CCIT Report #486 May 2004

Unbounded Uncertainty Estimation

Zvika Ben-Haim* Yonina C. Eldar†

Department of Electrical Engineering

Technion—Israel Institute of Technology

Haifa 32000, Israel

May 30, 2004

Abstract

We consider the problem of estimating a deterministic parameter vector $\mathbf{x}$ from observations $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where $\mathbf{H}$ is a known linear transformation and $\mathbf{w}$ is additive noise. Although the least-squares (LS) estimator is often used in such estimation problems, it does not necessarily minimize the mean-squared error (MSE) between $\mathbf{x}$ and its estimate $\hat{\mathbf{x}}$. In fact, with few additional assumptions, linear estimators can be constructed which outperform the LS estimator. For example, we show that if the parameter vector $\mathbf{x}$ is known to lie within some bounded parameter set $\mathcal{U}$, then a linear *minimax* estimator exists which has lower MSE than the LS estimator for *any* $\mathbf{x} \in \mathcal{U}$.

The minimax approach, in which the estimator is chosen to minimize the worst-case error, is suitable when data such as the parameter set bound and noise level are available. If this information is unavailable, then it may be more appropriate to seek the estimator which guarantees a required maximum error for as wide a range of conditions as possible. We refer to this approach as *unbounded uncertainty estimation,* and develop two types of estimators based on this criterion: estimators guaranteeing the required error for as large a parameter set as possible, and for as large a noise level as possible. We show a relation between each of these estimators and the minimax estimator, which allows us to efficiently compute many types of unbounded uncertainty estimators and, in some cases, to obtain an analytical expression for the estimators. We then demonstrate the use of the unbounded uncertainty estimator in a channel estimation application, in which an unknown channel is estimated using a preamble sequence, and show that unbounded uncertainty estimation achieves a lower bit error rate (BER) than classical LS estimation.

EDICS Keyword: 2-ESTM — Estimation Theory and Applications

*Corresponding author. Phone: +972-3-741-5291. Fax: +972-4-829-5757. E-mail: zvikabh@technion.ac.il

†Phone: +972-4-829-3256. Fax: +972-4-829-5757. E-mail: yonina@ee.technion.ac.il

1 Introduction

The problem of estimating an unknown parameter vector $\mathbf{x}$ based on noisy measurements $\mathbf{y}$ is a fundamental problem in science and engineering. A common formulation of this problem assumes that $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$ is a known linear transformation $\mathbf{H}$ of $\mathbf{x}$ with additive zero-mean random noise $\mathbf{w}$, whose covariance matrix is known, possibly up to a scaling factor. 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 a linear estimator $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ that minimizes the mean-squared error (MSE) $E\|\mathbf{x} - \hat{\mathbf{x}}\|^2$.

In some cases, the parameter vector $\mathbf{x}$ is random with known second-order statistics. Under these circumstances the well-known Wiener filter [23, 18] minimizes the MSE among all linear estimators. However, in many other cases the parameter vector $\mathbf{x}$ must be treated as a deterministic vector, either because it is inherently deterministic or because it is random with unknown statistics. The assumption of a deterministic parameter vector will be adopted 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 $\hat{\mathbf{x}}$. 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, i.e., linear estimators whose bias is zero. 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 [15]. The LS estimator has the additional property that it minimizes the measurement error $E\|\mathbf{y} - \hat{\mathbf{y}}\|^2$, 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 in Section 4.1 that for any bounded set $\mathcal{U}$, a biased estimator exists whose MSE is lower than the MSE of the least-squares 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 [21] (also known as ridge regression [9]) and the shrunken estimator [17]. However, like the LS estimator, these estimators are designed to minimize the measurement error, rather than the estimation error.

Design of an estimator requires knowledge of various *system properties,* such as the transformation matrix $\mathbf{H}$ and the noise covariance $\mathbf{C}_{\mathbf{w}}$. When this information is lacking, one approach is to minimize the maximum (worst-case) estimation error among all possible values of the system property. This *minimax* approach was first introduced for dealing with uncertain noise statistics [11, 12], and has since been applied in a variety of estimation problems [13, 14, 22]. 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 $\mathcal{U}$ [19, 6]. 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* [7], 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. For example, 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. Alternatively, the parameter set may be estimated from the measurements $\mathbf{y}$, but this results in a nonlinear estimator whose computational complexity is higher.

In some estimation problems, requirements on the maximum estimation error are more readily available than other system properties. For example, in communication systems, a minimum SNR may be required for data transmission to be possible, while bounds on the estimated parameters may not be known. For such cases, following the philosophy of information-gap decision theory [1, 2], we propose an *unbounded uncertainty estimation* approach; in general terms, this approach designs an estimator to guarantee the required error for the widest range of conditions possible. The unbounded uncertainty estimation strategy can be applied in several ways, depending on the uncertain system property. In Sections 2–5, we discuss the case in which the parameter set is uncertain, and describe the *unbounded parameter set (UPS) estimator* which maximizes the parameter set for which error requirements are maintained. In Section 6 we deal with the case in which the noise level is uncertain, and provide an *unbounded noise level (UNL) estimator* which maximizes the noise level for a given maximum error and a given parameter set.

The UPS estimator is more accurately defined as follows. For a given estimator, we find the largest set of $\mathbf{x}$'s for which performance is satisfactory, i.e., for which the estimation error does not exceed the maximum allowed error. A measure of the size of this set is called the *parameter robustness* of the estimator. For instance, the parameter robustness can be defined as the largest L for which satisfactory performance is guaranteed for all $\mathbf{x}$ such that $\|\mathbf{x}\| \leq L$. The UPS estimator is defined as the estimator which maximizes the parameter robustness. In other words, this estimator provides satisfactory performance over as large a range of vectors $\mathbf{x}$ as possible; and yet, it does not require choosing an arbitrary bound on the range of values of $\mathbf{x}$. In Section 2, the UPS estimator is presented for the most common case in which the MSE is minimized over an ellipsoidal class of parameter sets. A more general definition of the UPS estimator is given in Section 3.1.

The design criteria of UPS estimators are inherently different from those of their bounded (minimax) counterparts. Nevertheless, we show in Section 3.2 that in many cases there exists a mathematical equivalence between the two approaches. This equivalence does not imply that the methods are redundant, as they are designed to be used for different types of estimation problems. However, as we show in Section 4, the mathematical relation allows us to apply known properties of the minimax estimator to the UPS estimator, and thus to efficiently calculate UPS estimators — in some cases reaching a closed form of the estimator.

In Section 5, our results are demonstrated for a typical application of estimation theory: the problem of channel estimation using a training sequence. We show that UPS estimation substantially reduces estimation error compared to the classical maximum-likelihood estimator, without requiring additional knowledge of channel statistics.

Applying the concept of unbounded uncertainty estimation in a different setting, in Section 6 we consider the estimation problem when the noise covariance is known up to a constant, i.e., $E\mathbf{w}\mathbf{w}^* = \sigma^2\mathbf{C}_\mathbf{w}$, where σ^2 is unknown [16]. In this case, we assume that $\mathbf{x}$ lies in a *known* parameter set $\mathcal{U}$, and find an *unbounded noise level (UNL) estimator* which guarantees a required estimation error for as large a range of noise levels σ^2 as possible. Here again, we show that the obtained estimator generally has the same form as the minimax estimator, and derive a closed form for the estimator obtained when the error function is the MSE and the parameter set $\mathcal{U}$ is spherical.

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 $\mathbf{P} - \mathbf{Q} \succeq 0$.

2 Unbounded Parameter Set Estimation: A Useful Special Case

To demonstrate the main ideas of this paper, we begin by presenting an important special case of the unbounded parameter set (UPS) estimator. This example is generalized and formalized in Section 3.

Consider the system of measurements $\mathbf{y}$,

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

where $\mathbf{x}$ is an unknown deterministic vector, $\mathbf{H}$ 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 $\hat{\mathbf{x}} = \mathbf{Gy}$ 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 $E\|\hat{\mathbf{x}} - \mathbf{x}\|^2$ as the estimation error function; a general discussion follows in Section 3, in which any estimation error function $\epsilon(\hat{\mathbf{x}}, \mathbf{x})$ may be used.

The MSE is equal to the sum of the variance of $\hat{\mathbf{x}}$ and the squared norm of the bias of $\hat{\mathbf{x}}$ [15],

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

where

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

and

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

Since the bias $\mathbf{b}(\hat{\mathbf{x}})$ 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

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

The MSE of the LS estimator is

$$\gamma_0 = \mathrm{Tr}\big((\mathbf{H}^*\mathbf{C_w}^{-1}\mathbf{H})^{-1}\big). \tag{6}$$

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, attempting to deal with *any* value of $\mathbf{x}$ requires the use of unbiased estimators. 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 bounded parameter set $\mathcal{U}$, then the estimator minimizing the worst-case MSE among all values of $\mathbf{x}$ in $\mathcal{U}$ can be determined. This is the minimax MSE estimator [19, 6], given by

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

Many possibilities for defining the parameter set $\mathcal{U}$ exist [3]. One commonly used set is the ellipsoid

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

where $L > 0$ is a known constant and

$$\|\mathbf{x}\|_{\mathbf{T}}^2 = \mathbf{x}^*\mathbf{T}\mathbf{x} \tag{9}$$

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.

Since the parameter L describes the unknown vector $\mathbf{x}$, a suitable value of L is often difficult to determine. Indeed, even if a small amount of information about $\mathbf{x}$ is available, such as several past measurements, then these usually characterize typical values of $\mathbf{x}$, while 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 a UPS 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,

$$\hat{L}(\hat{\mathbf{x}}) = \max\{L : \; E\|\mathbf{x} - \hat{\mathbf{x}}\|^2 \leq \epsilon_m \quad \forall \mathbf{x}^*\mathbf{T}\mathbf{x} \leq L\}. \tag{10}$$

The UPS estimator is designed to maximize the robustness, i.e.,

$$\hat{\mathbf{x}}_{\mathrm{UP}} = \arg\max_{\hat{\mathbf{x}}} \hat{L}(\hat{\mathbf{x}}). \tag{11}$$

Suppose we wish to find the UPS estimator for maximum error ϵ_m equal to γ_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 γ_0 or greater. This implies that requiring a maximum error of γ_0 (or greater) yields the LS estimator as a UPS estimator. More interesting is the case $\epsilon_m < \gamma_0$, for which the LS estimator no longer achieves the required error, regardless of the value of $\mathbf{x}$. A UPS estimator $\hat{\mathbf{x}}$ for a given error level $\epsilon_m < \gamma_0$ 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, the UPS 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 the linear UPS estimator of (11) and (10) 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 $\{\mathbf{z} : \; f(\mathbf{z}) \leq \alpha\}$ are all convex. Quasiconvex problems can be efficiently solved, for example, using bisection [4]. In addition, as we shall see in Section 4.1, in many special cases a closed form for the UPS estimator can be obtained, by exploring its relation to the minimax MSE estimator.

Proposition 1. *The linear unbounded parameter set (UPS) estimator $\hat{\mathbf{x}}_{\mathrm{UP}} = \mathbf{G}\mathbf{y}$ defined by (11) and (10) can be found* *by solving the following quasiconvex optimization problem:*

$$\min_{\mathbf{G},\lambda,y} y/\lambda \tag{12}$$

$$\text{s.t.} \begin{cases} \begin{bmatrix} y+\epsilon_m & \mathbf{g}^* \\ \mathbf{g} & \mathbf{I} \end{bmatrix} \succeq 0 \\ \begin{bmatrix} \lambda\mathbf{I} & \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{GH})^* \\ (\mathbf{I}-\mathbf{GH})\mathbf{T}^{-1/2} & \mathbf{I} \end{bmatrix} \succeq 0 \end{cases}$$

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 a solution to (11) with $\hat{L}(\hat{\mathbf{x}})$ defined by (10), which is equivalent to solving the optimization problem

$$\begin{aligned} &\max_{\mathbf{G},L^2} L^2 \\ &\text{s.t.} \left(\max_{\|\mathbf{x}\|_{\mathbf{T}}^2 \le L^2} E\|\mathbf{x}-\hat{\mathbf{x}}\|^2 \right) \le \epsilon_m, \end{aligned} \tag{13}$$

where $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y} = \mathbf{G}(\mathbf{H}\mathbf{x}+\mathbf{w})$. Using (2)–(4), the MSE is given by

$$\epsilon = E\|\mathbf{x}-\hat{\mathbf{x}}\|^2 = \mathrm{Tr}(\mathbf{G}\mathbf{C}_{\mathbf{w}}\mathbf{G}^*) + \mathbf{x}^*(\mathbf{I}-\mathbf{GH})^*(\mathbf{I}-\mathbf{GH})\mathbf{x}. \tag{14}$$

Thus, for a given estimator $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$, we have

$$\max_{\|\mathbf{x}\|_{\mathbf{T}} \le L} E\|\mathbf{x}-\hat{\mathbf{x}}\|^2 = \mathrm{Tr}(\mathbf{G}\mathbf{C}_{\mathbf{w}}\mathbf{G}^*) + \max_{\|\mathbf{x}\|_{\mathbf{T}} \le L} \mathbf{x}^*(\mathbf{I}-\mathbf{GH})^*(\mathbf{I}-\mathbf{GH})\mathbf{x}. \tag{15}$$

However,

$$\begin{aligned} \max_{\|\mathbf{x}\|_{\mathbf{T}} \le L} \mathbf{x}^*(\mathbf{I}-\mathbf{GH})^*(\mathbf{I}-\mathbf{GH})\mathbf{x} &= \max_{\mathbf{z}^*\mathbf{z} \le L^2} \mathbf{z}^*\mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{GH})^*(\mathbf{I}-\mathbf{GH})\mathbf{T}^{-1/2}\mathbf{z} \\ &= \lambda_{\max} L^2, \end{aligned} \tag{16}$$

where $\lambda_{\max}$ is the maximum eigenvalue of $\mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{GH})^*(\mathbf{I}-\mathbf{GH})\mathbf{T}^{-1/2}$, and is the solution to the semidefinite problem

$$\begin{aligned} &\min_{\lambda} \lambda \\ &\text{s.t. } \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{GH})^*(\mathbf{I}-\mathbf{GH})\mathbf{T}^{-1/2} \preceq \lambda\mathbf{I}. \end{aligned} \tag{17}$$

Consider the problem

$$\max_{\mathbf{G},\lambda,L^2} L^2 \tag{18}$$
$$\text{s.t.} \begin{cases} \mathrm{Tr}(\mathbf{G}\mathbf{C}_{\mathbf{w}}\mathbf{G}^*) + \lambda L^2 \le \epsilon_m & \text{(a)} \\ \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{G}\mathbf{H})^*(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{T}^{-1/2} \preceq \lambda\mathbf{I}. & \text{(b)} \end{cases}$$

We claim that the optimal solution to this problem always has $\lambda = \lambda_{\max}$. Suppose this were not the case, and $\lambda > \lambda_{\max}$ for the optimal solution. Then, λ can be decreased while still maintaining (18b). As a result, (18a) is no longer tight, so that L^2 can be increased, contradicting the assumption that λ was the optimal solution. Thus, the optimal solution for (18) always has $\lambda = \lambda_{\max}$, and therefore, by (15) and (16), the optimal solution of (18) satisfies

$$\max_{\|\mathbf{x}\|_{\mathbf{T}}^2 \le L^2} E\|\mathbf{x}-\hat{\mathbf{x}}\|^2 = \mathrm{Tr}(\mathbf{G}\mathbf{C}_{\mathbf{w}}\mathbf{G}^*) + \lambda L^2, \tag{19}$$

so that (18) and (13) 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 [10, p. 472], it is shown in [6] that (18a) and (18b) are equivalent to the following matrix inequalities:

$$\begin{bmatrix} \epsilon_m - \lambda L^2 & \mathbf{g}^* \\ \mathbf{g} & \mathbf{I} \end{bmatrix} \succeq 0, \tag{20a}$$

$$\begin{bmatrix} \lambda\mathbf{I} & \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{G}\mathbf{H})^* \\ (\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{T}^{-1/2} & \mathbf{I} \end{bmatrix} \succeq 0. \tag{20b}$$

Defining $r = -L^2$, (20a) becomes

$$r\begin{bmatrix} \lambda & 0 \\ 0 & 0 \end{bmatrix} \succeq -\begin{bmatrix} \epsilon_m & \mathbf{g}^* \\ \mathbf{g} & \mathbf{I} \end{bmatrix}. \tag{21}$$

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

$$\min_{\mathbf{G},\lambda,r,y} r \tag{22}$$
$$\text{s.t.} \begin{cases} r\lambda \ge y \\ \begin{bmatrix} y & 0 \\ 0 & 0 \end{bmatrix} \succeq -\begin{bmatrix} \epsilon_m & \mathbf{g}^* \\ \mathbf{g} & \mathbf{I} \end{bmatrix} \\ \begin{bmatrix} \lambda\mathbf{I} & \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{G}\mathbf{H})^* \\ (\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{T}^{-1/2} & \mathbf{I} \end{bmatrix} \succeq 0. \end{cases}$$

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 (12). The objective function of (12) is quasiconvex, and all its constraints are convex, so that this is a quasiconvex optimization problem [4]. □

In the next section we generalize the discussion to UPS estimators which optimize various error functions over different parameter sets. We also demonstrate a relation between UPS estimation and minimax estimation, which provides further insight into the idea of UPS estimation and yields an alternative method for finding UPS estimators. In particular, this leads to a closed form for the UPS estimator of (11) and (10) when the weighting matrix $\mathbf{T}$ commutes with $\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}$, which occurs, for example, when $\mathbf{T} = \mathbf{I}$.

3 General Form of Unbounded Parameter Set Estimators

The example presented in Section 2 is a special case of a UPS estimator, which can be generalized to include different error functions and parameter sets. In Section 3.1, we provide definitions which construct the general form of the UPS estimator. In Section 3.2, we use these definitions to prove a useful relation between bounded and unbounded parameter set estimators. This relation will be used in Section 4 to find efficient algorithms for identifying UPS estimators, and in some cases, allows us to derive a closed form for the UPS estimator.

3.1 Definitions

Let $\mathbf{x}$ be an unknown vector in $\mathbb{C}^n$ and let $\mathbf{w}$ be a zero-mean random vector in $\mathbb{C}^m$ whose covariance is known. Suppose $\mathbf{H}$ is a known full-rank $m \times n$ matrix, and let $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$. An estimator $\hat{\mathbf{x}}$ is a function of $\mathbf{y}$ which returns an n-vector close to $\mathbf{x}$ in some sense. Using this notation, the following set of definitions constructs the UPS estimator.

Definition 1. The *system properties* required for the design of an unbounded parameter set (UPS) 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 which must be known to the designer. The UPS estimator seeks to maximize the range of values of $\mathbf{x}$ for which the maximum error is guaranteed.

3. A *class of parameter sets* $\{\mathcal{U}_L \subseteq \mathbb{C}^n : \quad L \in [0, \infty)\}$ which define feasible values of $\mathbf{x}$ under varying parameter set bounds L. The parameter sets obey two basic properties:

   - As L increases, more values of $\mathbf{x}$ become feasible, so that the sets $\mathcal{U}_L$ are nested:

$$\mathcal{U}_{L_1} \subseteq \mathcal{U}_{L_2} \quad \Longleftrightarrow \quad L_1 \leq L_2. \tag{23}$$

- The parameter L is chosen so that it describes a linear expansion of the parameter sets: For all $L_1, L_2 > 0$,

$$\mathcal{U}_{L_1} = \frac{L_1}{L_2}\mathcal{U}_{L_2} = \left\{ \mathbf{x} : \frac{L_2}{L_1}\mathbf{x} \in \mathcal{U}_{L_2} \right\}. \tag{24}$$

  This requirement implies that the parameter sets are centered on the origin, an assumption which we adopt without loss of generality.

Most common bounds fulfill the requirements for the class of parameter sets above. The weighted norm $\mathcal{U}_L = \{\mathbf{x} : \mathbf{x}^*\mathbf{T}\mathbf{x} \le L^2\}$ used in Section 2 is one example. Another example is the box bound, $\mathcal{U}_L = \{\mathbf{x} : |x_i| \le Lb_i, \ \forall i\}$, 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 maximum parameter set bound L for which the maximum error is guaranteed, namely,

$$\hat{L}(\hat{\mathbf{x}}) = \max\{L : \epsilon(\hat{\mathbf{x}}, \mathbf{x}) \le \epsilon_m, \quad \forall \mathbf{x} \in \mathcal{U}_L\}. \tag{25}$$

Definition 3. An *unbounded parameter set (UPS) estimator* (among estimators of class $\mathcal{E}$) is an estimator $\hat{\mathbf{x}}_{\text{UP}}$ maximizing the parameter robustness $\hat{L}$ for given design parameters $\epsilon(\mathbf{x}, \hat{\mathbf{x}})$, ϵ_m, and $\mathcal{U}_L$; namely,

$$\hat{\mathbf{x}}_{\text{UP}} = \arg\max_{\hat{\mathbf{x}} \in \mathcal{E}} \hat{L}(\hat{\mathbf{x}}). \tag{26}$$

The set $\mathcal{E}$ is a given set of estimators among which the UPS estimator is chosen. For example, this set may be chosen to include only linear estimators, in which case $\hat{\mathbf{x}}_{\text{UP}}$ is called a linear UPS estimator.

The estimator presented in Section 2 is a special case of a UPS 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.

Although the minimax and UPS estimators are based on different design parameters, they are related in an important sense: in many cases, the set of minimax estimators equals the set of UPS estimators. This relation is formalized in Section 3.2. The minimax estimator is defined as follows.

Definition 4. A *minimax* (or *bounded parameter set*) *estimator* (among estimators of class $\mathcal{E}$) for the parameter set $\mathcal{U}$ is an estimator $\hat{\mathbf{x}}_{\text{M}}$ minimizing the worst-case error in $\mathcal{U}$:

$$\hat{\mathbf{x}}_{\text{M}} = \arg\min_{\hat{\mathbf{x}} \in \mathcal{E}} \max_{\mathbf{x} \in \mathcal{U}} \epsilon(\hat{\mathbf{x}}, \mathbf{x}). \tag{27}$$

3.2 Bounded and Unbounded Parameter Set Estimators

An interesting and useful relation exists between the UPS estimator $\hat{\mathbf{x}}_{\text{UP}}$ and the minimax estimator $\hat{\mathbf{x}}_{\text{M}}$: the UPS estimator maximizes the parameter robustness L within a range defined by the known value of ϵ, while the minimax estimator minimizes the worst-case error ϵ within a range defined by the known value of L. The relation is stated formally in Proposition 2 below. One implication of this proposition is that in many cases, UPS estimators can be efficiently found when an algorithm for finding the corresponding bounded parameter set estimator is known.

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

Proposition 2. *Let $\epsilon(\hat{\mathbf{x}}, \mathbf{x})$ be an error function, and let $\{\mathcal{U}_L : \ L \geq 0\}$ be a class of parameter sets, as defined in Definition 1.*

(a) *Let $\hat{\mathbf{x}}_{\text{UP}}$ be an unbounded parameter set (UPS) estimator with maximum error ϵ_m, and let $\hat{L}$ be the parameter robustness of $\hat{\mathbf{x}}_{\text{UP}}$. Then $\hat{\mathbf{x}}_{\text{UP}}$ is a minimax estimator for the parameter set $\mathcal{U}_{\hat{L}}$.*

(b) *Let $\{\hat{\mathbf{x}}_{\text{M}}(L) : L \geq 0\}$ be minimax estimators for the parameter sets $\{\mathcal{U}_L : \ L \geq 0\}$, respectively, and let $e(L)$ be the worst-case error function of $\hat{\mathbf{x}}_{\text{M}}(L)$, namely,*

$$e(L) = \max_{\mathbf{x} \in \mathcal{U}_L} \epsilon(\hat{\mathbf{x}}_{\text{M}}(L), \mathbf{x}). \tag{28}$$

If $e(L)$ is strictly monotonically increasing, then, for all L, $\hat{\mathbf{x}}_{\text{M}}(L)$ is a UPS estimator with maximum error $e(L)$.

Before proving Proposition 2, let us study the behavior of the worst-case error function $e(L)$ defined in (28). This function defines a region of achievable estimators (Figure 1). For any parameter set $\mathcal{U}_L$, an estimator may be constructed for which the worst-case error is $e(L)$; this is the minimax estimator $\hat{\mathbf{x}}_{\text{M}}(L)$. No estimator can be constructed with lower worst-case error, since this would contradict the fact that $\hat{\mathbf{x}}_{\text{M}}(L)$ is a minimax estimator.

Clearly, the worst-case error function is non-decreasing in L. To see this, suppose by contradiction that $e(L) > e(M)$ for some $L < M$. Then, the worst-case error of $\hat{\mathbf{x}}_{\text{M}}(M)$ over $\mathcal{U}_L$ is lower than the worst-case error of $\hat{\mathbf{x}}_{\text{M}}(L)$ over $\mathcal{U}_L$, contradicting the fact that $\hat{\mathbf{x}}_{\text{M}}(L)$ is a minimax estimator for $\mathcal{U}_L$. As Proposition 2(b) shows, if $e(L)$ is *strictly* increasing, then $\hat{\mathbf{x}}_{\text{M}}(L)$ is not only the set of minimax estimators, but also the set of

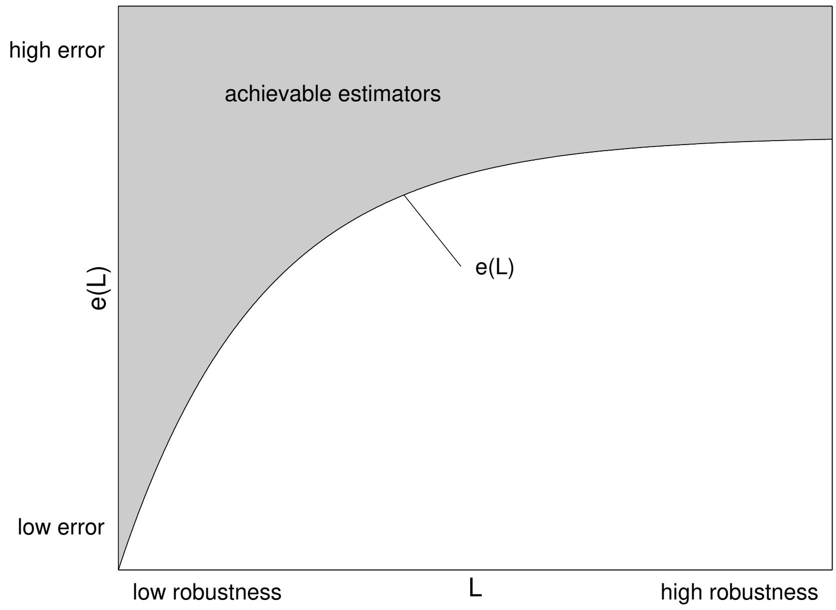

Figure 1: The worst-case error function and the region of achievable estimators

UPS estimators for varying maximum errors. As we shall 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. Also, if the error function is constant, at least for some range of values of $\mathbf{x}$, then increasing the parameter set may leave the worst-case error unchanged.

The worst-case error function can also be used to find the parameter robustness of a UPS estimator. Suppose we are given a maximum error ϵ_m and wish to find the robustness of the resulting UPS estimator. For any L satisfying $e(L) \leq \epsilon_m$, the estimator $\hat{\mathbf{x}}_{\mathrm{M}}(L)$ has worst-case error not exceeding ϵ_m. Also, for any L satisfying $e(L) > \epsilon_m$, there does not exist an estimator for which the worst-case error does not exceed ϵ_m. Hence, the parameter robustness of the UPS estimator is the largest L for which $e(L) \leq \epsilon_m$.

When describing UPS estimators, $e(L)$ represents a trade-off between performance and robustness to uncertainty in the parameter set. When performance requirements are modest (ϵ_m is large), performance can be guaranteed for large variations in the parameter set ($\mathcal{U}_L$ is large); when performance requirements are stringent, satisfactory performance can only be guaranteed for a small parameter set. As Proposition 2 shows, this is equivalent, in many cases, to the minimax trade-off, which states that when the parameter set is increased ($\mathcal{U}_L$ is large), the achieved performance is decreased (the worst-case error is increased).

Proof of Proposition 2. (a) Suppose $\hat{\mathbf{x}}_{\mathrm{UP}}$ is a UPS estimator with parameter robustness $\hat{L}$. Assume by contradiction that it is not a minimax estimator over $\mathcal{U}_{\hat{L}}$. Then, by Definition 4, there exists an estimator $\hat{\mathbf{x}}_{\mathrm{M}}$ such that

$$\max_{\mathbf{x}\in\mathcal{U}_{\hat{L}}} \epsilon(\mathbf{x},\hat{\mathbf{x}}_{\mathrm{M}}) < \max_{\mathbf{x}\in\mathcal{U}_{\hat{L}}} \epsilon(\mathbf{x},\hat{\mathbf{x}}_{\mathrm{UP}}) \leq \epsilon_m. \tag{29}$$

By Definition 1, the parameter sets expand linearly, so that for sufficiently small $\alpha > 1$, the parameter set $\mathcal{U}_{\alpha L}$ contains values of $\mathbf{x}$ which are all arbitrarily close to some value of $\mathbf{x}$ in $\mathcal{U}_L$. Furthermore, by Definition 1, ϵ is continuous, so that sufficiently small changes in $\mathbf{x}$ yield arbitrarily small changes in $\epsilon(\hat{\mathbf{x}}_{\mathrm{UP}},\mathbf{x})$. Hence, there exists a sufficiently small $\alpha > 1$ for which

$$\max_{\mathbf{x}\in\mathcal{U}_{\alpha\hat{L}}} \epsilon(\mathbf{x},\hat{\mathbf{x}}_{\mathrm{M}}) \leq \epsilon_m. \tag{30}$$

Thus the parameter robustness of $\hat{\mathbf{x}}_{\mathrm{M}}$ is at least $\alpha\hat{L}$, which is greater than the parameter robustness of $\hat{\mathbf{x}}_{\mathrm{UP}}$. This contradicts the fact that $\hat{\mathbf{x}}_{\mathrm{UP}}$ is a UPS estimator.

(b) Given L, define the required maximum error as $\epsilon_m = e(L)$. Assume by contradiction that $\hat{\mathbf{x}}_{\mathrm{M}}(L)$ is not a UPS estimator for the maximum error ϵ_m. Then, there exists an $\hat{\mathbf{x}}_{\mathrm{UP}}$ for which

$$\hat{L}(\hat{\mathbf{x}}_{\mathrm{UP}}) > \hat{L}(\hat{\mathbf{x}}_{\mathrm{M}}(L)) \geq L. \tag{31}$$

Therefore,

$$\max_{\mathbf{x}\in\mathcal{U}_{\hat{L}(\hat{\mathbf{x}}_{\mathrm{UP}})}} \epsilon(\hat{\mathbf{x}}_{\mathrm{UP}},\mathbf{x}) \leq \epsilon_m = e(L) < e(\hat{L}(\hat{\mathbf{x}}_{\mathrm{UP}})). \tag{32}$$

However, by definition of the worst-case error function $e(L)$,

$$\max_{\mathbf{x}\in\mathcal{U}_{\hat{L}(\hat{\mathbf{x}}_{\mathrm{UP}})}} \epsilon(\hat{\mathbf{x}}_{\mathrm{M}}(\hat{L}(\mathbf{x}_u)),\mathbf{x}) = e(\hat{L}(\hat{\mathbf{x}}_{\mathrm{UP}})). \tag{33}$$

Hence $\hat{\mathbf{x}}_{\mathrm{UP}}$ achieves a lower worst-case error over $\mathcal{U}_{\hat{L}(\hat{\mathbf{x}}_{\mathrm{UP}})}$ than the minimax estimator of $\mathcal{U}_{\hat{L}(\hat{\mathbf{x}}_{\mathrm{UP}})}$, which is a contradiction. We conclude that $\hat{\mathbf{x}}_{\mathrm{M}}(L)$ must be a UPS estimator. □

We have shown that under appropriate conditions, a UPS estimator is equivalent to a minimax estimator with a matching choice of a parameter set. These conditions will be shown to hold for several cases of interest, including linear MSE estimators, in the following section. This equivalence can be used to efficiently find the UPS estimator, if an algorithm for finding a minimax estimator is known, 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 be found using a logarithmic number of calculations of the value of $e(L)$. From Proposition 2(b), the minimax estimator achieving the required worst-case error is also a UPS estimator.

In some cases, such as the linear MSE estimator presented in Section 2, an algorithm for directly finding the UPS estimator is known, and may be more efficient than the bisection method. However, for arbitrary design parameters, a direct algorithm for finding a UPS estimator may not be known. Furthermore, in many cases the minimax estimator can be determined very efficiently, and in some cases a closed form is known for the minimax estimator. Under these conditions the bisection method may be more efficient than the direct algorithm, even if one is available. If, in addition, $e(L) = \epsilon_m$ can be solved analytically for a particular value of ϵ_m, then a closed form can be found for the UPS estimator. One example in which this occurs is presented in Section 4.1.

4 Estimators for Various Error Functions

As we have shown in Section 3.2, the UPS estimator is equivalent to the minimax estimator, provided that simple conditions on the error function are maintained. We now prove that these conditions hold for two cases of interest, the MSE estimator (Section 4.1) and the regret estimator (Section 4.2).

4.1 Linear MSE Estimators

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

Proposition 3. *Let $\{\mathcal{U}_L : L \geq 0\}$ be a class of parameter sets as defined in Definition 1. Suppose that the error function of interest is the MSE, $\epsilon(\hat{\mathbf{x}}, \mathbf{x}) = E\|\hat{\mathbf{x}} - \mathbf{x}\|^2$. For all L, let $\hat{\mathbf{x}}_{\mathrm{M}}(L) = \mathbf{G}_L \mathbf{y}$ be a linear minimax MSE estimator for the parameter set $\mathcal{U}_L$, and let $e(L)$ be the worst-case error function of $\hat{\mathbf{x}}_{\mathrm{M}}(L)$, namely,*

$$e(L) = \max_{\mathbf{x} \in \mathcal{U}_L} E\|\hat{\mathbf{x}}_{\mathrm{M}}(L) - \mathbf{x}\|^2. \tag{34}$$

Then, $\hat{\mathbf{x}}_{\mathrm{M}}(L)$ is an unbounded parameter set (UPS) estimator with maximum error $e(L)$.

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

Lemma 1. *Given any bounded parameter set $\mathcal{U}$, there exists a linear biased estimator $\hat{\mathbf{x}}_b$ 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 M such that $\mathcal{U}$ is bounded within a sphere of radius M. The linear minimax MSE estimator for this sphere is given by [6]

$$\hat{\mathbf{x}}_b = \mathbf{G}\mathbf{y} = \frac{M^2}{M^2 + \gamma_0} (\mathbf{H}^* \mathbf{C}_{\mathbf{w}}^{-1} \mathbf{H})^{-1} \mathbf{H}^* \mathbf{C}_{\mathbf{w}}^{-1} \mathbf{y} \tag{35}$$

where $\gamma_0 = \mathrm{Tr}((\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H})^{-1})$ is the MSE of the unbiased estimator. We now show that $\hat{\mathbf{x}}_b$ achieves a lower MSE than the LS estimator for all $\mathbf{x} \in \mathcal{U}$. The bias of $\hat{\mathbf{x}}_b$ is given by

$$\mathbf{b}(\hat{\mathbf{x}}_b) = E(\hat{\mathbf{x}}_b - \mathbf{x}) = E\left(\beta\mathbf{x} + \beta(\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{w} - \mathbf{x}\right) = (\beta - 1)\mathbf{x}, \tag{36}$$

where $\beta = \frac{M^2}{M^2+\gamma_0}$. The variance of $\hat{\mathbf{x}}_b$ is

$$\begin{aligned} v(\hat{\mathbf{x}}_b) &= \mathrm{Tr}(\mathbf{G}\mathbf{C}_{\mathbf{w}}\mathbf{G}^*) \\ &= \mathrm{Tr}\left(\beta(\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{C}_{\mathbf{w}}\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}(\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H})^{-1}\beta\right) \\ &= \beta^2\mathrm{Tr}((\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H})^{-1}) \\ &= \beta^2\gamma_0. \end{aligned} \tag{37}$$

Since $\mathrm{MSE}(\hat{\mathbf{x}}_b) = v(\hat{\mathbf{x}}_b) + \|\mathbf{b}(\hat{\mathbf{x}}_b)\|^2$, we have

$$\begin{aligned} \mathrm{MSE}(\hat{\mathbf{x}}_b) &= \beta^2\gamma_0 + (1-\beta)^2\|\mathbf{x}\|^2 && (38) \\ &\leq \beta^2\gamma_0 + (1-\beta)^2 M^2 && (39) \\ &= \left(\frac{M^2}{M^2+\gamma_0}\right)\gamma_0 && (40) \\ &< \gamma_0. && (41) \end{aligned}$$

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

Proof of Proposition 3. By Proposition 2(b), it is sufficient to show that $e(L)$ is strictly monotonically increasing. From the definition of $e(L)$ and from (2)–(4),

$$e(L) = \mathrm{Tr}(\mathbf{G}_L\mathbf{C}_{\mathbf{w}}\mathbf{G}_L^*) + \max_{\mathbf{x}\in\mathcal{U}_L} \mathbf{x}^*(\mathbf{I} - \mathbf{G}_L\mathbf{H})^*(\mathbf{I} - \mathbf{G}_L\mathbf{H})\mathbf{x}. \tag{42}$$

Lemma 1 states that there exists a biased estimator which achieves lower MSE than the LS estimator. 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., $\mathbf{G}_L\mathbf{H} \neq \mathbf{I}$. Thus, as long as $\mathbf{x}$ is not on the boundary of $\mathcal{U}_L$, $\mathbf{x}^*(\mathbf{I} - \mathbf{G}_L\mathbf{H})^*(\mathbf{I} - \mathbf{G}_L\mathbf{H})\mathbf{x}$ can be increased by replacing $\mathbf{x}$ with $(1+\delta)\mathbf{x}$ for some small $\delta > 0$. Hence, the maximum of $\mathbf{x}^*(\mathbf{I} - \mathbf{G}_L\mathbf{H})^*(\mathbf{I} - \mathbf{G}_L\mathbf{H})\mathbf{x}$ over $\mathcal{U}_L$ is obtained *only* on the boundary of $\mathcal{U}_L$. Therefore, by shrinking the parameter set, the worst-case error must decrease: for any $L < M$,

$$\max_{\mathbf{x}\in\mathcal{U}_L} E\|\hat{\mathbf{x}}_{\mathrm{M}}(M) - \mathbf{x}\|^2 < \max_{\mathbf{x}\in\mathcal{U}_M} E\|\hat{\mathbf{x}}_{\mathrm{M}}(M) - \mathbf{x}\|^2. \tag{43}$$

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

$$\max_{\mathbf{x}\in\mathcal{U}_L} E\|\hat{\mathbf{x}}_{\mathrm{M}}(L) - \mathbf{x}\|^2 \leq \max_{\mathbf{x}\in\mathcal{U}_L} E\|\hat{\mathbf{x}}_{\mathrm{M}}(M) - \mathbf{x}\|^2. \tag{44}$$

Together with (43), this implies that $e(L) < e(M)$ for all $L < M$. Thus, by Proposition 2(b), $\hat{\mathbf{x}}_{\mathrm{M}}(L)$ is a UPS estimator with maximum error $e(L)$, for all L. $\square$

As we have seen, using the MSE as an error function, the set of minimax estimators equals the set of UPS estimators for a given class of parameter sets. Thus, finding a UPS estimator for a given maximum error ϵ_m becomes simply a matter of finding the 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 the UPS estimator as well. This is the case for the class of ellipsoidal parameter sets, as demonstrated by the following proposition.

Proposition 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,*

$$\hat{\mathbf{x}}_{\mathrm{LS}} = (\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{y}, \tag{45}$$

and let $\gamma_0 = \mathrm{Tr}\left((\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H})^{-1}\right)$ *be the MSE of* $\hat{\mathbf{x}}_{\mathrm{LS}}$.

(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 unbounded parameter set (UPS) estimator for a given maximum error* ϵ_m *is given by*

$$\hat{\mathbf{x}}_{\mathrm{UP}} = \begin{cases} \mathbf{V}\begin{bmatrix} \mathbf{0} & \\ & \mathbf{I}_{m-k} \end{bmatrix}\mathbf{V}^*(\mathbf{I} - \alpha\mathbf{T}^{1/2})\hat{\mathbf{x}}_{\mathrm{LS}}, & \epsilon_m < \gamma_0 \\ \hat{\mathbf{x}}_{\mathrm{LS}}, & \epsilon_m \ge \gamma_0, \end{cases} \tag{46}$$

where

$$\alpha = \frac{\sum_{i=k+1}^{m} \frac{1}{\sigma_i} - \epsilon_m}{\sum_{i=k+1}^{m} \frac{\lambda_i^{1/2}}{\sigma_i}} \tag{47}$$

and

$$k = \min\left\{i : \ \alpha\lambda_{i+1}^{1/2} < 1\right\}. \tag{48}$$

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

$$\hat{\mathbf{x}}_{\mathrm{UP}} = \begin{cases} \frac{\epsilon_m}{\gamma_0}\hat{\mathbf{x}}_{\mathrm{LS}}, & \epsilon_m < \gamma_0 \\ \hat{\mathbf{x}}_{\mathrm{LS}}, & \epsilon_m \ge \gamma_0. \end{cases} \tag{49}$$

The parameter robustness of this estimator is given by

$$\hat{L}(\hat{\mathbf{x}}_{\mathrm{UP}}) = \begin{cases} \sqrt{\frac{\gamma_0\epsilon_m}{\gamma_0 - \epsilon_m}}, & \epsilon_m < \gamma_0 \\ \infty, & \epsilon_m \ge \gamma_0. \end{cases} \tag{50}$$

Proof. (a) We seek an estimator which guarantees an error not exceeding ϵ_m for as large a parameter set as possible. We begin with the case $\epsilon_m \geq \gamma_0$. In this case, the allowed error is larger than γ_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 a UPS estimator for this trivial case.

We now consider the case $\epsilon_m < \gamma_0$. From Proposition 3, the UPS estimator is also a minimax estimator. It is shown in [6] that the minimax MSE estimator for a given parameter set $\mathcal{U}_L$ is given by

$$\hat{\mathbf{x}}_{\mathrm{M}} = \mathbf{V} \begin{bmatrix} \mathbf{0} & \\ & \mathbf{I}_{m-k} \end{bmatrix} \mathbf{V}^*(\mathbf{I} - \alpha \mathbf{T}^{1/2})\hat{\mathbf{x}}_{\mathrm{LS}}, \tag{51}$$

where

$$\alpha = \frac{\sum_{i=k+1}^{m} \frac{\lambda_i^{1/2}}{\sigma_i}}{L^2 + \sum_{i=k+1}^{m} \frac{\lambda_i}{\sigma_i}}, \tag{52}$$

and k is defined in (48). The worst-case error for this estimator is

$$\sum_{i=k+1}^{m} \frac{1 - \alpha \lambda_i^{1/2}}{\sigma_i}. \tag{53}$$

We require a value of L for which the worst-case error equals ϵ_m. Equating (53) with ϵ_m, we arrive at (47).

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

$$\sum_{i=1}^{m} \frac{1}{\sigma_i} = \mathrm{Tr}(\mathbf{\Sigma}^{-1}) = \mathrm{Tr}\left((\mathbf{H}^* \mathbf{C}_{\mathbf{w}}^{-1} \mathbf{H})^{-1} \right) = \gamma_0, \tag{54}$$

and thus

$$\alpha = \frac{\gamma_0 - \epsilon_m}{\gamma_0}. \tag{55}$$

Substituting these results into (46) yields the required estimator (49). We have already seen that the parameter robustness when $\epsilon_m \geq \gamma_0$ is infinite. To find the parameter robustness when $\epsilon_m < \gamma_0$, notice that (52) is now

$$\alpha = \frac{\gamma_0}{L^2 + \gamma_0}. \tag{56}$$

Combining this with (55) yields

$$L^2 = \frac{\gamma_0 \epsilon_m}{\gamma_0 - \epsilon_m}, \tag{57}$$

which is the required result (50). □

It is sometimes useful to find the actual MSE obtained by the UPS estimator. The MSE can be calculated for the matching minimax estimator. For example, it has been shown (38) that the MSE of the minimax estimator for a spherical parameter set $\mathbf{x}^*\mathbf{x} \leq L^2$ is given by

$$\mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{UP}}) = \left(\frac{L^2}{L^2 + \gamma_0} \right)^2 \gamma_0 + \left(\frac{\gamma_0}{L^2 + \gamma_0} \right)^2 \|\mathbf{x}\|^2. \tag{58}$$

Substituting the value of L^2 from (50), we have

$$\text{MSE}(\hat{\mathbf{x}}_{\text{UP}}) = \begin{cases} \frac{\epsilon_m^2}{\gamma_0} + \left(\frac{\gamma_0 - \epsilon_m}{\gamma_0}\right)^2 \|\mathbf{x}\|^2, & \epsilon_m < \gamma_0 \\ \gamma_0, & \epsilon_m \geq \gamma_0 \end{cases} \tag{59}$$

Thus, the MSE of the unbounded spherical parameter set estimator is a linear function of $\|\mathbf{x}\|^2$. This result is useful for comparing the performance of the UPS estimator with other estimators, as we demonstrate in Section 5.

4.2 Linear Regret Estimators

We now present a different example of a UPS estimator, one which guarantees a worst-case *regret*. 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 [7] that the regret is given by

$$\epsilon(\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}, \mathbf{x}) = \text{Tr}(\mathbf{G}\mathbf{C}_{\mathbf{w}}\mathbf{G}^*) + \mathbf{x}^*(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{x} - \frac{\mathbf{x}^*\mathbf{x}}{1 + \mathbf{x}^*\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}\mathbf{x}}. \tag{60}$$

In this section, we limit our discussion to parameter sets of the form $\mathcal{U}_L = \{\mathbf{x} : \mathbf{x}^*\mathbf{T}\mathbf{x} \leq L^2\}$, 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 $\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}$ have the same eigenvectors. We show that, under these assumptions, the linear UPS regret estimator is equivalent to the linear minimax regret estimator. It follows that the UPS 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 [7].

Proposition 5. *Let $\mathcal{U}_L = \{\mathbf{x} : \ \mathbf{x}^*\mathbf{T}\mathbf{x} \leq L^2\}$ 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 $\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}$. Consider the regret (60) as the error function of interest. For all L, let $\hat{\mathbf{x}}_{\text{M}}(L) = \mathbf{G}_L\mathbf{y}$ be a linear minimax regret estimator for the parameter set $\mathcal{U}_L$, and let $e(L)$ be the worst-case regret of $\hat{\mathbf{x}}_{\text{M}}(L)$, namely,*

$$e(L) = \max_{\mathbf{x} \in \mathcal{U}_L} \epsilon(\hat{\mathbf{x}}_{\text{M}}(L), \mathbf{x}). \tag{61}$$

Then, $\hat{\mathbf{x}}_{\text{M}}(L)$ is a linear unbounded parameter set (UPS) regret estimator with maximum error $e(L)$.

Proof. As in the proof of Proposition 3, we will show that $e(L)$ is strictly increasing with L. By Proposition 2(b), this suffices to show that $\hat{\mathbf{x}}_{\text{M}}(L)$ is a UPS regret estimator.

It has been shown in [7, Theorem 1] that under the conditions of Proposition 5, the minimax regret estimator is given by

$$\mathbf{G}_L = \mathbf{V}\mathbf{D}\mathbf{V}^*(\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}, \tag{62}$$

where $\mathbf{D}$ is a diagonal matrix whose diagonal elements d_i are the solution to the optimization problem

$$\begin{aligned} &\min_{\tau,\mathbf{d}} \tau \\ &\text{s.t.} \begin{cases} F_1(\mathbf{d}) \leq \tau & \text{(a)} \\ F_2(\mathbf{d},\mathbf{s}) \leq \tau \quad \forall\, \mathbf{s} \in \mathcal{S}, & \text{(b)} \end{cases} \end{aligned} \tag{63}$$

where

$$F_1(\mathbf{d}) = \sum_{i=1}^{m} \frac{d_i^2}{\sigma_i}, \tag{64}$$

$$F_2(\mathbf{d},\mathbf{s}) = \sum_{i=1}^{m} \frac{d_i^2}{\sigma_i} + \sum_{i=1}^{m}(1-d_i)^2 s_i - \frac{\sum_{i=1}^{m} s_i}{1+\sum_{i=1}^{m}\sigma_i s_i}, \tag{65}$$

$$\mathcal{S} = \left\{ \mathbf{s}:\ s_i \geq 0,\ \sum_{i=1}^{m} \lambda_i s_i = L^2 \right\}, \tag{66}$$

and σ_i are eigenvectors of $\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}$ such that $\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H} = \mathbf{V}\mathbf{\Sigma}\mathbf{V}^*$, with $\mathbf{\Sigma} = \operatorname{diag}(\sigma_1,\dots\sigma_m)$. In (63), the optimal value of τ is the worst-case regret $e(L)$.

We begin our proof by showing that (63b) is an active constraint in the optimization problem. Assume by contradiction that (63b) is inactive. Then, by the Karush-Kuhn-Tucker conditions for optimality [4, §5.5.3], (63) is equivalent to

$$\min_{\mathbf{d},\tau} \tau \quad \text{s.t.} \sum \frac{d_i^2}{\sigma_i} \leq \tau, \tag{67}$$

for which the optimal solution is $\mathbf{d} = \mathbf{0}, \tau = 0$. However, for any $\mathbf{s}$,

$$F_2(\mathbf{0},\mathbf{s}) > 0 = \tau, \tag{68}$$

contradicting the fact that (63b) is inactive. Thus, for the optimal value of τ and $\mathbf{d}$, there exists at least one active $\mathbf{s} \in \mathcal{S}$ for which $F_2(\mathbf{d},\mathbf{s}) = \tau$. From the definition of F_1 and F_2, we have

$$\sum(1-d_i)^2 s_i - \frac{\sum s_i}{1+\sum \sigma_i s_i} \geq 0 \quad \text{for any active } \mathbf{s}. \tag{69}$$

In the above discussion, the dependence on L was implicit and resulted from the requirement that $\sum \lambda_i s_i = L^2$. To make this dependence explicit, we define

$$r_i = \frac{s_i}{L^2}, \tag{70}$$

so that the requirement $\mathbf{s} \in \mathcal{S}$ is equivalent to $r_i \geq 0, \sum \lambda_i r_i = 1$. Rewriting (69) using this notation, we have

$$g(\mathbf{d}, \mathbf{r}, L^2) \triangleq L^2 \sum (1-d_i)^2 r_i - \frac{L^2 \sum r_i}{1 + L^2 \sum \sigma_i r_i} \geq 0 \quad \text{for any active } \mathbf{r}. \tag{71}$$

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

$$\frac{\partial g}{\partial L^2} = \sum (1-d_i)^2 r_i - \frac{\sum r_i}{(1 + L^2 \sum \sigma_i r_i)^2}. \tag{72}$$

Hence, at points for which $g(\mathbf{d}, \mathbf{r}, L^2) \geq 0$, we have

$$\begin{aligned} \left.\frac{\partial g}{\partial L^2}\right|_{g \geq 0} &\geq \frac{\sum r_i}{1 + L^2 \sum \sigma_i r_i} - \frac{\sum r_i}{(1 + L^2 \sum \sigma_i r_i)^2} \\ &= \frac{\sum r_i}{1 + L^2 \sum \sigma_i r_i} \left(1 - \frac{1}{1 + L^2 \sum \sigma_i r_i}\right) \\ &> 0. \end{aligned} \tag{73}$$

Thus, for any active $\mathbf{r}$, $g(\mathbf{d}, \mathbf{r}, L^2)$ is strictly increasing with L. Therefore, if L is decreased, then $F_2(\mathbf{d}, \mathbf{s}) = F_1(\mathbf{d}) + g(\mathbf{d}, \mathbf{s}/L^2, L^2)$ is decreased for all active $\mathbf{s}$, and the constraint (63b) 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. From Proposition 2(b), this indicates that $\hat{\mathbf{x}}_\mathrm{M}(L)$ is a UPS regret estimator. □

Estimators based on different error functions were examined in this section, in an attempt to identify conditions for equivalence between bounded and unbounded parameter set estimation. We have seen (Section 4.1) that when the error function is the MSE, minimax and UPS estimators are equivalent. For the regret as an error function, the proof of equivalence requires several assumptions for analytical tractability (Section 4.2). The relation between minimax and UPS estimation led to a closed form for the UPS estimator, in the special case of ellipsoidal parameter sets with the MSE error function (Proposition 4). An application of the UPS estimator, which makes use of this closed-form estimator, is presented in the next section.

5 Application: Channel Estimation

As an application of the UPS 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) [20].

Let $\mathbf{c} = (c_0, \dots c_{N_c-1})$ denote the unknown channel impulse response of length N_c, and let $\mathbf{p} = (p_{-N_c+1}, p_{-N_c+2}, \dots p_0, \dots p_{N_p-N_c})^T$ denote the known vector of preamble symbols of length N_p. The corresponding received symbols are given by

$$r_k = \sum_{l=0}^{N_c-1} c_l p_{k-l} + w_k, \quad k = 0, 1, \dots N'_p - 1, \tag{74}$$

where w_k is additive white noise with variance σ_w^2, and $N'_p = N_p - N_c + 1$. We assume that the channel length N_c is known. We also assume that the channel consists of a direct transmission element c_0, which is normalized to 1, and multipath echoes $\mathbf{c}' = (c_1, \dots c_{N_c-1})^T$. Substituting the known value of c_0, we obtain

$$\mathbf{r} - \mathbf{p}' = \mathbf{H}\mathbf{c}' + \mathbf{w}, \tag{75}$$

where $\mathbf{p}' = (p_0, \dots p_{N_p-N_c})^T$, and

$$\mathbf{H} = \begin{bmatrix} p_{-1} & p_{-2} & \cdots & p_{-N_c+1} \\ p_0 & p_{-1} & \cdots & p_{-N_c+2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N'_p-2} & p_{N'_p-3} & \cdots & p_{-N_c+N'_p} \end{bmatrix}. \tag{76}$$

The classical approach to channel estimation using a preamble is least-squares estimation of the unknown, deterministic vector $\mathbf{c}'$ from the measurements $\mathbf{r}$ [5, 8, 20]. The estimated channel in this case is

$$\hat{\mathbf{c}}' = \mathbf{G}_{\mathrm{LS}}\mathbf{r} = (\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*\mathbf{r}. \tag{77}$$

This estimator minimizes the squared error of the measurements, i.e.,

$$\mathbf{G}_{\mathrm{LS}} = \arg\min_{\mathbf{G}} \left\|(\mathbf{r} - \mathbf{p}') - \mathbf{H}\mathbf{G}(\mathbf{r} - \mathbf{p}')\right\|^2. \tag{78}$$

However, in this context there is no advantage in minimizing the *measurement* error; rather, we are interested in minimizing the channel estimation error $\epsilon = \|\mathbf{c} - \hat{\mathbf{c}}\|^2$, as the channel estimate is used for further processing (e.g., detection of payload data). For example, in [5], 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 parameter $\mathbf{c}'$, so 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 $\mathcal{U}$. But although we may believe that $\|\mathbf{c}'\|$ is generally small compared with c_0, we generally cannot explicitly determine a bound on $\|\mathbf{c}'\|$.

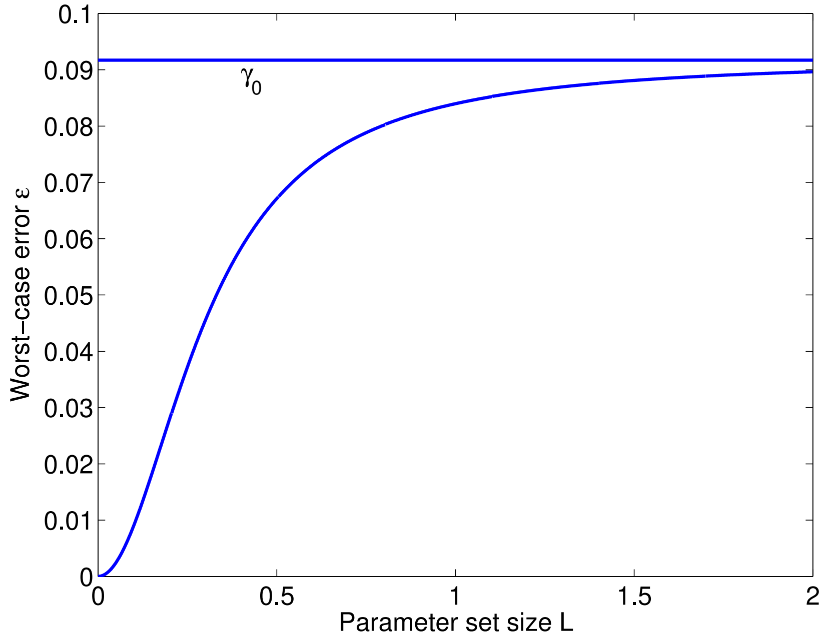

Figure 2: The worst-case error for the channel estimation problem

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. In this case, the estimation error requirement is a design parameter; it is to be chosen together with other system properties such as receiver SINR requirements. An informed choice may be made regarding the desired estimation error. We can use the UPS estimator to maximize the set of channels for which a required estimation error is achieved. The reasoning behind such an approach could state that a given maximum estimation error is critical for system operation; such an estimation error should be guaranteed for as wide a range of channels as possible.

Consider a class of nested parameter sets $\mathcal{U}_L = \{\mathbf{c}' : \|\mathbf{c}'\| \leq L\}$ and suppose we require an MSE not exceeding ϵ_m. We seek the estimator guaranteeing estimation error of ϵ_m or less, for as large an parameter set as possible. By Proposition 4, the maximum error ϵ_m must first be compared with $\gamma_0 = \mathrm{Tr}((\mathbf{H}^*\mathbf{H})^{-1})$, the MSE of the LS estimator. If $\epsilon_m \geq \gamma_0$, then an error of γ_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, when $\epsilon_m < \gamma_0$, a different estimator is required. It was shown in Proposition 4(b) that in this case, the UPS estimator is given by

$$\hat{\mathbf{c}}_u = \frac{\epsilon_m}{\gamma_0}(\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*\mathbf{r}, \quad \epsilon_m \leq \gamma_0. \tag{79}$$

To compare the performance of the LS and UPS channel estimators, we consider the problem of estimating

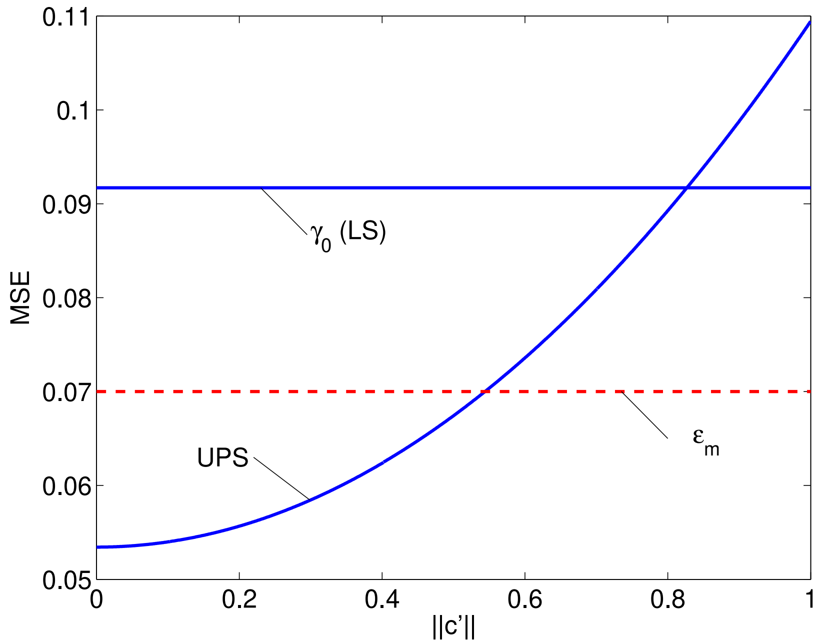

Figure 3: MSE for various values of $\|\mathbf{c}'\|$

a 7-tap channel using a 14-symbol BPSK preamble, and use the optimal preamble suggested in [5], given by

$$[-1,\ -1,\ -1,\ +1,\ -1,\ -1,\ +1,\ -1,\ -1,\ -1,\ +1,\ +1,\ +1,\ -1]. \tag{80}$$

We assume that the noise variance is $\sigma_w^2 = 0.1$. For these parameters, the resulting MSE of the LS estimator is $\gamma_0 = 0.0917$. Using this value and (58), the worst-case error of various minimax MSE estimators is plotted in Figure 2. All of these estimators are also UPS 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. For instance, if a maximum error level of 0.1 is acceptable, then clearly the optimal choice is the LS estimator, which guarantees an MSE of $\gamma_0 < 0.1$ for all channels. However, in some cases the detector degradation provided by such an estimation error may be prohibitive. The engineer may choose a lower maximum channel estimation error while taking into consideration the reduced set of channels for which estimation would be successful. For instance, a maximum error in the range of 50% to 90% of the value of γ_0 is a reasonable choice, as this range covers a reasonably-sized parameter set while still reducing the worst-case error.

Suppose we choose to design our system such that a channel estimation error of $\epsilon_m = 0.75\gamma_0$ is to be tolerated, i.e., we require an error not exceeding 75% of the MSE obtained by the LS estimator. This implies appropriate design steps in the system, which allow the receiver to handle estimation errors up to ϵ_m (for example, error correction suitable for such noise levels). Once the choice of ϵ_m has been made, we would

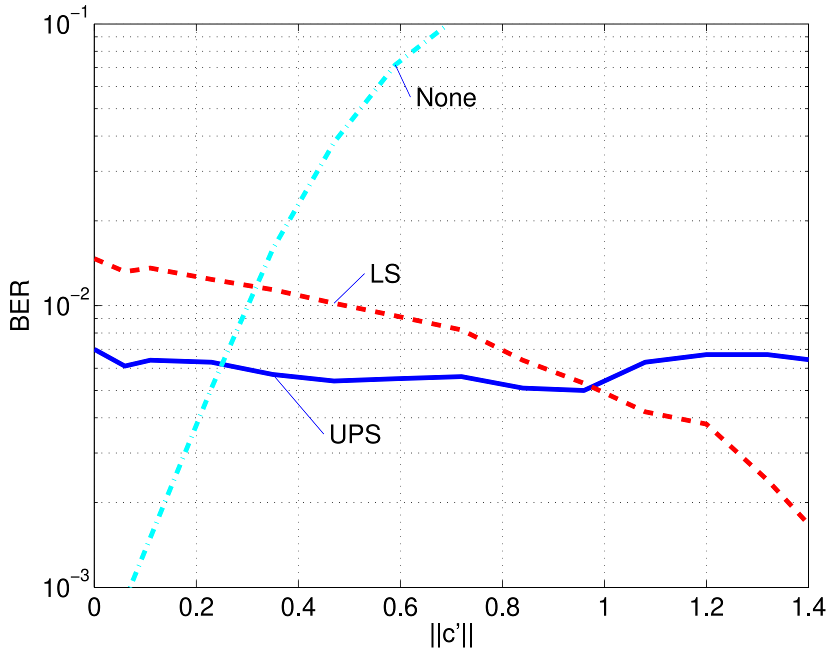

Figure 4: Bit error rate for various channels with the LS and UPS channel estimators

like to maximize the set of channels for which the estimation error does not exceed the specified limit; this is achieved by the UPS estimator (79). The MSE (59) of the UPS estimator is compared with the MSE γ_0 of the LS estimator in Figure 3. The LS estimator never achieves the required error level ϵ_m, while the UPS estimator achieves this error level for some channels. The drawback of the UPS estimator is poor performance for other channels. However, no linear estimator can achieve the required error level for a larger parameter set.

To verify that the reduced estimation error results in improved detection performance, a simulation of BPSK detection was performed. A received signal (74) was simulated from a BPSK transmission containing the 14 preamble symbols (80) and 100 random data symbols. A basic 7-tap channel, given by

$$\mathbf{c} = [1.00,\ -0.20\alpha,\ 0.15\alpha,\ 0.10\alpha,\ -0.08\alpha,\ 0.05\alpha,\ 0.02\alpha], \tag{81}$$

was used, where the parameter α was varied between 0 and 5 to obtain values of $\|\mathbf{c}'\|$ between 0 and 1.4. (The coefficient c_0 is not multiplied by α as this coefficient is assumed to be normalized to 1.) The channel was estimated using both the LS and UPS estimators described above, and the resulting channel estimate was used for MLSE detection of the data symbols. The experiment was repeated 2500 times to obtain an estimate of the bit error rate (BER).

The results are presented in Figure 4. For comparison, a null estimator is also plotted; this "estimator" assumes that $\|\mathbf{c}'\| = 0$, i.e., only direct transmission is present.

These results demonstrate that in terms of BER, UPS estimation outperforms standard LS estimation for a range of channels. The UPS estimator maintains a BER level around 0.6% in the measured channel range, while the LS estimator results in BER levels above 1% for many common channels.[1]

The UPS 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 $\mathbf{c}' = 0$. UPS estimators provide a spectrum 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.

In summary, we presented the use of UPS estimators for channel estimation problems. We have seen that channel estimation error can be reduced if a maximum estimation error is specified by the designer, and this leads to lower BER for a wide range of channels. This is achieved without increasing estimation complexity: the UPS estimator is a linear estimator and can be calculated as easily as the least-squares estimator.

6 Unbounded Noise Level Estimation

Throughout this 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. Following [16], in this section we consider the case where

$$E(\mathbf{w}\mathbf{w}^*) = \sigma^2 \mathbf{C}_\mathbf{w}, \tag{82}$$

for some unknown deterministic *noise level* σ^2, and some known covariance matrix $\mathbf{C}_\mathbf{w}$. For example, when the noise is i.i.d., $\mathbf{C}_\mathbf{w} = \mathbf{I}$ and σ^2 is the noise variance. The estimation techniques used so far require complete knowledge of the noise covariance: LS, minimax or UPS 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 $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, 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 which 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 *unbounded noise level (UNL) estimator.* As we shall show, the UNL estimator is related to

[1] The reason for the improved BER achieved by the LS estimator for large $\|\mathbf{c}'\|$ is not a decrease in channel estimation error. Rather, it is a result of the large amount of energy found in the multipath channel coefficients; this energy can be used to improve detection SNR. For the UPS estimator, this increase is roughly cancelled by the increased channel estimation error.

the minimax estimator, allowing us to efficiently find the UNL 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 $\epsilon_{\sigma^2}(\hat{\mathbf{x}}, \mathbf{x}) \leq \epsilon_m$ to be satisfied over the entire range $\mathbf{x} \in \mathcal{U}$. We can now define a new type of unbounded uncertainty estimator, in a manner analogous to the definition of the UPS in Section 3.1, 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:

$$\hat{\sigma}^2(\hat{\mathbf{x}}) = \max\left\{\sigma^2 : \max_{\mathbf{x}\in\mathcal{U}} \epsilon_{\sigma^2}(\hat{\mathbf{x}}, \mathbf{x}) \leq \epsilon_m\right\}. \tag{83}$$

Definition 6. The *unbounded noise level (UNL) estimator* $\hat{\mathbf{x}}_{\text{UN}}$ (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:

$$\hat{\mathbf{x}}_{\text{UN}} = \arg\max_{\hat{\mathbf{x}}\in\mathcal{E}} \hat{\sigma}^2(\hat{\mathbf{x}}). \tag{84}$$

We now show that, if the error function ϵ_{σ^2} is continuous in σ^2, then the UNL estimator is a minimax estimator. The error function is indeed continuous for many cases of interest, such as the MSE and the regret. This result enables us to efficiently find UNL estimators, including obtaining a closed form for the estimator in some important cases.

Proposition 6. *Suppose the error function ϵ of interest is continuous in σ^2. Then, the unbounded noise level (UNL) estimator $\hat{\mathbf{x}}_{\text{UN}}$ is a minimax estimator for the parameter set $\mathcal{U}$, with noise level $\sigma_1^2 = \hat{\sigma}^2(\hat{\mathbf{x}}_{\text{UN}})$.*

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

$$\max_{\mathbf{x}\in\mathcal{U}} \epsilon_{\sigma_1^2}(\hat{\mathbf{x}}_{\text{M}}, \hat{\mathbf{x}}) < \max_{\mathbf{x}\in\mathcal{U}} \epsilon_{\sigma_1^2}(\hat{\mathbf{x}}_{\text{UN}}, \mathbf{x}) \leq \epsilon_m. \tag{85}$$

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

$$\max_{\mathbf{x}\in\mathcal{U}} \epsilon_{\sigma_2^2}(\hat{\mathbf{x}}_{\text{M}}, \hat{\mathbf{x}}) \leq \epsilon_m. \tag{86}$$

Hence, $\hat{\sigma}^2(\hat{\mathbf{x}}_{\text{M}}) \geq \sigma_2^2 > \sigma_1^2 = \hat{\sigma}^2(\hat{\mathbf{x}}_{\text{UN}})$, contradicting the fact that $\hat{\mathbf{x}}_{\text{UN}}$ is a UNL estimator. □

A consequence of this proposition is that the UNL estimator can be found if an algorithm for finding the 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 proposition demonstrates, a closed form for the linear UNL estimator can be identified when a closed form for the minimax estimator is known.

Proposition 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 unbounded noise level (UNL) estimator is given by*

$$\hat{\mathbf{x}}_{\text{UN}} = \begin{cases} \frac{L^2-\epsilon_m}{L^2}\hat{\mathbf{x}}_{\text{LS}}, & L^2 > \epsilon_m \\ \mathbf{0}, & L^2 \leq \epsilon_m, \end{cases} \tag{87}$$

where $\hat{\mathbf{x}}_{\text{LS}} = (\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{y}$ is the LS estimator.

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 a 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 a UNL 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 Proposition 6, $\hat{\mathbf{x}}_{\text{UN}}$ 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 [6]

$$\hat{\mathbf{x}}_{\text{M}}(\sigma^2) = \frac{L^2}{L^2 + \sigma^2\gamma_0}(\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{y}, \tag{88}$$

where

$$\gamma_0 = \text{Tr}\left((\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H})^{-1}\right) \tag{89}$$

is the MSE of the LS estimator with $\sigma^2 = 1$. As we have seen in Eq. (40) of Lemma 1, the worst-case error for this estimator within the set $\mathcal{U}$ is given by

$$\max_{\mathbf{x}\in\mathcal{U}} \epsilon_{\sigma^2}(\hat{\mathbf{x}}_{\text{M}}(\sigma^2), \mathbf{x}) = \frac{L^2\sigma^2\gamma_0}{L^2 + \sigma^2\gamma_0}. \tag{90}$$

Requiring that this value not exceed ϵ_m, we obtain

$$\sigma^2\gamma_0(L^2 - \epsilon_m) \leq \epsilon_m L^2. \tag{91}$$

Since $L^2 > \epsilon_m$, this leads to

$$\sigma^2\gamma_0 \leq \frac{\epsilon_m L^2}{L^2 - \epsilon_m}. \tag{92}$$

Combining this with (89), we find that the maximum noise level for which performance requirements are satisfied is

$$\hat{\sigma}^2 = \frac{\epsilon_m L^2}{L^2 - \epsilon_m} \frac{1}{\text{Tr}\left((\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H})^{-1}\right)}. \tag{93}$$

Substituting this value of σ^2 into (88) yields the required estimator (87). □

It is instructive to compare the closed forms obtained for the UPS estimator (Proposition 4(b)) and the UNL estimator (Proposition 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 [17]. They can thus be viewed as a compromise between the least-squares estimator and the zero estimator. However, for the UPS estimator, the shrinkage factor increases with the maximum allowed error ϵ_m; while for the UNL 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 $\mathbf{x}$); while 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.

7 Discussion

Least-squares estimation, despite being very common, is not necessarily the optimal estimator in terms of mean-squared error. With very little additional information, alternative estimators can be found which outperform the LS estimator. For instance, in Lemma 1 we have shown that if the parameter $\mathbf{x}$ to be estimated is known to lie within some bounded parameter set, then there exists a (minimax MSE) estimator which outperforms LS for *any* value of $\mathbf{x}$ in this set.

Alternatively, it is often possible to define the maximum allowed estimation error. For instance, a system may be designed to function with a known and tolerable estimation error. In this case, one is interested in maximizing the parameter set for which the allowed error is guaranteed; this results in the unbounded parameter set (UPS) estimator. As we have seen in Propositions 3 and 5, in many cases, the UPS estimator is equivalent in form to minimax estimation. In other words, if a maximum allowed error ϵ_m is known, then the UPS estimator equals the minimax estimator whose worst-case error is ϵ_m.

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 Figure 2) 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 a UPS 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.

An alternative estimation strategy, explored in Section 6, attempts to maximize the noise level for which a required error level is maintained, for any $\mathbf{x}$ in a known parameter set $\mathcal{U}$. This approach leads to the unbounded noise level (UNL) estimator. As we have seen in Proposition 6, the UNL estimator also has the same form as the minimax estimator.

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 in Section 5, can often result in improved performance.

8 Acknowledgement

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] Y. Ben-Haim, *Information-Gap Decision Theory: Decisions under Severe Uncertainty.* Academic Press, 2001.

[2] Y. Ben-Haim, "Set-models of information-gap uncertainty: axioms and an inference scheme," J. Franklin Inst., 336: 1093–1117, 1999.

[3] A. Ben-Tal and A. Nemirovski, "Robust optimization: methodology and applications," to appear in Mathematical Programming (Series B).

[4] S. Boyd and L. Vanderberghe, *Convex Optimization.* Cambridge, 2004 (in print). http://www.stanford.edu/~boyd/cvxbook.html

[5] S. N. Crozier, D. D. Falconer, and S. A. Mahmoud, "Least sum of square errors (LSSE) channel estimation," IEE Proc.-F, 138 (4): 371–378, Aug. 1991.

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

[7] Y. C. Eldar, A. Ben-Tal and A. Nemirovski, "Linear minimax regret estimation with bounded data uncertainties," to appear in IEEE Trans. Signal Processing.

[8] C. Fragouli, N. Al-Dhahir, and W. Turin, "Training-based channel estimation for multiple-antenna broadband transmissions," IEEE Trans. Wireless Comm., 2 (2): 384–391, 2003.

[9] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," Technometrics, 12: 55–67, 1970.

[10] R. A. Horn and C. R. Johnson, *Matrix Analysis.* Cambridge University Press, 1985.

[11] P. J. Huber, "Robust estimation of a location parameter," Ann. Math. Statist., 35 (1): 73–101, 1964.

[12] P. J. Huber, *Robust Statistics.* Wiley, 1981.

[13] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: a survey," Proc. IEEE, 73: 433-481, 1985.

[14] S. A. Kassam and H. V. Poor, "Robust signal processing for communication systems," IEEE Comm. Mag., 21 (1): 20–28, 1983.

[15] S. M. Kay, *Fundumentals of Statistical Signal Processing: Estimation Theory.* Prentice Hall, 1993.

[16] K.-C. Li, "Minimaxity of the method of regularization on stochastic processes," Ann. Statist., 10 (3): 937–942, 1982.

[17] L. S. Mayer and T. A. Wilke, "On biased estimation in linear models," Technometrics, 15: 497–508, 1973.

[18] A. Papoulis, *Probability, Random Variables, and Stochastic Processes,* 3rd ed. McGraw-Hill, 1991.

[19] M. S. Pinsker, "Optimal filtering of square-integrable signals in Gaussian noise," Problems in Inform. Transmission, 16: 120-133, 1980.

[20] T. Rappaport, *Wireless Communications: Principles and Practice.* Prentice Hall, 1996.

[21] A. N. Tichonov and V. Y. Arsenin, *Solution of Ill-Posed Problems.* V. H. Winston, 1977.

[22] S. Verdú and H. V. Poor, "On minimax robustness: A general approach and applications," IEEE Trans. Inform. Theory, 30 (2): 328–340, 1984.

[23] N. Wiener, *The Extrapolation, Interpolation and Smoothing of Stationary Time Series.* Wiley, 1949.