GENERALIZED BARANKIN-TYPE LOWER BOUNDS FOR MISSPECIFIED MODELS

M. L. Diong (1)*, E.Chaumette*(1) *and F. Vincent*(1)

(1) Isae-Supaero, Université de Toulouse, 10 av. Ed. Belin, Toulouse, France ([mouhamadou.diong, eric.chaumette, francois.vincent]@isae.fr)

ABSTRACT

When the assumed probability distribution of the observations differs from the true distribution, the model is said to be misspecified. The key results on maximum-likelihood estimation of misspecified models have been introduced in the limit of large sample support and depend on a parameters vector solution of a computationally expensive non-linear optimization problem. As a possible strategy to circumvent these limitations, we extend the approach lately proposed by Fritsche et al [1]. It is shown that the lower bound derived in [1] is a representative of a family of lower bounds deriving from a misspecified unbiasedness constraint leading to *generalized* Barankin-type lower bounds. For future use, we derive the standard representative of the "Small Errors" and "Large Errors" bounds, namely the *generalized* CRB and the *generalized* McAulay-Seidman bound.

***Index Terms*—** Maximum likelihood estimation, misspecified model, Cramér-Rao bound, Barankin bound

1. INTRODUCTION

In deterministic parameters estimation, the widespread use of maximum likelihood estimators (MLEs) originates from the fact that, under reasonably general conditions on the probabilistic observation model [2][3], MLEs are, in the limit of large sample support, Gaussian distributed and consistent[1]. However, a fundamental assumption underlying the above classical results on the properties of MLEs is that the probability distribution function (p.d.f.) which determines the behavior of the observations, for instance T i.i.d. M-dimensional complex random vectors $\{\mathbf{x}_t\}_{t=1}^{T}$, is assumed to lie within a specified parametric family of p.d.f. denoted $f_{\boldsymbol{\theta}}(\mathbf{x}_t) \triangleq f(\mathbf{x}_t|\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \mathbb{R}^P$. In other words, the probability model is assumed to be "correctly specified". Actually, in many (if not most) circumstances, a certain amount of mismatch between the true p.d.f. of the observations denoted $p(\mathbf{x}_t)$ and the probability model $f_{\boldsymbol{\theta}}(\mathbf{x}_t)$ that we assume is present. As a consequence, it is natural to investigate what happens to the properties of MLEs if the probability model is misspecified, i.e. not correctly specified. Huber [5] explored in detail the performance of MLEs (1a) in the limit of large sample support under very general assumptions on misspecification, proved consistency, normality, and derived the MLEs asymptotic covariance that is often referred to as the Huber's "sandwich covariance" in literature (1c). Later, Akaike [6] observed that when the true distribution is unknown, MLEs are natural estimators for the parameters vector $\boldsymbol{\theta}_f$ (1b) which minimizes the Kullback-Leibler information criterion (KLIC) [7] between the true and the assumed probability model. Last, White [8] provided simple conditions under which MLEs are strongly consistent estimators for the parameters vector which minimizes the KLIC. Interestingly enough, a covariance matrix is the tightest lower bound (LB) on itself since it satisfies the covariance inequality (1d) [9][10], so-called the Huber "sandwich" inequality in the present case. It is probably the reason why the derivation of additional misspecified LBs have received little consideration in the literature [11], apart from the misspecified Cramér-Rao Bound (MCRB) [12]-[14]. However, any LB deriving from the Huber sandwich inequality, including the asymptotic covariance matrix and the MCRB, depends on $\boldsymbol{\theta}_f$. Thus, its numerical evaluation requires to solve a non-linear multidimensional optimization problem for each value of $\boldsymbol{\theta}$, a procedure suffering from a large computational cost as the dimension of $\boldsymbol{\theta}$ increases. Moreover, all the results mentioned above hold only in the limit of large sample support.

A possible strategy to circumvent these limitations is the alternative approach proposed in [1], where a so-called Cramér-Rao bound (CRB) is derived for the class of estimators that are unbiased or that have a specified bias (gradient) w.r.t. the assumed model $f_{\boldsymbol{\theta}}(\mathbf{x}_t)$, in the restricted case where $p(\mathbf{x}_t|\,)$ and $f(\mathbf{x}_t|\,)$ share the same parameterization: $p(\mathbf{x}_t) \triangleq p_{\boldsymbol{\theta}}(\mathbf{x}_t) \triangleq p(\mathbf{x}_t|\boldsymbol{\theta})$.

In the present paper, it is shown that the LB derived in [1] is a particular case of a family of LBs deriving from a misspecified unbiasedness (or biasedness) constraint leading to *generalized* Barankin-type LBs. The *generalized* LBs hold even if the true parametric model is unknown, i.e. we do not have prior information on the particular parameterization of the true distribution, and in any region of operation of MLEs yielding unbiased estimates w.r.t. the assumed model $f_{\boldsymbol{\theta}}(\mathbf{x}_t)$. In particular, in the limit of large sample support and in some additional asymptotic regions of operation at finite sample support when the observation model is Gaussian [15]-[19]. Actually, the LB derived in [1] is not the *generalized* CRB but a looser *generalized* LB. Last, we examine the relevance of the *generalized* LBs by computing standard representative of the "Small Errors" and "Large Errors" bounds, namely the *generalized* CRB and the *generalized* McAulay-Seidman bound, in the case of linear models.

1.1. Relation to prior work

It is shown that the LB proposed in [1] is a representative of a family of LBs deriving from a misspecified unbiasedness (or biasedness) constraint leading to *generalized* Barankin-type LBs, valid even if the true parametric model is unknown. For future use, we derive the standard representative of the "Large Errors" and "Small Errors" bounds, namely the *generalized* McAulay-Seidman bound and the *generalized* CRB, which is actually a tighter bound than the bound released in [1].

2. BACKGROUND ON MLES UNDER MISSPECIFICATION

As mentioned in the introduction, several authors [5][6][8] has contributed to show that, under mild regularity conditions given in [8] (and summarized in [14, Section II.A]), the misspecified MLE (MMLE) defined as:

This work has been partially supported by the DGA/MRIS (2015.60.0090.00.470.75.01).

[1] If the MLEs are consistent then they are also asymptotically efficient [4].

$$\widehat{\boldsymbol{\theta}}(\overline{\mathbf{x}}) = \arg\max_{\boldsymbol{\theta}} \left\{ f_{\boldsymbol{\theta}}(\overline{\mathbf{x}}) = \prod_{t=1}^{T} f_{\boldsymbol{\theta}}(\mathbf{x}_t) \right\}, \tag{1a}$$

is, in the limit of large sample support ($T \to \infty$), a strongly consistent estimator for the parameters vector which minimizes the KLIC:

$$\widehat{\boldsymbol{\theta}}(\overline{\mathbf{x}}) \overset{a.s.}{\to} \boldsymbol{\theta}_f = \arg\min_{\boldsymbol{\theta}} \left\{ E_p \left[\ln(p(\mathbf{x}_t)) - \ln(f_{\boldsymbol{\theta}}(\mathbf{x}_t)) \right] \right\} \tag{1b}$$

where $\overline{\mathbf{x}} = \left(\mathbf{x}_1^T, \ldots, \mathbf{x}_T^T\right)^T$, $p(\overline{\mathbf{x}}) = \prod_{t=1}^{T} p(\mathbf{x}_t)$ and $E_p[\mathbf{g}(\overline{\mathbf{x}})] = \int_{\mathbb{C}^{M\times T}} \mathbf{g}(\overline{\mathbf{x}})\, p(\overline{\mathbf{x}})\, d\overline{\mathbf{x}}$. Moreover $\widehat{\boldsymbol{\theta}}(\overline{\mathbf{x}})$ is asymptotically normal: $\widehat{\boldsymbol{\theta}}(\overline{\mathbf{x}}) \overset{A}{\sim} \mathcal{N}\left(\boldsymbol{\theta}_f, \mathbf{C}_{\widehat{\boldsymbol{\theta}}}\right)$, $\mathbf{C}_{\widehat{\boldsymbol{\theta}}} \overset{a.s.}{\to} \mathbf{C}_{HS}(\boldsymbol{\theta}_f)$, where the asymptotic covariance matrix $\mathbf{C}_{HS}(\boldsymbol{\theta}_f)$, the so-called Huber's "sandwich covariance", is given by:

$$T\mathbf{C}_{HS}(\boldsymbol{\theta}_f) = E_p\left[\frac{\partial^2 \ln f(\mathbf{x}_t|\boldsymbol{\theta}_f)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right]^{-1} \times \tag{1c}$$
$$E_p\left[\frac{\partial \ln f(\mathbf{x}_t|\boldsymbol{\theta}_f)}{\partial \boldsymbol{\theta}} \frac{\partial \ln f(\mathbf{x}_t|\boldsymbol{\theta}_f)}{\partial \boldsymbol{\theta}^T}\right] E_p\left[\frac{\partial^2 \ln f(\mathbf{x}_t|\boldsymbol{\theta}_f)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right]^{-1}.$$

A covariance matrix is the tightest LB on itself since it satisfies the covariance inequality [9][10]. Thus $\forall \boldsymbol{\eta}(\overline{\mathbf{x}})$:

$$\mathbf{C}_{HS}(\boldsymbol{\theta}_f) \geq E_p\left[\left(\widehat{\boldsymbol{\theta}}(\overline{\mathbf{x}}) - \boldsymbol{\theta}_f\right) \boldsymbol{\eta}(\overline{\mathbf{x}})^T\right] E_p\left[\boldsymbol{\eta}(\overline{\mathbf{x}})\, \boldsymbol{\eta}(\overline{\mathbf{x}})^T\right]^{-1}$$
$$\times E_p\left[\boldsymbol{\eta}(\overline{\mathbf{x}}) \left(\widehat{\boldsymbol{\theta}}(\overline{\mathbf{x}}) - \boldsymbol{\theta}_f\right)^T\right], \tag{1d}$$

also called the Huber's "sandwich" (covariance) inequality. Note that $\mathbf{C}_{HS}(\boldsymbol{\theta}_f)$ (1c) is obtained for $\boldsymbol{\eta}(\overline{\mathbf{x}}) = \frac{\partial \ln f(\overline{\mathbf{x}}|\boldsymbol{\theta}_f)}{T \partial \boldsymbol{\theta}}$ [13]. However, any lower bound deriving from the Huber sandwich inequality, including (1c), depends on $\boldsymbol{\theta}_f$. As a consequence, its numerical evaluation requires to solve a non-linear multidimensional optimization problem (1b) for each value of $\boldsymbol{\theta}$, a procedure suffering from a large computational cost when the dimension of $\boldsymbol{\theta}$ increases.

3. GENERALIZED BARANKIN-TYPE LOWER BOUNDS FOR MISSPECIFIED MODELS

For the sake of legibility, we primarily focus on the estimation of a single unknown real deterministic parameter θ, although the results are easily extended to the estimation of multiple functions of multiple parameters [20][21] (see the generalized CRB below for an example). Let us denote $E_\theta[\mathbf{g}(\overline{\mathbf{x}})] \triangleq E_{f_\theta}[\mathbf{g}(\overline{\mathbf{x}})] = \int_{\mathbb{C}^{M\times T}} \mathbf{g}(\overline{\mathbf{x}})\, f_\theta(\overline{\mathbf{x}})\, d\overline{\mathbf{x}}$.

3.1. On Lower Bounds and Norm Minimization

In this subsection the assumed p.d.f. $f_\theta(\overline{\mathbf{x}})$ coincides with the true p.d.f. of the observations $\overline{\mathbf{x}}$.

In the search for a LB on the mean square error (MSE) of unbiased estimators, two fundamental properties of the problem at hand, introduced by Barankin [22], must be noticed. The first property is that the MSE of a particular estimator $\widehat{\theta^0}$ of θ^0, $\widehat{\theta^0} \triangleq \widehat{\theta^0}(\overline{\mathbf{x}}) \in L^2\left(\mathbb{C}^{M\times T}\right)$, where θ^0 is a selected value of the parameter θ, is a norm associated with a particular scalar product $\langle u(\overline{\mathbf{x}}) \mid v(\overline{\mathbf{x}}) \rangle_\theta = E_\theta[u(\overline{\mathbf{x}})\, v(\overline{\mathbf{x}})]$:

$$MSE_{\theta^0}\left[\widehat{\theta^0}\right] = \left\| \widehat{\theta^0}(\overline{\mathbf{x}}) - \theta^0 \right\|_{\theta^0}^2. \tag{2}$$

The second property is that an unbiased estimator $\widehat{\theta^0}$ of θ should be uniformly unbiased, i.e. if Θ denotes the parameter space:

$$\forall \theta \in \Theta : E_\theta\left[\widehat{\theta^0}(\overline{\mathbf{x}})\right] = \theta. \tag{3a}$$

If the support of $f_\theta(\overline{\mathbf{x}})$ does not depend on θ, i.e. $\Omega(\theta) = \left\{\overline{\mathbf{x}} \in \mathbb{C}^{M\times T} \mid f(\overline{\mathbf{x}}|\theta) > 0\right\} \triangleq \Omega$, then (3a) can be recasted as:

$$\forall \theta \in \Theta : E_{\theta^0}\left[\left(\widehat{\theta^0}(\overline{\mathbf{x}}) - \theta^0\right) \upsilon_{\theta^0}(\overline{\mathbf{x}}; \theta)\right] = \theta - \theta^0, \tag{3b}$$

where $\upsilon_{\theta^0}(\overline{\mathbf{x}}; \theta) = \frac{f_\theta(\overline{\mathbf{x}})}{f_{\theta^0}(\overline{\mathbf{x}})}$ denotes the likelihood ratio (LR). As a consequence, the locally-best (at θ^0) unbiased estimator is the solution of a norm minimization under linear constraints:

$$\min\left\{\left\|\widehat{\theta^0}(\overline{\mathbf{x}}) - \theta^0\right\|_{\theta^0}^2\right\} \text{ under}$$
$$\forall \theta \in \Theta : \left\langle \widehat{\theta^0}(\overline{\mathbf{x}}) - \theta^0 \mid \upsilon_{\theta^0}(\overline{\mathbf{x}}; \theta) \right\rangle_{\theta^0} = \theta - \theta^0. \tag{4}$$

Unfortunately, if Θ contains a continuous subset of $\mathbb{R}$, then (4) leads to an integral equation with no analytical solution in general [20][23]. Therefore, since the seminal work of Barankin [22], many studies quoted in [10][20][21]have been dedicated to the derivation of "computable" LBs approximating the Barankin bound (BB), i.e. the MSE of the locally-best unbiased estimator. All these approximations derive from sets of discrete or integral linear transform of the "Barankin" constraint (3b) and can be easily obtained using the following well known norm minimization lemma [20]. Let $\mathbb{U}$ be an Euclidean vector space on the body of real numbers $\mathbb{R}$ which has a scalar product $\langle \mid \rangle$. Let $(\mathbf{c}_1, \ldots, \mathbf{c}_K)$ be a free family of K vectors of $\mathbb{U}$ and $\mathbf{v} \in \mathbb{R}^K$. The problem of the minimization of $\|\mathbf{u}\|^2$ under the K linear constraints $\langle \mathbf{u} \mid \mathbf{c}_k \rangle = v_k$, $k \in [1, K]$, then has the solution:

$$\min\left\{\|\mathbf{u}\|^2\right\} = \mathbf{v}^T \mathbf{R}^{-1} \mathbf{v}, \quad \mathbf{R}_{n,k} = \langle \mathbf{c}_k \mid \mathbf{c}_n \rangle. \tag{5}$$

3.2. On Lower Bounds under Unbiasedness Misspecification

Under reasonably general conditions on the parametric p.d.f. $f_\theta(\overline{\mathbf{x}})$ [2, pp. 500-503], the MLE $\widehat{\theta^0_{ML}}$ of θ^0 is, in the limit of large sample support, consistent, uniformly unbiased with respect to $f_\theta(\overline{\mathbf{x}})$:

$$\forall \theta \in \Theta : E_\theta\left[\widehat{\theta^0_{ML}}(\overline{\mathbf{x}})\right] = \theta, \tag{6a}$$

Gaussian distributed and efficient. Additionally, if the observation model is Gaussian, some additional asymptotic regions of operation yielding Gaussian, consistent and uniformly unbiased (6a) MLEs have also been identified at finite sample support [15][16][17][18][19].

However if $f(\overline{\mathbf{x}}|\theta)$ is not the true p.d.f. of the observations, then (6a) is no longer the uniform unbiasedness constraint (3a) but a given linear constraint:

$$\int_\Omega \widehat{\theta^0_{ML}}(\overline{\mathbf{x}})\, f_\theta(\overline{\mathbf{x}})\, d\overline{\mathbf{x}} = \theta, \ \int_\Omega f_\theta(\overline{\mathbf{x}})\, d\overline{\mathbf{x}} = 1, \tag{6b}$$

where the constraint vector $f_\theta(\overline{\mathbf{x}})$ has a normalized integral. By the way, as $f_\theta(\overline{\mathbf{x}})$ is a p.d.f., it makes sense to regard (6a-6b) as a misspecification of the uniform unbiasedness property.

Additionally, if $\left\{\overline{\mathbf{x}} \in \mathbb{C}^{M\times T} \mid p(\overline{\mathbf{x}}) > 0\right\} \triangleq \Omega$, i.e., if the supports of the assumed p.d.f. $f_\theta(\overline{\mathbf{x}})$ and the true p.d.f. $p(\overline{\mathbf{x}})$ are identical, then, any estimator $\widehat{\theta^0}$ verifying (6a) satisfies,

$$\forall\theta\in\Theta: E_p\left[\left(\widehat{\theta^0}(\overline{\mathbf{x}})-\theta^0\right)\omega_p(\overline{\mathbf{x}};\theta)\right]=\theta-\theta^0, \quad (6c)$$

where $\omega_p(\overline{\mathbf{x}};\theta)=\frac{f_\theta(\overline{\mathbf{x}})}{p(\overline{\mathbf{x}})}$, and (4) becomes:

$$\min\left\{MSE_p\left[\widehat{\theta^0}\right]=\left\|\widehat{\theta^0}(\overline{\mathbf{x}})-\theta^0\right\|_p^2\right\} \text{ under}$$
$$\forall\theta\in\Theta:\left\langle\widehat{\theta^0}(\overline{\mathbf{x}})-\theta^0 \mid \omega_p(\overline{\mathbf{x}};\theta)\right\rangle_p=\theta-\theta^0, \quad (6d)$$

where $\|u(\overline{\mathbf{x}})\|_p^2 = \langle u(\overline{\mathbf{x}}) \mid u(\overline{\mathbf{x}})\rangle_p$, and $\langle u(\overline{\mathbf{x}}) \mid v(\overline{\mathbf{x}})\rangle_p = E_p[u(\overline{\mathbf{x}})v(\overline{\mathbf{x}})]$. Then, we can build upon the rationale introduced in [20, Section II.B] to propose a *generalization* of the BB under unbiasedness misspecification[2] and an unknown true parametric p.d.f. model $p(\overline{\mathbf{x}})$.

Let $\boldsymbol{\theta}^N=\left(\theta^1,\ldots,\theta^N\right)^T\in\Theta^N$ be a vector of N selected values of the parameter θ (aka test points), $\boldsymbol{\xi}^N=\left(\theta^1-\theta^0,\ldots,\theta^N-\theta^0\right)^T$ and $\boldsymbol{\omega}_p\left(\overline{\mathbf{x}};\boldsymbol{\theta}^N\right)=\left(\omega_p\left(\overline{\mathbf{x}};\theta^1\right),\ldots,\omega_p\left(\overline{\mathbf{x}};\theta^N\right)\right)^T$. Any estimator $\widehat{\theta^0}$ verifying (6a)(6c) must comply with:

$$E_p\left[\left(\widehat{\theta^0}(\overline{\mathbf{x}})-\theta^0\right)\boldsymbol{\omega}_p\left(\overline{\mathbf{x}};\boldsymbol{\theta}^N\right)\right]=\boldsymbol{\xi}^N, \quad (7a)$$

and with any subsequent linear transformation of (7a). Thus, any given set of K ($K\leq N$) independent linear transformations of (7a):

$$E_p\left[\left(\widehat{\theta^0}(\overline{\mathbf{x}})-\theta^0\right)\mathbf{h}_k^T\boldsymbol{\omega}_p\left(\overline{\mathbf{x}};\boldsymbol{\theta}^N\right)\right]=\mathbf{h}_k^T\boldsymbol{\xi}^N, \quad (7b)$$

$\mathbf{h}_k\in\mathbb{R}^N$, $k\in[1,K]$, provides with a LB on the MSE (5):

$$MSE_p\left[\widehat{\theta^0}\right]\geq\left(\boldsymbol{\xi}^N\right)^T\mathbf{R}_{\mathbf{H}_K}^{\dagger}\boldsymbol{\xi}^N, \quad (7c)$$
$$\mathbf{R}_{\mathbf{H}_K}^{\dagger}=\mathbf{H}_K\left(\mathbf{H}_K^T\mathbf{R}_{\boldsymbol{\omega}_p}\mathbf{H}_K\right)^{-1}\mathbf{H}_K^T, \quad \mathbf{H}_K=[\mathbf{h}_1\ \ldots\ \mathbf{h}_K],$$
$$\left(\mathbf{R}_{\boldsymbol{\omega}_p}\right)_{n,n'}=E_p\left[\omega_p\left(\overline{\mathbf{x}};\theta^{n'}\right)\omega_p\left(\overline{\mathbf{x}};\theta^n\right)\right].$$

The *generalized* BB is obtained by taking the supremum of (7c) over all the existing degrees of freedom $\left(N,\boldsymbol{\theta}^N,K,\mathbf{H}_K\right)$. Moreover, for a given vector of test points $\boldsymbol{\theta}^N$, the LB (7c) reaches its maximum $\left(\boldsymbol{\xi}^N\right)^T\mathbf{R}_{\boldsymbol{\omega}_p}^{-1}\boldsymbol{\xi}^N$ if, and only if, the matrix $\mathbf{H}_K$ is invertible ($K=N$) [27][28, Lemma 3], which represents a bijective transformation of the set of the N initial constraints (7a). A *generalized* form of any known bound on the MSE can be obtained with the appropriate instantiation of (7c) [29]. If $p(\overline{\mathbf{x}})\triangleq p_\theta(\overline{\mathbf{x}})\triangleq f_\theta(\overline{\mathbf{x}})$, then any *generalized* BB approximation reduces to its standard form associated to correctly specified uniformly unbiased estimators (3a). However, the *generalized* form of existing BB approximations commonly used in practice must be derived again as illustrated below.

• The generalized McAulay-Seidman bound (GMSB)

Actually (7a) yields the GMSB, that is (7c) where $K=N$ and $\mathbf{H}_N=\mathbf{I}_N$: $GMSB=\boldsymbol{\xi}\left(\boldsymbol{\theta}^N\right)^T\mathbf{R}_{\boldsymbol{\omega}_p}^{-1}\boldsymbol{\xi}\left(\boldsymbol{\theta}^N\right)$. A particular case of interest occurs when $\left(\boldsymbol{\theta}^{N+1}\right)^T=\left(\theta^0,\left(\boldsymbol{\theta}^N\right)^T\right)$. Indeed, then $\left(\boldsymbol{\xi}^{N+1}\right)^T=\left(0,\left(\boldsymbol{\xi}^N\right)^T\right)$ and, by resorting to the inverse of a block matrix, one obtains the following practical form :

$$GMSB=\left(\boldsymbol{\xi}^N\right)^T\left(\mathbf{R}_{\boldsymbol{\omega}_{\theta^0}}-\frac{\mathbf{e}_p^N\left(\mathbf{e}_p^N\right)^T}{E_p\left[\omega_p^2\left(\overline{\mathbf{x}};\theta^0\right)\right]}\right)^{-1}\boldsymbol{\xi}^N, \quad (8)$$

where $\mathbf{e}_p^N=E_p\left[\omega_p\left(\overline{\mathbf{x}};\theta^0\right)\boldsymbol{\omega}_p\left(\overline{\mathbf{x}};\boldsymbol{\theta}^N\right)\right]$, which allows to incorporate the selected value θ^0 into the set of test points while keeping the matrix and vectors dimension (N) unchanged. As announced above, (8) cannot be directly extrapolated from the practical form of the MSB: $MSB=\left(\boldsymbol{\xi}^N\right)^T\left(\mathbf{R}_{\boldsymbol{\upsilon}_{\theta^0}}-\mathbf{1}_N\mathbf{1}_N^T\right)^{-1}\boldsymbol{\xi}^N$ obtained where $p(\overline{\mathbf{x}})\triangleq p_\theta(\overline{\mathbf{x}})\triangleq f_\theta(\overline{\mathbf{x}})$ [24, (3)].

• The generalized Cramér-Rao bound (GCRB)

The GCRB is associated to the limiting form (7a) with 2 test points:

$$E_p\left[\left(\widehat{\theta^0}(\overline{\mathbf{x}})-\theta^0\right)\begin{pmatrix}\omega_p\left(\overline{\mathbf{x}};\theta^0\right)\\ \omega_p\left(\overline{\mathbf{x}};\theta^0+d\theta\right)\end{pmatrix}\right]=\begin{pmatrix}0\\ d\theta\end{pmatrix} \quad (9a)$$

where $d\theta\to0$ [22][25][26], leading to [28, Lemma 3]:

$$E_p\left[\left(\widehat{\theta^0}(\overline{\mathbf{x}})-\theta^0\right)\begin{pmatrix}\omega_p\left(\overline{\mathbf{x}};\theta^0\right)\\ \frac{\partial\omega_p\left(\overline{\mathbf{x}};\theta^0\right)}{d\theta}\end{pmatrix}\right]=\begin{pmatrix}0\\1\end{pmatrix} \quad (9b)$$

If $\boldsymbol{\theta}$ is a vector of P unknown parameters, then (9b) becomes [29]:

$$\begin{cases}E_p\left[\left(\widehat{\boldsymbol{\theta}^0}(\overline{\mathbf{x}})-\boldsymbol{\theta}^0\right)\omega_p\left(\overline{\mathbf{x}};\boldsymbol{\theta}^0\right)\right]=\mathbf{0}\\ E_p\left[\left(\widehat{\boldsymbol{\theta}^0}(\overline{\mathbf{x}})-\boldsymbol{\theta}^0\right)\frac{\partial\omega_p\left(\overline{\mathbf{x}};\boldsymbol{\theta}^0\right)}{\partial\boldsymbol{\theta}^T}\right]=\mathbf{I}_P\end{cases}. \quad (9c)$$

By resorting to the generalization of (5) to a vector of estimators [20, Lemma 1], (9c) yields:

$$\mathbf{GCRB}_p\left(\boldsymbol{\theta}^0\right)=\begin{bmatrix}\mathbf{0} & \mathbf{I}_P\end{bmatrix}\mathbf{R}_{\boldsymbol{\theta}^0}^{-1}\begin{bmatrix}\mathbf{0} & \mathbf{I}_P\end{bmatrix}^T, \quad (10a)$$
$$\mathbf{R}_{\boldsymbol{\theta}^0}=E_p\begin{bmatrix}\omega_p\left(\overline{\mathbf{x}};\boldsymbol{\theta}^0\right)^2 & \omega_p\left(\overline{\mathbf{x}};\boldsymbol{\theta}^0\right)\frac{\partial\omega_p\left(\overline{\mathbf{x}};\boldsymbol{\theta}^0\right)}{\partial\boldsymbol{\theta}^T}\\ \frac{\partial\omega_p\left(\overline{\mathbf{x}};\boldsymbol{\theta}^0\right)}{\partial\boldsymbol{\theta}}\omega_{\boldsymbol{\theta}^0}\left(\overline{\mathbf{x}};\boldsymbol{\theta}^0\right) & \frac{\partial\omega_p\left(\overline{\mathbf{x}};\boldsymbol{\theta}^0\right)}{\partial\boldsymbol{\theta}}\frac{\partial\omega_p\left(\overline{\mathbf{x}};\boldsymbol{\theta}^0\right)}{\partial\boldsymbol{\theta}^T}\end{bmatrix}.$$

Then, a few additional lines of calculus allow to show that [29]:

$$\mathbf{GCRB}_p\left(\boldsymbol{\theta}^0\right)=\left(\mathbf{F}_p\left(\boldsymbol{\theta}^0\right)-\frac{\mathbf{r}_{21}\left(\boldsymbol{\theta}^0\right)\mathbf{r}_{21}^T\left(\boldsymbol{\theta}^0\right)}{r_{11}\left(\boldsymbol{\theta}^0\right)}\right)^{-1}, \quad (10b)$$

where $\mathbf{F}_p(\boldsymbol{\theta})=E_p\left[\frac{\partial\omega_p(\overline{\mathbf{x}};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\omega_p(\overline{\mathbf{x}};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right]$ is a *generalized* Fisher information matrix (GFIM), $r_{11}(\boldsymbol{\theta})=E_p\left[\omega_p(\overline{\mathbf{x}};\boldsymbol{\theta})^2\right]$, $\mathbf{r}_{21}(\boldsymbol{\theta})=E_p\left[\frac{\partial\omega_p(\overline{\mathbf{x}};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\omega_p(\overline{\mathbf{x}};\boldsymbol{\theta})\right]$. Additionally if $p(\overline{\mathbf{x}})\triangleq p_{\boldsymbol{\theta}}(\overline{\mathbf{x}})\triangleq f_{\boldsymbol{\theta}}(\overline{\mathbf{x}})$ then $\omega_p(\overline{\mathbf{x}};\boldsymbol{\theta})=\upsilon_{\boldsymbol{\theta}^0}(\overline{\mathbf{x}};\boldsymbol{\theta})$, $\frac{\partial\omega_p\left(\overline{\mathbf{x}};\boldsymbol{\theta}^0\right)}{\partial\boldsymbol{\theta}}=\frac{\partial\ln p_{\boldsymbol{\theta}^0}(\overline{\mathbf{x}})}{\partial\boldsymbol{\theta}}$ and $\mathbf{r}_{21}\left(\boldsymbol{\theta}^0\right)=E_{\boldsymbol{\theta}^0}\left[\frac{\partial\ln p_{\boldsymbol{\theta}^0}(\overline{\mathbf{x}})}{\partial\boldsymbol{\theta}}\right]=\mathbf{0}$, leading to:

$$\mathbf{GCRB}_p\left(\boldsymbol{\theta}^0\right)=\mathbf{F}_p\left(\boldsymbol{\theta}^0\right)^{-1}=\mathbf{F}\left(\boldsymbol{\theta}^0\right)^{-1}=\mathbf{CRB}\left(\boldsymbol{\theta}^0\right), \quad (11)$$

where $\mathbf{F}(\boldsymbol{\theta})=E_{\boldsymbol{\theta}}\left[\frac{\partial\ln p_{\boldsymbol{\theta}}(\overline{\mathbf{x}})}{\partial\boldsymbol{\theta}}\frac{\partial\ln p_{\boldsymbol{\theta}}(\overline{\mathbf{x}})}{\partial\boldsymbol{\theta}^T}\right]$ is the usual FIM.

As a consequence, the LB derived in [1, (8)] in the particular case where $p(\overline{\mathbf{x}})\triangleq p_{\boldsymbol{\theta}}(\overline{\mathbf{x}})$:

$$\mathbf{LB}\left(\boldsymbol{\theta}^0\right)=\mathbf{F}_p\left(\boldsymbol{\theta}^0\right)^{-1} \quad (12)$$

is not the CRB under unbiasedness misspecification (10b) since it only takes into account the constraints: $E_p\left[\left(\widehat{\boldsymbol{\theta}^0}(\overline{\mathbf{x}})-\boldsymbol{\theta}^0\right)\frac{\partial\omega_p\left(\overline{\mathbf{x}};\boldsymbol{\theta}^0\right)}{\partial\boldsymbol{\theta}^T}\right]=\mathbf{I}_P$. It is however a LB under unbiasedness misspecification, but looser than the GCRB.

[2] Actually $f_\theta(\overline{\mathbf{x}})$ can be any real-valued function of $\overline{\mathbf{x}}$ parameterized by θ with support Ω and a normalized integral, not necessarily positive.

4. APPLICATION TO LINEAR GAUSSIAN MODELS

As in [1], the following true linear Gaussian model is considered: $\mathbf{x} = \mathbf{d}_p\theta + \mathbf{n}$, $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_p)$, where $\mathbf{d}_p \in \mathbb{R}^M$ and $\mathbf{C}_p \in \mathbb{R}^{M\times M}$ are supposed to be known. For any selected value θ^0 of the parameter θ, the true p.d.f. of the observation is then $p(\mathbf{x}) \triangleq p_{\theta^0}(\mathbf{x}) \triangleq p_{\mathcal{N}}\left(\mathbf{x}|\mathbf{d}_p\theta^0, \mathbf{C}_p\right)$. Even if the linear structure and the noise p.d.f. is known, generally $\mathbf{d}_p$ and $\mathbf{C}_p$ are not accurately known and are replaced by assumed values $\mathbf{d}_f$ and $\mathbf{C}_f$, leading to the following assumed p.d.f. $f_\theta(\mathbf{x}) \triangleq p_{\mathcal{N}}(\mathbf{x}|\mathbf{d}_f\theta, \mathbf{C}_f)$. In order to illustrate the theoretical results of the previous section, we compare using two examples, the MSE of MMLE:

$$\widehat{\theta^0}(\mathbf{x}) = \mathbf{w}_f^T\mathbf{x}, \quad \mathbf{w}_f = \mathbf{C}_f^{-1}\mathbf{d}_f \;/\; \mathbf{d}_f^T\mathbf{C}_f^{-1}\mathbf{d}_f, \tag{13}$$

the Huber's MSE prediction computed from the "sandwich covariance" (1c), the GMSB (8) with 2 test points, aka the *generalized* Hammersley-Chapman-Robbins bound (GHCRB) [25][26], the GCRB (10b), the *generalized* bound (12) derived by Fritsche et al. [1] (referred to as FB) and the CRB (11) associated to the true and assumed p.d.f.s. The results are displayed in Fig. 1. Due to page length limitation, the results hereafter, which are not difficult to derive, are provided without proofs (included in [29]).

• The Huber's "sandwich covariance"

For the Gaussian p.d.f.s $p_{\theta^0}(\mathbf{x})$ and $f_\theta(\mathbf{x})$, the KLIC (1b) is:

$$\begin{aligned} KLIC \propto tr(\mathbf{C}_f^{-1}\mathbf{C}_p) + \ln\left|\mathbf{C}_f^{-1}\mathbf{C}_p\right| \\ + \left(\mathbf{d}_f\theta - \mathbf{d}_p\theta^0\right)^T \mathbf{C}_f^{-1}(\mathbf{d}_f\theta - \mathbf{d}_p\theta^0) \end{aligned} \tag{14a}$$

leading to:

$$\theta_f^0 = \mathbf{w}_f^T\mathbf{d}_p\theta_0 = E_p\left[\widehat{\theta^0}(\mathbf{x})\right]. \tag{14b}$$

Moreover, since $\frac{\partial \ln f_\theta(\mathbf{x})}{\partial\theta} = \mathbf{d}_f^T\mathbf{C}_f^{-1}(\mathbf{x} - \mathbf{d}_f\theta)$, $E_p\left[\frac{\partial \ln f_{\theta_f^0}(\mathbf{x})}{\partial\theta}^2\right] = \mathbf{d}_f^T\mathbf{C}_f^{-1}\mathbf{C}_p\mathbf{C}_f^{-1}\mathbf{d}_f$, and $\frac{\partial^2 \ln f_\theta(\mathbf{x})}{\partial^2\theta} = -\mathbf{d}_f^T\mathbf{C}_f^{-1}\mathbf{d}_f$, then:

$$\mathbf{C}_{HS}(\boldsymbol{\theta}_f) = \frac{\mathbf{d}_f^T\mathbf{C}_f^{-1}\mathbf{C}_p\mathbf{C}_f^{-1}\mathbf{d}_f}{(\mathbf{d}_f^T\mathbf{C}_f^{-1}\mathbf{d}_f)^2} = Var_p\left[\widehat{\theta^0}\right], \tag{14c}$$

that is the Huber's MSE prediction coincides with the MSE of the MMLE (hence a single line plot "MML" in Fig. 1).

• The GHCRB and the GCRB

If $\boldsymbol{\theta}^N \triangleq \theta^1 = \theta^0 + d\theta$, then from (8):

$$GHCRB = \sup_{d\theta} \frac{d\theta^2 R_{\theta^0,\theta^0}}{R_{\theta^0+d\theta,\theta^0+d\theta}R_{\theta^0,\theta^0} - R^2_{\theta^0+d\theta,\theta^0}}, \tag{15a}$$

where, if $\mathbf{C}_p > \frac{1}{2}\mathbf{C}_f$ [20, (33-34)]:

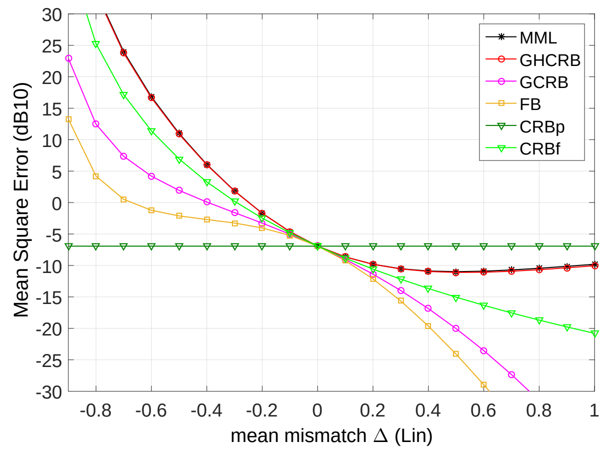
$$\begin{aligned} R_{\theta^0,\theta^1} &= E_p\left[\omega_p\left(\mathbf{x};\theta^0\right)\omega_p\left(\mathbf{x};\theta^1\right)\right] \\ &= \frac{\sqrt{\left|\overline{\mathbf{C}}\right|\left|\mathbf{C}_p\right|}}{\left|\mathbf{C}_f\right|} e^{\frac{1}{2}\left(\mathbf{m}_{\theta^0,\theta^1}^T\overline{\mathbf{C}}\mathbf{m}_{\theta^0,\theta^1} - \delta_{\theta^0,\theta^1}\right)} \end{aligned} \tag{15b}$$

and $\overline{\mathbf{C}} = \left(2\mathbf{C}_f^{-1} - \mathbf{C}_p^{-1}\right)^{-1}$, $\mathbf{m}_{\theta^0,\theta^1} = \mathbf{C}_f^{-1}\mathbf{d}_f\left(\theta^0 + \theta^1\right) - \mathbf{C}_p^{-1}\mathbf{d}_p\theta^0$, $\delta_{\theta^0,\theta^1} = \left(\left(\theta^0\right)^2 + \left(\theta^1\right)^2\right)\mathbf{d}_f^T\mathbf{C}_f^{-1}\mathbf{d}_f - \left(\theta^0\right)^2\mathbf{d}_p^T\mathbf{C}_p^{-1}\mathbf{d}_p$. The GCRB (10b) is the limiting case of the GHCRB (15a) where $d\theta \to 0$, that is:

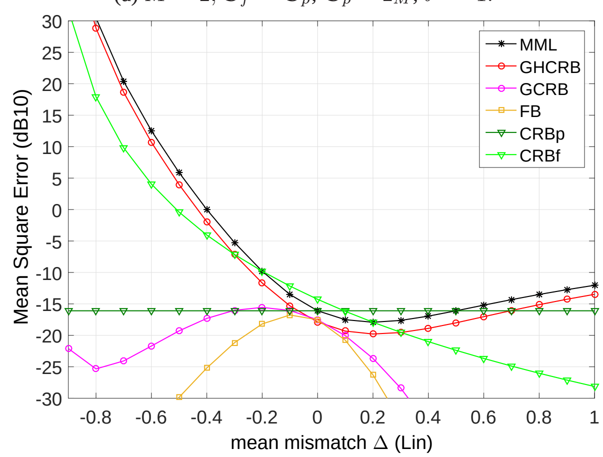
$$GCRB(\theta_0) = \frac{e^{-\frac{1}{2}\left(\mathbf{m}_{\theta^0,\theta^0}^T\overline{\mathbf{C}}\mathbf{m}_{\theta^0,\theta^0} - \delta_{\theta^0,\theta^0}\right)}}{\mathbf{d}_f^T\mathbf{C}_f^{-1}\overline{\mathbf{C}}\mathbf{C}_f^{-1}\mathbf{d}_f} \tag{16}$$

Last, the FB (12) is obtained from [1, (10)] and the CRBs are $CRB_f = (\mathbf{d}_f^T\mathbf{C}_f^{-1}\mathbf{d}_f)^{-1}$, $CRB_p = (\mathbf{d}_p^T\mathbf{C}_p^{-1}\mathbf{d}_p)^{-1}$.

We consider two examples similar to those used in [1]. We assume



(a) $M = 2$, $\mathbf{C}_f = \mathbf{C}_p$, $\mathbf{C}_p = \mathbf{I}_M$, $\theta^0 = 1$.



(b) $M = 5$, $\mathbf{C}_f = 1.2 \times \mathbf{C}_p$, $\mathbf{C}_p = \mathbf{I}_M$, $\theta^0 = 1$.

Fig. 1. MSE vs. Δ of (a) Example 1 and (b) Example 2.

that $\mathbf{d}_p = \mathbf{1}_M$, $\mathbf{d}_f = (1 + \Delta)\,\mathbf{d}_p$ where Δ is varied in the interval $[-1, 1]$, $\theta^0 = 1$, and $\mathbf{C}_p = \mathbf{I}_M$ (unit noise power). In the first example, $M = 2$, and we assume that $\mathbf{C}_f = \mathbf{C}_p$, that is the true noise power is accurately known, whereas in the second example, $M = 5$, and the true noise power is known up to a scalar factor, which is assumed to be 1.2: $\mathbf{C}_f = 1.2 \times \mathbf{C}_p$. Fig. 1b) exemplify again (as in [1]) the fact that the standard CRBs, CRB_f and CRB_p, no longer provide a lower bound on estimation performance whatever the misspecification considered. As expected, in both examples, the FB is looser than the GCRB. Last, it appears that the *generalized* "Small Errors" bounds (FB and GCRB) are unlikely to be informative in a large domain of misspecification values $(\Delta, \mathbf{C}_f)$, since they become overly optimistic as soon as the misspecification on $(\Delta, \mathbf{C}_f)$ increases. Fortunately, the behavior of the GHCRB suggests that the use of *generalized* "Large Errors" bounds (*generalized* Barankin-Type LBs) will allow to increase the domain of misspecification values $(\Delta, \mathbf{C}_f)$ where such LBs remain tight enough to be relevant, which is clearly a topic to be investigated in future research.

5. REFERENCES

[1] C. Fritsche et al, ”On the Cramér-Rao lower bound under model mismatch”, in Proceedings of ICASSP 2015, pp 3986-3990

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

[3] E. L. Lehmann and G. Casella, *Theory of Point Estimation (2nd ed.)*. Springer, 1998

[4] R.A. Fisher, ”Theory of statistical Estimation”, Proceedings of the Cambridge Philosophical Society, 22:700-725, 1925

[5] P. J. Huber, “The behavior of maximum likelihood estimates under nonstandard conditions,” in Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability. Berkeley, CA, USA: Univ. of California, 1967.

[6] H. Akaike, Information Theory and an Extension of the Maximum Likelihood Principle, in Proceeding of the Second International Symposium on Information Theory, B.N. Petrovand F. Caski, eds., Akademiai Kiado, Budapest, 1973, 267-281

[7] S. Kullback and A. Leibler, ”On information and sufficiency”, Annals of Mathematical Statistics, 22: 79-86, 1951

[8] H. White, “Maximum likelihood estimation of misspecified models,” Econometrica, vol. 50, pp. 1-25, Jan. 1982.

[9] C. R. Blyth, ”Necessary and sufficient conditions for inequalities of Cramér-Rao type,” Ann. Math. Stat., 2(3): 464-473, 1974

[10] J.S. Abel, ”A bound on mean-square-estimate error”, IEEE Trans. on IT, 39(5): 1675-1680, 1993

[11] T. B. Fomby and R. C. Hill, *Maximum-Likelihood Estimation of Misspecified Models: Twenty Years Later*. Oxford, U.K.: Elsevier, Ltd., 2003.

[12] Q. H. Vuong, “Cramér-Rao bounds for misspecified models,” Div. of he Humanities and Social Sci., California Inst. of Technol., Pasadena, CA, USA, Working Paper 652, Oct. 1986 [Online]. Available: https://www.hss.caltech.edu/content/cramer-rao-bounds-misspecified-models

[13] C.D. Richmond and L.L. Horowitz, ”Parameter Bounds on Estimation Accuracy Under Model Misspecification”, IEEE Trans. on SP, 63(9): 2263-2278, 2015

[14] S. Fortunati, F. Gini, and M. S. Greco, “The misspecified Cramér-Rao bound and its application to the scatter matrix estimation in complex elliptically symmetric distributions,” in IEEE Trans. Signal Process, 64(9): 2387-2399, 2016

[15] P. Stoica, A. Nehorai, ”Performances study of conditional and unconditional direction of arrival estimation”, IEEE Trans. on SP, 38(10): 1783-1795, 1990

[16] M. Viberg and B. Ottersten, ”Sensor array processing based on subspace fitting”, IEEE Trans. on SP, 39(5): 1110-1121, 1991

[17] B. Ottersten, M. Viberg, P. Stoica and A. Nehorai, ”Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing”, in *Radar Array Prvocessing*, S. Haykin, J. Litva, and T. J. Shepherd, Eds. Berlin, Germany: Springer-Verlag, ch. 4, pp. 99-151, 1993

[18] J. Li and R. T. Compton, ”Maximum likelihood angle estimation for signals with known waveforms”, IEEE Trans. on SP, 41(9): 2850-2862, 1993

[19] A. Renaux, P. Forster, E. Chaumette, and P. Larzabal, ”On the high snr cml estimator full statistical characterization”, IEEE Trans. on SP, 54(12): 4840-4843, 2006

[20] E. Chaumette, J. Galy, A. Quinlan, P. Larzabal, ”A New Barankin Bound Approximation for the Prediction of the Threshold Region Performance of Maximum-Likelihood Estimators”, IEEE Trans. on SP, 56(11): 5319-5333, 2008

[21] K. Todros and J. Tabrikian, ”General Classes of Performance Lower Bounds for Parameter Estimation-Part I: Non-Bayesian Bounds for Unbiased Estimators”, IEEE Trans. on IT, 56(10): 5064-5082, 2010

[22] E.W. Barankin, ”Locally best unbiased estimates”, Ann. Math. Stat., 20(4): 477-501, 1949.

[23] P. Forster, P. Larzabal, ”On lower Bounds For Deterministic Parameter Estimation”, in Proc. IEEE ICASSP, 2002

[24] I. Reuven and H. Messer, “A Barankin-type lower bound on the estimation error of a hybrid parameter vector,” IEEE Trans. on IT, 43(3): 1084-1093, 1997

[25] J.M. Hammersley, ”Estimating Restricted Parameters”, Journal of Roy. Stat. Soc., 12(2): 192-240, 1950

[26] D.G. Chapman, H. Robbins, ”Minimum variance estimation without regularity assumptions”, Ann. Math. Stat., 22: 581-586, 1951

[27] R. McAulay and L.P. Seidman, ”A useful form of the Barankin lower bound and its application to PPM threshold analysis”, IEEE Trans. on IT, 15(2): 273-279, 1969

[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. on SP, 60(3): 1032-1049, 2012

[29] M. L. Diong, E. Chaumette and F. Vincent, ”Generalized Lower Bounds and its Application to Misspecified Models”, in preparation for IEEE Trans. on Signal Processing