Robust Mean-Squared Error Estimation in the Presence of Model Uncertainties

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

***Abstract*—We consider the problem of estimating an unknown parameter vector $\mathbf{x}$ in a linear model that may be subject to uncertainties, where the vector $\mathbf{x}$ is known to satisfy a weighted norm constraint. We first assume that the model is known exactly and seek the linear estimator that minimizes the worst-case mean-squared error (MSE) across all possible values of $\mathbf{x}$. We show that for an arbitrary choice of weighting, the optimal minimax MSE estimator can be formulated as a solution to a semidefinite programming problem (SDP), which can be solved very efficiently. We then develop a closed form expression for the minimax MSE estimator for a broad class of weighting matrices and show that it coincides with the shrunken estimator of Mayer and Willke, with a specific choice of shrinkage factor that explicitly takes the prior information into account.**

Next, we consider the case in which the model matrix is subject to uncertainties and seek the robust linear estimator that minimizes the worst-case MSE across all possible values of $\mathbf{x}$ *and* all possible values of the model matrix. As we show, the robust minimax MSE estimator can also be formulated as a solution to an SDP.

Finally, we demonstrate through several examples that the minimax MSE estimator can significantly increase the performance over the conventional least-squares estimator, and when the model matrix is subject to uncertainties, the robust minimax MSE estimator can lead to a considerable improvement in performance over the minimax MSE estimator.

***Index Terms*—Data uncertainty, linear estimation, mean squared error estimation, minimax estimation, robust estimation.**

I. Introduction

The problem of estimating a set of unknown deterministic parameters $\mathbf{x}$ observed through a linear transformation $\mathbf{H}$, and corrupted by additive noise $\mathbf{w}$, arises in a large variety of areas in science and engineering, e.g., communication, economics, signal processing, seismology, and control.

Owing to the lack of statistical information about the parameters $\mathbf{x}$, often, the estimated parameters are chosen to optimize a criterion based on the observed signal $\mathbf{y}$. The celebrated least-squares (LS) estimator [1]–[3], which was first used by Gauss to predict movements of planets [4], seeks the linear 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 LS sense, to the given data vector $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$ so that $\hat{\mathbf{x}}$ is chosen to minimize the Euclidian norm of the data error $\hat{\mathbf{y}} - \mathbf{y}$. 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 $\hat{\mathbf{y}} - \mathbf{y}$.

To develop an estimation method that is based directly on the estimation error, we may seek the estimator $\hat{\mathbf{x}}$ that minimizes the mean-squared error (MSE), where the MSE of an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ is the expected value of the squared norm of the estimation error and is equal to the sum of the variance and the squared norm of the bias. Since the bias generally depends on the unknown parameters $\mathbf{x}$, we cannot choose an estimator to directly minimize the MSE. A common approach is to restrict the estimator to be linear and unbiased and then seek the estimator of this form that minimizes the variance or the MSE. It is well known that the LS estimator minimizes the variance in the estimate $\hat{\mathbf{x}}$ among all *unbiased* linear estimators. However, this does not imply that the LS estimator leads to a small variance or a small mean-squared error (MSE). A difficulty often encountered in this estimation problem is that the resulting variance can be very large, particularly in nonorthogonal and ill-conditioned problems.

Various modifications of the LS estimator for the case in which the data model holds i.e., $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$ with $\mathbf{H}$ and $\mathbf{y}$ known exactly, have been proposed. Among the alternatives are Tikhonov regularization [5], which is also known in the statistical literature as the ridge estimator [6], the shrunken estimator [7], and the covariance shaping LS estimator [8], [9]. In general, these LS alternatives attempt to reduce the MSE in estimating $\mathbf{x}$ by allowing for a bias. However, each of the estimators above is designed to optimize an objective which does not depend directly on the MSE, but rather depends on the data error $\hat{\mathbf{y}} - \mathbf{y}$.

In many engineering applications, the model matrix $\mathbf{H}$ is also subject to uncertainties. For example, the matrix $\mathbf{H}$ may be estimated from noisy data, in which case, $\mathbf{H}$ may not be known exactly. If the actual data matrix deviates from the one assumed, then the performance of an estimator designed based on $\mathbf{H}$ alone may deteriorate considerably. Various methods have been proposed to account for uncertainties in $\mathbf{H}$. The Total LS method [10], [11] seeks the parameters $\mathbf{x}$ and the minimum perturbation to the model matrix $\mathbf{H}$ that minimize the data error. Although the total LS method allows for uncertainties in $\mathbf{H}$, in many cases, it results in correction terms that are unnecessarily large. In particular, when the model matrix $\mathbf{H}$ is square, the total LS method recuse to the conventional LS method, which does not take the uncertainties into account. Recently, several methods [12]–[14] have been developed to treat the case in which the perturbation to the model matrix $\mathbf{H}$ is bounded. These methods seek the parameters that minimize the worst-case data error across

Manuscript received January 30, 2003; revised January 29, 2004. The work of Y. C. Eldar was supported in part by the Taub Foundation and by a Technion V.P.R. Fund. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jean Pierre Delmas.

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 the 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).

Digital Object Identifier 10.1109/TSP.2004.838933

all bounded perturbations of $\mathbf{H}$ and possibly bounded perturbations of the data vector. In [15], the authors seek the estimator that minimizes the best possible data error over all possible perturbations of $\mathbf{H}$. Here again, the above objectives depend on the data error $\hat{\mathbf{y}} - \mathbf{y}$ and not on the estimation error or the MSE.

In this paper, we consider the case in which the (possibly weighted) norm of the unknown vector $\mathbf{x}$ is bounded and develop robust estimators $\hat{\mathbf{x}}$ of $\mathbf{x}$, whose performance is reasonably good across all possible values of $\mathbf{x}$ in the region of uncertainty, by minimizing objectives that depend explicitly on the MSE.

We first consider the case in which $\mathbf{H}$ is known exactly and develop a minimax linear robust estimator that minimizes the worst-case MSE across all possible bounded values of $\mathbf{x}$, i.e., over all values of $\mathbf{x}$ such that $\mathbf{x}^*\mathbf{T}\mathbf{x} \le L^2$ for some constant L and weighting matrix $\mathbf{T}$. The minimax MSE estimator for the special case in which $\mathbf{H} = \mathbf{I}$ and the covariance matrix $\mathbf{C}_w$ of the noise vector $\mathbf{w}$ is given by $\mathbf{C}_w = \sigma^2\mathbf{I}$ for some $\sigma^2 > 0$ has been developed in [16]. Here, we extend the results to arbitrary $\mathbf{H}$ and $\mathbf{C}_w$ and show that the minimax MSE estimator can be formulated as the solution to a semidefinite programming problem (SDP) [17]–[19], which is a tractable convex optimization problem that can be solved efficiently, e.g., using interior point methods [19], [20]. We then develop a closed-form solution to the minimax estimation problem for the case in which the weighting $\mathbf{T}$ and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have the same eigenvector matrix. In particular, when $\mathbf{T} = \mathbf{I}$, we show that the optimal estimator is a shrunken estimator proposed by Mayer and Willke [7], with a specific choice of shrinkage factor, that explicitly takes the prior information into account. We demonstrate through simulations, that the minimax MSE estimator can increase the performance over the conventional LS approach.

We then consider the case in which the model matrix is not known exactly, but is rather given by $\mathbf{H} + \delta\mathbf{H}$, where $\mathbf{H}$ is known, and $\delta\mathbf{H}$ is a bounded perturbation matrix. Under this model, we seek a robust linear estimator that minimizes the worst-case MSE across all possible values of $\mathbf{x}$ and $\delta\mathbf{H}$. Here again, we show that the optimal estimator can be found by solving an SDP. In the special case in which the weighting $\mathbf{T}$ and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have the same eigenvector matrix and $\mathbf{C}_w$ and $\mathbf{H}\mathbf{H}^*$ have the same eigenvector matrix, we show that the minimax MSE estimator can be found by solving a convex optimization problem in two unknowns, regardless of the problem dimension.

We note that an alternative way to account for bounds on $\mathbf{x}$ is through regularization methods, such as Tikhonov regularization [5]. A more general regularization method that takes uncertainties in $\mathbf{H}$, as well as possibly other data uncertainties, into account, was developed in [21]. However, these methods are based on minimizing a weighted data error, whereas our approach directly minimizes the estimation error. In a companion paper [27], we consider a minimax regret approach that also depends explicitly on the MSE rather than the data error.

The paper is organized as follows. In Section II, we consider the case in which $\mathbf{H}$ is known and develop an SDP formulation of the linear minimax MSE estimator that minimizes the worst-case MSE across all possible bounded parameters $\mathbf{x}$. In Section III, we develop a closed-form expression for the minimax linear estimator in the case in which the weighting $\mathbf{T}$ and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have the same eigenvector matrix. In Section IV, we consider the case in which both $\mathbf{x}$ and the model matrix $\mathbf{H}$ are subject to uncertainties and show that the minimax MSE estimator that minimizes the worst-case MSE in the region of uncertainty can again be formulated as an SDP. We then consider, in Section V, the special case in which $\mathbf{T}$ and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have the same eigenvector matrix and $\mathbf{C}_w$ and $\mathbf{H}\mathbf{H}^*$ have the same eigenvector matrix. Examples illustrating the performance advantage of the minimax MSE estimator over the LS estimator, and the advantage of the robust minimax MSE estimator over the minimax MSE estimator in the presence of model uncertainties, are discussed in Section VI.

II. Minimax MSE Estimation with Known $\mathbf{H}$

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 $\widehat{(\cdot)}$ denotes an estimated vector or matrix. The notation $\mathbf{A} \preceq \mathbf{B}$ means that $\mathbf{B} - \mathbf{A}$ is positive semidefinite.

Consider the problem of estimating the unknown deterministic parameters $\mathbf{x}$ 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 $\|\mathbf{x}\|_{\mathbf{T}} \le L$ for some positive definite matrix $\mathbf{T}$ 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 MSE of the estimator $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ is given by

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

The second term in (2) (the squared norm of the bias $B(\hat{\mathbf{x}})$) depends on the unknown parameters $\mathbf{x}$; thus, in general, we cannot construct an estimator to directly minimize the MSE. Instead, we seek the linear estimator that minimizes the worst-case MSE across all possible values of $\mathbf{x}$ satisfying $\|\mathbf{x}\|_{\mathbf{T}} \le L$. Thus, we consider the problem

$$\begin{aligned} &\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\|\mathbf{x}\|_{\mathbf{T}} \le L} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) = \\ &\min_{\mathbf{G}} \max_{\|\mathbf{x}\|_{\mathbf{T}} \le L} \left\{ \mathbf{x}^*(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{x} + \mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) \right\}. \end{aligned} \tag{3}$$

To develop the solution to (3), we first determine the worst possible parameters $\mathbf{x}$, i.e., the parameters that are the solution to the inner problem in (3):

$$\max_{\|\mathbf{x}\|_{\mathbf{T}} \le L} \mathbf{x}^*(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{x}. \tag{4}$$

By introducing the change of variable $\mathbf{z} = \mathbf{T}^{1/2}\mathbf{x}$, we have that

$$\begin{aligned} &\max_{\mathbf{x}^*\mathbf{T}\mathbf{x} \le L^2} \mathbf{x}^*(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{x} \\ &\quad= \max_{\mathbf{z}^*\mathbf{z} \le L^2} \mathbf{z}^*\mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{T}^{-1/2}\mathbf{z} \\ &\quad= L^2\lambda_{\max} \end{aligned} \tag{5}$$

where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{GH})^*(\mathbf{I}-\mathbf{GH})\mathbf{T}^{-1/2}$. We can express $\lambda_{\max}$ as the solution to

$$\min_{\lambda} \lambda \tag{6}$$

subject to

$$\mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{GH})^*(\mathbf{I}-\mathbf{GH})\mathbf{T}^{-1/2} \preceq \lambda\mathbf{I}. \tag{7}$$

From (5)–(7), it follows that the problem (3) can be reformulated as

$$\min_{\mathbf{G},\lambda} \left\{\mathrm{Tr}\left(\mathbf{G}\mathbf{C}_w\mathbf{G}^*\right) + L^2\lambda\right\} \tag{8}$$

subject to (7), which in turn is equivalent to

$$\min_{\tau,\mathbf{G},\lambda} \tau \tag{9}$$

subject to

$$\begin{aligned} \mathrm{Tr}\left(\mathbf{G}\mathbf{C}_w\mathbf{G}^*\right) + L^2\lambda &\le \tau \\ \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{GH})^*(\mathbf{I}-\mathbf{GH})\mathbf{T}^{-1/2} &\preceq \lambda\mathbf{I}. \end{aligned} \tag{10}$$

We now show that the problem of (9) and (10) can be formulated as a standard semidefinite program (SDP) [17]–[19], which is the problem of minimizing a linear objective subject to linear matrix inequality (LMI) constraints. An LMI is a matrix constraint of the form $\mathbf{A}(\mathbf{x}) \succeq 0$, where the matrix $\mathbf{A}$ depends linearly on $\mathbf{x}$. The advantage if this formulation is that it readily lends itself to efficient computational methods. Indeed, by exploiting the many well known algorithms for solving SDPs [17], [18], e.g., interior point methods[1] [19], [20], the optimal estimator can be computed efficiently in polynomial time. Furthermore, SDP-based algorithms are guaranteed to converge to the global optimum.

A. Semidefinite Programming Formulation of the Estimation Problem

We now establish our claim that the problem of (9) and (10) can be formulated as an SDP. To this end, let $\mathbf{g} = \mathrm{vec}\left(\mathbf{G}\mathbf{C}_w^{1/2}\right)$, where $\mathbf{m} = \mathrm{vec}(\mathbf{M})$ denotes the vector obtained by stacking the columns of $\mathbf{M}$. With this notation, the constraints (10) become

$$\begin{aligned} \mathbf{g}^*\mathbf{g} + L^2\lambda &\le \tau \\ \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{GH})^*(\mathbf{I}-\mathbf{GH})\mathbf{T}^{-1/2} &\preceq \lambda\mathbf{I}. \end{aligned} \tag{11}$$

The constraints (11) are not in the form of an LMI because of the terms $\mathbf{g}^*\mathbf{g}$ and $\mathbf{T}^{-1/2}\mathbf{H}^*\mathbf{G}^*\mathbf{G}\mathbf{H}\mathbf{T}^{-1/2}$ in which the elements $\mathbf{G}(i,j)$ of $\mathbf{G}$ do not appear linearly. To express these inequalities as LMIs in the variables $\mathbf{G}(i,j)$, λ, and τ we rely on the following lemma [22, p. 472]:

[1]Interior point methods are iterative algorithms that terminate once a prespecified accuracy has been reached. A worst-case analysis of interior point methods shows that the effort required to solve an SDP to a given accuracy grows no faster than a polynomial of the problem size. In practice, the algorithms behave much better than predicted by the worst-case analysis, and in fact, in many cases, the number of iterations is almost constant in the size of the problem.

Lemma 1 (Schur's Complement): Let

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

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

From Lemma 1, it follows that the constraints (11) are satisfied if and only if

$$\begin{aligned} \begin{bmatrix} \tau - L^2\lambda & \mathbf{g}^* \\ \mathbf{g} & \mathbf{I} \end{bmatrix} &\succeq 0 \\ \begin{bmatrix} \lambda\mathbf{I} & \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{GH})^* \\ (\mathbf{I}-\mathbf{GH})\mathbf{T}^{-1/2} & \mathbf{I} \end{bmatrix} &\succeq 0. \end{aligned} \tag{12}$$

Note that the constraints (12) are indeed LMIs in the variables $\mathbf{G}$, λ, and τ. We conclude that the problem of (3) is equivalent to the SDP defined by (9) and (12).

In the next section, we develop an explicit expression for the optimal estimator that minimizes the worst-case estimation error in the case in which the weighting matrix $\mathbf{T}$ and the matrix $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have the same eigenvector matrix.

III. Minimax MSE Estimator for $\mathbf{T}$ and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ Jointly Diagonalizable

We now consider the case in which $\mathbf{T}$ 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 $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\Sigma\mathbf{V}^*$, where $\mathbf{V}$ is a unitary matrix and Σ is a diagonal matrix, then $\mathbf{T} = \mathbf{V}\Lambda\mathbf{V}^*$ for some diagonal matrix Λ. We then have the following proposition.

Proposition 1: Let $\mathbf{x}$ denote the deterministic unknown parameters 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 definitive covariance $\mathbf{C}_w$. Let $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\Sigma\mathbf{V}^*$ where Σ is a diagonal matrix with diagonal elements $\sigma_i > 0$, and let $\mathbf{T} = \mathbf{V}\Lambda\mathbf{V}^*$, where Λ is a diagonal matrix with diagonal elements $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_m > 0$. Then, the solution to the problem

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

is given by

$$\hat{\mathbf{x}} = \mathbf{P}_{\mathcal{V}_k}\left(\mathbf{I} - \alpha\mathbf{T}^{1/2}\right)\left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}$$

where

$$\mathbf{P}_{\mathcal{V}_k} = \mathbf{V}\begin{bmatrix} \mathbf{0} & \\ & \mathbf{I}_{m-k} \end{bmatrix}\mathbf{V}^* \tag{13}$$

is an orthogonal projection onto the space $\mathcal{V}_k$ spanned by the last $m-k$ columns of $\mathbf{V}$

$$\alpha = \frac{\sum_{i=k+1}^{m}\left(\frac{\lambda_i^{1/2}}{\sigma_i}\right)}{L^2 + \sum_{i=k+1}^{m}\left(\frac{\lambda_i}{\sigma_i}\right)} \tag{14}$$

and k is the smallest index such that $0 \le k \le m-1$ and

$$\alpha\lambda_{k+1}^{1/2} < 1. \tag{15}$$

Before proving Proposition 1, we note that there always exists a $0 \leq k \leq m-1$ satisfying (15). Indeed, for $k = m-1$, we have that

$$\alpha = \frac{\sum_{i=k+1}^{m}\left(\frac{\lambda_i^{1/2}}{\sigma_i}\right)}{L^2 + \sum_{i=k+1}^{m}\left(\frac{\lambda_i}{\sigma_i}\right)} = \frac{\frac{\lambda_m^{1/2}}{\sigma_m}}{L^2 + \frac{\lambda_m}{\sigma_m}} < \lambda_m^{-1/2} \tag{16}$$

so that $\alpha\lambda_m^{1/2} < 1$. For particular values of λ_i and σ_i, there may be smaller values of k for which (15) is satisfied.

Proof: The proof of Proposition 1 is comprised of three parts. First, we show that the optimal $\mathbf{G}$ minimizing the worst-case MSE has the form

$$\mathbf{G} = \mathbf{V}\mathbf{D}\mathbf{V}^* \left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^*\mathbf{C}_w^{-1} \tag{17}$$

for some $m \times m$ matrix $\mathbf{D}$. We then show that $\mathbf{D}$ can be chosen as a diagonal matrix. Finally, we derive the optimal values of the diagonal elements of $\mathbf{D}$.

We begin by showing that the optimal $\mathbf{G}$ has the form given by (17). To this end, note that the MSE of (2) depends on $\mathbf{G}$ only through $\mathbf{GH}$ and $\mathrm{Tr}\,(\mathbf{G}\mathbf{C}_w\mathbf{G}^*)$. Now, for any choice of $\mathbf{G}$

$$\begin{aligned}\mathrm{Tr}\,(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) &= \mathrm{Tr}\left(\mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{1/2}\mathbf{G}^*\right)\\ &\quad + \mathrm{Tr}\left(\mathbf{G}\mathbf{C}_w^{1/2}(\mathbf{I}-\mathbf{P})\mathbf{C}_w^{1/2}\mathbf{G}^*\right)\\ &\geq \mathrm{Tr}\left(\mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{1/2}\mathbf{G}^*\right)\end{aligned} \tag{18}$$

where

$$\mathbf{P} = \mathbf{C}_w^{-1/2}\mathbf{H}\left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^*\mathbf{C}_w^{-1/2} \tag{19}$$

is the orthogonal projection onto the range space of $\mathbf{C}_w^{-1/2}\mathbf{H}$. 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 $\mathrm{Tr}\,(\mathbf{G}\mathbf{C}_w\mathbf{G}^*)$, it is sufficient to consider matrices $\mathbf{G}$ that satisfy

$$\mathbf{G}\mathbf{C}_w^{1/2} = \mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}. \tag{20}$$

Substituting (19) into (20), we have

$$\begin{aligned}\mathbf{G} &= \mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{-1/2} = \mathbf{GH}\left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\\ &= \mathbf{B}\left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\end{aligned} \tag{21}$$

for some $m \times m$ matrix $\mathbf{B}$. Denoting by $\mathbf{D}$ the matrix $\mathbf{D} = \mathbf{V}^*\mathbf{B}\mathbf{V}$, (21) reduces to (17).

We now show that $\mathbf{D}$ can be chosen as a diagonal matrix. Since $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\Sigma\mathbf{V}^*$, we can express the constraints (10) as

$$\mathrm{Tr}\left(\mathbf{D}^*\mathbf{D}\Sigma^{-1}\right) + L^2\lambda \leq \tau \tag{22}$$

and

$$\mathbf{V}\Lambda^{-1/2}(\mathbf{I}-\mathbf{D})^*(\mathbf{I}-\mathbf{D})\Lambda^{-1/2}\mathbf{V}^* \preceq \lambda\mathbf{I} \tag{23}$$

which is equivalent to

$$\Lambda^{-1/2}(\mathbf{I}-\mathbf{D})^*(\mathbf{I}-\mathbf{D})\Lambda^{-1/2} \preceq \lambda\mathbf{I}. \tag{24}$$

Thus, our problem reduces to finding $\mathbf{D}$, τ, and λ that minimize τ subject to (22) and (24).

Let $\mathbf{J}$ be any diagonal matrix with diagonal elements equal to ± 1. If $\mathbf{D}$ satisfies the constraints (22) and (24), then so does $\mathbf{JDJ}$. Indeed

$$\begin{aligned}\mathrm{Tr}(\mathbf{J}^*\mathbf{D}^*\mathbf{J}^*\mathbf{JDJ}\Sigma^{-1}) &= \mathrm{Tr}(\mathbf{J}\mathbf{J}^*\mathbf{D}^*\mathbf{D}\Sigma^{-1})\\ &= \mathrm{Tr}(\mathbf{D}^*\mathbf{D}\Sigma^{-1})\end{aligned} \tag{25}$$

where we used the fact that diagonal matrices commute and that $\mathbf{J}^*\mathbf{J} = \mathbf{J}^2 = \mathbf{I}$. Similarly

$$\begin{aligned}&\Lambda^{-1/2}(\mathbf{I}-\mathbf{JDJ})^*(\mathbf{I}-\mathbf{JDJ})\Lambda^{-1/2}\\ &\quad = \Lambda^{-1/2}\mathbf{J}(\mathbf{I}-\mathbf{D})^*\mathbf{J}^*\mathbf{J}(\mathbf{I}-\mathbf{D})\mathbf{J}\Lambda^{-1/2}\\ &\quad = \mathbf{J}\Lambda^{-1/2}(\mathbf{I}-\mathbf{D})^*(\mathbf{I}-\mathbf{D})\Lambda^{-1/2}\mathbf{J}.\end{aligned} \tag{26}$$

Since $\Lambda^{-1/2}(\mathbf{I}-\mathbf{D})^*(\mathbf{I}-\mathbf{D})\Lambda^{-1/2} \preceq \lambda\mathbf{I}$ if and only if $\mathbf{J}\Lambda^{-1/2}(\mathbf{I}-\mathbf{D})^*(\mathbf{I}-\mathbf{D})\Lambda^{-1/2}\mathbf{J} \preceq \lambda\mathbf{I}$, we conclude that if $\mathbf{D}$ satisfies (24), then so does $\mathbf{JDJ}$. Therefore, if $\mathbf{D}$ is an optimal matrix that minimizes τ subject to (22) and (24), then $\mathbf{JDJ}$ is also an optimal solution. Now, since the problem of minimizing τ subject to (22) and (24) is convex, the set of optimal solutions is also convex [23], 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 $\mathbf{D}'$ is a diagonal matrix. Therefore, we have shown that there exists an optimal diagonal solution $\mathbf{D}$.

Denote by d_i the diagonal elements of $\mathbf{D}$. Then, our problem reduces to

$$\min_{d_i,\lambda}\left\{\sum_{i=1}^{m}\frac{d_i^2}{\sigma_i} + L^2\lambda\right\} \tag{27}$$

subject to

$$(1-d_i)^2 - \lambda_i\lambda \leq 0, \quad 1 \leq i \leq m. \tag{28}$$

Since the problem of (27) and (28) is a convex optimization problem, from Lagrange duality theory [24], it follows that the value of the minimum in (27), which we denote by A, is equal to the optimal value of the dual problem, namely

$$A = \max_{\beta_i \geq 0}\min_{d_i,\lambda}\mathcal{L}(d_i,\lambda) \tag{29}$$

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

$$\begin{aligned}\mathcal{L}(d_i,\lambda) &= \sum_{i=1}^{m}\frac{d_i^2}{\sigma_i} + L^2\lambda + \sum_{i=1}^{m}\left(\beta_i(1-d_i)^2 - \beta_i\lambda_i\lambda\right)\\ &= \sum_{i=1}^{m}\left(\frac{d_i^2}{\sigma_i} + \beta_i(1-d_i)^2\right) + \left(L^2 - \beta_i\lambda_i\right)\lambda.\end{aligned} \tag{30}$$

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

$$d_i = \frac{\beta_i \sigma_i}{1+\beta_i\sigma_i}. \quad (31)$$

For this choice of d_i

$$\frac{d_i^2}{\sigma_i} + \beta_i(1-d_i)^2 = \frac{\beta_i}{1+\beta_i\sigma_i} \quad (32)$$

so that

$$\min_{d_i,\lambda} \mathcal{L}(d_i,\lambda) = \begin{cases} \sum_{i=1}^m \frac{\beta_i}{1+\beta_i\sigma_i}, & \sum_{i=1}^m \lambda_i\beta_i = L^2 \\ -\infty, & \text{otherwise.} \end{cases} \quad (33)$$

The dual problem associated with (27) and (28) is therefore

$$\max_{\beta_i} \sum_{i=1}^m \frac{\beta_i}{1+\beta_i\sigma_i} \quad (34)$$

subject to

$$\beta_i \geq 0, \quad 1 \leq i \leq m \quad (35)$$

$$\sum_{i=1}^m \lambda_i\beta_i = L^2. \quad (36)$$

To solve (34) subject to (35) and (36), we form the Lagrangian

$$\mathcal{L} = -\sum_{i=1}^m \frac{\beta_i}{1+\beta_i\sigma_i} - \sum_{i=1}^m \zeta_i\beta_i + \mu\left(\sum_{i=1}^m \lambda_i\beta_i - L^2\right) \quad (37)$$

where from the Karush–Kuhn–Tucker (KKT) conditions [24], $\zeta_i \geq 0$. Differentiating $\mathcal{L}$ with respect to β_i and equating to 0

$$\frac{1}{(1+\beta_i\sigma_i)^2} + \zeta_i = \mu\lambda_i, \quad 1 \leq i \leq m \quad (38)$$

so that $\mu > 0$. If $\mu\lambda_i > 1$, then from (38), $\zeta_i > 0$, and from the KKT conditions, $\beta_i = 0$. If $\mu\lambda_i \leq 1$, then to satisfy (38), we must have that $\zeta_i = 0$. In this case

$$\beta_i = \frac{1}{\sigma_i}\left(\frac{1}{\sqrt{\lambda_i\mu}} - 1\right). \quad (39)$$

Thus, the optimal value of β_i is

$$\beta_i(\mu) = \begin{cases} \frac{1}{\sigma_i}\left(\frac{1}{\sqrt{\lambda_i\mu}} - 1\right), & \mu < \frac{1}{\lambda_i} \\ 0, & \mu \geq \frac{1}{\lambda_i} \end{cases} \quad (40)$$

where μ is chosen to satisfy (36).

We now show that there is a unique $\mu > 0$ satisfying (36). Defining

$$\mathcal{G}(\mu) = \sum_{i=1}^m \lambda_i\beta_i(\mu) - L^2 \quad (41)$$

μ is a root of $\mathcal{G}(\mu)$. Clearly, $\mathcal{G}(\mu)$ is monotonically decreasing for $0 < \mu \leq \lambda_m$, where $\lambda_m = \min_i \lambda_i$. In addition, $\mathcal{G}(\mu) > 0$ for $\mu \to 0$, and $\mathcal{G}(\mu) = -L^2$ for $\mu \geq \lambda_m$. Therefore, there is a unique $0 < \mu < \lambda_m$ satisfying (36).

Substituting the optimal value of β_i into (31), we have that

$$d_i = \begin{cases} 1 - \sqrt{\mu}\sqrt{\lambda_i}, & i > k \\ 0, & i \leq k \end{cases} \quad (42)$$

where

$$\sqrt{\mu} = \frac{\sum_{i=k+1}^m \left(\frac{\lambda_i^{1/2}}{\sigma_i}\right)}{L^2 + \sum_{i=k+1}^m \left(\frac{\lambda_i}{\sigma_i}\right)} \quad (43)$$

and k is chosen as the smallest index for which $\sqrt{\mu}\lambda_{k+1}^{1/2} < 1$. Denoting $\alpha = \sqrt{\mu}$ completes the proof of the proposition.

In the special case in which $\mathbf{T} = \mathbf{I}$, we can immediately verify that $k = 0$ so that the optimal estimator $\hat{\mathbf{x}}$ reduces to

$$\hat{\mathbf{x}} = \frac{L^2}{L^2+\gamma_0}\left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}^*\right)^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y} \quad (44)$$

where

$$\gamma_0 = \sum_{i=1}^m \frac{1}{\sigma_i} = \mathrm{Tr}\left(\left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1}\right) \quad (45)$$

is the variance corresponding to the LS estimator. The estimator given by (44) is a shrunken estimator proposed by Mayer and Willke [7], which is simply a scaled version of the LS 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 $\mathbf{x}$ in the linear model (1) such that $\|\mathbf{x}\| \leq L$, it minimizes the worst-case estimation error.

As we expect intuitively, when $L \to \infty$, $\hat{\mathbf{x}}$ of (44) reduces to the LS estimator. Indeed, when the norm of $\mathbf{x}$ can be made arbitrarily large, the MSE will also be arbitrarily large unless the bias is equal to zero. Therefore, in this limit, the worst-case estimation error is minimized by choosing an estimator with zero bias that minimizes the variance, which leads to the LS estimator.

We summarize our results in the following theorem.

Theorem 1: Let $\mathbf{x}$ denote the deterministic unknown parameters in the model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where $\mathbf{H}$ is a known $n \times m$ matrix with rank m, and $\mathbf{w}$ is a zero-mean random vector with positive definite covariance $\mathbf{C}_w$. Then, the problem

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

is equivalent to the semidefinite programming problem

$$\min_{\tau,\mathbf{G},\lambda} \tau$$

subject to

$$\begin{bmatrix} \tau - L^2\lambda & \mathbf{g}^* \\ \mathbf{g} & \mathbf{I} \end{bmatrix} \succeq 0$$

$$\begin{bmatrix} \lambda\mathbf{I} & \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{G}\mathbf{H})^* \\ (\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{T}^{-1/2} & \mathbf{I} \end{bmatrix} \succeq 0$$

where $\mathbf{g} = \mathrm{vec}\left(\mathbf{G}\mathbf{C}_w^{1/2}\right)$. In addition, we have the following.

1) If $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ and $\mathbf{T}$ have the same eigenvector matrix so that $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\Sigma\mathbf{V}^*$, where Σ is diagonal with diagonal elements σ_i and $\mathbf{T} = \mathbf{V}\Lambda\mathbf{V}^*$, where Λ is diagonal with diagonal elements $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m > 0$, then

$$\hat{\mathbf{x}} = \mathbf{P}_{\mathcal{V}_k}(\mathbf{I} - \alpha\mathbf{T}^{1/2})\left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}$$

where $\mathbf{P}_{\mathcal{V}_k}$ is an orthogonal projection onto the space $\mathcal{V}_k$ spanned by the last $m-k$ columns of $\mathbf{V}$ and is defined by (13), α is defined by (14), and k is the smallest index such that $0 \leq k \leq m-1$ and $\alpha\lambda_{k+1}^{1/2} < 1$.

2) If $\mathbf{T} = \mathbf{I}$, then

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

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

In Section VI, we provide several examples that illustrate the performance advantage of the minimax MSE estimator over the conventional LS estimator. In [28], we prove analytically that the MSE of the minimax MSE estimator is smaller than that of the LS estimator *for all* $\|\mathbf{x}\|_{\mathbf{T}} \leq \mathbf{L}$.

Before proceeding to the case in which $\mathbf{H}$ is also subject to uncertainties, we note that in Theorem 1, the bound L on the norm of $\mathbf{x}$ is assumed to be known. If L is not given *a priori*, then one possibility is to choose L to be equal to the norm of the LS estimator of $\mathbf{x}$, namely

$$L^2 = \mathbf{y}^*\mathbf{C}_w^{-1}\mathbf{H}\left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-2}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}. \tag{46}$$

IV. Minimax Estimation with Unknown $\mathbf{H}$

In the previous section, we developed the optimal estimator that minimizes the worst-case estimation error across all possible values of $\mathbf{x}$ that are bounded. In our development, we assumed that the model matrix $\mathbf{H}$ is known exactly. However, in many engineering applications, the model matrix $\mathbf{H}$ is subject to uncertainties, for example, it may have been estimated from noisy data, in which case, $\mathbf{H}$ is an approximation to some nominal underlying matrix. If the true data matrix is $\mathbf{H} + \delta\mathbf{H}$ for some unknown perturbation matrix $\delta\mathbf{H}$, then the actual performance of an estimator designed based on $\mathbf{H}$ alone may perform poorly.

In this section, we consider robust estimators that explicitly take uncertainties in $\mathbf{H}$ into account. Specifically, suppose now that the model matrix $\mathbf{H}$ is not known exactly but is rather given by $\mathbf{H} + \delta\mathbf{H}$, where $\|\delta\mathbf{H}\| \leq \rho$, and $\|\cdot\|$ denotes the matrix spectral norm [22], i.e., the largest singular value of the corresponding matrix. Our problem then is

$$\begin{aligned}&\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\|\mathbf{x}\|_{\mathbf{T}}\leq L, \|\delta\mathbf{H}\|\leq\rho} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)\\ &= \min_{\mathbf{G}}\left\{\max_{\|\mathbf{x}\|_{\mathbf{T}}\leq L, \|\delta\mathbf{H}\|\leq\rho}\left\{\mathbf{x}^*(\mathbf{I} - \mathbf{G}(\mathbf{H}+\delta\mathbf{H}))^*\right.\right.\\ &\qquad\left.\left.\times(\mathbf{I} - \mathbf{G}(\mathbf{H}+\delta\mathbf{H}))\mathbf{x}\right\} + \mathrm{Tr}\left(\mathbf{G}\mathbf{C}_w\mathbf{G}^*\right)\right\}.\end{aligned} \tag{47}$$

To develop the solution to (47), we first consider the inner maximization problem

$$\max_{\|\mathbf{x}\|_{\mathbf{T}}\leq L, \|\delta\mathbf{H}\|\leq\rho} \mathbf{x}^*(\mathbf{I}-\mathbf{G}(\mathbf{H}+\delta\mathbf{H}))^*(\mathbf{I}-\mathbf{G}(\mathbf{H}+\delta\mathbf{H}))\mathbf{x}. \tag{48}$$

We note that

$$\begin{aligned}&\max_{\|\mathbf{x}\|_{\mathbf{T}}\leq L, \|\delta\mathbf{H}\|\leq\rho} \mathbf{x}^*(\mathbf{I} - \mathbf{G}(\mathbf{H}+\delta\mathbf{H}))^*(\mathbf{I} - \mathbf{G}(\mathbf{H}+\delta\mathbf{H}))\mathbf{x}\\ &= \max_{\|\delta\mathbf{H}\|\leq\rho}\max_{\|\mathbf{x}\|\leq L} \mathbf{x}^*\mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{G}(\mathbf{H}+\delta\mathbf{H}))^*\\ &\quad\times(\mathbf{I} - \mathbf{G}(\mathbf{H}+\delta\mathbf{H}))\mathbf{T}^{-1/2}\mathbf{x}\\ &= \max_{\|\delta\mathbf{H}\|\leq\rho} L^2\lambda_{\max}(\delta\mathbf{H})\end{aligned} \tag{49}$$

where $\lambda_{\max}(\delta\mathbf{H})$ is the largest eigenvalue of the matrix $\mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{G}(\mathbf{H}+\delta\mathbf{H}))^*(\mathbf{I} - \mathbf{G}(\mathbf{H}+\delta\mathbf{H}))\mathbf{T}^{-1/2}$. We can express (49) as the solution to the problem

$$\min_{\tau} L^2\tau \tag{49}$$

subject to

$$\mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{G}(\mathbf{H}+\delta\mathbf{H}))^*(\mathbf{I} - \mathbf{G}(\mathbf{H}+\delta\mathbf{H}))\mathbf{T}^{-1/2} \preceq \tau\mathbf{I} \quad \forall\,\delta\mathbf{H} : \|\delta\mathbf{H}\| \leq \rho. \tag{51}$$

Using Lemma 1 we can rewrite the constraint (51) as

$$\begin{bmatrix}\tau\mathbf{I} & \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{G}(\mathbf{H}+\delta\mathbf{H}))^*\\ (\mathbf{I}-\mathbf{G}(\mathbf{H}+\delta\mathbf{H}))\mathbf{T}^{-1/2} & \mathbf{I}\end{bmatrix} \succeq 0 \quad \forall\,\delta\mathbf{H} : \|\delta\mathbf{H}\| \leq \rho \tag{52}$$

which is equivalent to

$$\mathbf{A}(\tau, \mathbf{G}) \succeq \mathbf{B}^*(\mathbf{G})\delta\mathbf{H}\mathbf{C} + \mathbf{C}^*(\delta\mathbf{H})^*\mathbf{B}(\mathbf{G}), \forall\,\delta\mathbf{H} : \|\delta\mathbf{H}\| \leq \rho \tag{53}$$

where

$$\begin{aligned}\mathbf{A}(\tau, \mathbf{G}) &= \begin{bmatrix}\tau\mathbf{I} & \mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{G}\mathbf{H})^*\\ (\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{T}^{-1/2} & \mathbf{I}\end{bmatrix}\\ \mathbf{B}(\mathbf{G}) &= [\mathbf{0} \quad \mathbf{G}^*]\\ \mathbf{C} &= [\mathbf{T}^{-1/2} \quad \mathbf{0}].\end{aligned} \tag{54}$$

We now exploit the following proposition, the proof of which is provided in the Appendix.

Lemma 2: Given matrices $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{A}$ with $\mathbf{A} = \mathbf{A}^*$

$$\mathbf{A} \succeq \mathbf{P}^*\mathbf{X}\mathbf{Q} + \mathbf{Q}^*\mathbf{X}^*\mathbf{P}, \quad \forall\,\mathbf{X} : \|\mathbf{X}\| \leq \rho$$

if and only if there exists a $\lambda \geq 0$ such that

$$\begin{bmatrix}\mathbf{A} - \lambda\mathbf{Q}^*\mathbf{Q} & -\rho\mathbf{P}^*\\ -\rho\mathbf{P} & \lambda\mathbf{I}\end{bmatrix} \succeq 0.$$

From Proposition 2, it follows that (53) is satisfied if and only if there exists a $\lambda \geq 0$ such that

$$\begin{bmatrix}\tau\mathbf{I} - \mathbf{T}^{-1} & \mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{G}\mathbf{H})^* & \mathbf{0}\\ (\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{T}^{-1/2} & \mathbf{I} & -\rho\mathbf{G}\\ \mathbf{0} & -\rho\mathbf{G}^* & \lambda\mathbf{I}\end{bmatrix} \succeq 0. \tag{55}$$

We conclude that the problem (47) can be expressed as

$$\min_{t,\mathbf{G},\lambda,\tau} t \tag{56}$$

subject to the LMI (55) and

$$\mathrm{Tr}\,(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + L^2\tau \le t. \tag{57}$$

Using Lemma 1, we can express (57) as the LMI

$$\begin{bmatrix} t - L^2\tau & \mathbf{g}^* \\ \mathbf{g} & \mathbf{I} \end{bmatrix} \succeq 0 \tag{58}$$

where $\mathbf{g} = \mathrm{vec}\left(\mathbf{G}\mathbf{C}_w^{1/2}\right)$, so that the problem of (56) subject to (57) and (55) can be formulated as an SDP.

We summarize our results in the following theorem.

Theorem 2: Let $\mathbf{x}$ denote the deterministic unknown parameters in the model $\mathbf{y} = (\mathbf{H} + \delta\mathbf{H})\mathbf{x} + \mathbf{w}$, where $\mathbf{H}$ is a known $n \times m$ matrix with rank m, $\delta\mathbf{H}$ is an unknown matrix satisfying $\|\delta\mathbf{H}\| \le \rho$, and $\mathbf{w}$ is a zero-mean random vector with positive definitive covariance $\mathbf{C}_w$. Then, the problem

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\|\mathbf{x}\|_{\mathbf{T}} \le L, \|\delta\mathbf{H}\| \le \rho} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)$$

is equivalent to the semidefinite programming problem

$$\min_{t,\mathbf{G},\lambda,\tau} t$$

subject to

$$\begin{bmatrix} t - L^2\tau & \mathbf{g}^* \\ \mathbf{g} & \mathbf{I} \end{bmatrix} \succeq 0$$

$$\begin{bmatrix} \tau\mathbf{I} - \mathbf{T}^{-1} & \mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{G}\mathbf{H})^* & \mathbf{0} \\ (\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{T}^{-1/2} & \mathbf{I} & -\rho\mathbf{G} \\ \mathbf{0} & -\rho\mathbf{G}^* & \lambda\mathbf{I} \end{bmatrix} \succeq 0$$

where $\mathbf{g} = \mathrm{vec}\left(\mathbf{G}\mathbf{C}_w^{1/2}\right)$.

In Theorem 2, the bound ρ on the norm of $\delta\mathbf{H}$ is assumed to be known. If the value of ρ is not specified, then one possibility is to choose ρ to be equal to the norm of the perturbation matrix resulting from the total LS estimator.

In Section VI, we demonstrate that the minimax MSE estimator of Theorem 2 explicitly takes the uncertainties in $\mathbf{H}$ into account and can in some cases significantly outperform the minimax MSE estimator of Theorem 1, which does not account for uncertainties in the model matrix.

V. Minimax Estimator for Jointly Diagonalizable Matrices

Suppose now that $\mathbf{T}$ and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have the same $m \times m$ eigenvector matrix $\mathbf{V}$, and $\mathbf{C}_w$ and $\mathbf{H}\mathbf{H}^*$ have the same $n \times n$ unitary eigenvector matrix $\mathbf{U}$. In this case

$$\begin{aligned} \mathbf{C}_w^{-1/2}\mathbf{H} &= \mathbf{U}\Sigma\mathbf{V}^* \\ \mathbf{C}_w &= \mathbf{U}\Theta\mathbf{U}^* \\ \mathbf{T} &= \mathbf{V}\Lambda\mathbf{V}^* \end{aligned} \tag{59}$$

where Σ and Θ are $n \times m$ diagonal matrices with diagonal elements $\sigma_i > 0, 1 \le i \le m$ and $\theta_i > 0, 1 \le i \le m$, respectively, and Λ is an $m \times m$ diagonal matrix with diagonal elements $\lambda_i > 0, 1 \le i \le m$. Note that in the case $\mathbf{T} = \mathbf{I}$ and $\mathbf{C}_w = \sigma^2\mathbf{I}$ for some $\sigma^2 > 0$, (59) is satisfied.

The assumption (59) is made for analytical tractability. If $\mathbf{H}$ and $\mathbf{T}$ represent convolution with some filter, and $\mathbf{w}$ is a stationary process, then $\mathbf{C}_w$, $\mathbf{H}$, and $\mathbf{T}$ will be Toeplitz matrices and are therefore approximately diagonalized by Fourier transform matrices of appropriate dimensions so that in this case, (59) is approximately satisfied [25].

We now show that in the case of jointly diagonalizable matrices as in (59), the minimax MSE estimator of Theorem 2 reduces to a simple convex optimization problem in two unknowns and can therefore be solved very efficiently, for example, using the Ellipsoidal method (see, e.g., [17, Ch. 5.2]). Specifically, we have the following theorem.

Theorem 3: Let $\mathbf{x}$ denote the deterministic unknown parameters in the model $\mathbf{y} = (\mathbf{H} + \delta\mathbf{H})\mathbf{x} + \mathbf{w}$, where $\mathbf{H}$ is a known $n \times m$ matrix with rank m, $\delta\mathbf{H}$ is an unknown matrix satisfying $\|\delta\mathbf{H}\| \le \rho$, and $\mathbf{w}$ is a zero-mean random vector with positive definitive covariance $\mathbf{C}_w$. Let $\mathbf{C}_w^{-1/2}\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^*$, where Σ is a diagonal matrix with diagonal elements $\sigma_i > 0$, let $\mathbf{C}_w = \mathbf{U}\Theta\mathbf{U}^*$, where Θ is a diagonal matrix with diagonal elements $\theta_i > 0$, and let $\mathbf{T} = \mathbf{V}\Lambda\mathbf{V}^*$, where Λ is a diagonal matrix with diagonal elements $\lambda_i > 0$. Then, the solution to the problem

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\|\mathbf{x}\|_{\mathbf{T}} \le L, \|\delta\mathbf{H}\| \le \rho} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)$$

is given by

$$\hat{\mathbf{x}} = \mathbf{V}\mathbf{Z}\mathbf{V}^*\left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1/2}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}.$$

Here $\mathbf{Z}$ is an $m \times m$ diagonal matrix with diagonal elements $z_i = f_i(\lambda, \tau)$, where

$$f_i(\lambda,\tau) = \frac{\sigma_i\lambda\theta_i - \sqrt{\lambda\theta_i(\tau\lambda_i - \lambda)\left(\sigma_i^2\lambda\theta_i - \rho^2(1 + \lambda - \tau\lambda_i)\right)}}{(\tau\lambda_i - \lambda)\rho^2 + \sigma_i^2\lambda\theta_i}$$

and λ and τ are the solution to the convex optimization problem

$$\min_{\tau,\lambda}\left\{\sum_{i=1}^m f_i^2(\lambda,\tau) + L^2\right\}$$

subject to

$$\begin{aligned} \lambda\theta_i\sigma_i^2 &\ge \rho^2(1 + \lambda - \lambda_i\tau), \quad 1 \le i \le m \\ \lambda &\ge 0 \\ \tau &\ge \frac{\lambda}{\min_i \lambda_i}. \end{aligned} \tag{60}$$

Proof: The proof is comprised of three parts. First, we show that the optimal $\mathbf{G}$ minimizing the worst-case MSE has the form

$$\mathbf{G} = \mathbf{V}\mathbf{Z}\mathbf{V}^*\left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1/2}\mathbf{H}^*\mathbf{C}_w^{-1} \tag{61}$$

for some $m \times m$ matrix $\mathbf{Z}$. We then show that $\mathbf{Z}$ can be chosen as a diagonal matrix. Finally, we derive the optimal values of the diagonal elements of $\mathbf{Z}$.

We begin by showing that the optimal $\mathbf{G}$ has the form (61). From Theorem 2, it follows that the minimax MSE estimator with a bounded uncertainty in $\mathbf{H}$ can be expressed as the solution to

$$\min_{\tau,\mathbf{G},\lambda} \left\{ \mathrm{Tr}\left(\mathbf{G}\mathbf{C}_w\mathbf{G}^*\right) + L^2\tau \right\} \tag{62}$$

subject to

$$\mathbf{M} \triangleq \begin{bmatrix} \tau\mathbf{I} - \lambda\mathbf{T}^{-1} & \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{G}\mathbf{H})^* & \mathbf{0} \\ (\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{T}^{-1/2} & \mathbf{I} & -\rho\mathbf{G} \\ \mathbf{0} & -\rho\mathbf{G}^* & \lambda\mathbf{I} \end{bmatrix} \succeq 0. \tag{63}$$

The constraint (63) is equivalent to $\mathbf{Q}\mathbf{M}\mathbf{Q}^* \succeq 0$ for any invertible $\mathbf{Q}$. Choosing

$$\mathbf{Q} = \begin{bmatrix} \Lambda^{1/2}\mathbf{V}^* & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{U}^*\mathbf{C}_w^{1/2} \end{bmatrix} \tag{64}$$

and using (59), (63) becomes

$$\begin{bmatrix} \tau\Lambda - \lambda\mathbf{I} & \mathbf{V}^*(\mathbf{I}-\mathbf{G}\mathbf{H})^*\mathbf{V} & \mathbf{0} \\ \mathbf{V}(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{V}^* & \mathbf{I} & -\rho\mathbf{V}^*\mathbf{G}\mathbf{C}_w^{1/2}\mathbf{U} \\ \mathbf{0} & -\rho\mathbf{U}^*\mathbf{C}_w^{1/2}\mathbf{G}^*\mathbf{V} & \lambda\Theta \end{bmatrix} \succeq 0. \tag{65}$$

Define

$$\mathbf{C} \triangleq \mathbf{V}^*\mathbf{G}\mathbf{C}_w^{1/2}\mathbf{U} \tag{66}$$

so that

$$\mathbf{G} = \mathbf{V}\mathbf{C}\mathbf{U}^*\mathbf{C}_w^{-1/2}. \tag{66}$$

Then, the problem of (62) and (65) can be expressed in terms of $\mathbf{C}$ as

$$\min_{\tau,\mathbf{C},\lambda} \left\{ \mathrm{Tr}(\mathbf{C}^*\mathbf{C}) + L^2\tau \right\} \tag{68}$$

subject to

$$\begin{bmatrix} \tau\Lambda - \lambda\mathbf{I} & (\mathbf{I}-\mathbf{C}\Sigma)^* & \mathbf{0} \\ \mathbf{I}-\mathbf{C}\Sigma & \mathbf{I} & -\rho\mathbf{C} \\ \mathbf{0} & -\rho\mathbf{C}^* & \lambda\Theta \end{bmatrix} \succeq 0. \tag{69}$$

Let $\mathbf{C} = [\,\mathbf{Z} \quad \mathbf{C}_2\,]$, where $\mathbf{Z}$ is the $m \times m$ matrix consisting of the first m columns of $\mathbf{C}$, let Σ_1 denote the $m \times m$ matrix with diagonal elements $\sigma_i, 1 \le i \le m$, let Θ_1 denote the $m \times m$ matrix with diagonal elements $\theta_i, 1 \le i \le m$, and let Θ_2 denote the $(n-m)\times(n-m)$ matrix with diagonal elements $\theta_i, m+1 \le i \le n$. Then, we can express the constraint (69) as

$$\mathbf{B}(\mathbf{C}) \triangleq \begin{bmatrix} \tau\Lambda - \lambda\mathbf{I} & (\mathbf{I}-\mathbf{Z}\Sigma_1)^* & \mathbf{0} & \mathbf{0} \\ \mathbf{I}-\mathbf{Z}\Sigma_1 & \mathbf{I} & -\rho\mathbf{Z} & -\rho\mathbf{C}_2 \\ \mathbf{0} & -\rho\mathbf{Z}^* & \lambda\Theta_1 & \mathbf{0} \\ \mathbf{0} & -\rho\mathbf{C}_2^* & \mathbf{0} & \lambda\Theta_2 \end{bmatrix} \succeq 0. \tag{70}$$

Clearly, if (70) is satisfied, then

$$\mathbf{A}(\mathbf{Z}) \triangleq \begin{bmatrix} \tau\Lambda - \lambda\mathbf{I} & (\mathbf{I}-\mathbf{Z}\Sigma_1)^* & \mathbf{0} \\ \mathbf{I}-\mathbf{Z}\Sigma_1 & \mathbf{I} & -\rho\mathbf{Z} \\ \mathbf{0} & -\rho\mathbf{Z}^* & \lambda\Theta_1 \end{bmatrix} \succeq 0. \tag{71}$$

TABLE I
SIGNAL PARAMETERS

ℓ	a_ℓ	$\omega_{\ell,1}$	$\omega_{\ell,2}$	ϕ_ℓ
1	1.3936	0.1473	0.0982	5.8777
2	0.5579	0.0982	0.0982	5.7611
3	0.8529	0.0491	0.0982	2.5778

Now, let $\mathbf{C} = [\,\mathbf{Z} \quad \mathbf{C}_2\,]$ be any matrix satisfying (70), and define $\widetilde{\mathbf{C}} = [\,\mathbf{Z} \quad \mathbf{0}\,]$. Then

$$\mathbf{B}(\widetilde{\mathbf{C}}) = \begin{bmatrix} \mathbf{A}(\mathbf{Z}) & \mathbf{0} \\ \mathbf{0} & \lambda\Theta_2 \end{bmatrix} \succeq 0 \tag{72}$$

since $\mathbf{A}(\mathbf{Z}) \succeq 0$. In addition

$$\begin{aligned} \mathrm{Tr}(\widetilde{\mathbf{C}}^*\widetilde{\mathbf{C}}) &= \mathrm{Tr}(\mathbf{Z}^*\mathbf{Z}) \le \mathrm{Tr}(\mathbf{Z}^*\mathbf{Z}) + \mathrm{Tr}(\mathbf{C}_2^*\mathbf{C}_2) \\ &= \mathrm{Tr}(\mathbf{C}^*\mathbf{C}). \end{aligned} \tag{73}$$

Therefore, the optimal value of $\mathbf{C}$ satisfies $\mathbf{C}_2 = \mathbf{0}$ so that the problem of (68) and (69) reduces to

$$\min_{\tau,\mathbf{Z},\lambda} \left\{ \mathrm{Tr}(\mathbf{Z}^*\mathbf{Z}) + L^2\tau \right\} \tag{74}$$

subject to (71). Once we find the optimal $\mathbf{Z}$, the optimal $\mathbf{G}$ can be found from (66) as

$$\begin{aligned} \mathbf{G} &= \mathbf{V}\mathbf{Z}[\,\mathbf{I} \quad \mathbf{0}\,]\,\mathbf{U}^*\mathbf{C}_w^{-1/2} \\ &= \mathbf{V}\mathbf{Z}\mathbf{V}^* \left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1/2} \mathbf{H}^*\mathbf{C}_w^{-1} \end{aligned} \tag{75}$$

which is equivalent to (61), thus completing the first part of the proof.

We now show that the optimal value of $\mathbf{Z}$ can be chosen as a diagonal matrix. To this end, we first note that if $\mathbf{Z}$ satisfies (71), then

$$\begin{aligned} &\begin{bmatrix} \mathbf{J} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{J} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J} \end{bmatrix} \\ &\quad \times \begin{bmatrix} \tau\Lambda - \lambda\mathbf{I} & (\mathbf{I}-\mathbf{Z}\Sigma_1)^* & \mathbf{0} \\ \mathbf{I}-\mathbf{Z}\Sigma_1 & \mathbf{I} & -\rho\mathbf{Z} \\ \mathbf{0} & -\rho\mathbf{Z}^* & \lambda\Theta_1 \end{bmatrix} \begin{bmatrix} \mathbf{J} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{J} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J} \end{bmatrix} \\ &= \begin{bmatrix} \tau\Lambda - \lambda\mathbf{I} & (\mathbf{I}-\mathbf{J}\mathbf{Z}\mathbf{J}\Sigma_1)^* & \mathbf{0} \\ \mathbf{I}-\mathbf{J}\mathbf{Z}\mathbf{J}\Sigma_1 & \mathbf{I} & -\rho\mathbf{J}\mathbf{Z}\mathbf{J} \\ \mathbf{0} & -\rho\mathbf{J}\mathbf{Z}^*\mathbf{J} & \lambda\Theta_1 \end{bmatrix} \succeq 0. \end{aligned} \tag{76}$$

Here, $\mathbf{J}$ is any diagonal matrix with diagonal elements ± 1. It follows from (76) that $\mathbf{A}(\mathbf{Z}_1) \succeq 0$ for any $\mathbf{J}$, where $\mathbf{Z}_1 = \mathbf{J}\mathbf{Z}\mathbf{J}$. In addition, we have that $\mathrm{Tr}\left(\mathbf{Z}_1^*\mathbf{Z}_1\right) = \mathrm{Tr}(\mathbf{Z}^*\mathbf{Z})$. Therefore, if $\mathbf{Z}$ is an optimal solution, then so is $\mathbf{J}\mathbf{Z}\mathbf{J}$. Since our problem is convex, the set of optimal solutions is also convex [23], which implies that the diagonal matrix $\mathbf{Z}' = (1/2^m)\sum_{\mathbf{J}} \mathbf{J}\mathbf{Z}\mathbf{J}$ is also a solution, where the summation is over all 2^m diagonal matrices $\mathbf{J}$ with diagonal elements ± 1. Therefore, we have shown that there exists an optimal diagonal solution $\mathbf{Z}$.

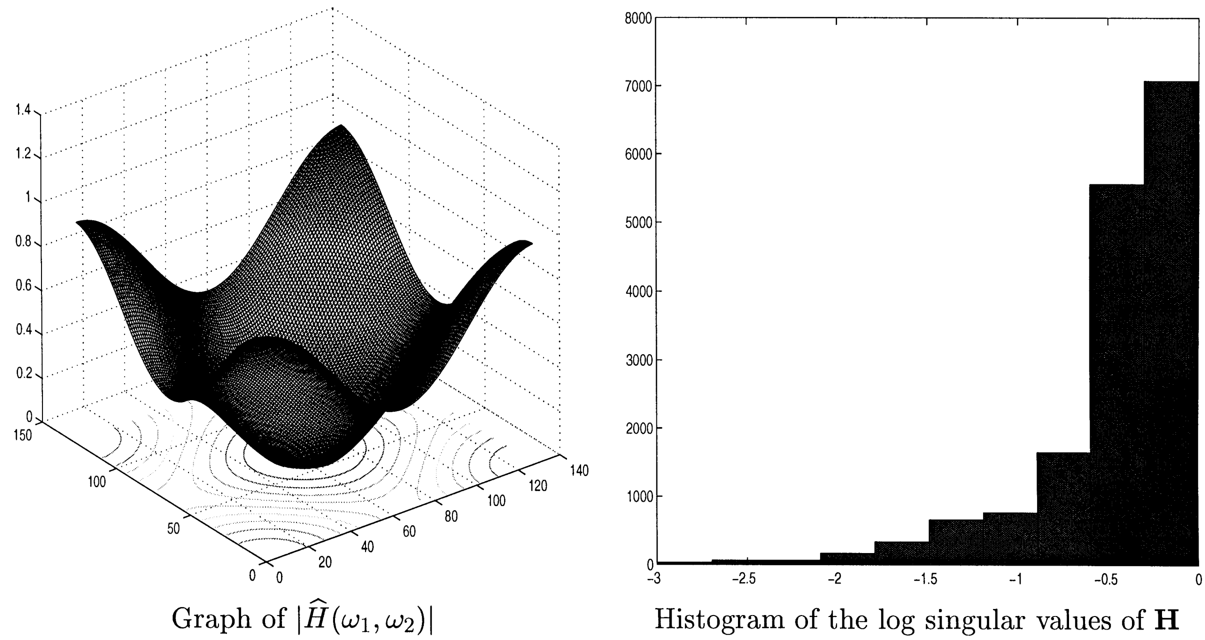


Fig. 1. Nominal blurring kernel $H(z_1, z_2)$.

Denote the diagonal elements of $\mathbf{Z}$ by z_i, and let $\mathrm{diag}(\alpha_1,\ldots\alpha_m)$ denote the $m \times m$ diagonal matrix with diagonal elements α_i. Then, the constraint $\mathbf{A}(\mathbf{Z}) \succeq 0$ can be written as (77), shown at the bottom of the page. By permuting the rows and the columns of the matrix in (77), we can transform it into a block diagonal matrix, where the ith block is

$$\begin{bmatrix} \tau\lambda_i - \lambda & 1 - z_i\sigma_i & 0 \\ 1 - z_i\sigma_i & 1 & -\rho z_i \\ 0 & -\rho z_i & \lambda\theta_i \end{bmatrix} \tag{78}$$

so that (77) is satisfied if and only if each of the matrices (78) is positive semidefinite. Thus, the problem of (74) and (71) become

$$\min_{\tau, z_i, \lambda} \left\{ \sum_{i=1}^{m} z_i^2 + L^2\tau \right\} \tag{79}$$

subject to

$$\begin{bmatrix} \tau\lambda_i - \lambda & 1 - z_i\sigma_i & 0 \\ 1 - z_i\sigma_i & 1 & -\rho z_i \\ 0 & -\rho z_i & \lambda\theta_i \end{bmatrix} \succeq 0, \quad 1 \le i \le m. \tag{80}$$

We now show that the problem of (79) subject to (80) can be further simplified. First, we note that to satisfy (80), we must have that

$$\tau \ge \frac{\lambda}{\lambda_i}, \quad 1 \le i \le m. \tag{81}$$

Suppose first that $\tau\lambda_i > \lambda$. In this case, using Lemma 1, (80) is equivalent to

$$\begin{bmatrix} 1 & -\rho z_i \\ -\rho z_i & \lambda\theta_i \end{bmatrix} - \frac{1}{\tau\lambda_i - \lambda} \begin{bmatrix} 1 - z_i\sigma_i \\ 0 \end{bmatrix} \begin{bmatrix} 1 - z_i\sigma_i & 0 \end{bmatrix} = \begin{bmatrix} 1 - \frac{(1 - z_i\sigma_i)^2}{\tau\lambda_i - \lambda} & -\rho z_i \\ -\rho z_i & \lambda\theta_i \end{bmatrix} \succeq 0. \tag{82}$$

Now, a 2×2 matrix is positive semidefinite if and only if the diagonal elements and the determinant are non-negative. Therefore, (82) is equivalent to the conditions

$$\lambda \ge 0 \tag{83}$$

$$\tau\lambda_i - \lambda \ge (1 - z_i\sigma_i)^2 \tag{84}$$

$$\lambda\theta_i \left(1 - \frac{(1 - z_i\sigma_i)^2}{\tau\lambda_i - \lambda}\right) - \rho^2 z_i^2 \ge 0. \tag{85}$$

Clearly, (84) and (82) together imply (84). Furthermore, we can express (84) as

$$z_i^2\left((\lambda - \tau\lambda_i)\rho^2 - \sigma_i^2\lambda\theta_i\right) + 2z_i\sigma_i\lambda\theta_i + \lambda\theta_i(\tau\lambda_i - \lambda - 1) \ge 0. \tag{86}$$

Since the coefficient multiplying z_i^2 in (86) is negative, it follows that there exists a z_i satisfying (86) if and only if the discriminant is non-negative, i.e., if and only if

$$\sigma_i^2\lambda\theta_i + \left((\tau\lambda_i - \lambda)\rho^2 + \sigma_i^2\lambda\theta_i\right)(\tau\lambda_i - \lambda - 1) \ge 0 \tag{87}$$

which, using the fact that $\tau\lambda_i - \lambda > 0$, is equivalent to

$$\lambda\theta_i\sigma_i^2 \ge \rho^2(1 + \lambda - \lambda_i\tau). \tag{88}$$

$$\begin{bmatrix} \mathrm{diag}(\tau\lambda_1 - \lambda, \ldots, \tau\lambda_m - \lambda) & \mathrm{diag}(1 - z_1\sigma_1, \ldots, 1 - z_m\sigma_m) & \mathbf{0} \\ \mathrm{diag}(1 - z_1\sigma_1, \ldots, 1 - z_m\sigma_m) & \mathbf{I} & -\rho\,\mathrm{diag}(z_1, \ldots, z_m) \\ \mathbf{0} & -\rho\,\mathrm{diag}(z_1, \ldots, z_m) & \lambda\,\mathrm{diag}(\theta_1, \ldots, \theta_m) \end{bmatrix} \succeq 0. \tag{77}$$

TABLE II
MSE USING THE LS, MINIMAX MSE, AND ROBUST MINIMAX MSE ESTIMATORS

Uncertainty level	Perturbation type	Estimator, $\alpha = 0.5$			Estimator, $\alpha = 1.0$		
		Robust	Minimax	LS	Robust	Minimax	LS
$\rho = 0.0$	1	0.519/0.525†	0.519	10.6	0.385/0.386	0.385	10.7
	2	0.527/0.525	0.527	10.1	0.386/0.386	0.386	10.8
	3	0.524/0.525	0.524	10.7	0.386/0.386	0.386	10.7
$\rho = 0.1$	1	0.345/0.654	0.600	11.2	0.319/0.459	0.428	11.0
	2	0.366/0.654	0.545	10.8	0.320/0.459	0.421	9.49
	3	0.406/0.654	0.587	10.4	0.320/0.459	0.422	9.66
$\rho = 0.2$	1	0.263/0.787	0.775	9.07	0.221/0.570	0.522	10.7
	2	0.264/0.787	0.773	11.3	0.220/0.570	0.444	9.43
	3	0.431/0.787	0.692	11.3	0.352/0.570	0.502	10.3
$\rho = 0.3$	1	0.227/0.909	0.991	11.4	0.175/0.686	0.653	11.4
	2	0.430/0.909	0.674	9.61	0.179/0.686	0.389	12.2
	3	0.472/0.909	0.771	11.0	0.399/0.686	0.578	11.6
$\rho = 0.4$	1	0.208/1.023	1.245	9.18	0.147/0.804	0.797	10.4
	2	0.523/1.023	0.827	10.4	0.147/0.804	0.404	11.2
	3	0.526/1.023	0.838	10.5	0.648/0.804	0.512	11.8
$\rho = 0.6$	1	0.457/1.230	1.773	12.2	0.113/1.041	1.120	10.0
	2	0.457/1.230	1.165	10.6	0.629/1.041	0.782	11.2
	3	0.706/1.230	0.920	10.9	0.950/1.041	0.724	11.0
$\rho = 0.9$	1	0.729/1.499	2.594	13.8	0.083/1.432	1.620	9.56
	2	0.733/1.499	1.600	10.4	0.082/1.432	1.568	11.3
	3	0.918/1.499	0.982	10.8	1.423/1.432	1.092	11.3

† The number after the slash is the robust optimal value.

If (88) is satisfied, then the set of z_i's satisfying (86) are $z_i^- \leq z_i \leq z_i^+$, where $z_i^- \leq z_i^+$ are the roots of the quadratic function in (86). Since we would like to choose z_i to minimize (79), it follows that the optimal z_i is

$$z_i = f_i(\tau, \lambda) = \frac{\sigma_i \lambda \theta_i - \sqrt{\lambda \theta_i (\tau \lambda_i - \lambda)\left(\sigma_i^2 \lambda \theta_i - \rho^2 (1 + \lambda - \tau \lambda_i)\right)}}{(\tau \lambda_i - \lambda)\rho^2 + \sigma_i^2 \lambda \theta_i}. \tag{89}$$

Thus, if $\tau\lambda_i > \lambda$, then the optimal value of z_i is given by (89), where in addition, conditions (88) and (84) must be satisfied.

Next, suppose that $\tau\lambda_i = \lambda$. In this case, to ensure that (80) is satisfied, we must have that

$$z_i = \frac{1}{\sigma_i} \tag{90}$$

$$\lambda \geq \frac{\rho^2}{\sigma_i^2 \theta_i}. \tag{91}$$

We can immediately verify that (90) and (91) are special cases of (89) and (88) with $\tau\lambda_i = \lambda$. We therefore conclude that the optimal value of z_i is given by (89), subject to (88) and (84). Substituting the optimal value of z_i into (79), our problem becomes

$$\min_{\tau, \lambda} \left\{ \sum_{i=1}^{m} f_i^2(\lambda, \tau) + L^2 \tau \right\} \tag{92}$$

subject to (60).

Since the problem of (79) subject to (80) is convex, and the reduced problem (91) subject to (92) is obtained by minimizing over one of the variables in (79), the reduced problem is also convex, completing the proof of the theorem. □

In Section VI, we illustrate the performance of the estimators of Theorems 1 and 3.

VI. EXAMPLES

The purpose of this section is to illustrate the performance advantage of the minimax MSE estimator of Theorem 1 over the conventional LS estimator and to demonstrate the fact that in the presence of uncertainties in the model matrix $\mathbf{H}$, robust estimation, which explicitly takes these uncertainties into account, can recover the signal much better, as compared with the total LS method and (nonrobust) minimax MSE estimation, particularly in those cases where the latter is expected to perform poorly.

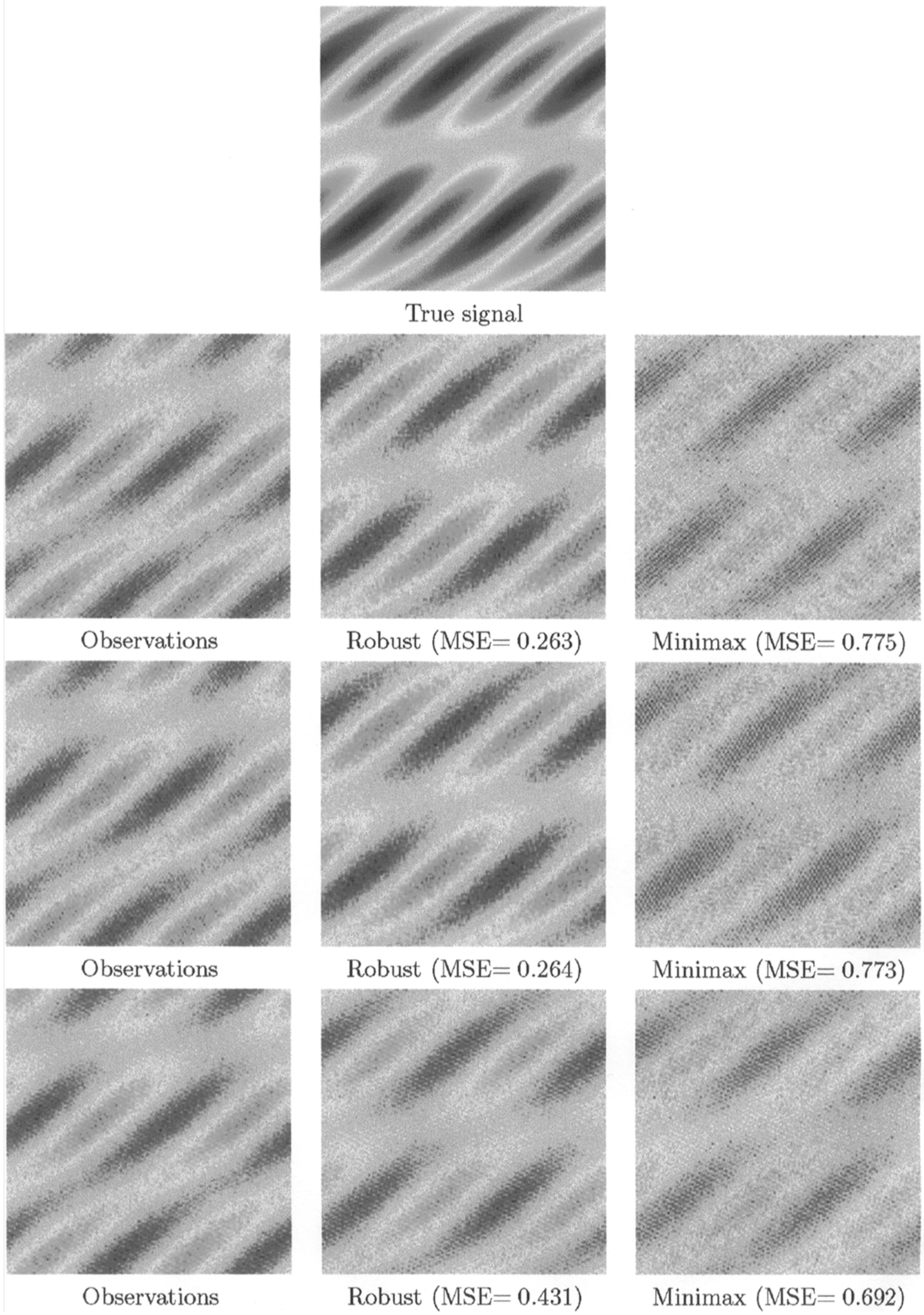


Fig. 2. Minimax MSE and robust minimax MSE estimation with $\rho = 0.2$ for the three different types of perturbations.

In the simulations below, 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) that may not be known exactly, and adding random Gaussian noise.

Specifically, we generate an image $x(z_1, z_2)$ which is the sum of three harmonic oscillations:

$$x(z_1, z_2) = \sum_{\ell=1}^{3} a_\ell \cos(\omega_{\ell,1} z_1 + \omega_{\ell,2} z_2 + \phi_\ell) \qquad (93)$$

where

$$\omega_{\ell,i} = \frac{2\pi k_{\ell,i}}{n} \qquad (94)$$

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 $\{x(z_1, z_2) : 0 \leq z_1, z_2 \leq n-1\}$. In the experiments, $n = 128$, and the amplitudes and frequencies of the harmonic components of $x(z_1, z_2)$

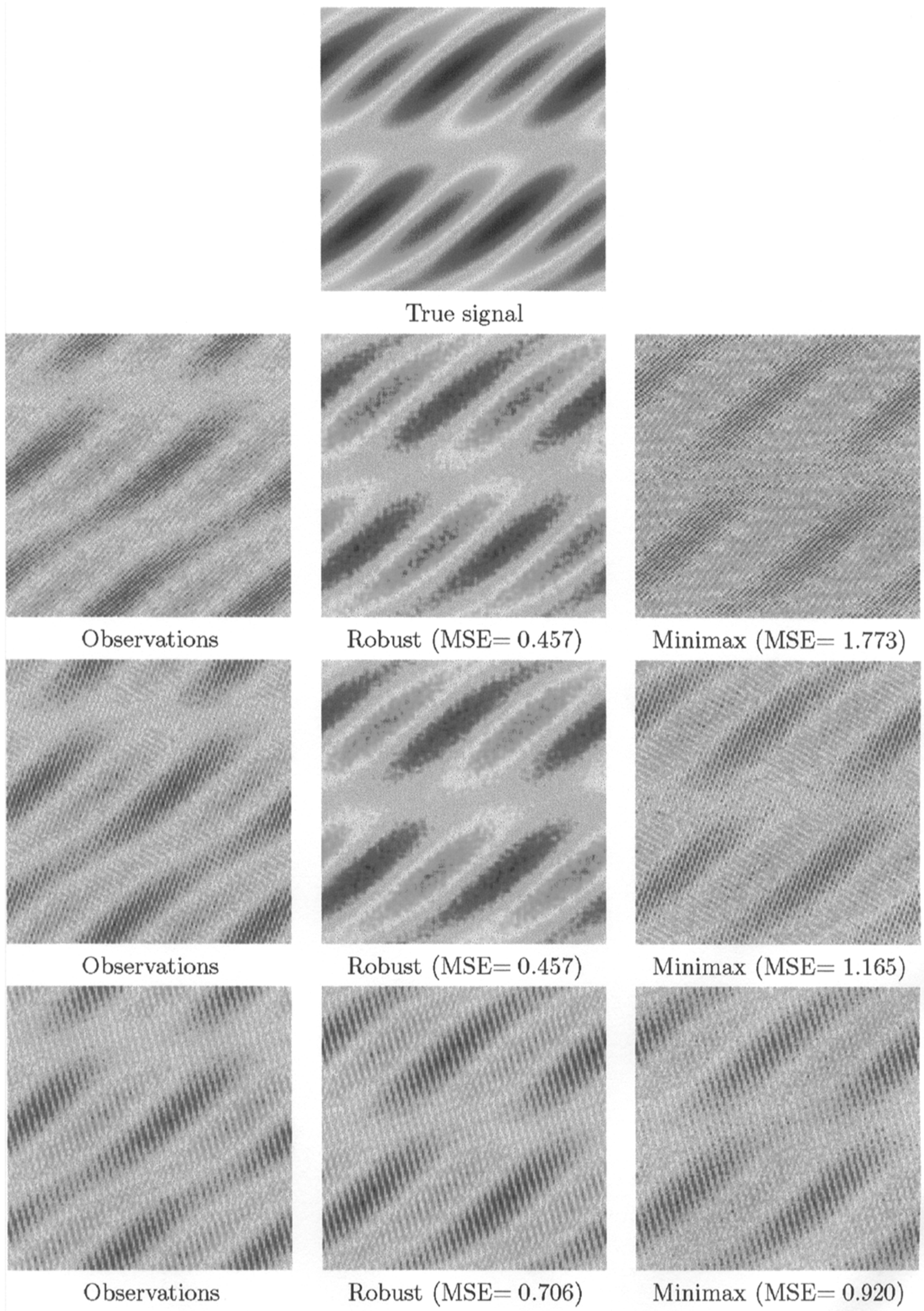


Fig. 3. Minimax MSE and robust minimax MSE estimation with $\rho = 0.6$ for the three different types of perturbations.

are given in Table I. Note that the *typical magnitude* of an entry in $x(z_1, z_2)$, i.e.,

$$\sqrt{n^{-2} \sum_{0 \le z_1, z_2 < n} x^2(z_1, z_2)}$$

is 1.578.

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

$$y(z_1, z_2) = \sum_{\tau_1, \tau_2} K(\tau_1, \tau_2) x(z_1 - \tau_1 - d_1, z_2 - \tau_2 - d_2) + \sigma w(z_1, z_2), \quad 0 \le z_1, z_2 \le n - 1 \tag{95}$$

where d_1 and d_2 are randomly chosen shifts (to make $y(z_1, z_2)$ visually distinct from $x(z_1, z_2)$), $w(z_1, z_2)$ is an independent, zero-mean, Gaussian noise process so that for each z_1 and z_2, $w(z_1, z_2)$ is $\mathcal{N}(0, 1)$, and $\sigma = 0.25$ is the noise variance. The convolutional kernel $K(z_1, z_2)$ is given by

$$K(z_1, z_2) = H(z_1, z_2) + \delta H(z_1, z_2) \tag{96}$$

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

$$H(z_1, z_2) = \frac{1}{\sum_{0 \le t_1, t_2 < n} L(t_1, t_2)} L(z_1, z_2) \tag{97}$$

with

$$L(z_1,z_2) = \max\left(1 - \frac{n\sqrt{(z_1-d_1)^2+(z_2-d_2)^2}}{1.4}, 0\right) \tag{98}$$

and $\delta H(z_1,z_2)$ is a bounded perturbation. The support of the convolution kernel $H(z_1,z_2)$ is, up to a shift, the five-point set $\Gamma = \{(0,0);(0,1);(0,-1);(1,0);(-1,0)\}$.

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 the matrices $\mathbf{H}$ and $\delta\mathbf{H}$ with the appropriate elements $H(z_1,z_2)$ and $\delta H(z_1,z_2)$, respectively, the observations $\mathbf{y}$ can be expressed in the form of an uncertain linear model $\mathbf{y} = (\mathbf{H}+\delta\mathbf{H})\mathbf{x}+\mathbf{w}$.

For our choice of $H(z_1,z_2)$, the norm (maximum singular value) of $\mathbf{H}$ is 1, and the condition number of $\mathbf{H}$ is 987.95. The magnitude of the Fourier transform $\widehat{H}(\omega_1,\omega_2)$, $0 \le \omega_1$, and $\omega_2 < n$ of $H(z_1,z_2)$ is shown in Fig. 1, together with a histogram of the log of the singular values of $\mathbf{H}$.

In the simulations below, we chose the weighting matrix $\mathbf{T} = (\mathbf{H}^*\mathbf{H})^{-\alpha}$ with $\alpha = 0.5$ or $\alpha = 1$. Note that $\mathbf{T}$ and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have the same eigenvector matrix since $\mathbf{C}_w = \sigma^2\mathbf{I}$. This choice of $\mathbf{T}$ reflects the fact that components of $\mathbf{x}$ corresponding to small singular values of $\mathbf{H}^*\mathbf{H}$ should receive a smaller weight then components corresponding to large singular values. For each choice of $\mathbf{T}$, we choose the bound L as $L = \mathbf{x}^*\mathbf{T}\mathbf{x}$. For the bound ρ on the norm of the perturbation matrix $\delta\mathbf{H}$, we used seven different values:

$$\rho = 0,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.6,\ 0.9.$$

The case $\rho = 0$ corresponds to the case in which the model matrix is equal to the nominal matrix $\mathbf{H}$. Therefore, in this case, the robust estimator coincides with the (nonrobust) minimax MSE estimator.

For each value of ρ, we consider three different choices of the perturbation matrix $\delta\mathbf{H}$, which are chosen as follows. For a given estimator $\mathbf{G}$, we define the worst-case perturbation $\delta\mathbf{H}$ as the one that maximizes

$$\begin{aligned}\delta\mathbf{H}_{\max} &= \arg\max_{\|\delta\mathbf{H}\|\le\rho} \|\, \mathbf{G}(\mathbf{H}+\delta\mathbf{H})\mathbf{x} - \mathbf{G}\mathbf{H}\mathbf{x}\,\| \\ &= \arg\max_{\|\delta\mathbf{H}\|\le\rho} \|\mathbf{G}\delta\mathbf{H}\mathbf{x}\|.\end{aligned} \tag{99}$$

Thus

$$\delta\mathbf{H}_{\max} = \frac{\rho}{\|\mathbf{x}\|}\mathbf{z}\mathbf{x}^* \tag{100}$$

where $\mathbf{z}$ is a unit singular vector of $\mathbf{G}$ corresponding to the largest singular value. In our simulations, we consider three special cases of $\delta\mathbf{H}_{\max}$.

1) worst case with respect to the minimax MSE estimator (i.e., $\mathbf{G}$ is chosen as the $\mathbf{G}$ of the minimax MSE estimator);
2) $\delta\mathbf{H}_{\max}$ given by (100) with a randomly chosen unit vector $\mathbf{z}$;
3) worst case with respect to the robust minimax MSE estimator.

For each choice of the perturbation matrix, we compute the MSE

$$\frac{1}{n}\sqrt{\sum_{0\le z_1,z_2<n} \left|x(z_1,z_2) - \hat{x}(z_1,z_2)\right|^2} \tag{101}$$

for the LS, minimax MSE, and robust minimax MSE estimators. Since, in our problem, $\mathbf{H}$ is square, the total LS method coincides with the LS method so that the LS performance is also the total LS performance. The MSE results are given in Table II.

As we expect, for $\rho = 0$, the minimax MSE and robust estimators coincide. We also see that the minimax MSE estimator can significantly outperform the LS estimator. For $\rho > 0$, the robust estimator that takes uncertainties in $\mathbf{H}$ into account can lead to improved performance over the LS, total LS, and minimax MSE estimators, which for large values of ρ can be quite significant. As we expect, the minimax MSE estimator performs best for case 3, while the robust estimator performs best for case 1. Note that even when the perturbation matrix is chosen to be worst for the robust estimator, the robust estimator still performs better than the minimax MSE estimator.

We note that we do not compare our results with those of [13] and [21] since the later methods require a prior bound on the norm of the data error, which we do not assume in our model.

In Figs. 2 and 3, we plot the original image, the observed image, and the estimated images using the minimax MSE estimator and the robust minimax MSE estimator for the three choices of perturbations, where in Fig. 2, $\rho = 0.2$, and in Fig. 3, $\rho = 0.6$. Since the error in the LS estimate is so large, we do not show the resulting image.

APPENDIX
PROOF OF PROPOSITION 2

To prove the proposition, we first note that

$$\mathbf{A} \succeq \mathbf{P}^*\mathbf{Z}\mathbf{Q} + \mathbf{Q}^*\mathbf{Z}^*\mathbf{P}, \quad \forall\, \mathbf{Z} : \|\mathbf{Z}\| \le \rho \tag{102}$$

if and only if for every $\mathbf{x}$

$$\begin{aligned}\mathbf{x}^*\mathbf{A}\mathbf{x} &\ge \max_{\|\mathbf{Z}\|\le\rho} \{\mathbf{x}^*\mathbf{P}^*\mathbf{Z}\mathbf{Q}\mathbf{x} + \mathbf{x}^*\mathbf{Q}^*\mathbf{Z}^*\mathbf{P}\mathbf{x}\} \\ &= 2\rho\|\mathbf{P}\mathbf{x}\|\|\mathbf{Q}\mathbf{x}\|.\end{aligned} \tag{103}$$

Using the Cauchy–Schwarz inequality, we can express (103) as

$$\mathbf{x}^*\mathbf{A}\mathbf{x} - 2\rho\mathbf{y}^*\mathbf{P}\mathbf{x} \ge 0, \quad \forall\, \mathbf{x},\mathbf{y} : \|\mathbf{y}\| \le \|\mathbf{Q}\mathbf{x}\|. \tag{104}$$

We now rely on the following lemma [26, p. 23].

Lemma 3: [$\mathcal{S}$-procedure] Let $P(\mathbf{z}) = \mathbf{z}^*\mathbf{A}\mathbf{z} + 2\mathbf{u}^*\mathbf{z} + v$ and $\mathcal{Q}(\mathbf{z}) = \mathbf{z}^*\mathbf{B}\mathbf{z} + 2\mathbf{x}^*\mathbf{z} + y$ be two quadratic functions of $\mathbf{z}$, where $\mathbf{A}$ and $\mathbf{B}$ are symmetric, and there exists a $\mathbf{z}_0$ satisfying $P(\mathbf{z}_0) > 0$. Then, the implication

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

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

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

Since $\|\mathbf{y}\| \leq \|\mathbf{Qx}\|$ is equivalent to $\mathbf{x}^*\mathbf{Q}^*\mathbf{Qx} - \mathbf{y}^*\mathbf{y} \geq 0$, we can use Lemma 3 to conclude that (104) is satisfied if and only if there exists a $\lambda \geq 0$ such that

$$\begin{bmatrix} \mathbf{A} - \lambda\mathbf{Q}^*\mathbf{Q} & -\rho\mathbf{P}^* \\ -\rho\mathbf{P} & \lambda\mathbf{I} \end{bmatrix} \succeq 0 \tag{105}$$

completing the proof.

REFERENCES

[1] T. Kailath, *Lectures On Linear Least-Squares Estimation*. New York: Springer, 1976.

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

[3] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1992.

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

[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, "Quantum Signal Processing," Ph.D. dissertation, Mass. Inst. Technol., Cambridge, MA, 2001, http://www.ee.technion.ac.il/Sites/People/YoninaEldar/Download/thesis.pdf.

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

[10] G. H. Golub and C. F. Van Loan, "An analysis of the total least-squares problem," *SIAM J. Numer. Anal.*, vol. 17, no. 6, pp. 883–893, Dec. 1980.

[11] S. Van Huffel and J. Vandewalle, *The Total Least-Squares Problem: Computational Aspects and Analysis*. Philadelphia, PA: SIAM, 1991, vol. 9, Frontier in Applied Mathematics.

[12] L. El Ghaoui and H. Lebret, "Robust solution to least-squares problems with uncertain data," *SIAM J. Matrix Anal. Appl.*, vol. 18, no. 4, pp. 1035–1064, 1997.

[13] S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed, "Parameter estimation in the presence of bounded data uncertainties," *SIAM J. Matrix Anal. Appl.*, vol. 19, no. 1, pp. 235–252, 1998.

[14] A. H. Sayed, V. H. Nascimento, and S. Chandrasekaran, "Estimation and control with bounded data uncertainties," *Linear Alg. Appl.*, vol. 284, no. 1, pp. 259–306, Nov. 1998.

[15] S. Chandrasekaran, M. Gu, A. H. Sayed, and K. E. Schubert, "The degenerate bounded error-in-variables model," *SIAM J. Matrix Anal. Appl.*, vol. 23, pp. 138–166, 2001.

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

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

[18] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 40–95, Mar. 1996.

[19] Y. Nesterov and A. Nemirovski, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA: SIAM, 1994.

[20] F. Alizadeh, "Combinatorial Optimization With Interior Point Methods and Semi-Definite Matrices," Ph.D. dissertation, Univ. Minnesota, Minneapolis, MN, 1991.

[21] A. H. Sayed, V. H. Nascimento, and F. A. M. Cipparrone, "A regularized robust design criterion for uncertain data," *SIAM J. Matrix Anal. Appl.*, vol. 23, no. 4, pp. 1120–1142, 2002.

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

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

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

[25] R. Gray, "Toeplitz and circulant matrices: A review," Inform. Sys. Lab., Stanford Univ., Stanford, CA, Tech. Rep. 6504-1, 1977.

[26] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. Philadelphia, PA: SIAM, 1994.

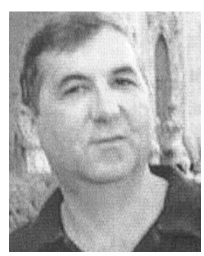
[27] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Linear Minimax regret estimation of deterministic parameters with bounded data uncertainties," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2177–2188, Aug. 2004.

[28] Z. Ben-Haim and Y. C. Eldar, "Maximum set estimators with bounded estimation error," *IEEE Trans. Signal Process.*, to be published.

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.

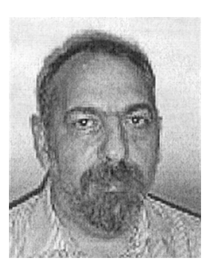
From January 2002 to July 2002, she was a Postdoctoral fellow at the Digital Signal Processing Group, MIT. She is currently a Senior Lecturer with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa, Israel. She is also a Research Affiliate with the Research Laboratory 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 and an Alon Fellow.



Aharon Ben-Tal received the Ph.D. degree in applied mathematics from Northwestern University, Evanston, 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, Denmark; 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 (continuous optimization) for *Mathematics of Operations Research* from 1993 to 1999, a member of Editorial Boards of the *SIAM Journal on Optimization* and *Mathematical Modeling and Numerical Analysis*, and currently is 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, Russia, in 1973.

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

Prof. 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.