MSE BOUNDS DOMINATING THE CRAMÉR-RAO BOUND

Yonina C. Eldar

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

ABSTRACT

Traditional Cramér-Rao type bounds provide benchmarks on the variance of any estimator of a deterministic parameter vector, while requiring a priori specification of a desired bias gradient. However, in applications, it is often not clear how to choose the required bias. A direct measure of the estimation error that takes both the variance and the bias into account is the mean-squared error (MSE). Here, we develop bounds on the MSE in estimating a deterministic vector $\mathbf{x}_0$ using estimators with linear bias vectors, which includes the traditional unbiased estimation as a special case. We show that there often exists linear bias vectors that result in an MSE bound that dominates the CRLB, so that it is smaller than the CRLB for all $\mathbf{x}_0$. Furthermore, we explicitly construct estimators that achieve these bounds by linearly transforming the maximum-likelihood estimator.

1. INTRODUCTION

One of the prime goals of statistical estimation theory is the development of bounds on the best achievable performance in estimating parameters of interest in a given model, as well as determining estimators that achieve these bounds. Such bounds provide benchmarks against which we can compare the performance of any proposed estimator, and insight into the fundamental limitations of the problem.

Here, we consider the class of estimation problems in which we seek to estimate a deterministic parameter vector $\mathbf{x}_0$ from measurements $\mathbf{y}$. The relationship between $\mathbf{y}$ and $\mathbf{x}_0$ is described by the probability density function (pdf) $p(\mathbf{y};\mathbf{x}_0)$ of $\mathbf{y}$ characterized by $\mathbf{x}_0$.

A classic performance bound is the Cramér-Rao lower bound (CRLB) [1], which characterizes the smallest achievable total variance of any *unbiased* estimator of $\mathbf{x}_0$. When the measurements $\mathbf{y}$ are related to the unknowns $\mathbf{x}_0$ through a linear Gaussian model, the maximum likelihood (ML) estimate of $\mathbf{x}_0$, which is given by the value of $\mathbf{x}$ that maximizes $p(\mathbf{y};\mathbf{x})$, achieves the CRLB. Furthermore, when $\mathbf{x}_0$ is estimated from independent identically distributed (iid) measurements, under suitable regularity assumptions, the ML estimator is asymptotically unbiased and achieves the CRLB [1].

Although the CRLB is a popular performance benchmark, it only provides a bound on the variance of the estimator assuming a zero bias. In many cases the variance can be made smaller at the expense of increasing the bias, while ensuring that the overall estimation error is reduced. The total variance of any estimator with a given bias is bounded by the *biased CRLB* [2]. The specification of the biased CRLB requires an a-priori choice of the bias gradient. However, in applications it is typically not obvious how to make such a choice. Instead, it would be desirable to obtain a bound directly on the estimation error $\hat{\mathbf{x}} - \mathbf{x}_0$, where $\hat{\mathbf{x}}$ is an estimate of $\mathbf{x}_0$, without having to pre-specify the bias. To characterize the best possible bias-variance tradeoff we may consider the mean-squared error (MSE), which is the sum of the total variance and the squared-norm of the bias. Using the biased CRLB which bounds the total variance, we can readily obtain a bound on the MSE of any estimator $\hat{\mathbf{x}}$ with a given bias vector: if $\hat{\mathbf{x}}$ is an arbitrary estimator of $\mathbf{x}_0$ with bias vector $\mathbf{b}(\mathbf{x}_0)$ then its MSE is bounded below by

$$\|\mathbf{b}(\mathbf{x}_0)\|^2 + \operatorname{Tr}\left((\mathbf{I}+\mathbf{D}(\mathbf{x}_0))\,\mathbf{J}^{-1}(\mathbf{x}_0)\,(\mathbf{I}+\mathbf{D}(\mathbf{x}_0))^*\right), \quad (1)$$

where $\mathbf{J}^{-1}(\mathbf{x}_0)$ is the Fisher information matrix[1], and $\mathbf{D}(\mathbf{x}_0) = \partial\mathbf{b}(\mathbf{x}_0)/\partial\mathbf{x}$ is the bias gradient matrix. Ideally, to obtain the tightest possible MSE bound, we would like to minimize (1) over all bias vectors $\mathbf{b}(\mathbf{x}_0)$. Unfortunately, since no limitations are imposed on $\hat{\mathbf{x}}$, the minimal bound is the trivial (zero) bound which can be achieved with $\hat{\mathbf{x}} = \mathbf{x}_0$.

Instead of attempting to minimize the MSE over all possible estimators, we may restrict attention to estimators with bias vectors that lie in a suitable class; the CRLB is an example where we consider only estimators with zero bias. More generally, we may allow for bias vectors that are linear in $\mathbf{x}_0$ so that $\mathbf{b}(\mathbf{x}_0) = \mathbf{M}\mathbf{x}_0$ for some matrix $\mathbf{M}$. The MSE bound of (1) then becomes

$$\begin{aligned}&\mathrm{MSEB}\,(\mathbf{M},\mathbf{x}_0)\\ &= \quad \mathbf{x}_0^*\mathbf{M}^*\mathbf{M}\mathbf{x}_0 + \operatorname{Tr}\left((\mathbf{I}+\mathbf{M})\mathbf{J}^{-1}(\mathbf{x}_0)(\mathbf{I}+\mathbf{M})^*\right). \quad (2)\end{aligned}$$

Our goal then is to find the smallest MSE achievable over all estimators with linear bias.

An advantage of this class of 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 estimator, *i.e.,* an unbiased estimator that achieves the CRLB (corresponding to $\mathbf{M} = 0$), then the MSE of the estimator $\hat{\mathbf{x}}_{\mathbf{b}} = (\mathbf{I}+\mathbf{M})\hat{\mathbf{x}}$ is equal to $\mathrm{MSEB}\,(\mathbf{M},\mathbf{x}_0)$. Therefore, if $\hat{\mathbf{x}}$ achieves the CRLB and we find an $\mathbf{M}$ such that $\mathrm{MSEB}\,(\mathbf{M},\mathbf{x}_0) < \mathrm{MSEB}\,(0,\mathbf{x}_0)$ for a suitable set of $\mathbf{x}_0$, then the MSE of $\hat{\mathbf{x}}_{\mathbf{b}}$ will be smaller than that of $\hat{\mathbf{x}}$ for all $\mathbf{x}_0$ in the set. This allows us to reduce the MSE by a simple linear transformation. In particular, since under suitable regularity conditions the ML estimator is asymptotically efficient, our results will allow us to improve upon the ML estimator asymptotically for *all choices of* $\mathbf{x}_0$ in a suitable set using a linear transformation.

In Section 2 we discuss cases in which the MSE bound can be minimized directly over all linear bias vectors. For the more general setting we propose, in Section 3, concrete methods for finding a linear bias vector such that the resulting MSE bound is smaller than the CRLB for all values $\mathbf{x}_0$. To this end we first show that such a bias vector can be obtained as a solution to a convex optimization problem. We then restrict our attention to estimation problems in which

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

[1] We assume throughout that the Fisher information matrix is invertible.

the CRLB is quadratic in $\mathbf{x}_0$, and analyze the resulting problem for two special cases: unrestricted choices of $\mathbf{x}_0$ (Section 4) and the case where $\mathbf{x}_0$ lies in a quadratic set (Section 5). In both settings we show that a linear bias vector exists such that the resulting MSE bound is smaller than the CRLB for all feasible values of $\mathbf{x}_0$. This vector can be found as a solution to a semidefinite programming problem (SDP) which is a tractable convex problem that can be solved very efficiently. We then develop necessary and sufficient optimality conditions which lead to further insight into the optimal solution and in some cases can be used to derive explicit closed-form solutions for the optimal bias vector. In Section 6 we demonstrate through an example that by a simple linear transformation of the ML estimator, we can reduce the MSE.

Proofs of results, which are omitted due to space limitations, can be found in [3].

2. MINIMAL MSE BOUND WITH LINEAR BIAS

We begin by discussing cases in which the bound of (2) can be minimized directly. Since $\text{MSEB}\,(\mathbf{M},\mathbf{x})$ is convex in $\mathbf{M}$, we can find the optimal value by setting the derivative to 0, which yields

$$\mathbf{M}(\mathbf{J}^{-1}(\mathbf{x})+\mathbf{x}\mathbf{x}^*) = -\mathbf{J}^{-1}(\mathbf{x}). \tag{3}$$

Using the matrix inversion lemma, $\mathbf{M}$ can be written explicitly as

$$\mathbf{M} = -\mathbf{I} + \frac{1}{1+\mathbf{x}^*\mathbf{J}(\mathbf{x})\mathbf{x}}\mathbf{x}\mathbf{x}^*\mathbf{J}(\mathbf{x}). \tag{4}$$

In general $\mathbf{M}$ will depend on $\mathbf{x}$ which is unknown. However, if (4) 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 1. *Let* $\mathbf{y}$ *denote measurements of a deterministic parameter* x_0 *with Fisher information* $J(x_0) = 1/(\alpha|x_0|^2)$ *for some* $\alpha > 0$. *Then the MSE of any estimate* $\hat{x}$ *of* x_0 *with linear bias satisfies*

$$E\left\{|\hat{x}-x_0|^2\right\} \geq \frac{\alpha}{1+\alpha}|x_0|^2. \tag{5}$$

Furthermore, if there exists an efficient estimator $\hat{x}$ *that achieves the CRLB* $J^{-1}(x_0) = \alpha|x_0|^2$, *then*

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

achieves the bound (5), and has smaller MSE than $\hat{x}$ *for all* $x_0 \neq 0$.

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 1 and the discussion in the remainder of the paper hold true when we replace the CRLB $\mathbf{J}^{-1}(\mathbf{x}_0)$ everywhere by the variance of an MVU estimator. In this case, by linearly transforming the MVU estimator, we can reduce its MSE uniformly.

We now consider some examples of Theorem 1.

Example 1. Suppose we are given N iid measurements y_i that are distributed uniformly on $[0, x_0]$, and we wish to estimate x_0. Although the CRLB in this problem is not defined, an MVU estimator exists and is given by $\hat{x} = (1+1/N)y_{\max}$, where $y_{\max} = \max_i y_i$. The MSE of $\hat{x}$, which is the minimum variance achievable with an unbiased estimator, is αx_0^2 with $\alpha = \frac{1}{N(N+2)}$. From Theorem 1, the estimator

$$\hat{x}_{\mathbf{b}} = \frac{N+2}{N+1}y_{\max} \tag{6}$$

has smaller MSE than $\hat{x}$ for all values of $x_0 > 0$.

Example 2. Consider the problem of estimating the variance σ^2 of a Gaussian random variable with unknown mean μ 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

$$\hat{x} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{y})^2 \tag{7}$$

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 1 to (7) we conclude that

$$\hat{x}_{\mathbf{b}} = \frac{1}{N+1}\sum_{i=1}^{N}(y_i-\bar{y})^2 \tag{8}$$

has smaller MSE for all values of μ and σ^2. This result has also been obtained in [1, p. 316]. The MSE of $\hat{x}_{\mathbf{b}}$ is $2\sigma^4/(N+1)$ which is smaller than the CRLB for all $\sigma^2 > 0$.

Example 3. Suppose we wish to estimate the mean x_0 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_0^2/N. From Theorem 1, the MSE of the estimator

$$\hat{x}_{\mathbf{b}} = \frac{1}{N+1}\sum_{i=1}^{N} y_i \tag{9}$$

is $x_0^2/(N+1)$, which is smaller than the CRLB for all $x_0 > 0$.

3. DOMINATING THE CRLB WITH LINEAR BIAS

We have seen in the previous section that in some special cases we can directly minimize the MSE over all linear bias vectors. Even when direct minimization is not possible, we may still be able to find a matrix $\mathbf{M}$ such that the resulting MSE bound is smaller than the unbiased CRLB $\mathbf{J}^{-1}(\mathbf{x}_0)$ for all possible values of $\mathbf{x}_0$.

Thus, our goal now is to find a matrix $\mathbf{M}$ such that

$$\text{MSEB}\,(\mathbf{M},\mathbf{x}_0) < \text{MSEB}\,(0,\mathbf{x}_0) = \mathbf{J}^{-1}(\mathbf{x}_0) \tag{10}$$

for all values of $\mathbf{x}_0$ in some set $\mathcal{U}$. If the matrix $\mathbf{M}$ satisfies (10), then we will say that $\mathbf{M}$ (strictly) *dominates* the CRLB on $\mathcal{U}$ [4]. This will ensure that if $\hat{\mathbf{x}}$ is an efficient estimator, then the estimator $\hat{\mathbf{x}}_{\mathbf{b}} = (\mathbf{I}+\mathbf{M})\hat{\mathbf{x}}$ will have smaller MSE than $\hat{\mathbf{x}}$ for all values of $\mathbf{x}_0 \in \mathcal{U}$. In addition to satisfying (10), we would like $\mathbf{M}$ to have the property that there is no other matrix $\mathbf{M}' \neq \mathbf{M}$ such that

$$\text{MSEB}\,(\mathbf{M}',\mathbf{x}_0) \leq \text{MSEB}\,(\mathbf{M},\mathbf{x}_0) \tag{11}$$

for all $\mathbf{x}_0$ in $\mathcal{U}$. Such a matrix $\mathbf{M}$ will be called *admissible* [4]. Our problem therefore is to find an admissible $\mathbf{M}$ that dominates the CRLB on $\mathcal{U}$. It turns out that an admissible dominating matrix can be found as a solution to a convex optimization problem:

Theorem 2. *Let* $\mathbf{y}$ *denote measurements of an unknown deterministic parameter vector* $\mathbf{x}_0$. *Let* $\text{MSEB}\,(\mathbf{M},\mathbf{x}_0)$ *of (2) denote the bound on the MSE of any estimate* $\hat{\mathbf{x}}$ *of* $\mathbf{x}_0$ *with linear bias* $\mathbf{b}(\mathbf{x}_0) = \mathbf{M}\mathbf{x}_0$, *and let* $\mathcal{U} \subseteq \mathbb{C}^m$. *Define*

$$\widehat{\mathbf{M}} = \arg\min_{\mathbf{M}}\sup_{\mathbf{x}\in\mathcal{U}}\left\{\text{MSEB}\,(\mathbf{M},\mathbf{x}) - \text{MSEB}\,(0,\mathbf{x})\right\}. \tag{12}$$

Then

1. $\widehat{\mathbf{M}}$ *is unique;*
2. $\widehat{\mathbf{M}}$ *is admissible on* $\mathcal{U}$*;*

3. If $\widehat{\mathbf{M}} \neq 0$*, then* $\text{MSEB}(\mathbf{M}, \mathbf{x}) < \text{MSEB}(0, \mathbf{x})$ *on* $\mathcal{U}$.

Note that the minimum in (12) is well defined since the objective is continuous and coercive [5].

Proof. The proof follows immediately from the proof of [6, Theorem 1] by noting that $\text{MSEB}(\mathbf{M}, \mathbf{x})$ is continuous, coercive and strictly convex in $\mathbf{M}$. □

It follows from Theorem 2 that if we find an $\widehat{\mathbf{M}} \neq 0$ that is the solution to (12), 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 $\mathbf{x}_0 \in \mathcal{U}$; furthermore, no other estimator with linear bias exists that has a smaller (or equal) MSE than $\hat{\mathbf{x}}_{\mathbf{b}}$ for all $\mathbf{x}_0 \in \mathcal{U}$.

For arbitrary forms of $\mathbf{J}^{-1}(\mathbf{x}_0)$, $\widehat{\mathbf{M}}$ can be found by using one of the many known algorithms for solving minimax problems, such as subgradient algorithms [7] or the prox method [8]. In the remainder of the paper we treat the problem (12) in the case in which $\mathbf{J}^{-1}(\mathbf{x}_0)$ has the quadratic form

$$\mathbf{J}^{-1}(\mathbf{x}) = \sum_{i=1}^{\ell} \mathbf{B}_i \mathbf{x}\mathbf{x}^* \mathbf{B}_i^* + \sum_{i=1}^{k} (\mathbf{C}_i \mathbf{x} \mathbf{z}_i^* + \mathbf{z}_i \mathbf{x}^* \mathbf{C}_i^*) + \mathbf{A}, \quad (13)$$

for some matrices $\mathbf{A} \succeq 0$, $\mathbf{B}_i, \mathbf{C}_i$ and vectors $\mathbf{z}_i$, where $\mathbf{A} \succeq 0$ means that $\mathbf{A}$ is Hermitian and positive semidefinite. As we show, this will allow for efficient solutions and in some cases, a closed-form expression for the optimal $\mathbf{M}$. Besides leading to analytically tractable solutions, there are many cases in which $\mathbf{J}^{-1}(\mathbf{x}_0)$ can be written in the form (13). Several examples are presented below.

Example 1 [9]. Consider estimating the mean μ and variance σ^2 of a Gaussian random variable from N iid measurements y_i. In this case $\mathbf{x} = [\mu \quad \sigma^2]^T$, and

$$\mathbf{J}^{-1}(\mathbf{x}) = \frac{\sigma^2}{N} \begin{bmatrix} 1 & 0 \\ 0 & 2\sigma^2 \end{bmatrix}, \quad (14)$$

which has the form (13). If μ is Gaussian with zero mean and variance σ_μ^2, then

$$\mathbf{J}^{-1}(\sigma_\mu^2) = 2\left(\sigma_\mu^2 + \frac{\sigma^2}{N}\right)^2, \quad (15)$$

which again can be written as in (13).

Example 2. Suppose that the observation vector $\mathbf{y}$ is a vector of counts with mean $\mathbf{g}(\mathbf{x}) = \mathbf{H}\mathbf{x} + \mathbf{c}$ for some invertible matrix $\mathbf{H}$ and constant vector $\mathbf{c}$. The elements y_i of $\mathbf{y}$ are assumed to be independent, with a Poisson distribution

$$\ln f(y_i; \mathbf{x}) = y_i \ln\Big(g_i(\mathbf{x})\Big) - g_i(\mathbf{x}) + a,$$

where a is a constant. This problem arises for example in emission-computed tomography [10]. In this case,

$$\mathbf{J}^{-1}(\mathbf{x}) = \mathbf{H}^{-1}\text{diag}\Big(g_1(\mathbf{x}), \ldots, g_m(\mathbf{x})\Big)(\mathbf{H}^{-1})^*, \quad (16)$$

which can also be expressed in the form (13).

Example 3. As a final example, suppose that $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where $\mathbf{w}$ is a zero-mean Gaussian vector with covariance $\mathbf{C}$. In this case, $\mathbf{J}^{-1}(\mathbf{x}) = (\mathbf{H}^*\mathbf{C}^{-1}\mathbf{H})^{-1}$, which also has the form (13).

In Section 4 we treat the problem (12) with $\mathbf{J}^{-1}(\mathbf{x})$ given by (13) when $\mathcal{U} = \mathbb{C}^m$ so that there is no restriction on $\mathbf{x}_0$. We show that the optimal $\mathbf{M}$ can be found as a solution to an SDP, which is a convex optimization problem that involves minimizing a linear function subject to linear matrix inequalities, *i.e.,* matrix inequalities of the form $\mathbf{G}(\mathbf{M}) \succeq 0$ where $\mathbf{G}(\mathbf{M})$ is linear in $\mathbf{M}$ [11]. Once a problem is formulated as an SDP, standard software packages, such as the Self-Dual-Minimization (SeDuMi) package [12], can be used to solve the problem in polynomial time within any desired accuracy. We also develop necessary and sufficient optimality conditions on $\mathbf{M}$ that lead to further insight into the solution.

In some settings, we may have additional information on $\mathbf{x}_0$ which can result in a lower MSE bound. For example, we may know that the norm of $\mathbf{x}_0$ is bounded. There may also be natural restrictions on the parameters, for example if x_0 represents the variance or the SNR, then $x_0 \geq 0$. More generally, x_0 may lie in a specified interval $\alpha \leq x_0 \leq \beta$. These constraints can all be viewed as special cases of the quadratic constraint $\mathbf{x}_0 \in \mathcal{Q}$ where

$$\mathcal{Q} = \{\mathbf{x} | \mathbf{x}^* \mathbf{A}_1 \mathbf{x} + 2\mathbf{b}_1^* \mathbf{x} + c_1 \leq 0\}, \quad (17)$$

for some $\mathbf{A}_1, \mathbf{b}_1$ and c_1. In Section 5, we discuss the scenario in which $\mathbf{x}_0 \in \mathcal{Q}$, and show that again an admissible dominating $\mathbf{M}$ can be found by solving an SDP. Note that we do not require that $\mathbf{A}_1 \succeq 0$ so that the constraint set (17) is not necessarily convex.

4. DOMINATING BOUND ON THE ENTIRE SPACE

We first treat the case in which $\mathcal{U} = \mathbb{C}^m$. With $\mathbf{J}^{-1}(\mathbf{x}_0)$ given by (13), the MSE bound of (2) can be written compactly as

$$\text{MSEB}(\mathbf{M}, \mathbf{x}_0) = \mathbf{x}^* \mathbf{A}_0(\mathbf{M})\mathbf{x} + 2\Re\{\mathbf{b}_0^*(\mathbf{M})\mathbf{x}\} + c_0(\mathbf{M}), \quad (18)$$

where we defined

$$\begin{aligned} \mathbf{A}_0(\mathbf{M}) &= \mathbf{M}^*\mathbf{M} + \sum_{i=1}^{\ell} \mathbf{B}_i^* \left((\mathbf{I}+\mathbf{M})^*(\mathbf{I}+\mathbf{M}) - \mathbf{I}\right)\mathbf{B}_i; \\ \mathbf{b}_0(\mathbf{M}) &= \sum_{i=1}^{k} \mathbf{C}_i^* \left((\mathbf{I}+\mathbf{M})^*(\mathbf{I}+\mathbf{M}) - \mathbf{I}\right)\mathbf{z}_i; \\ c_0(\mathbf{M}) &= \text{Tr}\left(((\mathbf{I}+\mathbf{M})^*(\mathbf{I}+\mathbf{M}) - \mathbf{I})\,\mathbf{A}\right). \end{aligned} \quad (19)$$

From Theorem 2, an admissible dominating matrix $\mathbf{M}$ can then be found as the solution to

$$\min_{\mathbf{M}} \max_{\mathbf{x}} \{\mathbf{x}^* \mathbf{A}_0(\mathbf{M})\mathbf{x} + 2\Re\{\mathbf{b}_0^*(\mathbf{M})\mathbf{x}\} + c_0(\mathbf{M})\}, \quad (20)$$

which can alternatively be written as $\min_{t,\mathbf{M}} t$ subject to

$$\mathbf{x}^* \mathbf{A}_0(\mathbf{M})\mathbf{x} + 2\Re\{\mathbf{b}_0^*(\mathbf{M})\mathbf{x}\} + c_0(\mathbf{M}) \leq t, \quad \text{for all } \mathbf{x}. \quad (21)$$

The constraint (21) is equivalent to [13, p. 163]

$$\mathbf{G}(\mathbf{M}) \stackrel{\triangle}{=} \begin{bmatrix} \mathbf{A}_0(\mathbf{M}) & \mathbf{b}_0(\mathbf{M}) \\ \mathbf{b}_0^*(\mathbf{M}) & c_0(\mathbf{M}) - t \end{bmatrix} \preceq 0. \quad (22)$$

Since $\mathbf{M} = 0, t = 0$ satisfies (22), our problem is always feasible. It can be shown that (22) is strictly feasible if and only if $\sum_{i=1}^{\ell} \mathbf{B}_i^* \mathbf{B}_i \succ 0$. If (22) is not strictly feasible then, as we show in [3], it can always be reduced to a strictly feasible problem with additional linear constraints on $\mathbf{M}$. A similar approach to that taken here can then be followed for the reduced problem. Therefore, we assume strict feasibility in the remainder of the paper.

4.1. SDP Formulation of the Problem

The constraint (22) is not written in convex form. Fortunately, in can be converted into a convex constraint, as incorporated in the following lemma:

Lemma 1. *The problem* $\min_{t,\mathbf{M}} \{t : \mathbf{G}(\mathbf{M}) \preceq 0\}$ *with* $\mathbf{G}(\mathbf{M})$ *given by (22) is equivalent to the convex problem*

$$\min_{t,\mathbf{M},\mathbf{X}} \{t : \mathbf{Z}(\mathbf{M},\mathbf{X}) \preceq 0,\ \mathbf{M}^*\mathbf{M} \preceq \mathbf{X}\}, \tag{23}$$

where

$$\mathbf{Z}(\mathbf{M},\mathbf{X}) = \begin{bmatrix} \mathbf{X} + \sum_{i=1}^{\ell} \mathbf{B}_i^*\Phi\mathbf{B}_i & \sum_{i=1}^{k} \mathbf{C}_i^*\Phi\mathbf{z}_i \\ \sum_{i=1}^{k} \mathbf{z}_i^*\Phi\mathbf{C}_i & Tr\,(\mathbf{A}\Phi) - t \end{bmatrix}, \tag{24}$$

and for brevity we denoted $\Phi = \mathbf{X} + \mathbf{M} + \mathbf{M}^*$.

The matrix $\mathbf{Z}(\mathbf{M},\mathbf{X})$ is linear in both $\mathbf{M}$ and $\mathbf{X}$ so that (24) is a linear matrix inequality. Using Schur's Lemma [13] the constraint $\mathbf{M}^*\mathbf{M} \preceq \mathbf{X}$ can be written as

$$\begin{bmatrix} \mathbf{X} & \mathbf{M}^* \\ \mathbf{M} & \mathbf{I} \end{bmatrix} \succeq 0, \tag{25}$$

which is also linear. Therefore, using (23) the optimal $\mathbf{M}$ can be found as a solution to an SDP.

4.2. Dual Problem

To gain more insight into the form of the optimal $\mathbf{M}$, and to provide an alternative method of solution which in some cases may admit a closed-from solution, we now rely on Lagrange duality theory.

Since the problem (23) is convex and strictly feasible, its optimal value is equal to the optimal value of the dual problem. To find the dual, we first write the Lagrangian associated with our problem:

$$\mathcal{L} = t + \mathrm{Tr}(\widetilde{\Pi}\mathbf{Z}(\mathbf{M},\mathbf{X})) + \mathrm{Tr}\,(\Delta(\mathbf{M}^*\mathbf{M} - \mathbf{X})) \tag{26}$$

where $\Delta \succeq 0$ and

$$\widetilde{\Pi} = \begin{bmatrix} \Pi & \mathbf{w} \\ \mathbf{w}^* & \pi \end{bmatrix} \succeq 0 \tag{27}$$

are the dual variables. Differentiating with respect to t and equating to 0, $\pi = 1$. Differentiating with respect to $\mathbf{X}$ and equating to 0,

$$\Delta = \Pi + \mathbf{S}(\Pi,\mathbf{w}), \tag{28}$$

where we defined

$$\mathbf{S}(\Pi,\mathbf{w}) = \sum_{i=1}^{\ell} \mathbf{B}_i\Pi\mathbf{B}_i^* + \sum_{i=1}^{k}(\mathbf{z}_i\mathbf{w}^*\mathbf{C}_i^* + \mathbf{C}_i\mathbf{w}\mathbf{z}_i^*) + \mathbf{A}. \tag{29}$$

Finally, the derivative with respect to $\mathbf{M}$ combined with (28) yields

$$\mathbf{M}\,(\mathbf{S}(\Pi,\mathbf{w}) + \Pi) = -\mathbf{S}(\Pi,\mathbf{w}). \tag{30}$$

The condition $\widetilde{\Pi} \succeq 0$ implies that $\Pi \succeq \mathbf{w}\mathbf{w}^*$. Therefore,

$$\mathbf{S}(\Pi,\mathbf{w}) \succeq \mathbf{S}(\mathbf{w}\mathbf{w}^*,\mathbf{w}) = \mathbf{J}^{-1}(\mathbf{w}) \succ 0, \tag{31}$$

and $\mathbf{S}$ is invertible. Thus, from (30),

$$\mathbf{M} = -\mathbf{S}(\Pi,\mathbf{w})\,(\mathbf{S}(\Pi,\mathbf{w}) + \Pi)^{-1}. \tag{32}$$

An important observation from (32) is that regardless of Π, $\mathbf{M}$ is not equal 0. Therefore, from Theorem 2 it follows that as long as the problem is strictly feasible, we can improve the CRLB for all values of $\mathbf{x}_0$ by a linear transformation. Using similar tools to those used here we can show that the optimal $\mathbf{M}$ is not zero also when the problem is not strictly feasible, as long as $\mathbf{B}_i \neq 0$ for some i.

Substituting (32) into the Lagrangian, the dual problem becomes

$$\min_{\mathbf{w},\Pi} \mathrm{Tr}\left(\mathbf{S}(\Pi,\mathbf{w})\,(\mathbf{S}(\Pi,\mathbf{w}) + \Pi)^{-1}\,\mathbf{S}(\Pi,\mathbf{w})\right), \tag{33}$$

subject to (27), which can be written as the SDP $\min_{\mathbf{Y},\mathbf{w},\Pi} \mathrm{Tr}\,(\mathbf{Y})$ subject to

$$\begin{aligned} \begin{bmatrix} \mathbf{Y} & \mathbf{S}(\Pi,\mathbf{w}) \\ \mathbf{S}(\Pi,\mathbf{w}) & \mathbf{S}(\Pi,\mathbf{w}) + \Pi \end{bmatrix} &\succeq 0; \\ \begin{bmatrix} \Pi & \mathbf{w} \\ \mathbf{w} & 1 \end{bmatrix} &\succeq 0. \end{aligned} \tag{34}$$

In some cases, the dual problem may admit a closed form solution, leading to an explicit expression for $\mathbf{M}$ via (32). As an example, suppose that $\mathbf{x} = x$ is a scalar and $J^{-1}(x) = a + b^2x^2$ for some $a > 0$. The dual problem becomes

$$\min_{\pi \geq 0} \frac{(a + b^2\pi)^2}{a + (b^2+1)\pi}. \tag{35}$$

The optimal solution can be shown to be

$$\pi = \max\left(\frac{a(1-b^2)}{b^2(b^2+1)}, 0\right), \tag{36}$$

leading to

$$\widehat{M} = \max\left(-\frac{2b^2}{b^2+1}, -1\right). \tag{37}$$

Therefore, if $\hat{x}$ achieves the CRLB, then the estimator

$$\hat{x}_b = \begin{cases} \frac{1-b^2}{1+b^2}\hat{x}, & |b| \leq 1; \\ 0, & |b| \geq 1 \end{cases} \tag{38}$$

achieves the MSE

$$\mathrm{MSEB}\,(\widehat{M}, x) = \begin{cases} a\frac{(1-b^2)^2}{(1+b^2)^2} + b^2x^2, & |b| \leq 1; \\ x^2, & |b| \geq 1, \end{cases} \tag{39}$$

which is smaller than $J^{-1}(x)$ for all x.

4.3. Necessary and Sufficient Optimality Conditions

To complete our description of the optimal $\mathbf{M}$, we can use the Karush-Kuhn-Tucker (KKT) theory [5] to develop necessary and sufficient optimality conditions. Using the facts that $\pi = 1$, $\Delta = \Pi + \mathbf{S}(\Pi,\mathbf{w})$ and $\mathbf{M}$ is given by (32), we can conclude that the matrix $\mathbf{M}$ is optimal if and only if there exists a matrix Π and a vector $\mathbf{w}$ such that $\Pi \succeq \mathbf{w}\mathbf{w}^*$ and the following conditions hold:

$$\begin{aligned} \mathbf{M} = -\mathbf{S}(\Pi,\mathbf{w})\,(\mathbf{S}(\Pi,\mathbf{w}) + \Pi)^{-1}; \\ \begin{bmatrix} \mathbf{A}_0(\mathbf{M}) & \mathbf{b}_0(\mathbf{M}) \\ \mathbf{b}_0^*(\mathbf{M}) & c_0(\mathbf{M}) - \mathrm{Tr}\,(\mathbf{M}\mathbf{S}(\Pi,\mathbf{w})) \end{bmatrix} \preceq 0, \end{aligned} \tag{40}$$

were $\mathbf{A}_0(\mathbf{M}), \mathbf{b}_0(\mathbf{M}), c_0(\mathbf{M})$ are defined by (19), and $\mathbf{S}(\Pi,\mathbf{w})$ is given by (29).

5. DOMINATING BOUND ON A QUADRATIC SET

We now treat the case in which $\mathbf{x}$ is restricted to the quadratic set $\mathcal{Q}$ of (17), which is assumed to be strictly feasible. To find an admissible dominating matrix in this case we need to solve the problem

$$\min_{\mathbf{M}} \max_{\mathbf{x}\in\mathcal{Q}} \{\mathrm{MSEB}\,(\mathbf{M},\mathbf{x}) - \mathrm{MSEB}\,(0,\mathbf{x})\}. \tag{41}$$

We first consider the inner maximization in (41) which, omitting the dependence on $\mathbf{M}$, has the form

$$\max_{\mathbf{x}} \{\mathbf{x}^*\mathbf{A}_0\mathbf{x} + 2\Re\{\mathbf{b}_0^*\mathbf{x}\} + c_0 : \mathbf{x}^*\mathbf{A}_1\mathbf{x} + 2\mathbf{b}_1^*\mathbf{x} + c_1 \leq 0\}. \tag{42}$$

The problem of (42) is a *trust region problem*, for which strong duality holds [11]. Thus, it is equivalent to

$$\min_{t,\lambda\geq 0,\mathbf{M}} \{t : \mathbf{G}(\mathbf{M}) \preceq \lambda\mathbf{F}\}, \quad (43)$$

where $\mathbf{G}(\mathbf{M})$ is defined in (22) and

$$\mathbf{F} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{b}_1 \\ \mathbf{b}_1^* & c_1 \end{bmatrix}. \quad (44)$$

The problem (43) is very similar to that of the previous section. Therefore, the development of the solution is analogous to the previous development (in particular, we also assume strict feasibility). We begin with the equivalent of Lemma 1, which shows that the optimal $\mathbf{M}$ can be found by solving an SDP:

Lemma 2. *The problem (43) is equivalent to the convex problem*

$$\min_{t,\lambda\geq 0,\mathbf{M},\mathbf{X}} \{t : \mathbf{Z}(\mathbf{M},\mathbf{X}) \preceq \lambda\mathbf{F},\ \mathbf{M}^*\mathbf{M} \preceq \mathbf{X}\}, \quad (45)$$

where $\mathbf{Z}(\mathbf{M},\mathbf{X})$ *is defined in (24).*

We now use Lagrange duality theory to gain more insight into the optimal $\mathbf{M}$. The Lagrangian associated with (45) is

$$\mathcal{L} = t + \mathrm{Tr}(\widetilde{\Pi}(\mathbf{Z}(\mathbf{M},\mathbf{X}) - \lambda\mathbf{F})) + \mathrm{Tr}\left(\Delta(\mathbf{M}^*\mathbf{M} - \mathbf{X})\right) \quad (46)$$

where $\Delta \succeq 0$ and Π is defined by (27). Since $\lambda \geq 0$, the minimum of the Lagrangian is finite only if

$$\mathrm{Tr}(\widetilde{\Pi}\mathbf{F}) = \mathrm{Tr}(\Pi\mathbf{A}_1) + 2\Re\{\mathbf{w}^*\mathbf{b}_1\} + c_1 \leq 0. \quad (47)$$

The optimal value is then obtained at $\lambda = 0$, and the Lagrangian becomes the same as that associated with the unconstrained problem (23). Thus, the dual problem of (45) is $\min_{\mathbf{Y},\mathbf{w},\Pi} \mathrm{Tr}\,(\mathbf{Y})$ subject to

$$\begin{aligned} \begin{bmatrix} \mathbf{Y} & \mathbf{S}(\Pi,\mathbf{w}) \\ \mathbf{S}(\Pi,\mathbf{w}) & \mathbf{S}(\Pi,\mathbf{w}) + \Pi \end{bmatrix} &\succeq 0; \\ \begin{bmatrix} \Pi & \mathbf{w} \\ \mathbf{w} & 1 \end{bmatrix} &\succeq 0; \\ \mathrm{Tr}(\Pi\mathbf{A}_1) + 2\Re\{\mathbf{w}^*\mathbf{b}_1\} + c_1 &\leq 0, \end{aligned} \quad (48)$$

which is again an SDP. The optimal matrix $\mathbf{M}$ is then given by (32), where Π and $\mathbf{w}$ are the solution to the dual problem.

Finally, in analogy to (40), using the KKT conditions we can show that the matrix $\mathbf{M}$ is optimal if and only if there exists a matrix Π and a vector $\mathbf{w}$ such that $\Pi \succeq \mathbf{w}\mathbf{w}^*$ and the following conditions hold:

$$\begin{gathered} \mathbf{M} = -\mathbf{S}(\Pi,\mathbf{w})\left(\mathbf{S}(\Pi,\mathbf{w}) + \Pi\right)^{-1}; \\ \mathrm{Tr}(\Pi\mathbf{A}_1) + 2\Re\{\mathbf{w}^*\mathbf{b}_1\} + c_1 \leq 0; \\ \lambda\left(\mathrm{Tr}(\Pi\mathbf{A}_1) + 2\Re\{\mathbf{w}^*\mathbf{b}_1\} + c_1\right) = 0; \\ \begin{bmatrix} \mathbf{A}_0(\mathbf{M}) & \mathbf{b}_0(\mathbf{M}) \\ \mathbf{b}_0^*(\mathbf{M}) & c_0(\mathbf{M}) - \mathrm{Tr}\left(\mathbf{M}\mathbf{S}(\Pi,\mathbf{w})\right) \end{bmatrix} \preceq \lambda \begin{bmatrix} \mathbf{A}_1 & \mathbf{b}_1 \\ \mathbf{b}_1^* & c_1 \end{bmatrix}. \end{gathered} \quad (49)$$

As an example, suppose that $\mathbf{J}^{-1}(\mathbf{x}) = \mathbf{A}$ and the set $\mathcal{Q}$ is defined by $\|\mathbf{x}\|^2 \leq c$. In this case, we can use the optimality conditions (49) to verify that a strictly dominating $\mathbf{M}$ is

$$\mathbf{M} = -\frac{\mathrm{Tr}(\mathbf{A})}{\mathrm{Tr}(\mathbf{A}) + c}\mathbf{I}. \quad (50)$$

The optimal dual variables are $\Pi = (c/\mathrm{Tr}(\mathbf{A}))\mathbf{A}$, $t = -\mathrm{Tr}^2(\mathbf{A})/(\mathrm{Tr}(\mathbf{A}) + c)$ and $\lambda = \mathrm{Tr}^2(\mathbf{A})/(\mathrm{Tr}(\mathbf{A}) + c)^2$. The corresponding MSE bound is

$$E\left\{\|\hat{\mathbf{x}} - \mathbf{x}_0\|^2\right\} \geq \frac{\mathrm{Tr}(\mathbf{A})}{(\mathrm{Tr}(\mathbf{A}) + c)^2}(\mathrm{Tr}(\mathbf{A})\mathbf{x}_0^*\mathbf{x}_0 + c^2). \quad (51)$$

Furthermore, if there exists an efficient estimator $\hat{\mathbf{x}}$, then

$$\hat{\mathbf{x}}_{\mathbf{b}} = \frac{c}{\mathrm{Tr}(\mathbf{A}) + c}\hat{\mathbf{x}} \quad (52)$$

achieves the bound (51), and has smaller MSE than $\hat{\mathbf{x}}$ for all $\|\mathbf{x}_0\|^2 \leq c$.

A special case is the linear Gaussian model in which $\mathbf{y} = \mathbf{H}\mathbf{x}_0 + \mathbf{w}$ where $\mathbf{H}$ is a known linear transformation with full rank, and $\mathbf{w}$ is a zero-mean Gaussian random vector with positive definite covariance $\mathbf{C}$. In this case, $\mathbf{J}^{-1}(\mathbf{x}_0) = (\mathbf{H}^*\mathbf{C}^{-1}\mathbf{H})^{-1}$. The estimator $\hat{\mathbf{x}}_{\mathbf{b}}$ of (52) reduces to the minimax MSE estimator developed in [14], which minimizes the worst-case MSE over all linear estimators. This estimator was shown in [15] to dominate the least-squares estimator for all $\|\mathbf{x}_0\|^2 \leq c$; since the least-squares estimator is efficient in this problem, our results coincide with those of [15].

6. EXAMPLE

In this section we demonstrate that our results can be used in practical settings even when an unbiased estimator is unknown. Specifically, we propose a linearly modified ML estimator in which we multiply the ML estimator by a strictly dominating matrix $\mathbf{M}$ on the appropriate set, even in cases when the ML estimator is not efficient. Since the ML estimator is asymptotically efficient, we are guaranteed that the MSE of the linearly modified ML estimator will be smaller than that of the ML estimator for all choices of parameters in the corresponding set *asymptotically*. We now demonstrate, through an example, that we often gain in performance even in the non-asymptotic regime.

Suppose we wish to estimate the SNR of a constant signal in Gaussian noise, from N iid measurements $y_i = \mu + w_i, 1 \leq i \leq N$ where w_i is a zero-mean Gaussian random variable with variance σ^2, and the SNR is defined by $x = \mu^2/\sigma^2$. The ML estimator of the SNR is $\hat{x} = \hat{\mu}^2/\hat{\sigma}^2$, where

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} y_i, \quad \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{\mu})^2. \quad (53)$$

The CRLB is this case is $J^{-1}(x) = (1/N)(4x + 2x^2)$. Note, that in general $\hat{x}$ is biased and does not achieve the CRLB; however, it is efficient asymptotically.

Since $x \geq 0$ for all μ and σ^2, to obtain a lower bound than the CRLB we may seek the $\hat{M}$ that is the solution to

$$\min_{M}\max_{x\geq 0}\left\{x^2M^2 + ((1 + M)^2 - 1)J^{-1}(x)\right\}. \quad (54)$$

The optimal value of M can then be found using the SDP formulation of Section 5. For our estimator, we then use the linearly transformed ML estimator which is given by $(1 + \hat{M})\hat{x}$.

In Fig. 1 we compare the MSE of the ML estimator and the linear ML estimator as a function of N for an SNR of $x = 2$. For each N, the MSE is averaged over 10000 noise realizations. As can be seen from the figure, the MSE of the linear ML is smaller than that of the ML estimator for all values of N.

In some cases we may have prior information on the range of SNR values possible, which can be exploited to further improve the performance. Suppose that we know that $\alpha \leq x \leq \beta$ for some

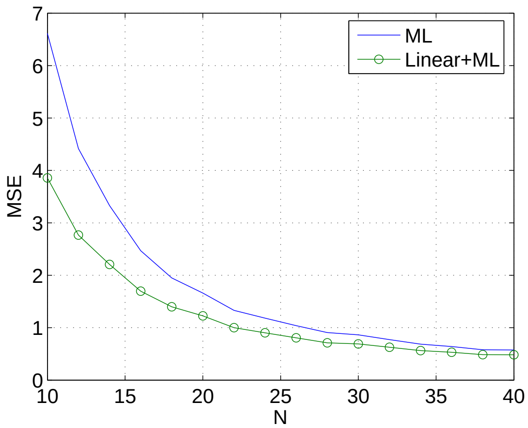

Fig. 1. MSE in estimating the SNR as a function of N for $x = 2$.

values of α and β. The ML estimator in this case is

$$\hat{x}_c = \begin{cases} \hat{x}, & \alpha \leq \hat{x} \leq \beta; \\ \alpha, & \hat{x} \leq \alpha; \\ \beta, & \hat{x} \geq \beta, \end{cases} \tag{55}$$

where $\hat{x} = \hat{\mu}^2/\hat{\sigma}^2$. To develop a linear modification of the ML estimator we note that the constraint $\alpha \leq x \leq \beta$ can be written as

$$(x - \alpha)(x - \beta) = x^2 - (\alpha + \beta)x + \alpha\beta \leq 0. \tag{56}$$

We can now solve the problem (54) with respect to the set (56).

In Fig. 2 we compare the MSE of the constrained ML and the linear ML estimators for $x = 2$ and SNR bounds $\alpha = 1$ and $\beta = 5$. As can be seen from the figure, the linearly modified ML estimator performs significantly better than the ML estimator.

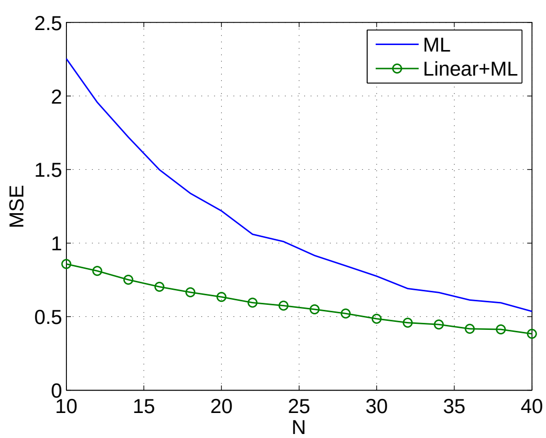

Fig. 2. MSE in estimating the SNR as a function of N for $x = 2$ subject to the constraint (56).

7. CONCLUSION

In this paper we introduced a general framework for obtaining bounds on the MSE performance of estimators with linear bias vectors. The bounds we developed dominate the conventional CRLB so that they are uniformly lower for all feasible values of the unknown parameter vector $\mathbf{x}_0$. The key idea we proposed is that a dominating bound can be obtained by solving a certain minimax optimization problem. We then analyzed the resulting minimax problem in the case in which the CRLB is quadratic in $\mathbf{x}_0$.

8. REFERENCES

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

[2] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, John Willey and Sons, Inc., 1968.

[3] Y. C. Eldar, "Uniformly improving the Cramér-Rao bound and maximum-likelihood estimation," submitted to *IEEE Trans. Signal Processing*.

[4] E. L. Lehmann and G. Casella, *Theory of point estimation*, New York, NY: Springer-Verlag, Inc., second edition, 1998.

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

[6] Y. C. Eldar, "Comparing between estimation approaches: Admissible and dominating linear estimators," to appear in *IEEE Trans. Signal Processing*.

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

[8] 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.

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

[10] J. O. Hero and J. A. Fessler, "A recursive algorithm for computing cr-type bounds on estimator covariance," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1205–1210, July 1994.

[11] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

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

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

[14] 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.

[15] Z. Ben-Haim and Y. C. Eldar, "Maximum set estimators with bounded estimation error," to appear in *IEEE Trans. Signal Processing*.