# Lower Bounds For Parametric Estimation with Constraints

JOHN D. GORMAN, STUDENT MEMBER, IEEE, AND ALFRED O. HERO, MEMBER IEEE

*Abstract* —**A Chapman–Robbins form of the Barankin bound is used to derive a multiparameter Cramér–Rao (CR) type lower bound on estimator error covariance when the parameter $\boldsymbol{\theta} \in \mathbb{R}^n$ is constrained to lie in a subset of the parameter space. A simple form for the constrained CR bound is obtained when the constraint set $\Theta_C$ can be expressed as a smooth functional inequality constraint, $\Theta_C = \{\boldsymbol{\theta}: \mathscr{G}_{\boldsymbol{\theta}} \leq 0\}$. We show that the constrained CR bound is identical to the unconstrained CR bound at the regular points of $\Theta_C$, i.e. where no equality constraints are active. On the other hand, at those points $\boldsymbol{\theta} \in \Theta_C$ where pure equality constraints are active the full-rank Fisher information matrix in the unconstrained CR bound must be replaced by a rank-reduced Fisher information matrix obtained as a projection of the full-rank Fisher matrix onto the tangent hyperplane of the constraint set at $\boldsymbol{\theta}$. A necessary and sufficient condition involving the forms of the constraint and the likelihood function is given for the bound to be achievable, and examples for which the bound is achieved are presented. In addition to providing a useful generalization of the CR bound, our results permit analysis of the gain in achievable mse performance due to the imposition of particular constraints on the parameter space without the need for a global reparameterization. For the purpose of illustration, we apply the constrained bound to problems involving linear constraints and quadratic constraints. Specific examples considered include: linear constraints for Gaussian linear models, object support constraints in image reconstruction, signal subspace constraints in sensor array processing, and average power constraints in spectral estimation and signal extraction.**

*Index Terms* —**Constrained estimation, Cramér–Rao bounds, multiple parameter estimation, spectrum estimation.**

## I. Introduction

THE MULTIPLE PARAMETER Cramér–Rao (CR) lower bound is widely used to investigate the fundamental limits on estimator performance in multidimensional parameter estimation problems, and in single parameter estimation problems involving unknown nuisance parameters. The CR bound on estimator error covariance is computed as the inverse of the Fisher information matrix premultiplied and postmultiplied by the gradient of the mean vector of the estimator. Although elementary derivations, for instance [27, Section 2.4], may not explicitly make the assumption, the CR bound is typically derived under the assumption that the parameter space is an open subset of $\mathbb{R}^n$ [13, Section I.7]. Frequently, however, the parameter is constrained to lie in a proper non-open subset of the original parameter space. Some examples are: bandwidth, support, and positivity constraints in phase retrieval [5], [9] and tomographic reconstruction [24], [29]; kernel-sieve constraints in probability-density estimation [25]; array geometry constraints in estimation of coupled times-of-arrival across multiple-sensor arrays [28]; and auto-correlation lag constraints in maximum-entropy spectral analysis and image reconstruction [23]. Constraints restrict the allowable parameter variations and hence the local structure of the log-likelihood function over the constrained parameter space may be changed. Specifically, the average curvature of the log-likelihood function, and in particular the Fisher information matrix, may be affected, thereby invalidating the unconstrained CR bound.

We present a multiparameter CR type bound for parametric estimators when the vector parameter $\boldsymbol{\theta}$ is constrained to lie in a subset $\Theta_C$ of $\mathbb{R}^n$. We refer to this bound as a *constrained* CR bound. The constrained CR bound is derived directly from a version of the Barankin bound: the multiple parameter Chapman–Robbins bound. The tightest such Barankin bound is nonincreasing as $\Theta_C$ decreases. Thus, in general, a bound reduction occurs as a result of incorporating constraints. When $\boldsymbol{\theta}$ is a nonisolated point in a locally convex region of $\Theta_C$, and the log-likelihood function is smooth, the constrained CR bound depends on $\Theta_C$ only through the linear span of a set of basis vectors for the region. When the constraints on the parameter take the form of smooth functional inequality constraints $\mathscr{G}_{\boldsymbol{\theta}} \leq 0$ more explicit results are obtained. Specifically, let the inequality constraint be decomposed into a finite vector of equality constraints $\boldsymbol{G}_{\boldsymbol{\theta}} = 0$ and a finite vector of pure inequality constraints $\boldsymbol{H}_{\boldsymbol{\theta}} \leq 0$ (defined in Section II-C). Then the constrained CR bound is obtained by implementing the classical unconstrained CR bound with a different "constrained" Fisher matrix. The structure of the constrained Fisher matrix depends on whether or not $\boldsymbol{\theta}$ is a regular point of $\Theta_C$, where a regular point is a point where no equality constraints are active. As examples, points on the interior and points on the boundary of open regions in $\Theta_C$ are

Manuscript received April 26, 1989; revised November 27, 1989. This work was supported in part by the Office of Naval Research under contract N00014-86-C-0587 and in part by the National Cancer Institute of the National Institutes of Health, DHHS, under PHS Grant R01-CA46622-01. This work was presented in part at the Fourth Annual ASSP Workshop on Spectrum Estimation and Modeling, August 1988.

J. D. Gorman is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 and also with Environmental Research Institute of Michigan, Box 8618, Ann Arbor, MI 48107-8618.

A. O. Hero is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109.

IEEE Log Number 9038001.

0018-9448/90/1100-1285$01.00 

regular points. It is shown that if $\boldsymbol{\theta}$ is a regular point then the constrained Fisher matrix is identical to the unconstrained Fisher matrix for that point. Conversely, if $\boldsymbol{\theta}$ is not a regular point, the constrained Fisher matrix is the product of the unconstrained Fisher matrix and a $\boldsymbol{\theta}$-dependent, rank-deficient, idempotent matrix whose columns span a hyperplane that is tangent to the constraint set at $\boldsymbol{\theta}$.

The constrained CR bound presented here has the following attributes.

- For range constraints, orthant constraints, positivity constraints, and any other constraint sets $\Theta_C$ with no isolated boundaries, the constrained CR bound is identical to the unconstrained CR bound restricted to $\Theta_C$. Hence the incorporation of these types of constraints provides no CR bound reduction.
- For constraints which restrict $\boldsymbol{\theta}$ to a lower-dimensional manifold of parameter space, e.g., through active equality constraints of the form $\boldsymbol{G}_{\boldsymbol{\theta}} = 0$, the unconstrained CR bound is invalid and a reduced-rank Fisher matrix must be used.
- While an equivalent lower-dimensional unconstrained parameter estimation problem can sometimes be specified via a reparameterization of parameter space, such a global reparameterization is not necessary for the computation of the constrained CR bound. Rather, the constrained CR bound only depends on the local properties of the constraint set through its tangent hyperplanes. Since the tangent hyperplanes can typically be computed much more easily than can a global reparameterization of parameter space, the amount of bound reduction due to particular constraints is more easily analyzed.
- Conditions under which the constrained CR bound is achieved are similar to those required for achievement of the unconstrained CR bound. Examples are provided for which the constrained CR bound is achievable.

The following geometrical interpretation is helpful in interpreting the effect of constraints on the CR bound. The Fisher information matrix $J_{\boldsymbol{\theta}}$ being the expected value of the Hessian matrix of the ($n$-dimensional) log-likelihood surface at $\boldsymbol{\theta}$, can be related to the average curvature of the log-likelihood surface at $\boldsymbol{\theta}$ along $n$ different directions in $\mathbb{R}^n$. Thus the unconstrained CR bound is a function of the variation of the likelihood surface over an n-dimensional neighborhood of $\boldsymbol{\theta}$. When the parameter constraint $\boldsymbol{G}_{\boldsymbol{u}} = 0$, $\boldsymbol{u} \in \mathbb{R}^n$, is introduced, local parameter variations will generally be restricted to lie in a lower dimensional neighborhood. This neighborhood is contained in the linear vector space which is tangent to the constraint set $\{\boldsymbol{u}\colon \boldsymbol{G}_{\boldsymbol{u}} = 0\}$ at the point $\boldsymbol{u} = \boldsymbol{\theta}$. As the parameter varies over the lower-dimensional neighborhood, only certain "constrained" trajectories are traversed on the likelihood surface. Thus the average curvature of the surface appears different for the constrained parameter, as compared to the unconstrained parameter for which all local trajectories are allowed. This results in a change in the associated Fisher information matrix and a different CR bound. This constrained CR bound depends on the constraint set only through its tangent space at the point $\boldsymbol{\theta}$.

It is interesting to note that tangent space approximations to subsets of parameter space arise in general asymptotic statistical theory [15], [19] and specific applications have appeared in the statistical literature. For example tangent spaces arise in: the study [7] of the asymptotic distribution of the likelihood ratio for testing composite hypotheses involving smooth boundaries; the study [18] of the asymptotic distribution of a specific estimator arising in a composite detection problem with inequality constraints on the unknown parameter; the study [4] of asymptotic efficiency of estimators in partially parametric models; the study [1] of the asymptotic distribution of maximum likelihood estimators subject to equality constraints. While the study of finite sample CR bounds and the study of asymptotic properties of estimators have points in common, it is important to distinguish between the results of this paper and the aforementioned references. First, our result is a general finite sample CR lower bound on estimator covariance for fully parametric models. Second, the bound is of a simple and explicit form which is useful for studying the impact of particular parameter constraints on estimation error covariance. Third, while the CR bound holds for any estimator whose mean is smooth, the CR bound is not applicable to cases where the estimator has a nondifferentiable mean, such as the estimator considered in [18]. Furthermore, since the bound is a finite sample bound on covariance, methods of large sample theory are not needed for our derivation permitting a more elementary, and therefore more accessible, presentation.

To illustrate the utility of the constrained CR bound, we investigate the effect of constraints on the achievable estimator error for several representative problems in signal processing. First we consider the problem of estimation of parameters subject to linear constraints in the general linear Gaussian model. For this problem the tangent hyperplanes of the constraint set are functionally independent of the parameter $\boldsymbol{\theta}$, and hence the constrained CR lower bound can be achieved by projecting the unconstrained minimum variance unbiased (MVU) estimator onto the tangent hyperplane. The amount of bound reduction depends on the rank of the projection of the covariance matrix of the unconstrained MVU onto the linear constraint subspace.

Second, we consider the problem of image reconstruction subject to support constraints on the image. The constrained CR bound is equal to the pseudo-inverse of a constrained Fisher matrix, obtained by zeroing out the rows and columns of the unconstrained. Fisher information matrix which are associated with estimator errors outside of the region of support. It is significant that this is not generally the same as zeroing out rows and columns

of the unconstrained CR bound, unless the image pixels are statistically independent. This establishes that, if an efficient estimator of the unconstrained image exists, zeroing the unconstrained efficient estimator outside of the support region does not, in general, provide an efficient constrained estimator.

Third, power spectral density (PSD) estimation subject to average power constraints over disjoint frequency intervals, called frequency bands, is considered. For the case where the unconstrained Fisher information matrix is diagonal, corresponding to large observation time, it is shown that the constrained Fisher matrix is block diagonal. This means that average power constraints effectively couple the PSD estimation errors over a particular frequency band, but do not couple errors across different frequency bands. Within a particular frequency band where average power constraints are active, our results indicate that bound reduction is greatest over frequency bands where there are highly resolved spectral peaks, while there is virtually no reduction over bands where the true spectrum is smooth. This suggests that average power constraints make peaks easier to estimate but have little impact on the estimation of the rest of the spectrum.

Fourth, the estimation of the eigenvalues of a structured covariance matrix subject to signal subspace constraints is considered. We put this problem in the context of estimating the eigenvalues and eigenvectors of the array covariance matrix when it is known *a priori* that $p$ of the eigenvalues, the "signal dependent eigenvalues," are larger than the remaining eigenvalues, the "noise eigenvalues," and that these latter eigenvalues are identical. When the unconstrained Fisher matrix is block diagonal, the constrained CR bound can be achieved by averaging the noise eigenvalues of an efficient unconstrained estimator, if one exists.

Finally, we consider the problem of estimation of a deterministic time varying signal, and its Fourier transform, subject to average power constraints applied to its spectrum (squared Fourier magnitudes). Unlike the PSD estimation problem previously mentioned, here the constraints on the parameters (the signal) are nonlinear. Nonetheless, it is shown that if the unconstrained Fisher information is an identity matrix, e.g., corresponding to observation of the signal in additive-white-Gaussian noise, the structure of the constrained Fisher matrix is identical to the structure found in the PSD estimation problem, with the signal spectrum taking the place of the PSD.

An outline of the paper is as follows. Section II is divided into several subsections. In Section II-A a Barankin lower bound on the estimator covariance is given for general constrained parameters. In Section II-B the constrained CR bound is derived from this Barankin bound for locally convex regions of the constrained parameter space $\Theta_C$. In Section II-C the constrained CR bound of Section II-B is extended to the case of smooth nonlinear functional inequality constraints. In Section III, examples of the implementation of the constrained CR bound are presented.

## II. Lower Bounds on the Error Covariance

Throughout the paper the notation $\boldsymbol{\theta}$ and $[\theta_i]_{i=1,\cdots,n}$ will denote a column vector, $[\theta_1,\cdots,\theta_n]^T$, of unknown parameters contained in the unconstrained parameter space $\Theta=\mathbb{R}^n$. For a particular value of the vector $\boldsymbol{\theta}$ we specify a probability distribution $P_{\boldsymbol{\theta}}$ governing the observations $X$, taking values $x$ in a sample space $\Omega$. The collection of probability spaces $\mathscr{E}=\{(\Omega,\mathscr{F},P_{\boldsymbol{\theta}})\}_{\boldsymbol{\theta}\in\Theta}$ defines a $\boldsymbol{\theta}$-indexed set of possible models for $X$, and is called a *statistical experiment* over $\Theta$. If it is known that $\boldsymbol{\theta}$ is restricted to a subset of $\Theta$, called the constrained parameter space $\Theta_C$, the relevant statistical experiment becomes the reduced set of models $\mathscr{E}_C=\{(\Omega,\mathscr{F},P_{\boldsymbol{\theta}})\}_{\boldsymbol{\theta}\in\Theta_C}$. In this context, the constrained parameter estimation problem can be stated as follows: given a statistical experiment $\mathscr{E}_C$, a random variable $X$ is observed which has distribution $P_{\boldsymbol{\theta}}$; the objective is to specify an estimator $\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}(X)\in\Theta$ for the parameter vector $\boldsymbol{\theta}$. Define the vector mean $\boldsymbol{m}_{\boldsymbol{\theta}}\overset{\text{def}}{=}E_{\boldsymbol{\theta}}\{\hat{\boldsymbol{\theta}}\}$ of $\hat{\boldsymbol{\theta}}$, where $E_{\boldsymbol{\theta}}$ denotes expectation with respect to the distribution $P_{\boldsymbol{\theta}}$. The objective of this paper is to investigate the impact of parameter constraints on bounds for the minimum estimation error, where error is measured by the covariance matrix

$$\Sigma_{\boldsymbol{\theta}}\overset{\text{def}}{=}E_{\boldsymbol{\theta}}\left\{(\hat{\boldsymbol{\theta}}-\boldsymbol{m}_{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}-\boldsymbol{m}_{\boldsymbol{\theta}})^T\right\}. \tag{1}$$

We say that a matrix $B$ is a lower bound on a matrix $A$ if $A\geq B$ in the sense that $A-B$ is nonnegative definite.

### A. A Multiple Parameter Barankin Bound

We first present a Chapman–Robbins version of the multiple parameter Barankin lower bound on the covariance matrix $\Sigma_{\boldsymbol{\theta}}$ for the case where $\boldsymbol{\theta}\in\Theta_C$. Unlike the CR bound, the Barankin bound requires no regularity conditions on the distribution $P_{\boldsymbol{\theta}}$. To achieve a unified treatment of the cases of continuous and discrete random variables $X$, we let $P_{\boldsymbol{\theta}}$ have a density function $f_{\boldsymbol{\theta}}=f_{\boldsymbol{\theta}}(x)$ with respect to some reference measure $\mu$: $P_{\boldsymbol{\theta}}(A)=\int_A f_{\boldsymbol{\theta}}\,d\mu$, where $P_{\boldsymbol{\theta}}(A)$ is the probability that $X\in A$, $A\in\mathscr{F}$. For a continuous sample space $\Omega$ the previous integral can be interpreted as the standard (Lebesgue) integral over $A$, while for discrete $\Omega$, $\mu$ is the counting measure and the integral can be interpreted as a sum over elements $x\in A$.

For arbitrary vectors $\boldsymbol{\nu}_1,\cdots,\boldsymbol{\nu}_k\in\mathbb{R}^n$ and scalars $\Delta_1,\cdots,\Delta_k\in\mathbb{R}$, define the scalar and vector finite differences, $\delta_i f_{\boldsymbol{\theta}}$ and $\delta_i\boldsymbol{m}_{\boldsymbol{\theta}}$, of the density function and of the mean vector for $\hat{\boldsymbol{\theta}}$, respectively, which are produced by a change in the underlying parameter from the point $\boldsymbol{\theta}$ to the point $\boldsymbol{\theta}+\Delta_i\boldsymbol{\nu}_i$:

$$\delta_i f_{\boldsymbol{\theta}}\overset{\text{def}}{=}\frac{f_{\boldsymbol{\theta}+\Delta_i\boldsymbol{\nu}_i}-f_{\boldsymbol{\theta}}}{\Delta_i}, \tag{2}$$

$$\delta_i\boldsymbol{m}_{\boldsymbol{\theta}}\overset{\text{def}}{=}\frac{\boldsymbol{m}_{\boldsymbol{\theta}+\Delta_i\boldsymbol{\nu}_i}-\boldsymbol{m}_{\boldsymbol{\theta}}}{\Delta_i}. \tag{3}$$

These finite differences are the variations in $f_{\boldsymbol{\theta}}$ and $\boldsymbol{m}_{\boldsymbol{\theta}}$

along the directions of the vectors $\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_k$; a set of vectors which are henceforth referred to as *direction vectors*. Define the row vector of $k$ finite differences,

$$\delta f_{\boldsymbol{\theta}} \stackrel{\text{def}}{=} [\delta_1 f_{\boldsymbol{\theta}}, \cdots, \delta_k f_{\boldsymbol{\theta}}], \tag{4}$$

and the $n \times k$ matrix of finite differences

$$\delta \boldsymbol{m}_{\boldsymbol{\theta}} \stackrel{\text{def}}{=} [\delta_1 \boldsymbol{m}_{\boldsymbol{\theta}}, \cdots, \delta_k \boldsymbol{m}_{\boldsymbol{\theta}}]. \tag{5}$$

With these definitions we have the following multiple parameter Chapman–Robbins version of the Barankin bound [6], [17] when $\boldsymbol{\theta}$ is constrained to lie in the set $\Theta_C$.

*Proposition 1:* Let the $k+1$ vectors $\boldsymbol{\theta}, \boldsymbol{\theta} + \Delta_1 \boldsymbol{\nu}_1, \cdots, \boldsymbol{\theta} + \Delta_k \boldsymbol{\nu}_k$ be arbitrary points contained in the constrained parameter set $\Theta_C \subset \mathbb{R}^n$. Then for any estimator $\hat{\boldsymbol{\theta}}$ having mean $\boldsymbol{m}_{\boldsymbol{\theta}}$, the estimator error covariance matrix $\Sigma_{\boldsymbol{\theta}}$ satisfies the matrix inequality

$$\Sigma_{\boldsymbol{\theta}} \geq B_b \tag{6}$$

where

$$B_b = [\delta \boldsymbol{m}_{\boldsymbol{\theta}}] \cdot \left( E_{\boldsymbol{\theta}} \left[ \frac{\delta f_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}} \right]^T \left[ \frac{\delta f_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}} \right] \right)^+ \cdot [\delta \boldsymbol{m}_{\boldsymbol{\theta}}]^T, \tag{7}$$

and the plus sign denotes pseudo-inverse. Equality holds in (6) if and only if there exists a nonrandom $n \times k$ matrix $\Gamma$ such that the estimator $\hat{\boldsymbol{\theta}}$ satisfies

$$\hat{\boldsymbol{\theta}} - \boldsymbol{m}_{\boldsymbol{\theta}} = \Gamma \left[ \frac{\delta f_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}} \right]^T \quad \text{(w.p.1)}. \tag{8}$$

In Proposition 1, the pseudo-inverse of a matrix $A$ is defined as the unique matrix $A^+$ that satisfies the Moore–Penrose conditions [2, Ch. 3], [21, Section 1.b5]:

1) $AA^+$ and $A^+A$ are symmetric,
2) $AA^+A = A$,
3) $A^+AA^+ = A^+$. (9)

The conditions 1)–3) are a statement of the fact that $AA^+$ and $A^+A$ are projection operators onto the range of $A$ and $A^+$, respectively. Pseudo-inverses always exist, are continuous under certain conditions [26], and if $A$ is invertible $A^+ = A^{-1}$.

Before proving Proposition 1, we make the following observations. Since only a pseudo-inverse is required for the bound $B_b$ of Proposition 1, the covariance matrix, $E_{\boldsymbol{\theta}}[\delta f_{\boldsymbol{\theta}}/f_{\boldsymbol{\theta}}]^T[\delta f_{\boldsymbol{\theta}}/f_{\boldsymbol{\theta}}]$, of the finite difference vector does not have to be invertible. This general form is necessary for the present application since parameter constraints can reduce the rank of the covariance matrix. In view of the definition (4) of the finite difference vector $\delta f_{\boldsymbol{\theta}}$ the bound (6) is a measure of the variation of the probability density $f_{\boldsymbol{\theta}}$ relative to the set of "test" points $\boldsymbol{\theta} + \Delta_1 \boldsymbol{\nu}_1, \cdots, \boldsymbol{\theta} + \Delta_k \boldsymbol{\nu}_k$, which are arbitrarily specified in the constrained parameter space $\Theta_C$. On the other hand, since $\Theta_C \subset \Theta$, it is obvious that

$$\max_{\Theta} B_b \geq \max_{\Theta_C} B_b,$$

where each maximization is performed over the set of admissible test points in the parameter space. Hence constraining the parameter space can only reduce the (greatest) lower bound of the form (6). Thus it is clear that some bound reduction can occur due to incorporation of parameter constraints. Due to the difficulty in finding the best test vectors for (6), however, the amount of bound reduction is difficult to quantify in general. In the next section we will derive a constrained CR bound as a limiting form of the bound (6) for which the impact of constraints will be much easier to evaluate.

The proof of Proposition 1 depends on the following generalized version of the Cauchy–Schwarz inequality.

*Lemma 1:* Let $\boldsymbol{U} \in \mathbb{R}^n$ and $\boldsymbol{V} \in \mathbb{R}^k$ be random column vectors. Then

$$E_{\boldsymbol{\theta}}\{\boldsymbol{U}\boldsymbol{U}^T\} \geq E_{\boldsymbol{\theta}}\{\boldsymbol{U}\boldsymbol{V}^T\} \left[ E_{\boldsymbol{\theta}}\{\boldsymbol{V}\boldsymbol{V}^T\} \right]^+ E_{\boldsymbol{\theta}}\{\boldsymbol{V}\boldsymbol{U}^T\}, \tag{10}$$

where the plus sign denotes pseudo-inverse. Moreover, equality holds if and only if there is an $n \times k$ nonrandom matrix $\Gamma$ such that $\boldsymbol{U} = \Gamma \boldsymbol{V}$ w.p.1.

Note that if the $k \times k$ matrix $E_{\boldsymbol{\theta}}\{\boldsymbol{V}\boldsymbol{V}^T\}$ is nonsingular, the matrix inequality (10) is the standard Cauchy–Schwarz inequality for random vectors.

*Proof of Lemma 1:* Define the $\mathbb{R}^{n+k}$ vector $\boldsymbol{Z} = [\boldsymbol{U}^T \boldsymbol{V}^T]^T$. Then $E_{\boldsymbol{\theta}}\{\boldsymbol{Z}\boldsymbol{Z}^T\} \geq 0$ implies the matrix inequality

$$E_{\boldsymbol{\theta}}\{\boldsymbol{Z}\boldsymbol{Z}^T\} = E_{\boldsymbol{\theta}} \left\{ \begin{bmatrix} \boldsymbol{U}\boldsymbol{U}^T & \boldsymbol{U}\boldsymbol{V}^T \\ \boldsymbol{V}\boldsymbol{U}^T & \boldsymbol{V}\boldsymbol{V}^T \end{bmatrix} \right\} \geq 0.$$

Let $D$ be the $n \times (n+k)$ partitioned matrix

$$D = \left[ I \mid - E_{\boldsymbol{\theta}}\{\boldsymbol{U}\boldsymbol{V}^T\} \left[ E_{\boldsymbol{\theta}}\{\boldsymbol{V}\boldsymbol{V}^T\} \right]^+ \right],$$

where $I$ is the $n \times n$ identity. Since $E_{\boldsymbol{\theta}}\{\boldsymbol{Z}\boldsymbol{Z}^T\}$ is symmetric and nonnegative-definite, it has a nonnegative square root: $E_{\boldsymbol{\theta}}\{\boldsymbol{Z}\boldsymbol{Z}^T\} = E_{\boldsymbol{\theta}}^{1/2}\{\boldsymbol{Z}\boldsymbol{Z}^T\} E_{\boldsymbol{\theta}}^{T/2}\{\boldsymbol{Z}\boldsymbol{Z}^T\}$. Thus, $DE_{\boldsymbol{\theta}}\{\boldsymbol{Z}\boldsymbol{Z}^T\} D^T = [DE_{\boldsymbol{\theta}}^{1/2}\{\boldsymbol{Z}\boldsymbol{Z}^T\}][DE_{\boldsymbol{\theta}}^{1/2}\{\boldsymbol{Z}\boldsymbol{Z}^T\}]^T \geq 0$, and use of property 3) of (9) results in

$$E_{\boldsymbol{\theta}}\{\boldsymbol{U}\boldsymbol{U}^T\} - E_{\boldsymbol{\theta}}\{\boldsymbol{U}\boldsymbol{V}^T\} \left[ E_{\boldsymbol{\theta}}\{\boldsymbol{V}\boldsymbol{V}^T\} \right]^+ E_{\boldsymbol{\theta}}\{\boldsymbol{V}\boldsymbol{U}^T\} \geq 0.$$

This equation can be reexpressed as $E_{\boldsymbol{\theta}}\{(\boldsymbol{U} - \Gamma \boldsymbol{V})(\boldsymbol{V} - \Gamma \boldsymbol{V})^T\} \geq 0$, where $\Gamma \stackrel{\text{def}}{=} E_{\boldsymbol{\theta}}\{\boldsymbol{U}\boldsymbol{V}^T\}[E_{\boldsymbol{\theta}}\{\boldsymbol{V}\boldsymbol{V}^T\}]^+$. Equality holds if and only if the eigenvalues, $\lambda_i$, of the matrix $E_{\boldsymbol{\theta}}\{(\boldsymbol{U} - \Gamma \boldsymbol{V})(\boldsymbol{U} - \Gamma \boldsymbol{V})^T\}$ are zero. Furthermore, the nonnegative definiteness of this matrix implies that $\lambda_1 = \cdots = \lambda_n = 0$ if and only if $0 = \Sigma \lambda_i = \operatorname{tr}[E_{\boldsymbol{\theta}}\{(\boldsymbol{U} - \Gamma \boldsymbol{V})(\boldsymbol{V} - \Gamma \boldsymbol{V})^T\}] = E_{\boldsymbol{\theta}}\{(\boldsymbol{U} - \Gamma \boldsymbol{V})^T(\boldsymbol{U} - \Gamma \boldsymbol{V})\}$. Hence, equality holds in (10) if and only if $\boldsymbol{U} = \Gamma \boldsymbol{V}$ w.p.1. □

Using the previous Lemma, Proposition 1 is proven next.

*Proof of Proposition 1:* Define the $n$-vector $\boldsymbol{U}$ and the $k$-vector $\boldsymbol{V}$

$$\boldsymbol{U} \stackrel{\text{def}}{=} \hat{\boldsymbol{\theta}} - \boldsymbol{m}_{\boldsymbol{\theta}},$$

$$\boldsymbol{V} \stackrel{\text{def}}{=} \left[ \frac{\delta f_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}} \right]^T,$$

where $\boldsymbol{m}_{\boldsymbol{\theta}}$ is the mean vector of $\hat{\boldsymbol{\theta}}$ and $\delta f_{\boldsymbol{\theta}}$ is the vector of

finite differences defined in (4). With these definitions, application of Lemma 1 gives a lower bound involving the pseudo-inverse of the $k \times k$ matrix $E_{\boldsymbol{\theta}}\{\boldsymbol{V}\boldsymbol{V}^T\}$ and the $k \times n$ and $n \times k$ matrices $E_{\boldsymbol{\theta}}\{\boldsymbol{V}\boldsymbol{U}^T\}$ and $E_{\boldsymbol{\theta}}\{\boldsymbol{U}\boldsymbol{V}^T\}$, respectively. If it can be shown that $E_{\boldsymbol{\theta}}\{\boldsymbol{U}\boldsymbol{V}^T\} = \delta \boldsymbol{m}_{\boldsymbol{\theta}}$ Proposition 1 would be established. Consider the $j$th column of $E_{\boldsymbol{\theta}}\{\boldsymbol{U}\boldsymbol{V}^T\}$ and recall the definition (4) of $\delta f_{\boldsymbol{\theta}}$,

$$\begin{aligned}
\left[E_{\boldsymbol{\theta}}\{\boldsymbol{U}\boldsymbol{V}^T\}\right]_{*j} &= E_{\boldsymbol{\theta}}\left\{\left[\hat{\boldsymbol{\theta}} - \boldsymbol{m}_{\boldsymbol{\theta}}\right]\frac{\delta_j f_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}}\right\} \\
&= E_{\boldsymbol{\theta}}\left\{\left[\hat{\boldsymbol{\theta}} - \boldsymbol{m}_{\boldsymbol{\theta}}\right]\frac{f_{\boldsymbol{\theta}+\Delta_j \boldsymbol{\nu}_j} - f_{\boldsymbol{\theta}}}{\Delta_j f_{\boldsymbol{\theta}}}\right\} \\
&= \frac{E_{\boldsymbol{\theta}+\Delta_j \boldsymbol{\nu}_j}\{\hat{\boldsymbol{\theta}} - \boldsymbol{m}_{\boldsymbol{\theta}}\} - E_{\boldsymbol{\theta}}\{\hat{\boldsymbol{\theta}} - \boldsymbol{m}_{\boldsymbol{\theta}}\}}{\Delta_j} \\
&= \frac{\boldsymbol{m}_{\boldsymbol{\theta}+\Delta_j \boldsymbol{\nu}_j} - \boldsymbol{m}_{\boldsymbol{\theta}}}{\Delta_j} \\
&= \delta_j \boldsymbol{m}_{\boldsymbol{\theta}}. \qquad \square
\end{aligned}$$

### B. The Constrained CR bound

We first obtain a constrained CR bound for locally convex $\Theta_C$ directly from the bound (6). We then show that the same bound holds for points $\boldsymbol{\theta} \in \Theta_C$ at which $\Theta_C$ can be approximated by a union of locally convex sets. These results are then used in Section II-C to construct CR bounds when $\Theta_C$ is specified by continuously differentiable functional constraints.

Let $\boldsymbol{\theta}$ and the $k$ linearly independent test vectors $\boldsymbol{\theta} + \Delta_1 \boldsymbol{\nu}_1, \cdots, \boldsymbol{\theta} + \Delta_k \boldsymbol{\nu}_k$ be contained in the reduced parameter space $\Theta_C$ for all sufficiently small $\Delta_i$, $i = 1, \cdots, k$. Such test vectors can always be found for points $\boldsymbol{\theta}$ that are in locally convex regions of $\Theta_C$ with dimension at least $k$. Assuming the exchange of limiting and expectation operations is valid, the limit of the bound $B_b$, (6) of Proposition 1, as $\Delta_i \to 0$, $i = 1, \cdots, k$, gives a bound which depends only on the directional derivatives, $\lim_{\Delta_i \to 0} \delta_i f_{\boldsymbol{\theta}}$ and $\lim_{\Delta_i \to 0} \delta_i \boldsymbol{m}_{\boldsymbol{\theta}}$, of $f_{\boldsymbol{\theta}}$ and the mean vector, $\boldsymbol{m}_{\boldsymbol{\theta}}$, along the directions of the vectors $\boldsymbol{\nu}_i$, $i = 1, \cdots, k$, at the point $\boldsymbol{\theta}$. Specifically, by the chain rule we would have: $\lim_{\Delta_1, \cdots, \Delta_k \to 0} \delta f_{\boldsymbol{\theta}} = \nabla f_{\boldsymbol{\theta}} K$ and $\lim_{\Delta_1, \cdots, \Delta_k \to 0} \delta \boldsymbol{m}_{\boldsymbol{\theta}} = \nabla \boldsymbol{m}_{\boldsymbol{\theta}} K$, where $K = [\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_k]$ is the $n \times k$ matrix of direction vectors; $\nabla f_{\boldsymbol{\theta}}$ is the $1 \times n$ (row-vector) gradient of $f_{\boldsymbol{\theta}}$; and $\nabla \boldsymbol{m}_{\boldsymbol{\theta}}$ is the $n \times n$ matrix whose rows are the gradient vectors associated with each scalar component of $\boldsymbol{m}_{\boldsymbol{\theta}}$. If we could substitute the above limiting expressions into the right-hand side of (6) we would obtain

$$\Sigma_{\boldsymbol{\theta}} \geq [\nabla \boldsymbol{m}_{\boldsymbol{\theta}}] K \left[K^T J_{\boldsymbol{\theta}} K\right]^+ K^T [\nabla \boldsymbol{m}_{\boldsymbol{\theta}}]^T, \tag{11}$$

where

$$\begin{aligned}
J_{\boldsymbol{\theta}} &= E_{\boldsymbol{\theta}}\left\{\left[\frac{\nabla f_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}}\right]^T \left[\frac{\nabla f_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}}\right]\right\} \\
&= E_{\boldsymbol{\theta}}\{[\nabla \ln f_{\boldsymbol{\theta}}]^T [\nabla \ln f_{\boldsymbol{\theta}}]\}
\end{aligned} \tag{12}$$

is the $n \times n$ Fisher information matrix. Under appropriate regularity conditions [13, Lemma 8.1], [27, Section 2.4], the Fisher matrix is equivalent to

$$J_{\boldsymbol{\theta}} = -E_{\boldsymbol{\theta}}\{\nabla^2 \ln f_{\boldsymbol{\theta}}\}, \tag{13}$$

where $\nabla^2 \ln f_{\boldsymbol{\theta}}$ is the Hessian matrix of partial derivatives of $\ln f_{\boldsymbol{\theta}}$ with respect to elements of $\boldsymbol{\theta}$. This motivates the following lemma.

*Lemma 2:* Let the vector $\boldsymbol{\theta}$ be in the constrained parameter space $\Theta_C \subset \mathbb{R}^n$, and let $\{\boldsymbol{\nu}_i\}_{i=1}^k$ be $k$ linearly independent vectors such that $\boldsymbol{\theta} + \Delta_i \boldsymbol{\nu}_i \in \Theta_C$ for all sufficiently small $\Delta_i > 0$, $i = 1, \cdots, k$. Then for any estimator $\hat{\boldsymbol{\theta}}$ having mean $\boldsymbol{m}_{\boldsymbol{\theta}}$, the estimator error covariance matrix $\Sigma_{\boldsymbol{\theta}}$ satisfies the matrix inequality

$$\Sigma_{\boldsymbol{\theta}} \geq B_c \stackrel{\text{def}}{=} \limsup_{\Delta_1, \cdots, \Delta_k \to 0} B_b, \tag{14}$$

where $B_b$ is the bound (6) of Proposition 1. If in addition the following four regularity conditions hold:

- $\hat{\boldsymbol{\theta}}$ has finite variance: $\operatorname{var}\{\hat{\boldsymbol{\theta}}_i\} < \infty$; (15)
- $f_{\boldsymbol{\theta}}$ has continuous partial derivatives; (16)
- $E_{\boldsymbol{\theta}}\left\{\left|\frac{\partial \ln f_{\boldsymbol{\theta}}}{\partial \theta_i} \frac{\partial \ln f_{\boldsymbol{\theta}}}{\partial \theta_j}\right|\right\} < \infty$; (17)
- the matrix $E_{\boldsymbol{\theta}}\{[\nabla \ln f_{\boldsymbol{\theta}}]^T [\nabla \ln f_{\boldsymbol{\theta}}]\}$ is positive definite; (18)

then

$$B_c = [\nabla \boldsymbol{m}_{\boldsymbol{\theta}}] A \left[A^T J_{\boldsymbol{\theta}} A\right]^+ A^T [\nabla \boldsymbol{m}_{\boldsymbol{\theta}}]^T, \tag{19}$$

where $J_{\boldsymbol{\theta}}$ is the positive definite $n \times n$ Fisher matrix (12), and $A$ is any $n \times n$ matrix whose column space equals $\operatorname{span}\{\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_k\}$. Under these regularity conditions, equality is achieved in the lower bound (14) if and only if there exists a non-random $n \times n$ matrix $\Gamma$ such that:

$$\hat{\boldsymbol{\theta}} - \boldsymbol{m}_{\boldsymbol{\theta}} = \Gamma A^T [\nabla \ln f_{\boldsymbol{\theta}}]^T \quad \text{(w.p.1)}. \tag{20}$$

If such an estimator $\hat{\boldsymbol{\theta}}$ exists, this estimator is called an efficient constrained estimator.

*Proof of Lemma 2:* By assumption, $\boldsymbol{\theta} + \Delta_1 \boldsymbol{\nu}_1, \cdots, \boldsymbol{\theta} + \Delta_k \boldsymbol{\nu}_k$ are contained in $\Theta_C$ for all $\Delta_i$ sufficiently small, $i = 1, \cdots, k$, and the bound (14) follows directly from the Barankin bound of Proposition 1.

The regularity conditions (15)–(17) ensure that the Fisher matrix $J_{\boldsymbol{\theta}}$ (12) exists and has bounded elements [13, Section I.7], and condition (18) says that $J_{\boldsymbol{\theta}}$ is positive definite.

We first derive the limits as $\Delta_1, \cdots, \Delta_k \to 0$ of the matrices $E_{\boldsymbol{\theta}}\left[\frac{\delta f_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}}\right]^T \left[\frac{\delta f_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}}\right]$ and $\delta \boldsymbol{m}_{\boldsymbol{\theta}}$ under the stated regularity conditions of Lemma 2. Define $\Delta \stackrel{\text{def}}{=} \max_i |\Delta_i|$. Let $K$ be the $n \times k$ matrix with columns $\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_k$. By condition (16) and the chain rule

$$\begin{aligned}
\lim_{\Delta_1, \cdots, \Delta_k \to 0} \frac{\delta f_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}} &= \frac{1}{f_{\theta}} \nabla f_{\boldsymbol{\theta}} K \\
&= \nabla \ln f_{\boldsymbol{\theta}} K.
\end{aligned}$$

From this, and the stated continuity of $\nabla \ln f_\theta$, condition (16), the $ij$th element of $\left[\frac{\delta f_\theta}{f_\theta}\right]^T\left[\frac{\delta f_\theta}{f_\theta}\right]$ is dominated by $\Sigma_{l,m=1}^n K_{il}\left|\left[\frac{\partial \ln f_\theta}{\partial \theta_l}\frac{\partial \ln f_\theta}{\partial \theta_m}\right]\right| K_{mj} + O(\Delta)$, which has finite expectation by condition (17). Hence, by dominated convergence [3, Theorem 16.4], we have the finite limit

$$\begin{aligned}\lim_{\Delta_1,\cdots,\Delta_k\to 0} & E_\theta\left[\frac{\delta f_\theta}{f_\theta}\right]^T\left[\frac{\delta f_\theta}{f_\theta}\right] \\ &= K^T E_\theta\{[\nabla \ln f_\theta]^T[\nabla \ln f_\theta]\}K \\ &= K^T J_\theta K. \end{aligned} \tag{21}$$

Next consider the $n\times k$ matrix

$$\begin{aligned}\delta \boldsymbol{m}_\theta &= \left[\frac{\boldsymbol{m}_{\theta+\Delta_j \boldsymbol{\nu}_j} - \boldsymbol{m}_\theta}{\Delta_j}\right]_{j=1,\cdots,k} \\ &= \left[\frac{E_{\theta+\Delta_j\boldsymbol{\nu}_j}\{\hat{\boldsymbol{\theta}}\} - E_\theta\{\hat{\boldsymbol{\theta}}\}}{\Delta_j}\right]_{j=1,\cdots,k} \\ &= \left[E_\theta\left\{\hat{\boldsymbol{\theta}}\frac{f_{\theta+\Delta_j\boldsymbol{\nu}_j} - f_\theta}{\Delta_j}\frac{1}{f_\theta}\right\}\right]_{j=1,\cdots,k} \\ &= \left[E_\theta\left\{\hat{\boldsymbol{\theta}}\frac{\delta_j f_\theta}{f_\theta}\right\}\right]_{j=1,\cdots,k} = E_\theta\left\{\hat{\boldsymbol{\theta}}\frac{\delta f_\theta}{f_\theta}\right\}, \\ &= E_\theta\left\{(\hat{\boldsymbol{\theta}} - \boldsymbol{m}_\theta)\frac{\delta f_\theta}{f_\theta}\right\},\end{aligned}$$

where the last equality results from the identity $E_\theta\left[\frac{\delta f_\theta}{f_\theta}\right] = 0$. Now from condition (16) the elements of the $n\times k$ matrix $(\hat{\boldsymbol{\theta}} - \boldsymbol{m}_\theta)\delta f_\theta / f_\theta$ are equal to the elements of $(\hat{\boldsymbol{\theta}} - \boldsymbol{m}_\theta)\nabla \ln f_\theta K$ to order $O(\Delta)$. The Schwarz inequality and the regularity conditions (15) and (17) can be used to establish that the elements of the latter matrix have finite absolute expectation

$$\begin{aligned}|E_\theta\{(\hat{\boldsymbol{\theta}} - \boldsymbol{m}_\theta)\nabla \ln f_\theta\}_{i,j}| &= \left|E_\theta\left\{(\hat{\theta}_i - [\boldsymbol{m}_\theta]_i)\frac{\partial \ln f_\theta}{\partial \theta_j}\right\}\right| \\ &\le \operatorname{var}^{1/2}\{\hat{\theta}_i\}E_\theta^{1/2}\left\{\left|\frac{\partial \ln f_\theta}{\partial \theta_j}\right|^2\right\} \\ &< \infty.\end{aligned}$$

Hence, by dominated convergence, the limit

$$\lim_{\Delta_1,\cdots,\Delta_k\to 0} \delta \boldsymbol{m}_\theta$$

exists and is equal to the finite matrix

$$\begin{aligned}\lim_{\Delta_1,\cdots,\Delta_k\to 0} \delta \boldsymbol{m}_\theta &= E_\theta\left\{\hat{\boldsymbol{\theta}}\frac{\nabla f_\theta}{f_\theta}\right\}K \\ &= \nabla E_\theta\{\hat{\boldsymbol{\theta}}\}K \\ &= \nabla \boldsymbol{m}_\theta K.\end{aligned} \tag{22}$$

Since the columns, $\{\boldsymbol{\nu}_i\}_{i=1}^k$, of $K$ are linearly independent, by condition (18) $K^T J_\theta K$ is a full rank invertible matrix and $[K^T J_\theta K]^+ = [K^T J_\theta K]^{-1}$. Since the matrix $K^T J_\theta K$ is symmetric and positive definite the eigenvalues of the perturbed matrix $K^T J_\theta K + E$ are positive for a sufficiently small matrix perturbation $E$ [12, Corollary 6.3.4]. This implies that the inverse of $K^T J_\theta K$ is continuous in perturbations of its elements

$$\begin{aligned}\left(E_\theta\left[\frac{\delta f_\theta}{f_\theta}\right]^T\left[\frac{\delta f_\theta}{f_\theta}\right]\right)^+ &= \left(K^T J_\theta K + \underline{\underline{O}}(\Delta)\right)^+ \\ &= [K^T J_\theta K]^+ + \underline{\underline{o}}(1),\end{aligned} \tag{23}$$

where $\underline{\underline{O}}(\Delta)$ and $\underline{\underline{o}}(1)$ are matrices whose elements are of order $O(\Delta)$ and of order $o(1)$, respectively. In view of (21) we therefore have

$$\begin{aligned}\limsup B_b = & \lim_{\Delta_1,\cdots,\Delta_k\to 0}[\delta \boldsymbol{m}_\theta] \\ & \cdot \lim_{\Delta_1,\cdots,\Delta_k\to 0}\left(E_\theta\left[\frac{\delta f_\theta}{f_\theta}\right]^T\left[\frac{\delta f_\theta}{f_\theta}\right]\right)^+ \\ & \cdot \lim_{\Delta_1,\cdots,\Delta_k\to 0}[\delta \boldsymbol{m}_\theta]^T \\ &= [\nabla \boldsymbol{m}_\theta]K[K^T J_\theta K]^+ K^T[\nabla \boldsymbol{m}_\theta]^T.\end{aligned} \tag{24}$$

It remains to show that the bound (24) depends only on the range space of $K = [\boldsymbol{\nu}_1,\cdots,\boldsymbol{\nu}_k]$. Let $A$ be an $n\times n$ matrix whose column span is identical to the span of $\boldsymbol{\nu}_1,\cdots,\boldsymbol{\nu}_k$. Since the column spaces of $A$ and $K$ are identical, there exists an invertible $n\times n$ matrix $T$ such that

$$[K \quad O_1]T = A,$$

where $O_1$ is an $n\times(n-k)$ matrix of zeros. Let $O_2$ and $O_3$ be $(n-k)\times(n-k)$ and $k\times(n-k)$ matrices of zeros, respectively. Then,

$$\begin{aligned}A[A^T J_\theta A]^+ A^T & \\ &= [K \quad O_1]T\left[T^T\begin{bmatrix}K^T\\ O_1^T\end{bmatrix}J_\theta[K \quad O_1]T\right]^+ T^T\begin{bmatrix}K^T\\ O_1^T\end{bmatrix} \\ &= [K \quad O_1]\left[\begin{bmatrix}K^T\\ O_1^T\end{bmatrix}J_\theta[K \quad O_1]\right]^+\begin{bmatrix}K^T\\ O_1^T\end{bmatrix} \\ &= [K \quad O_1]\begin{bmatrix}K^T J_\theta K & O_3\\ O_3^T & O_2\end{bmatrix}^+\begin{bmatrix}K^T\\ O_1^T\end{bmatrix} \\ &= [K \quad O_1]\begin{bmatrix}[K^T J_\theta K]^+ & O_3\\ O_3^T & O_2\end{bmatrix}\begin{bmatrix}K^T\\ O_1^T\end{bmatrix} \\ &= K[K^T J_\theta K]^+ K^T,\end{aligned}$$

where the second equality follows from (65) of Lemma 5 in the Appendix.

The condition for equality in the bound (14), under the regularity conditions (15)–(17), can be obtained by mak-

ing the identifications $U=(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$, $V=K^T[\nabla \ln f_{\boldsymbol{\theta}}]^T$ in Lemma 1, verifying that the right side of the resultant bound (10) is identical to the right side of the bound (14) and invoking the necessary and sufficient condition for equality in (10): $U=\Gamma V$ for some $k\times n$ matrix $\Gamma$. This gives:

$$\hat{\boldsymbol{\theta}}-\boldsymbol{m}_{\boldsymbol{\theta}}=\Gamma K^T[\nabla \ln f_{\boldsymbol{\theta}}]^T \quad \text{(w.p.1)}.$$

Since $A$ has the identical column span as $K$, the above is equivalent to condition (20). □

The constrained CR bound (19) of Lemma 2 is in a general form that is applicable to nonisolated points $\boldsymbol{\theta}$ in locally convex regions of the parameter space $\Theta_C$. It is significant that, unlike the Barankin bound of Proposition 1, the constrained CR bound (19) only depends on the test points through the span of the set $\{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_k\}$. In particular, when $\Theta_C$ is only $p$-dimensional in the neighborhood of $\boldsymbol{\theta}$, and $p<n$, all $p$-dimensional sets of test points are equivalent in the sense that the limit (19) of the Barankin bound is the same.

The construction of Lemma 2 requires that $\Theta_C$ be locally convex or star-shaped in the neighborhood of $\boldsymbol{\theta}$. Lemma 2 can be extended to include nonisolated points in regions of $\Theta_C$ that have the property that local neighborhoods can be approximated to order $o(\Delta)$ by locally convex neighborhoods. The result is the following lemma.

*Lemma 3:* Let the vector $\boldsymbol{\theta}$ be in the constrained parameter space $\Theta_C\subset\mathbb{R}^n$, and let $\{\boldsymbol{v}_i\}_{i=1}^k$ be $k$ linearly independent vectors such that $\boldsymbol{\theta}+\Delta_i\boldsymbol{v}_i+\boldsymbol{o}(\Delta_i)\in\Theta_C$ for all $\Delta_i$ sufficiently small, $i=1,\cdots,k$, where $\boldsymbol{o}(\Delta_i)$ is a $\mathbb{R}^n$ vector whose length is of order $o(\Delta_i)$. Then the conclusions of Lemma 2 remain valid when the vectors $\boldsymbol{\theta}+\Delta_i\boldsymbol{v}_i$ are replaced by $\boldsymbol{\theta}+\Delta_i\boldsymbol{v}+\boldsymbol{o}(\Delta_i)$, $i=1,\cdots,k$.

*Proof of Lemma 3:* Similarly to (2), let $\delta' f_{\boldsymbol{\theta}}$ denote the $k$-length vector of scalar differences $\delta' f_{\boldsymbol{\theta}}=[\delta_1' f_{\boldsymbol{\theta}},\cdots,\delta_k' f_{\boldsymbol{\theta}}]$ where

$$\delta_i' f_{\boldsymbol{\theta}} \stackrel{\text{def}}{=} \frac{f_{\boldsymbol{\theta}+\Delta_i\boldsymbol{v}_i+o(\Delta_i)}-f_{\boldsymbol{\theta}}}{\Delta_i}. \tag{25}$$

Define $\delta'\boldsymbol{m}_{\boldsymbol{\theta}}$ similarly. Let $B_b'$ denote the Barankin bound of Proposition 1 formed with the $k$ test points $\{\boldsymbol{\theta}+\Delta_1\boldsymbol{v}_1+\boldsymbol{o}(\Delta_1),\cdots,\boldsymbol{\theta}+\Delta_k\boldsymbol{v}_k+\boldsymbol{o}(\Delta_k)\}$. We need to establish that the limits $\limsup_{\Delta_1,\cdots,\Delta_k\to 0} B_b'$ and $\limsup_{\Delta_1,\cdots,\Delta_k\to 0} B_b$ (14) are identical.

By assumption (16) $f_{\boldsymbol{\theta}}$ is continuous and therefore: $f_{\boldsymbol{\theta}+\Delta_i\boldsymbol{v}_i+\boldsymbol{o}(\Delta_i)}=f_{\boldsymbol{\theta}+\Delta_i\boldsymbol{v}_i}+\boldsymbol{o}(\Delta_i)$. In view of (25) this implies

$$\begin{aligned}\frac{\delta' f_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}}&=\frac{1}{f_{\boldsymbol{\theta}}}\left[\frac{f_{\boldsymbol{\theta}+\Delta_j\boldsymbol{v}_j}-f_{\boldsymbol{\theta}}}{\Delta_j}+\frac{\boldsymbol{o}(\Delta_j)}{\Delta_j}\right]_{j=1,\cdots,k}\\&=\frac{\delta f_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}}+\frac{1}{f_{\boldsymbol{\theta}}}\left[\frac{\boldsymbol{o}(\Delta_j)}{\Delta_j}\right]_{j=1,\cdots,k}\\&=\nabla\ln f_{\boldsymbol{\theta}}K+\boldsymbol{o}(1).\end{aligned}$$

Using the definition of the Fisher matrix and the continuity of the inverse of the full rank matrix $K^TJ_{\boldsymbol{\theta}}K$,

$$\begin{aligned}\left(E_{\boldsymbol{\theta}}\left[\frac{\delta' f_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}}\right]^T\left[\frac{\delta' f_{\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}}}\right]\right)^+&=\left(K^TJ_{\boldsymbol{\theta}}K+\underline{\underline{o}}(1)\right)^+\\&=\left[K^TJ_{\boldsymbol{\theta}}K\right]^{-1}+\underline{\underline{o}}(1),\end{aligned} \tag{26}$$

where $\underline{\underline{o}}(1)$ is a matrix that has $o(1)$ entries that go to zero as the $\Delta_i$'s go to zero. In a similar manner it can be shown that $\delta'\boldsymbol{m}_{\boldsymbol{\theta}}=\nabla\boldsymbol{m}_{\boldsymbol{\theta}}K+\underline{\underline{o}}(1)$, which, when taken with (26), implies $B_b'=B_c+\underline{\underline{o}}(1)$. This establishes the lemma. □

### C. Functional Constraints

Often the constrained parameter space $\Theta_C$ can be defined in terms of an implicit functional inequality constraint of the form

$$\mathscr{G}_{\boldsymbol{\theta}}\le 0, \tag{27}$$

where $\mathscr{G}=[\mathscr{G}^1,\cdots,\mathscr{G}^q]^T$ is a vector function on $\mathbb{R}^n$, $\mathscr{G}$: $\mathbb{R}^n\to\mathbb{R}^q$, and the inequality is to be interpreted element by element. We will assume that the inequality constraints are consistent, i.e., there exists at least one $\boldsymbol{\theta}\in\mathbb{R}^n$ that satisfies (27), and that $\mathscr{G}$ is continuously differentiable in the sense that the $q\times n$ gradient matrix

$$\nabla\mathscr{G}_{\boldsymbol{\theta}}=\begin{bmatrix}\nabla\mathscr{G}_{\boldsymbol{\theta}}^1\\ \vdots\\ \nabla\mathscr{G}_{\boldsymbol{\theta}}^q\end{bmatrix}=\begin{bmatrix}\dfrac{\partial\mathscr{G}_{\boldsymbol{\theta}}^1}{\partial\theta_1} & \cdots & \dfrac{\partial\mathscr{G}_{\boldsymbol{\theta}}^1}{\partial\theta_n}\\ \vdots & & \vdots\\ \dfrac{\partial\mathscr{G}_{\boldsymbol{\theta}}^q}{\partial\theta_1} & \cdots & \dfrac{\partial\mathscr{G}_{\boldsymbol{\theta}}^q}{\partial\theta_n}\end{bmatrix}, \tag{28}$$

exists and has continuous elements.

With the parameterization (27) of $\Theta_C$, the boundary of $\Theta_C$ is defined as the set of points where at least one component, $\mathscr{G}_{\boldsymbol{\theta}}^i$, of the vector function $\mathscr{G}_{\boldsymbol{\theta}}$ is equal to zero. The interior of $\Theta_C$ is defined as the set $\{\boldsymbol{\theta}\colon \mathscr{G}_{\boldsymbol{\theta}}<0\}$, where the strict inequality means $\mathscr{G}^i<0$, for each $i=1,\cdots,q$.

Note that equality constraints can be imbedded in (27) by letting $\mathscr{G}_{\boldsymbol{\theta}}^i=-\mathscr{G}_{\boldsymbol{\theta}}^j$ for some $i,j$, $i\ne j$. It is customary to extract the equality constraints from the inequality constraints (27), denoting what remains as *pure inequality constraints*. This yields the equivalent description of $\Theta_C$

$$\boldsymbol{G}_{\boldsymbol{\theta}}=0, \tag{29}$$

$$\boldsymbol{H}_{\boldsymbol{\theta}}\le 0, \tag{30}$$

where $\boldsymbol{G}=[G^1,\cdots,G^k]^T$ and $\boldsymbol{H}=[H^1,\cdots,H^l]^T$ are vector functions of $\boldsymbol{\theta}$, $\boldsymbol{G}$: $\mathbb{R}^n\to\mathbb{R}^k$, $\boldsymbol{H}$: $\mathbb{R}^n\to\mathbb{R}^l$. We will say that the equality constraint (29) is *active* if it restricts $\boldsymbol{\theta}$ to a lower dimensional subset of $\mathbb{R}^n$. Otherwise the equality constraint is said to be *inactive*.

The decomposition (29) and (30) is accomplished by partitioning the constraint set $\Theta_C$ into a set of *regular points* and *nonregular points*.

*Definition [16, Section 9.4]:* The point $\boldsymbol{\theta}_o \in \mathbb{R}^n$ is called a regular point of the inequality $\mathscr{G}_{\boldsymbol{\theta}} \le 0$ (a regular point of the constraint set $\Theta_C$) if: $\mathscr{G}_{\boldsymbol{\theta}_o} \le 0$ and if there exists a $\boldsymbol{\nu} \in \mathbb{R}^n$ such that $\mathscr{G}_{\boldsymbol{\theta}_o} + \nabla \mathscr{G}_{\boldsymbol{\theta}_o} \boldsymbol{\nu} < 0$.

There can be no active equality constraints at a regular point $\boldsymbol{\theta}_o$. Specifically, it can be shown that $\boldsymbol{\theta}_o$ is a regular point of $\Theta_C$ if and only if $\mathscr{G}_{\boldsymbol{\theta}_o + \Delta \boldsymbol{\nu}} < 0$ for some $\boldsymbol{\nu} \in \mathbb{R}^n$ and all sufficiently small $\Delta > 0$ (see proof of Lemma 4). This implies that there exits a sequence of interior points (e.g., $\{\boldsymbol{\theta}_o + \frac{1}{n}\boldsymbol{\nu}\}_n$) that converge to $\boldsymbol{\theta}_o$. Hence regular points are points that are in the closure of the interior of $\Theta_C$. In particular, all interior points of $\Theta_C$ are regular points and points on the boundary of pure inequality constraints $\boldsymbol{H}_{\boldsymbol{\theta}} \le 0$ are regular points. See Figs. 1 and 2 for graphical illustrations.

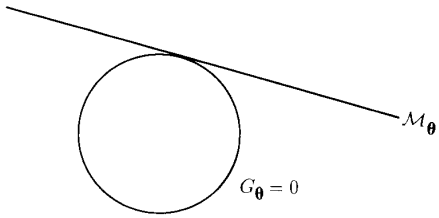


Fig. 1. Equality constraint $G_{\boldsymbol{\theta}} \stackrel{\text{def}}{=} (\theta_1 - \theta_1^o)^2 + (\theta_2 - \theta_2^o)^2 - a^2 = 0$. Here $\boldsymbol{\theta}$ can only vary along boundary of disk. Set of admissible directions, $\{\boldsymbol{\nu}_i\}$, in which parameter can move must lie on tangent hyperplane $\mathscr{M}_{\boldsymbol{\theta}}$. Since $\Theta_C$ has no interior points, there are no regular points of constraint set.

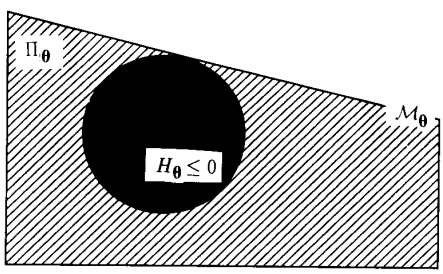


Fig. 2. Inequality-constraint $H_{\boldsymbol{\theta}} \le 0$, where $H_{\boldsymbol{\theta}} \stackrel{\text{def}}{=} (\theta_1 - \theta_1^o)^2 + (\theta_2 - \theta_2^o)^2 - a^2$. Here $\boldsymbol{\theta}$ can move into interior of disk. Set of admissible directions is contained in half-space $\Pi_{\boldsymbol{\theta}}$ that is supported by tangent hyperplane $\mathscr{M}_{\boldsymbol{\theta}}$. Since any point $\boldsymbol{\theta} \in \Theta_C$ can be represented as a limit of interior points, all points in $\Theta_C$ are regular points.

The following Lemma shows that if $\boldsymbol{\theta}$ is a regular point of $\Theta_C$ the constrained CR bound is identical to the unconstrained CR bound.

*Lemma 4:* Assume that the conditions (15)–(18) of Lemma 2 hold. Let the parameter space $\Theta_C$ be defined by the general inequality constraint $\mathscr{G}_{\boldsymbol{\theta}} \le 0$ where the vector function $\mathscr{G} = [\mathscr{G}^1, \cdots, \mathscr{G}^q]^T$ is differentiable. Let $\boldsymbol{\theta}$ be a regular point of $\Theta_C$. Then for any estimator $\hat{\boldsymbol{\theta}}$ having mean $\boldsymbol{m}_{\boldsymbol{\theta}}$, the estimator error covariance matrix $\Sigma_{\boldsymbol{\theta}}$ satisfies the classical unconstrained CR matrix inequality

$$\Sigma_{\boldsymbol{\theta}} \ge B_u, \tag{31}$$

where

$$B_u \stackrel{\text{def}}{=} [\nabla \boldsymbol{m}_{\boldsymbol{\theta}}] J_{\boldsymbol{\theta}}^{-1} [\nabla \boldsymbol{m}_{\boldsymbol{\theta}}]^T, \tag{32}$$

and $J_{\boldsymbol{\theta}}$ is the Fisher matrix (12). Equality holds in (31) if and only if there exists an $n \times n$ matrix $\Gamma$ such that

$$\hat{\boldsymbol{\theta}} - \boldsymbol{m}_{\boldsymbol{\theta}} = \Gamma [\Delta \ln f_{\boldsymbol{\theta}}]^T. \tag{33}$$

If such an estimator $\hat{\boldsymbol{\theta}}$ exists, it is called an efficient unconstrained estimator.

*Proof of Lemma 4:* Since $\boldsymbol{\theta}$ is a regular point, there exists a $\boldsymbol{\nu} \in \mathbb{R}^n$ such that for all $\Delta$, $0 < \Delta < 1$, we have: $(1-\Delta)\mathscr{G}_{\boldsymbol{\theta}} < 0$ and $\Delta[\mathscr{G}_{\boldsymbol{\theta}} + \nabla \mathscr{G}_{\boldsymbol{\theta}} \boldsymbol{\nu}] < 0$. Hence $(1-\Delta)\mathscr{G}_{\boldsymbol{\theta}} + \Delta[\mathscr{G}_{\boldsymbol{\theta}} + \nabla \mathscr{G}_{\boldsymbol{\theta}} \boldsymbol{\nu}] = \mathscr{G}_{\boldsymbol{\theta}} + \nabla \mathscr{G}_{\boldsymbol{\theta}} \boldsymbol{\nu} \Delta < 0$. Since for fixed $\boldsymbol{\nu}$

$$\|\mathscr{G}_{\boldsymbol{\theta} + \Delta \boldsymbol{\nu}} - \mathscr{G}_{\boldsymbol{\theta}} - \nabla \mathscr{G}_{\boldsymbol{\theta}} \boldsymbol{\nu} \Delta\| = o(\Delta),$$

it follows that for all sufficiently small $\Delta$, $\mathscr{G}_{\boldsymbol{\theta} + \Delta \boldsymbol{\nu}} < 0$. In a similar manner, it can be verified that there exists an $\epsilon > 0$ such that for all $\boldsymbol{\xi} \in \mathbb{R}^n$ with length $\|\boldsymbol{\xi}\| \le 1$

$$\mathscr{G}_{\boldsymbol{\theta} + \Delta(\boldsymbol{\nu} + \epsilon \boldsymbol{\xi})} < 0, \qquad \text{for all sufficiently small } \Delta > 0, \tag{34}$$

that is, $\boldsymbol{\theta} + \Delta \boldsymbol{\nu}$ is an interior point of $\Theta_C$. Choose $n$ linearly independent unit length vectors $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_n$, and define $\boldsymbol{\nu}_i = \boldsymbol{\nu} + \epsilon \boldsymbol{\xi}_i$, $i = 1, \cdots, n$. Then, using (34) it is seen that $\{\boldsymbol{\theta} + \Delta \boldsymbol{\nu}_i\}_{i=1}^n$ is a set of $n$ linearly independent vectors contained $\Theta_C$ for all sufficiently small $\Delta > 0$. Application of Lemma 2 thus gives the lower bound on the covariance matrix

$$B_c = [\nabla \boldsymbol{m}_{\boldsymbol{\theta}}] A [A^T J_{\boldsymbol{\theta}} A]^+ A^T [\nabla \boldsymbol{m}_{\boldsymbol{\theta}}]^T,$$

where $A$ is any $n \times n$ matrix with identical column space as $[\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_n]$. But the column space of this latter matrix is identical to $\mathbb{R}^n$, by linear independence of the $\boldsymbol{\nu}_i$'s, so taking $A = I$ in the previous equation for $B_c$ we obtain

$$B_c = B_u = [\nabla \boldsymbol{m}_{\boldsymbol{\theta}}] J_{\boldsymbol{\theta}}^{-1} [\nabla \boldsymbol{m}_{\boldsymbol{\theta}}]^T. \qquad \square$$

The bound (31) of Lemma 4 is identical to the classical multiparameter unconstrained CR bound [21], [27]. Since no equality constraints can be active at the regular points of $\Theta_C$, the Lemma establishes that pure inequality constraints on $\boldsymbol{\theta}$ do not affect the CR bound on the error covariance of estimators having a given mean gradient $\nabla \boldsymbol{m}_{\boldsymbol{\theta}}$. A number of parameter estimation problems have parameter constraint sets for which all of the points are regular. Examples include: orthant constraints, e.g., positivity of each of the elements $\theta_i$ in the parameter vector $\boldsymbol{\theta}$; range constraints, e.g., magnitude of $\theta_i$ less than 1; length constraints, e.g., $\sum_{i=1}^n \theta_i^2 \le 1$. For these types of constraints the classical unconstrained CR bound applies to all points in $\Theta_C$.

On the other hand, many estimation problems are formulated with parameter constraint sets for which some or all of the points are not regular. In particular, as previously mentioned, for the case of active equality constraints (29), if $\Theta_C$ is a $k$-dimensional surface, $k < n$, then $\Theta_C$ contains no regular points. Examples of these problems are provided in Section III of this paper. For this case, the classical CR bound is invalid and bound reduction occurs due to the constraints.

We now consider the construction of a CR bound under continuously differentiable equality constraints. As-

sume the equality constraint $\boldsymbol{G}_{\boldsymbol{\theta}} = 0$ (29) is active at $\boldsymbol{\theta}$. Define the $k \times n$ gradient matrix, $\nabla \boldsymbol{G}_{\boldsymbol{\theta}}$, of the function $\boldsymbol{G}$. Also define the hyperplane, $\mathscr{M}_{\boldsymbol{\theta}}$, tangent to the constraint set $\Theta_C$ at the point $\boldsymbol{\theta}$:

$$\mathscr{M}_{\boldsymbol{\theta}} = \{ y \in \mathbb{R}^n : \nabla \boldsymbol{G}_{\boldsymbol{\theta}} y = 0 \}. \tag{35}$$

If $\boldsymbol{G}$ is a linear function, e.g., $\boldsymbol{G}_{\boldsymbol{\theta}} = F\boldsymbol{\theta}$ for some $n \times k$ matrix $F$, $\mathscr{M}_{\boldsymbol{\theta}} \equiv \Theta_C$. Otherwise, when $\boldsymbol{G}$ is a continuously differentiable function, any set of points in $\Theta_C$ that are in the local $\Delta$-neighborhood of the point $\boldsymbol{\theta} \in \Theta_C$ are approximated to $o(\Delta)$ by a set of points in the tangent hyperplane $\mathscr{M}_{\boldsymbol{\theta}}$. Using Lemma 3 this implies that the constrained CR bound $B_c(\boldsymbol{\theta})$ depends on the equality-constraint function $\boldsymbol{G}$ only through its associated tangent hyperplane at the point $\boldsymbol{\theta}$.

The constrained CR bound for smooth inequality constraints is given in the following theorem.

*Theorem 1:* Let the regularity conditions (15)–(18) of Lemma 2 be satisfied. Let the parameter space $\Theta_C \subset \mathbb{R}^n$ be defined by the consistent set of equality and pure inequality constraints: $\boldsymbol{G}_{\boldsymbol{\theta}} = 0$, $\boldsymbol{H}_{\boldsymbol{\theta}} \leq 0$, where the vector functions $\boldsymbol{G} = [G^1, \cdots, G^k]^T$ and $\boldsymbol{H} = [H^1, \cdots, H^l]^T$ are continuously differentiable. Assume that the $k \times n$ gradient matrix $\nabla \boldsymbol{G}_{\boldsymbol{\theta}}$ has rank $p, p \leq k$. Then for any estimator $\hat{\boldsymbol{\theta}}$ having mean $\boldsymbol{m}_{\boldsymbol{\theta}}$, the estimator error covariance matrix $\Sigma_{\boldsymbol{\theta}}$ satisfies the matrix inequality

$$\Sigma_{\boldsymbol{\theta}} \geq B_c, \tag{36}$$

where

$$B_c = [\nabla \boldsymbol{m}_{\boldsymbol{\theta}}] Q_{\boldsymbol{\theta}} J_{\boldsymbol{\theta}}^{-1} [\nabla \boldsymbol{m}_{\boldsymbol{\theta}}]^T, \tag{37}$$

and $Q_{\boldsymbol{\theta}}$ is the $n \times n$, idempotent, rank $n - p$ matrix

$$Q_{\boldsymbol{\theta}} = \begin{cases} I, & \text{if } \boldsymbol{\theta} \text{ is a regular point of } \Theta_C \\ I - J_{\boldsymbol{\theta}}^{-1} [\nabla \boldsymbol{G}_{\boldsymbol{\theta}}]^T \{ [\nabla \boldsymbol{G}_{\boldsymbol{\theta}}] J_{\boldsymbol{\theta}}^{-1} [\nabla \boldsymbol{G}_{\boldsymbol{\theta}}]^T \}^+ [\nabla \boldsymbol{G}_{\boldsymbol{\theta}}], & \text{otherwise.} \end{cases} \tag{38}$$

Furthermore, equality holds in (36) if and only if there exists an $n \times n$ matrix $\Gamma$ such that

$$\hat{\boldsymbol{\theta}} - \boldsymbol{m}_{\boldsymbol{\theta}} = \Gamma Q_{\boldsymbol{\theta}}^T [\nabla \ln f_{\boldsymbol{\theta}}]^T \quad \text{(w.p.1)}. \tag{39}$$

If such an estimator $\hat{\boldsymbol{\theta}}$ exists, it is called an efficient constrained estimator.

*Proof of Theorem 1:* For the case that $\boldsymbol{\theta}$ is a regular point, in view of Lemma 4, there is nothing left to prove. Conversely, suppose that $\boldsymbol{\theta}$ is not a regular point. We will show that any sequence of test points in $\Theta_C$ that converges to $\boldsymbol{\theta}$ approximates an equivalent sequence in $\mathscr{M}_{\boldsymbol{\theta}}$. Then, for $0 < k \leq n - p$, we define $k$ sequences of test points in $\Theta_C$ whose associated approximating sequences in $\mathscr{M}_{\boldsymbol{\theta}}$ converge to $\boldsymbol{\theta}$ along linearly independent line paths $\boldsymbol{\theta} + \Delta_1 \boldsymbol{\nu}_1, \cdots, \boldsymbol{\theta} + \Delta_k \boldsymbol{\nu}_k, \Delta_1, \cdots, \Delta_k \to 0$, $\nu_i \in \mathscr{M}_{\boldsymbol{\theta}}$. Finally, with $B_b$ the Barankin bound (7), we show that $\limsup B_b$ is equal to the expression (37) for $B_c$, where the "lim sup" is taken over all such sequences of test points.

Let $\boldsymbol{\xi} = \boldsymbol{\xi}(\Delta)$ be a vector such that $\|\boldsymbol{\xi}(\Delta)\| \leq \Delta \to 0$ and assume that $\boldsymbol{\theta} + \boldsymbol{\xi}$ is a vector in $\Theta_C$ that converges to $\boldsymbol{\theta} \in \Theta_C$. By the assumed continuous differentiability of $\boldsymbol{G}_{\boldsymbol{\theta}}, \nabla \boldsymbol{G}_{\boldsymbol{\theta}} \boldsymbol{\xi} = \boldsymbol{o}(\Delta)$:

$$\begin{aligned} 0 &= \boldsymbol{G}_{\boldsymbol{\theta}+\boldsymbol{\xi}} - \boldsymbol{G}_{\boldsymbol{\theta}} \\ &= \nabla \boldsymbol{G}_{\boldsymbol{\theta}} \boldsymbol{\xi} + \boldsymbol{o}(\|\boldsymbol{\xi}\|) \\ &= \nabla \boldsymbol{G}_{\boldsymbol{\theta}} \boldsymbol{\xi} + \boldsymbol{o}(\Delta), \end{aligned} \tag{40}$$

where $\boldsymbol{o}(\Delta)$ is a vector of length $o(\Delta)$. Now define $P_{\mathscr{M}_{\boldsymbol{\theta}}} = I - \nabla \boldsymbol{G}_{\boldsymbol{\theta}}^T [\nabla \boldsymbol{G}_{\boldsymbol{\theta}} \nabla \boldsymbol{G}_{\boldsymbol{\theta}}^T]^+ \nabla \boldsymbol{G}_{\boldsymbol{\theta}}$. $P_{\mathscr{M}_{\boldsymbol{\theta}}}$ is an orthogonal-projection operator onto the null space of $\nabla \boldsymbol{G}_{\boldsymbol{\theta}}$, i.e., onto $\mathscr{M}_{\boldsymbol{\theta}}$ [21, Section 1c.4]. This induces an orthogonal decomposition of $\boldsymbol{\xi} = \boldsymbol{\xi}(\Delta)$ relative to $\mathscr{M}_{\boldsymbol{\theta}}$: $\boldsymbol{\xi} = P_{\mathscr{M}_{\boldsymbol{\theta}}} \boldsymbol{\xi} + [I - P_{\mathscr{M}_{\boldsymbol{\theta}}}] \boldsymbol{\xi}$. From (40), $[I - P_{\mathscr{M}_{\boldsymbol{\theta}}}] \boldsymbol{\xi} = \nabla \boldsymbol{G}_{\boldsymbol{\theta}}^T [\nabla \boldsymbol{G}_{\boldsymbol{\theta}} \nabla \boldsymbol{G}_{\boldsymbol{\theta}}^T]^+ \nabla \boldsymbol{G}_{\boldsymbol{\theta}} \boldsymbol{\xi} = \boldsymbol{o}(\Delta)$ so that

$$\boldsymbol{\xi} = P_{\mathscr{M}_{\boldsymbol{\theta}}} \boldsymbol{\xi} + \boldsymbol{o}(\Delta). \tag{41}$$

Hence to order $\Delta, \boldsymbol{\xi}$ is equal to the vector $P_{\mathscr{M}_{\boldsymbol{\theta}}} \boldsymbol{\xi}$ that is contained in $\mathscr{M}_{\boldsymbol{\theta}}$.

Now let $\{\boldsymbol{\theta} + \boldsymbol{\xi}_i(\Delta_i)\}_{i=1}^k$ be $k$ sequences in $\Theta_C$ indexed by $\Delta_1, \cdots, \Delta_k$ such that $P_{\mathscr{M}_{\boldsymbol{\theta}}} \boldsymbol{\xi}_i(\Delta_i) = \Delta_i \boldsymbol{\nu}_i$, $i = 1, \cdots, k$, where $\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_k$ are fixed linearly independent vectors and $0 < k \leq n - p$. Since $\boldsymbol{G}_{\boldsymbol{\theta}}$ is continuously differentiable and $\mathscr{M}_{\boldsymbol{\theta}}$ has dimension $n - p$, such sequences exist [8, Prop. 26.1]. Hence, in view of (41), for fixed $\Delta_1, \cdots, \Delta_k$ the $k$ test points $\boldsymbol{\theta} + \boldsymbol{\xi}_1(\Delta_1), \cdots, \boldsymbol{\theta} + \boldsymbol{\xi}_k(\Delta_k)$ are equal to $\boldsymbol{\theta} + \Delta_1 \boldsymbol{\nu}_1 + \boldsymbol{o}(\Delta_1), \cdots, \boldsymbol{\theta} + \Delta_k \boldsymbol{\nu}_k + \boldsymbol{o}(\Delta_k)$. Define $B_b(\boldsymbol{\theta} + \boldsymbol{\xi}_1(\Delta_1), \cdots, \boldsymbol{\theta} + \boldsymbol{\xi}_k(\Delta_k))$ the Barankin bound of Proposition 1 evaluated at these test points and define $B_c(\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_k)$ the CR bound of Lemma 2 evaluated with the direction vectors $\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_k$. Lemma 3 implies

$$\begin{aligned} &B_b(\boldsymbol{\theta} + \boldsymbol{\xi}_1(\Delta_1), \cdots, \boldsymbol{\theta} + \boldsymbol{\xi}_k(\Delta_k)) \\ &\quad = B_c(\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_k) + \underline{\underline{o}}(1) \\ &\quad = [\nabla \boldsymbol{m}_{\boldsymbol{\theta}}] A [A^T J_{\boldsymbol{\theta}} A]^+ A^T [\nabla \boldsymbol{m}_{\boldsymbol{\theta}}]^T + \underline{\underline{o}}(1), \end{aligned} \tag{42}$$

where $\underline{\underline{o}}(1)$ is a matrix of $o(1)$ elements that go to zero as the $\Delta_i$'s go to zero, and $A$ is an $n \times n$ matrix with column space equal to the span of $\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_k$.

Next we show that if $\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_k$ and $\boldsymbol{\nu}_1', \cdots, \boldsymbol{\nu}_k'$ are sets of vectors in $\mathscr{M}_{\boldsymbol{\theta}}$ such that $\operatorname{span}\{\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_k\} \supset \operatorname{span}\{\boldsymbol{\nu}_1', \cdots, \boldsymbol{\nu}_k'\}$ then $A[A^T J_{\boldsymbol{\theta}} A]^+ A^T \geq B[B^T J_{\boldsymbol{\theta}} B]^+ B^T$, where $A$ and $B$ are $n \times n$ matrices which have identical column spaces as $\operatorname{span}\{\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_k\}$ and $\operatorname{span}\{\boldsymbol{\nu}_1', \cdots, \boldsymbol{\nu}_k'\}$, respectively. Since by definition $\boldsymbol{\nu}_i \in \mathscr{M}_{\boldsymbol{\theta}}$, $i = 1, \cdots, k$, this will establish that the matrix $[\nabla \boldsymbol{m}_{\boldsymbol{\theta}}] A [A^T J_{\boldsymbol{\theta}} A]^+ A^T [\nabla \boldsymbol{m}_{\boldsymbol{\theta}}]^T$ on the right of (42) is maximized when the column space of $A$ is equal to $\mathscr{M}_{\boldsymbol{\theta}}$. With $J_{\boldsymbol{\theta}}^{-1/2}$ the positive square root matrix corresponding to $J_{\boldsymbol{\theta}}^{-1}$, the previous relation between the two spans holds if and only if $\operatorname{span}\{J_{\boldsymbol{\theta}}^{-1/2} \boldsymbol{\nu}_1, \cdots, J_{\boldsymbol{\theta}}^{-1/2} \boldsymbol{\nu}_k\} \supset \operatorname{span}\{J_{\boldsymbol{\theta}}^{-1/2} \boldsymbol{\nu}_1', \cdots, J_{\boldsymbol{\theta}}^{-1/2} \boldsymbol{\nu}_k'\}$. Hence it is sufficient to show that $A[A^T A]^+ A^T \geq B[B^T B]^+ B^T$ when the column space of $A$ contains the column space of $B$. Now $A[A^T A]^+ A^T$ and $I - B[B^T B]^+ B^T$ are idempotent,

symmetric, orthogonal-projection matrices onto the column space of $A$ and the null space of $B$ [21, Section 1c.4], respectively. Therefore, since the column space of $A$ contains the column space of $B$: $A[A^TA]^+A^TB = B$ and $B^TA[A^TA]^+A^T = B^T$. Since idempotent matrices are nonnegative definite, it follows that $A[A^TA]^+A^T - B[B^TB]^+B^T = A[A^TA]^+A^T[I - B[B^TB]^+B^T] = A[A^TA]^+A^T[I - B[B^TB]^+B^T]A[A^TA]^+A^T$, which is nonnegative-definite. Therefore we have from (42)

$$\limsup B_b = [\nabla \boldsymbol{m}_\theta]A\left[A^TJ_\theta A\right]^+ A^T[\nabla \boldsymbol{m}_\theta]^T, \quad (43)$$

where $A$ is a matrix whose column span equals $\mathscr{M}_\theta$.

Finally we show that the column span of $Q_\theta$ (38) is equal to $\mathscr{M}_\theta$ and that, setting $A = Q_\theta$ in (43), we obtain (37). Since $\nabla \boldsymbol{G}_\theta$ has rank $p$, there exists a row-echelon representation

$$\nabla \boldsymbol{G}_\theta = T\begin{bmatrix} B \\ O_1 \end{bmatrix}$$

where $T$ is a nonsingular $k \times k$ matrix, $B$ is a $p \times n$ full-row-rank matrix, and $O_1$ is a $(k-p)\times n$ matrix of zeros. Let $O_2$, $O_3$ and $O_4$ denote matrices of zeros having dimensions $(k-p)\times(k-p)$, $(k-p)\times p$ and $k \times n$, respectively. Use of (38) and (65) of Lemma 5 in the Appendix results in

$$\begin{aligned}
\nabla \boldsymbol{G}_\theta Q_\theta &= T\begin{bmatrix} B \\ O_1 \end{bmatrix}\Bigg[I - J_\theta^{-1}\begin{bmatrix} B^T & O_1^T \end{bmatrix} \\
&\quad \cdot T^T\left\{T\begin{bmatrix} B \\ O_1 \end{bmatrix}J_\theta^{-1}\begin{bmatrix} B^T & O_1^T \end{bmatrix}T^T\right\}^+ T\begin{bmatrix} B \\ O_1 \end{bmatrix}\Bigg] \\
&= T\Bigg[\begin{bmatrix} B \\ O_1 \end{bmatrix} - \begin{bmatrix} B \\ O_1 \end{bmatrix}J_\theta^{-1}\begin{bmatrix} B^T & O_1^T \end{bmatrix} \\
&\quad \cdot\left\{\begin{bmatrix} B \\ O_1 \end{bmatrix}J_\theta^{-1}\begin{bmatrix} B^T & O_1^T \end{bmatrix}\right\}^+\begin{bmatrix} B \\ O_1 \end{bmatrix}\Bigg] \\
&= T\Bigg[\begin{bmatrix} B \\ O_1 \end{bmatrix} - \begin{bmatrix} BJ_\theta^{-1}B^T & O_3^T \\ O_3 & O_2 \end{bmatrix} \\
&\quad \cdot\begin{bmatrix} BJ_\theta^{-1}B^T & O_3^T \\ O_3 & O_2 \end{bmatrix}^+\begin{bmatrix} B \\ O_1 \end{bmatrix}\Bigg] \\
&= T\left[\begin{bmatrix} B \\ O_1 \end{bmatrix} - \begin{bmatrix} I & O_3^T \\ O_3 & O_2 \end{bmatrix}\begin{bmatrix} B \\ O_1 \end{bmatrix}\right] \\
&= O_4,
\end{aligned}$$

where the invertibility of the full rank $p \times p$ matrix $BJ_\theta^{-1}B^T$ has been used on the third line of this equation. This establishes that the columns of $Q_\theta$ are contained in the hyperplane $\mathscr{M}_\theta$. A straightforward calculation shows that $Q_\theta Q_\theta = Q_\theta$ and $Q_\theta^T Q_\theta^T = Q_\theta^T$, i.e., both $Q_\theta$ and $Q_\theta^T$ are idempotent. Hence the rank of $Q_\theta$ is equal to its trace

$$\begin{aligned}
&\operatorname{rank}\{Q_\theta\} \\
&\quad = \operatorname{tr}\{Q_\theta\} \\
&\quad = \operatorname{tr}\left\{I - J_\theta^{-1}[\nabla \boldsymbol{G}_\theta]^T\left\{[\nabla \boldsymbol{G}_\theta]J_\theta^{-1}[\nabla \boldsymbol{G}_\theta]^T\right\}^+[\nabla \boldsymbol{G}_\theta]\right\} \\
&\quad = n - \operatorname{tr}\left\{[\nabla \boldsymbol{G}_\theta]J_\theta^{-1}[\nabla \boldsymbol{G}_\theta]^T\left\{[\nabla \boldsymbol{G}_\theta]J_\theta^{-1}[\nabla \boldsymbol{G}_\theta]^T\right\}^+\right\} \\
&\quad = n - p,
\end{aligned}$$

and $Q_\theta$ has $n-p$ linearly independent columns. Since these columns are contained in $\mathscr{M}_\theta$ and since $n - \operatorname{rank}\{\nabla \boldsymbol{G}_\theta\} = n - p$ is the dimension of $\mathscr{M}_\theta$, this establishes that the column space of $Q_\theta$ is identical to $\mathscr{M}_\theta$. Hence, using $A = Q_\theta$ in Lemma 2, we obtain the bound

$$B_c = [\nabla \boldsymbol{m}_\theta]Q_\theta\left[Q_\theta^TJ_\theta Q_\theta\right]^+ Q_\theta^T[\nabla \boldsymbol{m}_\theta]^T.$$

Now it is evident from symmetry that $Q_\theta J_\theta^{-1} = J_\theta^{-1}Q_\theta^T$. Define $J_G \overset{\text{def}}{=} Q_\theta^TJ_\theta Q_\theta$. One can verify that the matrix $Q_\theta J_\theta^{-1} = J_\theta^{-1}Q_\theta^T$ satisfies the Penrose conditions (9) for the pseudo-inverse, $J_G^+$, of $J_G$. Using these results and the fact that $Q_\theta$ and $Q_\theta^T$ are idempotent results in

$$\begin{aligned}
Q_\theta\left[Q_\theta^TJ_\theta Q_\theta\right]^+ Q_\theta^T &= Q_\theta J_G^+ Q_\theta^T \\
&= Q_\theta\left[Q_\theta J_\theta^{-1}\right]Q_\theta^T \\
&= Q_\theta J_\theta^{-1}Q_\theta^T \\
&= Q_\theta Q_\theta J_\theta^{-1} \\
&= Q_\theta J_\theta^{-1}.
\end{aligned}$$

Hence (37) is established. □

In reference to Theorem 1 we make the following remarks.

*Remark 1:* If the set of constraints $\boldsymbol{G}_\theta = 0$ is defined so that the rows of $\nabla \boldsymbol{G}_\theta$ are linearly independent, the $k \times k$ matrix $[\nabla \boldsymbol{G}_\theta]J_\theta^{-1}[\nabla \boldsymbol{G}_\theta]^T$ will be of full rank and $Q_\theta$ (38) will only involve the more familiar inverse matrix $\{[\nabla \boldsymbol{G}_\theta]J_\theta^{-1}[\nabla \boldsymbol{G}_\theta]^T\}^{-1}$. Although a reformulation eliminating redundant constraints can always be accomplished, frequently the most natural description of a constraint involves a rank-deficient $\nabla \boldsymbol{G}_\theta$, e.g., see Example 4 of Section III. In this case the general result of Theorem 1 is applicable.

*Remark 2:* Comparison between the bound of Lemma 4 and the bound of Theorem 1 indicates that the presence of constraints on the parameter space has the effect of reducing the rank of the Fisher information matrix. In particular if the $k$ equality constraints $\boldsymbol{G}_\theta = 0$ reduce the dimension of the parameter space from $n$ to $n-p$ then the rank $n$ inverse Fisher information $J_\theta^{-1}$ becomes the rank $n-p$ inverse *constrained Fisher information* $Q_\theta J_\theta^{-1}$. Hence active equality constraints have the effect of reducing the rank of the Fisher information matrix. In the proof of Theorem 1 it was shown that the column span of $Q_\theta$ is the tangent hyperplane $\mathscr{M}_\theta$, and that $Q_\theta J_\theta^{-1} = Q_\theta[Q_\theta^TJ_\theta Q_\theta]^+Q_\theta^T$. Furthermore, by Lemma 2,

$$Q_\theta\left[Q_\theta^TJ_\theta Q_\theta\right]^+ Q_\theta^T = A\left[A^TJ_\theta A\right]^+ A^T$$

if $A$ has the same column span as $Q_\theta$. Using these facts we have

$$Q_\theta J_\theta^{-1} = P_{\mathscr{M}_\theta}\left[P_{\mathscr{M}_\theta}^TJ_\theta P_{\mathscr{M}_\theta}\right]^+ P_{\mathscr{M}_\theta}^T,$$

where $P_{\mathscr{M}_\theta} = I - [\nabla \boldsymbol{G}_\theta]^T\{[\nabla \boldsymbol{G}_\theta][\nabla \boldsymbol{G}_\theta]^T\}^+[\nabla \boldsymbol{G}_\theta]$ is the $n \times n$ orthogonal-projection matrix that projects vectors in $\mathbb{R}^n$ onto $\mathscr{M}_\theta$. Hence the inverse constrained Fisher matrix $Q_\theta J_\theta^{-1}$ is obtained from a projection of the rows and columns of the unconstrained Fisher matrix $J_\theta$ onto the tangent hyperplanes of the constraint set.

*Remark 3:* The matrix $B_c$ (37) in Theorem 1 can be represented as the quantity

$$B_c = E\{[\nabla \boldsymbol{m}_\theta P_{\mathscr{M}_\theta}]\}\left(E\{[\nabla \ln f_\theta P_{\mathscr{M}_\theta}]^T[\nabla \ln f_\theta P_{\mathscr{M}_\theta}]\}\right)^+ \cdot E\{[\nabla \boldsymbol{m}_\theta P_{\mathscr{M}_\theta}]^T\},$$

where $P_{\mathscr{M}_\theta}$ is the projection operator defined in Remark 2. The vectors $\nabla \boldsymbol{m}_\theta P_{\mathscr{M}_\theta}$ and $\nabla \ln f_\theta P_{\mathscr{M}_\theta}$ are the projections of the unconstrained gradients of the mean and log-likelihood (score) functions onto the constraint tangent hyperplane $\mathscr{M}_\theta$, that is, these vectors correspond to constrained gradient vectors. In [10] these constrained gradient vectors were used along with Lemma 1 to give an alternative derivation of the inequality $\Sigma_\theta \geq B_c$.

*Remark 4:* Theorem 1 indicates that a certain bound reduction is induced by adding constraints on $\theta$. In particular, it is easy to show that the constrained CR bound $B_c$ of Theorem 1 is always less than the unconstrained CR bound $B_u$ in the sense that $B_u - B_c$ is nonnegative definite. This follows from: 1) the idempotence of $I - Q_\theta$; 2) the symmetry of $J_\theta^{-1}$ and $Q_\theta J_\theta^{-1}$, which imply that $(I - Q_\theta)J_\theta^{-1} = J_\theta^{-1}(I - Q_\theta)^T$; and 3) the nonnegative definiteness of $J_\theta^{-1}$. In particular, for unbiased estimators $\nabla \boldsymbol{m}_\theta = I$ and

$$\begin{aligned} B_c &= Q_\theta J_\theta^{-1} \\ &= J_\theta^{-1} - (I - Q_\theta)J_\theta^{-1} \\ &= J_\theta^{-1} - (I - Q_\theta)(I - Q_\theta)J_\theta^{-1} \\ &= J_\theta^{-1} - (I - Q_\theta)J_\theta^{-1}(I - Q_\theta)^T \\ &\leq J_\theta^{-1} = B_u. \end{aligned} \tag{44}$$

An important implication of (44) is that the incorporation of constraints can only reduce the CR bound on the component error variances.

*Remark 5:* In many examples of interest $Q_\theta$ is nondiagonal, accounting for the functional relationships between individual components of $\theta$ introduced by the constraint. Thus even if $J_\theta$ is diagonal, suggesting uncorrelated unconstrained estimator errors, the rank-reduced inverse Fisher information $Q_\theta J_\theta^{-1}$ in Theorem 1 can have off-diagonal terms, suggesting correlated constrained estimator errors.

*Remark 6:* A result of Lemma 4 and Theorem 1 is that pure inequality constaints $\boldsymbol{H}_\theta \leq 0$ do not affect the CR bound on error covariance of estimators with a given mean gradient $\nabla \boldsymbol{m}_\theta$. This is true even when $\theta$ is on the boundary of this set. An interpretation of this fact is obtained by recalling that the Fisher information matrix $J_\theta$ (12) is a function of the gradient of the likelihood surface at $\theta$. For a smooth surface, the gradient of the surface at $\theta$ is completely determined by the set of directional derivatives along directions contained in a convex cone with vertex at $\theta$, e.g., the half-space indicated in Fig. 2. In the case of one-dimensional differentiable functions, this simply reflects the equivalence of right and left derivatives. Therefore, the restriction of allowable local variations of a parameter at the boundary of $\boldsymbol{H}_\theta \leq 0$ does not affect the CR bound.

*Remark 7:* While Theorem 1 is stated as a lower bound on the estimator error covariance matrix, it can be used to specify a bound on the mean-square error (mse) matrix, $E_\theta\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T\}$. Specifically, since the mse matrix is equal to $\Sigma_\theta + (\boldsymbol{m}_\theta - \theta)(\boldsymbol{m}_\theta - \theta)^T$, application of the theorem gives a constrained CR bound on mse:

$$E_\theta\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T\} \geq B_c + (\boldsymbol{m}_\theta - \theta)(\boldsymbol{m}_\theta - \theta)^T,$$

where $B_c$ is given by (37).

*Remark 8:* Remarks 6 and 7 notwithstanding, when $\Theta_C$ corresponds to a pure inequality constraint Theorem 1 does not imply that improvement in mse is impossible. Indeed the minimum-distance projection of an unconstrained estimator $\hat{\theta}_u$ onto $\Theta_C$ can yield an estimator with lower mse than that of $\hat{\theta}_u$. Such an estimator arises in the example studied in [18]. However, if the estimators differ the projected estimator may have a different mean from that of $\hat{\theta}_u$ which generally is not differentiable, whereas Theorem 1 applies to classes of estimators with identical differentiable means $\boldsymbol{m}_\theta$.

*Remark 9:* In the course of proof of Theorem 1 it was established that the lower bound $B_c$ (36) is the tightest bound of the form (14) in the sense that $B_c = \limsup_{\Delta_1,\cdots,\Delta_k \to 0} B_b(\theta + \xi_1(\Delta_1),\cdots,\theta + \xi_k(\Delta_k))$ where $\{\theta + \xi_i(\Delta_i)\}_{i=1}^k$ are $k$ arbitrary sequences converging to $\theta$ along paths whose projections onto the tangent plane $\mathscr{M}_\theta$ are $k$ linearly independent line segments, $0 < k \leq n - p$. For linear constraints and exponential families of $f_\theta$ more can be proven: $B_c$ is the "lim sup" of the Barankin bound $B_b$ (7) with respect to arbitrary sequences of test points converging to $\theta$, i.e., $B_c$ is the tightest local Barankin bound.

## III. Applications

In this section we illustrate the application of the constrained CR bound (37) by specializing to the cases of linear and quadratic constraints.

*Example 1) Linearly Constrained Gauss-Markov Problem:* Let $F$ be an $m \times n$ matrix of rank $n$, $n \leq m$, and suppose that one observes the vector $\boldsymbol{X} \in \mathbb{R}^m$,

$$\boldsymbol{X} = F\theta + \eta,$$

where $\theta \in \mathbb{R}^n$, $\eta \in \mathbb{R}^m$ and $\eta$ is a zero-mean Gaussian vector with nonsingular $m \times m$ covariance matrix $K = E_\theta\{\eta\eta^T\}$. Since the model is linear and Gaussian, the Fisher information matrix is simply calculated as $J \stackrel{\text{def}}{=} J_\theta = F^T K^{-1} F$, which is independent of $\theta$. Furthermore, by the Gauss-Markov theorem [21, Ch. 4], the minimum variance unbiased (MVU) estimator $\hat{\theta}_u$ is a linear function of $\boldsymbol{X}$,

$$\hat{\theta}_u = J^{-1}F^T K^{-1}\boldsymbol{X}.$$

The error covariance of $\hat{\theta}_u$ is

$$\Sigma_\theta^u = J^{-1}.$$

Thus $\hat{\theta}_u$ achieves the unconstrained CR bound, (31) of

Lemma 2, for unbiased estimators. (Recall that for unbiased estimators, $\nabla \boldsymbol{m}_{\boldsymbol{\theta}} = I$.)

Consider, however, the problem of estimating $\boldsymbol{\theta}$ subject to the $k$ linear equality constraints $\boldsymbol{G}_{\boldsymbol{\theta}} = A\boldsymbol{\theta} = 0$, where $A$ is a $k \times n$ matrix, $k \leq n$. Using the fact that $\nabla \boldsymbol{G}_{\boldsymbol{\theta}} = A$, Theorem 1 gives the constrained CR bound: $B_c = [\nabla \boldsymbol{m}_{\boldsymbol{\theta}}] Q J^{-1} [\nabla \boldsymbol{m}_{\boldsymbol{\theta}}]^T$, where

$$Q \stackrel{\text{def}}{=} Q_{\boldsymbol{\theta}} = I - J^{-1}A^T[AJ^{-1}A^T]^+ A.$$

Since the matrix $Q$ is independent of $\boldsymbol{\theta}$, one can define the estimator

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= QJ^{-1}F^TK^{-1}X \\ &= Q\hat{\boldsymbol{\theta}}_u.\end{aligned} \tag{45}$$

Due to the constraint $A\boldsymbol{\theta} = 0$, $\hat{\boldsymbol{\theta}}$ is unbiased

$$E_{\boldsymbol{\theta}}\{\hat{\boldsymbol{\theta}}\} = \{I - J^{-1}A^T[AJ^{-1}A^T]^+ A\}\boldsymbol{\theta} = \boldsymbol{\theta}.$$

The error covariance of $\hat{\boldsymbol{\theta}}$ can be calculated directly from (45) using the idempotence of $Q$:

$$\begin{aligned}\Sigma_{\boldsymbol{\theta}} &= QJ^{-1}Q^T \\ &= QQJ^{-1} \\ &= QJ^{-1} \\ &= B_c,\end{aligned}$$

where $B_c$ is the constrained CR bound, (37) of Theorem 1, for unbiased estimation. This establishes that: 1) the estimator $\hat{\boldsymbol{\theta}}$ of (45) is the MVU constrained estimator, and 2) the constrained CR bound of Theorem 1 is achievable for the Gaussian linear model with linear constraints.

*Example 2) Image Reconstruction with a Support Constraint:* Support constraints are frequently used in image reconstruction problems such as those arising in tomographic imaging [24], [29] and phase retrieval [5], [9]. Suppose that the parameter vector of interest consists of a sampled two-dimensional image that is represented by a complex-valued vector with elements $\theta_{(k_1,k_2)}$, $k_1, k_2 = 0, 1, \cdots, M-1$. We will represent the parameter vector $\boldsymbol{\theta}$ as the $\mathbb{R}^{2M \times M}$ vector

$$\boldsymbol{\theta} = \left[\theta^R_{(0,0)}, \theta^I_{(0,0)}, \theta^R_{(0,1)}, \theta^I_{(0,1)}, \cdots, \theta^R_{(M-1,M-1)}, \theta^I_{(M-1,M-1)}\right]^T,$$

where the superscripts $R$ and $I$ denote respectively the real and imaginary parts of $\theta_{(k_1,k_2)}$.

If the support of the object is known, it can be used as a constraint in the estimation of $\boldsymbol{\theta}$. Let $S$ be the support of $\boldsymbol{\theta}$,

$$S = \{(k_1, k_2): \theta_{(k_1,k_2)} \neq 0;\ k_1, k_2 = 0, 1, \cdots, M-1\}.$$

Let $\mathbf{1}_S$ denote the $2M^2 \times 2M^2$ diagonal matrix with $[\mathbf{1}_S]_{ii} = 1$ if the $i$th element of $\boldsymbol{\theta}$ lies inside the support set $S$ and $[\mathbf{1}_S]_{ii} = 0$ otherwise, i.e., $\mathbf{1}_S$ is the matrix indicator function of $S$. The support constraint then has the form $\boldsymbol{G}_{\boldsymbol{\theta}} = [I - \mathbf{1}_S]\boldsymbol{\theta} = 0$. From Theorem 1 we have the constrained CR bound $B_c = [\nabla \boldsymbol{m}_{\boldsymbol{\theta}}] Q_{\boldsymbol{\theta}} J_{\boldsymbol{\theta}}^{-1} [\nabla \boldsymbol{m}_{\boldsymbol{\theta}}]^T$. Using $\nabla \boldsymbol{G}_{\boldsymbol{\theta}} = [I - \mathbf{1}_S]$ it is easy to verify:

$$\begin{aligned}Q_{\boldsymbol{\theta}} J_{\boldsymbol{\theta}}^{-1} &= J_{\boldsymbol{\theta}}^{-1} - J_{\boldsymbol{\theta}}^{-1}[I - \mathbf{1}_S]^T \\ &\quad \cdot \{[I - \mathbf{1}_S] J_{\boldsymbol{\theta}}^{-1} [I - \mathbf{1}_S]^T\}^+ [I - \mathbf{1}_S] J_{\boldsymbol{\theta}}^{-1} \\ &= T\Big\{\mathscr{T}^{-1} - \mathscr{T}^{-1}[I - \mathbf{1}_S]^T \\ &\quad \cdot \{[I - \mathbf{1}_S]\mathscr{T}^{-1}[I - \mathbf{1}_S]^T\}^+ [I - \mathbf{1}_S]\mathscr{T}^{-1}\Big\} T^T,\end{aligned} \tag{46}$$

where $\mathscr{T} = T^T J_{\boldsymbol{\theta}} T$ and $T$ is an orthogonal matrix such that

$$\mathbf{1}_S = T \begin{bmatrix} I & O_1 \\ O_1^T & O_2 \end{bmatrix} T^T, \tag{47}$$

where $O_1$ and $O_2$ are zero matrices. In other words, $T$ is a transformation that rearranges the image pixels so that the support is in the upper left hand corner of the image. Now let $\mathscr{T}$ and $\mathscr{T}^{-1}$ have the partitions

$$\mathscr{T} = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \tag{48}$$

$$\mathscr{T}^{-1} = \begin{bmatrix} K & L \\ L^T & M \end{bmatrix}, \tag{49}$$

where $A$ and $K$ are matrices of the same dimension as the identity matrix $I$ on the right-hand side of (47). With this notation $[I - \mathbf{1}_S]\mathscr{T}^{-1}[I - \mathbf{1}_S]^T$ is the partitioned matrix $\begin{bmatrix} O_0 & O_1 \\ O_1^T & M \end{bmatrix}$, where $O_0$ is a zero matrix of the appropriate dimensions. Therefore the pseudo-inverse on the right-hand side of (46) is simply $\begin{bmatrix} O_0 & O_1 \\ O_1^T & M^{-1} \end{bmatrix}$. Performing the rest of the matrix algebra indicated on the right-hand side of (46) we obtain

$$Q_{\boldsymbol{\theta}} J_{\boldsymbol{\theta}}^{-1} = T \begin{bmatrix} K - LM^{-1}L^T & O_1 \\ O_1^T & O_2 \end{bmatrix} T^T.$$

Using identities for the inverse of a partitioned matrix [11, Theorem 8.2.1] and the definitions of $A, B, C$ and $K, L, M$, (48) and (49), the matrix $K - LM^{-1}L^T$ can be identified as the inverse of the block matrix $A$. Hence,

$$\begin{aligned}Q_{\boldsymbol{\theta}} J_{\boldsymbol{\theta}}^{-1} &= T \begin{bmatrix} A^{-1} & O_1 \\ O_1^T & O_2 \end{bmatrix} T^T \\ &= T \begin{bmatrix} A & O_1 \\ O_1^T & O_2 \end{bmatrix}^+ T^T \\ &= T \left\{ \begin{bmatrix} I & O_1 \\ O_1^T & O_2 \end{bmatrix} \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} I & O_1 \\ O_1^T & O_2 \end{bmatrix} \right\}^+ T^T \\ &= \{\mathbf{1}_S J_{\boldsymbol{\theta}} \mathbf{1}_S\}^+,\end{aligned} \tag{50}$$

where the last equality follows by the orthogonality of $T$, the application of (47), (48), and the identification $T\mathscr{T}T^T = J_{\boldsymbol{\theta}}$. For the case of unbiased estimation $\nabla \boldsymbol{m}_{\boldsymbol{\theta}} = I$ and (50) is the constrained CR bound. Comparing the constrained CR bound (50) to the unconstrained CR bound

$J_\theta^{-1}$ it is evident that the incorporation of support constraints has the effect of zeroing out those rows and columns of the Fisher information matrix corresponding to image pixels $\theta$ for which it is known *a priori* that the pixel values are zero.

It is useful to compare the covariance of the estimator errors within the support region for the unconstrained cases. Using the same transformation $T$ (47) as before, we can assume without loss of generality that the support is in the upper left corner of image, i.e., the support matrix indicator function is $\mathbf{1}_S = \begin{bmatrix} I & O_1 \\ O_1^T & O_2 \end{bmatrix}$. In this case the unconstrained bound within the support region is $(A - BC^{-1}B^T)^{-1}$, which is the upper left block element of the inverse matrix $J_\theta^{-1} = \mathscr{T}^{-1}$ (48), while the constrained CR bound for these pixels is $A^{-1}$. If the Fisher matrix is block diagonal then $B$ is a matrix of zeros in (48), indicating that the errors of an unbiased efficient estimator of pixels inside and outside of the support region are uncorrelated; in this case the constrained CR bound is identical to the unconstrained CR bound. If the Fisher matrix is not block diagonal, however, there may be substantial reduction in the constrained CR bound over the support region. It is also significant that, unless $J_\theta$ is block diagonal, setting the pixels of an efficient (CR bound achieving) unconstrained estimator to zero outside the image support region does not produce an estimator that achieves the constrained CR bound. This is in contrast to the results obtained in [5] for diagonal $J_\theta$.

*Example 3) Spectrum Estimation with Power Constraints:* When there is prior information on the power of a random process over some regions of frequency, it is reasonable to expect that the achieveable error covariance of spectral estimators will be affected. This example quantifies the effect of such prior information on the constrained CR bound.

Let $\{X_i\}_{i=1}^N$ be a segment of a real wide sense stationary random process with power spectral density (PSD) $\{\mathscr{P}(f)\}_{f\in[-1/2,1/2]}$. The objective is to estimate the PSD, $\theta_i \overset{\text{def}}{=} \mathscr{P}(f_i)$, at $n$ distinct frequencies $f_1,\cdots,f_n$. Let the average power of $\{X_i\}$ be known over $P$ nonoverlapping frequency bands

$$\sum_{S_p} \theta_k = E_p, \qquad p = 1,\cdots,P, \tag{51}$$

where $S_p$ is the index set of the $p$th frequency band, and $E_p$ is the known average power of $\{X_i\}$ over this frequency band. The equations (51) correspond to $P$ linear constraints on the unknown PSD, known as the *P-point constraint* in robust Wiener filtering theory [20]. The concatenation of the $P$ equalities (51) gives the $P$ equations

$$G_\theta = \begin{bmatrix} \chi_1^T \\ \vdots \\ \chi_P^T \end{bmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} = \begin{bmatrix} E_1 \\ \vdots \\ E_P \end{bmatrix} \tag{52}$$

where $\chi_p$ is an $n\times 1$ column vector with $i$th element equal to 1 if $i \in S_p$ and 0 otherwise, i.e., $\chi_p$ is the vector indicator function of $S_p$. The gradient matrix $\nabla G_\theta$ is given by $\nabla G_\theta = [\chi_1,\cdots,\chi_P]^T$, resulting in

$$Q_\theta J_\theta^{-1} = J_\theta^{-1} - J_\theta^{-1}[\chi_1 \cdots \chi_P] \cdot \begin{bmatrix} \chi_1^T J_\theta^{-1}\chi_1 & \cdots & \chi_1^T J_\theta^{-1}\chi_P \\ \vdots & \ddots & \vdots \\ \chi_P^T J_\theta^{-1}\chi_1 & \cdots & \chi_P^T J_\theta^{-1}\chi_P \end{bmatrix}^+ \begin{bmatrix} \chi_1^T \\ \vdots \\ \chi_P^T \end{bmatrix} J_\theta^{-1}. \tag{53}$$

The structure of $Q_\theta J_\theta^{-1}$ is considerably simplified when $J_\theta$ is the diagonal matrix:

$$J_\theta = \operatorname{diag}_i\{\theta_i^{-2}\},$$

which is appropriate for the case of Gaussian observations $\{X_i\}_{i=1}^N$ and large $N$. Since the frequency bands $\{S_i\}$ are nonoverlapping the pseudo-inverse on the right-hand side of (53) becomes the pseudo-inverse of a diagonal matrix and

$$Q_\theta J_\theta^{-1} = J_\theta^{-1} - \sum_{i=1}^{P} \frac{J_\theta^{-1}\chi_i\chi_i^T J_\theta^{-1}}{\chi_i^T J_\theta^{-1}\chi_i}. \tag{54}$$

Let $e_l = [0,\cdots,0,1,0,\cdots,0]^T$ denote the $l$th standard basis vector in $\mathbb{R}^n$. Let $l$ be an index in the constraint set $S_p$. Then for an unbiased estimator, $\hat{\theta}$, the constrained CR bound on the variance of the $l$th component, $\hat{\theta}_l$, is obtained from (54)

$$\begin{aligned} [B_c]_{ll} &= e_l^T B_c e_l \\ &= e_l^T \left\{ J_\theta^{-1} - \sum_{i=1}^{P} \frac{J_\theta^{-1}\chi_i\chi_i^T J_\theta^{-1}}{\chi_i^T J_\theta^{-1}\chi_i} \right\} e_l \\ &= \left[J_\theta^{-1}\right]_{ll} \left( 1 - \frac{[J_\theta]_{ll}^{-1}}{\sum_{\{i\in S_p\}} \left[J_\theta^{-1}\right]_{ii}} \right). \end{aligned} \tag{55}$$

Using the unconstrained CR bound $[B_u]_{ll} = \theta_l^2 = \mathscr{P}^2(f_l)$, we obtain the relative reduction in the CR bound due to the constraint

$$\frac{[B_c]_{ll}}{[B_u]_{ll}} = 1 - \frac{1}{1 + \sum_{\{i\in S_p, i\neq l\}} \frac{\mathscr{P}^2(f_i)}{\mathscr{P}^2(f_l)}}. \tag{56}$$

Since the term on the right hand side of (56) is between 0 and 1, the average power constraint induces a CR bound reduction on the component PSD estimation errors. The bound reduction factor (56) is independent of the other constraint sets $S_k$, $k = 1,\cdots,P$, $k \neq p$, and therefore average power constraints over $S_p$ do not affect PSD estimator errors at frequencies outside of $S_p$. The amount of bound reduction depends on two factors: 1) the relative magnitude of the spectral component of interest, $\mathscr{P}^2(f_l)$, compared to the magnitude of the other frequency components within the frequency band $S_p$; and 2) the length, $|S_p|$ = number of indices, of $S_p$. In particular, little or no reduction in the variance bound occurs for the case where $\mathscr{P}^2(f_l)/\mathscr{P}^2(f_i)$ is small for all $i \in S_p, i \neq l$. However, when $\mathscr{P}^2(f_l)$ is large compared to the other $\mathscr{P}^2(f_i)$,

$i \in S_p$, a substantial reduction in the bound occurs. This implies that the most bound reduction will be achieved over those constraint regions $S_i$ where the PSD has a high dynamic range, i.e., large peaks. The particular dynamic range required for a significant bound reduction is proportional to $|S_p|$. As a rule of thumb, for a reduction in the CR bound at frequency $f_l$ by a factor $\alpha$ or more, the ratio of $\mathscr{P}(f_l)$ to the root mean-squared value of the remaining spectral components in $S_p$,

$$\bar{\mathscr{P}} \stackrel{\text{def}}{=} \sqrt{\frac{1}{|S_p|-1} \sum_{i \in S_p, i \neq l} \mathscr{P}^2(f_i)},$$

must satisfy

$$\frac{\mathscr{P}(f_l)}{\bar{\mathscr{P}}} \geq \sqrt{\frac{1-\alpha}{\alpha}\left[|S_p|-1\right]}.$$

*Example 4) Signal Subspace Constraints:* Signal subspace constraints are used in sensor array processing estimation problems to take account of a particular structure of the array covariance matrix [14]. Specifically, assume that $p$ zero-mean Gaussian signals arrive at different angles of incidence on an $m$-sensor array having a zero-mean, spatially incoherent array noise of power $\sigma^2$. Further, assume that $p < m$. Then the covariance matrix of the set of sensor outputs has the singular value decomposition

$$R = \sum_{i=1}^{p} \lambda_i^s \boldsymbol{v}_i \boldsymbol{v}_i^H + \sigma^2 I = \sum_{i=1}^{m} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^H,$$

where $\{\boldsymbol{v}_i\}_{i=1}^m$ are the eigenvectors of $R$ and $\{\lambda_i\}_{i=1}^m$ are the eigenvalues:

$$\lambda_i = \begin{cases} \lambda_i^s + \sigma^2, & i = 1, \cdots, p \\ \sigma^2, & i = p+1, \cdots, m \end{cases}$$

and $\{\lambda_i^s\}_{i=1}^p$ denote the signal-dependent eigenvalues of $R$. The span of $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_p$ is called the *signal subspace*.

Consider the problem of estimating the eigenvalues of $R$ when $p$ is known but all of the other parameters are unknown. This partial knowledge induces the following constraints on the $\lambda_i$:

$$\begin{aligned} &\text{a)} \quad \lambda_j > 0, && j = 1, \cdots, m \\ &\text{b)} \quad \lambda_j \geq \frac{1}{m-p} \sum_{i=p+1}^{m} \lambda_i, && j = 1, \cdots, m \\ &\text{c)} \quad \lambda_j - \frac{1}{m-p} \sum_{i=p+1}^{m} \lambda_i = 0, && j = p+1, \cdots, m \end{aligned} \tag{57}$$

where constraint a) arises from the assumed positive-definiteness of $R$, constraint b) takes account of the positivity of the signal eigenvalues $\{\lambda_i^s\}_{i=1}^p$, and constraint c) reflects the equality of the $m-p$ noise eigenvalues.

Let each unknown eigenvector $\boldsymbol{v}_i \in \mathbb{R}^n$ be parameterized by its $m-1$ direction cosines, $\boldsymbol{\rho}_i = [\rho_{i,1}, \cdots, \rho_{i,m-1}]^T$, $i = 1, \cdots, m$. The combination of the $m$ unknown eigenvalues and the $m(m-1)$ unknown direction cosines yields the $n = m^2$ element parameter vector $\boldsymbol{\theta} = [\lambda_m, \cdots, \lambda_1, \boldsymbol{\rho}_m^T, \cdots, \boldsymbol{\rho}_1^T]^T$. The constraint c) can be then be expressed as the $(m-p) \times n$ matrix constraint

$$\boldsymbol{G}_{\boldsymbol{\theta}} = \left[ I_{m-p} - \frac{1}{m-p} \mathbf{1}\mathbf{1}^T O_1 \right] \boldsymbol{\theta},$$

where $I_k$ denotes a $k \times k$ identity matrix, $O_1$ is a $(m-p) \times (n-m+p)$ matrix of zero entries, and $\mathbf{1}$ is a $(m-p)$-vector of ones.

Observe that the rows of $\nabla \boldsymbol{G}_{\boldsymbol{\theta}}$ are not linearly independent due to the fact that there is one redundant constraint in c) of (57). Observe also that the equality constraint c) creates a dimension $n-m+p+1$ linear subspace in the unconstrained parameter space $\mathbb{R}^n$. Hence, despite the presence of inequality constaints a) and b), the constrained parameter space $\Theta_C$ contains no regular points, and, by Theorem 1, the constraints a), b) do not impact the form of the constrained CR bound.

As in Example 2, partition $J_{\boldsymbol{\theta}}$ according to

$$J_{\boldsymbol{\theta}} = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix},$$

where $A$ is $(m-p) \times (m-p)$, $B$ is $(m-p) \times (n-m+p)$, and $C$ is $(n-m+p) \times (n-m+p)$. Then the $n \times n$ inverse constrained Fisher matrix, $Q_{\boldsymbol{\theta}} J_{\boldsymbol{\theta}}^{-1}$, of Theorem 1 is given by

$$Q_{\boldsymbol{\theta}} J_{\boldsymbol{\theta}}^{-1} = J_{\boldsymbol{\theta}}^{-1} - J_{\boldsymbol{\theta}}^{-1} \begin{bmatrix} Z^+ & O_1 \\ O_1^T & O_2 \end{bmatrix} J_{\boldsymbol{\theta}}^{-1},$$

where $O_1$ and $O_2$ are zero matrices of dimensions $(m-p) \times (n-m+p)$ and $(n-m+p) \times (n-m+p)$, respectively, and $Z$ is the $(m-p) \times (m-p)$ matrix

$$\begin{aligned} Z &= \nabla \boldsymbol{G}_{\boldsymbol{\theta}} J_{\boldsymbol{\theta}}^{-1} [\nabla \boldsymbol{G}_{\boldsymbol{\theta}}]^T \\ &= \left[ I_{m-p} - \frac{1}{m-p} \mathbf{1}\mathbf{1}^T \right] [A - BC^{-1}B^T] \\ &\quad \cdot \left[ I_{m-p} - \frac{1}{m-p} \mathbf{1}\mathbf{1}^T \right]. \end{aligned} \tag{58}$$

As a simple example, consider the case where the Fisher information matrix is block diagonal with: $B = O_1$ and $A = \sigma_\lambda^{-2} I_{m-p}$. Then $Z = \sigma_\lambda^{-2}[I_{m-p} - \frac{1}{m-p}\mathbf{1}\mathbf{1}^T]$. Using condition 3) of (9) it is easy to show that $Z^+ = \sigma_\lambda^2[I_{m-p} - \frac{1}{m-p}\mathbf{1}\mathbf{1}^T]$. This results in

$$Q_{\boldsymbol{\theta}} J_{\boldsymbol{\theta}}^{-1} = \begin{bmatrix} \sigma_\lambda^2 \frac{1}{m-p} \mathbf{1}\mathbf{1}^T & O_1 \\ O_1^T & C^{-1} \end{bmatrix}. \tag{59}$$

Suppose there exists an efficient unbiased estimator $\hat{\boldsymbol{\theta}}_u$ for the eigenvalues and eigenvectors which satisfies constraints a) and b), and assume that the Fisher information is block diagonal as previously specified. The right-hand side of (59) is then the covariance matrix of the estimator obtained by replacing each of the $m-p$ noise eigenvalue estimates in $\hat{\boldsymbol{\theta}}_u$ by their average $\frac{1}{m-p}\sum_{i=p+1}^m \hat{\lambda}_i$. Hence, if an efficient unconstrained estimator of the eigenvalues

can be found that has positive elements, the estimator obtained by averaging over the $m-p$ smallest eigenvalues of the efficient estimator achieves the constrained CR bound.

*Example 5) Signal Estimation with Power Constraints:* Consider the problem of estimating the discrete-time signal waveform, $\theta_1, \cdots, \theta_n$, subject to constraints on the squared-modulus of the DFT of $\boldsymbol{\theta}$. Here, the sum of the squared moduli over each of $P$ nonoverlapping frequency intervals is constrained to be equal to known constants $E_p$, $p=1,\cdots,P$. While similar to the case studied in Example 3, this problem involves nonlinear quadratic constraints on the parameters, and time rather than frequency domain estimation is performed.

Let $W=[\boldsymbol{W}_1,\cdots,\boldsymbol{W}_n]$ denote the $n\times n$ unitary matrix of orthonormal discrete Fourier transform columns: $\boldsymbol{W}_l = 1/\sqrt{N}[1, e^{-j2\pi l/n},\cdots, e^{-j2\pi(n-1)l/n}]^T$, $l=1,\cdots,n$. Now suppose that for $P\le n$ the constraint takes the form

$$\sum_{i\in S_p} |[W\boldsymbol{\theta}]_i|^2 = E_p, \qquad p=1,2,\cdots,P. \tag{60}$$

Here, $S_p$ denotes the index set of the $p$th interval and $[W\boldsymbol{\theta}]_i$ is $i$th component of the $n$-point DFT of $\boldsymbol{\theta}$. When $P=n$, (60) specifies the modulus Fourier transform of $\boldsymbol{\theta}$. As in Example 2, we let $\mathbf{1}_p$ denote the $n\times n$ diagonal matrix with $[\mathbf{1}_p]_{ii}=1$ if $i\in S_p$ and $[\mathbf{1}_p]_{ii}=0$ otherwise. Then the constraint (60) can be written as the set of $P$ equations

$$\boldsymbol{G}_{\boldsymbol{\theta}} = \begin{bmatrix} \boldsymbol{\theta}^T W^H \mathbf{1}_1 W\boldsymbol{\theta} \\ \vdots \\ \boldsymbol{\theta}^T W^H \mathbf{1}_P W\boldsymbol{\theta} \end{bmatrix} - \begin{bmatrix} E_1 \\ \vdots \\ E_P \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix},$$

where the superscript $H$ denotes hermitian transpose. The gradient $\nabla \boldsymbol{G}_{\boldsymbol{\theta}}$ is the $P\times n$ matrix

$$\nabla \boldsymbol{G}_{\boldsymbol{\theta}} = \begin{bmatrix} 2\boldsymbol{\theta}^T W^H \mathbf{1}_1 W \\ \vdots \\ 2\boldsymbol{\theta}^T W^H \mathbf{1}_P W \end{bmatrix}. \tag{61}$$

We now specialize to the linear observation model:

$$X_i = \theta_i + \eta_i, \qquad i=1,\cdots,n$$

where $\eta_i$ is a zero-mean Gaussian white noise with variance $\sigma^2$. Recalling Example 1, $J_{\boldsymbol{\theta}}$ can be seen to be the scaled identity matrix $\sigma^{-2}I$. Let $O$ denote the $n\times n$ zero matrix. Using (61) and the fact that the intervals $S_p$ are nonoverlapping $\mathbf{1}_i\mathbf{1}_j = O$, $i\ne j$, the inverse constrained Fisher matrix of Theorem 1 is the $n\times n$ matrix

$$Q_{\boldsymbol{\theta}} J_{\boldsymbol{\theta}}^{-1} = \sigma^2 W^H \left[ I - \sum_{i=1}^{P} \frac{\mathbf{1}_i W\boldsymbol{\theta}\boldsymbol{\theta}^T W^H \mathbf{1}_i}{\boldsymbol{\theta}^T W^H \mathbf{1}_i W\boldsymbol{\theta}} \right] W. \tag{62}$$

Since $W$ is the (linear) DFT operator, the matrix $\sigma^2[I-\Sigma(\cdot)]$ on the right-hand side of (62) is the inverse constrained Fisher information matrix for estimation of the DFT $W\boldsymbol{\theta}$. As in Example 3, let the index $l$ be constrained in $S_p$. Then the ratio between the constrained and unconstrained CR bounds on the variance, $\mathrm{var}\{[W\hat{\boldsymbol{\theta}}]_l\} = E_{\boldsymbol{\theta}}\{|[W\hat{\boldsymbol{\theta}}]_l - E_{\boldsymbol{\theta}}[W\boldsymbol{\theta}]_l|^2\}$, is obtained by evaluating the quadratic forms $\mathbf{e}_l^T W^H Q_{\boldsymbol{\theta}} J_{\boldsymbol{\theta}}^{-1} W\mathbf{e}_l$ and $\mathbf{e}_l^T W^H J_{\boldsymbol{\theta}}^{-1} W\mathbf{e}_l$

$$\frac{[B_c]_{ll}}{[B_u]_{ll}} = 1 - \frac{1}{1+\sum_{i\in S_p, i\ne l} \frac{|[W\boldsymbol{\theta}]_i|^2}{|[W\boldsymbol{\theta}]_l|^2}}. \tag{63}$$

This is of identical form to the expression obtained for constrained PSD estimation, (55) of Example 3, when the power spectral density, $\mathscr{P}(f_i)$, is identified with the magnitude spectrum $|[W\boldsymbol{\theta}]_i|$, $i=1,\cdots,n$. For unbiased estimators, a bound on the total mean-squared error in estimating the time domain signal $\theta_i$'s can be determined from (63) by using the unitary property of the DFT matrix $W$ (Parseval's Theorem):

$$\begin{aligned} \sum_{i=1}^{n} \|\theta_i - \hat{\theta}_i\|^2 &= \mathrm{tr}\{\Sigma_{\boldsymbol{\theta}}\} \\ &\ge \mathrm{tr}\{Q_{\boldsymbol{\theta}} J_{\boldsymbol{\theta}}^{-1}\} \\ &= \sigma^2 \,\mathrm{tr}\left\{ W^H \left[ I - \sum_{i=1}^{P} \frac{\mathbf{1}_i W\boldsymbol{\theta}\boldsymbol{\theta}^T W^H \mathbf{1}_i}{\boldsymbol{\theta}^T W^H \mathbf{1}_i W\boldsymbol{\theta}} \right] W \right\} \\ &= \sigma^2 \,\mathrm{tr}\left\{ I - \sum_{i=1}^{P} \frac{\mathbf{1}_i W\boldsymbol{\theta}\boldsymbol{\theta}^T W^H \mathbf{1}_i}{\boldsymbol{\theta}^T W^H \mathbf{1}_i W\boldsymbol{\theta}} \right\} \\ &= \sigma^2[n-P]. \end{aligned}$$

Therefore, on the average, the constraints produce a factor of $1-P/n$ reduction in the CR bound on the variances of unbiased estimators of the $\theta_i$'s.

## IV. Conclusion

A constrained Cramér–Rao (CR) lower bound on the error covariance of estimators of multidimensional parameters has been obtained. The constrained CR bound was derived from a limiting form of a multiparameter Barankin-type bound. For constraint sets defined by a general smooth functional inequality constraint of the form $\mathscr{G}_{\boldsymbol{\theta}} \le 0$, the constrained CR bound is equivalent to the unconstrained CR bound evaluated with a "constrained" Fisher information matrix. This constrained Fisher matrix was shown to be identical to the classical unconstrained Fisher matrix at all regular points of the constraint set, e.g., at interior points. However at nonregular points, such as points governed by equality constraints, the constrained Fisher matrix is a rank-deficient matrix. This constrained Fisher matrix is equivalent to a matrix of orthogonal projections of the rows and columns of the unconstrained Fisher matrix onto the tangent hyperplanes of the constraint set. The simple form of the constrained CR bound allows the effect of particular equality and inequality constraints to be easily studied through comparisons between the constrained and unconstrained CR bounds. It was shown that the incorporation of functional constraints necessarily decreases the CR bound for unbiased estimators. Not surprisingly, the constrained bound was shown to be achievable for the lin-

early-constrained Gauss–Markov problem. To illustrate the application of the constrained CR bound, several applications in the area of signal processing were considered. These included support constraints in image reconstruction, signal subspace constraints in array processing, and average power constraints in spectral estimation and in signal estimation.

In their present form, the results obtained in this paper only directly apply to a finite dimensional parameter space and a non-stochastic constraint. A generalization of these results to infinite dimensional parameter spaces would be useful for the study of constraints in filtering, prediction, and smoothing problems. Theorem 1 could perhaps be applied to complete separable infinite-dimensional parameter spaces, e.g., a separable Hilbert space, by taking the formal limit of the elements of the matrix bound (37) as the dimension of the indicated matrices goes to infinity. Stochastic constraints are of interest when the constraint depends on the particular realization of the statistical experiment, and they provide a model for partially-known constraints. A main difficulty in obtaining a generalization of the constrained CR bound to differentiable stochastic constraints is that the column space of the constraint equality gradient matrix, $\nabla \mathscr{G}_\theta$, is in general a random set and therefore Lemma 2 cannot be applied. On the other hand, a tractible analysis may be possible for simple stochastic constraints such as constraints obtained from random perturbations of the constraint function $\mathscr{G}_\theta$.

## Acknowledgment

The authors would like to thank Dr. John Jayne of M.I.T. Lincoln Laboratory for pointing out an error in an early draft of this manuscript and for helpful discussions concerning this paper.

## Appendix

*Lemma 5:* Let $Q$ be an arbitrary $n \times m$ matrix and $T$ be any $m \times m$ invertible matrix. Then

$$QT[T^TQ^TQT]^+T^TQ^T = Q[Q^TQ]^+Q^T, \tag{64}$$

where the plus sign denotes (Moore–Penrose) pseudo-inverse. As a consequence, if $R$ is an arbitrary $m \times n$ matrix, $J$ is an $m \times m$ positive definite matrix, and $T$ is an invertible $n \times n$ matrix, then:

$$RT[T^TR^TJRT]^+T^TR^T = R[R^TJR]^+R^T. \tag{65}$$

*Proof of Lemma 5:* Let the left and right sides of the identity (64) be denoted as the $n \times n$ matrices $P_1$ and $P_2$, respectively. It is easily verified that $P_1$ and $P_2$ are symmetric and idempotent. Therefore $P_1$ and $P_2$ are orthogonal projections onto respective subsets, $\mathscr{M}_1$ and $\mathscr{M}_2$ say, of $\mathbb{R}^m$ [22, Section 105]. Furthermore, using properties 1)–3) of (9), it is easily verified that $P_2P_1P_2 = P_2$ and $P_1P_2P_1 = P_1$. Equivalently, since $P_1P_2$ and $P_2P_1$ are projections onto respective subsets of $\mathscr{M}_1 \cap \mathscr{M}_2$: $P_2P_1 = P_1$ and $P_1P_2 = P_2$. However, $P_iP_j = P_j$ implies $P_i \geq P_j$ [22, Prop. d of Section 104], and hence $P_1 = P_2$.

To show (65), first observe that, due to positive definiteness, there exists an invertible matrix $J^{1/2}$ such that $J^{T/2}J^{1/2}$. Define $Q = J^{1/2}R$. Then (65) reads

$$J^{-1/2}QT[T^TQ^TQT]^+T^TQ^TJ^{-T/2} = J^{-1/2}Q[Q^TQ]^+Q^TJ^{-T/2},$$

which follows directly from (64). This finishes the proof of Lemma 5. □

## References

[1] J. Aitchison and S. D. Silvey, "Maximum-likelihood estimation of parameters subject to restraints," *Ann. Math. Stat.*, vol. 29, pp. 813–828, 1958.
[2] A. Albert, *Regression and the Moore–Penrose Pseudo-Inverse*. New York: Academic Press, 1972.
[3] P. Billingsley, *Probability and Measure* 2nd ed. New York: Wiley, 1986.
[4] J. M. Begun, W. J. Hall, W.-M. Huang, and J. A. Wellner, "Information and asymptotic efficiency in parametric–nonparametric models," *Ann. Statist.*, vol. 11, no. 2, pp. 432–452, 1983.
[5] J. N. Cederquist and C. C. Wackerman, "Phase-retrieval error: A lower bound," *J. Opt. Soc. Amer. A*, vol. 4, pp. 1788–1792, Sept. 1987.
[6] D. G. Chapman and H. Robbins, "Minimum variance estimation without regularity assumptions," *Ann. Math. Statist.*, vol. 22, pp. 581–586, 1951.
[7] H. R. Chernoff, "On the distribution of the likelihood ratio," *Ann. Math. Statist.*, vol. 25, pp. 573–578, 1954.
[8] K. Deimling, *Nonlinear Functional Analysis*. New York: Springer-Verlag, 1985.
[9] J. R. Fienup, "Reconstruction and synthesis applications of an iterative algorithm," in *Transformations in Optical Signal Processing*, W. T. Rhodes, J. R. Fienup, and B. E. A. Saleh, Eds., *Proc. SPIE*, vol. 373, pp. 147–160, 1981.
[10] J. D. Gorman and A. O. Hero, "Lower bounds on parametric estimators with constraints," in *Proc. 4th ASSP Workshop on Spectrum Estimation and Modeling*, Minneapolis, MN, August 3–5, 1988, pp. 223–228.
[11] F. A. Graybill, *Matrices with Applications in Statistics*. Belmont, CA: Wadsworth, 1969.
[12] R. A. Horn and C. G. Johnson, *Matrix Analysis*. Cambridge, MA: Cambridge Univ. Press. 1985.
[13] I. A. Ibragimov and R. Z. Has'minskii, *Statistical Estimation—Asymptotic Theory*. New York: Springer-Verlag, 1981.
[14] R. Kumaresan and D. W. Tufts, "Estimating the angles of arrival of multiple plane waves," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-19, pp. 134–139, Jan. 1983.
[15] L. LeCam, *Asymptotic Methods in Statistical Decision Theory*. New York: Springer-Verlag, 1984.
[16] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.
[17] R. J. McAulay and E. M. Hofstetter, "Barankin bounds on parameter estimation," *IEEE Trans. Inform. Theory*, vol. IT-27, no. 6, pp. 669–675, Nov. 1971.
[18] P. A. P. Moran, "Maximum-likelihood estimation in non-standard conditions," *Proc. Camb. Phil. Soc.*, vol. 70, 1971, pp. 441–450.
[19] J. Pfanzagl and W. Wefelmeyer *Contributions to a General Asymptotic Statistical Theory*, *Lecture Notes in Statistics*, vol. 13. New York: Springer-Verlag, 1982.
[20] K. S. Vastola and V. Poor, "Robust Wiener–Kolmogorov theory," *IEEE Trans. Inform. Theory*, vol. IT-30, no. 2, pp. 316–327, Mar. 1984.
[21] C. R. Rao, *Linear Statistical Inference and its Applications*, 2nd ed. New York, Wiley, 1973.
[22] F. Riesz and B. Sz.-Nagy, *Functional Analysis*, 2nd ed. New York: Frederick Ungar Pub., New York, 1978.
[23] S. E. Roucos and D. G. Childers, "A two-dimensional maximum entropy spectral estimator," *IEEE Trans. on Inform. Theory*, vol. IT-26, no. 5, pp. 554–560, Sept. 1980.

[24] M. I. Sezan and H. Stark, "Applications of convex projection theory to image recovery in tomography and related areas," in *Image Recovery: Theory and Application*, H. Stark Ed. Orlando, FL: Academic Press, 1987.

[25] D. L. Snyder, M. I. Miller, T. J. Schulz, and J. D. O'Sullivan, "Constrained probability-density estimation from noisy data," in *Proc. 22nd Ann. Conf. Inform. Sci. Syst.*, Princeton Univ., Mar. 1988.

[26] G. W. Stewart, "On the continuity of the generalized inverse," *SIAM J. Appl. Math.*, vol. 17, no. 1, pp. 33–45, Jan. 1969.

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

[28] P. A. Yansouni and R. J. Inkol, "The use of linear constraints to reduce the variance of time of arrival difference estimates for source location," *IEEE Trans. Acoust. Speech and Signal Processing*, vol. ASSP-32, pp. 907–912, Aug. 1984.

[29] D. C. Youla, "Mathematical theory of image restoration by the method of convex projections," in *Image Recovery: Theory and Application*, H. Stark Ed. Orlando, FL: Academic Press, 1987.