Minimax Estimation of Deterministic Parameters in Linear Models With a Random Model Matrix

Yonina C. Eldar, *Member, IEEE*

Abstract—We consider the problem of estimating an unknown deterministic parameter vector in a linear model with a random model matrix, with known second-order statistics. We first seek the linear estimator that minimizes the worst-case mean-squared error (MSE) across all parameter vectors whose (possibly weighted) norm is bounded above. We show that the minimax MSE estimator can be found by solving a semidefinite programming problem and develop necessary and sufficient optimality conditions on the minimax MSE estimator. Using these conditions, we derive closed-form expressions for the minimax MSE estimator in some special cases. We then demonstrate, through examples, that the minimax MSE estimator can improve the performance over both a Baysian approach and a least-squares method. We then consider the case in which the norm of the parameter vector is also bounded below. Since the minimax MSE approach cannot account for a nonzero lower bound, we consider, in this case, a minimax regret method in which we seek the estimator that minimizes the worst-case difference between the MSE attainable using a linear estimator that does not know the parameter vector, and the optimal MSE attained using a linear estimator that knows the parameter vector. For analytical tractability, we restrict our attention to the scalar case and develop a closed-form expression for the minimax regret estimator.

Index Terms—Linear models, minimax mean-squared error (MSE), mean-squared error (MSE) estimation, random model matrix, regret.

I. INTRODUCTION

A GENERIC estimation problem that has received much attention in the estimation literature is that of estimating an unknown, deterministic vector parameter $\mathbf{x}$ in the linear model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where $\mathbf{H}$ is a linear transformation and $\mathbf{w}$ is a noise vector. The importance of this problem stems from the fact that a wide range of problems in communications, array processing, and many other areas of signal processing and statistics can be cast in this form.

In the simplest case, it is assumed that the model matrix $\mathbf{H}$ is completely specified, and the goal is to estimate $\mathbf{x}$ from $\mathbf{y}$ using a linear estimator. A variety of different linear estimators have been proposed for this problem, including the celebrated least-squares estimator, Tikhonov regularization [1], the linear shrunken estimator [2], and the covariance shaping least-squares

estimator [3]. In many cases of practical interest, the model matrix $\mathbf{H}$ may not be known precisely. If $\mathbf{H}$ deviates from the assumed matrix, then the performance of the estimator based on the assumed transformation can deteriorate considerably. Various methods have been proposed to account for deterministic uncertainties in $\mathbf{H}$, including the well-known total least-squares method [4], [5]. Recently, several methods have been developed to treat the case in which the perturbation to the model matrix $\mathbf{H}$ is bounded [6], [7].

Each of the estimators above can be formulated as a solution to an optimization problem that depends on the data error $\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}$, where $\hat{\mathbf{x}}$ is an estimate of $\mathbf{x}$, and not directly on the estimation error, or the mean-squared error (MSE). In an estimation context, we typically would like to minimize the *estimation error* $\hat{\mathbf{x}} - \mathbf{x}$, rather than the data error. However, since no prior statistics on $\mathbf{x}$ are available, the estimation error will generally depend on the unknown vector $\mathbf{x}$, and therefore cannot be minimized. In this case, it is desirable to design a robust estimator, whose performance is reasonably good across all possible values of $\mathbf{x}$, where in this paper we assume that $\mathbf{x}$ is known to satisfy a weighted norm constraint. The most common approach for designing robust estimators is the minimax MSE approach, in which we seek the estimator that minimizes the worst-case MSE in the region of uncertainty. This approach has been applied both to the case in which $\mathbf{x}$ is deterministic [8], [9] and to the case in which $\mathbf{x}$ is assumed to be random, but the statistics of $\mathbf{x}$ are not completely specified [10]–[14]. It has also been applied to the case in which the model matrix $\mathbf{H}$ is subject to bounded perturbations [9] and to the case of multichannel estimation [15]. In [16], an alternative approach is suggested for the case in which $\mathbf{H}$ is known, in which the vector parameter is chosen to minimize the worst-case regret, which is the difference between the MSE of the linear estimator which does not know $\mathbf{x}$, and the smallest attainable MSE with a linear estimator that knows $\mathbf{x}$. The minimax regret concept has also been applied to the case in which $\mathbf{x}$ is random with an unknown covariance matrix [14], [17], [18].

In this paper we consider the case in which the model matrix $\mathbf{H}$ is not known exactly, where the uncertainties are assumed to be random, so that we model $\mathbf{H}$ as a random matrix with known second-order statistics. One example in which such a model arises is in multiple-input multiple-output (MIMO) communication channels. In this example, $\mathbf{x}$ represents the transmitted data, and $\mathbf{H}$ represents the MIMO channel matrix, which is often modeled as a random matrix with known second-order statistics (see, e.g., [19]). Another example is beamforming in which $\mathbf{x}$ is a scalar representing a desired signal amplitude, and $\mathbf{H}$ is a vector that represents the steering vector, which, in many

practical scenarios, is modeled as a random vector with known second-order statistics (see, e.g., [20] and references therein). Other examples include channel equalization, in which $\mathbf{H}$ represents a linear-time invariant filter, which is estimated from training data.

Following the popular minimax approach, in Section II, we seek the linear estimator that minimizes the worst-case MSE over all possible values of $\mathbf{x}$ that satisfy a weighted norm constraint of the form $\mathbf{x}^*\mathbf{T}\mathbf{x} \leq U^2$ for some positive definite matrix $\mathbf{T}$ and constant U. Here the MSE is computed by averaging over the noise *and* the model matrix. We first develop an explicit expression for the minimax MSE estimator in the case in which $\mathbf{x} = x$ is a scalar. We then show that in the general vector case, the minimax MSE estimator can be found by solving a semidefinite programming problem (SDP) [21]–[23], which is the problem of minimizing a linear functional subject to linear matrix inequalities (LMIs), i.e., matrix inequalities in which the matrices depend *linearly* on the unknowns. The main advantage of the SDP formulation is that it readily lends itself to efficient computational methods. Specifically, by exploiting the many well-known algorithms for solving SDPs [22], [21], e.g., interior point methods [23], which are guaranteed to converge to the global optimum, the optimal estimator can be computed very efficiently in polynomial time.

In Section IV, we develop necessary and sufficient conditions for optimality on the minimax MSE estimator. Although these conditions are difficult to solve in general, they can be used to verify a solution and lead to further insight into the optimal estimator. Furthermore, using these conditions, we derive a closed-form expression for the minimax MSE estimator in the case in which the noise vector is white, and the elements of $\mathbf{H}$ are uncorrelated. We also consider conditions under which the minimax MSE estimator is equal to $\mathbf{0}$.

In Section V, we propose two additional (naive) estimators for our problem and compare their performance with that of the minimax MSE estimator. To eliminate the dependence of the MSE on $\mathbf{x}$, we first assume that $\mathbf{x}$ is random with covariance $(U^2/m)\mathbf{T}^{-1}$, where m is the length of x, and design a linear estimator that minimizes the MSE. We then consider a least-squares approach in which the estimator is designed to minimize the expected data error, where the expectation is taken with respect to the model matrix. We demonstrate through several examples that the minimax MSE approach often performs better than both of these estimators.

We next treat the case in which there is both an upper and lower bound on the norm of $\mathbf{x}$, so that $L^2 \leq \mathbf{x}^*\mathbf{T}\mathbf{x} \leq U^2$ for some $0 \leq L \leq U$. The minimax MSE estimator in this case is the same for all values of L, and therefore this approach cannot account for a lower bound. To derive an estimator that depends both on L and on U, following the approach taken in [16], [14], we seek the minimax regret estimator that minimizes the worst-case difference between the MSE attainable with a linear estimator that does not know $\mathbf{x}$ and the minimal attainable MSE when $\mathbf{x}$ is assumed to be known. Note that, as shown in Section 6, 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. Developing the minimax regret estimator turns out to be a difficult mathematical problem. Therefore, we

restrict our attention to the case in which $\mathbf{x} = x$ is an unknown scalar and develop a closed-form expression for the scalar minimax regret estimator.

In the sequel, we denote vectors in $\mathcal{C}^m$ by boldface lower case letters and matrices in $\mathcal{C}^{n \times m}$ by boldface upper case 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. The ith column of a matrix $\mathbf{A}$ is denoted by $[\mathbf{A}]_i$, the ith row is denoted by $\mathbf{a}_i$, and the ijth element is denoted by $[\mathbf{A}]_{ij}$. The ith component of a vector $\mathbf{z}$ is denoted by z_i. The inequality $\mathbf{X} \succeq 0$ means that the matrix $\mathbf{X}$ is positive semidefinite, and the notation $\mathbf{C} \preceq \mathbf{D}$ means that $\mathbf{D} - \mathbf{C} \succeq 0$.

II. MINIMAX MSE ESTIMATOR

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 $\mathbf{w}$ is a zero-mean random vector with positive definite covariance matrix $\mathbf{C}_w$, and $\mathbf{H}$ is a random matrix, independent of $\mathbf{w}$, with known first- and second-order statistics. Specifically

$$E\{\mathbf{H}\} = \mathbf{M}, \quad E\{(\mathbf{h}_i - \mathbf{m}_i)^*(\mathbf{h}_j - \mathbf{m}_j)\} = \mathbf{C}_{ij} \tag{2}$$

where $\mathbf{h}_i$ and $\mathbf{m}_i$ denote the ith rows of $\mathbf{H}$ and $\mathbf{M}$, respectively. We assume that $\mathbf{x}$ is known to satisfy the weighted norm constraint $\|\mathbf{x}\|_\mathbf{T} \leq U$ for some positive definite matrix $\mathbf{T}$ and scalar $U > 0$, 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 $\mathbf{G}$. We would like to design an estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ to minimize the MSE, which is given by

$$E\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\} =$$
$$\mathbf{x}^*E\{(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\}\mathbf{x} + \mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*). \tag{3}$$

To express the MSE in terms of $\mathbf{M}$ and $\mathbf{C}_{ij}$, we note that

$$E\{(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\}$$
$$= E\{(\mathbf{I} - \mathbf{G}\mathbf{M} - \mathbf{G}(\mathbf{H} - \mathbf{M}))^*(\mathbf{I} - \mathbf{G}\mathbf{M} - \mathbf{G}(\mathbf{H} - \mathbf{M}))\}$$
$$= (\mathbf{I} - \mathbf{G}\mathbf{M})^*(\mathbf{I} - \mathbf{G}\mathbf{M}) + E\{(\mathbf{H} - \mathbf{M})^*\mathbf{G}^*\mathbf{G}(\mathbf{H} - \mathbf{M})\}. \tag{4}$$

Now, for any Hermitian matrix $\mathbf{Z}$ and any matrix $\mathbf{D}$

$$\mathbf{D}\mathbf{Z}\mathbf{D}^* = \sum_i [\mathbf{D}]_i [\mathbf{D}\mathbf{Z}]_i^* = \sum_i [\mathbf{D}]_i \sum_j [\mathbf{Z}]_{ji}^* [\mathbf{D}]_j^*$$
$$= \sum_{ij} [\mathbf{Z}]_{ij} [\mathbf{D}]_i [\mathbf{D}]_j^*. \tag{5}$$

Using (5), we have that

$$E\{(\mathbf{H} - \mathbf{M})^*\mathbf{G}^*\mathbf{G}(\mathbf{H} - \mathbf{M})\}$$
$$= \sum_{ij} [\mathbf{G}^*\mathbf{G}]_{ij} E\{(\mathbf{h}_i - \mathbf{m}_i)^*(\mathbf{h}_j - \mathbf{m}_j)\}$$
$$= \sum_{ij} [\mathbf{G}^*\mathbf{G}]_{ij} \mathbf{C}_{ij}. \tag{6}$$

Substituting (4) and (6) into (3), the MSE of $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ is given by

$$E\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\} = \mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*)$$
$$+ \mathbf{x}^*\left((\mathbf{I} - \mathbf{G}\mathbf{M})^*(\mathbf{I} - \mathbf{G}\mathbf{M}) + \sum_{ij}[\mathbf{G}^*\mathbf{G}]_{ij}\mathbf{C}_{ij}\right)\mathbf{x}. \quad (7)$$

Since the MSE (7) depends explicitly on the unknown vector $\mathbf{x}$, we cannot choose an estimate to directly minimize the MSE. Instead, we seek the estimator that minimizes the worst-case MSE over all possible choices of $\mathbf{x}$ such that $\|\mathbf{x}\|_{\mathbf{T}} \leq U$. Thus, we seek the matrix $\mathbf{G}$ that is the solution to the problem

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\|\mathbf{x}\|_{\mathbf{T}} \leq U} E\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\}$$
$$= \min_{\mathbf{G}}\{\mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \mathcal{D}(\mathbf{G})\} \quad (8)$$

where

$$\mathcal{D}(\mathbf{G}) =$$
$$\max_{\|\mathbf{x}\|_{\mathbf{T}} \leq U} \mathbf{x}^*\left((\mathbf{I} - \mathbf{G}\mathbf{M})^*(\mathbf{I} - \mathbf{G}\mathbf{M}) + \sum_{ij}[\mathbf{G}^*\mathbf{G}]_{ij}\mathbf{C}_{ij}\right)\mathbf{x}. \quad (9)$$

To develop the solution to (8) we first determine $\mathcal{D}(\mathbf{G})$. To this end we note that for any nonnegative definite matrix $\mathbf{Q}$

$$\max_{\mathbf{x}^*\mathbf{T}\mathbf{x}\leq U^2} \mathbf{x}^*\mathbf{Q}\mathbf{x} = \max_{\mathbf{x}^*\mathbf{x}\leq U^2} \mathbf{x}^*\mathbf{T}^{-1/2}\mathbf{Q}\mathbf{T}^{-1/2}\mathbf{x}$$
$$= U^2\lambda_{\max}(\mathbf{T}^{-1/2}\mathbf{Q}\mathbf{T}^{-1/2}) \quad (10)$$

where $\lambda_{\max}(\mathbf{B})$ denotes the largest eigenvalue of the matrix $\mathbf{B}$. Furthermore, $\lambda_{\max}(\mathbf{T}^{-1/2}\mathbf{Q}\mathbf{T}^{-1/2})$ can be expressed as the solution to the problem

$$\min \lambda \quad (11)$$

subject to

$$\mathbf{T}^{-1/2}\mathbf{Q}\mathbf{T}^{-1/2} \preceq \lambda\mathbf{I} \quad (12)$$

or, equivalently, subject to

$$\mathbf{Q} \preceq \lambda\mathbf{T}. \quad (13)$$

Since for any choice of $\mathbf{G}$

$$\mathbf{A} \triangleq (\mathbf{I} - \mathbf{G}\mathbf{M})^*(\mathbf{I} - \mathbf{G}\mathbf{M}) + \sum_{ij}[\mathbf{G}^*\mathbf{G}]_{ij}\mathbf{C}_{ij}$$
$$= E\{(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\} \succeq 0 \quad (14)$$

we have that

$$\mathcal{D}(\mathbf{G}) = U^2\lambda_{\max}(\mathbf{T}^{-1/2}\mathbf{A}\mathbf{T}^{-1/2}) \quad (15)$$

and $\lambda_{\max}(\mathbf{T}^{-1/2}\mathbf{A}\mathbf{T}^{-1/2})$ can be expressed as the solution to the problem

$$\min \lambda \quad (16)$$

subject to

$$(\mathbf{I} - \mathbf{G}\mathbf{M})^*(\mathbf{I} - \mathbf{G}\mathbf{M}) + \sum_{ij}[\mathbf{G}^*\mathbf{G}]_{ij}\mathbf{C}_{ij} \preceq \lambda\mathbf{T}. \quad (17)$$

Thus, the problem (8) reduces to

$$\min_{\mathbf{G},\lambda}\{\mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + U^2\lambda\} \quad (18)$$

subject to

$$(\mathbf{I} - \mathbf{G}\mathbf{M})^*(\mathbf{I} - \mathbf{G}\mathbf{M}) + \sum_{ij}[\mathbf{G}^*\mathbf{G}]_{ij}\mathbf{C}_{ij} \preceq \lambda\mathbf{T}. \quad (19)$$

Since the objective function in (18) is strictly convex and the constraint set defined by (19) is convex, there is a unique minimizer $\mathbf{G}$.

Although the problem of (18) and (19) is convex, in the general case, it is difficult to find an explicit solution to the optimal matrix $\mathbf{G}$. In the next section, we consider the case in which $\mathbf{x} = x$ is a scalar and develop a closed-from solution for the minimax MSE estimator. To develop fast numerical algorithms for finding the minimax MSE estimator for the vector case, in Section III, we formulate the problem of (18) and (19) as an SDP. In Section IV, we develop necessary and sufficient optimality conditions on the matrix $\mathbf{G}$ and consider some special cases in which a closed-form solution exists to the minimax MSE estimator.

A. Scalar Minimax Estimator

We now develop the optimal estimator when $\mathbf{x} = x$ is a scalar and $\mathbf{H} = \mathbf{h}$ is a length-n random vector with mean $E\{\mathbf{h}\} = \mathbf{m}$ and covariance $E\{(\mathbf{h} - \mathbf{m})(\mathbf{h} - \mathbf{m})^*\} = \mathbf{C}$.

In this case, the estimator $\hat{x}$ has the form $\hat{x} = \mathbf{g}^*\mathbf{y}$ for some length-n vector $\mathbf{g}$, and from (7) the MSE of the estimator is

$$E\{\|\hat{x} - x\|^2\} = |x|^2(|1 - \mathbf{g}^*\mathbf{m}|^2 + \mathbf{g}^*\mathbf{C}\mathbf{g}) + \mathbf{g}^*\mathbf{C}_w\mathbf{g}. \quad (20)$$

Using (18) and (19), our problem then is to choose $\mathbf{g}$ as the solution to

$$\min_{\mathbf{g}}\{U^2(|1 - \mathbf{g}^*\mathbf{m}|^2 + \mathbf{g}^*\mathbf{C}\mathbf{g}) + \mathbf{g}^*\mathbf{C}_w\mathbf{g}\}. \quad (21)$$

Differentiating (21) with respect to $\mathbf{g}$ and equating to zero

$$\mathbf{g} = U^2(\mathbf{C}_w + U^2\mathbf{C} + U^2\mathbf{m}\mathbf{m}^*)^{-1}\mathbf{m}. \quad (22)$$

Using the Matrix Inversion Lemma [24], (22) reduces to

$$\mathbf{g} = \frac{U^2}{1 + U^2\mathbf{m}^*(\mathbf{C}_w + U^2\mathbf{C})^{-1}\mathbf{m}}(\mathbf{C}_w + U^2\mathbf{C})^{-1}\mathbf{m}. \quad (23)$$

If the covariance matrix of $\mathbf{h}$ is equal to zero, so that $\mathbf{h} = \mathbf{m}$ (with probability 1), then (23) becomes

$$\mathbf{g} = \frac{U^2}{1 + U^2\mathbf{m}^*\mathbf{C}_w^{-1}\mathbf{m}}\mathbf{C}_w^{-1}\mathbf{m}. \quad (24)$$

The estimator (23) is equivalent to the minimax MSE estimator developed in [9] for the problem of estimating the scalar x in the model $\mathbf{y} = x\mathbf{m} + \mathbf{w}$ with $|x| \leq U$.

III. SDP Formulation of the Minimax MSE Estimator

We now formulate the minimax MSE estimator as an SDP, which is the problem of minimizing a linear functional subject to LMIs. The advantage of the SDP formulation is that it readily lends itself to fast numerical algorithms that are guaranteed to converge to the global optimum in polynomial time within any desired accuracy. In particular, the solution of an SDP can be computed in Matlab using standard software packages, such as the Self-Dual-Minimization (SeDuMi) package [25], [26].

To formulate the problem of (18) and (19) as an SDP, we first note that this problem can be written as

$$\min_{\mathbf{G}, \lambda, \mathbf{X}} \{\mathrm{Tr}(\mathbf{X}) + U^2 \lambda\} \tag{25}$$

subject to

$$\mathbf{G}\mathbf{C}_w\mathbf{G}^* \preceq \mathbf{X}; \tag{26}$$
$$(\mathbf{I} - \mathbf{G}\mathbf{M})^*(\mathbf{I} - \mathbf{G}\mathbf{M}) + \sum_{ij}[\mathbf{G}^*\mathbf{G}]_{ij}\mathbf{C}_{ij} \preceq \lambda\mathbf{T}. \tag{27}$$

The objective in (25) is linear in the unknowns. Thus, it remains to show that (26) and (27) can be formulated as LMIs.

The constraint (26) can be converted into an LMI by exploiting the following lemma [24, p. 472]:

Lemma 1 (Schur's Complement): Let

$$\mathbf{Z} = \begin{bmatrix} \mathbf{P} & \mathbf{S}^* \\ \mathbf{S} & \mathbf{Q} \end{bmatrix}$$

be a Hermitian matrix. Then with $\mathbf{Q} \succ 0$, $\mathbf{Z} \succeq 0$ if and only if $\mathbf{P} - \mathbf{S}^*\mathbf{Q}^{-1}\mathbf{S} \succeq 0$.

Using Lemma 1, (26) is equivalent to the LMI

$$\begin{bmatrix} \mathbf{X} & \mathbf{G}^* \\ \mathbf{G} & \mathbf{C}_w^{-1} \end{bmatrix} \succeq 0. \tag{28}$$

To express (27) as an LMI, we note that this constraint can equivalently be written as

$$(\mathbf{I} - \mathbf{G}\mathbf{M})^*(\mathbf{I} - \mathbf{G}\mathbf{M}) + \sum_{ij}[\mathbf{B}]_{ij}\mathbf{C}_{ij} \preceq \lambda\mathbf{T} \tag{29}$$

where $\mathbf{B}$ is chosen such that

$$\sum_{ij}[\mathbf{G}^*\mathbf{G}]_{ij}\mathbf{C}_{ij} \preceq \sum_{ij}[\mathbf{B}]_{ij}\mathbf{C}_{ij}. \tag{30}$$

To develop a constraint on $\mathbf{B}$ such that (30) is satisfied, we rely on the following lemma.

Lemma 2: Let $\mathbf{D}$ and $\mathbf{B}$ be arbitrary matrices such that $\mathbf{D} \preceq \mathbf{B}$, and let $\mathbf{C}_{ij} = E\{\mathbf{h}_i^*\mathbf{h}_j\}$. Then, $\sum_{ij}[\mathbf{D}]_{ij}\mathbf{C}_{ij} \preceq \sum_{ij}[\mathbf{B}]_{ij}\mathbf{C}_{ij}$.

Proof: To show that $\sum_{ij}[\mathbf{D}]_{ij}\mathbf{C}_{ij} \preceq \sum_{ij}[\mathbf{B}]_{ij}\mathbf{C}_{ij}$ we must show that for any vector $\mathbf{y}$

$$\sum_{ij}([\mathbf{B}]_{ij} - [\mathbf{D}]_{ij})\mathbf{y}^*\mathbf{C}_{ij}\mathbf{y} \geq 0. \tag{31}$$

Denote by Γ the matrix with ijth element $[\Gamma]_{ij} = \mathbf{y}^*\mathbf{C}_{ji}\mathbf{y}$. Then

$$\sum_{ij}([\mathbf{B}]_{ij} - [\mathbf{D}]_{ij})\mathbf{y}^*\mathbf{C}_{ij}\mathbf{y} = \sum_{ij}([\mathbf{B}]_{ij} - [\mathbf{D}]_{ij})[\Gamma]_{ji}$$
$$= \mathrm{Tr}((\mathbf{B} - \mathbf{D})\Gamma). \tag{32}$$

Now, we can immediately show that $\Gamma \succeq 0$. Indeed, for any $\mathbf{z}$

$$\mathbf{z}^*\Gamma\mathbf{z} = \sum_{ij} z_i^*[\Gamma]_{ij}z_j = \sum_{ij} z_j\mathbf{y}^* E\{\mathbf{h}_j^*\mathbf{h}_i\}\mathbf{y}z_i^*$$
$$= E\left\{\left|\sum_i t_i\right|^2\right\} \geq 0 \tag{33}$$

where $t_i = \mathbf{h}_i\mathbf{y}z_i^*$. By assumption we also have that $\mathbf{B} - \mathbf{D} \succeq 0$. Thus, $\mathrm{Tr}((\mathbf{B} - \mathbf{D})\Gamma) \geq 0$, completing the proof of the lemma. ∎

It follows from Lemma 2 that if $\mathbf{G}^*\mathbf{G} \preceq \mathbf{B}$, then (30) is satisfied. Therefore, the constraint (27) can be replaced by the constraints

$$(\mathbf{I} - \mathbf{G}\mathbf{M})^*(\mathbf{I} - \mathbf{G}\mathbf{M}) + \sum_{ij}[\mathbf{B}]_{ij}\mathbf{C}_{ij} \preceq \lambda\mathbf{T}$$
$$\mathbf{G}^*\mathbf{G} \preceq \mathbf{B} \tag{34}$$

which from Lemma 1 are equivalent to the LMIs

$$\begin{bmatrix} \lambda\mathbf{T} - \sum_{ij}[\mathbf{B}]_{ij}\mathbf{C}_{ij} & (\mathbf{I} - \mathbf{G}\mathbf{M})^* \\ \mathbf{I} - \mathbf{G}\mathbf{M} & \mathbf{I} \end{bmatrix} \succeq 0$$
$$\begin{bmatrix} \mathbf{B} & \mathbf{G}^* \\ \mathbf{G} & \mathbf{I} \end{bmatrix} \succeq 0. \tag{35}$$

To conclude, the minimax MSE estimator can be found as the solution to the SDP

$$\min_{\mathbf{X}, \lambda, \mathbf{G}, \mathbf{B}} \{\mathrm{Tr}(\mathbf{X}) + U^2\lambda\} \tag{36}$$

subject to

$$\begin{bmatrix} \mathbf{X} & \mathbf{G}^* \\ \mathbf{G} & \mathbf{C}_w^{-1} \end{bmatrix} \succeq 0$$
$$\begin{bmatrix} \lambda\mathbf{T} - \sum_{ij}[\mathbf{B}]_{ij}\mathbf{C}_{ij} & (\mathbf{I} - \mathbf{G}\mathbf{M})^* \\ \mathbf{I} - \mathbf{G}\mathbf{M} & \mathbf{I} \end{bmatrix} \succeq 0$$
$$\begin{bmatrix} \mathbf{B} & \mathbf{G}^* \\ \mathbf{G} & \mathbf{I} \end{bmatrix} \succeq 0. \tag{37}$$

IV. Optimality Conditions

We now use Lagrange duality theory to develop necessary and sufficient conditions on the minimax MSE estimator [27]. Although these conditions are difficult to solve in general, they can be used to verify a solution and, as we show, can lead to further insight in the optimal estimator. Furthermore, in certain

special cases, they can be used to develop a closed-form solution for the minimax MSE estimator.

In Section IV-A, we develop necessary and sufficient conditions on the minimax MSE matrix $\mathbf{G}$. Then, in Section IV-B, we use these conditions to develop closed-form expressions for the minimax MSE estimator in two special cases. In Section IV-B-1), we consider the case in which the noise is white and the elements of $\mathbf{H}$ are uncorrelated, and in Section IV-B-2), we consider the case in which $\mathbf{M}$ is rank deficient.

A. Necessary and Sufficient Conditions

In the previous section, we have shown that the problem of (8) can be expressed as

$$\min_{\mathbf{G}, \lambda}\{\mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + U^2\lambda\} \qquad (38)$$

subject to

$$(\mathbf{I} - \mathbf{G}\mathbf{M})^*(\mathbf{I} - \mathbf{G}\mathbf{M}) + \sum_{ij}[\mathbf{G}^*\mathbf{G}]_{ij}\mathbf{C}_{ij} \preceq \lambda\mathbf{T}. \qquad (39)$$

We have seen already that the optimal matrix $\mathbf{G}$ is unique. Since the objective (38) and the constraint set defined by (39) are convex, the unique solution to (38) and (39) can be found by forming the Lagrangian

$$\begin{aligned}\mathcal{L} = \mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + U^2\lambda \\ + \mathrm{Tr}\Big(\Pi\big((\mathbf{I}-\mathbf{G}\mathbf{M})^*(\mathbf{I}-\mathbf{G}\mathbf{M}) + \sum_{ij}[\mathbf{G}^*\mathbf{G}]_{ij}\mathbf{C}_{ij}-\lambda\mathbf{T}\big)\Big)\end{aligned}$$
$$(40)$$

where from the Karush–Kuhn–Tucker (KKT) conditions [27], $\Pi \succeq 0$. Differentiating $\mathcal{L}$ with respect to λ and equating to 0

$$\mathrm{Tr}(\Pi\mathbf{T}) = U^2. \qquad (41)$$

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

$$\mathbf{G} = \Pi\mathbf{M}^*(\mathbf{C}_w + \mathbf{X} + \mathbf{M}\Pi\mathbf{M}^*)^{-1} \qquad (42)$$

where $[\mathbf{X}]_{ij} = \mathrm{Tr}(\Pi\mathbf{C}_{ji})$. Here, we used the fact that

$$\mathrm{Tr}\Big(\sum_{ij}[\mathbf{G}^*\mathbf{G}]_{ij}\Pi\mathbf{C}_{ij}\Big) = \sum_{ij}[\mathbf{G}^*\mathbf{G}]_{ij}[\mathbf{X}]_{ji} = \mathrm{Tr}(\mathbf{G}^*\mathbf{G}\mathbf{X}). \qquad (43)$$

Note that we can alternatively express $\mathbf{X}$ as

$$\mathbf{X} = \sum_{ij}[\Pi]_{ij}\mathbf{K}_{ij} \qquad (44)$$

where $\mathbf{K}_{ij} = E\{[\mathbf{H} - \mathbf{M}]_i[\mathbf{H} - \mathbf{M}]_j^*\}$, so that $\mathbf{K}_{ij}(\ell, s) = \mathbf{C}_{s\ell}(j, i)$.

It follows that $\mathbf{G}$ is optimal if and only if it is given by (42), where $\Pi \succeq 0$ is any matrix satisfying $\mathrm{Tr}(\Pi\mathbf{T}) = U^2$ and such

that there exists a λ that together with $\mathbf{G}$ satisfies (39), and the complementary slackness condition

$$\mathrm{Tr}\Big(\Pi\big((\mathbf{I} - \mathbf{G}\mathbf{M})^*(\mathbf{I} - \mathbf{G}\mathbf{M}) + \sum_{ij}[\mathbf{G}^*\mathbf{G}]_{ij}\mathbf{C}_{ij} - \lambda\mathbf{T}\big)\Big)$$
$$= 0. \qquad (45)$$

A sufficient condition for $\mathbf{G}$ to be optimal is that it is given by (42), where $\Pi \succeq 0$ is any matrix satisfying $\mathrm{Tr}(\Pi\mathbf{T}) = U^2$ and such that there exists a λ that together with $\mathbf{G}$ satisfies

$$(\mathbf{I} - \mathbf{G}\mathbf{M})^*(\mathbf{I} - \mathbf{G}\mathbf{M}) + \sum_{ij}[\mathbf{G}^*\mathbf{G}]_{ij}\mathbf{C}_{ij} = \lambda\mathbf{T}. \qquad (46)$$

Indeed, if $\mathbf{G}$ and λ satisfy (46), then clearly they satisfy (39) and (45), and therefore, from the necessary and sufficient conditions above, $\mathbf{G}$ is optimal.

B. Closed-Form Solution for Special Cases

We now consider two special cases in which a closed-form solution can be obtained, using the necessary and sufficient conditions derived in the previous section.

1) White Noise and Uncorrelated $\mathbf{H}$ Elements: We first treat the case in which $\mathbf{T} = \mathbf{I}$, the noise $\mathbf{w}$ is white so that $\mathbf{C}_w = \sigma_w^2\mathbf{I}$ for some $\sigma_w^2 > 0$, and the elements of $\mathbf{H}$ are uncorrelated, with variance $\sigma_h^2 \geq 0$ so that $\mathbf{C}_{ij} = \sigma_h^2\delta_{ij}\mathbf{I}$. We further assume that $\mathbf{M}$ has full column rank.

Under these assumptions

$$\sum_{ij}[\mathbf{G}^*\mathbf{G}]_{ij}\mathbf{C}_{ij} = \sigma_h^2\sum_i[\mathbf{G}^*\mathbf{G}]_{ii}\mathbf{I} = \sigma_h^2\mathrm{Tr}(\mathbf{G}^*\mathbf{G})\mathbf{I} \qquad (47)$$

and from (44)

$$\mathbf{X} = \sum_{ij}[\Pi]_{ij}\mathbf{K}_{ij} = \sigma_h^2\mathrm{Tr}(\Pi)\mathbf{I} = \sigma_h^2 U^2\mathbf{I}. \qquad (48)$$

Substituting (47) and (48) into (42), $\mathbf{G}$ is given by

$$\mathbf{G} = \Pi\mathbf{M}^*(\alpha\mathbf{I} + \mathbf{M}\Pi\mathbf{M}^*)^{-1} \qquad (49)$$

where we defined

$$\alpha = \sigma_w^2 + \sigma_h^2 U^2. \qquad (50)$$

With $\mathbf{G}$ given by (49), and using the Matrix Inversion Lemma [24]

$$\begin{aligned}\mathbf{I} - \mathbf{G}\mathbf{M} &= \mathbf{I} - \Pi\mathbf{M}^*(\alpha\mathbf{I} + \mathbf{M}\Pi\mathbf{M}^*)^{-1}\mathbf{M} \\ &= (\mathbf{I} + (1/\alpha)\Pi\mathbf{M}^*\mathbf{M})^{-1}. \end{aligned} \qquad (51)$$

Using the sufficient conditions derived in the previous section, it then follows that if can choose Π and λ such that

$$\begin{aligned}(\mathbf{I} + (1/\alpha)\mathbf{M}^*\mathbf{M}\Pi)^{-1}(\mathbf{I} + (1/\alpha)\Pi\mathbf{M}^*\mathbf{M})^{-1} \\ + \sigma_h^2\mathrm{Tr}(\mathbf{M}\Pi^2\mathbf{M}^*(\alpha\mathbf{I} + \mathbf{M}\Pi\mathbf{M}^*)^{-2})\mathbf{I} = \lambda\mathbf{I} \end{aligned} \qquad (52)$$

and $\text{Tr}(\Pi) = U^2$, then $\mathbf{G}$ given by (49) with the corresponding choice of Π is minimax optimal.

It is easy to see that if Π is proportional to $(\mathbf{M}^*\mathbf{M})^{-1}$, then (52) will be satisfied with an appropriate choice of λ. Since we also must have that $\text{Tr}(\Pi) = U^2$, the optimal choice of Π is

$$\Pi = \frac{U^2}{\text{Tr}((\mathbf{M}^*\mathbf{M})^{-1})}(\mathbf{M}^*\mathbf{M})^{-1} \triangleq \beta(\mathbf{M}^*\mathbf{M})^{-1} \tag{53}$$

where

$$\beta = \frac{U^2}{\text{Tr}((\mathbf{M}^*\mathbf{M})^{-1})}. \tag{54}$$

With Π given by (53), $\mathbf{G}$ of (49) becomes

$$\mathbf{G} = \beta(\mathbf{M}^*\mathbf{M})^{-1}\mathbf{M}^*(\alpha\mathbf{I} + \beta\mathbf{P})^{-1} \tag{55}$$

where $\mathbf{P}$ is the orthogonal projection onto the range space of $\mathbf{M}$. Let $\mathbf{M}$ have a singular value decomposition (SVD)

$$\mathbf{M} = \mathbf{U}\begin{bmatrix} \Sigma \\ 0 \end{bmatrix}\mathbf{V}^* \tag{56}$$

where $\mathbf{U}$ is an $n \times n$ unitary matrix, Σ is an $m \times m$ diagonal matrix with positive diagonal elements, and $\mathbf{V}$ is an $m \times m$ unitary matrix. Substituting the SVD (56) into (55), and using the fact that

$$\mathbf{P} = \mathbf{U}\begin{bmatrix} \mathbf{I}_m & 0 \\ 0 & 0 \end{bmatrix}\mathbf{U}^* \tag{57}$$

$\mathbf{G}$ reduces to

$$\begin{aligned} \mathbf{G} &= \beta\mathbf{V}[\Sigma^{-1}\quad 0]\begin{bmatrix} (\alpha+\beta)\mathbf{I} & 0 \\ 0 & \alpha\mathbf{I} \end{bmatrix}^{-1}\mathbf{U}^* \\ &= \beta\mathbf{V}[((\alpha+\beta)\Sigma)^{-1}\quad 0]\mathbf{U}^* \\ &= \frac{\beta}{\alpha+\beta}(\mathbf{M}^*\mathbf{M})^{-1}\mathbf{M}^* \\ &= \frac{U^2}{U^2+(\sigma_w^2+\sigma_h^2 U^2)\,\text{Tr}((\mathbf{M}^*\mathbf{M})^{-1})}(\mathbf{M}^*\mathbf{M})^{-1}\mathbf{M}^*. \end{aligned} \tag{58}$$

Note that if $\sigma_h^2 = 0$, so that $\mathbf{H} = \mathbf{M}$ with probability 1, then $\mathbf{G}$ of (58) reduces to

$$\mathbf{G} = \frac{U^2}{U^2+\sigma_w^2\,\text{Tr}((\mathbf{M}^*\mathbf{M})^{-1})}(\mathbf{M}^*\mathbf{M})^{-1}\mathbf{M}^*. \tag{59}$$

The estimator (59) is equivalent to the minimax MSE estimator developed in [9], for the problem of estimating the vector $\mathbf{x}$ in the model $\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{w}$ with $\|\mathbf{x}\| \leq U$, where $\mathbf{w}$ is a white noise vector with covariance $\sigma_w^2\mathbf{I}$.

2) Rank-Deficient $\mathbf{M}$: From the necessary and sufficient conditions for optimality, it follows that the optimal $\mathbf{G}$ is $\mathbf{G} = \mathbf{0}$ only if $\mathbf{M}$ does not have full column rank, i.e., $\text{rank}(\mathbf{M}) < m$. Indeed, from (42) we have that $\mathbf{G} = \mathbf{0}$ only if

$$\mathbf{M}\Pi = \mathbf{0}. \tag{60}$$

Since $\text{Tr}(\Pi\mathbf{T}) = U^2$, we must have that $\Pi \neq \mathbf{0}$ so that (60) can be satisfied for some Π only if $\text{rank}(\mathbf{M}) < m$.

Conversely, suppose that $\text{rank}(\mathbf{M}) < m$. Then, the optimal choice of $\mathbf{G}$ is $\mathbf{G} = \mathbf{0}$ if there exists a vector $\mathbf{z}$ such that $\mathbf{M}\mathbf{z} = \mathbf{0}$, and $\mathbf{z}$ is an eigenvector of $\mathbf{T}$ with eigenvalue $\lambda = \lambda_{\min}(\mathbf{T})$. Indeed, in this case, let

$$\Pi = \frac{U^2}{\mathbf{z}^*\mathbf{T}\mathbf{z}}\mathbf{z}\mathbf{z}^* = \frac{U^2}{\lambda_{\min}(\mathbf{T})\mathbf{z}^*\mathbf{z}}\mathbf{z}\mathbf{z}^*. \tag{61}$$

We have immediately that $\mathbf{M}\Pi = \mathbf{0}, \Pi \succeq 0$ and $\text{Tr}(\Pi\mathbf{T}) = U^2$. In addition, if we choose $\lambda = 1/\lambda_{\min}(\mathbf{T})$, then both (39) and (45) are satisfied. In particular, if $\mathbf{T} = \mathbf{I}$, then the optimal $\mathbf{G}$ is $\mathbf{G} = \mathbf{0}$ if and only if $\text{rank}(\mathbf{M}) < m$.

C. Summary of the Minimax MSE Estimator

We summarize our results on the minimax MSE estimator in the following theorem.

Theorem 1 (Minimax MSE Estimator): Let $\mathbf{x}$ denote the unknown deterministic vector in the model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where $\mathbf{H}$ is an $n \times m$ random matrix with known mean $\mathbf{M}$ and covariance $E\{(\mathbf{h}_i - \mathbf{m}_i)(\mathbf{h}_j - \mathbf{m}_j)^*\} = \mathbf{C}_{ij}$ where $\mathbf{h}_i$ and $\mathbf{m}_i$ denote the ith rows of $\mathbf{H}$ and $\mathbf{M}$, respectively, and $\mathbf{w}$ is a length-n, zero-mean random vector with covariance $\mathbf{C}_w$, independent of $\mathbf{H}$. Then, the problem

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\|\mathbf{x}\|_{\mathbf{T}}\leq U} E\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\}$$

is equivalent to the semidefinite programming problem

$$\min_{\mathbf{X},\lambda,\mathbf{G},\mathbf{B}}\{\text{Tr}(\mathbf{X}) + U^2\lambda\}$$

subject to

$$\begin{bmatrix} \mathbf{X} & \mathbf{G}^* \\ \mathbf{G} & \mathbf{C}_w^{-1} \end{bmatrix} \succeq 0$$

$$\begin{bmatrix} \lambda\mathbf{T} - \sum_{ij}[\mathbf{B}]_{ij}\mathbf{C}_{ij} & (\mathbf{I}-\mathbf{G}\mathbf{M})^* \\ \mathbf{I} - \mathbf{G}\mathbf{M} & \mathbf{I} \end{bmatrix} \succeq 0$$

$$\begin{bmatrix} \mathbf{B} & \mathbf{G}^* \\ \mathbf{G} & \mathbf{I} \end{bmatrix} \succeq 0.$$

The matrix $\mathbf{G}$ is minimax optimal if and only if

$$\mathbf{G} = \Pi\mathbf{M}^*(\mathbf{C}_w + \mathbf{X} + \mathbf{M}\Pi\mathbf{M}^*)^{-1}$$

where $\mathbf{X} = \sum_{ij}[\Pi]_{ij}\mathbf{K}_{ij}$ with $\mathbf{K}_{ij} = E\{[\mathbf{H}-\mathbf{M}]_i[\mathbf{H}-\mathbf{M}]_j^*\}$, and $\Pi \succeq 0$ is any matrix satisfying $\text{Tr}(\Pi\mathbf{T}) = U^2$, and such that there exists a λ that together with $\mathbf{G}$ satisfy

$$(\mathbf{I}-\mathbf{G}\mathbf{M})^*(\mathbf{I}-\mathbf{G}\mathbf{M}) + \sum_{ij}[\mathbf{G}^*\mathbf{G}]_{ij}\mathbf{C}_{ij} \preceq \lambda\mathbf{T}$$

$$\text{Tr}\left(\Pi\left((\mathbf{I}-\mathbf{G}\mathbf{M})^*(\mathbf{I}-\mathbf{G}\mathbf{M}) + \sum_{ij}[\mathbf{G}^*\mathbf{G}]_{ij}\mathbf{C}_{ij} - \lambda\mathbf{T}\right)\right)$$
$$= 0.$$

In addition

1) if $\mathbf{C}_w = \sigma_w^2 \mathbf{I}$ and $\mathbf{C}_{ij} = \sigma_h^2 \delta_{ij} \mathbf{I}$, then

$$\hat{\mathbf{x}} = \frac{U^2}{U^2 + (\sigma_w^2 + \sigma_h^2 U^2) \operatorname{Tr}((\mathbf{M}^*\mathbf{M})^{-1})} (\mathbf{M}^*\mathbf{M})^{-1} \mathbf{M}^* \mathbf{y}$$

2) if $\mathbf{x} = x$ is a scalar and $\mathbf{H} = \mathbf{h}$ is a vector with mean $\mathbf{m}$ and covariance $\mathbf{C}$, then

$$\hat{x} = \frac{U^2}{1 + U^2 \mathbf{m}^*(\mathbf{C}_w + U^2\mathbf{C})^{-1}\mathbf{m}} \mathbf{m}^*(\mathbf{C}_w + U^2\mathbf{C})^{-1}\mathbf{y}$$

3) if there exists a vector $\mathbf{z}$ such that $\mathbf{Mz} = \mathbf{0}$, and $\mathbf{z}$ is an eigenvector of $\mathbf{T}$ with eigenvalue $\lambda = \lambda_{\min}(\mathbf{T})$, then $\hat{\mathbf{x}} = \mathbf{0}$;
4) if $\mathbf{T} = \mathbf{I}$, then $\hat{\mathbf{x}} = \mathbf{0}$ if and only if $\operatorname{rank}(\mathbf{M}) < m$.

V. ALTERNATIVE ESTIMATORS AND EXAMPLES

In Sections V-A and V-B, we propose two additional estimators for our problem. The first estimator is based on the assumption that $\mathbf{x}$ is random, while the second estimator is based on a least-squares criterion. In Section V- C, we compare the performance of the minimax MSE estimator with those developed in Sections V-A and V-B and show that in many cases the performance of the minimax estimator is superior.

A. Minimum MSE Estimator for Random $\mathbf{x}$

In designing optimal estimators, one possible approach is to minimize the MSE. However, as we have seen, in the case of a deterministic parameter vector $\mathbf{x}$, the MSE depends explicitly on $\mathbf{x}$ and therefore cannot be minimized. We therefore proposed seeking estimators that minimize a worst-case function of the MSE. Alternatively, we may assume that $\mathbf{x}$ is a random vector, independent of $\mathbf{H}$ and $\mathbf{w}$, with known second-order statistics, and then explicitly compute the MSE.

Suppose that $\mathbf{x}$ is a zero-mean random vector with covariance $\mathbf{C}_x = (U^2/m)\mathbf{T}^{-1}$. In this case, $\operatorname{Tr}(\mathbf{C}_x \mathbf{T}) = U^2$, which is the stochastic equivalent of the condition $\mathbf{x}^*\mathbf{Tx} = \operatorname{Tr}(\mathbf{xx}^*\mathbf{T}) \leq U^2$. With this choice of $\mathbf{C}_x$, the MSE is given by

$$\begin{aligned}
E\{&\|\hat{\mathbf{x}} - \mathbf{x}\|^2\} \\
&= \operatorname{Tr}(E\{(\mathbf{I} - \mathbf{GH})^*(\mathbf{I} - \mathbf{GH})\}\mathbf{C}_x) + \operatorname{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) \\
&= \frac{U^2}{m}\operatorname{Tr}\left(\mathbf{T}^{-1}\left((\mathbf{I} - \mathbf{GM})^*(\mathbf{I} - \mathbf{GM})\right.\right. \\
&\qquad\qquad \left.\left. + \sum_{i,j}[\mathbf{G}^*\mathbf{G}]_{ij}\mathbf{C}_{ij}\right)\right) + \operatorname{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*).
\end{aligned}$$

$$\text{(62)}$$

Differentiating (62) with respect to $\mathbf{G}$ and equating to 0, the optimal $\mathbf{G}$ is

$$\mathbf{G} = \mathbf{T}^{-1}\mathbf{M}^*\left(\frac{m}{U^2}\mathbf{C}_w + \sum_{ij}[\mathbf{T}^{-1}]_{ij}\mathbf{K}_{ij} + \mathbf{M}\mathbf{T}^{-1}\mathbf{M}^*\right)^{-1}$$

$$\text{(63)}$$

where $\mathbf{K}_{ij} = E\{[\mathbf{H} - \mathbf{M}]_i[\mathbf{H} - \mathbf{M}]_j^*\}$. For future reference, we refer to the estimator $\hat{\mathbf{x}} = \mathbf{Gy}$ with $\mathbf{G}$ given by (63) as a Wiener estimator.

Note that $\mathbf{G}$ of (63) has the same form as the minimax MSE estimator of (42), where the Lagrange multiplier Π in (42) is replaced with $(U^2/m)\mathbf{T}^{-1}$ in (63).

In the case in which $\mathbf{x} = x$ is a scalar, $\mathbf{T} = 1, m = 1$, and (63) reduces to the minimax estimator $\mathbf{g} = U^2(U^2(\mathbf{C} + \mathbf{mm}^*) + \mathbf{C}_w)^{-1}\mathbf{m}$. However, as we demonstrate in Section V-C, in higher dimensions the two estimators are different, and the minimax MSE estimator often performs better than the Wiener estimator for all values of $\mathbf{x}$.

B. Least-Squares Estimator

Another method for deriving optimal estimators is based on a least-squares approach. When the matrix $\mathbf{H}$ is known, the (weighted) least-squares estimator seeks the estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ that minimizes the error

$$\|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|_{\mathbf{C}_w^{-1}}^2 = (\mathbf{y} - \mathbf{H}\hat{\mathbf{x}})^*\mathbf{C}_w^{-1}(\mathbf{y} - \mathbf{Hx}). \quad \text{(64)}$$

Note, that in this approach, the vector $\mathbf{y}$ is treated as a known vector, since no expectation with respect to $\mathbf{w}$ is taken.

In our problem, $\mathbf{H}$ is random with known statistics. Generalizing the least-squares approach, we may seek the estimator $\hat{\mathbf{x}}$ that minimizes the error

$$E_\mathbf{H}\{\|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|_{\mathbf{C}_w^{-1}}^2\} = E\{(\mathbf{y} - \mathbf{H}\hat{\mathbf{x}})^*\mathbf{C}_w^{-1}(\mathbf{y} - \mathbf{Hx})\} \quad \text{(65)}$$

where the expectation is taken only with respect to $\mathbf{H}$ and not with respect to the noise $\mathbf{w}$. Expanding (65), we have

$$\begin{aligned}
E_\mathbf{H}\{\|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|_{\mathbf{C}_w^{-1}}^2\} &= (\mathbf{y} - \mathbf{M}\hat{\mathbf{x}})^*\mathbf{C}_w^{-1}(\mathbf{y} - \mathbf{M}\hat{\mathbf{x}}) \\
&\quad + \hat{\mathbf{x}}^* E\{(\mathbf{H} - \mathbf{M})^*\mathbf{C}_w^{-1}(\mathbf{H} - \mathbf{M})\}\hat{\mathbf{x}} \\
&= (\mathbf{y} - \mathbf{M}\hat{\mathbf{x}})^*\mathbf{C}_w^{-1}(\mathbf{y} - \mathbf{M}\hat{\mathbf{x}}) + \hat{\mathbf{x}}^* \sum_{ij}[\mathbf{C}_w^{-1}]_{ij}\mathbf{C}_{ij}\hat{\mathbf{x}}.
\end{aligned}$$

$$\text{(66)}$$

Differentiating (66) with respect to $\hat{\mathbf{x}}$ and equating to 0

$$\hat{\mathbf{x}} = \left(\sum_{ij}[\mathbf{C}_w^{-1}]_{ij}\mathbf{C}_{ij} + \mathbf{M}^*\mathbf{C}_w^{-1}\mathbf{M}\right)^{-1}\mathbf{M}^*\mathbf{C}_w^{-1}\mathbf{y}. \quad \text{(67)}$$

Note that if $\mathbf{C}_{ij} = \mathbf{0}$, then the estimator of (67) reduces to the conventional least-squares estimator

$$\hat{\mathbf{x}} = (\mathbf{M}^*\mathbf{C}_w^{-1}\mathbf{M})\mathbf{M}^*\mathbf{C}_w^{-1}\mathbf{y}. \quad \text{(68)}$$

C. Examples

We now present several examples illustrating the minimax MSE estimator of Theorem 1. The purpose of these examples is to demonstrate the performance advantage in using the minimax MSE estimator, rather than a detailed practical application, which is beyond the scope of the paper. An application to beamforming is developed in [28]–[30].
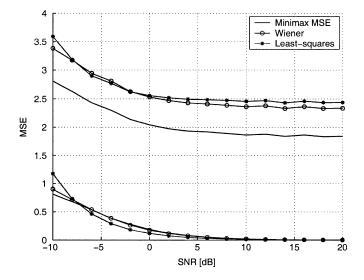
Fig. 1.   MSE in estimating $\mathbf{x}$ as a function of signal-to-noise ratio (SNR) using the minimax MSE, Wiener, and least-squares estimators for $\sigma_h^2 = 1$ and $U = 3$. The upper curves correspond to the worst-case $\mathbf{x}$ and the lower curves correspond to the best-case $\mathbf{x}$.

Fig. 2.   MSE in estimating $\mathbf{x}$ as a function of SNR using the minimax MSE, Wiener, and least-squares estimators for $\sigma_h^2 = 1$ and $U = 1$. The upper curves correspond to the worst-case $\mathbf{x}$ and the lower curves correspond to the best-case $\mathbf{x}$.

In the examples below, we assume that $\mathbf{T} = \mathbf{I}, \mathbf{C}_w = \sigma_w^2 \mathbf{I}$, and $\mathbf{C}_{ij} = \sigma_h^2 \delta_{ij} \mathbf{I}$. Under these assumptions, the minimax MSE estimator $\hat{\mathbf{x}}_{\text{MX}}$, the Wiener estimator $\hat{\mathbf{x}}_{\text{WN}}$, and the least-squares estimator $\hat{\mathbf{x}}_{\text{LS}}$ are given by

$$\hat{\mathbf{x}}_{\text{MX}} = \frac{U^2}{U^2 + (\sigma_w^2 + \sigma_h^2 U^2)\text{Tr}((\mathbf{M}^*\mathbf{M})^{-1})}(\mathbf{M}^*\mathbf{M})^{-1}\mathbf{M}^*\mathbf{y}$$

$$\hat{\mathbf{x}}_{\text{WN}} = \left(\mathbf{M}^*\mathbf{M} + \frac{m}{U^2}\left(\sigma_w^2 + \sigma_h^2 U^2\right)\mathbf{I}\right)^{-1}\mathbf{M}^*\mathbf{y}$$

$$\hat{\mathbf{x}}_{\text{LS}} = (\mathbf{M}^*\mathbf{M} + n\sigma_h^2\mathbf{I})^{-1}\mathbf{M}^*\mathbf{y}. \tag{69}$$

Clearly, the performance of each of the estimators will depend on the choice of $\mathbf{M}$ and the true parameter vector $\mathbf{x}$. In the simulations below, we compute the MSE averaged over random choices of the mean matrix $\mathbf{M}$. Specifically, we consider 1000 random choices of $\mathbf{M}$, where for each choice of $\mathbf{M}$ we compute the MSE by averaging the error $\|\hat{\mathbf{x}} - \mathbf{x}\|$ over 1000 noise realizations. The elements of $\mathbf{M}$ are chosen as zero-mean, independent Gaussian random variables with unit variance.

For each choice of $\mathbf{M}$ and estimator $\hat{\mathbf{x}}$, the value of the MSE will depend on the choice of $\mathbf{x}$. The best-case (lowest) MSE is attained when $\mathbf{x} = \mathbf{0}$. The worst-case (highest) MSE is attained when $\mathbf{x}$ is a vector with norm U in the direction of the eigenvector of $(\mathbf{I} - \mathbf{GH})^*(\mathbf{I} - \mathbf{GH})$ that corresponds to the largest eigenvalue. For any other choice of $\mathbf{x}$, the value of the MSE will lie between these two bounds. Therefore, to evaluate the performance of the estimators of (69), we compute the best-case and worst-case MSE of each of the estimators.

In Fig. 1, we plot the MSE of the minimax MSE, Wiener, and least-squares estimators as a function of the SNR defined by $-10 \log \sigma_w^2$, for $U = 3, n = 5, m = 3$, and $\sigma_h^2 = 1$. In the figure, the lower and upper curves correspond to the best-case MSE and the worst-case MSE, respectively, of each of the estimators.

In Fig. 2, we plot the MSE of the minimax MSE, Wiener, and least-squares estimators as a function of the SNR for $U = 1, n = 5, m = 3$ and $\sigma_h^2 = 1$.
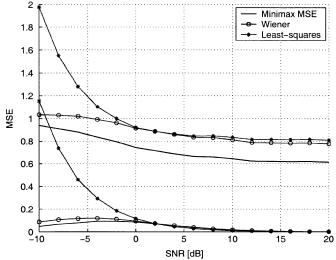
As we expect, for the worst-case choice of $\mathbf{x}$, the performance of the minimax MSE estimator is significantly better than that of the Wiener and least-squares estimators. The interesting observation is that even in the best case, we do not lose much by using the minimax MSE estimator, and in fact, in some cases, the minimax MSE estimator has the best performance even for the best-case choice of $\mathbf{x}$.

VI. MINIMAX REGRET ESTIMATOR

In the development of the minimax MSE estimator, we assumed that an upper bound on the (weighted) norm of $\mathbf{x}$ is available, so that $\mathbf{x}$ is known to satisfy the constraint $\|\mathbf{x}\|_{\mathbf{T}} \leq U$. In some applications, we may also have a lower bound on the norm of $\mathbf{x}$ so that $\mathbf{x}$ satisfies

$$L \leq \|\mathbf{x}\|_{\mathbf{T}} \leq U. \tag{70}$$

It is easy to see that the minimax MSE estimator that minimizes the worst-case MSE subject to (70) for any $L \leq U$ is the same as the minimax MSE estimator corresponding to $L = 0$. This follows from the fact that the worst-case $\mathbf{x}$ satisfies $\|\mathbf{x}\|_{\mathbf{T}} = U$. Therefore, the minimax MSE approach cannot take a lower bound into account. To develop a minimax estimator that depends both on L and on U, we may consider a minimax regret approach in which we seek the linear estimator that minimizes the worst-case regret [16], [14], [17], [18]. It turns out that finding the linear minimax regret estimator in the case of a random model matrix is a difficult problem. Therefore, in this section, we restrict our attention to the case in which $\mathbf{x} = x$ is a scalar, and $\mathbf{H} = \mathbf{h}$ is a length-n vector with mean $\mathbf{m}$ and covariance $\mathbf{C}$. In this case, the estimator has the form $\hat{x} = \mathbf{g}^*\mathbf{y}$ for some length-n vector $\mathbf{g}$.

Thus, we now design an estimator to minimize the worst-case regret $\mathcal{R}(x, \mathbf{g})$, which is defined as the difference between the MSE using an estimator $\hat{x} = \mathbf{g}^*\mathbf{y}$ and the smallest possible MSE attainable with an estimator of the form $\hat{x} = \mathbf{g}^*(x)\mathbf{y}$ when

the parameter x is known, which we denote by MSE^o. To develop an explicit expression for MSE^o, we first determine the estimator $\hat{x} = \mathbf{g}^*(x)\mathbf{y}$ that minimizes the MSE when x is known. To this end, we differentiate the MSE of (20) with respect to $\mathbf{g}$ and equate to 0, which results in

$$\mathbf{g}(x) = |x|^2(|x|^2(\mathbf{C} + \mathbf{mm}^*) + \mathbf{C}_w)^{-1}\mathbf{m}$$
$$= \frac{|x|^2}{1 + |x|^2\mathbf{m}^*(\mathbf{C}_w + |x|^2\mathbf{C})^{-1}\mathbf{m}}(\mathbf{C}_w + |x|^2\mathbf{C})^{-1}\mathbf{m}. \tag{71}$$

Note that $\mathbf{g}(U)$ is equal to the minimax estimator of (23). Thus, we can view the minimax estimator as an approximation of the optimal estimator, where the unknown norm $|x|^2$ is replaced by its upper bound U^2. Substituting $\mathbf{g}(x)$ back into (20), MSE^0 is given by

$$\mathrm{MSE}^o = |x|^2 - |x|^4\mathbf{m}^*(|x|^2\mathbf{R} + \mathbf{C}_w)^{-1}\mathbf{m} \tag{72}$$

where

$$\mathbf{R} = \mathbf{C} + \mathbf{mm}^*. \tag{73}$$

Since x is unknown, we cannot implement the optimal estimator (71). Instead, we seek the estimator $\hat{x} = \mathbf{g}^*\mathbf{y}$ that minimizes the worst-case regret $\mathcal{R}(x, \mathbf{g})$ subject to $L \leq |x| \leq U$, with

$$\mathcal{R}(x, \mathbf{g}) = E\{|\mathbf{g}^*\mathbf{y} - x|^2\} - \mathrm{MSE}^o = h(|x|^2) + \mathbf{g}^*\mathbf{C}_w\mathbf{g} \tag{74}$$

where we defined

$$h(s) = s(-\mathbf{g}^*\mathbf{m} - \mathbf{m}^*\mathbf{g} + \mathbf{g}^*\mathbf{R}\mathbf{g})$$
$$+ s^2\mathbf{m}^*(s\mathbf{R} + \mathbf{C}_w)^{-1}\mathbf{m}. \tag{75}$$

Thus, we seek the vector $\mathbf{g}$ that is the solution to

$$\min_{\mathbf{g}} \max_{L \leq |x| \leq U} \mathcal{R}(x, \mathbf{g}). \tag{76}$$

To solve (76), we first consider the inner maximization

$$\max_{L \leq |x| \leq U} h(|x|^2) = \max_{L^2 \leq z \leq U^2} h(z) \tag{77}$$

were we denoted $z = |x|^2$. Lemma 3 below asserts that $h(z)$ is convex in z, which immediately implies that the maximum value of $h(z)$ over the closed interval $[L^2, U^2]$ is attained at $z = L^2$ or $z = U^2$.

Lemma 3: Let $h(s)$ be defined as in (75) for some given $\mathbf{g}, \mathbf{m}, \mathbf{R} \succeq 0$ and $\mathbf{C}_w \succ 0$. Then, $h(s)$ is convex in s.

Proof: The first term in $h(s)$ is linear, and therefore convex, in s. Thus, to prove the lemma it remains to show that

$$\Phi(s) = s^2\mathbf{m}^*(s\mathbf{R} + \mathbf{C}_w)^{-1}\mathbf{m} \tag{78}$$

is convex in s. To this end, let $\mathbf{C}_w^{-1/2}\mathbf{R}\mathbf{C}_w^{-1/2}$ have an eigendecomposition

$$\mathbf{C}_w^{-1/2}\mathbf{R}\mathbf{C}_w^{-1/2} = \mathbf{U}\Lambda\mathbf{U}^* \tag{79}$$

where $\mathbf{U}$ is an $n \times n$ unitary matrix, and Λ is an $n \times n$ diagonal matrix with diagonal elements $\lambda_i \geq 0$. Using (79), we can express $\Phi(s)$ as

$$\Phi(s) = s^2\mathbf{m}^*\mathbf{C}_w^{-1/2}\left(\mathbf{I} + s\mathbf{C}_w^{-1/2}\mathbf{R}\mathbf{C}_w^{-1/2}\right)^{-1}\mathbf{C}_w^{-1/2}\mathbf{m}$$
$$= s^2\mathbf{z}^*(\mathbf{I} + s\Lambda)^{-1}\mathbf{z}$$
$$= \sum_{i=1}^{n} \frac{s^2|z_i|^2}{1 + s\lambda_i}$$
$$= \sum_{i=1}^{n}\left((s\lambda_i - 1)\frac{|z_i|^2}{\lambda_i^2} + \frac{1}{\lambda_i^2}\frac{|z_i|^2}{1 + s\lambda_i}\right) \tag{80}$$

where $\mathbf{z} = \mathbf{U}^*\mathbf{C}_w^{-1/2}\mathbf{m}$, and z_i is the ith component of $\mathbf{z}$. Since each of the terms $|z_i|^2/(1 + s\lambda_i)$ is convex in s, $\Phi(s)$ is convex, completing the proof of the lemma. ∎

From Lemma 3, it follows that the problem (76) is equivalent to

$$\min_{\mathbf{g}}\{\max(h(L^2), h(U^2)) + \mathbf{g}^*\mathbf{C}_w\mathbf{g}\} \tag{81}$$

which can be expressed as

$$\min_{\tau, \mathbf{g}} \tau \tag{82}$$

subject to

$$h(L^2) + \mathbf{g}^*\mathbf{C}_w\mathbf{g} \leq \tau \tag{83}$$
$$h(U^2) + \mathbf{g}^*\mathbf{C}_w\mathbf{g} \leq \tau. \tag{84}$$

To find the optimal $\mathbf{g}$, we form the Lagrangian

$$\mathcal{L} = \tau + \alpha(h(L^2) + \mathbf{g}^*\mathbf{C}_w\mathbf{g} - \tau)$$
$$+ \beta(h(U^2) + \mathbf{g}^*\mathbf{C}_w\mathbf{g} - \tau) \tag{85}$$

where from the KKT conditions, $\alpha, \beta \geq 0$. Differentiating $\mathcal{L}$ with respect to τ and equating to 0

$$\alpha + \beta = 1. \tag{86}$$

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

$$(\alpha + \beta)\mathbf{C}_w\mathbf{g} + (\alpha L^2 + \beta U^2)\mathbf{R}\mathbf{g} = (\alpha L^2 + \beta U^2)\mathbf{m} \tag{87}$$

where we used the fact that

$$\frac{dh(s)}{d\mathbf{g}} = 2s(\mathbf{R}\mathbf{g} - \mathbf{m}). \tag{88}$$

Since $\alpha + \beta = 1$, it follows from (87) that

$$\mathbf{g} = (\alpha L^2 + \beta U^2)(\mathbf{C}_w + (\alpha L^2 + \beta U^2)\mathbf{R})^{-1}\mathbf{m}$$
$$= \gamma(\mathbf{C}_w + \gamma\mathbf{R})^{-1}\mathbf{m}$$
$$= \gamma\mathbf{C}_w^{-1/2}\mathbf{U}(\mathbf{I} + \gamma\Lambda)^{-1}\mathbf{z} \tag{89}$$

where we denoted

$$\gamma = \alpha L^2 + \beta U^2 = L^2 + \beta(U^2 - L^2) \tag{90}$$

and $\mathbf{z} = \mathbf{U}^* \mathbf{C}_w^{-1/2} \mathbf{m}$, and we used the eigendecomposition (79). Since $0 \le \beta \le 1$, it follows from (90) that $L^2 \le \gamma \le U^2$.

With $\mathbf{g}$ given by (89), $h(s)$ of (75) becomes

$$h(s) = s\gamma^2 \mathbf{z}^* (\mathbf{I} + \gamma\Lambda)^{-1} \Lambda (\mathbf{I} + \gamma\Lambda)^{-1} \mathbf{z}$$
$$-2s\gamma\mathbf{z}^*(\mathbf{I}+\gamma\Lambda)^{-1}\mathbf{z} + s^2\mathbf{z}^*(s\Lambda+\mathbf{I})^{-1}\mathbf{z} \quad (91)$$

which, after some algebraic manipulations, can be shown to be equal to

$$h(s) = \sum_{i=1}^{n} \frac{s|z_i|^2(s - 2\gamma - \gamma^2\lambda_i)}{(1+\gamma\lambda_i)^2(1+s\lambda_i)}. \quad (92)$$

In the case $L = U$, it follows from (90) that $\gamma = U^2$. To determine γ for $L < U$, we now use the KKT conditions. If $\gamma = U^2$ so that $\alpha = 0$ and $\beta = 1$, then from the KKT conditions it follows that (84) must be satisfied with equality, so that $\tau = h(U^2) + \mathbf{g}^* \mathbf{C}_w \mathbf{g}$. To satisfy (83) in this case, we must have that $h(L^2) \le h(U^2)$. For every $L^2 \le \gamma \le U^2$, let

$$\mathcal{G}(\gamma) = h(U^2) - h(L^2). \quad (93)$$

Using (92), it can be shown that

$$\mathcal{G}(\gamma) = \sum_{i=1}^{n} \frac{|z_i|^2(U^2 - L^2)}{(1+U^2\lambda_i)(1+L^2\lambda_i)\lambda_i}$$
$$\times \left(\frac{(1+U^2\lambda_i)(1+L^2\lambda_i)}{(1+\gamma\lambda_i)^2} - 1 \right). \quad (94)$$

It follows from (94) that if $\gamma = U^2$, then $h(U^2) < h(L^2)$, which violates (83). Similarly, if $\gamma = L^2$, then $\alpha = 1$ and $\beta = 0$, which implies from the KKT conditions that (83) must be satisfied with equality so that $\tau = h(L^2) + \mathbf{g}^* \mathbf{C}_w \mathbf{g}$. To satisfy (84) in this case, we must have that $h(U^2) \le h(L^2)$. However, from (94) with $\gamma = L^2, h(U^2) > h(L^2)$, which violates (84). We therefore conclude that for $L < U$, $L^2 < \gamma < U^2$, which implies from the KKT conditions that both (83) and (84) must be satisfied with equality. Thus, γ must be chosen such that $h(L^2) = h(U^2)$, or equivalently, such that $\mathcal{G}(\gamma)$, defined by (94), is equal to zero.

We now show that there is exactly one value $L^2 < \gamma < U^2$ such that $\mathcal{G}(\gamma) = 0$. Indeed, it can easily be seen from (94) that $\mathcal{G}(\gamma)$ is monotonically decreasing in γ. In addition, we have seen already that $\mathcal{G}(U^2) < 0$ and $\mathcal{G}(L^2) > 0$. It therefore follows that there is a unique $L^2 < \gamma < U^2$ such that $\mathcal{G}(\gamma) = 0$.

We conclude that the minimax regret estimator is equal to $\hat{\mathbf{x}} = \mathbf{g}^* \mathbf{y}$, where

$$\mathbf{g} = \gamma(\mathbf{C}_w + \gamma\mathbf{C} + \gamma\mathbf{m}\mathbf{m}^*)^{-1}\mathbf{m}$$
$$= \frac{\gamma}{1 + \gamma\mathbf{m}^*(\mathbf{C}_w + \gamma\mathbf{C})^{-1}\mathbf{m}}(\mathbf{C}_w + \gamma\mathbf{C})^{-1}\mathbf{m} \quad (95)$$

and $L^2 \le \gamma \le U^2$ is the unique root of $\mathcal{G}(\gamma)$ defined by (94) in the interval $[L^2, U^2]$.

A. Special Case

We now consider a special case in which we can obtain a closed-form expression for the minimax regret estimator. Specifically, we assume that the covariance $\mathbf{C}$ of $\mathbf{h}$ is proportional to the covariance $\mathbf{C}_w$ of $\mathbf{w}$ so that

$$\mathbf{C} = \alpha\mathbf{C}_w \quad (96)$$

for some $\alpha \ge 0$. One example in which (96) is satisfied is when the noise is white so that $\mathbf{C}_w = \sigma_w^2 \mathbf{I}$ for some $\sigma_w^2 > 0$, and the elements of $\mathbf{h}$ are uncorrelated, in which case $\mathbf{C} = \sigma_h^2 \mathbf{I}$ for some σ_h^2. Another case is when $\mathbf{C} = \mathbf{0}$ so that $\mathbf{h} = \mathbf{m}$ (with probability one).

Under the assumption (96), $\mathbf{R} = \alpha\mathbf{C}_w + \mathbf{m}\mathbf{m}^*$, and

$$\mathbf{C}_w^{-1/2}\mathbf{R}\mathbf{C}_w^{-1/2} = \alpha\mathbf{I} + \mathbf{C}_w^{-1/2}\mathbf{m}\mathbf{m}^*\mathbf{C}_w^{-1/2}. \quad (97)$$

Therefore, we can choose the eigenvector matrix $\mathbf{U}$ of $\mathbf{C}_w^{-1/2}\mathbf{R}\mathbf{C}_w^{-1/2}$ such that its first column is equal to $\mathbf{C}_w^{-1/2}\mathbf{m}/\|\mathbf{C}_w^{-1/2}\mathbf{m}\|$, and its remaining columns are a set of $n - 1$ orthonormal vectors that are orthogonal to $\mathbf{C}_w^{-1/2}\mathbf{m}$. The eigenvalues λ_i of $\mathbf{C}_w^{-1/2}\mathbf{R}\mathbf{C}_w^{-1/2}$ are then given by

$$\lambda_1 = \alpha + \mathbf{m}^* \mathbf{C}_w^{-1} \mathbf{m} \quad (98)$$

and $\lambda_i = \alpha, 2 \le i \le n$. With this choice of $\mathbf{U}$

$$\mathbf{z} = \mathbf{U}^* \mathbf{C}_w^{-1/2} \mathbf{m} = \left\| \mathbf{C}_w^{-1/2} \mathbf{m} \right\| \mathbf{e}_1 \quad (99)$$

where $\mathbf{e}_1$ is the vector whose first element is equal to 1 and whose remaining elements are equal to 0. Since $z_i = 0, 2 \le i \le n$, it follows from (94) that

$$\mathcal{G}(\gamma) = \frac{\mathbf{m}^* \mathbf{C}_w^{-1} \mathbf{m}(U^2 - L^2)}{(1+U^2\lambda_1)(1+L^2\lambda_1)\lambda_1}$$
$$\times \left(1 - \frac{(1+U^2\lambda_1)(1+L^2\lambda_1)}{(1+\gamma\lambda_1)^2} \right) \quad (100)$$

where λ_1 is given by (98). Therefore, $\mathcal{G}(\gamma) = 0$ if and only if

$$(1+\gamma\lambda_1)^2 = (1+U^2\lambda_1)(1+L^2\lambda_1) \quad (101)$$

or

$$\gamma = \frac{1}{\lambda_1}(\sqrt{(1+U^2\lambda_1)(1+L^2\lambda_1)} - 1). \quad (102)$$

The optimal value of $\mathbf{g}$ then follows from (95) and (98), as

$$\mathbf{g} = \frac{\gamma}{1 + \gamma(\alpha + \mathbf{m}^* \mathbf{C}_w^{-1} \mathbf{m})} \mathbf{C}_w^{-1} \mathbf{m}$$
$$= \frac{1}{\zeta}\left(1 - \frac{1}{\sqrt{(1+U^2\zeta)(1+L^2\zeta)}} \right) \mathbf{C}_w^{-1} \mathbf{m} \quad (103)$$

where $\zeta = \alpha + \mathbf{m}^*\mathbf{C}_w^{-1}\mathbf{m}$.

In the case that $\mathbf{C} = \sigma_h^2\mathbf{I}$ and $\mathbf{C}_w = \sigma_w^2\mathbf{I}$, the vector $\mathbf{g}$ of (103) becomes

$$\mathbf{g} = \frac{\sigma_w^2}{\zeta}\left(1 - \frac{1}{\sqrt{(\sigma_w^2 + U^2\zeta)(\sigma_w^2 + L^2\zeta)}}\right)\mathbf{m} \qquad (104)$$

where $\zeta = \sigma_h^2 + \mathbf{m}^*\mathbf{m}$. Finally, if $\mathbf{h} = \mathbf{m}$ with probability one so that $\alpha = 0$, then $\mathbf{g}$ of (103) is

$$\mathbf{g} = \frac{1}{\mathbf{m}^*\mathbf{C}_w^{-1}\mathbf{m}}$$
$$\times \left(1 - \frac{1}{\sqrt{\left(1 + U^2\mathbf{m}^*\mathbf{C}_w^{-1}\mathbf{m}\right)\left(1 + L^2\mathbf{m}^*\mathbf{C}_w^{-1}\mathbf{m}\right)}}\right)\mathbf{C}_w^{-1}\mathbf{m}.$$
$$(105)$$

Note that in this case our problem reduces to finding the minimax regret estimator for the model $\mathbf{y} = x\mathbf{m} + \mathbf{w}$. The solution to this problem was developed in [28] and is equal to (105).

We summarize our results in the following theorem.

Theorem 2: Let x denote an unknown parameter in the model $\mathbf{y} = x\mathbf{h} + \mathbf{w}$, where $\mathbf{h}$ is an unknown, length-n, random vector with mean $\mathbf{m}$ and covariance matrix $\mathbf{C}$, and $\mathbf{w}$ is a length-n, zero-mean random vector with covariance $\mathbf{C}_w$, independent of $\mathbf{h}$. Then, the solution to the problem

$$\min_{\hat{x} = \mathbf{g}^*\mathbf{y}} \max_{L \leq |x| \leq U}\left\{E\{|\hat{x} - x|^2\} - \min_{\hat{x} = \mathbf{g}^*(x)\mathbf{y}} E\{|\hat{x} - x|^2\}\right\}$$

is

$$\hat{x} = \frac{\gamma}{1 + \gamma\mathbf{m}^*(\mathbf{C}_w + \gamma\mathbf{C})^{-1}\mathbf{m}}\mathbf{m}^*(\mathbf{C}_w + \gamma\mathbf{C})^{-1}\mathbf{y}$$

where $L^2 \leq \gamma \leq U^2$ is the unique root of $\mathcal{G}(\gamma)$ defined by

$$\mathcal{G}(\gamma) = \sum_{i=1}^{n} \frac{|z_i|^2(U^2 - L^2)}{(1 + U^2\lambda_i)(1 + L^2\lambda_i)\lambda_i}$$
$$\times \left(\frac{(1 + U^2\lambda_i)(1 + L^2\lambda_i)}{(1 + \gamma\lambda_i)^2} - 1\right).$$

Here, z_i is the ith component of $\mathbf{z} = \mathbf{V}^*\mathbf{C}_w^{-1/2}\mathbf{m}$, $\mathbf{V}$ is the unitary matrix in the eigendecomposition of $\mathbf{A} = \mathbf{C}_w^{-1/2}(\mathbf{C} + \mathbf{mm}^*)\mathbf{C}_w^{-1/2}$, and λ_i is the ith eigenvalue of $\mathbf{A}$. In addition, if $\mathbf{C} = \alpha\mathbf{C}_w$, then

$$\hat{x} = \frac{1}{\zeta}\left(1 - \frac{1}{\sqrt{(1 + U^2\zeta)(1 + L^2\zeta)}}\right)\mathbf{m}^*\mathbf{C}_w^{-1}\mathbf{y}$$

where $\zeta = \alpha + \mathbf{m}^*\mathbf{C}_w^{-1}\mathbf{m}$.

VII. CONCLUSION

We treated the problem of estimating a random vector $\mathbf{x}$ in the linear model $\mathbf{y} = \mathbf{Hx} + \mathbf{w}$, where $\mathbf{H}$ is a random matrix with known second-order statistics. We first considered the minimax MSE estimator which minimizes the worst-case MSE across all bounded vectors $\mathbf{x}$ satisfying $\|\mathbf{x}\|_{\mathbf{T}} \leq U$. We showed that the minimax MSE estimator can be computed very efficiently in polynomial time using standard software packages. We then derived necessary and sufficient optimality conditions and used these conditions to develop a closed-form expression for the minimax MSE estimator for some special cases. For the scalar case, we also developed the minimax regret estimator that minimizes the worst-case regret over all parameters $L \leq |x| \leq U$. As we demonstrated, the minimax MSE approach can increase the performance over other more straightforward methods.

REFERENCES

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

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

[3] Y. C. Eldar and A. V. Oppenheim, "Covariance shaping least-squares estimation," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 686–697, Mar. 2003.

[4] 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.

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

[6] 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.

[7] 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.

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

[9] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 168–181, Jan. 2005.

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

[11] H. V. Poor, "On robust Wiener filtering," *IEEE Trans. Autom. Control*, vol. AC-25, no. 3, pp. 521–526, Jun. 1980.

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

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

[14] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation of random parameters," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1931–1946, Jul. 2004.

[15] A. Beck, Y. C. Eldar, and A. Ben-Tal, "Minimax mean-squared error estimation of multichannel signals," *SIAM J. Matrix Anal. Appl.*, submitted for publication.

[16] 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.

[17] Y. C. Eldar and N. Merhav, "Robust linear estimation under a minimax MSE—Ratio criterion," The Technion—Israel Institute of Technology, Haifa, Israel, Tech. Rep. CCIT 449, Oct. 2003.

[18] Y. C. Eldar, "Robust competitive estimation with signal and noise covariance uncertainties," *IEEE Trans. Inform. Theory*, to be published.

[19] D. Gesbert, M. Shafi, S. Da-shan, P. J. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 281–302, Apr. 2003.

[20] S. Shahbazpanahi, A. B. Gershman, Z.-Q. Luo, and K. M. Wong, "Robust adaptive beamforming for general-rank signal models," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2257–2269, Sep. 2003.

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

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

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

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

[25] J. F. Strum, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods Softw.*, vol. 11–12, pp. 625–653, 1999.

[26] D. Peaucelle, D. Henrion, and Y. Labit. *Users Guide for SeDuMi Interface 1.03* [Online]. Available: http://www.laas.fr/peaucell/SeDuMilnt.html

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

[28] Y. C. Eldar and A. Nehorai, "Competitive mean-squared error beamforming," presented at the 12th Annu. Workshop Adaptive Sensor Array Processing, Lexington, MA, Mar. 2004.

[29] ——, "Uniformly robust mean-squared error beamforming," presented at the 3rd Sensor Array Multichannel Signal Processing Workshop, Barcelona, Catalonia, Spain, Jul. 2004.

[30] ——, "Mean-squared error beamforming for signal estimation: A competitive approach," in *Robust Adaptive Beamforming*, J. Li and P. Stoica, Eds.   New York: Wiley, to be published.

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

From January 2002 to July 2002, she was a Postdoctoral Fellow at the Digital Signal Processing Group, MIT. She is currently an Associate Professor with the Department of Electrical Engineering at the Technion—Israel Institute of Technology, Haifa. 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. In 2004, she was awarded the Wolf Foundation Krill Prize, and in 2005 the Andre and Bella Meyer Lectureship. She is a member of the IEEE Signal Processing Theory and Methods technical committee and an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING.