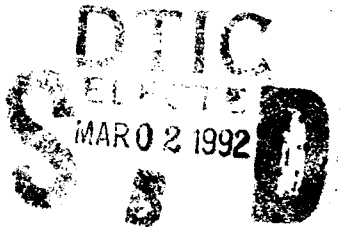
ESD-TR-91-104

Technical Report
890

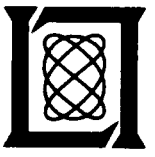
A Cramer-Rao Type Lower Bound for Essentially Unbiased Parameter Estimation



A.O. Hero

3 January 1992

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS



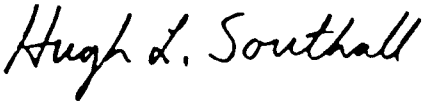
Prepared for the Department of the Air Force
under Contract F19628-90-C-0002.

This report is based on studies performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. The work was sponsored by the Department of the Air Force under Contract F19628-90-C-0002.

The ESD Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER



Hugh L. Southall, Lt. Col., USAF
Chief, ESD Lincoln Laboratory Project Office

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

A CRAMER-RAO TYPE LOWER BOUND FOR ESSENTIALLY UNBIASED PARAMETER ESTIMATION

A.O. HERO
Group 44

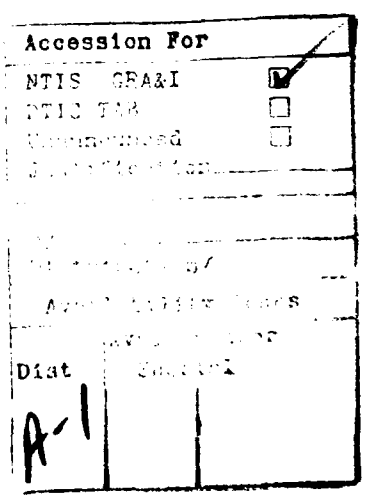
TECHNICAL REPORT 890

3 JANUARY 1992

LEXINGTON MASSACHUSETTS

ABSTRACT

In this report a new Cramer-Rao (CR) type lower bound is derived which takes into account a user-specified constraint on the length of the gradient of estimator bias with respect to the set of underlying parameters. If the parameter space is bounded, the constraint on bias gradient translates into a constraint on the magnitude of the bias itself; the bound reduces to the standard unbiased form of the CR bound for unbiased estimation. In addition to its usefulness as a lower bound that is insensitive to small biases in the estimator, the rate of change of the new bound provides a quantitative bias "sensitivity index" for the general bias-dependent CR bound. An analytical form for this sensitivity index is derived which indicates that small estimator biases can make the new bound significantly less than the unbiased CR bound when important but difficult-to-estimate nuisance parameters exist. This implies that the application of the CR bound is unreliable for this situation due to severe bias sensitivity. As a practical illustration of these results, the problem of estimating elements of the 2×2 covariance matrix associated with a pair of independent identically distributed (IID) zero-mean Gaussian random sequences is presented.



ACKNOWLEDGMENTS

I would like to thank Larry Horowitz and John Jayne for their numerous insightful comments and suggestions offered during the preparation of this report.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

LIST OF ILLUSTRATIONS

1. INTRODUCTION

This report deals with the problem of bounding the variance of parameter estimators under the constraint of small bias. In multiple parameter estimation problems, the variance of the estimates of a single parameter can appear to violate the unbiased Cramer-Rao (CR) lower bound due to the presence of extremely small biases; that is, the actual variance of the estimator is lower than that predicted by the CR bound (for a particularly simple example of bound violation see Stoica and Moses [1]). This indicates that the unbiased CR bound may be an unreliable predictor of performance even when biases are otherwise insignificant. On the other hand, the application of the general CR bound for biased estimators depends on knowledge of the particular bias of the estimator; in particular, it is necessary to know the gradient of the bias with respect to the vector of unknown parameters. However, the precise evaluation of estimator bias is frequently difficult and not of direct interest when bias is small. Furthermore, for performance comparisons, a useful lower bound should apply to the entire class of estimators with acceptably small bias.

In this report a new CR-type lower bound is derived which takes into account a user-specified constraint on the length of the gradient of estimator bias with respect to the vector of unknown parameters. Alternatively, the bound takes into account a constraint on the actual estimator bias as the unknown parameters range over a specified ellipsoid. This bound is uniform with respect to a special class of biased estimators: those whose bias gradient has a length of less than or equal to $\delta < 1$. In addition to its usefulness as a bias-insensitive lower bound, the slope of the new bound as a function of δ provides a characterization of the bias sensitivity of the general CR bound on estimator variance. For a given estimation problem, an overly large magnitude of bias sensitivity provides a warning against use of the unbiased CR bound. If an upper bound on the bias gradient of the estimator is specified, our lower bound on estimator variance can subsequently be applied.

The specific results developed herein follow.

1. A geometric point of view provides some insight into the behavior of the general CR bound;
2. A functional minimization is performed to arrive at the new bound based on the Fisher information matrix;
3. Results of an asymptotic analysis of the new bound as bias $\rightarrow 0$ indicate important factors controlling bias sensitivity of general CR bounds;
4. The asymptotic analysis suggests a "bias-sensitivity index," which is the slope of the new bound as a function of the length δ of the bias gradient. This index indicates the impact of difficult-to-estimate "nuisance" parameters on the magnitude of the general CR bound;
5. The form of the new bound is suggestive of "superefficient," essentially unbiased estimator structures which could outperform absolutely unbiased estimators in the sense of mean-squared-error;

6. Sensitivity results are obtained for estimation of the elements of the 2×2 covariance matrix associated with a pair of independent identically distributed (IID) zero-mean Gaussian random sequences.

The report is organized as follows. Section 2.1 gives the notation. Section 2.2 is a summary of useful vector and matrix relations. Section 3 defines the class of essentially unbiased estimators. Section 4 is a geometric interpretation of the CR bound in terms of its bias dependency. The new bound is derived in Section 5. In Section 6 the slope of the bound is derived and an asymptotic approximation to the new bound is given. Section 7 is a discussion of the results and an interpretation of the new bound in terms of the joint "estimability" of the multiple parameters. Finally, Section 8 applies the new bound to covariance estimation for a pair of IID Gaussian sequences.

2. PRELIMINARIES

2.1 Notation

General notational conventions are as follows. If $\{(y, z) : z = g(y)\}$ is the graph of a function, then g denotes the function and $g(y)$ denotes a functional evaluation at the point y. An exception to this convention occurs when g is used to denote $g(y)$ for compactness of notation. In general, an uppercase letter near the end of the alphabet, e.g., X, denotes a random variable, random vector, or random process and the corresponding lowercase, e.g., x, denotes its realization. An uppercase letter near the beginning of the alphabet, e.g., F, denotes a matrix. The ith–jth element of a matrix F is denoted F_{ij} or $((F))_{ij}$. An underbar denotes a column vector, e.g., $\underline{d}$, and a superscript T denotes the transpose, e.g., $\underline{d}^T$. For vectors, subscripts index over the elements, e.g., $\underline{\theta} = [\theta_1, \ldots, \theta_n]^T$, while superscripts discriminate between different vector quantities, e.g., $\underline{\theta}^o = [\theta_1^o, \ldots, \theta_n^o]^T$. The gradient operator $\nabla_{\underline{\theta}}$ is, by convention, a row vector of partial derivatives $[\frac{\partial}{\partial \theta_1}, \ldots, \frac{\partial}{\partial \theta_n}]$. For convenience, when there is no risk of confusion the simplified notation $\nabla g(\underline{\theta}) \stackrel{\text{def}}{=} \nabla_{\underline{u}} g(\underline{u})|_{\underline{u}=\underline{\theta}}$ will be used for the gradient of a scalar or vector valued function $g(\underline{u})$ at a point $\underline{u} = \underline{\theta}$.

Some particularly useful definitions:

- X : the generic observation; e.g., a set of snapshots of the data outputs from multiple sensors.
- x : a realization of X.
- Θ : the n-dimensional parameter space.
- $\underline{\theta}$, $\underline{u}$: parameter vectors.
- $\|\underline{d}\|$: the Euclidean norm of vector $\underline{d}$, $\|\underline{d}\| = \sqrt{\underline{d}^T \underline{d}}$.
- $f(x; \underline{\theta}^o)$: the probability density function of X evaluated at $X = x$, $\underline{\theta} = \underline{\theta}^o$.
- $\hat{\underline{\theta}} = \hat{\underline{\theta}}(X)$: an estimator of $\underline{\theta}$.
- $\underline{m}(\underline{\theta})$: the mean vector $E_{\underline{\theta}}(\hat{\underline{\theta}})$ of $\hat{\underline{\theta}}$, where $\underline{\theta}$ is the true underlying parameter.
- $\underline{b}(\underline{\theta})$: the bias, $\underline{m}(\underline{\theta}) - \underline{\theta}$, of $\hat{\underline{\theta}}$.
- $\text{cov}_{\underline{\theta}}(\hat{\underline{\theta}})$: the $n \times n$ covariance matrix of $\hat{\underline{\theta}}$, $E_{\underline{\theta}}[(\hat{\underline{\theta}} - m(\underline{\theta}))(\hat{\underline{\theta}} - m(\underline{\theta}))^T]$.
- $\text{var}_{\underline{\theta}}(\hat{\theta}_1)$: the variance of $\hat{\theta}_1$.
- $F(\underline{\theta})$: the $n \times n$ Fisher information matrix associated with estimators of $\underline{\theta}$. This matrix will always be assumed to have bounded elements and to be invertible with a bounded inverse $F^{-1}(\underline{\theta})$ over any domain of $\underline{\theta}$ of interest.

- a, $\underline{b}$, F_s : the Fisher information for θ_1, the information coupling between θ_1 and $\theta_2, \ldots, \theta_n$, and the $(n-1) \times (n-1)$ Fisher information matrix for $\theta_2, \ldots, \theta_n$. Note that $F = \begin{bmatrix} a & \underline{b}^T \\ \underline{b} & F_s \end{bmatrix}$, where a is positive and F_s is invertible by assumption.
- $l(\underline{\theta})$: the log-likelihood function, $\ln f(X;\underline{\theta})$.
- $\bar{l}(\underline{\theta}, \underline{\theta}^o)$: the mean log-likelihood (ambiguity) function, $E_{\underline{\theta}^o}[\ln f(X;\underline{\theta})]$.
- δ : a user-specified upper bound on the length of the bias-gradient vector.
- $B_{b_1}(\underline{\theta})$: the general CR lower bound on $\mathrm{var}_{\underline{\theta}}(\hat{\theta_1})$ for estimators with bias $b_1(\underline{\theta})$ (28).
- $B(\underline{\theta}, \delta)$: the new lower bound on $\mathrm{var}_{\underline{\theta}}(\hat{\theta_1})$ for estimators with bias $b_1(\underline{\theta})$ such that $\|\nabla b_1\| \leq \delta$ (56).
- $\Delta B(\underline{\theta}, \delta)$: the normalized difference between the unbiased CR bound and the new bound, $\frac{B(\underline{\theta},0) - B(\underline{\theta},\delta)}{B(\underline{\theta},0)}$.
- $\underline{d}^I_{min}$: the minimizing bias-gradient vector which characterizes $B(\underline{\theta}, \delta)$ (65).
- λ : a scaling constant determined by the solution to the constraint equation on the bias gradient (57).
- η : the sensitivity index of the general CR bound, derived from $B(\underline{\theta}, \delta)$.

2.2 Identities

Some vector and matrix identities to be used in the sequel are given here.

- Let A be an invertible $m \times m$ matrix which has the partition

$$A = \begin{bmatrix} a & \underline{c}^T \\ \underline{c} & A_s \end{bmatrix} \quad , \tag{1}$$

where a is a nonzero scalar, $\underline{c}$ is an $(m-1) \times 1$ vector, and A_s is an $(m-1) \times (m-1)$ invertible matrix. The inverse of A can be expressed in terms of the partition elements a, $\underline{c}$, and A_s ([2], Theorem 8.2.1):

$$A^{-1} = \begin{bmatrix} \frac{1}{a - \underline{c}^T A_s^{-1} \underline{c}} & -\underline{c}^T A_s^{-1} \frac{1}{a - \underline{c}^T A_s^{-1} \underline{c}} \\ -A_s^{-1} \underline{c} \frac{1}{a - \underline{c}^T A_s^{-1} \underline{c}} & A_s^{-1} + A_s^{-1} \underline{c} \underline{c}^T A_s^{-1} \frac{1}{(a - \underline{c}^T A_s^{-1} \underline{c})} \end{bmatrix} \quad . \tag{2}$$

- Let A be an $m \times m$ invertible matrix and let U and V be $m \times k$ matrices, respectively. If the matrix $[A + UV^T]$ is nonsingular, the Sherman-Morrison-Woodbury identity gives the inverse as ([3], Section 0.7.4)

$$[A + UV^T]^{-1} = A^{-1} - A^{-1}U[I + V^T A^{-1} U]^{-1} V^T A^{-1} \quad . \tag{3}$$

- For the gradient of quadratic forms and inner products with respect to a vector, we have the following identities ([2], Section 10.8):

$$\nabla_{\underline{x}} \underline{x}^T A \underline{x} = 2\underline{x}^T A, \quad (A \text{ symmetric}) \tag{4}$$

$$\nabla_{\underline{x}} \underline{y}^T \underline{x} = \underline{y}^T \quad . \tag{5}$$

- If A and B are symmetric matrices which possess identical eigenvectors, then $AB = BA$ ([3], Theorem 4.1.6). If, in addition, A and B are positive-definite, then AB is positive-definite ([4], p. 350, Exercise 23).
- Let A, B, and C be matrices and assume B is symmetric and positive-definite. If λ is a scalar, a singular value decomposition of B establishes the following for positive integer k:

$$\frac{d}{d\lambda} \underline{x}^T A[I + \lambda B]^{-k} C \underline{x} = -k \underline{x}^T AB[I + \lambda B]^{-(k+1)} C \underline{x} \quad . \tag{6}$$

3. ESSENTIALLY UNBIASED ESTIMATORS

This section deals with the following general setup of our estimation problem. Let an observation X have the probability density function $f(x;\underline{\theta})$, where

$$\underline{\theta} = (\theta_1, \ldots, \theta_n)^T \tag{7}$$

is a real, nonrandom parameter vector residing in an open subset $\Theta = \Theta_1 \times \cdots \times \Theta_n$ of the n-dimensional space $\mathbb{R}^n$. Suitably modified, the theory herein can be applied to more general Θ (e.g., subsets of $\mathbb{R}^n$ which are defined by differentiable functional inequality and equality constraints) by replacing the Fisher information matrix $F(\underline{\theta})$ (21), used throughout the report, with a reduced-rank Fisher matrix [5]. We use the conventional notation for expectation of a random variable Z, with respect to $f(x;\underline{\theta})$,

$$E_{\underline{\theta}}[Z] \stackrel{\text{def}}{=} \int_{\text{supp}f(\bullet;\underline{\theta})} Z(x) f(x;\underline{\theta}) dx \quad , \tag{8}$$

where $\text{supp}f(\bullet;\underline{\theta}) = \{x : f(x;\underline{\theta}) > 0\}$ is the support of the probability density function (PDF) f for fixed $\underline{\theta}$.

Let $\hat{\theta}_1$ be an estimator of θ_1 with mean $E_{\underline{\theta}}(\hat{\theta}_1) = m_1(\underline{\theta})$ and bias

$$b_1(\underline{\theta}) \stackrel{\text{def}}{=} m_1(\underline{\theta}) - \theta_1. \tag{9}$$

The bias is said to be "globally removable" if b_1 is a constant independent of $\underline{\theta}$ and "locally removable over a region $\underline{\theta} \in \mathcal{D}$" if b_1 is constant over the region $\mathcal{D}$. By convention, when we refer to a "region $\mathcal{D}$ in Θ" we mean a nonempty, open, connected subset of Θ. The estimator $\hat{\theta}_1$ is said to be globally (locally) unbiased if the bias is globally (locally) removable. In the sequel we will address the problem of lower bounding the variance of $\hat{\theta}_1$, given that $\hat{\theta}_1$ is "essentially unbiased" in the sense that for a prespecified constant $\delta \in [0, 1]$

$$\|\nabla b_1(\underline{\theta})\|^2 = \|(\frac{\partial b_1(\underline{\theta})}{\partial \theta_1}, \ldots, \frac{\partial b_1(\underline{\theta})}{\partial \theta_n})\|^2 \leq \delta^2 \quad , \quad \forall \underline{\theta} \in \Theta \quad . \tag{10}$$

Bias gradients contained in the constraint set $\{\underline{d} : \underline{d}^T \underline{d} \leq \delta^2\}$ for all values of $\underline{\theta}$ are called admissible bias gradients. Note, however, that vector functions $\underline{d}(\underline{\theta})$ exist which satisfy the constraint in (10) for all $\underline{\theta}$ but are not valid gradient functions, and therefore are not admissible bias gradients. The importance of the bias gradient in (10) in lower bounding the variance of $\hat{\theta}_1$ will be seen in Equation (28).

The restriction $\delta \in [0,1]$ in (10) is sufficiently general, since for $\delta \geq 1$ the gradient can be taken as $\nabla b_1 = [-1, 0, \ldots, 0]$. This bias gradient corresponds to the trivial estimator $\hat{\theta}_1 = constant$, which has zero variance. Observe that for Inequality (10) to be well defined the bias must be differentiable. Ibragimov and Has'minskii [6], Chapter 1, Lemma 7.2, shows that under essentially the same regularity conditions which guarantee the existence of the Fisher information, the bias exists and is differentiable regardless of the estimator $\hat{\theta}_1$. Hence, the differentiability property is only dependent on the underlying distribution of the observations, not on the particular form of the estimator. Differentiability is therefore not a restrictive assumption in characterizing classes of biased estimators for a particular estimation problem.

More significantly, note that Inequality (10) is a constraint on the rate of change of the bias and not on the bias itself. However, the definition of an acceptable range of the bias is typically more natural than the definition of an acceptable range of the bias gradient; this issue will be discussed presently. For sufficiently small δ, (10) implies that in a practical sense the bias is locally removable over any prespecified finite region; therefore the estimator is locally unbiased. On the other hand, for bounded parameters (10) can be related to global unbiasedness.

The bound presented here is applicable if, for example, a user is interested in a lower bound on estimator variance which applies to a class of estimators permitted to have small, perhaps "acceptable," biases over a parameter range of interest. As mentioned previously, it is generally more natural to specify an acceptable range of biases than an acceptable range of bias gradients. We will now show how the former can be converted to the latter. Assume that the user specifies an ellipsoid of parameters centered at some parameter $\underline{\theta} = \underline{\nu}$ and a maximal allowable variation in the bias over the ellipsoid. This requirement is stated mathematically as

$$|b_1(\underline{\theta}) - b_1(\underline{\theta}^o)| \leq \gamma, \quad \forall \underline{\theta}, \underline{\theta}^o \in \{\underline{u} : \|\mathrm{diag}(K_i)(\underline{u} - \underline{\nu})\| \leq 1\} \quad , \tag{11}$$

where $\mathrm{diag}(K_i)$ is a diagonal matrix of positive constants and the user-specified quantities $K_1, \ldots, K_n$ and γ determine the ellipsoid and the maximal allowable bias variation, respectively. The ellipsoid of (11) also reflects the user's choice of units to represent each of the parameters.

To standardize the analysis, it is convenient to normalize the ellipsoid to a sphere via a coordinate transformation (scaling) of the parameters. This coordinate transformation is implemented by premultiplying parameter vectors in the original coordinates by the diagonal matrix $\mathrm{diag}(K_i)$ in (11). The reader may verify that the result of this transformation is to replace the quantities $[\underline{\theta}, \underline{\theta}^o, \underline{\nu}, b_1(\underline{\theta}), b_1(\underline{\theta}^o), \gamma]$ (which are parameterized in the original coordinates) in (11) with the quantities $[\mathrm{diag}(K_i^{-1})\underline{\theta}, \mathrm{diag}(K_i^{-1})\underline{\theta}^o, \mathrm{diag}(K_i^{-1})\underline{\nu}, K_1^{-1}b_1(\underline{\theta}), K_1^{-1}b_1(\underline{\theta}^o), K_1^{-1}\gamma]$ (where $[\underline{\theta}, \underline{\theta}^o, \underline{\nu}, b_1(\underline{\theta}), b_1(\underline{\theta}^o), \gamma]$ are parameterized in the new coordinates). It is then seen that (11) becomes equivalent to a bias constraint over a displaced unit sphere in the new coordinates

$$|b_1(\underline{\theta}) - b_1(\underline{\theta}^o)| \leq \gamma, \quad \forall \underline{\theta}, \underline{\theta}^o \in \{\underline{u} : \|\underline{u} - \underline{\nu}\| \leq 1\} \quad . \tag{12}$$

Throughout the rest of the report it is assumed that the user ellipse has been normalized to the standard spherical region (12).

The following proposition translates the user constraint on the bias (12) to the constraint on the bias gradient (10).

Proposition 1 *Let $b_1(\underline{\theta})$ be a differentiable scalar (bias) function, with (bias) gradient ∇b_1, over the spherical region $\underline{\theta} \in \mathcal{S}^n \stackrel{\text{def}}{=} \{\underline{u} : \|\underline{u} - \underline{\nu}\| \leq 1\}$. Then the set of n dimensional vectors*

$$\mathcal{D}_\gamma \stackrel{\text{def}}{=} \{\underline{d} : \|\underline{d}\| \leq \gamma/2\} \tag{13}$$

defines the largest region in $\mathbf{R}^n$ containing gradients ∇b_1 for which $b_1(\bullet)$ satisfies the requirement (12), in the sense that:

1. *If $\nabla b_1 \in \mathcal{D}_\gamma$, $\underline{\theta} \in \mathcal{S}^n$, then $b_1(\bullet)$ satisfies the requirement (12).*
2. *If $\mathcal{D}'$ is such that $\mathcal{D}_\gamma \subset \mathcal{D}'$ (strictly proper subset) then $\mathcal{D}'$ contains a vector ∇b_1, $\underline{\theta} \in \mathcal{S}^n$, which is the gradient of a function $b_1(\bullet)$ that violates the requirement (12).*

Proof of Proposition 1. Because $b_1(\underline{\theta})$ is differentiable and the sphere $\mathcal{S}^n$ is a convex set, we have from Rudin [7], Theorem 9.19,

$$|b_1(\underline{\theta}) - b_1(\underline{\theta}^o)| \leq M\|\underline{\theta} - \underline{\theta}^o\|, \quad \forall \underline{\theta}, \underline{\theta}^o \in \mathcal{S}^n \quad , \tag{14}$$

where M is an upper bound on $\|\nabla b_1\|$ over $\mathcal{S}^n$. Because the maximal distance between any two vectors $\underline{\theta}, \underline{\theta}^o$ in the unit sphere $\mathcal{S}^n$ is 2, the inequality (14) can be replaced by

$$|b_1(\underline{\theta}) - b_1(\underline{\theta}^o)| \leq 2M, \quad \forall \underline{\theta}, \underline{\theta}^o \in \mathcal{S}^n \quad . \tag{15}$$

Now, if $\nabla b_1 \in \mathcal{D}_\gamma$, then $\|\nabla b_1\| \leq \gamma/2$ so that, using $M = \gamma/2$ in (15),

$$|b_1(\underline{\theta}) - b_1(\underline{\theta}^o)| \leq \gamma, \quad \forall \underline{\theta}, \underline{\theta}^o \in \mathcal{S}^n \quad , \tag{16}$$

which proves Assertion 1 of the proposition. However, if $\mathcal{D}'$ contains $\mathcal{D}_\gamma$ as a proper subset, a constant vector $\underline{d} \in \mathcal{D}'$ exists such that $\|\underline{d}\| > \gamma/2$. Let the gradient vector ∇b_1 be defined as the constant $\underline{d}^T$. Because $\nabla b_1 = \underline{d}^T$ is independent of $\underline{\theta}$, we have $b_1(\underline{\theta}) = \underline{d}^T\underline{\theta} + C$ for some constant C. Consider the two vectors

$$\underline{\theta} = \underline{\nu} + \frac{1}{\|\underline{d}\|}\underline{d} \quad , \tag{17}$$

$$\underline{\theta}^o = \underline{\nu} - \frac{1}{\|\underline{d}\|}\underline{d} \quad . \tag{18}$$

These vectors are on the boundary of the sphere $\mathcal{S}^n$ because $\|\underline{\theta} - \underline{\theta}^o\| = 2$; furthermore,

$$\begin{aligned} |b_1(\underline{\theta}) - b_1(\underline{\theta}^o)| &= |\underline{d}^T[\underline{\theta} - \underline{\theta}^o]| \\ &= |\underline{d}^T \left[\frac{2}{\|\underline{d}\|}\underline{d}\right]| = \|\underline{d}\|2 \\ &> (\gamma/2)2 = \gamma \quad , \end{aligned} \tag{19}$$

so that the requirement (12) is violated. This establishes Assertion 2 and completes the proof of Proposition 1.

Proposition 1 asserts that to satisfy the bias requirement (12), the constraint $\|\nabla b_1\| < \delta$, with $\delta = \gamma/2$, is the weakest possible gradient constraint which satisfies that requirement and is independent of $\underline{\theta}$. It must be emphasized that before the proposition can be used the user ellipsoid $\{\underline{\theta} : \|\mathrm{diag}(K_1)[\underline{\theta} - \underline{\nu}]\| \leq 1\}$ has to be transformed to a sphere via the coordinate transformation described in the paragraph following requirement (11).

It is important to note that Proposition 1 does not address the existence of estimators having the bias function b_1 prescribed by Assertion 2 and violating the requirement (12). Specifically, in proving Assertion 2 we produced a function b_1 and its gradient ∇b_1, which violate constraints on function variation and constraints on gradient magnitude, respectively. While this shows a certain topological equivalence between these two types of constraints, there is no guarantee that the function b_1 is the bias $E_{\underline{\theta}}[\underline{\hat{\theta}}] - \underline{\theta}$ of any physically realizable estimator $\underline{\hat{\theta}}(X)$.

4. INTERPRETATIONS OF THE CR BOUND

Define the vector of estimators $\underline{\hat{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_n)^T$ of parameters in the vector $\underline{\theta}$. Assume that the PDF of the observations is "regular" ([6], Chapter 1, Section 7) and that $E_{\underline{\theta}}[\hat{\theta}_i^2]$ is bounded, $i = 1, \ldots, n$; then the gradient of the mean $m_i(\underline{\theta}) = E_{\underline{\theta}}[\hat{\theta}_i]$ exists, $i = 1, \ldots, n$, and is continuous, and the covariance matrix of $\underline{\hat{\theta}}$ satisfies the matrix CR lower bound $B_{\underline{b}}(\underline{\theta})$ ([6], Chapter 1, Theorem 7.3):

$$\mathrm{cov}_{\underline{\theta}}(\underline{\hat{\theta}}) \geq B_b(\underline{\theta}) = [\nabla \underline{m}(\underline{\theta})] F^{-1}(\underline{\theta}) [\nabla \underline{m}(\underline{\theta})]^T \quad . \tag{20}$$

In (20), $F = F(\underline{\theta})$ is the nonsingular $n \times n$ Fisher information matrix

$$F(\underline{\theta}) \stackrel{\mathrm{def}}{=} E_{\underline{\theta}}[\nabla_{\underline{u}} \ln f(X; \underline{u})|_{\underline{u}=\underline{\theta}}]^T [\nabla_{\underline{u}} \ln f(X; \underline{u})|_{\underline{u}=\underline{\theta}}] \quad , \tag{21}$$

and $\nabla \underline{m} = \nabla \underline{m}(\underline{\theta})$ is an $n \times n$ matrix whose rows are the gradient vectors ∇m_i, $i = 1, \ldots, n$. Under additional assumptions ([6], Chapter 1, Lemma 8.1) the Fisher information matrix is equivalent to the Hessian, or "curvature," matrix of the mean of $\ln f(X; \underline{u})$:

$$F(\underline{\theta}) = -E_{\underline{\theta}} \nabla_{\underline{u}}^T \nabla_{\underline{u}} \ln f(X; \underline{u})|_{\underline{u}=\underline{\theta}} = -\nabla_{\underline{u}}^T \nabla_{\underline{u}} E_{\underline{\theta}} \ln f(X; \underline{u})|_{\underline{u}=\underline{\theta}} \quad . \tag{22}$$

If the vector estimator $\underline{\hat{\theta}}$ is locally unbiased, then $\nabla \underline{m}(\underline{\theta}) = I$ and the lower bound (20) becomes the unbiased CR bound

$$\mathrm{cov}_{\underline{\theta}}(\underline{\hat{\theta}}) \geq F^{-1}(\underline{\theta}) \quad . \tag{23}$$

Comparison between the right-hand sides of the general CR bound (20) and the unbiased CR bound (23) suggests defining the biased Fisher information matrix F_b

$$F_b(\underline{\theta}) \stackrel{\mathrm{def}}{=} [\nabla \underline{m}(\underline{\theta})]^{-T} F(\underline{\theta}) [\nabla \underline{m}(\underline{\theta})]^{-1} \quad , \tag{24}$$

where it has been assumed that the matrix $\nabla \underline{m}(\underline{\theta})$ is invertible. With the definition (24) the general CR bound (20) becomes

$$\mathrm{cov}_{\underline{\theta}}(\underline{\hat{\theta}}) \geq F_b^{-1} \quad . \tag{25}$$

Because $\mathrm{var}_{\underline{\theta}}(\hat{\theta}_1)$ is the (1,1) element of $\mathrm{cov}_{\underline{\theta}}(\underline{\hat{\theta}})$, the matrix bound (20) gives the following bound on the variance of $\hat{\theta}_1$:

$$\mathrm{var}_{\underline{\theta}}(\hat{\theta}_1) \geq \underline{e}_1^T [\nabla \underline{m}(\underline{\theta})] F^{-1}(\underline{\theta}) [\nabla \underline{m}(\underline{\theta})]^T \underline{e}_1 \quad , \tag{26}$$

where $\underline{e}_1$ is the unit (column) vector

$$\underline{e}_1 = [1, 0, \ldots, 0]^T \quad . \tag{27}$$

Note that, concerning the CR bound on $\hat{\theta}_1$, only the first row of $\nabla \underline{m}$ is important. The following, denoted the general CR bound in the sequel, is equivalent to (26):

$$\begin{aligned} \mathrm{var}_{\underline{\theta}}(\hat{\theta}_1) \geq B_{b_1}(\underline{\theta}) &\stackrel{\mathrm{def}}{=} [\nabla m_1(\underline{\theta})] F^{-1}(\underline{\theta}) [\nabla m_1(\underline{\theta})]^T \\ &= [\underline{e}_1 + \nabla b_1(\underline{\theta})^T]^T F^{-1}(\underline{\theta}) [\underline{e}_1 + \nabla b_1(\underline{\theta})^T] \quad , \end{aligned} \tag{28}$$

where, in the second equality of (28), the relation (9) has been used.

Observe that a lower bound on the mean-squared error (MSE) of $\hat{\theta}_1$, $MSE_{\underline{\theta}}(\hat{\theta}_1) \stackrel{\mathrm{def}}{=} E_{\underline{\theta}}[\hat{\theta}_1 - \theta_1]^2$, can be obtained from the variance lower bound (28) by using the relation $MSE_{\underline{\theta}}(\hat{\theta}_1) = \mathrm{var}_{\underline{\theta}}(\hat{\theta}_1) + b_1^2(\underline{\theta})$:

$$MSE_{\underline{\theta}}(\hat{\theta}_1) \geq B_{b_1}(\underline{\theta}) + b_1^2(\underline{\theta}) \quad . \tag{29}$$

Any lower bound on the variance is also a lower bound on the MSE, as the second term on the right of (29) is non-negative.

For locally unbiased estimators of θ_1, the gradient vector $\nabla m_1(\underline{\theta})$ is the unit row vector $\underline{e}_1$ and the CR bound is the (1,1) element of the inverse Fisher matrix

$$\mathrm{var}_{\underline{\theta}}(\hat{\theta}_1) \geq \underline{e}_1^T F^{-1}(\underline{\theta}) \underline{e}_1 \quad , \tag{30}$$

which will be called the unbiased form of the CR bound on $\hat{\theta}_1$.

The following interpretations are helpful in understanding the influence of bias on the CR bound.

4.1 CR Bounds and Unbiased Estimation of the Mean

The general matrix CR bound (20) on the covariance of a biased estimator $\hat{\underline{\theta}}$ of $\underline{\theta}$ can be equivalently interpreted as an unbiased CR bound, just as in (23), on the covariance of $\hat{\underline{\theta}}$ viewed as a "differentially unbiased" estimator of its mean $\underline{m}(\underline{\theta})$.

Fix a point $\underline{\theta}^o \in \Theta$. An estimator $\hat{\underline{\theta}}$ with mean $\underline{m}(\underline{\theta})$ is defined to be differentially unbiased at the point $\underline{\theta} = \underline{\theta}^o$ if $\nabla \underline{m}(\underline{\theta}^o) = I$, where I is the $n \times n$ identity matrix. Note that, under the assumption of differentiability of $\underline{m}(\underline{\theta})$, a locally unbiased estimator is necessarily differentially unbiased. As $\underline{m}(\underline{\theta})$ is a differentiable function of $\underline{\theta}$,

$$\underline{m}(\underline{\theta}) - \underline{m}(\underline{\theta}^o) = \nabla \underline{m}(\underline{\theta}^o)(\underline{\theta} - \underline{\theta}^o) + o(\|\underline{\theta} - \underline{\theta}^o\|) \tag{31}$$

so that $\underline{m}(\underline{\theta})$ is a locally linear transformation in the neighborhood of $\underline{\theta}^o$. Assuming the matrix $\nabla \underline{m}(\underline{\theta}^o)$ to be invertible, this permits a local reparameterization of Θ by the values $\underline{\nu}$ taken on by the linear approximation to the function $\underline{m}(\underline{\theta})$ over this neighborhood:

$$\underline{\nu} = \underline{\nu}(\underline{\theta}) \stackrel{\text{def}}{=} \nabla m(\underline{\theta}^o)(\underline{\theta} - \underline{\theta}^o) + \underline{m}(\underline{\theta}^o) \quad , \tag{32}$$

and

$$\underline{\theta} = \underline{\theta}(\underline{\nu}) \stackrel{\text{def}}{=} [\nabla \underline{m}(\underline{\theta}^o)]^{-1}(\underline{\nu} - \underline{m}(\underline{\theta}^o)) + \underline{\theta}^o \quad . \tag{33}$$

Using (33) and the chain rule of vector differentiation,

$$\begin{aligned} \nabla_{\underline{\nu}} E_{\underline{\theta}(\underline{\nu})}[\hat{\underline{\theta}}] &= \nabla_{\underline{\nu}} \underline{m}(\underline{\theta}(\underline{\nu})) \\ &= \nabla_{\underline{\nu}} \underline{m}([\nabla \underline{m}(\underline{\theta}^o)]^{-1}(\underline{\nu} - \underline{m}(\underline{\theta}^o)) + \underline{\theta}^o) \\ &= \nabla_{\underline{u}} \underline{m}(\underline{u})|_{\underline{u}=([\nabla \underline{m}(\underline{\theta}^o)]^{-1}(\underline{\nu}-\underline{m}(\underline{\theta}^o))+\underline{\theta}^o)} [\nabla \underline{m}(\underline{\theta}^o)]^{-1} \\ &= \nabla \underline{m}(\underline{\theta})[\nabla \underline{m}(\underline{\theta}^o)]^{-1} \quad . \end{aligned} \tag{34}$$

When $\underline{\theta} = \underline{\theta}^o$, the last line of the above is the identity matrix so that $\hat{\underline{\theta}}$ is a differentially unbiased estimator of the transformed parameter $\underline{\nu}$ at the point $\underline{\nu} = \underline{\nu}(\underline{\theta}^o) = \underline{m}(\underline{\theta}^o)$.

Because $\hat{\underline{\theta}}$ is a differentially unbiased estimator of $\underline{\nu} = \underline{m}(\underline{\theta}^o)$, the CR bound on the covariance of $\hat{\underline{\theta}}$ at $\underline{\theta}^o$ is given by (23) with Fisher matrix $F'(\underline{\theta}^o)$

$$F'(\underline{\theta}^o) = E_{\underline{\theta}^o}[\nabla_{\underline{\nu}} \ln f(X; \underline{\theta}(\underline{\nu}))|_{\underline{\nu}=\underline{m}(\underline{\theta}^o)}]^T [\nabla_{\underline{\nu}} \ln f(X; \underline{\theta}(\underline{\nu}))|_{\underline{\nu}=\underline{m}(\underline{\theta}^o)}] \quad . \tag{35}$$

Use of the relation (33) and application of the chain rule yields

$$\begin{aligned}\nabla_{\underline{\nu}} \ln f(X;\underline{\theta}(\underline{\nu}))|_{\underline{\nu}=\underline{m}(\underline{\theta}^o)} &= \nabla_{\underline{\nu}} \ln f(X;[\nabla \underline{m}(\underline{\theta}^o)]^{-1}(\underline{\nu}-\underline{m}(\underline{\theta}^o))+\underline{\theta}^o))|_{\underline{\nu}=\underline{m}(\underline{\theta}^o)} \\ &= \nabla_{\underline{u}} \ln f(X;\underline{u})|_{\underline{u}=\underline{\theta}^o} [\nabla \underline{m}(\underline{\theta}^o)]^{-1} \quad . \end{aligned} \tag{36}$$

Substitution of the above into (35) yields the form

$$\begin{aligned} F'(\underline{\theta}^o) &= [\nabla \underline{m}(\underline{\theta}^o)]^{-T} E_{\underline{\theta}^o}[\nabla_{\underline{u}} \ln f(X;\underline{u})|_{\underline{u}=\underline{\theta}^o}]^T [\nabla_{\underline{u}} \ln f(X;\underline{u})|_{\underline{u}=\underline{\theta}^o}] [\nabla \underline{m}(\underline{\theta}^o)]^{-1} \\ &= F_b(\underline{\theta}^o) \quad . \end{aligned} \tag{37}$$

Hence, the Fisher matrix $F'(\underline{\theta})$ (37) for (differentially) unbiased estimation of $\underline{m}(\underline{\theta})$ is identical to the biased Fisher matrix $F_b(\underline{\theta})$ (24) for biased estimation of $\underline{\theta}$.

We can therefore conclude that there are two equivalent ways of interpreting the biased Fisher information, alternately the CR bound (20): a measure of the accuracy with which the mean $\underline{m}(\underline{\theta})$ of $\hat{\underline{\theta}}$ can be estimated without bias; and a measure of the accuracy with which the parameter $\underline{\theta}$ can be estimated with bias.

4.2 CR Bounds and Sensitivity of the Ambiguity Function

Define the log-likelihood function l

$$l(\underline{\theta}) \stackrel{\text{def}}{=} \ln f(X;\underline{\theta}) \quad , \tag{38}$$

and the ambiguity function

$$\bar{l}(\underline{u},\underline{\theta}) \stackrel{\text{def}}{=} E_{\underline{\theta}}[\ln f(X;\underline{u})] \quad . \tag{39}$$

For a fixed value $\underline{\theta} = \underline{\theta}^o$ of the parameter, the ambiguity function is simply the mean log-likelihood function. Although the arguments $\underline{u}$ and $\underline{\theta}$ reside in the same space Θ, it is useful to distinguish between the search parameter $\underline{u}$ and the true parameter $\underline{\theta}$.

Two important properties of the ambiguity function are 1. $\bar{l}(\underline{u},\underline{\theta}^o)$ has a global maximum over $\underline{u}$ at $\underline{u} = \underline{\theta}^o$ and consequently, if $\nabla_{\underline{u}}\bar{l}(\underline{u},\underline{\theta}^o)|_{\underline{u}=\underline{\theta}^o}$ exists, $\bar{l}(\underline{u},\underline{\theta}^o)$ has a stationary point at $\underline{u} = \underline{\theta}^o$: $\nabla_{\underline{u}}\bar{l}(\underline{u},\underline{\theta}^o)|_{\underline{u}=\underline{\theta}^o} = 0$, and 2. the sharpness of this maximum is related to the Fisher information matrix $F(\underline{\theta}^o)$. Due to the latter property, the general CR bound (26) can be investigated through a study of the smoothness of the ambiguity function.

To see 1. as defined in the preceding paragraph, observe that for $\underline{\theta} = \underline{\theta}^o$ and arbitrary $\underline{u} \in \Theta$,

$$\bar{l}(\underline{\theta},\underline{\theta}) - \bar{l}(\underline{u},\underline{\theta}) = E_{\underline{\theta}}[\ln f(X;\underline{\theta}) - \ln f(X;\underline{u})]$$

$$= \int_{\text{supp}\, f(\bullet;\underline{\theta})} f(x;\underline{\theta})[\ln f(x;\underline{\theta}) - \ln f(x;\underline{u})]dx \quad , \tag{40}$$

where $\text{supp} f(\bullet;\underline{\theta}) = \{x : f(x;\underline{\theta}) > 0\}$. Using the elementary inequality $\ln y \leq y - 1$, $y \geq 0$ ([8], Theorem 150), we have the following:

$$\begin{aligned}
\bar{l}(\underline{\theta},\underline{\theta}) - \bar{l}(\underline{u},\underline{\theta}) &= -\int_{\text{supp}\, f(\bullet;\underline{\theta})} f(x;\underline{\theta}) \ln \frac{f(x;\underline{u})}{f(x;\underline{\theta})} dx \\
&\geq -\int_{\text{supp}\, f(\bullet;\underline{\theta})} f(x;\underline{\theta}) \left[\frac{f(x;\underline{u})}{f(x;\underline{\theta})} - 1\right] dx \\
&= \int_{\text{supp}\, f(\bullet;\underline{\theta})} f(x;\underline{\theta})dx - \int_{\text{supp}\, f(\bullet;\underline{\theta})} f(x;\underline{u})dx \\
&= 1 - p \quad , \qquad (41)
\end{aligned}$$

where $p \in [0,1]$. Hence, since $1 - p \geq 0$ in (41), $\bar{l}(\underline{\theta},\underline{\theta}) \geq \bar{l}(\underline{u},\underline{\theta})$, $\forall \underline{u}$; thus, $\bar{l}(\underline{u},\underline{\theta})$ has a global maximum at $\underline{u} = \underline{\theta}$.

For 2., observe that the incremental variation $\Delta_{\underline{u}}\bar{l}$, in $\bar{l}(\underline{u},\underline{\theta}^o)$, which is produced by an incremental change in $\underline{u}$ from $\underline{u} = \underline{\theta}^o$ to $\underline{u} = \underline{\theta}^o + \Delta\underline{u}$, is given by the Taylor formula

$$\begin{aligned}
\Delta_{\underline{u}}\bar{l} &\stackrel{\text{def}}{=} \bar{l}(\underline{\theta}^o + \Delta\underline{u},\underline{\theta}^o) - \bar{l}(\underline{\theta}^o,\underline{\theta}^o) \\
&= \nabla_{\underline{u}}\bar{l}(\underline{u},\underline{\theta}^o)|_{\underline{u}=\underline{\theta}^o}\Delta\underline{u} + \frac{1}{2}\Delta\underline{u}^T\left[\nabla_{\underline{u}}^T\nabla_{\underline{u}}\bar{l}(\underline{u},\underline{\theta}^o)|_{\underline{u}=\underline{\theta}^o}\right]\Delta\underline{u} + \epsilon \quad . \qquad (42)
\end{aligned}$$

In (42), ϵ is a remainder that falls off to zero as $o(\|\Delta\underline{u}\|^2)$. Use the fact that $\nabla_{\underline{u}}\bar{l}(\underline{u},\underline{\theta}^o)|_{\underline{u}=\underline{\theta}^o} = 0$ and identify the Fisher matrix F (22) in the quadratic form on the right-hand side of Equation (42) to obtain

$$\Delta_{\underline{u}}\bar{l} = -\frac{1}{2}\Delta\underline{u}^T F(\underline{\theta}^o)\Delta\underline{u} + \epsilon \quad . \tag{43}$$

From (43) it is clear that a small variation, $\Delta\underline{u}$, in the search parameter $\underline{u}$ produces a quadratic variation in $\bar{l}(\underline{u},\underline{\theta}^o)$, with F playing the role of a gain or sensitivity matrix. Let the difference, $\Delta\underline{u}$, between the search parameter and the fixed true parameter $\underline{\theta}^o$ vary over some differential region defined with respect to the standard orthonormal basis for $\mathbb{R}$. If the Fisher information is a high-gain matrix, e.g., F has large eigenvalues, then the ambiguity function will have a large variation in the corresponding eigenvector directions. In view of the dependence of the unbiased CR bound (30) on F, this suggests that the sharpness of the peak of the ambiguity function in the standard coordinates is directly related to the CR bound on unbiased estimators of $\underline{\theta}$.

Now, let $\nabla\underline{m}(\underline{\theta}^o)$ be the gradient matrix (Jacobian) of the mean $\underline{m}(\underline{\theta}^o)$ of a biased estimator and assume that $\nabla\underline{m}(\underline{\theta}^o)$ is invertible. Using the identity $[\nabla\underline{m}(\underline{\theta}^o)][\nabla\underline{m}(\underline{\theta}^o)]^{-1} = I$ in (43), we

obtain, by regrouping terms,

$$\begin{aligned}\Delta_{\underline{u}}\bar{l} &= -\frac{1}{2}\Delta\underline{u}^T[\nabla\underline{m}(\underline{\theta}^o)]^T[\nabla\underline{m}(\underline{\theta}^o)]^{-T}F(\underline{\theta}^o)[\nabla\underline{m}(\underline{\theta}^o)]^{-1}[\nabla\underline{m}(\underline{\theta}^o)]\Delta\underline{u}+\epsilon \\ &= -\frac{1}{2}[\nabla\underline{m}(\underline{\theta}^o)\Delta\underline{u}]^T\left\{[\nabla\underline{m}(\underline{\theta}^o)]^{-T}F(\underline{\theta}^o)[\nabla\underline{m}(\underline{\theta}^o)]^{-1}\right\}[\nabla\underline{m}(\underline{\theta}^o)\Delta\underline{u}]+\epsilon \\ &= -\frac{1}{2}\Delta\underline{\nu}^T F_b(\underline{\theta}^o)\Delta\underline{\nu}+\epsilon \quad , \end{aligned} \tag{44}$$

where F_b is the biased Fisher information (24) at $\underline{\theta}=\underline{\theta}^o$ and

$$\Delta\underline{\nu} \stackrel{\text{def}}{=} \nabla\underline{m}(\underline{\theta}^o)\Delta\underline{u} \tag{45}$$

is the differential parameter variation $\Delta\underline{\nu}=\underline{\nu}-\underline{\nu}(\underline{\theta}^o)$ in the new coordinates induced by the local transformation (32), $\underline{\nu}=\nabla\underline{m}(\underline{\theta}^o)\Delta\underline{u}+\underline{m}(\underline{\theta}^o)$. The relation (44) is similar to the relation (43) in that they both relate variations in the search parameter, $\Delta\underline{u}$ and $\Delta\underline{\nu}$, respectively, to variations in the ambiguity function $\bar{l}$ via a gain matrix, $F(\underline{\theta}^o)$ and $F_b(\underline{\theta}^o)$, respectively. If we fix a differential region of variation for $\Delta\underline{u}$ and $\Delta\underline{\nu}$ in (43), $F(\underline{\theta}^o)$ is the gain associated with variation $\Delta\underline{u}$ over this differential region in the standard coordinates, while in (44) $F_b(\underline{\theta}^o)$ is the gain associated with variation $\Delta\underline{\nu}$ over this differential region in the transformed coordinates (45). In light of the dependence on F_b of the general biased-estimation CR bound (25), this suggests that for biased estimators the sharpness of the peak of the ambiguity function at $\Delta\underline{u}=0$ in the transformed coordinates (which give a locally linear approximation to the mean function) is directly related to the general CR bound, which, as noted previously, is the bound which applies to unbiased estimation of the mean function of the estimator. Comparing this observation to that made after (43), we see that in either the standard or transformed parameter coordinates the sharpness of the peak of the ambiguity function is directly related to the CR bound that applies to unbiased estimation in these coordinates. The relation between the general CR bound and the variation of $\bar{l}$ is explained in greater detail in the following paragraphs.

Using (42) and (45), the variation of the ambiguity function as $\Delta\underline{\nu}$ varies in the transformed coordinates of (45) can be explicitly given in terms of the gradient matrix $\nabla\underline{m}(\underline{\theta}^o)$,

$$\begin{aligned}\Delta_{\underline{u}}\bar{l} &= \bar{l}(\underline{\theta}^o+\Delta\underline{u})-\bar{l}(\underline{\theta}^o) \\ &= \bar{l}(\underline{\theta}^o+[\nabla\underline{m}(\underline{\theta}^o)]^{-1}\Delta\underline{\nu})-\bar{l}(\underline{\theta}^o) \quad . \end{aligned} \tag{46}$$

In (46) $[\nabla\underline{m}(\underline{\theta}^o)]^{-1}\Delta\underline{\nu}$ is the differential in the standard coordinates that is induced by the differential $\Delta\underline{\nu}$ in the transformed coordinates.

For purposes of illustration, consider Figure 1, denoting a (spherical) volume element $\{\Delta\underline{\nu}: \|\Delta\underline{\nu}\|=\Delta\}$ in the transformed coordinates $\Delta\underline{\nu}=[\nabla\underline{m}(\underline{\theta}^o)]\Delta\underline{u}$, and Figure 2, denoting the induced

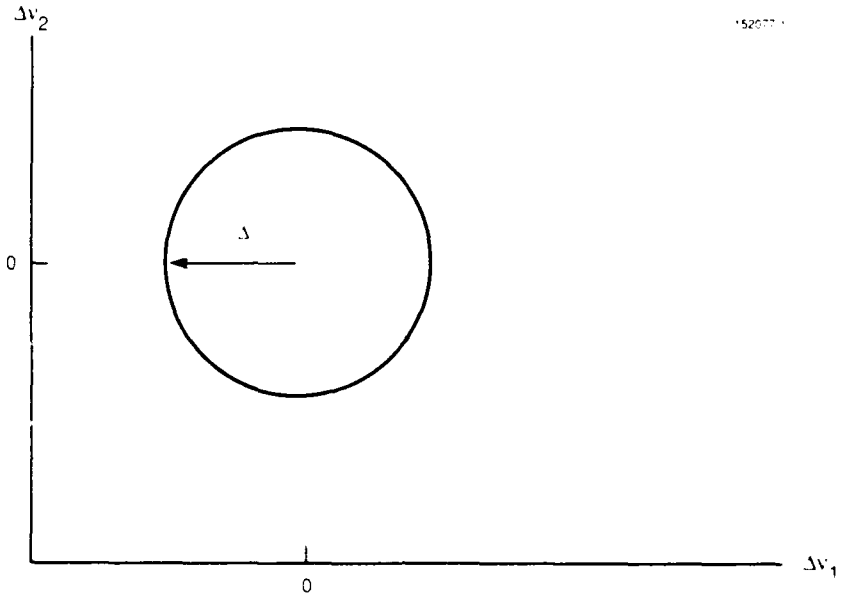


Figure 1. A spherical volume element $\{\Delta\underline{v} : \|\Delta\underline{v}\| = \Delta\}$ *in the transformed coordinates* $\Delta\underline{v} = [\nabla\underline{m}(\underline{\theta}^o)]\Delta\underline{u}$.

(ellipsoidal) volume element $\{\Delta\underline{u} : \|[\nabla\underline{m}(\underline{\theta}^o)]\Delta\underline{u}\| = \Delta\}$ in the standard coordinates. Figure 2 corresponds to the case

$$\nabla\underline{m} \stackrel{\text{def}}{=} \begin{bmatrix} 1-a & b \\ 0 & 1 \end{bmatrix} ,$$

where $a \ll 1$ and $b \geq 0$. In Figure 2 the angle of the principal axis can be shown (through considerable algebra) to be

$$\phi = \tan^{-1}\left(\frac{r(1-q^2)-(1-a)^2}{(1-a)b}\right) ,$$

and the positive parameters r and q^2 are given by

$$r \stackrel{\text{def}}{=} \frac{(1-a)^2+b^2+1}{2}$$

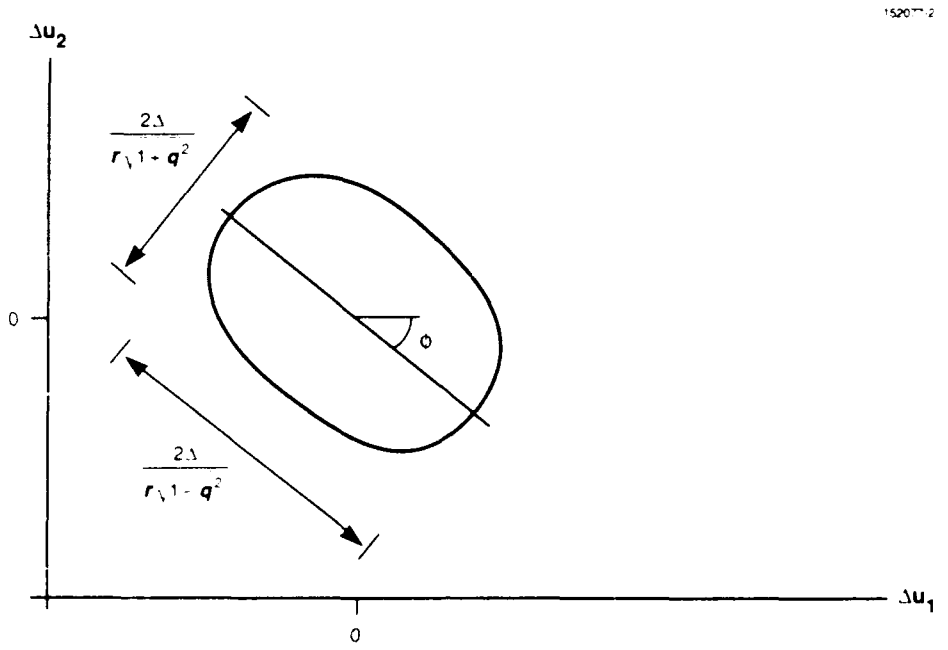


Figure 2. The induced differential volume element $\{\Delta\underline{u} : \|[\nabla\underline{m}(\underline{\theta}^o)]\Delta\underline{u}\| = \Delta\}$ *in the standard coordinates is an ellipsoid.*

$$q^2 \stackrel{\text{def}}{=} \frac{\sqrt{r^2 - (1-a)^2}}{r}$$

Referring to Figures 1 and 2, let $\Delta\underline{\nu}$ vary over the radius-Δ n-dimensional sphere $\{\underline{z} : \|\underline{z}\| \leq \Delta\}$ (Figure 1). For unbiased estimation $[\nabla\underline{m}(\underline{\theta}^o)] = I$ and the variation in the argument $\Delta\underline{u} = [\nabla\underline{m}(\underline{\theta}^o)]^{-1}\Delta\underline{\nu} = \Delta\underline{\nu}$, of $\Delta_{\underline{u}}\bar{l}$ (46) is over the same n-dimensional sphere. For biased estimation $\nabla\underline{m}(\underline{\theta}^o)$ is not an identity matrix and the variation in the argument $\Delta\underline{u}$ is over the n-dimensional ellipsoid $\{\underline{z} : \|[\nabla\underline{m}(\underline{\theta}^o)]\underline{z}\| \leq \Delta\}$ (Figure 2). Under a set of bias constraints of the type (10) on each of the rows of $\nabla\underline{m}(\underline{\theta}^o)$, this ellipsoid can only be a small perturbation of a sphere. Nonetheless, if the ambiguity function has an unstable characteristic locally in the standard coordinates, such as a sharp ridge, then the differential variation, $\Delta_{\underline{u}}\bar{l}$, of $\bar{l}$ over the ellipsoid in $\Delta\underline{u}$ can be made much larger than the differential variation of $\bar{l}$ over the sphere in $\Delta\underline{\nu}$ by judicious choice of $\nabla\underline{m}(\underline{\theta}^o)$ (see Figure 3). The variation of $\bar{l}$ over the sphere of radius Δ in $\Delta\nu$ can be used to bound from below the general CR bound on variance (26) for estimators of θ_1 by using the following fact.

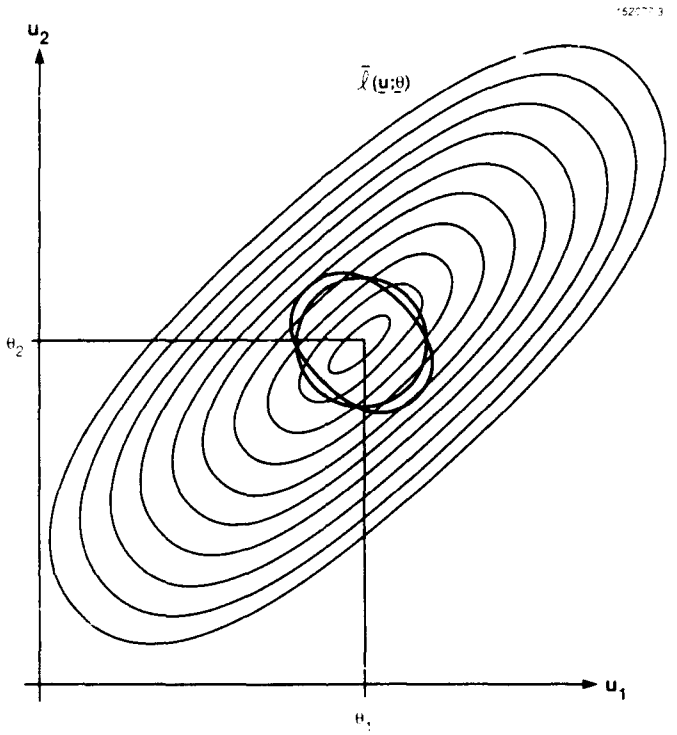


Figure 3. The constant contours of a hypothetical ambiguity function $\bar{l}(\underline{u},\underline{\theta}^o)$, over $\underline{u} \in \Theta$ for fixed $\underline{\theta} = \underline{\theta}^o$, and the superimposed induced differential volume elements of Figures 1 and 2. Because the ellipsoidal region includes a greater number of contour lines, the maximum variation of $\bar{l}$ over the ellipsoidal region is greater than the variation over the spherical region.

FACT:

$$B_{b_1} \geq \frac{\Delta^2}{2}\Big[\max_{\|\Delta\underline{\nu}\|^2 \leq \Delta^2} |\Delta_{\underline{u}}\bar{l}|\Big]^{-1} + \epsilon \quad , \tag{47}$$

where B_{b_1} *is the CR bound (28) and* ϵ *is* $o(\|\Delta\underline{\nu}\|^2) = o(\Delta^2)$.

Hence, if the variation of $\bar{l}$ is small, i.e, $\bar{l}$ has a broad peak, in the transformed coordinates, then (47) asserts that B_{b_1} must be large, implying poor variance performance of estimators of θ_1 with mean gradient matrix $\nabla\underline{m}(\underline{\theta})$. More important, if a bias-induced coordinate transformation on the search parameter $\Delta\underline{\nu} = [\nabla\underline{m}(\underline{\theta})]\Delta\underline{u}$ can be found for which the local variation of the ambiguity function over the ellipsoid $\{\Delta\underline{u} : \|[\nabla\underline{m}(\underline{\theta})]\Delta\underline{u}\| \leq \Delta\}$ is large, then a reduction in the CR bound may be possible.

Proof of Fact. Fix a parameter value $\underline{\theta}$. Define $\Delta\underline{\nu}$:

$$\Delta\underline{\nu} = \frac{F_b^{-\frac{1}{2}}\underline{e}_1}{\sqrt{\underline{e}_1^T F_b^{-1}\underline{e}_1}}\Delta \quad , \tag{48}$$

where $F_b^{-\frac{1}{2}}$ is a square-root factor of the (positive-definite) matrix F_b^{-1} [see (24)].

It is shown that $\Delta\underline{\nu}$ (48) is a vector contained in the radius-Δ sphere $\{\underline{z} : \|\underline{z}\| \leq \Delta\}$, because the norm squared of $\Delta\underline{\nu}$ is

$$\begin{aligned} \|\Delta\underline{\nu}\|^2 &= \frac{\underline{e}_1^T F_b^{-\frac{1}{2}} F_b^{-\frac{1}{2}}\underline{e}_1}{\underline{e}_1^T F_b^{-1}\underline{e}_1}\Delta^2 \\ &= \frac{\underline{e}_1^T F_b^{-1}\underline{e}_1}{\underline{e}_1^T F_b^{-1}\underline{e}_1}\Delta^2 = \Delta^2 \quad . \end{aligned} \tag{49}$$

Furthermore, substituting $\Delta\underline{\nu} = \frac{F_b^{-\frac{1}{2}}\underline{e}_1}{\sqrt{\underline{e}_1^T F_b^{-1}\underline{e}_1}}\Delta$ into (42) and (44), $\Delta_{\underline{u}}\bar{l}$ has the form

$$\bar{l}(\underline{\theta} + [\nabla\underline{m}(\underline{\theta})]^{-1}\Delta\underline{\nu}) - \bar{l}(\underline{\theta}) = -\frac{1}{2}\Delta\underline{\nu}^T F_b \Delta\underline{\nu} + o(\Delta^2)$$

$$= -\frac{1}{2}\left[\frac{F_b^{-\frac{1}{2}}\underline{e}_1}{\sqrt{\underline{e}_1^T F_b^{-1}\underline{e}_1}}\Delta\right]^T F_b \left[\frac{F_b^{-\frac{1}{2}}\underline{e}_1}{\sqrt{\underline{e}_1^T F_b^{-1}\underline{e}_1}}\Delta\right] + o(\Delta^2)$$

$$= -\frac{\Delta^2}{2}\frac{\underline{e}_1^T\underline{e}_1}{\underline{e}_1^T F_b^{-1}\underline{e}_1} + o(\Delta^2)$$

$$= -\frac{\Delta^2}{2}\frac{1}{\underline{e}_1^T F_b^{-1}\underline{e}_1} + o(\Delta^2) \quad . \tag{50}$$

Now (50) gives the value for $\Delta_{\underline{u}}\bar{l}$ evaluated at a point $\Delta\underline{\nu}$ (48) contained in the radius-Δ sphere. Hence, the maximum magnitude of $\Delta_{\underline{u}}\bar{l}$ over the radius-Δ sphere must be at least as great as the magnitude of the right-hand side of (50). This, and the elementary inequality $|x-y| \geq |x| - |y|$, gives the bound

$$\max_{\|\Delta\underline{\nu}\|^2 \leq \Delta^2} |\Delta_{\underline{u}}\bar{l}| \quad \geq \quad \frac{\Delta^2}{2}\frac{1}{B_{b_1}(\underline{\theta})} + o(\Delta^2) \quad . \tag{51}$$

Multiplying this inequality through by $B_{b_1} / \max_{\|\Delta\underline{\nu}\|^2 \leq \Delta^2} |\Delta_{\underline{u}}\bar{l}|$ gives the statement of the fact (47).

5. A NEW CR BOUND FOR ESSENTIALLY UNBIASED ESTIMATORS

Here we obtain a new lower bound which is applicable for essentially unbiased estimators, i.e., those whose bias gradient is small. Assume that the bias b_1 is such that $\|\nabla b_1(\underline{\theta})\|^2 \leq \delta^2$ for all $\underline{\theta} \in \Theta$. The starting point for the lower bound is the obvious inequality [see (28)]

$$\mathrm{var}_{\underline{\theta}}(\hat{\theta}_1) \geq B_{b_1}(\underline{\theta}) \geq \min_{b_1 : \|\nabla_{\underline{\theta}} b_1\|^2 \leq \delta^2} B_{b_1}(\underline{\theta}) \geq \min_{\underline{d} : \|\underline{d}\|^2 \leq \delta^2} [\underline{e}_1 + \underline{d}]^T F^{-1} [\underline{e}_1 + \underline{d}] \quad . \tag{52}$$

This gives the lower bound valid for estimators $\hat{\underline{\theta}}$, which satisfy the bias-gradient constraint (10)

$$\mathrm{var}_{\underline{\theta}}(\hat{\theta}_1) \geq B(\underline{\theta}, \delta) \quad , \tag{53}$$

where

$$B(\underline{\theta}, \delta) \stackrel{\mathrm{def}}{=} \min_{\underline{d} : \|\underline{d}\|^2 \leq \delta^2} [\underline{e}_1 + \underline{d}]^T F^{-1} [\underline{e}_1 + \underline{d}] \quad . \tag{54}$$

Note that by definition $B(\underline{\theta}, 0)$ is just the unbiased CR bound. The bound (53) is independent of the particular bias, b_1, of the estimator as long as the bias constraint is satisfied. The normalized difference

$$\Delta B(\underline{\theta}, \delta) = \frac{B(\underline{\theta}, 0) - B(\underline{\theta}, \delta)}{B(\underline{\theta}, 0)} \tag{55}$$

is the potential improvement achievable over an absolutely unbiased estimator, i.e., $\Delta B(\underline{\theta}, \delta)$ measures the bias sensitivity of the unbiased CR bound. The sensitivity is a real number between 0 and 1, and increased sensitivity corresponds to a larger difference.

The new bound is specified by the solution of the minimization problem (54). We find the solution in the following theorem and corollary.

Theorem 1 *Let $\hat{\theta}_1$ be an essentially unbiased estimator of θ_1 with bias $b_1(\underline{\theta})$ which satisfies the constraint (10) $\|\nabla_{\underline{\theta}} b_1\|^2 \leq \delta^2 < 1$. The lower bound $B(\underline{\theta}, \delta)$ (53) is equal to*

$$B(\underline{\theta}, \delta) = B(\underline{\theta}, 0) - \lambda \delta^2 - \underline{e}_1^T [I + \lambda F]^{-1} F^{-1} \underline{e}_1 \quad , \tag{56}$$

where λ is given by the unique non-negative solution of the following equation involving the monotone decreasing convex function $g(\lambda) \in [0, 1]$:

$$g(\lambda) \stackrel{\text{def}}{=} \underline{e}_1^T [I + \lambda F]^{-2} \underline{e}_1 = \delta^2 \quad , \quad \lambda \geq 0 \qquad . \tag{57}$$

Corollary 1 *An alternative form for the bound (56) is*

$$B(\underline{\theta}, \delta) = [\underline{e}_1 + \underline{d}_{min}]^T F^{-1} [\underline{e}_1 + \underline{d}_{min}] \qquad , \tag{58}$$

where $\underline{d}_{min}$ is the vector which minimizes the quadratic form $[\underline{e}_1 + \underline{d}]^T F^{-1} [\underline{e}_1 + \underline{d}]$ in (54) over the constraint set $\{\underline{d} : \|\underline{d}\| \leq \delta\}$

$$\underline{d}_{min}^T = -\underline{e}_1^T [I + \lambda F]^{-1} \qquad . \tag{59}$$

In (59), λ is the solution to (57).

Observe that the vector $\underline{d}_{min}$ (59) of Corollary 1 is related to the function $g(\lambda)$ (57) of Theorem 1 by the identity

$$\underline{d}_{min}^T \underline{d}_{min} = \underline{e}_1^T [I + \lambda F]^{-2} \underline{e}_1 = g(\lambda) \qquad . \tag{60}$$

Because $g(\lambda) = \delta^2$ in Theorem 1, $\underline{d}_{min}$ is on the boundary of the bias-gradient constraint set.

Proof of Theorem 1. The objective is to show that the right-hand side of (54) is equal to the right-hand side of (56):

$$\begin{aligned} \min_{\underline{d}:\|\underline{d}\|^2 \leq \delta^2} Q(\underline{d}) &= Q(\underline{d}_{min}) \\ &= B(\underline{\theta}, 0) - \lambda\delta^2 - \underline{e}_1^T [I + \lambda F]^{-1} F^{-1} \underline{e}_1 \qquad , \end{aligned} \tag{61}$$

where the general CR bound (28) has been denoted by the quadratic form $Q(\nabla b_1)$ to make evident the quadratic dependence on the bias gradient ∇b_1

$$Q(\underline{d}) \stackrel{\text{def}}{=} [\underline{e}_1 + \underline{d}]^T F^{-1} [\underline{e}_1 + \underline{d}] \qquad . \tag{62}$$

We take a geometric point of view which is easily formalized by using standard Lagrange multiplier theory. Specifically, $Q(\underline{d})$ is a convex-upwards paraboloid centered at coordinates $-\underline{e}_1 = -[1, 0, \ldots, 0]^T$, and $Q(-\underline{e}_1) = 0$. The problem is to find the vector $\underline{d} = \underline{d}_{min}$ within the radius δ ball, $\underline{d}^T \underline{d} \leq \delta^2$, for which $Q(\underline{d})$ is a minimum. Observe that by the assumption $\delta < 1$ the absolute minimum of Q is not attained within the radius δ ball. Therefore, as there are no local minima, the minimizing vector $\underline{d}_{min}$ must be on the boundary of the ball. By inspection of the constant

contours of $Q(\underline{d})$ (Figure 4), the minimum is attained when the contour of $Q(\underline{d})$ is tangent to the boundary of the radius δ ball, i.e., the gradients are collinear and of opposite sign;

$$\nabla_{\underline{d}} Q(\underline{d}) = -\lambda \nabla_{\underline{d}} \ \underline{d}^T \underline{d} \tag{63}$$

for some $\lambda \geq 0$. Using the rules of vector differentiation for quadratic forms, (4) and (5), (63) is equivalent to

$$F^{-1}[\underline{e}_1 + \underline{d}] = -\lambda \underline{d} \quad . \tag{64}$$

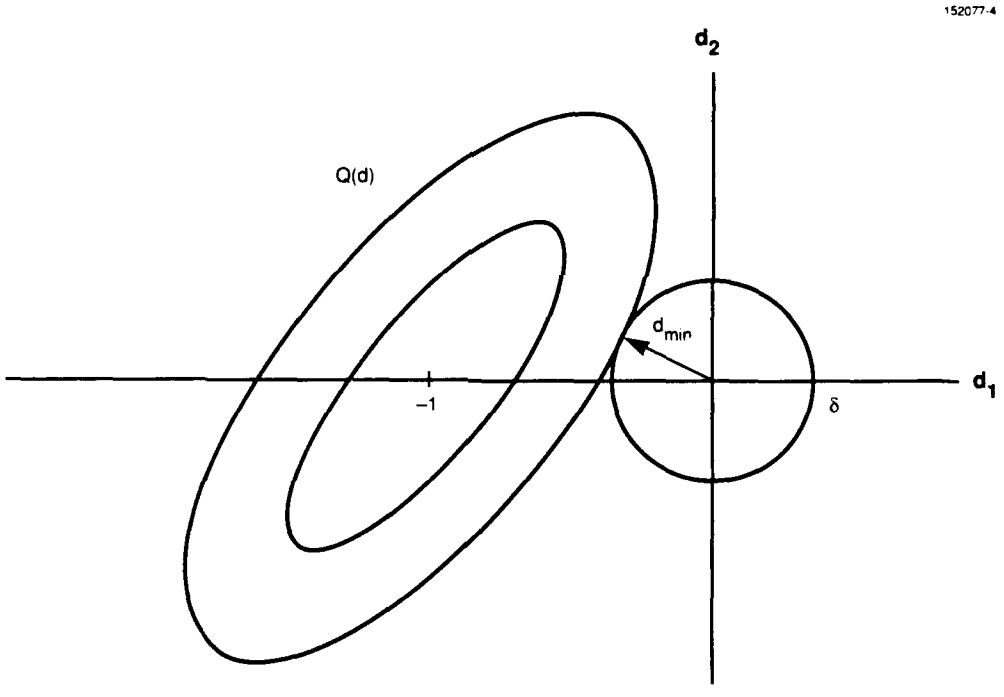


Figure 4. Plot of the constant contours of $Q(\underline{d})$ and the domain of the constraint $\underline{d}^T \underline{d} \leq \delta^2$. The minimum of Q is achieved at the point indicated by the vector $\underline{d}_{min}$ which is normal to the tangent plane between $Q(\underline{d})$ and $\underline{d}^T \underline{d} = \delta^2$.

Solving (64) for $\underline{d} = \underline{d}_{min}$ gives

$$\underline{d} = \underline{d}_{min} = -[I + \lambda F]^{-1}\underline{e}_1 \quad . \tag{65}$$

Note that the matrix inverse $[I + \lambda F]^{-1}$ exists, since λ is non-negative. The scaling constant λ has to be chosen so that $\underline{d}$ is on the boundary of the radius δ ball $\underline{d} : \|\underline{d}\|^2 \leq \delta^2$, from which we obtain equation (57) for λ,

$$g(\lambda) \stackrel{\text{def}}{=} \underline{d}_{min}^T \underline{d}_{min} = \underline{e}_1^T [I + \lambda F]^{-2} \underline{e}_1 = \delta^2 \quad . \tag{66}$$

Observe that $g(0) = 1$, $g(\infty) = 0$, and because F is positive-definite, $g(\lambda) \geq 0$ for $\lambda \geq 0$. Application of the differentiation identity (6) to $g(\lambda)$ gives $g'(\lambda) = -2\underline{e}_1^T F[I + \lambda F]^{-3}\underline{e}_1$. Due to positive-definitness of F, $|g'(\lambda)| < \infty$ so that g is continuous at all points $\lambda \geq 0$. Furthermore, since $[I + \lambda F]^{-3}$ is symmetric positive-definite with identical eigenvectors as F, $F[I + \lambda F]^{-3}$ is positive-definite (see Section 2.2) for $\lambda \geq 0$. Hence, g is monotone decreasing over $\lambda \geq 0$ with values $g(\lambda) \in [0, 1]$. In a similar manner, the second derivative g'' can be shown to be positive, which establishes that g is a convex (upwards) function.

It remains to show that the minimizing solution $\underline{d}_{min}$ gives the bound (56). This is established by substitution of $\underline{d}_{min}$ (65) into $Q(\underline{d})$ (62):

$$\begin{aligned} B(\underline{\theta}, \delta) &= Q(\underline{d}_{min}) \\ &= [\underline{e}_1 + \underline{d}_{min}]^T F^{-1} [\underline{e}_1 + \underline{d}_{min}] \\ &= \underline{e}_1^T F^{-1} \underline{e}_1 - \underline{e}_1^T [I + \lambda F]^{-1} F^{-1} \underline{e}_1 \\ &- \underline{e}_1^T \left[F^{-1}[I + \lambda F]^{-1} - [I + \lambda F]^{-1} F^{-1} [I + \lambda F]^{-1} \right] \underline{e}_1 \quad . \end{aligned} \tag{67}$$

In (67) we have made use of the property of symmetry of the Fisher matrix. The first and second terms on the right-hand side of (67) are simply the unbiased CR bound (30), $B(\underline{\theta}, 0)$, and the final term on the right-hand side of (56), respectively. Using the constraint (66), the third term in (67) is seen to be equal to

$$\begin{aligned} & \underline{e}_1^T \left[F^{-1}[I + \lambda F]^{-1} - [I + \lambda F]^{-1} F^{-1} [I + \lambda F]^{-1} \right] \underline{e}_1 \\ = \; & \underline{e}_1^T [I + \lambda F]^{-1} \left[[I + \lambda F] F^{-1} - F^{-1} \right] [I + \lambda F]^{-1} \underline{e}_1 \\ = \; & \underline{e}_1^T [I + \lambda F]^{-1} [\lambda I] [I + \lambda F]^{-1} \underline{e}_1 \\ = \; & \lambda \underline{e}_1^T [I + \lambda F]^{-2} \underline{e}_1 = \lambda \delta^2 \quad . \end{aligned} \tag{68}$$

This establishes the theorem.

As stated, Theorem 1 gives a bound $B(\underline{\theta},\delta)$ for which the (1,1) element of the inverses of the $n \times n$ matrices $F[I + \lambda F]$ and $[I + \lambda F]^2$ must be computed. These calculations can be implemented by sequential partitioning [9]. A more explicit version of $B(\underline{\theta},\delta)$ will be of interest for the approximations of Section 6.

Let the Fisher information matrix F (21) have the partition

$$F = \begin{bmatrix} a & \underline{c}^T \\ \underline{c} & F_s \end{bmatrix} \quad , \tag{69}$$

where a is a scalar, $\underline{c}$ is a $(n-1) \times 1$ vector, and F_s is a $(n-1) \times (n-1)$ submatrix. Define the inverse matrices

$$F^{-1} \stackrel{\text{def}}{=} \begin{bmatrix} \alpha & \underline{\beta}^T \\ \underline{\beta} & \Gamma \end{bmatrix} \quad , \tag{70}$$

$$[I + \lambda F]^{-1} \stackrel{\text{def}}{=} \begin{bmatrix} \alpha_\lambda & \underline{\beta}_\lambda^T \\ \underline{\beta}_\lambda & \Gamma_\lambda \end{bmatrix} \quad . \tag{71}$$

Using the partitioned-matrix inverse identity (2), α, α_λ, Γ, $\underline{\beta}$, and $\underline{\beta}_\lambda$ have the following expressions in terms of the elements of F (69):

$$\begin{aligned} \alpha &= (a - \underline{c}^T F_s^{-1} \underline{c})^{-1} \\ \underline{\beta} &= -\alpha F_s^{-1} \underline{c} \\ \Gamma &= F_s^{-1} + \alpha F_s^{-1} \underline{c}\underline{c}^T F_s^{-1} \end{aligned} \tag{72}$$

$$\begin{aligned} \alpha_\lambda &= (1 + \lambda a - \lambda^2 \underline{c}^T [I + \lambda F_s]^{-1} \underline{c})^{-1} \\ \underline{\beta}_\lambda &= -\lambda \alpha_\lambda [I + \lambda F_s]^{-1} \underline{c} \quad . \end{aligned} \tag{73}$$

The quantity Γ_λ will not be explicitly needed and is omitted.

In (69), $\underline{c}$ represents the information coupling between estimates of θ_1 and estimates of the other parameters $\theta_2, \ldots \theta_n$. This can be seen from the fact that for unbiased estimation the CR bound on $\hat{\theta}_1$ (30) is $(F^{-1})_{11} = \alpha = (a - \underline{c}^T F_s^{-1} \underline{c})^{-1}$, which is identical to a^{-1} in the case $\underline{c} = \underline{0}$.

Theorem 2 *In terms of the partitioned Fisher information matrix (69), the bound $B(\underline{\theta},\delta)$ and the associated constraint equation (57) of Theorem 1 have the expressions*

$$B(\underline{\theta},\delta) = B(\underline{\theta},0) - \lambda\delta^2 - \alpha_\lambda\alpha - \underline{\beta}_\lambda^T\underline{\beta} \quad , \tag{74}$$

and λ is the unique positive solution of

$$g(\lambda) = \frac{1 + \lambda^2\underline{c}^T[I + \lambda F_s]^{-2}\underline{c}}{(1 + \lambda a - \lambda^2\underline{c}^T[I + \lambda F_s]^{-1}\underline{c})^2} = \delta^2 \quad . \tag{75}$$

Furthermore, the minimizing vector (65) of Corollary 1 is given by

$$\underline{d}_{min} = [-\alpha_\lambda, -\underline{\beta}_\lambda^T]^T = \alpha_\lambda[-1, \lambda\underline{c}^T[I + \lambda F_s]^{-1}]^T \quad . \tag{76}$$

In (74), (75), and (76) α, α_λ, $\underline{\beta}$, and $\underline{\beta}_\lambda$ are the quantities given in (72) and (73).

Proof of Theorem. Substitution of the inverses F^{-1} (70) and $[I + \lambda F]^{-1}$ (71) into the expressions for $B(\underline{\theta},\delta)$ (56) of Theorem 1 gives

$$\begin{aligned} B(\underline{\theta},\delta) &= B(\underline{\theta},0) - \lambda\delta^2 - \underline{e}_1^T \begin{bmatrix} \alpha_\lambda & \underline{\beta}_\lambda^T \\ \underline{\beta}_\lambda & \Gamma_\lambda \end{bmatrix} \begin{bmatrix} \alpha & \underline{\beta}^T \\ \underline{\beta} & \Gamma \end{bmatrix} \underline{e}_1 \\ &= B(\underline{\theta},0) - \lambda\delta^2 - [\alpha_\lambda, \underline{\beta}_\lambda^T][\alpha, \underline{\beta}^T]^T \\ &= B(\underline{\theta},0) - \lambda\delta^2 - \alpha_\lambda\alpha - \underline{\beta}_\lambda^T\underline{\beta} \quad , \end{aligned} \tag{77}$$

which is expression (74). Similarly, substitution of (71) into the expression for $\underline{d}_{min}$ (65) of Corollary 1 gives

$$\begin{aligned} \underline{d}_{min} &- -[I + \lambda F]^{-1}\underline{c}_1 \\ &= -\begin{bmatrix} \alpha_\lambda & \underline{\beta}_\lambda^T \\ \underline{\beta}_\lambda & \Gamma_\lambda \end{bmatrix} \underline{e}_1 = [-\alpha_\lambda, -\underline{\beta}_\lambda^T]^T \\ &= \alpha_\lambda[-1,\ \lambda\underline{c}^T[I + \lambda F_s]^{-1}]^T \quad , \end{aligned} \tag{78}$$

where, in the last line, the identity for $\underline{\beta}_\lambda$ (73) has been used.

Squaring the matrix $[I + \lambda F]^{-1}$ (71), we find that the constraint equation (57) of Theorem 1 has the following expression:

$$g(\lambda) = \underline{\epsilon}_1^T [I + \lambda F]^{-2} \underline{\epsilon}_1 = \alpha_\lambda^2 + \underline{\beta}_\lambda^T \underline{\beta}_\lambda = \delta^2 \quad . \tag{79}$$

Application of the identities (73) to the previous equation gives the equivalent expression for g,

$$\begin{aligned} g(\lambda) &= \alpha_\lambda^2 + \lambda^2 \alpha_\lambda^2 \underline{c}^T [I + \lambda F_s]^{-2} \underline{c} \\ &= \alpha_\lambda^2 [1 + \lambda^2 \underline{c}^T [I + \lambda F_s]^{-2} \underline{c}] \\ &= \frac{1 + \lambda^2 \underline{c}^T [I + \lambda F_s]^{-2} \underline{c}}{(1 + \lambda a - \lambda^2 \underline{c}^T [I + \lambda F_s]^{-1} \underline{c})^2} \quad , \end{aligned} \tag{80}$$

which is equivalent to the expression in (75). This completes the proof of Theorem 2.

The bound in Theorem 2 does not have an analytic form in general because the solution λ to (75) is not given explicitly. Numerical polynomial root-finding techniques can be used to solve (75), or equivalently (57), for λ. In particular, as the Fisher matrix F_s is positive-definite and symmetric it has the representation $F_s = Q \Phi Q^T$, where Q is an orthogonal matrix with columns $\underline{q}_i$ and Φ is a diagonal matrix with diagonal elements $\phi_i > 0$, $i = 1, \ldots, n$. Hence, we have

$$\begin{aligned} \underline{\epsilon}_1^T [I + \lambda F]^{-2} \underline{\epsilon}_1 &= \underline{\epsilon}_1^T [I + \lambda Q \Phi Q^T]^{-2} \underline{\epsilon}_1 \\ &= \underline{\epsilon}_1^T \left(Q [I + \lambda \Phi] Q^T \right)^{-2} \underline{\epsilon}_1 \\ &= \underline{\epsilon}_1^T Q [I + \lambda \Phi]^{-2} Q^T \underline{\epsilon}_1 \\ &= \sum_{i=1}^n \frac{1}{(1 + \lambda \phi_i)^2} (\underline{\epsilon}_1^T \underline{q}_i)^2 \quad . \end{aligned} \tag{81}$$

Therefore, inserting the expression into the left-hand side of (57) and multiplying both sides by $\prod_{i=1}^n [1 + \lambda \phi_i]^2$, we obtain a polynomial equation for λ:

$$\begin{aligned} \sum_{i=1}^n \frac{1}{(1 + \lambda \phi_i)^2} (\underline{\epsilon}_1^T \underline{q}_i)^2 \prod_{j=1}^n [1 + \lambda \phi_j]^2 &= \delta^2 \prod_{j=1}^n [1 + \lambda \phi_j]^2 \\ \sum_{i=1}^n \prod_{j \neq i} [1 + \lambda \phi_j]^2 (\underline{\epsilon}_1^T \underline{q}_i)^2 &= \delta^2 \prod_{j=1}^n [1 + \lambda \phi_j]^2 \quad . \end{aligned} \tag{82}$$

Subtraction of the right-hand side of (82) from the left-hand side gives a polynomial in λ of degree $2n$ which must be set to zero.

For the special case of zero information coupling, $\underline{c} = \underline{0}$, an explicit expression for λ can be found and the bound $B(\underline{\theta}, \delta)$ of Theorem 2 has an explicit form. From the definitions of α, α_λ, $\underline{\beta}$ and $\underline{\beta}_\lambda$ (72) and (73), $\underline{c} = \underline{0}$ implies that $\alpha = a^{-1} = B(\underline{\theta}, 0)$, $\alpha_\lambda = (1 + \lambda a)^{-1}$, and $\underline{\beta} = \underline{\beta}_\lambda = \underline{0}$. Furthermore, from (75)

$$\delta^2 = g(\lambda) = \frac{1}{(1+\lambda a)^2} = \alpha_\lambda^2 \quad , \tag{83}$$

and, consequently,

$$\lambda = a^{-1}(\delta^{-1} - 1) \quad . \tag{84}$$

Hence, using (83) in (76)

$$\underline{d}_{min} = [-\alpha_\lambda, \underline{0}^T]^T = [-\delta, \underline{0}^T]^T = -\delta \underline{e}_1 \quad , \tag{85}$$

and, from (58), the bound is

$$\begin{aligned} B(\underline{\theta}, \delta) &= [\underline{e}_1 + \underline{d}_{min}]^T F^{-1} [\underline{e}_1 + \underline{d}_{min}] \\ &= [\underline{e}_1 - \delta \underline{e}_1]^T F^{-1} [\underline{e}_1 - \delta \underline{e}_1] = (1-\delta)^2 \underline{e}_1^T F^{-1} \underline{e}_1 \\ &= (1-\delta)^2 B(\underline{\theta}, 0) = (1-\delta)^2 a^{-1} \quad . \end{aligned} \tag{86}$$

Observe that zero information coupling implies two important facts: small bias gradients have very little effect on the general CR bound, as the difference (55) $\Delta B(\underline{\theta}, \delta) = 1 - (1-\delta)^2 \approx 0$; and coupling of the bias, b_1, of $\hat{\theta}_1$ to the other parameters $\theta_2, \ldots, \theta_n$ is not likely to significantly reduce the CR bound, as the CR bound's minimizing vector $\underline{d}_{min}$ (85) makes the bias of $\hat{\theta}_1$ independent of the other parameters.

In the following sections we will assume that $\underline{c} \neq \underline{0}$.

6. A BIAS-SENSITIVITY INDEX AND A SMALL-δ APPROXIMATION

In the previous section the form of a CR-type bound $B(\underline{\theta}, \delta)$ was given in terms of an undetermined multiplier λ given by the solution of $g(\lambda) = \delta^2$ (75). For the case of zero information coupling, $g(\lambda)$ has a simple form, giving an explicit solution for λ and $B(\underline{\theta}, \delta)$ (86). Otherwise, an exact analytic form for $B(\underline{\theta}, \delta)$ is difficult to obtain and the solution λ to (75) might, for example, be calculated using numerical polynomial root-finding techniques applied to (82). In this section an $o(\delta)$ approximation is developed for $B(\underline{\theta}, \delta)$ which converges to the true solution as δ approaches zero. Associated with the approximation is an approximate minimizing bias-gradient vector $\underline{d}_{min}$, which achieves the $o(\delta)$ approximation to $B(\underline{\theta}, \delta)$.

While the approximations offer little computational advantage relative to the implementation of the exact computation indicated in Theorems 1 and 2, the approximate analytical forms provide some insight into the important factors underlying the bias sensitivity of the CR bound. In particular, the bias sensitivity of the CR bound is characterized by the slope of $B(\underline{\theta}, \delta)$, at $\delta = 0$. A large slope implies that a small amount of bias can substantially decrease the nominally unbiased CR bound, corresponding to high sensitivity. In the sequel, this slope will be related to a bias-sensitivity index.

For convenience, formula (75) is repeated here:

$$g(\lambda) = \frac{1 + \lambda^2 \underline{c}^T [I + \lambda F_s]^{-2} \underline{c}}{(1 + \lambda a - \lambda^2 \underline{c}^T [I + \lambda F_s]^{-1} \underline{c})^2} = \delta^2 \quad . \tag{87}$$

The idea behind the derivation of the $o(\delta)$ approximation follows. Theorem 1 established that $g(\lambda)$ is convex and monotone decreasing over $\lambda \geq 0$, $g(0) = 1$ and $g(\infty) = 0$. Therefore, for a sufficiently small value of δ, the solution λ to $g(\lambda) = \delta^2$ is sufficiently large so that simultaneously

$$\lambda \underline{c}^T [I + \lambda F_s]^{-1} \underline{c} \approx \underline{c}^T F_s^{-1} \underline{c} \text{ and } \lambda^2 \underline{c}^T [I + \lambda F_s]^{-2} \underline{c} \approx \underline{c}^T F_s^{-2} \underline{c} \quad . \tag{88}$$

If (88) holds, the constraint to be satisfied (87) becomes the simpler equation

$$\frac{1 + \underline{c}^T F_s^{-2} \underline{c}}{(1 + \lambda [a - \underline{c}^T F_s^{-1} \underline{c}])^2} = \delta^2 \quad . \tag{89}$$

from which the solution λ is simply computed by taking the square root of both sides of (89) and solving for λ. This solution can be plugged back into (65) and (56) to obtain approximations to $\underline{d}_{min}$ and $B(\underline{\theta}, \delta)$.

The following proposition puts precise asymptotic conditions on the solution to the constraint equation (87) to guarantee the validity of the approximation (89).

Proposition 2 *Let λ^* be given by the non-negative solution of (89)*

$$\lambda^* \stackrel{\text{def}}{=} \left(-1 + \frac{\sqrt{1 + \underline{c}^T F_s^{-2} \underline{c}}}{\delta} \right) [a - \underline{c}^T F_s^{-1} \underline{c}]^{-1} \quad , \tag{90}$$

and assume $\underline{c} \neq \underline{0}$. Corresponding to λ^*, define the vector $\underline{d}^*_{min}$;

$$\underline{d}^*_{min} \stackrel{\text{def}}{=} -[I + \lambda^* F]^{-1} \underline{e}_1 \quad . \tag{91}$$

With these definitions, λ^* approximates the actual solution λ of $g(\lambda) = \delta^2$, $\lambda \geq 0$ (75), in the sense that

1. *$0 \leq g(\lambda^*) \leq \delta^2$, so that $\underline{d}^*_{min}$ does not violate the constraint (10) on the bias gradient. (Recall, from (60), that $g(\lambda^*) = \|\underline{d}^*_{min}\|^2$.)*
2. *$\lambda\delta \longrightarrow \lambda^*\delta = \alpha\sqrt{1 + \underline{c}^T F_s^{-2} \underline{c}} + O(\delta)$ as $\delta \longrightarrow 0$.*

As a consequence of 2., $\lambda\delta = O(1)$ and also $\frac{1}{\lambda} = O(\delta)$.

To prove Proposition 2 we will use the following:

Lemma 1 *Let $\underline{c} \neq \underline{0}$ and define the positive quantity q;*

$$q \stackrel{\text{def}}{=} \min \left\{ \frac{\underline{c}^T F_s^{-1} \underline{c}}{\underline{c}^T F_s^{-2} \underline{c}} \; , \; \frac{\underline{c}^T F_s^{-2} \underline{c}}{\underline{c}^T F_s^{-3} \underline{c}} \right\} \quad . \tag{92}$$

Then $\lambda q \longrightarrow \infty$ as $\delta \longrightarrow 0$. Furthermore, the following bounds are valid for all $\lambda > 0$:

$$\underline{c}^T F_s^{-1} \underline{c} \left(1 - \frac{1}{\lambda q} \right) \leq \lambda \underline{c}^T [I + \lambda F_s]^{-1} \underline{c} \leq \underline{c}^T F_s^{-1} \underline{c} \quad ; \tag{93}$$

$$\underline{c}^T F_s^{-2} \underline{c} \left(1 - \frac{2}{\lambda q} \right) \leq \lambda^2 \underline{c}^T [I + \lambda F_s]^{-2} \underline{c} \leq \underline{c}^T F_s^{-2} \underline{c} \quad . \tag{94}$$

A weaker set of bounds is obtained by replacement of q in (93) and (94) by the minimum eigenvalue $\lambda^{F_s}_{min}$ of the matrix F_s. Furthermore, from (93) and (94), $\lambda \underline{c}^T [I + \lambda F_s]^{-1} \underline{c} = \underline{c}^T F_s^{-1} \underline{c} + O(\frac{1}{\lambda})$ and $\lambda^2 \underline{c}^T [I + \lambda F_s]^{-2} \underline{c} = \underline{c}^T F_s^{-2} \underline{c} + O(\frac{1}{\lambda})$.

Proof of Lemma 1. Recall from Theorem 1 that g is a continuous, monotonically decreasing function with $\lim_{\lambda \to \infty} g(\lambda) = 0$. This implies that the inverse function g^{-1} is continuous monotonically decreasing with $\lim_{\delta \to 0} g^{-1}(\delta) = \infty$; therefore, if q is fixed and nonzero, for sufficiently small δ the quantity λq can be made arbitrarily large. Hence, $\lambda q \longrightarrow \infty$ as $\delta \longrightarrow 0$ as claimed.

The right-hand inequalities in (93) and (94) follow from the inequality for a positive-definite matrix A and arbitrary vector $\underline{x}$:

$$\underline{x}^T[I + A]^{-k}\underline{x} \leq \underline{x}^T A^{-k}\underline{x} \tag{95}$$

for any integer $k \geq 0$. This can be proven via an eigen-decomposition of A. The left-hand inequalities in (93) and (94) are established by application of the Sherman-Morrison-Woodbury identity [Equation (3)]:

$$\begin{aligned}
[I + \lambda F_s]^{-1} &= \frac{1}{\lambda}F_s^{-1} - \frac{1}{\lambda}F_s^{-1}[I + \frac{1}{\lambda}F_s^{-1}]^{-1}\frac{1}{\lambda}F_s^{-1} \\
&= \frac{1}{\lambda}F_s^{-1} - \frac{1}{\lambda}[I + \lambda F_s]^{-1}F_s^{-1} \quad .
\end{aligned} \tag{96}$$

Application of (95) and (96) to $\underline{c}^T[I + \lambda F_s]^{-1}\underline{c}$ gives directly

$$\begin{aligned}
\lambda\underline{c}^T[I + \lambda F_s]^{-1}\underline{c} &= \underline{c}^T F_s^{-1}\underline{c} - \underline{c}^T[I + \lambda F_s]^{-1}F_s^{-1}\underline{c} \\
&= \underline{c}^T F_s^{-1}\underline{c} - \underline{c}^T F_s^{-\frac{1}{2}}[I + \lambda F_s]^{-1}F_s^{-\frac{1}{2}}\underline{c} \\
&\geq \underline{c}^T F_s^{-1}\underline{c} - \frac{1}{\lambda}\underline{c}^T F_s^{-2}\underline{c} \\
&= \underline{c}^T F_s^{-1}\underline{c}\left(1 - \frac{1}{\lambda}\frac{\underline{c}^T F_s^{-2}\underline{c}}{\underline{c}^T F_s^{-1}\underline{c}}\right) \quad .
\end{aligned} \tag{97}$$

Application of (95) and (96) to $\underline{c}^T[I + \lambda F_s]^{-2}\underline{c}$ yields

$$\begin{aligned}
\lambda^2\underline{c}^T[I + \lambda F_s]^{-2}\underline{c} &= \lambda^2\underline{c}^T[I + \lambda F_s]^{-1}[I + \lambda F_s]^{-1}\underline{c} \\
&= \lambda^2\underline{c}^T[\frac{1}{\lambda}F_s^{-1} - \frac{1}{\lambda}[I + \lambda F_s]^{-1}F_s^{-1}][\frac{1}{\lambda}F_s^{-1} - \frac{1}{\lambda}[I + \lambda F_s]^{-1}F_s^{-1}]\underline{c} \\
&= \underline{c}^T F_s^{-2}\underline{c} - 2\underline{c}^T[I + \lambda F_s]^{-1}F_s^{-2}\underline{c} + \underline{c}^T[I + \lambda F_s]^{-2}F_s^{-2}\underline{c} \\
&\geq \underline{c}^T F_s^{-2}\underline{c} - 2\underline{c}^T[I + \lambda F_s]^{-1}F_s^{-2}\underline{c} \\
&= \underline{c}^T F_s^{-2}\underline{c} - 2\underline{c}^T F_s^{-1}[I + \lambda F_s]^{-1}F_s^{-1}\underline{c} \\
&\geq \underline{c}^T F_s^{-2}\underline{c} - \frac{2}{\lambda}\underline{c}^T F_s^{-3}\underline{c}
\end{aligned}$$

$$= \underline{c}^T F_s^{-2} \underline{c} \left(1 - \frac{2}{\lambda} \frac{\underline{c}^T F_s^{-3} \underline{c}}{\underline{c}^T F_s^{-2} \underline{c}} \right) \quad . \tag{98}$$

In obtaining the inequalities (97) and (98) we have used the fact that $[I + \lambda F_s]^{-1}$, F_s^{-1}, and $F_s^{-\frac{1}{2}}$ are positive-definite matrices with identical eigen-decompositions; hence (see Section 2.2), these matrices commute and $[I + \lambda F_s]^{-2} F_s^{-2}$ is positive-definite. By the definition (92), $1/q \geq \frac{\underline{c}^T F_s^{-2} \underline{c}}{\underline{c}^T F_s^{-1} \underline{c}}$ and $1/q \geq \frac{\underline{c}^T F_s^{-3} \underline{c}}{\underline{c}^T F_s^{-2} \underline{c}}$, so that the right-hand sides of the inequalities (97) and (98) are underbounded by the right-hand sides of the inequalities (93) and (94) in the statement of the lemma.

Recall the variational inequality ([3], Theorem 4.2.2) for any compatible vector $\underline{z}$ and symmetric matrix A:

$$\frac{\underline{z}^T A \underline{z}}{\underline{z}^T \underline{z}} \geq \lambda_{min}^A \quad , \tag{99}$$

where λ_{min}^A is the minimum eigenvalue of A. Apply (99) to the lower inequality in (97) along with the definition $\underline{z} \stackrel{\text{def}}{=} F_s^{-1} \underline{c}$:

$$\begin{aligned} \underline{c}^T F_s^{-1} \underline{c} \left(1 - \frac{1}{\lambda} \frac{\underline{c}^T F_s^{-2} \underline{c}}{\underline{c}^T F_s^{-1} \underline{c}} \right) &= \underline{c}^T F_s^{-1} \underline{c} \left(1 - \frac{1}{\lambda} \frac{\underline{z}^T \underline{z}}{\underline{z}^T F_s \underline{z}} \right) \\ &\geq \underline{c}^T F_s^{-1} \underline{c} \left(1 - \frac{1}{\lambda \lambda_{min}^{F_s}} \right) \quad . \end{aligned} \tag{100}$$

Likewise, the definition $\underline{z} \stackrel{\text{def}}{=} F_s^{-\frac{3}{2}} \underline{c}$ and (98) give

$$\begin{aligned} \underline{c}^T F_s^{-2} \underline{c} \left(1 - \frac{2}{\lambda} \frac{\underline{c}^T F_s^{-3} \underline{c}}{\underline{c}^T F_s^{-2} \underline{c}} \right) &= \underline{c}^T F_s^{-2} \underline{c} \left(1 - \frac{2}{\lambda} \frac{\underline{z}^T \underline{z}}{\underline{z}^T F_s \underline{z}} \right) \\ &\geq \underline{c}^T F_s^{-2} \underline{c} \left(1 - \frac{2}{\lambda \lambda_{min}^{F_s}} \right) \quad . \end{aligned} \tag{101}$$

This establishes the lemma.

Proof of Proposition 2. Application of the inequalities in the lemma to the expression for $g(\lambda)$ (87):

$$g(\lambda) = \frac{1 + \lambda^2 \underline{c}^T [I + \lambda F_s]^{-2} \underline{c}}{(1 + \lambda a - \lambda^2 \underline{c}^T [I + \lambda F_s]^{-1} \underline{c})^2} \tag{102}$$

evaluated at $\lambda = \lambda^*$ gives

$$\frac{1+\underline{c}^T F_s^{-2}\underline{c}(1-2(\lambda^* q)^{-1})}{[1+\lambda^*(a-\underline{c}^T F_s^{-1}\underline{c}(1-(\lambda^* q)^{-1}))]^2} \leq g(\lambda^*) \leq \frac{1+\underline{c}^T F_s^{-2}\underline{c}}{[1+\lambda^*(a-\underline{c}^T F_s^{-1}\underline{c})]^2} = \delta^2 \quad . \tag{103}$$

The right-hand inequality in (103) establishes statement 1. in the statement of the proposition: $g(\lambda^*) \leq \delta^2$. Because the lower bound in (103) approaches the upper bound δ^2 in (103) as $\lambda^* q \longrightarrow \infty$, $g(\lambda^*)$ is forced to δ^2 as $\lambda^* q \longrightarrow \infty$, or equivalently, by the definition of λ^* (90), as $\delta \longrightarrow 0$. To establish statement 2., recall that $g(\lambda) = \delta^2$ (75) and consider the following:

$$\begin{aligned} \lambda\delta &= \lambda\sqrt{g(\lambda)} \qquad (104)\\ &= \lambda\sqrt{\frac{1+\lambda^2\underline{c}^T[I+\lambda F_s]^{-2}\underline{c}}{(1+\lambda a-\lambda^2\underline{c}^T[I+\lambda F_s]^{-1}\underline{c})^2}} \\ &= \frac{\sqrt{1+\lambda^2\underline{c}^T[I+\lambda F_s]^{-2}\underline{c}}}{\frac{1}{\lambda}+a-\lambda\underline{c}^T[I+\lambda F_s]^{-1}\underline{c}} \quad . \end{aligned}$$

Now, by Lemma 1, (104) becomes

$$\begin{aligned} \lambda\delta &= \frac{\sqrt{1+\underline{c}^T F_s^{-2}\underline{c}+O(\frac{1}{\lambda})}}{a-\underline{c}^T F_s^{-1}\underline{c}+O(\frac{1}{\lambda})} \\ &= \frac{\sqrt{1+\underline{c}^T F_s^{-2}\underline{c}}}{a-\underline{c}^T F_s^{-1}\underline{c}}+O(\frac{1}{\lambda}) \quad . \qquad (105) \end{aligned}$$

Because $\lambda \longrightarrow \infty$ as $\delta \longrightarrow 0$, and identifying a (72), we obtain the limit

$$\lim_{\delta\to 0}\lambda\delta = a\sqrt{1+\underline{c}^T F_s^{-2}\underline{c}} \quad . \tag{106}$$

That $\lim_{\delta\to 0}\lambda^*\delta$ is identically the right-hand side of (106) follows directly from the form of λ^* (90) and the identity (72).

We next derive an explicit expression for the slope of $B(\underline{\theta},\delta)$ (56) at $\delta = 0$. This expression will be used to develop an $o(\delta)$ approximation to $B(\underline{\theta},\delta)$.

Theorem 3 *The derivatives of the bound $B(\underline{\theta},\delta)$ and of the normalized difference between the unbiased CR bound and this bound, $\Delta B(\underline{\theta},\delta) = \frac{B(\underline{\theta},0)-B(\underline{\theta},\delta)}{B(\underline{\theta},0)}$, exist and are given by*

$$\frac{dB(\underline{\theta},\delta)}{d\delta}|_{\delta=0} = -2B(\underline{\theta},0)\sqrt{1+\eta^2} \quad , \quad \text{and} \quad \frac{d\Delta B(\underline{\theta},\delta)}{d\delta}|_{\delta=0} = 2\sqrt{1+\eta^2} \quad , \tag{107}$$

where η is the bias-sensitivity index defined by

$$\eta^2 \stackrel{\text{def}}{=} \|F_s^{-1}\underline{c}\|^2 = \underline{c}^T F_s^{-2}\underline{c} \quad . \tag{108}$$

Proof of Theorem 3. The existence of the derivatives will follow from the existence of $\lim_{\delta\to 0}\frac{\Delta B(\underline{\theta},\delta)}{\delta}$, which is derived in the following expressions.

For convenience we repeat (74):

$$B(\underline{\theta},\delta) = B(\underline{\theta},0) - \lambda\delta^2 - \alpha_\lambda\alpha - \underline{\beta}_\lambda^T\underline{\beta} \quad , \tag{109}$$

where

$$\begin{aligned} \alpha &= (a - \underline{c}^T F_s^{-1}\underline{c})^{-1} = B(\underline{\theta},0) \\ \underline{\beta} &= -\alpha F_s^{-1}\underline{c} \quad , \end{aligned} \tag{110}$$

and

$$\begin{aligned} \alpha_\lambda &= (1 + \lambda a - \lambda^2\underline{c}^T[I + \lambda F_s]^{-1}\underline{c})^{-1} \\ \underline{\beta}_\lambda &= -\lambda\alpha_\lambda[I + \lambda F_s]^{-1}\underline{c} \quad . \end{aligned} \tag{111}$$

The use of the identities for α, α_λ, $\underline{\beta}$, and $\underline{\beta}_\lambda$, (110) and (111), gives the following equation for the difference $B(\underline{\theta},0) - B(\underline{\theta},\delta)$:

$$\begin{aligned} B(\underline{\theta},0) - B(\underline{\theta},\delta) &= \lambda\delta^2 + \alpha_\lambda\alpha + \underline{\beta}_\lambda^T\underline{\beta} \\ &= \lambda\delta^2 + \alpha_\lambda\alpha + \lambda\alpha_\lambda\underline{c}^T[I + \lambda F_s]^{-1}F_s^{-1}\underline{c}\alpha \\ &= \left[\lambda\delta + \alpha(1 + \lambda\underline{c}^T[I + \lambda F_s]^{-1}F_s^{-1}\underline{c})\frac{\alpha_\lambda}{\delta}\right]\delta \quad , \end{aligned} \tag{112}$$

so that

$$\frac{B(\underline{\theta},0) - B(\underline{\theta},\delta)}{\delta} = \lambda\delta + \alpha(1 + \lambda\underline{c}^T[I + \lambda F_s]^{-1}F_s^{-1}\underline{c})\frac{\alpha_\lambda}{\delta} \quad . \tag{113}$$

Next we develop the following facts:

- From (93) of Lemma 1 and the forms of α_λ (111) and α (110),

$$\frac{\alpha_\lambda}{\delta} = \frac{1}{1 + \lambda a - \lambda^2\underline{c}^T[I + \lambda F_s]^{-1}\underline{c}}\left(\frac{1}{\delta}\right) \tag{114}$$

$$= \frac{1}{1+\lambda(a-\underline{c}^T F_s^{-1}\underline{c})+\lambda O(\frac{1}{\lambda})}\left(\frac{1}{\delta}\right)$$

$$= \frac{1}{1+\frac{\lambda}{\alpha}+O(1)}\left(\frac{1}{\delta}\right)$$

$$= \frac{1}{\frac{\lambda\delta}{\alpha}+O(\delta)} \quad .$$

- A completely analogous argument as used in proving (93) of Lemma 1 establishes

$$\lambda\underline{c}^T[I+\lambda F_s]^{-1}F_s^{-1}\underline{c} = \underline{c}^T F_s^{-2}\underline{c}+O(\frac{1}{\lambda})$$

$$= \underline{c}^T F_s^{-2}\underline{c}+O(\delta) \quad . \tag{115}$$

Substitute relations (114) and (115) into the right-hand side of (113) to obtain

$$\frac{B(\underline{\theta},0)-B(\underline{\theta},\delta)}{\delta} \tag{116}$$

$$= \lambda\delta+\alpha\left(1+\underline{c}^T F_s^{-2}\underline{c}+O(\delta)\right)\frac{1}{\frac{\lambda\delta}{\alpha}+O(\delta)} \quad .$$

Recalling the result of Proposition 2,

$$\lambda\delta \rightarrow \alpha\sqrt{1+\underline{c}^T F_s^{-2}\underline{c}} \quad , \text{ as } \delta \rightarrow 0 \quad , \tag{117}$$

take the limit of (116) as $\delta \rightarrow 0$ to obtain

$$-\frac{dB(\underline{\theta},\delta)}{d\delta}|_{\delta=0} = \lim_{\delta\rightarrow 0}\frac{B(\underline{\theta},0)-B(\underline{\theta},\delta)}{\delta}$$

$$= \alpha\sqrt{1+\underline{c}^T F_s^{-2}\underline{c}}+\alpha(1+\underline{c}^T F_s^{-2}\underline{c})\frac{1}{\frac{\alpha\sqrt{1+\underline{c}^T F_s^{-2}\underline{c}}}{\alpha}}$$

$$= 2\alpha\sqrt{1+\underline{c}^T F_s^{-2}\underline{c}} \quad . \tag{118}$$

Because $\alpha = B(\underline{\theta},0)$, the first identity of (107) is established. The additional observation that

$$\frac{d\Delta B(\underline{\theta},\delta)}{d\delta} = \frac{d}{d\delta}\left[\frac{B(\underline{\theta},0)-B(\underline{\theta},\delta)}{B(\underline{\theta},0)}\right] = -\frac{dB(\underline{\theta},\delta)}{d\delta}\frac{1}{B(\underline{\theta},0)} \tag{119}$$

establishes Theorem 3.

Theorem 3 can be used to approximate the form of the bound $B(\underline{\theta},\delta)$ up to $o(\delta)$ accuracy; in particular, $B(\underline{\theta},\delta) \approx B(\underline{\theta},0) + \delta\frac{dB(\underline{\theta},\delta)}{d\delta}|_{\delta=0}$ for sufficiently small δ. We have the following:

Theorem 4 *The lower bound $B(\underline{\theta},\delta)$ (56) on the variance of $\hat{\theta}_1$ has the representation*

$$B(\underline{\theta},\delta) = B^*(\underline{\theta},\delta) + o(\delta) \quad , \tag{120}$$

where

$$B^*(\underline{\theta},\delta) = B(\underline{\theta},0)\ \left(1 - 2\delta\sqrt{1 + \underline{c}^T F_s^{-2}\underline{c}}\right) \quad . \tag{121}$$

Define the vector $\underline{d}^{**}_{min}$:

$$\underline{d}^{**}_{min} = \frac{\delta}{\sqrt{1 + \underline{c}^T F_s^{-2}\underline{c}}}[-1, \underline{c}^T F_s^{-1}]^T \quad . \tag{122}$$

$\underline{d}^{**}_{min}$ is an $o(\delta)$ approximation to the minimizing vector of $B(\underline{\theta},\delta)$ in the sense that

$$[\underline{e}_1 + \underline{d}^{**}_{min}]^T F^{-1}[\underline{e}_1 + \underline{d}^{**}_{min}] = B(\underline{\theta},\delta) + o(\delta) \quad . \tag{123}$$

Proof of Theorem 4. The relation (120) is just a consequence of the form of the derivative (107) of $B(\underline{\theta},\delta)$ at $\delta = 0$, given in Theorem 3, and the Taylor expansion

$$\begin{aligned} B(\underline{\theta},\delta) &= B(\underline{\theta},0) + \delta\frac{dB(\underline{\theta},\delta)}{d\delta}|_{\delta=0} + o(\delta) \\ &= B(\underline{\theta},0) - 2B(\underline{\theta},0)\delta\sqrt{1 + \underline{c}^T F_s^{-2}\underline{c}} + o(\delta) \quad . \end{aligned} \tag{124}$$

To establish the second part of the theorem, (123) must be shown to hold. For notational convenience, define

$$\alpha_\infty \stackrel{def}{=} \frac{\delta}{\sqrt{1 + \underline{c}^T F_s^{-2}\underline{c}}} \tag{125}$$

so that $\underline{d}^{**}_{min} = \alpha_\infty[-1, \underline{c}^T F_s^{-1}]^T$. The substitution of $\underline{d}^{**}_{min}$ into the left-hand side of (123) and identification of the inverse Fisher matrix (70) gives

$$
\begin{aligned}
& [\underline{e}_1 + \underline{d}^{**}_{min}]^T F^{-1} [\underline{e}_1 + \underline{d}^{**}_{min}] \qquad (126) \\
&= [1 - \alpha_\infty, \alpha_\infty \underline{c}^T F_s^{-1}] \begin{bmatrix} \alpha & \underline{\beta}^T \\ \underline{\beta} & \Gamma \end{bmatrix} [1 - \alpha_\infty, \alpha_\infty \underline{c}^T F_s^{-1}]^T \\
&= (1 - \alpha_\infty)^2 \alpha + 2\alpha_\infty (1 - \alpha_\infty) \underline{c}^T F_s^{-1} \underline{\beta} + \alpha_\infty^2 \underline{c}^T F_s^{-1} \Gamma F_s^{-1} \underline{c} \quad .
\end{aligned}
$$

Now recall the definitions (72) for α, $\underline{\beta}$ and Γ:

$$
\begin{aligned}
\alpha &= (a - \underline{c}^T F_s^{-1} \underline{c})^{-1} = B(\underline{\theta}, 0) \qquad (127) \\
\underline{\beta} &= -\alpha F_s^{-1} \underline{c} \\
\Gamma &= F_s^{-1} + \alpha F_s^{-1} \underline{c}\underline{c}^T F_s^{-1} \quad .
\end{aligned}
$$

Substitution of (127) into (126) gives

$$
\begin{aligned}
& [\underline{e}_1 + \underline{d}^{**}_{min}]^T F^{-1} [\underline{e}_1 + \underline{d}^{**}_{min}] \qquad (128) \\
&= \alpha(1 - \alpha_\infty)^2 - 2\alpha\alpha_\infty (1 - \alpha_\infty) \underline{c}^T F_s^{-2} \underline{c} \\
&\quad + \alpha_\infty^2 \underline{c}^T F_s^{-1} [F_s^{-1} + \alpha F_s^{-1} \underline{c}\underline{c}^T F_s^{-1}] F_s^{-1} \underline{c} \\
&= \alpha[(1 - \alpha_\infty)^2 - 2\alpha_\infty (1 - \alpha_\infty) \underline{c}^T F_s^{-2} \underline{c} + \alpha_\infty^2 (\underline{c}^T F_s^{-2} \underline{c})^2] \\
&\quad + \alpha_\infty^2 \underline{c}^T F_s^{-3} \underline{c} \\
&= \alpha[1 - \alpha_\infty - \alpha_\infty \underline{c}^T F_s^{-2} \underline{c}]^2 + \alpha_\infty^2 \underline{c}^T F_s^{-3} \underline{c} \\
&= \alpha[1 - \alpha_\infty (1 + \underline{c}^T F_s^{-2} \underline{c})]^2 + \alpha_\infty^2 \underline{c}^T F_s^{-3} \underline{c} \quad .
\end{aligned}
$$

Finally, recalling the definition (125) of α_∞,

$$
\begin{aligned}
& [\underline{e}_1 + \underline{d}^{**}_{min}]^T F^{-1} [\underline{e}_1 + \underline{d}^{**}_{min}] \qquad (129) \\
&= \alpha[1 - \frac{\delta}{\sqrt{1 + \underline{c}^T F_s^{-2} \underline{c}}} (1 + \underline{c}^T F_s^{-2} \underline{c})]^2 + \delta^2 \frac{\underline{c}^T F_s^{-3} \underline{c}}{1 + \underline{c}^T F_s^{-2} \underline{c}} \\
&= B(\underline{\theta}, 0)[1 - \delta\sqrt{1 + \underline{c}^T F_s^{-2} \underline{c}}]^2 + o(\delta) \\
&= B^*(\underline{\theta}, \delta) + o(\delta) \\
&= B(\underline{\theta}, \delta) + o(\delta) \quad ,
\end{aligned}
$$

where in (129) we have used the facts $\alpha = B(\underline{\theta}, 0)$ (70), and $B(\underline{\theta}, \delta) = B^*(\underline{\theta}, \delta) + o(\delta)$ (120). This finishes the proof.

The following corollary to Theorem 4 will be needed for the sequel.

Corollary 2 *If for a given point $\underline{\theta}^o$ the Fisher matrix $F(\underline{\theta})$ is invertible at $\underline{\theta} = \underline{\theta}^o$ and its elements are uniformly continuous over an open neighborhood U of $\underline{\theta}^o$, then an open neighborhood $V \subset U$ of $\underline{\theta}^o$ exists such that in (120) $\frac{1}{\delta}o(\delta)$ converges to zero uniformly over $\underline{\theta} \in V$ as $\delta \to 0$.*

Proof of Corollary 2. Denote by $\|A\|$ the norm of the square matrix A ([3], Section 5.6) and define $E(\underline{\theta}) = F(\underline{\theta}) - F(\underline{\theta}^o)$. Because the elements of $F(\underline{\theta})$ are uniformly continuous over a neighborhood U of $\underline{\theta}^o$, $\|E(\underline{\theta})\|$ converges uniformly to zero as $\underline{\theta} \to \underline{\theta}^o$, and since $F(\underline{\theta}^o)$ is invertible, an open neighborhood $W \subset U$ of $\underline{\theta}^o$ exists such that $F(\underline{\theta}) = F(\underline{\theta}^o) + E(\underline{\theta})$ is invertible over $\underline{\theta} \in W$. Therefore, using the inequality $\|F^{-1}(\underline{\theta}) - F^{-1}(\underline{\theta}^o)\| \leq \|F^{-1}(\underline{\theta})\| \cdot \|F^{-1}(\underline{\theta}^o)\| \cdot \|E(\underline{\theta})\|$ ([3], p. 341, Exercise 13) $F^{-1}(\underline{\theta})$ converges to $F^{-1}(\underline{\theta}^o)$ uniformly over the neighborhood W of $\underline{\theta}^o$. Hence, $F(\underline{\theta})$ and $F^{-1}(\underline{\theta})$ are both uniformly continuous over the neighborhood W of $\underline{\theta}^o$. Now recall the definition (57) of the function $g(\lambda) = g_{\underline{\theta}}(\lambda)$: $g_{\underline{\theta}}(\lambda) \stackrel{\text{def}}{=} \underline{e}_1^T [I + \lambda F(\underline{\theta})]^{-2} \underline{e}_1$. Since $I + \lambda F(\underline{\theta})$ is invertible and uniformly continuous over the neighborhood U of $\underline{\theta}^o$, a similar argument establishes that $g_{\underline{\theta}}(\lambda)$ is uniformly continuous in $\underline{\theta}$ over a neighborhood W' of $\underline{\theta}^o$, where without loss of generality we can take $W' = W$. It can also be verified that for all $\delta > 0$ the function $B^*(\underline{\theta}, \delta)$ (121) is uniformly continuous in $\underline{\theta}$ over the neighborhood W of $\underline{\theta}^o$. Define $f(\underline{\xi}, \lambda) = g_{\underline{\theta}}(\lambda) - \delta^2$ where $\underline{\xi} \stackrel{\text{def}}{=} [\delta, \underline{\theta}^T]^T$. From the uniform continuity of $g_{\underline{\theta}}(\lambda)$ it follows that $f(\underline{\xi}, \lambda)$ is uniformly continuous in $\underline{\xi} \in \mathbb{R}^+ \times W$. Furthermore, applying the identity (6) to $f(\underline{\xi}, \lambda)$, the derivative $\nabla_\lambda f(\underline{\xi}, \lambda) = -2\underline{e}_1^T F(\underline{\theta})[I + \lambda F(\underline{\theta})]^{-3} \underline{e}_1$ is nonzero and continuous in λ for $\lambda > 0$. Define $\underline{\xi}^o = [\delta, \underline{\theta}^{oT}]^T$. By Theorem 1, for any $\delta > 0$ a unique point λ^o exists such that $f(\underline{\xi}^o, \lambda^o) = 0$. We can now apply the implicit function theorem ([10], Chapter 4, Theorem 15.1) to assert that an open neighborhood $V \subset W$ exists such that the solution $\lambda = T(\delta, \underline{\theta})$ to the equation $g_{\underline{\theta}}(\lambda) = \delta^2$ is uniformly continuous in $\underline{\theta}$ over $\underline{\theta} \in V$. Therefore, in view of the functional form (56), the bound $B(\underline{\theta}, \delta)$ is uniformly continuous in $\underline{\theta}$ over $\underline{\theta} \in V$. The remainder term $o(\delta)$ in (120) is thus equal to the difference $B(\underline{\theta}, \delta) - B^*(\underline{\theta}, \delta)$ of two uniformly continuous functions; therefore $\frac{1}{\delta}o(\delta)$ converges to zero uniformly over $\underline{\theta} \in V$. This establishes the corollary.

7. DISCUSSION

In this section issues related to the bound of Theorems 1 and 2 will be briefly discussed. Some general issues are of importance: Is the bound of Theorem 1 achievable with any practical estimator? What does the bound (56) imply about the inherent performance limitations of unbiased estimators in the presence of nuisance parameters?

7.1 Achievability of New Bound

If for all $\underline{\theta}$ in a region $D \subseteq \Theta$ an unbiased estimator achieves the (unbiased) CR bound, $B(\underline{\theta},0)$, on MSE (recall that MSE equals variance for unbiased estimators), then the estimator is called efficient over the region. An estimator that by virtue of its bias has MSE which is less than or equal to $B(\underline{\theta},0)$ for all $\underline{\theta}$ in a region $D \subseteq \Theta$, and strictly less than $B(\underline{\theta},0)$ for at least one $\underline{\theta} \in D$, is called superefficient over D. Assume that D is a finite rectangular region $D = D_1 \times \cdots \times D_n$, where D_k is an open interval, $k = 1,\ldots,n$, and also assume that $F(\underline{\theta})$ satisfies the uniform continuity properties over $U = D$ assumed in the Corollary to Theorem 4. While achievement of the bound $B(\underline{\theta},\delta)$ is not necessary for superefficiency, if for a sufficiently small value of δ (to be specified below) the variance of an essentially unbiased estimator $\hat{\theta}_1$ achieves $B(\underline{\theta},\delta)$ for all $\underline{\theta}$ in an open region D, then an open region $V \subset D$ exists such that $\hat{\theta}_1$ is a superefficient estimator over $\underline{\theta} \in V$. To be specific, because the bias gradient of $\hat{\theta}_1$ satisfies the condition (10), for all $\underline{\theta},\underline{\theta}^o \in D$

$$
\begin{aligned}
&|b_1(\underline{\theta}) - b_1(\underline{\theta}^o)| \\
&= |b_1(\underline{\theta}) - b_1(\theta_1^o,\theta_2,\ldots,\theta_n) + b_1(\theta_1^o,\theta_2,\ldots,\theta_n) - b_1(\theta_1^o,\theta_2^o,\ldots,\theta_n) \\
&\quad + b_1(\theta_1^o,\theta_2^o,\ldots,\theta_n) - \cdots - b_1(\theta_1^o,\ldots,\theta_{n-1}^o,\theta_n) + b_1(\theta_1^o,\ldots,\theta_{n-1}^o,\theta_n) - b_1(\underline{\theta}^o)| \\
&< |b_1(\underline{\theta}) - b_1(\theta_1^o,\theta_2,\ldots,\theta_n)| + |b_1(\theta_1^o,\theta_2,\ldots,\theta_n) - b_1(\theta_1^o,\theta_2^o,\ldots,\theta_n)| \\
&\quad + \cdots + |b_1(\theta_1^o,\ldots,\theta_{n-1}^o,\theta_n) - b_1(\underline{\theta}^o)| \\
&= \left|\int_{\theta_1^o}^{\theta_1} \frac{\partial b_1(u_1,\theta_2,\ldots,\theta_n)}{\partial u_1} du_1\right| + \left|\int_{\theta_2^o}^{\theta_2} \frac{\partial b_1(\theta_1^o,u_2,\theta_3,\ldots,\theta_n)}{\partial u_2} du_2\right| \\
&\quad + \cdots + \left|\int_{\theta_n^o}^{\theta_n} \frac{\partial b_1(\theta_1^o,\ldots,\theta_{n-1}^o,u_n)}{\partial u_n} du_n\right| \\
&\leq \delta \sum_{i=1}^{n} |\theta_i - \theta_i^o| \\
&\leq \delta \sum_{i=1}^{n} \max_{D_i} |\theta_i - \theta_i^o| \\
&= \delta M \quad , \qquad (130)
\end{aligned}
$$

where M is a positive constant independent of specific values $\underline{\theta}, \underline{\theta}^o \in D$:

$$M \stackrel{\text{def}}{=} \sum_{i=1}^{n} \max_{D_i} |\theta_i - \theta_i^o| \quad . \tag{131}$$

As D is finite, $\max_{D_i} |\theta_i - \theta_i^o| < \infty$ and M is finite. Assume without loss of generality that a point $\underline{\theta}^o \in D$ exists such that $b_1(\underline{\theta}^o) = 0$ (if no such point exists, pick an arbitrary $\underline{\theta}^o \in D$ and redefine $\hat{\theta}_1$ to be $\hat{\theta}_1 - b_1(\underline{\theta}^o)$). Assume $\text{var}_{\underline{\theta}}(\hat{\theta}_1) = B(\underline{\theta}, \delta)$ for all $\underline{\theta} \in D$. While it follows from Theorem 2 that for any $\delta > 0$, $B(\underline{\theta}, \delta) < B(\underline{\theta}, 0)$, $\forall \underline{\theta}$, we need to show that $MSE_{\underline{\theta}}(\hat{\theta}_1) \leq B(\underline{\theta}, 0)$, where strict inequality holds for at least one point $\underline{\theta}$. The form (120) of $B(\underline{\theta}, \delta)$ in Theorem 4 gives for all $\underline{\theta} \in D$

$$\begin{aligned} MSE_{\underline{\theta}}(\hat{\theta}_1) &= B(\underline{\theta}, \delta) + b_1^2(\underline{\theta}) \\ &= B(\underline{\theta}, 0)\left(1 - 2\delta\sqrt{1 + \underline{c}^T F_s^{-2} \underline{c}}\right) + o(\delta) + b_1^2(\underline{\theta}) \quad . \end{aligned} \tag{132}$$

Therefore, subtracting $B(\underline{\theta}, 0)$ from both sides of (132) and using the bound (130),

$$\begin{aligned} MSE_{\underline{\theta}}(\hat{\theta}_1) - B(\underline{\theta}, 0) &= -\delta[2B(\underline{\theta}, 0)\sqrt{1 + \underline{c}^T F_s^{-2} \underline{c}} - \frac{1}{\delta} o(\delta)] + b_1^2(\underline{\theta}) \\ &\leq -\delta[2B(\underline{\theta}, 0)\sqrt{1 + \underline{c}^T F_s^{-2} \underline{c}} - \frac{1}{\delta} o(\delta)] + M^2 \delta^2 \\ &= -\delta[2B(\underline{\theta}, 0)\sqrt{1 + \underline{c}^T F_s^{-2} \underline{c}} - \frac{1}{\delta} o(\delta) - M^2 \delta] \quad . \end{aligned} \tag{133}$$

Now, assuming uniform continuity and invertibility of $F(\underline{\theta})$ over $\underline{\theta} \in D$, by Corollary 2 a region $V \subset D$ exists such that the quantity $\frac{1}{\delta} o(\delta)$ converges to zero uniformly over $\underline{\theta} \in V$. Hence, since $2B(\underline{\theta}, 0)\sqrt{1 + \underline{c}^T F_s^{-2} \underline{c}} > 0$, a sufficiently small positive δ exists which is independent of $\underline{\theta}$, such that

$$MSE_{\underline{\theta}}(\hat{\theta}_1) < B(\underline{\theta}, 0), \quad \forall \underline{\theta} \in V \quad . \tag{134}$$

Therefore, $\hat{\theta}_1$ is a superefficient estimator over V.

The condition that the variance of $\hat{\theta}_1$ achieves the variance bound $B(\underline{\theta}, \delta)$ everywhere in D is unnecessarily restrictive. There are two necessary conditions for the achievability of $B(\underline{\theta}, \delta)$. First, a real bias function $b_1(\underline{\theta})$ must exist that has the minimizing vector $\underline{d}_{min}(\underline{\theta})$ as its gradient over D. If $\underline{d}_{min}$ is the gradient of b_1, the gradient $\nabla_{\underline{\theta}} \underline{d}_{min}$ is equal to the Hessian matrix of $b_1(\underline{\theta})$, which is always symmetric; therefore, the first necessary condition requires that $\nabla_{\underline{\theta}} \underline{d}_{min}$ be a symmetric matrix. Second, since $B(\underline{\theta}, \delta)$ is a CR-type bound, the sufficient condition for equality in the Schwarz inequality underlying the derivation of the CR bound must be satisfied. This latter condition can only be satisfied in the case of parameter estimation for exponential families of distributions ([11],

Theorem 1, [6], Chapter 1, Section 7). As these two necessary conditions cannot always be satisfied, the variance bound $B(\underline{\theta}, \delta)$ is generally not achievable over any region D.

If for some point $\underline{\theta}^o$ an efficient estimator exists for the vector $\underline{\theta}$ over an open neighborhood of $\underline{\theta}^o$, then an essentially unbiased locally superefficient estimator can be constructed. $\hat{\theta}_1$ is called an essentially unbiased locally superefficient estimator at the point $\underline{\theta}^o$ if $\hat{\theta}_1$ is essentially unbiased and if $MSE_{\underline{\theta}^o}(\hat{\theta}_1) < B(\underline{\theta}^o, 0)$ while $MSE_{\underline{\theta}}(\hat{\theta}_1) \leq B(\underline{\theta}, 0)$ for $\underline{\theta}$ over an open neighborhood of $\underline{\theta}^o$. Local superefficiency is a weaker property than global superefficiency because a locally superefficient estimator only requires superefficient performance in the neighborhood of a particular point $\underline{\theta}^o$ and it may have MSE which exceeds $B(\underline{\theta}, 0)$ outside this neighborhood.

Let $\underline{\theta}^o$ be some fixed point in Θ and let a $\delta > 0$ be specified. As in Corollary 2, assume that $F(\underline{\theta}^o)$ is invertible and that $F(\underline{\theta})$ is uniformly continuous over an open neighborhood U of $\underline{\theta}^o$. This assumption implies that an open neighborhood $V \subset U$ of $\underline{\theta}^o$ exists such that $F^{-1}(\underline{\theta})$ exists and is uniformly continuous over V. Now assume that $\hat{\underline{\theta}}^{eff}$ is an efficient estimator for $\underline{\theta}$ over the neighborhood V, i.e., $\hat{\underline{\theta}}^{eff}$ is unbiased over V and

$$E_{\underline{\theta}}[(\hat{\underline{\theta}}^{eff} - \underline{\theta})(\hat{\underline{\theta}}^{eff} - \underline{\theta})^T] = F^{-1}(\underline{\theta}), \quad \underline{\theta} \in V \qquad . \tag{135}$$

The claim is that the following estimator is essentially unbiased, locally superefficient in an open neighborhood of $\underline{\theta}^o$:

$$\hat{\theta}_1 \stackrel{\text{def}}{=} \theta_1^o + [\underline{e}_1 + \underline{d}_{min}(\underline{\theta}^o)]^T(\hat{\underline{\theta}}^{eff} - \underline{\theta}^o) \tag{136}$$

where $\underline{d}_{min}(\underline{\theta})$ is the vector given in Theorem 1. Because $E_{\underline{\theta}}[\hat{\underline{\theta}}^{eff}] = \underline{\theta}$, we have for the bias of $\hat{\theta}_1$

$$\begin{aligned} b_1(\underline{\theta}) &= E_{\underline{\theta}}[\hat{\theta}_1 - \theta_1] \qquad (137) \\ &= E_{\underline{\theta}}[\theta_1^o + [\underline{e}_1 + \underline{d}_{min}(\underline{\theta}^o)]^T(\hat{\underline{\theta}}^{eff} - \underline{\theta}^o) - \theta_1] \\ &= [\underline{e}_1 + \underline{d}_{min}(\underline{\theta}^o)]^T(\underline{\theta} - \underline{\theta}^o) - (\theta_1 - \theta_1^o) \\ &= [\underline{e}_1 + \underline{d}_{min}(\underline{\theta}^o)]^T(\underline{\theta} - \underline{\theta}^o) - \underline{e}_1^T(\underline{\theta} - \underline{\theta}^o) \\ &= \underline{d}_{min}^T(\underline{\theta}^o)(\underline{\theta} - \underline{\theta}^o), \quad \forall \underline{\theta} \in V \qquad . \end{aligned}$$

Hence, the bias gradient $\nabla b_1(\underline{\theta}^o)$ is simply $\underline{d}_{min}^T(\underline{\theta}^o)$, which by Theorem 1 satisfies the constraint $\underline{d}_{min}^T \underline{d}_{min} \leq \delta^2$ and the estimator $\hat{\theta}_1$ (136) is essentially unbiased over $\underline{\theta} \in V$. Furthermore, the MSE of $\hat{\theta}_1$ is equal to

$$MSE_{\underline{\theta}}(\hat{\theta}_1) = E_{\underline{\theta}}[(\hat{\theta}_1 - \theta_1)^2]$$

$$
\begin{aligned}
&= E_{\underline{\theta}}[(\theta_1^o + [\underline{e}_1 + \underline{d}_{min}(\underline{\theta}^o)]^T(\hat{\underline{\theta}}^{eff} - \underline{\theta}^o) - \theta_1)^2] \\
&= E_{\underline{\theta}}[([\underline{e}_1 + \underline{d}_{min}(\underline{\theta}^o)]^T(\hat{\underline{\theta}}^{eff} - \underline{\theta}^o) - \underline{e}_1^T[\underline{\theta} - \underline{\theta}^o])^2] \\
&= E_{\underline{\theta}}[([\underline{e}_1 + \underline{d}_{min}(\underline{\theta}^o)]^T(\hat{\underline{\theta}}^{eff} - \underline{\theta}) + [\underline{e}_1 + \underline{d}_{min}(\underline{\theta}^o)]^T[\underline{\theta} - \underline{\theta}^o] - \underline{e}_1^T[\underline{\theta} - \underline{\theta}^o])^2] \\
&= E_{\underline{\theta}}[([\underline{e}_1 + \underline{d}_{min}(\underline{\theta}^o)]^T(\hat{\underline{\theta}}^{eff} - \underline{\theta}) + \underline{d}_{min}^T(\underline{\theta}^o)[\underline{\theta} - \underline{\theta}^o])^2] \\
&= [\underline{e}_1 + \underline{d}_{min}(\underline{\theta}^o)]^T F^{-1}(\underline{\theta})[\underline{e}_1 + \underline{d}_{min}(\underline{\theta}^o)] + (\underline{d}_{min}^T(\underline{\theta}^o)[\underline{\theta} - \underline{\theta}^o])^2 \\
&= [\underline{e}_1 + \underline{d}_{min}(\underline{\theta}^o)]^T F^{-1}(\underline{\theta})[\underline{e}_1 + \underline{d}_{min}(\underline{\theta}^o)] + b_1^2(\underline{\theta}), \quad \forall \underline{\theta} \in V \quad . \qquad (138)
\end{aligned}
$$

Define the function $e(\underline{\theta})$

$$
\begin{aligned}
e(\underline{\theta}) &\stackrel{\text{def}}{=} MSE_{\underline{\theta}}(\hat{\theta}_1) - B(\underline{\theta}, 0) \qquad (139) \\
&= [\underline{e}_1 + \underline{d}_{min}(\underline{\theta}^o)]^T F^{-1}(\underline{\theta})[\underline{e}_1 + \underline{d}_{min}(\underline{\theta}^o)] - \underline{e}_1^T F^{-1}(\underline{\theta})\underline{e}_1 + b_1^2(\underline{\theta}) \quad .
\end{aligned}
$$

Note that when $\underline{\theta} = \underline{\theta}^o$, $MSE_{\underline{\theta}^o}(\hat{\theta}_1) = B(\underline{\theta}^o, \delta)$ so that $e(\underline{\theta}^o) < 0$ for all positive δ. Furthermore, $e(\underline{\theta})$ is uniformly continuous over $\underline{\theta} \in V$ because $F^{-1}(\underline{\theta})$ is uniformly continuous over $\underline{\theta} \in V$ and $b_1(\underline{\theta})$ (137) is linear. Hence, for any $\gamma > 0$ and any $\underline{\theta} \in V$ an $\epsilon = \epsilon(\gamma)$ independent of $\underline{\theta}$ exists such that

$$
|e(\underline{\theta}) - e(\underline{\theta}^o)| < \gamma \qquad (140)
$$

whenever $\|\underline{\theta} - \underline{\theta}^o\| < \epsilon(\gamma)$. Relation (140) implies

$$
e(\underline{\theta}) < e(\underline{\theta}^o) + \gamma \quad .
$$

As $e(\underline{\theta}^o) < 0$ a $\gamma = \gamma'$ exists sufficiently small so that $e(\underline{\theta}) < 0$ for all $\underline{\theta}$ in the neighborhood $\mathcal{O} = \|\underline{\theta} - \underline{\theta}^o\| < \epsilon(\gamma')$. Consequently, $MSE_{\underline{\theta}}(\hat{\theta}_1) < B(\underline{\theta}, 0)$ for all $\underline{\theta} \in \mathcal{O}$ and $\hat{\theta}_1$ is locally superefficient.

7.2 Properties of the New Bound

The behavior of the new bound will be treated by listing some of its properties. For convenience we repeat equation (107) for the slope of the difference $\Delta B(\underline{\theta}, \delta)$:

$$
\frac{d\Delta B(\underline{\theta}, \delta)}{d\delta}|_{\delta=0} = 2\sqrt{1 + \underline{c}^T F_s^{-2} \underline{c}} \quad , \qquad (141)
$$

where

$$\Delta B(\underline{\theta},\delta) = \frac{B(\underline{\theta},0) - B(\underline{\theta},\delta)}{B(\underline{\theta},0)} \quad . \tag{142}$$

Properties:

1. The slope (141) is positive (because F_s^{-2} is positive-definite), as expected because $B(\underline{\theta},\delta)$ is less than the unbiased CR bound $B(\underline{\theta},0)$. More significant is the fact that to $o(\delta)$ approximation the relation between $B(\underline{\theta},\delta)$ and $B(\underline{\theta},0)$ is multiplicative as a function of δ: $B(\underline{\theta},\delta) \approx B(\underline{\theta},0)(1 - 2\delta\sqrt{1 + \underline{c}^T F_s^{-2}\underline{c}})$ (recall Theorem 4). This provides indirect evidence for a potential for severe bias sensitivity of the CR bound.

2. The slope (141) of $\Delta B(\underline{\theta},\delta)$ is characterized by the length of the vector $F_s^{-1}\underline{c}$, a quantity which has been identified in Theorem 3 as the sensitivity index η,

$$\eta \stackrel{\text{def}}{=} \|F_s^{-1}\underline{c}\| = \sqrt{\underline{c}^T F_s^{-2}\underline{c}} \quad . \tag{143}$$

   η is a dimensionless quantity which measures the inherent sensitivity of the unbiased CR bound to bias in the sense that $-2\sqrt{1+\eta^2}$ is the per δ decrease of $B(\underline{\theta},\delta)$ relative to the unbiased CR bound $B(\underline{\theta},0)$. A useful form for η is given in terms of the elements of the inverse Fisher matrix F^{-1} (70) and (72)

$$\eta = \frac{\|\underline{\beta}\|}{\alpha} \quad . \tag{144}$$

3. The quantity η (143) can be interpreted in terms of the "joint estimability" of the set of parameters $\underline{\theta} = (\theta_1,\ldots,\theta_n)^T$ by recalling the CR matrix inequality (23) for the covariance of an unbiased vector estimator $\underline{\tilde{\theta}} = (\tilde{\theta}_1,\ldots,\tilde{\theta}_n)^T$,

$$\mathrm{cov}_{\underline{\theta}}(\underline{\tilde{\theta}}) \geq \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1n} \\ \vdots & \vdots & & \vdots \\ \sigma_{n1} & \sigma_{n2} & \ldots & \sigma_n^2 \end{bmatrix} \stackrel{\text{def}}{=} F^{-1} \quad . \tag{145}$$

   This gives a correspondence between η and the covariance between a set of optimal unbiased estimators, i.e., an efficient estimator $\hat{\theta}_i^{eff}$, $i = 1,\ldots,n$,

$$\begin{aligned} \|\underline{c}^T F_s^{-1}\| = \frac{\|\underline{\beta}\|}{\alpha} &= \|[\frac{\sigma_{12}}{\sigma_1^2},\ldots,\frac{\sigma_{1n}}{\sigma_1^2}]\| \\ &= \|[\rho_{12}\frac{\sigma_2}{\sigma_1},\ldots,\rho_{1n}\frac{\sigma_n}{\sigma_1}]\| \\ &= \left(\sum_{i=2}^{n}\rho_{1i}^2\frac{\sigma_i^2}{\sigma_1^2}\right)^{\frac{1}{2}} \quad , \end{aligned} \tag{146}$$

   where ρ_{ij} is the correlation coefficient between $\hat{\theta}_i^{eff}$ and $\hat{\theta}_j^{eff}$;

$$\rho_{ij} \stackrel{\text{def}}{=} \frac{\sigma_{ij}}{\sigma_i \sigma_j} \quad . \tag{147}$$

In light of (146), the set of important but difficult-to-estimate parameters is of particular significance; these are information-coupled with θ_1 ($\rho_{1j} \neq 0$) but their best unbiased estimators have large variance ($var(\hat{\theta}_j^{eff}) \approx \sigma_j^2 \gg 0$). η (143) is a measure of the propensity of these parameters to influence the estimator of θ_1.

4. Note that the slope (141) of $\Delta B(\underline{\theta}, \delta)$ is minimized for the case $\eta^2 = \underline{c}^T F_s^{-2} \underline{c} = 0$, as

$$\frac{d\Delta B(\underline{\theta}, \delta)}{d\delta}|_{\delta=0} = 2\sqrt{1 + \underline{c}^T F_s^{-2} \underline{c}} \quad \geq \quad 2 \quad . \tag{148}$$

Therefore, $B(\underline{\theta}, \delta)$ is the most stable with respect to bias when there is no information coupling between θ_1 and the other parameters so that $\underline{c} = \underline{0}$ and $\eta = 0$. Furthermore, if $\eta^2 = \underline{c}^T F_s^{-2} \underline{c} = 0$, $F_s^{-1} \underline{c}$ must be the zero vector and, from (85), the minimizing bias gradient has the form

$$\underline{d}_{min} = [-\delta, 0, \ldots, 0]^T \quad . \tag{149}$$

From the form of the gradient (149) the "best" biased estimator has mean

$$E_{\underline{\theta}}(\hat{\theta}_1) = (1 - \delta)\theta_1 + const \quad . \tag{150}$$

This mean can be obtained by a simple "shrinkage" of an efficient estimator $\hat{\theta}_1^{eff}$, if such exists;

$$\hat{\theta}_1 = (1 - \delta)(\hat{\theta}_1^{eff} - \theta_1^o) + \theta_1^o \quad . \tag{151}$$

and as discussed in Section 7.1, $\hat{\theta}_1$ is locally superefficient in a neighborhood of $\underline{\theta}^o$.

5. Let the Fisher information be such that the ith element of $\underline{c}^T F_s^{-1}$ dominates the other elements: $\underline{c}^T F_s^{-1} \approx [0, \ldots, 0, \zeta, 0, \ldots, 0]$, where as in (146) $\zeta \stackrel{\text{def}}{=} \rho_{1i}\sigma_i/\sigma_1$. This can occur if θ_i is a coupled but extremely hard-to-estimate parameter. Then, the sensitivity index is given by

$$\eta^2 = \underline{c}^T F_s^{-2} \underline{c} \approx \zeta^2 \quad . \tag{152}$$

and, under the additional assumption $\zeta >> 1$, the minimizing bias gradient (122) takes the approximate form

$$\underline{d}_{min} = [0, \ldots, 0, \delta, 0, \ldots, 0]^T \quad . \tag{153}$$

i.e.,

$$b_1(\underline{\theta}) = \delta\theta_i + const \quad . \tag{154}$$

In this case, a small bias due to information coupling between θ_1 and θ_i can give a substantial decrease in the general CR bound.

7.3 Practical Implementation of the Bound

From (146), the sensitivity index η is apparently dependent on the units used to represent the parameters $\theta_2, \ldots, \theta_n$. For example, if θ_1 is represented in μrads and θ_2 is represented in MW, the quantity σ_2/σ_1 in (146) will be much smaller than if θ_1 is expressed in rads and θ_2 is expressed in μW. Thus, a parameter may be rendered difficult to estimate solely by virtue of the choice of units used to represent the parameter. The choice of units is equivalent to the specification of a coordinate system to represent the parameters (see Section 3). This unit dependency is due to the fact that when taken alone the constraint on the bias gradient is not tied to a particular choice of coordinates for Θ and therefore does not adequately describe a user constraint on the bias. In order to use the results of this report, the coordinates for Θ must be specified, e.g., through the specification of a constraint ellipsoid (11) in Θ over which the bias $b_1(\underline{\theta})$ is allowed to vary by at most γ. This bias constraint (11) will naturally reflect the user's choice of units for the parameters. In Section 3 we transformed the user's units to make the constraint ellipsoid a sphere (12). The results derived in Sections 5 and 6 are valid in these transformed coordinates. The reason that we chose to work with spherical constraint regions is that for a nonspherical region we could not have simultaneously achieved properties 1. and 2. of Proposition 1 for our form of bias-gradient constraint (10). Thus, expression of the bias gradient in a transformation-induced spherical-constraint region guarantees maximal bound reduction, i.e., the greatest freedom on the value of $\nabla m_1(\underline{\theta})$ subject to the requirement (11), as compared with any other choice of coordinates.

8. APPLICATIONS

In this section we apply the previous results to estimation of standard deviation and estimation of the correlation coefficient, based on measurements of a correlated pair of IID zero-mean Gaussian random sequences. Analytical expressions are given for the unbiased CR bound $B(\underline{\theta},0)$ the sensitivity index η, and the approximation $\underline{d}^{**}_{min}$ in Theorem 4 for both of these estimation problems. Then, for the estimation of standard deviation, the exact bound $B(\underline{\theta},\delta)$ (56) of Theorem 1 is numerically evaluated and its behavior is compared to the behavior of the sensitivity index.

Consider the following covariance estimation problem. Available for observation are a pair of random sequences

$$\begin{aligned} \{X_{1i}\}_{i=1}^{m} &= X_{11},\ldots,X_{1m} \\ \{X_{2i}\}_{i=1}^{m} &= X_{21},\ldots,X_{2m} \end{aligned} \quad , \tag{155}$$

where $\{[X_{1i},X_{2i}]^T\}$ are IID Gaussian random vectors with mean zero and unknown covariance matrix Λ:

$$\Lambda \stackrel{\text{def}}{=} \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \quad . \tag{156}$$

Define the correlation coefficient

$$\rho \stackrel{\text{def}}{=} \frac{\sigma_{12}}{\sigma_1\sigma_2} \quad , \tag{157}$$

where σ_1 and σ_2 are the positive square roots of σ_1^2 and σ_2^2, respectively.

8.1 Estimation of Standard Deviation

The objective here is to estimate the standard deviation σ_1 of X_{1i} under the assumption that σ_2 and ρ are unknown. The likelihood function corresponding to this estimation problem has the form (A.2) from Appendix A

$$l(\sigma_1,\sigma_2,\rho) = -\frac{m}{2}\ln((2\pi)^2\sigma_1^2\sigma_2^2[1-\rho^2]) - \frac{m}{2(1-\rho^2)}\left(\frac{\overline{X_1^2}}{\sigma_1^2} - 2\rho\frac{\overline{X_1X_2}}{\sigma_1\sigma_2} + \frac{\overline{X_2^2}}{\sigma_2^2}\right) \quad , \tag{158}$$

where in (158) $\overline{X_1^2}$ is the sample variance of $\{X_{i1}\}_{i=1}^m$ and $\overline{X_1 X_2}$ is the sample covariance between $\{X_{i1}\}$ and $\{X_{i2}\}$ for known mean zero. The Fisher information matrix for σ_1, σ_2, and ρ is derived in Appendix A, Equation (A.11)

$$F = \frac{m}{1-\rho^2} \begin{bmatrix} \frac{2-\rho^2}{\sigma_1^2} & -\frac{\rho^2}{\sigma_1\sigma_2} & -\frac{\rho}{\sigma_1} \\ -\frac{\rho^2}{\sigma_1\sigma_2} & \frac{2-\rho^2}{\sigma_2^2} & -\frac{\rho}{\sigma_2} \\ -\frac{\rho}{\sigma_1} & -\frac{\rho}{\sigma_2} & \frac{1+\rho^2}{1-\rho^2} \end{bmatrix} \quad . \tag{159}$$

In the notation of (7), identify the parameter vector $\underline{\theta}$: $\theta_1 \stackrel{\text{def}}{=} \sigma_1$, $\theta_2 \stackrel{\text{def}}{=} \sigma_2$, and $\theta_3 \stackrel{\text{def}}{=} \rho$. The unbiased CR bound $B(\underline{\theta},0)$ (30), the sensitivity index η (143), and the $o(\delta)$ approximation $\underline{d}_{min}^{**}$ (122) are derived in Appendix A, Equations (A.24), (A.25), and (A.26), respectively,

$$B(\underline{\theta},0) = \frac{\sigma_1^2}{2m} \quad , \tag{160}$$

$$\eta^2 = \left(\frac{\sigma_2^2}{\sigma_1^2}\rho^2\right)\rho^2 + \left(\frac{\rho^2(1-\rho^2)}{\sigma_1^2}\right)(1-\rho^2) \quad , \tag{161}$$

$$\underline{\ell}_{min}^{**} = \frac{\delta}{\sqrt{1+\eta^2}} \left[-1, \; -\frac{\sigma_2}{\sigma_1}\rho^2, \; -\frac{\rho}{\sigma_1}(1-\rho^2)\right]^T \quad . \tag{162}$$

The following comments are of interest:

1. The CR bound $B(\underline{\theta},0)$ (160) on the variance of an unbiased estimator $\hat{\sigma}_1$ is functionally independent of the variance σ_2^2 of X_{2i} and of the correlation ρ between X_{1i} and X_{2i}.

2. If $\rho = 0$, then $\eta = 0$, and the general CR bound is minimally sensitive to bias-gradient length δ.

3. The squared sensitivity η^2 (161) is the convex combination of two terms: $\frac{\sigma_2^2}{\sigma_1^2}\rho^2$ and $\frac{\rho^2(1-\rho^2)}{\sigma_1^2}$. Since η^2 is inversely proportional to σ_1^2, the sensitivity of the CR bound to small bias can be significant when σ_1^2 is small.

4. For ρ^2 close to 0, the ratio of the first and second terms of (161) is approximately $\sigma_2^2\rho^2$. In this case, if $\sigma_2^2 \gg \frac{1}{\rho^2}$ the first term dominates η^2, while if $\sigma_2^2 \ll \frac{1}{\rho^2}$ the second term dominates. For ρ^2 close to 1, the ratio of the first to second term is approximately $\sigma_2^2/(1-\rho^2)^2$. In this case, if $\sigma_2^2 \gg (1-\rho^2)^2$ the first term dominates,

while if $\sigma_2^2 \ll (1-\rho^2)^2$ the second term dominates. In any case, it is seen that $\sigma_2^2 \gg 1$ brings about a much higher sensitivity index than $\sigma_2^2 \ll 1$, particularly for $\rho^2 \gg 0$.

A related problem is the estimation of the correlation coefficient ρ for the above pair of Gaussian measurements.

8.2 Estimation of Correlation Coefficient

Consider the estimation of $\theta_1 \stackrel{\text{def}}{=} \rho$, when $\theta_2 \stackrel{\text{def}}{=} \sigma_1$ and $\theta_3 \stackrel{\text{def}}{=} \sigma_2$ are unknown. For this case we have, from Appendix A, Equations (A.39), (A.40), and (A.41), respectively:

$$B(\underline{\theta},0) = \frac{1}{m}(1-\rho^2)^2 \quad , \tag{163}$$

$$\eta^2 = \frac{\sigma_1^2+\sigma_2^2}{2}\frac{\rho^2}{2(1-\rho^2)^2} \quad , \tag{164}$$

$$\underline{d}_{min}^{**} = \frac{\delta}{\sqrt{1+\eta^2}}\left[-1,\ -\frac{\rho}{2(1-\rho^2)}\sigma_1,\ -\frac{\rho}{2(1-\rho^2)}\sigma_2\right]^T \quad . \tag{165}$$

We make the following comments concerning the problem of estimation of ρ:

1. The CR bound $B(\underline{\theta},0)$ (163) on the variance of an unbiased estimator $\hat{\rho}$ has a form which is independent of the standard deviations σ_1 and σ_2. Hence, the MSE performance of an unbiased efficient estimator of ρ is also invariant to these quantities.
2. As occurs in estimation of σ_1, the case $\rho = 0$ corresponds to a CR bound which is minimally sensitive to small biases.
3. The form of the sensitivity index η (164) indicates that the CR bound is sensitive to small biases if the average $(\sigma_1^2+\sigma_2^2)/2$ is large and if there is significant correlation $|\rho| \gg 0$. This, along with the form of $\underline{d}_{min}^{**}$ (165), suggests that substantial improvement in estimator variance may be possible by using an estimator whose bias is not invariant to the standard deviations σ_1 and σ_2.

8.3 Numerical Evaluations

Surfaces for the normalized difference $\Delta B(\underline{\theta},\delta)$ (55) and the sensitivity index η were generated numerically for the variance estimation problem of Section 8.1. The matrix F (159) was input to a computer program which computes $\Delta B(\underline{\theta},\delta)$ and η as the parameter vector $\underline{\theta} = (\sigma_1,\sigma_2,\rho)$ ranges over the set $\{1\} \times [0,1000] \times [-1,1]$. For this example, δ was set to 0.001 and $m = 1$. In Figure 5 the quantity $\Delta B(\underline{\theta},\delta)$ is plotted over (σ_2,ρ) and in Figure 6 the sensitivity index is plotted over the same range of (σ_2,ρ). A comparison between Figures 5 and 6 supports the approximate small

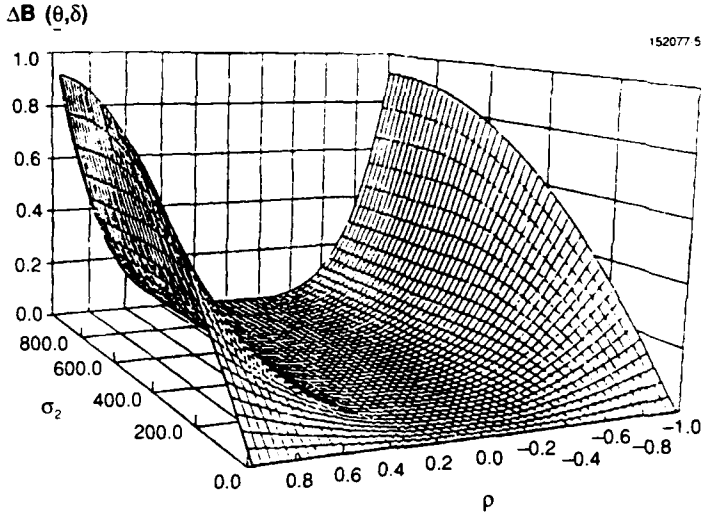


Figure 5. The normalized difference $\Delta B(\underline{\theta},\delta)$ plotted as a function of σ_2 and ρ for RMS power estimation. Increasing values of $\Delta B(\underline{\theta},\delta)$ correspond to increased sensitivity of the CR bound to small bias. Surface is plotted for $\delta = 0.001$ and $\sigma_1 = 1.0$.

δ analysis in Section 8.1, which was based on an investigation of the sensitivity index. Note that the region of ρ for which the bound $B(\underline{\theta},\delta)$ differs significantly from the unbiased bound becomes increasingly large as the standard deviation σ_2 of X_{2i} increases. Further, when $\rho = 0$ then $\eta = 0$ in (161), and we see minimal bias sensitivity. In the limit as $\sigma_2 \to \infty$ the surface $\Delta B(\underline{\theta},\delta)$ (Figure 5) becomes a deep wedge centered along the line $\rho = 0$; substantial bound reduction is achieved for all nonzero ρ. Figure 5 indicates that for small σ_2 the surface $\Delta B(\underline{\theta},\delta)$ is close to zero for all ρ. A simple calculation shows that $\rho^2(1-\rho^2)^2$ attains its maximum for $\rho^2 = 1/3$. Hence, in view of the expression (161), $\eta^2 \leq \sigma_2^2/\sigma_1^2 + 4/(27\sigma_1^2)$. Thus, for the present example where $\sigma_1 = 1.0$, $\eta^2 \leq \sigma_2^2 + 4/27$. Recalling that $\Delta B(\underline{\theta},\delta) \approx \delta \left(\frac{d\Delta B(\underline{\theta},\delta)}{d\delta}\right) = 2\delta\sqrt{1+\eta^2}$ (107), this implies that little bound reduction occurs for small σ_2 and $\delta = 0.001$.

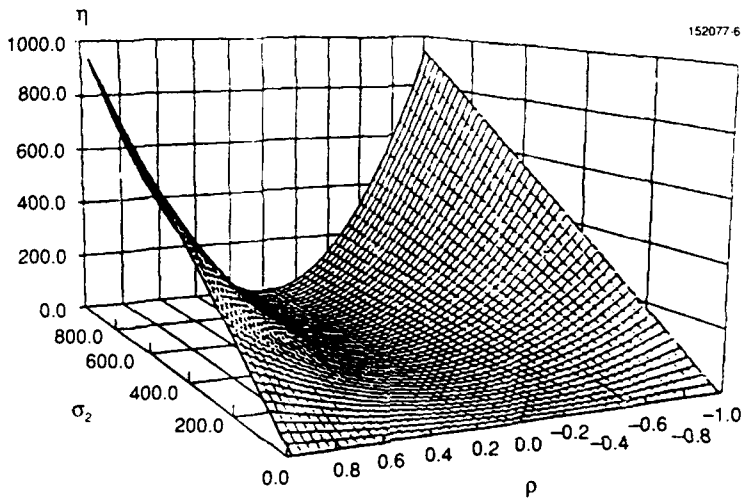


Figure 6. The bias-sensitivity index η plotted over the same range of the parameters as in Figure 5.

9. CONCLUSION

A new lower bound on estimator variance for almost unbiased estimators in the presence of nuisance parameters has been derived as a function of the Fisher information matrix and the unbiased CR bound. The new bound has the form (56) which involves finding the positive root λ of a convex function (57). The bound (56) is valid over all estimators whose bias gradient has length less than or equal to a prespecified constant δ. It reduces to the standard CR bound on unbiased estimators for $\delta = 0$. The sensitivity of the general CR bound to small bias has been characterized by the slope of the normalized difference between the new bound and the CR bound. This slope is monotone in a sensitivity index η (108). The form of the sensitivity index indicates that the new bound is significantly less than the unbiased CR bound when important but difficult-to-estimate nuisance parameters exist. This implies that the application of the CR bound is unreliable for this situation due to severe bias sensitivity.

We obtained numerical and analytical results for the problem of estimation of the standard deviations and the correlation coefficient for a pair of IID zero-mean Gaussian sequences. For estimation of the standard deviation of the first sequence it was shown that a small estimator bias can significantly affect the CR bound when the variance of the second sequence dominates the variance of the first and the correlation coefficient of the two sequences has high magnitude. For correlation estimation, it was shown that a small estimator bias can significantly affect the CR bound when the average of the two variances is high and the correlation coefficient has large magnitude.

APPENDIX

Here the Fisher matrix F for the estimation problem of Section 8 is derived and the quantities $B(\underline{\theta},0)$, η and $\underline{d}^{**}_{min}$ of Theorem 3 and Theorem 4 are calculated.

A.1 Fisher Information Matrix for Estimation of 2×2 Covariance Matrix

Since $\underline{X}_i = [X_{i1}, X_{i2}]^T$ are IID bivariate Gaussian random vectors with mean $\underline{0}$ and covariance Λ

$$\Lambda \stackrel{\text{def}}{=} \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \quad , \tag{A.1}$$

the log-likelihood function for σ_1, σ_2 and $\rho \stackrel{\text{def}}{=} \sigma_{12}/(\sigma_1\sigma_2)$, given the observations $\underline{X}_1, \ldots, \underline{X}_m$, is

$$\begin{aligned} l(\sigma_1,\sigma_2,\rho) &= \ln f(\underline{X}_1,\ldots,\underline{X}_m;\sigma_1,\sigma_2,\rho) \qquad (A.2) \\ &= -\frac{m}{2}\ln((2\pi)^2\sigma_1^2\sigma_2^2(1-\rho^2)) - \frac{1}{2(1-\rho^2)}\sum_{i=1}^{m}\left(\frac{X_{i1}^2}{\sigma_1^2} - \frac{2\rho}{\sigma_1\sigma_2}X_{i1}X_{i2} + \frac{X_{i2}^2}{\sigma_2^2}\right) \\ &= -\frac{m}{2}\ln((2\pi)^2\sigma_1^2\sigma_2^2(1-\rho^2)) - \frac{m}{2(1-\rho^2)}\left(\frac{\overline{X_1^2}}{\sigma_1^2} - \frac{2\rho}{\sigma_1\sigma_2}\overline{X_1X_2} + \frac{\overline{X_2^2}}{\sigma_2^2}\right) \quad , \end{aligned}$$

where $\overline{X_1^2}$, $\overline{X_2^2}$ and $\overline{X_1X_2}$ are the sample variances and the sample covariance associated with $\{X_{i1}\}$ and $\{X_{i2}\}$, respectively.

Next, the elements F_{ij} of the Fisher matrix F (22) are computed:

$$F = ((F_{ij})) = E\begin{bmatrix} -\frac{\partial^2 l}{\partial{\sigma_1}^2} & -\frac{\partial^2 l}{\partial\sigma_1\partial\sigma_2} & -\frac{\partial^2 l}{\partial\sigma_1\partial\rho} \\ -\frac{\partial^2 l}{\partial\sigma_2\partial\sigma_1} & -\frac{\partial^2 l}{\partial{\sigma_2}^2} & -\frac{\partial^2 l}{\partial\sigma_2\partial\rho} \\ -\frac{\partial^2 l}{\partial\rho\partial\sigma_1} & -\frac{\partial^2 l}{\partial\rho\partial\sigma_2} & -\frac{\partial^2 l}{\partial\rho^2} \end{bmatrix} \quad . \tag{A.3}$$

Using (A.2), the partial derivatives in F are simply computed. The results are

$$-\frac{\partial^2 l}{\partial{\sigma_1}^2} = -\frac{m}{\sigma_1^2}\left(1 - \frac{1}{1-\rho^2}\left[\frac{3\overline{X_1^2}}{\sigma_1^2} - 2\rho\frac{\overline{X_1X_2}}{\sigma_1\sigma_2}\right]\right) \tag{A.4}$$

$$-\frac{\partial^2 l}{\partial\sigma_1\partial\sigma_2} = -m\frac{\rho}{1-\rho^2}\frac{\overline{X_1X_2}}{\sigma_1^2\sigma_2^2} \tag{A.5}$$

$$-\frac{\partial^2 l}{\partial\sigma_1\partial\rho} = \frac{m}{\sigma_1(1-\rho^2)}\left(\frac{2\rho}{1-\rho^2}\left[\rho\overline{\frac{X_1X_2}{\sigma_1\sigma_2}} - \overline{\frac{X_1^2}{\sigma_1^2}}\right] + \overline{\frac{X_1X_2}{\sigma_1\sigma_2}}\right) \tag{A.6}$$

$$-\frac{\partial^2 l}{\partial{\sigma_2}^2} = -\frac{m}{\sigma_2^2}\left(1 - \frac{1}{1-\rho^2}\left[\overline{\frac{3X_2^2}{\sigma_2^2}} - 2\rho\overline{\frac{X_1X_2}{\sigma_1\sigma_2}}\right]\right) \tag{A.7}$$

$$-\frac{\partial^2 l}{\partial\sigma_2\partial\rho} = \frac{m}{\sigma_2(1-\rho^2)}\left(\frac{2\rho}{1-\rho^2}\left[\rho\overline{\frac{X_1X_2}{\sigma_1\sigma_2}} - \overline{\frac{X_2^2}{\sigma_2^2}}\right] + \overline{\frac{X_1X_2}{\sigma_1\sigma_2}}\right) \tag{A.8}$$

$$-\frac{\partial^2 l}{\partial\rho^2} = \frac{-m}{(1-\rho^2)^2}\left((1+\rho^2) - \frac{1+3\rho^2}{1-\rho^2}\left[\overline{\frac{X_1^2}{\sigma_1^2}} - 2\rho\overline{\frac{X_1X_2}{\sigma_1\sigma_2}} + \overline{\frac{X_2^2}{\sigma_2^2}}\right] + 4\rho\overline{\frac{X_1X_2}{\sigma_1\sigma_2}}\right) \quad , \tag{A.9}$$

where the remaining terms are omitted due to symmetry of F. As the sample averages in (A.4)–(A.9) are unbiased,

$$\begin{aligned} E[\overline{\frac{X_1^2}{\sigma_1^2}}] &= 1 \\ E[\overline{\frac{X_2^2}{\sigma_2^2}}] &= 1 \\ E[\overline{\frac{X_1X_2}{\sigma_1\sigma_2}}] &= \rho \quad ; \end{aligned} \tag{A.10}$$

the elements of F are easily computed by taking the expectation over the quantities in (A.4)–(A.9). The results are

$$\begin{aligned} F_{11} &= E[-\frac{\partial^2 l}{\partial{\sigma_1}^2}] = \frac{m}{1-\rho^2}\frac{2-\rho^2}{\sigma_1^2} \\ F_{12} &= E[-\frac{\partial^2 l}{\partial\sigma_1\partial\sigma_2}] = \frac{m}{1-\rho^2}\frac{(-\rho^2)}{\sigma_1\sigma_2} \\ F_{13} &= E[-\frac{\partial^2 l}{\partial\sigma_1\partial\rho}] = \frac{m}{1-\rho^2}\frac{(-\rho)}{\sigma_1} \\ F_{21} &= F_{12} \\ F_{22} &= E[-\frac{\partial^2 l}{\partial{\sigma_2}^2}] = \frac{m}{1-\rho^2}\frac{2-\rho^2}{\sigma_2^2} \\ F_{23} &= E[-\frac{\partial^2 l}{\partial\sigma_2\partial\rho}] = \frac{m}{1-\rho^2}\frac{(-\rho)}{\sigma_2} \end{aligned}$$

$$
\begin{aligned}
F_{31} &= F_{13} \\
F_{32} &= F_{23} \\
F_{33} &= E[-\frac{\partial^2 l}{\partial \rho^2}] = \frac{m}{1-\rho^2}\frac{1+\rho^2}{1-\rho^2} \quad .
\end{aligned} \tag{A.11}
$$

A.2 Bound Derivations for Estimation of 2×2 Covariance Matrix

Next, the sensitivity index η and the approximation to the minimizing bias-gradient vector $\underline{d}^{**}_{min}$ are derived for estimation of the standard deviation, $\theta_1 = \sigma_1$, and for correlation estimation, $\theta_1 = \rho$, respectively.

Estimation of Standard Deviation

In the case where σ_1 is of interest, identify the partition

$$
F = \begin{bmatrix} a & \underline{c}^T \\ \underline{c} & F_s \end{bmatrix} \quad , \tag{A.12}
$$

i.e., $a = F_{11}$, $\underline{c} = [F_{12}, F_{13}]^T$, and

$$
F_s = \begin{bmatrix} F_{22} & F_{23} \\ F_{32} & F_{33} \end{bmatrix} \quad . \tag{A.13}
$$

Specifically, in terms of (A.11),

$$
a = \frac{m}{1-\rho^2}\frac{2-\rho^2}{\sigma_1^2} \tag{A.14}
$$

$$
\underline{c} = -\frac{m}{1-\rho^2}[\frac{\rho^2}{\sigma_1\sigma_2}, \frac{\rho}{\sigma_1}]^T \tag{A.15}
$$

$$
F_s = \frac{m}{1-\rho^2}\begin{bmatrix} \frac{2-\rho^2}{\sigma_2^2} & \frac{-\rho}{\sigma_2} \\ \frac{-\rho}{\sigma_2} & \frac{1+\rho^2}{1-\rho^2} \end{bmatrix} \quad . \tag{A.16}
$$

Only the inverse F_s^{-1} is needed to compute expressions for η (143) and $\underline{d}^{**}_{min}$ (122):

$$
\eta^2 = \underline{c}^T F_s^{-2} \underline{c} = \|F_s^{-1}\underline{c}\|^2 \quad , \tag{A.17}
$$

$$\underline{d}^{**}_{min} = \frac{\delta}{\sqrt{1 + \underline{c}^T F_s^{-2} \underline{c}}} [-1, \underline{c}^T F_s^{-1}]^T \quad . \tag{A.18}$$

Using Cramer's rule, the inverse of F_s is, from (A.16),

$$F_s^{-1} = \frac{1}{|F_s|} \frac{m}{1-\rho^2} \begin{bmatrix} \frac{1+\rho^2}{1-\rho^2} & \frac{\rho}{\sigma_2} \\ \frac{\rho}{\sigma_2} & \frac{2-\rho^2}{\sigma_2^2} \end{bmatrix} \quad , \tag{A.19}$$

where $|F_s|$ is the determinant of F_s;

$$|F_s| = \left(\frac{m}{1-\rho^2}\right)^2 \frac{2}{\sigma_2^2} \frac{1}{1-\rho^2} \quad . \tag{A.20}$$

Now, in view of (A.15) and (A.19), the vector $\underline{c}^T F_s^{-1}$ is

$$\begin{aligned} \underline{c}^T F_s^{-1} &= -\frac{m}{1-\rho^2} \left[\frac{\rho^2}{\sigma_1\sigma_2}, \frac{\rho}{\sigma_1}\right] \frac{1}{|F_s|} \frac{m}{1-\rho^2} \begin{bmatrix} \frac{1+\rho^2}{1-\rho^2} & \frac{\rho}{\sigma_2} \\ \frac{\rho}{\sigma_2} & \frac{2-\rho^2}{\sigma_2^2} \end{bmatrix} \\ &= -\frac{1}{|F_s|} \left(\frac{m}{1-\rho^2}\right)^2 \left[\frac{\rho^2(1+\rho^2)}{\sigma_1\sigma_2(1-\rho^2)} + \frac{\rho^2}{\sigma_1\sigma_2}, \frac{\rho^3}{\sigma_1\sigma_2^2} + \frac{\rho(2-\rho^2)}{\sigma_1\sigma_2^2}\right] \\ &= -\frac{1}{|F_s|} \left(\frac{m}{1-\rho^2}\right)^2 \frac{1}{\sigma_1\sigma_2^2} \left[2\frac{\sigma_2\rho^2}{1-\rho^2}, 2\rho\right] \quad . \end{aligned} \tag{A.21}$$

Use the expression (A.20) for $|F_s|$ in (A.21) to obtain

$$\underline{c}^T F_s^{-1} = -\frac{1}{\sigma_1} \left[\sigma_2\rho^2, \rho(1-\rho^2)\right] \quad . \tag{A.22}$$

Now the quantity $\underline{c}^T F_s^{-1} \underline{c}$ is simply

$$\begin{aligned} \underline{c}^T F_s^{-1} \underline{c} &= -\frac{1}{\sigma_1} \left[\sigma_2\rho^2, \rho(1-\rho^2)\right] \begin{bmatrix} \frac{-\rho^2}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1} \end{bmatrix} \frac{m}{1-\rho^2} \\ &= \frac{m}{(1-\rho^2)\sigma_1} \left(\sigma_2\rho^2 \frac{\rho^2}{\sigma_1\sigma_2} + \rho(1-\rho^2)\frac{\rho}{\sigma_1}\right) \\ &- \frac{m}{(1-\rho^2)\sigma_1^2} \left(\rho^4 + \rho^2(1-\rho^2)\right) \end{aligned}$$

$$= \frac{m}{1-\rho^2}\frac{\rho^2}{\sigma_1^2} \quad . \tag{A.23}$$

Using (A.23), the unbiased CR bound $B(\underline{\theta},0)$ is next derived. From (30), the definition (A.12), and the identity (2),

$$\begin{aligned} B(\underline{\theta},0) &= \underline{e}_1^T F^{-1} \underline{e}_1 = \frac{1}{a - \underline{c}^T F_s^{-1} \underline{c}} \\ &= \frac{1}{\frac{m}{1-\rho^2}\frac{2-\rho^2}{\sigma_1^2} - \frac{m}{1-\rho^2}\frac{\rho^2}{\sigma_1^2}} \\ &= \frac{\sigma_1^2}{2m} \quad . \end{aligned} \tag{A.24}$$

The sensitivity index η^2 (A.17) is given by the squared magnitude of (A.22)

$$\begin{aligned} \eta^2 &= \underline{c}^T F_s^{-2} \underline{c} = \|\underline{c}^T F_s^{-1}\|^2 \\ &= \frac{1}{\sigma_1^2}\left\|\left[\sigma_2\rho^2,\ \rho(1-\rho^2)\right]\right\|^2 \\ &= \left(\frac{\sigma_2^2}{\sigma_1^2}\rho^2\right)\rho^2 + \left(\frac{\rho^2(1-\rho^2)}{\sigma_1^2}\right)(1-\rho^2) \quad . \end{aligned} \tag{A.25}$$

Finally, using (A.22) in (A.18), we have the following expression for $\underline{d}_{min}^{**}$:

$$\underline{d}_{min}^{**} = \frac{\delta}{\sqrt{1+\eta^2}}\left[-1,\ -\frac{\sigma_2}{\sigma_1}\rho^2,\ -\frac{\rho}{\sigma_1}(1-\rho^2)\right]^T \quad . \tag{A.26}$$

Estimation of Correlation

In the case where ρ is of interest, identify the partition

$$F = \begin{bmatrix} F_s^\rho & \underline{c}^\rho \\ \underline{c}^{\rho T} & a^\rho \end{bmatrix} \quad , \tag{A.27}$$

i.e., $a^\rho = F_{33}$, $\underline{c}^\rho = [F_{13}, F_{23}]^T$, and

$$F_s^\rho = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} \quad . \tag{A.28}$$

Specifically, in terms of the identities (A.11)

$$a^\rho = \frac{m}{1-\rho^2}\frac{1+\rho^2}{1-\rho^2} \tag{A.29}$$

$$\underline{c}^\rho = \frac{m}{1-\rho^2}[\frac{-\rho}{\sigma_1}, \frac{-\rho}{\sigma_2}]^T \tag{A.30}$$

$$F_s^\rho = \frac{m}{1-\rho^2}\begin{bmatrix} \frac{2-\rho^2}{\sigma_1^2} & \frac{-\rho^2}{\sigma_1\sigma_2} \\ \frac{-\rho^2}{\sigma_1\sigma_2} & \frac{2-\rho^2}{\sigma_2^2} \end{bmatrix} \quad . \tag{A.31}$$

Only the inverse $[F_s^\rho]^{-1}$ is needed to compute expressions (143) for η, (122) for $\underline{d}_{min}^{**}$, and (30) for $B(\underline{\theta},0)$ (Note: we use a reversal of the previous parameter ordering);

$$\eta^2 = [\underline{c}^\rho]^T[F_s^\rho]^{-2}\underline{c}^\rho \quad , \tag{A.32}$$

$$\underline{d}_{min}^{**} = \frac{\delta}{\sqrt{1+[\underline{c}^\rho]^T[F_s^\rho]^{-2}\underline{c}^\rho}}[-1, [\underline{c}^\rho]^T[F_s^\rho]^{-1}]^T \quad . \tag{A.33}$$

Using Cramer's rule, the inverse of $[F_s^\rho]$ is, from (A.31),

$$[F_s^\rho]^{-1} = \frac{1}{|F_s^\rho|}\frac{m}{1-\rho^2}\begin{bmatrix} \frac{2-\rho^2}{\sigma_2^2} & \frac{\rho^2}{\sigma_1\sigma_2} \\ \frac{\rho^2}{\sigma_1\sigma_2} & \frac{2-\rho^2}{\sigma_1^2} \end{bmatrix} \quad , \tag{A.34}$$

where $|F_s^\rho|$ is the determinant of F_s^ρ;

$$|F_s^\rho| = \left(\frac{m}{1-\rho^2}\right)^2 \frac{4}{\sigma_1^2\sigma_2^2}(1-\rho^2) \quad . \tag{A.35}$$

As a final form, combining (A.34) and (A.35) we obtain

$$[F_s^\rho]^{-1} = \frac{1}{4m}\begin{bmatrix} \sigma_1^2(2-\rho^2) & \rho^2\sigma_1\sigma_2 \\ \sigma_1\sigma_2\rho^2 & \sigma_2^2(2-\rho^2) \end{bmatrix} \quad . \tag{A.36}$$

Now, in view of (A.30) and (A.36), the vector $[\underline{c}^{\rho}]^T[F_s^{\rho}]^{-1}$ has the form

$$\begin{aligned}
[\underline{c}^{\rho}]^T[F_s^{\rho}]^{-1} &= \frac{m}{1-\rho^2}\left[\frac{-\rho}{\sigma_1}, \frac{-\rho}{\sigma_2}\right]\begin{bmatrix} \sigma_1^2(2-\rho^2) & \rho^2\sigma_1\sigma_2 \\ \rho^2\sigma_1\sigma_2 & \sigma_2^2(2-\rho^2) \end{bmatrix}\frac{1}{4m} \\
&= \frac{1}{4(1-\rho^2)}\left[-\sigma_1\rho(2-\rho^2)-\rho^3\sigma_1,\ -\sigma_2\rho(2-\rho^2)-\rho^3\sigma_2\right] \\
&= \frac{-\rho}{2(1-\rho^2)}[\sigma_1,\sigma_2] \quad . \qquad \text{(A.37)}
\end{aligned}$$

Now the quantity $[\underline{c}^{\rho}]^T[F_s^{\rho}]^{-1}\underline{c}^{\rho}$ is simply

$$\begin{aligned}
[\underline{c}^{\rho}]^T[F_s^{\rho}]^{-1}\underline{c}^{\rho} &= \frac{-\rho}{2(1-\rho^2)}[\sigma_1,\sigma_2]\begin{bmatrix} \frac{-\rho}{\sigma_1} \\ \frac{-\rho}{\sigma_2} \end{bmatrix}\frac{m}{1-\rho^2} \\
&= \frac{m\rho^2}{(1-\rho^2)^2} \quad . \qquad \text{(A.38)}
\end{aligned}$$

Using (A.38), the unbiased CR bound $B(\underline{\theta},0)$ is next derived. From (30), the definition (A.27), and an identity analogous to (2);

$$\begin{aligned}
B(\underline{\theta},0) &= \underline{e}_3^T F^{-1}\underline{e}_3 = \frac{1}{a^{\rho} - [\underline{c}^{\rho}]^T[F_s^{\rho}]^{-1}\underline{c}^{\rho}} \\
&= \frac{1}{\frac{m}{1-\rho^2}\frac{1+\rho^2}{1-\rho^2} - \frac{m\rho^2}{(1-\rho^2)^2}} \\
&= \frac{(1-\rho^2)^2}{m} \quad . \qquad \text{(A.39)}
\end{aligned}$$

The sensitivity index η^2 (A.32) is given by the squared magnitude of (A.37)

$$\begin{aligned}
\eta^2 &= \|[\underline{c}^{\rho}]^T[F_s^{\rho}]^{-1}\|^2 \\
&= \frac{\rho^2}{4(1-\rho^2)^2}\|[\sigma_1,\sigma_2]\|^2 \\
&= \frac{\sigma_1^2+\sigma_2^2}{2}\frac{\rho^2}{2(1-\rho^2)^2} \quad . \qquad \text{(A.40)}
\end{aligned}$$

Finally, using (A.37) in (A.33), we have the following expression for $\underline{d}^{**}_{min}$:

$$\underline{d}^{**}_{min} = \frac{\delta}{\sqrt{1+\eta^2}} \left[-1, \; -\sigma_1 \frac{\rho}{2(1-\rho^2)}, \; -\sigma_2 \frac{\rho}{2(1-\rho^2)} \right]^T \quad . \tag{A.41}$$

REFERENCES

1. P. Stoica and R.L. Moses, "On biased estimators and the Cramer-Rao lower bound," *Signal Processing*, Vol. 21, 349–351 (1990).
2. F.A. Graybill, *Matrices with Applications in Statistics*, Belmont, CA: Wadsworth (1983).
3. R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge, UK: Cambridge University Press (1988).
4. I.N. Herstein, *Topics in Algebra*, New York, NY: Wiley (1975).
5. J.D. Gorman and A.O. Hero, "Lower bounds on parametric estimation with constraints," *IEEE Transactions on Information Theory*, Vol. 36, No. 6, 1285–1301 (1990).
6. I.A. Ibragimov and R.Z. Has'minskii, *Statistical Estimation: Asymptotic Theory*, New York, NY: Springer-Verlag (1981).
7. W. Rudin, *Principles of Mathematical Analysis*, New York, NY: McGraw-Hill (1976).
8. G.H. Hardy, J.E. Littlewood, and G. Polya, *Inequalities*, Cambridge, UK: Cambridge University Press (1952).
9. A. Kuruc, "Lower bounds on multiple-source direction finding in the presence of direction-dependent antenna-array-calibration errors," Technical Report 799, MIT Lincoln Laboratory, Lexington, MA (November 1989). DTIC AD A215825.
10. K. Deimling, *Nonlinear Functional Analysis*, New York, NY: Springer-Verlag (1985).
11. A.V. Fend, "On the attainment of Cramer-Rao and Bhattacharyya bounds for the variance of an estimate," *Ann. Math. Stat.*, Vol. 30, 381–388 (1959).
12. C.R. Rao, *Linear Statistical Inference and Its Applications*, New York, NY: Wiley (1973).

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	3 January 1992	Technical Report

4. TITLE AND SUBTITLE
A Cramer-Rao Type Lower Bound for
Essentially Unbiased Parameter Estimation

5. FUNDING NUMBERS
C — F19628-90-C-0002
PE — 33401G
PR — 281

6. AUTHOR(S)
Alfred O. Hero

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Lincoln Laboratory, MIT
P.O. Box 73
Lexington, MA 02173-9108

8. PERFORMING ORGANIZATION REPORT NUMBER
TR-890

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Air Force
ESD/ENKL
Hanscom AFB, MA 01731-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ESD-TR-91-104

11. SUPPLEMENTARY NOTES
None

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

In this report a new Cramer-Rao (CR) type lower bound is derived which takes into account a user-specified constraint on the length of the gradient of estimator bias with respect to the set of underlying parameters. If the parameter space is bounded, the constraint on bias gradient translates into a constraint on the magnitude of the bias itself; the bound reduces to the standard unbiased form of the CR bound for unbiased estimation. In addition to its usefulness as a lower bound that is insensitive to small biases in the estimator, the rate of change of the new bound provides a quantitative bias "sensitivity index" for the general bias-dependent CR bound. An analytical form for this sensitivity index is derived which indicates that small estimator biases can make the new bound significantly less than the unbiased CR bound when important but difficult-to-estimate nuisance parameters exist. This implies that the application of the CR bound is unreliable for this situation due to severe bias sensitivity. As a practical illustration of these results, the problem of estimating elements of the 2×2 covariance matrix associated with a pair of independent identically distributed (IID) zero-mean Gaussian random sequences is presented.

14. SUBJECT TERMS		15. NUMBER OF PAGES
Cramer-Rao bounds	biased estimates	78
estimation theory	covariance	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR