A Lower Bound on the Bayesian MSE Based on the Optimal Bias Function

Zvika Ben-Haim, *Student Member, IEEE*, and Yonina C. Eldar, *Senior Member, IEEE*

***Abstract*—A lower bound on the minimum mean-squared error (MSE) in a Bayesian estimation problem is proposed in this paper. This bound utilizes a well-known connection to the deterministic estimation setting. Using the prior distribution, the bias function which minimizes the Cramér–Rao bound can be determined, resulting in a lower bound on the Bayesian MSE. The bound is developed for the general case of a vector parameter with an arbitrary probability distribution, and is shown to be asymptotically tight in both the high and low signal-to-noise ratio (SNR) regimes. A numerical study demonstrates several cases in which the proposed technique is both simpler to compute and tighter than alternative methods.**

***Index Terms*—Bayesian bounds, Bayesian estimation, minimum mean-squared error (MSE) estimation, optimal bias, performance bounds.**

I. Introduction

THE goal of estimation theory is to infer the value of an unknown parameter based on observations. A common approach to this problem is the Bayesian framework, in which the estimate is constructed by combining the measurements with prior information about the parameter [1]. In this setting, the parameter $\boldsymbol{\theta}$ is random, and its distribution describes the *a priori* knowledge of the unknown value. In addition, measurements $\boldsymbol{x}$ are obtained, whose conditional distribution, given $\boldsymbol{\theta}$, provides further information about the parameter. The objective is to construct an estimator $\hat{\boldsymbol{\theta}}$, which is a function of the measurements, so that $\hat{\boldsymbol{\theta}}$ is close to $\boldsymbol{\theta}$ in some sense. A common measure of the quality of an estimator is its mean-squared error (MSE), given by $E\{\|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\|^2\}$.

It is well-known that the posterior mean $E\{\boldsymbol{\theta}|\boldsymbol{x}\}$ is the technique minimizing the MSE. Thus, from a theoretical perspective, there is no difficulty in finding the minimum MSE (MMSE) estimator in any given problem. In practice, however, the complexity of computing the posterior mean is often prohibitive. As a result, various alternatives, such as the maximum *a posteriori* probability (MAP) technique, have been developed [2]. The purpose of such methods is to approach the performance of the MMSE estimator with a computationally efficient algorithm.

Manuscript received April 28, 2008; revised April 03, 2009. Current version published October 21, 2009. This work was supported in part by the Israel Science Foundation under Grant 1081/07 and by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM++ under Contract 216715.

The authors are with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: zvikabh@ee.technion.ac.il; yonina@ee.technion.ac.il).

Communicated by L. Tong, Associate Editor for Detection and Estimation.
Digital Object Identifier 10.1109/TIT.2009.2030451

An important goal is to quantify the performance degradation resulting from the use of these suboptimal techniques. One way to do this is to compare the MSE of the method used in practice with the MMSE. Unfortunately, computation of the MMSE is itself infeasible in many cases. This has led to a large body of work seeking to find simple lower bounds on the MMSE in various estimation problems [3]–[12].

Generally speaking, previous bounds can be divided into two categories. The Weiss–Weinstein family is based on a covariance inequality and includes the Bayesian Cramér–Rao bound [3], the Bobrovski–Zakai bound [8], and the Weiss–Weinstein bound [9], [10]. The Ziv–Zakai family of bounds is based on comparing the estimation problem to a related detection scenario. This family includes the Ziv–Zakai bound [4] and its improvements, notably the Bellini–Tartara bound [6], the Chazan–Zakai–Ziv bound [7], and the generalization of Bell *et al.* [11]. Recently, Renaux *et al.* have combined both approaches [12].

The accuracy of the bounds described above is usually tested numerically in particular estimation settings. Few of the previous results provide analytical proofs of accuracy, even under asymptotic conditions. Bellini and Tartara [6] briefly discuss performance of their bound at high signal-to-noise ratio (SNR), and Bell *et al.* [11] prove that their bound converges to the true value at low SNR for a particular family of Gaussian-like probability distributions. To the best of our knowledge, there are no other results concerning the asymptotic performance of Bayesian bounds.

A different estimation setting arises when one considers $\boldsymbol{\theta}$ as a *deterministic* unknown parameter. In this case, too, a common goal is to construct an estimator having low MSE. However, the term MSE has a very different meaning in the deterministic setting, since in this case, the expectation is taken only over the random variable $\boldsymbol{x}$. One elementary difference with far-reaching implications is that in the Bayesian case, the MSE is a single real number, whereas the deterministic MSE is a function of the unknown parameter $\boldsymbol{\theta}$ [13]–[15].

Many lower bounds have been developed for the deterministic setting, as well. These include classical results such as the Cramér–Rao [16], [17], Hammersley–Chapman–Robbins [18], [19], Bhattacharya [20], and Barankin [21] bounds, as well as more recent results [22]–[27]. By far, the simplest and most commonly used of these approaches is the Cramér–Rao bound (CRB). Like most other deterministic bounds, the CRB deals explicitly with unbiased estimators, or, equivalently, with estimators having a specific, prespecified bias function. Although many later bounds are sharper [14], [25], [28], the CRB is known to be *asymptotically* tight in many cases.

0018-9448/$26.00 © 2009 IEEE

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on November 17, 2009 at 07:06 from IEEE Xplore. Restrictions apply.

While the deterministic and Bayesian settings stem from different points of view, there exist insightful relations between the two approaches. The basis for this connection is the fact that by adding a prior distribution for $\boldsymbol{\theta}$, any deterministic problem can be transformed to a corresponding Bayesian setting. Several theorems relate the performance of corresponding Bayesian and deterministic scenarios [13]. As a consequence, numerous bounds have both a deterministic and a Bayesian version [3], [10], [12], [29].

The simplicity and asymptotic tightness of the deterministic CRB motivate its use in problems in which $\boldsymbol{\theta}$ is random. Such an application was described by Young and Westerberg [5], 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 function for use in the biased CRB, and thus obtained a Bayesian bound. It should be noted that this result differs from the Bayesian CRB of Van Trees [3]; the two bounds are compared in Section II-C. We refer to the result of Young and Westerberg as the optimal-bias bound (OBB), since it is based on choosing the bias function which optimizes the CRB using the given prior distribution.

This paper provides an extension and a deeper analysis of the OBB. Specifically, we generalize the bound to an arbitrary n-dimensional estimation setting [30]. The bound is determined by finding the solution to a certain partial differential equation. Using tools from functional analysis, we demonstrate that a unique solution exists for this differential equation. Under suitable symmetry conditions, it is shown that the method can be reduced to the solution of an ordinary differential equation and, in some cases, presented in closed form.

The mathematical tools employed in this paper are also used for characterizing the performance of the OBB. Specifically, it is demonstrated analytically that the proposed bound is asymptotically tight for both high and low SNR values. Furthermore, the OBB is compared with several other bounds; in the examples considered, the OBB is both simpler computationally and more accurate than all relevant alternatives.

The remainder of this paper is organized as follows. In Section II, we derive the OBB for a vector parameter. Section III discusses some mathematical concepts required to ensure the existence of the OBB. In Section IV, a practical technique for calculating the bound is developed using variational calculus. In Section V, we demonstrate some properties of the OBB, including its asymptotic tightness. Finally, in Section VI, we compare the performance of the bound with that of other relevant techniques.

II. The Optimal-Bias Bound

In this section, we derive the OBB for the general vector case. To this end, we first examine the relation between the Bayesian and deterministic estimation settings (Section II-A). Next, we focus on the deterministic case and review the basic properties of the CRB (Section II-B). Finally, the OBB is derived from the CRB (Section II-C).

The focus of this paper is the Bayesian estimation problem, but the bound we propose stems from the theory of deterministic estimation. To avoid confusion, we will indicate that a particular quantity refers to the deterministic setting by appending the symbol $\boldsymbol{\theta}$ to it. For example, the notation $E\{\cdot\}$ denotes expectation over *both* $\boldsymbol{\theta}$ and $\boldsymbol{x}$, i.e., expectation in the Bayesian sense, while expectation solely over $\boldsymbol{x}$ (in the deterministic setting) is denoted by $E\{\cdot;\boldsymbol{\theta}\}$. The notation $E\{\cdot \mid \boldsymbol{\theta}\}$ indicates Bayesian expectation conditioned on $\boldsymbol{\theta}$.

Some further notation used throughout the paper is as follows. Lower case boldface letters signify vectors and upper case boldface letters indicate matrices. The ith component of a vector $\boldsymbol{v}$ is denoted v_i, while $\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}, \ldots$ signifies a sequence of vectors. The derivative $\partial f/\partial \boldsymbol{v}$ of a function $f(\boldsymbol{v})$ is a vector function whose ith element is $\partial f/\partial v_i$. Similarly, given a vector function $\boldsymbol{b}(\boldsymbol{\theta})$, the derivative $\partial \boldsymbol{b}/\partial \boldsymbol{\theta}$ is defined as the matrix function whose (i,j)th entry is $\partial b_i/\partial \theta_j$. The squared Euclidean norm $\boldsymbol{v}^T\boldsymbol{v}$ of a vector $\boldsymbol{v}$ is denoted $\|\boldsymbol{v}\|^2$, while the squared Frobenius norm $\mathrm{Tr}(\boldsymbol{M}\boldsymbol{M}^T)$ of a matrix $\boldsymbol{M}$ is denoted $\|\boldsymbol{M}\|_F^2$. In Section III, we will also define some functional norms, which will be of use later in the paper.

A. The Bayesian-Deterministic Connection

We now review a fundamental relation between the Bayesian and deterministic estimation settings. Let $\boldsymbol{\theta}$ be an unknown random vector in $\mathbb{R}^n$ and let $\boldsymbol{x}$ be a measurement vector. The joint probability density function (pdf) of $\boldsymbol{\theta}$ and $\boldsymbol{x}$ is $p_{\boldsymbol{x},\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{\theta}) = p_{\boldsymbol{x}|\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{\theta})p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$, where $p_{\boldsymbol{\theta}}$ is the prior distribution of $\boldsymbol{\theta}$ and $p_{\boldsymbol{x}|\boldsymbol{\theta}}$ is the conditional distribution of $\boldsymbol{x}$ given $\boldsymbol{\theta}$. For later use, define the set Θ of feasible parameter values by

$$\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^n : p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) > 0\}. \tag{1}$$

Suppose $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\boldsymbol{x})$ is an estimator of $\boldsymbol{\theta}$. Its (Bayesian) MSE is given by

$$\begin{aligned}\mathrm{MSE} &= E\left\{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2\right\} \\ &= \int \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 p_{\boldsymbol{x},\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{\theta})d\boldsymbol{x}d\boldsymbol{\theta}.\end{aligned} \tag{2}$$

By the law of total expectation, we have

$$\begin{aligned}\mathrm{MSE} &= \int \left(\int \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 p_{\boldsymbol{x}|\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{\theta})d\boldsymbol{x}\right) p_{\boldsymbol{\theta}}(\boldsymbol{\theta})d\boldsymbol{\theta} \\ &= E\left\{E\left\{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 \middle| \boldsymbol{\theta}\right\}\right\}.\end{aligned} \tag{3}$$

Now consider a deterministic estimation setting, i.e., suppose $\boldsymbol{\theta}$ is a deterministic unknown which is to be estimated from random measurements $\boldsymbol{x}$. Let the distribution $p_{\boldsymbol{x};\boldsymbol{\theta}}$ of $\boldsymbol{x}$ (as a function of $\boldsymbol{\theta}$) be given by $p_{\boldsymbol{x};\boldsymbol{\theta}}(\boldsymbol{x};\boldsymbol{\theta}) = p_{\boldsymbol{x}|\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{\theta})$, i.e., the distribution of $\boldsymbol{x}$ in the deterministic case equals the conditional distribution in the corresponding Bayesian problem.

The estimator $\hat{\boldsymbol{\theta}}$ defined above is simply a function of the measurements, and can therefore be applied in the deterministic case as well. Its deterministic MSE is given by

$$E\left\{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2;\boldsymbol{\theta}\right\} = \int \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 p_{\boldsymbol{x};\boldsymbol{\theta}}(\boldsymbol{x};\boldsymbol{\theta})d\boldsymbol{x}. \tag{4}$$

Since $p_{\boldsymbol{x};\boldsymbol{\theta}}(\boldsymbol{x};\boldsymbol{\theta}) = p_{\boldsymbol{x}|\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{\theta})$, we have

$$E\left\{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2;\boldsymbol{\theta}\right\} = E\left\{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 \middle| \boldsymbol{\theta}\right\}. \tag{5}$$

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on November 17, 2009 at 07:06 from IEEE Xplore. Restrictions apply.

Combining this fact with (3), we find that the Bayesian MSE equals the expectation of the MSE of the corresponding deterministic problem, i.e.,

$$E\left\{\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|^2\right\}=E\left\{E\left\{\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|^2;\boldsymbol{\theta}\right\}\right\}. \tag{6}$$

This relation will be used to construct the OBB in Section II-C.

B. The Deterministic CRB

Before developing the OBB, we review some basic results in the deterministic estimation setting. Suppose $\boldsymbol{\theta}$ is a deterministic parameter vector and let $\boldsymbol{x}$ be a measurement vector having pdf $p_{\boldsymbol{x};\boldsymbol{\theta}}(\boldsymbol{x};\boldsymbol{\theta})$. Denote by $\Theta\subseteq\mathbb{R}^n$ the set of all possible values of $\boldsymbol{\theta}$. We assume for technical reasons that Θ is an open set.[1]

Let $\hat{\boldsymbol{\theta}}$ be an estimator of $\boldsymbol{\theta}$ from the measurements $\boldsymbol{x}$. We require the following regularity conditions to ensure that the CRB holds [31, Sec. 3.1.3].

1) $p_{\boldsymbol{x};\boldsymbol{\theta}}(\boldsymbol{x};\boldsymbol{\theta})$ is continuously differentiable with respect to $\boldsymbol{\theta}$. This condition is required to ensure the existence of the Fisher information.
2) The Fisher information matrix $\boldsymbol{J}(\boldsymbol{\theta})$, defined by

$$[\boldsymbol{J}(\boldsymbol{\theta})]_{ij}=E\left\{\frac{\partial\log p_{\boldsymbol{x};\boldsymbol{\theta}}}{\partial\theta_i}\frac{\partial\log p_{\boldsymbol{x};\boldsymbol{\theta}}}{\partial\theta_j};\boldsymbol{\theta}\right\} \tag{7}$$

is bounded and positive definite for all $\boldsymbol{\theta}\in\Theta$. This ensures that the measurements contain data about the unknown parameter.
3) Exchanging the integral and derivative in the equation

$$\int t(\boldsymbol{x})\frac{\partial}{\partial\theta_i}p_{\boldsymbol{x};\theta}(\boldsymbol{x};\theta)d\boldsymbol{x}=\frac{\partial}{\partial\theta_i}\int t(\boldsymbol{x})p_{\boldsymbol{x};\theta}(\boldsymbol{x};\theta)d\boldsymbol{x} \tag{8}$$

is justified for any measurable function $t(\boldsymbol{x})$, in the sense that, if one side exists, then the other exists and the two sides are equal. A sufficient condition for this to hold is that the support of $p_{\boldsymbol{x};\boldsymbol{\theta}}$ does not depend on $\boldsymbol{\theta}$.
4) All estimators $\hat{\boldsymbol{\theta}}$ are Borel measurable functions which satisfy

$$\left\|\frac{\partial p_{\boldsymbol{x};\boldsymbol{\theta}}}{\partial\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^T\right\|_F\leq g(\boldsymbol{x}),\ \text{for all }\boldsymbol{\theta} \tag{9}$$

for some integrable function $g(\boldsymbol{x})$. This technical requirement is needed in order to exclude certain pathological estimators whose statistical behavior is insufficiently smooth to allow the application of the CRB.

The bias of an estimator $\hat{\boldsymbol{\theta}}$ is defined as

$$\boldsymbol{b}(\boldsymbol{\theta})=E\left\{\hat{\boldsymbol{\theta}};\boldsymbol{\theta}\right\}-\boldsymbol{\theta}. \tag{10}$$

[1]This is required in order to ensure that one can discuss differentiability of $p_{\boldsymbol{x};\boldsymbol{\theta}}$ with respect to $\boldsymbol{\theta}$ at any point $\boldsymbol{\theta}\in\Theta$. In the Bayesian setting to which we will return in Section II-C, Θ is defined by (1); in this case, adding a boundary to Θ essentially leaves the setting unchanged, as long as the prior probability for $\boldsymbol{\theta}$ to be on the boundary of Θ is zero. Therefore, this requirement is of little practical relevance.

Under the above assumptions, it can be shown that the bias of any estimator is continuously differentiable [5, Lemma 2]. Furthermore, under these assumptions, the CRB holds, and thus, for any estimator having bias $\boldsymbol{b}(\boldsymbol{\theta})$, we have

$$\begin{aligned}E\left\{\|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\|^2;\boldsymbol{\theta}\right\}&\geq\mathrm{CRB}[\boldsymbol{b},\boldsymbol{\theta}]\\&\triangleq\mathrm{Tr}\left[\left(\boldsymbol{I}+\frac{\partial\boldsymbol{b}}{\partial\boldsymbol{\theta}}\right)\boldsymbol{J}^{-1}(\boldsymbol{\theta})\left(\boldsymbol{I}+\frac{\partial\boldsymbol{b}}{\partial\boldsymbol{\theta}}\right)^T\right]+\|\boldsymbol{b}(\boldsymbol{\theta})\|^2.\end{aligned} \tag{11}$$

A more common form of the CRB is obtained by restricting attention to unbiased estimators (i.e., techniques for which $\boldsymbol{b}(\boldsymbol{\theta})=\boldsymbol{0}$). Under the unbiasedness assumption, the bound simplifies to $\mathrm{MSE}\geq\mathrm{Tr}(\boldsymbol{J}^{-1}(\boldsymbol{\theta}))$. However, in the sequel we will make use of the general form (11).

C. A Bayesian Bound From the CRB

The OBB of Young and Westerberg [5] is based on applying the Bayesian-deterministic connection described in Section II-A to the deterministic CRB (11). Specifically, returning now to the Bayesian setting, one can combine (6) and (11) to obtain that, for any estimator $\hat{\boldsymbol{\theta}}$ with bias function $\boldsymbol{b}(\boldsymbol{\theta})$

$$E\left\{\|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\|^2\right\}\geq Z[\boldsymbol{b}]\triangleq\int_\Theta\mathrm{CRB}[\boldsymbol{b},\boldsymbol{\theta}]\,p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \tag{12}$$

where the expectation is now performed over both $\boldsymbol{\theta}$ and $\boldsymbol{x}$. Note that (12) describes the Bayesian MSE as a function of a deterministic property (the bias) of $\hat{\boldsymbol{\theta}}$. Since any estimator has *some* bias function, and since all bias functions are continuously differentiable in our setting, minimizing $Z[\boldsymbol{b}]$ over all continuously differentiable functions $\boldsymbol{b}$ yields a lower bound on the MSE of any Bayesian estimator. Thus, under the regularity conditions of Section II-B, a lower bound on the Bayesian MSE is given by

$$\begin{aligned}s=\inf_{\boldsymbol{b}\in C^1}\int_\Theta\Bigg[&\|\boldsymbol{b}(\boldsymbol{\theta})\|^2\\&+\mathrm{Tr}\left(\left(\boldsymbol{I}+\frac{\partial\boldsymbol{b}}{\partial\boldsymbol{\theta}}\right)\boldsymbol{J}^{-1}(\boldsymbol{\theta})\left(\boldsymbol{I}+\frac{\partial\boldsymbol{b}}{\partial\boldsymbol{\theta}}\right)^T\right)\Bigg]p_{\boldsymbol{\theta}}(d\boldsymbol{\theta})\end{aligned} \tag{13}$$

where C^1 is the space of continuously differentiable functions $\boldsymbol{f}:\Theta\rightarrow\mathbb{R}^n$.

Note that the OBB differs from the Bayesian CRB of Van Trees [3]. Van Trees' result is based on applying the Cauchy–Schwarz inequality to the joint pdf $p_{\boldsymbol{x},\boldsymbol{\theta}}$, whereas the deterministic CRB is based on applying a similar procedure to $p_{\boldsymbol{x};\boldsymbol{\theta}}$. As a consequence, the regularity conditions required for the Bayesian CRB are stricter, requiring that $p_{\boldsymbol{x},\boldsymbol{\theta}}$ be twice differentiable with respect to $\boldsymbol{\theta}$. By contrast, the OBB requires differentiability only of the conditional pdf $p_{\boldsymbol{x}|\boldsymbol{\theta}}$. An example in which this difference is important is the case in which the prior distribution $p_{\boldsymbol{\theta}}$ is discontinuous, e.g., when $p_{\boldsymbol{\theta}}$ is uniform. The performance of the OBB in this setting will be examined in Section VI.

In the next section, we will see that it is advantageous to perform the minimization (13) over a somewhat modified class of functions. This will allow us to prove the unique existence of a solution to the optimization problem, a result which will be

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on November 17, 2009 at 07:06 from IEEE Xplore. Restrictions apply.

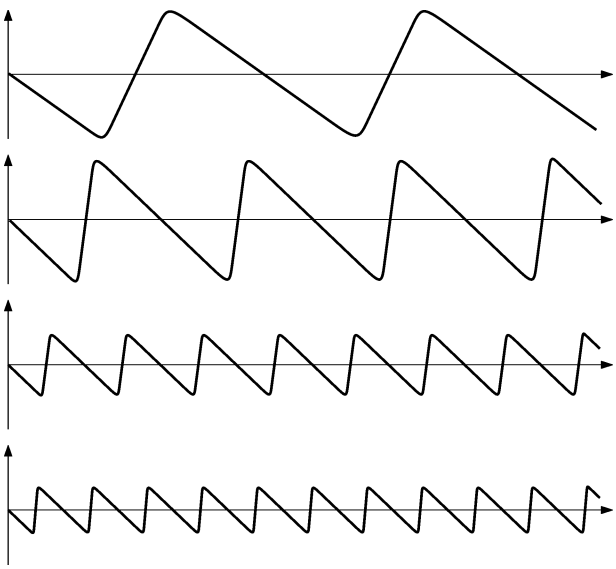


Fig. 1. A sequence of continuous functions for which both $|b(\theta)|^2$ and $|1 + b'(\theta)|^2$ tend to zero for almost every value of θ.

of use when examining the properties of the bound later in the paper.

III. Mathematical Safeguards

As we have seen, a lower bound on the MMSE can be obtained by solving the minimization problem (13). However, at this point, we have no guarantee that the solution s of (13) is anywhere near the true value of the MMSE. Indeed, at first sight, it may appear that $s = 0$ for any estimation setting. To see this, note that $Z[\boldsymbol{b}]$ is a sum of two components, a bias gradient part and a squared bias part. Both parts are nonnegative, but the former is zero when the bias gradient is $-\boldsymbol{I}$, while the latter is zero when the bias is zero. No differentiable function $\boldsymbol{b}$ satisfies these two constraints simultaneously for all $\boldsymbol{\theta}$, since if the squared bias is everywhere zero, then the bias gradient is also zero. However, it is possible to construct a sequence of functions $\boldsymbol{b}^{(i)}$ for which both the bias gradient and the squared bias norm tend to zero for *almost* every value of $\boldsymbol{\theta}$. An example of such a sequence in a one-dimensional setting is plotted in Fig. 1. Here, a sequence $\boldsymbol{b}^{(i)}$ of smooth, periodic functions is presented. The function period tends to zero, and the percentage of the cycle in which the derivative equals -1 increases as i increases. Thus, the pointwise limit of the function sequence is zero almost everywhere (a.e.), and the pointwise limit of the derivative is -1 a.e.

In the specific case depicted in Fig. 1, it can be shown that the value of $Z[\boldsymbol{b}^{(i)}]$ does not tend to zero; in fact, $Z[\boldsymbol{b}^{(i)}]$ tends to infinity in this situation. However, our example illustrates that care must be taken when applying concepts from finite-dimensional optimization problems to variational calculus.

The purpose of this section is to show that $s > 0$, so that the bound is meaningful, for any problem setting satisfying the regularity conditions of Section II-B. (This question was not addressed by Young and Westerberg [5].) While doing so, we develop some abstract concepts which will also be used when analyzing the asymptotic properties of the OBB in Section V.

As often happens with variational problems, it turns out that the minimum of (13) is not necessarily achieved by any continuously differentiable function. In order to guarantee an achievable minimum, one must instead minimize (13) over a slightly modified space, which is defined below. As explained in Section II-B, all bias functions are continuously differentiable, so that the minimizing function ultimately obtained, if it is not differentiable, will not be the bias of any estimator. However, as we will see, the minimum value of our new optimization problem is identical to the infimum of (13). Furthermore, this approach allows us to demonstrate several important theoretical properties of the OBB.

Let L^2 be the space of $p_{\boldsymbol{\theta}}$-measurable functions $\boldsymbol{b} : \Theta \longrightarrow \mathbb{R}^n$ such that

$$\int_{\Theta} \|\boldsymbol{b}(\boldsymbol{\theta})\|^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) < \infty. \tag{14}$$

Define the associated inner product

$$\left\langle \boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)} \right\rangle_{L^2} \triangleq \sum_{i=1}^{n} \int_{\Theta} b_i^{(1)}(\boldsymbol{\theta}) b_i^{(2)}(\boldsymbol{\theta}) p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \tag{15}$$

and the corresponding norm $\|\boldsymbol{b}\|_{L^2}^2 \triangleq \langle \boldsymbol{b}, \boldsymbol{b} \rangle_{L^2}$. Any function $\boldsymbol{b} \in L^2$ has a derivative in the distributional sense, but this derivative might not be a function. For example, discontinuous functions have distributional derivatives which contain a Dirac delta. If, for every i, the distributional derivative $\partial b_i / \partial \boldsymbol{\theta}$ of $\boldsymbol{b}$ is a function in L^2, then $\boldsymbol{b}$ is said to be weakly differentiable [32], and its weak derivative is the matrix function $\partial \boldsymbol{b} / \partial \boldsymbol{\theta}$. Roughly speaking, a function is weakly differentiable if it is continuous and its derivative exists a.e..

The space of all weakly differentiable functions in L^2 is called the first-order Sobolev space [32], and is denoted H^1. Define an inner product on H^1 as

$$\left\langle \boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)} \right\rangle_{H^1} \triangleq \left\langle \boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)} \right\rangle_{L^2} + \sum_{j=1}^{n} \left\langle \frac{\partial b_j^{(1)}}{\partial \boldsymbol{\theta}}, \frac{\partial b_j^{(2)}}{\partial \boldsymbol{\theta}} \right\rangle_{L^2}. \tag{16}$$

The associated norm is $\|\boldsymbol{b}\|_{H^1}^2 \triangleq \langle \boldsymbol{b}, \boldsymbol{b} \rangle_{H^1}$. An important property which will be used extensively in our analysis is that H^1 is a Hilbert space.

Note that since Θ is an open set, not all functions in C^1 are in H^1. For example, in the case $\Theta = \mathbb{R}^n$, the function $\boldsymbol{b}(\boldsymbol{\theta}) = \boldsymbol{k}$, for some nonzero constant $\boldsymbol{k}$, is continuously differentiable but not integrable. Thus, $\boldsymbol{b}$ is in C^1 but not in H^1, nor even in L^2. However, any measurable function which is not in H^1 has $\|\boldsymbol{b}\|_{H^1} = \infty$, meaning that either $\boldsymbol{b}$ or $\partial \boldsymbol{b} / \partial \boldsymbol{\theta}$ has infinite L^2 norm. Consequently, either the bias norm part or the bias gradient part of $Z[\boldsymbol{b}]$ is infinite. It follows that performing the minimization (13) over $C^1 \cap H^1$, rather than over C^1, does not change the minimum value. On the other hand, $C^1 \cap H^1$ is dense in H^1, and $Z[\boldsymbol{b}]$ is continuous, so that minimizing (13) over H^1 rather than $C^1 \cap H^1$ also does not alter the minimum. Consequently, we will henceforth consider the problem

$$s = \inf_{\boldsymbol{b} \in H^1} Z[\boldsymbol{b}]. \tag{17}$$

The advantage of including weakly differentiable functions in the minimization is that a unique minimizer can now be guaranteed, as demonstrated by the following result.

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on November 17, 2009 at 07:06 from IEEE Xplore. Restrictions apply.

Proposition 1: Consider the problem

$$\bar{\boldsymbol{b}} = \arg\min_{\boldsymbol{b}\in H^1} Z[\boldsymbol{b}] \tag{18}$$

where $Z[\boldsymbol{b}]$ is given by (12) and $\boldsymbol{J}(\boldsymbol{\theta})$ is positive definite and bounded with probability 1. This problem is well defined, i.e., there exists a unique $\bar{\boldsymbol{b}} \in H^1$ which minimizes $Z[\boldsymbol{b}]$. Furthermore, the minimum value $s = Z[\bar{\boldsymbol{b}}]$ is finite and nonzero.

Proving the unique existence of a minimizer for (17) is a technical exercise in functional analysis which can be found in Appendix B. However, once the existence of such a minimizer is demonstrated, it is not difficult to see that $0 < s < \infty$. To see that $s < \infty$, we must find a function $\boldsymbol{b}$ for which $Z[\boldsymbol{b}] < \infty$. One such function is $\boldsymbol{b} = \boldsymbol{0}$, for which $Z[\boldsymbol{b}]$ is finite since $\boldsymbol{J}(\boldsymbol{\theta})$ is bounded. Now suppose by contradiction that $s = 0$, which implies that there exists a function $\bar{\boldsymbol{b}} \in H^1$ such that $Z[\bar{\boldsymbol{b}}] = 0$. Therefore, both the bias gradient and the squared bias parts of $Z[\bar{\boldsymbol{b}}]$ are zero. In particular, since the squared bias part equals zero, we have $\|\bar{\boldsymbol{b}}\|_{L^2} = 0$. Hence, $\bar{\boldsymbol{b}} = \boldsymbol{0}$, because L^2 is a normed space. But then, by the definition (12) of $Z[\cdot]$,

$$Z[\bar{\boldsymbol{b}}] = \int_\Theta \mathrm{Tr}(\boldsymbol{J}^{-1}(\boldsymbol{\theta}))p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \tag{19}$$

which is positive; this is a contradiction.

Note that functions in H^1 are defined up to changes on a set having zero measure. In particular, the fact that $\boldsymbol{b}^{(0)}$ is unique does not preclude functions which are identical to $\boldsymbol{b}^{(0)}$ a.e. (which obviously have the same value $Z[\boldsymbol{b}]$).

Summarizing the discussion of the last two sections, we have the following theorem.

Theorem 1: Let $\boldsymbol{\theta}$ be an unknown random vector with pdf $p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) > 0$ over the open 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 the regularity conditions of Section II-B hold. Then, for any estimator $\hat{\boldsymbol{\theta}}$,

$$E\left\{\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2\right\} \geq \min_{\boldsymbol{b}\in H^1} \int_\Theta \mathrm{CRB}[\boldsymbol{b}, \boldsymbol{\theta}]p_{\boldsymbol{\theta}}(\boldsymbol{\theta})d\boldsymbol{\theta}. \tag{20}$$

The minimum in (20) is nonzero and finite. Furthermore, this minimum is achieved by a function $\bar{\boldsymbol{b}} \in H^1$, which is unique up to changes having zero probability.

Two remarks are in order concerning Theorem 1. First, the function $\boldsymbol{b}$ solving (20) might not be the bias of any estimator; indeed, under our assumptions, all bias functions are continuously differentiable, whereas $\boldsymbol{b}$ need only be weakly differentiable. Nevertheless, (20) is still a lower bound on the MMSE. Another important observation is that Theorem 1 arises from the deterministic CRB; hence, there are no requirements on the prior distribution $p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$. In particular, $p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ can be discontinuous or have bounded support. By contrast, many previous Bayesian bounds do not apply in such circumstances.

IV. 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 convex functional optimization problems such as (20), the optimum is given by the solution of a set of differential equations. The following theorem, whose proof can be found in Appendix C, specifies the differential equation relevant to our optimization problem.

In this section and in the remainder of the paper, we will consider the case in which the set $\Theta = \{\boldsymbol{\theta} : p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) > 0\}$ is bounded. From a practical point of view, even when Θ consists of the entire set $\mathbb{R}^n$, it can be approximated by a bounded set containing only those values of $\boldsymbol{\theta}$ for which $p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) > \epsilon$.

Theorem 2: Under the conditions of Theorem 1, suppose Θ is a bounded subset of $\mathbb{R}^n$ with a smooth boundary Λ. Then, the optimal $\boldsymbol{b}(\boldsymbol{\theta})$ of (20) is given by the solution to the system of partial differential equations

$$p_{\boldsymbol{\theta}}(\boldsymbol{\theta})b_i(\boldsymbol{\theta}) = p_{\boldsymbol{\theta}}(\boldsymbol{\theta})\sum_{j,k}\frac{\partial^2 b_i}{\partial\theta_j\partial\theta_k}(\boldsymbol{J}^{-1})_{jk} + \sum_{j,k}\left(\delta_{ik} + \frac{\partial b_i}{\partial\theta_k}\right)\left((\boldsymbol{J}^{-1})_{jk}\frac{\partial p_{\boldsymbol{\theta}}}{\partial\theta_j} + p_{\boldsymbol{\theta}}(\boldsymbol{\theta})\frac{\partial(\boldsymbol{J}^{-1})_{jk}}{\partial\theta_j}\right) \tag{21}$$

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

$$\left(\boldsymbol{I} + \frac{\partial\boldsymbol{b}}{\partial\boldsymbol{\theta}}\right)\boldsymbol{J}^{-1}\boldsymbol{\nu}(\boldsymbol{\theta}) = \boldsymbol{0} \tag{22}$$

for all points $\boldsymbol{\theta} \in \Lambda$. Here, $\boldsymbol{\nu}(\boldsymbol{\theta})$ is a normal to the boundary at $\boldsymbol{\theta}$. All derivatives in this system of equations are to be interpreted in the weak sense.

Note that Theorem 1 guarantees the existence of a unique solution in H^1 to the differential equation (21) with the boundary conditions (22).

The bound of Young and Westerberg [5] 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 bounded interval in $\mathbb{R}$. Then, the bias function $b(\theta)$ minimizing (20) is a solution to the second-order ordinary differential equation

$$J(\theta)b(\theta) = b''(\theta) + (1 + b'(\theta))\left(\frac{d\log p_{\boldsymbol{\theta}}}{d\theta} - \frac{d\log J}{d\theta}\right) \tag{23}$$

within the range $\theta \in \Theta$, subject to the boundary conditions $b'(\theta_0) = b'(\theta_1) = -1$.

Theorem 2 can be solved numerically, thus obtaining a bound for any problem satisfying the regularity conditions. However, directly solving (21) becomes increasingly complex as the dimension of the problem increases. Instead, in many cases, symmetry relations in the problem can be used to simplify the solution. As an example, the following spherically symmetric case can be reduced to a one-dimensional setting equivalent to that of Corollary 1. The proof of this theorem can be found in Appendix D.

Theorem 3: Under the setting of Theorem 1, suppose that $\Theta = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| < r\}$ is a sphere centered on the origin, $p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = q(\|\boldsymbol{\theta}\|)$ is spherically symmetric, and $\boldsymbol{J}(\boldsymbol{\theta}) = J(\|\boldsymbol{\theta}\|)\boldsymbol{I}$, where

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on November 17, 2009 at 07:06 from IEEE Xplore. Restrictions apply.

$J : \mathbb{R} \rightarrow \mathbb{R}$ is a scalar function. Then, the optimal-bias bound (20) is given by

$$E\left\{\|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\|^2\right\} \geq \frac{2\pi^{n/2}}{\Gamma(n/2)}\int_0^r \left[b^2(\rho)+\frac{(1+b'(\rho))^2}{J(\rho)} + \frac{n-1}{J(\rho)}\left(1+\frac{b(\rho)}{\rho}\right)^2\right] q(\rho)\rho^{n-1}d\rho. \quad (24)$$

Here, $\Gamma(\cdot)$ is the Gamma function, and $b(\rho)$ is a solution to the ODE

$$J(\theta)b(\theta) = b''(\theta) + (n-1)\left(\frac{b'(\theta)}{\theta} - \frac{b(\theta)}{\theta^2}\right) + (1+b'(\theta))\left(\frac{d\log q}{d\theta} - \frac{d\log J}{d\theta}\right) \quad (25)$$

subject to the boundary conditions $b(0) = 0, b'(r) = -1$. The bias function for which the bound is achieved is given by

$$\boldsymbol{b}(\boldsymbol{\theta}) = b(\|\boldsymbol{\theta}\|)\frac{\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|}. \quad (26)$$

In this theorem, the requirement $\boldsymbol{J}(\boldsymbol{\theta}) = J(\|\boldsymbol{\theta}\|)\boldsymbol{I}$ indicates that the Fisher information matrix is diagonal and that its components are spherically symmetric. Parameters having a diagonal matrix $\boldsymbol{J}$ are sometimes referred to as *orthogonal*. The simplest case of orthogonality occurs when, to each parameter θ_i, there corresponds a measurement x_i, in such a way that the random variables $x_i|\boldsymbol{\theta}$ are independent. Other orthogonal scenarios can often be constructed by an appropriate parametrization [33].

The requirement that $\boldsymbol{J}$ have spherically symmetric components occurs, for example, in location problems, i.e., situations in which the measurements have the form $\boldsymbol{x} = \boldsymbol{\theta} + \boldsymbol{w}$, where $\boldsymbol{w}$ is additive noise which is independent of $\boldsymbol{\theta}$. Indeed, under such conditions, $\boldsymbol{J}$ is constant in $\boldsymbol{\theta}$ [31, Sec. 3.1.3]. If, in addition, the noise components are independent, then this setting also satisfies the orthogonality requirement, and thus application of Theorem 3 is appropriate. Note that this estimation problem is not separable, since the components of $\boldsymbol{\theta}$ are correlated; thus, the MMSE in this situation is lower than the sum of the components' MMSE. An example of such a setting is presented in Section VI.

V. Properties

In this section, we examine several properties of the OBB. We first demonstrate that the optimal bias function has zero mean, a property which also characterizes the bias function of the MMSE estimator. Next, we prove that, under very general conditions, the resulting bound is tight at both low and high SNR values. This is an important result, since a desirable property of a Bayesian bound is that it provides an accurate estimate of the ambiguity region between high and low SNR [11]. Reliable estimation at the two extremes increases the likelihood that the transition between these two regimes will be correctly identified.

A. Optimal Bias Has Zero Mean

In any Bayesian estimation problem, the bias of the MMSE estimator $\hat{\boldsymbol{\theta}}_{\text{opt}} = E\{\boldsymbol{\theta}|\boldsymbol{x}\}$ has zero mean. Indeed

$$E\left\{\hat{\boldsymbol{\theta}}_{\text{opt}}\right\} = E\{E\{\boldsymbol{\theta}|\boldsymbol{x}\}\} = E\{\boldsymbol{\theta}\} \quad (27)$$

so that

$$E\left\{\boldsymbol{b}(\hat{\boldsymbol{\theta}}_{\text{opt}})\right\} = E\{E\{\boldsymbol{\theta}|\boldsymbol{x}\} - \boldsymbol{\theta}\} = \boldsymbol{0}. \quad (28)$$

Thus, it is interesting to ask whether the optimal bias which minimizes (20) also has zero mean. This does in fact occur, as shown by the following theorem.

Theorem 4: Let $\boldsymbol{b}(\boldsymbol{\theta})$ be the solution to (20). Then

$$E\{\boldsymbol{b}(\boldsymbol{\theta})\} = \boldsymbol{0}. \quad (29)$$

Proof: Assume by contradiction that $\boldsymbol{b}(\boldsymbol{\theta})$ has nonzero mean $E\{\boldsymbol{b}(\boldsymbol{\theta})\} = \boldsymbol{\mu} \neq \boldsymbol{0}$. Define $\boldsymbol{b}_0(\boldsymbol{\theta}) \triangleq \boldsymbol{b}(\boldsymbol{\theta}) - \boldsymbol{\mu}$. From (11), we then have

$$\begin{aligned}\text{CRB}[\boldsymbol{b}_0,\boldsymbol{\theta}] - \text{CRB}[\boldsymbol{b},\boldsymbol{\theta}] &= \|\boldsymbol{b}_0(\boldsymbol{\theta})\|^2 - \|\boldsymbol{b}(\boldsymbol{\theta})\|^2 \\ &= \|\boldsymbol{\mu}\|^2 - 2\boldsymbol{\mu}^T\boldsymbol{b}(\boldsymbol{\theta}).\end{aligned} \quad (30)$$

Using the functional $Z[\cdot]$ defined in (12), we obtain

$$\begin{aligned}Z[\boldsymbol{b}_0] - Z[\boldsymbol{b}] &= E\left\{\|\boldsymbol{\mu}\|^2 - 2\boldsymbol{\mu}^T\boldsymbol{b}(\boldsymbol{\theta})\right\} \\ &= \|\boldsymbol{\mu}\|^2 - 2\boldsymbol{\mu}^T E\{\boldsymbol{b}(\boldsymbol{\theta})\} \\ &= -\|\boldsymbol{\mu}\|^2 < 0.\end{aligned} \quad (31)$$

Thus, $Z[\boldsymbol{b}_0] < Z[\boldsymbol{b}]$, contradicting the fact that $\boldsymbol{b}(\boldsymbol{\theta})$ minimizes (20). □

B. Tightness at Low SNR

Bell *et al.* [11] examined the performance of the extended Ziv–Zakai bound at low SNR and demonstrated that, for a particular family of distributions, the extended Ziv–Zakai bound achieves the MSE of the optimal estimator as the SNR tends to 0. We now examine the low-SNR performance of the OBB, and demonstrate tightness for a much wider range of problem settings.

Bell *et al.* did not define the general meaning of a low-SNR value, and only stated that "[a]s observation time and/or SNR become very small, the observations become useless ... [and] the minimum MSE estimator converges to the *a priori* mean." This statement clearly does not apply to all estimation problems, since it is not always clear what parameter corresponds to the observation time or the SNR. We propose to define the zero-SNR case more generally as any situation in which $\boldsymbol{J}(\boldsymbol{\theta}) = \boldsymbol{0}$ with probability 1. This definition implies that the measurements do not contain information about the unknown parameter, which is the usual informal meaning of zero SNR. In the case $\boldsymbol{J}(\boldsymbol{\theta}) = \boldsymbol{0}$, it can be shown that the MMSE estimator is the prior mean, so that our definition implies the statement of Bell *et al.*.

The OBB is inapplicable when $\boldsymbol{J}(\boldsymbol{\theta}) = \boldsymbol{0}$, since the CRB is based on the assumption that $\boldsymbol{J}(\boldsymbol{\theta})$ is positive definite. To avoid

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on November 17, 2009 at 07:06 from IEEE Xplore. Restrictions apply.

this singularity, we consider a sequence of estimation settings which converge to zero SNR. More specifically, we require all eigenvalues of $\boldsymbol{J}(\boldsymbol{\theta})$ to decrease monotonically to zero for $p_{\boldsymbol{\theta}}$-almost all $\boldsymbol{\theta}$. The following theorem, the proof of which can be found in Appendix E, demonstrates the tightness of the OBB in this low-SNR setting.

Theorem 5: Let $\boldsymbol{\theta}$ be a random vector whose pdf $p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ is nonzero over an open set $\Theta \subseteq \mathbb{R}^n$. Let $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots$ be a sequence of observation vectors having finite Fisher information matrices $\boldsymbol{J}^{(1)}(\boldsymbol{\theta}), \boldsymbol{J}^{(2)}(\boldsymbol{\theta}), \ldots$, respectively. Suppose that, for all N, the matrix $\boldsymbol{J}^{(N)}(\boldsymbol{\theta})$ is positive definite for $p_{\boldsymbol{\theta}}$-almost all $\boldsymbol{\theta}$, and that all eigenvalues of $\boldsymbol{J}^{(N)}(\boldsymbol{\theta})$ decrease monotonically to zero as $N \longrightarrow \infty$ for $p_{\boldsymbol{\theta}}$-almost all $\boldsymbol{\theta}$. Let β_N denote the optimal-bias bound for estimating $\boldsymbol{\theta}$ from $\boldsymbol{x}^{(N)}$. Then

$$\lim_{N \longrightarrow \infty} \beta_N = E\left\{\|\boldsymbol{\theta} - E\{\boldsymbol{\theta}\}\|^2\right\}. \tag{32}$$

C. Tightness at High SNR

We now examine the performance of the OBB for high SNR values. To formally define the high-SNR regime, we consider a sequence of measurements $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots$ of a single parameter vector $\boldsymbol{\theta}$. It is assumed that, when conditioned on $\boldsymbol{\theta}$, all measurements $\boldsymbol{x}^{(i)}$ are independent and identically distributed (i.i.d.). Furthermore, we assume that the Fisher information matrix of a single observation $\bar{\boldsymbol{J}}(\boldsymbol{\theta})$ is well defined, positive definite, and finite for $p_{\boldsymbol{\theta}}$-almost all $\boldsymbol{\theta}$. We consider the problem of estimating $\boldsymbol{\theta}$ from the set of measurements $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}\}$, for a given value of N. The high-SNR regime is obtained when N is large.

When N tends to infinity, the MSE of the optimal estimator tends to zero. An important question, however, concerns the rate of convergence of the minimum MSE. More precisely, given the optimal estimator $\hat{\boldsymbol{\theta}}^{(N)}$ of $\boldsymbol{\theta}$ from $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}\}$, one would like to determine the asymptotic distribution of $\sqrt{N}(\hat{\boldsymbol{\theta}}^{(N)} - \boldsymbol{\theta})$, conditioned on $\boldsymbol{\theta}$. A fundamental result of asymptotic estimation theory can be loosely stated as follows [28, Sec. III.3], [13, Sec. 6.8]. Under some fairly mild regularity conditions, the asymptotic distribution of $\sqrt{N}(\hat{\boldsymbol{\theta}}^{(N)} - \boldsymbol{\theta})$, conditioned on $\boldsymbol{\theta}$, does not depend on the prior distribution $p_{\boldsymbol{\theta}}$; rather, $\sqrt{N}(\hat{\boldsymbol{\theta}}^{(N)} - \boldsymbol{\theta}) \mid \boldsymbol{\theta}$ converges in distribution to a Gaussian random vector with mean zero and covariance $\boldsymbol{J}^{-1}(\boldsymbol{\theta})$. It follows that

$$\lim_{N \longrightarrow \infty} NE\left\{\|\hat{\boldsymbol{\theta}}^{(N)} - \boldsymbol{\theta}\|^2\right\} = E\left\{\mathrm{Tr}[\boldsymbol{J}^{-1}(\boldsymbol{\theta})]\right\}. \tag{33}$$

Since the minimum MSE tends to zero at high SNR, any lower bound on the minimum MSE must also tend to zero as $N \longrightarrow \infty$. However, one would further expect a good lower bound to follow the behavior of (33). In other words, if β_N represents the lower bound for estimating $\boldsymbol{\theta}$ from $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}\}$, a desirable property is $N\beta_N \longrightarrow E\{\mathrm{Tr}[\boldsymbol{J}^{-1}(\boldsymbol{\theta})]\}$. The following theorem, whose proof is found in Appendix E, demonstrates that this is indeed the case for the OBB.

Except for a very brief treatment by Bellini and Tartara [6], no previous Bayesian bound has shown such a result. Although it appears that the Ziv–Zakai and Weiss–Weinstein bounds may also satisfy this property, this has not been proven formally. It is also known that the Bayesian CRB is *not* asymptotically tight in this sense [34, eqs. (37)–(39)].

Theorem 6: Let $\boldsymbol{\theta}$ be a random vector whose pdf $p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ is nonzero over an open set $\Theta \subseteq \mathbb{R}^n$. Let $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots$ be a sequence of measurement vectors, such that $\boldsymbol{x}^{(1)}|\boldsymbol{\theta}, \boldsymbol{x}^{(2)}|\boldsymbol{\theta}, \ldots$ are i.i.d. Let $\boldsymbol{J}(\boldsymbol{\theta})$ be the Fisher information matrix for estimating $\boldsymbol{\theta}$ from $\boldsymbol{x}^{(1)}$, and suppose $\boldsymbol{J}(\boldsymbol{\theta})$ is finite and positive definite for $p_{\boldsymbol{\theta}}$-almost all $\boldsymbol{\theta}$. Let β_N be the optimal-bias bound (20) for estimating $\boldsymbol{\theta}$ from the observation sequence $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}\}$. Then

$$\lim_{N \longrightarrow \infty} N\beta_N = E\left\{\mathrm{Tr}(\boldsymbol{J}^{-1}(\boldsymbol{\theta}))\right\}. \tag{34}$$

Note that for Theorem 6 to hold, we require only that $\boldsymbol{J}(\boldsymbol{\theta})$ be finite and positive definite. By contrast, the various theorems guaranteeing asymptotic efficiency of Bayesian estimators all require substantially stronger regularity conditions [28, Sec. III.3], [13, Sec. 6.8]. One reason for this is that asymptotic efficiency describes the behavior of $\hat{\boldsymbol{\theta}}$ conditioned on each possible value of $\boldsymbol{\theta}$, and is thus a stronger result than the asymptotic Bayesian MSE of (33).

VI. Example: Uniform Prior

The original bound of Young and Westerberg [5] predates most Bayesian bounds, and, surprisingly, it has never been cited by or compared with later results. In this section, we measure the performance of the original bound and of its extension to the vector case against that of various other techniques. We consider the case in which $\boldsymbol{\theta}$ is uniformly distributed over an n-dimensional open ball $\Theta = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| < r\} \subseteq \mathbb{R}^n$, so that

$$p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \frac{1}{V_n(r)} \mathbb{1}_{\Theta} \tag{35}$$

where $\mathbb{1}_S$ equals 1 when $\boldsymbol{\theta} \in S$ and 0 otherwise, and

$$V_n(r) = \frac{\pi^{n/2} r^{n-1}}{\Gamma\left(1 + \frac{n}{2}\right)} \tag{36}$$

is the volume of an n-ball of radius r [35]. We further assume that

$$\boldsymbol{x} = \boldsymbol{\theta} + \boldsymbol{w} \tag{37}$$

where $\boldsymbol{w}$ is zero-mean Gaussian noise, independent of $\boldsymbol{\theta}$, having covariance $\sigma^2 \boldsymbol{I}$. We are interested in lower bounds on the MSE achievable by an estimator of $\boldsymbol{\theta}$ from $\boldsymbol{x}$.

We begin by developing the OBB for this setting, as well as some alternative bounds. We then compare the different approaches in a one-dimensional and a three-dimensional setting.

The Fisher information matrix for the given estimation problem is given by $\boldsymbol{J}(\boldsymbol{\theta}) = \sigma^{-2}\boldsymbol{I}$, so that the conditions of Theorem 3 hold. It follows that the optimal bias function is given by $\boldsymbol{b}(\boldsymbol{\theta}) = b(\|\boldsymbol{\theta}\|)\boldsymbol{\theta}/\|\boldsymbol{\theta}\|$, where $b(\cdot)$ is a solution to the differential equation

$$\frac{b}{\sigma^2} = b'' + (n-1)\left(\frac{b'}{\theta} - \frac{b}{\theta^2}\right) \tag{38}$$

with boundary conditions $b(0) = 0$, $b'(r) = -1$. The general solution to this differential equation is given by

$$b(\theta) = C_1 \theta^{1-n/2} I_{n/2}\left(\frac{\theta}{\sigma}\right) + C_2 \theta^{1-n/2} K_{n/2}\left(\frac{\theta}{\sigma}\right) \tag{39}$$

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on November 17, 2009 at 07:06 from IEEE Xplore. Restrictions apply.

where $I_\alpha(z)$ and $K_\alpha(z)$ are the modified Bessel functions of the first and second types, respectively [36]. Since $K_\alpha(z)$ is singular at the origin, the requirement $b(0)=0$ leads to $C_2=0$. Differentiating (39) with respect to θ, we obtain

$$b'(\theta) = C_1\theta^{-n/2}\left(I_{n/2}\left(\frac{\theta}{\sigma}\right) + \frac{\theta}{\sigma}I_{1+n/2}\left(\frac{\theta}{\sigma}\right)\right) \tag{40}$$

so that the requirement $b'(r)=-1$ leads to

$$C_1 = -\frac{r^{n/2}}{I_{n/2}\left(\frac{r}{\sigma}\right) + \frac{r}{\sigma}I_{1+n/2}\left(\frac{r}{\sigma}\right)}. \tag{41}$$

Substituting this value of $b(\cdot)$ into (24) yields the OBB, which can be computed by evaluating a single one-dimensional integral. Alternatively, in the one-dimensional case, the integral can be computed analytically, as will be shown below.

Despite the widespread use of finite-support prior distributions [4], [10], the regularity conditions of many bounds are violated by such prior pdf functions. Indeed, the Bayesian CRB of Van Trees [3], the Bobrovski–Zakai bound [8], and the Bayesian Abel bound [12] all assume that $p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ has infinite support, and thus cannot be applied in this scenario.

Techniques from the Ziv–Zakai family are applicable to constrained problems. An extension of the Ziv–Zakai bound for vector parameter estimation was developed by Bell *et al.* [11]. From [11, Property 4], the MSE of the ith component of $\boldsymbol{\theta}$ is bounded by

$$E\left\{(\theta_i-\hat{\theta}_i)^2\right\} \geq \int_0^\infty V\left\{\max_{\boldsymbol{\delta}:\boldsymbol{e}_i^T\boldsymbol{\delta}=h} A(\boldsymbol{\delta})P_{\min}(\boldsymbol{\delta})\right\} h\,dh \tag{42}$$

where $\boldsymbol{e}_i$ is a unit vector in the direction of the ith component, $V\{\cdot\}$ is the valley-filling function defined by

$$V\{f(h)\} = \max_{\eta\geq 0} f(h+\eta), \tag{43}$$

$$A(\boldsymbol{\delta}) \triangleq \int_{\mathbb{R}^n} \min\left(p_{\boldsymbol{\theta}}(\boldsymbol{\theta}), p_{\boldsymbol{\theta}}(\boldsymbol{\theta}+\boldsymbol{\delta})\right) d\boldsymbol{\theta}, \tag{44}$$

and $P_{\min}(\boldsymbol{\delta})$ is the minimum probability of error for the problem of testing hypothesis $H_0:\boldsymbol{\theta}=\boldsymbol{\theta}_0$ versus $H_1:\boldsymbol{\theta}=\boldsymbol{\theta}_0+\boldsymbol{\delta}$. In the current setting, $P_{\min}(\boldsymbol{\delta})$ is given by $P_{\min}(\boldsymbol{\delta}) = Q\left(\|\boldsymbol{\delta}\|/2\sigma\right)$, where $Q(z) = (2\pi)^{-1/2}\int_z^\infty e^{-t^2/2}dt$ is the tail function of the normal distribution. Also, in the present setting we have

$$A(\boldsymbol{\delta}) = \frac{V_n^C(r,\|\boldsymbol{\delta}\|)}{V_n(r)} \tag{45}$$

where

$$V_n^C(r,h) = \int_{\mathbb{R}^n} \mathbb{1}_\Theta \mathbb{1}_{\Theta+h\boldsymbol{e}_1} d\boldsymbol{\theta} \tag{46}$$

and $\Theta + h\boldsymbol{e}_1 = \{\boldsymbol{\theta}+h\boldsymbol{e}_1 : \boldsymbol{\theta}\in\Theta\}$. Thus, $V_n^C(r,h)$ is the volume of the intersection of two n-balls whose centers are at a distance of h units from one another. Substituting these results into (42), we have

$$E\left\{(\theta_i-\hat{\theta}_i)^2\right\} \geq \int_0^\infty V\left\{\max_{\boldsymbol{\delta}:\boldsymbol{e}_i^T\boldsymbol{\delta}=h} \frac{V_n^C(r,\|\boldsymbol{\delta}\|)}{V_n(r)} Q\left(\frac{\|\boldsymbol{\delta}\|}{2\sigma}\right)\right\} h\,dh. \tag{47}$$

Note that both $V_n^C(r,\|\boldsymbol{\delta}\|)$ and $Q(\|\boldsymbol{\delta}\|/2\sigma)$ decrease with $\|\boldsymbol{\delta}\|$. Therefore, the maximum in (47) is obtained for $\boldsymbol{\delta}=h\boldsymbol{e}_i$. Also, since the argument of $V\{\cdot\}$ is monotonically decreasing, the valley-filling function has no effect and can be removed. Finally, since $V_n^C(r,h)=0$ for $h>2r$, the integration can be limited to the range $[0,2r]$. Thus, the extended Ziv–Zakai bound is given by

$$E\left\{\|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\|^2\right\} \geq \int_0^{2r} n\frac{V_n^C(r,h)}{V_n(r)} Q\left(\frac{h}{2\sigma}\right) h\,dh. \tag{48}$$

We now compute the Weiss–Weinstein bound for the setting at hand. This bound is given by

$$E\left\{\|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\|^2\right\} \geq \mathrm{Tr}(\boldsymbol{H}\boldsymbol{G}^{-1}\boldsymbol{H}^T) \tag{49}$$

where $\boldsymbol{H} = [\boldsymbol{h}_1,\ldots,\boldsymbol{h}_m]$ is a matrix containing an arbitrary number m of test vectors and $\boldsymbol{G}$ is a matrix whose elements are given by

$$G_{ij} = \frac{E\left\{r(\boldsymbol{x},\boldsymbol{\theta};\boldsymbol{h}_i,s_i)r(\boldsymbol{x},\boldsymbol{\theta};\boldsymbol{h}_j,s_j)\right\}}{E\left\{L^{s_i}(\boldsymbol{x};\boldsymbol{\theta}+\boldsymbol{h}_i,\boldsymbol{\theta})\right\}E\left\{L^{s_j}(\boldsymbol{x};\boldsymbol{\theta}+\boldsymbol{h}_j,\boldsymbol{\theta})\right\}} \tag{50}$$

in which

$$r(\boldsymbol{x},\boldsymbol{\theta};\boldsymbol{h}_i,s_i) \triangleq L^{s_i}(\boldsymbol{x};\boldsymbol{\theta}+\boldsymbol{h}_i,\boldsymbol{\theta}) - L^{1-s_i}(\boldsymbol{x};\boldsymbol{\theta}-\boldsymbol{h}_i,\boldsymbol{\theta}) \tag{51}$$

and

$$L(\boldsymbol{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \triangleq \frac{p_{\boldsymbol{\theta}}(\boldsymbol{\theta}_1)p_{\boldsymbol{x}|\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{\theta}_1)}{p_{\boldsymbol{\theta}}(\boldsymbol{\theta}_2)p_{\boldsymbol{x}|\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{\theta}_2)}. \tag{52}$$

The vectors $\boldsymbol{h}_1,\ldots,\boldsymbol{h}_m$ and the scalars $s_1,\ldots,s_m$ are arbitrary, and can be optimized to maximize the bound (49). To avoid a multidimensional nonconvex optimization problem, we restrict attention to $m=n$, $\boldsymbol{h}_i = h\boldsymbol{e}_i$, and $s_i = 1/2$, as suggested by [10]. This results in a dependency on a single scalar parameter h.

Under these conditions, G_{ij} can be written as

$$G_{ij} = \frac{1}{M(\boldsymbol{h}_i)M(\boldsymbol{h}_j)}\left[\tilde{M}(\boldsymbol{h}_i-\boldsymbol{h}_j,-\boldsymbol{h}_j) + \tilde{M}(\boldsymbol{h}_i-\boldsymbol{h}_j,\boldsymbol{h}_i) - \tilde{M}(\boldsymbol{h}_i+\boldsymbol{h}_j,\boldsymbol{h}_j) - \tilde{M}(\boldsymbol{h}_i+\boldsymbol{h}_j,\boldsymbol{h}_i)\right] \tag{53}$$

where

$$M(\boldsymbol{h}) \triangleq E\left\{L^{1/2}(\boldsymbol{x};\boldsymbol{\theta}+\boldsymbol{h},\boldsymbol{\theta})\right\} \tag{54}$$

and

$$\tilde{M}(\boldsymbol{h}_1,\boldsymbol{h}_2) \triangleq E\left\{L^{1/2}(\boldsymbol{x};\boldsymbol{\theta}+\boldsymbol{h}_1,\boldsymbol{\theta})\mathbb{1}_{\Theta+\boldsymbol{h}_2}\right\}. \tag{55}$$

Note that we have used the corrected version of the Weiss–Weinstein bound [37]. Substituting the probability distribution of $\boldsymbol{x}$ and $\boldsymbol{\theta}$ into the definitions of $M(\boldsymbol{h})$ and $\tilde{M}(\boldsymbol{h}_1,\boldsymbol{h}_2)$, we have

$$\begin{aligned} M(\boldsymbol{h}) &= E\left\{e^{-\|\boldsymbol{\theta}+\boldsymbol{h}-\boldsymbol{x}\|^2/4\sigma^2}e^{\|\boldsymbol{\theta}-\boldsymbol{x}\|^2/4\sigma^2}\mathbb{1}_{\Theta+\boldsymbol{h}}\right\} \\ &= \frac{V_n^C(r,\|\boldsymbol{h}\|)}{V_n(r)}e^{-\|\boldsymbol{h}\|^2/8\sigma^2} \end{aligned} \tag{56}$$

and similarly

$$\tilde{M}(\boldsymbol{h}_1,\boldsymbol{h}_2) = \frac{e^{-\|\boldsymbol{h}_1\|^2/8\sigma^2}}{V_n(r)}\int \mathbb{1}_\Theta \mathbb{1}_{\Theta+\boldsymbol{h}_1}\mathbb{1}_{\Theta+\boldsymbol{h}_2} d\boldsymbol{\theta}. \tag{57}$$

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on November 17, 2009 at 07:06 from IEEE Xplore. Restrictions apply.

Thus, $M(\boldsymbol{h})$ is a function only of $\|\boldsymbol{h}\|$, and $\tilde{M}(\boldsymbol{h}_1, \boldsymbol{h}_2)$ is a function only of $\|\boldsymbol{h}_1\|$, $\|\boldsymbol{h}_2\|$, and $\|\boldsymbol{h}_1 - \boldsymbol{h}_2\|$. Since $\boldsymbol{h}_i = h\boldsymbol{e}_i$, it follows that, for $i \neq j$, the numerator of (53) vanishes. Thus, $\boldsymbol{G}$ is a diagonal matrix, whose diagonal elements equal

$$G_{ii} = 2\frac{\tilde{M}(0, h\boldsymbol{e}_1) - \tilde{M}(2h\boldsymbol{e}_1, h\boldsymbol{e}_1)}{M^2(h\boldsymbol{e}_1)}. \tag{58}$$

The Weiss–Weinstein bound is given by substituting this result into (49) and maximizing over h, i.e.,

$$E\left\{\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2\right\} \geq \max_{h\in[0,2r]} \frac{nh^2 M^2(h\boldsymbol{e}_1)}{2[\tilde{M}(0, h\boldsymbol{e}_1) - \tilde{M}(2h\boldsymbol{e}_1, h\boldsymbol{e}_1)]}. \tag{59}$$

The value of h yielding the tightest bound can be determined by performing a grid search.

To compare the OBB with the alternative approaches developed above, we first consider the one-dimensional case in which θ is uniformly distributed in the range $\Theta = (-r, r)$. Let $x = \theta + w$ be a single noisy observation, where w is zero-mean Gaussian noise, independent of θ, with variance σ^2. We wish to bound the MSE of an estimator of θ from x.

The optimal bias function is given by (39). Using the fact that $I_{1/2}(t) = \sqrt{2/\pi}\sinh(t)/\sqrt{t}$, we obtain

$$b(\theta) = -\sigma\frac{\sinh\left(\frac{\theta}{\sigma}\right)}{\cosh\left(\frac{r}{\sigma}\right)} \tag{60}$$

which also follows [5] from Corollary 1. Substituting this expression into (20), we have that, for any estimator $\hat{\theta}$

$$E\left\{(\theta - \hat{\theta})^2\right\} \geq \sigma^2\left(1 - \frac{\tanh\left(\frac{r}{\sigma}\right)}{\frac{r}{\sigma}}\right). \tag{61}$$

Apart from the reduction in computational complexity, the simplicity of (61) 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, the asymptotic properties demonstrated in Theorems 5 and 6 can be easily verified. For $r \gg \sigma$, the bound converges to the noise variance σ^2, corresponding to an uninformative prior whose optimal estimator is $\hat{\theta} = x$; whereas, for $\sigma \gg r$, a Taylor expansion of $\tanh(z)/z$ immediately shows that 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 (61) is tight both for very low and for very high SNR, as expected.

In the one-dimensional case, we have $V_1(r) = 2r$ and $V_1^C(r, h) = \max(2r - h, 0)$, so that the extended Ziv–Zakai bound (48) and the Weiss–Weinstein bound (59) can also be simplified somewhat. In particular, the extended Ziv–Zakai bound (48) can be written as

$$E\left\{\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2\right\} \geq \int_0^{2r}\left(1 - \frac{h}{2r}\right)hQ\left(\frac{h}{2\sigma}\right)dh. \tag{62}$$

Using integration by parts, (62) becomes

$$E\left\{\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2\right\} \geq \frac{2r^2}{3}Q\left(\frac{r}{\sigma}\right) + \sigma^2\left[\Gamma_{3/2}\left(\frac{r^2}{2\sigma^2}\right) - \frac{8}{3\sqrt{2\pi}}\frac{\sigma}{r}\Gamma_2\left(\frac{r^2}{2\sigma^2}\right)\right] \tag{63}$$

where $\Gamma_a(z) = (1/\Gamma(a))\int_0^z e^{-t}t^{a-1}dt$ is the incomplete Gamma function. Like the expression (61) for the OBB, this bound can be shown to converge to the noise variance σ^2 when $r \gg \sigma$ and to the prior variance $r^2/3$ when $\sigma \gg r$. However, while the convergence of the OBB to these asymptotic values has been demonstrated in general in Theorems 5 and 6, the asymptotic tightness of the Ziv–Zakai bound in the general case remains an open question.

The Weiss–Weinstein bound (59) can likewise be simplified further in the one-dimensional case, yielding

$$E\left\{\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2\right\} \geq \max_{h\in[0,2r]} \frac{h^2 e^{-h^2/4\sigma^2}\left(1 - \frac{h}{2r}\right)^2}{2\left(1 - \frac{h}{2r} - \max\left(0, 1 - \frac{h}{r}\right)e^{-h^2/2\sigma^2}\right)}. \tag{64}$$

However, calculating this bound still requires a numerical search for the optimal value of h.

These bounds are compared with the exact value of the MMSE in Fig. 2. In this figure, the SNR is defined as

$$\begin{aligned}\text{SNR (dB)} &= 10\log_{10}\left(\frac{\mathrm{Var}(\theta)}{\mathrm{Var}(w)}\right)\\ &= 10\log_{10}\left(\frac{r^2}{3\sigma^2}\right).\end{aligned} \tag{65}$$

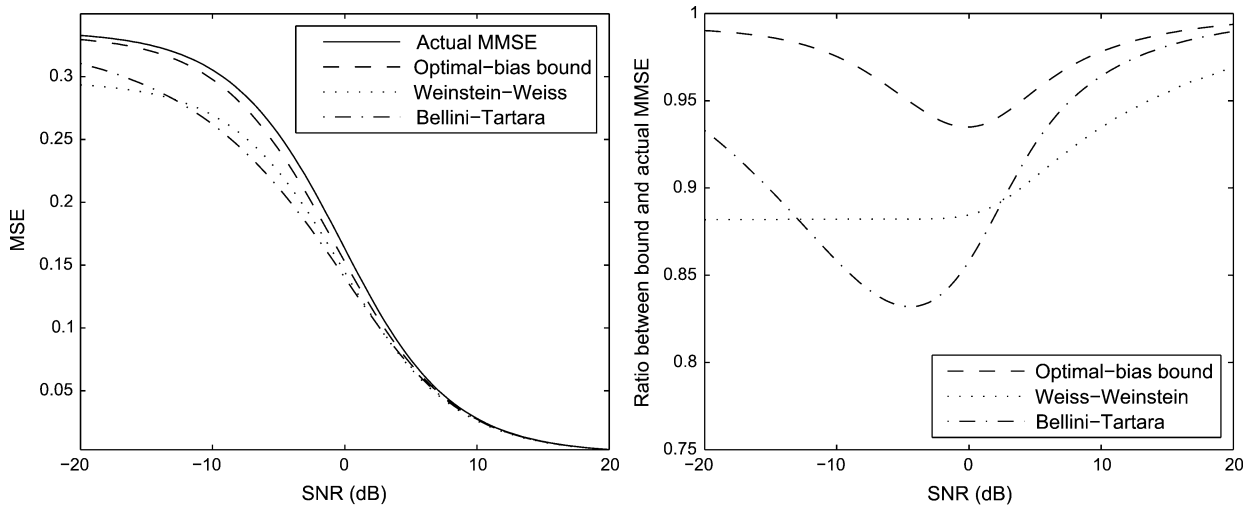
The MMSE was computed by Monte Carlo approximation of the error of the optimal estimator $E\{\theta|x\}$, which was itself computed by numerical integration. Fig. 2(a) plots the MMSE and the values obtained by the aforementioned bounds, while Fig. 2(b) plots the ratio between each of the bounds and the actual MMSE in order to emphasize the difference in accuracy between the various bounds. As can be seen from this figure, the OBB is closer to the true MSE than all other bounds, for all tested SNR values.

The improvements provided by the OBB continue to hold in higher dimensions as well, although in this case it is not possible to provide a closed form for any of the bounds. For example, Fig. 3 compares the aforementioned bounds with the true MMSE in the three-dimensional case. In this case, the SNR is given by

$$\begin{aligned}\text{SNR (dB)} &= 10\log_{10}\left(\frac{\mathrm{Var}(\boldsymbol{\theta})}{\mathrm{Var}(\boldsymbol{w})}\right)\\ &= 10\log_{10}\left(\frac{r^2}{5\sigma^2}\right).\end{aligned} \tag{66}$$

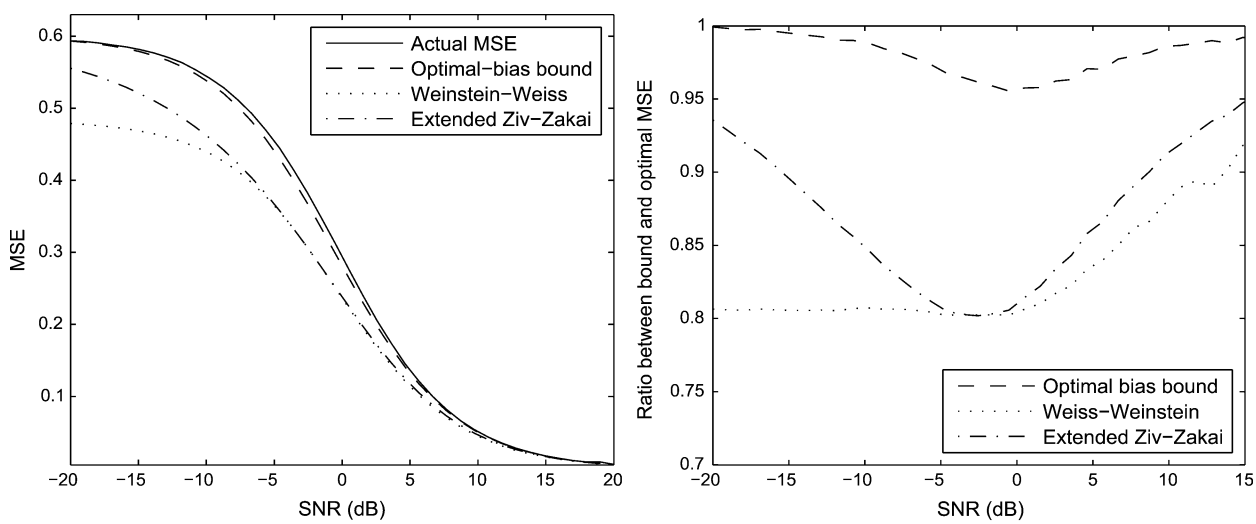
Here, computation of the minimum MSE requires multidimensional numerical integration, and is by far more computationally complex than the calculation of the bounds. Again, it is evident from this figure that the OBB is a very tight bound in all ranges

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on November 17, 2009 at 07:06 from IEEE Xplore. Restrictions apply.



(a) (b)

Fig. 2. Comparison of the MSE bounds and the minimum achievable MSE in a one-dimensional setting for which $\theta \sim U[-r, r]$ and $x|\theta \sim N(\theta, \sigma^2)$.



(a) (b)

Fig. 3. Comparison of the MSE bounds and the minimum achievable MSE in a three-dimensional setting for which $\boldsymbol{\theta}$ is uniformly distributed over a ball of radius r and $\boldsymbol{x}|\boldsymbol{\theta} \sim N(\boldsymbol{\theta}, \sigma^2 \boldsymbol{I})$.

of operation, and is considerably closer to the true value than either of the alternative approaches.

VII. 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 application of this deterministic bound to the problem of estimating the minimum Bayesian MSE results in a Bayesian bound which is provably tight at both high and low SNR values. Numerical simulation of the location estimation problem demonstrates that the technique is both simpler and tighter than alternative approaches.

Appendix A
Some Technical Lemmas

The proof of several theorems in the paper relies on the following technical results.

Lemma 1: Consider the minimization problems

$$M_\ell = \inf_{\boldsymbol{b} \in S} Z_\ell[\boldsymbol{b}], \quad \ell = 1, 2, 3 \tag{67}$$

where $\boldsymbol{J}(\boldsymbol{\theta})$ is positive definite and bounded a.e. $(p_{\boldsymbol{\theta}})$

$$\begin{aligned} Z_1[\boldsymbol{b}] &\triangleq \int_\Theta \|\boldsymbol{b}(\boldsymbol{\theta})\|^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \\ Z_2[\boldsymbol{b}] &\triangleq \int_\Theta \mathrm{Tr}\left(\left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}} \right) \boldsymbol{J}^{-1}(\boldsymbol{\theta}) \left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}} \right)^T \right) p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \\ Z_3[\boldsymbol{b}] &\triangleq Z_1[\boldsymbol{b}] + Z_2[\boldsymbol{b}] \end{aligned} \tag{68}$$

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on November 17, 2009 at 07:06 from IEEE Xplore. Restrictions apply.

and $S \subset H^1$ is convex, closed, and bounded under the H^1 norm (16). Then, for each ℓ, there exists a function $\boldsymbol{b}^{(0)} \in S$ such that $Z[\boldsymbol{b}^{(0)}] = M_\ell$. If $\ell = 1$ or $\ell = 3$, then the minimizer of (67) is unique.

Note that $Z_3[\boldsymbol{b}]$ equals $Z[\boldsymbol{b}]$ of (12); the notation $Z_3[\boldsymbol{b}]$ is introduced for simplicity. Also note that under mild regularity assumptions on $\boldsymbol{J}(\boldsymbol{\theta})$, uniqueness can be demonstrated for $\ell = 2$ as well, but this is not necessary for our purposes.

Proof: The space H^1 is a Cartesian product of n Sobolev spaces $H^1(\Theta)$, each of which is a separable Hilbert space [38, Section 3.7.1]. Therefore, H^1 is also a separable Hilbert space. It follows from the Banach-Alaoglu theorem [39, Section 3.17] that all bounded sequences in H^1 have weakly convergent subsequences [32, Section 2.18]. Recall that a sequence $\boldsymbol{f}^{(1)}, \boldsymbol{f}^{(2)}, \ldots \in H^1$ is said to converge weakly to $\boldsymbol{f}^{(0)} \in H^1$ (denoted $\boldsymbol{f}^{(i)} \rightharpoonup \boldsymbol{f}^{(0)}$) if

$$L[\boldsymbol{f}^{(j)}] \longrightarrow L[\boldsymbol{f}^{(0)}] \tag{69}$$

for all continuous linear functionals $L[\cdot]$ [32, Section 2.9].

Given a particular value $\ell \in \{1,2,3\}$, let $\boldsymbol{b}^{(i)}$ be a sequence of functions in S such that $Z_\ell[\boldsymbol{b}^{(i)}] \longrightarrow M_\ell$. This is a bounded sequence since S is bounded, and therefore there exists a subsequence $\boldsymbol{b}^{(i_k)}$ which converges weakly to some $\boldsymbol{b}_{\text{opt}}^{(\ell)} \in H^1$. Furthermore, since S is closed,[2] we have $\boldsymbol{b}_{\text{opt}}^{(\ell)} \in S$. We will now prove that $Z_\ell[\boldsymbol{b}_{\text{opt}}^{(\ell)}] = M_\ell$.

To this end, we will first show that $Z_\ell[\cdot]$ is weakly lower semicontinuous, i.e., for any sequence $\boldsymbol{f}^{(i)} \in H^1$ which converges weakly to $\boldsymbol{f}^{(0)} \in H^1$, we must show that

$$Z_\ell[\boldsymbol{f}^{(0)}] \le \liminf_{i \longrightarrow \infty} Z_\ell[\boldsymbol{f}^{(i)}]. \tag{70}$$

Consider a weakly convergent sequence $\boldsymbol{f}^{(j)} \rightharpoonup \boldsymbol{f}^{(0)}$. Then, (69) holds for any continuous linear functional $L[\cdot]$. Specifically, choose the continuous linear functional

$$L_1[\boldsymbol{f}] = \int_\Theta \boldsymbol{f}^{(0)}(\boldsymbol{\theta}) \boldsymbol{f}(\boldsymbol{\theta}) p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}). \tag{71}$$

We then have

$$\begin{aligned} Z_1[\boldsymbol{f}^{(0)}] &= L_1[\boldsymbol{f}^{(0)}] \\ &= \lim_{j \longrightarrow \infty} L_1[\boldsymbol{f}^{(j)}] \\ &= \lim_{j \longrightarrow \infty} \int_\Theta \sum_{i=1}^n f_i^{(0)}(\boldsymbol{\theta}) f_i^{(j)}(\boldsymbol{\theta}) p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \\ &\le \liminf_{j \longrightarrow \infty} \sqrt{\int_\Theta \|\boldsymbol{f}^{(0)}(\boldsymbol{\theta})\|^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \cdot \int_\Theta \|\boldsymbol{f}^{(j)}(\boldsymbol{\theta})\|^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta})} \\ &= \sqrt{Z_1[\boldsymbol{f}^{(0)}]} \liminf_{j \longrightarrow \infty} \sqrt{Z_1[\boldsymbol{f}^{(j)}]} \end{aligned} \tag{72}$$

where we have used the Cauchy–Schwarz inequality. It follows that

$$\sqrt{Z_1[\boldsymbol{f}^{(0)}]} \le \liminf_{j \longrightarrow \infty} \sqrt{Z_1[\boldsymbol{f}^{(j)}]} \tag{73}$$

and therefore $Z_1[\boldsymbol{f}^{(0)}] \le \liminf_{j \longrightarrow \infty} Z_1[\boldsymbol{f}^{(j)}]$, so that $Z_1[\cdot]$ is weakly lower semicontinuous.

Similarly, consider the continuous linear functional

$$L_2[\boldsymbol{f}] = \int_\Theta \operatorname{Tr}\left(\left(\boldsymbol{I} + \frac{\partial \boldsymbol{f}^{(0)}}{\partial \boldsymbol{\theta}}\right) \boldsymbol{J}^{-1}(\boldsymbol{\theta}) \left(\boldsymbol{I} + \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\theta}}\right)^T\right) p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \tag{74}$$

for which we have

$$\begin{aligned} Z_2[\boldsymbol{f}^{(0)}] &= L_2[\boldsymbol{f}^{(0)}] \\ &= \lim_{j \longrightarrow \infty} L_2[\boldsymbol{f}^{(j)}] \\ &= \lim_{j \longrightarrow \infty} \int_\Theta \operatorname{Tr}\left[\left(\boldsymbol{I} + \frac{\partial \boldsymbol{f}^{(0)}}{\partial \boldsymbol{\theta}}\right) \boldsymbol{J}^{-1}(\boldsymbol{\theta}) \cdot \left(\boldsymbol{I} + \frac{\partial \boldsymbol{f}^{(j)}}{\partial \boldsymbol{\theta}}\right)^T\right] p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}). \end{aligned} \tag{75}$$

Note that, for any positive definite matrix $\boldsymbol{W}$, $\operatorname{Tr}(\boldsymbol{A}\boldsymbol{W}\boldsymbol{B}^T)$ is an inner product of the two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. Therefore, by the Cauchy–Schwarz inequality

$$\operatorname{Tr}(\boldsymbol{A}\boldsymbol{W}\boldsymbol{B}^T) \le \sqrt{\operatorname{Tr}(\boldsymbol{A}\boldsymbol{W}\boldsymbol{A}^T)\operatorname{Tr}(\boldsymbol{B}\boldsymbol{W}\boldsymbol{B}^T)}. \tag{76}$$

Applying this to (75), we get (77) shown at the bottom of the page. Once again using the Cauchy–Schwarz inequality results in

$$Z_2[\boldsymbol{f}^{(0)}] \le \liminf_{j \longrightarrow \infty} \sqrt{Z_2[\boldsymbol{f}^{(0)}] Z_2[\boldsymbol{f}^{(j)}]} \tag{78}$$

and therefore $Z_2[\boldsymbol{f}^{(0)}] \le \liminf_{j \longrightarrow \infty} Z_2[\boldsymbol{f}^{(j)}]$, so that $Z_2[\cdot]$ is weakly lower semicontinuous. Since $Z_3[\boldsymbol{f}] = Z_1[\boldsymbol{f}] + Z_2[\boldsymbol{f}]$, it follows that $Z_3[\cdot]$ is also weakly lower semicontinuous.

Now recall that $\boldsymbol{b}^{(i_k)} \rightharpoonup \boldsymbol{b}_{\text{opt}}^{(\ell)}$ and $Z_\ell[\boldsymbol{b}^{(i_k)}] \longrightarrow M_\ell$. By the definition (70) of lower semicontinuity, it follows that

$$Z_\ell[\boldsymbol{b}_{\text{opt}}^{(\ell)}] \le \liminf_{k \longrightarrow \infty} Z_\ell[\boldsymbol{b}^{(i_k)}] = M_\ell \tag{79}$$

and since M_ℓ is the infimum of $Z_\ell[\boldsymbol{b}]$, we obtain $Z[\boldsymbol{b}_{\text{opt}}^{(\ell)}] = M$. Thus, $\boldsymbol{b}_{\text{opt}}^{(\ell)}$ is a minimizer of (67).

[2]In fact, we require that S be "weakly closed" in the sense that weakly convergent sequences in S converge to an element in S. However, since S is convex, this notion is equivalent to the ordinary definition of closure [39, Section 3.13].

$$Z_2[\boldsymbol{f}^{(0)}] \le \liminf_{j \longrightarrow \infty} \int_\Theta \sqrt{\operatorname{Tr}\left(\left(\boldsymbol{I} + \frac{\partial \boldsymbol{f}^{(0)}}{\partial \boldsymbol{\theta}}\right) \boldsymbol{J}^{-1}(\boldsymbol{\theta}) \left(\boldsymbol{I} + \frac{\partial \boldsymbol{f}^{(0)}}{\partial \boldsymbol{\theta}}\right)^T\right)} \sqrt{\operatorname{Tr}\left(\left(\boldsymbol{I} + \frac{\partial \boldsymbol{f}^{(j)}}{\partial \boldsymbol{\theta}}\right) \boldsymbol{J}^{-1}(\boldsymbol{\theta}) \left(\boldsymbol{I} + \frac{\partial \boldsymbol{f}^{(j)}}{\partial \boldsymbol{\theta}}\right)^T\right)} p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}). \tag{77}$$

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on November 17, 2009 at 07:06 from IEEE Xplore. Restrictions apply.

It remains to show that for $\ell \in \{1,3\}$, the minimizer of (67) is unique. We will demonstrate this by showing that $Z_1[\cdot]$ and $Z_3[\cdot]$ are strictly convex. Let $b^{(0)}, b^{(1)} \in S$ be two essentially different functions, i.e.,

$$p_{\boldsymbol{\theta}}\left(\left\{\boldsymbol{\theta} \in \Theta : \boldsymbol{b}^{(0)}(\boldsymbol{\theta}) \neq \boldsymbol{b}^{(1)}(\boldsymbol{\theta})\right\}\right) > 0. \tag{80}$$

Let $\boldsymbol{b}^{(2)}(\boldsymbol{\theta}) = \lambda \boldsymbol{b}^{(0)}(\boldsymbol{\theta}) + (1-\lambda)\boldsymbol{b}^{(1)}(\boldsymbol{\theta})$ for some $0 < \lambda < 1$, so that $\boldsymbol{b}^{(2)} \in S$ by convexity. We then have

$$\begin{aligned} Z_1[\boldsymbol{b}^{(2)}] &= \int_Q \left\|\lambda \boldsymbol{b}^{(0)}(\boldsymbol{\theta}) + (1-\lambda)\boldsymbol{b}^{(1)}(\boldsymbol{\theta})\right\|^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \\ &\quad + \int_{\Theta\setminus Q} \left\|\lambda \boldsymbol{b}^{(0)}(\boldsymbol{\theta}) + (1-\lambda)\boldsymbol{b}^{(1)}(\boldsymbol{\theta})\right\|^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \\ &< \int_Q \left[\lambda \|\boldsymbol{b}^{(0)}(\boldsymbol{\theta})\|^2 + (1-\lambda)\|\boldsymbol{b}^{(1)}(\boldsymbol{\theta})\|^2\right] p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \\ &\quad + \int_{\Theta\setminus Q} \left[\lambda \|\boldsymbol{b}^{(0)}(\boldsymbol{\theta})\|^2 + (1-\lambda)\|\boldsymbol{b}^{(1)}(\boldsymbol{\theta})\|^2\right] p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \\ &= \lambda Z_1[\boldsymbol{b}^{(0)}] + (1-\lambda) Z_2[\boldsymbol{b}^{(1)}] \end{aligned} \tag{81}$$

where the inequality follows from strict convexity of the squared Euclidean norm $\|\boldsymbol{x}\|^2$. Thus, $Z_1[\cdot]$ is strictly convex, and hence has a unique minimum.

Note that $Z_3[\boldsymbol{b}] = Z_1[\boldsymbol{b}] + Z_2[\boldsymbol{b}]$. Since $Z_1[\cdot]$ is strictly convex and $Z_2[\cdot]$ is convex, it follows that $Z_3[\cdot]$ is strictly convex, and thus also has a unique minimum. This completes the proof. □

The following lemma can be thought of as a triangle inequality for a normed space of matrix functions over Θ.

Lemma 2: Let $p_{\boldsymbol{\theta}}$ be a probability measure over Θ, and let $\boldsymbol{M} : \Theta \longrightarrow \mathbb{R}^{n\times n}$ be a matrix function. Suppose

$$\int_\Theta \|\boldsymbol{I} + \boldsymbol{M}(\boldsymbol{\theta})\|_F^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \leq \alpha \tag{82}$$

for some constant α. It follows that

$$\int_\Theta \|\boldsymbol{M}(\boldsymbol{\theta})\|_F^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \leq (\sqrt{\alpha} + \sqrt{n})^2. \tag{83}$$

Proof: By the triangle inequality

$$\begin{aligned} \|\boldsymbol{M}(\boldsymbol{\theta})\|_F &= \|\boldsymbol{M}(\boldsymbol{\theta}) + \boldsymbol{I} - \boldsymbol{I}\|_F \\ &\leq \|\boldsymbol{M}(\boldsymbol{\theta}) + \boldsymbol{I}\|_F + \|\boldsymbol{I}\|_F. \end{aligned} \tag{84}$$

Since $\|\boldsymbol{I}\|_F^2 = n$, we have

$$\begin{aligned} &\int_\Theta \|\boldsymbol{M}(\boldsymbol{\theta})\|_F^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \\ &\quad \leq \int_\Theta \left[\|\boldsymbol{I}+\boldsymbol{M}(\boldsymbol{\theta})\|_F^2 + n + 2\sqrt{n}\,\|\boldsymbol{I}+\boldsymbol{M}(\boldsymbol{\theta})\|_F\right] p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}). \end{aligned} \tag{85}$$

Using the fact that

$$\int_\Theta \|\boldsymbol{I} + \boldsymbol{M}(\boldsymbol{\theta})\|_F\, p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \leq \sqrt{\int_\Theta \|\boldsymbol{I} + \boldsymbol{M}(\boldsymbol{\theta})\|_F^2\, p_{\boldsymbol{\theta}}(d\boldsymbol{\theta})} \tag{86}$$

and combining with (82), it follows that

$$\int_\Theta \|\boldsymbol{M}(\boldsymbol{\theta})\|_F^2\, p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \leq \alpha + n + 2\sqrt{n\alpha} \tag{87}$$

which completes the proof. □

Appendix B
Proof of Proposition 1

The following proof of Proposition 1 makes use of the results developed in Appendix A.

Proof of Proposition 1: Recall that $Z_3[\boldsymbol{b}]$ of (68) equals $Z[\boldsymbol{b}]$. Thus, we would like to apply Lemma 1 (with $\ell = 3$) to prove the unique existence of a minimizer of (17). However, Lemma 1 requires that the minimization be performed over a closed, bounded, and convex set S, whereas (17) is performed over the unbounded set H^1. To resolve this issue, we must show that the minimization (17) can be reformulated as a minimization over a closed, bounded, and convex set S.

To this end, note that

$$Z[\mathbf{0}] = \int_\Theta \mathrm{Tr}(\boldsymbol{J}^{-1}(\boldsymbol{\theta})) p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \triangleq U \tag{88}$$

and therefore $M \leq U < \infty$. Thus, it suffices to perform the minimization (17) over those functions for which $Z[\boldsymbol{b}] \leq U$. We now show that this can be achieved by minimizing over a closed, bounded, and convex set S. First, note that $Z[\boldsymbol{b}] \geq \|\boldsymbol{b}\|_{L^2}^2$, so that one may choose to minimize (17) only over functions $\boldsymbol{b}$ for which

$$\|\boldsymbol{b}\|_{L^2}^2 \leq U. \tag{89}$$

Similarly, we have

$$Z[\boldsymbol{b}] \geq \int_\Theta \mathrm{Tr}\left(\left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}}\right) \boldsymbol{J}^{-1}(\boldsymbol{\theta}) \left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}}\right)^T\right) p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \tag{90}$$

so that it suffices to minimize (17) over functions $\boldsymbol{b}$ for which

$$\int_\Theta \mathrm{Tr}\left(\left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}}\right) \boldsymbol{J}^{-1}(\boldsymbol{\theta}) \left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}}\right)^T\right) p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \leq U. \tag{91}$$

Note that $\boldsymbol{J}(\boldsymbol{\theta})$ is bounded a.e., and therefore $\lambda_{\min}(\boldsymbol{J}^{-1}) \geq 1/K$ a.e., for some constant K. It follows that

$$\begin{aligned} &\mathrm{Tr}\left(\left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}}\right) \boldsymbol{J}^{-1}(\boldsymbol{\theta}) \left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}}\right)^T\right) \\ &\qquad \geq \frac{1}{K} \left\|\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}}\right\|_F^2 \text{ a.e. } (p_{\boldsymbol{\theta}}). \end{aligned} \tag{92}$$

Combining with (91) yields

$$\int_\Theta \left\|\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}}\right\|_F^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \leq KU. \tag{93}$$

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on November 17, 2009 at 07:06 from IEEE Xplore. Restrictions apply.

From Lemma 2, we then have

$$\int_\Theta \left\| \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}} \right\|_F^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \leq \left(\sqrt{n} + \sqrt{KU}\right)^2. \tag{94}$$

From (89) and (94) it follows that the minimization (17) can be limited to the closed, bounded, convex set

$$S = \left\{ \boldsymbol{b} \in H^1 : \|\boldsymbol{b}\|_{H^1}^2 \leq U + \left(\sqrt{KU} + \sqrt{n}\right)^2 \right\}. \tag{95}$$

Applying Lemma 1 proves the unique existence of a minimizer of (17). The proof that $0 < s < \infty$ appears immediately after the statement of Proposition 1. □

Appendix C
Proof of Theorem 2

The following is the proof of Theorem 2 concerning the calculation of the OBB.

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

$$Z[\boldsymbol{b}] = \int_\Theta F[\boldsymbol{b}, \boldsymbol{\theta}] d\boldsymbol{\theta} \tag{96}$$

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 bounded set with a smooth boundary Λ. 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 $\boldsymbol{h} : \Theta \to \mathbb{R}^n$ [40].

By a standard technique [40, Sec. 35], it can be shown that

$$\begin{aligned}\delta Z[\boldsymbol{h}] = \epsilon \sum_i \int_\Theta \left(\frac{\partial F}{\partial b_i} - \sum_j \frac{\partial}{\partial \theta_j} \frac{\partial F}{\partial b_i^{(j)}} \right) h_i(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ + \epsilon \sum_i \int_\Lambda \left(\frac{\partial F}{\partial b_i^{(1)}}, \ldots, \frac{\partial F}{\partial b_i^{(n)}} \right)^T \boldsymbol{\nu}(\boldsymbol{\theta})\, h_i(\boldsymbol{\theta})\, d\sigma\end{aligned} \tag{97}$$

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 $\boldsymbol{\theta} \in \Lambda$. 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 Λ. In this case, the second integral vanishes, and we obtain the Euler–Lagrange equations

$$\forall i, \quad \frac{\partial F}{\partial b_i} - \sum_j \frac{\partial}{\partial \theta_j} \frac{\partial F}{\partial b_i^{(j)}} = 0. \tag{98}$$

Substituting this result back into (97), and again using the fact that $\delta Z[\boldsymbol{h}] = 0$ for all $\boldsymbol{h}$, we obtain the boundary condition

$$\forall i,\ \forall \boldsymbol{\theta} \in \Lambda, \quad \left(\frac{\partial F}{\partial b_i^{(1)}}, \ldots, \frac{\partial F}{\partial b_i^{(n)}} \right)^T \boldsymbol{\nu}(\boldsymbol{\theta}) = 0. \tag{99}$$

Plugging $F[\boldsymbol{b}, \boldsymbol{\theta}] = \mathrm{CRB}[\boldsymbol{b}, \boldsymbol{\theta}] p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ into (98) and (99) provides the required result. □

Appendix D
Proof of Theorem 3

Before proving Theorem 3, we provide the following two lemmas, which demonstrate some symmetry properties of the CRB.

Lemma 3: Under the conditions of Theorem 3, the functional $Z[\boldsymbol{b}]$ of (12) is rotation and reflection invariant, i.e., $Z[\boldsymbol{b}] = Z[\boldsymbol{U}\boldsymbol{b}]$ for any unitary matrix $\boldsymbol{U}$.

Proof: We first demonstrate that $Z[\boldsymbol{b}]$ is rotation invariant. From the definitions of $Z[\boldsymbol{b}]$ and $\mathrm{CRB}[\boldsymbol{b}, \boldsymbol{\theta}]$, we have

$$\begin{aligned}Z[\boldsymbol{b}] = \int_\Theta \mathrm{Tr}\left[\left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}} \right) \left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}} \right)^T \right] \frac{q(\|\boldsymbol{\theta}\|)}{J(\|\boldsymbol{\theta}\|)} d\boldsymbol{\theta} \\ + \int_\Theta \|\boldsymbol{b}(\boldsymbol{\theta})\|^2 q(\|\boldsymbol{\theta}\|) d\boldsymbol{\theta}.\end{aligned} \tag{100}$$

The second integral is clearly rotation invariant, since a rotation of $\boldsymbol{b}$ does not alter its norm. It remains to show that the first integral, which we denote by $I_1[\boldsymbol{b}]$, does not change when $\boldsymbol{b}$ is rotated. To this end, we begin by considering a rotation about the first two coordinates, such that $\boldsymbol{b}$ is transformed to $\tilde{\boldsymbol{b}} \triangleq \boldsymbol{R}_\phi \boldsymbol{b}$, where the rotation matrix $\boldsymbol{R}_\phi$ is defined such that

$$\boldsymbol{R}_\phi \boldsymbol{b} = (b_1 \cos\phi + b_2 \sin\phi, -b_1 \sin\phi + b_2 \cos\phi, b_3, \ldots, b_n)^T. \tag{101}$$

We must thus show that $I_1[\boldsymbol{b}] = I_1[\tilde{\boldsymbol{b}}]$. Let us perform the change of variables $\boldsymbol{\theta} \mapsto \tilde{\boldsymbol{\theta}}$, where $\tilde{\boldsymbol{\theta}} = \boldsymbol{R}_{(-\phi)} \boldsymbol{\theta}$. Rewriting the trace in (100) as a sum, we have

$$I_1[\tilde{\boldsymbol{b}}] = \int_\Theta \sum_{i,j} \left(\delta_{ij} + \frac{\partial \tilde{b}_i}{\partial \theta_j} \right)^2 \frac{q(\|\tilde{\boldsymbol{\theta}}\|)}{J(\|\tilde{\boldsymbol{\theta}}\|)} d\tilde{\boldsymbol{\theta}} \tag{102}$$

where we have used the facts that $\|\boldsymbol{\theta}\| = \|\tilde{\boldsymbol{\theta}}\|$ and that Θ does not change under the change of variables.

We now demonstrate some properties of the transformation of $\boldsymbol{b}$ and $\boldsymbol{\theta}$. First, we have, for any j

$$\begin{aligned}\left(\frac{\partial \tilde{b}_1}{\partial \theta_j} \right)^2 + \left(\frac{\partial \tilde{b}_2}{\partial \theta_j} \right)^2 &= \left(\frac{\partial b_1}{\partial \theta_j} \cos\phi + \frac{\partial b_2}{\partial \theta_j} \sin\phi \right)^2 \\ &\quad + \left(-\frac{\partial b_1}{\partial \theta_j} \sin\phi + \frac{\partial b_2}{\partial \theta_j} \cos\phi \right)^2 \\ &= \left(\frac{\partial b_1}{\partial \theta_j} \right)^2 + \left(\frac{\partial b_2}{\partial \theta_j} \right)^2.\end{aligned} \tag{103}$$

Also, for any i

$$\begin{aligned}\left(\frac{\partial b_i}{\partial \tilde{\theta}_1} \right)^2 + \left(\frac{\partial b_i}{\partial \tilde{\theta}_2} \right)^2 &= \left(\frac{\partial b_i}{\partial \theta_1} \frac{\partial \theta_1}{\partial \tilde{\theta}_1} + \frac{\partial b_i}{\partial \theta_2} \frac{\partial \theta_2}{\partial \tilde{\theta}_1} \right)^2 \\ &\quad + \left(\frac{\partial b_i}{\partial \theta_1} \frac{\partial \theta_1}{\partial \tilde{\theta}_2} + \frac{\partial b_i}{\partial \theta_2} \frac{\partial \theta_2}{\partial \tilde{\theta}_2} \right)^2 \\ &= \left(\frac{\partial b_i}{\partial \theta_1} \right)^2 + \left(\frac{\partial b_i}{\partial \theta_2} \right)^2\end{aligned} \tag{104}$$

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on November 17, 2009 at 07:06 from IEEE Xplore. Restrictions apply.

where we used the fact that $\boldsymbol{\theta} = \boldsymbol{R}_\phi \tilde{\boldsymbol{\theta}}$. Third, we have

$$\begin{aligned}\frac{\partial \tilde{b}_1}{\partial \theta_1} &= \frac{\partial b_1}{\partial \tilde{\theta}_1}\cos^2\phi + \frac{\partial b_1}{\partial \tilde{\theta}_2}\sin\phi\cos\phi \\ &\quad + \frac{\partial b_2}{\partial \tilde{\theta}_1}\sin\phi\cos\phi + \frac{\partial b_2}{\partial \tilde{\theta}_2}\sin^2\phi, \\ \frac{\partial \tilde{b}_2}{\partial \theta_2} &= \frac{\partial b_1}{\partial \tilde{\theta}_1}\sin^2\phi - \frac{\partial b_1}{\partial \tilde{\theta}_2}\sin\phi\cos\phi \\ &\quad - \frac{\partial b_2}{\partial \tilde{\theta}_1}\sin\phi\cos\phi + \frac{\partial b_2}{\partial \tilde{\theta}_2}\cos^2\phi \end{aligned} \tag{105}$$

so that

$$\frac{\partial \tilde{b}_1}{\partial \theta_1} + \frac{\partial \tilde{b}_2}{\partial \theta_2} = \frac{\partial b_1}{\partial \tilde{\theta}_1} + \frac{\partial b_2}{\partial \tilde{\theta}_2}. \tag{106}$$

We now show that

$$\sum_{i,j}\left(\delta_{ij} + \frac{\partial \tilde{b}_i}{\partial \theta_j}\right)^2 = \sum_{i,j}\left(\delta_{ij} + \frac{\partial b_i}{\partial \tilde{\theta}_j}\right)^2. \tag{107}$$

For terms with $i, j \geq 3$, we have $b_i = \tilde{b}_i$ and $\theta_j = \tilde{\theta}_j$, so that replacing $\tilde{\boldsymbol{b}}$ with $\boldsymbol{b}$ and $\boldsymbol{\theta}$ with $\tilde{\boldsymbol{\theta}}$ does not change the result. The terms with $i = 1, 2$ and $j \geq 3$ do not change because of (103), while the terms with $i \geq 3$ and $j = 1, 2$ do not change because of (104). It remains to show that the terms $i, j = 1, 2$ do not modify the sum. To this end, we write out these four terms as

$$\begin{aligned}&\left(1 + \frac{\partial \tilde{b}_1}{\partial \theta_1}\right)^2 + \left(1 + \frac{\partial \tilde{b}_2}{\partial \theta_2}\right)^2 + \left(\frac{\partial \tilde{b}_1}{\partial \theta_2}\right)^2 + \left(\frac{\partial \tilde{b}_2}{\partial \theta_1}\right)^2 \\ &=2 + 2\frac{\partial \tilde{b}_1}{\partial \theta_1} + 2\frac{\partial \tilde{b}_2}{\partial \theta_2} \\ &\quad + \left(\frac{\partial \tilde{b}_1}{\partial \theta_1}\right)^2 + \left(\frac{\partial \tilde{b}_1}{\partial \theta_2}\right)^2 + \left(\frac{\partial \tilde{b}_2}{\partial \theta_1}\right)^2 + \left(\frac{\partial \tilde{b}_2}{\partial \theta_2}\right)^2 \\ &=2 + 2\frac{\partial b_1}{\partial \tilde{\theta}_1} + 2\frac{\partial b_2}{\partial \tilde{\theta}_2} \\ &\quad + \left(\frac{\partial b_1}{\partial \tilde{\theta}_1}\right)^2 + \left(\frac{\partial b_1}{\partial \tilde{\theta}_2}\right)^2 + \left(\frac{\partial b_2}{\partial \tilde{\theta}_1}\right)^2 + \left(\frac{\partial b_2}{\partial \tilde{\theta}_2}\right)^2 \\ &= \left(1 + \frac{\partial b_1}{\partial \tilde{\theta}_1}\right)^2 + \left(1 + \frac{\partial b_2}{\partial \tilde{\theta}_2}\right)^2 + \left(\frac{\partial b_1}{\partial \tilde{\theta}_2}\right)^2 + \left(\frac{\partial b_2}{\partial \tilde{\theta}_1}\right)^2\end{aligned} \tag{108}$$

where, in the second transition, we have used (103), (104), and (106). It follows that $I_1[\tilde{\boldsymbol{b}}]$ of (102) is equal to $I_1[\boldsymbol{b}]$, and hence $Z[\boldsymbol{b}] = Z[\tilde{\boldsymbol{b}}]$. The result similarly holds for rotations about any other pair of coordinates. 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

$$\tilde{\boldsymbol{b}} \triangleq (-b_1(\boldsymbol{\theta}), b_2(\boldsymbol{\theta}), \ldots, b_n(\boldsymbol{\theta}))^T \tag{109}$$

be the reflection of $\boldsymbol{b}$, and consider the corresponding change of variables

$$\tilde{\boldsymbol{\theta}} \triangleq (-\theta_1, \theta_2, \ldots, \theta_n)^T. \tag{110}$$

By the symmetry assumptions, $p_{\boldsymbol{\theta}}$ 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 $Z[\boldsymbol{b}] = Z[\tilde{\boldsymbol{b}}]$. □

Lemma 4: Suppose $\boldsymbol{b}(\boldsymbol{\theta})$ is radial and rotation invariant, i.e., $\boldsymbol{b}(\boldsymbol{\theta}) = t(\|\boldsymbol{\theta}\|^2)\boldsymbol{\theta}$ for some function $t \in H^1$. Also suppose that $\boldsymbol{J}(\boldsymbol{\theta}) = J(\|\boldsymbol{\theta}\|)\boldsymbol{I}$, where $J(\cdot)$ is a scalar function. Then, $\mathrm{CRB}[\boldsymbol{b}, \boldsymbol{\theta}]$ of (11) is rotation invariant in $\boldsymbol{\theta}$, i.e., $\mathrm{CRB}[\boldsymbol{b}, \boldsymbol{R}\boldsymbol{\theta}] = \mathrm{CRB}[\boldsymbol{b}, \boldsymbol{\theta}]$ for any rotation matrix $\boldsymbol{R}$.

Proof: We will show that $\mathrm{CRB}[\boldsymbol{b}, \boldsymbol{\theta}]$ depends on $\boldsymbol{\theta}$ only through $\|\boldsymbol{\theta}\|^2$, and is therefore rotation invariant. For the given value of $\boldsymbol{b}(\boldsymbol{\theta})$ and $\boldsymbol{J}(\boldsymbol{\theta})$, we have

$$\begin{aligned}\mathrm{CRB}[\boldsymbol{b}, \boldsymbol{\theta}] &= \|\boldsymbol{b}(\boldsymbol{\theta})\|^2 + \mathrm{Tr}\left[\left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}}\right)\boldsymbol{J}^{-1}(\boldsymbol{\theta})\left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}}\right)^T\right] \\ &= t^2\|\boldsymbol{\theta}\|^2 + \frac{1}{J(\|\boldsymbol{\theta}\|)}\mathrm{Tr}\left[\left(\boldsymbol{I} + \frac{\partial t\boldsymbol{\theta}}{\partial \boldsymbol{\theta}}\right)\left(\boldsymbol{I} + \frac{\partial t\boldsymbol{\theta}}{\partial \boldsymbol{\theta}}\right)^T\right]\end{aligned} \tag{111}$$

where, for notational convenience, we have omitted the dependence of t on $\|\boldsymbol{\theta}\|^2$. It remains to show that the trace in the above expression is a function of $\boldsymbol{\theta}$ only through $\|\boldsymbol{\theta}\|^2$. To this end, we note that

$$\frac{\partial b_i}{\partial \theta_j} = t\delta_{ij} + t'\theta_i\frac{\partial\|\boldsymbol{\theta}\|^2}{\partial \theta_j} = t\delta_{ij} + 2t'\theta_i\theta_j \tag{112}$$

where δ_{ij} is the Kronecker delta. It follows that

$$\left(\delta_{ij} + \frac{\partial b_i}{\partial \theta_j}\right)^2 = (1+t)^2\delta_{ij} + 4(1+t)t'\theta_i\theta_j\delta_{ij} + 4t'^2\theta_i^2\theta_j^2. \tag{113}$$

Therefore

$$\begin{aligned}&\mathrm{Tr}\left[\left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}}\right)\left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}}\right)^T\right] \\ &= \sum_{i,j}\left(\delta_{ij} + \frac{\partial b_i}{\partial \theta_j}\right)^2 \\ &= n(1+t)^2 + 4t'^2\sum_{i,j}\theta_i^2\theta_j^2 + 4(1+t)t'\sum_i\theta_i^2 \\ &= n(1+t)^2 + 4t'^2\|\boldsymbol{\theta}\|^4 + 4(1+t)t'\|\boldsymbol{\theta}\|^2.\end{aligned} \tag{114}$$

Thus, $\mathrm{CRB}[\boldsymbol{b}, \boldsymbol{\theta}]$ depends on $\boldsymbol{\theta}$ only through $\|\boldsymbol{\theta}\|^2$, completing the proof. □

Proof of Theorem 3: We have seen in Theorem 2 that the solution of (20) is unique. Now suppose that the optimum $\boldsymbol{b}$ is not rotation invariant, i.e., there exists a rotation matrix $\boldsymbol{R}$ such that $\boldsymbol{R}\boldsymbol{b}(\boldsymbol{\theta})$ is not identical to $\boldsymbol{b}(\boldsymbol{\theta})$. By Lemma 3, $\boldsymbol{R}\boldsymbol{b}(\boldsymbol{\theta})$ is also optimal, which is a contradiction.

Furthermore, suppose that $\boldsymbol{b}$ is not radial, i.e., for some value of $\boldsymbol{\theta}$, $\boldsymbol{b}(\boldsymbol{\theta})$ 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 3, The reflection of $\boldsymbol{b}$ through this hyperplane is also an optimal solution of (20), which is again a contradiction. Therefore, the optimum $\boldsymbol{b}$ is spherically symmetric and radial, so that it can be written as

$$\boldsymbol{b}(\boldsymbol{\theta}) = b(\|\boldsymbol{\theta}\|)\frac{\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|} \tag{115}$$

where $b(\cdot)$ is a scalar function.

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on November 17, 2009 at 07:06 from IEEE Xplore. Restrictions apply.

To determine the value of $b(\cdot)$, it suffices to analyze the differential equation (21) along a straight line from the origin to the boundary. We choose a line along the θ_1 axis, and begin by calculating the derivatives of $b_1(\boldsymbol{\theta})$, $q(\|\boldsymbol{\theta}\|)$, and $\boldsymbol{J}(\|\boldsymbol{\theta}\|)$ along this axis. The derivative of $q(\|\boldsymbol{\theta}\|)$ is given by

$$\frac{\partial q}{\partial \theta_j} = q'(\rho)\frac{\theta_j}{\rho} \tag{116}$$

where we have denoted $\rho = \|\boldsymbol{\theta}\|$, so that ρ is weakly differentiable and

$$\frac{\partial \rho}{\partial \theta_j} = \frac{\theta_j}{\rho}. \tag{117}$$

Along the θ_1 axis, we have $\theta_1 = \rho$ while $\theta_2 = \cdots = \theta_n = 0$, so that

$$\left.\frac{\partial q}{\partial \theta_j}\right|_{\boldsymbol{\theta}=\rho\boldsymbol{e}_1} = q'(\rho)\delta_{j1}. \tag{118}$$

Similarly, since $\boldsymbol{J}(\boldsymbol{\theta}) = J(\rho)\boldsymbol{I}$

$$\frac{\partial(\boldsymbol{J}^{-1})_{jk}}{\partial \theta_j} = -\frac{J'(\rho)}{J^2(\rho)}\frac{\theta_j}{\rho}\delta_{jk} \tag{119}$$

so that along the θ_1 axis

$$\left.\frac{\partial(\boldsymbol{J}^{-1})_{jk}}{\partial \theta_j}\right|_{\boldsymbol{\theta}=\rho\boldsymbol{e}_1} = -\frac{J'(\rho)}{J^2(\rho)}\delta_{jk}\delta_{j1}. \tag{120}$$

From (115), we have

$$\frac{\partial b_i}{\partial \theta_j} = b'(\rho)\frac{\theta_i\theta_j}{\rho^2} + \frac{b(\rho)}{\rho}\left(\delta_{ij} - \frac{\theta_i\theta_j}{\rho^2}\right). \tag{121}$$

Thus, on the θ_1 axis, we have

$$\left.\frac{\partial b_1}{\partial \theta_j}\right|_{\boldsymbol{\theta}=\rho\boldsymbol{e}_1} = b'(\rho)\delta_{j1}. \tag{122}$$

The second derivative of $b_i(\boldsymbol{\theta})$ can be shown to equal

$$\frac{\partial^2 b_i}{\partial\theta_j\partial\theta_k} = b''(\rho)\frac{\theta_i\theta_j\theta_k}{\rho^3} + \left(\frac{b'(\rho)}{\rho} - \frac{b(\rho)}{\rho^2}\right)\left(\frac{\theta_i}{\rho}\delta_{jk} + \frac{\theta_j}{\rho}\delta_{ik} + \frac{\theta_k}{\rho}\delta_{ij} - 3\frac{\theta_i\theta_j\theta_k}{\rho^3}\right). \tag{123}$$

Therefore, on the θ_1 axis

$$\begin{aligned}
\left.\frac{\partial^2 b_1}{\partial\theta_1^2}\right|_{\boldsymbol{\theta}=\rho\boldsymbol{e}_1} &= b''(\rho)\\
\left.\frac{\partial^2 b_1}{\partial\theta_j^2}\right|_{\boldsymbol{\theta}=\rho\boldsymbol{e}_1} &= \frac{b'(\rho)}{\rho} - \frac{b(\rho)}{\rho^2}\ (j \neq 1)\\
\left.\frac{\partial^2 b_1}{\partial\theta_j\partial\theta_k}\right|_{\boldsymbol{\theta}=\rho\boldsymbol{e}_1} &= 0\ (j,k \neq 1).
\end{aligned} \tag{124}$$

Substituting these derivatives into (21), we obtain

$$q(\rho)b(\rho) = \frac{q(\rho)}{J(\rho)}\left(b''(\rho) + (n-1)\frac{b'(\rho)}{\rho} - (n-1)\frac{b(\rho)}{\rho^2}\right) + (1+b'(\rho))\left(\frac{q'(\rho)}{J(\rho)} - q(\rho)\frac{J'(\rho)}{J^2(\rho)}\right) \tag{125}$$

which is equivalent to (25).

To obtain the boundary conditions, observe that Lemma 3 implies $\boldsymbol{b}(\boldsymbol{0}) = \boldsymbol{0}$, whence, we conclude that $b(0) = 0$. Next, evaluate the boundary condition (22) at boundary point $\boldsymbol{\theta} = r\boldsymbol{e}_1$, where the surface normal $\boldsymbol{\nu}(\boldsymbol{\theta})$ equals $\boldsymbol{e}_1$, so that

$$1 + b'(\rho) = 1 + \frac{\partial b_1}{\partial \theta_1} = 0, \quad \boldsymbol{\theta} = r\boldsymbol{e}_1 \tag{126}$$

which is equivalent to the boundary condition $b'(r) = -1$.

To find the OBB (24), we must now calculate $Z[\boldsymbol{b}]$ for the obtained bias function (115). To this end, note that, by Lemma 4, $\mathrm{CRB}[\boldsymbol{b},\boldsymbol{\theta}]$ is rotation invariant in $\boldsymbol{\theta}$ for the required $\boldsymbol{b}(\boldsymbol{\theta})$. Thus, the integrand $\mathrm{CRB}[\boldsymbol{b},\boldsymbol{\theta}]q(\|\boldsymbol{\theta}\|)$ is constant on any $(n-1)$-sphere centered on the origin, so that

$$Z[\boldsymbol{b}] = \int_0^r \mathrm{CRB}[\boldsymbol{b},\rho\boldsymbol{e}_1]q(\rho)S_n(\rho)d\rho \tag{127}$$

where

$$S_n(\rho) = \frac{2\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)}\rho^{n-1} \tag{128}$$

is the hypersurface area of an $(n-1)$-sphere of radius ρ [35]. It thus suffices to calculate the value of $\mathrm{CRB}[\boldsymbol{b},\boldsymbol{\theta}]$ at points along the θ_1 axis. From (121), it follows that

$$\left.\frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\rho\boldsymbol{e}_1} = diag\left(b'(\rho), \frac{b(\rho)}{\rho}, \ldots, \frac{b(\rho)}{\rho}\right). \tag{129}$$

Substituting this into the definition of $\mathrm{CRB}[\boldsymbol{b},\boldsymbol{\theta}]$, we obtain

$$\mathrm{CRB}[\boldsymbol{b},\rho\boldsymbol{e}_1] = b^2(\rho) + \frac{1}{J(\rho)}(1+b'(\rho))^2 + \frac{n-1}{J(\rho)}\left(1 + \frac{b(\rho)}{\rho}\right)^2. \tag{130}$$

Combining (130) with (127) yields (24), as required. □

Appendix E
Proofs of Asymptotic Properties

Theorems 5 and 6 demonstrate asymptotic tightness of the OBB. The proofs of these two theorems follow.

Proof of Theorem 5: We begin the proof by studying a certain optimization problem, whose relevance will be demonstrated shortly. Let $t \geq 0$ be a constant and consider the problem

$$u(t) = \inf_{\boldsymbol{b}\in H^1} \int_\Theta \left\|\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}}\right\|_F^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \quad \text{s.t.} \int_\Theta \|\boldsymbol{b}(\boldsymbol{\theta})\|^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \leq t. \tag{131}$$

Notice that $u(t) \leq n$ for all t, since an objective having a value of n is achieved by the function $\boldsymbol{b}(\boldsymbol{\theta}) = \boldsymbol{0}$. Thus, it suffices to perform the minimization (131) over functions $\boldsymbol{b} \in H^1$ satisfying

$$\int_\Theta \left\|\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}}\right\|_F^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \leq n. \tag{132}$$

It follows from Lemma 2 that such functions also satisfy

$$\int_\Theta \left\|\frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}}\right\|_F^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \leq (2\sqrt{n})^2 = 4n. \tag{133}$$

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on November 17, 2009 at 07:06 from IEEE Xplore. Restrictions apply.

Therefore, (131) is equivalent to the minimization

$$u(t) = \inf_{\boldsymbol{b}\in S_t} \int_\Theta \left\| \boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}} \right\|_F^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \tag{134}$$

where

$$S_t = \left\{ \boldsymbol{b} \in H^1 : \int_\Theta \|\boldsymbol{b}(\boldsymbol{\theta})\|^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \le t, \int_\Theta \left\| \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}} \right\|_F^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \le 4n \right\}. \tag{135}$$

The set S_t is convex, closed, and bounded in H^1. Applying Lemma 1 (with $\ell = 2$) implies that there exists a function $\boldsymbol{b}_{\rm opt} \in S_t$ which minimizes (134), and hence also minimizes (131).

Note that the objective in (131) is zero if and only if

$$\frac{\partial \boldsymbol{b}_{\rm opt}}{\partial \boldsymbol{\theta}} = -\boldsymbol{I} \quad \text{a.e. } (p_{\boldsymbol{\theta}}). \tag{136}$$

The only functions in H^1 satisfying this requirement are the functions

$$\boldsymbol{b}(\boldsymbol{\theta}) = \boldsymbol{k} - \boldsymbol{\theta} \quad \text{a.e.}(p_{\boldsymbol{\theta}}) \tag{137}$$

for some constant $\boldsymbol{k} \in \mathbb{R}^n$. Let $\boldsymbol{\mu} \triangleq E\{\boldsymbol{\theta}\}$ and define

$$v \triangleq E\left\{ \|\boldsymbol{\theta} - E\{\boldsymbol{\theta}\}\|^2 \right\}. \tag{138}$$

For functions of the form (137), the constraint of (131) is given by

$$\begin{aligned} \int_\Theta \|\boldsymbol{k} - \boldsymbol{\theta}\|^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) &= \int_\Theta \|\boldsymbol{k} - \boldsymbol{\mu} + \boldsymbol{\mu} - \boldsymbol{\theta}\|^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \\ &= \|\boldsymbol{k} - \boldsymbol{\mu}\|^2 + v \\ &\ge v. \end{aligned} \tag{139}$$

In (139), equality is obtained if and only if $\boldsymbol{k} = \boldsymbol{\mu}$. Therefore, if $t < v$, no functions satisfying (136) are feasible, and thus

$$\begin{aligned} u(t) &= 0, \quad \text{if } t \ge v \\ u(t) &> 0, \quad \text{if } t < v. \end{aligned} \tag{140}$$

We now return to the setting of Theorem 5. We must show that $\beta_N \to v$ as $N \to \infty$. We denote functions corresponding to the problem of estimating $\boldsymbol{\theta}$ from $\boldsymbol{x}^{(N)}$ with a superscript (N). Thus, for example, $Z^{(N)}[\boldsymbol{b}]$ denotes the functional $Z[\boldsymbol{b}]$ of (12) for the problem corresponding to the measurement vector $\boldsymbol{x}^{(N)}$.

Since all eigenvalues of $\boldsymbol{J}^{(N)}(\boldsymbol{\theta})$ decrease monotonically with N for $p_{\boldsymbol{\theta}}$-almost all $\boldsymbol{\theta}$, we have

$$\text{CRB}^{(N)}[\boldsymbol{b}, \boldsymbol{\theta}] \le \text{CRB}^{(N+1)}[\boldsymbol{b}, \boldsymbol{\theta}] \tag{141}$$

for any $\boldsymbol{b} \in H^1$, for $p_{\boldsymbol{\theta}}$-almost all $\boldsymbol{\theta}$, and for all N. Therefore

$$Z^{(N)}[\boldsymbol{b}] \le Z^{(N+1)}[\boldsymbol{b}] \tag{142}$$

for any $\boldsymbol{b} \in H^1$ and for all N. It follows that for all N

$$\beta_N = \min_{\boldsymbol{b}\in H^1} Z^{(N)}[\boldsymbol{b}] \le \min_{\boldsymbol{b}\in H^1} Z^{(N+1)}[\boldsymbol{b}] = \beta_{N+1} \tag{143}$$

so that β_N is a nondecreasing sequence. Furthermore, note that

$$Z^{(N)}[\boldsymbol{\mu} - \boldsymbol{\theta}] = v \quad \text{for all } N \tag{144}$$

where v is given by (138). Therefore, $\beta_N \le v$ for all N. Thus, β_N converges to some value q, and we have

$$\beta_N \le q \le v \quad \text{for all } N. \tag{145}$$

To prove the theorem, it remains to show that $q = v$.

Let $\boldsymbol{b}^{(N)}$ be the minimizer of (17) when $\boldsymbol{\theta}$ is estimated from $\boldsymbol{x}^{(N)}$; this minimizer exists by virtue of Proposition 1. We then have

$$\beta_N = Z^{(N)}[\boldsymbol{b}^{(N)}] \le q \tag{146}$$

and therefore

$$\int_\Theta \|\boldsymbol{b}^{(N)}(\boldsymbol{\theta})\|^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \le q. \tag{147}$$

It follows that $\boldsymbol{b}^{(N)}$ satisfies the constraint of the optimization problem (131) with $t = q$. As a consequence, we have

$$\int_\Theta \left\| \boldsymbol{I} + \frac{\partial \boldsymbol{b}^{(N)}}{\partial \boldsymbol{\theta}} \right\|_F^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \ge u(q). \tag{148}$$

Define

$$\lambda_N \triangleq \operatorname*{ess\,sup}_{\boldsymbol{\theta}\in\Theta} \lambda_{\max}(\boldsymbol{J}^{(N)}(\boldsymbol{\theta})) \tag{149}$$

and note that $\lambda_N > 0$ for all N, since $\boldsymbol{J}^{(N)}(\boldsymbol{\theta})$ is positive definite. Thus

$$\begin{aligned} Z^{(N)}[\boldsymbol{b}^{(N)}] &\ge \int_\Theta \text{Tr}\left[\left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}^{(N)}}{\partial \boldsymbol{\theta}} \right) \left(\boldsymbol{J}^{(N)}(\boldsymbol{\theta}) \right)^{-1} \cdot \left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}^{(N)}}{\partial \boldsymbol{\theta}} \right)^T \right] p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \\ &\ge \frac{1}{\lambda_N} \int_\Theta \left\| \boldsymbol{I} + \frac{\partial \boldsymbol{b}^{(N)}}{\partial \boldsymbol{\theta}} \right\|_F^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \\ &\ge \frac{u(q)}{\lambda_N}. \end{aligned} \tag{150}$$

Assume by contradiction that $q < v$. From (140), it then follows that $u(q) > 0$. Since all eigenvalues of $\boldsymbol{J}^{(N)}(\boldsymbol{\theta})$ decrease to zero, we have $\lambda_N \to 0$, and thus

$$\beta_N \ge \frac{u(q)}{\lambda_N} \to \infty. \tag{151}$$

This contradicts the fact (145) that $\beta_N \le v$. We conclude that $q = v$, as required. □

Proof of Theorem 6: The structure of this proof is similar to that of Theorem 5. We begin by considering the optimization problem

$$\begin{aligned} &\inf_{\boldsymbol{b}\in H^1} \int_\Theta \|\boldsymbol{b}(\boldsymbol{\theta})\|^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \\ &\text{s.t.} \int_\Theta \text{Tr}\left(\left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}} \right) \boldsymbol{J}^{-1}(\boldsymbol{\theta}) \left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}} \right)^T \right) p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \le t \end{aligned} \tag{152}$$

for some constant $t \ge 0$. Denote the solution of (152) by $w(t)$. Let $\boldsymbol{\mu} = E\{\boldsymbol{\theta}\}$ and note that $\boldsymbol{b}(\boldsymbol{\theta}) = \boldsymbol{\mu} - \boldsymbol{\theta}$ satisfies the constraint

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on November 17, 2009 at 07:06 from IEEE Xplore. Restrictions apply.

in (152) for any $t \geq 0$, and has an objective equal to v of (138). Thus, to determine $w(t)$, it suffices to minimize (152) over the set

$$S_t = \left\{ \boldsymbol{b} \in H^1 : \int_\Theta \|\boldsymbol{b}(\boldsymbol{\theta})\|^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \leq v, \right.$$
$$\left. \int_\Theta \mathrm{Tr}\left(\left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}} \right) \boldsymbol{J}^{-1}(\boldsymbol{\theta}) \left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}} \right)^T \right) p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \leq t \right\}.$$

Define

$$\lambda \triangleq \operatorname*{ess\,sup}_{\boldsymbol{\theta} \in \Theta} \lambda_{\max}(\boldsymbol{J}(\boldsymbol{\theta})). \tag{153}$$

Since $\boldsymbol{J}(\boldsymbol{\theta})$ is positive definite a.e., we have $\lambda > 0$. For any $\boldsymbol{b} \in S_t$, we have

$$\frac{1}{\lambda} \int_\Theta \left\| \boldsymbol{I} + \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}} \right\|_F^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \leq t \tag{154}$$

and, therefore, by Lemma 2

$$\int_\Theta \left\| \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}} \right\|_F^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \leq \left(\sqrt{t\lambda} + \sqrt{n} \right)^2. \tag{155}$$

Hence, for any $\boldsymbol{b} \in S_t$

$$\begin{aligned} \|\boldsymbol{b}\|_{H^1}^2 &= \int_\Theta \|\boldsymbol{b}(\boldsymbol{\theta})\|^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) + \int_\Theta \left\| \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\theta}} \right\|_F^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \\ &\leq v + \left(\sqrt{t\lambda} + \sqrt{n} \right)^2. \end{aligned} \tag{156}$$

Thus, S_t is bounded for all t. It is straightforward to show that S_t is also closed and convex. Therefore, employing Lemma 1 (with $\ell = 1$) ensures that there exists a (unique) $\boldsymbol{b}_{\mathrm{opt}} \in S_t$ minimizing (152).

Note that the objective in (152) is 0 if and only if $\boldsymbol{b}_{\mathrm{opt}}(\boldsymbol{\theta}) = \mathbf{0}$ a.e. So, if $\mathbf{0} \in S_t$, we have $w(t) = 0$, and otherwise $w(t) > 0$. Let us define

$$s \triangleq E\left\{ \mathrm{Tr}(\boldsymbol{J}^{-1}(\boldsymbol{\theta})) \right\} \tag{157}$$

and note that $\mathbf{0} \in S_t$ if and only if $t \geq s$. Thus

$$\begin{aligned} w(t) &= 0, \quad \text{for } t \geq s \\ w(t) &> 0, \quad \text{otherwise.} \end{aligned} \tag{158}$$

Let us now return to the setting of Theorem 6. For simplicity, we denote functions corresponding to the problem of estimating $\boldsymbol{\theta}$ from $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}\}$ with a superscript (N). For example, from the additive property of the Fisher information [2, Sec. 3.4], we have

$$\boldsymbol{J}^{(N)}(\boldsymbol{\theta}) = N\boldsymbol{J}(\boldsymbol{\theta}). \tag{159}$$

It follows that

$$(N+1)\mathrm{CRB}^{(N+1)}[\boldsymbol{b}, \boldsymbol{\theta}] \geq N\mathrm{CRB}^{(N)}[\boldsymbol{b}, \boldsymbol{\theta}] \tag{160}$$

for all $\boldsymbol{b} \in H^1$, all $\boldsymbol{\theta} \in \Theta$, and all N. Therefore

$$(N+1)Z^{(N+1)}[\boldsymbol{b}] \geq NZ^{(N)}[\boldsymbol{b}] \tag{161}$$

for all $\boldsymbol{b} \in H^1$, and hence

$$\begin{aligned} (N+1)\beta_{N+1} &= \min_{\boldsymbol{b} \in H^1} \left((N+1)Z^{(N+1)}[\boldsymbol{b}] \right) \\ &\geq \min_{\boldsymbol{b} \in H^1} \left(NZ^{(N)}[\boldsymbol{b}] \right) \\ &= N\beta_N. \end{aligned} \tag{162}$$

Thus, $\{N\beta_N\}$ is a nondecreasing sequence. Furthermore, we have

$$NZ^{(N)}[\mathbf{0}] = s \tag{163}$$

so that $N\beta_N \leq s$ for all N. It follows that $\{N\beta_N\}$ is nondecreasing and bounded, and therefore converges to some value r such that

$$N\beta_N \leq r \leq s, \quad \text{for all } N. \tag{164}$$

To prove the theorem, we must show that $r = s$.

Let $\boldsymbol{b}^{(N)} \in H^1$ denote the minimizer of (17) when $\boldsymbol{\theta}$ is estimated from $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}\}$ (the existence of $\boldsymbol{b}^{(N)}$ is guaranteed by Proposition 1). We then have $N\beta_N = NZ^{(N)}[\boldsymbol{b}^{(N)}] \leq r$, so that

$$\int_\Theta \mathrm{Tr}\left(\left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}^{(N)}}{\partial \boldsymbol{\theta}} \right) \boldsymbol{J}^{-1}(\boldsymbol{\theta}) \left(\boldsymbol{I} + \frac{\partial \boldsymbol{b}^{(N)}}{\partial \boldsymbol{\theta}} \right)^T \right) p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \leq r. \tag{165}$$

Thus, $\boldsymbol{b}^{(N)}$ satisfies the constraint of (152) with $t = r$. As a consequence, we have

$$\int_\Theta \|\boldsymbol{b}^{(N)}(\boldsymbol{\theta})\|^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \geq w(r) \tag{166}$$

and therefore

$$\begin{aligned} N\beta_N &= NZ^{(N)}[\boldsymbol{b}^{(N)}] \\ &\geq N \int_\Theta \|\boldsymbol{b}^{(N)}(\boldsymbol{\theta})\|^2 p_{\boldsymbol{\theta}}(d\boldsymbol{\theta}) \\ &\geq Nw(r). \end{aligned} \tag{167}$$

Now suppose by contradiction that $r < s$. It follows from (158) that $w(r) > 0$. Hence, by (167), $N\beta_N \to \infty$, which contradicts the fact that $N\beta_N$ is bounded. We conclude that $r = s$, as required. □

ACKNOWLEDGMENT

The authors are grateful to Dr. Volker Pohl for fruitful discussions concerning many of the mathematical aspects of the paper. The authors would also like to thank the anonymous reviewers for their constructive comments.

REFERENCES

[1] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. New York: Springer-Verlag, 1985.

[2] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on November 17, 2009 at 07:06 from IEEE Xplore. Restrictions apply.

[3] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1968, vol. 1.
[4] J. Ziv and M. Zakai, "Some lower bounds on signal parameter estimation," *IEEE Trans. Inf. Theory*, vol. IT-15, no. 3, pp. 386–391, May 1969.
[5] T. Y. Young and R. A. Westerberg, "Error bounds for stochastic estimation of signal parameters," *IEEE Trans. Inf. Theory*, vol. IT-17, no. 5, pp. 549–557, Sep. 1971.
[6] S. Bellini and G. Tartara, "Bounds on error in signal parameter estimation," *IEEE Trans. Commun.*, vol. COM-22, no. 3, pp. 340–342, Mar. 1974.
[7] D. Chazan, M. Zakai, and J. Ziv, "Improved lower bounds on signal parameter estimation," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 1, pp. 90–93, Jan. 1975.
[8] B. Z. Bobrovski and M. Zakai, "A lower bound on the estimation error for certain diffusion problems," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 45–52, Jan. 1976.
[9] A. J. Weiss and E. Weinstein, "A lower bound on the mean-square error in random parameter estimation," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 5, pp. 680–682, Sep. 1985.
[10] 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.
[11] K. L. Bell, Y. Steinberg, Y. Ephraim, and H. L. Van Trees, "Extended Ziv–Zakai lower bound for vector parameter estimation," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 624–637, Mar. 1997.
[12] A. Renaux, P. Forster, P. Larzabal, and C. Richmond, "The Bayesian Abel bound on the mean square error," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP 2006)*, Toulouse, France, May 2006, vol. III, pp. 9–12.
[13] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed. New York: Springer, 1998.
[14] Y. C. Eldar, "Rethinking biased estimation: Improving maximum likelihood and the Cramér-Rao bound," *Foundations and Trends in Signal Processing*, vol. 1, no. 4, pp. 305–449, 2008.
[15] S. M. Kay and Y. C. Eldar, "Rethinking biased estimation," *IEEE Signal Process. Mag.*, vol. 25, no. 3, pp. 133–136, May 2008.
[16] H. Cramér, "A contribution to the theory of statistical estimation," *Skand. Akt. Tidskr.*, vol. 29, pp. 85–94, 1945.
[17] C. R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, vol. 37, pp. 81–91, 1945.
[18] J. M. Hammersley, "On estimating restricted parameters," *J. Roy. Statist. Soc. B*, vol. 12, no. 2, pp. 192–240, 1950.
[19] D. G. Chapman and H. Robbins, "Minimum variance estimation without regularity assumptions," *Ann. Math. Statist.*, vol. 22, no. 4, pp. 581–586, Dec. 1951.
[20] P. K. Bhattacharya, "Estimating the mean of a multivariate normal population with general quadratic loss function," *Ann. Math. Statist.*, vol. 37, no. 6, pp. 1819–1824, Dec. 1966.
[21] E. W. Barankin, "Locally best unbiased estimates," *Ann. Math. Statist.*, vol. 20, no. 4, pp. 477–501, Dec. 1949.
[22] J. S. Abel, "A bound on mean-square-estimate error," *IEEE Trans. Inf. Theory*, vol. 39, no. 5, pp. 1675–1680, Sep. 1993.
[23] A. O. Hero, J. A. Fessler, and M. Usman, "Exploring estimator bias-variance tradeoffs using the uniform CR bound," *IEEE Trans. Signal Process.*, vol. 44, no. 8, pp. 2026–2041, Aug. 1996.
[24] P. Forster and P. Larzabal, "On lower bounds for deterministic parameter estimation," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP 2002)*, Orlando, FL, May 2002, vol. 2, pp. 1137–1140.
[25] Y. C. Eldar, "Minimum variance in biased estimation: Bounds and asymptotically optimal estimators," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1915–1930, Jul. 2004.
[26] Y. C. Eldar, "Uniformly improving the Cramér-Rao bound and maximum-likelihood estimation," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 2943–2956, Aug. 2006.
[27] Y. C. Eldar, "MSE bounds with affine bias dominating the Cramér-Rao bound," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3824–3836, Aug. 2008.
[28] I. A. Ibragimov and R. Z. Has'minskii, *Statistical Estimation: Asymptotic Theory*. New York: Springer, 1981.
[29] A. Renaux, "Contribution à l'analyse des performances d'estimation en traitement statistique du signal," Ph.D. dissertation, École Normale Supérieure de Cachan, Cachan, France, 2006.
[30] Z. Ben-Haim and Y. C. Eldar, "A Bayesian estimation bound based on the optimal bias function," in *Proc. 2nd Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP 2007)*, St. Thomas, U.S. Virgin Islands, Dec. 2007, pp. 45–48.
[31] J. Shao, *Mathematical Statistics*, 2nd ed. New York: Springer, 2003.
[32] E. H. Lieb and M. Loss, *Analysis*, 2nd ed. Providence, RI: AMS, 2001.
[33] D. R. Cox and N. Reid, "Parameter orthogonality and approximate conditional inference," *J. Roy. Statist. Soc. B*, vol. 49, no. 1, pp. 1–39, 1987.
[34] H. L. Van Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. New York: Wiley, 2007.
[35] , M. Vinogradov, Ed., *Encyclopaedia of Mathematics*. Dordrecht, The Netherlands: Kluwer, 1995.
[36] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1964.
[37] Z. Ben-Haim and Y. C. Eldar, "A comment on the use of the Weiss–Weinstein bound with constrained parameter sets," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4682–4684, Oct. 2008.
[38] L. P. Lebedev and M. J. Cloud, *The Calculus of Variations and Functional Analysis*. Singapore: World Scientific, 2003.
[39] W. Rudin, *Functional Analysis*. New York: McGraw-Hill, 1973.
[40] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*. Mineola, NY: Dover, 2000.

Zvika Ben-Haim (S'05) received the B.Sc. degree in electrical engineering in 2000, and the B.A. degree in physics as well as the M.Sc. degree in electrical engineering in 2005, all *summa cum laude* from the Technion–Israel Institute of Technology, Haifa, Israel.

From 1996 to 2000, he was a member of the Technion Excellence Program. From 2000 to 2006, he served as a researcher in the Israel Defense Forces. He is currently working towards the Ph.D. degree in the Department of Electrical Engineering at the Technion. His research interests include theoretical aspects of signal processing, estimation theory, and statistical signal processing.

Yonina C. Eldar (S'98–M'02–SM'07) received the B.Sc. degree in physics in 1995 and the B.Sc. degree in electrical engineering in 1996 both from Tel-Aviv University (TAU), Tel-Aviv, Israel, and the Ph.D. degree in electrical engineering and computer science in 2001 from the Massachusetts Institute of Technology (MIT), Cambridge.

From January 2002 to July 2002, she was a Postdoctoral Fellow at the Digital Signal Processing Group at MIT. She is currently an Associate Professor in the Department of Electrical Engineering at the Technion–Israel Institute of Technology, Haifa, Israel. She is also a Research Affiliate with the Research Laboratory of Electronics at MIT. Her research interests are in the general areas of statistical signal processing, sampling theory, and computational biology.

Dr. Eldar was in the program for outstanding students at TAU from 1992 to 1996. In 1998, she held the Rosenblith Fellowship for study in Electrical Engineering at MIT, and in 2000, she held an IBM Research Fellowship. From 2002 to 2005, she was a Horev Fellow of the Leaders in Science and Technology program at the Technion and an Alon Fellow. In 2004, she was awarded the Wolf Foundation Krill Prize for Excellence in Scientific Research, in 2005 the Andre and Bella Meyer Lectureship, in 2007 the Henry Taub Prize for Excellence in Research, and in 2008 the Hershel Rich Innovation Award, the Award for Women with Distinguished Contributions, and the Muriel & David Jacknow Award for Excellence in Teaching. She is a member of the IEEE Signal Processing Theory and Methods technical committee and the Bio Imaging Signal Processing technical committee, an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the *EURASIP Journal of Signal Processing*, the *SIAM Journal on Matrix Analysis and Applications*, and the *SIAM Journal on Imaging Sciences*, and on the Editorial Board of *Foundations and Trends in Signal Processing*.

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on November 17, 2009 at 07:06 from IEEE Xplore. Restrictions apply.