Error Intensity Measures for Multi-Parameter Tracking and Passive Bearing Estimation

Joong K. Kim and Alfred O. Hero

Department of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor, MI 48109

ABSTRACT

We present a general method for approximating the global error performance of ML type multi-parameter estimators. The method is a multi-dimensional extension of the approaches of [2] and [1]. Comparisons between the global approximation to the MSE for known and unknown "nuisance" parameters indicates the degree to which nuisance parameters degrade the performance of ML estimators. The results of a numerical study illustrate the implementation of our method for time delay estimation in cases of coherent interference and doppler.

I. Introduction

This paper presents an extension of the method [1] [2] of error intensity analysis of maximum likelihood (ML) type estimators to the case of multiple unknown parameters. The method involves calculating the intensity function of a multi-dimensional point process, called an error candidate point process, which is associated with the local maximum of the ambiguity surface. Under a Poisson model for the point process and an equally likely model for the distribution of the global maxima over the set of local maxima, the large error MSE of a ML type estimator of one of the parameters is given by the radius of gyration of the marginal one dimensional intensity function obtained by integrating the multi-dimensional intensity function over all other parameters. As in [1], the global MSE is obtained by taking a convex combination of the large error MSE and the local MSE, e.g. calculated from the *multidimensional* Cramer-Rao (CR) bound. We will illustrate the method for problems associated with active and passive radar/sonar. Specifically, we will consider the effects of unknown doppler and the presence of an unknown interference on the large error probability and the MSE of ML type time delay estimators.

Based on our investigation, the following general results can be stated. The principal effect of nuisance parameters on the intensity function of single parameter estimation errors is a broadening and smoothing of the marginal intensity function, which is accompanied by an increase in its radius of gyration. The degree to which the radius of gyration increases is controlled by the size of the *a priori* region for the nuisance parameters, and by the coupling among the parameters through the statistics of the ambiguity function. Since the radius of gyration of the intensity is directly related to the MSE, our approximation behaves in a manner consistent with expectations that nuisance parameter uncertainty can only worsen the MSE performance.

II. Background

An observation $\underline{x}$, whose p.d.f., $f(\underline{x}|\underline{\theta})$, is parameterized by a column vector of K unknown but deterministic signal parameters $\underline{\theta} = [\theta_1, \cdots, \theta_K]^T$, is taken over a finite interval of time $t \in [0, T]$. The unknown parameter vector $\underline{\theta}$ is restricted to lie within a K-dimensional *a priori* region $\Theta_i = \times_{i=1}^{K} \Theta_i$ which is a continuous subset of $\mathbb{R}^K$. An ML type estimate $\hat{\underline{\theta}}$ of $\underline{\theta}$ is any estimator obtained by maximizing a statistic, called the ambiguity function, $\ell(\underline{\theta})$ within the *a priori* region $\underline{\Theta}$. Although not absolutely necessary for the following [4], it will be assumed that $\ell(\underline{\theta})$ has derivatives of all orders less than three with respect to each component θ_i, $i = 1, \cdots, K$.

We say that $\hat{\underline{\theta}}$ incurs a small error when $|\hat{\theta}_i - \theta_i^{true}| < \delta_i$, $i = 1, \ldots, K$, for small positive δ_i. Otherwise the error is large. The small and large error regions will be denoted $\underline{\Theta}^S$ and $\underline{\Theta}^L$ respectively. By the total probability law, the MSE matrix $\Omega_{\underline{\theta}}$ can be expressed in terms of the conditional MSE matrices $\Omega_{\underline{\theta}^S}$ and $\Omega_{\underline{\theta}^L}$ given $\hat{\underline{\theta}} \in \underline{\Theta}^S$ and $\hat{\underline{\theta}} \in \underline{\Theta}^L$ respectively:

$$\begin{aligned} \Omega_{\underline{\theta}} &= E[(\hat{\underline{\theta}} - \underline{\theta}^{true})(\hat{\underline{\theta}} - \underline{\theta}^{true})^T] \\ &= \Omega_{\underline{\theta}^S} \cdot (1 - P_e) + \Omega_{\underline{\theta}^L} \cdot P_e, \end{aligned} \tag{1}$$

$P_e = P(\hat{\underline{\theta}} \in \underline{\Theta}^L)$ is the probability of large error.

While the expression (1) is valid for any size of $\underline{\Theta}^S$, for sufficiently small δ_i, $i = 1, \ldots, K$, $\Omega_{\underline{\theta}^S}$ can be replaced with a local approximation to the MSE matrix in the neighborhood of the true parameter using standard linearization methods. For true ML estimation, the matrix form of the CR bound can be used in place of $\Omega_{\underline{\theta}}^S$ [4]. In the rest of this paper we focus on the development of approximations to the large error probability, P_e, and the large error MSE matrix $\Omega_{\underline{\theta}^L}$.

Similarly to [2], define the "ambiguity process" $\Delta\ell(\underline{\theta}) \stackrel{\text{def}}{=} \ell(\underline{\theta}) - m_\delta$, where m_δ is the maximum value of $\ell(\underline{\theta})$ over $\underline{\Theta}^S$. Observe that since the global maximum location of the ambiguity process is the ML (type) estimate of the parameters, the occurrence times and relative magnitudes of the positive local maxima of the ambiguity process provides an equivalent description of large error. In general, we define the "set of candidate locations for (large) error" as any set of random variables $\{\underline{\vartheta}_i\}_{i=1}^M$ such that $\underline{\theta} \in \{\underline{\vartheta}_i\}_{i=1}^M$ with probability one, where M is a random variable representing the number of candidate locations within the large error region. The $\{\underline{\vartheta}_i\}_{i=1}^M$ correspond to a set of "sampling points" of the surface of the ambiguity function, which form the points of a multi-dimensional point process N. The point process (complete) intensity thus characterizes the statistics of the sampling points.

CH2673-2/89/0000-2641 $1.00 © 1989 IEEE

III. Error Intensity Measure

Since the set of error candidate locations $\{\underline{\ell}_i\}_{i=1}^M$ contains $\hat{\underline{\ell}}$ with probability one, an exact representation of any risk, or average loss, $\Re \stackrel{def}{=} E[g(\underline{\ell}^{true}, \hat{\underline{\ell}})]$, associated with the estimator is obtained via iterated expectations:

$$\begin{aligned} \Re(\hat{\underline{\ell}}) &= E[E[g(\underline{\ell}^{true}, \hat{\underline{\ell}})|\hat{\underline{\ell}} \in \{\underline{\ell}_i\}_{i=1}^M]] \qquad (2) \\ &\approx E[\sum_{j=1}^{M} g(\underline{\ell}^{true}, \underline{\ell}_j) \cdot P(\hat{\underline{\ell}} = \underline{\ell}_j | \hat{\underline{\ell}} \in \{\underline{\ell}_i\}_{i=1}^M)], \end{aligned}$$

where $g(\bullet,\bullet)$, in (2) is a general loss function. The special cases $g(\underline{u},\underline{v}) = \Pi_{i=1}^K I(|u_i - v_i| > \delta_i)$, $g(\underline{u},\underline{v}) = I(|u_1 - v_1| > \delta_1)$, and $g(\underline{u},\underline{v}) = (\underline{\ell} - \hat{\underline{\ell}})(\underline{\ell} - \hat{\underline{\ell}})^T$ give the large error probability $P_e(\hat{\underline{\ell}})$, the large error probability $P_e(\hat{\theta}_1)$, and the large error MSE matrix $\Omega_{\underline{\Theta}^L}$, respectively.

As can be seen from (2), to make use of the exact point process representation, we need to specify the joint statistics of the candidate locations $\{\underline{\ell}_i\}_{i=1}^M$,or equivalently the point process N, and the conditional probability distribution of $\hat{\underline{\ell}}$ over the candidate error locations: $P_j \stackrel{def}{=} P(\hat{\underline{\ell}} = \underline{\ell}_j | \{\underline{\ell}_i\}_{i=1}^M)$ for $j = 1,\ldots,M$. Analogously to [2] and [1] we model the points $\{\underline{\ell}\}$ as a multi-dimensional Poisson process. We consider, as examples, the following three models for P_j, $j = 1,\ldots,M$:

1. *Exact Model*: $\{\underline{\ell}_i\}$ = random point process.

$$P_j = \int_{-\infty}^{\infty} f_j(x) \cdot F_{1,\ldots,j-1,j+1,\ldots M|j}(z,\ldots,x|x)dx, \qquad (3)$$

2. *Equi-partition Model*: $\{\underline{\ell}\}$ = uniform grid.

$$P_j = \int_{-\infty}^{\infty} f_j(x) \cdot [\prod_{k \neq j}^{M} F_{k|j}(x|x)]dx, \qquad (4)$$

3. *Uniform Model*: $\{\underline{\ell}_i\}$ = random point process.

$$P_j = \frac{1}{M}. \qquad (5)$$

In the exact model (3), $f_j(\bullet)$ is the marginal probability density function of $\Delta\ell(\underline{\ell}_j)$ and $F_{1,\ldots,j-1,j+1,\ldots M|j}$ is the conditional joint distribution function of $\{\Delta\ell(\underline{\ell}_i)\}_{i\neq j}^M$ given $\Delta\ell(\underline{\ell}_j)$. Even though P_j in (3) is in a well defined form, it is difficult to use this model for analysis due to the complexity involved in computing joint distributions $F_{1,\ldots,j-1,j+1,\ldots M|j}$. The equi-partition model, on the other hand, fixes the values of $\{\underline{\ell}_i\}_{i=1}^M$ on a uniform grid, and imposes the assumption of conditional independence on the samples $\{\Delta\ell(\underline{\ell}_i)\}_{i=1}^M$ given $\Delta\ell(\underline{\ell}_j)$. The equi-partition model is a generalization of the method proposed by Ianniello [3] for large error performance analysis of time delay estimators. In (4), $F_j(\bullet)$ represents the marginal cumulative probability distribution of $\Delta\ell(\underline{\ell}_j)$. The uniform model (5) for P_j assigns equal probabilities to $\hat{\underline{\ell}}$ over each of the candidate error points $\underline{\ell}_j$.

Under the Poisson approximation to the point process $\{\underline{\ell}\}_{i=1}^M$, and the uniform model (5), the large error probability P_e and the MSE matrix $\Omega_{\underline{\Theta}^L}$ can be derived [4]:

$$\begin{aligned} P_e &= 1 - e^{-\Lambda}, \qquad (6) \\ \Omega_{\underline{\Theta}^L} &= \int_{\underline{\Theta}^L} (\underline{\ell} - \underline{\ell}^{true})(\underline{\ell} - \underline{\ell}^{true})^T \cdot \tilde{\lambda}(\underline{\ell})d\underline{\ell}, \qquad (7) \end{aligned}$$

where $\tilde{\lambda}(\underline{\ell})$ is a normalized version of the intensity, $\lambda(\underline{\ell})$, of the point process $\{\underline{\ell}_i\}_{i=1}^M$ over $\underline{\Theta}^L$ ($\int_{\underline{\Theta}} \tilde{\lambda}(\underline{\ell})d\underline{\ell} = 1$), and Λ is the average number of points $E[M]$. In view of (7) and (6), the Poisson process model for $\{\underline{\ell}_i\}_{i=1}^M$ and the uniform model (5) for $\{P_i\}_{i=1}^M$ induce an approximation to the joint probability density of the vector estimator $(\hat{\theta}_1,\cdots,\hat{\theta}_K)$ by the normalized intensity $\tilde{\lambda}(\underline{\ell})$, and the propensity for large error is measured by the average number, Λ, of zero exceeding local maxima of the ambiguity process.

The $(\ )_{11}$ element of the MSE matrix (7) can be considered as the MSE of the estimator $\hat{\theta}_1$ in the presence of $K-1$ unknown "nuisance" parameters $\{\theta_i\}_{i=2}^K$ when ML estimation is used:

$$MSE(\hat{\theta}_1) = \int_{\underline{\Theta}_1^L} (\theta_1 - \theta_1^{true})^2 \cdot \tilde{\lambda}(\theta_1)d\theta_1, \qquad (8)$$

where $\tilde{\lambda}(\theta_1)$ is the marginal normalized intensity corresponding to errors in $\hat{\theta}_1$

$$\tilde{\lambda}(\theta_1) \stackrel{def}{=} \int_{\Theta_2} d\theta_2 \cdots \int_{\Theta_K} d\theta_K \tilde{\lambda}(\underline{\ell}). \qquad (9)$$

From relation (8), we can deduce some implications of the unknown nuisance parameters on the error associated with $\hat{\theta}_1$. By virtue of the $K-1$ integration in (9), the marginal normalized intensity is in general smoother than the θ_1 cross-section of the multi-dimensional intensity $\tilde{\lambda}(\underline{\ell})$ for $\{\theta_i\}_{i=2}^M$ in the neighborhood of the true values $\{\theta_i^{true}\}_{i=2}^M$. This smoothing causes a reduction in the concentration of the intensity about the true parameter and a corresponding increase in MSE. The degree to which this smoothing occurs depends on two factors: the size of the *a priori* search regions $\{\theta_i\}_{i=2}^M$; and the coupling between θ_1 and the nuisance parameters manifested in the multi-dimensional intensity function $\tilde{\lambda}(\underline{\ell})$.

In the context of estimation of θ_1, it is worthwhile to consider the difference between known parameters $\{\theta_i\}_{i=2}^K$ and unknown nuisance parameters, treated above. When these parameters are known, the ambiguity function evolves over the line $\Theta_1 \times \{\theta_2^{true}\} \times \cdots \times \{\theta_K^{true}\}$. Hence the local maxima intensity, denoted $\lambda(\theta_1|\theta_2^{true},\ldots,\theta_K^{true})$, is a one-dimensional function parameterized by $\{\theta_i^{true}\}_{i=2}^K$. The MSE approximation (8) for $\hat{\theta}_1$ is therefore directly implemented by using the above intensity in place of $\lambda(\theta_1)$. It is important to note that the marginal intensity under multiple unknown nuisance parameters $\lambda(\theta_1)$ (9) has no relation to the one-dimensional intensity for multiple known parameters $\lambda(\theta_1|\theta_2,\ldots,\theta_K)$ since the point process representations evolve over spaces of different dimension. In particular, the analog to the classical conditional probability relation $\lambda(\theta_1|\theta_2,\ldots,\theta_K) = \lambda(\theta_1,\ldots,\theta_K)/\lambda(\theta_2,\ldots,\theta_K)$ is not generally valid.

IV. Applications

In this section we will illustrate the implementation of the intensity measures for large error, presented in the previous section. Although, in general, the local maxima on the interior and on the boundary of $\underline{\Theta}^L$ must be considered separately [4], the boundary local maxima are neglected in what follows. The intensity function for the positive interior local maxima, including saddlepoints, of the ambiguity process $\Delta\ell(\underline{\ell}) = \ell(\underline{\ell}) - m_\ell$ has the exact form:

$$\tilde{\lambda}(\underline{\ell}) = \int_0^\infty dx \int_0^\infty dz_1 \int_0^\infty dz_K \; z_1 \cdots z_K f(x, \underline{0}, -\underline{z}), \qquad (10)$$

where $f(x,\underline{y},\underline{z})$ is the joint distribution of the $2K+1$-vector consisting of $\Delta\ell(\underline{\ell})$ and its first and second derivatives (excluding mixed partials). For the numerical computations below, the MSE and P_e were computed using a Gaussian approximation to $f(x,\underline{y},\underline{z})$. In the figures the MSE is nor-

malized with respect to a uniform variable over the *a priori* region. Further details can be found in [4].

Passive Time Delay Estimation with Interference

Available for observation over a time interval $[0,T]$ are two sensor outputs with a common delayed Gaussian signal s_1 and Gaussian interference s_2, with PSD's G_{s_1} and G_{s_2}, in independent white Gaussian noises n_1 and n_2:

$$\begin{aligned} x_1(t) &= s_1(t) + s_2(t) + n_1(t) \\ x_2(t) &= s_1(t - D_1) + s_2(t - D_2) + n_2(t). \end{aligned} \quad (11)$$

In the notation of the previous section $\theta_1^{true} \stackrel{def}{=} D_1$ and $\theta_2^{true} \stackrel{def}{=} D_2$. The ML estimator of D_1 and D_2 is the peak location of the two-dimensional log-likelihood surface, which is derived in [4] under the assumptions that T and $|D_1 - D_2|$ are large relative to the signal decorrelation times:

$$\begin{aligned} \ell(\theta_1,\theta_2) &= \int_{-\infty}^{\infty} \frac{X_1(\omega)X_2^*(\omega)}{G_{s_1}(\omega)} W(\omega) e^{-j\omega\theta_1} d\omega \quad (12) \\ &\quad + \int_{-\infty}^{\infty} \frac{X_1(\omega)X_2^*(\omega)}{G_{s_1}(\omega)} \frac{G_{s_2}}{G_{s_1}} W(\omega) e^{-j\omega\theta_2} d\omega, \end{aligned}$$

where $X_i(w)$ is the finite time Fourier transform of $x_i(t)$ over $[0,T]$, and $W(w)$ is a weighting function independent of θ_1 and θ_2.

Figures 1-4 show numerical results for the case of flat G_{s_1} and G_{s_2} with identical $20(Hz)$ bandwidths. The observation interval is $T = 8(sec)$, $\theta_1 \in [-0.5, 0.5]$, $\theta_2 \in [0.5, 1.5]$, the true values of the parameters are $D_1 = 0(sec)$, $D_2 = 1(sec)$, the small error region is $\Theta^S = [-0.05, 0.05] \times [0.95, 1.05]$, and the SIR is 3(dB). These parameter values represent a scenario for which a broadband moderate power interferer impinges on a two sensor array from a known sector of azimuthal angle, but with unknown position. In Fig. 1, the mean ambiguity surface $E\{\Delta\ell\ (\theta_1,\theta_2)\}$ (12) is displayed, and in Fig. 2 the intensity surface $\lambda(\theta_1,\theta_2)$ is displayed for a SNR of $-10(dB)$. From these figures the intensity surface can be seen to take a maximum value at the true parameters D_1 and D_2 and display a suppressed but similar sidelobe structure to the ambiguity surface. The relatively higher concentration of the error intensity over θ_1 as compared to over θ_2 is indicative of better ML estimator performance for the stronger signal. In Fig. 3 the one dimensional intensity functions for the ML estimator of D_1 are plotted for unknown D_2, obtained by integrating Fig. 2 over θ_2, and for known D_2, obtained directly from (10) with $K = 1$. From the plots of normalized $MSE(\hat{\theta}_1)$ Fig. 4, obtained from the radius of gyration of the intensities of Fig. 2, we note that, consistent with expectation, ML estimation error for $\hat{\theta}_1$ is larger on the average for unknown D_2 than it is for known D_2.

Active Time Delay Estimation with Doppler

For this case the observation model is :

$$x(t) = s(d(t - D)) + n(t), \quad (13)$$

where $s(t)$ is a known deterministic signal, D is an unknown time delay, d is a positive "doppler ratio" and $n(t)$ is a white Gaussian noise. In the notation of the previous section $\theta_1^{true} \stackrel{def}{=} D$ and $\theta_2^{true} \stackrel{def}{=} d$. The log-likelihood function for θ_1 and θ_2 is proportional to:

$$\ell(\theta_1,\theta_2) = \int_{-\infty}^{\infty} X(w) S^*(\frac{w}{\theta_2}) e^{jw\theta_1} dw, \quad (14)$$

where $S(w)$ is the Fourier transform of $s(t)$.

Figures 5-8 show numerical results for the case of a Gaussian envelope r.f. signal wth center frequency $f_o = 2.5(Hz)$ and bandwidth $B = 0.25(Hz)$. In Figs. 5-7, $\theta_1 \in [-4,4]$, $\theta_2 \in [0.8, 1.2]$, the true values of the parameters are $D = 0(sec)$ and $d = 1$, and the small error region is $\Theta^S = [-0.01, 0.01] \times [0.9, 1.1]$. These values correspond to radar ranging of a target which is known *a priori* to be slowly moving. Figure 8 was obtained for the same parameters as above except the doppler has a larger *a priori* region $\theta_2 \in [0.1, 1.9]$. Figures 5 and 6 show the ambiguity surface and the intensity surface for SNR $= -10(dB)$. Note that for a fixed value of the time delay variable, θ_1, near D, the intensity surface has no sidelobes over values of the doppler variable θ_2. On the other hand, for a fixed value of θ_2 near d, the intensity over θ_1 has high sidelobes. This implies that, for a small *a priori* region of doppler, the ML time delay estimator is more prone to peak ambiguity errors than the ML doppler estimator. This penchant for peak ambiguity is reflected in the presence of multiple SNR thresholds in $MSE(\hat{\theta}_1)$ (Fig. 7), equal to the radius of gyration of the marginal intensity $\bar{\lambda}(\theta_1) = \int_{\Theta_2} \lambda(\theta_1,\theta_2) d\theta_2$. On the other hand, for a larger *a priori* doppler region, the sidelobes in $\bar{\lambda}(\theta_1)$ become significantly smoother, due to the diverging ridges on $\lambda(\theta_1,\theta_2)$, and hence $MSE(\hat{\theta}_1)$ (Fig. 8), loses the multiple thresholds.

V. Conclusion

In this paper we have presented an extension of the error intensity measure analysis of [1] to multiple parameters. The critical component of the analysis is the computation of the intensity function of the local maxima of the multi-dimensional ambiguity surface over the parameter space. For the case of Gaussian ambiguity surfaces this calculation is straightforward. The technique should be valuable in investigating the global performance of ML type estimators in the presence of unknown parameters.

References

[1] A.O. Hero. Applications of error intensity measures to bearing estimation. *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, April, 1987.

[2] A.O. Hero and S.C. Schwartz. Poisson models and mean-square-error for correlator estimators of time delay. *IEEE Trans. Information Theory*, IT-34:287-303, March, 1988.

[3] J.P. Ianniello. Time delay estimation via cross-correlation in the presence of large estimation errors. *IEEE Trans. Acoustics, Speech, and Signal Processing.*, (6):998-1003, Dec. 1982.

[4] J.K. Kim and A.O. Hero. Error intensity measures for multi-dimensional maximum likelihood type estimators. Technical Report in preparation, Communications and Signal Processing Laboratory, Dept. EECS, University of Michigan, Ann Arbor.

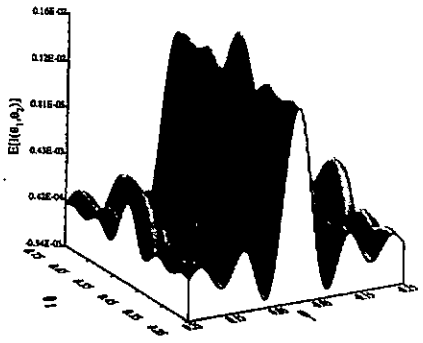

Fig. 1: Mean ambiguity function $E[\ell(\theta_1,\theta_2)]$ for passive time delay estimation with interference. $\theta_1^{true} = D_1, \theta_2^{true} = D_2$, $SNR = -10dB$.

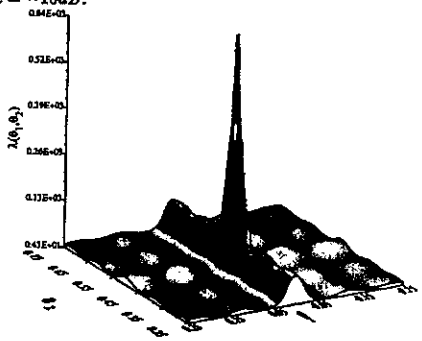

Fig. 2: Error intensity, $\bar{\lambda}(\theta_1,\theta_2)$, corresponding to Fig. 1. $SNR = -10dB$.

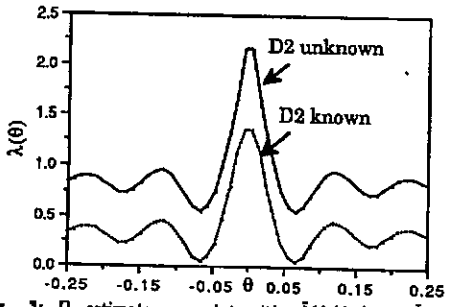

Fig. 3: D_1 estimator error intensities $\bar{\lambda}(\theta_1|D_2)$ and $\bar{\lambda}(\theta_1)$ for known and unknown D_2.

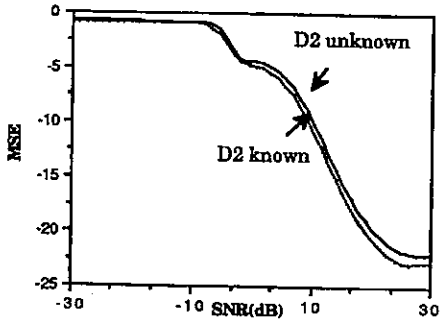

Fig. 4: Normalized $MSE(\hat{\theta}_1)$, computed from radius of gyration of $\bar{\lambda}(\theta_1|D_2)$ and $\bar{\lambda}(\theta_1)$ (Fig. 3), respectively.

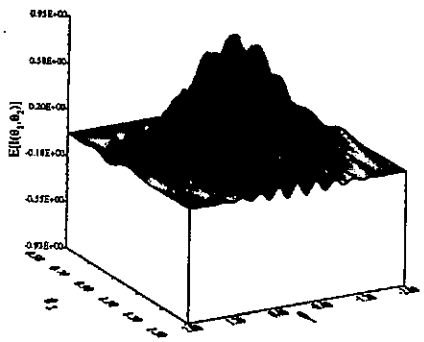

Fig. 5: $E[\ell(\theta_1,\theta_2)]$ for active time delay (D) estimation in the presence of doppler (d). $\theta_1^{true} = D, \theta_2^{true} = d$, $SNR = -10dB$.

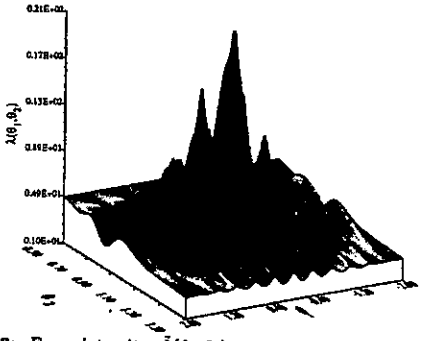

Fig. 6: Error intensity, $\bar{\lambda}(\theta_1,\theta_2)$, corresponding to Fig. 5. $SNR = -10dB$.

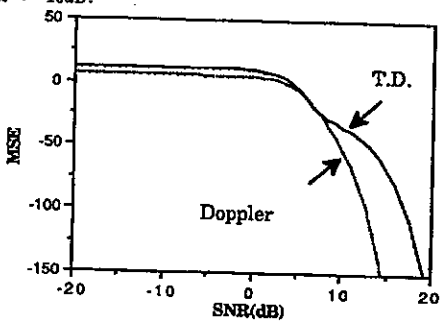

Fig. 7: Normalized $MSE(\theta_1)$ and $MSE(\theta_2)$ obtained from the radius of gyration of the marginal intensities, $\bar{\lambda}(\theta_1)$ and $\bar{\lambda}(\theta_2)$ obtained from Fig. 6, with small *a priori* doppler region.

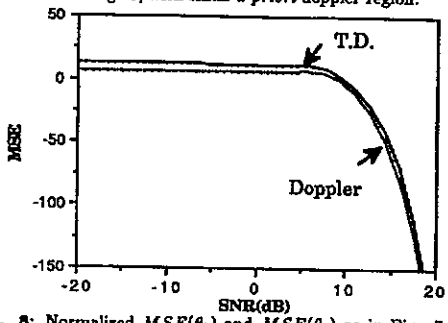

Fig. 8: Normalized $MSE(\theta_1)$ and $MSE(\theta_2)$ as in Fig. 7 but with large *a priori* doppler region.