COMPUTABLE LOWER BOUNDS FOR DETERMINISTIC PARAMETER ESTIMATION.

[1]*Eric Chaumette,* [2]*Jerome Galy,* [3]*Francois Vincent,* [4]*Pascal Larzabal*

(1) ONERA, DEMR, BP 72, 92322 Chatillon, France - chaumette@onera.fr
(2) LIRMM, Montpellier, France - galy@lirmm.fr
(3) ENSICA, Toulouse, France - vincent@ensica.fr
(4) SATIE, ENS Cachan, CNRS, Universud, France - larzabal@satie.ens-cachan.fr

ABSTRACT

This paper is primarily tutorial in nature and presents a simple approach (norm minimization under linear constraints) for deriving computable lower bounds on the MSE of deterministic parameter estimators with a clear interpretation of the bounds. We also address the issue of lower bounds tightness in comparison with the MSE of ML estimators and their ability to predict the SNR threshold region. Last, as many practical estimation problems must be regarded as joint detection-estimation problems, we remind that the estimation performance must be conditional on detection performance, leading to the open problem of the fundamental limits of the joint detection-estimation performance.

Index Terms— Estimation, MSE lower bounds, Signal detection

1. INTRODUCTION

Lower bounds on the mean square error (MSE) in estimating a set of deterministic parameters [4] from noisy observations provide the best performance of any estimators in terms of the MSE. They allow to investigate fundamental limits of a parameter estimation problem or to assess the relative performance of a specific estimator. All existing bounds on the MSE of unbiased estimators are different solutions of the same norm minimization problem under sets of appropriate linear constraints defining approximations of unbiasness in the Barankin sense (2) (§2). The *weakest* and the *strongest* definition of unbiasness (§2.2) leads respectively to the Cramér-Rao bound (CRB) and to the Barankin bound (BB), which are, the lowest (non trivial) and the highest lower bound on the MSE of unbiased estimators. Therefore, the CRB and BB can be regarded as key representative of two general classes of bounds, respectively the Small-Error bounds and the Large-Error bounds. Indeed, in non-linear estimation problems three distinct regions of operation can be observed. In the asymptotic region, the MSE of estimators is small and, in many cases, close to the Small-Error bounds. In the *a priori* performance region where the number of independent snapshots and/or the signal-to-noise ratio (SNR) are very low, the observations provide little information and the MSE is close to that obtained from the prior knowledge about the problem. Between these two extremes, there is an additional ambiguity region, also called the transition region. In this region, the MSE of maximum likelihood estimators (MLEs) deteriorates rapidly with respect to Small-Error bounds and generally exhibits a threshold behavior corresponding to a "performance breakdown" highlighted by Large-Error bounds (§2.3). Additionally, in nearly all fields of science and engineering, a wide variety of processing requires a binary detection step (detector) designed to decide if a signal is present or not in noise. As a detector restricts the set of observations available for parameter estimation, any accurate MSE lower bound must take into account this initial statistical conditioning. If the derivation of any lower bound with statistical conditioning is straightforward for *realizable* detectors (which do not depend on the true parameter values) by resorting to the norm minimization approach (§3.1), it remains an open problem for *clairvoyant* detectors (which depend on the true parameters value) (§3.2), including optimal detectors (Bayes or Neyman-Pearson criteria). As a consequence, it is not yet possible to compute the fundamental limits of the joint detection-estimation problem, such as, for example, lower bounds on the MSE conditioned by the optimal detector.

2. LOWER BOUNDS AND NORM MINIMIZATION

For the sake of simplicity we will focus on the estimation of a single real function $g(\theta)$ of a single unknown real deterministic parameter θ. In the following, unless otherwise stated, $\mathbf{x}$ denotes the random observations vector, Ω the observation space, and $p(\mathbf{x};\theta)$ the probability density function (p.d.f.) of observations depending on $\theta \in \Theta$, where Θ denotes the parameter space. Let $\mathbb{F}_\Omega$ be the real vector space of square integrable functions over Ω. A fundamental property of the MSE of a particular estimator $\widehat{g(\theta^0)}(\mathbf{x}) \in \mathbb{F}_\Omega$ of $g(\theta^0)$, where θ^0 is a selected value of the parameter θ, is that it is a norm associated with a particular scalar product $\langle \,|\,\rangle_\theta$:

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

$$\langle g(\mathbf{x}) \,|\, h(\mathbf{x})\rangle_{\theta^0} = E_{\theta^0}[g(\mathbf{x})h(\mathbf{x})] = \int g(\mathbf{x})h(\mathbf{x})p(\mathbf{x};\theta^0)\,d\mathbf{x}.$$

In the search for a lower bound on the MSE, this property allows the use of two equivalent fundamental results: the generalization of the Cauchy-Schwartz inequality to Gram matrices (generally referred to as the "covariance inequality") and the minimization of a norm under linear constraints. Nevertheless, we shall prefer the "norm minimization" form as its use provides a better understanding of the hypotheses associated with the different lower bounds on the MSE. Then, let $\mathbb{U}$ be a Euclidean vector space of any dimension (finite or infinite) on the body of real numbers $\mathbb{R}$ which has a scalar product $\langle \,|\,\rangle$. Let $(\mathbf{c}_1,\ldots,\mathbf{c}_K)$ be a free family of K vectors of $\mathbb{U}$ and $\mathbf{v} = (v_1,\ldots,v_K)^T$ a vector of $\mathbb{R}^K$. The problem of the minimization of $\|\mathbf{u}\|^2$ under the K linear constraints $\langle \mathbf{u} \,|\, \mathbf{c}_k\rangle = v_k$, $k \in [1, K]$ then has the solution:

$$\min\left\{\|\mathbf{u}\|^2\right\} = \mathbf{v}^T\mathbf{G}^{-1}\mathbf{v} \quad \text{for} \quad \mathbf{u}_{opt} = \sum_{k=1}^{K}\alpha_k\mathbf{c}_k$$

$$(\alpha_1,\ldots,\alpha_K)^T = \boldsymbol{\alpha} = \mathbf{G}^{-1}\mathbf{v}, \quad \mathbf{G}_{n,k} = \langle \mathbf{c}_k \,|\, \mathbf{c}_n\rangle \tag{1}$$

2.1. Highest lower bound for unbiased estimators

An unbiased estimator $\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)$ of $g\left(\theta\right)$ is an estimator unbiased for all possible values of the unknown parameter ($\forall \theta \in \Theta$):

$$E_\theta\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right] = g\left(\theta\right). \tag{2}$$

Thus, the locally-best (at θ^0) unbiased estimator is the solution of:

$$\min\left\{MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\right]\right\} \text{ under } E_\theta\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right] = g\left(\theta\right). \tag{3}$$

This problem can be solved by applying the work of Barankin based on a point discretization of (2). Indeed, any unbiased estimator $\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)$ of $g\left(\theta\right)$ must verify, $\forall \theta^n \in \Theta$:

$$E_{\theta^n}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right] = g\left(\theta^n\right) = \int \widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right) p\left(\mathbf{x};\theta^n\right) d\mathbf{x},$$

and more generally, $\forall \mathbf{w} \in \mathbb{R}^N$:

$$E_{\theta^0}\left[\left(\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right) - g\left(\theta^0\right)\right)\left(\mathbf{w}^T\boldsymbol{\pi}\left(\mathbf{x};\theta^0\right)\right)\right] = \mathbf{w}^T\Delta\mathbf{g}$$

where $\left(\Delta\mathbf{g}\right)_n = g\left(\theta^n\right) - g\left(\theta^0\right)$ and $\left(\boldsymbol{\pi}\left(\mathbf{x};\theta^0\right)\right)_n = \frac{p\left(\mathbf{x};\theta^n\right)}{p\left(\mathbf{x};\theta^0\right)}$.
Therefore, according to (1), the minimization of $MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\right]$ under the constraint as above - valid for any subset of test points $\{\theta^n\}^{[1,N]}$ of Θ and $\mathbf{w} \in \mathbb{R}^N$ - implies:

$$MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\right] \geq \lim_{N\to\infty} \sup_{\mathbf{w},\{\theta^n\}^{[1,N]}} \frac{\left(\mathbf{w}^T\Delta\mathbf{g}\right)^2}{\mathbf{w}^T\mathbf{R}\mathbf{w}}$$

where $\mathbf{R} = E_{\theta^0}\left[\boldsymbol{\pi}\left(\mathbf{x};\theta^0\right)\boldsymbol{\pi}\left(\mathbf{x};\theta^0\right)^T\right]$, which is the original form of the BB. From a computational point of view, a more efficient form (McAulay-Seidman) can be derived by noting that $\frac{\left(\mathbf{w}^T\Delta\mathbf{g}\right)^2}{\mathbf{w}^T\mathbf{R}\mathbf{w}}$ $\leq \Delta\mathbf{g}^T\mathbf{R}^{-1}\Delta\mathbf{g}$ with equality if $\mathbf{w} = \lambda\mathbf{R}^{-1}\Delta\mathbf{g}$. Therefore:

$$MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\right] \geq \lim_{N\to\infty} \sup_{\{\theta^n\}^{[1,N]}} \left\{\Delta\mathbf{g}^T\mathbf{R}^{-1}\Delta\mathbf{g}\right\} \tag{4}$$

It is then worth noting that (4) is also the solution of, $\{\theta^n\}^{[1,N]} \in \Theta$:

$$\min\left\{MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\right]\right\} \text{ under } E_{\theta^n}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right] = g\left(\theta^n\right)$$

Finally, the locally best unbiased estimator $\widehat{g\left(\theta^0\right)}_{opt}$ satisfies (4):

$$\lim_{N\longrightarrow\infty} \left|\begin{array}{c} \mathbf{R}\left(\frac{\mathbf{w}}{\lambda}\right) = \Delta\mathbf{g} \\ \widehat{g\left(\theta^0\right)}_{opt}\left(\mathbf{x}\right) - g\left(\theta^0\right) = \sum_{n=1}^{N} \frac{w_n}{\lambda} \frac{p\left(\mathbf{x};\theta^n\right)}{p\left(\mathbf{x};\theta^0\right)} \\ MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\right] \geq \Delta\mathbf{g}^T\mathbf{R}^{-1}\Delta\mathbf{g} = \Delta\mathbf{g}^T\left(\frac{\mathbf{w}}{\lambda}\right) \end{array}\right.$$

that leads to, defining $\frac{1}{\lambda} = d\theta = \theta_{n+1} - \theta^n$, :

$$\int K\left(\theta,\theta'\right) w\left(\theta'\right) d\theta' = g\left(\theta\right) - g\left(\theta^0\right) \tag{5a}$$

$$K\left(\theta,\theta'\right) = \int \frac{p\left(\mathbf{x};\theta\right) p\left(\mathbf{x};\theta'\right)}{p\left(\mathbf{x};\theta^0\right)} d\mathbf{x} \tag{5b}$$

$$\widehat{g\left(\theta^0\right)}_{opt}\left(\mathbf{x}\right) - g\left(\theta^0\right) = \int \frac{p\left(\mathbf{x};\theta\right)}{p\left(\mathbf{x};\theta^0\right)} w\left(\theta\right) d\theta \tag{5c}$$

$$MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\right] \geq \int \left(g\left(\theta\right) - g\left(\theta^0\right)\right) w\left(\theta\right) d\theta \tag{5d}$$

Unfortunately it is generally impossible to find either the limit of (4) or an analytical solution of (5a) to obtain an explicit form of $\widehat{g\left(\theta^0\right)}_{opt}$ and of the lower bound on the MSE. Therefore, the search for an easily computable but tight approximation of the BB is a subject of great theoretical and practical importance.

2.2. Computable lower bounds: approximations of the BB

All existing computable bounds on the MSE of unbiased estimators are different solutions of the same norm minimization problem under sets of appropriate linear constraints defining approximations of unbiasness (2). Indeed, let us consider that both $p\left(\mathbf{x};\theta\right)$ and $g\left(\theta\right)$ can be approximated by piecewise Taylor series expansions of order L^n ($\Theta = \cup I^n$):

$$g(\theta^n + d\theta) = g(\theta^n) + \sum_{l=1}^{L_n} \frac{\partial^l g(\theta^n)}{\partial^l \theta} \frac{d\theta^l}{l!} + o\left(d\theta^{L_n}\right), \quad \theta^n + d\theta \in I^n$$

$$p\left(\mathbf{x};\theta^n + d\theta\right) = p\left(\mathbf{x};\theta^n\right) + \sum_{l=1}^{L_n} \frac{\partial^l p\left(\mathbf{x};\theta^n\right)}{\partial^l \theta} \frac{d\theta^l}{l!} + o_\mathbf{x}\left(d\theta^{L_n}\right)$$

Then, under the required regularity conditions to allow order of integration and differentiation interchange, a possible local approximation of unbiasedness (2) on every sub-interval I_n is:

$$E_{\theta^n + d\theta}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right] = g\left(\theta^n + d\theta\right) + o\left(d\theta^{L_n}\right) \tag{6}$$

provided the $L_n + 1$ linear constraints are verified:

$$\int \widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right) \frac{\partial^l p\left(\mathbf{x};\theta^n\right)}{\partial^l \theta} d\mathbf{x} = \frac{\partial^l g\left(\theta^n\right)}{\partial^l \theta}, \quad l \in [0, L_n] \tag{7}$$

Thus, the set of $\sum_{n=1}^{N} \left(L_n + 1\right)$ constraints (7) deriving from the N piecewise local approximation of (2) defines a given approximation of the BB denoted by $\mathrm{BB}_{L_1,...,L_N}^{I_1,...,I_N}$ (1):

$$\mathrm{BB}_{L_1,...,L_N}^{I_1,...,I_N} = \mathbf{v}^T\mathbf{G}^{-1}\mathbf{v} \tag{8}$$

$$\mathbf{v} = \left[\mathbf{v}_1^T, ..., \mathbf{v}_N^T\right]^T, \quad \mathbf{G} = E_{\theta^0}\left[\mathbf{c}\mathbf{c}^T\right], \quad \mathbf{c} = \left[\mathbf{c}_1^T, ..., \mathbf{c}_N^T\right]^T$$
$$\mathbf{c}_n = \left[p\left(\mathbf{x};\theta^n\right), \frac{\partial p\left(\mathbf{x};\theta^n\right)}{\partial\theta}, ..., \frac{\partial^{L_n} p\left(\mathbf{x};\theta^n\right)}{\partial^{L_n}\theta}\right]$$
$$\mathbf{v}_n = \left[g\left(\theta^n\right) - g\left(\theta^0\right), \frac{\partial g\left(\theta^n\right)}{\partial\theta}, ..., \frac{\partial^{L_n} g\left(\theta^n\right)}{\partial^{L_n}\theta}\right]$$

Moreover, if $\min\left\{L_1, ..., L_N\right\}$ tends to infinity, $\mathrm{BB}_{L_1,...,L_N}^{I_1,...,I_N}$ converges in mean-square to the BB. An immediate generalization of expression (8) consists of taking its supremum over existing degrees of freedom (sub-interval definitions and series expansion orders). Lastly, it is worth noting that the set of $\mathrm{BB}_{L_1,...,L_N}^{I_1,...,I_N}$ allows exploration of the unbiasedness assumption from its *weakest* to its *strongest* formulation.
Designating the BB approximations as:
• N-piecewise BB approximation of homogeneous order L, if on all sub-intervals I_n the series expansions are of the same order L,
• N-piecewise BB approximation of heterogeneous orders $\{L_1, ..., L_N\}$, if otherwise,
• the **CRB** [4] is a 1-piecewise BB approximation of homogeneous order 1, since the constraints are:

$$E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right] = g\left(\theta^0\right), \ E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\frac{\frac{\partial p\left(\mathbf{x};\theta^0\right)}{\partial\theta}}{p\left(\mathbf{x};\theta^0\right)}\right] = \frac{\partial g\left(\theta^0\right)}{\partial\theta}$$

• the Bhattacharyya bound [4] of order L is a 1-piecewise BB approximation of homogeneous order L, since the constraints are:

$$E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right] = g\left(\theta^0\right), \ E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\frac{\frac{\partial^l p\left(\mathbf{x};\theta^0\right)}{\partial^l\theta}}{p\left(\mathbf{x};\theta^0\right)}\right] = \frac{\partial^l g\left(\theta^0\right)}{\partial^l\theta}$$

• the Hammersley-Chapman-Robbins bound [1] (**HCRB**) is the supremum of a 2-piecewise BB approximation of homogeneous order 0, over a set of constraints of type:

$$E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right] = g\left(\theta^0\right), \ E_{\theta^1}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right] = g\left(\theta^1\right)$$

• the McAulay-Seidman bound [1] (**MSB**^{N}) with N test points is an

$N + 1$-piecewise BB approximation of homogeneous order 0, since the constraints are:

$$E_{\theta^0} \left[\widehat{g\left(\theta^0\right)}(\mathbf{x}) \right] = g\left(\theta^0\right), \ E_{\theta^n} \left[\widehat{g\left(\theta^0\right)}(\mathbf{x}) \right] = g\left(\theta^n\right)$$

• the Hybrid Barankin-Bhattacharyya bound (Abel Bound) [1] ($\mathbf{HBB}_L^N$) is an $N + 1$-piecewise BB approximation of heterogeneous order $\{L, 0, ..., 0\}$, since the constraints are:

$$E_{\theta^0} \left[\widehat{g\left(\theta^0\right)}(\mathbf{x}) \right] = g\left(\theta^0\right), \ E_{\theta^0} \left[\widehat{g\left(\theta^0\right)}(\mathbf{x}) \frac{\frac{\partial^l p\left(\mathbf{x};\theta^0\right)}{\partial^l \theta}}{p\left(\mathbf{x};\theta^0\right)} \right] = \frac{\partial^l g\left(\theta^0\right)}{\partial^l \theta}$$

$$E_{\theta^n} \left[\widehat{g\left(\theta^0\right)}(\mathbf{x}) \right] = g\left(\theta^n\right)$$

• the $\mathbf{BB}_1^N$ [3] is an $N + 1$-piecewise BB approximation of homogeneous order 1, since the constraints are:

$$E_{\theta^0} \left[\widehat{g\left(\theta^0\right)}(\mathbf{x}) \right] = g\left(\theta^0\right), \ E_{\theta^0} \left[\widehat{g\left(\theta^0\right)}(\mathbf{x}) \frac{\frac{\partial p\left(\mathbf{x};\theta^0\right)}{\partial \theta}}{p\left(\mathbf{x};\theta^0\right)} \right] = \frac{\partial g\left(\theta^0\right)}{\partial \theta}$$

$$E_{\theta^n} \left[\widehat{g\left(\theta^0\right)}(\mathbf{x}) \right] = g\left(\theta^n\right), \ E_{\theta^n} \left[\widehat{g\left(\theta^0\right)}(\mathbf{x}) \frac{\frac{\partial p\left(\mathbf{x};\theta^n\right)}{\partial \theta}}{p\left(\mathbf{x};\theta^n\right)} \right] = \frac{\partial g\left(\theta^n\right)}{\partial \theta}$$

2.3. Lower bounds and threshold region determination

In non-linear estimation problems, ML estimators exhibit a threshold effect, i.e. a rapid deterioration of estimation accuracy below a certain SNR or number of snapshots. This effect is caused by outliers and is not captured by standard techniques such as the CRB. The search of the SNR threshold value (where the CRB becomes unreliable for prediction of ML estimator variance) can be achieved with the help of the BB approximations introduced above. For example, let us consider the single tone estimation problem:

$$\mathbf{x} = \mathbf{s}_{\boldsymbol{\theta}} + \mathbf{n}, \ \ \mathbf{s}_{\boldsymbol{\theta}} = a\boldsymbol{\psi}(\theta), \ \ \boldsymbol{\psi}(\theta) = \left[1, e^{j2\pi\theta}, ..., e^{j2\pi(M-1)\theta} \right]^T,$$

where $\theta \in \,]-0.5, 0.5[$, a^2 is the SNR $(a > 0)$ and $\mathbf{n}$ is a complex circular zero mean white Gaussian noise ($\mathbf{C_x} = \mathbf{Id}$). Then $\widehat{\theta}_{ML} = \max_{\theta} \left\{ \text{Re} \left[\boldsymbol{\psi}(\theta)^H \mathbf{x} \right] \right\}$. The MSB^N, HBB_1^N, BB_1^N are computed as supremum over the possible values of $\{\theta^n\}^{[1,3]} = \left\{\theta^0, \theta^0 + d\theta, \theta^0 - d\theta\right\}$ where $\theta^0 = 0$. Curves with solid lines of figure (1) shows the evolution of the various bounds as a function of SNR in the case of $M = 10$ samples. The MSE of the MLE is also shown in order to compare the threshold behaviour of the bounds.

2.4. Lower bounds for biased estimators

Additionally, these curves highlight that the achievable performance predicted by any lower bounds for unbiased estimators becomes less informative as the SNR decreases, since most realizable estimators (including MLEs) can not remain unbiased at low SNR. To overcome this limitation, one can resort to biased lower bounds where (2) becomes:

$$E_{\theta} \left[\widehat{g\left(\theta^0\right)}(\mathbf{x}) \right] = g\left(\theta\right) + b\left(\theta\right)$$

where $b\left(\theta\right)$ is the bias function. It is an attractive theoretical refinement if analytical expression of the bias is available [2] (figure (2)). Unfortunately the bias depends on the specific estimator and furthermore is hardly ever known in practice.

3. CONDITIONAL LOWER BOUNDS

In many practical problems of interest, the observations vector $\mathbf{x}$ can be modelled as a mixture of a signal of interest $\mathbf{s}_{\boldsymbol{\theta}}$ and a noise $\mathbf{n}$ ($\mathbf{x} = \mathbf{s}_{\boldsymbol{\theta}} + \mathbf{n}$) where the signal of interest $\mathbf{s}_{\boldsymbol{\theta}}$ is not always present. Such problems require first a binary detection step (decision rule) to decide if the signal of interest $\mathbf{s}_{\boldsymbol{\theta}}$ is present (H_1) or not (H_0) in the noise before running any estimation scheme [2]:

$$H_0 : \mathbf{x} = \mathbf{n}$$
$$H_1 : \mathbf{x} = \mathbf{s}_{\boldsymbol{\theta}} + \mathbf{n}$$

The derivation of optimal decision rules [4] require knowledge of the p.d.f. of observations under each hypothesis and the *a priori* probability of each hypothesis ($P\left(H_0\right), P\left(H_1\right)$), if known (Bayes criterion). If no *a priori* probability of hypotheses is available, then the Neyman-Pearson criterion is often used:

$$\max \left\{P_D = P\left(D \mid H_1\right)\right\} \text{ under } P_{FA} = P\left(D \mid H_0\right) = \alpha,$$

where D denotes the event of detection of $\mathbf{s}_{\boldsymbol{\theta}}$. Both criteria lead to the likelihood ratio test (LRT)

$$\frac{P(\mathbf{x}|H_1)}{P(\mathbf{x}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} T$$

which is generally not *realizable* since it almost always depend at least on one of the unknown parameters $\boldsymbol{\theta}$. Therefore, a common approach to designing *realizable* tests is to replace the unknown parameters by estimates, the detection problem becoming a composite hypothesis testing problem (CHTP) [4]. Although not necessarily optimal for detection performance, the estimates are generally chosen in the maximum likelihood sense, thereby obtaining the generalized likelihood ratio test (GLRT). Additionally, as a detection step restricts the set of observations available for parameter estimation, any MSE lower bound must take this statistical conditioning into account, which is straightforward for *realizable* test by resorting to the norm minimization approach.

3.1. Lower Bounds conditioned by a realizable detection test

If D is a *realizable* conditioning event (detection test) with probability $P_D\left(\boldsymbol{\theta}\right) = \int_D p\left(\mathbf{x}; \boldsymbol{\theta}\right) d\mathbf{x}$, the conditional lower bounds are obtained by substituting D and $p\left(\mathbf{x} \mid D; \boldsymbol{\theta}\right) = \frac{p(\mathbf{x};\boldsymbol{\theta})}{P_D(\boldsymbol{\theta})}$ for Ω and $p\left(\mathbf{x}; \boldsymbol{\theta}\right)$ in the MSE norm definition:

$$MSE_{\theta^0|D} \left[\widehat{g\left(\theta^0\right)} \right] = \left\| \widehat{g\left(\theta^0\right)}(\mathbf{x}) - g\left(\theta^0\right) \right\|_{\theta^0|D}^2$$
$$\langle g\left(\mathbf{x}\right) \mid h\left(\mathbf{x}\right) \rangle_{\theta^0|D} = E_{\theta^0} \left[g\left(\mathbf{x}\right) h\left(\mathbf{x}\right) \mid D \right]$$
$$= \int_D g\left(\mathbf{x}\right) h\left(\mathbf{x}\right) p\left(\mathbf{x} \mid D; \theta^0\right) d\mathbf{x}.$$

As a result, the Conditional Fisher Information Matrix (CFIM) is:

$$\mathbf{F}\left(\boldsymbol{\theta} \mid D\right)_{i,j} = E_{\boldsymbol{\theta}} \left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_j} \mid D \right] - \frac{\partial \ln P_D(\boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \ln P_D(\boldsymbol{\theta})}{\partial \theta_j}$$
$$\mathbf{F}\left(\boldsymbol{\theta} \mid D\right)_{i,j} = -E_{\boldsymbol{\theta}} \left[\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \mid D \right] + \frac{\partial^2 \ln P_D(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}$$

The possible influence of the detection step on parameter estimation performance can be illustrated by the study of the influence of the energy detector:

$$\mathbf{x}^H \mathbf{x} \underset{H_0}{\overset{H_1}{\gtrless}} T$$

on the single tone estimation problem (§2.3) [3] and on the estimation of the direction of arrival (DOA) of a signal source by means of a 2 sensors array called monopulse antenna [2]. This high-precision technique is widely used in tracking systems where:

$$\mathbf{x} = \beta \mathbf{g} + \mathbf{n}, \ \ \mathbf{g} = [1, r\left(\theta\right)]^T$$

θ is the deviation angle from array boresight, $r\left(\theta\right)$ is the monopulse ratio. If β is of Rayleigh type, then the p.d.f. of $\widehat{r\left(\theta\right)} = \frac{(\mathbf{x})_2}{(\mathbf{x})_1}$ without conditioning follows a Student distribution with mean value 0 and a smoothly increasing variance [2] as the SNR decreases. It is the alternative case where the transition region is smooth when the detection threshold effect is negligible. Intuitively, the detection step is expected to modify MSE behavior mainly in the transition region
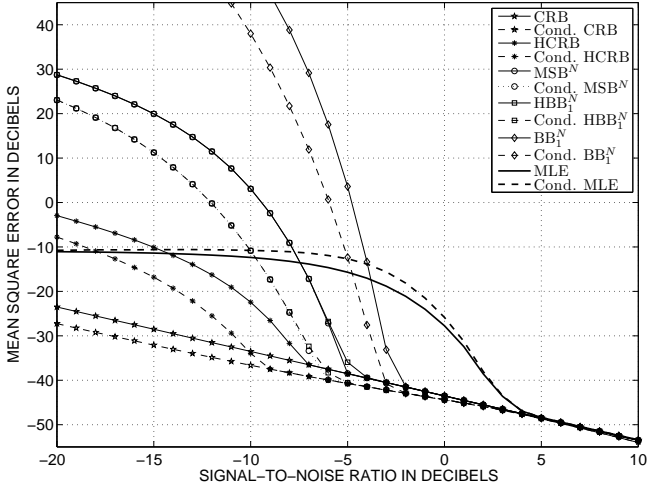
Fig. 1. Single Tone Estimation: MSE conditioned or not by the Energy Detector versus SNR, $L = 10$, $P_{FA} = 10^{-4}$
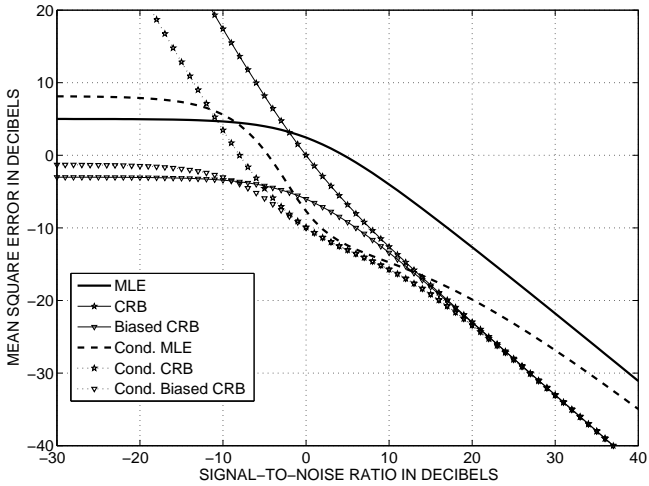
Fig. 2. Monopulse Ratio Estimation: MSE conditioned or not by the Energy Detector versus SNR, $P_{FA} = 10^{-4}$

where it plays a crucial role in selecting instances with relatively high signal energy - sufficient to exceed the detection threshold - and disregarding instances mainly consisting of noise that deteriorate the MSE. The former analysis is confirmed theoretically by the lower bounds behavior in both figures (1 - dot curves) (2). As a consequence, such a detection step is expected to improve the lower bounds tightness in the transition region and to significantly modify the conditions required to attain the CRB and thus to obtain an efficient estimator (figure (2)).

A more unexpected and non intuitive result highlighted by figure (1) is the increase of the MSE of the MLE in the transition region as the detection threshold increases (as the P_{FA} decreases). Indeed, if we consider the stochastic case, i.e. $a \sim \mathcal{CN}_1(0, snr)$, then $\widehat{\theta}_{ML} = \max_{\theta} \left\{ \left| \boldsymbol{\psi}(\theta)^H \mathbf{x} \right|^2 \right\}$ and one can check that the behavior of its MSE is the opposite and true to the common intuition.

3.2. Lower Bounds conditioned by a clairvoyant detection test

If the conditioning event D_θ is *clairvoyant*, the MSE becomes:

$$MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)} \mid D_{\theta^0}\right] = \int_{D_{\theta^0}} \left(\widehat{g\left(\theta^0\right)}(\mathbf{x}) - g\left(\theta^0\right)\right)^2 p\left(\mathbf{x} \mid D_{\theta^0}; \theta^0\right) d\mathbf{x}$$

where $p\left(\mathbf{x} \mid D_\theta; \theta\right) = \frac{p(\mathbf{x};\theta)}{P_{D_\theta}} \Pi_{D_\theta}(\mathbf{x})$, $\Pi_D(\mathbf{x})$ being the indicatrice function over the subset D. Unfortunately the linear transformation $\langle \mid \rangle_{\theta^0|D_{\theta^0}}$ based on $p\left(\mathbf{x} \mid D_{\theta^0}; \theta^0\right)$ is a scalar product only over D_{θ^0} as it is no longer definite over any D_θ not included in D_{θ^0}, which prevents from rewriting the unbiased estimator constraint (2)

$$g(\theta) = E_\theta\left[\widehat{g\left(\theta^0\right)}(\mathbf{x}) \mid D_\theta\right] = \int_{D_\theta} \widehat{g\left(\theta^0\right)}(\mathbf{x}) p\left(\mathbf{x} \mid D_\theta; \theta\right) d\mathbf{x},$$

by resorting to $\langle \mid \rangle_{\theta^0|D_{\theta^0}}$. Nevertheless, the above difficulty can be partially overcome if we restrict our search to best locally-unbiased estimator of $g\left(\theta^0\right)$. Let us denote $\mathcal{U} = D_{\theta^1} \cup D_{\theta^0}$. Then the lower bound of $\int_{\mathcal{U}} \left(\widehat{g\left(\theta^0\right)}(\mathbf{x}) - g\left(\theta^0\right)\right)^2 \frac{p(\mathbf{x};\theta^0)}{P_{D_{\theta^0}}} d\mathbf{x}$ under the constraints:

$$\int_{\mathcal{U}} \left(\widehat{g\left(\theta^0\right)}(\mathbf{x}) - g\left(\theta^0\right)\right) p\left(\mathbf{x} \mid D_{\theta^1}; \theta^1\right) d\mathbf{x} = g\left(\theta^1\right) - g\left(\theta^0\right)$$

$$\int_{\mathcal{U}} \left(\widehat{g\left(\theta^0\right)}(\mathbf{x}) - g\left(\theta^0\right)\right) p\left(\mathbf{x} \mid D_{\theta^0}; \theta^0\right) d\mathbf{x} = 0$$

exists (1). Under the smooth condition of a set of events D_θ satisfying $\lim_{\theta' \to \theta} D_{\theta'} = D_\theta, \forall \{\theta, \theta'\} \in \Theta$, the above minimization problem, where $\theta^1 \to \theta^0$, converges to:

$$\min\left\{MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)} \mid D_{\theta^0}\right]\right\} \text{ under } \begin{matrix} E_{\theta^1}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x}) \mid D_{\theta^1}\right] = g\left(\theta^1\right) \\ E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x}) \mid D_{\theta^0}\right] = g\left(\theta^0\right) \end{matrix}$$

whose lower bound is the HCRB where $\theta^1 \to \theta^0$, that is the CRB,:

$$MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)} \mid D_{\theta^0}\right] \geq \frac{\left[\frac{\partial g\left(\theta^0\right)}{\partial \theta}\right]^2}{\lim_{\theta \to \theta^0} \frac{z(\theta;\theta^0)}{(\theta - \theta^0)^2}}$$

$$z\left(\theta; \theta^0\right) = \int \frac{p(\mathbf{x};\theta)^2 P_{D_{\theta^0}}}{p(\mathbf{x};\theta^0) P_{D_\theta}^2} \Pi_{D_\theta}(\mathbf{x}) d\mathbf{x} - \left(\int \frac{p(\mathbf{x};\theta)}{P_{D_\theta}} \Pi_{D_\theta \cap D_{\theta^0}}(\mathbf{x}) d\mathbf{x}\right)^2$$

The above inequality provides the most general form of the CRB. Examination of the function $z\left(\theta; \theta^0\right)$ shows that it is:
• second order in $\left(\theta - \theta^0\right)$ if the conditioning event is *realizable* $(D_\theta = D, \forall \theta \in \Theta)$ leading to a non-trivial bound (§3.1),
• and first order in $\left(\theta - \theta^0\right)$ if the conditioning event is *clairvoyant* leading unfortunately to the trivial zero bound.
Therefore the derivation and the computation of a **non-trivial** estimation lower bound conditioned by a *clairvoyant* optimal test, remains an open problem of great importance.

4. REFERENCES

[1] J.S. Abel, "A bound on mean-square-estimate error", IEEE Trans., IT vol 39, pp. 1675-1680, Sept. 1993

[2] E. Chaumette, P. Larzabal, P. Forster, "On the Influence of a Detection Step on Lower Bounds for Deterministic Parameters Estimation", IEEE Trans. on SP, vol 53, pp 4080-4090, 2005

[3] E. Chaumette, J. Galy, F. Vincent, A. Renaux, P. Larzabal, "MSE lower bounds conditioned by the energy detector", EU-SIPCO Conf., Sept 2007

[4] H.L. Van Trees, "Detection, Estimation and Modulation Theory, Part 1", New York, Wiley, 1968