MSE Bounds With Affine Bias Dominating the Cramér–Rao Bound

Yonina C. Eldar, *Senior Member, IEEE*

***Abstract*—In continuation to an earlier work, we further develop bounds on the mean-squared error (MSE) when estimating a deterministic parameter vector $\boldsymbol{\theta}_0$ in a given estimation problem, as well as estimators that achieve the optimal performance. The traditional Cramér–Rao (CR) type bounds provide benchmarks on the variance of any estimator of $\boldsymbol{\theta}_0$ under suitable regularity conditions, while requiring *a priori* specification of a desired bias gradient. To circumvent the need to choose the bias, which is impractical in many applications, it was suggested in our earlier work to directly treat the MSE, which is the sum of the variance and the squared-norm of the bias. While previously we developed MSE bounds assuming a *linear* bias vector, here we study, in the same spirit, *affine* bias vectors. We demonstrate through several examples that allowing for an affine transformation can often improve the performance significantly over a linear approach. Using convex optimization tools we show that in many cases we can choose an affine bias that results in an MSE bound that is smaller than the unbiased CR bound for all values of $\boldsymbol{\theta}_0$. Furthermore, we explicitly construct estimators that achieve these bounds in cases where an efficient estimator exists, by performing an affine transformation of the standard maximum likelihood (ML) estimator. This leads to estimators that result in a smaller MSE than ML for all possible values of $\boldsymbol{\theta}_0$.**

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

I. Introduction

THE Cramér–Rao lower bound (CRLB) [1], [2] is a classical performance measure which is used pervasively in a wide range of applications. Given a set of measurements $\mathbf{y}$, the CRLB characterizes the smallest achievable total variance of any unbiased estimator of a deterministic parameter vector $\boldsymbol{\theta}_0$, where the relationship between $\mathbf{y}$ and $\boldsymbol{\theta}_0$ is described by the probability density function (pdf) $p(\mathbf{y};\boldsymbol{\theta}_0)$ of $\mathbf{y}$ parameterized by $\boldsymbol{\theta}_0$.

When the measurements $\mathbf{y}$ are related to the unknowns $\boldsymbol{\theta}_0$ through a linear Gaussian model, the maximum likelihood (ML) estimate of $\boldsymbol{\theta}_0$, which is given by the value of $\boldsymbol{\theta}$ that maximizes $p(\mathbf{y};\boldsymbol{\theta})$, is efficient, i.e., it achieves the CRLB. Furthermore, when $\boldsymbol{\theta}_0$ is estimated from independent identically distributed (i.i.d.) measurements, under suitable regularity assumptions, the ML estimator is asymptotically unbiased and efficient [2]. 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 [3]–[5]. Furthermore, in some problems, restricting attention to unbiased approaches leads to unreasonable estimators; see [6] and [7] for several examples. To allow for a nonzero bias, the CRLB has been extended to characterize the total variance of any estimator with a given bias [1]. However, the specification of the biased CRLB requires an *a priori* choice of the bias gradient, which in typical applications is not obvious.

Instead of dealing separately with the bias and variance, we can use the biased CRLB to develop a bound on the mean-squared error (MSE), which is the sum of the total variance and the squared-norm of the bias [9]. This bound depends in general on the specific bias vector and on the true unknown parameter vector $\boldsymbol{\theta}_0$. To improve the CRLB it was suggested to restrict attention to choices of bias that are linear in $\boldsymbol{\theta}_0$, and then optimize the resulting MSE bound over all linear possibilities. A similar strategy was introduced in [10] for certain scalar estimation problems, and later extended in [11] to vector-valued estimates with bias gradient matrix proportional to the identity, and restricted parameter values. Two minimization strategies were developed depending on the structure of the CRLB: In some simple settings it was shown that the MSE bound can be directly minimized. The more typical scenario is that the optimal linear choice depends on $\boldsymbol{\theta}_0$ and therefore cannot be implemented. To overcome this difficulty, a minimax framework was proposed in which the linear bias is determined such that it minimizes the maximum difference between the MSE bound and the CRLB. The MSE using this minimax linear bias was then shown to be smaller than the CRLB for all possible values of $\boldsymbol{\theta}_0$. Furthermore, the minimax bias was used to construct a new estimator which is a linear transformation of the ML strategy. When the ML solution is efficient, this approach was shown to achieve the minimax MSE bound and as such to have smaller MSE than ML for all $\boldsymbol{\theta}_0$.

Here we extend the framework developed in [9] to allow for affine bias vectors. Although the ideas we present are rooted in [9], the addition of an affine term renders the mathematics associated with both problems somewhat different, as becomes evident from the derivations in this paper. Furthermore, as we show, an affine choice can in some cases significantly improve the performance over linear bias vectors and lead to bounds that are intuitively more appealing.

Manuscript received November 23, 2007; revised March 17, 2008. This work was supported in part by the Israel Science Foundation by Grant 1081/07 and by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM++ (Contract 216715). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Erik G. Larsson.

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

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2008.925584

We begin in Section II by summarizing the relevant results from [9] and introducing the affine MSE bound. In Section III we motivate the use of an affine bias in contrast to a linear one via a simple example. Based on ideas of [9] and [12] we present, in Section IV, a minimax optimization problem that provides a concrete method for finding an affine bias vector such that the resulting MSE bound is smaller than the CRLB for all values of $\boldsymbol{\theta}_0$. We then restrict our attention to estimation problems in which the CRLB is quadratic in $\boldsymbol{\theta}_0$, and analyze the resulting problem using convex analysis methods for two special cases: In Section V, we consider the case in which $\boldsymbol{\theta}_0$ is not restricted. In Section VI, we treat the case in which $\boldsymbol{\theta}_0$ 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 an affine bias vector exists such that the resulting MSE bound is smaller than the CRLB for all possible values of $\boldsymbol{\theta}_0$. Such a bias can be found as a solution to a semidefinite programming problem (SDP) which is a tractable convex problem that can be solved very efficiently [13], [14]. We then develop necessary and sufficient optimality conditions in both scenarios 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. As an example, we discuss the minimax MSE estimator for the linear Gaussian model and show how our general results apply to this problem. In Section VII, we demonstrate through an example that by an affine transformation of the ML estimator, we can reduce the MSE for all values of $\boldsymbol{\theta}_0$, and improve the performance over a linear approach.

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 $\mathbf{I}$, $\mathrm{diag}_{m,n}(\delta_1,\ldots,\delta_m)$ is an $m\times n$ diagonal matrix with diagonal elements δ_i and $\hat{(\cdot)}$ denotes an estimated vector or matrix. For a given matrix $\mathbf{A}$, $\mathbf{A}^*$ and $\mathbf{A}^\dagger$ are the Hermitian conjugate and Moore-Penrose pseudoinverse, and x_i is the ith component of the vector $\boldsymbol{\theta}$. The true value of an unknown vector parameter $\boldsymbol{\theta}$ is denoted by $\boldsymbol{\theta}_0$. The gradient of a vector $\partial\mathbf{b}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}$ is a matrix, with ijth element equal to $\partial b_i(\boldsymbol{\theta}_0)/\partial x_j$. 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 $\mathbf{A}-\mathbf{B}\succeq 0$. The notation $\Re\{a\}$ denotes the real part of the variable a.

II. MSE Bound

We treat the problem of estimating a deterministic parameter vector $\boldsymbol{\theta}_0\in\mathbb{C}^m$ from a given measurement vector $\mathbf{y}\in\mathbb{C}^n$, that is related to $\boldsymbol{\theta}_0$ through the pdf $p(\mathbf{y};\boldsymbol{\theta}_0)$. As our performance measure, we focus on the MSE which is defined by

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

where $\mathbf{b}(\boldsymbol{\theta}_0)=E\{\hat{\boldsymbol{\theta}}\}-\boldsymbol{\theta}_0$ is the bias vector of $\hat{\boldsymbol{\theta}}$ and

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}}=E\left\{[\hat{\boldsymbol{\theta}}-E\{\hat{\boldsymbol{\theta}}\}][\hat{\boldsymbol{\theta}}-E\{\hat{\boldsymbol{\theta}}\}]^*\right\} \tag{2}$$

is its covariance matrix.

Under suitable regularity conditions on $p(\mathbf{y};\boldsymbol{\theta})$ (see, e.g., [1], [2], and [15]), the MSE of any unbiased estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_0$ is bounded below by the CRLB

$$E\left\{\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|^2\right\}\geq\mathrm{Tr}\left(\mathbf{J}^{-1}(\boldsymbol{\theta}_0)\right) \tag{3}$$

where $\mathbf{J}(\boldsymbol{\theta}_0)$ is the Fisher information matrix

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

and is assumed throughout to be nonsingular. An estimator achieving the CRLB is referred to as efficient, and has minimum variance among all unbiased strategies. There are a variety of estimation problems in which the CRLB cannot be achieved, but nonetheless a minimum variance unbiased (MVU) approach exsits. The discussion in the remainder of the paper also holds true when we replace the CRLB $\mathbf{J}^{-1}(\boldsymbol{\theta}_0)$ everywhere by the variance of an MVU estimator. In this case, the proposed estimators are affine transformations of the corresponding MVU strategy.

Using a biased version of the CLRB [1] it can be readily shown (see, e.g., [9]) that the MSE of any estimator with bias $\mathbf{b}(\boldsymbol{\theta}_0)$ is bounded by

$$\|\mathbf{b}(\boldsymbol{\theta}_0)\|^2+\mathrm{Tr}\left((\mathbf{I}+\mathbf{D}(\boldsymbol{\theta}_0))\,\mathbf{J}^{-1}(\boldsymbol{\theta}_0)\,(\mathbf{I}+\mathbf{D}(\boldsymbol{\theta}_0))^*\right) \tag{5}$$

where $\mathbf{D}(\boldsymbol{\theta}_0)=\partial\mathbf{b}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}$. If the bias vector $\mathbf{b}(\boldsymbol{\theta}_0)$ is not restricted, then it can be chosen such that the MSE bound is equal to 0; thus, minimizing the bound over all bias vectors leads to a useless result. Instead, it was proposed in [9] to restrict attention to linear bias vectors of the form $\mathbf{b}(\boldsymbol{\theta}_0)=\mathbf{M}\boldsymbol{\theta}_0$ for some matrix $\mathbf{M}$, in which case the bound of (5) becomes

$$\mathrm{MSEB}(\mathbf{M},\boldsymbol{\theta}_0)=\boldsymbol{\theta}_0^*\mathbf{M}^*\mathbf{M}\boldsymbol{\theta}_0+\mathrm{Tr}\left((\mathbf{I}+\mathbf{M})\mathbf{J}^{-1}(\boldsymbol{\theta}_0)(\mathbf{I}+\mathbf{M})^*\right) \tag{6}$$

and then optimizing (6) over all matrices $\mathbf{M}$.

The advantage of considering linear bias vectors is twofold: The first reason is computational. The second is statistical and relates to the fact that once we find a lower bound we would also like to determine a method that achieves it at least in some special cases. With linear bias vectors this is possible. Specifically, if $\hat{\boldsymbol{\theta}}$ is an efficient unbiased estimator, then the MSE of $\hat{\boldsymbol{\theta}}_{\mathrm{b}}=(\mathbf{I}+\mathbf{M})\hat{\boldsymbol{\theta}}$ is given by (6). Therefore, if we find an $\mathbf{M}$ such that $\mathrm{MSEB}\,(\mathbf{M},\boldsymbol{\theta}_0)<\mathrm{Tr}(\mathbf{J}^{-1}(\boldsymbol{\theta}_0))$ for a suitable set of $\boldsymbol{\theta}_0$, then the MSE of $\hat{\boldsymbol{\theta}}_{\mathrm{b}}$ will be smaller than that of $\hat{\boldsymbol{\theta}}$ for all $\boldsymbol{\theta}_0$ in the set. In contrast, if we consider more general nonlinear bias vectors, then even if we find a bias that leads to 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.

Two approaches to optimizing $\mathbf{M}$ in (6) were discussed in [9]. When estimating a scalar with Fisher information satisfying $J^{-1}(\theta_0)=\alpha\theta_0^2$ for some $\alpha>0$, it was shown that the linear bound can be minimized for all θ_0. In more general settings, the optimal choice of $\mathbf{M}$ depends on $\boldsymbol{\theta}_0$, and therefore cannot be implemented. Instead, it was suggested to consider an admissible and dominating choice [16], i.e., a matrix $\mathbf{M}$ such that $\mathrm{MSEB}\,(\mathbf{M},\boldsymbol{\theta}_0)<\mathrm{Tr}(\mathbf{J}^{-1}(\boldsymbol{\theta}_0))$ for all values of $\boldsymbol{\theta}_0$ in a desired class, with the additional property that no other matrix $\mathbf{M}'\neq\mathbf{M}$

satisfies $\mathrm{MSEB}(\mathbf{M}', \boldsymbol{\theta}_0) \leq \mathrm{MSEB}(\mathbf{M}, \boldsymbol{\theta}_0)$ for all $\boldsymbol{\theta}_0$. Using ideas developed in [12] in the context of estimation in linear models, a minimax convex optimization problem was formulated whose solution $\mathbf{M}$ is both admissible and dominating over a bounded set $\mathcal{U}$. Specifically

$$\widehat{\mathbf{M}} = \arg\min_{\mathbf{M}} \sup_{\boldsymbol{\theta}\in\mathcal{U}} \{\mathrm{MSEB}(\mathbf{M}, \boldsymbol{\theta}) - \mathrm{MSEB}(0, \boldsymbol{\theta})\}. \quad (7)$$

Various techniques were then derived to solve (7).

Here we extend these results to affine bias vectors of the form

$$\mathbf{b}(\boldsymbol{\theta}_0) = \mathbf{M}\boldsymbol{\theta}_0 + \mathbf{u}. \quad (8)$$

Although the ideas in this paper stem from [9] we will see that the mathematics associated with the affine bias leads to some challenges and are not a direct extension of the linear bias case. Furthermore, we demonstrate throughout the paper via several examples that allowing for an affine bias can often improve the performance significantly over a linear choice, and leads to more meaningful bounds.

When the bias has an affine form as in (8), the MSE bound of (5) becomes

$$\begin{aligned}\mathrm{MSEB}(\mathbf{M}, \mathbf{u}, \boldsymbol{\theta}_0) = (\mathbf{M}\boldsymbol{\theta}_0 + \mathbf{u})^*(\mathbf{M}\boldsymbol{\theta}_0 + \mathbf{u}) \\ + \mathrm{Tr}((\mathbf{I}+\mathbf{M})\mathbf{J}^{-1}(\boldsymbol{\theta}_0)(\mathbf{I}+\mathbf{M})^*).\end{aligned} \quad (9)$$

With a slight abuse of notation we refer to this bound as the affine MSE bound. Note that when $\mathbf{M} = 0$, $\mathbf{u} = 0$, we have $\mathrm{MSEB}(0, 0, \boldsymbol{\theta}_0) = \mathrm{Tr}(\mathbf{J}^{-1}(\boldsymbol{\theta}_0))$ which is equal to the CRLB. As in [9], our goal is to find an admissible dominating pair $\mathbf{M}, \mathbf{u}$ such that

$$\begin{aligned}\mathrm{MSEB}(\mathbf{M}, \mathbf{u}, \boldsymbol{\theta}_0) &< \mathrm{MSEB}(0, 0, \boldsymbol{\theta}_0), &&\text{for all } \boldsymbol{\theta}_0 \in \mathcal{U} \\ \mathrm{MSEB}(\mathbf{M}', \mathbf{u}', \boldsymbol{\theta}_0) &\leq \mathrm{MSEB}(\mathbf{M}, \mathbf{u}, \boldsymbol{\theta}_0), &&\text{for all } \boldsymbol{\theta}_0 \in \mathcal{U} \\ &\Rightarrow (\mathbf{M}', \mathbf{u}') = (\mathbf{M}, \mathbf{u}).\end{aligned} \quad (10)$$

If $\mathbf{M}, \mathbf{u}$ satisfy (10) and $\hat{\boldsymbol{\theta}}$ is efficient, then the estimator

$$\hat{\boldsymbol{\theta}}_{\mathrm{b}} = (\mathbf{I} + \mathbf{M})\hat{\boldsymbol{\theta}} + \mathbf{u} \quad (11)$$

achieves the affine MSE bound and will have smaller MSE than $\hat{\boldsymbol{\theta}}$ for all values of $\boldsymbol{\theta}_0 \in \mathcal{U}$. Thus an admissible and dominating $\mathbf{M}, \mathbf{u}$ lead to an improved estimation strategy with lower MSE.

Before we discuss how to find such a pair $\mathbf{M}, \mathbf{u}$ in the next section we provide motivation for affine bias vectors by considering a simple example.

III. Motivation for Affine Bias

Consider the problem of estimating a constant θ from noisy measurements

$$y[n] = \theta + w[n], \quad 1 \leq n \leq N \quad (12)$$

where $w[n]$ are iid, zero-mean Gaussian random variables with variance σ^2. The CRLB for estimating θ is $J^{-1}(\theta) = \sigma^2/N = A$ and is achieved by the empirical average $\overline{\theta} = (1/N)\sum_{n=1}^{N} y[n]$.

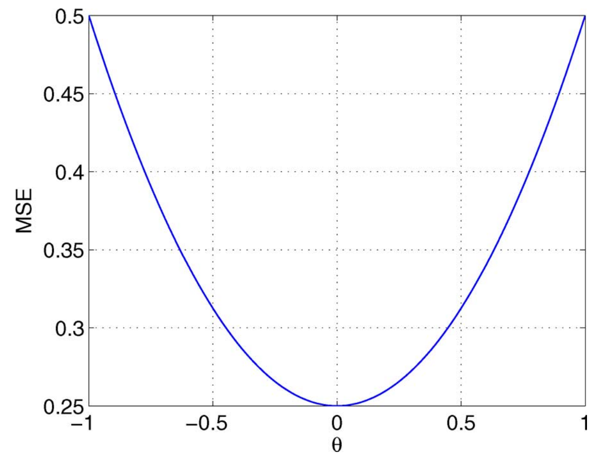

Fig. 1. Minimax MSE bound for estimating θ in the model (12) assuming that $\theta \in [-1, 1]$ and $A = 1$.

Suppose in addition that θ lies in the interval [−1,1]. In order to improve the MSE of $\overline{\theta}$ we multiply it by $1 + M$ where M is the solution to (7). This guarantees that the MSE of $(1 + M)\overline{\theta}$ is smaller than that of $\overline{\theta}$ for all $\theta \in [-1, 1]$. Adapting (7) to our model, M is determined by the minimax problem

$$\min_{M} \max_{\theta\in[-1,1]} \left\{\theta^2 M^2 + \left((1+M)^2 - 1\right) A\right\}. \quad (13)$$

Since the objective is monotonically increasing in $|\theta|$, the maximum is obtained at $\theta = \pm 1$, and (13) reduces to a simple quadratic minimization problem whose solution is

$$M = -\frac{A}{A+1}. \quad (14)$$

The resulting linear MSE bound of (6), which is the MSE of the modified estimator, becomes

$$\mathrm{MSEB}_1 = \frac{A^2\theta^2 + A}{(A+1)^2}. \quad (15)$$

In Fig. 1 we plot MSEB_1 as a function of θ in the interval [−1,1] with $A = 1$. As can be seen from the figure, the linear modification of $\overline{\theta}$ reduces the MSE for all values of θ (the MSE of $\overline{\theta}$ is equal to A, i.e., 1).

Now, suppose instead that θ is restricted to the interval [1,3]. We can express the unknown value of θ as $\theta = \theta' + 2$ where $\theta' \in [-1, 1]$. Defining the modified observations $y'[n] = y[n] - 2$, we have from (12) that

$$y'[n] = \theta' + w[n], \quad 0 \leq n \leq N-1. \quad (16)$$

Evidently, estimating θ from $y[n]$ is equivalent to estimating θ' from $y'[n]$. Consequently, we expect the MSE bound for this setup to have the same form as MSEB_1, shifted by 2. Following the same steps as before we can evaluate the minimax linear M over the interval [1,3] which results in $M = -A/(A+9)$. The corresponding MSE bound is plotted in Fig. 2 as a function of θ for $A = 1$. Comparing Figs. 1 and 2 we see that counter to our intuition, the form of the bounds is very different. As we now show, the reason for this behavior is that we constrain the

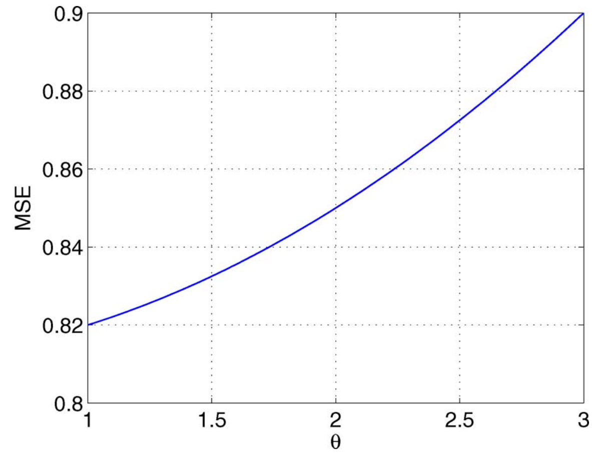

Fig. 2. Minimax MSE bound for estimating θ in the model (12) assuming that $\theta \in [1, 3]$ and $A = 1$.

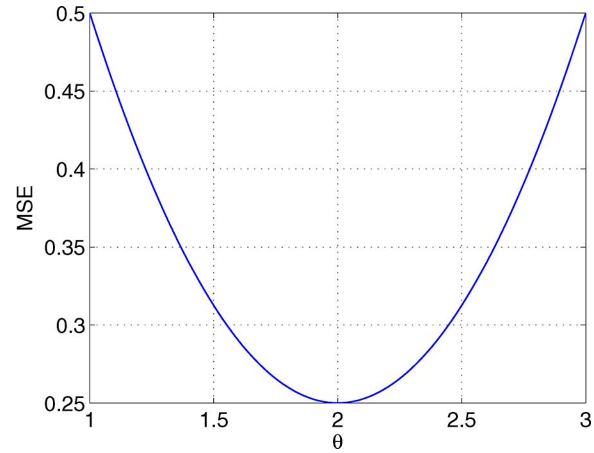

Fig. 3. Affine minimax MSE bound for estimating θ in the model (12) assuming that $\theta \in [1, 3]$ and $A = 1$.

bias to be linear as opposed to affine. Allowing for an affine bias will render the shape of the resulting bounds independent of the specific region chosen.

To see this, suppose we seek M and u that optimize the affine bound (9) in a minimax sense. Specifically, they are chosen as the solution to the problem (20) which we discuss in the next section. From Proposition 4, developed later on in Section VI, it follows that the minimax M and u over the interval $\theta \in [L, U]$ are given by

$$M = -\frac{A}{A + \frac{(U-L)^2}{4}}, \quad u = \frac{A}{A + \frac{(U-L)^2}{4}} \frac{U+L}{2} \tag{17}$$

and the corresponding MSE bound is

$$\mathrm{MSEB}_2 = \frac{A\left(A\left(\theta - \frac{(U+L)}{2}\right)^2 + \frac{(U-L)^4}{16}\right)}{\left(A + \frac{(U-L)^2}{4}\right)^2}. \tag{18}$$

When $L = -1, U = 1$ we have that $u = 0$ so that the optimal M of (17) and the corresponding bound (18) are equivalent to (14) and (15) which are obtained by optimizing only over M. For $L = 1, U = 3$, (17) becomes

$$M = -\frac{A}{A+1}, \quad u = \frac{2A}{A+1} \tag{19}$$

so that M is now the same as in the previous case and only u changes. Therefore, the shape of the MSE bound is the same but it is shifted around 2, as can be seen in Fig. 3.

This simple example demonstrates that allowing for an affine bias is necessary in order to obtain meaningful results. In the next section we show how to find an admissible and dominating pair $\mathbf{M}, \mathbf{u}$ by solving a minimax optimization problem.

IV. Dominating the CRLB With Affine Bias

In analogy to the linear case, it turns out that an admissible dominating pair $\mathbf{M}, \mathbf{u}$ can be found as a solution to a convex optimization problem, as incorporated in the following theorem.

Theorem 1: Let $\mathbf{y}$ *be a random vector with pdf* $p(\mathbf{y}; \boldsymbol{\theta}_0)$*, and let*

$$\begin{aligned}\mathrm{MSEB}\,(\mathbf{M}, \mathbf{u}, \boldsymbol{\theta}_0) = &(\mathbf{M}\boldsymbol{\theta}_0 + \mathbf{u})^*(\mathbf{M}\boldsymbol{\theta}_0 + \mathbf{u}) \\ &+ \mathrm{Tr}\left((\mathbf{I}+\mathbf{M})\mathbf{J}^{-1}(\boldsymbol{\theta}_0)(\mathbf{I}+\mathbf{M})^*\right)\end{aligned}$$

be a bound on the MSE of any estimate $\hat{\boldsymbol{\theta}}$ *of* $\boldsymbol{\theta}_0$ *with affine bias* $\mathbf{b}(\boldsymbol{\theta}_0) = \mathbf{M}\boldsymbol{\theta}_0 + \mathbf{u}$*, where* $\mathbf{J}(\boldsymbol{\theta}_0)$ *is the Fisher information matrix. Define*

$$(\widehat{\mathbf{M}}, \hat{\mathbf{u}}) = \arg\min_{\mathbf{M},\mathbf{u}} \sup_{\boldsymbol{\theta}\in\mathcal{U}} \{\mathrm{MSEB}\,(\mathbf{M}, \mathbf{u}, \boldsymbol{\theta}) - \mathrm{MSEB}\,(0, 0, \boldsymbol{\theta})\} \tag{20}$$

where $\mathcal{U} \subseteq \mathbb{C}^m$*. Then*

1. $\widehat{\mathbf{M}}$ *and* $\hat{\mathbf{u}}$ *are unique;*
2. $\widehat{\mathbf{M}}$ *and* $\hat{\mathbf{u}}$ *are admissible on* $\mathcal{U}$*;*
3. *If* $\widehat{\mathbf{M}} \neq 0$ *or* $\hat{\mathbf{u}} \neq 0$ *, then* $\mathrm{MSEB}\,(\widehat{\mathbf{M}}, \hat{\mathbf{u}}, \boldsymbol{\theta}) < \mathrm{MSEB}\,(0, 0, \boldsymbol{\theta})$ *for all* $\boldsymbol{\theta} \in \mathcal{U}$*.*

Note that the minimum in (20) is well defined since the objective is continuous and coercive [17].

Proof: The proof follows immediately from the proof of [12, Theorem 1] by noting that $\mathrm{MSEB}\,(\mathbf{M}, \mathbf{u}, \boldsymbol{\theta})$ is continuous, coercive and strictly convex in $\mathbf{M}$ and $\mathbf{u}$. □

From Theorem 1 we conclude that if we find a pair $(\widehat{\mathbf{M}}, \hat{\mathbf{u}}) \neq 0$ that is the solution to (20), and if $\hat{\boldsymbol{\theta}}$ achieves the CRLB, then the MSE of

$$\hat{\boldsymbol{\theta}}_{\mathrm{b}} = (\mathbf{I} + \widehat{\mathbf{M}})\hat{\boldsymbol{\theta}} + \hat{\mathbf{u}} \tag{21}$$

is smaller than that of $\hat{\boldsymbol{\theta}}$ for all $\boldsymbol{\theta}_0 \in \mathcal{U}$; furthermore, no other estimator with affine bias exists that has a smaller (or equal) MSE than $\hat{\boldsymbol{\theta}}_{\mathrm{b}}$ for all values of $\boldsymbol{\theta}_0 \in \mathcal{U}$.

For arbitrary forms of $\mathbf{J}^{-1}(\boldsymbol{\theta}_0)$ we can solve (20) by known iterative algorithms for solving minimax problems, such as subgradient algorithms [18] or the prox method [19]. To obtain more efficient solutions, we follow the same path as in [9] and restrict $\mathbf{J}^{-1}(\boldsymbol{\theta}_0)$ to the quadratic form

$$\mathbf{J}^{-1}(\boldsymbol{\theta}) = \sum_{i=1}^{\ell} \mathbf{B}_i\boldsymbol{\theta}\boldsymbol{\theta}^*\mathbf{B}_i^* + \sum_{i=1}^{k}(\mathbf{C}_i\boldsymbol{\theta}\mathbf{z}_i^* + \mathbf{z}_i\boldsymbol{\theta}^*\mathbf{C}_i^*) + \mathbf{A} \tag{22}$$

for some matrices $\mathbf{A} \succeq 0$, $\mathbf{B}_i$, $\mathbf{C}_i$ and vectors $\mathbf{z}_i$ chosen such that $\mathbf{J}^{-1}(\boldsymbol{\theta}) \succ 0$ for all $\boldsymbol{\theta}_0 \in \mathcal{U}$. [Alternatively, when considering MVU estimators, we assume that the minimum variance has the form (22)]. The motivation for studying the class (22) is that there are many cases of interest in which the inverse Fisher information can be written in this form; several examples are presented in [9] and throughout the paper. As we show in the ensuing sections, assuming this structure we can obtain efficient representations of (20) and in many cases even closed form solutions. Specifically, we prove that (20) can be converted into an SDP, which is a broad class of convex problems for which polynomial-time algorithms exist [13], [14]. These are optimization problems that involve minimizing a linear function subject to linear matrix inequalities (LMIs), 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 [20], 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 which often lead to closed from solutions.

In the next section, we demonstrate these ideas for the case in which $\boldsymbol{\theta}_0$ is not restricted. In some settings, we may have additional information on $\boldsymbol{\theta}_0$ which can result in a lower MSE bound. The set $\mathcal{U}$ is then chosen to capture these properties of $\boldsymbol{\theta}_0$. For example, we may know that the norm of $\boldsymbol{\theta}_0$ is bounded: $\boldsymbol{\theta}_0^*\boldsymbol{\theta}_0 \leq U$ for some $U > 0$. There are also examples where there are natural restrictions on the parameters, for example if θ_0 represents the variance or the SNR of a random variable, then $\theta_0 > 0$. More generally, θ_0 may lie in a specified interval $\alpha \leq \theta_0 \leq \beta$. These constraints can all be viewed as special cases of the quadratic restriction $\boldsymbol{\theta}_0 \in \mathcal{Q}$ where

$$\mathcal{Q} = \{\boldsymbol{\theta} | \boldsymbol{\theta}^*\mathbf{A}_1\boldsymbol{\theta} + 2\mathbf{b}_1^*\boldsymbol{\theta} + c_1 \leq 0\} \tag{23}$$

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 (23) is not necessarily convex. In Section VI, we discuss the scenario in which $\boldsymbol{\theta}_0 \in \mathcal{Q}$, and show that again admissible and dominating $\mathbf{M}, \mathbf{u}$ can be found by solving an SDP. Using the results of [21], the ideas we develop can also be generalized to the case of two quadratic constraints of the form $\mathcal{Q}$.

V. Dominating Bound on the Entire Space

We first treat the case in which $\mathcal{U} = \mathbb{C}^m$ so that $\boldsymbol{\theta}_0$ is not restricted. As we will show, if $\mathbf{B}_i \neq 0$ for some i in (22), then a strictly dominating pair $\mathbf{M}, \mathbf{u}$ over the entire space can always be found. This implies that under this condition, the CRLB can always be improved on uniformly.

With $\mathbf{J}^{-1}(\boldsymbol{\theta}_0)$ given by (22), the difference between the MSE bound and the CRLB can be written compactly as $\boldsymbol{\theta}_0^*\mathbf{A}_0(\mathbf{M})\boldsymbol{\theta}_0 + 2\Re\{\mathbf{b}_0^*(\mathbf{M},\mathbf{u})\boldsymbol{\theta}_0\} + c_0(\mathbf{M},\mathbf{u})$, where we defined

$$\begin{aligned}
\mathbf{A}_0(\mathbf{M}) &\triangleq \mathbf{M}^*\mathbf{M} + \sum_{i=1}^{\ell} \mathbf{B}_i^* \left((\mathbf{I}+\mathbf{M})^*(\mathbf{I}+\mathbf{M}) - \mathbf{I}\right)\mathbf{B}_i;\\
\mathbf{b}_0(\mathbf{M},\mathbf{u}) &\triangleq \sum_{i=1}^{k} \mathbf{C}_i^* \left((\mathbf{I}+\mathbf{M})^*(\mathbf{I}+\mathbf{M}) - \mathbf{I}\right)\mathbf{z}_i + \mathbf{M}^*\mathbf{u};\\
c_0(\mathbf{M},\mathbf{u}) &\triangleq \mathrm{Tr}\left(\left((\mathbf{I}+\mathbf{M})^*(\mathbf{I}+\mathbf{M}) - \mathbf{I}\right)\mathbf{A}\right) + \mathbf{u}^*\mathbf{u}.
\end{aligned} \tag{24}$$

From Theorem 1, an admissible dominating matrix $\mathbf{M}$ and vector $\mathbf{u}$ can then be found as the solution to

$$\min_{\mathbf{M},\mathbf{u}} \max_{\boldsymbol{\theta}} \left\{\boldsymbol{\theta}^*\mathbf{A}_0(\mathbf{M})\boldsymbol{\theta} + 2\Re\left\{\mathbf{b}_0^*(\mathbf{M},\mathbf{u})\boldsymbol{\theta}\right\} + c_0(\mathbf{M},\mathbf{u})\right\} \tag{25}$$

or equivalently

$$\min_{t,\mathbf{M},\mathbf{u}} \{t : \boldsymbol{\theta}^*\mathbf{A}_0(\mathbf{M})\boldsymbol{\theta} + 2\Re\{\mathbf{b}_0^*(\mathbf{M},\mathbf{u})\boldsymbol{\theta}\} + c_0(\mathbf{M},\mathbf{u}) \leq t, \forall \boldsymbol{\theta}\}. \tag{26}$$

Problem (26) can be written in matrix form as [22, p. 163]

$$\min_{t,\mathbf{M},\mathbf{u}} \{t : \mathbf{G}(\mathbf{M},\mathbf{u}) \preceq 0\} \tag{27}$$

where

$$\mathbf{G}(\mathbf{M},\mathbf{u}) \triangleq \begin{bmatrix} \mathbf{A}_0(\mathbf{M}) & \mathbf{b}_0(\mathbf{M},\mathbf{u}) \\ \mathbf{b}_0^*(\mathbf{M},\mathbf{u}) & c_0(\mathbf{M},\mathbf{u}) - t \end{bmatrix}. \tag{28}$$

Since the choice of parameters $\mathbf{M} = 0$, $\mathbf{u} = 0$, $t = 0$ satisfies (28), our problem is always feasible.

In our development below, we consider the case in which the constraint (28) is strictly feasible, i.e., there exist $\mathbf{M}$ and $\mathbf{u}$ such that $\mathbf{G}(\mathbf{M},\mathbf{u}) \prec 0$. Conditions for strict feasibility are given in the following lemma.

Lemma 1: [9] The constraint (28) is strictly feasible if and only if $\sum_{i=1}^{\ell} \mathbf{B}_i^*\mathbf{B}_i \succ 0$.

If (28) is not strictly feasible then it is shown in [9, App. II] that 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.

Next, we show that the optimal $\mathbf{M}, \mathbf{u}$ 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 V-C. Finally, in Section V-D we derive a set of necessary and sufficient optimality conditions on $\mathbf{M}$ and $\mathbf{u}$, which are then used, in Section V-E, to develop closed form solutions for some special cases.

A. SDP Formulation of the Problem

In order to apply standard convex algorithms or Lagrange duality theory to find the optimal $\mathbf{M}, \mathbf{u}$ the constraint (28) must be written in convex form. Fortunately, this nonconvex constraint can be converted into a convex one, as incorporated in the following lemma.

Lemma 2: The problem (27) with $\mathbf{G}(\mathbf{M},\mathbf{u})$ *given by (28) is equivalent to the convex problem*

$$\min_{t,\mathbf{W},\mathbf{X}} \{t : \mathbf{X} + \mathbf{Z}(\mathbf{W},\mathbf{X}) \preceq 0,\ \mathbf{W}^*\mathbf{W} \preceq \mathbf{X}\} \tag{29}$$

where $\mathbf{W} = [\mathbf{M}\ \mathbf{u}]$

$$\mathbf{Z}(\mathbf{W},\mathbf{X}) = \begin{bmatrix} \sum_{i=1}^{\ell} \mathbf{B}_i^*\Phi\mathbf{B}_i & \sum_{i=1}^{k} \mathbf{C}_i^*\Phi\mathbf{z}_i \\ \sum_{i=1}^{k} \mathbf{z}_i^*\Phi\mathbf{C}_i & \mathrm{Tr}(\mathbf{A}\Phi) - t \end{bmatrix} \tag{30}$$

and for brevity we denoted $\Phi = \widetilde{\mathbf{I}}^*\mathbf{X}\widetilde{\mathbf{I}} + \mathbf{W}\widetilde{\mathbf{I}} + \widetilde{\mathbf{I}}^*\mathbf{W}^*$ *where* $\widetilde{\mathbf{I}} = \mathrm{diag}_{m+1,m}(1,\ldots,1)$.

Note that $\mathbf{W}\widetilde{\mathbf{I}} = \mathbf{M}$ and $\widetilde{\mathbf{I}}^*\mathbf{X}\widetilde{\mathbf{I}}$ is the upper left $m \times m$ block of the matrix $\mathbf{X}$.

Proof: See Appendix A. □

From Lemma 2 we see that (27) can be written as a convex problem. Moreover, the optimal $\mathbf{M}$ and $\mathbf{u}$ can be found using standard software packages by noting that (29) can be written as an SDP. Indeed, the matrix $\mathbf{Z}(\mathbf{W},\mathbf{X})$ is linear in both $\mathbf{W}$ and $\mathbf{X}$, and the inequality $\mathbf{W}^*\mathbf{W} \preceq \mathbf{X}$ is equivalent to the LMI

$$\begin{bmatrix} \mathbf{X} & \mathbf{W}^* \\ \mathbf{W} & \mathbf{I} \end{bmatrix} \succeq 0. \tag{31}$$

This is a direct consequence of the following result, referred to as Schur's lemma:

Lemma 3: [23, 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* $\mathbf{Y}(\mathbf{I} - \mathbf{X}\mathbf{X}^\dagger) = 0$.

B. Dual Problem

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

Since (29) is convex and strictly feasible, the optimal value of t is equal to the optimal value of the dual problem. The dual is derived in Appendix B in which we show that it is given by

$$\min_{\Pi \succeq \mathbf{w}\mathbf{w}^*} \mathrm{Tr}\left(\mathbf{S}(\Pi,\mathbf{w})\left(\mathbf{S}(\Pi,\mathbf{w}) + \Pi - \mathbf{w}\mathbf{w}^*\right)^{-1}\mathbf{S}(\Pi,\mathbf{w})\right) \tag{32}$$

where we defined

$$\mathbf{S}(\Pi,\mathbf{w}) \triangleq \sum_{i=1}^{\ell} \mathbf{B}_i\Pi\mathbf{B}_i^* + \sum_{i=1}^{k}(\mathbf{z}_i\mathbf{w}^*\mathbf{C}_i^* + \mathbf{C}_i\mathbf{w}\mathbf{z}_i^*) + \mathbf{A}. \tag{33}$$

Using Lemma 3, (32) can be written as an SDP

$$\begin{array}{ll} \min_{\mathbf{Y},\mathbf{w},\Pi} & \mathrm{Tr}\,(\mathbf{Y}) \\ \text{s.t.} & \begin{bmatrix} \mathbf{Y} & \mathbf{S}(\Pi,\mathbf{w}) & 0 \\ \mathbf{S}(\Pi,\mathbf{w}) & \mathbf{S}(\Pi,\mathbf{w}) + \Pi & \mathbf{w} \\ 0 & \mathbf{w}^* & 1 \end{bmatrix} \succeq 0; \\ & \begin{bmatrix} \Pi & \mathbf{w} \\ \mathbf{w} & 1 \end{bmatrix} \succeq 0. \end{array} \tag{34}$$

We also show that the optimal $\mathbf{M}$ and $\mathbf{u}$ are related to the dual solutions by

$$\mathbf{M} = -\mathbf{S}(\Pi,\mathbf{w})\left(\mathbf{S}(\Pi,\mathbf{w}) + \Pi - \mathbf{w}\mathbf{w}^*\right)^{-1} \tag{35}$$

$$\mathbf{u} = \frac{1}{1 - \mathbf{w}^*(\mathbf{S}(\Pi,\mathbf{w}) + \Pi)^{-1}\mathbf{w}}\mathbf{S}(\Pi,\mathbf{w})\left(\mathbf{S}(\Pi,\mathbf{w}) + \Pi\right)^{-1}\mathbf{w}. \tag{36}$$

Note that since $\Pi \succeq \mathbf{w}\mathbf{w}^*$

$$\mathbf{S}(\Pi,\mathbf{w}) + \Pi - \mathbf{w}\mathbf{w}^* \succeq \mathbf{S}(\mathbf{w}\mathbf{w}^*,\mathbf{w}) = \mathbf{J}^{-1}(\mathbf{w}) \succ 0 \tag{37}$$

so that the inverses in (35), (36) are well defined. Furthermore, (37) implies that $\mathbf{S}(\Pi,\mathbf{w}) + \Pi \succ \mathbf{w}\mathbf{w}^*$, or

$$\mathbf{I} \succ (\mathbf{S}(\Pi,\mathbf{w}) + \Pi)^{-1/2}\mathbf{w}\mathbf{w}^*(\mathbf{S}(\Pi,\mathbf{w}) + \Pi)^{-1/2} \tag{38}$$

which is equivalent to $\mathbf{w}^*(\mathbf{S}(\Pi,\mathbf{w}) + \Pi)^{-1}\mathbf{w} < 1$.

An important observation from (35) is that regardless of Π, $\mathbf{M}$ is not equal 0. Furthermore, $\mathbf{u}$ of (36) is 0 only if $\mathbf{w} = 0$. If $\mathbf{z}_i = 0$ or $\mathbf{C}_i = 0$ so that $\mathbf{J}^{-1}$ does not include a linear term, then it is easy to see that $\mathbf{w} = 0$ and consequently $\mathbf{u} = 0$. Indeed, in this case $\mathbf{S}(\Pi,\mathbf{w}) = \mathbf{S}(\Pi)$ is only a function of Π. Therefore, if $\widehat{\mathbf{Y}}, \widehat{\Pi}, \widehat{\mathbf{w}}$ are an optimal solution to the dual problem (34), then $\widehat{\mathbf{Y}}, \widehat{\Pi}, \mathbf{w} = 0$ also satisfy the constraints of (34) and are therefore also optimal. Since the solution is unique we conclude that $\mathbf{w} = 0$ is optimal.

C. Necessary and Sufficient Optimality Conditions

To complete our description of the optimal $\mathbf{M}$ and $\mathbf{u}$, we now use the Karush-Kuhn-Tucker (KKT) theory [17], [22] to develop necessary and sufficient optimality conditions.

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

1. $d\mathcal{L}/d\mathbf{X} = 0$, $d\mathcal{L}/d\mathbf{W} = 0$ and $d\mathcal{L}/dt = 0$ where $\mathcal{L}$ is the Lagrangian defined by (80);
2. Feasibility: $\mathbf{X} + \mathbf{Z}(\mathbf{W},\mathbf{X}) \preceq 0$ where $\mathbf{Z}(\mathbf{W},\mathbf{X})$ is defined by (30);
3. Complementary slackness: $\mathrm{Tr}(\widetilde{\Pi}(\mathbf{X} + \mathbf{Z}(\mathbf{W},\mathbf{X}))) = 0$ and $\mathrm{Tr}\,(\Delta(\mathbf{W}^*\mathbf{W} - \mathbf{X})) = 0$.

From the derivations in Appendix B it follows that the first condition results in $\pi = 1$, $\mathbf{M}, \mathbf{u}$ given by (35) and (36), and

$$\Delta = \Pi + \widetilde{\mathbf{I}}\mathbf{S}(\Pi,\mathbf{w})\widetilde{\mathbf{I}}^* = \begin{bmatrix} \Pi + \mathbf{S}(\Pi,\mathbf{w}) & \mathbf{w} \\ \mathbf{w}^* & 1 \end{bmatrix}. \tag{39}$$

We also showed that $\Pi + \mathbf{S}(\Pi,\mathbf{w}) \succ 0$ and $\mathbf{w}^*(\Pi + \mathbf{S}(\Pi,\mathbf{w}))^{-1}\mathbf{w} < 1$ which from Lemma 3 leads to $\Delta \succ 0$. The second complementary slackness condition then becomes $\mathbf{X} = \mathbf{W}^*\mathbf{W}$. Using the fact that

$$\mathrm{Tr}(\widetilde{\Pi}\mathbf{Z}) = 2\mathrm{Tr}\left(\mathbf{W}\widetilde{\mathbf{I}}\mathbf{S}(\Pi,\mathbf{w})\right) + \mathrm{Tr}\left(\widetilde{\mathbf{I}}^*\mathbf{X}\widetilde{\mathbf{I}}\mathbf{S}(\Pi,\mathbf{w})\right) - t \tag{40}$$

and $\mathbf{W}\widetilde{\Pi} = -\mathbf{S}(\Pi,\mathbf{w})\widetilde{\mathbf{I}} - \mathbf{M}\mathbf{S}(\Pi,\mathbf{w})\widetilde{\mathbf{I}}^*$ where we substituted $\mathbf{W}\widetilde{\mathbf{I}} = \mathbf{M}$, the first complementary slackness condition reduces to

$$\mathrm{Tr}(\widetilde{\Pi}(\mathbf{X} + \mathbf{Z}(\mathbf{M},\mathbf{X}))) = \mathrm{Tr}(\mathbf{M}\mathbf{S}(\Pi,\mathbf{w})) - t = 0. \tag{41}$$

Thus, the matrix $\mathbf{M}$ and the vector $\mathbf{u}$ are 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:

$$\begin{gathered} \mathbf{M} = -\mathbf{S}(\Pi,\mathbf{w})\left(\mathbf{S}(\Pi,\mathbf{w}) + \Pi - \mathbf{w}\mathbf{w}^*\right)^{-1} \\ \mathbf{u} = \mathbf{S}(\Pi,\mathbf{w})\left(\mathbf{S}(\Pi,\mathbf{w}) + \Pi - \mathbf{w}\mathbf{w}^*\right)^{-1}\mathbf{w} \\ \begin{bmatrix} \mathbf{A}_0(\mathbf{M}) & \mathbf{b}_0(\mathbf{M},\mathbf{u}) \\ \mathbf{b}_0^*(\mathbf{M},\mathbf{u}) & c_0(\mathbf{M},\mathbf{u}) - \mathrm{Tr}\,(\mathbf{M}\mathbf{S}(\Pi,\mathbf{w})) \end{bmatrix} \preceq 0 \end{gathered} \tag{42}$$

were $\mathbf{A}_0(\mathbf{M}), \mathbf{b}_0(\mathbf{M}, \mathbf{u}), c_0(\mathbf{M}, \mathbf{u})$ are defined by (24), and $\mathbf{S}(\Pi, \mathbf{w})$ is given by (33).

D. Special Cases

Using conditions (42) we can derive explicit expressions for the optimal $\mathbf{M}$ and $\mathbf{u}$ in some special cases.

We first assume that $\boldsymbol{\theta} = \theta$ is a scalar with inverse Fisher information

$$J^{-1}(\theta) = b^2\theta^2 + 2c\theta + a \tag{43}$$

where b, c are real constants and $a > 0$. In this case the dual program can be solved directly. Instead of presenting the detailed derivation of the dual solution, we prove optimality by showing that the proposed solution satisfies the necessary and sufficient conditions of (42).

Proposition 1: Let $\mathbf{y}$ be a random vector with pdf $p(\mathbf{y}; \theta_0)$ such that the Fisher information with respect to θ_0 has the form (43). Then the minimax M and u that are the solution to (20) with $\mathcal{U} = \mathbb{R}$ are given by

$$M = \begin{cases} -\frac{2b^2}{1+b^2}, & |b| < 1 \\ -1, & |b| \geq 1, \end{cases} \qquad u = \begin{cases} -\frac{2c}{1+b^2}, & |b| < 1 \\ -\frac{c}{b^2}, & |b| \geq 1. \end{cases} \tag{44}$$

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

$$\hat{\theta}_{\mathrm{b}} = \begin{cases} \frac{1-b^2}{1+b^2}\hat{\theta} - \frac{2c}{1+b^2} & |b| < 1 \\ -\frac{c}{b^2}, & |b| \geq 1 \end{cases}$$

has smaller MSE than $\hat{\theta}$ for all θ_0.

Proof: If $|b| \geq 1$, then it is easy to see that the optimality conditions are satisfied with $\pi = \omega^2$ and $\omega = -c/b^2$. If, on the other hand, $|b| < 1$, then the optimality conditions are satisfied with

$$\pi = \frac{2c^2 + (ab^2 - 2c^2)(1 - b^2)}{b^4(1 + b^2)}, \quad \omega = -\frac{c}{b^2}. \tag{45}$$

To see this, note that with this choice

$$S(\pi, w) = \frac{2(ab^2 - c^2)}{b^2(1 + b^2)}, \quad \pi - \omega^2 = \frac{(1 - b^2)(ab^2 - c^2)}{b^4(1 + b^2)} \tag{46}$$

completing the proof. □

We next treat the case in which the optimal solution is $\mathbf{M} = -\mathbf{I}$ which results in a constant estimator: $\hat{\boldsymbol{\theta}}_{\mathrm{b}} = \mathbf{u}$. The conditions for this solution are stated in the following proposition.

Proposition 2: Let $\mathbf{y}$ be a random vector with pdf $p(\mathbf{y}; \boldsymbol{\theta}_0)$ such that the Fisher information with respect to $\boldsymbol{\theta}_0$ has the form (22). Then $\mathbf{M} = -\mathbf{I}$ is the solution to (20) with $\mathcal{U} = \mathbb{C}^m$ if and only if

$$\sum_{i=1}^{\ell} \mathbf{B}_i^* \mathbf{B}_i \succeq \mathbf{I}. \tag{47}$$

In this case, the optimal $\mathbf{u}$ is $\mathbf{u} = -(\sum_{i=1}^{\ell} \mathbf{B}_i^* \mathbf{B}_i)^{-1} \sum_{i=1}^{k} \mathbf{C}_i^* \mathbf{z}_i$. Furthermore, if there exists an efficient estimator $\hat{\boldsymbol{\theta}}$, then the estimator $\hat{\boldsymbol{\theta}}_{\mathrm{b}} = \mathbf{u}$ has smaller MSE for all $\boldsymbol{\theta}_0$.

Proof: See Appendix C. □

VI. Dominating Bound on a Quadratic Set

We now treat the case in which the parameter vector $\boldsymbol{\theta}$ is restricted to the quadratic set $\mathcal{Q}$ of (23). To find an admissible dominating matrix in this case we need to solve

$$\min_{\mathbf{M}, \mathbf{u}} \max_{\boldsymbol{\theta} \in \mathcal{Q}} \{\mathrm{MSEB}\,(\mathbf{M}, \mathbf{u}, \boldsymbol{\theta}) - \mathrm{MSEB}\,(0, 0, \boldsymbol{\theta})\}. \tag{48}$$

We assume that the set $\mathcal{Q}$ is not empty, and that there exists an $\boldsymbol{\theta}$ 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 (48) which, omitting the dependence on $\mathbf{M}$ and $\mathbf{u}$, has the form

$$\max_{\boldsymbol{\theta}} \{\boldsymbol{\theta}^* \mathbf{A}_0 \boldsymbol{\theta} + 2\Re\{\mathbf{b}_0^* \boldsymbol{\theta}\} + c_0 : \boldsymbol{\theta}^* \mathbf{A}_1 \boldsymbol{\theta} + 2\mathbf{b}_1^* \boldsymbol{\theta} + c_1 \leq 0\}. \tag{49}$$

The problem of (49) is a *trust region problem*, for which strong duality holds (assuming that there is a strictly feasible point) [24]. Thus, it is equivalent to

$$\begin{aligned} &\min_{\lambda \geq 0, t, \mathbf{M}, \mathbf{u}} \quad t \\ &\text{s.t.} \quad \begin{bmatrix} \lambda \mathbf{A}_1 & \lambda \mathbf{b}_1 \\ \lambda \mathbf{b}_1^* & \lambda c_1 + t \end{bmatrix} \succeq \begin{bmatrix} \mathbf{A}_0(\mathbf{M}) & \mathbf{b}_0(\mathbf{M}, \mathbf{u}) \\ \mathbf{b}_0^*(\mathbf{M}, \mathbf{u}) & c_0(\mathbf{M}, \mathbf{u}) \end{bmatrix}. \end{aligned} \tag{50}$$

It is easy to see that (50) is always feasible, since both matrices in (50) can be made equal to 0 by choosing $\mathbf{M} = 0$, $\mathbf{u} = 0$, and $\lambda = t = 0$. A necessary and sufficient condition for strict feasibility is given in the following proposition.

Proposition 3: The constraint in (50) is strictly feasible if and only if

$$\sum_{i=1}^{\ell} \mathbf{B}_i^* \mathbf{B}_i \mathbf{v} = 0, \ \mathbf{v} \neq 0 \quad \Rightarrow \quad \mathbf{v}^* \mathbf{A}_1 \mathbf{v} > 0. \tag{51}$$

In particular, (50) is strictly feasible if $\sum_{i=1}^{\ell} \mathbf{B}_i^* \mathbf{B}_i \succ 0$ or $\mathbf{A}_1 \succ 0$.

Proof: See Appendix D. □

We assume in the remainder of this section that (50) is strictly feasible.

Our minimization problem is very similar to that of (29). Indeed, it can be written compactly as

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

where $\mathbf{G}(\mathbf{M}, \mathbf{u})$ is defined in (28) and

$$\mathbf{F} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{b}_1 \\ \mathbf{b}_1^* & c_1 \end{bmatrix}. \tag{53}$$

Therefore, the development 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 4: The problem (52) with $\mathbf{G}(\mathbf{M}, \mathbf{u})$ and $\mathbf{F}$ given by (28) and (53) respectively, is equivalent to the convex problem

$$\min_{t, \lambda \geq 0, \mathbf{W}, \mathbf{X}} \{t : \mathbf{X} + \mathbf{Z}(\mathbf{M}, \mathbf{X}) \preceq \lambda \mathbf{F}, \ \mathbf{W}^* \mathbf{W} \preceq \mathbf{X}\} \tag{54}$$

where $\mathbf{Z}(\mathbf{W}, \mathbf{X})$ is defined in (30).

A. Dual Problem

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

The Lagrangian associated with (54) is

$$\mathcal{L} = t + \mathrm{Tr}(\widetilde{\Pi}(\mathbf{X} + \mathbf{Z}(\mathbf{M}, \mathbf{X}) - \lambda \mathbf{F})) + \mathrm{Tr}\left(\Delta(\mathbf{W}^*\mathbf{W} - \mathbf{X})\right) \tag{55}$$

where $\Delta \succeq 0$ and $\widetilde{\Pi}$ is defined by (81). Since $\lambda \geq 0$, the minimum of the Lagrangian is finite only if

$$\mathrm{Tr}(\widetilde{\Pi}\mathbf{F}) = \mathrm{Tr}(\Pi \mathbf{A}_1) + 2\Re\{\mathbf{w}^*\mathbf{b}_1\} + c_1 \leq 0. \tag{56}$$

The optimal value is then obtained at $\lambda = 0$, and the Lagrangian becomes the same as that associated with the unconstrained problem (29).

We conclude that the optimal $\mathbf{M}$ and $\mathbf{u}$ are given by (35) and (36), where Π and $\mathbf{w}$ are the solution to

$$\begin{aligned} \min_{\mathbf{Y},\mathbf{w},\Pi} \; & \mathrm{Tr}\left(\mathbf{Y}\right) \\ \text{s.t.} \; & \begin{bmatrix} \mathbf{Y}\mathbf{S}(\Pi, \mathbf{w}) & 0 & \\ \mathbf{S}(\Pi, \mathbf{w}) & \mathbf{S}(\Pi, \mathbf{w}) + \Pi & \mathbf{w} \\ 0 & \mathbf{w}^* & 1 \end{bmatrix} \succeq 0 \\ & \begin{bmatrix} \Pi & \mathbf{w} \\ \mathbf{w} & 1 \end{bmatrix} \succeq 0 \\ & \mathrm{Tr}(\Pi \mathbf{A}_1) + 2\Re\{\mathbf{w}^*\mathbf{b}_1\} + c_1 \leq 0. \end{aligned} \tag{57}$$

B. Necessary and Sufficient Optimality Conditions

Following the same steps as in Section VI-C we can show, using the KKT conditions, that $\mathbf{M}$ and $\mathbf{u}$ are 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:

$$\begin{aligned} &\mathbf{M} = -\mathbf{S}(\Pi, \mathbf{w})\left(\mathbf{S}(\Pi, \mathbf{w}) + \Pi - \mathbf{w}\mathbf{w}^*\right)^{-1} \\ &\mathbf{u} = \mathbf{S}(\Pi, \mathbf{w})\left(\mathbf{S}(\Pi, \mathbf{w}) + \Pi - \mathbf{w}\mathbf{w}^*\right)^{-1}\mathbf{w} \\ &\mathrm{Tr}(\Pi \mathbf{A}_1) + 2\Re\{\mathbf{w}^*\mathbf{b}_1\} + c_1 \leq 0 \\ &\lambda\left(\mathrm{Tr}(\Pi \mathbf{A}_1) + 2\Re\{\mathbf{w}^*\mathbf{b}_1\} + c_1\right) = 0 \\ &\begin{bmatrix} \mathbf{A}_0(\mathbf{M}) & \mathbf{b}_0(\mathbf{M}, \mathbf{u}) \\ \mathbf{b}_0^*(\mathbf{M}, \mathbf{u}) & c_0(\mathbf{M}, \mathbf{u}) - \mathrm{Tr}\left(\mathbf{M}\mathbf{S}(\Pi, \mathbf{w})\right) \end{bmatrix} \preceq \lambda \begin{bmatrix} \mathbf{A}_1 & \mathbf{b}_1 \\ \mathbf{b}_1^* & c_1 \end{bmatrix} \end{aligned} \tag{58}$$

were $\mathbf{A}_0(\mathbf{M}), \mathbf{b}_0(\mathbf{M}, \mathbf{u}), c_0(\mathbf{M}, \mathbf{u})$ are defined by (24), and $\mathbf{S}(\Pi, \mathbf{w})$ is given by (33).

The optimality conditions can be used to verify that a solution is optimal, as illustrated in the following proposition for $\mathbf{J}^{-1}(\boldsymbol{\theta}) = \mathbf{A}$.

Proposition 4: Let $\mathbf{y}$ *be a random vector with pdf* $p(\mathbf{y}; \boldsymbol{\theta}_0)$. *Assume that the Fisher information with respect to* $\boldsymbol{\theta}_0$ *has the form* $\mathbf{J}(\boldsymbol{\theta}_0) = \mathbf{A}^{-1}$, *and that* $\boldsymbol{\theta}_0 \in \mathcal{Q}$ *with*

$$\mathcal{Q} = \left\{\boldsymbol{\theta}_0 : \boldsymbol{\theta}_0^*\boldsymbol{\theta}_0 + 2\Re\{\mathbf{b}_1^*\boldsymbol{\theta}_0\} + c_1 \leq 0\right\}. \tag{59}$$

Then the optimal $\mathbf{M}$ *and* $\mathbf{u}$ *that are the solution to (48) are*

$$\mathbf{M} = -\frac{\mathrm{Tr}(\mathbf{A})}{\mathrm{Tr}(\mathbf{A}) + \mathbf{b}_1^*\mathbf{b}_1 - c_1}\mathbf{I}, \; \mathbf{u} = -\frac{\mathrm{Tr}(\mathbf{A})}{\mathrm{Tr}(\mathbf{A}) + \mathbf{b}_1^*\mathbf{b}_1 - c_1}\mathbf{b}_1. \tag{60}$$

The corresponding affine MSE bound is

$$\begin{aligned} &E\left\{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2\right\} \\ &\geq \frac{\mathrm{Tr}(\mathbf{A})}{(\mathrm{Tr}(\mathbf{A}) + \mathbf{b}_1^*\mathbf{b}_1 - c_1)^2}\left(\mathrm{Tr}(\mathbf{A})\|\boldsymbol{\theta}_0 + \mathbf{b}_1\|^2 + (\mathbf{b}_1^*\mathbf{b}_1 - c)^2\right). \end{aligned} \tag{61}$$

Furthermore, if there exists an efficient estimator $\hat{\boldsymbol{\theta}}$, *then*

$$\hat{\boldsymbol{\theta}}_{\mathrm{b}} = \frac{1}{\mathrm{Tr}(\mathbf{A}) + \mathbf{b}_1^*\mathbf{b}_1 - c_1}\left((\mathbf{b}_1^*\mathbf{b}_1 - c_1)\hat{\boldsymbol{\theta}} - \mathrm{Tr}(\mathbf{A})\mathbf{b}_1\right)$$

achieves the bound (61), and has smaller MSE than $\hat{\boldsymbol{\theta}}$ *for all* $\boldsymbol{\theta}_0 \in \mathcal{Q}$ *with* $\mathcal{Q}$ *given by (59).*

Proof: The proof follows from noting that since $\ell = k = 0$, the choice

$$\begin{aligned} \Pi &= \frac{\mathbf{b}_1^*\mathbf{b}_1 - c_1}{\mathrm{Tr}(\mathbf{A})}\mathbf{A} + \mathbf{b}_1\mathbf{b}_1^*, \quad \mathbf{w} = -\mathbf{b}_1 \\ \lambda &= \frac{\mathrm{Tr}^2(\mathbf{A})}{\left(\mathrm{Tr}(\mathbf{A}) + \mathbf{b}_1^*\mathbf{b}_1 - c_1\right)^2} \end{aligned} \tag{62}$$

satisfies (58). □

Closed form expressions for $\hat{\boldsymbol{\theta}}_{\mathrm{b}}$ when $\mathbf{J}(\boldsymbol{\theta}_0) = \mathbf{A}^{-1}$ can also be obtained in the case when $\mathbf{A}_1 = \mathbf{T}$ for certain choices of $\mathbf{T} \succ 0$ using similar techniques as those used in [12].

It is interesting to consider the relative improvement over the CRLB afforded by using the optimal affine bias. Denoting by $r(\mathbf{A}, \boldsymbol{\theta}_0)$ the ratio between the affine bound (61) and the CRLB (which is equal to $\mathrm{Tr}(\mathbf{A})$) we have

$$r(\mathbf{A}, \boldsymbol{\theta}_0) = \frac{\mathrm{Tr}(\mathbf{A})\|\boldsymbol{\theta}_0 + \mathbf{b}_1\|^2 + (\mathbf{b}_1^*\mathbf{b}_1 - c)^2}{(\mathrm{Tr}(\mathbf{A}) + \mathbf{b}_1^*\mathbf{b}_1 - c_1)^2}. \tag{63}$$

It is easy to see that the derivative of $r(\mathbf{A}, \boldsymbol{\theta}_0)$ with respect to $\mathrm{Tr}(\mathbf{A})$ is negative, as long as

$$\|\boldsymbol{\theta}_0 + \mathbf{b}_1\|^2 < 2(\mathbf{b}_1^*\mathbf{b}_1 - c_1). \tag{64}$$

Since $\boldsymbol{\theta}_0 \in \mathcal{Q}$ with $\mathcal{Q}$ defined by (59) we have that $\|\boldsymbol{\theta}_0 + \mathbf{b}_1\|^2 \leq \mathbf{b}_1^*\mathbf{b}_1 - c_1$ and therefore (64) is satisfied as long as $\mathbf{b}_1^*\mathbf{b}_1 - c_1 > 0$, or equivalently, as long as there is more than one possible value of $\boldsymbol{\theta}_0$, which is our standing assumption. Thus, $r(\mathbf{A}, \boldsymbol{\theta}_0)$ is monotonically decreasing in $\mathrm{Tr}(\mathbf{A})$ and consequently the relative improvement is more pronounced when the CRLB is large. This makes intuitive sense: When the estimation problem is difficult (such as small sample size, low SNR), we can benefit from biased methods.

We now discuss some special cases of the set $\mathcal{Q}$ defined by (59). Suppose first that $\|\boldsymbol{\theta} - \mathbf{v}\|^2 \leq c$ for some vector $\mathbf{v}$ and $c > 0$. From (60) we have that

$$\mathbf{M} = -\frac{\mathrm{Tr}(\mathbf{A})}{\mathrm{Tr}(\mathbf{A}) + c}\mathbf{I}, \quad \mathbf{u} = \frac{\mathrm{Tr}(\mathbf{A})}{\mathrm{Tr}(\mathbf{A}) + c}\mathbf{v}. \tag{65}$$

The minimax linear choice of $\mathbf{M}$ for this setting with $\mathbf{v} = 0$ was derived in [9]. The resulting optimal value of $\mathbf{M}$ coincides with that given by (65). Therefore, the effect of shifting the center of the set is to shift the estimator in the direction of the center, with magnitude that takes into account both the set (via c) and the Fisher information (via $\mathbf{A}$).

Another interesting case is when

$$\mathcal{Q} = \{\boldsymbol{\theta} : (\boldsymbol{\theta} - \mathbf{v}_1)^*(\boldsymbol{\theta} - \mathbf{v}_2) \le 0\} \tag{66}$$

where $\mathbf{v}_1$ and $\mathbf{v}_2$ are arbitrary vectors such that $\mathbf{v}_1^*\mathbf{v}_2$ is real. This choice of $\mathcal{Q}$ arises for example when we have an interval constraint on a scalar parameter $L \le x \le U$, as in the example in Section III, which corresponds to (66) with $\mathbf{v}_1 = L$ and $\mathbf{v}_2 = U$. In this case $\mathbf{b}_1 = -(\mathbf{v}_1+\mathbf{v}_2)/2$ and $c_1 = \mathbf{v}_1^*\mathbf{v}_2$. Using the fact that $\mathbf{b}_1^*\mathbf{b}_1 - c_1 = \|\mathbf{v}_1 - \mathbf{v}_2\|^2/4$, the optimal choice of $\mathbf{M}$ depends only on the norm of the difference $\|\mathbf{v}_1 - \mathbf{v}_2\|$. Furthermore, the MSE bound can be written as

$$\frac{\mathrm{Tr}(\mathbf{A})}{\left(\mathrm{Tr}(\mathbf{A}) + \frac{\|\mathbf{v}_1-\mathbf{v}_2\|^2}{4}\right)^2} \cdot \left(\mathrm{Tr}(\mathbf{A}) \left\|\boldsymbol{\theta}_0 + \frac{(\mathbf{v}_1+\mathbf{v}_2)}{2}\right\|^2 + \frac{\|\mathbf{v}_1-\mathbf{v}_2\|^2}{4}\right). \tag{67}$$

Therefore, for fixed $\mathbf{v}_1 - \mathbf{v}_2$ the form of the MSE bound is the same; the only effect is a shift towards the mean value $(\mathbf{v}_1 + \mathbf{v}_2)/2$.

C. Minimax MSE Estimation in the Linear Gaussian Model

A special case of Proposition 4 is the linear Gaussian model in which

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta}_0 + \mathbf{w} \tag{68}$$

where $\mathbf{H}$ is a known full-rank matrix, and $\mathbf{w}$ is a zero-mean Gaussian random vector with covariance $\mathbf{C} \succ 0$. The Fisher information in this setting is independent of $\boldsymbol{\theta}_0$ and given by $\mathbf{J}^{-1}(\boldsymbol{\theta}_0) = (\mathbf{H}^*\mathbf{C}^{-1}\mathbf{H})^{-1}$. For this model, the ML solution is the well-known least-squares estimator

$$\hat{\boldsymbol{\theta}}_{\mathrm{LS}} = (\mathbf{H}^*\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}^{-1}\mathbf{y}. \tag{69}$$

Since $\hat{\boldsymbol{\theta}}_{\mathrm{LS}}$ is efficient, Proposition 4 implies that $\hat{\boldsymbol{\theta}}_{\mathrm{b}}$ will dominate the least-squares approach for all feasible parameter vectors. This result has been proven for arbitrary closed feasible sets in [25].

Noting that $\mathbf{J}^{-1}(\hat{\boldsymbol{\theta}}_0)$ is independent of $\boldsymbol{\theta}_0$, it is easy to show that the minimax problem (48) is equivalent to minimizing the worst-case MSE of an affine estimator $\hat{\boldsymbol{\theta}} = \mathbf{G}\mathbf{y} + \mathbf{u}$ where $\mathbf{M}$ and $\mathbf{G}$ are related through $\mathbf{I} - \mathbf{G}\mathbf{H} = \mathbf{M}$. The minimax MSE estimator for this model in the special case of $\mathbf{b}_1 = 0$ has been treated in [26] and [12]. Earlier results for $\mathbf{H} = \mathbf{I}$ and white noise can be found in [27]; minimax MSE estimation with a rank-one weighting was discussed in [28]. Using Proposition 4 the minimax MSE strategy can be extended to arbitrary quadratic constraint sets.

As an example, suppose we know that $\|\boldsymbol{\theta} - \mathbf{v}\|^2 \le c$ for some $\mathbf{v}$ and $c > 0$. From Proposition 4 the minimax MSE estimator under this constraint is

$$\hat{\boldsymbol{\theta}} = \frac{c}{c + \mathrm{Tr}((\mathbf{H}^*\mathbf{C}^{-1}\mathbf{H})^{-1})}\hat{\boldsymbol{\theta}}_{\mathrm{LS}} + \frac{\mathrm{Tr}((\mathbf{H}^*\mathbf{C}^{-1}\mathbf{H})^{-1})}{c + \mathrm{Tr}((\mathbf{H}^*\mathbf{C}^{-1}\mathbf{H})^{-1})}\mathbf{v}. \tag{70}$$

For $\mathbf{v} = 0$, (70) is a simple shrinkage estimate [29]. In Fig. 4 we compare the MSE of the affine modification (70) which is given by the corresponding affine MSE bound, with the MSE of the minimax linear estimator. Note that the resulting transformations $\mathbf{M}$ are different in both cases. We also plot the CRLB which is the MSE of the least-squares method. We assume that $\mathbf{C} = \sigma^2\mathbf{I}$ where σ^2 is varied to achieve the desired SNR, defined by

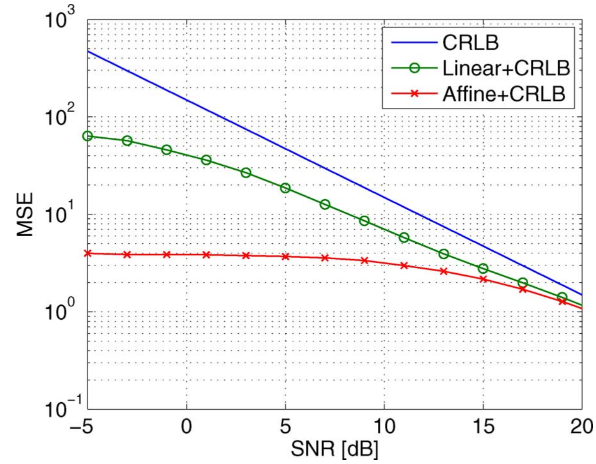

Fig. 4. MSE in estimating $\boldsymbol{\theta}$ in a linear Gaussian model as a function of the SNR using the least-squares, linear modification and affine modifications of the least-squares estimator.

$$\mathrm{SNR}\ [\mathrm{dB}] = 10\log\frac{\|\boldsymbol{\theta}\|^2}{\sigma^2}. \tag{71}$$

In this particular example, we chose $n = 5$, $m = 4$, $\mathbf{v} = [1,\dots,1]^T$, $c = 4$, $\boldsymbol{\theta} = 2\mathbf{v}$ and $\mathbf{H}^*\mathbf{H}$ was generated randomly. As can be seen from the figure, allowing for an affine transformation improves the performance significantly. It is also apparent that as σ^2 increases, the relative improvement in performance is more pronounced. This follows from our general analysis in which we have shown that the relative advantage increases when the CRLB is large.

VII. Example

Up until this point we have shown *analytically* that the CRLB can be uniformly improved upon using an affine bias. We also discussed how to construct an estimator whose MSE is uniformly lower than a given efficient method. Here we demonstrate that these results can be used in practical settings even when an efficient approach is unknown. Specifically, we propose an affine modification of the ML estimator regardless of whether the ML strategy is efficient. Furthermore, we illustrate the possible performance advantage when considering an affine modification in contrast to a linear choice. To this end, we consider the same example that was introduced in [9].

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

$$y_i = \mu + w_i, \quad 1 \le i \le N \tag{72}$$

where w_i is a zero-mean Gaussian random variable with variance σ^2, and the SNR is defined by $\theta = \mu^2/\sigma^2$. In addition, suppose the SNR satisfies $\alpha \le \theta \le \beta$ for some values of α and β. The ML solution is

$$\hat{\theta}_c = \begin{cases} \hat{\theta}, & \alpha \le \frac{\hat{\mu}^2}{\hat{\sigma}^2} \le \beta\,; \\ \alpha, & \frac{\hat{\mu}^2}{\hat{\sigma}^2} \le \alpha\,; \\ \beta, & \frac{\hat{\mu}^2}{\hat{\sigma}^2} \ge \beta \end{cases} \tag{73}$$

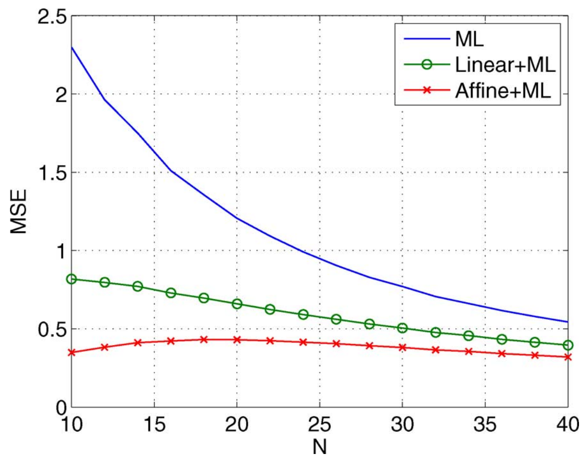

Fig. 5. MSE in estimating the SNR as a function of the number of observations N for an SNR of 2 using the ML, linearly transformed ML and affine transformed ML estimators subject to the constraint (76).

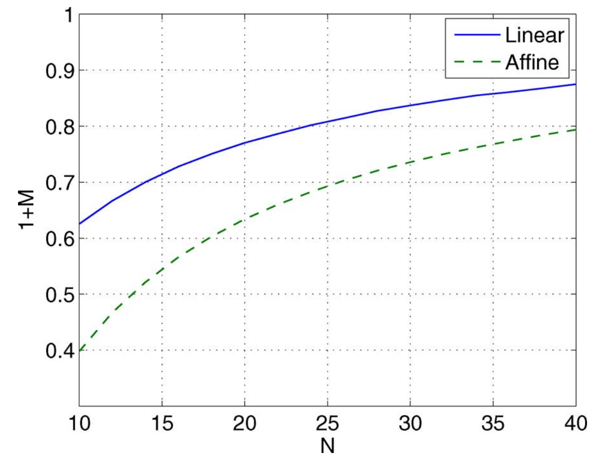

Fig. 6. $1+M$ as a function of the number of observations N when estimating an SNR of 2 using the linear and affine modifications subject to the constraint (76).

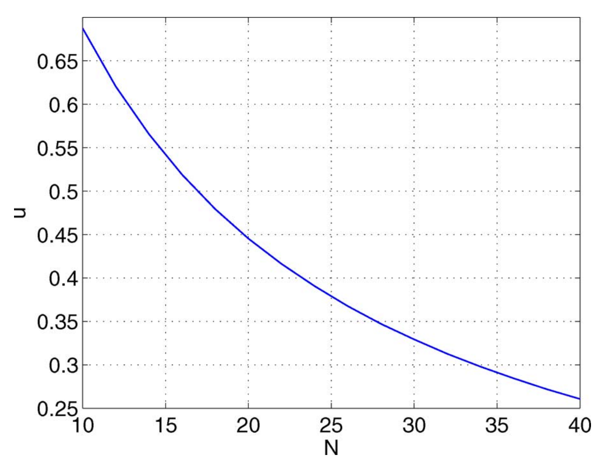

Fig. 7. u as a function of the number of observations N when estimating an SNR of 2 subject to the constraint (76).

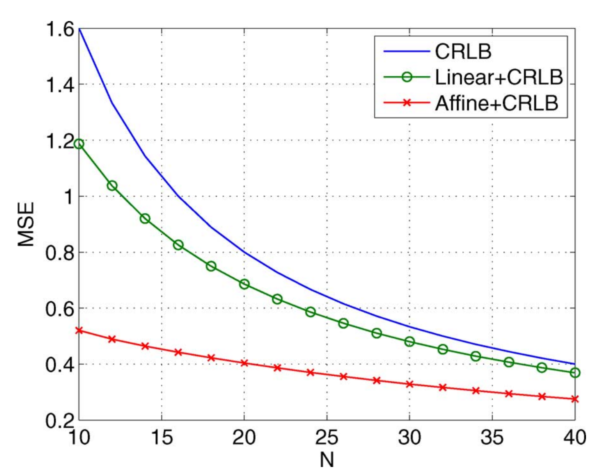

Fig. 8. MSE bound in estimating the SNR as a function of the number of observations N for an SNR of 2: CRLB, linear minimax and affine minimax.

where

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} y_i, \quad \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{\mu})^2. \tag{74}$$

In general $\hat{\theta}$ is biased and does not achieve the CRLB which is given by

$$J^{-1}(\theta) = \frac{1}{N}(4\theta + 2\theta^2). \tag{75}$$

To develop an affine modification of ML we note that the constraint $\alpha \leq \theta \leq \beta$ can be written as

$$(\theta - \alpha)(\theta - \beta) = \theta^2 - (\alpha + \beta)\theta + \alpha\beta \leq 0. \tag{76}$$

Since the constraint is quadratic, the optimal M and u can be found using the SDP formulation of Section VI.

In Fig. 5, we compare the MSE of the ML, the linear ML and affine ML estimators subject to (76), for an SNR of $\theta = 2$ 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 affine modification of the ML estimator performs significantly better than the ML approach and also better than the linearly transformed ML method. It is also interesting to note that in this particular example, the affine modification always lies in the interval $[\alpha, \beta]$. This was not the case for the linear correction which often resulted in an estimate outside this region. When this happens, in principle, we can always project the estimated value onto the given interval to further reduce the MSE. Here, however, projecting the linearly modified ML solution has only a minor impact on the MSE performance.

In Figs. 6 and 7, we plot the values of $1+M$ and u as a function of N. Note that the values of M are different in the linear and affine strategies. In Fig. 8 we plot the CRLB, the linear MSE and affine MSE bounds as a function of N. As we expect, the affine bound is much lower than the other two. Note, however, that in the presence of constraints, the ML estimate is typically biased so that the CRLB does not necessarily bound its performance. Nonetheless, as Fig. 5 shows, our general approach is useful in deriving an estimate with improved performance despite the fact that the ML strategy is no longer efficient, and in fact may have lower MSE than that predicted by the CRLB.

APPENDIX A
PROOF OF LEMMA 2

Here, we prove that (27) is equivalent to (29).

Substituting $\mathbf{X} = \mathbf{W}^*\mathbf{W}$ and $\mathbf{M} = \mathbf{W}\widetilde{\mathbf{I}}$ into $\mathbf{G}(\mathbf{M},\mathbf{u})$, and noting that $\mathbf{W}^*\mathbf{W} + \mathbf{Z}(\mathbf{W},\mathbf{W}^*\mathbf{W}) = \mathbf{G}(\mathbf{M},\mathbf{u})$ (27) can be written as

$$\min_{t,\mathbf{M},\mathbf{X}} \{t : \mathbf{X} + \mathbf{Z}(\mathbf{W},\mathbf{X}) \preceq 0, \mathbf{X} = \mathbf{W}^*\mathbf{W}\}. \tag{77}$$

The problem is that the constraint $\mathbf{X} = \mathbf{W}^*\mathbf{W}$ is not convex. To obtain a convex problem we would like to relax this constraint to the convex form $\mathbf{X} \succeq \mathbf{W}^*\mathbf{W}$, leading to (29). 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 $\hat{t}, \widehat{\mathbf{W}}, \widehat{\mathbf{X}}$ are optimal for (29), then we can achieve the same value $\hat{t}$ with $\mathbf{X} = \widehat{\mathbf{W}}^*\widehat{\mathbf{W}}$.

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

$$\begin{aligned}&\widehat{\mathbf{X}} + \mathbf{Z}(\widehat{\mathbf{W}},\widehat{\mathbf{X}}) - \widehat{\mathbf{W}}^*\widehat{\mathbf{W}} - \mathbf{Z}(\widehat{\mathbf{W}},\widehat{\mathbf{W}}^*\widehat{\mathbf{W}})\\ &= \mathbf{Y} + \begin{bmatrix} \sum_{i=1}^{\ell} \mathbf{B}_i^*\widetilde{\mathbf{I}}^*\mathbf{Y}\widetilde{\mathbf{I}}\mathbf{B}_i & \sum_{i=1}^{k} \mathbf{C}_i^*\widetilde{\mathbf{I}}^*\mathbf{Y}\widetilde{\mathbf{I}}\mathbf{z}_i \\ \sum \mathbf{z}_i^*\widetilde{\mathbf{I}}^*\mathbf{Y}\widetilde{\mathbf{I}}\mathbf{C}_i & \mathrm{Tr}(\mathbf{A}\widetilde{\mathbf{I}}^*\mathbf{Y}\widetilde{\mathbf{I}}) \end{bmatrix} \triangleq \mathbf{Y} + \mathbf{F}(\widetilde{\mathbf{I}}^*\mathbf{Y}\widetilde{\mathbf{I}})\end{aligned} \tag{78}$$

where we defined

$$\mathbf{Y} \triangleq \widehat{\mathbf{X}} - \widehat{\mathbf{W}}^*\widehat{\mathbf{W}}. \tag{79}$$

In [9, App. III] it was shown that $\mathbf{F}(\mathbf{Q}) \succeq 0$ for all $\mathbf{Q} \succeq 0$. Since $\widehat{\mathbf{X}}$ and $\widehat{\mathbf{W}}$ are feasible, $\mathbf{Y} \succeq 0$, which implies that $\widetilde{\mathbf{I}}^*\mathbf{Y}\widetilde{\mathbf{I}} \succeq 0$, completing the proof.

APPENDIX B
DERIVATION OF THE DUAL (32)

To find the dual of (29), we first write the Lagrangian associated with our problem

$$\mathcal{L} = t + \mathrm{Tr}(\widetilde{\Pi}(\mathbf{X} + \mathbf{Z}(\mathbf{W},\mathbf{X}))) + \mathrm{Tr}\left(\Delta(\mathbf{W}^*\mathbf{W} - \mathbf{X})\right) \tag{80}$$

where $\Delta \succeq 0$ and

$$\widetilde{\Pi} = \begin{bmatrix} \Pi & \mathbf{w} \\ \mathbf{w}^* & \pi \end{bmatrix} \succeq 0 \tag{81}$$

are the dual variables.

Differentiating the Lagrangian with respect to t and equating to 0, we have $\pi = 1$. Setting the derivative with respect to $\mathbf{X}$ to 0 results in,

$$\begin{aligned}\Delta &= \widetilde{\Pi} + \widetilde{\mathbf{I}}\sum_{i=1}^{\ell} \mathbf{B}_i\Pi\mathbf{B}_i^*\widetilde{\mathbf{I}}^* + \widetilde{\mathbf{I}}\sum_{i=1}^{k}\left(\mathbf{z}_i\mathbf{w}^*\mathbf{C}_i^* + \mathbf{C}_i\mathbf{w}\mathbf{z}_i^*\right)\widetilde{\mathbf{I}}^* + \widetilde{\mathbf{I}}\mathbf{A}\widetilde{\mathbf{I}}^*\\ &= \widetilde{\Pi} + \widetilde{\mathbf{I}}\mathbf{S}(\Pi,\mathbf{w})\widetilde{\mathbf{I}}^*\end{aligned} \tag{82}$$

where we defined

$$\mathbf{S}(\Pi,\mathbf{w}) \triangleq \sum_{i=1}^{\ell} \mathbf{B}_i\Pi\mathbf{B}_i^* + \sum_{i=1}^{k}(\mathbf{z}_i\mathbf{w}^*\mathbf{C}_i^* + \mathbf{C}_i\mathbf{w}\mathbf{z}_i^*) + \mathbf{A}. \tag{83}$$

Finally, the derivative with respect to $\mathbf{W}$ yields $\mathbf{W}\Delta = -\mathbf{S}(\Pi,\mathbf{w})\widetilde{\mathbf{I}}^*$, which after substituting the value of Δ from (82), becomes

$$\mathbf{W}\left(\widetilde{\Pi} + \widetilde{\mathbf{I}}\mathbf{S}(\Pi,\mathbf{w})\widetilde{\mathbf{I}}^*\right) = -\mathbf{S}(\Pi,\mathbf{w})\widetilde{\mathbf{I}}^*. \tag{84}$$

Using the definitions of $\mathbf{W}$, $\widetilde{\mathbf{I}}$ and $\widetilde{\Pi}$, (84) can be written more explicitly as

$$[\mathbf{M} \quad \mathbf{u}]\begin{bmatrix} \mathbf{S}(\Pi,\mathbf{w}) + \Pi & \mathbf{w} \\ \mathbf{w}^* & 1 \end{bmatrix} = -\left[\mathbf{S}(\Pi,\mathbf{w}) \quad 0\right] \tag{85}$$

which is equivalent to

$$\mathbf{M}\mathbf{w} = -\mathbf{u} \tag{86}$$

$$\mathbf{M}\left(\mathbf{S}(\Pi,\mathbf{w}) + \Pi\right) + \mathbf{u}\mathbf{w}^* = -\mathbf{S}(\Pi,\mathbf{w}). \tag{87}$$

Substituting (86) into (87) we have

$$\mathbf{M}\left(\mathbf{S}(\Pi,\mathbf{w}) + \Pi - \mathbf{w}\mathbf{w}^*\right) = -\mathbf{S}(\Pi,\mathbf{w}). \tag{88}$$

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

$$\mathbf{S}(\Pi,\mathbf{w}) \succeq \mathbf{S}(\mathbf{w}\mathbf{w}^*,\mathbf{w}) = \mathbf{J}^{-1}(\mathbf{w}) \succ 0 \tag{89}$$

and $\mathbf{S}(\Pi,\mathbf{w}) + \Pi - \mathbf{w}\mathbf{w}^* \succ 0$. Thus, from (88) and (86)

$$\mathbf{M} = -\mathbf{S}(\Pi,\mathbf{w})\left(\mathbf{S}(\Pi,\mathbf{w}) + \Pi - \mathbf{w}\mathbf{w}^*\right)^{-1} \tag{90}$$

$$\mathbf{u} = \mathbf{S}(\Pi,\mathbf{w})\left(\mathbf{S}(\Pi,\mathbf{w}) + \Pi - \mathbf{w}\mathbf{w}^*\right)^{-1}\mathbf{w}. \tag{91}$$

To simplify the expression for $\mathbf{u}$ we would like to apply the matrix inversion lemma, which states that for an invertible matrix $\mathbf{A}$

$$(\mathbf{A} - \mathbf{b}\mathbf{b}^*)^{-1} = \mathbf{A}^{-1} + \frac{1}{1 - \mathbf{b}^*\mathbf{A}^{-1}\mathbf{b}}\mathbf{A}^{-1}\mathbf{b}\mathbf{b}^*\mathbf{A}^{-1} \tag{92}$$

as long as $\mathbf{b}^*\mathbf{A}^{-1}\mathbf{b} < 1$. Using the fact that $\mathbf{S} + \Pi \succ \mathbf{w}\mathbf{w}^*$, we have $\mathbf{I} \succ (\mathbf{S} + \Pi)^{-1/2}\mathbf{w}\mathbf{w}^*(\mathbf{S} + \Pi)^{-1/2}$, which is equivalent to $\mathbf{w}^*(\mathbf{S} + \Pi)^{-1}\mathbf{w} < 1$, where for brevity we denoted $\mathbf{S} = \mathbf{S}(\Pi,\mathbf{w})$. Therefore we can use (92) to reduce (91) to

$$\mathbf{u} = \frac{1}{1 - \mathbf{w}^*(\mathbf{S}(\Pi,\mathbf{w}) + \Pi)^{-1}\mathbf{w}}\mathbf{S}(\Pi,\mathbf{w})\left(\mathbf{S}(\Pi,\mathbf{w}) + \Pi\right)^{-1}\mathbf{w}. \tag{93}$$

Substituting (90) and (93) into the Lagrangian, the dual problem becomes

$$\min_{\Pi \succeq \mathbf{w}\mathbf{w}^*} \mathrm{Tr}\left(\mathbf{S}(\Pi,\mathbf{w})\left(\mathbf{S}(\Pi,\mathbf{w}) + \Pi - \mathbf{w}\mathbf{w}^*\right)^{-1}\mathbf{S}(\Pi,\mathbf{w})\right). \tag{94}$$

APPENDIX C
PROOF OF PROPOSITION 2

To prove the proposition, note that with $\mathbf{M} = -\mathbf{I}$

$$\begin{aligned} \mathbf{A}_0(\mathbf{M}) &= \mathbf{I} - \sum_{i=1}^{\ell} \mathbf{B}_i^* \mathbf{B}_i; \\ \mathbf{b}_0(\mathbf{M}, \mathbf{u}) &= -\sum_{i=1}^{k} \mathbf{C}_i^* \mathbf{z}_i - \mathbf{u}; \\ c_0(\mathbf{M}, \mathbf{u}) &= -\operatorname{Tr}(\mathbf{A}) + \mathbf{u}^* \mathbf{u}. \end{aligned} \tag{95}$$

Since we must have that $\mathbf{A}_0(\mathbf{M}) \preceq 0$ this implies (47), so that this condition is necessary.

We now establish that (47) is also sufficient by showing that the conditions (42) are satisfied with $\mathbf{w} = \mathbf{u}$ and $\Pi = \mathbf{w}\mathbf{w}^*$. The first two equalities clearly hold. It remains to prove that the matrix in (42) is nonnegative. From Lemma 3, this is equivalent to

$$\mathbf{A}_0(-\mathbf{I}) \preceq 0; \tag{96}$$

$$c_0(-\mathbf{I}, \mathbf{u}) + \operatorname{Tr}(\mathbf{S}(\mathbf{u}\mathbf{u}^*, \mathbf{u})) \leq \mathbf{b}_0^*(-\mathbf{I}, \mathbf{u}) \mathbf{A}_0^\dagger(-\mathbf{I}) \mathbf{b}_0^*(-\mathbf{I}, \mathbf{u}); \tag{97}$$

$$(\mathbf{I} - \mathbf{A}_0(-\mathbf{I}) \mathbf{A}_0^\dagger(-\mathbf{I})) \mathbf{b}_0(-\mathbf{I}, \mathbf{u}) = 0. \tag{98}$$

The first inequality (96) holds from (47). To prove (97), we first note that

$$\begin{aligned} & c_0(-\mathbf{I}, \mathbf{u}) + \operatorname{Tr}(\mathbf{S}(\mathbf{u}\mathbf{u}^*, \mathbf{u})) \\ &\quad = \sum_{i=1}^{k} \mathbf{z}_i^* \mathbf{C}_i \left(\mathbf{F}^{-1} (\mathbf{F} + \mathbf{I}) \mathbf{F}^{-1} - 2\mathbf{F}^{-1} \right) \sum_{i=1}^{k} \mathbf{C}_i^* \mathbf{z}_i \\ & \mathbf{b}_0^*(-\mathbf{I}, \mathbf{u}) \mathbf{A}_0^\dagger(-\mathbf{I}) \mathbf{b}_0^*(-\mathbf{I}, \mathbf{u}) \\ &\quad = \sum_{i=1}^{k} \mathbf{z}_i^* \mathbf{C}_i (\mathbf{I} - \mathbf{F}^{-1}) (\mathbf{I} - \mathbf{F})^\dagger (\mathbf{I} - \mathbf{F}^{-1}) \sum_{i=1}^{k} \mathbf{C}_i^* \mathbf{z}_i \end{aligned} \tag{99}$$

where for brevity we denoted $\mathbf{F} = \sum_{i=1}^{\ell} \mathbf{B}_i^* \mathbf{B}_i$. Using (99), the condition (97) is satisfied if

$$\mathbf{F}^{-1}(\mathbf{F}+\mathbf{I})\mathbf{F}^{-1} - 2\mathbf{F}^{-1} \preceq (\mathbf{I}-\mathbf{F}^{-1})(\mathbf{I}-\mathbf{F})^\dagger(\mathbf{I}-\mathbf{F}^{-1}). \tag{100}$$

Multiplying (100) on the left and on the right by $\mathbf{F} \succ 0$, this requirement is equivalent to

$$\mathbf{I} - \mathbf{F} \preceq (\mathbf{I} - \mathbf{F})(\mathbf{I} - \mathbf{F})^\dagger (\mathbf{I} - \mathbf{F}). \tag{101}$$

Using the fact that for any matrix $\mathbf{Z}$, $\mathbf{Z}\mathbf{Z}^\dagger\mathbf{Z} = \mathbf{Z}$ establishes (101).

It remains to prove (98). From the properties of the pseudoinverse, this condition is equivalent to

$$\mathbf{b}_0(-\mathbf{I}, \mathbf{u}) = (\mathbf{F}^{-1} - \mathbf{I}) \sum_{i=1}^{k} \mathbf{C}_i^* \mathbf{z}_i \in \mathcal{R}(\mathbf{A}_0(-\mathbf{I})) = \mathcal{R}(\mathbf{I} - \mathbf{F}). \tag{102}$$

Clearly $\mathbf{b}_0(-\mathbf{I}, \mathbf{u}) \in \mathcal{R}(\mathbf{F}^{-1} - \mathbf{I})$. Since $\mathbf{F}^{-1} - \mathbf{I} = (\mathbf{I} - \mathbf{F})\mathbf{F}^{-1}$ we have immediately that $\mathcal{R}(\mathbf{F}^{-1} - \mathbf{I}) = \mathcal{R}(\mathbf{I} - \mathbf{F})$, establishing (102).

APPENDIX D
PROOF OF PROPOSITION 3

Using arguments similar to those used in proving [9, Lemma 1] we can show that strict feasibility is equivalent to the condition that there exist $\mathbf{M}$ and $\lambda > 0$ such that

$$\mathbf{A}_0(\mathbf{M}) - \lambda \mathbf{A}_1 \prec 0. \tag{103}$$

We now show that (103) is equivalent to (51). Suppose first that (51) does not hold; this implies that there exists a vector $\mathbf{v}$ such that $\mathbf{F}\mathbf{v} = 0$ and $\mathbf{v}^* \mathbf{A}_1 \mathbf{v} \leq 0$, where we denoted $\mathbf{F} = \sum_{i=1}^{\ell} \mathbf{B}_i^* \mathbf{B}_i$. Since $\mathbf{F}$ is a sum of positive semidefinite matrices, $\mathbf{F}\mathbf{v} = 0$ 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

$$\mathbf{v}^* (\mathbf{A}_0(\mathbf{M}) - \lambda \mathbf{A}_1) \mathbf{v} = \mathbf{v}^* \mathbf{M}^* \mathbf{M} \mathbf{v} - \lambda \mathbf{v}^* \mathbf{A}_1 \mathbf{v} \geq 0 \tag{104}$$

for all $\lambda \geq 0$, so that (103) cannot be satisfied.

Next, suppose that (51) is satisfied. Since $\mathbf{A}_1$ and $\mathbf{F}$ are Hermitian and $\mathbf{F} \succeq 0$, there exists an invertible matrix $\mathbf{Q}$ such that

$$\mathbf{F} = \mathbf{Q}^* \operatorname{diag}(d_1, \ldots, d_m) \mathbf{Q}, \quad \mathbf{A}_1 = \mathbf{Q}^* \operatorname{diag}(\sigma_1, \ldots, \sigma_m) \mathbf{Q} \tag{105}$$

where $d_1, \ldots, d_r = 0$ with $r = \dim \mathcal{N}(\mathbf{F})$, and $0 < d_{r+1} \leq d_{r+2}, \ldots, d_m$. Now, any vector in $\mathcal{N}(\mathbf{F})$ can be written as

$$\mathbf{v} = \sum_{k=1}^{r} a_i \mathbf{q}_i \tag{106}$$

where a_i are arbitrary constants and $\mathbf{q}_i$ is the ith column of $\mathbf{Q}^{-1}$. From (51), $\mathbf{v}^* \mathbf{A}_1 \mathbf{v} > 0$ for any $\mathbf{v}$ of the form (106). Since $\mathbf{v}^* \mathbf{A}_1 \mathbf{v} = \sum_{k=1}^{r} \sigma_i |a_i|^2$, this implies that $\sigma_i > 0, 1 \leq i \leq r$.

If $d_m = 0$, then $\sigma_i > 0$ for all i which implies that $\mathbf{A}_1 \succ 0$. In this case (103) is satisfied with $\mathbf{M} = 0$ and any $\lambda > 0$. If $r = 0$, then $\mathbf{F} \succ 0$. Choosing $\lambda = 0$ it remains to show that there exists an $\mathbf{M}$ such that $\mathbf{A}_0(\mathbf{M}) \prec 0$; this is shown in [9, Lemma 1].

Next suppose that $r > 0, d_m > 0$ and $\mathbf{A}_1$ is not positive definite. Let $\beta = 1/\lambda_{\min}(\mathbf{Q}^* \mathbf{Q})$, $\kappa_+ = \arg\min_{i:\sigma_i > 0} \sigma_i$, and $\kappa_- = \arg\max_{i:\sigma_i \leq 0} |\sigma_i|$. Our assumptions imply that κ_+ and κ_- exist. We now establish that (103) is satisfied with $\lambda = \alpha^2(d_m + \beta)/\sigma_{\kappa_+}$ and $\mathbf{M} = \alpha \mathbf{I}$ where $\alpha = -d_{r+1}/(d_{r+1} + \beta - (d_m + \beta)\sigma_{\kappa_-}/\sigma_{\kappa_+})$. With this choice

$$\begin{aligned} \mathbf{A}_0(\mathbf{M}) - \lambda \mathbf{A}_1 &= \alpha^2 \mathbf{I} + \mathbf{Q}^* ((\alpha^2 + 2\alpha) \operatorname{diag}(d_1, \ldots, d_m) \\ &\qquad - \lambda \operatorname{diag}(\sigma_1, \ldots, \sigma_m)) \mathbf{Q} \\ &\preceq \mathbf{Q}^* (\alpha^2 \beta \mathbf{I} + (\alpha^2 + 2\alpha) \operatorname{diag}(d_1, \ldots, d_m) \\ &\qquad - \lambda \operatorname{diag}(\sigma_1, \ldots, \sigma_m)) \mathbf{Q} \end{aligned} \tag{107}$$

where we used the fact that $\mathbf{I} \preceq \beta \mathbf{Q}^* \mathbf{Q}$. Thus, to prove (103) it is sufficient to establish that

$$\alpha^2 \left(\beta + d_i - \left(\frac{\sigma_i}{\sigma_{\kappa_+}} \right) (d_m + \beta) \right) + 2\alpha d_i < 0, \; r+1 \leq i \leq m \tag{108}$$

where we substituted the value of λ. For indices i such that $\sigma_i > 0$, (108) is trivially satisfied since $\sigma_i / \sigma_{\kappa_+} \geq 1$, $d_i \leq d_m$, $d_m >$

0 and $\alpha < 0$. Consider next indices $r+1 \leq i \leq m$ such that $\sigma_i \leq 0$. For these values, (108) is equivalent to

$$\alpha > -\frac{2d_i}{\beta + d_i - \frac{\sigma_i}{\sigma_{\kappa_+}}(d_m + \beta)}. \tag{109}$$

Since the right-hand expression is monotonically decreasing in d_i and $\sigma_i < 0$ our choice of α satisfies (109).

References

[1] H. Cramer, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton Univ. Press, 1946.

[2] C. R. Rao, *Linear Statistical Inference and Its Applications*, second ed. New York: Wiley, 1973.

[3] B. Efron, "Biased versus unbiased estimation," *Adv.Math.*, vol. 16, pp. 259–277, 1975.

[4] 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, pp. 2026–2041, Aug. 1996.

[5] Y. C. Eldar, "Minimum variance in biased estimation: Bounds and asymptotically optimal estimators," *IEEE Trans. Signal Process.*, vol. 52, pp. 1915–1930, Jul. 2004.

[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] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1968.

[9] Y. C. Eldar, "Uniformly improving the Cramér-Rao bound and maximum-likelihood estimation," *IEEE Trans. Signal Process.*, vol. 54, pp. 2943–2956, Aug. 2006.

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

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

[12] Y. C. Eldar, "Comparing between estimation approaches: Admissible and dominating linear estimators," *IEEE Trans. Signal Process.*, vol. 54, pp. 1689–1702, May 2006.

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

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

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

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

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

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

[19] 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.

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

[21] A. Beck and Y. C. Eldar, "Strong duality in nonconvex quadratic optimization with two quadratic constraints," *SIAM J. Optim.*, vol. 17, no. 3, pp. 844–860, 2006.

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

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

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

[25] Z. Ben-Haim and Y. C. Eldar, "Maximum set estimators with bounded estimation error," *IEEE Trans. Signal Process.*, vol. 53, pp. 3172–3182, Aug. 2005.

[26] 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, pp. 168–181, Jan. 2005.

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

[28] J. Kuks, "A minimax estimator of regression coefficients," *Iswestija Akademija Nauk Estonskoj SSR*, vol. 21, pp. 73–78, 1972.

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

Yonina C. Eldar (S'98–M'02–SM'07) received the B.Sc. degree in physics in 1995 and the B.Sc. degree in electrical engineering in 1996, both from Tel-Aviv University (TAU), Tel-Aviv, Israel, and the Ph.D. degree in electrical engineering and computer science in 2001 from the Massachusetts Institute of Technology (MIT), Cambridge.

From January 2002 to July 2002, she was a Postdoctoral Fellow with the Digital Signal Processing Group, MIT. She is currently an Associate Professor in 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 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. From 2002 to 2005, she was a Horev Fellow of the Leaders in Science and Technology program at the Technion and an Alon Fellow. In 2004, she was awarded the Wolf Foundation Krill Prize for Excellence in Scientific Research, in 2005 the Andre and Bella Meyer Lectureship, in 2007 the Henry Taub Prize for Excellence in Research, and in 2008, the Hershel Rich Innovation Award, the Award for Women with Distinguished Contributions, and the Muriel and David Jacknow Award for Excellence in Teaching. She is a member of the IEEE Signal Processing Theory and Methods Technical Committee, an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the *EURASIP Journal of Signal Processing*, and the *SIAM Journal on Matrix Analysis and Applications*, and on the Editorial Board of *Foundations and Trends in Signal Processing*.