Uniformly Improving the Cramér-Rao Bound and Maximum-Likelihood Estimation

Yonina C. Eldar, *Member, IEEE*

***Abstract*—An important aspect of estimation theory is characterizing the best achievable performance in a given estimation problem, as well as determining estimators that achieve the optimal performance. The traditional Cramér–Rao type bounds provide benchmarks on the variance of any estimator of a deterministic parameter vector under suitable regularity conditions, while requiring a-priori specification of a desired bias gradient. 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), which is the sum of the variance and the squared-norm of the bias. Here, we develop bounds on the MSE in estimating a deterministic parameter vector $\mathbf{x}_0$ over all bias vectors that are linear in $\mathbf{x}_0$, which includes the traditional unbiased estimation as a special case. In some settings, it is possible to minimize the MSE over all linear bias vectors. More generally, direct minimization is not possible since the optimal solution depends on the unknown $\mathbf{x}_0$. Nonetheless, we show that in many cases, we can find bias vectors that result in an MSE bound that is smaller than the Cramér–Rao lower bound (CRLB) for all values of $\mathbf{x}_0$. Furthermore, we explicitly construct estimators that achieve these bounds in cases where an efficient estimator exists, by performing a simple linear transformation on the standard maximum likelihood (ML) estimator. This leads to estimators that result in a smaller MSE than the ML approach for all possible values of $\mathbf{x}_0$.**

***Index Terms*—Biased estimation, Cramér–Rao bound, dominating estimators, maximum likelihood, mean-squared error (MSE) bounds, minimax bounds.**

I. 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 an unknown deterministic parameter vector $\mathbf{x}_0$ from measurements $\mathbf{y}$, where 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 $\mathbf{x}_0$.

Manuscript received March 10, 2005; revised September 17, 2005. This work was supported in part by the Israel Science Foundation under Grant No. 536/04 and by the EU 6th framework programme via the NEWCOM network of excellence. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Carlos H. Muravchik.

The author is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: yonina@ee.technion.ac.il).

Digital Object Identifier 10.1109/TSP.2006.877648

A classic performance bound is the Cramér–Rao lower bound (CRLB) [1]–[3], which characterizes the smallest achievable total variance of any *unbiased* estimator of $\mathbf{x}_0$. Although other variance bounds exist in the literature, this bound is relatively easy to determine, and can often be achieved. Specifically, in the case in which 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 (i.i.d.) measurements, under suitable regularity assumptions on the pdf $p(\mathbf{y};\mathbf{x}_0)$, the ML estimator is asymptotically unbiased and achieves the CRLB [2], [4].

Although the CRLB is a popular performance benchmark, it only provides a bound on the variance of the estimator assuming 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 [5]. Furthermore, in some problems, restricting attention to unbiased approaches leads to unreasonable estimators that may, for example, be independent of the problem parameters; see [6] and [7] for some examples. Biased estimation methods are used extensively in a variety of different signal-processing applications, such as image restoration [8] where the bias corresponds to spatial resolution, smoothing techniques in time series analysis [9], [10], and spectrum estimation [11]. Thus, the design of estimators is typically subject to a tradeoff between variance and bias.

The total variance of any estimator with a given bias is bounded by the *biased CRLB* [12], which is an extension of the CRLB for unbiased estimators. 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. In [13] and [14], the *uniform CRLB* was developed which is a bound on the smallest attainable variance that can be achieved using any estimator with bias gradient whose norm is bounded by a constant. Although the uniform CRLB requires fixing only one parameter, and is therefore often more practical than the biased CRLB, it is still not clear in general how to optimally choose the bias gradient norm.

Evidently, standard CR-type bounds require specification of a bias measure, which is often not practical. 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 prespecify the bias. To characterize the best possible bias–variance tradeoff, we may consider the mean squared error (MSE) which is the average of the squared-norm error $\|\hat{\mathbf{x}} - \mathbf{x}_0\|^2$ and is equal to the sum of the variance and the squared norm of the bias. Thus, ideally, we would like to obtain a bound on the smallest possible MSE in a given estimation problem. Note that since $\mathbf{x}_0$

is deterministic in our setting, the MSE will in general depend on $\mathbf{x}_0$ itself. Unfortunately, since no limitations are imposed on $\hat{\mathbf{x}}$, the minimal bound is the trivial (zero) bound which can be achieved with $\hat{\mathbf{x}} = \mathbf{x}_0$.

For specific estimation problems, estimators have been developed that have smaller MSE than the CRLB for all values of $\mathbf{x}_0$. A classical example is the estimation of the mean of a Gaussian random vector for which the CRLB is achieved by using the well-known least squares estimator (which is also the ML estimator for this problem). James and Stein showed that the MSE can be reduced by using a biased estimator that is a nonlinear shrinkage of the ML method [15], [16]. Subsequently, a variety of other linear [17], [18] and nonlinear [19]–[22] approaches have been developed for this problem that outperform ML for all values of $\mathbf{x}_0$.

To treat the general problem, 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 $\mathbf{x}_0$. Our goal then is to find the smallest MSE achievable over all estimators with linear bias. A similar strategy was introduced in [23] for certain scalar estimation problems, and later extended in [24] to vector-valued estimates with bias gradient matrix proportional to the identity and restricted parameter values. The advantage of considering linear bias vectors is twofold. Analytically, this case is easier than other bias forms. From a practical perspective, since the bias vector is linear, we can easily construct estimators achieving the corresponding bound, based on efficient unbiased estimators. Specifically, if we find an optimal bias vector that minimizes the suggested MSE bound, and if in our problem an efficient estimator exists (*i.e.,* an unbiased estimator achieving the CRLB), then we can obtain an estimator achieving the corresponding MSE bound by simply multiplying the efficient estimator by a linear transformation constructed from the optimal bias vector.

It turns out that it is not always possible to minimize the proposed MSE bound over all linear bias vectors, since the optimal bias typically depends on $\mathbf{x}_0$ itself. In cases when direct minimization cannot be achieved, we may still be able to find a bias vector such that the resulting MSE is smaller than that of the CRLB for all possible values of $\mathbf{x}_0$. We are then guaranteed that if an efficient estimator exists, then there also exists an estimator, which is a simple linear transformation of the efficient method, whose MSE is smaller than the CRLB for all $\mathbf{x}_0$.

We begin in Section II by developing a bound on the MSE that depends on the bias of the estimator and the Fisher information. In Section III, we discuss cases in which the suggested MSE bound can be minimized directly over linear bias vectors. For the more general setting, we propose, in Section IV, concrete methods for finding a linear bias vector such that the resulting MSE bound is smaller than the CRLB for all values of $\mathbf{x}_0$. To this end we first show that such a bias vector can be obtained as a solution to a certain convex optimization problem. We then restrict our attention, in Section V, to estimation problems in which the CRLB is quadratic in $\mathbf{x}_0$ and analyze the resulting problem for two special cases: In Section VI we consider the case in which $\mathbf{x}_0$ is not restricted. In Section VII, we treat the case in which $\mathbf{x}_0$ lies in a quadratic set; this includes the scenario in which we seek to estimate a nonnegative parameter such as the variance or the signal-to-noise ratio (SNR). In both settings we show that a linear bias vector exists such that the resulting MSE bound is smaller than the CRLB for all possible 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 [25], [26]. We then develop necessary and sufficient optimality conditions in both settings which lead to further insight into the solution and in some cases can be used to derive closed-form expressions for the optimal bias vector. In Section VIII, we demonstrate through an example that by a linear transformation of the ML estimator, we can reduce the MSE for all values of $\mathbf{x}_0$.

In the sequel, we denote vectors in $\mathbb{C}^m$ (m arbitrary) by boldface lowercase letters and matrices in $\mathbb{C}^{n\times m}$ by boldface uppercase letters. The identity matrix of appropriate dimension is denoted by $\mathbf{I}$, $\widehat{(\cdot)}$ denotes an estimated vector or matrix, $(\cdot)^*$ is the Hermitian conjugate of the corresponding matrix, and x_i is the ith component of the vector $\mathbf{x}$. The true value of an unknown vector parameter $\mathbf{x}$ is denoted by $\mathbf{x}_0$, and the true value of an unknown scalar parameter x is written as x_0. The gradient of a vector $\partial\mathbf{b}(\mathbf{x}_0)/\partial\mathbf{x}$ is a matrix, with the ijth element equal to $\partial b_i(\mathbf{x}_0)/\partial x_j$. For a square matrix $\mathbf{A}$, $\mathrm{Tr}(\mathbf{A})$ is the trace of $\mathbf{A}$, $\mathbf{A}\succ 0$ ($\mathbf{A}\succeq 0$) means that $\mathbf{A}$ is Hermitian and positive (nonnegative) definite, and $\mathbf{A}\succeq\mathbf{B}$ means that $\mathbf{A}-\mathbf{B}\succeq 0$.

II. MSE Bound

We treat the problem of estimating a deterministic parameter vector $\mathbf{x}_0 \in \mathbb{C}^m$ from a given measurement vector $\mathbf{y}\in\mathbb{C}^n$ that is related to $\mathbf{x}_0$ through the pdf $p(\mathbf{y};\mathbf{x}_0)$.

A popular measure of estimator performance is the MSE, which is defined as

$$E\left\{\|\hat{\mathbf{x}}-\mathbf{x}_0\|^2\right\} = \|\mathbf{b}(\mathbf{x}_0)\|^2 + \mathrm{Tr}(\mathbf{C}_{\hat{\mathbf{x}}}). \tag{1}$$

Here

$$\mathbf{b}(\mathbf{x}_0) = E\left\{\hat{\mathbf{x}}\right\} - \mathbf{x}_0 \tag{2}$$

is the bias vector of $\hat{\mathbf{x}}$ and

$$\mathbf{C}_{\hat{\mathbf{x}}} = E\left\{[\hat{\mathbf{x}}-E\left\{\hat{\mathbf{x}}\right\}][\hat{\mathbf{x}}-E\left\{\hat{\mathbf{x}}\right\}]^*\right\} \tag{3}$$

is its covariance matrix.

Under suitable regularity conditions on $p(\mathbf{y};\mathbf{x})$ (see, e.g., [1], [2], and [27]), the covariance of any unbiased estimator $\hat{\mathbf{x}}$ of $\mathbf{x}_0$ is bounded below by the CRLB, which is given by $\mathbf{J}^{-1}(\mathbf{x}_0)$, where $\mathbf{J}(\mathbf{x}_0)$ is the Fisher information matrix

$$\mathbf{J}(\mathbf{x}_0) = E\left\{\left[\frac{\partial \log p(\mathbf{y};\mathbf{x}_0)}{\partial\mathbf{x}}\right]^*\left[\frac{\partial \log p(\mathbf{y};\mathbf{x}_0)}{\partial\mathbf{x}}\right]\right\} \tag{4}$$

and is assumed to be nonsingular. Thus, the MSE of any unbiased estimator satisfies

$$E\left\{\|\hat{\mathbf{x}}-\mathbf{x}_0\|^2\right\} \geq \mathrm{Tr}\left(\mathbf{J}^{-1}(\mathbf{x}_0)\right). \tag{5}$$

An estimator achieving the CRLB has minimum variance among all unbiased estimators. There are a variety of estimation problems in which the CRLB cannot be achieved, but nonetheless a minimum variance unbiased (MVU) estimator can be found. An example is when $p(y;x_0)$ is the uniform distribution on [0,x_0]. For this problem, the CRLB is not defined (see [27]); however, an MVU estimator of x_0 exists. The discussion in the remainder of this paper also holds true when we replace the CRLB $\mathbf{J}^{-1}(\mathbf{x}_0)$ everywhere by the variance of an MVU estimator. In this case, the proposed methods are linear transformations of the corresponding MVU estimator.

Often, the MSE can be further reduced by using a *biased* estimator. If $\hat{\mathbf{x}}$ is an estimator of $\mathbf{x}_0$ with bias vector $\mathbf{b}(\mathbf{x}_0)$, then the biased CRLB states that [12]

$$\mathbf{C}_{\hat{\mathbf{x}}} \succeq (\mathbf{I}+\mathbf{D}(\mathbf{x}_0))\,\mathbf{J}^{-1}(\mathbf{x}_0)\,(\mathbf{I}+\mathbf{D}(\mathbf{x}_0))^* \tag{6}$$

where $\mathbf{D}(\mathbf{x}_0)=\partial\mathbf{b}(\mathbf{x}_0)/\partial\mathbf{x}$ is the bias gradient matrix. Substituting (6) into (1), it follows that the MSE of any estimator with bias $\mathbf{b}(\mathbf{x}_0)$ is bounded below by

$$\|\mathbf{b}(\mathbf{x}_0)\|^2+\mathrm{Tr}\left((\mathbf{I}+\mathbf{D}(\mathbf{x}_0))\,\mathbf{J}^{-1}(\mathbf{x}_0)\,(\mathbf{I}+\mathbf{D}(\mathbf{x}_0))^*\right). \tag{7}$$

Ideally, to obtain the tightest possible MSE bound, we would like to minimize the bound (7) over all bias vectors $\mathbf{b}(\mathbf{x}_0)$. For every fixed value of $\mathbf{x}_0$, the minimum can be achieved with $\mathbf{b}(\mathbf{x})=\mathbf{x}_0-\mathbf{x}$; 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. Thus, in general, we cannot minimize (7) for all $\mathbf{x}_0$. Nonetheless, in some cases, we may be able to minimize the bound over all bias vectors in a suitable class. When the bound cannot be minimized directly, it still may be possible to find a bias $\mathbf{b}(\mathbf{x}_0)$ such that the resulting MSE bound is smaller than the unbiased CRLB for all possible values of $\mathbf{x}_0$. Our goal therefore is to minimize the MSE bound over all bias vectors in a suitable class (which includes the zero bias), when possible. Otherwise, we aim at finding a bias vector such that the resulting MSE bound is smaller than the unbiased CRLB for all values of $\mathbf{x}_0$ in a predefined set.

For our class of bias vectors, we consider linear bias vectors of the form

$$\mathbf{b}(\mathbf{x})=\mathbf{M}\mathbf{x} \tag{8}$$

for some $m\times m$ matrix $\mathbf{M}$. With this choice of bias, the MSE bound (MSEB) of (7) becomes

$$\mathrm{MSEB}(\mathbf{M},\mathbf{x}_0)=\mathbf{x}_0^*\mathbf{M}^*\mathbf{M}\mathbf{x}_0+\mathrm{Tr}\left((\mathbf{I}+\mathbf{M})\mathbf{J}^{-1}(\mathbf{x}_0)(\mathbf{I}+\mathbf{M})^*\right). \tag{9}$$

If $\mathbf{M}=0$, then as we expect, the bound coincides with the CRLB: MSEB $(0,\mathbf{x}_0)=\mathrm{Tr}(\mathbf{J}^{-1}(\mathbf{x}_0))$.

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, i.e., an estimator that achieves the CRLB, then the MSE of

$$\hat{\mathbf{x}}_{\mathrm{b}}=(\mathbf{I}+\mathbf{M})\hat{\mathbf{x}} \tag{10}$$

is equal to MSEB $(\mathbf{M},\mathbf{x}_0)$. To see this, since $E\{\hat{\mathbf{x}}\}=\mathbf{x}_0$

$$\mathbf{b}(\hat{\mathbf{x}}_{\mathrm{b}})=(\mathbf{I}+\mathbf{M})E\{\hat{\mathbf{x}}\}-\mathbf{x}_0=\mathbf{M}\mathbf{x}_0. \tag{11}$$

Using the fact that $\hat{\mathbf{x}}_{\mathrm{b}}-E\{\hat{\mathbf{x}}_{\mathrm{b}}\}=(\mathbf{I}+\mathbf{M})(\hat{\mathbf{x}}-\mathbf{x}_0)$ and $\mathbf{C}_{\hat{\mathbf{x}}}=\mathbf{J}^{-1}(\mathbf{x}_0)$

$$\begin{aligned}\mathbf{C}_{\hat{\mathbf{x}}_{\mathrm{b}}}&=(\mathbf{I}+\mathbf{M})E\{[\hat{\mathbf{x}}-\mathbf{x}_0][\hat{\mathbf{x}}-\mathbf{x}_0]^*\}(\mathbf{I}+\mathbf{M})^*\\&=(\mathbf{I}+\mathbf{M})\mathbf{J}^{-1}(\mathbf{x}_0)(\mathbf{I}+\mathbf{M})^*\end{aligned} \tag{12}$$

so that the MSE of $\hat{\mathbf{x}}_{\mathrm{b}}$ is given by MSEB $(\mathbf{M},\mathbf{x}_0)$. Therefore, if $\hat{\mathbf{x}}$ achieves the CRLB and we find an $\mathbf{M}$ such that MSEB $(\mathbf{M},\mathbf{x}_0)<$ MSEB $(0,\mathbf{x}_0)$ for a suitable set of $\mathbf{x}_0$, then the MSE of $\hat{\mathbf{x}}_{\mathrm{b}}$ 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. The important point is that this improvement is for *all choices of* $\mathbf{x}_0$ in a suitable set (which can be the entire space $\mathbb{C}^m$).

In contrast, if we consider more general nonlinear bias vectors, then even if we find a bias that results in an MSE bound that is lower than the CRLB and an efficient estimator exists, it is still unclear in general how to obtain an estimator achieving the resulting MSE bound.

III. Minimal MSE Bound With Linear Bias

We begin by discussing cases in which the bound (9) can be minimized directly.

Since the objective in (9) is convex in $\mathbf{M}$, we can find the minimal value by setting the derivative to zero, which yields

$$\mathbf{M}(\mathbf{J}^{-1}(\mathbf{x})+\mathbf{x}\mathbf{x}^*)=-\mathbf{J}^{-1}(\mathbf{x}) \tag{13}$$

where for brevity we denoted $\mathbf{x}=\mathbf{x}_0$. Using the matrix inversion lemma, the optimal $\mathbf{M}$ can be written as

$$\widehat{\mathbf{M}}=-\mathbf{I}+\frac{1}{1+\mathbf{x}^*\mathbf{J}(\mathbf{x})\mathbf{x}}\mathbf{x}\mathbf{x}^*\mathbf{J}(\mathbf{x}). \tag{14}$$

In general $\mathbf{M}$ will depend on $\mathbf{x}$, which is unknown, so that there is no constant value of $\mathbf{M}$ that minimizes the bound. However, if (14) is independent of $\mathbf{x}$, then this choice of $\mathbf{M}$ minimizes the bound for all possible values of $\mathbf{x}$. This occurs when $\mathbf{x}=x$ is a scalar, and $J^{-1}(x)=\alpha|x|^2$ for some $\alpha>0$. In this case, the optimal choice of $\mathbf{M}=\widehat{M}$ follows from (14) as

$$\widehat{M}=-\frac{\alpha}{1+\alpha} \tag{15}$$

and the corresponding bound is

$$\mathrm{MSEB}\,(\widehat{M},x)=\frac{\alpha}{1+\alpha}|x|^2=\frac{1}{1+\alpha}J^{-1}(x)<\mathrm{MSEB}\,(0,x) \tag{16}$$

for all x such that $J^{-1}(x)>0$. If $\hat{x}$ achieves the CRLB, then an estimator achieving MSEB $(\widehat{M},x)$ can be found using (10), which leads to the following theorem.

Theorem 1: Let $\mathbf{y}$ denote measurements of a deterministic parameter x_0 with pdf $p(\mathbf{y}; x_0)$. Assume that the Fisher information with respect to x_0 has the form $J(x_0) = 1/(\alpha|x_0|^2)$ for some $\alpha > 0$. Then the MSE of any estimate $\hat{x}$ of x_0 with linear bias satisfies

$$E\left\{|\hat{x} - x_0|^2\right\} \geq \frac{\alpha}{1+\alpha}|x_0|^2. \tag{17}$$

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

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

achieves the bound (17) and has smaller MSE than $\hat{x}$ for all $x_0 \neq 0$.

We now consider some examples illustrating the results of Theorem 1.

1) Example 1: Suppose that we are given N i.i.d. measurements $y_i, 1 \leq i \leq N$ that are each distributed uniformly on $[0, x_0]$ and we wish to estimate x_0. As mentioned in Section II, in this case the CRLB is not defined; however, an MVU estimator exists and is given by $\hat{x} = (1 + 1/N)y_{\max}$, where $y_{\max} = \max_i y_i$ [28, p. 108]. The minimum variance achievable with an unbiased estimator is equal to the MSE of $\hat{x}$ and is given by $(1/N(N+2))x_0^2$. Since the MSE has the form αx_0^2 with $\alpha = 1/N(N+2)$, we can use Theorem 1 to conclude that the estimator

$$\hat{x}_{\mathrm{b}} = \frac{N+2}{N+1}y_{\max} \tag{18}$$

has smaller MSE for values of x_0. The same estimator was shown in [6] to minimize the MSE among invariant estimates with the property that $\hat{x}(cy_1, \ldots, cy_N) = c\hat{x}(y_1, \ldots, y_N)$ for all $c > 0$.

2) Example 2: Consider the problem of estimating the variance σ^2 of a Gaussian random variable with known mean μ from N i.i.d. measurements $y_i, 1 \leq i \leq N$. An efficient estimate of σ^2 achieving the unbiased CRLB $J^{-1}(x) = 2\sigma^4/N$ is

$$\hat{x} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \mu)^2. \tag{19}$$

From Theorem 1, it follows that the estimator

$$\hat{x}_{\mathrm{b}} = \frac{1}{N+2}\sum_{i=1}^{N}(y_i - \mu)^2 \tag{20}$$

has smaller MSE than $\hat{x}$ for all values of $\sigma^2 > 0$. Note that this estimator has been proposed previously in [29].

If μ is not known, then the CRLB cannot be achieved. However, in this case the estimator

$$\hat{x} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2 \tag{21}$$

with $\bar{y} = (1/N)\sum_{i=1}^{N} y_i$ is an MVU estimator with MSE $2\sigma^4/(N-1)$. Applying Theorem 1 to (21), we conclude that

$$\hat{x}_{\mathrm{b}} = \frac{1}{N+1}\sum_{i=1}^{N}(y_i - \bar{y})^2 \tag{22}$$

has smaller MSE for all values of μ and σ^2. This result has also been obtained in [30], [2, p. 316], and [31]. The MSE of $\hat{x}_{\mathrm{b}}$ is $2\sigma^4/(N+1)$, which is smaller than the CRLB for unbiased estimators for known, as well as unknown, μ.

3) Example 3: As a final example, suppose we wish to estimate the mean x_0 of an exponential random variable from N i.i.d. measurements $y_i, 1 \leq i \leq N$, where

$$p(y; x_0) = \frac{1}{x_0}e^{-y/x_0}, \quad x_0 \geq 0. \tag{23}$$

An efficient estimator in this case 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

$$\hat{x}_{\mathrm{b}} = \frac{1}{N+1}\sum_{i=1}^{N} y_i \tag{24}$$

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

IV. Dominating the CRLB With Linear Bias

We have seen in the previous section that in some special cases, we can 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 for all possible values of the true parameter $\mathbf{x}_0$.

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

$$\mathrm{MSEB}\,(\mathbf{M}, \mathbf{x}_0) < \mathrm{MSEB}\,(0, \mathbf{x}_0) \tag{25}$$

for all values of $\mathbf{x}_0$ in some set $\mathcal{U}$. If the matrix $\mathbf{M}$ satisfies (25), then we will say that $\mathbf{M}$ (strictly) *dominates* [32] the CRLB on $\mathcal{U}$. This will ensure that if $\hat{\mathbf{x}}$ is an efficient estimator, then the estimator $\hat{\mathbf{x}}_{\mathrm{b}} = (\mathbf{I} + \mathbf{M})\hat{\mathbf{x}}$ will have smaller MSE than $\hat{\mathbf{x}}$ for all values of $\mathbf{x}_0 \in \mathcal{U}$. In addition to satisfying (25), we would like $\mathbf{M}$ to have the property that there is no other matrix $\mathbf{M}' \neq \mathbf{M}$ such that

$$\mathrm{MSEB}\,(\mathbf{M}', \mathbf{x}_0) \leq \mathrm{MSEB}\,(\mathbf{M}, \mathbf{x}_0) \tag{26}$$

for all $\mathbf{x}_0$ in $\mathcal{U}$. Such a matrix $\mathbf{M}$ will be called *admissible* [32]. 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, as incorporated in the following theorem.

Theorem 2: Let $\mathbf{y}$ denote measurements of a deterministic parameter vector $\mathbf{x}_0$ with pdf $p(\mathbf{y};\mathbf{x}_0)$. Let

$$\mathrm{MSEB}(\mathbf{M},\mathbf{x}_0)=\mathbf{x}_0^*\mathbf{M}^*\mathbf{M}\mathbf{x}_0+\mathrm{Tr}\left((\mathbf{I}+\mathbf{M})\mathbf{J}^{-1}(\mathbf{x}_0)(\mathbf{I}+\mathbf{M})^*\right),$$

be a bound on the MSE of any estimate $\hat{\mathbf{x}}$ of $\mathbf{x}_0$ with linear bias $\mathbf{b}(\mathbf{x}_0)=\mathbf{M}\mathbf{x}_0$, where $\mathbf{J}(\mathbf{x}_0)$ is the Fisher information matrix, and let $\mathcal{U}\subseteq\mathbb{C}^m$. Define

$$\widehat{\mathbf{M}}=\arg\min_{\mathbf{M}}\sup_{\mathbf{x}\in\mathcal{U}}\left\{\mathrm{MSEB}\left(\mathbf{M},\mathbf{x}\right)-\mathrm{MSEB}\left(0,\mathbf{x}\right)\right\}. \quad (27)$$

Then

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

Note that the minimum in (27) is well defined since the objective is continuous and coercive [33].

Proof: The proof follows immediately from the proof of [34, Th. 1] by noting that $\mathrm{MSEB}\left(\mathbf{M},\mathbf{x}\right)$ is continuous, coercive, and strictly convex in $\mathbf{M}$. ■

From Theorem 2, we conclude that if we find an $\widehat{\mathbf{M}}\neq 0$ that is the solution to (27), and if $\hat{\mathbf{x}}$ achieves the CRLB, then the MSE of $\hat{\mathbf{x}}_{\mathrm{b}}=(\mathbf{I}+\widehat{\mathbf{M}})\hat{\mathbf{x}}$ 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 $\hat{\mathbf{x}}_{\mathrm{b}}$ for all values of $\mathbf{x}_0\in\mathcal{U}$.

Problem (27) is convex in $\mathbf{M}$ for any constraint set $\mathcal{U}$ since the supremum of a convex function over any set $\mathcal{U}$ is convex. For arbitrary forms of $\mathbf{J}^{-1}(\mathbf{x}_0)$, we can solve (27) by using any one of the many known iterative algorithms for solving minimax problems, such as subgradient algorithms [35] or the prox method [36]. To obtain more efficient solutions, in the following sections we restrict the form of $\mathbf{J}^{-1}(\mathbf{x}_0)$ such that the resulting optimization problem can be converted into one of the standard convex problems for which very efficient software exists.

V. Quadratic Inverse Fisher Information

The most common tractable class of convex programs is the linear program which is the problem of minimizing a linear function subject to linear (affine) constraints. Recent advances in convex optimization lead to a generalization of the results and algorithms for linear programs to more complicated convex programs. A broad class of convex problems for which polynomial-time algorithms exists are semidefinite programs (SDPs) [25], [26]. These are optimization problems that involve 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 $\mathbf{M}$. Once a problem is formulated as an SDP, standard software packages, such as the self-dual-minimization (SeDuMi) package [37], can be used to solve the problem in polynomial time within any desired accuracy. Using principles of duality theory in vector space optimization, the SDP formulation can also be used to derive necessary and sufficient optimality conditions.

It turns out that for a large class of inverse Fisher information matrices, (27) can be reduced to an SDP. In the remainder of this paper, we treat the case in which $\mathbf{J}^{-1}(\mathbf{x}_0)$ has the quadratic form

$$\mathbf{J}^{-1}(\mathbf{x})=\sum_{i=1}^{\ell}\mathbf{B}_i\mathbf{x}\mathbf{x}^*\mathbf{B}_i^*+\sum_{i=1}^{k}(\mathbf{C}_i\mathbf{x}\mathbf{z}_i^*+\mathbf{z}_i\mathbf{x}^*\mathbf{C}_i^*)+\mathbf{A} \quad (28)$$

for some matrices $\mathbf{A}\succeq 0$, $\mathbf{B}_i$, $\mathbf{C}_i$ and vectors $\mathbf{z}_i$. (Alternatively, when considering MVU estimators, we assume that the minimum variance has the form (28).) As we will see, when the inverse Fisher information has such a quadratic form, and $\mathcal{U}$ is chosen appropriately, (27) can be reduced to an SDP. Besides leading to analytically tractable expressions, there are many cases in which the inverse Fisher information can be written in the form (28). Several examples are presented below.

1) Example 1: Suppose that we are given N i.i.d. measurements $y_i, 1\leq i\leq N$, where y_i has an exponential distribution as in (23). In this case, $J^{-1}(x)=(1/N)x^2$, which can be written in the form (28) with $\ell=1$, $B_1=1/\sqrt{N}$, and all the remaining parameters equal to zero. Another model that results in the same inverse Fisher information is when y_i is a Bernoulli trial with parameter x so that

$$y_i=\begin{cases}1, & \text{w.p. } x\\ 0, & \text{w.p. } 1-x.\end{cases} \quad (29)$$

2) Example 2: [27]. Consider the problem of estimating the mean μ and variance σ^2 of a Gaussian random variable from N i.i.d. measurements. In this case, $\mathbf{x}=[\mu \quad \sigma^2]^T$ and

$$\mathbf{J}^{-1}(\mathbf{x})=\frac{\sigma^2}{N}\begin{bmatrix}1 & 0\\ 0 & 2\sigma^2\end{bmatrix} \quad (30)$$

which has the form (28) with $\ell=1$, $k=1$

$$\mathbf{A}=0,\ \mathbf{C}_1=\frac{1}{N}\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix},\ \mathbf{z}_1=\begin{bmatrix}1\\ 0\end{bmatrix},\ \mathbf{B}_1=\sqrt{\frac{2}{N}}\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}. \quad (31)$$

Since the Fisher information matrix is diagonal, it follows that for μ known, the inverse Fisher information with respect to σ^2 is

$$\mathbf{J}^{-1}(\sigma^2)=\frac{2\sigma^4}{N} \quad (32)$$

in which case $\ell=1$, $B_1=\sqrt{2/N}$, and all the remaining parameters are equal to zero.

If σ^2 is known and μ is Gaussian with zero mean and unknown variance σ_μ^2, then

$$\mathbf{J}^{-1}(\sigma_\mu^2)=2\left(\sigma_\mu^2+\frac{\sigma^2}{N}\right)^2 \quad (33)$$

so that now $\ell=1$, $k=1$

$$\mathbf{A}=2\frac{\sigma^4}{N^2},\quad \mathbf{C}_1=4\frac{\sigma^2}{N},\quad \mathbf{z}_1=1,\quad \mathbf{B}_1=\sqrt{2}. \quad (34)$$

3) Example 3: Suppose that the observation vector $\mathbf{y}$ is a vector of counts with mean $\mathbf{g}(\mathbf{x})$, where

$$\mathbf{g}(\mathbf{x})=\mathbf{H}\mathbf{x}+\mathbf{c} \quad (35)$$

for some known invertible matrix $\mathbf{H}$ and known 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 known constant. This problem arises, for example, in emission-computed tomography [38]. The Fisher information in this case is given by [39]

$$\mathbf{J}(\mathbf{x}) = \mathbf{H}^*\mathrm{diag}\left(\frac{1}{g_1(\mathbf{x})},\ldots,\frac{1}{g_m(\mathbf{x})}\right)\mathbf{H} \tag{36}$$

and

$$\mathbf{J}^{-1}(\mathbf{x}) = \mathbf{H}^{-1}\mathrm{diag}\Big(g_1(\mathbf{x}),\ldots,g_m(\mathbf{x})\Big)\mathbf{H}^{-*} \tag{37}$$

where $\mathbf{H}^{-*} = (\mathbf{H}^{-1})^*$. We can express $\mathbf{J}^{-1}(\mathbf{x})$ of (37) in the form (28) with $\ell = m$, $k = 0$

$$\begin{aligned}\mathbf{A} &= \mathbf{H}^{-1}\mathrm{diag}\,(c_1,\ldots,c_m)\mathbf{H}^{-*}\\ \mathbf{C}_i &= \mathbf{H}^{-1}\mathbf{E}_i([\mathbf{H}]_i),\ \mathbf{z}_i = [\mathbf{H}^{-*}]_i, \quad 1 \le i \le m\end{aligned} \tag{38}$$

where $[\mathbf{H}]_i$ denotes the ith row of the matrix $\mathbf{H}$ and $\mathbf{E}_i(\mathbf{d})$ is the matrix whose ith row is equal to the vector $\mathbf{d}$ and whose remaining elements are equal zero.

4) Example 4: As another example, suppose that

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \tag{39}$$

where $\mathbf{w}$ is a Gaussian random vector with zero mean and known covariance $\mathbf{C}$. In this case, the inverse Fisher information matrix with respect to $\mathbf{x}$ is

$$\mathbf{J}^{-1}(\mathbf{x}) = (\mathbf{H}^*\mathbf{C}^{-1}\mathbf{H})^{-1} \tag{40}$$

which has the form (28) with $\ell = 0$, $k = 0$, and $\mathbf{A} = (\mathbf{H}^*\mathbf{C}^{-1}\mathbf{H})^{-1}$.

5) Example 5: As a final example, consider an unknown scalar μ in additive white Gaussian noise with unknown variance σ^2. The inverse Fisher information for estimating the SNR $\mathbf{x} = \mu^2/\sigma^2$ is

$$\mathbf{J}^{-1}(\mathbf{x}) = \frac{1}{N}(4\mathbf{x} + 2\mathbf{x}^2) \tag{41}$$

which has the form (28) with $\ell = 1$, $k = 1$

$$\mathbf{A} = 0, \quad \mathbf{C}_1 = \frac{4}{N}, \quad \mathbf{z}_1 = 1, \quad \mathbf{B}_1 = \sqrt{\frac{2}{N}}. \tag{42}$$

In Section VI, we treat the case in which $\mathcal{U} = \mathbb{C}^m$ so that $\mathbf{x}_0$ is not restricted, and show that with $\mathbf{J}^{-1}(\mathbf{x}_0)$ given by (28), the optimal $\mathbf{M}$ can be found as a solution to an SDP. 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 the parameter vector $\mathbf{x}_0$, which can result in a lower MSE bound. The set $\mathcal{U}$ is then chosen to capture these properties of $\mathbf{x}_0$. For example, we may know that the norm of $\mathbf{x}_0$ is bounded: $\mathbf{x}_0^*\mathbf{x}_0 \le U$ for some $U > 0$. There are also examples where there are natural restrictions on the parameters: for example, if x_0 represents the variance or the SNR of a random variable, then $x_0 > 0$. More generally, x_0 may lie in a specified interval $\alpha \le x_0 \le \beta$. These constraints can all be viewed as special cases of the quadratic constraint $\mathbf{x}_0 \in \mathcal{Q}$, where

$$\mathcal{Q} = \{\mathbf{x} | \mathbf{x}^*\mathbf{A}_1\mathbf{x} + 2\mathbf{b}_1^*\mathbf{x} + c_1 \le 0\} \tag{43}$$

for some $\mathbf{A}_1$, $\mathbf{b}_1$, and c_1. Note that we do not require that $\mathbf{A}_1 \succeq 0$ so that the constraint set (43) is not necessarily convex. In Section VII, 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. Using the results of [40], the ideas we develop can also be generalized to the case of two quadratic constraints of the form $\mathcal{Q}$.

Before proceeding to the detailed developments, it is important to note that even in cases where $\mathbf{M}$ is computed via an SDP, i.e., a closed-form solution does not exists, the calculation of $\mathbf{M}$ does not depend on the data $\mathbf{y}$. Therefore, $\mathbf{M}$ can be computed off line. Once the data is received, to implement the proposed approach, all that is needed is to multiply the unbiased estimator by the matrix $\mathbf{I} + \mathbf{M}$ so that the additional cost incurred is negligible.

VI. Dominating Bound on the Entire Space

We first treat the case in which $\mathcal{U} = \mathbb{C}^m$ so that $\mathbf{x}_0$ is not restricted. As we will show, if $\mathbf{B}_i \ne 0$ for some i, then a strictly dominating $\mathbf{M}$ over the entire space can always be found. This implies that under this condition, the CRLB can be improved on uniformly.

With $\mathbf{J}^{-1}(\mathbf{x}_0)$ given by (28), the MSE bound of (9) can be written compactly as

$$\mathrm{MSEB}\,(\mathbf{M},\mathbf{x}_0) = \mathbf{x}^*\mathbf{A}_0(\mathbf{M})\mathbf{x} + 2\Re\{\mathbf{b}_0^*(\mathbf{M})\mathbf{x}\} + c_0(\mathbf{M}) \tag{44}$$

where we defined

$$\begin{aligned}\mathbf{A}_0(\mathbf{M}) &= \mathbf{M}^*\mathbf{M} + \sum_{i=1}^{\ell}\mathbf{B}_i^*\left((\mathbf{I}+\mathbf{M})^*(\mathbf{I}+\mathbf{M}) - \mathbf{I}\right)\mathbf{B}_i\\ \mathbf{b}_0(\mathbf{M}) &= \sum_{i=1}^{k}\mathbf{C}_i^*\left((\mathbf{I}+\mathbf{M})^*(\mathbf{I}+\mathbf{M}) - \mathbf{I}\right)\mathbf{z}_i\\ c_0(\mathbf{M}) &= \mathrm{Tr}\left(\left((\mathbf{I}+\mathbf{M})^*(\mathbf{I}+\mathbf{M}) - \mathbf{I}\right)\mathbf{A}\right).\end{aligned} \tag{45}$$

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

$$\min_{\mathbf{M}}\max_{\mathbf{x}}\{\mathbf{x}^*\mathbf{A}_0(\mathbf{M})\mathbf{x} + 2\Re\{\mathbf{b}_0^*(\mathbf{M})\mathbf{x}\} + c_0(\mathbf{M})\} \tag{46}$$

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

$$\mathbf{x}^*\mathbf{A}_0(\mathbf{M})\mathbf{x} + 2\Re\{\mathbf{b}_0^*(\mathbf{M})\mathbf{x}\} + c_0(\mathbf{M}) \le t, \quad \text{for all } \mathbf{x}. \tag{47}$$

The constraint (47) is equivalent to [41, p. 163]

$$\mathbf{G}(\mathbf{M}) \triangleq \begin{bmatrix}\mathbf{A}_0(\mathbf{M}) & \mathbf{b}_0(\mathbf{M})\\ \mathbf{b}_0^*(\mathbf{M}) & c_0(\mathbf{M}) - t\end{bmatrix} \preceq 0. \tag{48}$$

Since the choice of parameters $\mathbf{M} = 0$, $t = 0$ satisfies the constraint (48), our problem is always feasible.

In our development below, we consider the case in which the constraint (48) is strictly feasible, i.e., there exists a matrix $\mathbf{M}$ such that $\mathbf{G}(\mathbf{M}) \prec 0$. Conditions for strict feasibility are given in the following lemma.

Lemma 1: The constraint (48) is strictly feasible if and only if

$$\mathbf{W} \triangleq \sum_{i=1}^{\ell} \mathbf{B}_i^* \mathbf{B}_i \succ 0. \tag{49}$$

Proof: We first show that strict feasibility is equivalent to

$$\mathbf{A}_0(\mathbf{M}) \prec 0 \text{ for some } \mathbf{M}. \tag{50}$$

Clearly, if (48) is strictly feasible, then (50) is satisfied. Conversely, if (50) holds, then we can always choose a large enough t such that (48) is strictly feasible. Indeed, using Schur's lemma (see Appendix A), strict feasibility reduces to the condition

$$\mathbf{A}_0(\mathbf{M}) \prec \frac{1}{c_0(\mathbf{M}) - t} \mathbf{b}_0(\mathbf{M}) \mathbf{b}_0(\mathbf{M})^*. \tag{51}$$

Since for some $\mathbf{M}$, $\mathbf{A}_0(\mathbf{M}) \prec 0$, there exists a small enough t such that (50) holds.

We now show that (49) is equivalent to (50). Suppose first that (49) is satisfied, and let $\mathbf{M} = \alpha \mathbf{I}$ with $\alpha = -\lambda_{\min}/(1 + \lambda_{\min})$ where $\lambda_{\min}$ is the smallest eigenvalue of $\mathbf{W}$. With this choice we can immediately verify that (50) is satisfied, and the problem is strictly feasible. Conversely, suppose that (49) does not hold; this implies that there exists a vector $\mathbf{v}$ such that $\mathbf{W}\mathbf{v} = 0$. Since $\mathbf{W}$ is a sum of positive semidefinite matrices, this is possible if and only if $\mathbf{B}_i^* \mathbf{B}_i \mathbf{v} = 0$ for all i, or equivalently, $\mathbf{B}_i \mathbf{v} = 0$ for all i. It then follows that $\mathbf{v}^* \mathbf{A}_0(\mathbf{M}) \mathbf{v} = \mathbf{v}^* \mathbf{M}^* \mathbf{M} \mathbf{v} \geq 0$, and (50) cannot hold. ■

If (48) is not strictly feasible then, as we show in Appendix B, 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. Due to the fact that any feasible problem can be reduced to a strictly feasible one, in the remainder of this section we assume that our problem is strictly feasible.

In the next section we show that the optimal $\mathbf{M}$ can be found as a solution to an SDP. We then develop an alternative SDP formulation via the dual program that also provides further insight into the solution in Section VI-C. In Section VI-D, we derive a set of necessary and sufficient optimality conditions on $\mathbf{M}$.

A. SDP Formulation of the Problem

The constraint (48) is not written in convex form, so that we cannot directly apply standard convex algorithms or Lagrange duality theory to find the optimal $\mathbf{M}$. Fortunately, this constraint can be converted into convex form, as incorporated in the following lemma.

Lemma 2: The problem

$$\min_{t,\mathbf{M}} \{t : \mathbf{G}(\mathbf{M}) \preceq 0\} \tag{52}$$

with $\mathbf{G}(\mathbf{M})$ given by (48) is equivalent to the convex problem

$$\min_{t,\mathbf{M},\mathbf{X}} \{t : \mathbf{Z}(\mathbf{M},\mathbf{X}) \preceq 0,\ \mathbf{M}^*\mathbf{M} \preceq \mathbf{X}\} \tag{53}$$

where

$$\mathbf{Z}(\mathbf{M},\mathbf{X}) = \begin{bmatrix} \mathbf{X} + \sum_{i=1}^{\ell} \mathbf{B}_i^* \Phi \mathbf{B}_i & \sum_{i=1}^{k} \mathbf{C}_i^* \Phi \mathbf{z}_i \\ \sum_{i=1}^{k} \mathbf{z}_i^* \Phi \mathbf{C}_i & \mathrm{Tr}(\mathbf{A}\Phi) - t \end{bmatrix} \tag{54}$$

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

Proof: See Appendix C. ■

From Lemma 2, we see that (52) can be written as a convex problem. Moreover, the optimal $\mathbf{M}$ can be found using standard software packages by noting that (53) can be written as an SDP. Indeed, the matrix $\mathbf{Z}(\mathbf{M},\mathbf{X})$ is linear in both $\mathbf{M}$ and $\mathbf{X}$; using Schur's lemma (Appendix A), the constraint $\mathbf{M}^*\mathbf{M} \preceq \mathbf{X}$ can be written as

$$\begin{bmatrix} \mathbf{X} & \mathbf{M}^* \\ \mathbf{M} & \mathbf{I} \end{bmatrix} \succeq 0 \tag{55}$$

which is also a linear matrix inequality.

B. 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 form expression, we now rely on Lagrange duality theory.

Since (53) is convex and strictly feasible, the optimal value of t is equal to the optimal value of the dual problem. To find the dual, we first write the Lagrangian associated with our problem

$$\mathcal{L} = t + \mathrm{Tr}(\widetilde{\Pi}\mathbf{Z}(\mathbf{M},\mathbf{X})) + \mathrm{Tr}(\Delta(\mathbf{M}^*\mathbf{M} - \mathbf{X})) \tag{56}$$

where $\Delta \succeq 0$ and

$$\widetilde{\Pi} = \begin{bmatrix} \Pi & \mathbf{w} \\ \mathbf{w}^* & \pi \end{bmatrix} \succeq 0 \tag{57}$$

are the dual variables.

Differentiating the Lagrangian with respect to t and equating to zero

$$\pi = 1. \tag{58}$$

Differentiating with respect to $\mathbf{X}$ and equating to zero

$$\begin{aligned} \Delta &= \Pi + \sum_{i=1}^{\ell} \mathbf{B}_i \Pi \mathbf{B}_i^* + \sum_{i=1}^{k} (\mathbf{z}_i \mathbf{w}^* \mathbf{C}_i^* + \mathbf{C}_i \mathbf{w} \mathbf{z}_i^*) + \mathbf{A} \\ &= \Pi + \mathbf{S}(\Pi, \mathbf{w}) \end{aligned} \tag{59}$$

where we defined

$$\mathbf{S}(\Pi, \mathbf{w}) = \sum_{i=1}^{\ell} \mathbf{B}_i \Pi \mathbf{B}_i^* + \sum_{i=1}^{k} (\mathbf{z}_i \mathbf{w}^* \mathbf{C}_i^* + \mathbf{C}_i \mathbf{w} \mathbf{z}_i^*) + \mathbf{A}. \tag{60}$$

Finally, the derivative with respect to $\mathbf{M}$ yields $\mathbf{M}\Delta = -\mathbf{S}(\Pi, \mathbf{w})$, which after substituting the value of Δ from (59) becomes

$$\mathbf{M}\left(\mathbf{S}(\Pi, \mathbf{w}) + \Pi\right) = -\mathbf{S}(\Pi, \mathbf{w}). \tag{61}$$

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

$$\mathbf{S}(\Pi, \mathbf{w}) \succeq \mathbf{S}(\mathbf{w}\mathbf{w}^*, \mathbf{w}) = \mathbf{J}^{-1}(\mathbf{w}) \succ 0 \tag{62}$$

and $\mathbf{S}$ is invertible. Thus, from (61)

$$\mathbf{M} = -\mathbf{S}(\Pi, \mathbf{w})\left(\mathbf{S}(\Pi, \mathbf{w}) + \Pi\right)^{-1}. \tag{63}$$

An important observation from (63) is that regardless of Π, $\mathbf{M}$ of (63) is not equal to zero. 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. In Appendix D, we show that $\mathbf{M}$ is also nonzero when the problem is not strictly feasible, as long as $\mathbf{B}_i \neq 0$ for some i. We therefore have the following proposition.

Proposition 1: Consider the setting of Theorem 2 with $\mathcal{U} = \mathbb{C}^m$ and $\mathbf{J}^{-1}(\mathbf{x}_0)$ given by (28). Then $\widehat{\mathbf{M}} = 0$ if and only if $\mathbf{B}_i = 0$ for all i.

Proof: If $\mathbf{B}_i = 0$ for all i, then $\mathbf{A}_0(\mathbf{M}) \preceq 0$ only if $\mathbf{M} = 0$. Since $\mathbf{M} = 0$ is feasible, we conclude that in this case $\mathbf{M} = 0$ is the optimal solution. The proof of the reverse implication is given in Appendix D. ∎

Substituting (63) into the Lagrangian, the dual problem becomes

$$\min_{\mathbf{w},\Pi} \operatorname{Tr}\left(\mathbf{S}(\Pi, \mathbf{w})\left(\mathbf{S}(\Pi, \mathbf{w}) + \Pi\right)^{-1}\mathbf{S}(\Pi, \mathbf{w})\right) \tag{64}$$

subject to (57). Using Schur's lemma, (64) can be written as

$$\min_{\mathbf{Y},\mathbf{w},\Pi} \operatorname{Tr}(\mathbf{Y}) \tag{65}$$

subject to

$$\begin{bmatrix} \mathbf{Y} & \mathbf{S}(\Pi, \mathbf{w}) \\ \mathbf{S}(\Pi, \mathbf{w}) & \mathbf{S}(\Pi, \mathbf{w}) + \Pi \end{bmatrix} \succeq 0$$
$$\begin{bmatrix} \Pi & \mathbf{w} \\ \mathbf{w} & 1 \end{bmatrix} \succeq 0 \tag{66}$$

which is again an SDP.

We conclude that the optimal matrix $\mathbf{M}$ is given by (63), where Π and $\mathbf{w}$ are the solution to the dual problem of minimizing (65) subject to (66). In some cases, the dual problem may admit a closed-form solution, leading to an explicit expression for $\mathbf{M}$ via (63). As an example, suppose that $\mathbf{x} = x$ is a scalar and $J^{-1}(x) = a + b^2x^2$, with $a > 0$. The dual problem becomes

$$\min_{\pi \geq 0} \frac{(a + b^2\pi)^2}{a + (b^2+1)\pi}. \tag{67}$$

The optimal solution can be shown to be

$$\pi = \max\left(\frac{a(1-b^2)}{b^2(b^2+1)}, 0\right) \tag{68}$$

leading to

$$\widehat{M} = \max\left(-\frac{2b^2}{b^2+1}, -1\right). \tag{69}$$

Therefore, if $\hat{x}$ achieves the CRLB $J^{-1}(x) = a + b^2x^2$, then the estimator

$$\hat{x}_{\mathrm{b}} = \begin{cases} \frac{1-b^2}{1+b^2}\hat{x}, & |b| \leq 1 \\ 0, & |b| \geq 1 \end{cases} \tag{70}$$

achieves the MSE

$$\operatorname{MSEB}(\widehat{M}, x) = \begin{cases} a\frac{(1-b^2)^2}{(1+b^2)^2} + b^2x^2, & |b| \leq 1 \\ x^2, & |b| \geq 1 \end{cases} \tag{71}$$

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

C. Necessary and Sufficient Optimality Conditions

To complete our description of the optimal $\mathbf{M}$, we now use the Karush–Kuhn–Tucker (KKT) theory [33] to develop necessary and sufficient optimality conditions.

The KKT conditions state that $\mathbf{M}$, $\mathbf{X}$, and t are optimal if and only if there exist matrices $\widetilde{\Pi}$, $\Delta \succeq 0$, such that

1) $d\mathcal{L}/d\mathbf{X} = 0$, $d\mathcal{L}/d\mathbf{M} = 0$, and $d\mathcal{L}/dt = 0$, where the Lagrangian $\mathcal{L}$ is defined by (56);
2) feasibility: $\mathbf{Z}(\mathbf{M}, \mathbf{X}) \preceq 0$, where $\mathbf{Z}(\mathbf{M}, \mathbf{X})$ is defined by (54);
3) complementary slackness: $\operatorname{Tr}(\widetilde{\Pi}\mathbf{Z}(\mathbf{M}, \mathbf{X})) = 0$ and $\operatorname{Tr}(\Delta(\mathbf{M}^*\mathbf{M} - \mathbf{X})) = 0$.

We have seen already that the first condition results in $\pi = 1$, $\Delta = \Pi + \mathbf{S}(\Pi, \mathbf{w})$, and $\mathbf{M}$ given by (63). Since $\Delta \succ 0$, we have immediately from the second complementary slackness condition that $\mathbf{X} = \mathbf{M}^*\mathbf{M}$. The first complementary slackness condition then becomes

$$\begin{aligned} 0 &= \operatorname{Tr}(\widetilde{\Pi}\mathbf{Z}(\mathbf{M}, \mathbf{X})) \\ &= \operatorname{Tr}\left((\mathbf{M}\Pi + (\mathbf{I} + \mathbf{M})\mathbf{S}(\Pi, \mathbf{w}))\mathbf{M}^*\right) + \operatorname{Tr}(\mathbf{M}\mathbf{S}(\Pi, \mathbf{w})) - t \\ &= \operatorname{Tr}(\mathbf{M}\mathbf{S}(\Pi, \mathbf{w})) - t. \end{aligned} \tag{72}$$

Thus, 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:

$$\begin{aligned} &\mathbf{M} = -\mathbf{S}(\Pi, \mathbf{w})\left(\mathbf{S}(\Pi, \mathbf{w}) + \Pi\right)^{-1} \\ &\begin{bmatrix} \mathbf{A}_0(\mathbf{M}) & \mathbf{b}_0(\mathbf{M}) \\ \mathbf{b}_0^*(\mathbf{M}) & c_0(\mathbf{M}) - \operatorname{Tr}(\mathbf{M}\mathbf{S}(\Pi, \mathbf{w})) \end{bmatrix} \preceq 0 \end{aligned} \tag{73}$$

where $\mathbf{A}_0(\mathbf{M})$, $\mathbf{b}_0(\mathbf{M})$, and $c_0(\mathbf{M})$ are defined by (45) and $\mathbf{S}(\Pi, \mathbf{w})$ is given by (60).

VII. Dominating Bound on a Quadratic Set

We now treat the case in which the parameter vector $\mathbf{x}$ is restricted to the quadratic set $\mathcal{Q}$ of (43). To find an admissible dominating matrix in this case, we need to solve the problem

$$\min_{\mathbf{M}} \max_{\mathbf{x}\in\mathcal{Q}} \{\text{MSEB}\,(\mathbf{M},\mathbf{x}) - \text{MSEB}\,(0,\mathbf{x})\}. \tag{74}$$

We assume that the set $\mathcal{Q}$ is not empty, and that there exists an $\mathbf{x}$ in the interior of $\mathcal{Q}$. However, we do not make any further assumptions on the parameters $\mathbf{A}_1$, $\mathbf{b}_1$ and c_1. In particular, we *do not* assume that $\mathbf{A}_1 \succeq 0$.

We first consider the inner maximization in (74), which, omitting the dependence on $\mathbf{M}$, has the form

$$\max_{\mathbf{x}} \{\mathbf{x}^*\mathbf{A}_0\mathbf{x} + 2\Re\{\mathbf{b}_0^*\mathbf{x}\} + c_0 : \mathbf{x}^*\mathbf{A}_1\mathbf{x} + 2\mathbf{b}_1^*\mathbf{x} + c_1 \le 0\}. \tag{75}$$

The problem of (75) is a *trust region problem* for which strong duality holds (assuming that there is a strictly feasible point) [42]. Thus, it is equivalent to

$$\min_{\lambda\ge 0, t, \mathbf{M}} t \tag{76}$$

subject to

$$\begin{bmatrix} \lambda\mathbf{A}_1 & \lambda\mathbf{b}_1 \\ \lambda\mathbf{b}_1^* & \lambda c_1 + t \end{bmatrix} \succeq \begin{bmatrix} \mathbf{A}_0(\mathbf{M}) & \mathbf{b}_0(\mathbf{M}) \\ \mathbf{b}_0^*(\mathbf{M}) & c_0(\mathbf{M}) \end{bmatrix}. \tag{77}$$

It is easy to see that (77) is always feasible, since both matrices in (77) can be made equal to zero by choosing $\mathbf{M} = 0$, and $\lambda = t = 0$. From Lemma 1, it follows that if $\sum_{i=1}^{\ell} \mathbf{B}_i^*\mathbf{B}_i \succ 0$, then the problem is strictly feasible with $\lambda = 0$. Since any feasible problem can be reduced to a strictly feasible one using the method described in Appendix B, we assume in the remainder of this section that the problem is strictly feasible.

The problem of minimizing (76) subject to (77) is very similar to that of (53). Indeed, our problem can be written compactly as

$$\min_{t,\lambda\ge 0,\mathbf{M}} \{t : \mathbf{G}(\mathbf{M}) \preceq \lambda\mathbf{F}\} \tag{78}$$

where $\mathbf{G}(\mathbf{M})$ is defined in (48) and

$$\mathbf{F} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{b}_1 \\ \mathbf{b}_1^* & c_1 \end{bmatrix}. \tag{79}$$

Therefore, the derivation of the solution is analogous to the development in the previous section. We begin with the equivalent of Lemma 2, which shows that the optimal $\mathbf{M}$ can be found by solving an SDP:

Lemma 3: The problem

$$\min_{t,\lambda\ge 0,\mathbf{M}} \{t : \mathbf{G}(\mathbf{M}) \preceq \lambda\mathbf{F}\} \tag{80}$$

with $\mathbf{G}(\mathbf{M})$ and $\mathbf{F}$ given by (48) and (79), respectively, is equivalent to the convex problem

$$\min_{t,\lambda\ge 0,\mathbf{M},\mathbf{X}} \{t : \mathbf{Z}(\mathbf{M},\mathbf{X}) \preceq \lambda\mathbf{F},\ \mathbf{M}^*\mathbf{M} \preceq \mathbf{X}\} \tag{81}$$

where $\mathbf{Z}(\mathbf{M},\mathbf{X})$ is defined in (54).

A. Dual Problem

We now can use Lagrange duality theory, as in Section VI-C, to gain more insight into the optimal $\mathbf{M}$.

The Lagrangian associated with the problem (81) is

$$\mathcal{L} = t + \text{Tr}(\widetilde{\Pi}(\mathbf{Z}(\mathbf{M},\mathbf{X}) - \lambda\mathbf{F})) + \text{Tr}\,(\Delta(\mathbf{M}^*\mathbf{M} - \mathbf{X})) \tag{82}$$

where $\Delta \succeq 0$ and Π is defined by (57). Since $\lambda \ge 0$, the minimum of the Lagrangian is finite only if

$$\text{Tr}(\widetilde{\Pi}\mathbf{F}) = \text{Tr}(\Pi\mathbf{A}_1) + 2\Re\{\mathbf{w}^*\mathbf{b}_1\} + c_1 \le 0. \tag{83}$$

The optimal value is then obtained at $\lambda = 0$, and the Lagrangian becomes the same as that associated with the unconstrained problem (53). Thus, the dual of (81) is

$$\min_{\mathbf{Y},\mathbf{w},\Pi} \text{Tr}\,(\mathbf{Y}) \tag{84}$$

subject to

$$\begin{gathered} \begin{bmatrix} \mathbf{Y} & \mathbf{S}(\Pi,\mathbf{w}) \\ \mathbf{S}(\Pi,\mathbf{w}) & \mathbf{S}(\Pi,\mathbf{w}) + \Pi \end{bmatrix} \succeq 0 \\ \begin{bmatrix} \Pi & \mathbf{w} \\ \mathbf{w} & 1 \end{bmatrix} \succeq 0 \\ \text{Tr}(\Pi\mathbf{A}_1) + 2\Re\{\mathbf{w}^*\mathbf{b}_1\} + c_1 \le 0 \end{gathered} \tag{85}$$

which is again an SDP.

We conclude that the optimal matrix $\mathbf{M}$ is given by (63), where Π and $\mathbf{w}$ are the solution to the dual problem of minimizing (84) subject to (85). Note that regardless of Π, $\mathbf{M}$ of (63) is not equal to zero. Therefore, from Theorem 2, it follows that as long as the problem is strictly feasible, we can improve the CRLB for all values of $\mathbf{x}_0 \in \mathcal{Q}$ by a linear transformation.

B. Necessary and Sufficient Optimality Conditions

Following the same steps as in Section VII-C we can show, using the KKT conditions, 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:

$$\begin{gathered} \mathbf{M} = -\mathbf{S}(\Pi,\mathbf{w})\,(\mathbf{S}(\Pi,\mathbf{w}) + \Pi)^{-1} \\ \text{Tr}(\Pi\mathbf{A}_1) + 2\Re\{\mathbf{w}^*\mathbf{b}_1\} + c_1 \le 0 \\ \lambda\,(\text{Tr}(\Pi\mathbf{A}_1) + 2\Re\{\mathbf{w}^*\mathbf{b}_1\} + c_1) = 0 \\ \begin{bmatrix} \mathbf{A}_0(\mathbf{M}) & \mathbf{b}_0(\mathbf{M}) \\ \mathbf{b}_0^*(\mathbf{M}) & c_0(\mathbf{M}) - \text{Tr}\,(\mathbf{M}\mathbf{S}(\Pi,\mathbf{w})) \end{bmatrix} \preceq \lambda \begin{bmatrix} \mathbf{A}_1 & \mathbf{b}_1 \\ \mathbf{b}_1^* & c_1 \end{bmatrix} \end{gathered} \tag{86}$$

where $\mathbf{A}_0(\mathbf{M})$, $\mathbf{b}_0(\mathbf{M})$, and $c_0(\mathbf{M})$ are defined by (45) and $\mathbf{S}(\Pi,\mathbf{w})$ is given by (60).

As an example, suppose that $\mathbf{J}^{-1}(\mathbf{x}) = \mathbf{A}$ and the set $\mathcal{Q}$ is defined by $\mathbf{x}^*\mathbf{x} \le c$. In this case, a strictly dominating $\mathbf{M}$ is given by

$$\mathbf{M} = -\frac{\mathrm{Tr}(\mathbf{A})}{\mathrm{Tr}(\mathbf{A}) + c}\mathbf{I}. \tag{87}$$

Indeed, since $\ell = k = 0$, and $\mathbf{b}_1 = 0$, the conditions in (86) are satisfied with

$$\Pi = \frac{c}{\mathrm{Tr}(\mathbf{A})}\mathbf{A}, \quad t = -\frac{\mathrm{Tr}^2(\mathbf{A})}{\mathrm{Tr}(\mathbf{A}) + c}, \quad \lambda = \frac{\mathrm{Tr}^2(\mathbf{A})}{(\mathrm{Tr}(\mathbf{A}) + c)^2}. \tag{88}$$

The corresponding MSE bound is

$$E\left\{\|\hat{\mathbf{x}} - \mathbf{x}_0\|^2\right\} \ge \frac{\mathrm{Tr}(\mathbf{A})}{(\mathrm{Tr}(\mathbf{A}) + c)^2}(\mathrm{Tr}(\mathbf{A})\mathbf{x}_0^*\mathbf{x}_0 + c^2). \tag{89}$$

This result is summarized in the following theorem.

Theorem 3: Let $\mathbf{y}$ denote measurements of a deterministic parameter vector $\mathbf{x}_0$ with pdf $p(\mathbf{y}; \mathbf{x}_0)$. Assume that the Fisher information with respect to $\mathbf{x}_0$ has the form $\mathbf{J}(\mathbf{x}_0) = \mathbf{A}^{-1}$ and that $\|\mathbf{x}_0\|^2 \le c$. If there exists an efficient estimator $\hat{\mathbf{x}}$, then the estimator

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

achieves the bound (89) and has smaller MSE than $\hat{\mathbf{x}}$ for all $\|\mathbf{x}_0\|^2 \le c$.

Closed-form expressions for $\hat{\mathbf{x}}_{\mathrm{b}}$ when $\mathbf{J}(\mathbf{x}_0) = \mathbf{A}^{-1}$ can also be obtained in the case of a weighted norm constraint of the form $\mathbf{x}_0^*\mathbf{T}\mathbf{x}_0 \le c$ for certain choices of $\mathbf{T} \succ 0$ using similar techniques as those used in [34].

The estimator $\hat{\mathbf{x}}_{\mathrm{b}}$ of Theorem 3 is a shrinkage estimator, i.e., a constant multiple of the unbiased estimator $\hat{\mathbf{x}}$. Estimators of this type have been used extensively in the literature [43], [19], [17], [21] following the seminal work of James and Stein [16].

A special case of Theorem 3 is the linear Gaussian model in which $\mathbf{y} = \mathbf{H}\mathbf{x}_0 + \mathbf{w}$, where $\mathbf{H}$ is a known full-rank matrix, and $\mathbf{w}$ is a zero-mean Gaussian random vector with covariance $\mathbf{C} \succ 0$. In this case, $\mathbf{J}^{-1}(\mathbf{x}_0) = (\mathbf{H}^*\mathbf{C}^{-1}\mathbf{H})^{-1}$. The estimator $\hat{\mathbf{x}}_{\mathrm{b}}$ resulting from the theorem is equal to the minimax MSE estimator developed in [17], which minimizes the worst case MSE over all linear estimators.[1] This estimator was shown in [18] to dominate the least squares estimator for all norm bounded parameters; since the least squares estimator is an efficient estimator for our problem, our results agree with those of [18].

VIII. Example

In the previous sections, we showed *analytically* that the CRLB can be uniformly improved upon using a linear bias. We also discussed how to construct an estimator whose MSE is uniformly lower than a given efficient estimator. Here we demonstrate that these results can be used in practical settings even when an efficient estimator is unknown.

Specifically, we propose a linearly modified ML estimator in which we multiply the ML solution by a strictly dominating matrix $\mathbf{M}$ on the appropriate set, even in cases when the ML estimator is not efficient. We now demonstrate, through an example, that we can often gain in performance by using such an approach.

Suppose we wish to estimate the SNR of a constant signal in Gaussian noise, from N i.i.d. measurements

$$y_i = \mu + w_i, \quad 1 \le i \le N \tag{90}$$

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

$$\hat{x} = \frac{\hat{\mu}^2}{\hat{\sigma}^2} \tag{91}$$

where

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} y_i, \quad \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{\mu})^2. \tag{92}$$

In general, $\hat{x}$ is biased and does not achieve the CRLB.

As we have seen in Example 5 in Section II, the inverse Fisher information in this case is

$$J^{-1}(x) = \frac{1}{N}(4x + 2x^2). \tag{93}$$

In addition, we know that $x \ge 0$ for all choices of μ and σ^2. Thus, to obtain a lower bound than the CRLB, we may seek the scalar $\hat{M}$ that is the solution to

$$\min_{M} \max_{x \ge 0} \left\{x^2M^2 + ((1+M)^2 - 1)J^{-1}(x)\right\}. \tag{94}$$

The optimal value of M can then be found using the SDP formulation of Section VII. Indeed, the constraint $x \ge 0$ can be written as $x \in \mathcal{Q}$, where $\mathcal{Q}$ is defined by (43) with $A = 0$, $b = 1$, $c = 0$. 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 and the linear ML estimators as a function of the number of observations N for an SNR of $x = 2$. For each value of N, the MSE is averaged over 10 000 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 Fig. 2, we plot the value of $1+M$ as a function of N for the example in Fig. 1.

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 we know that $\alpha \le x \le \beta$ for some values of α and β. The ML estimator in this case is

$$\hat{x}_c = \begin{cases} \hat{x}, & \alpha \le \hat{x} \le \beta \\ \alpha, & \hat{x} \le \alpha \\ \beta, & \hat{x} \ge \beta \end{cases} \tag{95}$$

where $\hat{x} = \hat{\mu}^2/\hat{\sigma}^2$.

[1]The minimax MSE estimator under a weighted norm constraint was also treated in [44]. In the unweighted case, the resulting estimator is equal to that of [17]. However, it was shown in [34] that there is an error in the derivations in [44], which leads to an erroneous expression for the minimax MSE estimator in the general case.

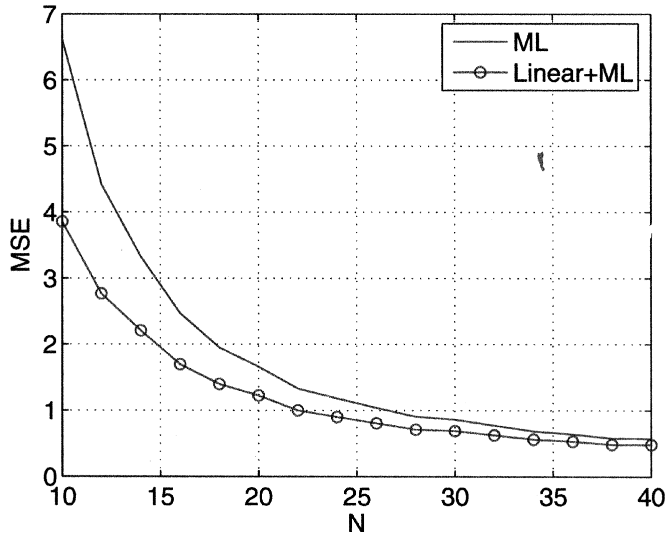

Fig. 1. MSE in estimating the SNR as a function of the number of observations N for an SNR of two using the ML and the linearly transformed ML estimators.

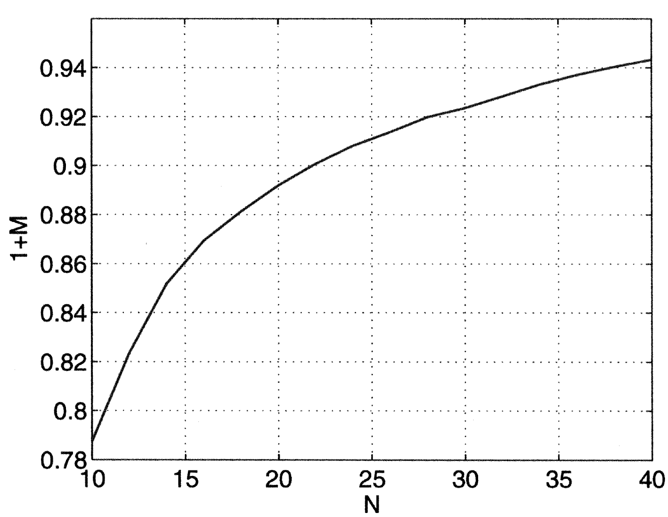

Fig. 2. $1+M$ as a function of the number of observations N when estimating an SNR of two.

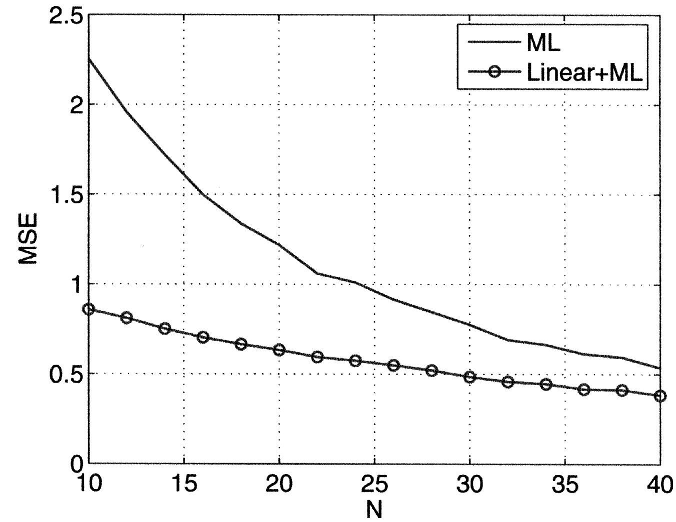

Fig. 3. MSE in estimating the SNR as a function of the number of observations N for an SNR of two using the ML and the linearly transformed ML estimators subject to the constraint (96).

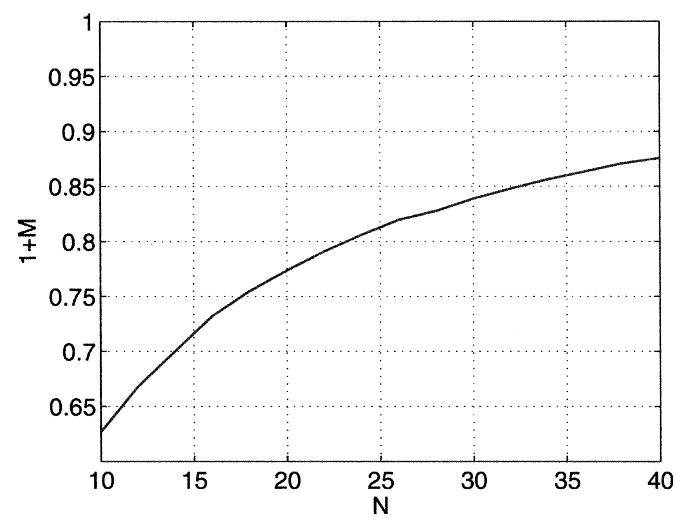

Fig. 4. $1+M$ as a function of the number of observations N when estimating an SNR of two subject to the constraint (96).

To develop a linear modification of the ML approach, we note that the constraint $\alpha \leq x \leq \beta$ can be written as

$$(x-\alpha)(x-\beta) = x^2 - (\alpha+\beta)x + \alpha\beta \leq 0. \tag{96}$$

We can now solve the optimization problem with respect to the set (96).

In Fig. 3, we compare the MSE of the constrained ML and the linear ML estimators subject to (96), for an SNR of $x = 2$ and SNR bounds $\alpha = 1$, $\beta = 5$. For each value of N, the MSE is averaged over 10 000 noise realizations. As can be seen from the figure, the linearly modified ML estimator performs significantly better than the ML methods. In Fig. 4, we plot the value of $1+M$ as a function of N.

IX. 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 $\mathbf{x}_0$. However, the basic concepts and tools we proposed are relevant in a more general context and can be used for other forms of the CRLB, as well as other classes of bias vectors.

It is well known that some of the existing modifications of ML, such as the James–Stein estimator, can be viewed in an empirical Bayes framework [45]. An interesting direction for future research is to explore the connection between the proposed methods and empirical Bayes approaches [46].

APPENDIX I
SCHUR'S LEMMA

The following result is referred to as Schur's lemma [47, p. 28]. Let

$$\mathbf{M} = \begin{bmatrix} \mathbf{X} & \mathbf{Y}^* \\ \mathbf{Y} & \mathbf{Z} \end{bmatrix}$$

be a Hermitian matrix. Then $\mathbf{M} \succeq (\succ)\ 0$ if and only if $\mathbf{Z} \succeq (\succ)\ 0$, $\mathbf{X} - \mathbf{Y}^*\mathbf{Z}^\dagger\mathbf{Y} \succeq (\succ)\ 0$, and $\mathbf{Y}^*(\mathbf{I} - \mathbf{Z}\mathbf{Z}^\dagger) = 0$. Equivalently, $\mathbf{M} \succeq (\succ)\ 0$ if and only if $\mathbf{X} \succeq (\succ)\ 0$, $\mathbf{Z} - \mathbf{Y}\mathbf{X}^\dagger\mathbf{Y}^* \succeq (\succ)\ 0$, and $\mathbf{Y}(\mathbf{I} - \mathbf{X}\mathbf{X}^\dagger) = 0$.

APPENDIX II
REDUCING A FEASIBLE PROBLEM TO A STRICTLY FEASIBLE PROBLEM

Consider the constraint

$$\mathbf{G}(\mathbf{M}) \triangleq \begin{bmatrix} \mathbf{A}_0(\mathbf{M}) & \mathbf{b}_0(\mathbf{M}) \\ \mathbf{b}_0^*(\mathbf{M}) & \mathbf{c}_0(\mathbf{M}) - t \end{bmatrix} \preceq 0 \tag{97}$$

and suppose that it is not strictly feasible. We now show how to reduce it to a strictly feasible constraint together with linear equalities.

Lemma 4: The constraint $\mathbf{G}(\mathbf{M}) \preceq 0$ is equivalent to

$$\begin{aligned} \mathbf{G}_2(\mathbf{M}) &\triangleq \begin{bmatrix} \mathbf{V}_2^*\mathbf{A}_0(\mathbf{M})\mathbf{V}_2 & \mathbf{V}_2^*\mathbf{b}_0(\mathbf{M}) \\ \mathbf{b}_0^*(\mathbf{M})\mathbf{V}_2 & c_0(\mathbf{M}) - t \end{bmatrix} \preceq 0 \\ \mathbf{M}\mathbf{V}_1 &= 0 \\ \mathbf{V}_1^*\mathbf{b}_0(\mathbf{M}) &= 0, \end{aligned} \tag{98}$$

where the r columns of $\mathbf{V}_1$ form an orthonormal basis for $\mathcal{N}(\mathbf{W})$ with $\mathbf{W} = \sum_{i=1}^{\ell} \mathbf{B}_i^*\mathbf{B}_i$ and $\mathbf{V}_2$ is an $m \times (m - r)$ matrix with orthonormal columns that are orthogonal to the columns of $\mathbf{V}_1$.

Proof: Suppose first that $\mathbf{G}(\mathbf{M}) \preceq 0$, and let $\widetilde{\mathbf{V}}$ be the unitary matrix defined by

$$\widetilde{\mathbf{V}} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{99}$$

Since $\widetilde{\mathbf{V}}$ is unitary, $\widetilde{\mathbf{V}}^*\mathbf{G}(\mathbf{M})\widetilde{\mathbf{V}} \preceq 0$. Now

$$\widetilde{\mathbf{V}}^*\mathbf{G}(\mathbf{M})\widetilde{\mathbf{V}} = \left[\begin{array}{c|c|c} \mathbf{V}_1^*\mathbf{M}^*\mathbf{M}\mathbf{V}_1 & \mathbf{V}_1^*\mathbf{M}^*\mathbf{M}\mathbf{V}_2 & \mathbf{V}_1^*\mathbf{b}_0(\mathbf{M}) \\ \hline \mathbf{V}_2^*\mathbf{M}^*\mathbf{M}\mathbf{V}_1 & \mathbf{V}_2^*\mathbf{A}_0(\mathbf{M})\mathbf{V}_2 & \mathbf{V}_2^*\mathbf{b}_0(\mathbf{M}) \\ \hline \mathbf{b}_0^*(\mathbf{M})\mathbf{V}_1 & \mathbf{b}_0^*(\mathbf{M})\mathbf{V}_2 & c_0(\mathbf{M}) - t \end{array}\right] \tag{100}$$

where we used the fact that $\mathbf{A}_0(\mathbf{M})\mathbf{V}_1 = \mathbf{M}^*\mathbf{M}\mathbf{V}_1$. To ensure that $\widetilde{\mathbf{V}}^*\mathbf{G}(\mathbf{M})\widetilde{\mathbf{V}} \preceq 0$, we must have $\mathbf{V}_1^*\mathbf{M}^*\mathbf{M}\mathbf{V}_1 \preceq 0$, which implies that $\mathbf{M}\mathbf{V}_1 = 0$ and

$$\widetilde{\mathbf{V}}^*\mathbf{G}(\mathbf{M})\widetilde{\mathbf{V}} = \left[\begin{array}{c|c|c} 0 & 0 & \mathbf{V}_1^*\mathbf{b}_0(\mathbf{M}) \\ \hline 0 & \mathbf{V}_2^*\mathbf{A}_0(\mathbf{M})\mathbf{V}_2 & \mathbf{V}_2^*\mathbf{b}_0(\mathbf{M}) \\ \hline \mathbf{b}_0^*(\mathbf{M})\mathbf{V}_1 & \mathbf{b}_0^*(\mathbf{M})\mathbf{V}_2 & c_0(\mathbf{M}) - t \end{array}\right]. \tag{101}$$

From (101), $\widetilde{\mathbf{V}}^*\mathbf{G}(\mathbf{M})\widetilde{\mathbf{V}} \preceq 0$ only if $\mathbf{V}_1^*\mathbf{b}_0(\mathbf{M}) = 0$ and $\mathbf{G}_2(\mathbf{M}) \preceq 0$.

Conversely, suppose that (98) holds. Then from (100) we have immediately that $\widetilde{\mathbf{V}}^*\mathbf{G}(\mathbf{M})\widetilde{\mathbf{V}} \preceq 0$, which implies that $\mathbf{G}(\mathbf{M}) \preceq 0$. ∎

It is easy to see that $\mathbf{G}_2(\mathbf{M})$ defined in Lemma 4 is always strictly feasible. Indeed, by our definition of $\mathbf{V}_2$

$$\mathbf{W}_2 = \sum_{i=1}^{\ell} \mathbf{V}_2^*\mathbf{B}_i^*\mathbf{B}_i\mathbf{V}_2 \succ 0. \tag{102}$$

Strict feasibility then follows in the same way as in Lemma 1.

APPENDIX III
PROOF OF LEMMA 2

In this Appendix, we prove that (52) is equivalent to (53).

We first note that (52) can be written as

$$\min_{t,\mathbf{M}} \{t : \mathbf{Z}(\mathbf{M}, \mathbf{X}) \preceq 0, \mathbf{X} = \mathbf{M}^*\mathbf{M}\}. \tag{103}$$

This follows from simply substituting $\mathbf{X} = \mathbf{M}^*\mathbf{M}$ into $\mathbf{G}(\mathbf{M})$ and noting that $\mathbf{Z}(\mathbf{M}, \mathbf{M}^*\mathbf{M}) = \mathbf{G}(\mathbf{M})$. The problem is that the constraint $\mathbf{X} = \mathbf{M}^*\mathbf{M}$ is not convex. To obtain a convex problem, we would like to relax this constraint to the convex form $\mathbf{X} \succeq \mathbf{M}^*\mathbf{M}$, leading to (53). As we now show, the original and relaxed problems have the same solution and are therefore equivalent. To this end, it is sufficient to show that if $\hat{t}$, $\widehat{\mathbf{M}}$, and $\widehat{\mathbf{X}}$ are optimal for (53), then we can achieve the same value $\hat{t}$ with $\mathbf{X} = \widehat{\mathbf{M}}^*\widehat{\mathbf{M}}$.

To prove the result, we therefore need to show that if $\mathbf{Z}(\widehat{\mathbf{M}}, \widehat{\mathbf{X}}) \preceq 0$, then $\mathbf{Z}(\widehat{\mathbf{M}}, \widehat{\mathbf{M}}^*\widehat{\mathbf{M}}) \preceq 0$. Now

$$\mathbf{Z}(\widehat{\mathbf{M}}, \widehat{\mathbf{X}}) - \mathbf{Z}(\widehat{\mathbf{M}}, \widehat{\mathbf{M}}^*\widehat{\mathbf{M}}) = \begin{bmatrix} \mathbf{Y} + \sum_{i=1}^{\ell} \mathbf{B}_i^*\mathbf{Y}\mathbf{B}_i & \sum_{i=1}^{k} \mathbf{C}_i^*\mathbf{Y}\mathbf{z}_i \\ \sum \mathbf{z}_i^*\mathbf{Y}\mathbf{C}_i & \mathrm{Tr}(\mathbf{A}\mathbf{Y}) \end{bmatrix} \triangleq \mathbf{W}(\mathbf{Y}) \tag{104}$$

where we defined

$$\mathbf{Y} = \widehat{\mathbf{X}} - \widehat{\mathbf{M}}^*\widehat{\mathbf{M}}. \tag{105}$$

Since $\widehat{\mathbf{X}}$ and $\widehat{\mathbf{M}}$ are feasible, $\mathbf{Y} \succeq 0$. We now show that $\mathbf{W}(\mathbf{Y}) \succeq 0$ for all $\mathbf{Y} \succeq 0$. This follows from the definition of

$$\min_{t,\mathbf{M},\mathbf{X}} \{t : \mathbf{Z}_2(\mathbf{M},\mathbf{X}) \preceq 0,\ \mathbf{X}\mathbf{V}_1 = 0, \mathbf{V}_1^*\mathbf{b}_0(\mathbf{M},\mathbf{X}) = 0,\ \mathbf{X} \succeq \mathbf{M}^*\mathbf{M}\} \tag{109}$$

$$\min_{t,\mathbf{M},\mathbf{X}} \{t : \mathbf{Z}_2(\mathbf{M},\mathbf{X}) \preceq 0,\ \mathbf{X}\mathbf{V}_1 = 0, \mathbf{V}_1^*\mathbf{b}_0(\mathbf{M},\mathbf{X}) = 0,\ \mathbf{X} = \mathbf{M}^*\mathbf{M}\} \tag{111}$$

$\mathbf{J}^{-1}(\mathbf{x})$ and the fact that $\mathbf{J}^{-1}(\mathbf{x}) \succ 0$. Since $\mathbf{J}^{-1}(\mathbf{x}) \succ 0$ for all $\mathbf{x}$

$$\mathrm{Tr}\left(\mathbf{Y}\mathbf{J}^{-1}(\mathbf{x})\right) \succeq 0 \text{ for all } \mathbf{x},\ \text{and all } \mathbf{Y} \succeq 0. \tag{106}$$

Now

$$\begin{aligned}\mathrm{Tr}(\mathbf{Y}\mathbf{J}^{-1}(\mathbf{x})) &= [\mathbf{x} \quad 1]^* \begin{bmatrix} \sum_{i=1}^{\ell} \mathbf{B}_i^*\mathbf{Y}\mathbf{B}_i & \sum_{i=1}^{k} \mathbf{C}_i^*\mathbf{Y}\mathbf{z}_i \\ \sum_{i=1}^{k} \mathbf{z}_i^*\mathbf{Y}\mathbf{C}_i & \mathrm{Tr}(\mathbf{Y}\mathbf{A}) \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} \\ &= [\mathbf{x} \quad 1]^* \, \mathbf{W}(\mathbf{Y}) \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}. \end{aligned} \tag{107}$$

Combining (107) with the fact that $\sum_{i=1}^{\ell} \mathbf{B}_i^*\mathbf{Y}\mathbf{B}_i \succeq 0$, we conclude that $\mathbf{c}^*\mathbf{W}(\mathbf{Y})\mathbf{c} \succeq 0$ for any vector $\mathbf{c}$, so that $\mathbf{W}(\mathbf{Y}) \succeq 0$ for any $\mathbf{Y} \succeq 0$.

APPENDIX IV
PROOF OF PROPOSITION 1

To prove the proposition, we first establish the following lemma, which is analogous to Lemma 2.

Lemma 5: The problem

$$\min_{t,\mathbf{M}} \{t : \mathbf{G}_2(\mathbf{M}) \preceq 0,\ \mathbf{M}\mathbf{V}_1 = 0,\ \mathbf{V}_1^*\mathbf{b}_0(\mathbf{M}) = 0\} \tag{108}$$

with $\mathbf{G}_2(\mathbf{M})$ given by (98) and $\mathbf{V}_1$ a matrix whose r columns form an orthonormal basis for $\mathcal{N}(\mathbf{W})$ with $\mathbf{W} = \sum_{i=1}^{\ell} \mathbf{B}_i^*\mathbf{B}_i$ is equivalent to the convex problem shown in (109) at the top of the page, where

$$\mathbf{Z}_2(\mathbf{M},\mathbf{X}) = \begin{bmatrix} \mathbf{V}_2^*\left(\mathbf{X} + \sum_{i=1}^{\ell} \mathbf{B}_i^*\Phi\mathbf{B}_i\right)\mathbf{V}_2 & \mathbf{V}_2^*\mathbf{b}_0(\mathbf{M},\mathbf{X}) \\ \mathbf{b}_0^*(\mathbf{M},\mathbf{X})\mathbf{V}_2 & \mathrm{Tr}\left(\mathbf{A}\Phi\right) - t \end{bmatrix} \tag{110}$$

in which for brevity we denoted $\Phi = \mathbf{X}+\mathbf{M}+\mathbf{M}^*$, $\mathbf{V}_2$ is an $m \times (m-r)$ matrix with orthonormal columns that are orthogonal to the columns of $\mathbf{V}_1$, and $\mathbf{b}_0(\mathbf{M},\mathbf{X}) = \sum_{i=1}^{k} \mathbf{C}_i^*\Phi\mathbf{z}_i$.

Proof: By substituting $\mathbf{X} = \mathbf{M}^*\mathbf{M}$ into (108), our problem can be written as shown in (111) at the top of the page. This follows from the fact that $\mathbf{Z}_2(\mathbf{M},\mathbf{M}^*\mathbf{M}) = \mathbf{G}_2(\mathbf{M})$, and that $\mathbf{M}\mathbf{V}_1 = 0$ if and only if $\mathbf{M}^*\mathbf{M}\mathbf{V}_1 = 0$. As in the proof of Lemma 2 (see Appendix C), we now show that we can relax the nonconvex constraint $\mathbf{X} = \mathbf{M}^*\mathbf{M}$ to the convex form $\mathbf{X} \succeq \mathbf{M}^*\mathbf{M}$, leading to (109). To this end it is sufficient to show that if $\hat{t}$, $\widehat{\mathbf{M}}$, and $\widehat{\mathbf{X}}$ are optimal for (109), then we can achieve the same value $\hat{t}$ with $\mathbf{X} = \widehat{\mathbf{M}}^*\widehat{\mathbf{M}}$.

We first note that

$$\mathbf{Z}_2(\widehat{\mathbf{M}},\widehat{\mathbf{X}}) - \mathbf{Z}_2(\widehat{\mathbf{M}},\widehat{\mathbf{M}}^*\widehat{\mathbf{M}}) = \mathbf{Q}^*\mathbf{W}(\mathbf{Y})\mathbf{Q} \tag{112}$$

where $\mathbf{W}(\mathbf{Y})$ and $\mathbf{Y}$ are defined in (104) and (105), respectively, and

$$\mathbf{Q} = \begin{bmatrix} \mathbf{V}_2 & 0 \\ 0 & 1 \end{bmatrix}. \tag{113}$$

In Appendix C, it was shown that $\mathbf{W}(\mathbf{Y}) \succeq 0$ for any $\mathbf{Y} \succeq 0$ from which it follows that $\mathbf{Q}^*\mathbf{W}(\mathbf{Y})\mathbf{Q} \succeq 0$ so that $\mathbf{G}_2(\widehat{\mathbf{M}}) = \mathbf{Z}_2(\widehat{\mathbf{M}},\widehat{\mathbf{M}}^*\widehat{\mathbf{M}}) \preceq \mathbf{Z}_2(\widehat{\mathbf{M}},\widehat{\mathbf{X}}) \preceq 0$. The last inequality follows from the fact that $\widehat{\mathbf{M}}$ and $\widehat{\mathbf{X}}$ are feasible for (109).

It remains to show that $\mathbf{X} = \widehat{\mathbf{M}}^*\widehat{\mathbf{M}}$ satisfies the conditions $\mathbf{X}\mathbf{V}_1 = 0$ and $\mathbf{V}_1^*\mathbf{b}_0(\widehat{\mathbf{M}},\mathbf{X}) = 0$. Since $\widehat{\mathbf{M}}$ and $\widehat{\mathbf{X}}$ are feasible and $\mathbf{B}_i\mathbf{V}_1 = 0$, we have that

$$\widetilde{\mathbf{V}}^*\mathbf{Z}(\widehat{\mathbf{M}},\widehat{\mathbf{X}})\widetilde{\mathbf{V}} = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & \mathbf{Z}_2(\widehat{\mathbf{M}},\widehat{\mathbf{X}}) \end{array}\right] \preceq 0 \tag{114}$$

where $\mathbf{Z}(\mathbf{M},\mathbf{X})$ is defined in (54) and $\widetilde{\mathbf{V}}$ is defined by (99). From (114) we conclude that $\mathbf{Z}(\widehat{\mathbf{M}},\widehat{\mathbf{X}}) \preceq 0$. Using the proof of Lemma 2 in Appendix C, we then have that $\mathbf{Z}(\widehat{\mathbf{M}},\widehat{\mathbf{M}}^*\widehat{\mathbf{M}}) \preceq 0$, which in turn implies from Lemma 4 that $\mathbf{X}\mathbf{V}_1 = 0$ and $\mathbf{V}_1^*\mathbf{b}_0(\widehat{\mathbf{M}},\mathbf{X}) = 0$. ■

Once we have shown that (108) is equivalent to the convex problem (109), we can use the KKT conditions to show that $\widehat{\mathbf{M}} \neq 0$, just as in Section VI-B.

REFERENCES

[1] H. Cramer, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton Univ. Press, 1946.
[2] C. R. Rao, *Linear Statistical Inference and its Applications*, 2nd ed. New York: Wiley, 1973.
[3] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1994.
[4] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed. New York: Springer-Verlag, 1998.
[5] B. Efron, "Biased versus unbiased estimation," *Adv. Math.*, vol. 16, pp. 259–277, 1975.
[6] E. L. Lehmann, "A general concept of unbiasedness," *Ann. Math. Statist.*, vol. 22, pp. 587–592, Dec. 1951.
[7] P. R. Halmos, "The theory of unbiased estimation," *Ann. Math. Statist.*, vol. 17, pp. 34–43, 1946.
[8] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems," *IEEE Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 2024–2036, Dec. 1989.
[9] D. M. Titterington, "Common structure of smoothing techniques in statistics," *Int. Statist. Rev.*, vol. 53, pp. 141–170, 1985.
[10] F. O'Sullivan, "A statistical perspective on ill-posed inverse problems," *Statist. Sci.*, vol. 1, no. 4, pp. 502–527, 1986.
[11] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1997.
[12] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1968.

[13] A. O. Hero, J. A. Fessler, and M. Usman, "Exploring estimator bias-variance tradeoffs using the uniform CR bound," *IEEE Trans. Signal Process.*, vol. 44, no. 8, pp. 2026–2041, Aug. 1996.

[14] Y. C. Eldar, "Minimum variance in biased estimation: Bounds and asymptotically optimal estimators," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1915–1930, Jul. 2004.

[15] C. Stein, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," in *Proc. 3rd Berkeley Symp. Math. Statist. Prob.*, Berkeley, CA, 1956, vol. 1, pp. 197–206.

[16] W. James and C. Stein, "Estimation of quadratic loss," in *Proc. 4th Berkeley Symp. Math. Statist. Prob.*, Berkeley, CA, 1961, vol. 1, pp. 361–379.

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

[18] Z. Ben-Haim and Y. C. Eldar, "Maximum set estimators with bounded estimation error," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3172–3182, Aug. 2005.

[19] M. E. Bock, "Minimax estimators of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 3, no. 1, pp. 209–218, Jan. 1975.

[20] J. O. Berger, "Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss," *Ann. Statist.*, vol. 4, no. 1, pp. 223–226, Jan. 1976.

[21] Z. Ben-Haim and Y. C. Eldar, "Blind minimax estimators: Improving on least squares estimation," presented at the IEEE Workshop Statistical Signal Process. (SSP'05) Bordeaux, France, Jul. 2005.

[22] ——, "Blind minimax estimation," Electrical Engineering Dept., Technion—Israel Inst. of Technology, Aug. 2005, CCIT Rep. 550.

[23] B. J. N. Blight, "Some general results on reduced mean square error estimation," *Amer. Statist.*, vol. 25, pp. 24–25, Jun. 1971.

[24] M. D. Perlman, "Reduced mean square error estimation for several parameters," *Sankhya*, vol. 34, pp. 89–92, 1972.

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

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

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

[28] S. Zacks, *The Theory of Statistical Inference*. New York: Wiley, 1971.

[29] P. Stoica and R. L. Moses, "On biased estimators and the unbiased Cramér-Rao lower bound," *Signal Process.*, vol. 21, pp. 349–350, 1990.

[30] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*. New York: Hafner, 1961.

[31] J. Beyerer, "Is it useful to know a nuisance parameter?," *Signal Process.*, vol. 68, pp. 107–111, 1998.

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

[33] D. P. Bertsekas, *Nonlinear Programming*, Second ed. Belmont, MA: Athena Scientific, 1999.

[34] Y. C. Eldar, "Comparing between estimation approaches: Admissible and dominating linear estimators," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1689–1702, May 2006.

[35] 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," *Eur. J. Oper. Res.*, vol. 151, no. 3, pp. 461–473, 2003.

[36] 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.

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

[38] D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space*. New York: Springer-Verlag, 1991.

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

[40] A. Beck and Y. C. Eldar, "Strong duality in nonconvex quadratic optimization with two quadratic constraints," *SIAM J. Optimiz.*, to be published.

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

[42] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

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

[44] J. Pilz, "Minimax linear regression estimation with symmetric parameter restrictions," *J. Stat. Plan. Infer.*, vol. 13, pp. 297–318, 1986.

[45] B. Efron and C. Morris, "Stein's estimation rule and its competitors: An empirical Bayes approach," *J. Amer. Statist. Assoc.*, vol. 68, pp. 117–130, 1973.

[46] B. P. Carlin and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*. New York: Chapman and Hall, 1996.

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

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

From January to July 2002, she was a Postdoctoral Fellow with the Digital Signal Processing Group, MIT. She is currently an Associate Professor with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa. She is also a Research Affiliate with the Research Laboratory of Electronics, MIT. Her current research interests are in the general areas of signal processing, statistical signal processing, and computational biology.

Dr. Eldar was in the program for outstanding students at TAU from 1992 to 1996. In 1998, she held the Rosenblith Fellowship for study in Electrical Engineering at MIT, and in 2000, she held an IBM Research Fellowship. She is currently a Horev Fellow of the Leaders in Science and Technology program at the Technion and an Alon Fellow. In 2004, she received the Wolf Foundation Krill Prize for Excellence in Scientific Research, and in 2005 the Andre and Bella Meyer Lectureship. She is a member of the IEEE Signal Processing Theory and Methods Technical Committee, and an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.