# DOMINATING AND ADMISSIBLE MSE-BOUNDS USING SADDLE-POINT METHODS

*Yonina C. Eldar*

Department of Electrical Engineering
Technion–Israel Institute of Technology, Haifa, Israel
Email: yonina@ee.technion.ac.il.

**ABSTRACT**

We treat the problem of evaluating the performance of estimators for estimating a deterministic parameter vector $\mathbf{x}$, with the mean-squared error (MSE) as the performance measure. Since the MSE depends on the unknown vector $\mathbf{x}$, direct comparison between estimators is a difficult problem. Here we consider a framework for examining the MSE of different approaches based on the concepts of admissible and dominating estimators. Using a saddle-point framework we reduce these abstract concepts to a concrete convex optimization problem, which can then by analyzed by utilizing the machinery of convex optimization. Our development considers both the case of linear estimation in linear models as well as more general nonlinear models.

## 1. INTRODUCTION

An important estimation problem that has been treated extensively in the literature is that of estimating a deterministic parameter vector $\mathbf{x}$ from observations $\mathbf{y}$ that are related through the regression model

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

Here $\mathbf{H}$ is a known $n \times m$ full-rank matrix and $\mathbf{w}$ is a zero-mean random vector with positive covariance $\mathbf{C}_w$. In an estimation context, the goal is to construct an estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ from the observations $\mathbf{y}$ that is close to $\mathbf{x}$ in some sense. A popular measure of estimator performance is the mean-squared error (MSE), which is the average squared-norm of the estimation error $\hat{\mathbf{x}} - \mathbf{x}$. Since $\mathbf{x}$ is fixed, the averaging is only over the noise, typically resulting in a parameter-dependent MSE, which cannot be minimized directly. Thus, alternative criteria for constructing estimators must be sought.

More generally, we may treat the class of estimation problems in which the relationship between $\mathbf{y}$ and $\mathbf{x}$ is described by the probability density function (pdf) $p(\mathbf{y};\mathbf{x})$ of $\mathbf{y}$ characterized by $\mathbf{x}$. In this broader setting, it is difficult to explicitly evaluate the MSE of a given estimator, and therefore bounds on the MSE are typically considered. Traditional Cramér-Rao type bounds provide benchmarks on the variance of any estimator of $\mathbf{x}$ while requiring a priori specification of a desired bias gradient [1]. However, in practice, it is often not clear how to choose an optimal bias vector.

We begin by treating the linear estimation problem (1). A myriad of linear estimators have been developed for this model with the common goal of leading to "good" MSE performance. The difficulty encountered in this estimation problem is that the MSE of a linear estimator depends generally on $\mathbf{x}$, rendering comparison between the different approaches a difficult (and often impossible) task. An important practical question is how to decide which method to use. Although in general this question is hard to answer, some strategies may be uniformly better than others in terms of MSE. An estimator $\hat{\mathbf{x}}$ is said to *dominate* a given estimator $\hat{\mathbf{x}}_0$ on a set $\mathcal{U}$ if its MSE is never larger than that of $\hat{\mathbf{x}}_0$ for all $\mathbf{x}$ in $\mathcal{U}$, and is strictly smaller for some $\mathbf{x}$ in $\mathcal{U}$ [2]; $\hat{\mathbf{x}}$ *strictly dominates* $\hat{\mathbf{x}}_0$ if its MSE is smaller than that of $\hat{\mathbf{x}}_0$ for all $\mathbf{x}$ in $\mathcal{U}$. An estimator that is not dominated by any other method is said to be *admissible* on $\mathcal{U}$. Clearly, a desirable property of an estimator is that it is admissible.

In Section 2 we develop a general procedure for determining whether an estimator is admissible on an arbitrary constraint set $\mathcal{U}$. We also propose a method for constructing admissible estimators that strictly dominate any given inadmissible strategy. Our approach is based on reducing the abstract concepts of admissible and dominating estimators to a convex optimization problem. The advantage of this formulation is that we can use duality theory to develop explicit necessary and sufficient admissibility and domination conditions. To demonstrate the details of our method we consider, in Section 3, the case in which $\mathcal{U}$ consists of vectors $\mathbf{x}$ satisfying the weighted norm constraint $\mathbf{x}^*\mathbf{T}\mathbf{x} \leq U^2$ for some $\mathbf{T} \succ 0$ and scalar $U$. For this setting we present explicit necessary and sufficient conditions on a linear estimator to be admissible which are easy to verify. Given an inadmissible strategy, we also develop closed form expressions for linear dominating estimators in many special cases. In the general setting, we show that a dominating estimator can be found by solving a semidefinite programming problem (SDP) [3], which is a convex optimization problem that can be solved efficiently.

We extend the ideas developed in the context of linear estimation to more general nonlinear estimation problems in Section 4. Using the biased CRLB we derive a bound on the MSE of any estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ with a linear bias vector. By choosing the bias appropriately, we construct admissible and dominating MSE bounds, i.e. bounds that are lower than the CRLB for all values of $\mathbf{x}$. When an efficient unbiased estimator exists, we show that using a simple linear transformation of the efficient estimator results in a biased estimator that achieves our lower MSE bound, thus allowing us to reduce the MSE for all values of $\mathbf{x}$ by a linear transformation.

Proofs of results, which are omitted due to space limitations, can be found in [4, 5].

In the sequel, the notation $\mathbf{A} \succ 0$ ($\mathbf{A} \succeq 0$) means that $\mathbf{A}$ is Hermitian and positive (nonnegative) definite. Given a Hermitian matrix $\mathbf{A}$, we denote by $\mathbf{A}^{1/2}$ the Hermitian square-root, and by $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ the largest and smallest eigenvalues, respectively.

## 2. ADMISSIBLE AND DOMINATING ESTIMATORS

We begin by treating the problem of estimating the deterministic parameter vector $\mathbf{x}$ in the model (1). Our goal is to design a *linear* estimator of the form $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ that minimizes the MSE:

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

This work was supported in part by the EU 6th framework programme, via the NEWCOM network of excellence, and the Israel Science Foundation.

Evidently, the MSE generally depends on $\mathbf{x}$ and therefore cannot be minimized directly. An alternative strategy is to restrict $\hat{\mathbf{x}}$ to be unbiased, so that $\mathbf{GH} = \mathbf{I}$, and then seek the estimator that minimizes the variance, which leads to the celebrated least-squares estimator. However, this method does not necessarily result in a small MSE.

Two important concepts in comparing the MSE of different estimators are those of domination and admissibility. An estimator $\hat{\mathbf{x}}_1$ dominates an estimator $\hat{\mathbf{x}}_2$ on $\mathcal{U}$ if

$$\begin{aligned} E\left\{\|\hat{\mathbf{x}}_1 - \mathbf{x}\|^2\right\} &\le E\left\{\|\hat{\mathbf{x}}_2 - \mathbf{x}\|^2\right\}, \text{ for all } \mathbf{x} \in \mathcal{U}; \\ E\left\{\|\hat{\mathbf{x}}_1 - \mathbf{x}\|^2\right\} &< E\left\{\|\hat{\mathbf{x}}_2 - \mathbf{x}\|^2\right\}, \text{ for some } \mathbf{x} \in \mathcal{U}. \end{aligned} \quad (3)$$

The estimator $\hat{\mathbf{x}}_1$ strictly dominates $\hat{\mathbf{x}}_2$ on $\mathcal{U}$ if

$$E\left\{\|\hat{\mathbf{x}}_1 - \mathbf{x}\|^2\right\} < E\left\{\|\hat{\mathbf{x}}_2 - \mathbf{x}\|^2\right\}, \text{ for all } \mathbf{x} \in \mathcal{U}. \quad (4)$$

A linear estimator $\hat{\mathbf{x}}$ is *admissible* if it is not dominated by any other linear method. If $\hat{\mathbf{x}}$ is inadmissible, then there exists another linear approach which leads to lower MSE on $\mathcal{U}$. Thus, although we cannot directly compare the performance of different strategies, we would like our estimator to at least be admissible. This raises two important questions: Can we easily verify whether a given estimator $\hat{\mathbf{x}}_0$ is admissible on $\mathcal{U}$? Furthermore, if $\hat{\mathbf{x}}_0$ is inadmissible, then can we develop a method to construct an admissible linear estimator that dominates $\hat{\mathbf{x}}_0$ on $\mathcal{U}$?

The following theorem shows that both questions can be addressed by examining the solution to a convex optimization problem.

**Theorem 1.** *Let $\hat{\mathbf{x}}_0 = \mathbf{G}_0\mathbf{y}$ be a given linear estimate of $\mathbf{x}$ in the model (1), let $\epsilon(\mathbf{G}, \mathbf{x}) = E\{\|\mathbf{G}\mathbf{y} - \mathbf{x}\|^2\}$, and let $\mathcal{U} \subseteq \mathbb{C}^m$. Define*

$$\widehat{\mathbf{G}} = \arg\min_{\mathbf{G}} \sup_{\mathbf{x}\in\mathcal{U}} \left\{\epsilon(\mathbf{G}, \mathbf{x}) - \epsilon(\mathbf{G}_0, \mathbf{x})\right\}, \quad (5)$$

*and let $\hat{\mathbf{x}} = \widehat{\mathbf{G}}\mathbf{y}$. Then*

1. *$\widehat{\mathbf{G}}$ is unique;*
2. *$\hat{\mathbf{x}}_0$ is admissible on $\mathcal{U}$ if and only if $\widehat{\mathbf{G}} = \mathbf{G}_0$;*
3. *If $\widehat{\mathbf{G}} \neq \mathbf{G}_0$ then $\hat{\mathbf{x}}$ is admissible and strictly dominates $\hat{\mathbf{x}}_0$ on $\mathcal{U}$.*

For general choices of $\mathcal{U}$, iterative procedures that are designed for saddle point problems can be utilized to solve (5), such as subgradient algorithms [6] or the prox method [7]. We can also use the many known techniques for dealing with convex problems in order to develop admissibility and domination conditions, as we show in the next section.

## 3. DOMINATION AND ADMISSIBILITY ON AN ELLIPSOID

We now demonstrate the details of our approach by considering ellipsoidal uncertainty sets of the form $\mathcal{U} = \left\{\mathbf{x} \in \mathbb{C}^m \left| \mathbf{x}^*\mathbf{T}\mathbf{x} \le U^2\right.\right\}$, with $U > 0$ and $\mathbf{T} \succ 0$.

### 3.1. Dominating Estimators on an Ellipsoid

The first problem we address is that of constructing an admissible estimator $\hat{\mathbf{x}} = \widehat{\mathbf{G}}\mathbf{y}$ that strictly dominates an inadmissible method. From Theorem 1, $\hat{\mathbf{x}}$ can be found by solving (5) with the appropriate choice of $\mathcal{U}$. Using Schur's lemma [8, p. 28] and several manipulations, we can show that $\widehat{\mathbf{G}} = \widehat{\mathbf{B}}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$ where $\widehat{\mathbf{B}}$ is the solution to the SDP

$$\min_{\tau,\mathbf{B},\lambda\ge 0} \tau \quad (6)$$

subject to

$$\begin{bmatrix} \tau - U^2\lambda & \mathbf{b}^* \\ \mathbf{b} & \mathbf{I} \end{bmatrix} \succeq 0$$
$$\begin{bmatrix} \lambda\mathbf{I} + \mathbf{T}^{-1/2}\mathbf{A}\mathbf{T}^{-1/2} & \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{B})^* \\ (\mathbf{I}-\mathbf{B})\mathbf{T}^{-1/2} & \mathbf{I} \end{bmatrix} \succeq 0, \quad (7)$$

with $\mathbf{b} = \text{vec}((\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1/2}\mathbf{B}^*)$ and $\mathbf{A} = (\mathbf{I} - \mathbf{G}_0\mathbf{H})^*(\mathbf{I} - \mathbf{G}_0\mathbf{H})$. Thus, $\widehat{\mathbf{G}}$ can be computed efficiently using standard software packages.

To obtain more insight into the structure and properties of $\widehat{\mathbf{G}}$, we can use the Karush-Kuhn-Tucker (KKT) theory [9] to develop necessary and sufficient optimality conditions, which are given in the following theorem.

**Theorem 2.** *Consider the problem of Theorem 1. Let $\mathbf{A} = (\mathbf{I} - \mathbf{G}_0\mathbf{H})^*(\mathbf{I} - \mathbf{G}_0\mathbf{H})$, $\mathbf{Q} = \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ and $\mathcal{U} = \{\mathbf{x} \in \mathbb{C}^m | \mathbf{x}^*\mathbf{T}\mathbf{x} \le U^2\}$. Then, $\widehat{\mathbf{G}}$ of (5) has the form $\widehat{\mathbf{G}} = \mathbf{B}\mathbf{Q}^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$ where $\mathbf{B}$ satisfies the conditions below for some $\lambda \ge 0$:*

1. *$\mathbf{QB} = \mathbf{B}^*\mathbf{Q}$;*
2. *$0 \preceq \mathbf{QB} \prec \mathbf{Q}$;*
3. *$(\mathbf{I} - \mathbf{B})^*(\mathbf{I} - \mathbf{B}) \preceq \lambda\mathbf{T} + \mathbf{A}$;*
4. *$(\mathbf{I} - \mathbf{B})^*(\mathbf{I} - \mathbf{B})\mathbf{B} = (\lambda\mathbf{T} + \mathbf{A})\,\mathbf{B}$;*
5. *$Tr(\mathbf{B}\,(\mathbf{I} - \mathbf{B})^{-1}\,\mathbf{Q}^{-1}\mathbf{T}) \le U^2$;*
6. *$\lambda(Tr(\mathbf{B}\,(\mathbf{I} - \mathbf{B})^{-1}\,\mathbf{Q}^{-1}\mathbf{T}) - U^2) = 0$.*

The conditions of Theorem 2 can be used to verify a proposed solution. There are also cases in which these conditions can be solved explicitly. Two such classes of problems are jointly diagonalizable matrices and high SNR, as given in the following theorems.

**Theorem 3.** *Consider the problem of Theorem 2. Let $\mathbf{Q} = \mathbf{V}\Sigma\mathbf{V}^*$ where $\mathbf{V}$ is a unitary matrix and $\Sigma = diag(\sigma_1, \dots, \sigma_m)$ with $\sigma_i > 0$. Suppose that $\mathbf{T} = \mathbf{V}\Lambda\mathbf{V}^*$ and $\mathbf{G}_0\mathbf{H} = \mathbf{V}\Delta\mathbf{V}^*$ where $\Lambda = diag(\lambda_1, \dots, \lambda_m)$ with $\lambda_i > 0$ and $\Delta = diag(\delta_1, \dots, \delta_m)$. Then*

$$\widehat{\mathbf{G}} = \mathbf{V}\mathbf{D}\mathbf{V}^*(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$$

*where $\mathbf{D} = diag(d_1, \dots, d_m)$ with*

$$d_i = \begin{cases} 1 - \sqrt{\eta_i}, & \eta_i < 1; \\ 0, & \eta_i \ge 1. \end{cases}$$

*Here $\eta_i = \lambda\lambda_i + |1 - \delta_i|^2$, and $\lambda$ is determined as follows. Define*

$$\mathcal{G}(\lambda) = \sum_{i:\eta_i<1} \frac{\lambda_i}{\sigma_i}\left(\frac{1}{\sqrt{\eta_i}} - 1\right) - U^2.$$

*Then, $\lambda = 0$ if $\mathcal{G}(0) \le 0$. Otherwise, $\lambda$ is the unique value in $(0, \alpha)$ for which $\mathcal{G}(\lambda) = 0$, where $\alpha = \max_i(1 - |1 - \delta_i|^2)/\lambda_i$.*

We next treat the case in which the matrices are not necessarily commuting, however $\mathbf{A} \preceq \mathbf{I}$ and the SNR exceeds a threshold.

**Theorem 4.** *Consider the problem of Theorem 2. Let*

$$\mathbf{B} = \mathbf{I} - \left(\mathbf{Q}^{-1}(\lambda\mathbf{T} + \mathbf{A})\mathbf{Q}^{-1}\right)^{1/2}\mathbf{Q}, \quad (8)$$

*where $\lambda$ is chosen as follows. If $\mathbf{A}^{-1}$ is defined, and*

$$Tr\left(\mathbf{Q}^{-1}\left(\mathbf{Q}\mathbf{A}^{-1}\mathbf{Q}\right)^{1/2}\mathbf{Q}^{-1}\mathbf{T}\right) \le Tr\left(\mathbf{Q}^{-1}\mathbf{T}\right) + U^2$$

*then $\lambda = 0$. Otherwise, $\lambda > 0$ is the unique solution to*

$$Tr\left(\mathbf{Q}^{-1}\left(\mathbf{Q}(\lambda\mathbf{T} + \mathbf{A})^{-1}\mathbf{Q}\right)^{1/2}\mathbf{Q}^{-1}\mathbf{T}\right) = Tr\left(\mathbf{Q}^{-1}\mathbf{T}\right) + U^2.$$

*We then have that if $\lambda_{\max}(\mathbf{Q}(\mathbf{Q}^{-1}(\lambda\mathbf{T} + \mathbf{A})\mathbf{Q}^{-1})^{1/2}) \le 1$, then $\widehat{\mathbf{G}} = \mathbf{B}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$ with $\mathbf{B}$ given by (8).*

Theorem 4 can be used to develop a closed form solution for the minimax MSE estimator [10, 11] $\hat{\mathbf{x}}_{\rm MX} = \mathbf{G}_{\rm MX}\mathbf{y}$ that minimizes the worst-case MSE on $\mathcal{U}$, for high SNR. Specifically, by definition $\mathbf{G}_{\rm MX} = \arg\min_{\mathbf{G}} \max_{\mathbf{x}\in\mathcal{U}} \epsilon(\mathbf{G},\mathbf{x})$. For the unbiased linear estimator $\mathbf{G}_{\rm UB} = \mathbf{Q}^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$ the MSE, denoted $\epsilon(\mathbf{G}_{\rm UB})$, does not depend on $\mathbf{x}$. Therefore,

$$\mathbf{G}_{\rm MX} = \arg\min_{\mathbf{G}} \max_{\mathbf{x}\in\mathcal{U}} \{\epsilon(\mathbf{G},\mathbf{x}) - \epsilon(\mathbf{G}_{\rm UB})\}. \quad (9)$$

Using Theorem 4 with $\mathbf{G}_0 = \mathbf{G}_{\rm UB}$, it follows that if $\lambda_{\min}(\mathbf{Q}^{-1}(\mathbf{Q}\mathbf{T}^{-1}\mathbf{Q})^{1/2}) \geq \alpha$ where

$$\alpha = \frac{\mathrm{Tr}\left((\mathbf{Q}^{-1}\mathbf{T}\mathbf{Q}^{-1})^{1/2}\right)}{U^2 + \mathrm{Tr}(\mathbf{Q}^{-1}\mathbf{T})}, \quad (10)$$

then $\mathbf{G}_{\rm MX} = (\mathbf{I} - \alpha\left(\mathbf{Q}^{-1}\mathbf{T}\mathbf{Q}^{-1}\right)^{1/2}\mathbf{Q})\mathbf{Q}^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$.

In [11] a closed form expression for the minimax MSE estimator was derived for the case of commuting matrices. The minimax MSE estimator for general matrices was developed under a certain condition by Pilz in [12]. However, the solution of [12] is incorrect in general, and leads to a larger worst-case MSE then our estimator (See [5] for a detailed discussion).

### 3.2. Admissible Estimators on an Ellipsoid

We now use Theorem 2 to develop necessary and sufficient conditions on $\hat{\mathbf{x}}_0$ to be admissible on an ellipsoidal set $\mathcal{U}$.

**Theorem 5.** *Let $\hat{\mathbf{x}}_0 = \mathbf{G}_0\mathbf{y}$ be a given linear estimate of $\mathbf{x}$ in the model (1), and let $\mathbf{Q} = \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$. Then $\hat{\mathbf{x}}_0$ is admissible on $\mathcal{U} = \{\mathbf{x} \in \mathbb{C}^m | \mathbf{x}^*\mathbf{T}\mathbf{x} \leq U^2\}$ if and only if $\mathbf{G}_0 = \mathbf{B}\mathbf{Q}^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$ for a matrix $\mathbf{B}$ satisfying:*

1. $\mathbf{Q}\mathbf{B} = \mathbf{B}^*\mathbf{Q}$;
2. $0 \preceq \mathbf{Q}\mathbf{B} \prec \mathbf{Q}$;
3. *Tr* $\left(\mathbf{B}(\mathbf{I}-\mathbf{B})^{-1}\mathbf{Q}^{-1}\mathbf{T}\right) \leq U^2$.

*Proof.* The proof follows from showing that $\mathbf{B} = \mathbf{G}_0\mathbf{H}$ satisfies the conditions of Theorem 2 with $\mathbf{A} = (\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B})$ for some $\lambda \geq 0$ if and only if it satisfies conditions 1-3. □

#### 3.2.1. Examples

A popular estimator is the Tikhonov estimator:

$$\hat{\mathbf{x}}_{\rm TIK} = (\mathbf{H}\mathbf{C}_w^{-1}\mathbf{H} + \mathbf{M})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}, \quad (11)$$

for some $\mathbf{M} \succ 0$. From Theorem 5 it follows that $\hat{\mathbf{x}}_{\rm TIK}$ is admissible on $\mathcal{U} = \{\mathbf{x} \in \mathbb{C}^m | \mathbf{x}^*\mathbf{T}\mathbf{x} \leq U^2\}$ if and only if $\mathrm{Tr}(\mathbf{T}\mathbf{M}^{-1}) \leq U^2$.

Another common estimator is the shrunken estimator,

$$\hat{\mathbf{x}}_{\rm SH} = \alpha\hat{\mathbf{x}}_{\rm LS} = \alpha(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y} \quad (12)$$

for some $0 \leq \alpha < 1$. For this estimator, $\mathbf{B} = \alpha\mathbf{I}$, and therefore $\hat{\mathbf{x}}_{\rm SH}$ is admissible if and only if $0 \leq \alpha \leq U^2/(U^2 + \mathrm{Tr}(\mathbf{Q}^{-1}\mathbf{T}))$.

## 4. MSE BOUND FOR NONLINEAR ESTIMATION

We now consider the more general problem in which $\mathbf{x}$ is related to $\mathbf{y}$ through the pdf $p(\mathbf{y};\mathbf{x})$.

Under suitable regularity conditions on $p(\mathbf{y};\mathbf{x})$ [1], the MSE of any estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ with bias vector $\mathbf{b}(\mathbf{x}) = E\{\hat{\mathbf{x}}\} - \mathbf{x}$ is bounded below by

$$\|\mathbf{b}(\mathbf{x})\|^2 + \mathrm{Tr}\left((\mathbf{I}+\mathbf{D}(\mathbf{x}))\,\mathbf{J}^{-1}(\mathbf{x})\,(\mathbf{I}+\mathbf{D}(\mathbf{x}))^*\right), \quad (13)$$

where $\mathbf{D}(\mathbf{x}) = \partial\mathbf{b}(\mathbf{x})/\partial\mathbf{x}$ is the bias gradient matrix. Ideally, to obtain the tightest MSE bound, we would like to minimize (13) over all bias vectors $\mathbf{b}(\mathbf{x})$. For every fixed value $\mathbf{x} = \mathbf{x}_0$ of $\mathbf{x}$, the bound can be minimized by choosing $\mathbf{b}(\mathbf{x}_0) = \mathbf{x} - \mathbf{x}_0$; for this choice $\mathbf{b}(\mathbf{x}_0) = 0$ and $\mathbf{D}(\mathbf{x}_0) = -\mathbf{I}$. The estimator achieving this bound is $\hat{\mathbf{x}} = \mathbf{x}_0$ which clearly cannot be implemented since $\mathbf{x}_0$ is unknown. Thus, in general we cannot minimize the bound for all $\mathbf{x}$. Instead, we restrict our attention to linear bias vectors of the form $\mathbf{b}(\mathbf{x}) = \mathbf{M}\mathbf{x}$, for some $m \times m$ matrix $\mathbf{M}$. With this choice of bias, the MSE bound of (13) becomes

$$\begin{aligned}&\mathrm{MSEB}\,(\mathbf{M},\mathbf{x}) \\ &\quad= \mathbf{x}^*\mathbf{M}^*\mathbf{M}\mathbf{x} + \mathrm{Tr}\left((\mathbf{I}+\mathbf{M})\mathbf{J}^{-1}(\mathbf{x})(\mathbf{I}+\mathbf{M})^*\right). \end{aligned} \quad (14)$$

If $\mathbf{M} = 0$, then as we expect the MSE bound coincides with the CRLB: $\mathrm{MSEB}\,(0,\mathbf{x}) = \mathrm{Tr}(\mathbf{J}^{-1}(\mathbf{x}))$.

Our goal is to minimize the bound (14) over all $\mathbf{M}$. When the bound cannot be minimized directly, we seek an admissible and dominating choice of $\mathbf{M}$ such that the resulting MSE bound dominates the unbiased CRLB, and is not dominated by any other linear bias vector.

An advantage of restricting attention to linear bias vectors is that we can use results on unbiased estimation to find estimators that achieve the corresponding MSE bound. Specifically, if $\hat{\mathbf{x}}$ is an efficient unbiased estimator achieving the CRLB, then the MSE of

$$\hat{\mathbf{x}}_{\mathbf{b}} = (\mathbf{I}+\mathbf{M})\hat{\mathbf{x}} \quad (15)$$

is equal to $\mathrm{MSEB}\,(\mathbf{M},\mathbf{x})$. Therefore, if $\hat{\mathbf{x}}$ achieves the CRLB and we find an $\mathbf{M}$ such that $\mathrm{MSEB}\,(\mathbf{M},\mathbf{x}) < \mathrm{MSEB}\,(0,\mathbf{x})$ for a suitable set of $\mathbf{x}$, then the MSE of $\hat{\mathbf{x}}_{\mathbf{b}}$ will be smaller than that of $\hat{\mathbf{x}}$ for all $\mathbf{x}$ in the set. This allows us to reduce the MSE by a simple linear transformation. The important point is that this improvement is for *all choices of* $\mathbf{x}$.

### 4.1. Minimal MSE Bound with Linear Bias

We begin by discussing cases in which (14) can be minimized directly. Since $\mathrm{MSEB}\,(\mathbf{M},\mathbf{x})$ is convex in $\mathbf{M}$, we can find the optimal value by setting the derivative to 0, which yields

$$\mathbf{M} = -\mathbf{I} + \frac{1}{1+\mathbf{x}^*\mathbf{J}(\mathbf{x})\mathbf{x}}\mathbf{x}\mathbf{x}^*\mathbf{J}(\mathbf{x}). \quad (16)$$

In general $\mathbf{M}$ will depend on $\mathbf{x}$ which is unknown. However, if (16) is independent of $\mathbf{x}$, then this choice of $\mathbf{M}$ minimizes the bound for all $\mathbf{x}$. This occurs when $\mathbf{x} = x$ is a scalar, and $J^{-1}(x) = \alpha|x|^2$ for some $\alpha > 0$, which leads to the following theorem.

**Theorem 6.** *Let $\mathbf{y}$ denote measurements of a deterministic parameter $x$ with Fisher information $J(x) = 1/(\alpha|x|^2)$ for some $\alpha > 0$. Then the MSE of any estimate $\hat{x}$ of $x$ with linear bias satisfies*

$$E\left\{|\hat{x}-x|^2\right\} \geq \frac{\alpha}{1+\alpha}|x|^2. \quad (17)$$

*Furthermore, if there exists an efficient estimator $\hat{x}$ that achieves the CRLB, then*

$$\hat{x}_{\mathbf{b}} = \frac{1}{1+\alpha}\hat{x}$$

*achieves the bound (17), and has smaller MSE than $\hat{x}$ for all $x \neq 0$.*

We note that an estimator achieving the CRLB has minimum variance among all unbiased estimators. There are a variety of estimation problems in which the CRLB is not defined or cannot be achieved, but nonetheless a minimum variance unbiased (MVU) estimator can be found. The results of Theorem 6 and the discussion in the remainder of the paper hold true when we replace the CRLB $\mathbf{J}^{-1}(\mathbf{x})$ everywhere by the variance of an MVU estimator.

We now consider some examples of Theorem 6.

*Example 1.* Suppose we are given $N$ iid measurements $y_i$ that are distributed uniformly on $[0, x]$, and we wish to estimate $x$. The MVU estimator for this problem is $\hat{x} = (1 + 1/N)y_{\max}$, where $y_{\max} = \max_i y_i$, and its MSE is $\alpha x^2$ with $\alpha = \frac{1}{N(N+2)}$. From Theorem 6, the estimator $\hat{x}_{\mathbf{b}} = (N+2)/(N+1)y_{\max}$ has smaller MSE than $\hat{x}$ for all values of $x > 0$.

*Example 2.* Consider estimating the variance $\sigma^2$ of a Gaussian random variable with unknown mean $\mu$ from $N$ iid measurements $y_i$. In this problem the CRLB, which is given by $2\sigma^4/N$, cannot be achieved, but the estimator

$$\hat{x} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2 \tag{18}$$

with $\bar{y} = (1/N)\sum_{i=1}^{N} y_i$ is an MVU estimator with MSE equal to $2\sigma^4/(N-1)$. Applying Theorem 6 to (18) we conclude that

$$\hat{x}_{\mathbf{b}} = \frac{1}{N+1}\sum_{i=1}^{N}(y_i - \bar{y})^2 \tag{19}$$

has smaller MSE for all values of $\mu$ and $\sigma^2$.

*Example 3.* As a final example, we consider estimating the mean $x$ of an exponential random variable from $N$ iid measurements $y_i$. An efficient estimator is the ensemble average $\hat{x} = (1/N)\sum_{i=1}^{N} y_i$, whose MSE is $x^2/N$. From Theorem 6, the MSE of the estimator $\hat{x}_{\mathbf{b}} = 1/(N+1)\sum_{i=1}^{N} y_i$ is $x^2/(N+1)$, which is smaller than the CRLB for all $x > 0$.

### 4.2. Dominating the CRLB with Linear Bias

We now treat the case in which direct minimization of the MSE bound is not possible. Our goal then is to find an admissible dominating matrix $\mathbf{M}$ such that the resulting MSE bound is smaller than the unbiased CRLB $\mathbf{J}^{-1}(\mathbf{x})$ for all possible values of $\mathbf{x}$.

By adopting Theorem 1 to our problem we can show that an admissible dominating matrix can be found as a solution to a convex optimization problem:

**Theorem 7.** *Let* $\mathbf{y}$ *denote measurements of a deterministic vector* $\mathbf{x}$. *Let* $\mathrm{MSEB}(\mathbf{M}, \mathbf{x})$ *of (14) be the bound on the MSE of any estimate* $\hat{\mathbf{x}}$ *of* $\mathbf{x}$ *with bias* $\mathbf{b}(\mathbf{x}) = \mathbf{M}\mathbf{x}$*, and let* $\mathcal{U} \subseteq \mathbb{C}^m$. *Define*

$$\widehat{\mathbf{M}} = \arg\min_{\mathbf{M}} \sup_{\mathbf{x}\in\mathcal{U}} \left\{\mathrm{MSEB}(\mathbf{M}, \mathbf{x}) - \mathrm{MSEB}(0, \mathbf{x})\right\}. \tag{20}$$

*Then*

1. $\widehat{\mathbf{M}}$ *is unique;*
2. $\widehat{\mathbf{M}}$ *is admissible on* $\mathcal{U}$*;*
3. *If* $\widehat{\mathbf{M}} \neq 0$*, then* $\mathrm{MSEB}(\mathbf{M}, \mathbf{x}) < \mathrm{MSEB}(0, \mathbf{x})$ *on* $\mathcal{U}$.

It follows from Theorem 7 that if we find an $\widehat{\mathbf{M}} \neq 0$ that is the solution to (20), and if $\hat{\mathbf{x}}$ achieves the CRLB (or is an MVU estimator), then the MSE of $\hat{\mathbf{x}}_{\mathbf{b}} = (\mathbf{I} + \widehat{\mathbf{M}})\hat{\mathbf{x}}$ is smaller than that of $\hat{\mathbf{x}}$ for all $\mathbf{x} \in \mathcal{U}$; furthermore, no other estimator with linear bias exists that has a smaller (or equal) MSE than $\hat{\mathbf{x}}_{\mathbf{b}}$ for all $\mathbf{x} \in \mathcal{U}$.

For arbitrary forms of $\mathbf{J}^{-1}(\mathbf{x})$, $\widehat{\mathbf{M}}$ can be found by using one of the many known algorithms for solving minimax problems [6, 7]. In [4] we show that in the case in which $\mathbf{J}^{-1}(\mathbf{x})$ has the quadratic form

$$\mathbf{J}^{-1}(\mathbf{x}) = \sum_{i=1}^{\ell}\mathbf{B}_i\mathbf{x}\mathbf{x}^*\mathbf{B}_i^* + \sum_{i=1}^{k}(\mathbf{C}_i\mathbf{x}\mathbf{z}_i^* + \mathbf{z}_i\mathbf{x}^*\mathbf{C}_i^*) + \mathbf{A}, \tag{21}$$

for some matrices $\mathbf{A} \succeq 0, \mathbf{B}_i, \mathbf{C}_i$ and vectors $\mathbf{z}_i$, more efficient solutions are possible via an SDP formulation, and for some choices closed-form expressions for the optimal $\widehat{\mathbf{M}}$ can be developed. Besides leading to analytically tractable solutions, there are many cases in which $\mathbf{J}^{-1}(\mathbf{x})$ can be written in the form (21); see [4] for some examples.

## 5. CONCLUSION

We addressed the important issue of comparing the MSE performance of different estimators by relying on the concepts of domination and admissibility. We showed that both admissibility and domination can be treated by considering a certain convex optimization problem. The machinery of convex optimization can then be utilized to analyze the MSE performance for specific uncertainty sets.

Our results can be used to help select an appropriate estimator from the multitude of methods proposed in the literature. We also suggested methods for designing new estimators with lower MSE in the case in which the conventional approaches are inadmissible.

## 6. REFERENCES


[1] C. R. Rao, *Linear Statistical Inference and Its Applications*, New York, NY: John Wiley & Sons, Inc., second edition, 1973.

[2] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, Springer, New York, 2nd edition, 1999.

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

[4] Y. C. Eldar, "Uniformly improving the Cramér-Rao bound and maximum-likelihood estimation," CCIT Report #567, Dec. 2005, EE Dept., Technion–Israel Institute of Technology; to appear in *IEEE Trans. Signal Processing*.

[5] Y. C. Eldar, "Comparing between estimation approaches: Admissible and dominating linear estimators," CCIT Report #547, August 2005, EE Dept., Technion–Israel Institute of Technology; to appear in *IEEE Trans. Signal Processing*.

[6] T. Larsson, M. Patriksson, and A.-B. Strömberg, "On the convergence of conditional $\epsilon$-subgradient methods for convex programs and convex-concave saddle-point problems," *European J. Oper. Res.*, vol. 151, no. 3, pp. 461–473, 2003.

[7] A. Nemirovski, "Prox-method with rate of convergence $o(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems," *SIAM J. Opt.*, vol. 15, pp. 229–251, 2004.

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

[9] D. P. Bertsekas, *Nonlinear Programming*, Belmont MA: Athena Scientific, second edition, 1999.

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

[11] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Processing*, vol. 53, pp. 168–181, Jan. 2005.

[12] J. Pilz, "Minimax linear regression estimation with symmetric parameter restrictions," *J. Stat. Planning and Inference*, vol. 13, pp. 297–318, 1986.