How to Derive Bias and Mean Square Error for an Estimator?

Hing Cheung So

http://www.ee.cityu.edu.hk/~hcso

Department of Electronic Engineering
City University of Hong Kong

Outline

- Introduction
- Bias and Mean Square Error Formulas for Scalar
- Examples for Scalar Estimation
- Bias and Mean Square Error Formulas for Vector
- Examples for Vector Estimation
- List of References

Introduction

What is Estimation?

Estimation refers to accurately finding the values of parameters of interest from the observed data which consist of two components, viz., signal and noise.

A generic model for estimation of a scalar x is:

$$\mathbf{r} = \mathbf{f}(x) + \mathbf{w}$$

where $\mathbf{r}$ is the observation vector, signal $\mathbf{f}(x)$ is a known function of x and noise $\mathbf{w}$ is an additive random process.

The estimation problem is to find x given $\mathbf{r}$.

Why Estimation is Needed?

Many science and engineering problems can be boiled down to parameter estimation:

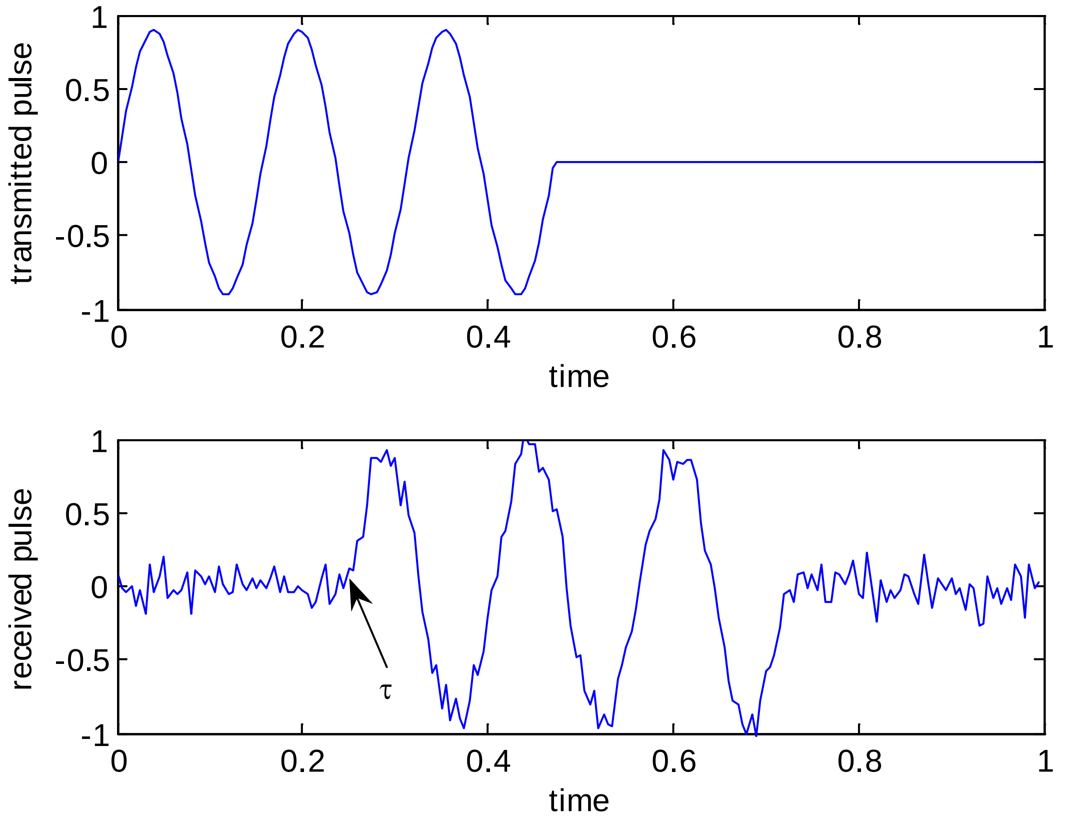
- Radar Ranging

Suppose a radar system transmits an electromagnetic pulse $s(t)$, which is then reflected by an object at a range of R, causing an echo to be received.

The received $r(t)$ is scaled, delayed and noisy version of $s(t)$:

$$r(t) = \alpha s(t-\tau) + w(t)$$

It is clear that the time delay $\tau > 0$ is round trip propagation time.

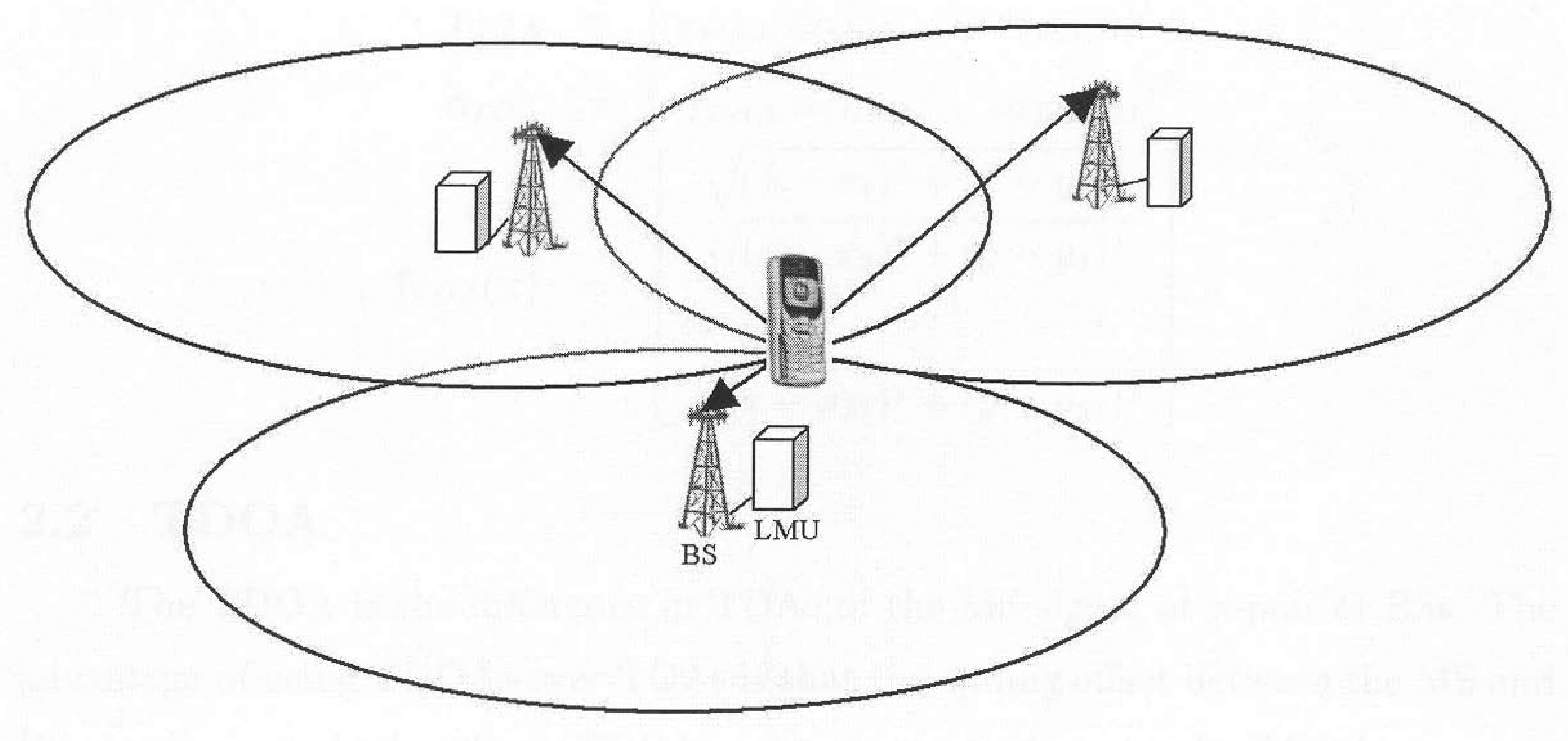


Via estimating τ, R can be obtained using the relationship:

$$\tau \cdot c = 2R$$

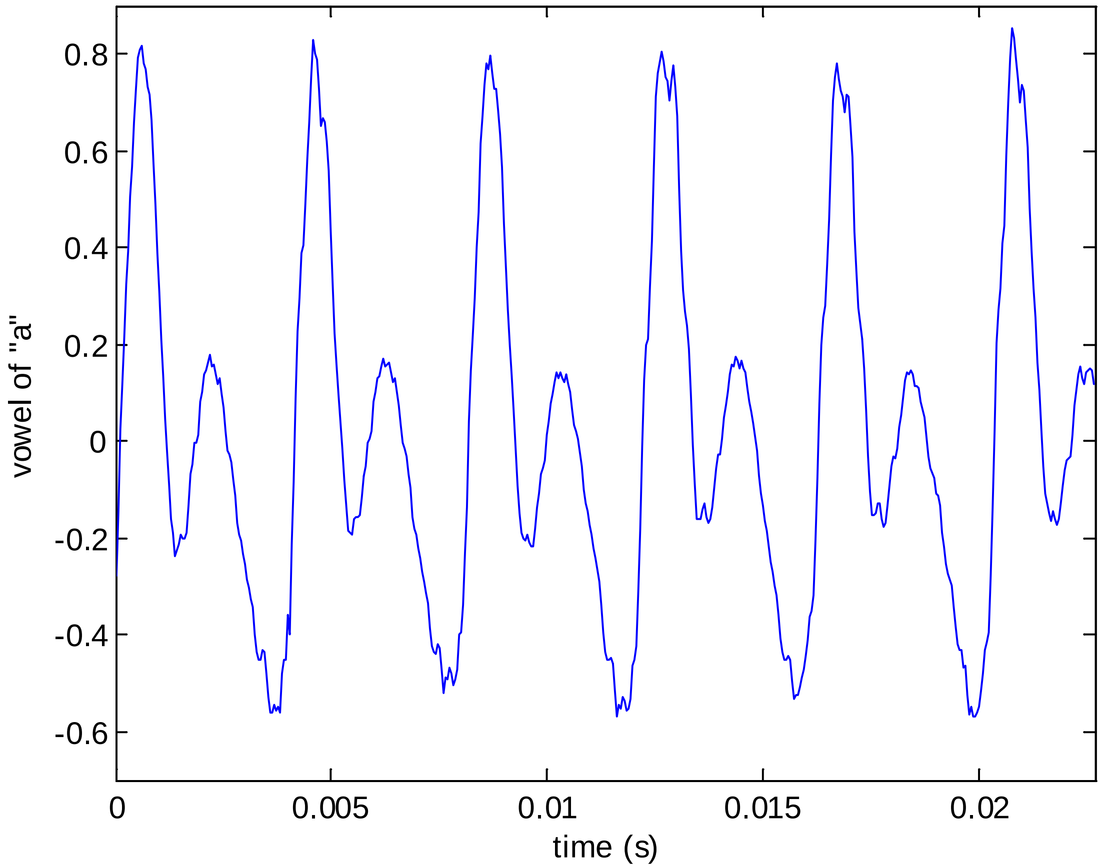
where c is the signal propagation speed.

➢ Mobile Positioning



If we know one-way propagation time of the signal traveling between mobile station and base station (BS), then the target position can be obtained using three BSs.

➢ Speech Analysis



For a voiced speech, it can be modeled as a periodic signal and it is important to estimate its pitch or fundamental frequency for analysis.

- **Image Processing**

Estimation of the position and orientation of an object from a camera image is useful when using a robot to pick it up, e.g., bomb-disposal.

- **Biomedical Engineering**

Estimation the heart rate of a fetus and the difficulty is that the measurements are corrupted by the mother's heart beat as well.

- **Seismology**

Estimation of the underground distance of an oil deposit based on sound reflection due to the different densities of oil and rock layers.

- **Astronomy**

Estimation of the periods of orbits.

How to Perform Estimation?

Least squares (LS) and maximum likelihood (ML) are two standard estimation approaches.

Consider the model of $\mathbf{r} = \mathbf{f}(x) + \mathbf{w}$.

The LS estimator does not require the probability density function (PDF) of $\mathbf{w}$, and its estimate is obtained by minimizing a sum of squared error:

$$\hat{x} = \arg\min_{\tilde{x}} (\mathbf{r} - \mathbf{f}(\tilde{x}))^T (\mathbf{r} - \mathbf{f}(\tilde{x})) = \arg\min_{\tilde{x}} \sum_{n=0}^{N-1} (r[n] - f_n(\tilde{x}))^2$$

where

$$\mathbf{r} = [r[0] \quad r[1] \quad \cdots \quad r[N-1]]^T$$
$$\mathbf{f}(\tilde{x}) = [f_0(\tilde{x}) \quad f_1(\tilde{x}) \quad \cdots \quad f_{N-1}(\tilde{x})]^T$$

To produce the ML estimator, the PDF of $\mathbf{w}$ is required.

Assuming that $\mathbf{w}$ is a zero-mean Gaussian noise, the PDF of the observed vector $\mathbf{r}$, which is parameterized by x, is

$$p(\mathbf{r};x) = \frac{1}{(2\pi)^{N/2}|\mathbf{C_w}|^{1/2}} e^{-\frac{1}{2}(\mathbf{r}-\mathbf{f}(x))^T \mathbf{C_w}^{-1}(\mathbf{r}-\mathbf{f}(x))}$$

where

$$\mathbf{C_w} = E\{(\mathbf{r}-\mathbf{f}(x))(\mathbf{r}-\mathbf{f}(x))^T\} = E\{\mathbf{w}\mathbf{w}^T\}$$

The ML estimate is:

$$\hat{x} = \arg\max_{\tilde{x}} p(\mathbf{r};\tilde{x})$$

When $\mathbf{w}$ is white with variance σ_w^2, the PDF reduces to

$$p(\mathbf{r};x) = \frac{1}{(2\pi\sigma_w^2)^{N/2}} e^{-\frac{1}{2\sigma_w^2}(\mathbf{r}-\mathbf{f}(x))^T(\mathbf{r}-\mathbf{f}(x))}$$

ML estimate is reduced to LS solution.

How to Assess Estimators?

Two standard performance measures for assessing accuracy of an estimator are bias and mean square error (MSE):

$$\mathrm{bias}(\hat{x}) = E\{\hat{x}\} - x$$

and

$$\mathrm{MSE}(\hat{x}) = E\{(\hat{x} - x)^2\}$$

It is desired that $\mathrm{bias}(\hat{x}) = 0$ or $E\{\hat{x}\} = x$, indicating that the estimator is unbiased, and MSE is as small as possible.

For an unbiased estimator, MSE is equal to variance:

$$\mathrm{var}(\hat{x}) = E\{(\hat{x} - E\{\hat{x}\})^2\} = E\{(\hat{x} - x)^2\} = \mathrm{MSE}(\hat{x})$$

In general:

$$\mathrm{MSE}(\hat{x}) = \mathrm{var}(\hat{x}) + (\mathrm{bias}(\hat{x}))^2$$

Consider a simple problem of estimating a DC level A from:

$$r[n] = A + w[n], \quad n = 0, 1, \cdots, N-1$$

where $w[n]$ has mean 0 and variance σ_w^2.

We easily suggest three estimators:

$$\hat{A}_1 = r[0]$$

$$\hat{A}_2 = \frac{1}{N}\sum_{n=0}^{N-1} r[n]$$

$$\hat{A}_3 = \frac{1}{N-1}\sum_{n=0}^{N-1} r[n]$$

It is easy to show:

$$E\{\hat{A}_1\} = A; \qquad \text{MSE}(\hat{A}_1) = \text{var}(\hat{A}_1) = E\{(r[0]-A)^2\} = E\{w^2[0]\} = \sigma_w^2$$

$$E\{\hat{A}_2\} = A; \qquad \text{MSE}(\hat{A}_2) = \text{var}(\hat{A}_2) = E\left\{\left(\frac{1}{N}\sum_{n=0}^{N-1} r[n] - A\right)^2\right\} = \frac{\sigma_w^2}{N}$$

$$E\{\hat{A}_3\} = \frac{1}{1-1/N}A; \qquad \text{MSE}(\hat{A}_3) = \left(\frac{A}{N-1}\right)^2 + \frac{\sigma_w^2}{N-1}$$

$\hat{A}_2$ is the best among the three because it has zero bias and minimum variance.

- Is $\hat{A}_2$ optimum?
- How to compute bias and MSE for more general cases?

Cramér-Rao Lower Bound (CRLB)

CRLB is performance bound in terms of minimum achievable variance provided by any unbiased estimators.

Its derivation requires knowledge of the noise PDF and the PDF must have closed-form.

Although there are other variance bounds, CRLB is simplest.

Suppose the PDF of $\mathbf{r} = \mathbf{f}(\mathbf{x}) + \mathbf{w}$ where $\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_L]^T$, is $p(\mathbf{r}; \mathbf{x})$.

The CRLB for $\mathbf{x}$ can be obtained in two steps:

- Compute the Fisher information matrix $\mathbf{I}(\mathbf{x})$.
- CRLB for x_l is the (l, l) entry of $\mathbf{I}^{-1}(\mathbf{x})$, $l = 1, 2, \cdots, L$.

$\mathbf{I}(\mathbf{x})$ has the form of:

$$\mathbf{I}(\mathbf{x}) = \begin{bmatrix} -E\left\{\frac{\partial^2 \ln p(\mathbf{r};\mathbf{x})}{\partial^2 x_1}\right\} & -E\left\{\frac{\partial^2 \ln p(\mathbf{r};\mathbf{x})}{\partial x_1 \partial x_2}\right\} & \cdots & -E\left\{\frac{\partial^2 \ln p(\mathbf{r};\mathbf{x})}{\partial x_1 \partial x_L}\right\} \\ -E\left\{\frac{\partial^2 \ln p(\mathbf{r};\mathbf{x})}{\partial x_2 \partial x_1}\right\} & -E\left\{\frac{\partial^2 \ln p(\mathbf{r};\mathbf{x})}{\partial^2 x_2}\right\} & \cdots & -E\left\{\frac{\partial^2 \ln p(\mathbf{r};\mathbf{x})}{\partial x_2 \partial x_L}\right\} \\ \vdots & \vdots & \vdots & \vdots \\ -E\left\{\frac{\partial^2 \ln p(\mathbf{r};\mathbf{x})}{\partial x_L \partial x_1}\right\} & \cdots & \cdots & -E\left\{\frac{\partial^2 \ln p(\mathbf{r};\mathbf{x})}{\partial^2 x_L}\right\} \end{bmatrix}$$

Consider $r[n] = A + w[n]$ with zero-mean white Gaussian noise:

$$p(\mathbf{r}; A) = \frac{1}{(2\pi\sigma_w^2)^{N/2}} e^{-\frac{1}{2\sigma_w^2}\sum_{n=0}^{N-1}(r[n]-A)^2}$$

$$\Rightarrow \ln p(\mathbf{r}; A) = -\ln((2\pi\sigma_w^2)^{N/2}) - \frac{1}{2\sigma_w^2}\sum_{n=0}^{N-1}(r[n] - A)^2$$

$$\Rightarrow \frac{\partial^2 \ln p(\mathbf{r}; A)}{\partial^2 A} = -\frac{N}{\sigma_w^2}$$

$$\Rightarrow \mathbf{I}(A) = -E\left\{\frac{\partial^2 \ln p(\mathbf{r}; A)}{\partial^2 A}\right\} = \frac{N}{\sigma_w^2}$$

$$\Rightarrow \mathbf{I}^{-1}(A) = \frac{\sigma_w^2}{N}$$

That is, $\hat{A}_2 = \frac{1}{N}\sum_{n=0}^{N-1} r[n]$ is the optimum estimator for A.

Bias and Mean Square Error Formulas for Scalar

Recall the signal model:

$$\mathbf{r} = \mathbf{f}(x) + \mathbf{w}$$

Suppose the scalar x is estimated by minimizing a differentiable cost function constructed from $\mathbf{r}$, $J(\tilde{x})$:

$$\hat{x} = \arg\min_{\tilde{x}} J(\tilde{x})$$

This implies

$$\left.\frac{dJ(\tilde{x})}{d\tilde{x}}\right|_{\tilde{x}=\hat{x}} = J'(\hat{x}) = 0$$

At small estimation error conditions, $\hat{x}$ is close to x. Applying Tayler series expansion yields:

$$J'(\hat{x}) \approx J'(x) + (\hat{x} - x)J''(x)$$

If $J''(\tilde{x})$ is sufficiently smooth around $\tilde{x} = x$, then

$$J''(x) \approx E\{J''(x)\}$$

Hence

$$0 = J'(\hat{x}) \approx J'(x) + (\hat{x} - x)E\{J''(x)\} \Rightarrow \hat{x} - x \approx -\frac{J'(x)}{E\{J''(x)\}}$$

$$\Rightarrow \text{bias}(\hat{x}) = E\{\hat{x}\} - x \approx -\frac{E\{J'(x)\}}{E\{J''(x)\}}$$

Similarly,

$$\text{MSE}(\hat{x}) = E\{(\hat{x} - x)^2\} \approx \frac{E\{(J'(x))^2\}}{(E\{J''(x)\})^2}$$

Note:

When $J(\tilde{x})$ is a quadratic function:

$$\text{bias}(\hat{x}) = -\frac{E\{J'(x)\}}{E\{J''(x)\}} \quad \text{and} \quad \text{MSE}(\hat{x}) = \frac{E\{(J'(x))^2\}}{(E\{J''(x)\})^2}$$

When $E\{J'(x)\} = 0$, $\hat{x}$ is an unbiased estimate of x.

For unbiased estimator:

$$\text{var}(\hat{x}) = \text{MSE}(\hat{x}) \approx \frac{E\{(J'(x))^2\}}{(E\{J''(x)\})^2}$$

Examples for Scalar Estimation

For simplicity, we assume that the noise is white Gaussian process with variance σ_w^2.

DC Level Estimation

Recall the model:

$$r[n] = A + w[n], \quad n = 0, 1, \cdots, N-1$$

Using the LS approach, the cost function to be minimized is

$$J(\tilde{A}) = \sum_{n=0}^{N-1} \left(r[n] - \tilde{A} \right)^2 \Rightarrow \hat{A} = \hat{A}_2 = \frac{1}{N} \sum_{n=0}^{N-1} r[n]$$

To apply the bias and MSE formulas, we compute:

$$J'(A) = -2\sum_{n=0}^{N-1}[r[n]-A] = -2\sum_{n=0}^{N-1} w[n] \Rightarrow E\{J'(A)\} = 0$$

$$(J'(A))^2 = 4\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} w[n]w[m] \Rightarrow E\{(J'(A))^2\} = 4E\left\{\sum_{n=0}^{N-1} w^2[n]\right\} = 4N\sigma_w^2$$

and

$$J''(A) = -2\sum_{n=0}^{N-1}(-1) = 2N = E\{J''(A)\}$$

Hence:

$$\mathrm{bias}(\hat{A}) = -\frac{E\{J'(A)\}}{E\{J''(A)\}} = 0 \quad \text{and} \quad \mathrm{var}(\hat{A}) = \frac{E\{(J'(A))^2\}}{(E\{J''(A)\})^2} = \frac{\sigma_w^2}{N}$$

which align with previous analysis.

Time-Difference-of-Arrival Estimation

The simplest model is to estimate the time-difference-of-arrival D between two signals:

$$r_1[n] = s[n] + w_1[n], \quad r_2[n] = s[n-D] + w_2[n], \quad n = 0, 1, \cdots, N-1$$

where $s[n]$, $w_1[n]$ and $w_2[n]$ are independent zero-mean white Gaussian variables with $E\{s^2[n]\} = \sigma_s^2$, $E\{w_1^2[n]\} = E\{w_2^2[n]\} = \sigma_w^2$.

It is clear that $r_1[n-D] = s[n-D] + w_1[n-D]$ is most similar to $r_2[n]$. As a result, D can be estimated by maximizing the cross-correlation between $r_1[n]$ and $r_2[n]$:

$$\hat{D} = \arg\max_{\tilde{D}} J(\tilde{D}), \quad J(\tilde{D}) = \sum_{n=0}^{N-1} r_1[n-\tilde{D}] r_2[n]$$

However, D is generally not an integer and thus $J(\tilde{D})$ is a continuous function of $\tilde{D}$.

Using the convolution theorem, $r_1[n-\tilde{D}]$ has the form of

$$r_1[n-\tilde{D}] = \sum_{k=-\infty}^{\infty} r_1[n-k]\mathrm{sinc}(k-\tilde{D}) = \sum_{k=-P}^{P} r_1[n-k]\mathrm{sinc}(k-\tilde{D})$$

Applying the bias and MSE formulas, we obtain:

$$E\{\hat{D}\} \approx D$$

and

$$\mathrm{var}(\hat{D}) \approx \frac{3\sigma_w^2\left(\sigma_w^2 + 2\sigma_s^2\right)}{\pi^2 N \sigma_s^4}$$

which is also the CRLB.

Frequency Estimation of a Complex Sinusoid

The signal model is:

$$x[n] = \alpha e^{j(\omega n + \phi)} + w[n], \quad n = 0, 1, \cdots, N-1$$

where $\alpha > 0$, $\omega \in (-\pi, \pi)$ and $\phi \in [0, 2\pi)$.

A conventional approach for estimating ω is to search the periodogram peak:

$$\hat{\omega} = \arg\max_{\tilde{\omega}} J(\tilde{\omega}), \quad J(\tilde{\omega}) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n] e^{-j\tilde{\omega}n}\right|^2$$

Applying the bias and MSE formulas, we obtain:

$$E\{\hat{\omega}\} = \omega \qquad \text{and} \qquad \mathrm{var}(\hat{\omega}) = \frac{6\sigma_w^2}{\alpha^2 N(N^2-1)}$$

which is also the CRLB.

Frequency Estimation of a Real Sinusoid

The signal model is:

$$x[n] = \alpha \cos(\omega n + \phi) + w[n], \quad n = 0, 1, \cdots, N-1$$

where $\alpha > 0$, $\omega \in (0, \pi)$ and $\phi \in [0, 2\pi)$

According to the linear prediction property:

$$\cos(\omega n + \phi) = \rho \cos(\omega(n-1) + \phi) - \cos(\omega(n-2) + \phi), \quad \rho = 2\cos(\omega)$$

A LS cost function for estimating ρ is then:

$$J(\tilde{\rho}) = \sum_{n=2}^{N-1} \left(x[n] + x[n-2] - \tilde{\rho} x[n-1]\right)^2$$

The LS estimate of ρ is:

$$\hat{\rho} = \frac{\displaystyle\sum_{n=2}^{N-1} x[n-1](x[n] + x[n-2])}{\displaystyle\sum_{n=2}^{N-1} x^2[n-1]}$$

Hence the frequency estimate is

$$\hat{\omega} = \cos^{-1}\left(\frac{\hat{\rho}}{2}\right)$$

which is known as the modified covariance method.

Applying the bias and MSE formulas, we obtain:

$$E\{J'(\rho)\} = 2\rho(N-2)\sigma_w^2$$

and

$$J''(\rho) = 2\sum_{n=2}^{N-1} x^2[n-1] \Rightarrow E\{J''(\rho)\} \approx (N-2)\alpha^2 + 2(N-2)\sigma_w^2$$

if N is sufficiently large.

Hence

$$\text{bias}(\hat{\rho}) = -\frac{E\{J'(\rho)\}}{E\{J''(\rho)\}} \approx -\frac{\rho}{\text{SNR}+1},$$

where $\text{SNR} = \alpha^2/(2\sigma_w^2)$.

With tedious calculation, we have

$$\mathrm{MSE}(\hat{\rho}) \approx \frac{4(N(N-2) - 4\mathrm{SNR}(N-3))\cos^2(\omega) + 2(2N-5)(2\mathrm{SNR}+1) + 8\mathrm{SNR}(N-3)\cos(2\omega)}{(\mathrm{SNR}+1)^2(N-2)^2}$$

Since

$$\hat{\omega} = \cos^{-1}\left(\frac{\hat{\rho}}{2}\right) \Rightarrow \cos(\hat{\omega}) = \frac{\hat{\rho}}{2}$$

$$\Rightarrow \cos(\hat{\omega}) - \cos(\omega) = -2\sin\left(\frac{\hat{\omega}+\omega}{2}\right)\sin\left(\frac{\hat{\omega}-\omega}{2}\right) = \frac{\hat{\rho}-\rho}{2}$$

$$-2\sin(\omega)\left(\frac{\hat{\omega}-\omega}{2}\right) \approx \frac{\hat{\rho}-\rho}{2} \Rightarrow 4\sin^2(\omega)\mathrm{MSE}(\hat{\omega}) \approx \mathrm{MSE}(\hat{\rho})$$

We have:

$$\mathrm{MSE}(\hat{\omega}) \approx \frac{2(N(N-2) - 4\mathrm{SNR}(N-3))\cos^2(\omega) + (2N-5)(2\mathrm{SNR}+1) + 4\mathrm{SNR}(N-3)\cos(2\omega)}{2(\mathrm{SNR}+1)^2(N-2)^2\sin^2(\omega)}$$

Bias and Mean Square Error Formulas for Vector

For estimation of a vector $\mathbf{x}$ from minimizing $J(\tilde{\mathbf{x}})$, the formulas are generalized as follows:

$$\mathbf{0} = \left.\frac{\partial J(\tilde{\mathbf{x}})}{\partial \tilde{\mathbf{x}}}\right|_{\tilde{\mathbf{x}}=\hat{\mathbf{x}}} \approx \left.\frac{\partial J(\tilde{\mathbf{x}})}{\partial \tilde{\mathbf{x}}}\right|_{\tilde{\mathbf{x}}=\mathbf{x}} + \left.\frac{\partial^2 J(\tilde{\mathbf{x}})}{\partial \tilde{\mathbf{x}} \partial \tilde{\mathbf{x}}^T}\right|_{\tilde{\mathbf{x}}=\mathbf{x}} (\hat{\mathbf{x}} - \mathbf{x})$$

$$\Rightarrow -\boldsymbol{\nabla}(J(\mathbf{x})) \approx \mathbf{H}(J(\mathbf{x}))\,(\hat{\mathbf{x}} - \mathbf{x})$$

and

$$\mathbf{H}(J(\mathbf{x})) \approx E\{\mathbf{H}(J(\mathbf{x}))\}$$

where $\boldsymbol{\nabla}(J(\mathbf{x}))$ is the gradient vector and $\mathbf{H}(J(\mathbf{x}))$ is the Hessian matrix:

$$\boldsymbol{\nabla}(J(\mathbf{x})) = \left[\frac{\partial J(\mathbf{x})}{\partial x_1} \quad \frac{\partial J(\mathbf{x})}{\partial x_2} \quad \cdots \quad \frac{\partial J(\mathbf{x})}{\partial x_L}\right]^T$$

$$\mathbf{H}(J(\mathbf{x})) = \begin{bmatrix} \frac{\partial^2 J(\mathbf{x})}{\partial^2 x_1} & \frac{\partial^2 J(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 J(\mathbf{x})}{\partial x_1 \partial x_L} \\ \frac{\partial^2 J(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 J(\mathbf{x})}{\partial^2 x_2} & \cdots & \frac{\partial^2 J(\mathbf{x})}{\partial x_2 \partial x_L} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 J(\mathbf{x})}{\partial x_L \partial x_1} & \frac{\partial^2 J(\mathbf{x})}{\partial x_L \partial x_2} & \cdots & \frac{\partial^2 J(\mathbf{x})}{\partial^2 x_L} \end{bmatrix}$$

As a result,

$$\mathrm{bias}(\hat{\mathbf{x}}) = E\{\hat{\mathbf{x}}\} - \mathbf{x} \approx -[E\{\mathbf{H}(J(\mathbf{x}))\}]^{-1} E\{\boldsymbol{\nabla}(J(\mathbf{x}))\}$$

Similarly, the covariance matrix is:

$$\mathbf{C}_{\hat{\mathbf{x}}} \approx [E\{\mathbf{H}(J(\mathbf{x}))\}]^{-1} E\{\boldsymbol{\nabla}(J(\mathbf{x}))\boldsymbol{\nabla}^T(J(\mathbf{x}))\}[E\{\mathbf{H}(J(\mathbf{x}))\}]^{-1}$$

The MSE of $\hat{x}_l$ is given by (l,l) entry of $\mathbf{C}_{\hat{\mathbf{x}}}$

Examples for Vector Estimation

Estimation of a Linear Model

The linear data model is:

$$\mathbf{b} = \mathbf{A}\boldsymbol{\theta} + \mathbf{w}$$

where $\mathbf{A}$ is known, $\boldsymbol{\theta}$ is unknown vector, and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_\mathbf{w})$.

Employing $\mathbf{C}_\mathbf{w}^{-1}$, the weighted LS cost function is:

$$J(\tilde{\boldsymbol{\theta}}) = \left(\mathbf{A}\tilde{\boldsymbol{\theta}} - \mathbf{b}\right)^T \mathbf{C}_\mathbf{w}^{-1} \left(\mathbf{A}\tilde{\boldsymbol{\theta}} - \mathbf{b}\right) \Rightarrow \hat{\boldsymbol{\theta}} = \left(\mathbf{A}^T\mathbf{C}_\mathbf{w}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{C}_\mathbf{w}^{-1}\mathbf{b}$$

Applying the bias and MSE formulas, we obtain:

$$E\{\hat{\boldsymbol{\theta}}\} = \boldsymbol{\theta} \qquad \text{and} \qquad \mathbf{C}_{\hat{\boldsymbol{\theta}}} = \left(\mathbf{A}^T\mathbf{C}_\mathbf{w}^{-1}\mathbf{A}\right)^{-1}$$

These align with the best linear unbiased estimator (BLUE).

Parameter Estimation of a Real Sinusoid

The signal model is:

$$x[n] = \alpha \cos(\omega n + \phi) + w[n], \quad n = 0, 1, \cdots, N-1$$

where $\alpha > 0$, $\omega \in (0, \pi)$ and $\phi \in [0, 2\pi)$, while $w[n]$ is a white Gaussian process with variance σ_w^2.

According to ML or LS, we construct:

$$J(\tilde{\boldsymbol{\theta}}) = \sum_{n=0}^{N-1} \left(x[n] - \tilde{\alpha} \cos(\tilde{\omega} n + \tilde{\phi}) \right)^2, \quad \tilde{\boldsymbol{\theta}} = \begin{bmatrix} \tilde{\alpha} & \tilde{\omega} & \tilde{\phi} \end{bmatrix}^T$$

Applying the bias and MSE formulas, we obtain:

$$E\{\hat{\boldsymbol{\theta}}\} = [\alpha \;\; \omega \;\; \phi]^T$$

$$\mathbf{C}_{\hat{\mathbf{x}}} \approx$$

$$\sigma_w^2 \begin{bmatrix} \frac{N}{2} & \frac{\alpha}{2}\sum_{n=0}^{N-1} n\sin(2\omega n + 2\phi) & \frac{\alpha}{2}\sum_{n=0}^{N-1} \sin(2\omega n + 2\phi) \\ \frac{\alpha}{2}\sum_{n=0}^{N-1} n\sin(2\omega n + 2\phi) & \alpha^2 \sum_{n=0}^{N-1} n^2 \left(\frac{1}{2} - \frac{1}{2}\cos(2\omega n + 2\phi)\right) & \alpha^2 \sum_{n=0}^{N-1} n\sin^2(\omega n + \phi) \\ \frac{\alpha}{2}\sum_{n=0}^{N-1} \sin(2\omega n + 2\phi) & \alpha^2 \sum_{n=0}^{N-1} n\sin^2(\omega n + \phi) & \alpha^2 \sum_{n=0}^{N-1} \sin^2(\omega n + \phi) \end{bmatrix}^{-1}$$

which is the inverse of the Fisher information matrix.

That is, the estimator provides the optimum performance.

Localization using Range Measurements

Consider positioning of a source at $\mathbf{x} = [x\ y]^T$ by $N \geq 3$ sensors at known coordinates $\mathbf{x}_n = [x_n\ y_n]^T$, $n = 1, 2, \ldots, N$.

If we have the one-way propagation time measurements, they can be easily converted to ranges:

$$r_n = d_n + w_n, \quad n = 1, 2, \ldots, N$$

where $d_n = \sqrt{(x - x_n)^2 + (y - y_n)^2}$ and $w_n \sim \mathcal{N}(0, \sigma_w^2)$ is white.

The ML or LS cost function is

$$J(\tilde{\mathbf{x}}) = \sum_{n=1}^{N} \left(r_n - \sqrt{(\tilde{x} - x_n)^2 + (\tilde{y} - y_n)^2} \right)^2$$

To determine the bias and MSE, the steps include:

$$\nabla(J(\mathbf{x})) = -\frac{2}{\sigma_w^2}\begin{bmatrix} \sum_{n=1}^{N} \frac{(r_n - d_n)(x - x_n)}{d_n} \\ \sum_{n=1}^{N} \frac{(r_n - d_n)(y - y_n)}{d_n} \end{bmatrix} \Rightarrow E\{\nabla(J(\mathbf{x}))\} = \mathbf{0}$$

because $E\{r_n\} = d_n$.

Similarly,

$$E\{\mathbf{H}(J(\mathbf{x}))\} = \frac{2}{\sigma_w^2}\begin{bmatrix} \sum_{n=1}^{N} \frac{(x - x_n)^2}{d_n^2} & \sum_{n=1}^{N} \frac{(x - x_n)(y - y_n)}{d_n^2} \\ \sum_{n=1}^{N} \frac{(x - x_n)(y - y_n)}{d_n^2} & \sum_{n=1}^{N} \frac{(y - y_n)^2}{d_n^2} \end{bmatrix}$$

As a result,

$$\text{bias}(\hat{\mathbf{x}}) \approx -[E\{\mathbf{H}(J(\mathbf{x}))\}]^{-1}E\{\boldsymbol{\nabla}(J(\mathbf{x}))\} = \mathbf{0}$$

With tedious calculation, we have

$$\mathbf{C}_{\hat{\mathbf{x}}} \approx \sigma_w^2 \begin{bmatrix} \sum_{n=1}^{N} \frac{(x-x_n)^2}{d_n^2} & \sum_{n=1}^{N} \frac{(x-x_n)(y-y_n)}{d_n^2} \\ \sum_{n=1}^{N} \frac{(x-x_n)(y-y_n)}{d_n^2} & \sum_{n=1}^{N} \frac{(y-y_n)^2}{d_n^2} \end{bmatrix}^{-1}$$

which is the inverse of the Fisher information matrix.

That is, the estimator provides the optimum performance.

Apart from the nonlinear approach, r_n can be linearized:

$$
\begin{aligned}
& r_n = \sqrt{(x - x_n)^2 + (y - y_n)^2} + w_n \\
& \Rightarrow r_n^2 = (x - x_n)^2 + (y - y_n)^2 + w_n^2 + 2w_n \sqrt{(x - x_n)^2 + (y - y_n)^2} \\
& \Rightarrow -2x_n x - 2y_n y + R + q_n = r_n^2 - x_n^2 - y_n^2
\end{aligned}
$$

where

$$
\begin{aligned}
R &= x^2 + y^2 \\
q_n &= w_n^2 + 2w_n d_n
\end{aligned}
$$

Hence the signal model is now linear:

$$
\mathbf{b} = \mathbf{A}\boldsymbol{\theta} + \mathbf{q}
$$

where

$$\mathbf{b} = \left[r_1^2 - x_1^2 - y_1^2 \quad r_2 - x_2^2 - y_2^2 \quad \cdots \quad r_N^2 - x_N^2 - y_N^2\right]^T$$

$$\mathbf{A} = \begin{bmatrix} -2x_1 & -2y_1 & 1 \\ -2x_2 & -2y_2 & 1 \\ \vdots & \vdots & \vdots \\ -2x_N & -2y_N & 1 \end{bmatrix}$$

$$\boldsymbol{\theta} = [x \quad y \quad R]^T$$

$$\mathbf{q} = [q_1 \quad q_2 \quad \cdots \quad q_N]^T$$

For sufficiently small noise conditions, we have

$$\begin{aligned} &\mathbf{q} \approx 2\left[w_1 d_1 \quad w_2 d_2 \quad \cdots \quad w_N d_N\right]^T \\ &\Rightarrow \mathbf{C}_{\mathbf{q}} \approx 4\sigma_q^2 \mathrm{diag}([d_1^2 \quad d_2^2 \quad \cdots \quad d_N^2]) \\ &\Rightarrow \mathbf{C}_{\mathbf{q}} \approx 4\sigma_q^2 \mathrm{diag}([r_1^2 \quad r_2^2 \quad \cdots \quad r_N^2]) \end{aligned}$$

The weighted LS cost function to be minimized is

$$J(\tilde{\boldsymbol{\theta}}) = \left(\mathbf{A}\tilde{\boldsymbol{\theta}} - \mathbf{b}\right)^T \mathbf{C}_{\mathbf{q}}^{-1} \left(\mathbf{A}\tilde{\boldsymbol{\theta}} - \mathbf{b}\right)$$

and the estimate is

$$\hat{\boldsymbol{\theta}} = \min_{\tilde{\boldsymbol{\theta}}} J(\tilde{\boldsymbol{\theta}}) = \left(\mathbf{A}^T \mathbf{C}_{\mathbf{q}}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^T \mathbf{C}_{\mathbf{q}}^{-1}$$

Applying the bias and MSE formulas, we obtain:

$$E\{\hat{\boldsymbol{\theta}}\} \approx [x \ y \ R]^T$$

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} \approx \left(\mathbf{A}^T \mathbf{C}_{\mathbf{q}}^{-1} \mathbf{A}\right)^{-1}$$

MSEs of $\hat{x}$ and $\hat{y}$ are given by $(1,1)$ and $(2,2)$ entries of $\mathbf{C}_{\hat{\boldsymbol{\theta}}}$.

To achieve higher accuracy, the information of $\hat{R}$ should be utilized, which results in a constrained optimization problem:

$$\min_{\tilde{\boldsymbol{\theta}}} J(\tilde{\boldsymbol{\theta}}) \quad \text{subject} \quad \text{to} \quad \tilde{R} = \tilde{x}^2 + \tilde{y}^2$$

The solution can be derived using the method of Lagrange multipliers.

To analyze the performance, the constrained problem can be converted to an unconstrained one by putting the relation of $\tilde{R} = \tilde{x}^2 + \tilde{y}^2$ into $J(\tilde{\boldsymbol{\theta}})$:

$$J(\tilde{\mathbf{x}}) = \sum_{n=1}^{N} \frac{1}{d_n^2} \left(-2x_l\tilde{x} - 2y_l\tilde{y} + \tilde{x}^2 + \tilde{y}^2 - r_l^2 + \tilde{x}^2 + \tilde{y}^2\right)^2$$

Applying the bias and MSE formulas, we obtain:

$$\text{bias}(\hat{\mathbf{x}}) \approx -[E\{\mathbf{H}(J(\mathbf{x}))\}]^{-1} E\{\boldsymbol{\nabla}(J(\mathbf{x}))\} = \mathbf{0}$$

$$\mathbf{C}_{\hat{\mathbf{x}}} \approx \sigma_w^2 \begin{bmatrix} \displaystyle\sum_{n=1}^{N} \frac{(x-x_n)^2}{d_n^2} & \displaystyle\sum_{n=1}^{N} \frac{(x-x_n)(y-y_n)}{d_n^2} \\ \displaystyle\sum_{n=1}^{N} \frac{(x-x_n)(y-y_n)}{d_n^2} & \displaystyle\sum_{n=1}^{N} \frac{(y-y_n)^2}{d_n^2} \end{bmatrix}^{-1}$$

which is the inverse of the Fisher information matrix.

That is, the estimator also provides optimum performance.

List of References

[1] H.C. So, Y.T. Chan, K.C. Ho and Y. Chen, "Simple formulas for bias and mean square error computation," *IEEE Signal Processing Magazine*, vol.30, no.4, pp.162-165, Jul. 2013.

[2] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, NJ: Englewood Cliffs, 1993.

[3] V.H. MacDonald and P.M. Schultheiss, "Optimum passive bearing estimation in a spatially incoherent noise environment,'" *Journal of the Acoustical Society of America*, vol.46, no.1, pt.1, pp.37-43, Jul. 1969.

[4] H.C. So, Y.T. Chan and F.K.W. Chan, "Closed-form formulae for optimum time difference of arrival based localization," *IEEE Transactions on Signal Processing*,

vol.56, no.6, pp.2614-2620, Jun. 2008.

[5] K.W.K. Lui and H.C. So, "Performance of modified covariance estimator for a single real tone," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol.E90-A, no.9, pp.2021-2023, Sep. 2007.

[6] K.W. Cheung, H.C. So, W.-K. Ma and Y.T. Chan, "Least squares algorithms for time-of-arrival based mobile location," *IEEE Transactions on Signal Processing*, vol.52, no.4, pp.1121-1128, Apr. 2004.

[7] H.C. So, "Source Localization: Algorithms and Analysis," *Handbook of Position Location: Theory, Practice and Advances*, Chapter 2, S.A. Zekavat and M. Buehrer, Eds., Wiley-IEEE Press, 2011.