A NEW LOWER BOUND ON THE MEAN-SQUARE ERROR OF UNBIASED ESTIMATORS

Koby Todros and Joseph Tabrikian

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

ABSTRACT

In this paper, a new class of lower bounds on the mean-square-error (MSE) of unbiased estimators of deterministic parameters is proposed. Derivation of the proposed class is performed by approximating each entry of the vector of estimation error in a closed Hilbert subspace of $\mathcal{L}_2$. This Hilbert subspace is spanned by a set of linear combinations of elements in the domain of an integral transform of the likelihood-ratio function. It is shown that some well known lower bounds on the MSE of unbiased estimators, can be derived from this class by inferring the integral transform. A new lower bound is derived from this class by choosing the Fourier transform. The bound is computationally manageable and provides better prediction of the signal-to-noise ratio (SNR) threshold region, exhibited by the maximum-likelihood estimator. The proposed bound is compared with other existing bounds in term of threshold SNR prediction in the problem of single tone estimation.

Index Terms— Parameter estimation, mean-square-error bounds, threshold SNR.

1. INTRODUCTION

Lower bounds on the mean-square error (MSE) in estimating a set of model parameters from noisy observations constitute the best performance that may be achieved by any estimator. Consequently, they are used as a benchmark against which the performance of estimators can be assessed and compared in the MSE sense.

Historically, the first lower bound on the MSE of any unbiased estimator of deterministic parameters was the Cramér-Rao bound (CRB) [1]. The CRB is widely used due to the following reasons. First, it is simple to calculate and obtain closed form expressions, which are useful for system analysis and design. Second, the maximum-likelihood estimator (MLE) attains the bound asymptotically. The main disadvantage of the CRB is the fact that it is not tight for "large" estimation errors, since it is derived using local statistical information of the observations only in the vicinity of the true parameters. Another disadvantage is that regularity conditions on the likelihood function are imposed. A tighter lower bound was proposed by Bhattacharyya [2]. In similar to the CRB, the Bhattacharyya bound (BHB) is not tight for "large" estimation errors due to the use of local statistical information. Regularity conditions on the likelihood function are required as well. Furthermore, derivation of the BHB is cumbersome due to high order derivatives of the log-likelihood function. Under the assumptions of uniform unbiasedness and finite second moments of the estimator, the tightest lower bound on the MSE of any unbiased estimator was derived by Barankin [3]. Derivation of the Barankin bound (BB) is performed by approximating the estimation error in a closed Hilbert subspace of $\mathcal{L}_2$, spanned by a set of likelihood-ratio (LR) functions [4]. Unfortunately, the exact BB is practically incomputable.

Therefore, numerous works were devoted to derive computationally manageable approximations of the BB [5]-[8]. In [9] it was shown that all the bounds derived in these works, including the CRB and the BHB, may be unified under one general class in which the BB is approximated via piecewise Taylor series expansions of the likelihood function and the function of the parameters to be estimated. Using this approach, a new computationally manageable and tighter BB approximation was derived in [9]. The BB approximation approach described in [9], has the following disadvantages. First, in subintervals, where only zero-order Taylor series expansion is used, a large amount of test points should be selected in order to achieve a good approximation of the BB. This might increase the computational complexity of the bound. Second, although the use of derivatives may reduce the number of test points and consequence better approximation of the BB, in some cases (especially in cases of multivariate functions), the derivatives are not simple to compute and not always the functions to be approximated are differentiable. Third, the use of derivatives impose regularity conditions on the likelihood function in the vicinity of each selected test point. Finally, there is no analytical procedure for optimal selection of test points. Therefore, numerical search methods, which become computationally cumbersome as the number of test points and the dimensionality of the parameters increase are used.

In this paper, a new class of lower bounds on the MSE of unbiased estimators is proposed. The proposed class is based on the approximation of each entry of the estimation error vector in a closed Hilbert subspace of $\mathcal{L}_2$. This Hilbert subspace is spanned by a set of linear combinations of elements in the domain of an integral transform of the LR function. The use of integral transform generalizes the traditional derivative and sampling operators used for computation of the bounds in [9]. By selection of the Fourier transform, a new computationally manageable and tight lower bound is obtained. The Fourier transform is easy to compute and in some cases has a strong "energy compaction" property. Hence, it is shown that in cases where most of the information in the frequency domain tends to be concentrated in a few frequency components of the transform easy selection of a small set of test frequencies, required for obtaining a tight and computationally manageable bound, is enabled. The proposed bound outperforms the family of bounds presented in [9] in terms of tightness, computational manageability and prediction of the transition region exhibited by the MLE.

The paper is organized as follows: In Section 2, a new class of lower bounds on the MSE of any unbiased estimator is derived. The relation of this class to the BB [3] is also discussed. In Section 3, some well known bounds are derived from the proposed class. In Section 4, a new bound is derived from the proposed class by selection of the Fourier transform. In Section 5, the applicability of the proposed bound and its superiority upon existing bounds, in term of SNR transition region prediction is exemplified in the problem of single tone estimation. Section 6, summarizes the main points of this contribution.

2. DERIVATION OF A NEW CLASS OF LOWER BOUNDS

Let $(\boldsymbol{\Theta}, \mathcal{D})$ denote a measurable space, where $\boldsymbol{\Theta} \subset \mathbb{R}^M$ and $\mathcal{D}$ denotes a σ-algebra on $\boldsymbol{\Theta}$. Consider the estimation of $\mathbf{g}(\boldsymbol{\theta}_0)$, where $\mathbf{g} : \boldsymbol{\Theta} \to \mathbb{R}^L$ denotes a Lebesgue measurable function and $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ is a deterministic unknown multivariate parameter. In this section, a new class of bounds on the MSE of any unbiased estimator of $\mathbf{g}(\boldsymbol{\theta}_0)$, having finite second statistical moments, is derived. The proposed class is derived by approximating each entry of the vector of estimation error in a closed Hilbert subspace of $\mathcal{L}_2$. This Hilbert subspace is spanned by a set of linear combinations of elements in the domain of an integral transform of the LR function.

Derivation of the proposed class is preceded by the following definitions and assumptions. Let $(\mathcal{X}, \mathcal{F}, \mathcal{P}_{\boldsymbol{\theta}})$ denote a complete probability space, where $\mathcal{X}$, $\mathcal{F}$ and $\mathcal{P}_{\boldsymbol{\theta}}$ denote an observation space of points $\mathbf{x}$, σ-algebra on $\mathcal{X}$ and a family of probability measures on the measurable space $(\mathcal{X}, \mathcal{F})$, parameterized by $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, respectively. The family $\mathcal{P}_{\boldsymbol{\theta}}$ is assumed to have densities $f(\mathbf{x}; \boldsymbol{\theta})$ relative to a σ-finite measure μ on $(\mathcal{X}, \mathcal{F})$, such that the probability of observing $A \in \mathcal{F}$ is $\mathcal{P}_{\boldsymbol{\theta}}(A) = \int_A f(\mathbf{x}; \boldsymbol{\theta}) \, d\mu(\mathbf{x})$. The Hilbert space of measurable functions $\zeta : \mathcal{X} \to \mathbb{C}$ with finite second moments w.r.t. $\mathcal{P}_{\boldsymbol{\theta}_0}$ is denoted by $\mathcal{L}_2(\mathcal{X}, \mathcal{F}, \mathcal{P}_{\boldsymbol{\theta}_0})$. Let $\hat{\mathbf{g}} : \mathcal{X} \to \mathbb{R}^L$ denote a uniformly unbiased estimator of $\mathbf{g}(\boldsymbol{\theta})$, such that

$$\mathrm{E}_{\mathbf{x};\boldsymbol{\theta}}[\hat{\mathbf{g}}(\mathbf{x})] = \int_{\mathcal{X}} \hat{\mathbf{g}}(\mathbf{x}) f(\mathbf{x}; \boldsymbol{\theta}) \, d\mu(\mathbf{x}) = \mathbf{g}(\boldsymbol{\theta}) \quad \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}. \tag{1}$$

It is assumed that $\hat{\mathbf{g}}(\mathbf{x})$ exists and $[\hat{\mathbf{g}}(\mathbf{x})]_l \in \mathcal{L}_2(\mathcal{X}, \mathcal{F}, \mathcal{P}_{\boldsymbol{\theta}_0}) \; \forall l = 1, \ldots, L$. Let $\nu(\mathbf{x}, \boldsymbol{\theta}) \triangleq \frac{f(\mathbf{x};\boldsymbol{\theta})}{f(\mathbf{x};\boldsymbol{\theta}_0)}$ denote the LR function, where it is assumed that $\nu(\mathbf{x}, \boldsymbol{\theta}) \in \mathcal{L}_2(\mathcal{X}, \mathcal{F}, \mathcal{P}_{\boldsymbol{\theta}_0}) \; \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$. A closed Hilbert subspace of $\mathcal{L}_2(\mathcal{X}, \mathcal{F}, \mathcal{P}_{\boldsymbol{\theta}_0})$ spanned by $S_\nu = \{\nu(\mathbf{x}, \boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ is denoted by $\mathcal{H}_\nu$.

Let $\mathbf{e}(\mathbf{x}) = \hat{\mathbf{g}}(\mathbf{x}) - \mathbf{g}(\boldsymbol{\theta}_0)$ denote the vector of estimation error. In [4] it was shown that the BB can be derived by approximating each entry of $\mathbf{e}(\mathbf{x})$ in $\mathcal{H}_\nu$. In this paper, the proposed class of bounds is derived by approximation of each entry of $\mathbf{e}(\mathbf{x})$ in a closed subspace of $\mathcal{H}_\nu$. This subspace is spanned by a set of linear combinations of elements in the domain of an integral transform, defined on S_ν.

An integral transform on S_ν is defined in the following manner. Let $(\boldsymbol{\Lambda}, \mathcal{D}')$ denote a measurable space, where $\boldsymbol{\Lambda} \subset \mathbb{R}^M$ and $\mathcal{D}'$ denotes a σ-algebra on $\boldsymbol{\Lambda}$. The space $(\boldsymbol{\Theta} \times \boldsymbol{\Lambda}, \mathcal{D} \times \mathcal{D}')$ denotes a product measurable space where $\mathcal{D} \times \mathcal{D}'$ is the σ-algebra on the cartesian product $\boldsymbol{\Theta} \times \boldsymbol{\Lambda}$. Let $\mathbf{h} : \boldsymbol{\Theta} \times \boldsymbol{\Lambda} \to \mathbb{C}^P$ denote a Lebesgue measurable function, an integral transform on S_ν is given by

$$(T_{\mathbf{h}}\nu)(\boldsymbol{\tau}) = \int_{\boldsymbol{\Theta}} \mathbf{h}(\boldsymbol{\tau}, \boldsymbol{\theta}) \, \nu(\mathbf{x}, \boldsymbol{\theta}) \, d\boldsymbol{\theta} = \boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\tau}), \tag{2}$$

where $\boldsymbol{\tau} \in \boldsymbol{\Lambda}$. The set $S_{\boldsymbol{\eta}} = \{\boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\tau}), \boldsymbol{\tau} \in \boldsymbol{\Lambda}\}$ denotes the domain of $(T_{\mathbf{h}}\nu)$.

A closed subspace of $\mathcal{H}_\nu$, spanned by a set of linear combinations of elements in $S_{\boldsymbol{\eta}}$ is constructed in the following manner. Let $\mathbf{a} : \boldsymbol{\Lambda} \to \mathbb{C}^P$ denote a Lebesgue measurable function, a linear combination of elements in $S_{\boldsymbol{\eta}}$ is given by

$$\varphi_{\mathbf{a},\mathbf{h}}(\mathbf{x}) = \int_{\boldsymbol{\Lambda}} \mathbf{a}^H(\boldsymbol{\tau}) \, \boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\tau}) \, d\boldsymbol{\tau}. \tag{3}$$

It is assumed that $\mathbf{a} \in \mathcal{Q}_{\mathbf{h}}$, where $\mathcal{Q}_{\mathbf{h}} = \{\mathbf{a} : \varphi_{\mathbf{a},\mathbf{h}}(\mathbf{x}) \in \mathcal{H}_\nu\}$. In [11] it is shown that a sufficient condition for $\varphi_{\mathbf{a},\mathbf{h}}(\mathbf{x}) \in \mathcal{H}_\nu$ is absolute integrability of $\mathbf{a}^H(\boldsymbol{\tau}) \, \mathbf{h}(\boldsymbol{\tau}, \boldsymbol{\theta}) \, \nu(\mathbf{x}, \boldsymbol{\theta})$ on $\boldsymbol{\Theta} \times \boldsymbol{\Lambda}$ for a.e. $\mathbf{x} \in \mathcal{X}$. Hence, a closed Hilbert subspace of $\mathcal{H}_\nu$, spanned by $\mathcal{S}_\varphi^{(\mathbf{h})} = \{\varphi_{\mathbf{a},\mathbf{h}}(\mathbf{x}), \mathbf{a} \in \mathcal{Q}_{\mathbf{h}}\}$ is denoted by $\mathcal{H}_\varphi^{(\mathbf{h})}$.

According to (1)-(3) and using the Hilbert projection theorem [10], in [11] it is shown that the best approximation of $[\mathbf{e}(\mathbf{x})]_l$, $l = 1, \ldots, L$ in $\mathcal{H}_\varphi^{(\mathbf{h})}$, in the sense of minimum norm of approximation error in $\mathcal{L}_2(\mathcal{X}, \mathcal{F}, \mathcal{P}_{\boldsymbol{\theta}_0})$, yields

$$\tilde{\mathbf{e}}(\mathbf{x}) = \int_{\boldsymbol{\Lambda}} \int_{\boldsymbol{\Lambda}} \boldsymbol{\Gamma}_{\mathbf{h}}^H(\boldsymbol{\tau}') \, \mathbf{G}_{\mathbf{h}}(\boldsymbol{\tau}', \boldsymbol{\tau}) \, \boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\tau}) \, d\boldsymbol{\tau}' d\boldsymbol{\tau}, \tag{4}$$

where $\tilde{\mathbf{e}}(\mathbf{x})$ denotes the approximation of $\mathbf{e}(\mathbf{x})$, $\boldsymbol{\Gamma}_{\mathbf{h}}(\boldsymbol{\tau}) = \int_{\boldsymbol{\Theta}} \mathbf{h}(\boldsymbol{\tau}, \boldsymbol{\theta}) \, \boldsymbol{\xi}^T(\boldsymbol{\theta}) \, d\boldsymbol{\theta}$ and $\boldsymbol{\xi}(\boldsymbol{\theta}) = \mathbf{g}(\boldsymbol{\theta}) - \mathbf{g}(\boldsymbol{\theta}_0)$. The matrix function $\mathbf{G}_{\mathbf{h}}(\cdot, \cdot)$ is defined in the following manner. Let $\mathbf{K}_{\mathbf{h}}(\boldsymbol{\tau}, \boldsymbol{\tau}') = \int_{\boldsymbol{\Theta}} \int_{\boldsymbol{\Theta}} \mathbf{h}(\boldsymbol{\tau}, \boldsymbol{\theta}) \, K(\boldsymbol{\theta}, \boldsymbol{\theta}') \, \mathbf{h}^H(\boldsymbol{\tau}', \boldsymbol{\theta}') \, d\boldsymbol{\theta} d\boldsymbol{\theta}'$, where $K(\boldsymbol{\theta}, \boldsymbol{\theta}') = \mathrm{E}_{\mathbf{x};\boldsymbol{\theta}_0}[\nu(\mathbf{x}, \boldsymbol{\theta}) \, \nu(\mathbf{x}, \boldsymbol{\theta}')]$. Provided that $\mathbf{K}_{\mathbf{h}}(\cdot, \cdot)$ is invertible, $\mathbf{G}_{\mathbf{h}}(\cdot, \cdot)$ is the inverse of $\mathbf{K}_{\mathbf{h}}(\cdot, \cdot)$ such that $\int_{\boldsymbol{\Lambda}} \mathbf{K}_{\mathbf{h}}(\boldsymbol{\tau}, \boldsymbol{\tau}') \, \mathbf{G}_{\mathbf{h}}(\boldsymbol{\tau}', \boldsymbol{\tau}'') \, d\boldsymbol{\tau}' = \delta(\boldsymbol{\tau} - \boldsymbol{\tau}'') \, \mathbf{I}_P$, where $\delta(\cdot)$ and $\mathbf{I}_P$ denote the Dirac's delta function and the P-dimensional identity matrix, respectively.

A lower bound on the MSE matrix, $E_{\mathbf{x};\boldsymbol{\theta}_0}\left[\mathbf{e}(\mathbf{x}) \, \mathbf{e}^T(\mathbf{x})\right]$ is derived in the following manner. Let $\mathbf{u}(\mathbf{x}) = \mathbf{e}(\mathbf{x}) - \tilde{\mathbf{e}}(\mathbf{x})$, denote the vector of approximation error. According to the Hilbert projection theorem, each entry of $\mathbf{u}(\mathbf{x})$ is orthogonal to each entry of $\tilde{\mathbf{e}}(\mathbf{x})$. Furthermore, the autocorrelation matrix of $\mathbf{u}(\mathbf{x})$ is Hermitian positive semidefinite. Therefore,

$$\begin{aligned} E_{\mathbf{x};\boldsymbol{\theta}_0}\left[\mathbf{u}(\mathbf{x}) \, \mathbf{u}^H(\mathbf{x})\right] &= E_{\mathbf{x};\boldsymbol{\theta}_0}\left[\mathbf{e}(\mathbf{x}) \, \mathbf{e}^T(\mathbf{x})\right] \\ &\quad - E_{\mathbf{x};\boldsymbol{\theta}_0}\left[\tilde{\mathbf{e}}(\mathbf{x}) \, \mathbf{e}^T(\mathbf{x})\right] \succeq \mathbf{0}. \end{aligned} \tag{5}$$

Hence, according to (1), (2), (4) and (5) a lower bound on the MSE matrix is given by

$$\mathbf{C}(\mathbf{h}) = \int_{\boldsymbol{\Theta}} \int_{\boldsymbol{\Theta}} \boldsymbol{\xi}(\boldsymbol{\theta}') \, \tilde{G}_{\mathbf{h}}(\boldsymbol{\theta}', \boldsymbol{\theta}) \, \boldsymbol{\xi}^T(\boldsymbol{\theta}) \, d\boldsymbol{\theta}' d\boldsymbol{\theta}, \tag{6}$$

where $\tilde{G}_{\mathbf{h}}(\boldsymbol{\theta}', \boldsymbol{\theta}) = \int_{\boldsymbol{\Lambda}} \int_{\boldsymbol{\Lambda}} \mathbf{h}^H(\boldsymbol{\tau}', \boldsymbol{\theta}') \, \mathbf{G}_{\mathbf{h}}(\boldsymbol{\tau}', \boldsymbol{\tau}) \mathbf{h}(\boldsymbol{\tau}, \boldsymbol{\theta}) d\boldsymbol{\tau}' d\boldsymbol{\tau}$. Observing (6), one can notice that numerous bounds can be derived by inference of $\mathbf{h}(\cdot, \cdot)$.

In the following, the relation of the proposed class of bounds in (6) to the BB is discussed. Approximation of each entry of $\mathbf{e}(\mathbf{x})$ in all $\mathcal{H}_\nu$ yields the following bound

$$\mathbf{C}' = \int_{\boldsymbol{\Theta}} \int_{\boldsymbol{\Theta}} \boldsymbol{\xi}(\boldsymbol{\theta}') \, G(\boldsymbol{\theta}', \boldsymbol{\theta}) \, \boldsymbol{\xi}^T(\boldsymbol{\theta}) \, d\boldsymbol{\theta}' d\boldsymbol{\theta}, \tag{7}$$

which is actually an integral form of the BB. The scalar function $G(\cdot, \cdot)$ is the inverse of $K(\cdot, \cdot)$, such that $\int_{\boldsymbol{\Theta}} K(\boldsymbol{\theta}, \boldsymbol{\theta}') \, G(\boldsymbol{\theta}', \boldsymbol{\theta}'') \, d\boldsymbol{\theta}' = \delta(\boldsymbol{\theta} - \boldsymbol{\theta}'')$. Due to the fact that $\mathcal{H}_\nu \supseteq \mathcal{H}_\varphi^{(\mathbf{h})}$ it is obvious that $\mathbf{C}' \succeq \mathbf{C}(\mathbf{h}) \, \forall \mathbf{h} : \mathbf{C}(\mathbf{h}) \prec \infty$. In [11], it is shown that for any invertible integral transform, $T_{\mathbf{h}}$, the bound $\mathbf{C}'$ is derived from (6). If $T_{\mathbf{h}}$ is non-invertible, $G(\boldsymbol{\theta}', \boldsymbol{\theta})$ is only approximated by $\tilde{G}_{\mathbf{h}}(\boldsymbol{\theta}', \boldsymbol{\theta})$. Since $\mathbf{C}' \succeq \mathbf{C}(\mathbf{h}) \, \forall \mathbf{h} : \mathbf{C}(\mathbf{h}) \prec \infty$, this will produce less tighter bounds than $\mathbf{C}'$. However, calculation of $\mathbf{C}'$ involves computation of the eigenfunctions and eigenvalues of $K(\boldsymbol{\theta}, \boldsymbol{\theta}')$ in order to derive its inverse, $G(\boldsymbol{\theta}', \boldsymbol{\theta})$. Unfortunately, in many cases this task is analytically impossible and consequently $\mathbf{C}'$ is practically incomputable. Hence, it is preferable to use non-invertible integral transforms in order to obtain computationally manageable bounds. In the following section, it is shown that some well known bounds on the MSE of unbiased estimators can be derived from (6) by inference of $\mathbf{h}(\cdot,\cdot)$. In Section 4, a new bound is derived by choosing the Fourier transform.

3. DERIVATION OF EXISTING BOUNDS VIA CERTAIN SELECTION OF INTEGRAL TRANSFORMS

In this section, it is shown that some well known bounds on the MSE of unbiased estimators can be derived from the class of bounds in (6) by inference of $\mathbf{h}(\cdot,\cdot)$. Due to space limit considerations, proofs are given only in the full paper [11].

1. The Bhattacharyya bound [2] is obtained by selecting $\mathbf{h}(\boldsymbol{\tau},\boldsymbol{\theta}) = \left[\begin{array}{ccc} \frac{\partial\delta(\boldsymbol{\tau}-\boldsymbol{\theta})}{\partial\boldsymbol{\tau}} & \dots & \frac{\partial^P\delta(\boldsymbol{\tau}-\boldsymbol{\theta})}{\partial\boldsymbol{\tau}^{\otimes P}}\end{array}\right]^T \delta(\boldsymbol{\tau}-\boldsymbol{\theta}_0)$, where $\frac{\partial^P}{\partial\boldsymbol{\tau}^{\otimes P}}$ denotes the vector of derivatives $\frac{\partial^P}{\partial\tau_{i_1}\partial\tau_{i_2}\cdots\partial\tau_{i_P}}; i_p = 1,\dots,M$. One can notice that by choosing $P=1$, the CRB [1] is obtained. In this paper it is assumed that $\delta(\boldsymbol{\tau}) = \lim_{\epsilon\to 0}\left(2\pi\epsilon^2\right)^{-\frac{M}{2}}\exp\left(-\frac{\|\boldsymbol{\tau}\|_2^2}{2\epsilon^2}\right)$, where $\|\cdot\|_2$ denotes the l_2 norm.
2. The McAulay-Seidman (MS) bound [6] is obtained by selecting $h(\boldsymbol{\tau},\boldsymbol{\theta}) = \left(\delta(\boldsymbol{\tau}-\boldsymbol{\theta}) - \delta(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\right)\sum_{n=1}^{N}\delta(\boldsymbol{\tau}-\boldsymbol{\theta}_n)$, where $\boldsymbol{\theta}_n \in \boldsymbol{\Theta}$, $n=1,\dots,N$ denotes a test point. One can notice that by choosing $N=1$ the Hammersley-Chapman-Robbins (HCR) bound [5] is obtained.
3. The McAulay-Hofstetter (MH) bound [7] is obtained by selecting $\mathbf{h}(\boldsymbol{\tau},\boldsymbol{\theta}) = \left[\begin{array}{c}\left(\frac{\partial\delta(\boldsymbol{\tau}-\boldsymbol{\theta})}{\partial\boldsymbol{\tau}}\right)^T\delta(\boldsymbol{\tau}-\boldsymbol{\theta}_0)\\ \sum_{n=1}^{N}\delta(\boldsymbol{\tau}-\boldsymbol{\theta}_n)\,\delta(\boldsymbol{\tau}-\boldsymbol{\theta})\end{array}\right]$.
4. The general form of bound offered by Quinlan et al. [9] is obtained by selecting $\mathbf{h}(\boldsymbol{\tau},\boldsymbol{\theta}) =$
$$\left[\begin{array}{c}\left[\begin{array}{ccc}\frac{\partial\delta(\boldsymbol{\tau}-\boldsymbol{\theta})}{\partial\boldsymbol{\tau}} & \dots & \frac{\partial^P\delta(\boldsymbol{\tau}-\boldsymbol{\theta})}{\partial\boldsymbol{\tau}^{\otimes P}}\end{array}\right]^{\mathrm{T}}\sum_{m=0}^{M}\delta(\boldsymbol{\tau}-\boldsymbol{\theta}_m)\\ \left(\delta(\boldsymbol{\tau}-\boldsymbol{\theta})-\delta(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\right)\sum_{n=1}^{N}\delta(\boldsymbol{\tau}-\boldsymbol{\theta}_n)\end{array}\right].$$
We note that in practice, it was offered in [9] to use $M=N$ and $P=1$. One can notice that by choosing $M=0$ the Abel bound [8] is obtained.

Quinlan et al. showed in [9] that all the bounds above may be unified under one general class of bounds in which the BB is approximated via piecewise Taylor series expansions of the likelihood function and the function of the parameters to be estimated. This approach has some disadvantages thoroughly discussed in the introduction section. In order to overcome these disadvantages, a new bound on the MSE of any unbiased estimator, derived by using the Fourier transform, is proposed in the following section.

4. DERIVATION OF A NEW BOUND VIA THE FOURIER TRANSFORM

The Fourier transform is easy to compute and in some cases has a strong "energy compaction" property. Therefore, in cases where most of the information in the frequency domain tends to be concentrated in a few frequency components, the Fourier transform is utilized for approximation and data compression. Motivated by these properties, a new bound is derived from (6) by choosing the Fourier transform. Let

$$\mathbf{h}(\boldsymbol{\tau},\boldsymbol{\theta}) = \left[\begin{array}{c}\left[\frac{\partial\delta(\boldsymbol{\tau}-\boldsymbol{\theta})}{\partial\boldsymbol{\tau}}\right]^{\mathrm{T}}\delta(\boldsymbol{\tau}-\boldsymbol{\theta}_0)\\ \sum_{j=1}^{J}\sum_{n=1}^{N}\delta(\boldsymbol{\tau}-\boldsymbol{\Omega}_j)\,\delta(\boldsymbol{\theta}-\boldsymbol{\theta}_n)\exp\left(-i\boldsymbol{\tau}^T\boldsymbol{\theta}\right)\end{array}\right], \quad (8)$$

where $\boldsymbol{\Omega}_j, j=1,\dots,J$ denotes a frequency test bin and $\boldsymbol{\theta}_n \in \boldsymbol{\Theta}, n=1,\dots,N$ denotes a test point. In [11] it is shown that substitution of (8) into (6) yields the following bound:

$$\mathbf{C}_{\text{Fourier}} = \boldsymbol{\Gamma}\mathbf{I}_{\text{FIM}}^{-1}\boldsymbol{\Gamma}^T + \mathbf{Q}\mathbf{W}^H\left(\mathbf{W}\mathbf{R}\mathbf{W}^H\right)^{-1}\mathbf{W}\mathbf{Q}^T, \quad (9)$$

where the Fisher information matrix is denoted by $\mathbf{I}_{\text{FIM}} = E_{\mathbf{x};\boldsymbol{\theta}_0}\left[\left(\left.\frac{\partial\log f(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)^{\mathrm{T}}\left(\left.\frac{\partial\log f(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\right]$ and $\boldsymbol{\Gamma} = \left.\frac{\partial\mathbf{g}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$. The matrix $\mathbf{Q} = \boldsymbol{\Gamma}\mathbf{I}_{\text{FIM}}^{-1}\mathbf{D} - \boldsymbol{\Phi}$, where $\boldsymbol{\Phi} = [\boldsymbol{\xi}(\boldsymbol{\theta}_1),\dots,\boldsymbol{\xi}(\boldsymbol{\theta}_N)]$ and $\mathbf{D} = [\mathbf{d}(\boldsymbol{\theta}_1),\dots,\mathbf{d}(\boldsymbol{\theta}_N)]$. The vector $\mathbf{d}(\boldsymbol{\theta}_n) = -\left(\left.\frac{\partial\mathrm{KLD}[f(\mathbf{x};\boldsymbol{\theta}_n)||f(\mathbf{x};\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)^{\mathrm{T}}$, where the term $\mathrm{KLD}\left[f(\mathbf{x};\boldsymbol{\theta}_n)||f(\mathbf{x};\boldsymbol{\theta})\right]$ is the Kullback-Leibler divergence [12] of $f(\mathbf{x};\boldsymbol{\theta})$ from $f(\mathbf{x};\boldsymbol{\theta}_n)$. The matrix $\mathbf{R} = \boldsymbol{\Psi} - \mathbf{D}^T\mathbf{I}_{\text{FIM}}^{-1}\mathbf{D}$, where the elements of $\boldsymbol{\Psi}$ are given by $[\boldsymbol{\Psi}]_{m,n} = K(\boldsymbol{\theta}_m,\boldsymbol{\theta}_n)$, $m,n=1,\dots,N$. Finally, the elements of the Fourier matrix are given by $[\mathbf{W}]_{j,k} = \exp\left(-i\boldsymbol{\Omega}_j^T\boldsymbol{\theta}_k\right)$, $j=1,\dots,J$, $k=1,\dots,N$.

We note that the bound in (9) is composed of the CRB supplemented by a positive semidefinite term. The positive-semidefiniteness of the supplemental term stems from the positive-semidefiniteness of $\mathbf{W}\mathbf{R}\mathbf{W}^H$[1]. Therefore, the only required regularity condition is the condition used in derivation of the CRB [1]. In cases where this condition is not satisfied, the first element of $\mathbf{h}(\cdot,\cdot)$ in (8) may be discarded and (9) becomes $\mathbf{C}'_{\text{Fourier}} = \boldsymbol{\Phi}\mathbf{W}^{\mathbf{H}}\left(\mathbf{W}\boldsymbol{\Psi}\mathbf{W}^{\mathbf{H}}\right)^{-1}\mathbf{W}\boldsymbol{\Phi}^{\mathbf{T}}$.

Observing (9), one can notice that the matrix $\mathbf{W}\mathbf{R}\mathbf{W}^H$ is the two-dimensional discrete Fourier transform of $\mathbf{R}$ and that the matrix $\mathbf{W}^H\left(\mathbf{W}\mathbf{R}\mathbf{W}^H\right)^{-1}\mathbf{W}$ is an approximation of $\mathbf{R}^{-1}$. Therefore, frequency bins for which the approximation error of $\mathbf{R}^{-1}$ is minimized should be selected when constructing the matrix $\mathbf{W}$. Hence, in cases where the power spectrum of $\mathbf{R}$ is concentrated in a subset, $\mathfrak{F}$, of the frequency space, $\{\boldsymbol{\Omega}_j\}_{j=1}^J$ is selected by uniform sampling of a bounded subset of $\mathfrak{F}^{\mathcal{C}}$, where $\mathcal{C}$ denotes a complement. Assuming that $\boldsymbol{\Theta}$ is bounded, the set $\{\boldsymbol{\theta}_n\}_{n=1}^N$ is selected by uniform sampling of $\boldsymbol{\Theta}$. Whenever $\boldsymbol{\Theta}$ is not bounded, uniform sampling of a bounded subset of $\boldsymbol{\Theta}$ is performed. In comparison to the bounds in [5]-[9], in which $\{\boldsymbol{\theta}_n\}_{n=1}^N$ is reselected for each evaluation of the bounds as a function of SNR or sample size, $\{\boldsymbol{\Omega}_j\}_{j=1}^J$ and $\{\boldsymbol{\theta}_n\}_{n=1}^N$ are selected here only once in order to compute the proposed bound.

In comparison to the bound in [9] the proposed bound avoids the use of derivatives in test points, other than the true parameter, and utilizes the Fourier transform instead. Therefore, the only regularity condition is the one required for derivation of the CRB [1]. Hence, the proposed bound is much easier to compute and as exemplified by simulations outperforms the family of bounds presented in [9] in terms of tightness, computational manageability and prediction of the transition region exhibited by the MLE in nonlinear estimation problems.

[1] Positive-semidefiniteness of $\mathbf{W}\mathbf{R}\mathbf{W}^H$ is proved in [11].

5. SIMULATIONS

In this subsection, the bound in (9) is compared to other existing bounds, described in [1], [5]-[9], in the problem of single tone estimation with Gaussian noise. The comparison criterion is prediction of the transition region exhibited by the MLE.

The observation model is given by $\mathbf{x} = s\mathbf{a}(\theta_0) + \mathbf{n}$, where $\mathbf{x}$ denotes an $L \times 1$ observation vector and $s \in \mathbb{C}$ is a known signal amplitude. The l^{th} element of $\mathbf{a}(\theta_0)$ is given by $[\mathbf{a}(\theta_0)]_l = \exp(i(l-1)2\pi\theta_0)$, where $\theta_0 \in [-0.5, 0.5]$ is the true tone to be estimated and $\mathbf{n}$ denotes an $L \times 1$ complex circular Gaussian noise vector, with zero mean and known covariance $\mathbf{C_n} = \sigma^2\mathbf{I}_L$. Hence, the terms composing (9) are given by $\mathbf{I}_{\text{FIM}} = 2\text{SNR}\| \dot{\mathbf{a}}^H(\theta_0)\|_2^2$, where $\dot{\mathbf{a}}(\theta_0) = \left.\frac{\partial \mathbf{a}(\theta)}{\partial\theta}\right|_{\theta=\theta_0}$ and $\text{SNR} = \frac{|s|^2}{\sigma^2}$, $\mathbf{\Gamma} = 1$,
$\mathbf{d}(\theta_n) = 2\text{SNR}\text{Re}\left\{\dot{\mathbf{a}}^H(\theta_0)[\mathbf{a}(\theta_n) - \mathbf{a}(\theta_0)]\right\}$,
$\mathbf{\Phi} = [\theta_1 - \theta_0, \ldots, \theta_n - \theta_0]$ and $K(\theta_m, \theta_n) =$
$\exp\left(2\text{SNR}\text{Re}\left\{[\mathbf{a}(\theta_n) - \mathbf{a}(\theta_0)]^H[\mathbf{a}(\theta_m) - \mathbf{a}(\theta_0)]\right\}\right)$.

The comparison was held under the following conditions. The values of θ_0 and L were set to 0 and 10, respectively. The parameter space, $\mathbf{\Theta} \subset \mathbb{R}$, was sampled uniformly, with sampling interval of $\Delta\theta = \frac{1}{N}$, where $N = 2^{10}$. Hence, $\theta_n = \frac{n-1}{N} - \frac{1}{2}$; $n = 1, \ldots, N$ and $[\mathbf{W}]_{j,n} = \exp\left(-i\frac{\omega_j}{\Delta\theta}\theta_n\right)$, where $\omega_j \in [0, 2\pi]$; $j = 1, \ldots, J$. Observing Fig. 1, in which the power spectrum of $\mathbf{R}$ is depicted, one can notice that due to the structure of $K(\cdot,\cdot)$ most of the power is concentrated in *low* frequencies. Therefore, according to the procedure described in Section 4, it is sufficient to select a small set of *high* test frequencies in order to obtain a good reconstruction of $\mathbf{R}^{-1}$ in the sense of minimum approximation error. Hence, for all values of SNR, the test frequencies were selected once by uniform sampling of the interval $\rho = \left[\frac{4\pi}{5}, \frac{6\pi}{5}\right]$. The proposed bound was evaluated twice, for $J = 4$ and $J = 32$ equally spaced frequency samples of ρ. For each SNR value, all other bounds were computed as supremum over the possible values of $\{\theta_n\}_{n=1}^3 = [0, m\Delta\theta, -m\Delta\theta], m = 1, \ldots, 2^9$ as described in [9].

Fig. 2 depicts the compared bounds on the root MSE (RMSE) as a function of SNR. The RMSE of the MLE is also depicted in order to compare the threshold behavior of the bounds. It is observed that the proposed bound is the tightest and allows better prediction of the SNR threshold value. The proposed bound exceeds the RMSE of the MLE for low SNR due to the the lack of *a priori* information in the proposed bound and the fact that in this region the MLE is biased.

6. CONCLUSIONS

In this paper, a novel class of lower bounds on the MSE of any unbiased estimator is proposed. It is shown that the bounds in this class are derived by approximating each entry of the vector of estimation error in a closed Hilbert subspace of $\mathcal{L}_2$. This Hilbert subspace is spanned by a set of linear combinations of elements in the domain of an integral transform of the LR function. Using the Fourier transform, a new computationally manageable and tight lower bound is derived from this class. It is shown that in cases where the power spectrum is concentrated in low frequencies a small set of high frequency test bins is sufficient in order to obtain a lower bound with superior computational manageability and tightness in comparison to the bounds in [1], [5]-[9]. Finally, finding some other integral transforms for which new computationally manageable and tight lower bounds could be derived from the proposed class in (5) should be included in a future research.

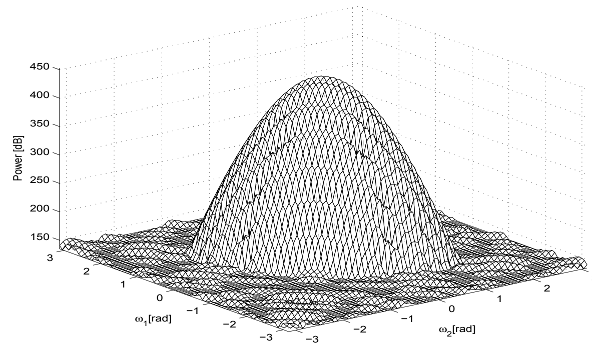

Fig. 1. The power spectrum of the matrix $\mathbf{R}$ in the scenario of single tone estimation, where the true tone, $\theta_0 = 0$, the number of spatial samples, $L = 10$ and SNR=0 dB.

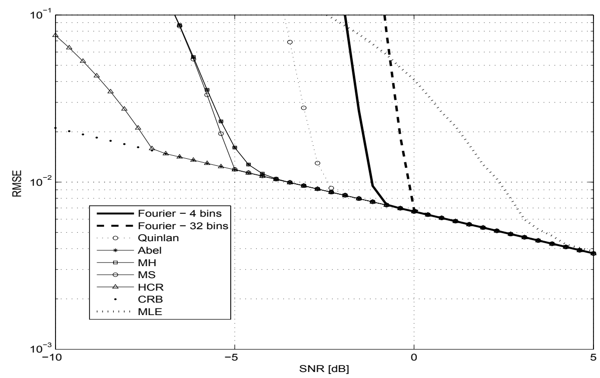

Fig. 2. Comparison of RMSE lower bounds versus SNR.

7. REFERENCES

[1] H. Cramér, "A contribution to the theory of statistical estimation," *Skand. Akt. Tidskr.,* vol. 29, pp. 85-94, 1945.

[2] A. Bhattacharyya, "On some analogous of the amount of information and their use in statistical estimation," *Shankya,* vol. 8, no. 1, pp. 1-14, 201-211, 211-218, 1946.

[3] E.W. Barankin, "Locally best unbiased estimates," *Ann. Math. Stat.,* vol. 20, pp. 477-501, 1946.

[4] J. Albuquerque, "The Barankin bound: A geometric interpretation," *IEEE Transactions on Information Theory,* vol. 19, pp. 559-561, 1973.

[5] D. G. Chapman and H. Robbins, "Minimum variance estimation without regularity assumptions," *Ann. Math. Stat.,* vol. 22, pp. 581-586, 1951.

[6] R. J. McAulay and L. P. Seidman, "A useful form of the Barnakin lower bound and its application to PPM threshold analysis," *IEEE Transactions on Information Theory,* vol. 15, pp. 273-279, 1969.

[7] R. J. McAulay and E. M. Hofstetter, "Barankin bound on parameter estimation," *IEEE Transactions on Information Theory,* vol. 17, no. 6, pp. 669-676, 1971.

[8] J. S. Abel, "A bound on mean-square-estimate error," *IEEE Transactions on Information Theory,* vol. 39, pp. 1675-1680, 1993.

[9] A. Quinlan, E. Chaumette and P. Larzabal, "A direct method to generate approximations of the barankin bound," *Proceedings of the ICASSP 2006,* vol. 3, pp. 808-811, 2006.

[10] G. B. Folland, *Real analysis*, John Wiley and Sons, 1984.

[11] K. Todros and J. Tabrikian, "A new class of lower bounds on the MSE of unbiased estimators," *to be submitted to the IEEE Transactions on signal processing*.

[12] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Stat.,* vol. 22, pp. 79-86, 1951.