A BAYESIAN ESTIMATION BOUND BASED ON THE OPTIMAL BIAS FUNCTION

Zvika Ben-Haim and Yonina C. Eldar

Department of Electrical Engineering
Technion—Israel Institute of Technology
Haifa 32000, Israel

ABSTRACT

We consider the problem of finding a lower bound on the minimum mean-squared error in a Bayesian estimation problem. The bound of Young and Westerberg, which is based on determining the optimal bias function, is extended to the case of a vector parameter. A numerical study demonstrates that the bound is both tighter and simpler to compute than alternative techniques.

Index Terms— Bayesian estimation, minimum mean-squared error estimation, performance bounds.

1. INTRODUCTION

Consider the Bayesian problem of estimating a random vector $\boldsymbol{\theta}$ from observations $\boldsymbol{x}$. It is well-known that the posterior mean $E\{\boldsymbol{\theta}|\boldsymbol{x}\}$ is the technique minimizing the mean-squared error (MSE). However, the complexity of computing the posterior mean is often prohibitive, and various approaches have been developed as alternatives. It is therefore of interest to determine the degradation in accuracy resulting from the use of suboptimal methods. Unfortunately, computation of the optimal MSE is itself infeasible in many cases. This has led to a large body of work seeking to find simple lower bounds for the minimum MSE in a given estimation problem [1–8]. Some of these are asymptotically tight, but in non-asymptotic cases there is usually a considerable gap between the bounds and the actual MSE obtained by the optimal estimator.

The situation is different for estimation of a deterministic parameter $\boldsymbol{\theta}$. The simplest deterministic bound, the Cramér–Rao bound (CRB), requires *a priori* specification of a bias function $\boldsymbol{b}(\boldsymbol{\theta})$ [9]. However, in many cases, any chosen bias $\boldsymbol{b}(\boldsymbol{\theta})$ yields a tight bound, in the sense that the CRB equals the MSE of the optimal estimator having bias $\boldsymbol{b}(\boldsymbol{\theta})$.

The tightness of the deterministic CRB motivates its application to problems in which $\boldsymbol{\theta}$ is random. Such an application was described by Young and Westerberg, who considered the case of a scalar θ constrained to the interval $[\theta_0, \theta_1]$. They used the prior distribution of θ to determine the optimal bias, and thus obtained a Bayesian bound from the CRB. In this paper, we extend the work of Young and Westerberg to a vector parameter constrained to a compact set. A general bound is given in the form of a solution to a partial differential equation. Under suitable symmetry conditions, we show that the vector problem can be reduced to a scalar setting. We compare the optimal bias technique to other Bayesian bounds, and show that, in some cases, the proposed method is not only simpler to compute, but is also tighter than all relevant bounds.

2. BAYESIAN ESTIMATION BOUND

Let $\boldsymbol{\theta}$ be a random vector which is to be estimated from observations $\boldsymbol{x}$. Denote the prior pdf of $\boldsymbol{\theta}$ by $\pi(\boldsymbol{\theta})$ and the conditional pdf of $\boldsymbol{x}$ given $\boldsymbol{\theta}$ by $p_{\boldsymbol{x}|\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{\theta})$. Let $\Theta \subseteq \mathbb{R}^n$ be the set of values of $\boldsymbol{\theta}$ for which $\pi(\boldsymbol{\theta}) > 0$.

Our goal is to develop a bound on the MSE of an estimator $\hat{\boldsymbol{\theta}}$. To this end, consider $\hat{\boldsymbol{\theta}}$ as an estimator of a *deterministic* parameter $\boldsymbol{\theta}$ from measurements distributed as $p_{\boldsymbol{x}|\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{\theta})$. As such, the bias of $\hat{\boldsymbol{\theta}}$ is defined by $\boldsymbol{b}(\boldsymbol{\theta}) = E\{\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}\} - \boldsymbol{\theta}$, and the Fisher information matrix is

$$[\boldsymbol{J}(\boldsymbol{\theta})]_{ij} = E\left\{ \frac{\partial \log p_{\boldsymbol{x}|\boldsymbol{\theta}}}{\partial \theta_i} \frac{\partial \log p_{\boldsymbol{x}|\boldsymbol{\theta}}}{\partial \theta_j} \middle| \boldsymbol{\theta} \right\}. \quad (1)$$

We assume that $\boldsymbol{J}(\boldsymbol{\theta})$ is finite and positive definite for all $\boldsymbol{\theta} \in \Theta$. We further assume that the regularity condition

$$\int t(\boldsymbol{x}) \frac{\partial}{\partial \theta_i} p_{\boldsymbol{x}|\theta}(\boldsymbol{x}|\theta) d\boldsymbol{x} = \frac{\partial}{\partial \theta_i} \int t(\boldsymbol{x}) p_{\boldsymbol{x}|\theta}(\boldsymbol{x}|\theta) d\boldsymbol{x} \quad (2)$$

holds for any function $t(\boldsymbol{x})$ such that $E\{|t(\boldsymbol{x})| \mid \boldsymbol{\theta}\} < \infty$. Roughly speaking, these regularity conditions require that the measurements contain data about the unknown parameter, and that the support of $p_{\boldsymbol{x}|\boldsymbol{\theta}}$ does not depend on $\boldsymbol{\theta}$ [9].

Under these assumptions, the CRB states that

$$E\left\{ (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^2 \middle| \boldsymbol{\theta} \right\} \geq \mathrm{CRB}[\boldsymbol{b}, \boldsymbol{\theta}]$$
$$\triangleq \mathrm{Tr}\left[\left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}} \right) \boldsymbol{J}^{-1} \left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}} \right)^T \right] + \|\boldsymbol{b}(\boldsymbol{\theta})\|^2. \quad (3)$$

Averaging (3) over $\boldsymbol{\theta}$, we obtain that, for any technique $\hat{\boldsymbol{\theta}}$ with bias function $\boldsymbol{b}(\boldsymbol{\theta})$,

$$E\left\{ (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^2 \right\} \geq Z[\boldsymbol{b}] \triangleq \int_{\Theta} \mathrm{CRB}[\boldsymbol{b}, \boldsymbol{\theta}]\, \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta} \quad (4)$$

978-1-4244-1714-8/07/$25.00 ©2007 IEEE

where the expectation is now performed over both $\boldsymbol{\theta}$ and $\boldsymbol{x}$. Note that any estimator has *some* bias function; furthermore, if Θ is bounded, then it can be shown that any reasonable estimator has a differentiable bias function. Thus, a lower bound on the Bayesian MSE of any estimator $\hat{\theta}$ is given by minimizing $Z[\boldsymbol{b}]$ over all functions $\boldsymbol{b} : \Theta \to \mathbb{R}^n$.

We summarize this result in the following theorem.

Theorem 1. *Let $\boldsymbol{\theta}$ be an unknown random vector with pdf $\pi(\boldsymbol{\theta}) > 0$ over the set $\Theta \subseteq \mathbb{R}^n$, and let $\boldsymbol{x}$ be a measurement vector whose pdf, conditioned on $\boldsymbol{\theta}$, is given by $p_{\boldsymbol{x}|\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{\theta})$. Assume that the deterministic Fisher information matrix $\boldsymbol{J}(\boldsymbol{\theta})$ is finite and positive definite for all $\boldsymbol{\theta}$, and that the regularity condition* (2) *holds. Then, for any estimator $\hat{\boldsymbol{\theta}}$,*

$$E\left\{\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2\right\} \geq \min_{\boldsymbol{b}(\boldsymbol{\theta})} \int_{\Theta} \text{CRB}[\boldsymbol{b}, \boldsymbol{\theta}]\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{5}$$

where the minimization is performed over all functions $\boldsymbol{b}$. If Θ is bounded, it suffices to minimize over all differentiable $\boldsymbol{b}$.

An important observation is that Theorem 1 arises from the deterministic CRB; hence, there are no requirements on the prior distribution $\pi(\boldsymbol{\theta})$. In particular, $\pi(\boldsymbol{\theta})$ can be discontinuous or have bounded support. As we will see, many previous Bayesian bounds do not apply in such circumstances.

3. CALCULATING THE BOUND

In finite-dimensional convex optimization problems, the requirement of a vanishing first derivative results in a set of equations, whose solution is the global minimum. Analogously, in the case of functional optimization problems such as (5), the optimum is given by the solution of a differential equation. The following theorem, whose proof can be found in Section 6, specifies the differential equation relevant to our optimization problem.

Theorem 2. *Under the conditions of Theorem 1, suppose Θ is a compact subset of $\mathbb{R}^n$ with a smooth boundary Γ. Then, a bias function $\boldsymbol{b}(\boldsymbol{\theta})$ is a global minimum of* (5) *if and only if it is a solution to the system of partial differential equations*

$$\pi b_i = \pi \sum_{j,k} \frac{\partial^2 b_i}{\partial\theta_j \partial\theta_k} (\boldsymbol{J}^{-1})_{jk} + \sum_{j,k} \left(\delta_{ik} + \frac{\partial b_i}{\partial\theta_k}\right) \left((\boldsymbol{J}^{-1})_{jk} \frac{\partial\pi}{\partial\theta_j} + \pi \frac{\partial(\boldsymbol{J}^{-1})_{jk}}{\partial\theta_j}\right) \tag{6}$$

for $i = 1, \ldots n$, within the range $\boldsymbol{\theta} \in \Theta$, which satisfies the boundary conditions

$$\left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}}\right) \boldsymbol{J}^{-1} \boldsymbol{\nu}(\boldsymbol{\theta}) = 0 \tag{7}$$

for all points $\boldsymbol{\theta} \in \Gamma$, where $\boldsymbol{\nu}(\boldsymbol{\theta})$ is a normal[1] *to the boundary at $\boldsymbol{\theta}$.*

[1] The normal vector $\boldsymbol{\nu}(\boldsymbol{\theta})$ of a surface given by $\{\boldsymbol{\theta} : q(\boldsymbol{\theta}) = 0\}$ can be obtained by computing the gradient of q at $\boldsymbol{\theta}$.

The bound of Young and Westerberg [3] is a special case of Theorem 2, and is given here for completeness.

Corollary 1. *Under the settings of Theorem 1, suppose $\Theta = [\theta_0, \theta_1]$ is a closed interval in $\mathbb{R}$. Then, any bias function $b(\theta)$ minimizing* (5) *is a solution to the differential equation*

$$J(\theta)b(\theta) = b''(\theta) + (1 + b'(\theta)) \left(\frac{d \log \pi}{d\theta} - \frac{d \log J}{d\theta}\right) \tag{8}$$

subject to the boundary conditions $b'(\theta_0) = b'(\theta_1) = -1$.

Theorem 2 can be solved numerically, e.g. using the Matlab `pde` toolbox, thus obtaining a bound for any problem satisfying the regularity conditions. However, in many cases, symmetry relations in the problem can be used to simplify the solution. For example, the following spherically symmetric case can be reduced to a problem similar to that of Corollary 1. The proof of this theorem can be found in Section 6.

Theorem 3. *Under the setting of Theorem 1, suppose that $\Theta = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| \leq r\}$ is a sphere centered on the origin, $\pi = \pi(\|\boldsymbol{\theta}\|)$ is spherically symmetric, and $\boldsymbol{J}(\boldsymbol{\theta}) = J(\|\boldsymbol{\theta}\|)\boldsymbol{I}$, where $J(\cdot)$ is a scalar function*[2]*. Then,* (5) *is solved by*

$$\boldsymbol{b} = b(\|\boldsymbol{\theta}\|) \frac{\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|} \tag{9}$$

where $b(\cdot)$ is a solution to the differential equation

$$J(\theta)b(\theta) = b''(\theta) + (1 + b'(\theta)) \left(\frac{d \log \pi}{d\theta} - \frac{d \log J}{d\theta}\right) + (n-1)\left(\frac{b'(\rho)}{\rho} - \frac{b(\rho)}{\rho^2}\right) \tag{10}$$

with boundary conditions $b(0) = 0$ and $b'(r) = -1$.

4. COMPARISON WITH OTHER BOUNDS

The original bound of Young and Westerberg [3] predates most Bayesian bounds, and, surprisingly, it has never been cited by or compared with later results. In this section, we measure its performance against that of various other techniques in a simple estimation scenario.

Consider the case in which θ is uniformly distributed in the range $\Theta = [-r, r]$. Let $x = \theta + w$ be a noisy observation, where w is zero-mean Gaussian noise, independent of θ, with variance σ^2. We wish to estimate θ from x. The Fisher information for this location problem is $J(\theta) = 1/\sigma^2$. Neither the optimal estimator of θ nor its MSE can be written as a closed form expression involving elementary functions. Thus, a simple expression bounding the optimal MSE is of interest.

It follows from Corollary 1 that, for any estimator $\hat{\theta}$ [3],

$$E\left\{(\theta - \hat{\theta})^2\right\} \geq \sigma^2 \left(1 - \frac{\tanh(r/\sigma)}{r/\sigma}\right). \tag{11}$$

[2] This form of $\boldsymbol{J}(\boldsymbol{\theta})$ occurs, for example, in the estimation of location parameters from independent, identically distributed measurements.

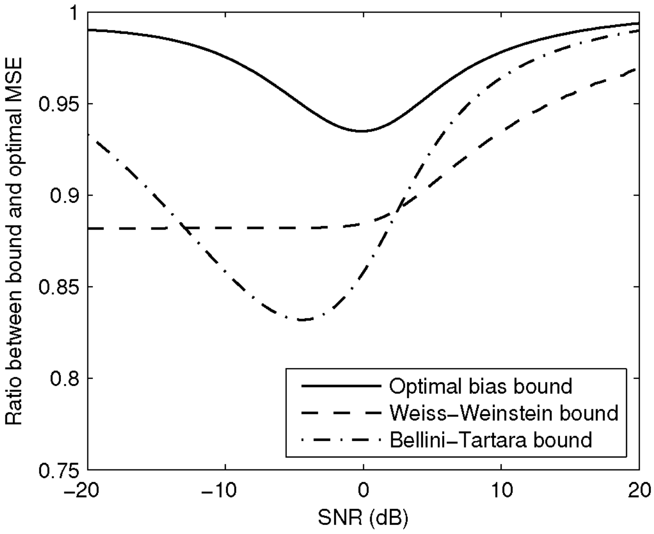

Fig. 1. Ratio between the MSE bounds and the optimal achievable MSE.

Despite the widespread use of finite-support prior distributions [2, 7], the regularity conditions of many bounds are violated by such prior pdf functions. Indeed, the Bayesian CRB of Van Trees [1], the Bobrovski–Zakai bound [6], and the Bayesian Abel bound [8] all assume that $\pi(\theta)$ has infinite support, and thus cannot be applied in this scenario.

Techniques from the Ziv–Zakai family are applicable to constrained problems. These include the Ziv–Zakai bound itself [2], the Bellini–Tartara bound [4], and the Chazan–Zakai–Ziv bound [5]. Of these, the Bellini–Tartara bound is known to be tightest. In the current setting, it is given by

$$E\left\{(\theta-\hat{\theta})^2\right\} \geq \int_0^{2r} \left(1-\frac{\xi}{2r}\right) \xi\, Q\!\left(\frac{\xi}{2\sigma}\right) d\xi \tag{12}$$

where $Q(z) = (2\pi)^{-1/2} \int_z^\infty e^{-t^2/2} dt$ is the tail function of the normal distribution.

We also compare the optimal bias bound with the result of Weinstein and Weiss [7], which is given by

$$E\left\{(\theta-\hat{\theta})^2\right\} \geq \max_{s\in(0,1),h\in\mathbb{R}} \frac{h^2 M^2(s,h)}{M(2s,h)+M(2s-1,h)-2M(s,2h)} \tag{13}$$

where

$$M(s,h) = \max\left(0,\, 1-\frac{|h|}{2r}\right) e^{-h^2 s(1-s)/2\sigma^2}. \tag{14}$$

We note that neither the Bellini–Tartara bound nor the Weiss–Weinstein bound can be expressed in a closed form involving only elementary functions. Yet, in this setting, the simple formula (11) of the optimal bias bound is substantially tighter. This is demonstrated in Fig. 1, which plots the ratio between the different bounds and the exact optimal MSE. To obtain this figure, the exact value of the optimal MSE was computed by numerical integration of $E\{(\boldsymbol{\theta} - E\{\boldsymbol{\theta}|\boldsymbol{x}\})^2\}$.

Apart from the reduction in computational complexity, the simplicity of (11) also emphasizes several features of the estimation problem. First, the dependence of the problem on the dimensionless quantity r/σ, rather than on r and σ separately, is clear. This is to be expected, as a change in units of measurement would multiply both r and σ by a constant. Second, for $r \gg \sigma$, the bound converges to the noise variance σ^2, corresponding to an uninformative prior whose optimal estimator is $\hat{\theta} = x$. Finally, for $\sigma \gg r$, the bound converges to $r^2/3$, corresponding to the case of uninformative measurements, where the optimal estimator is $\hat{\theta} = 0$. Thus, the bound (11) is tight both for very low and for very high SNR; this can also been seen from Fig. 1.

5. CONCLUSION

Although often considered distinct settings, there are insightful connections between the Bayesian and deterministic estimation problems. One such relation is the use of the deterministic CRB in a Bayesian problem. The combination of this generally tight deterministic bound with the well-defined Bayesian optimality criterion results in a tight Bayesian bound. Application to the location estimation problem demonstrates that the technique is both simpler and tighter than alternative approaches.

6. APPENDIX: PROOFS OF THEOREMS 2 AND 3

Proof of Theorem 2. Consider the more general problem of minimizing the functional

$$Z[\boldsymbol{b}] = \int_\Theta F[\boldsymbol{b},\boldsymbol{\theta}]d\boldsymbol{\theta} \tag{15}$$

where $F[\boldsymbol{b},\boldsymbol{\theta}]$ is smooth and convex in $\boldsymbol{b} : \Theta \to \mathbb{R}^n$, and $\Theta \subset \mathbb{R}^n$ is a compact set with a smooth boundary Γ. Then, $Z[\boldsymbol{b}]$ is also smooth and convex in $\boldsymbol{b}$, so that $\boldsymbol{b}$ is a global minimum of $Z[\boldsymbol{b}]$ if and only if the differential $\delta Z[\boldsymbol{h}]$ equals zero at $\boldsymbol{b}$ for all admissible functions $\boldsymbol{h} : \Theta \to \mathbb{R}^n$ [10].

By a standard technique [10, §35], it can be shown that

$$\delta Z[\boldsymbol{h}] = \epsilon \sum_i \int_\Theta \left(\frac{\partial F}{\partial b_i} - \sum_j \frac{\partial}{\partial\theta_j}\frac{\partial F}{\partial b_i^{(j)}}\right) h_i(\boldsymbol{\theta}) d\boldsymbol{\theta} + \epsilon \sum_i \int_\Gamma \left(\frac{\partial F}{\partial b_i^{(1)}},\ldots,\frac{\partial F}{\partial b_i^{(n)}}\right)^T \boldsymbol{\nu}(\boldsymbol{\theta})\, h_i(\boldsymbol{\theta})\, d\sigma \tag{16}$$

where ϵ is an infinitesimal quantity, $b_i^{(j)} = \partial b_i/\partial\theta_j$, and $\boldsymbol{\nu}(\boldsymbol{\theta})$ is an outward-pointing normal at the boundary point $\boldsymbol{\theta} \in \Gamma$. We now seek conditions for which $\delta Z[\boldsymbol{h}] = 0$ for all $\boldsymbol{h}(\boldsymbol{\theta})$.

Consider first functions $\boldsymbol{h}(\boldsymbol{\theta})$ which equal zero on the boundary Γ. In this case, the second integral vanishes, and we obtain the Euler–Lagrange equations

$$\forall i,\ \frac{\partial F}{\partial b_i} - \sum_j \frac{\partial}{\partial \theta_j} \frac{\partial F}{\partial b_i^{(j)}} = 0. \tag{17}$$

Substituting this result back into (16), and again using the fact that $\delta Z[\boldsymbol{h}] = 0$ for all $\boldsymbol{h}$, we obtain the boundary condition

$$\forall i,\ \forall \boldsymbol{\theta} \in \Gamma,\ \left(\frac{\partial F}{\partial b_i^{(1)}}, \ldots, \frac{\partial F}{\partial b_i^{(n)}} \right)^T \boldsymbol{\nu}(\boldsymbol{\theta}) = 0. \tag{18}$$

Plugging $F[\boldsymbol{b}, \boldsymbol{\theta}] = \mathrm{CRB}[\boldsymbol{b}, \boldsymbol{\theta}]\pi(\boldsymbol{\theta})$ into (17) and (18) provides the required result. □

The proof of Theorem 3 is based on the following lemma.

Lemma 1. *Under the conditions of Theorem 3, the functional* $Z[\boldsymbol{b}]$ *of* (4) *is rotation and reflection invariant, i.e., its value does not change if* $\boldsymbol{b}$ *is rotated about the origin or reflected through a hyperplane which contains the origin.*

Proof. We begin by considering a rotation about the first two coordinates, such that $\boldsymbol{b}$ is transformed to

$$\begin{aligned}\tilde{\boldsymbol{b}} \triangleq (&b_1(\boldsymbol{\theta}) \cos\phi + b_2(\boldsymbol{\theta}) \sin\phi, \\ &- b_1(\boldsymbol{\theta}) \sin\phi + b_2(\boldsymbol{\theta}) \cos\phi, b_3(\boldsymbol{\theta}), \ldots b_n(\boldsymbol{\theta}))^T.\end{aligned} \tag{19}$$

We now perform the change of variables $\boldsymbol{\theta} \mapsto \tilde{\boldsymbol{\theta}}$, where $\tilde{\boldsymbol{\theta}} \triangleq (\theta_1 \cos\phi + \theta_2 \sin\phi, -\theta_1 \sin\phi + \theta_2 \cos\phi, \theta_3, \ldots \theta_n)^T$, so that

$$Z[\tilde{\boldsymbol{b}}] = \int_\Theta \mathrm{CRB}[\tilde{\boldsymbol{b}}, \boldsymbol{\theta}]\pi(\boldsymbol{\theta})d\boldsymbol{\theta} = \int_\Theta \mathrm{CRB}[\tilde{\boldsymbol{b}}, \tilde{\boldsymbol{\theta}}]\pi(\tilde{\boldsymbol{\theta}})d\tilde{\boldsymbol{\theta}}. \tag{20}$$

By assumption, π, $\boldsymbol{J}$ and Θ are unaffected by the change of variables. Using the chain rule and some tedious manipulations, it can be verified that $\mathrm{CRB}[\tilde{\boldsymbol{b}}, \tilde{\boldsymbol{\theta}}] = \mathrm{CRB}[\boldsymbol{b}, \boldsymbol{\theta}]$, and therefore $Z[\boldsymbol{b}] = Z[\tilde{\boldsymbol{b}}]$. The result similarly holds for rotations about any other two coordinates of $\boldsymbol{\theta}$. Since any rotation can be decomposed into a sequence of two-coordinate rotations, we conclude that $Z[\boldsymbol{b}]$ is rotation invariant.

Next, we prove that $Z[\boldsymbol{b}]$ is invariant to reflections through hyperplanes containing the origin. Since $Z[\boldsymbol{b}]$ is invariant to rotations, it suffices to choose a single hyperplane, say $\{\boldsymbol{\theta} : \theta_1 = 0\}$. Let $\tilde{\boldsymbol{b}} \triangleq (-b_1(\boldsymbol{\theta}), b_2(\boldsymbol{\theta}), \ldots b_n(\boldsymbol{\theta}))^T$ be the reflection of $\boldsymbol{b}$ and consider the corresponding change of variables $\tilde{\boldsymbol{\theta}} \triangleq (-\theta_1, \theta_2, \ldots \theta_n)^T$. By the symmetry assumptions, π and $\boldsymbol{J}$ are unaffected by the change of variables; furthermore, $\partial\tilde{\boldsymbol{b}}/\partial\tilde{\boldsymbol{\theta}} = \partial\boldsymbol{b}/\partial\boldsymbol{\theta}$. It follows that $\mathrm{CRB}[\tilde{\boldsymbol{b}}, \tilde{\boldsymbol{\theta}}] = \mathrm{CRB}[\boldsymbol{b}, \boldsymbol{\theta}]$, and therefore $Z[\boldsymbol{b}] = Z[\tilde{\boldsymbol{b}}]$. □

Proof of Theorem 3. As noted previously, $Z[\boldsymbol{b}]$ is convex in $\boldsymbol{b}$; therefore, the set of optimal solutions of (5) is convex. Now suppose there exists an optimal solution $\boldsymbol{b}_o$ which is not rotation invariant. Then, by Lemma 1, its rotations about the origin are also optimal. By convexity, the average $\boldsymbol{b}_a$ of all such rotations, which is rotation invariant, is optimal as well.

Furthermore, suppose that $\boldsymbol{b}_a$ is not radial, i.e., for some value of $\boldsymbol{\theta}$, $\boldsymbol{b}_a(\boldsymbol{\theta})$ contains a component perpendicular to the vector $\boldsymbol{\theta}$. Consider a hyperplane passing through the origin, whose normal is the aforementioned perpendicular component. By Lemma 1, The reflection $\boldsymbol{b}_r$ of $\boldsymbol{b}_a$ through this hyperplane is also an optimal solution of (5), as is the average $\boldsymbol{b}_f = (\boldsymbol{b}_r + \boldsymbol{b}_a)/2$. However, $\boldsymbol{b}_f$ no longer has a component perpendicular to $\boldsymbol{\theta}$, and is thus radial. Therefore, $\boldsymbol{b}_f$, which is an optimal solution, is spherically symmetric and radial.

To determine $b(\cdot)$, it suffices to find a solution along a single line segment from the origin to the boundary of Θ. Choosing the segment along the θ_1 axis, we have $\boldsymbol{b}(\boldsymbol{\theta}) = b(\theta_1)\boldsymbol{e}_1$, where $\boldsymbol{e}_1$ is a unit vector in the direction of θ_1. Furthermore, $\boldsymbol{J}(\boldsymbol{\theta}) = J(\theta_1)\boldsymbol{I}$, and $\pi(\boldsymbol{\theta}) = \pi(\theta_1)$. Substitution of these values into (6) and (7), along with careful calculation of the derivatives, yields the required result. □

7. REFERENCES

[1] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. Wiley, 1968, vol. 1.

[2] J. Ziv and M. Zakai, “Some lower bounds on signal parameter estimation,” *IEEE Trans. Inf. Theory*, vol. 15, no. 3, pp. 386–391, May 1969.

[3] T. Y. Young and R. A. Westerberg, “Error bounds for stochastic estimation of signal parameters,” *IEEE Trans. Inf. Theory*, vol. 17, no. 5, pp. 549–557, Sep. 1971.

[4] S. Bellini and G. Tartara, “Bounds on error in signal parameter estimation,” *IEEE Trans. Commun.*, vol. 22, no. 3, pp. 340–342, 1974.

[5] D. Chazan, M. Zakai, and J. Ziv, “Improved lower bounds on signal parameter estimation,” *IEEE Trans. Inf. Theory*, vol. 21, no. 1, pp. 90–93, 1975.

[6] B. Z. Bobrovski and M. Zakai, “A lower bound on the estimation error for certain diffusion problems,” *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 45–52, Jan. 1976.

[7] E. Weinstein and A. J. Weiss, “A general class of lower bounds in parameter estimation,” *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 338–342, Mar. 1988.

[8] A. Renaux, P. Forster, P. Larzabal, and C. Richmond, “The Bayesian Abel bound on the mean square error,” in *ICASSP 2006*, vol. III, Toulouse, France, May 2006, pp. 9–12.

[9] J. Shao, *Mathematical Statistics*. Springer, 1998.

[10] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*. Dover, 2000.