ON MAXIMUM LIKELIHOOD ESTIMATION IN THE PRESENCE OF VANISHING INFORMATION MEASURE

Ori Landau and Anthony J. Weiss

Dept. Electrical Engineering - Systems
Tel Aviv University, Tel Aviv 69978, Israel
orila,ajw@eng.tau.ac.il

ABSTRACT

We analyze the parameter estimation Mean Square Error when the *Fisher Information Measure* is zero at some points within the parameter space. At these points the *Cramér-Rao Lower Bound* diverges and no unbiased estimator can achieve a finite Mean Square Error. Under mild regularity conditions the *Maximum Likelihood Estimator* is known to be asymptotically unbiased and therefore lower bounded by the *Cramér-Rao Lower Bound* [1]. It is therefore of interest to examine the Maximum Likelihood Estimator performance in the presence of vanishing Fisher Information Measure. We provide new theoretical and practical results. All results are corroborated by simulations.

1. INTRODUCTION

This work is motivated by the simple example of a linear sensor array used for measuring direction of arrival, θ, with respect to the array normal. The *Cramér-Rao Lower Bound* (CRLB) in this case is inversely proportional to $\cos^2(\theta)$ and therefore diverges as θ approaches $\pi/2$. The mean square error (MSE) of the Maximum Likelihood Estimator (MLE) is expected to diverge as well since the MLE is an asymptotically unbiased estimator with variance lower bounded by the CRLB [2]. However, both simulations and real systems show that this is not the case and the MSE is well behaved even when $\theta = \pi/2$. In this work we explore the MSE for points in the parameter space where the Fisher Information Measure (FIM) is zero and the CRLB diverges. Special attention is given to the vicinity of the parameter interval end points and to points in the vicinity of zero FIM.

2. PROBLEM FORMULATION AND DEFINITIONS

We wish to estimate a real deterministic parameter, θ, known to be within a closed interval $\Theta \equiv [a, b]$. The estimation is based on N measurements $y_n = h(\theta) + v_n, n = 1, \cdots, N$, where $\{v_n\}$ are independent, identically distributed (i.i.d.), zero mean, Gaussian random variables with variance σ_v^2 *i.e.*, $v_n \sim N\left(0, \sigma_v^2\right)$), and $\phi = h(\theta)$, is a non-decreasing function in the interval $\theta \in \Theta$. We define $A \triangleq h(a), B \triangleq h(b)$ so that $\phi \in \Phi \equiv [A, B]$. The inverse function $\theta = h^{-1}(\phi)$ exists in the interval $\phi \in \Phi$ and is non-decreasing, continuous, and smooth, except for a finite number of points where the derivative diverges (these are the zero FIM points). Define the sample mean: $\bar{y} \triangleq \frac{1}{N}\sum_{n=1}^{N} y_n$. By definition and the invariance property we have:

$$\hat{\phi}_{ML} \triangleq \underset{\phi\in\Phi}{\operatorname{argmax}}\, f_{\underline{y}}\left(\underline{y};\phi\right) = \begin{cases} \bar{y} & A \le \bar{y} \le B \\ A & \bar{y} < A \\ B & B < \bar{y} \end{cases} \tag{1}$$

and:

$$\hat{\theta}_{ML} \triangleq \underset{\theta\in\Theta}{\operatorname{argmax}}\, f_{\underline{y}}\left(\underline{y};\theta\right) = h^{-1}\left(\hat{\phi}_{ML}\right) \tag{2}$$

We extend, for convenience, the definition of the inverse function outside it's domain:

$$h^{-1}(x) \triangleq \begin{cases} h^{-1}(x) & A \le x \le B \\ a & x < A \\ b & B < x \end{cases} \tag{3}$$

Now, it is possible to express the estimator as a function of the sample mean:

$$\hat{\theta}_{ML} = h^{-1}(\bar{y}) \tag{4}$$

Let us define the equivalent noise as $\bar{v} \triangleq \frac{1}{N}\sum_{n=1}^{N} v_n$. Obviously, $\bar{v} \sim N\left(0, \sigma_N^2\right)$ where $\sigma_N^2 \triangleq \sigma_v^2/N$. Define a "centered" function that will enable expressing the estimation error as a function of the equivalent noise:

$$g_\theta(x) \triangleq h^{-1}(h(\theta) + x) - \theta \tag{5}$$

Thus the estimation error is:

$$\hat{\theta}_{ML} - \theta = g_\theta(\bar{v}) \tag{6}$$

and the MSE (which is a function of θ) is:

$$MSE\left\{\hat{\theta}_{ML}\right\} = E\left\{g_\theta(\bar{v})^2\right\} = \int_{-\infty}^{\infty} \frac{g_\theta(x)^2 e^{-\frac{x^2}{2\sigma_N^2}}}{\sqrt{2\pi\sigma_N^2}} dx \tag{7}$$

The function $g_\theta(x)$ has the properties:

$$\begin{aligned} &1.\, g_\theta(0) = 0 \\ &2.\, g_\theta^{(1)}(x) \equiv \frac{\partial g_\theta(x)}{\partial x} = \left.\frac{\partial h^{-1}(y)}{\partial y}\right|_{y=h(\theta)+x} \ge 0 \\ &3.\, g_{\theta_2}(x) = g_{\theta_1}\left(x + (h(\theta_2) - h(\theta_1))\right) - g_{\theta_1}\left(h(\theta_2) - h(\theta_1)\right) \end{aligned} \tag{8}$$

We can now show that the MSE of the MLE goes to zero as N increases to infinity. Recall that $\frac{1}{\sqrt{2\pi\sigma_N^2}} e^{-\frac{x^2}{2\sigma_N^2}} \overset{N\to\infty}{\to} \delta(x)$ *i.e.*, as N

increases to infinity, the Gaussian function becomes a Dirac's delta function. Substituting in (7) and using property 1 of (8) gives:

$$MSE\left\{\hat{\theta}_{ML}\right\} \overset{N\to\infty}{\to} \int_{-\infty}^{\infty} g_\theta(x)^2 \,\delta(x)\, dx = g_\theta(0)^2 = 0 \qquad (9)$$

The statement in (9) is rather obvious and not very informative. The question we want to address is how "fast" does the MSE decreases as a function of N at the vicinity of singular points where the FIM becomes zero. We explore this question directly by evaluating the MSE and not by analyzing the *Cramér-Rao Lower Bound* (CRLB). We provide analysis of the CRLB in another work.

3. POINTWISE ASYMPTOTIC EXPRESSIONS

Our first objective is to produce asymptotic expressions for the MSE of the MLE. We are interested in the dependence of the MSE on the number of measurements when θ is fixed, and the number of measurements is large. We replace $g_\theta(x)$ in (7), with an approximation, $\tilde{g}_\theta(x)$. A "good" approximation should satisfy:

$$\frac{\left|E\left\{g_\theta(\bar{v})^2\right\} - E\left\{\tilde{g}_\theta(\bar{v})^2\right\}\right|}{E\left\{g_\theta(\bar{v})^2\right\}} \overset{N\to\infty}{\to} 0 \qquad (10)$$

The normalization is required since all the terms in (10) decrease to zero as N increases to infinity. Since the Gaussian gets "narrower" as N increases, it is fairly easy to achieve (10). For brevity we will not justify our approximations using (10). However, all of the approximations in this contribution satisfy (10).

3.1. Non-Zero FIM Points

In this subsection we reproduce well known results in order to provide a complete discussion. "Regular" points (non-zero FIM points) exhibit a MSE which is inverse proportional to the number of measurements. Even though some points may exhibit a MSE which decreases "faster", these points are not interesting since the overall behavior of the domain is inverse proportional to N.

For "regular" points, the derivative $g_\theta^{(1)}(0)$ doesn't diverge and we can use Taylor expansion as our approximation $\tilde{g}_\theta(x)$ (second and third derivatives subtleties are not addressed here for brevity):

$$\begin{aligned} & g_\theta(x)^2 = g_\theta(0)^2 + 2g_\theta(0)\, g_\theta^{(1)}(0)\, x + \\ & + \tfrac{1}{2}\left(2g_\theta^{(1)}(0)^2 + 2g_\theta(0)\, g_\theta^{(2)}(0)\right) x^2 + ... = \\ & \overset{g_\theta(0)=0}{=} 0 + 0x + g_\theta^{(1)}(0)^2\, x^2 + ... \end{aligned}$$

If $g_\theta^{(1)}(0) = 0$, the MSE decreases "faster" than N^{-1}. If $g_\theta^{(1)}(0) \neq 0$, Substituting the approximation,

$$\tilde{g}_\theta(x)^2 = g_\theta^{(1)}(0)^2\, x^2 \qquad (11)$$

in (7), gives:

$$MSE\left(\hat{\theta}_{ML}\right) \approx E\left\{\tilde{g}_\theta(\bar{v})^2\right\} = g_\theta^{(1)}(0)^2\, \sigma_N^2 \qquad (12)$$

which is inverse proportional to N (because $\sigma_N^2 = \sigma_v^2/N$). More precisely, the interval's edges should be taken into account. Specifically, $g_a(x) = 0$ for $x < 0$ and $g_b(x) = 0$ for $x > 0$, so for $\theta = a$ or $\theta = b$, assuming these are "regular" points,

$$MSE\left(\hat{\theta}_{ML}\right) \approx \frac{1}{2} g_\theta^{(1)}(0)^2\, \sigma_N^2 \qquad (13)$$

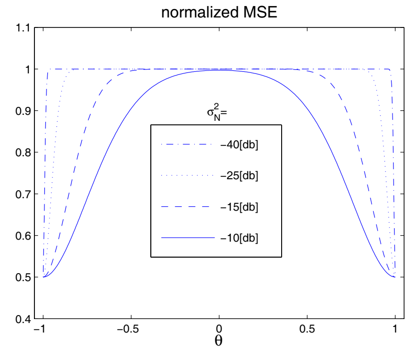


Fig. 1. $h(\theta) = \theta, A = -1, a = -1, B = b = 1, g_\theta^{(1)}(0) = 1$. Normalized MSE ($\frac{MSE\{\hat{\theta}_{ML}\}}{g_\theta^{(1)}(0)^2\sigma_N^2}$) as function of θ for different values of N.

Equation (13) implies a non-uniform convergence. We use an example to demonstrate the convergence: $h(\theta) = \theta, A = a = -1, B = b = 1$. Thus, $g_\theta^{(1)}(0) = 1$. Figure (1) depicts the normalized MSE, $\frac{MSE\{\hat{\theta}_{ML}\}}{g_\theta^{(1)}(0)^2\sigma_N^2}$, as function of θ for different values of σ_N^2. Asymptotically, the normalized MSE for every point in the domain converges to 1, except for the edges, where it converges to $\frac{1}{2}$.

3.2. Zero FIM Points

Let θ_1 be a zero FIM point. In order to approximate $g_{\theta_1}(x)$ about the origin, we assume that the derivative diverges according to:

$$g_{\theta_1}^{(1)}(x) \approx \kappa\, |x|^{-\beta}, \kappa > 0, 0 < \beta < 1 \qquad (14)$$

Although this assumption doesn't hold for all cases, it does hold for most real life applications. Integrating (14) we obtain

$$g_{\theta_1}(x) \approx \frac{\kappa}{1-\beta} |x|^{1-\beta}\, sign(x) \qquad (15)$$

Squaring (15) we get

$$g_{\theta_1}(x)^2 \approx \left(\frac{\kappa}{1-\beta}\right)^2 x^{2(1-\beta)} \qquad (16)$$

Substituting the approximation (16) in (7) yields,

$$MSE\left\{\hat{\theta}_{ML}\right\} \approx \int_{-\infty}^{\infty} \left(\frac{\kappa}{1-\beta}\right)^2 x^{2(1-\beta)} \frac{e^{-\frac{x^2}{2\sigma_N^2}}}{\sqrt{2\pi\sigma_N^2}} dx \qquad (17)$$

Define the function:

$$\begin{aligned} & F_\beta(\alpha) \triangleq |\alpha|^{2(1-\beta)} + \\ & + \tfrac{2^{1-\beta}}{\sqrt{\pi}} \left[\begin{array}{l} \Gamma\left(\frac{3}{2}-\beta\right) {}_1F_1\left(\beta-1, \frac{1}{2}, -\frac{\alpha^2}{2}\right) - \\ - |\alpha|^{2-\beta}\, 2^{1+\frac{\beta}{2}}\, \Gamma\left(\frac{3}{2}-\frac{\beta}{2}\right) {}_1F_1\left(\frac{\beta}{2}, \frac{3}{2}, -\frac{\alpha^2}{2}\right) \end{array} \right] \end{aligned} \qquad (18)$$

where $\Gamma(z)$ is the gamma function and ${}_1F_1(a, b, z)$ is the confluent hypergeometric function of the first kind. Using this definition, the right side of (17) can be shown to be $\left(\frac{\kappa}{1-\beta}\right)^2 F_\beta(0) \left(\sigma_N^2\right)^{1-\beta}$, thus for $\theta = \theta_1$:

$$MSE\left\{\hat{\theta}_{ML}\right\} \approx \left(\frac{\kappa}{1-\beta}\right)^2 F_\beta(0) \left(\sigma_N^2\right)^{1-\beta} \qquad (19)$$

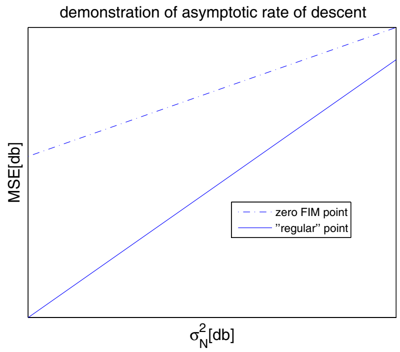


Fig. 2. Asymptotic MSE behavior as function of σ_N^2 for "regular" and zero FIM points. "Regular" points exhibit a slope of 1, whereas the zero FIM point exhibits a slope of $1-\beta$ (the "DC level" is irrelevant here).

The MSE decreases as $N^{-(1-\beta)}$ which is "slower" than N^{-1}. Figure (2) demonstrates this difference in the decreasing rate of the MSE as function of σ_N^2. For "regular" points, the slope is 1, whereas for zero FIM point the slope is $1-\beta$. We define another function (the purpose of the functions $F_\beta(\alpha)$ and $G_\beta(\alpha)$ will become clear in the next section):

$$\begin{aligned} G_\beta(\alpha) &\triangleq |\alpha|^{2(1-\beta)} + \\ &+ \frac{2^{\frac{1}{2}-\beta}}{\sqrt{\pi}} \left[\begin{array}{c} \frac{\Gamma\left(\frac{3}{2}-\beta\right)}{\sqrt{2}} {}_1F_1\left(\beta-1, \frac{1}{2}, -\frac{\alpha^2}{2}\right) + \\ + |\alpha| \Gamma(2-\beta) {}_1F_1\left(\beta-\frac{1}{2}, \frac{3}{2}, -\frac{\alpha^2}{2}\right) \end{array} \right] - \\ &- |\alpha|^{1-\beta} \frac{2^{1-\frac{\beta}{2}}}{\sqrt{\pi}} \left[\begin{array}{c} \frac{\Gamma\left(1-\frac{\beta}{2}\right)}{\sqrt{2}} {}_1F_1\left(\frac{\beta-1}{2}, \frac{1}{2}, -\frac{\alpha^2}{2}\right) + \\ + |\alpha| \Gamma\left(\frac{3}{2}-\frac{\beta}{2}\right) {}_1F_1\left(\frac{\beta}{2}, \frac{3}{2}, -\frac{\alpha^2}{2}\right) \end{array} \right] \end{aligned} \tag{20}$$

If $\theta_1 = a$ or $\theta_1 = b$ then for $\theta = \theta_1$:

$$MSE\left\{\hat{\theta}_{ML}\right\} \approx \left(\frac{\kappa}{1-\beta}\right)^2 G_\beta(0) \left(\sigma_N^2\right)^{1-\beta} \tag{21}$$

It can be easily verified that $G_\beta(0) = \frac{1}{2} F_\beta(0)$. This corresponds to the observation made in (13).

In summary, we obtained pointwise asymptotic expressions for the MSE of the MLE for regular points and zero FIM points. We showed that the MSE of a zero FIM point exhibits a slower decreasing rate. We define the zero FIM normalized MSE as $\frac{MSE\left\{\hat{\theta}_{ML}\right\}}{\left(\frac{\kappa}{1-\beta}\right)^2 F_\beta(0)\left(\sigma_N^2\right)^{1-\beta}}$. The zero FIM normalized MSE converges to 1 as $N \to \infty$ at $\theta = \theta_1$ or to $\frac{1}{2}$ if $\theta_1 = a$ or $\theta_1 = b$, but converges to zero elsewhere! This implies a non-uniform convergence. That means that there is no N' large enough to ensure that for any $N > N'$, the MSE over the whole domain satisfy the asymptotic behavior, *i.e.*, for any finite N, there is a region around the zero FIM point, which is "far" from reaching the N^{-1} asymptotic behavior. This is the reason we used the term "pointwise". A few interesting questions arise:

1. For a given N, is the largest MSE necessarily at $\theta = \theta_1$? If not, why, and what is the value of the worst MSE?
2. What is the highest ratio between the MSE and the asymptotic behavior? is there a bound on this ratio?

We examine these questions in the next section.

4. THE VICINITY OF THE ZERO FIM POINT

We keep the notation of θ_1 being a zero FIM point. Let θ_2 be a nearby point and define the normalized distance between the two points:

$$\alpha \triangleq \frac{h(\theta_2) - h(\theta_1)}{\sigma_N} \tag{22}$$

Using (15), property 3 of (8) in (7), it can be shown that for $\theta = \theta_2$:

$$MSE\left\{\hat{\theta}_{ML}\right\} \approx \left(\frac{\kappa}{1-\beta}\right)^2 F_\beta(\alpha) \left(\sigma_N^2\right)^{1-\beta} \tag{23}$$

which is a generalization of (19). Similarly, if $\theta_1 = a$ or $\theta_1 = b$, then for $\theta = \theta_2$:

$$MSE\left\{\hat{\theta}_{ML}\right\} \approx \left(\frac{\kappa}{1-\beta}\right)^2 G_\beta(\alpha) \left(\sigma_N^2\right)^{1-\beta} \tag{24}$$

which is a generalization of (21). The conclusions are:

1. Obviously, the worst MSE is in the vicinity of the zero FIM point but not necessarily at the point itself. This happens since the MSE is affected not only by the variance, but also by the bias. A nearby point, may have a greater bias than the bias at the zero FIM point, while having almost as large variance, thus having a greater MSE. By finding the maximum value and position of $F_\beta(\alpha)$ (or $G_\beta(\alpha)$), we get the worst case MSE. Further, when the position of the maximum is not at the origin, equations (23) and (24) indicate that this position is at fixed *normalized* distance from the zero FIM point.
2. Let θ_2 be at a small fixed *absolute* distance from θ_1. $g_{\theta_2}^{(1)}(0) = g_{\theta_1}^{(1)}(\alpha\sigma_N)$. Using that with (14) in (12), the asymptotic MSE for $\theta = \theta_2$ is $\left[\kappa\,|\alpha\sigma_N|^{-\beta}\right]^2 \sigma_N^2 = \kappa^2 |\alpha|^{-2\beta} \left(\sigma_N^2\right)^{1-\beta}$. Until reaching this asymptotic behavior, the MSE is given by (23) (or (24)). Define:

$$ratio_\beta(\alpha) \triangleq \left(\frac{1}{1-\beta}\right)^2 F_\beta(\alpha) |\alpha|^{2\beta} \tag{25}$$

Now note that the normalized MSE for $\theta = \theta_2$:

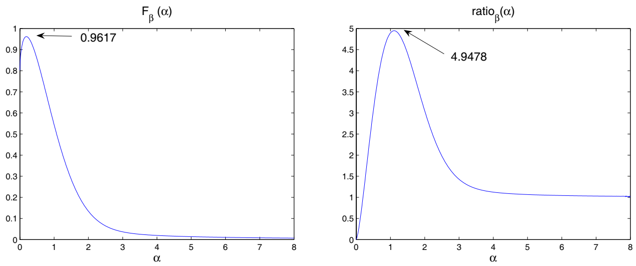
$$\frac{MSE\left\{\hat{\theta}_{ML}\right\}}{g_{\theta_2}^{(1)}(0)^2 \sigma_N^2} \approx \frac{\left(\frac{\kappa}{1-\beta}\right)^2 F_\beta(\alpha)\left(\sigma_N^2\right)^{1-\beta}}{\kappa^2 |\alpha|^{-2\beta} \left(\sigma_N^2\right)^{1-\beta}} = ratio_\beta(\alpha) \tag{26}$$

(if $\theta_1 = a$ or $\theta_1 = b$, use $G_\beta(\alpha)$ instead of $F_\beta(\alpha)$ in (25)). This suggests that $ratio_\beta(\alpha) \overset{\alpha\to\infty}{\to} 1$, which can be easily verified for either $F_\beta(\alpha)$ or $G_\beta(\alpha)$. The smaller the distance $|h(\theta_1) - h(\theta_2)|$, the larger the required N for asymptotic behavior.

5. EXAMPLES

5.1. Zero FIM point within the parameter's domain

Let $\phi = \theta^3, A = a = -1, B = b = 1$. It follows that $\frac{\partial h^{-1}(x)}{\partial x} = \frac{1}{3}|x|^{-\frac{2}{3}}$ (*i.e.* $\kappa = \frac{1}{3}, \beta = \frac{2}{3}$,see (14)) and $g_\theta^{(1)}(0) = \frac{1}{3}\left|\theta^3\right|^{-\frac{2}{3}} = \frac{1}{3\theta^2}$. The asymptotic MSE for "regular" points (12), is therefore: $g_\theta^{(1)}(0)^2 \sigma_N^2 = \frac{1}{9\theta^4}\sigma_N^2$ and the zero FIM point is $\theta_1 = 0$. Figure 3(a) depicts $F_{\frac{2}{3}}(\alpha)$. Note that the maximum is not at $\alpha = 0$. Figure 3(b) shows that for any point in the domain, the MSE is at most 5 (approximately) times the MSE predicted by the "regular" asymptotic expression. Figure 4 and 5 shows the normalized MSE and the zero FIM normalized MSE, respectively.



(a) $F_{\frac{2}{3}}(\alpha)$ (b) $ratio_{\frac{2}{3}}(\alpha)$

Fig. 3. Example 1. (a) $F_\beta(\alpha)$ and (b) $ratio_\beta(\alpha)$.

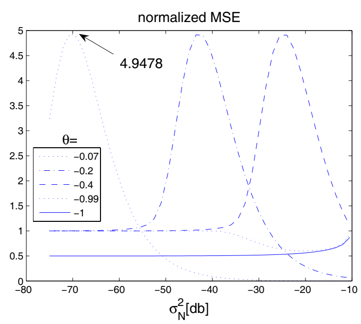


Fig. 4. Normalized MSE for Eaxmple 1. The closer the point to the zero FIM point, the larger the required N for asymptotic behvior.

5.2. Zero FIM point on the parameter's domain edges

When two spatially separated antennas receive the transmission from a far-field point source, the phase difference between the observed signals is proportional to the sine of the transmitter bearing with respect to the normal of the antennas baseline. This fact is frequently used for bearing estimation. Let $\phi = \sin(\theta), \theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right], \phi \in [-1, 1]$. It follows that $\frac{\partial h^{-1}(x)}{\partial x} = \frac{1}{\sqrt{1-x^2}} \overset{x=-1+\varepsilon}{=} \frac{1}{\sqrt{1-(-1+\varepsilon)^2}} \approx \frac{1}{\sqrt{1-(1-2\varepsilon)}} = \frac{1}{\sqrt{2}}\varepsilon^{-\frac{1}{2}}$ (*i.e.* $\kappa = \frac{1}{\sqrt{2}}, \beta = \frac{1}{2}$) and $g_\theta^{(1)}(0) = \frac{1}{\sqrt{1-\sin(\theta)^2}} = \frac{1}{|\cos(\theta)|}$. The asymptotic MSE for "regular" points (12), is therefore:$g_\theta^{(1)}(0)^2\sigma_N^2 = \frac{1}{\cos(\theta)^2}\sigma_N^2$ and the zero FIM point

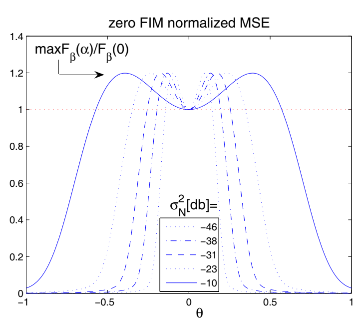
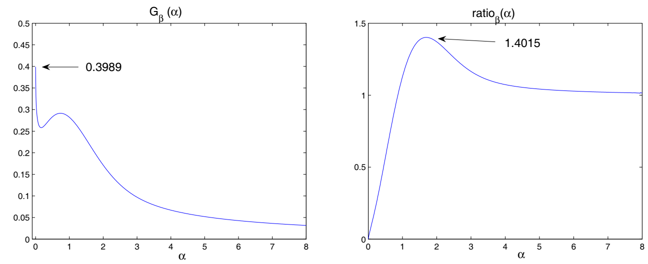


Fig. 5. zero FIM normalized MSE for Example 1. The point with the worst MSE gets closer to the zero FIM point as N increases.



(a) $G_{\frac{1}{2}}(\alpha)$ (b) $ratio_{\frac{1}{2}}(\alpha)$

Fig. 6. Example 2. (a) $G_\beta(\alpha)$ and (b) $ratio_\beta(\alpha)$.

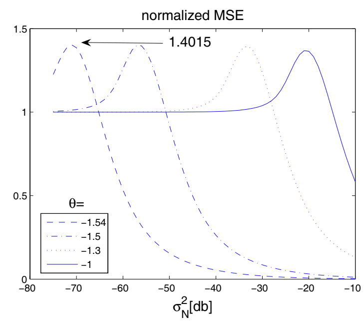


Fig. 7. Normalized MSE for Eaxmple 2. The closer the point to the zero FIM point, the larger the required N for asymptotic behvior.

is $\theta_1 = \pm\frac{\pi}{2}$. Figure 6 depicts $G_{\frac{1}{2}}(\alpha)$ and $ratio_{\frac{1}{2}}(\alpha)$. Figure 7 and 8 shows the normalized MSE and the zero FIM normalized MSE, respectively.

6. REFERENCES

[1] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol.1 Estimation Theory*. Prentice Hall, 1993.

[2] H. Cramer, *Mathematical Methods of Statistics*. Cambridge, 1946.

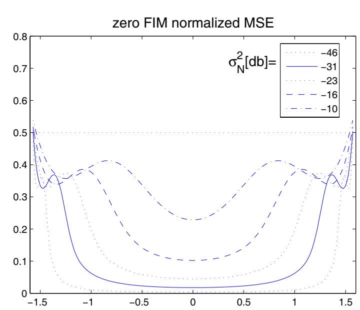


Fig. 8. zero FIM normalized MSE for Example 2. The high MSE region gets narrower as N increases.