To cite this document: Chaumette, Eric and Renaux, Alexandre and Larzabal, Pascal and Galy, Jerome and Vincent, François and Quinlan, A. *MSE lower bounds for deterministic parameter estimation.* (2010) In: 2nd SONDRA workshop on electromagnetic modeling, new concepts and signal processing for radar detection and remote sensing, 25-28 May 2010, Cargèse, France.

MSE LOWER BOUNDS FOR DETERMINISTIC PARAMETER ESTIMATION

Eric Chaumette et al. (1)

ONERA - DEMR/TSI, The French Aerospace Lab, Chemin de la Hunière, F-91120 Palaiseau, France

ABSTRACT

This paper presents a simple approach 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.

***Index Terms*—** Estimation, Signal detection

1. INTRODUCTION

Lower bounds on the mean square error (MSE) in estimating a set of deterministic parameters 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 (1). The *weakest* and the *strongest* definition of unbiasness 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. 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), it remains an open problem for *clairvoyant* detectors (which depend on the true parameters value), including optimal detectors (Bayes or Neyman-Pearson criteria).

(1) : A. Renaux, Université Paris-Sud 11, L2S, Supelec - P. Larzabal, SATIE, ENS Cachan, CNRS, UniverSud - J. Galy, LIRMM, Montpellier - F. Vincent, ISAE, Toulouse - A. Quinlan, Trinity College, Dublin

2. BARANKIN BOUND APPROXIMATIONS

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 observation vector of dimension M, Ω the observations space, and $p(\mathbf{x};\theta)$ the probability density function (p.d.f.) of $\mathbf{x}$ depending on $\theta \in \Theta$, where Θ denotes the parameter space. Let $L^2(\Omega)$ be the real Hilbert space of square integrable functions over Ω.

2.1. Lower bounds and norm minimization

In the search for a lower bound on the MSE of unbiased estimators, two fundamental properties of the problem at hand, introduced by Barankin [6], must be noticed. The first property is that the MSE of a particular estimator $\widehat{g(\theta^0)}(\mathbf{x}) \in L^2(\Omega)$ of $g(\theta^0)$, where θ^0 is a selected value of the parameter θ, is a norm associated with a particular scalar product $\langle \mid \rangle_\theta$:

$$\begin{aligned} MSE_{\theta^0}\left[\widehat{g(\theta^0)}\right] &= \left\|\widehat{g(\theta^0)}(\mathbf{x}) - g(\theta^0)\right\|^2_{\theta^0}, \\ \langle g(\mathbf{x}) \mid h(\mathbf{x})\rangle_{\theta^0} &= E_{\theta^0}\left[g(\mathbf{x})\,h(\mathbf{x})^*\right]. \end{aligned}$$

The second property is that an unbiased estimator $\widehat{g(\theta^0)}(\mathbf{x})$ of $g(\theta)$ should be uniformly unbiased, i.e. for all possible values of the unknown parameter $\theta \in \Theta$ it must verify:

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

where $\nu(\mathbf{x};\theta) = \frac{p(\mathbf{x};\theta)}{p(\mathbf{x};\theta^0)}$ 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:

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

solution that can be obtained by using the norm minimization lemma

$$\begin{aligned} \min\left\{\mathbf{u}^H\mathbf{u} \text{ under } \mathbf{c}_k^H\mathbf{u} = v_k, 1 \le k \le K\right\} = \mathbf{v}^H\mathbf{G}^{-1}\mathbf{v} \\ \mathbf{u}_{opt} = \sum_{k=1}^{K}\alpha_k\mathbf{c}_k, \quad \boldsymbol{\alpha} = \mathbf{G}^{-1}\mathbf{v}, \quad \mathbf{G}_{n,k} = \mathbf{c}_n^H\mathbf{c}_k \end{aligned} \quad . \quad (2)$$

Unfortunately, if Θ contains a continuous subset of $\mathbb{R}$, then the norm minimization under a set of an infinite number of linear constraints (1) leads to an integral equation (7) with no analytical solution in general.

2.2. Linear transformations

Therefore, since the original work of Barankin, many studies [6, and references therein][5] have been dedicated to the derivation of "computable" lower bounds approximating the MSE of the locally-best unbiased estimator (BB). All these approximations derive from sets of discrete or integral linear transform of the "Barankin" constraint (1), and accordingly of the LR, and can be obtained using the following simple rationale.

Let $\boldsymbol{\theta}^N = \left(\theta^1, \ldots, \theta^N\right)^T \in \Theta^N$ be a vector of N test points, $\nu\left(\mathbf{x};\boldsymbol{\theta}^N\right) = \left(\nu\left(\mathbf{x};\theta^1\right), \ldots, \nu\left(\mathbf{x};\theta^N\right)\right)^T$ be the vector of LR associated to $\boldsymbol{\theta}^N$, and $\boldsymbol{\xi}\left(\boldsymbol{\theta}^N\right) = \left(\xi\left(\theta^1\right), \ldots, \xi\left(\theta^N\right)\right)^T$ where $\xi(\theta) = g(\theta) - g\left(\theta^0\right)$.

Any unbiased estimator $\widehat{g\left(\theta^0\right)}(\mathbf{x})$ verifying (1) must comply with

$$E_{\theta^0}\left[\left(\widehat{g\left(\theta^0\right)}(\mathbf{x}) - g\left(\theta^0\right)\right)\boldsymbol{\nu}\left(\mathbf{x};\boldsymbol{\theta}^N\right)\right] = \boldsymbol{\xi}\left(\boldsymbol{\theta}^N\right), \quad (3)$$

and with any subsequent linear transformation of (3). Therefore, any given set of K $(K \leq N)$ independent linear transformations of (3):

$$E_{\theta^0}\left[\left(\widehat{g\left(\theta^0\right)}(\mathbf{x}) - g\left(\theta^0\right)\right)\mathbf{h}_k^T\boldsymbol{\nu}\left(\mathbf{x};\boldsymbol{\theta}^N\right)\right] = \mathbf{h}_k^T\boldsymbol{\xi}\left(\boldsymbol{\theta}^N\right), \quad (4)$$

$\mathbf{h}_k \in \mathbb{R}^N, k \in [1, K]$, provides with a lower bound on the MSE (2):

$$MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\right] \geq \boldsymbol{\xi}\left(\boldsymbol{\theta}^N\right)^T \widetilde{\mathbf{G}}_{\mathbf{H}_K}\boldsymbol{\xi}\left(\boldsymbol{\theta}^N\right), \quad (5)$$

where $\widetilde{\mathbf{G}}_{\mathbf{H}_K} = \mathbf{H}_K\left(\mathbf{H}_K^T\mathbf{R}_{\boldsymbol{\nu}}\mathbf{H}_K\right)^{-1}\mathbf{H}_K^T$, $\mathbf{H}_K = [\mathbf{h}_1 \ldots \mathbf{h}_K]$ and $\left(\mathbf{R}_{\boldsymbol{\nu}}\right)_{n,m} = E_{\theta^0}\left[\nu\left(\mathbf{x};\theta^n\right)\nu\left(\mathbf{x};\theta^m\right)\right]$. The BB is obtained by taking the supremum of (5) over all the existing degrees of freedom $\left(N, \boldsymbol{\theta}^N, K, \mathbf{H}_K\right)$. All known bounds on the MSE deriving from the Barankin Bound is a particular implementation of (5), including the most general formalism introduced lately in [5]. Indeed, the limit of (4) where $N \to \infty$ and $\boldsymbol{\theta}^N$ uniformly samples Θ leads to the linear integral constraint:

$$E_{\theta^0}\left[\left(\widehat{g\left(\theta^0\right)}(\mathbf{x}) - g\left(\theta^0\right)\right)\eta(\mathbf{x},\tau)\right] = \Gamma_h(\tau), \quad (6)$$

$$\eta(\mathbf{x},\tau) = \int_\Theta h(\tau,\theta)\,\nu(\mathbf{x};\theta)\,d\theta, \quad \Gamma_h(\tau) = \int_\Theta h(\tau,\theta)\,\xi(\theta)\,d\theta,$$

where each $\mathbf{h}_k = \left(h\left(\tau_k,\theta^1\right), \ldots, h\left(\tau_k,\theta^N\right)\right)^T$ is the vector of samples of a parametric function $h(\tau,\theta), \tau \in \Lambda \subset \mathbb{R}$, integrable over Θ. As the consequence, the integral form of (5) is [5][7]:

$$\begin{cases} MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)}_{lmvu}(\mathbf{x})\right] = \int_\Lambda \Gamma_h(\tau)\,\beta(\tau)\,d\tau \\ \widehat{g\left(\theta^0\right)}_{lmvu}(\mathbf{x}) - g\left(\theta^0\right) = \int_\Lambda \eta(\mathbf{x},\tau)\,\beta(\tau)\,d\tau \\ \int_\Lambda K_h(\tau',\tau)\,\beta(\tau)\,d\tau = \Gamma_h(\tau') \end{cases}, \quad (7)$$

$$K_h(\tau,\tau') = E_{\theta^0}\left[\eta(\mathbf{x},\tau)\,\eta(\mathbf{x},\tau')\right] = \iint_\Theta h(\tau,\theta)\,R_{\boldsymbol{\nu}}(\theta,\theta')\,h(\tau',\theta')\,d\theta d\theta',$$

$$R_{\boldsymbol{\nu}}(\theta,\theta') = E_{\theta^0}\left[\frac{p(\mathbf{x};\theta)}{p\left(\mathbf{x};\theta^0\right)}\frac{p\left(\mathbf{x};\theta'\right)}{p\left(\mathbf{x};\theta^0\right)}\right] = \int_\Omega \frac{p(\mathbf{x};\theta)\,p\left(\mathbf{x};\theta'\right)}{p\left(\mathbf{x};\theta^0\right)}d\mathbf{x},$$

Note that if $h(\tau,\theta) = \delta(\tau - \theta)$ (limit case of $\mathbf{H}_N = \mathbf{I}_N$ where $N = K \to \infty$) then $K_h(\tau,\tau') = R_{\boldsymbol{\nu}}(\tau,\tau')$ and (7) becomes the simplest expression of the exact Barankin Bound [6, (10)]. However, in most practical cases, it is impossible to find an analytical solution of (7) to obtain an explicit form of the exact Barankin Bound on the MSE, which somewhat limits its interest.

Nevertheless this formalism allows to use discrete (4) or integral (6) linear transforms of the LR, possibly non-invertible, possibly optimized for a set of p.d.f. (such as the Fourier transform in [5]) in order to get a tight approximation of the BB.

2.3. Non-linear transformations

At the opposite, the use of a non-linear transformation of the unbiasness definition (1) of type:

$$E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\,t\left(\nu(\mathbf{x};\theta)\right)\right] = h\left(g(\theta)\right) \quad (8)$$

is more obscure since it seems a difficult mathematical task to compute the bias transformation function $h(\,)$ as a function of the LR transformation function $t(\,)$ and of the LR. Nevertheless there is a class of estimation problems where non-linear transformations of the LR can be used to derive new lower bounds on the MSE. It is the class of estimation problems characterized by a p.d.f. $p(\mathbf{x};\theta)$ for which there exists at least one real valued function $t(\,)$ such that, the transformation of p.d.f. $p(\mathbf{x};\theta)$ by $t(\,)$ is still - up to a normalization constant, w.r.t. $\mathbf{x}$, $k(\theta,t)$ - a p.d.f. of the form $p(\mathbf{x};\,)$ but parameterized by a modified parameter value γ, function of the initial parameter θ and of the transformation $t(\,)$:

$$t\left(p(\mathbf{x};\theta)\right) = k(\theta,t)\,p\left(\mathbf{x};\gamma(\theta,t)\right), \quad k(\theta,t) = \int_\Omega t\left(p(\mathbf{x};\theta)\right)d\mathbf{x} \quad (9)$$

Then an unbiased estimator verifying (1) verifies as well,$\forall\theta \in \Theta$:

$$\int_\Omega \widehat{g\left(\theta^0\right)}(\mathbf{x})\,t\left(p(\mathbf{x};\theta)\right)d\mathbf{x} = k(\theta,t)\int_\Omega \widehat{g\left(\theta^0\right)}(\mathbf{x})\,p\left(\mathbf{x};\gamma(\theta,t)\right)d\mathbf{x} = k(\theta,t)\,g\left(\gamma(\theta,t)\right)$$

what implies,$\forall\theta \in \Theta$:

$$E_{\theta^0}\left[\left(\widehat{g\left(\theta^0\right)}(\mathbf{x}) - g\left(\theta^0\right)\right)\frac{t\left(p(\mathbf{x};\theta)\right)}{p\left(\mathbf{x};\theta^0\right)}\right] = k(\theta,t)\left[g\left(\gamma(\theta,t)\right) - g\left(\theta^0\right)\right].$$

In the most general case, if there exists a set of functions $t_\theta(\,)$ verifying (9), then any unbiased estimator also verifies,$\forall\theta \in \Theta$:

$$E_{\theta^0}\left[\left(\widehat{g\left(\theta^0\right)}(\mathbf{x}) - g\left(\theta^0\right)\right)\frac{t_\theta\left(p(\mathbf{x};\theta)\right)}{p\left(\mathbf{x};\theta^0\right)}\right] = k(\theta,t_\theta)\left[g\left(\gamma(\theta,t_\theta)\right) - g\left(\theta^0\right)\right].$$

Therefore, if we update the definition of $\nu(\mathbf{x};\theta)$ and $\xi(\theta)$ in (6) according to:

$$\nu(\mathbf{x};\theta) = \frac{t_\theta\left(p(\mathbf{x};\theta)\right)}{p\left(\mathbf{x};\theta^0\right)}, \quad \xi(\theta) = k(\theta,t_\theta)\left[g\left(\gamma(\theta,t_\theta)\right) - g\left(\theta^0\right)\right],$$

all the results released in the previous section still hold, the linear integral transformation becoming a mixture of linear and non-linear integral transformations:

$$\eta(\mathbf{x},\tau) = \int_\Theta h(\tau,\theta)\,\frac{t_\theta\left(p(\mathbf{x};\theta)\right)}{p\left(\mathbf{x};\theta^0\right)}d\theta,$$

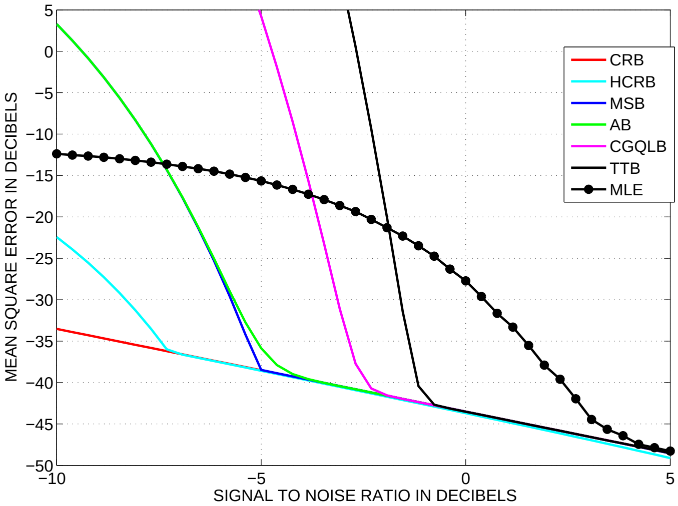

Fig. 1. Comparison of MSE lower bounds versus SNR

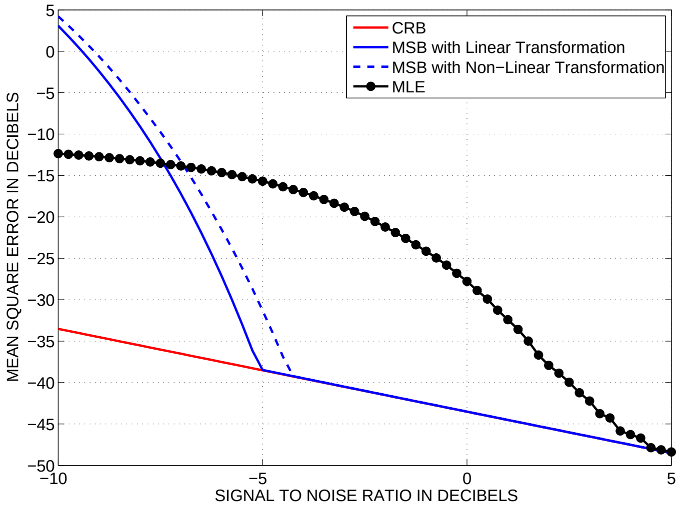

Fig. 2. Effect of the non-linear transformation (10) on MSE lower bounds tightness

$$\Gamma_h(\tau) = \int_{\Theta} h(\tau,\theta)\, k(\theta,t_\theta)\left[g(\gamma(\theta,t_\theta)) - g\left(\theta^0\right)\right] d\theta.$$

At first sight, the proposed rationale, which is a generalization of [1], does not seem appealing, since a non-linear transformation of type (8) or (9) is unlikely to exist whatever the form of the p.d.f., although the linear transformation of the LR (6) is always possible. However, it is applicable to M-dimensional complex circular Gaussian p.d.f.:

$$p(\mathbf{x};\boldsymbol{\theta}) = p(\mathbf{x};\mathbf{m}(\boldsymbol{\theta}),\mathbf{C}(\boldsymbol{\theta})) = \frac{e^{-(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}))^H \mathbf{C}(\boldsymbol{\theta})^{-1}(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}))}}{\pi^M |\mathbf{C}(\boldsymbol{\theta})|}$$

where the transformation $t_q(y) = y^q$ [1] and the observation model results from a mixture of deterministic and stochastic signals in presence of Gaussian interference.
Indeed, in this case $\mathbf{C}(\boldsymbol{\theta}) = \boldsymbol{\Psi}(\boldsymbol{\zeta})\mathbf{C}_\mathbf{s}\boldsymbol{\Psi}(\boldsymbol{\zeta})^H + \mathbf{C}_\mathbf{n}$, $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{m}(\boldsymbol{\varepsilon})$, $\boldsymbol{\theta} = \left[\boldsymbol{\varepsilon}^T, \boldsymbol{\zeta}^T, vec(\mathbf{C}_\mathbf{s})^T, vec(\mathbf{C}_\mathbf{n})^T\right]^T$ and:

$$\begin{aligned} t_q(p(\mathbf{x};\boldsymbol{\theta})) &= k(\boldsymbol{\theta},q)\, p(\mathbf{x};\boldsymbol{\gamma}(\boldsymbol{\theta},q)) \\ k(\boldsymbol{\theta},q) &= \frac{\pi^{M(1-q)}}{q^{qM}}\left|\frac{\mathbf{C}(\boldsymbol{\theta})}{q}\right|^{1-q} \\ \boldsymbol{\gamma}(\boldsymbol{\theta},q) &= \left[\boldsymbol{\varepsilon}^T, \boldsymbol{\zeta}^T, \frac{vec(\mathbf{C}_\mathbf{s})^T}{q}, \frac{vec(\mathbf{C}_\mathbf{n})^T}{q}\right]^T \end{aligned} \quad (10)$$

2.4. 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:

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

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 $(\mathbf{C}_\mathbf{x} = \mathbf{Id})$. Then $\widehat{\theta}_{ML} = \max_\theta \left\{\mathrm{Re}\left[\boldsymbol{\psi}(\theta)^H\mathbf{x}\right]\right\}$. Figure (1) shows the behaviour of various BB approximations as a function of SNR in the case of $M = 10$ samples: CRB, HCRB, MSB, AB, CGQLB described in [6] and TTB described in [5]. The MSE of the MLE is also shown in order to compare the threshold behaviour of the bounds. Figure (2) shows the benefits for tightness of the introduction of the non-linear transformation (10), illustrated in the case of the MSB $(M = 10)$.

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 require knowledge of the p.d.f. of observations under each hypothesis and the *a priori* probability of each hypothesis $(P(H_0), P(H_1))$, if known (Bayes criterion). If no *a priori* probability of hypotheses is available, then the Neyman-Pearson criterion is often used:

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

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). 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

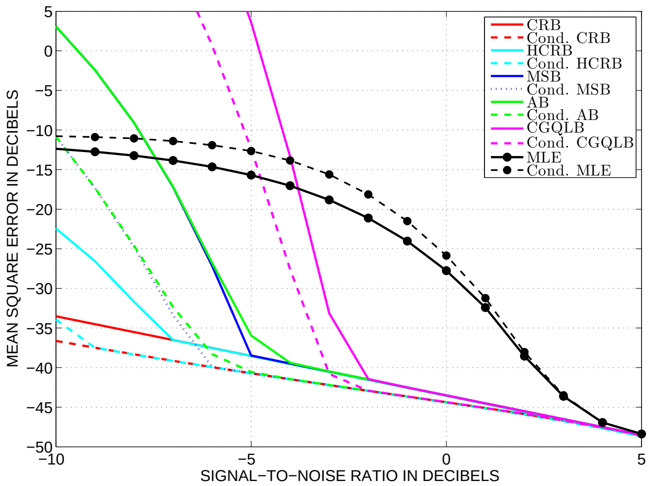

Fig. 3. Single Tone Estimation: MSE conditioned or not by the Energy Detector versus SNR, $L = 10$, $P_{FA} = 10^{-4}$

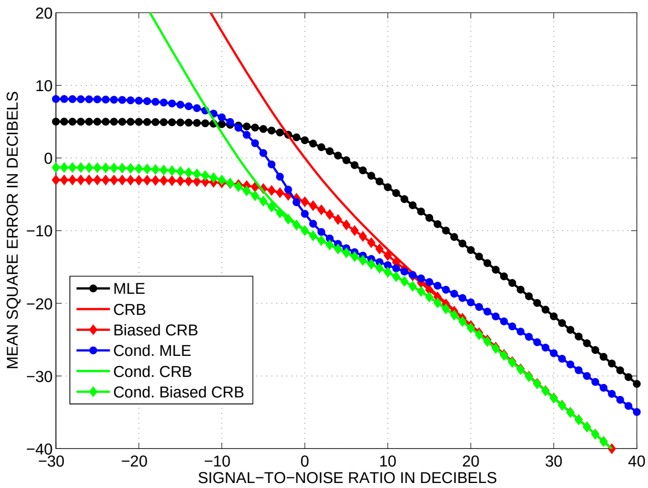

Fig. 4. Monopulse Ratio Estimation: MSE conditioned or not by the Energy Detector versus SNR, $P_{FA} = 10^{-4}$

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

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

As a result, the Conditional Fisher Information Matrix (CFIM) is [2]:

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

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 (§2.4)[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:

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

θ is the deviation angle from array boresight, $r(\theta)$ is the monopulse ratio. If β is of Rayleigh type, then the p.d.f. of $\widehat{r(\theta)} = \frac{x_2}{x_1}$ 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 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 (3)(4). 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 (4)). A more unexpected and non intuitive result highlighted by figure (1) is the increase of the MSE of the MLE in the transition region resulting from observations conditioning. Indeed, if we consider the stochastic case, i.e. $a \sim \mathcal{CN}_1(0, snr)$, then $\widehat{\theta}_{ML} = \max_{\theta}\left\{\left|\psi(\theta)^H\mathbf{x}\right|^2\right\}$ and one can check that the behavior of its MSE is the opposite and true to the common intuition.
Last, the derivation and the computation of a **non-trivial** estimation lower bound conditioned by a *clairvoyant* detectors (the conditional CRB is always 0) [4], including optimal detectors (Bayes or Neyman-Pearson criteria), remains an open problem.

4. REFERENCES

[1] W. R. Blischke, A. J. Truelove and P. B. Mundle, ”On non-regular estimation . I.Variance bounds for estimators of location parameters”, JASA, vol. 64, pp. 1056-1072, 1969.

[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”, EUSIPCO 2007

[4] E. Chaumette, J. Galy, F. Vincent, P. Larzabal, ”Computable Lower Bounds for Deterministic Parameter Estimation”, CAMSAP 2007

[5] K. Todros and J. Tabrikian, ”A new lower bound on the mean-square error of unbiased estimators”, ICASSP 2008

[6] 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, vol. 56, pp. 5319-5333, 2008

[7] E. Chaumette, A. Renaux, P. Larzabal, ”Lower bounds on the mean square error derived from mixture of linear and non-linear transformations of the unbiasness definition”, ICASSP 2009