shrinkage factor. We therefore conclude that this particular shrunken estimator has a strong optimality property: Among all linear estimators of $\mathbf{x}$ such that $\mathbf{x}^* \mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H} \mathbf{x} \leq L$, it minimizes the worst-case regret.

As we expect intuitively, when $L \to \infty$, the minimax regret estimator $\hat{\mathbf{x}}$ of Theorem 2 reduces to the least-squares estimator (5). Indeed, when the weighted norm of $\mathbf{x}$ can be made arbitrarily large, the MSE, and therefore the regret, will also be arbitrarily large, unless the bias is equal to zero. Therefore, in this limit, the worst-case regret is minimized by choosing an estimator with zero bias that minimizes the variance, which leads to the least-squares estimator.

V. MINIMAX REGRET ESTIMATOR WITH $\mathbf{T} = \mathbf{I}$

We now consider the case in which $\mathbf{T} = \mathbf{I}$. In this case, it follows from Theorem 1 that the optimal $\mathbf{G}$ that minimizes the worst-case regret has the form given by (13), where $\mathbf{D}$ is a diagonal matrix with diagonal elements d_i which are solution to the problem (Δ) and are defined as (58), shown at the bottom of the page. Here, $\sigma_i > 0$ are the eigenvalues of $\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H}$.

To develop a solution to (Δ), we note that

$$A = \max_{s_i \geq 0, \sum_{i=1}^{m} s_i = L^2} \left\{ \sum_{i=1}^{m} (1 - d_i)^2 s_i - \frac{L^2}{1 + \sum_{i=1}^{m} \sigma_i s_i} \right\}$$

$$= \max_{\rho, s_i \in \mathcal{P}} \left\{ \sum_{i=1}^{m} (1 - d_i)^2 s_i - \rho \right\} \qquad (59)$$

$$(\Delta) : \min_{\tau, d_i} \left\{ \tau : \begin{array}{c} \sum_{i=1}^{m} \frac{d_i^2}{\sigma_i} \leq \tau \\ \max_{s_i \geq 0, \sum_{i=1}^{m} s_i = L^2} \left\{ \sum_{i=1}^{m} (1 - d_i)^2 s_i - \frac{L^2}{1 + \sum_{i=1}^{m} \sigma_i s_i} \right\} + \sum_{i=1}^{m} \frac{d_i^2}{\sigma_i} \leq \tau \end{array} \right\}. \qquad (58)$$

where $\mathcal{P}$ is the set of vectors $\mathbf{s} \in \mathcal{R}^m$ and scalars $\mathbf{s} \in \mathcal{R}$ defined by

$$\mathcal{P} \triangleq \left\{ \mathbf{s} \geq 0, \sum_{i=1}^{m} s_i = L^2, \rho \geq \frac{L^2}{1 + \sum_{i=1}^{m} \sigma_i s_i} \right\} \quad (60)$$

or, equivalently

$$\mathcal{P} \triangleq \left\{ s_i \geq 0, \sum_{i=1}^{m} s_i = L^2 \rho \geq 0, 1 + \sum_{i=1}^{m} \sigma_i s_i - \frac{L^2}{\rho} \geq 0 \right\}. \quad (61)$$

Since $\mathcal{P}$ is a convex set, and the objective in (59) is linear, (59) is a convex optimization problem. From Lagrange duality theory [30], it then follows that A is equal to the optimal value of the dual problem, namely

$$A = \min_{\mu \geq 0, \lambda} \max_{s_i \geq 0, \rho \geq 0} \mathcal{L}(s_i, \rho, \lambda, \mu) \quad (62)$$

where the Lagrangian $\mathcal{L}$ is given by

$$\mathcal{L}(s_i, \rho, \lambda, \mu) = \sum_{i=1}^{m} (1 - d_i)^2 s_i - \rho + \lambda \left(L^2 - \sum_{i=1}^{m} s_i \right)$$
$$+ \mu \left(1 + \sum_{i=1}^{m} \sigma_i s_i - \frac{L^2}{\rho} \right). \quad (63)$$

Substituting (63) into (62), we have

$$A = \min_{\mu \geq 0, \lambda} \left\{ \lambda L^2 + \mu + \sum_{i=1}^{m} \max_{s_i \geq 0} \left\{ \left[(1 - d_i)^2 - \lambda + \mu \sigma_i \right] s_i \right\} \right.$$
$$\left. + \max_{\rho \geq 0} \left\{ -\rho - \frac{\mu L^2}{\rho} \right\} \right\}$$
$$= \min_{\mu \geq 0, \lambda} \left\{ \lambda L^2 + \mu - 2L\sqrt{\mu} : (1 - d_i)^2 + \mu \sigma_i \leq \lambda, 1 \leq i \leq m \right\} \quad (64)$$

where we used the fact that the optimal ρ is $\rho = L\sqrt{\mu}$. The problem (Δ) of (58) can therefore be expressed as

$$(\Gamma): \min_{\tau, d_i, \mu, \lambda} \left\{ \tau : \begin{array}{c} \sum_{i=1}^{m} \frac{d_i^2}{\sigma_i} \leq \tau \\ -2L\sqrt{\mu} + \mu + L^2\lambda + \sum_{i=1}^{m} \frac{d_i^2}{\sigma_i} \leq \tau \\ (1 - d_i)^2 + \sigma_i \mu \leq \lambda, \quad 1 \leq i \leq m \\ \mu \geq 0 \end{array} \right\}. \quad (65)$$

Since (Γ) is a convex optimization problem, we can find an optimal solution to Γ by forming the Lagrangian

$$\mathcal{L} = \tau + \alpha \left(\sum_{i=1}^{m} \frac{d_i^2}{\sigma_i} - \tau \right) + \beta \left(-2L\sqrt{\mu} + \mu + L^2\lambda + \sum_{i=1}^{m} \frac{d_i^2}{\sigma_i} - \tau \right)$$
$$+ \sum_{i=1}^{m} \gamma_i \left((1 - d_i)^2 + \sigma_i \mu - \lambda \right) \quad (66)$$

where, from the Karush–Kuhn–Tucker conditions [29], we must have that α, β, and $\gamma_i \geq 0$. Differentiating $\mathcal{L}$ with respect to τ and equating to 0

$$\alpha + \beta = 1. \quad (67)$$

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

$$d_i = \frac{\gamma_i}{\left(\frac{1}{\sigma_i} \right) (\alpha + \beta) + \gamma_i} = \frac{\gamma_i}{\frac{1}{\sigma_i} + \gamma_i} \quad (68)$$

from which we conclude that

$$0 \leq d_i \leq 1, \quad 1 \leq i \leq m. \quad (69)$$

From (65), d_i must satisfy

$$(1 - d_i)^2 + \sigma_i \mu \leq \lambda. \quad (70)$$

Suppose that we have equality in (70) for some $1 \leq j \leq m$. Then, to satisfy (69), we must have that

$$d_j = 1 - \sqrt{\lambda - \sigma_j \mu} \quad (71)$$

and

$$\lambda - \sigma_i \mu \leq 1. \quad (72)$$

If for some j we have inequality in (70) so that

$$(1 - d_j)^2 + \sigma_j \mu < \lambda \quad (73)$$

then by complementary slackness, we must have that $\gamma_j = 0$, which from (68) implies that $d_j = 0$.

Let $\sigma_1 \geq \ldots \geq \sigma_m > 0$. Then, by (68), we have that $d_1 \geq \ldots \geq d_m \geq 0$. Therefore, if $d_j = 0$ for some j, then $d_i = 0$ for all $i \geq j$. It follows that at an optimal solution, there exists a $1 \leq k \leq m$ such that

$$\begin{aligned} d_i &= 1 - \sqrt{\lambda - \sigma_i \mu}, \quad i \leq k \\ \lambda &\leq 1 + \sigma_k \mu \\ d_i &= 0, \quad i \geq k + 1 \\ &\text{if } k < m, \text{ then } \lambda \geq 1 + \sigma_{k+1} \mu. \end{aligned} \quad (74)$$

We conclude that (Γ) of (65) can be solved by first solving m problems (Γ_k), for $1 \leq k \leq m$ with three unknowns each, where we have (75), shown at the bottom of the page, and then choosing the value of k and the corresponding optimal values d_i given by (74), which result in the smallest possible value of τ.

Each problem (Γ_k) is a simple convex optimization problem involving three unknowns and can therefore be solved very efficiently, for example, using the Ellipsoidal method (see, e.g., [29, Ch. 5.2]).

$$(\Gamma_k): \min_{\tau, \mu, \lambda} \left\{ \tau : \begin{array}{c} \sum_{i=1}^{k} \frac{(1 - \sqrt{\lambda - \sigma_i \mu})^2}{\sigma_i} \leq \tau \\ -2L\sqrt{\mu} + \mu + L^2\lambda + \sum_{i=1}^{k} \frac{(1 - \sqrt{\lambda - \sigma_i \mu})^2}{\sigma_i} \leq \tau \\ \sigma_1 \mu \leq \lambda \leq 1 + \sigma_k \mu \\ \mu \geq 0 \\ \text{if } k < m, \text{ then } \lambda \geq 1 + \sigma_{k+1} \mu \end{array} \right\} \quad (75)$$

Following similar steps to those taken in (37)–(40), we can derive the dual problem of (Γ) given by (65), which results in

$$\max_{\gamma_i,\beta}\left\{\sum_{i=1}^{m}\frac{\gamma_i}{1+\sigma_i\gamma_i}-\frac{L^2\beta^2}{\beta+\sum_{i=1}^{m}\sigma_i\gamma_i}\right\} \quad (76)$$

subject to

$$\sum_{i=1}^{m}\gamma_i=\beta L^2$$
$$0\le\beta\le 1$$
$$\gamma_i\ge 0,\quad 1\le i\le m. \quad (77)$$

Suppose now that $\sigma_i=1$, $1\le i\le m$ so that $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}=\mathbf{I}$. In this case, the objective (76) can be written as

$$\max_{\gamma_i}\left\{\sum_{i=1}^{m}\frac{\gamma_i}{1+\gamma_i}-\frac{1}{1+L^2}\sum_{i=1}^{m}\gamma_i\right\} \quad (78)$$

where we used the fact that from (77), $\sum_{i=1}^{m}\gamma_i=\beta L^2$. Since β no longer appears in the objective, (77) can be expressed as

$$\sum_{i=1}^{m}\gamma_i\le L^2$$
$$\gamma_i\ge 0,\quad 1\le i\le m. \quad (79)$$

As we expect, the resulting dual problem of (78) and (79) is equivalent to the dual problem of (39) and (40) derived in Section IV for the case in which $\mathbf{T}=\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ when substituting $\sigma_i=1$. Indeed, if $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}=\mathbf{I}$, then the weighting matrices considered in Section IV and V are equal so that the corresponding optimization problems must coincide.

We summarize our results in the following theorem.

Theorem 3: 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 covariance $\mathbf{C}_w$. Let $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}=\mathbf{V}\Sigma\mathbf{V}^*$, where $\mathbf{V}$ is a unitary matrix, and Σ is an $m\times m$ diagonal matrix with diagonal elements $\sigma_1\ge\ldots\ge\sigma_m>0$. Then, the solution to the problem

$$\min_{\hat{\mathbf{x}}=\mathbf{Gy}}\max_{\|\mathbf{x}\|\le L}\left\{E\left(\|\hat{\mathbf{x}}-\mathbf{x}\|^2\right)-\min_{\hat{\mathbf{x}}=\mathbf{G}(\mathbf{x})\mathbf{y}}E\left(\|\hat{\mathbf{x}}-\mathbf{x}\|^2\right)\right\}$$

has the form

$$\hat{\mathbf{x}}=\mathbf{VDV}^*\left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}$$

where $\mathbf{D}$ is an $m\times m$ diagonal matrix with diagonal elements d_i that are given by

$$d_i=\begin{cases}1-\sqrt{\lambda-\sigma_i\mu}, & i\le k\\ 0, & i\ge k+1\end{cases}$$

with $k=\arg\min\tau_i$, $\mu=\mu_k$, and $\lambda=\lambda_k$. Here, τ_i, μ_i, and λ_i are the optimal solutions to the problem (Γ_i) given by (75).

We now show that as we expect intuitively, when $L\to\infty$, the minimax regret estimator $\hat{\mathbf{x}}$ of Theorem 3 reduces to the least-squares estimator (5). From (75), it follows that if $L\to\infty$ and $\lambda>0$, then $-2L\sqrt{\mu}+\mu+L^2\lambda+\sum_{i=1}^{k}(1-\sqrt{\lambda-\sigma_i\mu})^2/\sigma_i\to L^2\lambda$, which implies that $\tau\to\infty$. Therefore, to minimize τ, we must have that $\lambda=0$, which immediately implies that $\mu=0$

since we must have that $\lambda\ge\sigma_1\mu$ and $\mu\ge 0$. In addition, since for $k<m$, $\lambda\ge 1+\sigma_{k+1}\mu$, we must have that $k=m$. We conclude that for $L\to\infty$, $\lambda=\mu=0$ and $k=m$, which from Theorem 3 implies that $\mathbf{D}=\mathbf{I}$, and $\hat{\mathbf{x}}$ reduces to the least-squares estimator.

VI. EXAMPLES

We now present some examples that illustrate the performance advantage of the minimax regret estimator.

We consider the problem of estimating a two-dimensional (2-D) image from noisy observations, which are obtained by blurring the image with a blurring kernel (a 2-D filter) and adding random Gaussian noise. Specifically, we generate an image $x(z_1,z_2)$, which is the sum of m harmonic oscillations:

$$x(z_1,z_2)=\sum_{\ell=1}^{m}a_\ell\cos(\omega_{\ell,1}z_1+\omega_{\ell,2}z_2+\phi_\ell) \quad (80)$$

where

$$\omega_{\ell,i}=\frac{2\pi k_{\ell,i}}{n} \quad (81)$$

and $k_{\ell,i}\in\mathbb{Z}^2$ are given parameters. Clearly, the image $x(z_1,z_2)$ is periodic with period n. Therefore, we can represent the image by a length-n^2 vector $\mathbf{x}$, with components $\{x(z_1,z_2):0\le z_1,z_2\le n-1\}$.

The observed image $y(z_1,z_2)$ is given by

$$y(z_1,z_2)=\sum_{\tau_1,\tau_2}H(\tau_1,\tau_2)x(z_1-\tau_1-d_1,z_2-\tau_2-d_2)$$
$$+\sigma w(z_1,z_2),\quad 0\le z_1,z_2\le n-1 \quad (82)$$

where $H(z_1,z_2)$ is a blurring filter defined by

$$H(z_1,z_2)=\max\left(1-\frac{\sqrt{z_1^2+z_2^2}}{\rho},0\right) \quad (83)$$

for some parameter ρ, d_1 and d_2 are randomly chosen shifts, and $w(z_1,z_2)$ is an independent, zero-mean, Gaussian noise process so that for each z_1 and z_2, $w(z_1,z_2)$ is $\mathcal{N}(0,1)$.

By defining the vectors $\mathbf{y}$ and $\mathbf{w}$ with components $y(z_1,z_2)$ and $w(z_1,z_2)$, respectively, and defining a matrix $\mathbf{H}$ with the appropriate elements $H(z_1,z_2)$, the observations $\mathbf{y}$ can be expressed in the form of a linear model $\mathbf{y}=\mathbf{Hx}+\mathbf{w}$.

To evaluate the performance of the minimax regret estimator, we consider four different data sets, with parameters given by Table I. The filters used in all four simulations have, up to shifts, the same support $\{(0,0);(0,1);(0,-1);(1,0);(-1,0)\}$; however, the kernel used for the first data set is essentially different from the kernels used for data sets numbers 2–4, which are identical up to shifts of each other. The distributions of the singular values and the condition numbers of the kernels are shown in Fig. 2.

To estimate the image $x(z_1,z_2)$ from the noisy observations $y(z_1,z_2)$, we consider four different estimators: the least-squares (LS) estimator of (5), the minimax regret (REG) estimator of Theorem 3, and two other estimators—the deterministic Wiener estimator (WNR) and the minimax estimator (MMX)—which we now describe. We assume that $L=\|\mathbf{x}\|$ and the noise variance σ are known.
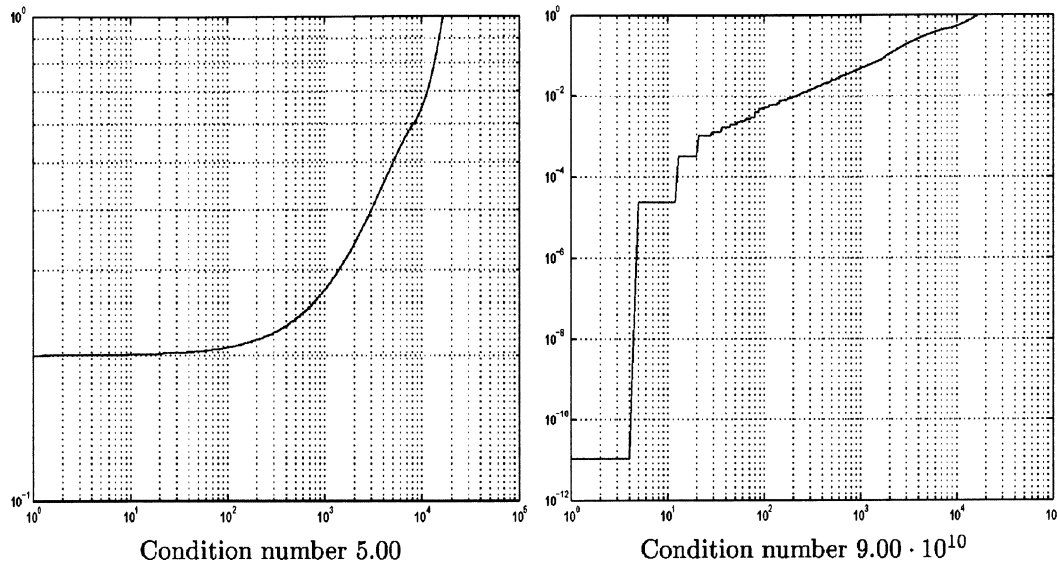
Fig. 2.   Distribution of the singular values of the $\mathbf{H}$-matrix for data set # 1 (left) and data sets # 2–4 (right).

TABLE I
PARAMETERS FOR THE FOUR DATA SETS

Data Set	m	n	σ	$n \cdot \rho$	$(k_{\ell,1}, k_{\ell,2})$	a_ℓ	ϕ_ℓ
1	5	128	0.50	1.200000	(2,1)	0.9235	1.7870
					(1,2)	1.0155	2.9482
					(1,1)	0.8340	0.4070
					(2,2)	0.9329	6.2099
					(1,3)	0.7259	3.6618
2	5	128	0.50	1.414214	(1,1)	1.0681	2.1438
					(2,1)	0.8704	3.3557
					(1,2)	1.2027	4.5686
					(2,2)	1.0466	1.9433
					(3,2)	0.9449	5.2684
3	5	128	0.50	1.414214	(16,11)	0.5784	5.0572
					(4,6)	1.1408	5.7076
					(13,24)	0.6909	1.4570
					(6,19)	1.3439	1.5036
					(14,14)	0.6739	0.3126
4	5	128	0.50	1.414214	(97,52)	1.1649	1.0072
					(51,107)	1.3704	5.4843
					(101,41)	0.5099	1.4946
					(6,66)	0.6370	4.0579
					(123,39)	1.3188	6.0751

TABLE II
RELATIVE ERROR FOR THE DATA SETS OF TABLE I

Data	Estimator	Relative Error
1	LS	0.748
	MMX	0.599
	WNR	0.731
	RGR	0.599
2	LS	5.0e8
	MMX	1.00
	WNR	6.17
	RGR	0.843
3	LS	5.0e8
	MMX	1.00
	WNR	6.65
	RGR	0.881
4	LS	5.0e8
	MMX	1.00
	WNR	6.16
	RGR	0.969

The least-squares estimator does not incorporate the prior knowledge on σ and the image norm $L = \|\mathbf{x}\|$. To develop an estimator that incorporates this knowledge, we may assume that $\mathbf{x}$ is a random vector with covariance $L^2 \mathbf{I}$ independent of the noise $\mathbf{w}$, and we may design an MMSE Wiener filter matched to this covariance. The resulting estimator is [2]

$$\hat{\mathbf{x}} = \left(\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H} + \mathbf{C}_x^{-1}\right)^{-1} \mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{y}$$
$$= \left(\mathbf{H}^* \mathbf{H} + \frac{\sigma^2}{L^2} \mathbf{I}\right)^{-1} \mathbf{H}^* \mathbf{y}. \qquad (84)$$

The minimax estimator is developed in [22] and is designed to minimize the worst-case MSE over all possible values of $\mathbf{x}$ such that $\mathbf{x}^* \mathbf{x} \leq L^2$, i.e., it is the solution to the problem

$$\min_{\hat{\mathbf{x}} = \mathbf{G} \mathbf{y}} \max_{\|\mathbf{x}\| \leq L} E\left(\|\hat{\mathbf{x}} - \mathbf{x}\|^2\right) \qquad (85)$$

and is given by

$$\hat{\mathbf{x}} = \frac{L^2}{L^2 + \gamma_0} \left(\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H}^*\right)^{-1} \mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{y} \qquad (86)$$

where $\gamma_0 = \mathrm{Tr}((\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H})^{-1})$ is the variance of the least-squares estimator.

In Table II, we report the relative error $\epsilon = \|\hat{\mathbf{x}} - \mathbf{x}\|/\|\mathbf{x}\|$ corresponding to the four estimators for each of the four data sets. As can be seen in Fig. 2, for the first data set, where the matrix $\mathbf{H}$ is perfectly conditioned, all of the methods work reasonably well. In contrast, for data sets number 2–4, where $\mathbf{H}$ is poorly conditioned, the performance of the least-squares, minimax, and Wiener estimators are severely degraded. The surprising result is that even though the matrix is ill-conditioned, the minimax regret estimator works pretty well, as can be seen from the results of Table II, as well as in Figs. 3 and 4.

In Figs. 3 and 4, we plot the original image, the observations, and the estimated image for data sets 2 and 3. Since the error in the least-squares estimate is so large, we do not show the
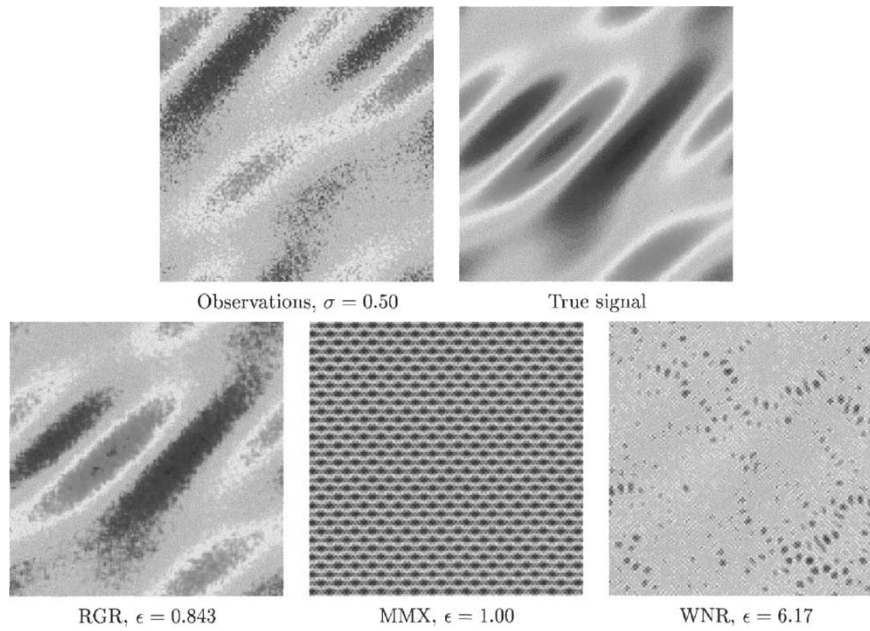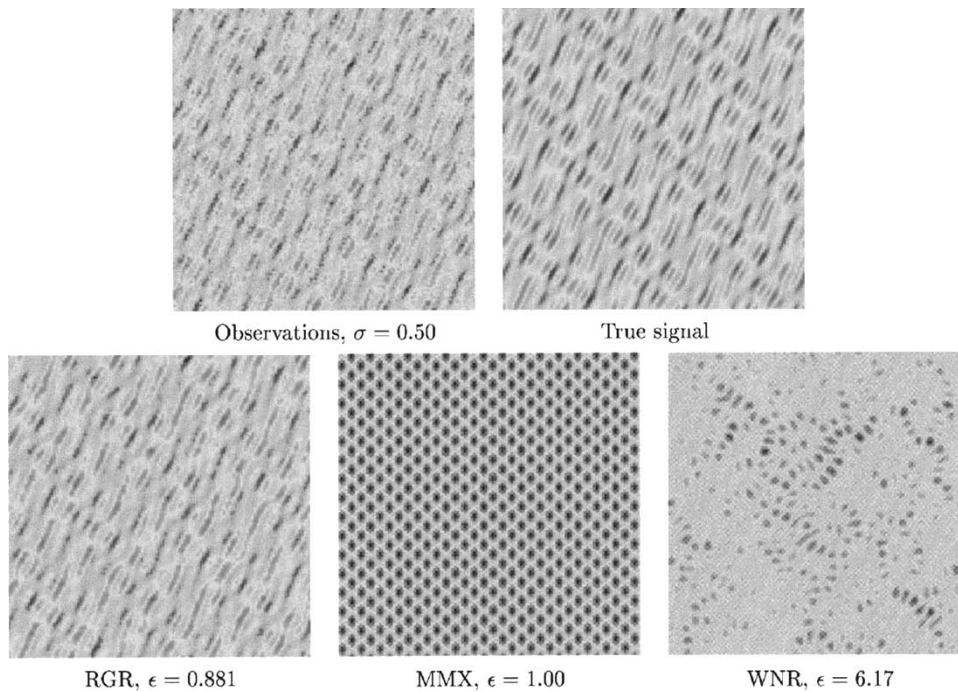
Fig. 3. Data set number 2.

Fig. 4. Data set number 3.

resulting image. In the images, the "more red" the image, the larger the signal value at that point.

VII. CONCLUSION

We considered the problem of estimating an unknown deterministic parameter vector $\mathbf{x}$ in the linear model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where $\mathbf{x}$ is known to be bounded so that $\|\mathbf{x}\|_{\mathbf{T}} \leq L$ for some weighting matrix $\mathbf{T}$. We developed a new linear estimator based on minimizing the worst-case regret, which is the difference between the MSE of the estimator and the best possible MSE attainable with a linear estimator that knows $\mathbf{x}$. As we demon-

strated, the minimax regret approach can significantly increase the performance over the traditional least-squares method, even in cases where the least-squares estimator, as well as other linear estimators, turn out to be completely useless.

There are, of course, examples where the minimax regret, as well as all other *linear* estimators, will perform poorly, in which case, one may need to consider nonlinear estimators.

In our development of the minimax regret, we assumed that $\mathbf{T}$ and $\mathbf{H}\mathbf{C}_w^{-1}\mathbf{H}$ have the same eigenvector matrix. An interesting direction for future research is to develop the minimax regret estimator for more general classes of weighting matrices $\mathbf{T}$, as well as in the presence of uncertainties in $\mathbf{H}$. It is also of interest

to consider the minimax regret estimator for the case in which the noise covariance $\mathbf{C}_w$ is subject to uncertainties.

ACKNOWLEDGMENT

REFERENCES

[1] N. Wiener, *The Extrapolation, Interpolation and Smoothing of Stationary Time Series*. New York: Wiley, 1949.

[2] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, Third ed. New York: McGraw-Hill, 1991.

[3] K. G. Gauss, *Theory of Motion of Heavenly Bodies*. New York: Dover, 1963.

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

[5] A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problems*. Washington, DC: V.H. Winston, 1977.

[6] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometr.*, vol. 12, pp. 55–67, Feb. 1970.

[7] L. S. Mayer and T. A. Willke, "On biased estimation in linear models," *Technometr.*, vol. 15, pp. 497–508, Aug. 1973.

[8] Y. C. Eldar and A. V. Oppenheim, "Covariance shaping least-squares estimation," *IEEE Trans. Signal Processing*, vol. 51, pp. 686–697, Mar. 2003.

[9] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Statist.*, vol. 35, pp. 73–101, 1964.

[10] ——, *Robust Statistics*. New York: Wiley, 1981.

[11] L. Breiman, "A note on minimax filtering," *Ann. Probab.*, vol. 1, pp. 175–179, 1973.

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

[13] H. V. Poor, "On robust Wiener filtering," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 521–526, June 1980.

[14] K. S. Vastola and H. V. Poor, "Robust Wiener–Kolmogorov theory," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 316–327, Mar. 1984.

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

[16] G. Moustakides and S. A. Kassam, "Minimax robust equalization for random signals through uncertain channels," in *Proc. 20th Annu. Allerton Conf. Commun., Contr. Comput.*, Oct. 1982, pp. 945–954.

[17] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation in linear models," *IEEE Trans. Signal Processing*, vol. 52, pp. 1931–1946, July 2004.

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

[19] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: a survey," *Proc. IEEE*, vol. 73, pp. 433–481, Mar. 1985.

[20] S. Verdú and H. V. Poor, "On minimax robustness: a general approach and applications," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 328–340, Mar. 1984.

[21] S. A. Kassam and H. V. Poor, "Robust signal processing for communication systems," *IEEE Commun. Mag.*, vol. 21, pp. 20–28, 1983.

[22] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, ""Robust mean-squared error estimation with bounded data uncertainties," *IEEE Trans. Signal Processing*, to be published.

[23] L. D. Davisson, "Universal noiseless coding," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 783–795, Nov. 1973.

[24] M. Feder and N. Merhav, "Universal composite hypothesis testing: a competitive minimax approach," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1504–1517, June 2002.

[25] E. Levitan and N. Merhav, "A competitive Neyman-Pearson approach to universal hypothesis testing with applications," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2215–2229, Aug. 2002.

[26] N. Merhav and M. Feder, "Universal prediction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2124–2147, Oct. 1998.

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

[28] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1968.

[29] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*, ser. MPS-SIAM Optimization, 2001.

[30] D. P. Bertsekas, *Nonlinear Programming*, Second ed. Belmont, MA: Athena Scientific, 1999.

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

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

Dr. Eldar was in the program for outstanding students at TAU from 1992 to 1996. In 1998, she held the Rosenblith Fellowship for study in Electrical Engineering at MIT, and in 2000, she held an IBM Research Fellowship. She is currently a Horev Fellow of the Leaders in Science and Technology program at the Technion as well as an Alon Fellow.
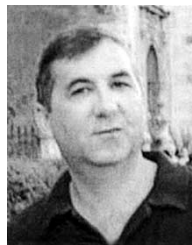
Aharon Ben-Tal received the Ph.D. degree in applied mathematics from Northwestern University, Chicago, IL, in 1973.

He is Professor of Operations Research and Head of Minerva Optimization Center at the Technion—Israel Institute of Technology, Haifa, Israel. He has been a Visiting Professor at the University of Michigan, Ann Arbor; University of Copenhagen, Copenhagen, Demark; and Delft University of Technology, Delft, The Netherlands. His recent interests are in continuous optimization, in particular, nonsmooth and large-scale problems, conic and robust optimization, as well as convex and nonsmooth analysis. In the last 10 years, he has devoted much effort to engineering applications of optimization methodology and computational schemes. He has published more than 90 papers and co-authored two monographs. He was Dean of the Faculty of Industrial Engineering and Management at the Technion from 1989 to 1992. He served as a Council Member of the Mathematical Programming Society from 1994 to 1997. He was Area Editor (for continuous optimization) of *Mathematics of Operations Research* from 1993 to 1999 and member of Editorial Boards of the *SIAM Journal on Optimization* and *Mathematical Modeling and Numerical Analysis*. He is currently Associate Editor of the *Journal of Convex Analysis* and *Mathematical Programming*.

Arkadi Nemirovski received the Ph.D. degree in mathematics from Moscow State University, Moscow, U.S.S.R., in 1973.

He is Professor of Operations Research and Deputy Head of the Minerva Optimization Center, the Technion—Israel Institute of Technology, Haifa, Israel. He has been a visiting professor at Delft Technical University, Delft, The Netherlands; and the Georgia Institute of Technology, Atlanta. His recent interests are in the area of convex optimization (complexity of convex programming. He has co-authored four monographs and over 80 papers. He has served as an Associate Editor of *Mathematics of Operations Research*.

Dr. Nemirovski received the Fulkerson Prize from the Mathematical Programming Society and the AMS in 1982, the Dantzig Prize from the MPS and SIAM in 1991, and the John von Neumann Theory Prize of INFORMS in 2003.