A RISK-UNBIASED APPROACH TO A NEW CRAMÉR-RAO BOUND

Shahar Bar and Joseph Tabrikian

Department of Electrical and Computer Engineering
Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel
Email: shahba@post.bgu.ac.il, joseph@bgu.ac.il

ABSTRACT

How accurately can one estimate a deterministic parameter subject to other unknown deterministic model parameters? The most popular answer to this question is given by the Cramér-Rao bound (CRB). The main assumption behind the derivation of the CRB is local unbiased estimation of all model parameters. The foundations of this work rely on doubting this assumption. Each parameter in its turn is treated as a single parameter of interest, while the other model parameters are treated as nuisance, as their mis-knowledge interferes with the estimation of the parameter of interest. Correspondingly, a new Cramér-Rao-type bound on the mean squared error (MSE) of non-Bayesian estimators is established with no unbiasedness condition on the nuisance parameters. Alternatively, Lehmann's concept of unbiasedness is imposed for a risk that measures the distance between the estimator and the locally best unbiased (LBU) estimator which assumes perfect knowledge of the nuisance parameters. The proposed bound is compared to the CRB and MSE of the maximum likelihood estimator (MLE). Simulations show that the proposed bound provides a tight lower bound for this estimator, compared with the CRB.

Index Terms— Cramér-Rao bound, Lehmann unbiasedness, risk-unbiasedness, nuisance parameters, MSE

1. INTRODUCTION

In many estimation problems one is interested in estimating parameters of interest, in the presence of other unknown parameters, referred to as nuisance. This issue has been addressed by Scharf [1, p. 231] and Kay [2, p. 431] for non-Bayesian parameter estimation problems. They have shown that the commonly used Cramér-Rao bound (CRB), introduced in [3, 4], is tighter when joint unbiased estimation of all the parameters is assumed. Gini [5] demonstrated how use of estimates of the nuisance parameters, instead of their true value, can improve the estimation mean squared error (MSE) of the parameters of interest. The improvement of the MSE was obtained by using a biased estimate of the parameter of interest. However, the same example was then used to show that by using the knowledge of the nuisance parameters, the MSE can be decreased by choosing a proper class of biased estimators, as intuition expects [6].

Due to its attractivity, variations of the CRB were derived for the Bayesian [7] and the hybrid [8] frameworks. Also, Cramér-Rao (CR)-type performance bounds for the problem of non-Bayesian parameter estimation in the presence of random nuisance parameters were vastly investigated, e.g. [8–

12]. In a recent research, a new Bayesian CR-type bound was developed for the problem of random parameter estimation in the presence of deterministic nuisance parameters [13, 14]. This bound, named risk-unbiased bound (RUB) is based on the novel concept of risk-unbiasedness, which combines two approaches [15]. The first is Lehmann-unbiasedness criterion [16], which generalizes the conventional mean-unbiasedness to arbitrary cost functions, e.g. for periodic cost functions [17–19] and for constrained parameter estimation [20]. The second is the concept of risk-unbiased prediction [21], which analyzes a new criterion for unbiasedness of Bayesian estimators when a deterministic nuisance parameter is involved.

In array processing, the effect of nuisance parameters has been mainly investigated in the context of direction-of-arrival (DOA) estimation [22, 23]. For the commonly used so called "deterministic" or "unconditional" model [24, 25], the number of parameters to be estimated grows with the number of data samples taken. Nevertheless, the MLE has been proved to be statistically efficient [26, Ch. 4].

In this paper, the problem of non-Bayesian multiple parameter estimation is addressed and a new CR-type lower bound on the MSE is derived using the covariance inequality [27, p. 113] and the concept of risk-unbiasedness. While the CRB assumes joint mean-unbiased estimation of all model parameters, a different approach is adopted in this work. Each parameter in its turn is treated as a single parameter of interest, while the other model parameters are treated as nuisance, as their misknowledge interferes with the estimation of the parameter of interest. Correspondingly, we are guided by two questions: (1) is it necessary to restrict unbiased estimation of the nuisance parameters? and (2) is there a more appropriate unbiasedness condition rather than the conventional mean-unbiasedness? These questions are answered using the concept of risk-unbiasedness.

The paper is organized as follows. Section 2 formulates the problem of deterministic parameter estimation in the presence of other deterministic nuisance parameters and the concept of risk-unbiased estimation is illustrated. This approach is then used in Section 3 to derive a new CR-type bound. Section 4 presents an example which demonstrates the main results. Finally, our conclusions appear in Section 5.

2. RISK-UNBIASEDNESS

Let $(\Omega_{\mathbf{x}}, \mathcal{F}, P_{\psi})$ denote a probability space where $\Omega_{\mathbf{x}} \subseteq \mathbb{R}^N$ is the observation space, $\mathcal{F}$ is the σ-algebra on $\Omega_{\mathbf{x}}$, and $\{P_{\psi}\}_{\psi \in \Psi}$ is a family of parameterized probability measures, such that the probability space has a finite second-order statistical moment w.r.t. P_{ψ}. The unknown deterministic vector parameter ψ is divided into two parts, such that (s.t.) $\psi = [\varphi, \boldsymbol{\theta}^T]^T$, where $\varphi \in \Phi \subseteq \mathbb{R}$ is the parameter of interest,

and $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^M$ is treated as the nuisance vector parameter. Note that, each parameter in its turn can be treated as a single parameter of interest, while the other model parameters are treated as nuisance, as their mis-knowledge interferes with the estimation of the parameter of interest. We are interested to estimate the parameter of interest φ based on the random observation vector $\mathbf{x} \in \Omega_{\mathbf{x}}$. Let $f_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\psi})$ denote the observation probability density function (pdf) of $\mathbf{x}$ parametrized by $\boldsymbol{\psi}$. $\mathrm{E}[\cdot]$ stands for the expectation operator w.r.t. $f_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\psi})$. Given a scalar b, dependent of $\mathbf{a} \in \mathbb{R}^K$, its gradient w.r.t. $\mathbf{a}$ is organized as a row vector, whose jth element is defined as $\left[\frac{db(\mathbf{a})}{d\mathbf{a}}\right]_j \triangleq \left.\frac{\partial b(\mathbf{c})}{\partial c_j}\right|_{\mathbf{c}=\mathbf{a}}$, and $\frac{db(\mathbf{a})}{d^T\mathbf{a}} \triangleq \left(\frac{db(\mathbf{a})}{d\mathbf{a}}\right)^T$. Given a vector $\mathbf{b}$, dependent of $\mathbf{a}$, its derivative w.r.t. $\mathbf{a}$ is a matrix whose j, k entry is defined as $\left[\frac{d\mathbf{b}}{d\mathbf{a}}\right]_{j,k} \triangleq \left.\frac{\partial b_j(\mathbf{c})}{\partial c_k}\right|_{\mathbf{c}=\mathbf{a}}$. The Hessian matrix of a scalar b w.r.t. $\mathbf{a}$ is a $K \times K$ matrix defined by $\frac{d^2 b(\mathbf{a})}{d\mathbf{a}d^T\mathbf{a}} \triangleq \frac{d}{d\mathbf{a}}\left(\frac{db}{d^T\mathbf{a}}\right)$. $\|\cdot\|$ indicates the Euclidean norm. We further assume the Fisher information matrix (FIM) for estimation of $\boldsymbol{\psi}$ exists and is non-singular. The FIM is defined and given in the block form

$$\mathbf{I}(\boldsymbol{\psi}) \triangleq \left[\begin{array}{cc} I_{\varphi\varphi}(\boldsymbol{\psi}) & \mathbf{I}_{\varphi\boldsymbol{\theta}}(\boldsymbol{\psi}) \\ \mathbf{I}_{\boldsymbol{\theta}\varphi}(\boldsymbol{\psi}) & \mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\psi}) \end{array}\right], \quad (1)$$

where $I_{\varphi\varphi}(\boldsymbol{\psi}) \triangleq \mathrm{E}\left[\ell_\varphi^2(\mathbf{x}; \boldsymbol{\psi})\right], \mathbf{I}_{\varphi\boldsymbol{\theta}}^T(\boldsymbol{\psi}) \triangleq \mathbf{I}_{\boldsymbol{\theta}\varphi}(\boldsymbol{\psi}) \triangleq$ $\mathrm{E}\left[\boldsymbol{\ell}_{\boldsymbol{\theta}}(\mathbf{x}; \boldsymbol{\psi})\ell_\varphi(\mathbf{x}; \boldsymbol{\psi})\right], \mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\psi}) \triangleq \mathrm{E}\left[\boldsymbol{\ell}_{\boldsymbol{\theta}}(\mathbf{x}; \boldsymbol{\psi})\boldsymbol{\ell}_{\boldsymbol{\theta}}^T(\mathbf{x}; \boldsymbol{\psi})\right]$, $\ell_\varphi(\mathbf{x}; \boldsymbol{\psi}) \triangleq \frac{\partial \log f_{\mathbf{x}}(\mathbf{x};\boldsymbol{\psi})}{\partial \varphi}$, and $\boldsymbol{\ell}_{\boldsymbol{\theta}}(\mathbf{x}; \boldsymbol{\psi}) \triangleq \frac{\partial \log f_{\mathbf{x}}(\mathbf{x};\boldsymbol{\psi})}{\partial^T \boldsymbol{\theta}}$. When $\boldsymbol{\theta}$ is known, the CRB for estimation of φ is given by $I_{\varphi\varphi}^{-1}(\boldsymbol{\psi})$, while for an unknown $\boldsymbol{\theta}$ it takes the form of

$$B_{CRB}(\boldsymbol{\psi}) \triangleq I_{\varphi\varphi}^{-1}(\boldsymbol{\psi}) +$$

$$\mathbf{I}_{\varphi\boldsymbol{\theta}}(\boldsymbol{\psi})\left(\mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\psi}) - \mathbf{I}_{\boldsymbol{\theta}\varphi}(\boldsymbol{\psi})I_{\varphi\varphi}^{-1}(\boldsymbol{\psi})\mathbf{I}_{\varphi\boldsymbol{\theta}}(\boldsymbol{\psi})\right)^{-1}\mathbf{I}_{\boldsymbol{\theta}\varphi}(\boldsymbol{\psi}), \quad (2)$$

satisfying $B_{CRB}(\boldsymbol{\psi}) \geq I_{\varphi\varphi}^{-1}(\boldsymbol{\psi})$ (see for example [1, p. 231] and [2, p. 431]). This bound assumes mean-unbiasedness for both the parameter of interest, φ, and the nuisance parameters, $\boldsymbol{\theta}$.

Denote by $\mathcal{L}^2$, the Hilbert space of absolutely square-integrable measurable functions w.r.t. $\mathcal{P}_{\boldsymbol{\psi}}$ for the measurable space $(\Omega_{\mathbf{x}}, \mathcal{F})$. The function $\hat{\varphi}(\mathbf{x})$ is an estimator of φ with estimation error $\epsilon = \hat{\varphi}(\mathbf{x}) - \varphi$. Under the MSE criterion the risk is defined as $L(\hat{\varphi}, \boldsymbol{\psi}) \triangleq \mathrm{E}[\epsilon^2]$. Let $\mathcal{U}_{\boldsymbol{\theta}} \subset \mathcal{L}^2$ denote the space of locally mean-unbiased estimators of φ, defined as (see (19)-(21) in [28] and (5),(6) in [29])

$$\mathcal{U}_{\boldsymbol{\theta}} \triangleq \left\{\hat{\varphi}(\mathbf{x}) \in \mathcal{L}^2 : \begin{array}{c} \mathrm{E}\left[\hat{\varphi}(\mathbf{x}) - \varphi\right] = 0, \\ \mathrm{E}\left[(\hat{\varphi}(\mathbf{x}) - \varphi)\,\ell_\varphi(\mathbf{x}; \boldsymbol{\psi})\right] = 1. \end{array}\right\} \quad (3)$$

We mark that the elements of $\mathcal{U}_{\boldsymbol{\theta}}$ may be functions of $\mathbf{x}$, $\boldsymbol{\theta}$, and the subjected value of φ, which is omitted from the notation for simplicity. Also, $\mathcal{U}_{\boldsymbol{\theta}}$ can be easily verified to be convex. When $\boldsymbol{\theta}$ is known, the optimal estimator in $\mathcal{U}_{\boldsymbol{\theta}}$, known as the locally best unbiased (LBU) estimator, is given by [3, 4]

$$\hat{\varphi}_{LBU}(\mathbf{x}, \boldsymbol{\theta}) = \varphi + I_{\varphi\varphi}^{-1}(\boldsymbol{\psi})\ell_\varphi(\mathbf{x}; \boldsymbol{\psi}). \quad (4)$$

The estimation error of the LBU estimator is $\epsilon_{LBU}(\mathbf{x}, \boldsymbol{\psi}) = \hat{\varphi}_{LBU}(\mathbf{x}, \boldsymbol{\psi}) - \varphi$. Using the LBU estimator, the MSE risk can

be rearranged as:

$$L(\hat{\varphi}, \boldsymbol{\theta}) = \mathrm{E}\left[\left((\hat{\varphi}(\mathbf{x}) - \hat{\varphi}_{LBU}(\mathbf{x}, \boldsymbol{\theta})) + (\hat{\varphi}_{LBU}(\mathbf{x}, \boldsymbol{\theta}) - \varphi)\right)^2\right]$$

$$= \mathrm{E}\left[\epsilon_{LBU}^2(\mathbf{x}, \boldsymbol{\theta})\right] + \mathrm{E}\left[(\hat{\varphi}(\mathbf{x}) - \hat{\varphi}_{LBU}(\mathbf{x}, \boldsymbol{\theta}))^2\right] +$$

$$2\mathrm{E}\left[\epsilon_{LBU}(\mathbf{x}, \boldsymbol{\theta})(\hat{\varphi}(\mathbf{x}) - \hat{\varphi}_{LBU}(\mathbf{x}, \boldsymbol{\theta}))\right]. \quad (5)$$

Since $\mathcal{U}_{\boldsymbol{\theta}}$ is a convex subspace of the Hilbert space $\mathcal{L}^2$, the Hilbert projection theorem [30, p. 79-80] states that the cross-term $\mathrm{E}\left[\epsilon_{LBU}(\mathbf{x}, \boldsymbol{\theta})(\hat{\varphi}(\mathbf{x}) - \hat{\varphi}_{LBU}(\mathbf{x}, \boldsymbol{\theta}))\right]$ in the right hand side (r.h.s.) of (5) is non-negative. Thus,

$$L(\hat{\varphi}, \boldsymbol{\theta}) \geq \mathrm{E}[\epsilon_{LBU}^2(\mathbf{x}, \boldsymbol{\theta})] + \mathrm{E}\left[(\hat{\varphi}(\mathbf{x}) - \hat{\varphi}_{LBU}(\mathbf{x}, \boldsymbol{\theta}))^2\right]. \quad (6)$$

The term $\mathrm{E}[\epsilon_{LBU}^2(\mathbf{x}, \boldsymbol{\theta})]$ in the r.h.s. of (6) is independent of $\hat{\varphi}(\mathbf{x})$ and is given by

$$\mathrm{E}[\epsilon_{LBU}^2(\mathbf{x}, \boldsymbol{\theta})] = I_{\varphi\varphi}^{-1}(\boldsymbol{\psi}). \quad (7)$$

Similar to the approach used in [13, 15] for the Bayesian framework, our focus is now turned to a modified risk, defined by the first term in the r.h.s. of (6), s.t.

$$R(\hat{\varphi}, \boldsymbol{\theta}) \triangleq \mathrm{E}[(\hat{\varphi}(\mathbf{x}) - \hat{\varphi}_{LBU}(\mathbf{x}, \boldsymbol{\theta}))^2]. \quad (8)$$

This modified risk redefines the estimation problem of φ via a measure of "closeness" between a valid estimator from the class $\mathcal{U}_{\boldsymbol{\theta}}$ and the optimal estimator in the class, which assumes perfect knowledge of $\boldsymbol{\theta}$. Correspondingly, the modified estimation error and the modified cost function are given by

$$z_{\hat{\varphi}}(\mathbf{x}, \boldsymbol{\theta}) \triangleq \hat{\varphi}(\mathbf{x}) - \hat{\varphi}_{LBU}(\mathbf{x}, \boldsymbol{\theta}) \quad (9a)$$

$$r_{\hat{\varphi}}(\mathbf{x}, \boldsymbol{\theta}) \triangleq z_{\hat{\varphi}}^2(\mathbf{x}, \boldsymbol{\theta}), \quad (9b)$$

respectively. We mark that the estimators $\hat{\varphi}(\mathbf{x}), \hat{\varphi}_{LBU}(\mathbf{x}, \boldsymbol{\theta}) \in \mathcal{U}_{\boldsymbol{\theta}}$ satisfy

$$\mathrm{E}\left[\hat{\varphi}(\mathbf{x}) - \varphi\right] = \mathrm{E}\left[\hat{\varphi}_{LBU}(\mathbf{x}, \boldsymbol{\theta}) - \varphi\right] = 0, \quad (10a)$$

$$\mathrm{E}\left[(\hat{\varphi}(\mathbf{x}) - \varphi)\,\ell_\varphi(\mathbf{x}; \boldsymbol{\psi})\right] =$$

$$\mathrm{E}\left[(\hat{\varphi}_{LBU}(\mathbf{x}, \boldsymbol{\theta}) - \varphi)\,\ell_\varphi(\mathbf{x}; \boldsymbol{\psi})\right] = 1. \quad (10b)$$

By subtracting the r.h.s. from the left-hand side (l.h.s.) of the first equality in each of (10a) and (10b) one obtains

$$\mathrm{E}\left[z_{\hat{\varphi}}(\mathbf{x}, \boldsymbol{\theta})\right] = 0, \quad (11a)$$

$$\mathrm{E}\left[z_{\hat{\varphi}}(\mathbf{x}, \boldsymbol{\theta})\ell_\varphi(\mathbf{x}; \boldsymbol{\psi})\right] = 0. \quad (11b)$$

In order to provide an appropriate unbiasedness criterion for the modified cost function, in addition to the constraints in (11) we utilize Lehmann's concept of unbiasedness, which was first introduced in the context of arbitrary cost functions in the non-Bayesian framework [16]. The Lehmann-unbiasedness definition [16] implies that an estimator is unbiased if on the average it is "closest" to the true parameter, φ, rather than to any other value in the parameter space, $\eta \in \Phi$. The measure of "closeness" between the estimator and the parameter is the cost function $C(\hat{\varphi}(\mathbf{x}), \varphi)$. It is shown in [17–20] that under the quadratic cost function, $C(\hat{\varphi}(\mathbf{x}), \varphi) = (\hat{\varphi}(\mathbf{x}) - \varphi)^2$, the Lehmann-unbiasedness is reduced to the conventional mean-unbiasedness, $\mathrm{E}[\hat{\varphi}(\mathbf{x})] = \varphi$. Applying Lehmann-unbiasedness condition to the risk in (8), leads to the following definition.

Definition 1. *The estimator $\hat{\varphi}(\mathbf{x}) \in \mathcal{U}_{\boldsymbol{\theta}}$ is said to be point-*

wise risk-unbiased at $\boldsymbol{\theta}$ if

$$\mathrm{E}_{\boldsymbol{\psi}}[r_{\hat{\varphi}}(\mathbf{x}, \boldsymbol{\theta})] \leq \mathrm{E}_{\boldsymbol{\psi}}[r_{\hat{\varphi}}(\mathbf{x}, \boldsymbol{\eta})] , \ \forall \boldsymbol{\eta} \in \boldsymbol{\Theta}, \qquad (12)$$

*where we used the subscript $\boldsymbol{\psi}$ in the notation of the expectation to mark the subjected value of $\boldsymbol{\psi}$. If (12) is satisfied for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and for $\boldsymbol{\theta} + \Delta\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\Delta\boldsymbol{\theta} \to \mathbf{0}_M$, then $\hat{\varphi}(\mathbf{x}) \in \mathcal{U}_{\boldsymbol{\theta}}$ is said to be **locally risk-unbiased** around $\boldsymbol{\theta}$. If (12) is satisfied $\forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$, then $\hat{\varphi}(\mathbf{x}) \in \mathcal{U}_{\boldsymbol{\theta}}$ is said to be **uniformly risk-unbiased** .*

The next theorem states some mild regularity conditions which simplify the above definition.

Theorem 1. If $\hat{\varphi}_{LBU}(\mathbf{x}, \boldsymbol{\eta})$ is once differentiable w.r.t. $\boldsymbol{\eta}$ for a.e. $\mathbf{x} \in \Omega_{\mathbf{x}}$, a necessary condition for the estimator $\hat{\varphi}(\mathbf{x})$ to be point-wise risk-unbiased at $\boldsymbol{\psi}$, is given by:

$$\mathrm{E}\left[z_{\hat{\varphi}}(\mathbf{x}, \boldsymbol{\theta})\mathbf{d}(\mathbf{x}, \boldsymbol{\psi})\right] = \mathbf{0}_M \qquad (13)$$

where $\mathbf{d}(\mathbf{x}, \boldsymbol{\psi}) \triangleq \frac{\partial \hat{\varphi}_{LBU}(\mathbf{x}, \boldsymbol{\theta})}{\partial^T \boldsymbol{\theta}}$ and $\mathbf{0}_M$ is a column vector of length M, whose entries are equal to 0. If, in addition, $\hat{\varphi}_{LBU}(\mathbf{x}, \boldsymbol{\eta})$ is twice differentiable w.r.t. $\boldsymbol{\eta}$ and $\mathrm{E}\left[r_{\hat{\varphi}}(\mathbf{x}, \boldsymbol{\eta})\right]$ is convex in $\boldsymbol{\eta}$, then the condition in (13) is also sufficient.

Proof. *The proof follows the proof of Theorem 2 in [13].*

Thus, equation (13) can be utilized as an alternative definition for point-wise risk-unbiasedness.

Following the conclusions of Theorem 1, the next proposition simplifies the condition for local risk-unbiasedness.

Proposition 2. An alternative condition for the estimator $\hat{\varphi}(\mathbf{x})$ to be locally risk-unbiased around $\boldsymbol{\theta}$ is given by (13) and

$$\mathrm{E}\left[z_{\hat{\varphi}}(\mathbf{x}, \boldsymbol{\theta})\mathbf{H}(\mathbf{x}, \boldsymbol{\theta})\right] = \mathbf{C}_{\mathbf{dd}}(\boldsymbol{\psi}), \qquad (14)$$

where

$$\mathbf{C}_{\mathbf{dd}}(\boldsymbol{\psi}) \triangleq \mathrm{E}\left[\mathbf{d}(\mathbf{x}, \boldsymbol{\psi})\mathbf{d}^T(\mathbf{x}, \boldsymbol{\psi})\right], \qquad (15a)$$

$$\mathbf{H}(\mathbf{x}, \boldsymbol{\psi}) \triangleq \frac{\partial^2 \hat{\varphi}_{LBU}(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial^T \boldsymbol{\theta}} + \mathbf{d}(\mathbf{x}, \boldsymbol{\psi})\boldsymbol{\ell}_{\boldsymbol{\theta}}^T(\mathbf{x}; \boldsymbol{\psi}), \qquad (15b)$$

are assumed to exist and to be well defined.

Proof. *The proof follows the proof of Proposition 3 in [13].*

3. MSE LOWER BOUND FOR RISK-UNBIASED ESTIMATORS

In this section, a non-Bayesian bound on the MSE of risk-unbiased estimators is derived. The covariance inequality [27, p. 113] is given by:

$$\mathrm{E}[u^2] \geq \mathrm{E}[u\mathbf{v}^T]\mathrm{E}^{-1}[\mathbf{v}\mathbf{v}^T]\mathrm{E}[\mathbf{v}u]. \qquad (16)$$

By setting $u = z_{\hat{\varphi}}(\mathbf{x}, \boldsymbol{\theta})$, the l.h.s. of (16) turns into the modified risk while the r.h.s. constitutes a lower bound. Thus, back-substitution of (16) with (7) into (6) results in a lower bound for the MSE of $\hat{\varphi}(\mathbf{x})$, given by

$$L(\hat{\varphi}, \boldsymbol{\theta}) \geq B_{RUCRB}(\varphi, \boldsymbol{\theta}) \triangleq$$
$$I_{\varphi\varphi}^{-1}(\boldsymbol{\psi}) + \mathrm{E}[u\mathbf{v}^T]\mathrm{E}^{-1}[\mathbf{v}\mathbf{v}^T]\mathrm{E}[\mathbf{v}u]. \qquad (17)$$

Denote $\mathbf{h}(\mathbf{x}, \boldsymbol{\theta}) = \text{vec}(\mathbf{H}(\mathbf{x}, \boldsymbol{\theta}))$, where $\text{vec}(\cdot)$ stands for the vectorization operation. Following the constraints on $z_{\hat{\varphi}}(\mathbf{x}, \boldsymbol{\theta})$ in (10), (13), and (14), we set $\mathbf{v} = \left[1, \ell_{\varphi}(\mathbf{x}; \boldsymbol{\psi}), \mathbf{d}^T(\mathbf{x}, \boldsymbol{\theta}), \mathbf{h}^T(\mathbf{x}, \boldsymbol{\theta})\right]^T$, s.t.

$$\mathrm{E}[u\mathbf{v}^T] = \left[0, 0, \mathbf{0}_M^T, vec^T(\mathbf{C}_{\mathbf{dd}}(\boldsymbol{\psi}))\right], \qquad (18)$$

and (19) on the bottom of this page, where $\mathbf{C}_{\mathbf{hd}}(\boldsymbol{\psi}) \triangleq \mathbf{C}_{\mathbf{dh}}^T(\boldsymbol{\psi}) \triangleq \mathrm{E}[\mathbf{h}(\mathbf{x}, \boldsymbol{\psi})\mathbf{d}^T(\mathbf{x}, \boldsymbol{\psi})]$, and $\mathbf{C}_{\mathbf{hh}}(\boldsymbol{\psi}) \triangleq \mathrm{E}[\mathbf{h}(\mathbf{x}, \boldsymbol{\psi})\mathbf{h}^T(\mathbf{x}, \boldsymbol{\psi})]$. Substituting (18) and (19) into (17) yields a lower bound on the MSE locally risk-unbiased estimators around $\boldsymbol{\theta}$, named risk-unbiased CRB (RUCRB).

The conventional CRB is based on the derivative of the observation likelihood function. It is used as a measure of the sensitivity of the observations distribution to the variations in the neighborhood of the deterministic parameters of interest. Equations (17)-(19) show that the RUCRB provides a small error bound, which is based on local variations. Like the CRB, it utilizes the derivative of the observation likelihood function. However, it also incorporates the first- and the second-order derivatives of the LBU estimator. These two additions are used as a measure of the sensitivity of the LBU estimator to the perturbations around the deterministic nuisance parameters. Hence, the proposed bound provides prediction of the performance for estimators that mimic the dependency of the LBU estimator on the deterministic nuisance parameters.

4. EXAMPLE - SIGNAL ESTIMATION

In this example, we examine the problem of source signal estimation using an array of P sensors. Consider the following observation model:

$$\mathbf{x}_n = \mathbf{v}s_n + \mathbf{w}_n, \ n = 1, \ldots, N, \qquad (20)$$

where $\mathbf{s} = [s_1, \ldots, s_N]^T \in \mathbb{C}^N$ is a sequence of unknown deterministic variables, $\mathbf{v} \in \mathbb{C}^P$ is an unknown deterministic normalized steering vector, $\{\mathbf{w}_n\}_{n=1}^N$ is a white complex proper Gaussian random noise vector sequence with a known covariance matrix $\sigma_w^2 \mathbf{I}_P$, and $\mathbf{I}_P$ is an identity matrix of size P. Since both $\mathbf{s}$ and $\mathbf{v}$ are unknown, this model consists of ambiguity, as these sizes can be estimated up to a complex scaling factor. Several approaches may be used to overcome this difficulty, such as assuming a constant Euclidean norm of $\mathbf{v}$. However, this would require using the tools of constrained parameter estimation. Instead, we choose to refer to $\mathbf{v}$ as being normalized by its first element, s.t. without loss of generality $v_1 = 1$. The vector of the parameters of interest is given by $\boldsymbol{\varphi} \triangleq \left[s_{1_r}, s_{1_q} \ldots, s_{N_r}, s_{N_q}\right]^T \in \mathbb{R}^{2N}$, where the subscripts r and q denote the real and imaginary parts, respectively. The RUCRB for estimation of $\boldsymbol{\varphi}$

$$\mathrm{E}\left[\mathbf{v}\mathbf{v}^T\right] = \begin{bmatrix} 1 & 0 & \mathrm{E}\left[\mathbf{d}^T(\mathbf{x}, \boldsymbol{\psi})\right] & \left[\mathbf{h}^T(\mathbf{x}, \boldsymbol{\psi})\right] \\ 0 & I_{\varphi\varphi}(\boldsymbol{\psi}) & \mathrm{E}\left[\mathbf{d}^T(\mathbf{x}, \boldsymbol{\psi})\ell_{\varphi}(\mathbf{x}; \boldsymbol{\psi})\right] & \left[\mathbf{h}^T(\mathbf{x}, \boldsymbol{\psi})\ell_{\varphi}(\mathbf{x}; \boldsymbol{\psi})\right] \\ \mathrm{E}\left[\mathbf{d}(\mathbf{x}, \boldsymbol{\psi})\right] & \mathrm{E}\left[\mathbf{d}(\mathbf{x}, \boldsymbol{\psi})\ell_{\varphi}(\mathbf{x}; \boldsymbol{\psi})\right] & \mathbf{C}_{\mathbf{dd}}(\boldsymbol{\psi}) & \mathbf{C}_{\mathbf{dh}}(\boldsymbol{\psi}) \\ \mathrm{E}\left[\mathbf{h}(\mathbf{x}, \boldsymbol{\psi})\right] & \mathrm{E}\left[\mathbf{h}(\mathbf{x}, \boldsymbol{\psi})\ell_{\varphi}(\mathbf{x}; \boldsymbol{\psi})\right] & \mathbf{C}_{\mathbf{hd}}(\boldsymbol{\psi}) & \mathbf{C}_{\mathbf{hh}}(\boldsymbol{\psi}) \end{bmatrix} \qquad (19)$$

is evaluated by summing over the RUCRBs obtained from setting φ as s_n, for each $n = 1, \ldots, N$. The LBU estimator of s_n from $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1^T, \ldots, \mathbf{x}_N^T \end{bmatrix}^T$ for a known value of $\boldsymbol{\theta} = \begin{bmatrix} v_{2_r}, v_{2_q} \ldots, v_{P_r}, v_{P_q} \end{bmatrix}^T \in \mathbb{R}^{2(P-1)}$ can be verified to be given by

$$\hat{s}_{n_{LBU}}(\mathbf{x}, \boldsymbol{\theta}) = \frac{\mathbf{v}^H \mathbf{x}}{\|\mathbf{v}\|^2}, \tag{21}$$

where the superscript ^{H} denotes the conjugate transpose operator. The logarithm of the observation pdf of $\mathbf{x}$ is

$$\log f(\mathbf{x}; \mathbf{s}, \mathbf{v}) = -NP \log(\pi \sigma_w^2) - \sum_{n=1}^{N} \frac{\|\mathbf{x}_n - \mathbf{v} s_n\|^2}{\sigma_w^2}. \tag{22}$$

Following [31], one can evaluate the FIM for estimation of $\boldsymbol{\psi}$. By taking the inverse of the FIM, the CRB for estimation of $\mathbf{s}$ can be verified to take the form

$$B_{CRB}(s_n) = \frac{\sigma_w^2}{1 - \frac{|s_n|^2}{\|\mathbf{s}\|^2}}, \; \forall n = 1, \ldots, N. \tag{23}$$

The proposed bound can be computed by using (17), (21), and (23). For each experiment, the MLE of $\mathbf{s}$ is obtained by maximizing $\log f_{\mathbf{x}}(\mathbf{x}; \mathbf{s}, \mathbf{v})$ w.r.t. $\mathbf{s}$ and $\mathbf{v}$, s.t.

$$\begin{aligned} \widehat{\mathbf{v}}_{ML} &= \arg \max_{\mathbf{v}: v_1 = 1} \sum_{n=1}^{N} \frac{|\mathbf{v}^H \mathbf{x}_n|^2}{\|\mathbf{v}\|^2}, \\ \hat{s}_{n_{ML}} &= \frac{\widehat{\mathbf{v}}_{ML}^H \mathbf{x}_n}{\|\widehat{\mathbf{v}}_{ML}\|^2}, \; \forall n = 1, \ldots, N. \end{aligned} \tag{24}$$

Fig. 1 presents the MSE of the MLE, the CRB, and the proposed RUCRB for estimation of s_1 versus N. The MSE was evaluated using 100,000 Monte-Carlo simulations with $\sigma_w^2 = 1$, $P = 2$, $\mathbf{s} = \mathbf{1}_N$, and $\mathbf{v} = \mathbf{1}_P$, where $\mathbf{1}_N$ is a column vector of length N, whose entries are equal to 1. The RUCRB provides a tight and a valid bound for all sample sizes. As the number of measurements increases, the CRB and the RUCRB coincide and provide a tight and asymptotically achievable lower bound for the MLE, as expected. However, when turn-
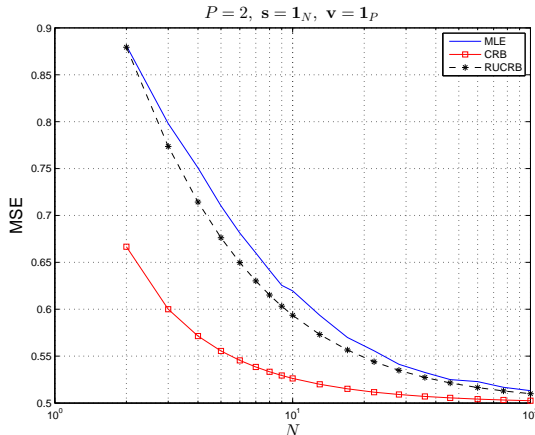

Fig. 1. The CRB, the proposed RUCRB, and the MSE of MLE for $SNR = 0dB$.

ing to the other side of asymptotic performance analysis, that is for large SNR values, a different picture is drawn. The normalized MSE (that is, the MSE divided by σ_w^2) of the MLE, the CRB, and the proposed RUCRB for estimation of $\mathbf{s}$ versus SNR are presented in Fig. 2, where $SNR \triangleq \frac{|s_1|^2}{\sigma_w^2}$. The

MSE was evaluated using 100,000 Monte-Carlo simulations with $N = 4$, $P = 2$, $\mathbf{s} = \mathbf{1}_N$, and $\mathbf{v} = \mathbf{1}_P$. While the CRB does not provide a tight bound for the MSE of the MLE, the RUCRB provides a tight and asymptotically achievable lower bound for the MLE, for all SNR values.
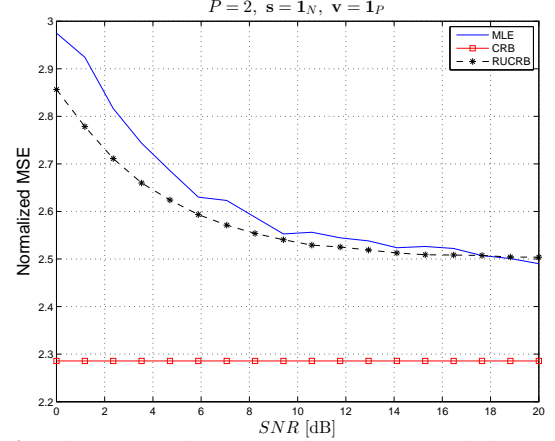

Fig. 2. The CRB, the proposed RUCRB, and the MSE of MLE for $N = 4$.

5. CONCLUSION

In this paper, the problem of multiple parameter estimation under the non-Bayesian framework is explored and a new Cramér-Rao-type bound for the MSE is developed. Unlike the CRB, the proposed bound does not assume unbiasedness for all the model parameters. Alternatively, the proposed RU-CRB treats a single parameter as the parameter of interest, and the other parameters as nuisance. The bound assumes risk-unbiased estimation which is more appropriate for the case of multiple parameters. It was shown that for the problem of source signal estimation using an array of sensors, the proposed bound provides a tight and valid bound on the performance of the MLE, while the CRB is not tight.

6. REFERENCES

[1] L. L. Scharf. *Statistical Signal Processing.*, vol. 98, Addison-Wesley, Reading, MA, 1991.

[2] S. M. Kay. *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory.*, Prentice-Hall, Upper Saddle River, NJ, 1993.

[3] H. Cramér. *Mathematical Methods of Statistics.*, Princeton Univ. Press, Princeton, NJ, 1946.

[4] C. R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, vol. 37, no. 3, pp. 81–91, 1945.

[5] F. Gini, "Estimation strategies in the presence of nuisance parameters," *Signal Process.*, vol. 55, no. 2, pp. 241 – 245, Dec. 1996.

[6] J. Beyerer, "Is it useful to know a nuisance parameter?," *Signal Process.*, vol. 68, no. 1, pp. 107 – 111, July 1998.

[7] H. L. Van Trees. *Detection, Estimation, and Modulation Theory—Part l. Detection, Estimation, and Linear Modulation Theory.*, John Wiley & Sons, 2001.

[8] Y. Noam and H. Messer, "Notes on the tightness of the hybrid Cramér-Rao lower bound," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2074–2084, June 2009.

[9] R. Miller and C. Chang, "A modified Cramér-Rao bound and its applications," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 398–400, May 1978.

[10] H. Messer and Y. Adar, "Lower bounds on bearing estimation errors of any number of narrowband far-field sources," in *Proc. ICASSP 1989*, May 1989, vol. 4, pp. 2113–2116.

[11] F. Gini, R. Reggiannini, and U. Mengali, "The modified Cramér-Rao bound in vector parameter estimation," *IEEE Trans. Commun.*, vol. 46, no. 1, pp. 52–60, Jan. 1998.

[12] M. Moeneclaey, "On the true and the modified Cramér-Rao bounds for the estimation of a scalar parameter in the presence of nuisance parameters," *IEEE Trans. Commun.*, vol. 46, no. 11, pp. 1536–1544, Nov. 1998.

[13] S. Bar and J. Tabrikian, "Bayesian estimation in the presence of deterministic nuisance parameters—Part I: Performance bounds," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6632–6646, 2015.

[14] S. Bar and J. Tabrikian, "Bayesian estimation in the presence of deterministic nuisance parameters—Part II: Estimation methods," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6647–6658, 2015.

[15] T. K. Nayak and M. Qin, "The concept of risk unbiasedness in statistical prediction," *Journal of Statistical Planning and Inference*, vol. 140, no. 7, pp. 1923–1938, Jul. 2010.

[16] E. L. Lehmann, "A general concept of unbiasedness," *The Annals of Mathematical Statistics*, vol. 22, no. 4, pp. 587–592, Dec. 1951.

[17] T. Routtenberg and J. Tabrikian, "Periodic CRB for non-Bayesian parameter estimation," in *Proc. ICASSP 2011*, May 2011, pp. 2448–2451.

[18] T. Routtenberg and J. Tabrikian, "Non-Bayesian periodic Cramér-Rao bound," *IEEE Trans. Signal Process.*, vol. 61, no. 4, pp. 1019–1032, Feb. 2013.

[19] T. Routtenberg and J. Tabrikian, "Cyclic Barankin-type bounds for non-Bayesian periodic parameter estimation," *IEEE Trans. Signal Process.*, vol. 62, no. 13, pp. 3321–3336, Jul. 2014.

[20] T. Routtenberg and J. Tabrikian, "Performance bounds for constrained parameter estimation," in *Proc. SAM 2012*, June 2012, pp. 513–516.

[21] Y. Xiao, "Linex unbiasedness in a prediction problem," *Annals of the Institute of Statistical Mathematics*, vol. 52, no. 4, pp. 712–721, Dec. 2000.

[22] P. Stoica and Nehorai A., "Music, maximum likelihood, and cramer-rao bound," *IEEE Trans. Audio, Speech, Language Process.*, vol. 37, no. 5, pp. 720–741, May 1989.

[23] H. Clergeot, S. Tressens, and A. Ouamri, "Performance of high resolution frequencies estimation methods compared to the cramer-rao bounds," *IEEE Trans. Audio, Speech, Language Process.*, vol. 37, no. 11, pp. 1703–1720, Nov 1989.

[24] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, 1996.

[25] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 38, no. 10, pp. 1783–1795, Oct. 1990.

[26] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai. *Exact and Large Sample Maximum Likelihood Techniques for Parameter Estimation and Detection in Array Processing*, Springer, 1993.

[27] E. L. Lehmann and G. Casella. *Theory of Point Estimation.*, Springer, 1998.

[28] T. Menni, E. Chaumette, P. Larzabal, and J.P. Barbot, "New results on deterministic Cramér-Rao bounds for real and complex parameters," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1032–1049, Mar. 2012.

[29] E. Chaumette, F. Vincent, J. Galy, and P. Larzabal, "On the influence of detection tests on deterministic parameters estimation," in *Proc. EUSIPCO 2006*, Sep. 2006, pp. 1–5.

[30] W. Rudin. *Real and Complex Analysis.*, McGraw-Hill, Inc., New York, NY, USA, third edition, 1987.

[31] B. Porat and B. Friedlander, "Computation of the exact information matrix of Gaussian time series with stationary random components," *IEEE Trans. Audio, Speech, Language Process.*, vol. 34, no. 1, pp. 118–130, Feb. 1986.