The Annals of Statistics
1987, Vol. 15, No. 4, 1421–1438

SOME CLASSES OF GLOBAL CRAMÉR–RAO BOUNDS

BY B. Z. BOBROVSKY, E. MAYER-WOLF AND M. ZAKAI[1]

Tel-Aviv University, Technion and Technion

This paper considers Cramér–Rao type bounds for the estimation error of a parameter in a Bayesian setup. This class of bounds, introduced by Van Trees, proved useful in various stochastic communications and control problems. Two issues are considered in this paper. The first deals with a comparison of the tightness of several different versions of the bound in the multivariate case. The second introduces several useful generalizations of the original version of the bound.

1. Introduction. The Cramér–Rao bound [10] was developed in the context of non-Bayesian estimation and provides a lower bound for the mean square error of any unbiased estimator of a parameter; this bound is, in general, a function of the parameter. In [12] Van Trees presented a global (or Bayesian) Cramér–Rao inequality,

$$E(x - E(x|y))^2 \geq \frac{1}{E[\partial/\partial x(\ln p(x, y))]^2}, \tag{1}$$

where $p(\xi, \eta)$ is the joint density of the random variables x and y. In the Bayesian setting, lower bounds on the estimation error are, in principle, less important since expressions for the least square estimator and error are available. It turns out, however, that in many cases (1) provides a tight and useful lower bound for the estimation error when the conditional expectation is difficult to find explicitly. For example, this bound was applied in [4] to compare the mean square estimation error for nonlinear filtering problems with that of related Gaussian problems (cf. also [6]).

A proof of (1) is included as a particular case of Proposition 2 in Section 4. It is noteworthy that unlike the classical Cramér–Rao bound, inequality (1) requires no uniform integrability condition on the density's derivative since the proof involves no interchange of derivatives and integrals.

This paper addresses basically two issues. The first one concerns the following problem. Consider the situation where two random variables x_1 and x_2 are to be estimated via a (scalar) measurement y. A straightforward extension of (1) (cf. [12]) is the matrix inequality

$$E\{(\mathbf{x} - E(\mathbf{x}|y))(\mathbf{x} - E(\mathbf{x}|y))^T\} \geq \mathbf{J}^{-1}, \tag{2}$$

Received April 1986; revised March 1987.
[1]Work supported by the Fund for Promotion of Research at the Technion.
AMS 1980 *subject classifications*. Primary 62F15; secondary 93E11.
Key words and phrases. Cramér-Rao inequality, Bayesian estimation.

where $\mathbf{J}$ is the Fisher information matrix

$$(\mathbf{J})_{ij} = E\left[\left(\frac{\partial}{\partial \xi_1} p(x_1, x_2, y) \frac{\partial}{\partial \xi_2} p(x_1, x_2, y)\right) \Big/ p(x_1, x_2, y)^2\right]$$
$$= -E\left[\left(\frac{\partial^2}{\partial \xi_1 \partial \xi_2} p(x_1, x_2, y)\right) \Big/ p(x_1, x_2, y)^2\right]$$

and $p(\xi_1, \xi_2, \eta)$ is now the joint density of x_1, x_2 and y. Two lower bounds thus arise for $E(x_1 - E(x_1|y))^2$: The first is $(\mathbf{J}^{-1})_{11}$ obtained by considering the first diagonal element in (2); the second is the right-hand side of (1) for $x = x_1$. We show in Section 3 that the bound obtained in (1) using the marginal density of x_1 and y is always greater than or equal to the bound provided by $(\mathbf{J}^{-1})_{11}$ in (2). This and related results will be obtained for the n-dimensional case. The measurement y will be assumed to take values in a general space. Still another bound can be obtained by considering x_2 as a "nuisance parameter" known to the observer, obtaining thus a bound depending on this parameter and then averaging over it. This approach has been suggested by Miller and Chang [11]. It is shown in Section 3 through examples that while this approach may yield bounds inferior to $(\mathbf{J}^{-1})_{11}$, surprisingly enough there are cases in which it yields tighter results than the right-hand side of (1) for $x = x_1$ (an application of this approach to the nonlinear filtering problem is given in [6]).

The second issue concerns generalizing the inequality (1) and its multidimensional counterparts. Several instances of such generalizations for the non-Bayesian situation are already known in the form of the Bhattacharya bound [3], the Barankin bound [2], etc. (cf. also [7]), and can easily be expressed in the Bayesian setting as well. In Section 4, the following generalization of (1) will be shown to hold:

$$(3) \qquad E(x - E(x|y))^2 \geq \frac{(Eq(x, y))^2}{E[q(x, y)\partial/\partial x(\ln p(x, y)q(x, y))]^2}$$

for a very large class of functions q, together with its multidimensional counterparts. Furthermore, it is shown in Section 4 that the supremum of the right-hand side of (3) over all admissible functions $q(\cdot, \cdot)$ achieves equality. Examples are given for which the use of a properly chosen q yields useful bounds, whereas for $q \equiv 1$ the bound (1) is trivial. Further generalizations of the global Cramér–Rao bound are derived in Section 5. In particular, it is shown that the densities $p(x, y)$ appearing in the bounds (1) and (3) can be replaced by Radon–Nikodym derivatives with respect to certain reference measures. This approach enables derivation of bounds for cases where expressions for $p(x, y)$ are untractable while Radon–Nikodym derivatives with respect to certain reference measures are available. An example where this approach is useful is the nonlinear filtering problem (cf. [4], [6]).

This paper is devoted to Cramér–Rao type bounds for Bayesian problems. It should, however, be pointed out that this is not the only method for deriving error bounds in the Bayesian setup (cf., e.g., [8], [14]).

2. Notation. Throughout $\mathbf{x} = (x_i)_{i=1}^n$ is a random vector satisfying

$$Ex_i^2 < \infty, \qquad 1 \le i \le n. \tag{4}$$

Let $\mathscr{Y}$ denote a subsigma algebra and assume the existence of a differentiable conditional density of $\mathbf{x}$ when conditioned on $\mathscr{Y}$ denoted $p_{\mathbf{x}|\mathscr{Y}}(\xi_1, \dots, \xi_n)$ or $p_{\mathbf{x}|\mathscr{Y}}(\boldsymbol{\xi})$ [in Section 5 the regularity conditions on $p_{\mathbf{x}|\mathscr{Y}}(\boldsymbol{\xi})$ will be relaxed]; $p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x})$ will denote the random variable obtained by setting $\boldsymbol{\xi} = \mathbf{x}$ in $p_{\mathbf{x}|\mathscr{Y}}(\boldsymbol{\xi})$. Note that if $p_{\mathbf{x}|\mathscr{Y}}(\cdot)$ is piecewise continuous with nonrandom points of discontinuity, then $p_{\mathbf{x}|\mathscr{Y}}(\xi(\omega))$ is indeed a measurable function of ω, namely a random variable. Let L be an inner product space and consider $\mathbf{u} = (u_i)_{i=1}^n$ and $\mathbf{v} = (v_i)_{i=1}^m$ which are, respectively, n and m vectors with elements in L. The Grammian $n \times m$ matrix $((\mathbf{u}, \mathbf{v}))$ is defined by $((\mathbf{u}, \mathbf{v}))_{ij} = (u_i, v_j)$, $1 \le i \le n$ and $1 \le j \le m$, where $(\ ,\ )$ is L's inner product and set $((\mathbf{u})) = ((\mathbf{u}, \mathbf{u}))$. Our interest lies in $L = L^2(\Omega)$, the space of second order random variables with inner product $(w, z) = Ewz$. We now denote

$$\mathbf{a}(\mathbf{x}|\mathscr{Y}) = \frac{\operatorname{grad} p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x})}{p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x})} = \operatorname{grad} \log p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x}), \tag{5}$$

where grad $= (D_i)_{i=1}^n$ is the gradient operator. It must be pointed out that the quantity defined in (5) is legitimate since $p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x}) > 0$ a.s. (indeed,

$$\begin{aligned} P\{p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x}) = 0\} &= E\left[E\left(\mathbf{1}_{\{0\}}\left(p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x})\right)|\mathscr{Y}\right)\right] \\ &= E\int_{\mathbb{R}^n} \mathbf{1}_{\{0\}}\left(p_{\mathbf{x}|\mathscr{Y}}(\boldsymbol{\xi})\right) p_{\mathbf{x}|\mathscr{Y}}(\boldsymbol{\xi})\, d\boldsymbol{\xi} = 0; \end{aligned}$$

this argument will frequently be implicit in the sequel).

We further assume that

$$E\left(a_i(\mathbf{x}|\mathscr{Y})\right)^2 < \infty, \qquad 1 \le i \le n, \tag{6}$$

and consequently define the Fisher information matrix

$$\mathbf{J}(\mathbf{x}|\mathscr{Y}) = ((\mathbf{a}(\mathbf{x}, y))) = \left[E\left(\frac{\partial}{\partial x_i}\log p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x}) \frac{\partial}{\partial x_j}\log p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x})\right)\right]_{i,j=1}^n \tag{7}$$

and also the error matrix

$$\boldsymbol{\Sigma}(\mathbf{x}, y) = ((\mathbf{x} - E(\mathbf{x}|\mathscr{Y}))) = \left[E\left(x_i - E(x_i|\mathscr{Y})\right)\left(x_j - E(x_j|\mathscr{Y})\right)\right]_{i,j=1}^n. \tag{8}$$

Whenever no confusion arises, one or both of the arguments of the expressions defined in (5), (7) and (8) may be omitted. For multiindices α, β and γ, i.e., ordered subsequences of $[n] =: \{1, 2, \dots, n\}$, the expressions $\mathbf{v}_\alpha$ and $\mathbf{M}_{\beta,\gamma}$ denote, respectively, the corresponding partial vector of $\mathbf{v} \in \mathbb{R}^n$ and submatrix of $\mathbf{M} \in \mathbb{R}^{n\times n}$. When n is clear from the context, α^* will denote the complementary multiindex of α in $[n]$. For an n-vector $\mathbf{u}$, diag($\mathbf{u}$) will denote the diagonal $n \times n$ matrix $\mathbf{D}$ with $\mathbf{D}_{ii} = u_i$.

Finally, for two symmetric matrices $\mathbf{M}_1$ and $\mathbf{M}_2$, the inequality $\mathbf{M}_1 \geq \mathbf{M}_2$ denotes that $\mathbf{M}_1 - \mathbf{M}_2$ is nonnegative definite.

3. Comparison among different multidimensional bounds. A global version of the Cramér–Rao inequality derived by Van Trees states that under the assumptions (4) and (6), $\boldsymbol{\Sigma}(\mathbf{x}|\mathscr{Y})\mathbf{J}(\mathbf{x}|\mathscr{Y}) \geq \mathbf{I}_n$ ($\mathbf{I}_n$ is the identity $n \times n$ matrix), which implies that $\mathbf{J}(\mathbf{x}|\mathscr{Y})$ is nonsingular and $\mathbf{J}^{-1}(\mathbf{x}|\mathscr{Y})$ constitutes a lower bound for the error matrix $\boldsymbol{\Sigma}(\mathbf{x}|\mathscr{Y})$; for the proof cf. [12] or Section 4 of this paper with $q \equiv 1$. Now suppose we are only interested in some of $\mathbf{x}$'s components, say $\mathbf{x}_\alpha$ for some multiindex α. Since nonnegative definiteness of a matrix is shared by all of its principal submatrices, we have

$$\boldsymbol{\Sigma}(\mathbf{x}_\alpha) = \left(\boldsymbol{\Sigma}(\mathbf{x})\right)_{\alpha,\alpha} \geq \left(\mathbf{J}^{-1}(\mathbf{x})\right)_{\alpha,\alpha}. \tag{9}$$

On the other hand, the global Cramér–Rao bound applied directly to $\mathbf{x}_\alpha$ yields

$$\boldsymbol{\Sigma}(\mathbf{x}_\alpha) \geq \mathbf{J}^{-1}(\mathbf{x}_\alpha). \tag{10}$$

The first inequality of the following proposition establishes that the lower bound appearing in (10) is tighter than the one in (9). In other words, as far as the Van Trees version of the global Cramér–Rao bound is concerned, no improvement can be obtained by embedding a given system in a larger system (with more components in the $\mathbf{x}$ vector).

PROPOSITION 1. *Under the assumptions* (4) *and* (6) *and for any multiindex* α (*subsequence of* $[n]$),

$$\mathbf{J}^{-1}(\mathbf{x}_\alpha) \geq \left(\mathbf{J}^{-1}(\mathbf{x})\right)_{\alpha,\alpha} \geq \left\{\left(\mathbf{J}(\mathbf{x})\right)_{\alpha,\alpha}\right\}^{-1}. \tag{11}$$

PROOF. First note that if A and B are two positive definite matrices, $A \geq B$ iff $B^{-1} \geq A^{-1}$. Thus, (11) is equivalent to

$$\left(\mathbf{J}(\mathbf{x})\right)_{\alpha,\alpha} \geq \left(\left(\mathbf{J}^{-1}(\mathbf{x})\right)_{\alpha,\alpha}\right)^{-1} \geq \mathbf{J}(\mathbf{x}_\alpha), \tag{11'}$$

which in itself shall be seen to be contained in the following chain of inequalities:

$$\begin{aligned}
\mathbf{J}_{\alpha,\alpha}(\mathbf{x}) &\geq \mathbf{J}_{\alpha,\alpha}(\mathbf{x}) - \mathbf{J}_{\alpha,\alpha^*}(\mathbf{x})\left(\mathbf{J}_{\alpha^*,\alpha^*}(\mathbf{x})\right)^{-1}\mathbf{J}_{\alpha^*,\alpha}(\mathbf{x}) \\
&= \left(\left(\mathbf{a}_\alpha(\mathbf{x}) - \mathbf{J}_{\alpha,\alpha^*}(\mathbf{x})\left(\mathbf{J}_{\alpha^*,\alpha^*}(\mathbf{x})\right)^{-1}\mathbf{a}_{\alpha^*}(\mathbf{x})\right)\right) \\
&\geq \left(\left(E\left(\mathbf{a}_\alpha(\mathbf{x}) - \mathbf{J}_{\alpha,\alpha^*}(\mathbf{x})\left(\mathbf{J}_{\alpha^*,\alpha^*}(\mathbf{x})\right)^{-1}\mathbf{a}_{\alpha^*}(\mathbf{x})|\mathbf{x}_\alpha, \mathscr{Y}\right)\right)\right) \\
&= \left(\left(\mathbf{a}(\mathbf{x}_\alpha)\right)\right) \\
&= \mathbf{J}(\mathbf{x}_\alpha).
\end{aligned} \tag{12}$$

The proof of (12) is as follows: The first, second and last lines follow by definition or simple verification. The third line follows from a direct extension of Pythagoras' formula to Grammians; namely, if $\mathbf{z}$ is a random vector and $\tilde{\mathbf{z}}$ the conditional expectation of $\mathbf{z}$ conditioned on some given subsigma algebra, then $(E\mathbf{z}\mathbf{z}^T - E\tilde{\mathbf{z}}\tilde{\mathbf{z}}^T) = E(\mathbf{z} - \tilde{\mathbf{z}})(\mathbf{z} - \tilde{\mathbf{z}})^T \geq 0$.

Finally, the fourth line in (12) is a consequence of

LEMMA 1.

$$E(\mathbf{a}_\alpha(\mathbf{x})|\mathbf{x}_\alpha, \mathscr{Y}) = \mathbf{a}(\mathbf{x}_\alpha) \quad a.s., \tag{13}$$

$$E(\mathbf{a}_{\alpha^*}(\mathbf{x})|\mathbf{x}_\alpha, \mathscr{Y}) = 0 \quad a.s. \tag{14}$$

PROOF. Let $1 \le i \le n$ and assume α has m components. Then, almost surely,

$$\begin{aligned} E(a_i(x)|\mathbf{x}_\alpha, \mathscr{Y}) &= \int_{\mathbb{R}^{n-m}} \frac{D_i p_{\mathbf{x}|\mathscr{Y}}(\boldsymbol{\xi})}{p_{\mathbf{x}|\mathscr{Y}}(\boldsymbol{\xi})} p_{\mathbf{x}_{\alpha^*}|\mathbf{x}_\alpha, \mathscr{Y}}(\boldsymbol{\xi}_{\alpha^*})\, d\boldsymbol{\xi}_{\alpha^*}\Big|_{\boldsymbol{\xi}_\alpha = \mathbf{x}_\alpha} \\ &= \frac{1}{p_{\mathbf{x}_\alpha|\mathscr{Y}}(\mathbf{x}_\alpha)} \left[\int_{\mathbb{R}^{n-m}} D_i p_{\mathbf{x}|\mathscr{Y}}(\boldsymbol{\xi})\, d\boldsymbol{\xi}_{\alpha^*}\Big|_{\boldsymbol{\xi}_\alpha = \mathbf{x}_\alpha} \right]. \end{aligned} \tag{15}$$

If i is one of α's components,

$$\int_{\mathbb{R}^{n-m}} D_i p_{\mathbf{x}|\mathscr{Y}}(\boldsymbol{\xi})\, d\boldsymbol{\xi}_{\alpha^*}\Big|_{\boldsymbol{\xi}_\alpha = \mathbf{x}_\alpha} = D_i \int_{\mathbb{R}^{n-m}} p_{\mathbf{x}|\mathscr{Y}}(\boldsymbol{\xi})\, d\boldsymbol{\xi}_{\alpha^*}\Big|_{\boldsymbol{\xi}_\alpha = \mathbf{x}_\alpha} = D_i p_{\mathbf{x}_\alpha|\mathscr{Y}}(\mathbf{x}_\alpha) \quad \text{a.s.},$$

yielding $E(a_i(\mathbf{x})|\mathbf{x}_\alpha, \mathscr{Y}) = \mathbf{a}(\mathbf{x}_\alpha)$ a.s., which is (13).

If i is not one of α's components, append i to α and call the augmented multiindex $\tilde{\alpha}$. Then

$$\int_{\mathbb{R}^{n-m}} D_i p_{\mathbf{x}|\mathscr{Y}}(\boldsymbol{\xi})\, d\boldsymbol{\xi}_{\alpha^*} = \int_{\mathbb{R}^{n-m-1}} d\boldsymbol{\xi}_{\tilde{\alpha}^*} \int_{-\infty}^{\infty} D_i p_{\mathbf{x}|\mathscr{Y}}(\boldsymbol{\xi})\, d\xi_i = 0. \tag{16}$$

To be precise, the preceding equality holds for (Lebesgue) almost every $\boldsymbol{\xi}_\alpha$ in $\mathbb{R}^m$. This can be justified by the following standard argument. Since

$$1 = \int_{\mathbb{R}^n} p_{\mathbf{x}|\mathscr{Y}}(\boldsymbol{\xi}) = \int_{\mathbb{R}^{n-1}} d\boldsymbol{\xi}_{i^*} \int_{-\infty}^{\infty} p_{\mathbf{x}|\mathscr{Y}}(\boldsymbol{\xi})\, d\xi_i,$$

the inner integral is finite for (Lebesgue) almost every $\boldsymbol{\xi}_{i^*} \in \mathbb{R}^{n-1}$. Thus for each such fixed $\boldsymbol{\xi}_{i^*}$, there exist sequences $s_k \downarrow -\infty$ and $t_k \uparrow \infty$ such that

$$\lim_{k\to\infty} p_{\mathbf{x}|\mathscr{Y}}(\boldsymbol{\xi})\Big|_{\xi_i = s_k} = \lim_{k\to\infty} p_{\mathbf{x}|\mathscr{Y}}(\boldsymbol{\xi})\Big|_{\xi_i = t_k} = 0$$

and thus

$$\int_{-\infty}^{\infty} D_i p_{\mathbf{x}|\mathscr{Y}}(\boldsymbol{\xi})\, d\xi_i = \lim_{k\to\infty} \int_{s_k}^{t_k} D_i p_{\mathbf{x}|\mathscr{Y}}(\boldsymbol{\xi})\, d\xi_i = 0.$$

Now combining (15) and (16) we obtain (14). This proves the lemma. □

Finally note that $(\mathbf{J}^{-1})_{\alpha,\alpha} = (\mathbf{J}_{\alpha,\alpha} - \mathbf{J}_{\alpha,\alpha^*}(\mathbf{J}_{\alpha^*,\alpha^*})^{-1}\mathbf{J}_{\alpha^*,\alpha})^{-1}$ so that to conclude the proof of Proposition 1, it only remains to collect the first, second and last terms in (12). □

REMARK. The right-hand side inequality in (11) holds for any positive definite matrix.

A different approach was adopted by Miller and Chang in [11] to provide a lower bound for $\Sigma(\mathbf{x}_\alpha)$ in the presence of $\mathbf{x}_{\alpha^*}$ (termed nuisance parameters in [11]). The bound of [11] and an immediately related result are summed up in the following lemma. To deal with the present situation the notation will be extended as follows. For any multiindex β and random vectors $\mathbf{u}$ and $\mathbf{v}$, $((\mathbf{u},\mathbf{v}))^{(\beta)}$ will be a random $n \times m$ matrix with entries $E((u_i, v_j)|\mathbf{x}_\beta)$. As before, $((\mathbf{u}))^{(\beta)}$ will stand for $((u,u))^{(\beta)}$. In particular, define $\mathbf{J}^{(\beta)}(\mathbf{x}|\mathscr{Y}) = ((\mathbf{a}(\mathbf{x}|\mathscr{Y})))^{(\beta)}$. Also grad_β will indicate the partial gradient operator along the components of β.

LEMMA 2.

$$\Sigma(\mathbf{x}_\alpha) \geq E\left\{\left(\mathbf{J}^{(\alpha^*)}(\mathbf{x})\right)_{\alpha,\alpha}\right\}^{-1} \geq \left\{\mathbf{J}(\mathbf{x})_{\alpha,\alpha}\right\}^{-1}. \tag{17}$$

REMARK. The left-hand side inequality is the bound of Miller and Chang and the right-hand side is the statement that it is tighter than the weakest (right-hand term) of the bounds in (11).

PROOF.

$$\begin{aligned}\Sigma(\mathbf{x}_\alpha) &= E\left[((\mathbf{x}_\alpha - E(\mathbf{x}_\alpha|\mathscr{Y})))^{(\alpha^*)}\right]\\ &\geq E\left[((\mathbf{x}_\alpha - E(\mathbf{x}_\alpha|\mathscr{Y},\mathbf{x}_{\alpha^*})))^{(\alpha^*)}\right]\\ &\geq E\left[\left\{\left(\left(\frac{\text{grad}_\alpha p_{\mathbf{x}_\alpha|\mathscr{Y},\mathbf{x}_{\alpha^*}}(\mathbf{x}_\alpha)}{p_{\mathbf{x}_\alpha|\mathscr{Y},\mathbf{x}_{\alpha^*}}(\mathbf{x}_\alpha)}\right)\right)^{(\alpha^*)}\right\}^{-1}\right]\\ &= E\left[\left\{\left(\left(\frac{\text{grad}_\alpha p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x})}{p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x})}\right)\right)^{(\alpha^*)}\right\}^{-1}\right] = E\left\{\left(\mathbf{J}^{(\alpha^*)}(\mathbf{x})\right)_{\alpha\alpha}\right\}^{-1}.\end{aligned}$$

The second inequality in (17) is an immediate consequence of the following general result on matrix valued random variables. □

LEMMA 3. *Let* $\mathbf{A}$ *be a random square matrix, almost surely positive definite; then*

$$(E\mathbf{A})^{-1} \leq E(\mathbf{A}^{-1}). \tag{18}$$

PROOF. First observe that any positive definite matrix $\mathbf{D}$ satisfies

$$\mathbf{D} + \mathbf{D}^{-1} \geq 2\mathbf{I} \tag{19}$$

[since $(\mathbf{D}^{1/2} - \mathbf{D}^{-1/2})^2 \geq 0$]. Now let $\mathbf{B}$ and $\mathbf{C}$ be any two positive definite matrices and set $\mathbf{D} = \mathbf{C}^{1/2}\mathbf{B}^{-1}\mathbf{C}^{1/2}$ in (19), from which one obtains

$$\mathbf{C}^{1/2}\mathbf{B}^{-1}\mathbf{C}^{1/2} - \mathbf{I} \geq \mathbf{I} - \mathbf{C}^{-1/2}\mathbf{B}\mathbf{C}^{-1/2}. \tag{20}$$

Multiplying (20) by $\mathbf{C}^{-1/2}$ on the left and on the right yields

$$\mathbf{B}^{-1} - \mathbf{C}^{-1} \geq \mathbf{C}^{-1}(\mathbf{C} - \mathbf{B})\mathbf{C}^{-1}. \tag{21}$$

For $\mathbf{B} = \mathbf{A}$ and $\mathbf{C} = E\mathbf{A}$, one has

$$\mathbf{A}^{-1} - (E\mathbf{A})^{-1} \geq (E\mathbf{A})^{-1}(E\mathbf{A} - \mathbf{A})(E\mathbf{A})^{-1} \tag{22}$$

and the lemma follows by taking expectation on both sides of (22). □

Until now, four lower bounds have been presented for the error matrix of a subvector $\mathbf{x}_\alpha$ of the random vector $\mathbf{x}$ conditioned on the subsigma algebra $\mathscr{Y}$, namely,

$$\mathbf{B}_1 = \mathbf{J}^{-1}(\mathbf{x}_\alpha) \qquad [\text{cf. } (10)],$$

$$\mathbf{B}_2 = \left(\mathbf{J}^{-1}(\mathbf{x})\right)_{\alpha,\alpha} \qquad [\text{cf. } (9)],$$

$$\mathbf{B}_3 = \left((\mathbf{J}(\mathbf{x}))_{\alpha,\alpha}\right)^{-1} \qquad [\text{cf.} (11)],$$

$$\mathbf{B}_4 = E\left\{\left((\mathbf{J}^{(\alpha^*)}(\mathbf{x}))\right)_{\alpha,\alpha}\right\}^{-1} \qquad (\text{the bound of Miller and Chang}).$$

In Proposition 1 it was shown that $\mathbf{B}_1 \geq \mathbf{B}_2 \geq \mathbf{B}_3$. Also, by Lemma 2, $\mathbf{B}_4 \geq \mathbf{B}_3$. The following two examples will show that no further ordering is possible. Indeed, in the first example, $\mathbf{B}_2 > \mathbf{B}_4$, while in the second, $\mathbf{B}_4 > \mathbf{B}_1$. For simplicity, in both examples, $\mathbf{x}$ will be independent of the subsigma algebra $\mathscr{Y}$ so that the error matrix is simply the covariance matrix.

EXAMPLE 3.1. Let $\mathbf{x} = (x_1, x_2)$ be a jointly Gaussian vector such that $x_i \sim N(0,1)$, $i = 1,2$, and $\operatorname{cov}(x_1, x_2) = \rho$, $-1 < \rho < 1$. Choosing $\alpha = \{1\}$, we are interested in bounding $\varepsilon^2 =: \operatorname{var}(x_1) = 1$. We have

$$\Sigma(\mathbf{x}) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

and since $\mathbf{x}$ is Gaussian, equality holds in the global Cramér–Rao inequality [12] so that $J^{-1}(\mathbf{x}) = \Sigma(\mathbf{x})$ and $B_2 = 1$, a tight bound for ε^2. On the other hand,

$$J^{(2)}(\mathbf{x})_{11} = E\left(\left(\frac{\partial \ln p_{x_1|x_2}(x_1)}{\partial x_1}\right)^2 \middle| x_2\right)$$

and since the law of x_1 conditioned on x_2 is $N(\rho x_2, 1 - \rho^2)$, it follows that $J^{(2)}(\mathbf{x})_{11} = 1/(1 - \rho^2)$ and thus $B_4 = 1 - \rho^2$.

EXAMPLE 3.2. For every $\beta > 2$, let $\mathbf{x} = (x_1, x_2)$ be distributed with density $p_\beta(x_1, x_2) = ((\beta - 1)/\pi)(1 + x_1^2 + x_2^2)^{-\beta}$ and choose again $\alpha = \{1\}$. For this example $\varepsilon^2 =: \mathrm{var}(x_1) = 1/(2\beta - 4)$, while

$$B_1 = \frac{\beta + 1/2}{(2\beta - 1)(\beta - 1)},$$

$$B_2 = B_3 = \frac{\beta + 1}{2\beta(\beta - 1)},$$

$$B_4 = \frac{(\beta + 1)(2\beta - 3)}{2\beta(\beta - 2)(2\beta - 1)}.$$

[These expressions are calculated making use of the integrals ($a > 0$)

$$\int_{-\infty}^{\infty} \frac{dt}{(a^2 + t^2)^\beta} = \frac{\sqrt{\pi}}{a^{2\beta-1}} \frac{\Gamma(\beta - 1/2)}{\Gamma(\beta)},$$

$$\int_{-\infty}^{\infty} \frac{t^2\, dt}{(a^2 + t^2)^\beta} = \frac{\sqrt{\pi}}{(2\beta - 3)a^{2\beta-3}} \frac{\Gamma(\beta - 1/2)}{\Gamma(\beta)}\Big].$$

It can be immediately verified that in this particular case, $\varepsilon^2 > B_4 > B_1 > B_2 = B_3$ for all $\beta > 2$.

4. Generalized bounds. Consider the scalar case $n = 1$. The Cauchy–Schwarz inequality in $L^2(\Omega)$ states that

$$\| u \| \geq \frac{|(u, v)|}{\| v \|}. \tag{23}$$

The right-hand side is the length of u's projection along v which must, of course, be not larger than u's length. The global Cramér–Rao inequality considers $u = x - E(x|\mathscr{Y})$ and $v = a(x|\mathscr{Y}) = (p'_{x|\mathscr{Y}}(x))/(p_{x|\mathscr{Y}}(x))$, its simplicity following from the fact that in this case $|(u, v)| = 1$. However, tighter bounds could be achieved by selecting v from an appropriate family of random variables. One such family of generalized bounds was considered in [13].

Consider, for example, the collection of $\mathscr{Y}$-measurable random variables $q(\xi) = q(\xi, \omega)$ depending smoothly on a real parameter ξ and define $a_q = [q(x)p_{x|\mathscr{Y}}(x)]'/(p_{x|\mathscr{Y}}(x))$. Assuming $q(x) \in L^1(\Omega)$ and $a_q \in L^2(\Omega)$, (23) yields

$$E(x - E(x))^2 \geq (Eq(x))^2 \Bigg/ \left(E\left\{ \frac{[q(x)p_{x|\mathscr{Y}}(x)]'}{p_{x|\mathscr{Y}}(x)} \right\}^2 \right). \tag{24}$$

[Note that for $q \equiv 1$, (24) reduces to the classical bound.]

Moreover, choosing $q^*(\xi, \omega) = (1/p_{x|\mathscr{Y}}(\xi))\int_\xi^\infty(\tau - E(\mathbf{x}|\mathscr{Y}))p_{x|\mathscr{Y}}(\tau)\, d\tau$, equality is achieved in (24). The details can be easily verified (the more general multidimensional case will be dealt with later). We remark that the family of bounds expressed by (24) includes other generalized bounds such as $E(x - E(\mathbf{x}|\mathscr{Y}))^2 \geq [E(h^{-1})'(z)]^2/(J(z|\mathscr{Y}))$, where $z = h(x)$ is an invertible transformation of x with inverse $x = h^{-1}(z)$. [Set $q(\xi) = 1/(h'(\xi))$, which is deterministic whenever h is.] Incidentally, for this one-dimensional case, (24) appears in [5, Equation (21)] with nonrandom q's.

We shall now obtain the multidimensional extension of (24). The next lemma is an extension of the Cauchy–Schwarz inequality to Grammians.

LEMMA 4. *Let L be an inner product space and* $\mathbf{u} = [u_1, \dots, u_n] \in L^n$ *and* $\mathbf{v} = [v_1, \dots, v_m] \in L^m$. *Assume* $((\mathbf{v}))$ *is nonsingular* [*which will be the case if, e.g.*, $\operatorname{rank}((\mathbf{u}, \mathbf{v})) = m$]. *Then,*

$$((\mathbf{u})) \geq ((\mathbf{u}, \mathbf{v}))((\mathbf{v}))^{-1}((\mathbf{v}, \mathbf{u})). \tag{25}$$

PROOF. Consider the appended vector $\mathbf{w} = [u_1, \dots, u_n, v_1, \dots, v_m] \in L^{n+m}$. We have

$$\mathbf{0} \leq ((\mathbf{w})) = \left[\begin{array}{c|c} ((\mathbf{u})) & ((\mathbf{u}, \mathbf{v})) \\ ((\mathbf{v}, \mathbf{u})) & ((\mathbf{v})) \end{array}\right],$$

where $\mathbf{0}$ is the $(n + m) \times (n + m)$ zero matrix. Also let $\mathbf{I}_n$ be the $n \times n$ identity matrix and let $\mathbf{D} \in R_{n \times (n+m)}$ be defined by

$$\mathbf{D} = \left[\mathbf{I}_n \middle| -((\mathbf{u}, \mathbf{v}))((\mathbf{v}))^{-1}\right].$$

Then we also have $\mathbf{D}((\mathbf{w}))\mathbf{D}^T \geq 0$ and (25) will follow by observing that

$$\mathbf{D}((\mathbf{w}))\mathbf{D}^T = ((\mathbf{u})) - ((\mathbf{u}, \mathbf{v}))((\mathbf{v}))^{-1}((\mathbf{v}, \mathbf{u})).$$ □

LEMMA 5. *Equality occurs in* (25) *in the ith diagonal element iff there exists a* $\gamma \in \mathbb{R}^m$ *such that* $u_i = \gamma^T \mathbf{v}$. *In this case,*

$$\gamma = ((u_i, \mathbf{v}))((\mathbf{v}))^{-1} \tag{26}$$

REMARK. In a nonnegative matrix, whenever a diagonal element is zero, so are the corresponding row and column. This justifies considering only the diagonal in Lemma 5.

PROOF. Define the multiindex $\alpha = [i, n + 1, \dots, n + m]$. Then, with the notation of Lemma 4,

$$((\omega_\alpha)) = ((\omega))_\alpha = \left[\begin{array}{c|c} ((\mathbf{u}))_{ii} & ((u_i, \mathbf{v})) \\ ((\mathbf{v}, u_i)) & ((\mathbf{v})) \end{array}\right].$$

Using the block expansion of a determinant leads to

$$\det\{((\omega_\alpha))\} = \det\{((\mathbf{v}))\}\left[((\mathbf{u})) - ((\mathbf{u}, \mathbf{v}))((\mathbf{v}))^{-1}((\mathbf{v}, \mathbf{u}))\right]_{i,i}.$$

It follows that equality occurs in (25) in the ith element of the diagonal iff the components of ω_α are linearly dependent. Since $((\mathbf{v}))$ is nonsingular, this implies that $u_i = \sum_{j=1}^m \gamma_j v_j$ as claimed; (26) follows by direct substitution. □

As a first application, we state the following

PROPOSITION 2. *Let* $\mathbf{q}(\xi) = \mathbf{q}(\xi, \omega) = (q_i(\xi, \omega))_{i=1}^n$ *be a collection of* n-*dimensional* $\mathscr{Y}$-*measurable random vectors depending on the parameter* $\xi \in \mathbb{R}^n$ *and satisfying for each* i:

(a) $q_i(\cdot, \omega) p_{\mathbf{x}|\mathscr{Y}}(\cdot) \in C^1(\mathbb{R}^n)$ *a.s.*
(b) $E|q_i(\mathbf{x}, \omega)| < \infty$.
(c) $Eq_i(\mathbf{x}, \omega) \neq 0$.

Furthermore, defining $\mathbf{a}_q(\mathbf{x}|\mathscr{Y}) = ((D_i[q_i(\mathbf{x}, \omega)p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x})])/(p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x})))_{i=1}^n$, *assume the components of* $\mathbf{a}_\mathbf{q}(\mathbf{x}|\mathscr{Y})$ *have finite second moment, and as a result, denote* $\mathbf{J}_\mathbf{q}(\mathbf{x}|\mathscr{Y}) = ((\mathbf{a}_\mathbf{q}(\mathbf{x}|\mathscr{Y})))$. *Then*

$$\Sigma(\mathbf{x}|\mathscr{Y}) \geq \operatorname{diag}[E\mathbf{q}(\mathbf{x}, \omega)]\mathbf{J}_\mathbf{q}^{-1}(\mathbf{x}|\mathscr{Y})\operatorname{diag}[E\mathbf{q}(\mathbf{x}, \omega)]. \tag{27}$$

PROOF. Inequality (27) will follow from Lemma 4 by choosing $\mathbf{u} = \mathbf{x} - E(\mathbf{x}|\mathscr{Y})$ and $\mathbf{v} = \mathbf{a}_\mathbf{q}(\mathbf{x}|\mathscr{Y})$. We only need to verify that

$$((\mathbf{u}, \mathbf{v}))_{ij} = E\big[(x_i - E(x_i|\mathscr{Y}))(\mathbf{a}_\mathbf{q}(\mathbf{x}|\mathscr{Y}))_j\big] = -(Eq_i(\mathbf{x}, \omega))\,\delta_{ij}. \tag{28}$$

Indeed,

$$E\big[(x_i - E(x_i|\mathscr{Y}))(\mathbf{a}_\mathbf{q}(\mathbf{x}|\mathscr{Y}))_j\big] = E\int_{\mathbb{R}^n}(\xi_i - E(x_i|\mathscr{Y}))D_j\big[q_j(\xi, \omega)p_{\mathbf{x}|\mathscr{Y}}(\xi)\big]\,d\xi.$$

If $i = j$,

$$\begin{aligned}((\mathbf{u}, \mathbf{v}))_{ii} &= E\int_{\mathbb{R}^{n-1}} d\xi_{i^*}\int_{-\infty}^{\infty}(\xi_i - E(x_i|\mathscr{Y}))D_i\big[q_i(\xi, \omega)p_{\mathbf{x}|\mathscr{Y}}(\xi)\big]\,d\xi_i \\ &= -E\int_{\mathbb{R}^{n-1}} d\xi_{i^*}\int_{-\infty}^{\infty} q_i(\xi, \omega)p_{\mathbf{x}|\mathscr{Y}}(\xi)\,d\xi_i = Eq_i(\mathbf{x}, \omega).\end{aligned}$$

The justification for dropping the boundary terms in the integration by parts is just the same as in Lemma 1. Here condition (b) is used.

If $i \neq j$,

$$((\mathbf{u}, \mathbf{v}))_{ij} = E\int_{\mathbb{R}^{n-1}}(\xi_i - E(x_i|\mathscr{Y}))\,d\xi_{j^*}\int_{-\infty}^{\infty} D_j\big[q_j(\xi, \omega)p_{\mathbf{x}|\mathscr{Y}}(\xi)\big]\,d\xi_j.$$

Again by similar arguments, the inner integral is zero for (Lebesgue) almost every ξ_{j^*} in $\mathbb{R}^{n-1}$; thus $((\mathbf{u}, \mathbf{v}))_{ij} = 0$, completing the proof. Note that in view of (28), condition (c) guarantees that $\operatorname{rank}((\mathbf{u}, \mathbf{v})) = n$; thus $((\mathbf{v}))$ is nonsingular. □

REMARK. If $\mathscr{Y}$ is the σ-algebra generated by a random vector $\mathbf{y} = (y_1, \dots, y_r)$ such that there exists a differentiable joint density $p_{\mathbf{x},\mathbf{y}}(\xi_1, \dots, \xi_n, \eta_1, \dots, \eta_r)$, then $p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x})$ may be replaced everywhere by $p_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \mathbf{y})$ so that (27) becomes

$$\begin{aligned}&\Sigma(\mathbf{x}, \mathbf{y}) \\ &\quad \geq \operatorname{diag}[E\mathbf{q}(\mathbf{x}, \mathbf{y})]\left\{\left(E\frac{D_i(q_i p_{\mathbf{x},\mathbf{y}})(\mathbf{x}, \mathbf{y})D_j(q_j p_{\mathbf{x},\mathbf{y}})(\mathbf{x}, \mathbf{y})}{p_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \mathbf{y})^2}\right)_{i,j=1}^n\right\}^{-1} \\ &\qquad \times \operatorname{diag}[E\mathbf{q}(\mathbf{x}, \mathbf{y})]\end{aligned} \tag{29}$$

(here D_k still denotes the derivative along x_k).

PROPOSITION 3. *Equality occurs in* (27) *if and only if for each* i, $1 \le i \le n$, *there exists a constant* $C_i \ne 0$ *such that for all* $\xi \in \mathbb{R}^n$ *satisfying* $p_{\mathbf{x}|\mathscr{Y}}(\xi) \ne 0$,

$$(30)\quad q_i(\xi) = \frac{C_i}{p_{\mathbf{x}|\mathscr{Y}}(\xi)} \int_{\xi_i}^{\infty} \left(\Sigma^{-1}(\mathbf{x}|\mathscr{Y})(\mathbf{t} - E(\mathbf{x}|\mathscr{Y}))\right)_i p_{\mathbf{x}|\mathscr{Y}}(\mathbf{t})\, dt_i \Big|_{\mathbf{t}_{i^*} = \xi_{i^*}} \quad a.s.,$$

or equivalently

$$(30')\quad q_i(\mathbf{x}) = \frac{C_i}{p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x})} \int_{x_i}^{\infty} \left(\Sigma^{-1}(\mathbf{x}|\mathscr{Y})(\mathbf{t} - E(\mathbf{x}|\mathscr{Y}))\right)_i p_{\mathbf{x}|\mathscr{Y}}(\mathbf{t})\, dt_i \Big|_{\mathbf{t}_{i^*} = \mathbf{x}_{i^*}} \quad a.s.$$

PROOF. This result is a direct application of Lemma 5. Note that since we are considering equality of matrices in (27) (and not of the individual diagonal entries), we may exchange the roles of $\mathbf{u}$ and $\mathbf{v}$ and conclude from Lemma 5 that equality occurs in (27) if and only if $\mathbf{a}_{\mathbf{q}}(\mathbf{x}|\mathscr{Y}) = \Gamma(\mathbf{x} - E(\mathbf{x}|\mathscr{Y}))$, while from (26) and (28) $\Gamma = -\operatorname{diag}(E\mathbf{q}(\mathbf{x}))\Sigma^{-1}(\mathbf{x}|\mathscr{Y})$. Thus, almost surely for each i and (Lebesgue) almost every ξ,

$$(31)\qquad D_i\left[\tilde{q}_i(\xi) p_{\mathbf{x}|\mathscr{Y}}(\xi)\right] = -p_{\mathbf{x}|\mathscr{Y}}(\xi)\left(\Sigma^{-1}(\mathbf{x}|\mathscr{Y})(\xi - E(\mathbf{x}|\mathscr{Y}))\right)_i,$$

where $\tilde{q}_i(\xi) = q_i(\xi)/Eq_i(\mathbf{x})$. Integrating (31) we obtain (30). Note that the constraint $E\tilde{q}_i(\mathbf{x}) = 1$ forces the integration constant to be nonzero. □

REMARK. In the scalar case ($n = 1$), the optimal function $q(\xi)$ given by (30) is sign preserving. To see this, note that the function $\int_{\xi}^{\infty}(t - E(x|\mathscr{Y}))p_{x|y}(t)\,dt$ tends to zero at $\pm\infty$ and its derivative changes sign exactly once. This means that in the scalar case, the bound (27) can be restricted to nonnegative functions q. When $n > 1$, it can be easily seen that this property no longer holds.

By denoting in Proposition 2

$$h_i(\xi) = \frac{q_i(\xi) p_{\mathbf{x}|\mathscr{Y}}(\xi)}{E(q_i(\mathbf{x}))}, \qquad i = 1, \dots, n,$$

one obtains a reformulation of Propositions 2 and 3 which is summarized in

PROPOSITION 4. *Let* $\mathbf{h}(\xi) = \mathbf{h}(\xi, \omega) = (h_i(\xi, \omega))_{i=1}^n$ *be a collection of* n-*dimensional* $\mathscr{Y}$-*measurable random vectors depending on the parameter* $\xi \in \mathbb{R}^n$ *and satisfying for each* i:

(a) $h_i(\cdot, \omega) \in C^1(\mathbb{R}^n)$ *a.s.*
(b) $h_i(\xi, \omega) \in L^1(\mathbb{R}^n \times \Omega)$ (*that is*, $E\int_{\mathbb{R}^n}|h_i(\xi, \omega)|\, d\xi < \infty$).
(c) $E\int_{\mathbb{R}^n} h_i(\xi, \omega)\, d\xi = 1$.

Then

$$(32)\qquad \Sigma(\mathbf{x}|\mathscr{Y}) \ge \left[\left(E\frac{D_i(h_i(\mathbf{x}, \omega))D_j(h_j(\mathbf{x}, \omega))}{p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x})^2}\right)_{i,j=1}^n\right]^{-1}.$$

Furthermore, equality is achieved in (32) *if and only if*

$$(33)\qquad h_i(\xi, \omega) = \int_{\xi_i}^{\infty} \left(\Sigma^{-1}(\mathbf{x}|\mathscr{Y})(\mathbf{t} - E(\mathbf{x}|\mathscr{Y}))\right)_i p_{\mathbf{x}|\mathscr{Y}}(\mathbf{t})\, dt_i \Big|_{\mathbf{t}_{i^*} = \xi_{i^*}} \qquad a.s.$$

We conclude this section with two simple examples in which the classical global Cramér–Rao bound ($q \equiv 1$ in Proposition 2) is trivial but by choosing q in an appropriate manner, useful bounds can be obtained. Both examples result from the following situation:

Let x be a random variable with density $p_x(\xi)$, r an independent Gaussian noise with mean 0 and variance ρ^2 and $g(\cdot)$ a continuous and almost everywhere differentiable function; x is to be estimated based on the observation $y = g(x) + r$. The joint density is

$$(34) \qquad p_{x,y}(\xi,\eta) = (2\pi\rho^2)^{-1/2} p_x(\xi) \exp\left[-(\eta - g(\xi))^2/2\rho^2\right].$$

Consider the bound (27) for $\varepsilon^2 = E(x - E(x|y))^2$ and for the sake of simplicity assume the function q is independent of the measurement y, that is, $q(x,y) = q(x)$. Then by inserting (34) in (29), a simple calculation yields

$$(35) \qquad \varepsilon^2 \ge (Eq(x))^2 \Bigg/ \left(E\left[\left(q'(x) + q(x)\frac{p_x'(x)}{p_x(x)} \right)^2 + \frac{1}{\rho^2} q^2(x) g'(x)^2 \right] \right).$$

EXAMPLE 4.1. Here x is an $N(0,1)$ random variable and $y = \operatorname{sgn}(x)$. It can be verified directly that $E(x|y) = (2/\pi)^{1/2} y$ and that $\varepsilon^2 = (1 - 2/\pi) \sim 0.3634$.

To be able to apply the bound (35), one must regularize the problem. Let z be another (independent) $N(0,1)$ variable and for each $\rho > 0$ and $\lambda > 0$ set $y = g_\lambda(x) + \rho z$, where $g_\lambda(\xi) = \xi/\lambda$ if $|\xi| \le \lambda$ and $g_\lambda(\xi) = \operatorname{sgn}(\xi)$ otherwise. Denoting by $\varepsilon^2_{\lambda,\rho}$ the mean square error of the perturbated problem, clearly $\varepsilon^2 = \lim_{\rho\to 0}\lim_{\lambda\to 0}\varepsilon^2_{\lambda,\rho}$. In particular, by (35),

$$(36) \qquad \varepsilon^2 \ge \frac{(Eq(x))^2}{E(q'(x) - xq(x))^2 + \lim_{\rho\to 0}(1/\rho^2)\lim_{\lambda\to 0} E\left(q^2(x) g_\lambda'(x)^2\right)}.$$

The classical bound (1) is obtained by choosing $q(\xi) \equiv 1$ in (36). Since $\lim_{\lambda\to 0} E(g_\lambda'(x)^2) = \infty$, this choice yields a trivial lower bound for ε^2. However, if $\lim_{\lambda\to 0} E(q^2(x) g_\lambda'(x)^2) = 0$, a significant (positive) bound is obtained. This will be the case, for example, if $q(\xi) = 0$ in a neighborhood of the origin and also for $q_\alpha(x) = x^2 e^{-(\alpha/2)x^2}$, $\alpha > -\frac{1}{2}$. This gives a family of lower bounds,

$$\varepsilon^2 \ge C_\alpha \equiv \frac{(2\alpha + 1)^{7/2}}{(\alpha + 1)^3 (7\alpha^2 + 10\alpha + 7)}, \qquad \alpha > -\frac{1}{2}.$$

The bound is approximately optimal when $\alpha_{\max} = 0.7$ and $C_{\max} \sim 0.25$. Note that $C_{\max}/\varepsilon^2 \sim 0.69$.

EXAMPLE 4.2. In this example, x is distributed uniformly on $[0,1]$ and $y = x + r$, where $r \sim N(0,1)$. Denoting by ϕ (resp. Φ) the standard Gaussian density function (resp. distribution function), we have that

$$E(x|y) = y + (\phi(y) - \phi(y-1))/(\Phi(y) - \Phi(y-1))$$

and the mean square error is given by $\varepsilon^2 = 1 - \sqrt{2\pi I}$, where

$$I = \int_{-\infty}^{\infty} (\phi(y) - \phi(y-1))^2 / (\Phi(y) - \Phi(y-1))\, dy.$$

Performing numerical integration one obtains $\varepsilon^2 \sim 0.07692$. To be able to apply the classical bound (1) ($q \equiv 1$), the problem must again be regularized, and we do so by choosing for any $k \in N$ the a priori density

$$p_k(\xi) = \begin{cases} \dfrac{k}{k+1}, & \text{if } \xi \in [0,1], \\ \dfrac{k}{k+1}(1 + \cos 2\pi k \xi), & \text{if } \xi \in \left[\dfrac{-1}{2k}, 0\right] \cup \left[1, 1 + \dfrac{1}{2k}\right], \\ 0, & \text{otherwise.} \end{cases}$$

Denoting by ε_k^2 the mean square error for this problem, clearly $\varepsilon^2 = \lim_{k\to\infty} \varepsilon_k^2$. But the lower bound (1) for ε_k^2 is $1/(1 + 2\pi^2[k^2/(k+1)])$, which becomes trivial when $k \to \infty$.

However, choosing an appropriate continuous function $q(\xi)$ whose support lies in $[0,1]$, we obtain from (35),

$$\varepsilon^2 \geq B_q \equiv \frac{\left(\int_0^1 q(\xi)\, d\xi\right)^2}{\int_0^1 \left(q^2(\xi) + q'^2(\xi)\right) d\xi}. \tag{37}$$

Defining

$$q_1(\xi) = \begin{cases} \xi - \xi^2, & \text{if } \xi \in [0,1], \\ 0, & \text{otherwise,} \end{cases}$$

a very tight bound is obtained, namely $B_{q_1} = 5/66 \sim 0.07576$. Actually the bound in (37) can be optimized for q; the goal is to find a function $q_0(\xi)$ in $C^1[0,1]$ which achieves

$$\min_q \int_0^1 \left(q^2(\xi) + q'^2(\xi)\right) d\xi$$

subject to the constraint $\int_0^1 q(\xi)\, d\xi = 1$ and satisfying the boundary conditions $q(0) = q(1) = 0$. This is a classical isoperimetric problem solvable by use of the method of Lagrange multipliers [1]. In this case the solution is

$$q_0(\xi) = \frac{e^{\xi} + e^{1-\xi} - (e+1)}{(e-3)},$$

yielding the bound $B_{q_1} \sim B_{q_0} = (3-e)/(e+1) \sim 0.07577 \leq \varepsilon^2 = 0.07692$.

5. Some further generalizations. The global lower bound (1) for the estimation error and its matrix counterparts in the vector case are particular applications of the abstract inner product space inequality (25), as well as the generalized bound of Proposition 2 (or equivalently Proposition 4).

In this section, a further use of (25) will yield a general family of bounds including the previous bounds, as well a bounds which are of the Barankin [2] and Bhattacharya [3] type (cf. also [12]). As a motivation, consider replacing the global Cramér–Rao inequality (1) by

$$E(x - E(x|y))^2 \geq 1 \Big/ \left(E\left[\frac{p(x+\delta, y) - p(x,y)}{\delta p(x,y)} \right]^2 \right) \tag{38}$$
$$= 1 \Big/ \left(\frac{1}{\delta^2} E\left[1 - \frac{p(x+\delta, y)}{p(x,y)} \right]^2 \right),$$

where the density's derivative has been replaced by a finite difference. Inequality (38), while easily verifiable, offers some additional advantages:

(i) No smoothness conditions are required for the density. Moreover, viewing $p(x+\delta, y)/p(x,y)$ as a Radon–Nikodym derivative, the density need not even exist (cf. [4], [6] for an application of this approach to the nonlinear filtering problem).

(ii) Even when a smooth density $p(x,y)$ does exist, it is easy to construct examples for which choosing δ appropriately, (38) provides a much tighter bound than (1).

This approach can be pushed further by considering weighted higher order finite differences of $p(\cdot, y)$ $[(\Sigma w_i p(x + \delta_i, y)$ with $\Sigma w_i = 0]$ or, ultimately, by convolving the density with an appropriate signed Borel measure (cf. [7] for a similar approach in the nonrandom parameter setup). Much of what follows simply formalizes the previous remarks.

In accordance with the preceding motivation, it is assumed from now on that the conditional density $p_{\mathbf{x}|\mathscr{Y}}(\cdot)$ is piecewise continuous with nonrandom points of discontinuity. Let μ be a finite signed Borel measure on $\mathbb{R}^n$ with compact support and let f be a function on $\mathbb{R}^n$ which is integrable on compacts. Consider the convolution $\mu * f$ which is a function on $\mathbb{R}^n$ defined by

$$\mu * f(\xi) = \int_{\mathbb{R}^n} f(\xi - \eta)\mu(d\eta). \tag{39}$$

Let α be a multiindex, i.e., an ordered subsequence of $\{1, 2, \dots, n\}$, of length $m \leq n$ and A a subset of $\mathbb{R}^n$. The projection A_α of A in the direction of α is the subset of $\mathbb{R}^n$: $A_\alpha = \{\xi_\alpha | \xi \in A\}$. The measure μ is said to be α-cylindrical if there exists a Borel measure $\tilde{\mu}$ on $\mathbb{R}^m$ such that $\mu(A) = \tilde{\mu}(A_\alpha)$ for each Borel set $A \subset \mathbb{R}^n$. If μ is α-cylindrical and denoting $f(\xi) = f(\xi_\alpha, \xi_{\alpha^*})$, we then have

$$\mu * f(\xi) = \int_{\mathbb{R}^m} f(\xi_\alpha - \eta_\alpha, \xi_{\alpha^*})\tilde{\mu}(d\eta_\alpha). \tag{40}$$

Note that μ is α-cylindrical as before if and only if $\mu = \tilde{\mu} \times \prod_{i \in \alpha^*} \delta_0^i$ (where δ_0^i is the Dirac δ function in the i-direction with $\tilde{\mu}$ acting on the components of α).

Now, for $i = 1, \dots, n$, let μ_i be an i-cylindrical signed measure on $\mathbb{R}^n$ with compact support, i.e., $\mu_i = \delta_0^1 \times \delta_0^2 \times \cdots \times \tilde{\mu}_i \times \cdots \times \delta_0^n$. For each i assume

$$\tilde{\mu}_i(\mathbb{R}) = 0, \tag{41a}$$

$$\lambda_i \equiv \int_{\mathbb{R}} \xi \tilde{\mu}_i(d\xi) \neq 0, \tag{41b}$$

$$E\left[\frac{(\mu_i * p_{\mathbf{x}|\mathscr{Y}})(\mathbf{x})}{p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x})}\right] < \infty. \tag{41c}$$

Consequently, denoting $\boldsymbol{\mu} = (\mu_i)_{i=1}^n$, define the $n \times n$ matrix $J_{\boldsymbol{\mu}}(\mathbf{x}|\mathscr{Y})$ by

$$(\mathbf{J}_{\boldsymbol{\mu}}(\mathbf{x}|\mathscr{Y}))_{i,j} = E\left[\frac{(\mu_i * p_{\mathbf{x}|\mathscr{Y}})(\mathbf{x})(\mu_j * p_{\mathbf{x}|\mathscr{Y}})(\mathbf{x})}{(p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x}))^2}\right], \qquad i, j = 1, \dots, n. \tag{42}$$

PROPOSITION 5. *Under the assumptions* (41), $\mathbf{J}_{\boldsymbol{\mu}}(\mathbf{x}|\mathscr{Y})$ *is a well-defined nonsingular matrix and denoting* $\boldsymbol{\lambda} = (\lambda_i)_{i=1}^n$,

$$\Sigma(\mathbf{x}|\mathscr{Y}) \geq \operatorname{diag}(\boldsymbol{\lambda})(\mathbf{J}_{\boldsymbol{\mu}}(\mathbf{x}|\mathscr{Y}))^{-1}\operatorname{diag}(\boldsymbol{\lambda}). \tag{43}$$

REMARK. Note that this proposition is formulated in terms of (41) and (42) and, therefore, could be based exclusively on Radon–Nikodym derivatives of measures without even having to assume the existence of a conditional density for $\mathbf{x}$ conditioned on $\mathscr{Y}$.

PROOF. In Lemma 4 substitute $\mathbf{u} = \mathbf{x} - E(\mathbf{x}|\mathscr{Y})$ and

$$\mathbf{v} = ((\boldsymbol{\mu} * p_{\mathbf{x}|\mathscr{Y}})(\mathbf{x}))/(p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x})).$$

To prove (43), it then suffices to verify that

$$((\mathbf{u}, \mathbf{v}))_{i,j} = E\left[(x_i - E(x_i|\mathscr{Y}))\left(\frac{(\mu_j * p_{\mathbf{x}|\mathscr{Y}})(\mathbf{x})}{p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x})}\right)\right] = \lambda_i \delta_{ij}.$$

Indeed [cf. (40)]

$$\begin{aligned} ((\mathbf{u}, \mathbf{v}))_{i,j} &= \int_{\mathbb{R}^n} d\xi \Big((\xi_i - E(x_i|\mathscr{Y})) \\ &\qquad \times \int_{-\infty}^{\infty} p_{\mathbf{x}|\mathscr{Y}}(\xi_1, \xi_2, \dots, \xi_j - \eta_j, \dots, \xi_n)\tilde{\mu}_j(d\eta_j)\Big) \\ &= \int_{-\infty}^{\infty} \tilde{\mu}_j(d\eta_j)\Big(\int_{\mathbb{R}^n} (\xi_i - E(x_i|\mathscr{Y})) \\ &\qquad \times p_{\mathbf{x}|\mathscr{Y}}(\xi_1, \xi_2, \dots, \xi_j - \eta_j, \dots, \xi_n)\, d\xi\Big) \end{aligned} \tag{44}$$

[assumption (41c) allows the change in order of integration]. If $i \neq j$, the inner integral is actually independent of η_j and thus $((\mathbf{u}, \mathbf{v}))_{i,j} = 0$ follows from (41a).

If $i = j$,

$$((\mathbf{u},\mathbf{v}))_{i,i} = \int_{-\infty}^{\infty} \tilde{\mu}_i(d\eta_i) \int_{\mathbb{R}^n} \left(\xi_i' + \eta_i - E(x_i|\mathscr{Y}) p_{\mathbf{x}|\mathscr{Y}}(\xi')\, d\xi'\right)$$
$$= \int_{-\infty}^{\infty} \eta_i \tilde{\mu}_i(d\eta_i) = \lambda_i. \qquad \square$$

A particular choice of the measures μ_i is obtained by setting

$$\tilde{\mu}_i = (1/\Delta_i)(\delta_{\Delta_i} - \delta_0),$$

where for each i, $\Delta_i \neq 0$, and δ_{Δ_i} is the point mass measure at Δ_i. In this case $\lambda_i = 1$. Moreover, if the conditional density is a.s. differentiable and $\Delta_i \to 0$, $\mathbf{J}_{\mu}$ tends to the standard Fisher information matrix so that (43) becomes the classical global Cramér–Rao inequality. This generalization was pointed out in [7]. Actually, the global Cramér–Rao inequality could be obtained as a particular case of (43) (rather than by a limiting procedure) as long as the measures μ_i are replaced by Schwarz distributions T_i; the definitions and requirements given in (39), (41) and (42) have their immediate generalizations to distributions [9] and Proposition 5 remains true. The global Cramér–Rao inequality then results by choosing $\tilde{T}_i$ to be the weak derivative of the Dirac measure at 0. We will not pursue this approach here.

Two unrelated generalizations of (2) were presented in Proposition 2 (or equivalently Proposition 4) and Proposition 5. For completeness sake, both results will be combined in

PROPOSITION 6. *Let* $\mathbf{h}(\xi) = (h_i(\xi,\omega))_{i=1}^n$ *be as in Proposition* 4 *and let* $\mu = (\mu_i)_{i=1}^n$ *and* $\lambda = (\lambda_i)_{i=1}^n$ *be as in Proposition* 5. *Define the* $n \times n$ *matrix* $\mathbf{J}_{\mu,\mathbf{h}}(\mathbf{x}|\mathscr{Y})$ *by*

$$(45) \qquad \left(\mathbf{J}_{\mu,\mathbf{h}}(\mathbf{x}|\mathscr{Y})\right)_{i,j} = E\left[\frac{(\mu_i * h_i)(\mathbf{x})(\mu_j * h_j)(\mathbf{x})}{\left(p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x})\right)^2}\right], \qquad i,j = 1,\dots,n.$$

Then $\mathbf{J}_{\mu,\mathbf{h}}(\mathbf{x}|\mathscr{Y})$ *is nonsingular and*

$$(46) \qquad \Sigma(\mathbf{x}|\mathscr{Y}) \geq \operatorname{diag}(\lambda)\mathbf{J}_{\mu,\mathbf{h}}^{-1}(\mathbf{x}|\mathscr{Y})\operatorname{diag}(\lambda).$$

The proof follows directly just as the proofs of Propositions 4 and 5 and is, therefore, omitted.

The last result will concern a generalization of the global Cramér–Rao inequality equivalent to the Bhattacharya bound [3] in the nonrandom parameter estimation setup.

Let $k \in \mathbb{N}$ and $\mathbf{M} = (\mu_{s,t})$, $s = 1,\dots,n$ and $t = 1,\dots,k$, be an $n \times k$ matrix of finite signed Borel measures on $\mathbb{R}^n$ with compact support such that for each $1 \leq s \leq n$ and $1 \leq t \leq k$, $\mu_{s,t}$ is s-cylindrical, that is $\mu_{s,t} = \delta_0 \times \delta_0 \times \cdots \times \tilde{\mu}_{s,t} \times \delta_0 \times \cdots \times \delta_0$ for some signed Borel measure $\tilde{\mu}_{s,t}$ on $\mathbb{R}$. Furthermore, assume

(47a) $$\tilde{\mu}_{s,t}(\mathbb{R}) = 0, \qquad 1 \le s \le n, 1 \le t \le k,$$

(47b) $$\int_{-\infty}^{\infty} \xi \tilde{\mu}_{s,t}(d\xi) = 0, \qquad 1 \le s \le n, 2 \le t \le k,$$

(47c) $$E\left[\frac{(\mu_{s,t} * h_s)(\mathbf{x})}{p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x})}\right]^2 < \infty, \qquad 1 \le s \le n, 1 \le t \le k.$$

Let also $\mathbf{h}(\mathbf{x}) = (h_i(\mathbf{x}, \omega))_{i=1}^n$ be as in Proposition 4 and define the $nk \times nk$ matrix $\mathbf{J}_{\mathbf{M},\mathbf{h}}(\mathbf{x}|\mathscr{Y})$ by

(48) $$(\mathbf{J}_{\mathbf{M},\mathbf{h}}(\mathbf{x}|\mathscr{Y}))_{(l-1)n+i,(m-1)n+j} = E\left[\frac{(\mu_{i,l} * h_i)(\mathbf{x})(\mu_{j,m} * h_j)(\mathbf{x})}{(p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x}))^2}\right],$$

for $i, j = 1, \dots, n$ and $l, m = 1, \dots, k$.

Proposition 7. *Under the assumption* (47) *and denoting* $\boldsymbol{\lambda} = (\int_{-\infty}^{\infty} \xi \tilde{\mu}_{s,1}(d\xi))_{s=1}^n$, *if* $\mathbf{J}_{\mathbf{M},\mathbf{h}}(\mathbf{x}|\mathscr{Y})$ *is nonsingular,*

(49) $$\Sigma(\mathbf{x}|\mathscr{Y}) \ge \operatorname{diag}(\boldsymbol{\lambda})(\mathbf{J}_{\mathbf{M},\mathbf{h}}^{-1}(\mathbf{x}|\mathscr{Y}))_{[n],[n]} \operatorname{diag}(\boldsymbol{\lambda}).$$

(*Recall that by the notation introduced in Section* 2, $\mathbf{A}_{[n],[n]}$ *is the upper left* $n \times n$ *submatrix of a matrix* $\mathbf{A}$.)

Proof. Inequality (49) follows from Lemma 4 by choosing $\mathbf{u} = \mathbf{x} - E(\mathbf{x}|\mathscr{Y})$ and $v_{(l-1)n+i} = (\mu_{i,l} * h_i(\mathbf{x}))/(p_{\mathbf{x}|\mathscr{Y}}(\mathbf{x}))$, $i = 1, \dots, n$ and $l = 1, \dots, k$. To apply (25) it only remains to verify that

(50) $$((u, v)) = \left[\begin{array}{ccc|c|c|c} \lambda_1 & & 0 & & & \\ & \ddots & & 0 & \cdots & 0 \\ 0 & & \lambda_s & & & \end{array}\right].$$

The calculations leading up to (50) are essentially the same as those appearing for example in the proof of Proposition 5 and will therefore be omitted here. □

Remark. Proposition 7 (and also Proposition 6 which is a particular case when $k = 1$) can be again generalized further by considering appropriate Schwarz distributions $T_{s,t}$ instead of the measures $\mu_{s,t}$. Note that while this extends the class of measures from which to choose $\mathbf{M} = (\mu_{s,t})$, condition (47c), must now be interpreted as implying that $T_{s,t} * h_s$ is indeed a function which in certain cases might restrict the class of admissible $\mathbf{h}$'s.

REFERENCES

[1] Akhiezer, N. (1962). *The Calculus of Variations*. Blaisdell, Waltham, Mass.
[2] Barankin, E. (1949). Locally best unbiased estimates. *Ann. Math. Statist.* **20** 477–501.
[3] Bhattacharya, A. (1946, 1947, 1948). On some analogues of the amount of information and their use in statistical estimation. *Sankhyā* **8** 1–14, 201–218, 315–328.

[4] BOBROVSKY, B. Z. and ZAKAI, M. (1976). A lower bound on the estimation error for certain diffusion processes. *IEEE Trans. Inform. Theory* **IT-22** 45–52.
[5] BOBROVSKY, B. Z. and ZAKAI, M. (1982). Asymptotic a priori estimates for the error in the nonlinear filtering problem. *IEEE Trans. Inform. Theory* **IT-28** 371–376.
[6] BOBROVSKY, B. Z., ZAKAI, M. and ZEITOUNI, O. (1986). Error bounds for the nonlinear filtering of signals with small diffusion coefficients. EE Publication No. 583, Faculty of Electrical Engineering, Technion, Israel.
[7] CHATTERJI, S. D. (1982). A remark on the Cramér–Rao bound. In *Statistics and Probability: Essays in Honor of C. R. Rao* (G. Kallianpur, P. R. Krishnaiah and J. K. Ghosh, eds.) 193–196. North-Holland, Amsterdam.
[8] CHAZAN, D., ZIV, J. and ZAKAI, M. (1975). Improved lower bounds on signal parameter estimation. *IEEE Trans. Inform. Theory* **IT-21** 90–93.
[9] DONOGHUE, W. F. (1969). *Distributions and Fourier Transforms*. Academic, New York.
[10] LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.
[11] MILLER, R. and CHANG, C. (1978). A modified Cramér–Rao bound and its applications. *IEEE Trans. Inform. Theory* **IT-24** 398–400.
[12] VAN TREES, H. (1968). *Detection, Estimation and Modulation Theory* **1**. Wiley, New York.
[13] WEISS, A. and WEINSTEIN, E. (1985). A lower bound on the mean square error in random parameter estimation. *IEEE Trans. Inform. Theory* **IT-31** 680–682.
[14] ZAKAI, M. and ZIV, J. (1975). A generalization of the rate-distortion theory and applications. In *Information Theory, New Trends and Open Problems* (G. Longo, ed.). *CISM Courses and Lectures* **219** 87–123. Springer, Vienna.

B. Z. BOBROVSKY
DEPARTMENT OF ELECTRONICS
SCHOOL OF ENGINEERING
TEL-AVIV UNIVERSITY
TEL-AVIV
ISRAEL

E. MAYER-WOLF
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MASSACHUSETTS
AMHERST, MASSACHUSETTS 01003

M. ZAKAI
DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION
TECHNION CITY, HAIFA 32000
ISRAEL