ON MAXIMUM LIKELIHOOD ESTIMATION AND ITS RELEVANCE TO TIME DELAY ESTIMATION

by

JEN-WEI KUO

B.S., National Taiwan Institute of Technology, 1978

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Electrical Engineering

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1982

Approved by:

Nasir Ahmed

Major Professor

ABSTRACT

The purpose of this report is to study the basic concepts of maximum likelihood (ML) estimation, and its relevance to time delay estimation. The ML estimate is derived for the difference between the times of arrival of a signal due to a source and received at two sensors. A lower bound for the variance of this estimate is also derived. Sufficient details related to the derivations are included so that this report can be used to good advantage for pedagogical purposes.

ILLEGIBLE

THE FOLLOWING DOCUMENT (S) IS ILLEGIBLE DUE TO THE PRINTING ON THE ORIGINAL BEING CUT OFF

ILLEGIBLE

TABLE OF CONTENTS

I. MAXIMUM LIKELIHOOD ESTIMATION

1.1 Estimation Model

A commonly used estimation model is shown in Fig. 1.1. The model has the following four components:

A. Parameter Space

The output of the information sources is a parameter (or variable), which is viewed as a point in a parameter space. The parameter space can be classified into two cases as follows:

i. The parameter is a random variable whose behavior is governed by a probability density function.

ii. The parameter is an unknown quantity A but not a random variable.

B. Probabilistic Mapping

This defines how a variable is mapped from the parameter space onto the observation space.

C. Observation Space

We can make n arbitrary observations which constitute a point in an n-dimensional space. We denote this point by the vector $\underline{R}$.

D. Estimation Rule

After observing $\underline{R}$, we wish to estimate the value of A, which we denote this as $\hat{a}(\underline{R})$. This mapping of the observation space into an estimate is called an estimation rule.

1.2 Methodology

Maximum likelihood (ML) estimation concerns the estimation of a parameter that causes an observation to be most likely to occur. For example, if an unknown parameter A is corrupted by zero-mean Gaussian noise, see Fig. 1.2(a), then the conditional density of an observation R is given by

$$p_{r/a}(R/A) = (\sqrt{2\pi}\,\sigma_n)^{-1} \exp\left[-\frac{1}{2\sigma_n^2}(R - A)^2\right]$$

which is shown in Fig. 1.2 (b). We note that $p_{r/a}(R/A)$ is maximum at an observed value which is most likely to occur, namely $A = R_{max}$. Thus we choose the value R_{max} to be the maximum likelihood estimate of A denoted as $\hat{a}_{ml}(R)$. It follows that we can use the following to calculate the value of $\hat{a}_{ml}(R)$.

(i) Single observation case

$$\left.\frac{\partial \ln p_{r/a}(R/A)}{\partial A}\right|_{A = \hat{a}_{ml}(R)} = 0$$

(ii) n observations case

$$\left.\frac{\partial \ln p_{\underline{r}/a}(\underline{R}/A)}{\partial A}\right|_{A = \hat{a}_{ml}(\underline{R})} = 0$$

where $\underline{R}$ is an n-vector.

In order to assess the effectiveness of the ML procedure, we may compute the variance of the ML estimate so obtained. However, it is usually difficult to compute the variance of an ML estimate. Thus a lower bound on the variance on any unbiased

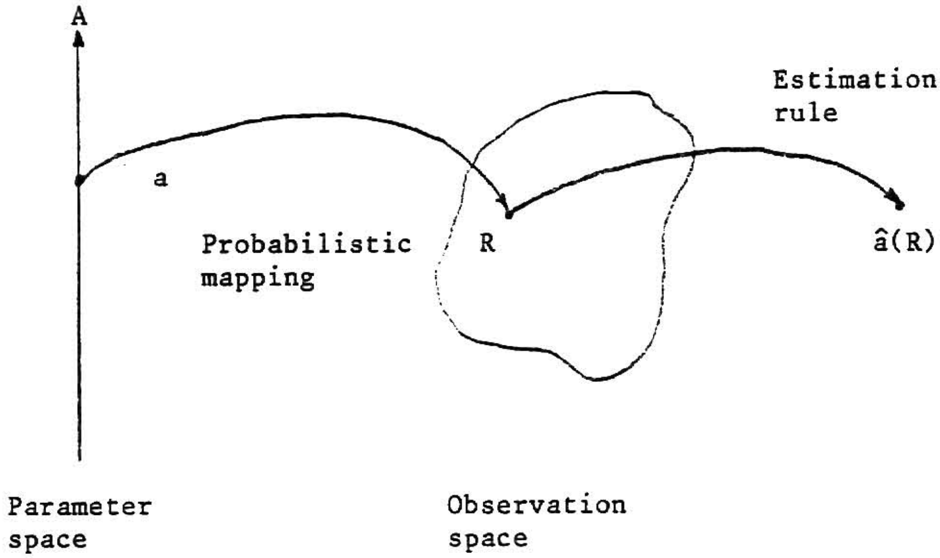


Fig. 1.1 Estimation Model [1]

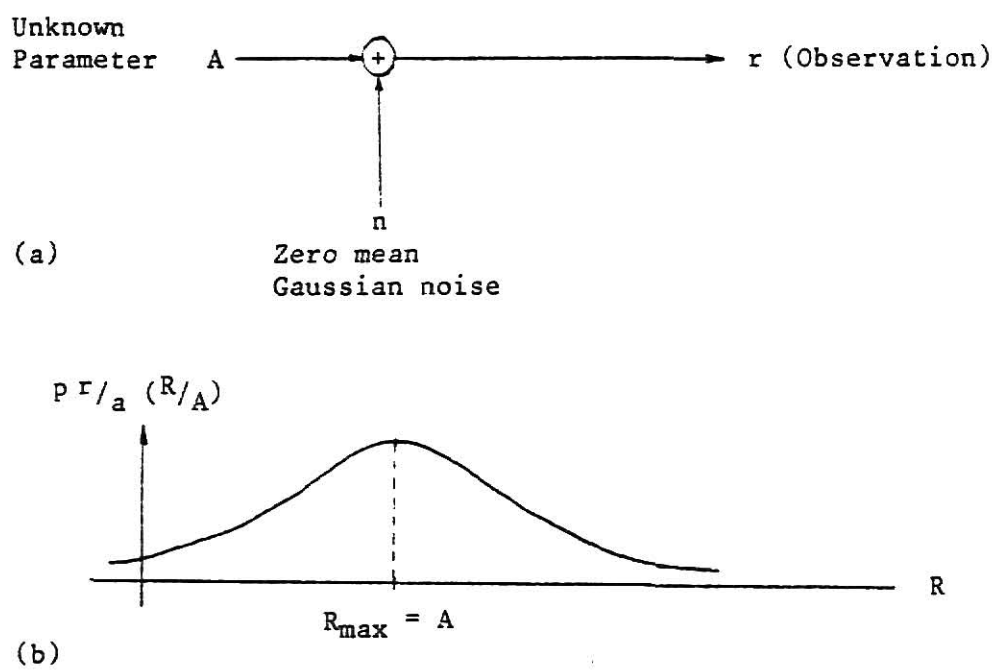


Fig. 1.2 Example of ML estimation

estimate is computed via the following Cramér-Rao lower bound (CRLB) [1]:

$$\sigma^2_{\hat{a}(\underline{R})} \geq \frac{-1}{E\left\{\dfrac{\partial^2 \ln p_{\underline{r}/a}(\underline{R}/A)}{\partial A^2}\right\}}$$

where $\hat{a}(\underline{R})$ is any unbiased estimate of A.

II. RELEVANCE TO TIME DELAY ESTIMATION

In this section we will study the relevance of ML estimation to the problem of time delay estimation (TDE). To this end we consider the two-sensor model.

$$\begin{aligned} x_1(t) &= s_1(t) + n_1(t) \\ x_2(t) &= \alpha\, s_1(t+D) + n_2(t) \end{aligned} \qquad (2\text{-}1)$$

where $s_1(t)$ is the signal, α is an attenuation factor, and $n_1(t)$ and $n_2(t)$ are additive noise processes. The Fourier Series (FS) expansion of $x_i(t)$, i=1,2, over a finite observation time T is given by

$$x_i(t) = \sum_{k=-\infty}^{\infty} X_i(kw_o)\, e^{jkw_o t}, \quad i = 1,2 \qquad (2\text{-}2)$$

where

$$X_i(kw_o) = \frac{1}{T}\int_0^T x_i(t)\, e^{-jkw_o t}\, dt$$

$i = 1,2$, $w_o = \frac{2\pi}{T}$, and $X_i(kw_o)$ is the k-th FS coefficient.

In practice, the FS information is obtained via the discrete Fourier transform (DFT). The relation between the DFT and FS is given by

$$X_i(k) = \frac{1}{T} X_i(kw_o) \tag{2-3}$$

where

$$X_i(k) = \frac{1}{N} \sum_{m=0}^{N-1} x_i(m)\, W^{km} \text{ is the k-th DFT coefficient,}$$

$x_i(m)$ is the m-th sampled point obtained by sampling $x_i(t)$,

N is the number of sample point in the observation period T,

$k = 0, 1, \ldots, N-1$ is the k-th harmonic, and

$W = e^{-j2\pi/N}$, $j = \sqrt{-1}$.

From [2] it is known that if the observation time $T >> [|D| + R_{s_1 s_1}(\tau)]$, then

$$E\,[X_1(k)\, X_2^*(l)] \cong \begin{cases} \frac{1}{T} G_{x_1 x_2}(kw_o), & k = l \\ 0, & k \neq l \end{cases} \tag{2-4}$$

where $G_{x_1x_2}$ denotes the cross power spectral density (PDS). If $x_i(t)$ is Gaussian and zero-mean, then DFT coefficient $X_i(k)$ in (2-3) is also Gaussian, and has a zero mean; that is

$$E\,[X_i(k)] = \frac{1}{T} \int_o^T E\,[x_i(t)]\, e^{-jkw_ot}\, dt$$

$$= 0 \,.$$

We now define the vector

$$\underline{X}(k) = \begin{bmatrix} X_1(k) \\ X_2(k) \end{bmatrix} .$$

Then the covariance matrix Σ_x of $\underline{X}(k)$ is given by [3],

$$\Sigma_x = E\,[\underline{X}(k)\ \underline{X}^{*\prime}(k)]$$

$$= E\begin{bmatrix} X_1(k)\ X_1^*(k) & X_1(k)\ X_2^*(k) \\ X_2(k)\ X_1^*(k) & X_2(k)\ X_2^*(k) \end{bmatrix}$$

where the asterisk denotes complex conjugate, and the prime denotes transpose.

From Σ_x and (2-4) it follows that

$$\Sigma_x = \frac{1}{T}\begin{bmatrix} G_{x_1x_1}(kw_o) & G_{x_1x_2}(kw_o) \\ G^*_{x_1x_2}(kw_o) & G_{x_2x_2}(kw_o) \end{bmatrix} \tag{2-5a}$$

where $G_{x_1x_1}(kw_o)$ and $G_{x_2x_2}(kw_o)$ denote auto spectra, and $G_{x_1x_2}(kw_o)$ denotes the cross spectrum. Alternately, (2-5a) can be written as

$$\Sigma_x \equiv \frac{1}{T}\quad Q_x(kw_o) \tag{2-5b}$$

where $Q_x(kw_o)$ is the spectral matrix of $\underline{X}(k)$.

We know that if $\underline{X}'(k) = [X_1(k)\quad X_2(k)]$ where $X_1(k)$ and $X_2(k)$ are zero mean Gaussian random variables, then the joint density function of $\underline{X}(k)$ for a given attenuation α and delay D is given by

$$f(\underline{X}(k)/\alpha, D) = \frac{\left|\Sigma_x^{-1}\right|^{1/2}}{2\pi}\cdot \exp\left\{-\frac{\underline{X}'(k)\ \Sigma_x^{-1}\ \underline{X}(k)}{2}\right\}$$

$$= h\cdot \exp\left\{-\frac{T}{2}\ \underline{X}'(k)\ Q_x^{-1}(kw_o)\ \underline{X}(k)\right\}$$

where

$$h = \frac{\left|\Sigma_x^{-1}\right|^{1/2}}{2\pi} \quad .$$

We recall that if $\underline{X}(0)$, $\underline{X}(1)$, ..., $\underline{X}(N-1)$ are zero-mean Gaussian and mutually uncorrelated, then the above joint density function simplifies to yield

$$
\begin{aligned}
p_{\underline{X}/\alpha,d}(\underline{X}/_{\alpha,D}) &= \prod_{k=0}^{N-1} f\,(\underline{X}(k)/_{\alpha,D}) \\
&= h^N \exp\left\{-\frac{T}{2}\left[\sum_{k=0}^{N-1} \underline{X}^{*\prime}(k)\,[Q_x(kw_o)]^{-1}\,\underline{X}(k)\right]\right\} \\
&= h^N \exp\left\{-\frac{1}{2} J_1\right\} \qquad (2\text{-}6)
\end{aligned}
$$

where

$$\underline{X}' = [\underline{X}(0)\ \underline{X}(1)\ \ldots(\underline{X}(N-1)],\quad \underline{X}'(k) = [X_1(k)\ X_2(k)]$$

$$h = \frac{\left|\Sigma_x^{-1}\right|^{1/2}}{2\pi} = \frac{\left|[Q_x(kw_o)]^{-1}\ T\right|^{1/2}}{2\pi}$$

and

$$J_1 = T\sum_{k=0}^{N-1} \underline{X}^{*\prime}(k)\ [Q_x(kw_o)]^{-1}\ \underline{X}(k)$$

From (2-3), it follows that J_1 can be written as

$$J_1 = \sum_{k=0}^{N-1} \underline{X}^{*\prime}(kw_o)\ [Q_x(kw_o)]^{-1}\ \underline{X}(kw_o)\ \frac{1}{T} \qquad (2\text{-}7a)$$

For large T, (2-7a) yields

$$J_1 \simeq \int_0^{\infty} \underline{X}^{*\prime}(f)\ Q_x^{-1}(f)\ \underline{X}(f)\ df \qquad (2\text{-}7b)$$

Since the integrand of (2-7b) is an even function of f, we have

$$J_1 = \frac{1}{2} \int_{-\infty}^{\infty} \underline{X}^{*\prime}(f)\, Q_x^{-1}(f)\, \underline{X}(f)\, df \tag{2-8}$$

where $\underline{X}(f) = F\{\underline{x}(t)\}$, $\underline{x}'(t) = [x_1(t)\ x_2(t)]$ and F denotes the Fourier transform. Again, from (2-5) it follows that

$$Q_x^{-1}(kw_o) \simeq Q_x^{-1}(f) = \frac{\begin{bmatrix} G_{x_2x_2}(f) & -G_{x_1x_2}(f) \\ -G^*_{x_1x_2}(f) & G_{x_1x_1}(f) \end{bmatrix}}{G_{x_1x_1}(f)\, G_{x_2x_2}(f) - |G_{x_1x_2}(f)|^2}$$

or

$$Q_x^{-1}(f) = \frac{1}{1 - C_{12}(f)} \begin{bmatrix} \dfrac{1}{G_{x_1x_1}(f)} & \dfrac{-G_{x_1x_2}(f)}{G_{x_1x_1}(f)\, G_{x_2x_2}(f)} \\[2ex] \dfrac{-G^*_{x_1x_2}(f)}{G_{x_1x_1}(f)\, G_{x_2x_2}(f)} & \dfrac{1}{G_{x_2x_2}(f)} \end{bmatrix} \tag{2-9}$$

where $C_{12}(f) = \dfrac{|G_{x_1x_2}(f)|^2}{G_{x_1x_1}(f)\, G_{x_2x_2}(f)}$ is called the <u>magnitude squared coherence</u> (MSC) function†. Thus

$$\left| Q_x^{-1}(f) \right| = \frac{1}{1 - C_{12}(f)} \left[\frac{1}{G_{x_1x_1}(f)\, G_{x_2x_2}(f)} - \frac{|G_{x_1x_2}(f)|^2}{G^2_{x_1x_1}(f)\, G^2_{x_2x_2}(f)} \right] \tag{2-10}$$

From Section 1.2 we know that the ML estimate of D, denoted by $\hat{D}_{ml}$, is obtained by maximizing $p_{\underline{X}/\alpha,d}(\underline{X}/\alpha,D)$ in (2-6) with respect

†The reader may refer to the Appendix for a discussion of the MSC function.

to D; that is

$$\left. \frac{\partial \ln \, p_{\underline{X}/\alpha,d}\,(\underline{X}/\alpha,D)}{\partial D} \right|_{D = \hat{D}_{ml}} = 0 \tag{2-11}$$

Combining (2-6) and (2-11) we obtain

$$\frac{\partial}{\partial D}\,(\ln h^N) - \frac{1}{2}\,\frac{\partial}{\partial D}\,(J_1) = 0 \tag{2-12a}$$

From (2-6) we note that h in (2-12a) is independent of D. Thus

(2-12a) simplifies to yield

$$\frac{\partial}{\partial D}\,(J_1) = 0 \tag{2-12b}$$

Now, from (2-8) and (2-10) we have

$$J_1 \cong \int_{-\infty}^{\infty} \underline{X}^{*\prime}(f)\; Q_x^{-1}(f)\; \underline{X}(f)\; df$$

$$= \int_{-\infty}^{\infty} [X_1^*(f)\;\; X_2^*(f)] \left[\frac{1}{1 - C_{12}(f)}\right] \begin{bmatrix} \dfrac{1}{G_{x_1x_1}(f)} & \dfrac{-G_{x_1x_2}(f)}{G_{x_1x_1}(f)\; G_{x_2x_2}(f)} \\ \dfrac{-G^*_{x_1x_2}(f)}{G_{x_1x_1}(f)\; G_{x_2x_2}(f)} & \dfrac{1}{G_{x_1x_2}(f)} \end{bmatrix} \begin{bmatrix} X_1(f) \\ X_2(f) \end{bmatrix} df$$

which yields

$$J_1 = \int_{-\infty}^{\infty} \frac{1}{1 - C_{12}(f)} \left[\frac{|X_1(f)|^2}{G_{x_1x_1}(f)} + \frac{|X_2(f)|^2}{G_{x_2x_2}(f)} \right] df$$

$$- \int_{-\infty}^{\infty} \frac{1}{1 - C_{12}(f)} \left[\frac{X_1(f)\; X_2^*(f)\; G^*_{x_1x_2}(f)}{G_{x_1x_1}(f)\; G_{x_2x_2}(f)} + \frac{X_1^*(f)\; X_2(f)\; G_{x_1x_2}(f)}{G_{x_1x_1}(f)\; G_{x_2x_2}(f)} \right] df$$

That is

$$J_1 = J_2 - J_3 \tag{2-13}$$

where

$$J_2 = \int_{-\infty}^{\infty} \frac{1}{1 - C_{12}(f)} \left[\frac{|X_1(f)|^2}{G_{x_1x_1}(f)} + \frac{|X_2(f)|^2}{G_{x_2x_2}(f)} \right] df$$

$$J_3 = \int_{-\infty}^{\infty} \frac{1}{1 - C_{12}(f)} \left[\frac{X_1(f)\ X_2{*}(f)\ G^*_{x_1x_2}(f)}{G_{x_1x_1}(f)\ G_{x_2x_2}(f)} + \frac{X_1{*}(f)\ X_2(f)\ G_{x_1x_2}(f)}{G_{x_1x_1}(f)\ G_{x_2x_2}(f)} \right] df$$

Noting that $G_{x_1x_2}(f) = |G_{x_1x_2}(f)|\, e^{j2\pi fD}$, it follows that J_2 in (2-13) is independent of D. Thus, D can be estimated by maximizing only J_3. Again, J_3 can be written as

$$\begin{aligned} J_3 &= \int_{-\infty}^{\infty} [A(f) + A{*}(f)] \quad df \\ &= \int_{-\infty}^{\infty} [A(f) + A(-f)] \quad df \\ &= 2 \int_{-\infty}^{\infty} A(f)\ df \end{aligned} \qquad (2\text{-}14)$$

where

$$\begin{aligned} A(f) &= \frac{1}{1 - C_{12}(f)} \left[\frac{X_1(f)\ X_2{*}(f)\ G_{x_1x_2}(f)}{G_{x_1x_1}(f)\ G_{x_2x_2}(f)} \right] \\ &= \left[\frac{X_1(f)\ X_2{*}(f)}{|G_{x_1x_2}(f)|} \right] \left[\frac{C_{12}(f)}{1 - C_{12}(f)} \right] e^{j2\pi fD} \\ &= \left[\frac{X_1(f)\ X_2{*}(f)}{|G_{x_1x_2}(f)|} \right] \left[\frac{C_{12}(f)}{1 - C_{12}(f)} \right] e^{j2\pi fD} \end{aligned}$$

and

$$G^*_{x_1x_2}(f) = |G_{x_1x_2}(f)|\ e^{-j2\pi fD}$$

Also, $G_{x_1x_2}(f)$, $X_1(f)$ and $X_2(f)$ can only be estimated. Thus, if their estimates are denoted as $\hat{G}_{x_1x_2}(f)$, $\tilde{X}_1(f)$, $\tilde{X}_2(f)$ respectively, then we have

$$A(f) = \left[\frac{\tilde{X}_1(f)\,\tilde{X}_2^*(f)}{|G_{x_1x_2}(f)|}\right] \left[\frac{C_{12}(f)}{1 - C_{12}(f)}\right] e^{j2\pi fD}$$

By letting

$$\hat{G}_{x_1x_2}(f) = \frac{1}{T} \tilde{X}_1(f)\,\tilde{X}_2^*(f)$$

we obtain

$$J_3 = 2T \int_{-\infty}^{\infty} \hat{G}_{x_1x_2}(f) \left[\frac{1}{|G_{x_1x_2}(f)|}\right] \left[\frac{C_{12}(f)}{1 - C_{12}(f)}\right] e^{j2\pi fD} \, df$$

$$= 2T \int_{-\infty}^{\infty} \hat{G}_{x_1x_2}(f)\, W_{ML}(f)\, e^{j2\pi fD} \, df \qquad (2\text{-}15)$$

where

$$W_{ML} = \frac{C_{12}(f)}{|G_{x_1x_2}(f)|\,[1 - C_{12}(f)]}$$

is the ML weighting function.

Inspection of J_3 shows that it can now be expressed as

$$J_3 = 2T\, \hat{R}_{x_1x_2}^{ML}(D) \qquad (2\text{-}16a)$$

where

$$\hat{R}_{x_1x_2}^{ML}(D) = \int_{-\infty}^{\infty} \hat{G}_{x_1x_2}(f)\, W_{ML}(f)\, e^{j2\pi fD} \, df \qquad (2\text{-}16b)$$

is considered to be the ML crosscorrelation function corresponding to the weighting function W_{ML} in (2-15).

Now, from (2-16) it is clear that the maximum value of J_3 equals the maximum value of $\hat{R}_{x_1x_2}^{ML}(D)$. The particular value of

D, say $\hat{D}_{ML}$, where $\hat{R}_{x_1x_2}^{ML}(D)$ attains its maximum is the desired ML estimate of D, as illustrated in Fig. 2.1.

In summary, (2-16) states that the ML estimate of the delay D in (2-1) can be obtained as follows:

1) Estimate the cross spectrum $\hat{G}_{x_1x_2}(f)$ where $x_1(t)$ and $x_2(t)$ are defined in (2-1).
2) Weight $\hat{G}_{x_1x_2}(f)$ with the ML weighting function W_{ML} defined in (2-15).
3) Find the inverse Fourier transform of the product of $[W_{ML} \cdot \hat{G}_{x_1x_2}(f)]$ to obtain the correlation function $\hat{R}_{x_1x_2}^{ML}(D)$.
4) The ML estimate for D in (2-1) is given by $\hat{D}_{ML}$ which is the value of D for which $\hat{R}_{x_1x_2}(D)$ attains its peak value.

In the foregoing discussion we have the ML weighting function defined as

$$W_{ML}(f) = \frac{1}{|G_{x_1x_2}(f)|} \left[\frac{C_{12}(f)}{1 - C_{12}(f)} \right]$$

which means that a priori knowledge of the signal and noise statistics is known. In many problems this information is not available. For example, in passive detection, unlike the usual communication problem, the source spectrum is unknown, or known only approximately. If this is the case, they can only be estimated to obtain the "approximate ML (AML)" weighting function $\hat{W}_{ML}$ which may be substituted for the true weighting functions W_{ML}.

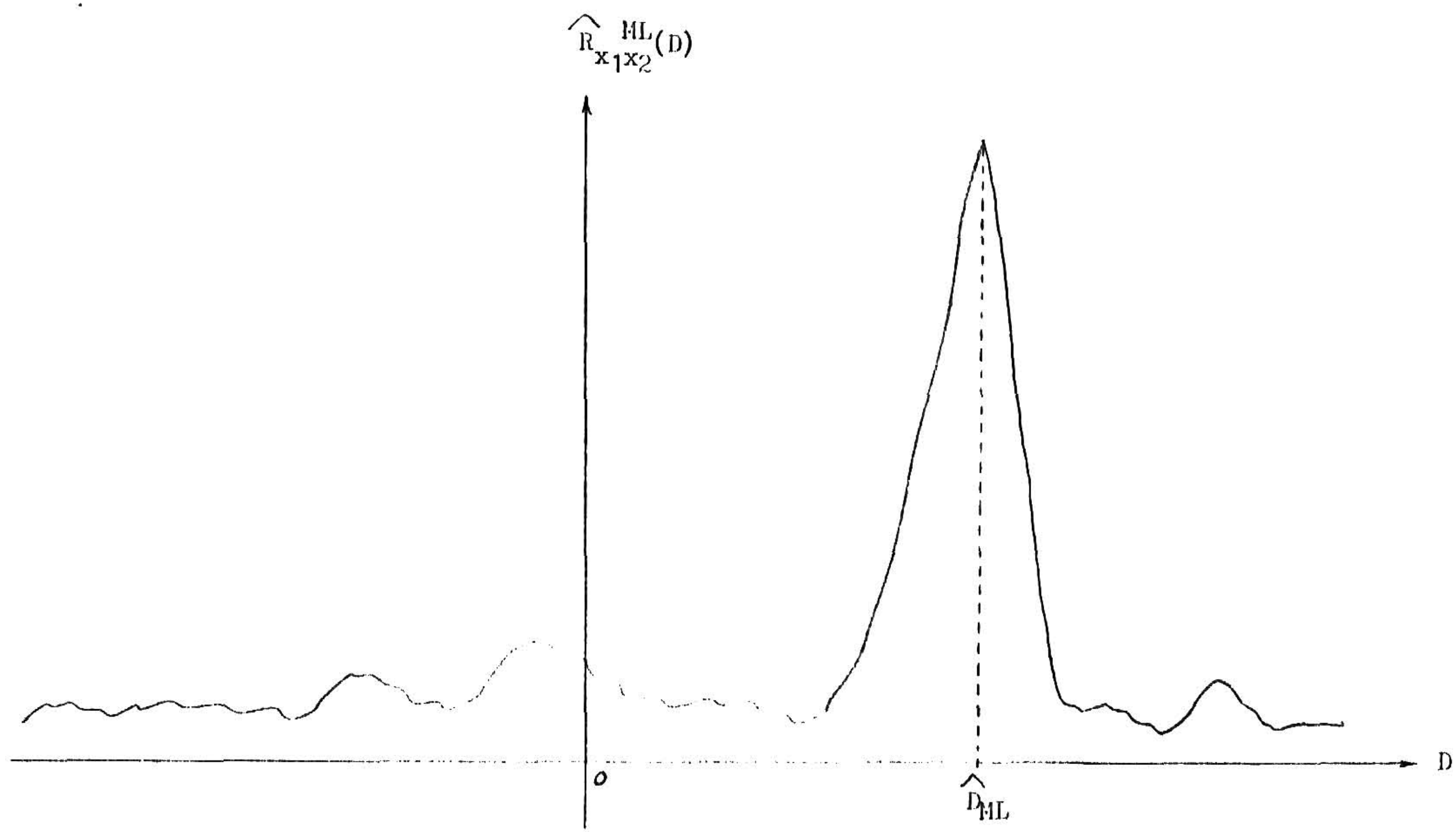


Fig. 2.1 The time corresponding to the maximum value of $\hat{R}_{x_1x_2}^{ML}(D)$ is the desired ML estimate of D.

III. VARIANCE OF ML ESTIMATOR

A lower bound on the variance for any unbiased estimator (which is not necessarily attainable) is given by the Cramér-Rao Lower Bound (CRLB), denoted by $\sigma_{\hat{D}}^2$. The CRLB is defined as [1]

$$\sigma_{\hat{D}}^2 \geq \frac{1}{E\left\{ \frac{\partial^2 \ln p_{\underline{X}/\alpha,d}(\underline{X}/\alpha,D)}{\partial D^2} \right\}} \quad (3\text{-}1)$$

It has been shown that ML estimation is unbiased [4]. Thus, (3-1) can be used for calculating the CRLB of an ML estimator.

We start from the joint probability of X for given α and D to find the variance $\sigma_{\hat{D}_{ML}}^2$ of the ML estimator. From (2-6) and (2-13) we have

$$\begin{aligned} p_{\underline{X}/\alpha,d}(\underline{X}/\alpha,D) &= h^n \exp\left[-1/2\ J_1\right] \\ &= h^n \exp\left[-1/2(J_2 - J_3)\right] , \end{aligned}$$

which yields

$$\ln p_{\underline{X}/\alpha,d}(\underline{X}/\alpha,D) = \ln h^n + \left[-1/2\ (J_2 - J_3)\right] . \quad (3\text{-}2)$$

Next, the derivatives in (3-2) yield

$$\begin{aligned} \frac{\partial^2 \ln p_{\underline{X}/\alpha,d}(\underline{X}/\alpha,D)}{\partial D^2} &= \frac{\partial^2}{\partial D^2}(h^n) + \frac{\partial^2}{\partial D^2}(-1/2\ J_2) + \frac{\partial}{\partial D}(1/2\ J_3) \\ &= \frac{\partial^2}{\partial D^2}(1/2\ J_3) \end{aligned}$$

since h and J_2 are independent of D.

Taking the expected value we obtain

$$E\left\{\frac{\partial^2 \ln p_{\underline{X}/\alpha,d}(\underline{X}/\alpha,D)}{\partial D^2}\right\} = 1/2 \; \frac{\partial^2}{\partial D^2}\left\{E[J_3]\right\}$$

$$= T\int_{-\infty}^{\infty} \frac{\partial^2}{\partial D^2} \; E\left[\hat{G}_{x_1x_2}(f)\, W_{ML}(f)\; e^{j2\pi fD}\right] \; df$$

$$= T\int_{-\infty}^{\infty} \frac{\partial^2}{\partial D^2} \; E\left[\hat{G}_{x_1x_2}(f)\right] W_{ML}(f)\; e^{j2\pi fD} \; df$$

$$= T\int_{-\infty}^{\infty} \frac{\partial^2}{\partial D^2} \; \left[G_{x_1x_2}(f)\, W_{ML}(f)\; e^{j2\pi fD}\right] \; df$$

where

$$E\left[\hat{G}_{x_1x_2}(f)\right] = G_{x_1x_1}(f).$$

Next, letting $G_{x_1x_2}(f) = |G_{x_1x_2}(f)|\, e^{j2\pi f\tau}$ we get

$$E\left\{\frac{\partial^2 \ln p_{\underline{X}/\alpha,d}(\underline{X}/\alpha,D)}{\partial D^2}\right\} = T\int_{-\infty}^{\infty} \frac{\partial^2}{\partial D^2}\left[\frac{G_{x_1x_2}(f)}{|G_{x_1x_2}(f)|}\;\frac{C_{12}(f)}{1 - C_{12}(f)}\; e^{j2\pi fD}\right] df$$

$$= T\int_{-\infty}^{\infty} (j2\pi f)^2 \;\frac{C_{12}(f)}{1 - C_{12}(f)}\; e^{j2\pi f(D+\tau)} \; df$$

$$= T\int_{-\infty}^{\infty} (2\pi f)^2 \;\frac{C_{12}(f)}{1 - C_{12}(f)}\; e^{j2\pi f(D+\tau)} \; df \qquad (3\text{-}3)$$

Since $0 < G_{12}(f) < 1$ and $(2\pi f)^2 \geq 0$, (3-3) attains its maximum value when $e^{j2\pi f(D+\tau)} = 1$; i.e. when $D = -\tau$, so that the minimum obtainable variance is

$$\left[T\int_{-\infty}^{-\infty} (2\pi f)^2 \;\frac{C_{12}(f)}{1 - C_{12}(f)}\; df\right]^{-1}$$

Thus, the CRLB for the ML estimate for D is given by (3-1) to be

$$\sigma^2_{\hat{D}_{ML}} \geq \left[T \int_{-\infty}^{\infty} (2\pi f)^2 \frac{C_{12}(f)}{1 - C_{12}(f)} \, df \right]^{-1} \tag{3-4}$$

ACKNOWLEDGEMENT

I would like to express my sincere appreciation to Dr. N. Ahmed and Mr. Kent Scarbrough for their patience and help throughout this report. Also I would like to thank Estella Creel for the wonderful typing.

REFERENCES

[1] H.L. Van Trees, Detection, Estimation, and Modulation Theory, Part I, Wiley, 1968.

[2] V.H. MacDonald and P.M. Schultheiss, "Optimum Passive Bearing Estimation", J. Acoust. Soc. Am., 46, pp. 37-43, 1969.

[3] Ahmed-Rao, Orthogonal Transforms for Digital Signal Processing, Springer-Verlag, New York, 1975.

[4] H. Cramer, Mathematical Methods of Statistics, Princeton University Press, Princeton, New Jersey, 1946.

[5] G.C. Carter, "Time Delay Estimation", University of Connecticut, Ph.D. Thesis, 1976.

[6] Hewlett-Packard, Application Note 245-2, "Measuring the Coherence Function with the HP3582A Spectrum Analyzer".

APPENDIX

Coherence Function

1.0 Introduction

In general, we use the crosscorrelation function to assess causality; that is, how much does the source influence the observed output. The coherence function also provides a measure of causality, but it has an additional advantage over the crosscorelation function. The crosscorrelation function is a function of time, and its maximum value corresponds to the approximate time delay (τ) between the source and the observed output. However, the coherence function is a function of frequency, and its maximum values occur at the frequencies where the greatest transfer of energy may be taking place. Techniques used to suppress interference (noise, vibration, etc.) depend on the frequency distribution of the interference. Hence the coherence function not only provides a measure of causality, but also provides ways of solving interference problems.

2.0 Coherence Function

The coherence function $\gamma_{x_1x_2}(f)$ of two jointly wide-sense stationary random processes $X_1(t)$ and $X_2(t)$, is a measure of the linear dependence between the processes. It is defined as

$$\gamma_{x_1x_2}(f) = \frac{G_{x_1x_2}(f)}{\sqrt{G_{x_1x_1}(f)\, G_{x_2x_2}(f)}} \qquad \text{(A-1)}$$

where $G_{x_1x_2}(f)$ is the cross power density spectrum between $X_1(t)$ and $X_2(t)$, $G_{x_1x_1}(f)$ and $G_{x_2x_2}(f)$ denote the auto power density spectra of $X_1(t)$ and $X_2(t)$, respectively.

3.0 Magnitude-Squared Coherence Function (MSC)

The MSC denoted as $C_{x_1x_2}(f)$, or more simply $C_{12}(f)$, is directly related to the coherence function, and is defined as

$$C_{x_1x_2}(f) \equiv \left| \gamma_{x_1x_2}(f) \right|^2 = \frac{\left|G_{x_1x_2}(f)\right|^2}{G_{x_1x_1}(f)\ G_{x_2x_2}(f)} \tag{A-2}$$

It is known that the power density spectrum matrix $Q_x(f)$ is positive semidefinite, i.e.,

$$Q_x(f) = \begin{vmatrix} G_{x_1x_1}(f) & G_{x_1x_2}(f) \\ G_{x_2x_1}(f) & G_{x_2x_2}(f) \end{vmatrix} \geqslant 0 \tag{A-3}$$

where $|\bullet|$ denotes the determinant of the matrix enclosed.

If the processes are real, then $G_{x_1x_2}(f) = G^*_{x_2x_1}(f)$. Thus (A-3) yeilds

$$G_{x_1x_1}(f)G_{x_2x_2}(f) - G_{x_1x_2}(f)^{\ 2} \geqslant 0 \tag{A-4}$$

or

$$G_{x_1x_1}(f)G_{x_2x_2}(f) \geqslant \left| G_{x_1x_2}(f) \right|^2 \tag{A-5}$$

Further, $G_{x_1x_1}(f)$ and $G_{x_2x_2}(f)$ are nonnegative and real function of f. Thus (A-5) can be divided through by $G_{x_1x_1}(f)\ G_{x_2x_2}(f)$ without changing the sense of the inequality. This division yields

$$1 \geqslant \frac{\left|G_{x_1x_2}(f)\right|^2}{G_{x_1x_1}(f)\ G_{x_2x_2}(f)} \geqslant 0 \quad , \quad \forall\ f$$

or

$$0 \leqslant C_{x_1x_2}(f) \leqslant 1.$$

Hence the MSC has the property that its value is always between 0 and 1. The boundary values correspond to the following cases:

(i) IF $X_1(t)$ and $X_2(t)$ are uncorrelated and zero mean wide-sense stationary processes, its crosscorrelation is zero, and hence $C_{x_1x_2}(f) = 0$.

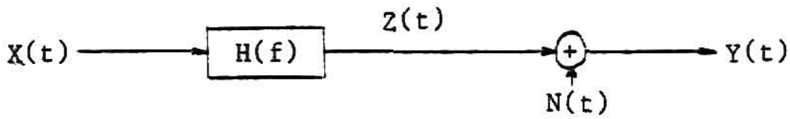
(ii) When $X_1(t)$ and $X_2(t)$ are linearly related and no noise contaminated, then

$\left|G_{x_1x_2}(f)\right|^2 = G_{x_1x_1}(f)\, G_{x_2x_2}(f)$ which means that

$C_{x_1x_2}(f) = 1.$

4.0 Signal-to-Noise Considerations

Consider the configuration shown below:



It follows that

$$Y_T(f) = X_T(f)\, H(f) + N_T(f)$$

where $Y_T(f)$, $X_T(f)$ and $N_T(f)$ are the Fourier transforms of sample function segments from random processes Y(t), X(t) and N(t); the length of each of these segements is 2T seconds. Then the auto power density spectrum of Y(t) is given by

$$G_{yy}(f) = \lim_{T\to\infty} \frac{E\left\{ \left|Y_T(f)\right|^2 \right\}}{2T}$$

$$= \lim_{T\to\infty} \frac{1}{2T}\, E\left\{ [H(f)\, X_T(f) + N_T(f)]\; [H(f)\, X_T(f) + N_T(f)]^* \right\}$$

or

$$G_{yy}(f) = |H(f)|^2\, G_{xx}(f) + G_{nn}(f) + H(f)\, G_{xn}(f) + H^*(f)\, G_{nx}(f) \qquad \text{(A-6)}$$

where $G_{xx}(f)$, $G_{nn}(f)$ and $G_{xn}(f)$, $G_{nx}(f)$ are auto and cross power density spectrums respectively.

Similarly,

$$G_{yx}(f) = \lim_{T\to\infty} \frac{E\left\{ Y_T(f)\ X_T^*(f) \right\}}{2T}$$

$$= \lim_{T\to\infty} \frac{E\left\{ [H(f)\ X_T(f) + N_T(f)] \cdot X_T^*(f) \right\}}{2T}$$

$$= H(f)\ G_{xx}(f) + G_{nx}(f) \qquad \text{(A-7)}$$

Next, if X(t) and N(t) are assumed to be zero mean uncorrelated processes, then (A-6) and (A-7) simplify to yield

$$G_{yy}(f) = |H(f)|^2\ G_{xx}(f) + G_{nn}(f) \qquad \text{(A-8)}$$

and

$$G_{yx}(f) = H(f)\ G_{xx}(f) \qquad \text{(A-9)}$$

Thus (A-2), (A-8), and (A-9) yield

$$C_{xy}(f) = \frac{|G_{yx}(f)|^2}{G_{xx}(f)\ G_{yy}(f)}$$

$$= \frac{|H(f)|^2\ G_{xx}^2(f)}{G_{xx}(f)\ [\,|H(f)|^2\ G_{xx}(f) + G_{nn}(f)]}$$

$$= \frac{|H(f)|^2\ G_{xx}(f)}{|H(f)|^2\ G_{xx}(f) + G_{nn}(f)} \qquad \text{(A-10)}$$

Thus the output signal-to-noise ratio (SNR) is given by

$$\text{SNR} = \frac{G_{zz}(f)}{G_{nn}(f)} = \frac{|H(f)|^2\ G_{xx}(f)}{G_{nn}(f)} = \frac{C_{xy}(f)}{1-C_{xy}(f)} \qquad \text{(A-11)}$$

Note that this is not the overall SNR, but the SNR at a specified value of frequency f. It states that the influence of noise decreases as $C_{xy}(f)$ approaches to 1.

We now summarize the following features of the MSC function [6]:

a) It is a dimensionless function in the frequency domain.

b) It takes values between 0 and 1.

c) At each frequency, it represents the fraction of the system output power to the total output power.

When the MSC is less than unity, at least one of the following conditions exists:

a) There is noise present in the measurements.

b) The system is nonlinear (i.e. energy is generated at additional frequencies).

c) Other inputs are present in the system.

ON MAXIMUM LIKELIHOOD ESTIMATION
AND ITS RELEVANCE TO TIME DELAY ESTIMATION

BY

JEN-WEI KUO

B.S., National Taiwan Institute of Technology, 1978

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Electrical Engineering

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1982

ABSTRACT

The purpose of this report is to study the basic concepts of maximum likelihood (ML) estimation, and its relevance to time delay estimation. The ML estimate is derived for the difference between the times of arrival of a signal due to a source and received at two sensors. A lower bound for the variance of this estimate is also derived. Sufficient details related to the derivations are included so that this report can be used to good advantage for pedagogical purposes.