

On the Constrained Cramér–Rao Bound With a Singular Fisher Information Matrix

Zvika Ben-Haim, *Student Member, IEEE*, and Yonina C. Eldar, *Senior Member, IEEE*

Abstract—The problem considered in this letter is to bound the performance of estimators of a deterministic parameter which satisfies given constraints. Specifically, previous work on the constrained Cramér–Rao bound (CRB) is generalized to include singular Fisher information matrices and biased estimators. A necessary and sufficient condition for the existence of a finite CRB is obtained. A closed form for an estimator achieving the CRB, if one exists, is also provided, as well as a necessary and sufficient condition for the existence of such an estimator. It is shown that biased estimators achieving the CRB can be constructed in situations for which no unbiased technique exists.

Index Terms—Constrained estimation, Cramér–Rao bound, parameter estimation.

I. INTRODUCTION

A CENTRAL goal in statistics and signal processing is to estimate unknown deterministic parameters from random measurements. The performance of estimators in such a setting is circumscribed by the well-known Cramér–Rao bound (CRB) [1]. Specifically, the CRB provides a lower limit on the variance obtainable by any technique as a function of the Fisher information matrix (FIM) and the estimator’s bias gradient.

A variant of the CRB for constrained estimation problems was developed by Gorman and Hero [2]. They considered the setting in which the parameter vector belongs to a known set. When this information is incorporated into the estimator, performance can be improved. As a consequence, the constrained CRB can be lower than the unconstrained version.

The derivation of Gorman and Hero assumed that the FIM is positive definite. Stoica and Ng [3] later extended the constrained CRB to the case in which the FIM is positive *semi*-definite, and may thus be singular. In an unconstrained problem, a singular FIM implies that unbiased estimation of the entire parameter vector is impossible [4]. However, Stoica and Ng demonstrated that, in some cases, one can obtain so-called constrained unbiased estimators, which are unbiased as long as the constraints hold.

Manuscript received November 16, 2008; revised February 01, 2009. Current version published April 24, 2009. This work was supported in part by the Israel Science Foundation under 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 Dr. Shahram Shahbazzpanahi.

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

Digital Object Identifier 10.1109/LSP.2009.2016831

The work of Stoica and Ng considers only unbiased estimation. Yet even when unbiased methods do not exist in a particular setting, biased techniques can still be found. As we will demonstrate, when the FIM is singular, estimators can be constructed by introducing a sufficient number of constraints, by specifying an appropriate bias function, or by a combination thereof.

More specifically, in this letter we generalize the above-mentioned bounds and obtain a biased CRB for constrained estimation with a positive semi-definite FIM. When an estimator achieving the CRB exists, we provide a closed form for it. We further derive a necessary and sufficient condition for the CRB to be infinite, indicating that no estimator exists in the given setting.

The following notation is used throughout the letter. Given a vector function $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^k$, we denote by $\partial\mathbf{f}/\partial\boldsymbol{\theta}$ the $k \times n$ matrix whose ij th element is $\partial f_i/\partial\theta_j$. Also, $\mathcal{R}(\mathbf{A})$, $\mathcal{N}(\mathbf{A})$, and \mathbf{A}^\dagger are, respectively, the range space, null space, and Moore–Penrose pseudoinverse of a matrix \mathbf{A} , and \mathcal{S}^\perp denotes the orthogonal complement of the subspace \mathcal{S} . Finally, $\mathbf{A} \succeq \mathbf{B}$ indicates that $\mathbf{A} - \mathbf{B}$ is positive semi-definite.

II. PROBLEM STATEMENT

Let \mathbf{y} be a measurement vector with pdf $p(\mathbf{y}; \boldsymbol{\theta})$, for some deterministic unknown parameter vector $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^n$. Suppose that $p(\mathbf{y}; \boldsymbol{\theta})$ is differentiable with respect to $\boldsymbol{\theta}$. The FIM $\mathbf{J}(\boldsymbol{\theta})$ is then defined as

$$\mathbf{J}(\boldsymbol{\theta}) = E\{\boldsymbol{\Delta}\boldsymbol{\Delta}^T\} \quad (1)$$

where

$$\boldsymbol{\Delta} = \frac{\partial \log p(\mathbf{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \quad (2)$$

We assume throughout that $\mathbf{J}(\boldsymbol{\theta})$ is finite for all $\boldsymbol{\theta} \in \Theta$.

Suppose that $\boldsymbol{\theta}$ is known to belong to a constraint set

$$\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^n : \mathbf{f}(\boldsymbol{\theta}) = \mathbf{0}\} \subseteq \mathbb{R}^n \quad (3)$$

where $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^k$ is a continuously differentiable function of $\boldsymbol{\theta}$ with $0 \leq k \leq n$. Note that we are assuming for simplicity that no inequality constraints are present, as it has been shown that such constraints have no effect on the CRB [2].

We further assume that the $k \times n$ matrix $\mathbf{F}(\boldsymbol{\theta}) = \partial\mathbf{f}/\partial\boldsymbol{\theta}$ has full row rank, which is equivalent to requiring that the constraints are not redundant. Thus, there exists an $n \times (n - k)$ matrix $\mathbf{U}(\boldsymbol{\theta})$ such that

$$\mathbf{F}(\boldsymbol{\theta})\mathbf{U}(\boldsymbol{\theta}) = \mathbf{0}, \quad \mathbf{U}^T(\boldsymbol{\theta})\mathbf{U}(\boldsymbol{\theta}) = \mathbf{I}. \quad (4)$$

Intuitively, $\mathcal{R}(\mathbf{U}(\boldsymbol{\theta}))$ is the set of feasible directions at $\boldsymbol{\theta}$, i.e., the set of directions in which an infinitesimal change does not

violate the constraints. For notational simplicity, in the sequel we will omit the dependence of \mathbf{U} and \mathbf{J} on $\boldsymbol{\theta}$.

Let $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{y})$ be an estimator of $\boldsymbol{\theta}$. We are interested in the performance of $\hat{\boldsymbol{\theta}}$ under the assumption that $\boldsymbol{\theta} \in \Theta$. Specifically, we derive a lower bound on the covariance matrix

$$\text{Cov}(\hat{\boldsymbol{\theta}}) = E\{(\hat{\boldsymbol{\theta}} - E\{\hat{\boldsymbol{\theta}}\})(\hat{\boldsymbol{\theta}} - E\{\hat{\boldsymbol{\theta}}\})^T\} \quad (5)$$

obtainable by any estimator $\hat{\boldsymbol{\theta}}$. The covariance matrix, as well as the CRB, are a function of $\boldsymbol{\theta}$; we are interested in bounding this matrix for all $\boldsymbol{\theta} \in \Theta$. To obtain a nontrivial bound, we assume that the desired bias $\mathbf{b}(\boldsymbol{\theta}) = E\{\hat{\boldsymbol{\theta}}\} - \boldsymbol{\theta}$ is specified for $\boldsymbol{\theta} \in \Theta$; the bias for $\boldsymbol{\theta} \notin \Theta$ is arbitrary.

Previous work on the constrained estimation setting [2], [3] assumed that the estimator $\hat{\boldsymbol{\theta}}$ satisfies the constraint $\hat{\boldsymbol{\theta}} \in \Theta$. However, it turns out that this requirement can be removed without altering the resulting bound. Furthermore, in some cases, the CRB can only be achieved by estimators violating the constraint. In this letter, the term ‘‘constrained estimator’’ refers to the situation in which the bias $\mathbf{b}(\boldsymbol{\theta})$ is specified only for $\boldsymbol{\theta} \in \Theta$, and the performance is evaluated when the true parameter value $\boldsymbol{\theta}$ belongs to the set Θ . The implications of this setting are discussed further in the next section.

III. CRAMÉR–RAO BOUND

A. Main Result

With the concepts developed in the previous section, our main result can be stated as follows.

Theorem 1: Let Θ be a constraint set of the form (3) with a corresponding matrix \mathbf{U} of (4). Let $\hat{\boldsymbol{\theta}}$ be an estimator of $\boldsymbol{\theta}$ whose bias is given by $\mathbf{b}(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Theta$, and define

$$\mathbf{A} = \mathbf{I} + \frac{\partial \mathbf{b}}{\partial \boldsymbol{\theta}}. \quad (6)$$

Assume that integration with respect to \mathbf{y} and differentiation with respect to $\boldsymbol{\theta}$ can be interchanged,¹ and suppose that

$$\mathcal{R}(\mathbf{U}\mathbf{U}^T\mathbf{A}^T) \subseteq \mathcal{R}(\mathbf{U}\mathbf{U}^T\mathbf{J}\mathbf{U}\mathbf{U}^T). \quad (7)$$

Then, the covariance of $\hat{\boldsymbol{\theta}}$ satisfies

$$\text{Cov}(\hat{\boldsymbol{\theta}}) \succeq \mathbf{A}\mathbf{U}(\mathbf{U}^T\mathbf{J}\mathbf{U})^\dagger\mathbf{U}^T\mathbf{A}^T \quad \text{for all } \boldsymbol{\theta} \in \Theta. \quad (8)$$

Equality is achieved in (8) if and only if

$$\hat{\boldsymbol{\theta}} = \mathbf{b}(\boldsymbol{\theta}) + \boldsymbol{\theta} + \mathbf{A}\mathbf{U}(\mathbf{U}^T\mathbf{J}\mathbf{U})^\dagger\mathbf{U}^T\boldsymbol{\Delta} \quad (9)$$

in the mean square sense, for all $\boldsymbol{\theta} \in \Theta$. Here, $\boldsymbol{\Delta}$ is given by (2). Conversely, if (7) does not hold, then there exists no finite-variance estimator with the given bias function.

It is illuminating to examine the influence of the constraints on the bound of Theorem 1. Recall that the CRB is a bound on the covariance of *all* estimators having a given bias function, at each specific point $\boldsymbol{\theta}$. The bound thus applies even to estimators which are designed for the specific point $\boldsymbol{\theta}$, a far more restrictive assumption than the knowledge that $\boldsymbol{\theta} \in \Theta$. How, then, can one expect to obtain a meaningful performance bound by imposing the constraint set Θ ?

¹This condition basically requires that the bounds of $p(\mathbf{y}; \boldsymbol{\theta})$ do not depend on $\boldsymbol{\theta}$. Such regularity conditions are assumed in all forms of the CRB.

The answer stems from the fact that the bias is specified in Theorem 1 only for $\boldsymbol{\theta} \in \Theta$. For example, consider constrained unbiased estimators, for which $\mathbf{b}(\boldsymbol{\theta}) = \mathbf{0}$ for all $\boldsymbol{\theta} \in \Theta$; the bias when $\boldsymbol{\theta} \notin \Theta$ is irrelevant and unspecified. This is a far larger class of estimators than those which are unbiased for all $\boldsymbol{\theta} \in \mathbb{R}^n$. Consequently, the bound (8) is lower than the unconstrained CRB. The weakened bias specification is apparent in Theorem 1 from the fact that the matrix \mathbf{A} only appears when multiplied by \mathbf{U} , which nullifies components in directions violating the constraints. Indeed, to calculate the bound, \mathbf{A} only needs to be specified in directions consistent with Θ . This issue will be discussed further in a forthcoming paper [5].

Condition (7) succinctly describes the possibilities for estimation under various values of the FIM. If \mathbf{J} is invertible, then (7) holds regardless of the constraint set and the bias gradient, implying that the CRB is always finite. The situation is more complicated when \mathbf{J} is singular. In this case, one option is to choose a matrix \mathbf{A} whose null space includes $\mathcal{N}(\mathbf{J})$; this implies that the estimator is insensitive to changes in elements of $\boldsymbol{\theta}$ for which there is no information. Another option is to provide external constraints for the unmeasurable elements of $\boldsymbol{\theta}$, thus changing \mathbf{U} in such a way as to ensure the validity of (7) for all \mathbf{A} . An example comparing these approaches will be presented in Section IV.

Theorem 1 encompasses several previous results as special cases. Most famously, when \mathbf{J} is nonsingular and no constraints are imposed, we obtain the standard CRB

$$\text{Cov}(\hat{\boldsymbol{\theta}}) \succeq \mathbf{A}\mathbf{J}^{-1}\mathbf{A}^T. \quad (10)$$

Several prior extensions [2]–[4] of (10) are also special cases of Theorem 1.

B. Proof of Theorem 1

The Proof of Theorem 1 is based on the following lemmas.

Lemma 1: Assuming that integration with respect to \mathbf{y} and differentiation with respect to $\boldsymbol{\theta}$ can be interchanged, we have

$$E\{(\hat{\boldsymbol{\theta}} - E\{\hat{\boldsymbol{\theta}}\})\boldsymbol{\Delta}^T\} = \mathbf{A} \quad (11)$$

for any estimator $\hat{\boldsymbol{\theta}}$. Here, $\boldsymbol{\Delta}$ is defined by (2), and \mathbf{A} is given by (6).

Proof: The proof is an extension of [6, Th. 1] to the case of a biased estimator. Using (2),

$$\begin{aligned} & E\{(\hat{\boldsymbol{\theta}} - E\{\hat{\boldsymbol{\theta}}\})\boldsymbol{\Delta}^T\} \\ &= \int (\hat{\boldsymbol{\theta}} - E\{\hat{\boldsymbol{\theta}}\}) \frac{1}{p(\mathbf{y}; \boldsymbol{\theta})} \frac{\partial p(\mathbf{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} p(\mathbf{y}; \boldsymbol{\theta}) d\mathbf{y} \\ &= \frac{\partial}{\partial \boldsymbol{\theta}} \int \hat{\boldsymbol{\theta}} p(\mathbf{y}; \boldsymbol{\theta}) d\mathbf{y} - E\{\hat{\boldsymbol{\theta}}\} \frac{\partial}{\partial \boldsymbol{\theta}} \int p(\mathbf{y}; \boldsymbol{\theta}) d\mathbf{y} \end{aligned} \quad (12)$$

where we interchanged the order of differentiation and integration, and used the fact that $\hat{\boldsymbol{\theta}}$ is a function of \mathbf{y} but not of $\boldsymbol{\theta}$. Noting that the second integral in (12) equals 1, we obtain

$$E\{(\hat{\boldsymbol{\theta}} - E\{\hat{\boldsymbol{\theta}}\})\boldsymbol{\Delta}^T\} = \frac{\partial E\{\hat{\boldsymbol{\theta}}\}}{\partial \boldsymbol{\theta}} = \mathbf{I} + \frac{\partial \mathbf{b}}{\partial \boldsymbol{\theta}} \quad (13)$$

which completes the proof. ■

The following lemma provides a family of bounds on $\text{Cov}(\hat{\boldsymbol{\theta}})$ for any estimator $\hat{\boldsymbol{\theta}}$ having a specified bias function. Theorem 1 is obtained by choosing an optimal member from this class.

Lemma 2: Let $\hat{\boldsymbol{\theta}}$ be an estimator of $\boldsymbol{\theta} \in \Theta$, and suppose its bias is $\mathbf{b}(\boldsymbol{\theta})$. Under the conditions of Lemma 1, for any $n \times n$ matrix \mathbf{W} , we have

$$\text{Cov}(\hat{\boldsymbol{\theta}}) \succeq \mathbf{W}\mathbf{U}\mathbf{U}^T\mathbf{A}^T + \mathbf{A}\mathbf{U}\mathbf{U}^T\mathbf{W}^T - \mathbf{W}\mathbf{U}\mathbf{U}^T\mathbf{J}\mathbf{U}\mathbf{U}^T\mathbf{W}^T. \quad (14)$$

Proof of Lemma 2: Let $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} - E\{\hat{\boldsymbol{\theta}}\}$ and note that

$$\begin{aligned} \mathbf{0} &\preceq E\left\{(\tilde{\boldsymbol{\theta}} - \mathbf{W}\mathbf{U}\mathbf{U}^T\boldsymbol{\Delta})(\tilde{\boldsymbol{\theta}} - \mathbf{W}\mathbf{U}\mathbf{U}^T\boldsymbol{\Delta})^T\right\} \\ &= \text{Cov}(\hat{\boldsymbol{\theta}}) - \mathbf{W}\mathbf{U}\mathbf{U}^T E\{\boldsymbol{\Delta}\tilde{\boldsymbol{\theta}}^T\} - E\{\tilde{\boldsymbol{\theta}}\boldsymbol{\Delta}^T\}\mathbf{U}\mathbf{U}^T\mathbf{W}^T \\ &\quad + \mathbf{W}\mathbf{U}\mathbf{U}^T E\{\boldsymbol{\Delta}\boldsymbol{\Delta}^T\}\mathbf{U}\mathbf{U}^T\mathbf{W}^T. \end{aligned} \quad (15)$$

Using (1) and Lemma 1, we obtain (14). \blacksquare

We recall the following properties of the pseudoinverse, which will be required for some further developments.

Lemma 3: Let \mathbf{M} and \mathbf{N} be arbitrary matrices and let $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. Then

$$(\mathbf{U}\mathbf{M}\mathbf{U}^T)^\dagger = \mathbf{U}\mathbf{M}^\dagger\mathbf{U}^T \quad (16)$$

$$(\mathbf{M}^T\mathbf{M})^\dagger = \mathbf{M}^\dagger\mathbf{M}^{T\dagger} \quad (17)$$

$$\mathbf{M}^\dagger = \mathbf{M}^\dagger\mathbf{M}^{T\dagger}\mathbf{M}^T \quad (18)$$

$$\mathbf{M} = \mathbf{M}^{T\dagger}\mathbf{M}^T\mathbf{M} \quad (19)$$

$$(\mathbf{M}\mathbf{N})^\dagger = (\mathbf{M}\mathbf{N})^\dagger\mathbf{M}^{T\dagger}\mathbf{M}^T. \quad (20)$$

Proof: Proofs for (16)–(19) can be found in [7, Th. 1.2.1], while (20) can be demonstrated by showing that $(\mathbf{M}\mathbf{N})^\dagger\mathbf{M}^{T\dagger}\mathbf{M}^T$ satisfies the Moore–Penrose conditions for the pseudoinverse of $\mathbf{M}\mathbf{N}$. \blacksquare

We are now ready to prove the main result.

Proof of Theorem 1: Our proof is based on that of Stoica and Ng [3]. Suppose first that (7) holds, and let

$$\mathbf{W} = \mathbf{A}\mathbf{U}\mathbf{U}^T(\mathbf{U}\mathbf{U}^T\mathbf{J}\mathbf{U}\mathbf{U}^T)^\dagger. \quad (21)$$

Applying Lemma 2 and using the Moore–Penrose condition $\mathbf{M}^\dagger\mathbf{M}\mathbf{M}^\dagger = \mathbf{M}^\dagger$, we obtain

$$\text{Cov}(\hat{\boldsymbol{\theta}}) \succeq \mathbf{A}\mathbf{U}\mathbf{U}^T(\mathbf{U}\mathbf{U}^T\mathbf{J}\mathbf{U}\mathbf{U}^T)^\dagger\mathbf{U}\mathbf{U}^T\mathbf{A}^T. \quad (22)$$

It follows from (16) that

$$(\mathbf{U}\mathbf{U}^T\mathbf{J}\mathbf{U}\mathbf{U}^T)^\dagger = \mathbf{U}^T(\mathbf{U}^T\mathbf{J}\mathbf{U})^\dagger\mathbf{U} = \mathbf{U}(\mathbf{U}^T\mathbf{J}\mathbf{U})^\dagger\mathbf{U}^T. \quad (23)$$

Substituting this into (22) yields (8), as required.

We now show that (9) holds if and only if

$$\text{Cov}(\hat{\boldsymbol{\theta}}) = \mathbf{A}\mathbf{U}(\mathbf{U}^T\mathbf{J}\mathbf{U})^\dagger\mathbf{U}^T\mathbf{A}^T \quad (24)$$

in the mean squared sense, for all $\boldsymbol{\theta} \in \Theta$. Note first that if $\hat{\boldsymbol{\theta}}$ satisfies (9), then the bias of $\hat{\boldsymbol{\theta}}$ is indeed $\mathbf{b}(\boldsymbol{\theta})$, since $E\{\boldsymbol{\Delta}\} = \mathbf{0}$. Furthermore

$$\text{Cov}(\hat{\boldsymbol{\theta}}) = \mathbf{A}\mathbf{U}(\mathbf{U}^T\mathbf{J}\mathbf{U})^\dagger\mathbf{U}^T E\{\boldsymbol{\Delta}\boldsymbol{\Delta}^T\}\mathbf{U}(\mathbf{U}^T\mathbf{J}\mathbf{U})^\dagger\mathbf{U}^T\mathbf{A}^T \quad (25)$$

which yields (24). Conversely, suppose that (24) holds, and let $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} - E\{\hat{\boldsymbol{\theta}}\}$. Using Lemma 1 and (24), it is straightforward to show that

$$\text{Cov}(\tilde{\boldsymbol{\theta}} - \mathbf{A}\mathbf{U}(\mathbf{U}^T\mathbf{J}\mathbf{U})^{-1}\mathbf{U}^T\boldsymbol{\Delta}) = \mathbf{0}. \quad (26)$$

Therefore, $\hat{\boldsymbol{\theta}} = E\{\hat{\boldsymbol{\theta}}\} + \mathbf{A}\mathbf{U}(\mathbf{U}^T\mathbf{J}\mathbf{U})^{-1}\mathbf{U}^T\boldsymbol{\Delta}$ in the mean square sense, as required.

It remains to show that if

$$\mathcal{R}(\mathbf{U}\mathbf{U}^T\mathbf{A}^T) \not\subseteq \mathcal{R}(\mathbf{U}\mathbf{U}^T\mathbf{J}\mathbf{U}\mathbf{U}^T) \quad (27)$$

then no finite variance estimator exists. Suppose that (27) holds. Since $\mathcal{R}(\mathbf{M}) = \mathcal{N}(\mathbf{M}^T)^\perp$ for any matrix \mathbf{M} , we have $\mathcal{N}(\mathbf{A}\mathbf{U}\mathbf{U}^T)^\perp \not\subseteq \mathcal{N}(\mathbf{U}\mathbf{U}^T\mathbf{J}\mathbf{U}\mathbf{U}^T)^\perp$, or equivalently, $\mathcal{N}(\mathbf{U}\mathbf{U}^T\mathbf{J}\mathbf{U}\mathbf{U}^T) \not\subseteq \mathcal{N}(\mathbf{A}\mathbf{U}\mathbf{U}^T)$. Thus, there exists a vector $\mathbf{v} \in \mathcal{N}(\mathbf{U}\mathbf{U}^T\mathbf{J}\mathbf{U}\mathbf{U}^T)$ for which $\mathbf{v} \notin \mathcal{N}(\mathbf{A}\mathbf{U}\mathbf{U}^T)$. Now, let $\mathbf{W} = \alpha\mathbf{A}\mathbf{U}\mathbf{U}^T\mathbf{v}\mathbf{v}^T$ for some scalar α to be defined below. From Lemma 2,

$$\begin{aligned} \text{Tr}(\text{Cov}(\hat{\boldsymbol{\theta}})) &\geq 2\text{Tr}(\mathbf{W}\mathbf{U}\mathbf{U}^T\mathbf{A}^T) - \text{Tr}(\mathbf{W}\mathbf{U}\mathbf{U}^T\mathbf{J}\mathbf{U}\mathbf{U}^T\mathbf{W}^T) \\ &= 2\alpha\text{Tr}(\mathbf{A}\mathbf{U}\mathbf{U}^T\mathbf{v}\mathbf{v}^T\mathbf{U}\mathbf{U}^T\mathbf{A}^T) \\ &\quad - \alpha^2\text{Tr}(\mathbf{A}\mathbf{U}\mathbf{U}^T\mathbf{v}\mathbf{v}^T\mathbf{U}\mathbf{U}^T\mathbf{J}\mathbf{U}\mathbf{U}^T\mathbf{v}\mathbf{v}^T\mathbf{U}\mathbf{U}^T\mathbf{A}^T). \end{aligned} \quad (28)$$

The second term in (28) is zero since $\mathbf{v} \in \mathcal{N}(\mathbf{U}\mathbf{U}^T\mathbf{J}\mathbf{U}\mathbf{U}^T)$, whereas the first term equals $2\alpha\|\mathbf{A}\mathbf{U}\mathbf{U}^T\mathbf{v}\|^2$, which is nonzero since $\mathbf{v} \notin \mathcal{N}(\mathbf{A}\mathbf{U}\mathbf{U}^T)$. Thus, by choosing α appropriately, $\text{Tr}(\text{Cov}(\hat{\boldsymbol{\theta}}))$ can be shown to be larger than any finite number. Therefore, there does not exist a finite-variance estimator with the required bias. \blacksquare

C. Choice of \mathbf{W}

The bound (8) of Theorem 1 is obtained from the more general Lemma 2 by choosing a specific value (21) for the matrix \mathbf{W} . We now show that this choice of \mathbf{W} is optimal, in that it results in the tightest bound obtainable from Lemma 2. Note that Lemma 2 provides a matrix inequality, so there does not necessarily exist a single maximum value of the bound (because the set of matrices is not totally ordered). However, in our case, such a maximum value does exist and results in the bound of Theorem 1.

The method of obtaining \mathbf{W} used in [3] does not seem to generalize to the case of biased estimators. Instead, let \mathbf{v} be an arbitrary vector in \mathbb{R}^m and observe that

$$\mathbf{v}^T(\mathbf{W}\mathbf{U}\mathbf{U}^T\mathbf{A}^T + \mathbf{A}\mathbf{U}\mathbf{U}^T\mathbf{W}^T - \mathbf{W}\mathbf{U}\mathbf{U}^T\mathbf{J}\mathbf{U}\mathbf{U}^T\mathbf{W}^T)\mathbf{v} \quad (29)$$

is concave in \mathbf{W} . Therefore, to maximize (29), it suffices to find a point \mathbf{W} at which the derivative is zero. Differentiating (29) with respect to \mathbf{W} , we obtain [8]

$$2\mathbf{v}\mathbf{v}^T\mathbf{A}\mathbf{U}\mathbf{U}^T - 2\mathbf{v}\mathbf{v}^T\mathbf{W}\mathbf{U}\mathbf{U}^T\mathbf{J}\mathbf{U}\mathbf{U}^T. \quad (30)$$

Thus, if there exists a matrix \mathbf{W} such that

$$\mathbf{U}\mathbf{U}^T\mathbf{J}\mathbf{U}\mathbf{U}^T\mathbf{W}^T = \mathbf{U}\mathbf{U}^T\mathbf{A}^T \quad (31)$$

then that value of \mathbf{W} maximizes (29) simultaneously for any choice of \mathbf{v} . Note that (31) can be written as a set of n vector equations

$$\mathbf{U}\mathbf{U}^T\mathbf{J}\mathbf{U}\mathbf{U}^T\mathbf{w}_i = \mathbf{u}_i, \quad i = 1, \dots, n \quad (32)$$

where \mathbf{w}_i is the i th row of \mathbf{W} and \mathbf{u}_i is the i th column of $\mathbf{U}\mathbf{U}^T\mathbf{A}^T$. Clearly, (32) has a solution \mathbf{w}_i if and only if $\mathbf{u}_i \in$

$\mathcal{R}(\mathbf{U}\mathbf{U}^T\mathbf{J}\mathbf{U}\mathbf{U}^T)$. This does indeed occur under the condition (7) of Theorem 1, and one such solution is given by

$$\mathbf{w}_i = (\mathbf{U}\mathbf{U}^T\mathbf{J}\mathbf{U}\mathbf{U}^T)^\dagger \mathbf{u}_i, \quad i = 1, \dots, n. \quad (33)$$

Combining these n equations, we obtain that the matrix \mathbf{W} chosen in (21) simultaneously maximizes (29) for all values of \mathbf{v} . Therefore, the bound of Theorem 1 is the tightest bound obtainable from Lemma 2.

IV. EXAMPLE

As an example of the applicability of Theorem 1, we consider an underdetermined linear regression setting. Let $\boldsymbol{\theta} \in \mathbb{R}^n$ be an unknown vector for which measurements $\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{v}$ are available. Here, \mathbf{v} is white Gaussian noise with variance σ^2 and \mathbf{H} is a known $p \times n$ matrix with $p < n$. Since there are fewer measurements than parameters, unbiased reconstruction of $\boldsymbol{\theta}$ is clearly impossible without additional assumptions. To see this formally, note that

$$\mathbf{J} = \sigma^{-2} \mathbf{H}^T \mathbf{H}. \quad (34)$$

Thus $\text{rank}(\mathbf{J}) \leq p < n$, so that the matrix \mathbf{J} is singular, and by Theorem 1, unconstrained unbiased estimation is impossible. This also follows from earlier results [3], [4].

In order to enable reconstruction of $\boldsymbol{\theta}$, additional assumptions are required. One possibility is to restrict $\boldsymbol{\theta}$ to some subset $\Theta \subset \mathbb{R}^n$, and then seek an unbiased estimator over this set. An alternative is to choose a reasonable value for $E\{\hat{\boldsymbol{\theta}}\}$, taking into account the lack of information. As we will see, both approaches result in the same estimator, but the latter implies optimality under wider conditions.

Beginning with the first approach, let us assume that $\boldsymbol{\theta} = \mathbf{W}\boldsymbol{\alpha}$ for a given $n \times k$ matrix \mathbf{W} and an unknown $\boldsymbol{\alpha} \in \mathbb{R}^k$. For example, \mathbf{W} can define a smoothness requirement on $\boldsymbol{\theta}$. We seek an unbiased estimator for such $\boldsymbol{\theta}$.

Choosing $\Theta = \mathcal{R}(\mathbf{W})$ results in $\mathbf{U}(\boldsymbol{\theta}) = \mathbf{W}$. Thus, it follows from Theorem 1 that if there exists a constrained unbiased estimator $\hat{\boldsymbol{\theta}}_c$ which achieves the CRB, then $\hat{\boldsymbol{\theta}}_c$ must satisfy, for all $\boldsymbol{\theta} \in \mathcal{R}(\mathbf{W})$,

$$\begin{aligned} \hat{\boldsymbol{\theta}}_c &= \boldsymbol{\theta} + \mathbf{W}(\mathbf{W}^T \mathbf{H}^T \mathbf{H} \mathbf{W})^\dagger \mathbf{W}^T \mathbf{H}^T (\mathbf{y} - \mathbf{H}\boldsymbol{\theta}) \\ &= \boldsymbol{\theta} + \mathbf{W}(\mathbf{H}\mathbf{W})^\dagger (\mathbf{H}\mathbf{W})^T (\mathbf{H}\mathbf{W})^T (\mathbf{y} - \mathbf{H}\boldsymbol{\theta}) \\ &= \boldsymbol{\theta} + \mathbf{W}(\mathbf{H}\mathbf{W})^\dagger (\mathbf{y} - \mathbf{H}\boldsymbol{\theta}) \end{aligned} \quad (35)$$

where we have used (17) in the first transition and (18) in the second. Since $\boldsymbol{\theta} \in \mathcal{R}(\mathbf{W})$, one may write $\boldsymbol{\theta} = \mathbf{W}\mathbf{d}$, for some vector \mathbf{d} . Thus

$$\hat{\boldsymbol{\theta}}_c = \boldsymbol{\theta} + \mathbf{W}(\mathbf{H}\mathbf{W})^\dagger \mathbf{y} - \mathbf{W}(\mathbf{H}\mathbf{W})^\dagger \mathbf{H}\mathbf{W}\mathbf{d}. \quad (36)$$

Suppose that

$$\mathcal{R}(\mathbf{W}) \cap \mathcal{N}(\mathbf{H}) = \{\mathbf{0}\}. \quad (37)$$

In this case, it is readily shown that $\mathcal{N}(\mathbf{H}\mathbf{W}) = \mathcal{N}(\mathbf{W})$, and consequently $\mathbf{W}(\mathbf{H}\mathbf{W})^\dagger \mathbf{H}\mathbf{W} = \mathbf{W}$. Thus

$$\hat{\boldsymbol{\theta}}_c = \mathbf{W}(\mathbf{H}\mathbf{W})^\dagger \mathbf{y} \quad (38)$$

is the constrained unbiased estimator achieving the CRB. In other words, $\hat{\boldsymbol{\theta}}_c$ has minimum MSE among all estimators which are unbiased over Θ .

On the other hand, suppose that $\mathcal{R}(\mathbf{W}) \cap \mathcal{N}(\mathbf{H}) \neq \{\mathbf{0}\}$. This implies that the constraints on $\boldsymbol{\theta}$ do not sufficiently compensate for the lack of information in the measurements \mathbf{y} . Indeed, in this case we have $\mathcal{R}(\mathbf{W}\mathbf{W}^T \mathbf{H}^T \mathbf{H}\mathbf{W}\mathbf{W}^T) \not\subseteq \mathcal{R}(\mathbf{W}\mathbf{W}^T)$, and it follows from Theorem 1 that no unbiased estimator exists. These conclusions can also be obtained from [3].

Observe that the expectation of $\hat{\boldsymbol{\theta}}_c$ is

$$E\{\hat{\boldsymbol{\theta}}_c\} = \mathbf{W}(\mathbf{H}\mathbf{W})^\dagger \mathbf{H}\boldsymbol{\theta}. \quad (39)$$

If (37) holds, then (39) is the oblique projection of $\boldsymbol{\theta}$ along $\mathcal{N}(\mathbf{H})$ onto $\mathcal{R}(\mathbf{W})$ [7]. Thus, if $\boldsymbol{\theta} \in \mathcal{R}(\mathbf{W})$, then $\mathbf{W}(\mathbf{H}\mathbf{W})^\dagger \mathbf{H}\boldsymbol{\theta} = \boldsymbol{\theta}$, so that $\hat{\boldsymbol{\theta}}_c$ is indeed unbiased under this constraint. As a generalization, let us seek an estimator whose expectation is given by (39), while removing the constraint on $\boldsymbol{\theta}$ and the assumption (37). If such an estimator existed, then its bias would be given by

$$\mathbf{b}(\boldsymbol{\theta}) = \mathbf{W}(\mathbf{H}\mathbf{W})^\dagger \mathbf{H}\boldsymbol{\theta} - \boldsymbol{\theta} \quad (40)$$

and therefore the matrix \mathbf{A} of (6) would equal

$$\mathbf{A} = \mathbf{W}(\mathbf{H}\mathbf{W})^\dagger \mathbf{H}. \quad (41)$$

Thus, we now seek an unconstrained but biased estimator. To find the minimum MSE estimator whose expectation is (39), we apply (9) of Theorem 1 with $\mathbf{U} = \mathbf{I}$ and \mathbf{A} given by (41). This yields

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \mathbf{W}(\mathbf{H}\mathbf{W})^\dagger \mathbf{H}\boldsymbol{\theta} + \mathbf{W}(\mathbf{H}\mathbf{W})^\dagger \mathbf{H}(\mathbf{H}^T \mathbf{H})^\dagger \mathbf{H}^T (\mathbf{y} - \mathbf{H}\boldsymbol{\theta}) \\ &= \mathbf{W}(\mathbf{H}\mathbf{W})^\dagger \mathbf{H}\boldsymbol{\theta} + \mathbf{W}(\mathbf{H}\mathbf{W})^\dagger \mathbf{H}^T \mathbf{H}^T \mathbf{y} \\ &\quad - \mathbf{W}(\mathbf{H}\mathbf{W})^\dagger \mathbf{H}\mathbf{H}^\dagger \mathbf{H}\boldsymbol{\theta} \\ &= \mathbf{W}(\mathbf{H}\mathbf{W})^\dagger \mathbf{H}^T \mathbf{H}^T \mathbf{y} \\ &= \mathbf{W}(\mathbf{H}\mathbf{W})^\dagger \mathbf{y} = \hat{\boldsymbol{\theta}}_c \end{aligned} \quad (42)$$

where we used (17) and (18) in the second line, and (20) in the last line.

Thus, $\hat{\boldsymbol{\theta}}_c$ of (38) is the approach achieving minimum MSE among all estimators whose expectation is (39). This implies that $\hat{\boldsymbol{\theta}}_c$ is a useful estimator under a wider range of settings than suggested by the unbiased approach. Indeed, among estimators having the required expectation, $\hat{\boldsymbol{\theta}}_c$ is optimal even if $\boldsymbol{\theta}$ does not satisfy the constraint $\boldsymbol{\theta} \in \mathcal{R}(\mathbf{W})$, and, furthermore, its optimality is guaranteed even if the intersection between $\mathcal{R}(\mathbf{W})$ and $\mathcal{N}(\mathbf{H})$ is nontrivial.

REFERENCES

- [1] J. Shao, *Mathematical Statistics*, 2nd ed. New York: Springer, 2003.
- [2] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inform. Theory*, vol. 26, no. 6, pp. 1285–1301, Nov. 1990.
- [3] P. Stoica and B. C. Ng, "On the Cramér–Rao bound under parametric constraints," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 177–179, 1998.
- [4] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 87–90, Jan. 2001.
- [5] Z. Ben-Haim and Y. C. Eldar, "The Cramér–Rao bound in sparse estimation settings," in preparation.
- [6] T. L. Marzetta, "A simple derivation of the constrained multiple parameter Cramér–Rao bound," *IEEE Trans. Signal Process.*, vol. 41, no. 6, pp. 2247–2249, June 1993.
- [7] S. L. Campbell and C. D. Meyer, Jr., *Generalized Inverses of Linear Transformations*. London, U.K.: Pitman, 1979.
- [8] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook* [Online]. Available: <http://matrixcookbook.com/> Feb. 2008