Maximum Set Estimators With Bounded Estimation Error

Zvika Ben-Haim, *Student Member, IEEE,* and Yonina C. Eldar, *Member, IEEE*

***Abstract*—We consider the linear regression problem of estimating a deterministic parameter vector $\mathbf{x}$ from observations $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where $\mathbf{H}$ is known, and $\mathbf{w}$ is additive noise. We seek an estimator whose estimation error does not exceed a given *maximum error* for as wide a range of conditions as possible. The maximum error is a design choice and is generally lower than the error provided by the well-known least-squares (LS) estimator. We develop estimators guaranteeing the required error for as large a parameter set as possible and for as large a noise level as possible. We discuss methods for finding these estimators and demonstrate that in many cases, the proposed estimators outperform the LS estimator.**

***Index Terms*—Deterministic parameter estimation, linear estimation, minimax mean squared error.**

I. Introduction

The problem of estimating an unknown deterministic parameter vector $\mathbf{x}$ based on noisy measurements $\mathbf{y}$ is a fundamental problem in science and engineering. It is often modeled as a linear regression problem, in which $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$ is a linear transformation of $\mathbf{x}$ with additive zero-mean noise $\mathbf{w}$. In an estimation context, we would like to design an estimator $\hat{\mathbf{x}}$ to be close to $\mathbf{x}$ in some sense. For example, we may seek an estimator that minimizes the mean-squared error (MSE) $E\{\|\mathbf{x}-\hat{\mathbf{x}}\|^2\}$. The goal of this paper is to develop estimators that guarantee a required estimation error for as large a range of operating conditions as possible. The required estimation error is known in many signal processing applications; for example, in communication systems, a minimum signal to noise ratio (SNR) may be necessaryfor data transmission to be possible.

Estimation problems may be grouped into two broad classes [1]. The *Bayesian* estimation approach is used when the parameter vector $\mathbf{x}$ is random with (partially) known statistics. For instance, when second-order statistics of $\mathbf{x}$ are known, the well-known Wiener filter [2], [3] minimizes the MSE among all linear estimators. The *deterministic* estimation approach, on the other hand, assumes the parameter vector is deterministic, which is an assumption we will adopt throughout this paper.

In the deterministic case, the MSE is the sum of the variance of $\hat{\mathbf{x}}$ and the squared norm of the bias of $\hat{\mathbf{x}}$ [4]. However, since the bias is a function of the unknown vector $\mathbf{x}$, direct minimization of the MSE is not possible. A common approach to designing MSE-based estimators is to choose the minimum MSE estimator among all linear *unbiased* estimators. For unbiased estimators, the MSE equals the estimator variance, which does not depend on the value of $\mathbf{x}$ and can therefore be minimized without knowledge of $\mathbf{x}$. Minimizing the variance for unbiased estimators results in the (weighted) least-squares (LS) estimator [4]. The LS estimator has the additional property that it minimizes the measurement error $\|\mathbf{y} - \hat{\mathbf{y}}\|^2$, where $\hat{\mathbf{y}} = \mathbf{H}\hat{\mathbf{x}}$ is the estimated measurement vector. However, in an estimation context, typically, the objective is to minimize the *estimation error*, i.e., a measure of the distance between $\mathbf{x}$ and $\hat{\mathbf{x}}$ such as the MSE, rather than the measurement error.

An unbiased estimator does not necessarily guarantee low MSE. Indeed, we show that for any bounded set $\mathcal{U}$, a biased estimator exists, whose MSE is lower than the MSE of the LS estimator for *all* $\mathbf{x}$ in $\mathcal{U}$. Several regularization techniques are aimed at improving estimation performance by introducing a bias; among these are Tichonov regularization [5] (also known as ridge regression [6]) and the shrunken estimator [7].

Estimator design is a function of various *system properties*, such as the transformation matrix $\mathbf{H}$ and the noise covariance $\mathbf{C_w}$. When these properties are uncertain, one approach is to minimize the maximum (worst-case) estimation error among all possible values. This *minimax* approach was first introduced for dealing with uncertain noise statistics [8] and has since been applied in a variety of estimation problems [9]–[12]. Of particular interest to us is the case of the *bounded parameter set*, in which the estimator is designed to minimize the worst-case estimation error for any parameter vector $\mathbf{x}$ in a given parameter set $\mathcal{U}$ [13], [14]. An important property of bounded parameter set estimation is that the analysis is performed on a particular (worst-case) value of $\mathbf{x}$ and can thus be used to minimize the estimation error, for example, by minimizing the worst-case MSE. Minimax estimators can also be constructed to minimize the worst case of other estimation error functions, such as the *regret* [15], which is defined as the difference between the estimator's MSE and the best possible MSE obtained using a linear estimator which has knowledge of the parameter vector $\mathbf{x}$.

The minimax approach assumes that bounds on various system properties are known. These bounds have considerable impact on the obtained estimator: If the parameter set is too small, then the estimator may receive values of $\mathbf{x}$ for which it was not designed, and the estimation error will be larger than expected. Yet the parameter set, which defines *extreme* parameter values, is sometimes difficult to characterize based on past experience, which contains mostly *nominal* parameter values. Furthermore, attempts to estimate the parameter set from the measurements $\mathbf{y}$ generally result in a nonlinear estimator, whose computational complexity is higher.

Manuscript received April 3, 2004; revised October 28, 2004. This work was supported in part by the Sixth EU framework programme via the NEWCOM network of excellence. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jian Li.

The authors are with the Technion-Israel Institute of Technology, Haifa 32000, Israel (e-mail: zvikabh@technion.ac.il; yonina@ee.technion.ac.il).

Digital Object Identifier 10.1109/TSP.2005.851113

We propose an alternative approach that is intended for estimation problems in which a maximum allowed estimation error is given. For such cases, following the philosophy of information-gap decision theory [16], [17], we propose a *maximum set estimation* approach [18]; in general terms, this approach designs an estimator to guarantee the required error for the widest range of conditions possible. The maximum set estimation strategy can be applied in several ways, depending on the uncertain system property. In Sections II–V, we discuss the case in which the parameter set is uncertain and describe the *maximum parameter set* (MPS) estimator, which maximizes the parameter set for which error requirements are maintained. Section II opens the discussion of MPS estimation with a concrete example, which is expanded to a general framework in Section III. Theorems for the construction of several types of MPS estimators are provided in Section IV, including, in some cases, closed forms of the estimators. An application of MPS estimation in the context of channel estimation is described in Section V.

Applying the concept of maximum set estimation in a different setting, in Section VI, we consider the estimation problem when the noise covariance is known up to a constant, i.e., $E\{\mathbf{w}\mathbf{w}^*\} = \sigma^2\mathbf{C_w}$, where σ^2 is unknown. In this case, we assume that $\mathbf{x}$ lies in a known parameter set $\mathcal{U}$ and find a *maximum noise level* (MNL) estimator that guarantees a required estimation error for as large a range of noise levels σ^2 as possible. Here again, we derive a closed form for the estimator obtained under the most common settings. We conclude with a discussion in Section VII.

Throughout the paper, matrices are denoted by boldface uppercase letters, and vectors are denoted by boldface lowercase letters. The Hermitian conjugate of a matrix $\mathbf{P}$ is denoted by $\mathbf{P}^*$. The notation $\mathbf{P} \succeq 0$ indicates that the matrix $\mathbf{P}$ is positive semidefinite, and the notation $\mathbf{P} \succeq \mathbf{Q}$ indicates that $\mathbf{P}-\mathbf{Q} \succeq 0$. For any matrix $\mathbf{P} \succeq 0$, $\mathbf{P}^{1/2}$ is the unique positive semidefinite matrix satisfying $(\mathbf{P}^{1/2})^2 = \mathbf{P}$.

II. MPS Estimation: Useful Special Case

To demonstrate the main ideas of this paper, we begin by presenting an important special case of the maximum parameter set (MPS) estimator. This example is generalized and formalized in Section III.

Consider the system of measurements $\mathbf{y} \in \mathbb{C}^n$, given by

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

where $\mathbf{x} \in \mathbb{C}^m$ is an unknown deterministic vector, $\mathbf{H} \in \mathbb{C}^{n\times m}$ is a known full-rank matrix, and $\mathbf{w}$ is a zero-mean random vector with positive definite covariance $\mathbf{C_w}$. We wish to construct a linear estimator $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ of $\mathbf{x}$, such that the estimate $\hat{\mathbf{x}}$ is close to the unknown parameter $\mathbf{x}$, i.e., the *estimation error* $\epsilon(\hat{\mathbf{x}}, \mathbf{x})$ is small in some sense. For clarity, this section makes use of the MSE as the estimation error function; a general discussion follows in Section III, in which any continuous error function $\epsilon(\hat{\mathbf{x}}, \mathbf{x})$ may be used.

The MSE is given by [4]

$$\epsilon(\hat{\mathbf{x}}, \mathbf{x}) = E\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\} = v(\hat{\mathbf{x}}) + \|\mathbf{b}(\hat{\mathbf{x}})\|^2 \tag{2}$$

where

$$v(\hat{\mathbf{x}}) = E\{\|\hat{\mathbf{x}} - E\{\hat{\mathbf{x}}\}\|^2\} = \mathrm{Tr}(\mathbf{G}\mathbf{C_w}\mathbf{G}^*) \tag{3}$$

is the variance of $\hat{\mathbf{x}}$, and

$$\mathbf{b}(\hat{\mathbf{x}}) = E\{\mathbf{x} - \hat{\mathbf{x}}\} = (\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{x} \tag{4}$$

is the bias of $\hat{\mathbf{x}}$. Since $\mathbf{b}(\hat{\mathbf{x}})$ depends on the unknown value of $\mathbf{x}$, direct minimization of the MSE is not possible. A common approach is to limit discussion to unbiased estimators, in which case the MSE no longer depends on $\mathbf{x}$, and then seek the linear estimator that minimizes the MSE. This results in the least-squares (LS) estimator, given by

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

The MSE of the LS estimator is

$$\epsilon_0 = \mathrm{Tr}\left(\left(\mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{H}\right)^{-1}\right). \tag{6}$$

Since the bias is a linear function of $\mathbf{x}$, a nonzero bias causes the MSE to tend to infinity as $\|\mathbf{x}\| \to \infty$. Thus, if one requires the ability to successfully estimate *any* value of $\mathbf{x}$, then an unbiased estimator must be used. However, in some cases, a reasonable assumption can be made regarding the size of $\mathbf{x}$. If $\mathbf{x}$ is known to lie within some compact parameter set $\mathcal{U}$, then an estimator minimizing the worst-case MSE among all values of $\mathbf{x}$ in $\mathcal{U}$ can be determined. Such an estimator is called a minimax MSE estimator [13], [14] and is defined as

$$\hat{\mathbf{x}}_{\mathrm{M}} = \arg\min_{\hat{\mathbf{x}}}\max_{\mathbf{x}\in\mathcal{U}} E\{\|\mathbf{x} - \hat{\mathbf{x}}\|^2\}. \tag{7}$$

Many possibilities for choosing the parameter set $\mathcal{U}$ exist [19]. One commonly used set is the ellipsoid

$$\mathcal{U}_L = \{\mathbf{x} : \|\mathbf{x}\|_\mathbf{T} \le L\} \tag{8}$$

where $L > 0$ is a known constant, and $\|\mathbf{x}\|_\mathbf{T}^2 = \mathbf{x}^*\mathbf{T}\mathbf{x}$ for a Hermitian positive-definite weighting matrix $\mathbf{T}$. For clarity, we continue the discussion in this section using ellipsoidal parameter sets. A more general discussion is presented in the following section.

A suitable value of L is often difficult to determine. Even if a small amount of information about $\mathbf{x}$ is available, such as several past measurements, then these usually characterize typical values of $\mathbf{x}$, whereas L is meant to characterize the extreme or rare values of $\mathbf{x}$. Thus, in some cases, it is our interest to find an estimator achieving "satisfactory" performance for as large a parameter set as possible. To this end, we assume that a maximum error ϵ_m is known; this is the maximum error allowed for satisfactory performance of the system. We aim to design an MPS estimator, for which satisfactory performance is achieved for as large a parameter set as possible.

Formally, the parameter robustness $\hat{L}$ of an estimator $\hat{\mathbf{x}}$ is defined as the largest L, for which performance is satisfactory:

$$\hat{L}(\hat{\mathbf{x}}) = \sup\{L : E\{\|\mathbf{x} - \hat{\mathbf{x}}\|^2\} \le \epsilon_m, \forall \mathbf{x}^*\mathbf{T}\mathbf{x} \le L^2\}. \tag{9}$$

An MPS estimator $\hat{\mathbf{x}}_{\mathrm{PS}}$ is an estimator achieving maximal parameter robustness, i.e., for any other linear estimator $\hat{\mathbf{x}}$,

$$\hat{L}(\hat{\mathbf{x}}_{\mathrm{PS}}) \ge \hat{L}(\hat{\mathbf{x}}). \tag{10}$$

Suppose we wish to find an MPS estimator for maximum error ϵ_m equal to ϵ_0 of (6), which is the MSE of the LS estimator. The LS estimator achieves this error regardless of the value of $\mathbf{x}$; thus, its parameter robustness is infinite when the maximum error is ϵ_0 or greater. This implies that requiring a maximum error of ϵ_0 (or greater) yields the LS estimator as an MPS estimator. More interesting is the case $\epsilon_m < \epsilon_0$, for which the LS estimator no longer achieves the required error, regardless of the value of $\mathbf{x}$. An MPS estimator $\hat{\mathbf{x}}$ for a given error level $\epsilon_m < \epsilon_0$ has finite robustness, but within the parameter set $\mathcal{U}_{\hat{L}(\hat{\mathbf{x}})}$, its worst-case error does not exceed ϵ_m. Thus, an MPS estimator outperforms the LS estimator for any $\mathbf{x} \in \mathcal{U}_{\hat{L}(\hat{\mathbf{x}})}$.

In the remainder of this section, we show that a linear MPS estimator can be found by solving a quasiconvex optimization problem. An optimization problem is quasiconvex if its constraints are convex and its objective function is quasiconvex; the function $f(\mathbf{z})$ is quasiconvex if the sublevel sets $\{\mathbf{z}: f(\mathbf{z}) \le \alpha\}$ are all convex. Quasiconvex problems can be efficiently solved, for example, using bisection [20]. In addition, as we will see in Section IV-A, in many special cases, a closed form for an MPS estimator can be obtained by exploring its relation to the minimax MSE estimator.

Theorem 1: A linear maximum parameter set (MPS) estimator $\hat{\mathbf{x}}_{\mathrm{PS}} = \mathbf{G}\mathbf{y}$ satisfying (10) can be found by solving the following quasiconvex optimization problem:

$$\min_{\mathbf{G},\lambda,y} \; y/\lambda$$
$$\text{s.t.} \begin{cases} \begin{bmatrix} y+\epsilon_m & \mathbf{g}^* \\ \mathbf{g} & \mathbf{I} \end{bmatrix} \succeq 0 \\ \begin{bmatrix} \lambda\mathbf{I} & \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{G}\mathbf{H})^* \\ (\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{T}^{-1/2} & \mathbf{I} \end{bmatrix} \succeq 0 \end{cases} \tag{11}$$

where $\mathbf{g}$ is the vector obtained by stacking the columns of $\mathbf{G}\mathbf{C}_{\mathbf{w}}^{1/2}$. The parameter robustness $\hat{L}$ of this estimator is given by $\sqrt{-y/\lambda}$ for the optimal values of y and λ.

Proof: We seek an estimator $\hat{\mathbf{x}}_{\mathrm{PS}}$ satisfying (10) with $\hat{L}(\hat{\mathbf{x}})$ defined by (9), which is equivalent to solving the optimization problem

$$\max_{\mathbf{G},L^2} L^2 \quad \text{s.t. } \epsilon_m \ge \max_{\|\mathbf{x}\|_{\mathbf{T}}^2 \le L^2} E\{\|\mathbf{x}-\hat{\mathbf{x}}\|^2\} \tag{12}$$

where $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$. Using (2)–(4), we find that for any given $\hat{\mathbf{x}}$

$$\max_{\|\mathbf{x}\|_{\mathbf{T}} \le L} E\{\|\mathbf{x}-\hat{\mathbf{x}}\|^2\} = \mathrm{Tr}(\mathbf{G}\mathbf{C}_{\mathbf{w}}\mathbf{G}^*) + \max_{\|\mathbf{x}\|_{\mathbf{T}} \le L} \|(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{x}\|^2. \tag{13}$$

However

$$\max_{\|\mathbf{x}\|_{\mathbf{T}} \le L} \|(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{x}\|^2 = \max_{\mathbf{z}^*\mathbf{z} \le L^2} \|\mathbf{P}\mathbf{z}\|^2 = \lambda_{\max} L^2 \tag{14}$$

where $\mathbf{P} = (\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{T}^{-1/2}$ and $\lambda_{\max}$ is the maximum eigenvalue of $\mathbf{P}^*\mathbf{P}$. We can express $\lambda_{\max}$ as the solution to the semidefinite problem

$$\min_{\lambda} \lambda \quad \text{s.t. } \mathbf{P}^*\mathbf{P} \preceq \lambda\mathbf{I}. \tag{15}$$

Consider the problem

$$\max_{\mathbf{G},\lambda,L^2} L^2 \tag{16}$$
$$\text{s.t.} \begin{cases} \mathrm{Tr}(\mathbf{G}\mathbf{C}_{\mathbf{w}}\mathbf{G}^*) + \lambda L^2 \le \epsilon_m & \text{(a)} \\ \mathbf{P}^*\mathbf{P} \preceq \lambda\mathbf{I}. & \text{(b)} \end{cases}$$

We claim that the optimal solution to this problem always has $\lambda = \lambda_{\max}$ so that (16) and (12) are equivalent.

Let $\mathbf{g}$ be the vector obtained by stacking the columns of $\mathbf{G}\mathbf{C}_{\mathbf{w}}^{1/2}$. Using Schur's Lemma [21, p. 472], it is shown in [14] that (16a) and (16b) are equivalent to the following matrix inequalities:

$$\begin{bmatrix} \epsilon_m - \lambda L^2 & \mathbf{g}^* \\ \mathbf{g} & \mathbf{I} \end{bmatrix} \succeq 0 \tag{17a}$$

$$\begin{bmatrix} \lambda\mathbf{I} & \mathbf{P}^* \\ \mathbf{P} & \mathbf{I} \end{bmatrix} \succeq 0. \tag{17b}$$

Defining $r = -L^2$, (17a) becomes

$$r\begin{bmatrix} \lambda & 0 \\ 0 & 0 \end{bmatrix} \succeq -\begin{bmatrix} \epsilon_m & \mathbf{g}^* \\ \mathbf{g} & \mathbf{I} \end{bmatrix}. \tag{18}$$

We now add a scalar optimization parameter y and note that the optimization problem is equivalent to

$$\min_{\mathbf{G},\lambda,r,y} r$$
$$\text{s.t.} \begin{cases} r\lambda \ge y \\ \begin{bmatrix} y & 0 \\ 0 & 0 \end{bmatrix} \succeq -\begin{bmatrix} \epsilon_m & \mathbf{g}^* \\ \mathbf{g} & \mathbf{I} \end{bmatrix} \\ \begin{bmatrix} \lambda\mathbf{I} & \mathbf{P}^* \\ \mathbf{P} & \mathbf{I} \end{bmatrix} \succeq 0. \end{cases} \tag{19}$$

It is evident that the optimal solution to this problem satisfies $r = y/\lambda$; substituting this into the above problem yields the required optimization problem (11). The objective function of (11) is quasiconvex, and all its constraints are convex; therefore, this is a quasiconvex optimization problem [20]. ■

In the next section, we generalize the discussion to MPS estimators that optimize various error functions over different parameter sets. We also demonstrate a relation between MPS estimation and minimax estimation, which provides further insight into the idea of MPS estimation and yields an alternative method for finding MPS estimators. In particular, this leads to a closed form for an MPS estimator when the weighting matrix $\mathbf{T}$ commutes with $\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}$, which occurs, for example, when $\mathbf{T} = \mathbf{I}$.

III. General Form of MPS Estimators

The example presented in Section II is a special case of an MPS estimator, which can be generalized to include different error functions and parameter sets. In Section III-A, we provide definitions that construct the general form of MPS estimators. In Section III-B, we use these definitions to prove a useful relation between minimax and MPS estimators. This relation will be used in Section IV to find efficient algorithms for identifying MPS estimators and, in some cases, allows us to derive closed forms for MPS estimators.

A. Definitions

Based on the system of measurements (1), the following set of definitions constructs the MPS estimator.

Definition 1: The *system properties* required for the design of a MPS estimator are the following three components:

1) An *error function* $\epsilon(\hat{\mathbf{x}}, \mathbf{x})$, which quantifies the degree to which an estimator $\hat{\mathbf{x}}$ misrepresents the specific value $\mathbf{x}$. The error function must be continuous.
2) A *maximum error* ϵ_m, which defines the error value required for successful operation of the system. This is a deterministic real number that must be known to the designer. An MPS estimator seeks to maximize the range of values of $\mathbf{x}$ for which the maximum error is guaranteed.
3) A *class of parameter sets* $\{\mathcal{U}_L \subseteq \mathbb{C}^m\}_{L \geq 0}$, which defines feasible values of $\mathbf{x}$ under varying parameter set bounds L. These sets obey three basic properties:
   a) As L increases, more values of $\mathbf{x}$ become feasible, so that the sets $\mathcal{U}_L$ are nested:
$$L_1 < L_2 \iff \mathcal{U}_{L_1} \subset \mathcal{U}_{L_2}. \tag{20}$$
   b) The parameter sets are linearly expanding: For all $L_1, L_2 > 0$
$$\mathcal{U}_{L_1} = \frac{L_1}{L_2}\mathcal{U}_{L_2} = \left\{\mathbf{x} : \frac{L_2}{L_1}\mathbf{x} \in \mathcal{U}_{L_2}\right\}. \tag{21}$$
   This requirement implies that the parameter sets are centered on the origin: an assumption we adopt without loss of generality.
   c) The sets $\mathcal{U}_L$ are compact (i.e., closed and bounded). This requirement ensures the existence of a maximum error for every parameter set.

Most common bounds fulfill the requirements for the class of parameter sets above. The weighted norm $\mathcal{U}_L = \{\mathbf{x} : \mathbf{x}^*\mathbf{T}\mathbf{x} \leq L^2\}$ used in Section II is one example. Another example is the box bound $\mathcal{U}_L = \{\mathbf{x} : |x_i| \leq Lb_i, \forall i\}$, where $b_i > 0$ are constants.

Definition 2: The *parameter robustness* $\hat{L}(\hat{\mathbf{x}})$ of an estimator $\hat{\mathbf{x}}$ (for particular system properties) is the largest parameter set bound L for which the maximum error is guaranteed, namely
$$\hat{L}(\hat{\mathbf{x}}) = \sup\{L : \forall \mathbf{x} \in \mathcal{U}_L, \epsilon(\hat{\mathbf{x}}, \mathbf{x}) \leq \epsilon_m\}. \tag{22}$$

Definition 3: An *MPS estimator* (among estimators of a given class $\mathcal{E}$) is an estimator $\hat{\mathbf{x}}_{\mathrm{PS}}$ such that, for any $\hat{\mathbf{x}} \in \mathcal{E}$
$$\hat{L}(\hat{\mathbf{x}}_{\mathrm{PS}}) \geq \hat{L}(\hat{\mathbf{x}}). \tag{23}$$

Definition 4: A *minimax* (or *bounded parameter set*) *estimator* for the compact parameter set $\mathcal{U}$, among estimators of a given class $\mathcal{E}$, is an estimator $\hat{\mathbf{x}}_{\mathrm{M}}$ minimizing the worst-case error in $\mathcal{U}$. In other words, for any $\hat{\mathbf{x}} \in \mathcal{E}$
$$\max_{\mathbf{x}\in\mathcal{U}} \epsilon(\hat{\mathbf{x}}_{\mathrm{M}}, \mathbf{x}) \leq \max_{\mathbf{x}\in\mathcal{U}} \epsilon(\hat{\mathbf{x}}, \mathbf{x}). \tag{24}$$

Note that the maxima in (24) are well-defined, since the continuous function ϵ obtains a maximum over any compact set $\mathcal{U}$. In addition, note that Definitions 3 and 4 do not imply the unique existence of MPS or minimax estimators. In fact, for some choices of ϵ_m, many estimators with infinite robustness exist. However, we will see that in many cases of interest, the MPS estimator exists and is unique.

The estimator presented in Section II is a special case of an MPS estimator, which makes use of a particular choice of the error function and of the class of parameter sets. Specifically, the MSE (2) is used as the error function, ellipsoids (8) of increasing size, and constant axis ratios are used as the nested parameter sets, and the estimator is restricted to being linear.

B. Minimax and MPS Estimators

An interesting and useful relation exists between the MPS estimator $\hat{\mathbf{x}}_{\mathrm{PS}}$ and the minimax estimator $\hat{\mathbf{x}}_{\mathrm{M}}$: The MPS estimator maximizes the parameter robustness L within a range defined by the known value of ϵ, whereas the minimax estimator minimizes the worst-case error ϵ within a range defined by the known value of L.

To formalize this relation, let $\{\mathcal{U}_L\}_{L\geq 0}$ be a class of parameter sets, and define the *worst-case error function*
$$e(L) = \max_{\mathbf{x}\in\mathcal{U}_L} \epsilon(\hat{\mathbf{x}}_{\mathrm{M}}(L), \mathbf{x}) \tag{25}$$
where $\hat{\mathbf{x}}_{\mathrm{M}}(L)$ is a minimax estimator for the parameter set $\mathcal{U}_L$. Clearly, $e(L)$ is nondecreasing since enlarging the parameter set cannot decrease the worst-case error. This tradeoff between parameter set size and worst-case error is applicable to MPS estimators as well. Indeed, if $e(L)$ is strictly increasing in L, there exists a one-to-one correspondence between the parameter set bound L and the worst-case error $e(L)$. In this case, it is intuitive to expect a one-to-one correspondence between minimax and MPS estimators. Thus, we have the following theorem.

Theorem 2: Consider an error function $\epsilon(\hat{\mathbf{x}}, \mathbf{x})$ and a class of parameter sets $\{\mathcal{U}_L\}_{L\geq 0}$, as defined in Definition 1. Assume that the worst-case error $e(L)$ of (25) is strictly increasing in L. For any L, an estimator $\hat{\mathbf{x}}$ is an MPS estimator with worst-case error $\epsilon_m = e(L)$ if, and only if, it is a minimax estimator over the uncertainty set $\mathcal{U}_L$.

Proof: Suppose first that $\hat{\mathbf{x}}_{\mathrm{PS}}$ is an MPS estimator with worst-case error $\epsilon_m = e(L_0)$ for a given L_0. Let $L_1 = \hat{L}(\hat{\mathbf{x}}_{\mathrm{PS}})$, and notice that $L_1 \geq L_0$ (we will show presently that $L_1 = L_0$). Assume by contradiction that $\hat{\mathbf{x}}_{\mathrm{PS}}$ is not a minimax estimator over $\mathcal{U}_{L_1}$. Then, by Definition 4, there exists an estimator $\hat{\mathbf{x}}_{\mathrm{M}}$ such that
$$\max_{\mathbf{x}\in\mathcal{U}_{L_1}} \epsilon(\hat{\mathbf{x}}_{\mathrm{M}}, \mathbf{x}) < \max_{\mathbf{x}\in\mathcal{U}_{L_1}} \epsilon(\hat{\mathbf{x}}_{\mathrm{PS}}, \mathbf{x}) \leq \epsilon_m. \tag{26}$$

By Definition 1, the parameter sets expand linearly, so that for sufficiently small $\alpha > 1$, each of the values in the parameter set $\mathcal{U}_{\alpha L_1}$ is arbitrarily close to some value in $\mathcal{U}_{L_1}$. Furthermore, by Definition 1, ϵ is continuous so that sufficiently small changes in $\mathbf{x}$ yield arbitrarily small changes in $\epsilon(\hat{\mathbf{x}}_{\mathrm{PS}}, \mathbf{x})$. Hence, there exists a sufficiently small $\alpha > 1$ for which
$$\max_{\mathbf{x}\in\mathcal{U}_{\alpha L_1}} \epsilon(\hat{\mathbf{x}}_{\mathrm{M}}, \mathbf{x}) \leq \epsilon_m. \tag{27}$$

Thus, the parameter robustness of $\hat{\mathbf{x}}_{\mathrm{M}}$ is at least $\alpha L_1 > L_1 = \hat{L}(\hat{\mathbf{x}}_{\mathrm{PS}})$, which contradicts the fact that $\hat{\mathbf{x}}_{\mathrm{PS}}$ is an MPS estimator. Hence, $\hat{\mathbf{x}}_{\rm PS}$ is a minimax estimator over $\mathcal{U}_{L_1}$, and its worst-case error is $e(L_1)$. However, from (26), the worst-case error of $\hat{\mathbf{x}}_{\rm PS}$ is $e(L_0)$. Since $e(L)$ is strictly increasing, this implies $L_0 = L_1$. We conclude that $\hat{\mathbf{x}}_{\rm PS}$ is minimax over $\mathcal{U}_{L_0}$.

We now prove that a minimax estimator is an MPS estimator. For any L_0, let $\hat{\mathbf{x}}_{\rm M}(L_0)$ be a minimax estimator for the uncertainty set $\mathcal{U}_{L_0}$. Assume by contradiction that $\hat{\mathbf{x}}_{\rm M}(L_0)$ is not an MPS estimator for the maximum error $\epsilon_m = e(L_0)$. Then, there exists an $\hat{\mathbf{x}}_{\rm PS}$ with robustness $L_1 = \hat{L}(\hat{\mathbf{x}}_{\rm PS})$ such that $L_1 > \hat{L}(\hat{\mathbf{x}}_{\rm M}(L_0)) \geq L_0$. Therefore

$$\max_{\mathbf{x}\in\mathcal{U}_{L_1}} \epsilon(\hat{\mathbf{x}}_{\rm PS}, \mathbf{x}) \leq \epsilon_m = e(L_0) < e(L_1). \quad (28)$$

However, by (25)

$$\max_{\mathbf{x}\in\mathcal{U}_{L_1}} \epsilon(\hat{\mathbf{x}}_{\rm M}(L_1), \mathbf{x}) = e(L_1). \quad (29)$$

Hence, $\hat{\mathbf{x}}_{\rm PS}$ achieves a lower worst-case error over $\mathcal{U}_{L_1}$ than the minimax estimator of $\mathcal{U}_{L_1}$, which is a contradiction. We conclude that $\hat{\mathbf{x}}_{\rm M}(L_0)$ must be an MPS estimator. ■

We have shown that when the worst-case error function $e(L)$ is strictly increasing in L, there is a one-to-one correspondence between minimax and MPS estimators. As we will see in the following sections, $e(L)$ is indeed strictly increasing for many important cases, such as the MSE error function. However, this is not always the case. For instance, if the error function decreases with $\|\mathbf{x}\|$, then increasing the parameter set will not increase the worst-case error.

Theorem 2 can be used to efficiently find an MPS estimator using known minimax estimators. This is done using bisection on the worst-case error function $e(L)$. Since the function is strictly monotonic, a value of L yielding $e(L)$, which equals ϵ_m to any desired accuracy, can efficiently be found. From Theorem 2, the minimax estimator $\hat{\mathbf{x}}_{\rm M}(L)$ equals the desired MPS estimator.

Similarities notwithstanding, minimax and MPS estimators differ qualitatively in the type of information on which their design is based. A minimax estimator requires that a bound on the uncertain parameter $\mathbf{x}$ be stated, whereas an MPS estimator requires knowledge of the maximum error under which the system still operates correctly. Thus, a proper choice of an estimator should depend on the nature of the information available to the designer.

IV. Estimators for Various Error Functions

We now use Theorem 2 to develop MPS estimators for two cases of interest: the MSE estimator (Section IV-A) and the regret estimator (Section IV-B).

A. Linear MSE Estimators

Consider the MPS estimation problem when the error function of interest is the MSE, and the estimator is restricted to being linear. In Theorem 3, we show that minimax and MPS criteria for optimality are equivalent in these circumstances. This allows us to find an MPS estimator whenever an algorithm for finding a minimax estimator is known. In particular, Theorem 4 derives a closed form for the estimator when the uncertainty sets are spherical.

Theorem 3: Suppose that the error function of interest is the MSE (2), let $\mathcal{E}$ be the class of linear estimators, and let $\{\mathcal{U}_L\}_{L\geq 0}$ be a class of parameter sets, as defined in Definition 1. An estimator $\hat{\mathbf{x}} \in \mathcal{E}$ is a linear minimax estimator over $\mathcal{U}_L$ if, and only if, it is a linear MPS estimator with maximum error ϵ_m equal to the worst-case error $e(L)$ of (25).

The proof of Theorem 3 is based on the following lemma.

Lemma 1: Given any bounded parameter set $\mathcal{U}$, there exists a linear biased estimator $\hat{\mathbf{x}}_{\rm M}$ whose MSE is lower than the MSE of the least-squares estimator, for *all* $\mathbf{x} \in \mathcal{U}$.

Proof: For any bounded $\mathcal{U}$, there exists a finite R such that $\mathcal{U}$ is bounded within a sphere of radius R. The linear minimax MSE estimator for this sphere is given by [14]

$$\hat{\mathbf{x}}_{\rm M} = \mathbf{G}\mathbf{y} = \frac{R^2}{R^2+\epsilon_0}\left(\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{y} \quad (30)$$

where ϵ_0 is the MSE of the unbiased estimator (6). We now show that $\hat{\mathbf{x}}_{\rm M}$ achieves a lower MSE than the LS estimator for all $\mathbf{x} \in \mathcal{U}$. The bias of $\hat{\mathbf{x}}_{\rm M}$ is given by

$$\mathbf{b}(\hat{\mathbf{x}}_{\rm M}) = E\{\hat{\mathbf{x}}_{\rm M} - \mathbf{x}\} = (\beta - 1)\mathbf{x} \quad (31)$$

where $\beta = R^2/(R^2+\epsilon_0)$. The variance of $\hat{\mathbf{x}}_{\rm M}$ is

$$v(\hat{\mathbf{x}}_{\rm M}) = \mathrm{Tr}(\mathbf{G}\mathbf{C}_{\mathbf{w}}\mathbf{G}^*) = \beta^2\epsilon_0. \quad (32)$$

Using (2), we have, for all $\mathbf{x} \in \mathcal{U}$

$$\begin{aligned} \mathrm{MSE}(\hat{\mathbf{x}}_{\rm M}) &= \beta^2\epsilon_0 + (1-\beta)^2\|\mathbf{x}\|^2 && (33)\\ &\leq \beta^2\epsilon_0 + (1-\beta)^2 R^2 && (34)\\ &= \left(\frac{R^2}{R^2+\epsilon_0}\right)\epsilon_0 && (35)\\ &< \epsilon_0. && (36)\end{aligned}$$

Hence, for all $\mathbf{x} \in \mathcal{U}$, the MSE using $\hat{\mathbf{x}}_{\rm M}$ is lower than the MSE for an unbiased estimator. ■

Proof of Theorem 3: By Theorem 2, it is sufficient to show that $e(L)$ is strictly increasing. Let $\hat{\mathbf{x}}_{\rm M}(L) = \mathbf{G}_L\mathbf{y}$ be a linear minimax MSE estimator over the set $\mathcal{U}_L$. From (2)–(4), we have

$$e(L) = \mathrm{Tr}(\mathbf{G}_L\mathbf{C}_{\mathbf{w}}\mathbf{G}_L^*) + \max_{\mathbf{x}\in\mathcal{U}_L}\|(\mathbf{I} - \mathbf{G}_L\mathbf{H})\mathbf{x}\|^2. \quad (37)$$

Lemma 1 states that there exists a biased estimator that achieves lower MSE than the LS estimator for any $\mathbf{x} \in \mathcal{U}_L$. Since the LS estimator achieves the lowest possible MSE among all unbiased estimators, it follows that the minimax MSE estimator must be biased, i.e., $\mathbf{G}_L\mathbf{H} \neq \mathbf{I}$.

We now show that $\max_{\mathbf{x}\in\mathcal{U}_L}\|(\mathbf{I} - \mathbf{G}_L\mathbf{H})\mathbf{x}\|^2$ is obtained only on the boundary of $\mathcal{U}_L$. Let $\mathbf{x}_0 \in \mathcal{U}_L$ be a point that is *not* on the boundary. Then, there exists a sufficiently small sphere S, centered on $\mathbf{x}_0$, such that $S \subset \mathcal{U}_L$. In particular, S necessarily includes a point $(1+\delta)\mathbf{x}_0$ (for a sufficiently small $\delta > 0$). Since $\mathbf{G}_L\mathbf{H} \neq \mathbf{I}$, we have

$$\|(\mathbf{I} - \mathbf{G}_L\mathbf{H})(1+\delta)\mathbf{x}_0\|^2 > \|(\mathbf{I} - \mathbf{G}_L\mathbf{H})\mathbf{x}_0\|^2. \quad (38)$$

Thus, $\max_{\mathbf{x}\in\mathcal{U}_L} \|(\mathbf{I}-\mathbf{G}_L\mathbf{H})\mathbf{x}\|^2$ is not obtained at $\mathbf{x}_0$; rather, the maximum is obtained only on the boundary of $\mathcal{U}_L$. Therefore, by shrinking the parameter set, the worst-case error must decrease: For any $L < M$,

$$\max_{\mathbf{x}\in\mathcal{U}_L} E\{\|\hat{\mathbf{x}}_{\mathrm{M}}(M)-\mathbf{x}\|^2\} < \max_{\mathbf{x}\in\mathcal{U}_M} E\{\|\hat{\mathbf{x}}_{\mathrm{M}}(M)-\mathbf{x}\|^2\}. \quad (39)$$

However, since $\hat{\mathbf{x}}_{\mathrm{M}}(L)$ is a minimax MSE estimator for $\mathcal{U}_L$

$$\max_{\mathbf{x}\in\mathcal{U}_L} E\{\|\hat{\mathbf{x}}_{\mathrm{M}}(L)-\mathbf{x}\|^2\} \le \max_{\mathbf{x}\in\mathcal{U}_L} E\{\|\hat{\mathbf{x}}_{\mathrm{M}}(M)-\mathbf{x}\|^2\}. \quad (40)$$

Together with (39), this implies that $e(L) < e(M)$ for all $L < M$, which completes the proof. ■

As we have seen, using the MSE as an error function, the set of minimax estimators equals the set of MPS estimators for a given class of parameter sets. Thus, finding an MPS estimator for a given maximum error ϵ_m becomes simply a matter of finding a minimax estimator whose worst-case error is ϵ_m. In particular, when a closed form is known for the set of minimax estimators and their worst-case errors, one can find a closed form for MPS estimators as well. This is the case for the class of ellipsoidal parameter sets, as demonstrated by the following theorem.

Theorem 4: Consider the MSE error function and define the ellipsoidal parameter sets $\mathcal{U}_L = \{\mathbf{x} : \mathbf{x}^*\mathbf{T}\mathbf{x} \le L^2\}$. Let $\hat{\mathbf{x}}_{\mathrm{LS}}$ be the LS estimator (5), and let ϵ_0 be the MSE of $\hat{\mathbf{x}}_{\mathrm{LS}}$ (6).

a) Suppose $\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}$ and $\mathbf{T}$ have the same unitary eigenvector matrix $\mathbf{V}$ so that $\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^*$, where $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1,\ldots\sigma_m)$, and $\mathbf{T} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^*$ where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1,\ldots\lambda_m)$ with $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_m > 0$. An MPS estimator for a given maximum error ϵ_m is given by

$$\hat{\mathbf{x}}_{\mathrm{PS}} = \begin{cases} \mathbf{P}(\mathbf{I}-\alpha\mathbf{T}^{1/2})\hat{\mathbf{x}}_{\mathrm{LS}}, & \epsilon_m < \epsilon_0 \\ \hat{\mathbf{x}}_{\mathrm{LS}}, & \epsilon_m \ge \epsilon_0 \end{cases} \quad (41)$$

where

$$\mathbf{P} = \mathbf{V}\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m-k}\end{bmatrix}\mathbf{V}^* \quad (42)$$

$$\alpha = \frac{\sum_{i=k+1}^{m}\frac{1}{\sigma_i} - \epsilon_m}{\sum_{i=k+1}^{m}\frac{\lambda_i^{1/2}}{\sigma_i}} \quad (43)$$

and

$$k = \min\left\{i : \alpha\lambda_{i+1}^{1/2} < 1\right\}. \quad (44)$$

b) Suppose $\mathbf{T} = \mathbf{I}$, i.e., the parameter sets are spherical. In this case, an MPS estimator is

$$\hat{\mathbf{x}}_{\mathrm{PS}} = \begin{cases} \frac{\epsilon_m}{\epsilon_0}\hat{\mathbf{x}}_{\mathrm{LS}}, & \epsilon_m < \epsilon_0 \\ \hat{\mathbf{x}}_{\mathrm{LS}}, & \epsilon_m \ge \epsilon_0. \end{cases} \quad (45)$$

The parameter robustness of this estimator is given by

$$\hat{L}(\hat{\mathbf{x}}_{\mathrm{PS}}) = \begin{cases} \sqrt{\frac{\epsilon_0\epsilon_m}{\epsilon_0-\epsilon_m}}, & \epsilon_m < \epsilon_0 \\ \infty, & \epsilon_m \ge \epsilon_0. \end{cases} \quad (46)$$

Proof: To prove a), we seek an estimator that guarantees an error not exceeding ϵ_m for as large a parameter set as possible. We begin with the case $\epsilon_m \ge \epsilon_0$. In this case, the allowed error is larger than ϵ_0: the MSE obtained by the LS estimator. Since the LS estimator guarantees this error for any value of $\mathbf{x}$, its parameter robustness is infinite; thus, $\hat{\mathbf{x}}_{\mathrm{LS}}$ is an MPS estimator for this trivial case. We now consider the case $\epsilon_m < \epsilon_0$. From Theorem 3, an MPS estimator is also a minimax estimator. It is shown in [14] that the minimax MSE estimator for a given parameter set $\mathcal{U}_L$ is given by

$$\hat{\mathbf{x}}_{\mathrm{M}} = \mathbf{V}\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m-k}\end{bmatrix}\mathbf{V}^*(\mathbf{I}-\alpha\mathbf{T}^{1/2})\hat{\mathbf{x}}_{\mathrm{LS}} \quad (47)$$

where

$$\alpha = \frac{\sum_{i=k+1}^{m}\frac{\lambda_i^{1/2}}{\sigma_i}}{L^2 + \sum_{i=k+1}^{m}\frac{\lambda_i}{\sigma_i}} \quad (48)$$

and k is defined in (44). The worst-case error for this estimator is

$$\sum_{i=k+1}^{m}\frac{1-\alpha\lambda_i^{1/2}}{\sigma_i}. \quad (49)$$

We require a value of L for which the worst-case error equals ϵ_m. Equating (49) with ϵ_m, we arrive at (43).

To prove part b), note that the case $\mathbf{T} = \mathbf{I}$ is a special case of a) in which $\mathbf{V}$ is unitary, and $\boldsymbol{\Lambda} = \mathbf{I}$. Substituting $\lambda_i = 1$ in the MPS estimator obtained for a), we observe that $\alpha < 1$, and thus, $k = 0$. Furthermore

$$\sum_{i=1}^{m}\frac{1}{\sigma_i} = \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}) = \mathrm{Tr}\left(\left(\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}\right)^{-1}\right) = \epsilon_0 \quad (50)$$

and thus

$$\alpha = \frac{\epsilon_0 - \epsilon_m}{\epsilon_0}. \quad (51)$$

Substituting these results into (41) yields the required estimator (45). We have already seen that the parameter robustness when $\epsilon_m \ge \epsilon_0$ is infinite. To find the parameter robustness when $\epsilon_m < \epsilon_0$, notice that (48) is now

$$\alpha = \frac{\epsilon_0}{L^2+\epsilon_0}. \quad (52)$$

Combining this with (51) yields

$$L^2 = \frac{\epsilon_0\epsilon_m}{\epsilon_0-\epsilon_m} \quad (53)$$

which is the required result (46). ■

It is sometimes useful to find the actual MSE obtained by an MPS estimator. The MSE can be calculated for the matching minimax estimator. For example, it has been shown in (33) that the MSE of the minimax estimator for a spherical parameter set $\mathbf{x}^*\mathbf{x} \le L^2$ is given by

$$\mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{PS}}) = \left(\frac{L^2}{L^2+\epsilon_0}\right)^2\epsilon_0 + \left(\frac{\epsilon_0}{L^2+\epsilon_0}\right)^2\|\mathbf{x}\|^2. \quad (54)$$

Substituting the value of L^2 from (46), we have

$$\mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{PS}}) = \begin{cases} \frac{\epsilon_m^2}{\epsilon_0} + \left(\frac{\epsilon_0-\epsilon_m}{\epsilon_0}\right)^2\|\mathbf{x}\|^2, & \epsilon_m < \epsilon_0 \\ \epsilon_0, & \epsilon_m \ge \epsilon_0. \end{cases} \quad (55)$$

Thus, the MSE of the maximum spherical parameter set estimator is a linear function of $\|\mathbf{x}\|^2$. This result is useful for comparing the performance of the MPS estimator with other estimators, as we demonstrate in Section V.

B. Linear Regret Estimators

We now present a different example of an MPS estimator: one that guarantees a worst-case *regret* [15]. The regret is defined as the difference between the MSE and the best MSE obtainable using a linear estimator $\hat{\mathbf{x}} = \mathbf{G}(\mathbf{x})\mathbf{y}$, which is a function of $\mathbf{x}$. Because we are limiting the discussion to linear estimators, even an estimator with knowledge of the value of $\mathbf{x}$ cannot achieve zero MSE. Minimizing the regret is intuitively appealing as it attempts to disregard errors resulting from limitations of linear estimators. It has been shown [15] that the regret is given by

$$\epsilon(\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}, \mathbf{x}) = \mathrm{Tr}(\mathbf{G}\mathbf{C_w}\mathbf{G}^*) + \|(\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{x}\|^2 - \frac{\mathbf{x}^*\mathbf{x}}{1 + \mathbf{x}^*\mathbf{H}^*\mathbf{C_w}^{-1}\mathbf{H}\mathbf{x}}. \quad (56)$$

In this section, we limit our discussion to parameter sets of the form $\mathcal{U}_L = \{\mathbf{x} : \mathbf{x}^*\mathbf{T}\mathbf{x} \le L^2\}$, where $\mathbf{T}$ is a Hermitian positive definite weighting matrix. For analytical tractability, we further restrict the discussion to the case where $\mathbf{T}$ and $\mathbf{H}^*\mathbf{C_w}^{-1}\mathbf{H}$ have the same eigenvectors. We show that under these assumptions, the linear MPS regret estimator is equivalent to the linear minimax regret estimator. It follows that the MPS estimator can be found as easily as the minimax estimator. In particular, closed-form solutions are known for some values of $\mathbf{T}$ and L [15].

Theorem 5: Suppose that the error function of interest is the regret of (56). Let $\mathcal{E}$ be the class of linear estimators, and let $\mathcal{U}_L = \{\mathbf{x} : \ \mathbf{x}^*\mathbf{T}\mathbf{x} \le L^2\}$ be a class of parameter sets, where $\mathbf{T} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^*$ is a Hermitian positive definite weighting matrix, $\mathbf{\Lambda}$ is a diagonal matrix with diagonal elements $\lambda_i > 0$, and $\mathbf{V}$ is an eigenvector matrix of $\mathbf{H}^*\mathbf{C_w}^{-1}\mathbf{H}$. An estimator $\hat{\mathbf{x}} \in \mathcal{E}$ is a linear minimax regret estimator over $\mathcal{U}_L$ if, and only if, it is a linear MPS regret estimator with maximum error ϵ_m equal to the worst-case error $e(L)$ of (25).

Proof: By Theorem 2, it is sufficient to show that $e(L)$ is strictly increasing with L. It has been shown in [15, Th. 1] that under the conditions of Theorem 5, the linear minimax regret estimator $\hat{\mathbf{x}}_{\mathrm{M}}(L)$ is given by

$$\hat{\mathbf{x}}_{\mathrm{M}}(L) = \mathbf{V}\mathbf{D}\mathbf{V}^* \left(\mathbf{H}^*\mathbf{C_w}^{-1}\mathbf{H}\right)^{-1} \mathbf{H}^*\mathbf{C_w}^{-1}\mathbf{y} \quad (57)$$

where $\mathbf{D}$ is a diagonal matrix whose diagonal elements $\mathbf{d} = [d_1, \ldots d_m]^T$ are the solution to the optimization problem

$$\min_{\tau,\mathbf{d}} \ \tau \quad (58)$$
$$\text{s.t.} \begin{cases} F_1(\mathbf{d}) \le \tau & \text{(a)} \\ F_2(\mathbf{d}, \mathbf{r}) \le \tau, \quad \forall \mathbf{r} \in \mathcal{S} & \text{(b)} \end{cases}$$

where

$$F_1(\mathbf{d}) = \sum_{i=1}^m \frac{d_i^2}{\sigma_i}$$
$$F_2(\mathbf{d}, \mathbf{r}) = \sum_{i=1}^m \frac{d_i^2}{\sigma_i} + L^2 \sum_{i=1}^m (1 - d_i)^2 r_i - \frac{L^2 \sum_{i=1}^m r_i}{1 + L^2 \sum_{i=1}^m \sigma_i r_i}$$
$$\mathcal{S} = \left\{\mathbf{r} : \ r_i \ge 0, \sum_{i=1}^m \lambda_i r_i = 1\right\} \quad (59)$$

and σ_i are eigenvectors of $\mathbf{H}^*\mathbf{C_w}^{-1}\mathbf{H}$ such that $\mathbf{H}^*\mathbf{C_w}^{-1}\mathbf{H} = \mathbf{V}\mathbf{\Sigma}\mathbf{V}^*$, with $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \ldots \sigma_m)$. In (58), the optimal value of τ is the worst-case regret $e(L)$. (Our notation differs from that of [15] in that we define $\mathbf{r} = \mathbf{s}/L^2$.)

We first show that (58b) is an active constraint in the optimization problem. Assume by contradiction that (58b) is inactive. Then, by the Karush–Kuhn–Tucker conditions for optimality [20, Sec. 5.5.3], (58) is equivalent to

$$\min_{\mathbf{d},\tau} \tau \quad \text{s.t.} \sum \frac{d_i^2}{\sigma_i} \le \tau \quad (60)$$

for which the optimal solution is $\mathbf{d} = \mathbf{0}, \tau = 0$. However, for any $\mathbf{r} \in \mathcal{S}$

$$F_2(\mathbf{0}, \mathbf{r}) > 0 = \tau \quad (61)$$

contradicting the fact that (58b) is inactive. Thus, for the optimal value of τ and $\mathbf{d}$, there exists at least one active $\mathbf{r} \in \mathcal{S}$ for which $F_2(\mathbf{d}, \mathbf{r}) = \tau$.

Next, define

$$g(\mathbf{d}, \mathbf{r}, L^2) \triangleq \sum (1 - d_i)^2 r_i - \frac{\sum r_i}{1 + L^2 \sum \sigma_i r_i}. \quad (62)$$

Let us study the behavior of $g(\mathbf{d}, \mathbf{r}, L^2)$ when L^2 is changed. Observe that

$$\frac{\partial g}{\partial L^2} = \frac{\sum \sigma_i r_i \sum r_i}{\left(1 + L^2 \sum \sigma_i r_i\right)^2} > 0. \quad (63)$$

Thus, $g(\mathbf{d}, \mathbf{r}, L^2)$ is strictly increasing with L. Therefore, if L is decreased, then $F_2(\mathbf{d}, \mathbf{r}) = F_1(\mathbf{d}) + L^2 g(\mathbf{d}, \mathbf{r}, L^2)$ is decreased for all active $\mathbf{r}$, and the constraint (58b) is relaxed, which implies that the optimal value of τ is also decreased. Since this value equals $e(L)$, we conclude that $e(L)$ is strictly monotonic in L, which completes the proof. ∎

In the following section, we make use of the estimators developed above in the context of a channel estimation problem.

V. Application: Channel Estimation

As an application of the MPS estimator, we now consider the problem of preamble-based channel estimation. Specifically, we seek to estimate the impulse response of an unknown channel using a training sequence (also called a preamble), which is transmitted along with payload data. The received symbols are compared to the known preamble sequence, and this information is used to obtain an estimate of the channel response. Knowledge of the channel response is required in many detection algorithms, for example, in maximum likelihood sequence estimation (MLSE) [22].

Let $\mathbf{c} = (c_0, \ldots c_{N_c-1})^T$ denote the unknown channel impulse response of known length N_c, and let

$$\mathbf{p} = (p_{-N_c+1}, p_{-N_c+2}, \ldots p_0, \ldots p_{N_p-N_c})^T \quad (64)$$

denote the known vector of preamble symbols of length N_p. The corresponding received symbols are given by

$$r_k = \sum_{l=0}^{N_c-1} c_l p_{k-l} + w_k, \quad k = 0, 1, \ldots N_p - N_c \quad (65)$$

where $\mathbf{w} = (w_0, \ldots w_{N_p - N_c})^T$ is additive white noise with variance σ_w^2. Defining

$$\mathbf{H} = \begin{bmatrix} p_0 & p_{-1} & \cdots & p_{-N_c+1} \\ p_1 & p_0 & \cdots & p_{-N_c+2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N_p-N_c} & p_{N_p-N_c-1} & \cdots & p_{N_p-1} \end{bmatrix} \quad (66)$$

we have

$$\mathbf{r} = \mathbf{Hc} + \mathbf{w}. \quad (67)$$

The classical approach to channel estimation using a preamble is least-squares estimation of the unknown, deterministic vector $\mathbf{c}$ from the measurements $\mathbf{r}$ [22]–[24]. The estimated channel in this case is

$$\hat{\mathbf{c}}_{\mathrm{LS}} = \mathbf{G}_{\mathrm{LS}}\mathbf{r} = (\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*\mathbf{r}. \quad (68)$$

This estimator minimizes the *measurement* error $\|\mathbf{r} - \mathbf{HGr}\|^2$. However, we are interested in minimizing the *estimation* error $\epsilon = E\{\|\mathbf{c} - \hat{\mathbf{c}}\|^2\}$, as the channel estimate is used for further processing (e.g., detection of payload data). For example, in [23], an increase in channel estimation error is assumed to be equivalent to an increase in noise level.

Unfortunately, the channel estimation error ϵ is a function of the unknown channel $\mathbf{c}$; therefore, direct minimization of ϵ is not possible. Were we to know that $\mathbf{c}$ lies within some bounded set $\mathcal{U}$, a minimax MSE approach would allow us to minimize the worst-case error among all possible channels within $\mathcal{U}$. However, we generally only have a vague understanding that channel dispersion is limited and that most of the energy in $\mathbf{c}$ lies in the first few components.

On the other hand, the desired channel estimation error is a parameter with known implications for the system designer. In particular, the maximum channel estimation error may be treated as an added noise source [23]. In this case, the estimation error requirement is a design parameter; it is to be chosen together with other system properties such as receiver signal to interference plus noise ratio (SINR) requirements. We can use the MPS estimator to maximize the set of channels for which a required estimation error ϵ_m is achieved, assuming that the given maximum estimation error is critical for system operation, and should be guaranteed for as wide a range of channels as possible.

Let $\mathbf{c}^0 = (1, 0, \ldots 0)^T$ be a perfect (nondispersive) channel, and let $\mathbf{c}' = \mathbf{c} - \mathbf{c}^0$. We construct a simple class of parameter sets by defining

$$\mathcal{U}_L = \{\mathbf{c}' : \|\mathbf{c}'\| \leq L\}. \quad (69)$$

This model assumes that most of the channel energy is concentrated in the first tap and that deviations from this nominal value are fairly uniform among the channel taps. More elaborate models may be constructed if additional information about the channel properties is known.

We seek an estimator guaranteeing estimation error of ϵ_m or less for as large a parameter set as possible. From (67), we have

$$\mathbf{r} - \mathbf{Hc}^0 = \mathbf{Hc}' + \mathbf{w}. \quad (70)$$

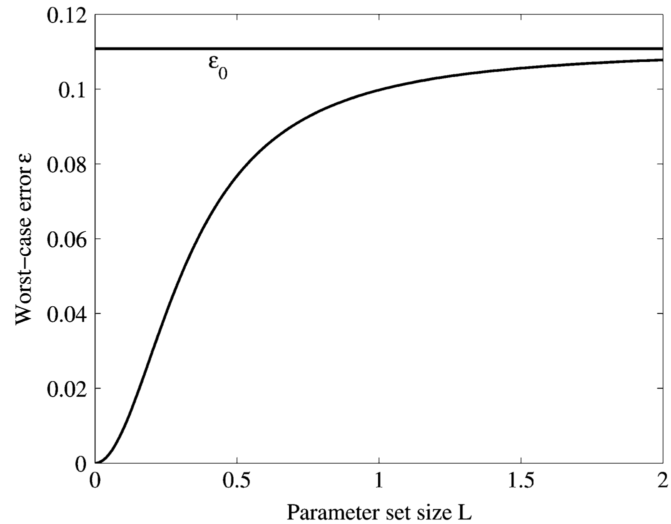

Fig. 1. Worst-case error of various minimax MSE channel estimators.

By Theorem 4, the maximum error ϵ_m must first be compared with $\epsilon_0 = \mathrm{Tr}((\mathbf{H}^*\mathbf{H})^{-1})$, which is the MSE of the LS estimator. If $\epsilon_m \geq \epsilon_0$, then an error of ϵ_0 is allowable. Such an error is guaranteed by the LS estimator for *any* value of $\mathbf{c}$ so that the LS estimator has infinite parameter robustness in this case. However, if $\epsilon_m < \epsilon_0$, then an MPS estimator is given by

$$\hat{\mathbf{c}}'_{\mathrm{PS}} = \frac{\epsilon_m}{\epsilon_0}(\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*(\mathbf{r} - \mathbf{Hc}^0) \quad (71)$$

and thus

$$\hat{\mathbf{c}}_{\mathrm{PS}} = \frac{\epsilon_m}{\epsilon_0}(\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*\mathbf{r} + \left(1 - \frac{\epsilon_m}{\epsilon_0}\right)\mathbf{c}^0. \quad (72)$$

To compare the performance of the LS and MPS channel estimators, we consider the problem of estimating a seven-tap channel using the optimal 14-symbol binary phase shift keying (BPSK) preamble suggested in [23]. We assume that the noise variance is $\sigma_w^2 = 0.1$. The worst-case error of various minimax MSE estimators is given by (33) and plotted in Fig. 1. By Theorem 4, all of these estimators are also MPS estimators. An engineer constructing a channel estimation system should use such a plot as a design tool, as it demonstrates the tradeoff between channel estimation error and the range of channels for which the error can be achieved.

Suppose we choose to design our system such that a channel estimation error of $\epsilon_m = (2/3)\epsilon_0$ is to be tolerated; this choice covers a reasonably sized parameter set while substantially reducing the estimation error. We note that the choice of ϵ_m is accompanied by appropriate design steps, which will allow the receiver to handle the resulting estimation errors (for example, error correction capabilities suitable for such noise levels). The MSE (55) of the resulting MPS estimator is compared with the MSE ϵ_0 of the LS estimator in Fig. 2.

To verify that the reduced estimation error resulting in improved detection performance, a BPSK detection scenario was simulated [25]. A signal containing the 14 preamble symbols and 100 random data symbols was generated. Channels were simulated by choosing each tap c_i $(1 \leq i \leq 7)$ to be an independent Rayleigh-distributed variate with parameter $A\beta^i$, where

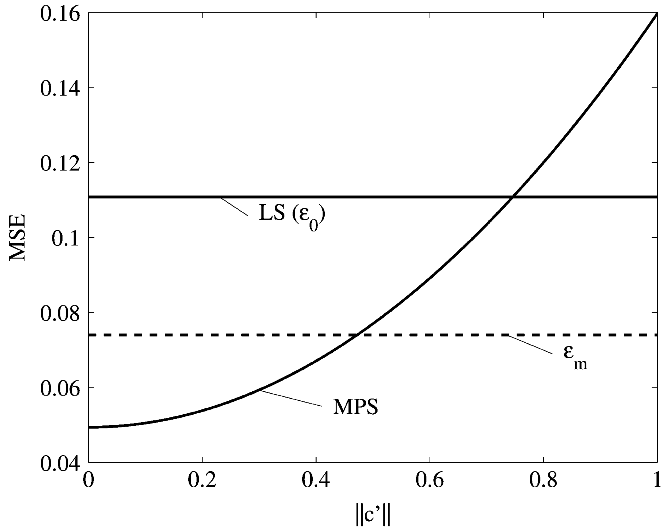

Fig. 2. Channel estimation error of MPS and LS estimators for various channels.

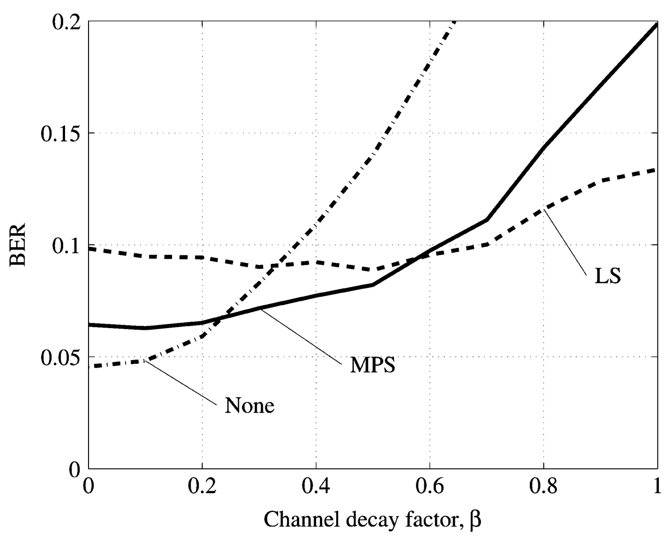

Fig. 3. BER for various channels with the LS and MPS channel estimators.

$0 \le \beta \le 1$ is the channel dispersion factor, and A is chosen so that $E\{\|\mathbf{c}\|^2\} = 1$. Thus, $\beta = 0$ results in a nondispersive channel, whereas $\beta = 1$ indicates a channel for which the taps are identically distributed (maximum channel dispersion). The channel was estimated using both the LS and MPS estimators described above, and the resulting channel estimate was used for MLSE detection of the data symbols. The simulation was repeated to obtain an estimate of the bit error rate (BER). The results are presented in Fig. 3. For comparison, a null estimator is also plotted; this "estimator" assumes a nondispersive channel, i.e., $\hat{\mathbf{c}} = \mathbf{c}^0$.

The MPS estimator is a compromise between the LS estimator and the null estimator: The LS estimator has modest estimation error requirements but achieves them for all values of $\mathbf{c}$; the null estimator can be viewed as an estimator requiring zero estimation error and achieves this requirement only for the nominal channel $\mathbf{c}^0$. MPS estimators provide a continuum of choices between these two extremes, allowing the designer to choose a point in the tradeoff between the estimation error requirement and the size of the parameter set for which the requirement is achieved. An appropriate choice of ϵ_m leads to an estimator that considerably outperforms the LS estimator for low- and moderate-dispersion channels and fails only when channel taps are nearly identically distributed.

VI. Maximum Noise Level Estimation

Throughout the paper, we have assumed that the noise covariance $E\{\mathbf{w}\mathbf{w}^*\}$ is known. In practice, this is rarely the case, and the covariance must itself often be estimated from measurements. In this section, we consider the case

$$E\{\mathbf{w}\mathbf{w}^*\} = \sigma^2 \mathbf{C_w} \tag{73}$$

for some unknown deterministic *noise level* σ^2 and some known covariance matrix $\mathbf{C_w}$ [26]. For example, when the noise is i.i.d., $\mathbf{C_w} = \mathbf{I}$, and σ^2 is the noise variance. The estimation techniques used so far require complete knowledge of the noise covariance. Thus, minimax or MPS approaches cannot be directly applied to this problem unless the noise parameters are estimated from the measurements; this increases computational complexity and may be unreliable in some situations.

As an alternative approach, we propose to estimate $\mathbf{x}$ from the observations $\mathbf{y} = \mathbf{Hx} + \mathbf{w}$ while guaranteeing maximum error requirements for as large a range of noise levels as possible. To this end, we assume that $\mathbf{x} \in \mathcal{U}$ for a known parameter set $\mathcal{U}$ and require a maximum error level ϵ_m. We seek the estimator that guarantees an error not exceeding ϵ_m for all $\mathbf{x} \in \mathcal{U}$, and for as large a noise level σ^2 as possible; this will be referred to as the *maximum noise level* (MNL) estimator. As we will show, the MNL estimator is related to the minimax estimator, allowing us to efficiently find the MNL estimator whenever the minimax estimator is known.

Formally, we define an error function $\epsilon_\sigma^2(\hat{\mathbf{x}}, \mathbf{x})$, such as the MSE or the regret, and require some level of performance $\epsilon_\sigma^2(\hat{\mathbf{x}}, \mathbf{x}) \le \epsilon_m$ to be satisfied over the entire range $\mathbf{x} \in \mathcal{U}$. We can now define a new type of maximum set estimator in a manner analogous to the definition of the MPS in Section III-A, as follows.

Definition 5: The *noise robustness* $\hat{\sigma}^2$ of an estimator $\hat{\mathbf{x}}$ is defined as the maximum σ^2 for which the performance requirement is satisfied:

$$\hat{\sigma}^2(\hat{\mathbf{x}}) = \max\left\{\sigma^2 : \max_{\mathbf{x}\in\mathcal{U}} \epsilon_\sigma^2(\hat{\mathbf{x}}, \mathbf{x}) \le \epsilon_m\right\}. \tag{74}$$

Definition 6: The MNL *estimator* $\hat{\mathbf{x}}_{\rm NL}$ (among a class of estimators $\mathcal{E}$) is the estimator maximizing the noise robustness among all estimators in $\mathcal{E}$, for given $\mathcal{U}$, $\epsilon_\sigma^2(\hat{\mathbf{x}}, \mathbf{x})$, and ϵ_m:

$$\hat{\mathbf{x}}_{\rm NL} = \arg\max_{\hat{\mathbf{x}}\in\mathcal{E}} \hat{\sigma}^2(\hat{\mathbf{x}}). \tag{75}$$

We now show that if the error function ϵ_σ^2 is continuous in σ^2, then the MNL estimator is a minimax estimator. The error function is indeed continuous for many cases of interest, such as the MSE and the regret.

Theorem 6: Suppose the error function ϵ of interest is continuous in σ^2. Then, the MNL estimator $\hat{\mathbf{x}}_{\rm NL}$ is a minimax estimator for the parameter set $\mathcal{U}$, with noise level $\sigma_1^2 = \hat{\sigma}^2(\hat{\mathbf{x}}_{\rm NL})$.

Proof: Assume by contradiction that $\hat{\mathbf{x}}_{\rm NL}$ is not a minimax estimator. Then, there exists $\hat{\mathbf{x}}_{\rm M} \neq \hat{\mathbf{x}}_{\rm NL}$ such that

$$\max_{\mathbf{x}\in\mathcal{U}} \epsilon_{\sigma_1^2}(\hat{\mathbf{x}}_{\rm M}, \hat{\mathbf{x}}) < \max_{\mathbf{x}\in\mathcal{U}} \epsilon_{\sigma_1^2}(\hat{\mathbf{x}}_{\rm NL}, \mathbf{x}) \leq \epsilon_m. \tag{76}$$

However, since ϵ_σ^2 is continuous in σ^2, a sufficiently small change in σ^2 causes an arbitrarily small change in ϵ_σ^2. Thus, there exists $\sigma_2^2 > \sigma_1^2$ such that

$$\max_{\mathbf{x}\in\mathcal{U}} \epsilon_{\sigma_2^2}(\hat{\mathbf{x}}_{\rm M}, \hat{\mathbf{x}}) \leq \epsilon_m. \tag{77}$$

Hence, $\hat{\sigma}^2(\hat{\mathbf{x}}_{\rm M}) \geq \sigma_2^2 > \sigma_1^2 = \hat{\sigma}^2(\hat{\mathbf{x}}_{\rm NL})$, contradicting the fact that $\hat{\mathbf{x}}_{\rm NL}$ is an MNL estimator. ■

A consequence of this theorem is that an MNL estimator can be found if an algorithm for finding a minimax estimator is known. This can be performed efficiently using a line search, in which minimax estimators are calculated for various noise levels, until a minimax estimator whose worst-case error equals ϵ_m is found. Alternatively, as the following theorem demonstrates, a closed form for a linear MNL estimator can be identified when a closed form for the minimax estimator is known.

Theorem 7: Let $\mathcal{U} = \{\mathbf{x} : \|\mathbf{x}\| \leq L\}$, and let $\epsilon_\sigma^2(\hat{\mathbf{x}}, \mathbf{x})$ be the MSE. For a given maximum error ϵ_m, a linear MNL estimator is given by

$$\hat{\mathbf{x}}_{\rm NL} = \begin{cases} \frac{L^2-\epsilon_m}{L^2}\hat{\mathbf{x}}_{\rm LS}, & L^2 > \epsilon_m \\ \mathbf{0}, & L^2 \leq \epsilon_m \end{cases} \tag{78}$$

where $\hat{\mathbf{x}}_{\rm LS}$ is the LS estimator (5).

Proof: We first consider the case $L^2 \leq \epsilon_m$. In this case, the performance requirements are extremely lax, and many estimators satisfy these requirements for *any* noise level. In particular, the estimator $\hat{\mathbf{x}} = \mathbf{0}$ has an MSE of $\|\mathbf{x}\|^2$, for which the worst case is $\max \|\mathbf{x}\|^2 = L^2 \leq \epsilon_m$; this is true regardless of the noise level. Thus, $\hat{\mathbf{x}} = \mathbf{0}$ is an MNL estimator (with infinite noise robustness) for the trivial case $L^2 \leq \epsilon_m$.

We now turn to the more interesting case $L^2 > \epsilon_m$. By Theorem 6, $\hat{\mathbf{x}}_{\rm NL}$ is a minimax estimator for some noise level σ^2. The minimax estimator for the parameter set $\mathcal{U}$ and for a given noise level σ^2 is given by [14]

$$\hat{\mathbf{x}}_{\rm M}(\sigma^2) = \frac{L^2}{L^2+\sigma^2\epsilon_0}\left(\mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{y} \tag{79}$$

where $\epsilon_0 = \mathrm{Tr}((\mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{H})^{-1})$ is the MSE of the LS estimator with $\sigma^2 = 1$. As we have seen in (35) of Lemma 1, the worst-case error for this estimator within the set $\mathcal{U}$ is given by

$$\max_{\mathbf{x}\in\mathcal{U}} \epsilon_\sigma^2(\hat{\mathbf{x}}_{\rm M}(\sigma^2), \mathbf{x}) = \frac{L^2\sigma^2\epsilon_0}{L^2+\sigma^2\epsilon_0}. \tag{80}$$

The critical value of σ^2 for which this value exactly equals ϵ_m is given by

$$\sigma^2 = \frac{\epsilon_m L^2}{L^2-\epsilon_m}\frac{1}{\mathrm{Tr}\left(\left(\mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{H}\right)^{-1}\right)}. \tag{81}$$

Substituting this value of σ^2 into (79) yields the required estimator (78). ■

It is instructive to compare the closed forms obtained for the MPS estimator [Theorem 4(b)] and the MNL estimator (Theorem 7) when spherical parameter sets are used. Both estimators take the form of a linear minimax MSE estimator for a spherical parameter set, and hence, they are shrunken least-squares estimators [7]. They can thus be viewed as a compromise between the least-squares estimator and the zero estimator. However, for the MPS estimator, the shrinkage factor increases with the maximum allowed error ϵ_m, whereas for the MNL estimator, the shrinkage factor decreases with ϵ_m. The reason for this is as follows. When the allowed error is increased, an increase in either the parameter set or noise level is allowed. However, a larger parameter set is achieved by an estimator closer to the LS estimator (which provides constant error for all $\mathbf{x}$), whereas a larger noise level is achieved by an estimator closer to the zero estimator (which provides zero error, regardless of noise level, for the nominal value $\mathbf{x} = \mathbf{0}$). Thus, increasing the maximum allowed error has opposite effects, depending on whether the goal is to increase the robustness to uncertainty in the parameter set or in the noise level.

VII. Discussion

In this paper, we considered the problem of parameter estimation given a maximum allowed estimation error. This is appropriate for systems designed to function with a known and tolerable error margin, such as communication systems designed for a certain SNR level. We have developed estimators that guarantee the required estimation error for as wide a range of operating conditions as possible. The goal of this paper has been to show that estimators that make use of given estimation error requirements outperform classical approaches such as the LS estimator.

The maximum set estimation concept was first applied to find the largest parameter set $\mathcal{U}_L$ such that performance is guaranteed for any parameter $\mathbf{x}$ in $\mathcal{U}_L$. This results in the MPS estimator. Next, the MNL estimator was developed; this estimator maximizes the range of noise variances for which the required estimation error is guaranteed.

As we have seen, in many cases, the maximum set estimator is equivalent in form to a matching minimax estimator: The maximum set estimator for a given error requirement ϵ_m equals a minimax estimator whose worst-case error is ϵ_m. However, while minimax estimators assume a given bound on the parameter set, maximum set estimators assume a requirement on the obtained estimation error. Thus, these estimators are used under different circumstances, and their similarity in form merely serves as a mathematical tool for finding maximum set estimators based on known results for minimax estimators.

The maximum allowed error is often a function of system design parameters and can be influenced by design decisions. In such cases, a plot of the worst-case error as a function of the size of the parameter set (as in Fig. 1) can be used as a design tool. Such a plot can be interpreted in two complementary ways. It describes the worst-case error obtained if a minimax estimator is used with a given parameter set bound. However, it also defines the size of the parameter set obtained if an MPS estimator is used with a given maximum error. Thus, such a plot can be used to select a meaningful value for the maximum error, based on the tradeoff between estimation error and parameter set bound.

The choice of an appropriate estimator for a given problem depends on the data available to the designer. Knowledge of the second-order statistics of the parameters $\mathbf{x}$, for example, leads to the well-known Wiener estimator, which is optimal in an MSE sense. However, partial information can also be used to improve estimation performance. The maximum allowed estimation error is an example of added information, which may be known to the designer and, as we have demonstrated, can often result in improved performance.

Acknowledgment

The authors are deeply grateful to Prof. Y. Ben-Haim for many helpful comments and suggestions provided throughout the work on this paper.

References

[1] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, Second ed. New York: Springer, 1998.
[2] N. Wiener, *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. New York: MIT Press, 1949.
[3] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, Third ed. New York: McGraw-Hill, 1991.
[4] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[5] A. N. Tichonov 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, 1970.
[7] L. S. Mayer and T. A. Wilke, "On biased estimation in linear models," *Technometr.*, vol. 15, pp. 497–508, 1973.
[8] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Statist.*, vol. 35, no. 1, pp. 73–101, Mar. 1964.
[9] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proc. IEEE*, vol. 73, pp. 433–481, 1985.
[10] S. Verdú and H. V. Poor, "On minimax robustness: A general approach and applications," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 2, pp. 328–340, Mar. 1984.
[11] 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.
[12] ——, "Minimax MSE-ratio estimation with signal covariance uncertainties," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1335–1347, Apr. 2005.
[13] M. S. Pinsker, "Optimal filtering of square-integrable signals in Gaussian noise," *Problems Inf. Transmission*, vol. 16, pp. 120–133, 1980.
[14] 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.
[15] ——, "Linear minimax regret estimation of deterministic parameters with bounded data uncertainties," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2177–2188, Aug. 2004.
[16] Y. Ben-Haim, *Information-Gap Decision Theory: Decisions Under Severe Uncertainty*. San Diego, CA: Academic, 2001.
[17] ——, "Set-models of information-gap uncertainty: Axioms and an inference scheme," *J. Franklin Inst.*, vol. 336, pp. 171–199, 2000.
[18] Z. Ben-Haim and Y. C. Eldar, "Estimation with maximum error requirements," in *Proc. IEEE Conv. Elect. Electron. Eng. Israel*, Tel-Aviv, Israel, Sep. 2004, pp. 416–419.
[19] A. Ben-Tal and A. Nemirovski, "Robust optimization: Methodology and applications," *Math. Program. B*, vol. 92, pp. 453–480, 2002.
[20] S. Boyd and L. Vanderberghe, *Convex Optimization*, Cambridge, U.K.: Cambridge Univ. Press, 2005. [Online] Available: http://www.stanford.edu/boyd/cvxbook.html.
[21] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[22] T. S. Rappaport, *Wireless Communications: Principles and Practice*, Second ed. Upper Saddle River, NJ: Prentice-Hall, 1996.
[23] S. N. Crozier, D. D. Falconer, and S. A. Mahmoud, "Least sum of square errors (LSSE) channel estimation," in *Proc. Inst. Elect. Eng. F*, vol. 138, Aug. 1991, pp. 371–378.
[24] C. Fragouli, N. Al-Dhahir, and W. Turin, "Training-based channel estimation for multiple-antenna broadband transmissions," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 384–391, Mar. 2003.
[25] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems*, Second ed. New York: Plenum, 1992.
[26] A. Beck, A. Ben-Tal, and Y. C. Eldar, "Robust mean-squared error estimation of multiple signals in linear systems affected by model and noise uncertainties," *Math. Programm.*, to be published.

Zvika Ben-Haim (S'05) received the B.Sc. degree in electrical engineering and the B.A. degree in physics, both *summa cum laude*, from the Technion—Israel Institute of Technology, Haifa, Israel, in 2000. He is currently pursuing the M.Sc. degree with the Department of Electrical Engineering at the Technion.

From 1996 to 2000, he was a member of the Technion Program for Exceptionally Gifted Students. His research interests include theoretical aspects of signal processing, estimation theory, and statistical signal processing.

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

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

Dr. Eldar was in the program for outstanding students at TAU from 1992 to 1996. In 1998, she held the Rosenblith Fellowship for study in Electrical Engineering at MIT, and in 2000, she held an IBM Research Fellowship. She is currently a Horev Fellow of the Leaders in Science and Technology program at the Technion as well as an Alon Fellow. In 2004, she was awarded the Wolf Foundaiton 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.