On the Cramér–Rao Bound Under Parametric Constraints

Petre Stoica, *Fellow, IEEE,* and Boon Chong Ng, *Student Member, IEEE*

***Abstract*— This letter presents a simple expression for the Cramér–Rao bound (CRB) for parametric estimation under differentiable, deterministic constraints on the parameters. In contrast to previous works, the constrained CRB presented here does not require that the Fisher information matrix (FIM) for the unconstrained problem be of full rank. This is a useful extension because, for several signal processing problems (such as blind channel identification), the unconstrained problem is unidentifiable. Our expression for the constrained CRB depends only on the unconstrained FIM and a basis of the nullspace of the constraint's gradient matrix. We show that our constrained CRB formula reduces to the known expression when the FIM for the unconstrained problem is nonsingular. A necessary and sufficient condition for the existence of the constrained CRB is also derived.**

***Index Terms*— Blind channel identification, Cramér–Rao bound, equality constraints, parametric estimation.**

I. Introduction

THE CRAMÉR-RAO bound (CRB) matrix provides a lower bound on the covariance matrix of any unbiased estimate of a nonrandom parameter vector. It is often used to investigate the optimality of parametric estimators. In some applications, the parameter space may be confined to a known subset of the Euclidean space through smooth functional constraints on the parameters. The CRB under the parametric constraints can be found by a reparameterization of the original problem to remove redundancies in the parameter vector. However, this approach may be difficult, and may also hinder insights into the original unconstrained problem.

A constrained CRB expression has been obtained in [1] and [2] under the assumption that the Fisher information matrix (FIM) for the original problem is nonsingular. Such an assumption may not hold in some problems. Some recent examples of such problems include blind channel identification [3] and blind symbol estimation [4], [5].

In this letter, we present a simple and general constrained CRB that does not require the FIM for the unconstrained problem to be full rank. We show that our constrained CRB formula reduces to the results in [1] and [2] when the unconstrained FIM is full rank. We also present a necessary and sufficient condition for the existence of the constrained CRB.

Manuscript received December 30, 1997. This work was supported in part by the Senior Individual Grant Program of the Swedish Foundation for Strategic Research, and by the Department of the Army, Army Research Office, under Grant DAAH04-95-1-0249. The work of P. Stoica was supported by a STINT fellowship. The work of B. C. Ng was supported by a DSO fellowship. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. K. Buckley.

P. Stoica is with the Systems and Control Group, Uppsala University, S-751 03 Uppsala, Sweden.

B. C. Ng is with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: nbooncho@rascals.stanford.edu).

Publisher Item Identifier S 1070-9908(98)05129-3.

II. Problem Statement and Preliminaries

Let $\boldsymbol{y}$ be the vector of observations and $\boldsymbol{\theta} \in \mathbb{R}^{\times 1}$ be the vector of nonrandom parameters to be estimated from $\boldsymbol{y}$. We restrict our attention to the class of unbiased estimators. We denote the estimate of $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$ and we require that $\hat{\boldsymbol{\theta}}$ satisfies k $(k < n)$ continuously differentiable constraints,

$$\boldsymbol{f}(\hat{\boldsymbol{\theta}}) = \boldsymbol{0}. \tag{1}$$

We assume that the set $\{\boldsymbol{\theta} | \boldsymbol{f}(\boldsymbol{\theta}) = \boldsymbol{0}\}$ is nonempty (i.e., the constraints are consistent). Let the $k \times n$ gradient matrix of the constraints be defined by

$$\boldsymbol{F}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}. \tag{2}$$

The gradient matrix $\boldsymbol{F}(\boldsymbol{\theta})$ is assumed to have full row rank for any $\boldsymbol{\theta}$ satisfying (1) (i.e., the constraints are nonredundant), and hence there exists a matrix $\boldsymbol{U} \in \mathbb{R}^{n \times (n-k)}$ whose columns form an orthonormal basis for the nullspace of $\boldsymbol{F}(\boldsymbol{\theta})$, that is,

$$\boldsymbol{F}(\boldsymbol{\theta})\boldsymbol{U} = \boldsymbol{0} \tag{3}$$

where $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$. Let $p(\boldsymbol{y}; \boldsymbol{\theta})$ be the likelihood function of the observed data and denote

$$\boldsymbol{\Delta} = \frac{\partial \log p(\boldsymbol{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \tag{4}$$

Then the FIM for the *unconstrained* parameter estimation problem is given by

$$\boldsymbol{J} = E(\boldsymbol{\Delta}\boldsymbol{\Delta}^T).$$

If $\boldsymbol{J}$ is nonsingular, then $\boldsymbol{J}^{-1}$ is the unconstrained CRB for the error covariance matrix of any unbiased estimate of $\boldsymbol{\theta}$. Furthermore, in such a case the constrained CRB is given by an expression derived in [1] and [2] [see also (11)].

Our objective in this letter is to find the CRB on the covariance matrix of any unbiased estimate satisfying the constraints (1), without imposing the nonsingularity condition on $\boldsymbol{J}$.

III. The Constrained CRB

To derive the constrained CRB, we make use of the following known fact (see [2, Th. 1]).

Fact: Let $\hat{\boldsymbol{\theta}}$ be an unbiased estimate of $\boldsymbol{\theta}$ satisfying (1) and let $\boldsymbol{U}$ be as defined in (3). Then under regularity conditions[1]

$$E((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\boldsymbol{\Delta}^T)\boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{U}\boldsymbol{U}^T. \tag{5}$$

We are now ready to state the main result.

Theorem 1: Let $\hat{\boldsymbol{\theta}}$ be an unbiased estimate of $\boldsymbol{\theta}$ satisfying (1) and let $\boldsymbol{U}$ be defined via (3). If $\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U}$ is nonsingular, then

$$E((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T) \geq \boldsymbol{U}(\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U})^{-1}\boldsymbol{U}^T \tag{6}$$

where equality is achieved if and only if

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \boldsymbol{U}(\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U})^{-1}\boldsymbol{U}^T\boldsymbol{\Delta} \text{ (in the mean square sense)}. \tag{7}$$

Proof: Let $\boldsymbol{P}_{\boldsymbol{U}} = \boldsymbol{U}\boldsymbol{U}^T$ be the projection matrix onto the column space of $\boldsymbol{U}$ and let $\boldsymbol{W} \in \mathbb{R}^{n\times n}$ be an arbitrary matrix. Also, let $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$. Then

$$\begin{aligned}&E(\tilde{\boldsymbol{\theta}} - \boldsymbol{W}\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{\Delta})(\tilde{\boldsymbol{\theta}} - \boldsymbol{W}\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{\Delta})^T\\ &\overset{(a)}{=} E(\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^T) - \boldsymbol{W}\boldsymbol{P}_{\boldsymbol{U}} - \boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{W}^T + \boldsymbol{W}\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{J}\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{W}^T \overset{(b)}{\geq} \boldsymbol{0}\end{aligned} \tag{8}$$

where equality (a) follows from (5) and inequality (b) is a consequence of the positive semi-definiteness of the covariance matrix of $\tilde{\boldsymbol{\theta}} - \boldsymbol{W}\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{\Delta}$. Thus, we have

$$E(\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^T) \geq \boldsymbol{W}\boldsymbol{P}_{\boldsymbol{U}} + \boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{W}^T - \boldsymbol{W}\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{J}\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{W}^T. \tag{9}$$

The *greatest* lower bound is obtained as the supremum over $\boldsymbol{W}$ of the right hand side of (9). Let $\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$ be the eigendecomposition of $\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U}$, where $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$. Since $\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U}$ is positive definite, $|\boldsymbol{\Lambda}| \neq 0$. It can be readily shown that

$$\begin{aligned}&\boldsymbol{W}\boldsymbol{P}_{\boldsymbol{U}} + \boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{W}^T - \boldsymbol{W}\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{J}\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{W}^T\\ &\quad= \boldsymbol{U}\boldsymbol{Q}\boldsymbol{\Lambda}^{-1}\boldsymbol{Q}^T\boldsymbol{U}^T - (\boldsymbol{W}\boldsymbol{U}\boldsymbol{Q} - \boldsymbol{U}\boldsymbol{Q}\boldsymbol{\Lambda}^{-1})\\ &\qquad\cdot \boldsymbol{\Lambda}(\boldsymbol{W}\boldsymbol{U}\boldsymbol{Q} - \boldsymbol{U}\boldsymbol{Q}\boldsymbol{\Lambda}^{-1})^T.\end{aligned}$$

Hence, the maximizing $\boldsymbol{W}$ satisfies

$$\boldsymbol{W}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{Q}\boldsymbol{\Lambda}^{-1}\boldsymbol{Q}^T. \tag{10}$$

Substituting (10) into (9) gives the result

$$E(\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^T) \geq \boldsymbol{U}(\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U})^{-1}\boldsymbol{U}^T.$$

If the equality in (6) holds true, then we see from (8) that

$$\begin{aligned}&E(\tilde{\boldsymbol{\theta}} - \boldsymbol{W}\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{\Delta})(\tilde{\boldsymbol{\theta}} - \boldsymbol{W}\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{\Delta})^T\\ &\quad= \boldsymbol{U}(\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U})^{-1}\boldsymbol{U}^T - \boldsymbol{W}\boldsymbol{P}_{\boldsymbol{U}} - \boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{W}^T\\ &\qquad+ \boldsymbol{W}\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{J}\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{W}^T\end{aligned}$$

which holds for every $\boldsymbol{W}$. In particular, if $\boldsymbol{W}$ satisfies (10), then $E(\tilde{\boldsymbol{\theta}} - \boldsymbol{W}\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{\Delta})(\tilde{\boldsymbol{\theta}} - \boldsymbol{W}\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{\Delta})^T = \boldsymbol{0}$ which implies (7) holds. On the other hand, if (7) holds, then it is trivial to show that $E(\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^T) = \boldsymbol{U}(\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U})^{-1}\boldsymbol{U}^T$. ■

The next corollary shows that the lower bound obtained in Theorem 1 reduces to the constrained CRB expression obtained in [1] and [2] when $\boldsymbol{J}$ is positive definite.[2]

Corollary 1: Assume $\boldsymbol{J}$ is positive definite and omit the argument of $\boldsymbol{F}(\boldsymbol{\theta})$ for notational convenience. Then

$$\boldsymbol{U}(\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U})^{-1}\boldsymbol{U}^T = \boldsymbol{J}^{-1} - \boldsymbol{J}^{-1}\boldsymbol{F}^T(\boldsymbol{F}\boldsymbol{J}^{-1}\boldsymbol{F}^T)^{-1}\boldsymbol{F}\boldsymbol{J}^{-1}. \tag{11}$$

Proof: Write $\boldsymbol{J} = \boldsymbol{J}^{T/2}\boldsymbol{J}^{1/2}$, and observe that

$$\begin{aligned}\boldsymbol{F}\boldsymbol{U} = \boldsymbol{0} \Leftrightarrow \boldsymbol{F}\boldsymbol{J}^{-(1/2)}\boldsymbol{J}^{1/2}\boldsymbol{U} = \boldsymbol{0} \Leftrightarrow \boldsymbol{P}_{\boldsymbol{J}^{1/2}\boldsymbol{U}}\\ = \boldsymbol{P}^{\perp}_{\boldsymbol{J}^{-(T/2)}\boldsymbol{F}^T}\end{aligned}$$

where $\boldsymbol{P}^{\perp}_{\boldsymbol{J}^{-(T/2)}\boldsymbol{F}^T}$ is the projection matrix onto the orthogonal complement of the column space of $\boldsymbol{J}^{-(T/2)}\boldsymbol{F}^T$. Thus

$$\begin{aligned}\boldsymbol{U}(\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U})^{-1}\boldsymbol{U}^T &= \boldsymbol{J}^{-(1/2)}\boldsymbol{P}_{\boldsymbol{J}}^{1/2}\boldsymbol{U}\boldsymbol{J}^{-(T/2)}\\ &= \boldsymbol{J}^{-(1/2)}\boldsymbol{P}^{\perp}_{\boldsymbol{J}^{-(T/2)}\boldsymbol{F}^T}\boldsymbol{J}^{-(T/2)}\end{aligned}$$

which yields (11). ■

The next proposition shows that the condition $\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U} > \boldsymbol{0}$ in Theorem 1 is necessary and sufficient for the existence of the (finite) constrained CRB matrix. The lower bound in Theorem 1 is thus the most general form of the constrained CRB matrix.

Proposition 1: A necessary and sufficient condition for the existence of a (finite) constrained CRB matrix is

$$|\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U}| \neq 0. \tag{12}$$

Proof: The sufficiency part follows from Theorem 1. To show the necessity of (12), suppose that (12) does not hold. Then $\mathrm{col}\,(\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{J}\boldsymbol{P}_{\boldsymbol{U}}) \subset \mathrm{col}\,(\boldsymbol{P}_{\boldsymbol{U}})$ or equivalently $\mathrm{null}\,(\boldsymbol{P}_{\boldsymbol{U}}) \subset \mathrm{null}\,(\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{J}\boldsymbol{P}_{\boldsymbol{U}})$. Let $\boldsymbol{\rho} \in \mathrm{null}\,(\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{J}\boldsymbol{P}_{\boldsymbol{U}}) - \mathrm{null}\,(\boldsymbol{P}_{\boldsymbol{U}})$ and let $\boldsymbol{W} = \frac{\alpha}{2}\boldsymbol{\rho}\boldsymbol{\rho}^T$ where $||\boldsymbol{\rho}|| = 1$. Then it follows from (9) that

$$\mathrm{Tr}\,(E(\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^T)) \geq \alpha||\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{\rho}||^2; \quad ||\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{\rho}|| > 0.$$

Since α can be chosen arbitrarily large, no unbiased estimate of $\boldsymbol{\theta}$ with bounded variance exists in such a case. ■

The constrained CRB expression in (6) can be applied to problems where the unconstrained estimation problem leads to a singular FIM. Possible applications include blind and semi-blind channel identification and equalization, directions of arrival estimation in array processing, and spectrum estimation. The result in Proposition 1 can also be used to check the *identifiability* of the constrained problem.

IV. A Blind Channel Identification Example

In this section, we compute the constrained CRB for a blind channel identification example (more details on this application will be presented elsewhere [8]). The data model is

$$\boldsymbol{x}_k = \boldsymbol{H}\boldsymbol{s}_k + \boldsymbol{v}_k \tag{13}$$

where $\boldsymbol{x}_k$ is the received signal at the kth time instant, $\boldsymbol{H} \in \mathbb{C}^{P\times L}$ is the channel matrix, $\boldsymbol{s}_k = [s_k, \cdots, s_{k-L+1}]^T$ is the $L \times 1$ vector of data symbols, and $\boldsymbol{v}_k$ is a vector of additive white Gaussian noise with covariance $\sigma^2\boldsymbol{I}$. The

[1] The regularity conditions are required for the interchange of certain integration and differentiation operators (see [6] for details).

[2] An alternative algebraic proof of (11) as a matrix identity was also given in [7, Lemma 1].

$$\boldsymbol{H} = \begin{bmatrix} 0.3079 + j0.0698 & 0.1657 + j0.2304 & 0.0198 - j0.3823 & 0.0929 - j0.1853 \\ -0.1841 + j0.3294 & 0.4484 - j0.1689 & 0.0156 + j0.1526 & 0.4750 - j0.0952 \end{bmatrix}$$

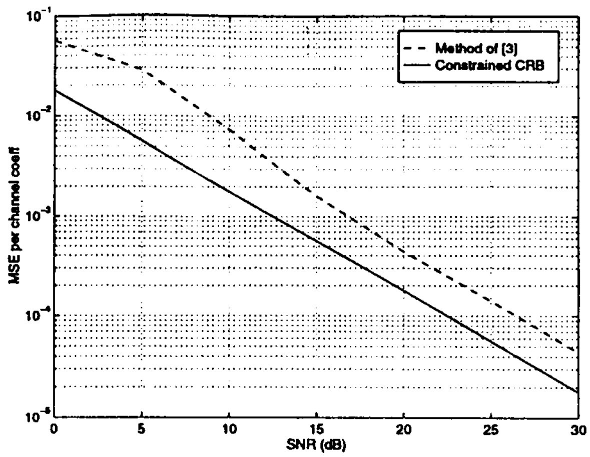

Fig. 1. Blind channel identification example.

unknown parameters are the elements of $\boldsymbol{H}$ and the data $\{s_k\}$ which are here assumed to be real-valued.

The blind channel identification algorithm used is the one proposed in [3] with a smoothing factor of eight. The constraint on the channel is $||\boldsymbol{h}||^2 = 1$, where $\boldsymbol{h} = \text{vec}\,(\boldsymbol{H})$. The number of data symbols is 50 and the symbols are generated as binary phase shift keying symbols (BPSK) (i.e., $s_k \in \{\pm 1\}$). The symbols are randomly chosen from a Bernoulli $(\frac{1}{2})$ distribution and then fixed throughout the numerical example. The channel matrix $\boldsymbol{H}$ is given by the expression shown at the top of the page. Fig. 1 compares the mean square error (MSE) of the channel parameter estimates obtained via the method of [3] with the constrained CRB, for various values of signal-to-noise ratio (SNR). The MSE (per real channel coefficient) is defined as $\text{MSE} = (1/16N_m)\,\Sigma_{i=1}^{N_m}\,||\hat{\boldsymbol{h}}_i - \boldsymbol{h}||^2$ where $\hat{\boldsymbol{h}}_i$ is the channel estimate obtained in the ith trial and N_m is the total number of trials. In our example, $N_m = 100$. The SNR is defined as $\text{SNR} = 10\log_1 0\,(1/\sigma^2)$. We observe that as the SNR increases, the method's MSE approaches the constrained CRB.

References

[1] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inform. Theory*, vol. 26, pp. 1285–1301, Nov 1990.
[2] T. L. Marzetta, "A simple derivation of the constrained multiple parameter Cramér-Rao bound," *IEEE Trans. Signal Processing*, vol. 41, pp. 2247–2249, June 1993.
[3] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.
[4] H. Liu and G. Xu, "Closed-form blind symbol estimation in digital communications," *IEEE Trans. Signal Processing*, vol. 43, pp. 2714–2723, Nov. 1995.
[5] A. J van der Veen, S. Talwar, and A. Paulraj, "A subspace approach to blind space-time signal processing for wireless communication systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 173–190, Jan. 1997.
[6] S. Zacks, *The Theory of Statistical Inference*. New York: Wiley, 1971.
[7] C. G. Khatri, "A note on a MANOVA model applied to problems in growth curve," *Ann. Inst. Stat. Math.*, vol. 18, pp. 75–86, 1966.
[8] B. C. Ng, P. Stoica, and A. Paulraj, "Performance study of semi-blind and blind channel estimation methods," in preparation.