TERM PAPER REPORT
ON

# LINEAR MODELS IN STATISTICAL SIGNAL PROCESSING

ANIMA MISHRA SHARMA, Y8104001
BISHWAJIT SHARMA, Y8104015

Course Instructor
DR. RAJESH HEDGE
Electrical Engineering Department
IIT Kanpur

# CONTENTS

# 1. INTRODUCTION

Linear estimators are much more appealing in practice due to reduced implementational complexity and relative simplicity of performance analysis. The determination of the MVU estimator is usually a difficult task. Fortunately a large number of signal processing estimation problems can be represented by a linear model that allows us to easily determine this estimator. Optimally structuring the problem in the linear model form is key to the derivation of optimal estimator. Once the linear model has been identified the MVU estimator is immediately evident.

# 2. DEFINITION AND PROPERTIES

In the problem to fit a straight line through noise corrupted data, our choice of model of the data is

$$x[n] = A + Bn + w[n], \quad n = 0,1,..., N - 1$$

where w[n] is WGN and the slope B and intercept A are to be estimated. The model is written more compactly in matrix notation as

$$\boldsymbol{x} = \boldsymbol{H\theta} + \boldsymbol{w} \quad \text{.........................................................(1)}$$

where,

$$\begin{aligned}
\mathbf{x} &= [x[0]\, x[1] \ldots x[N-1]]^T \\
\mathbf{w} &= [w[0]\, w[1] \ldots w[N-1]]^T \\
\boldsymbol{\theta} &= [A\, B]^T
\end{aligned}$$

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & N-1 \end{bmatrix}$$

Here, the matrix H, called observation matrix, is a known matrix of dimension N x 2. The data x are observed after theta is operated upon by H. Note also that the noise vector has the statistical characterization $\boldsymbol{w} \sim \mathbf{N}$ (0, σ2**I**). The data model in (1) is termed the linear model. In defining the linear model we assume that the noise vector is Gaussian. It is sometimes possible to determine the MVU estimator if the equality constraints of the CRLB theorem are satisfied. From Theorem 3.2 [Kay], g(x) will be the MVU estimator if

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{I}(\boldsymbol{\theta})(\mathbf{g}(\mathbf{x}) - \boldsymbol{\theta}) \quad \text{...................................(2)}$$

for some function **g**. Furthermore, the covariance matrix of the estimator will be given by $\mathbf{I}^{-1}(\boldsymbol{\theta})$.

To determine if this condition is satisfied for the linear model of (1), we have

$$\begin{aligned}
\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{\partial}{\partial \boldsymbol{\theta}}\left[-\ln(2\pi\sigma^2)^{\frac{N}{2}} - \frac{1}{2\sigma^2}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})\right] \\
&= -\frac{1}{2\sigma^2}\frac{\partial}{\partial \boldsymbol{\theta}}\left[\mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{H}\boldsymbol{\theta} + \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}\right].
\end{aligned}$$

Using the identities

$$\frac{\partial \mathbf{b}^T\boldsymbol{\theta}}{\partial\boldsymbol{\theta}} = \mathbf{b}$$
$$\frac{\partial \boldsymbol{\theta}^T\mathbf{A}\boldsymbol{\theta}}{\partial\boldsymbol{\theta}} = 2\mathbf{A}\boldsymbol{\theta} \qquad \text{.............................................(3)}$$

for **A** a symmetric matrix, and assuming that $H^TH$ is invertible, we have

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \frac{\mathbf{H}^T\mathbf{H}}{\sigma^2}[(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} - \boldsymbol{\theta}] \qquad \text{......................(4)}$$

which is exactly in the form of (2) with

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} \qquad \text{.........................................(5)}$$

$$\mathbf{I}(\boldsymbol{\theta}) = \frac{\mathbf{H}^T\mathbf{H}}{\sigma^2} \qquad \text{..................................(6)}$$

hence the MVU estimator is given by (5), and its covariance matrix is

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \sigma^2(\mathbf{H}^T\mathbf{H})^{-1} \qquad \text{.............................................(7)}$$

Moreover, the MVU estimator for the linear model is efficient, i.e., it attains the CRLB. Substituting **H** for the line fitting problem into (4), the result of line fitting problem can now be obtained. The only detail that requires closer scrutiny is the invertibility of $\mathbf{H}^T\mathbf{H}$. For the line fitting example a direct calculation will verify that the inverse exists. If the columns of H are not linearly independent, **x** lies in the range space of **H**. Therefore, even in the absence of noise, the model parameters will not be identifiable. Although rarely occurring in practice, this degeneracy sometimes occurs when $\mathbf{H}^T\mathbf{H}$ is ill-conditioned.

A generalization of the line fitting problem is summarized by the following theorem.

**Theorem 1 (Minimum Variance Unbiased Estimator for the Linear Model)**

If the data observed can be modeled as

$$\boldsymbol{x} = \boldsymbol{H\theta} + \boldsymbol{w} \qquad \text{.................................................................(8)}$$

where **x** is an N x 1 vector of observations, **H** is a known N x p observation matrix (with N > p) and rank p, $\boldsymbol{\theta}$ is a p x 1 vector of parameters to be estimated, and **w** is an N X 1 noise vector with PDF $\mathcal{N}(0, \sigma 2\mathbf{I})$, then the MVU estimator is

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} \qquad \text{.............................................(9)}$$

and, the covariance matrix of $\hat{\boldsymbol{\theta}}$ is

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \sigma^2(\mathbf{H}^T\mathbf{H})^{-1} \qquad \text{..................................................(10)}$$

For the linear model the MVU estimator is efficient in that it attains the CRLB.

By substituting (8) into (9), it is easily seen that the estimator is unbiased. Also, its statistical performance is completely specified, because $\hat{\boldsymbol{\theta}}$ is a linear transformation of a Gaussian vector **x** and hence

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, \sigma^2(\mathbf{H}^T\mathbf{H})^{-1}) \qquad \text{.................................................(11)}$$

The Gaussian nature of the MVU estimator for the linear model allows us to determine the exact statistical performance.

.

# 3. LINEAR MODEL EXAMPLES

We have already seen how the problem of line fitting is easily handled once we recognize it as a linear model. A simple extension is to the problem of fitting a curve to experimental data.

## a. Curve Fitting

In many experimental situations we seek to determine an empirical relationship between a pair of variables. For instance, in Figure 1, we present the results of a experiment in which voltage measurements are taken at the time instants $t = t_0, t_1, ..., t_{N-1}$. By plotting the measurements one can observe that the underlying voltage may be a quadratic function of time. Due to experimental error or noise, the points do not lie exactly on a curve. Hence, a reasonable model for the data is

$$x(t_n) = \theta_1 + \theta_2 t_n + \theta_3 t_n^2 + w(t_n) \quad n = 0, 1, \ldots, N-1.$$

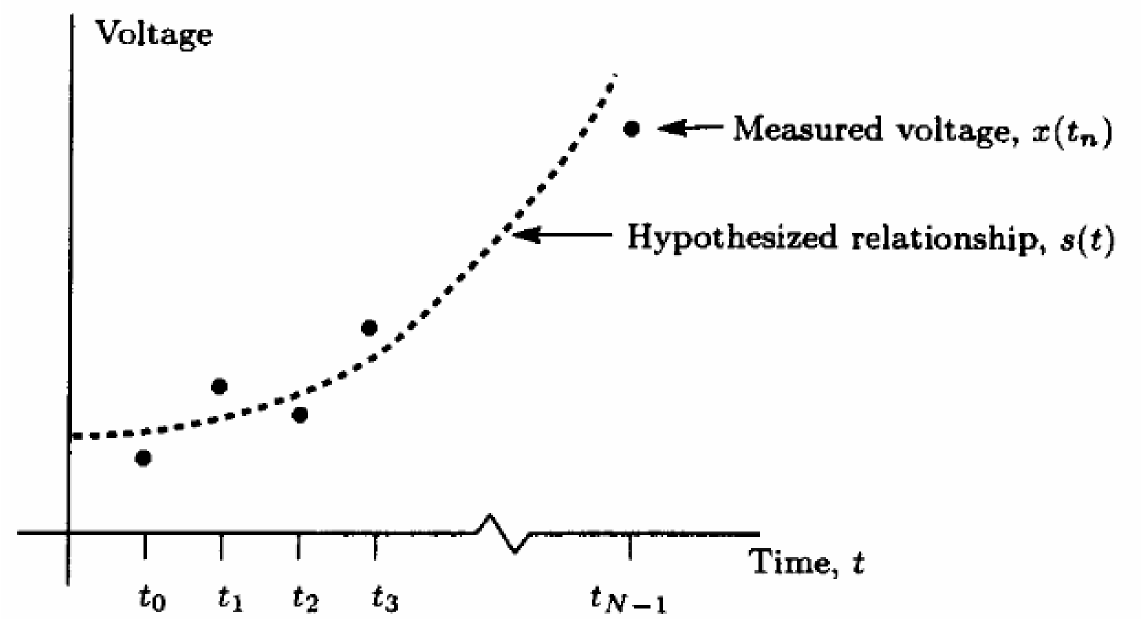


**Figure 1. Experimental Data for quadratic curve fitting**

To avail ourselves of the utility of the linear model we assume that $\mathbf{w}(t_n)$ are IID Gaussian random variables with zero mean and variance $\sigma^2$, or that they are WGN samples. Then, we have the usual linear model form

$\boldsymbol{x} = \boldsymbol{H\theta} + \boldsymbol{w}$

where,

$$\begin{aligned} \mathbf{x} &= [x(t_0)\, x(t_1) \ldots x(t_{N-1})]^T \\ \boldsymbol{\theta} &= [\theta_1\, \theta_2\, \theta_3]^T \end{aligned}$$

$$\mathbf{H} = \begin{bmatrix} 1 & t_0 & t_0^2 \\ 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{N-1} & t_{N-1}^2 \end{bmatrix}$$

In general, if we seek to fit a (p-l) order polynomial to experimental data, we will have

$$x(t_n) = \theta_1 + \theta_2 t_n + \cdots + \theta_p t_n^{p-1} + w(t_n) \qquad n = 0, 1, \ldots, N-1.$$

The MVU estimator follows from (9) as

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$$

where

$$\mathbf{x} = [x(t_0)\, x(t_1) \ldots x(t_{N-1})]^T$$

$$\mathbf{H} = \begin{bmatrix} 1 & t_0 & \ldots & t_0^{p-1} \\ 1 & t_1 & \ldots & t_1^{p-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t_{N-1} & \ldots & t_{N-1}^{p-1} \end{bmatrix}$$

The observation matrix for this example has the special form of a Vandermonde matrix.

## b. Fourier Analysis

Fourier analysis is commonly employed for analyzing signals that exhibit cyclical behavior. It is common practice to determine the presence of strong cyclical components by employing a Fourier analysis. Large Fourier coefficients are indicative of strong components. In this example we show that a Fourier analysis is actually just an estimation of the linear model parameters. Consider a data model consisting of sinusoids in white Gaussian noise:

$$x[n] = \sum_{k=1}^{M} a_k \cos\left(\frac{2\pi kn}{N}\right) + \sum_{k=1}^{M} b_k \sin\left(\frac{2\pi kn}{N}\right) + w[n] \qquad n = 0, 1, \ldots, N-1 \tag{12}$$

where **w**[n] is WGN. The frequencies are assumed to be harmonically related or multiples of the fundamental $f_1 = 1/N$ as $f_k = k/N$.

The amplitudes $a_k$,$b_k$ of the cosines and sines are to be estimated. To reformulate the problem in terms of the linear model we let

$$\boldsymbol{\theta} = [a_1\, a_2 \ldots a_M\, b_1\, b_2 \ldots b_M]^T$$

and

$$\mathbf{H} = \begin{bmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 \\ \cos\left(\frac{2\pi}{N}\right) & \cdots & \cos\left(\frac{2\pi M}{N}\right) & \sin\left(\frac{2\pi}{N}\right) & \cdots & \sin\left(\frac{2\pi M}{N}\right) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \cos\left[\frac{2\pi(N-1)}{N}\right] & \cdots & \cos\left[\frac{2\pi M(N-1)}{N}\right] & \sin\left[\frac{2\pi(N-1)}{N}\right] & \cdots & \sin\left[\frac{2\pi M(N-1)}{N}\right] \end{bmatrix}$$

Note that **H** has dimension N x 2M, where p = 2M. Hence, for **H** to satisfy N > p we require M < N/2. In determining the MVU estimator we can simplify the computations by noting that the columns of **H** are orthogonal. Let **H** be represented in column form as

$$\mathbf{H} = [\mathbf{h}_1\ \mathbf{h}_2 \ldots\ldots \mathbf{h}_{2M}]$$

where $\mathbf{h}_i$ denotes the $i^{th}$ column of **H**. Then, it follows that $\mathbf{h}_i^T\mathbf{h}_j=0$ for i not equal to j.

This property is quite useful since $\mathbf{H}^T\mathbf{H}$ becomes a diagonal matrix which is easily inverted. The orthogonality of the columns results from the discrete Fourier transform (DFT) relationships [Kay].

Using this property we have $\mathbf{H}^T\mathbf{H} = (N/2)\mathbf{I}$, so the MVU estimator of the amplitude is

$$\hat{\boldsymbol{\theta}} = \frac{2}{N}\mathbf{H}^T\mathbf{x} = \frac{2}{N}\begin{bmatrix} \mathbf{h}_1^T \\ \vdots \\ \mathbf{h}_{2M}^T \end{bmatrix}\mathbf{x} = \begin{bmatrix} \frac{2}{N}\mathbf{h}_1^T\mathbf{x} \\ \vdots \\ \frac{2}{N}\mathbf{h}_{2M}^T\mathbf{x} \end{bmatrix}$$

which means,

$$\hat{a}_k = \frac{2}{N}\sum_{n=0}^{N-1} x[n]\cos\left(\frac{2\pi kn}{N}\right) \qquad \hat{b}_k = \frac{2}{N}\sum_{n=0}^{N-1} x[n]\sin\left(\frac{2\pi kn}{N}\right) \qquad \ldots(13)$$

These are recognized as Discrete Fourier Transform coefficients. From the properties of Linear model we can immediately conclude that the covariance matrix is $(2\sigma^2/N)\mathbf{I}$.

It can be seen that, because the estimator is Gaussian and the covariance matrix is diagonal, the amplitude estimates are independent. It is seen from this example that a key ingredient in simplifying the computation of the MVU estimator and its covariance matrix is the orthogonality of the columns of H. Note that this property does not hold if the frequencies are arbitrarily chosen.

# 4. GENERALIZED LINEAR MODEL

A more general form of the linear model allows for noise that is not white. The general linear model assumes that $\mathbf{w} \sim \mathcal{N}(0,\mathbf{C})$, where $\mathbf{C}$ is not necessarily a scaled identity matrix. To determine the MVU estimator, we could have repeated the derivation in Section 2. Alternatively, we can use a whitening approach as follows. Since $\mathbf{C}$ is assumed to be positive definite, $\mathbf{C}^{-1}$ is positive definite and so can be factored as

$$\mathbf{C}^{-1} = \mathbf{D}^T\mathbf{D} \quad (14)$$

where $\mathbf{D}$ is an NxN invertible matrix. The matrix D acts as a whitening transformation when applied to $\mathbf{w}$ since

$E[(\mathbf{Dw})(\mathbf{Dw})^T] = \mathbf{DCD}^T = \mathbf{DD}^{-1}(\mathbf{D}^T)^{-1}\mathbf{D}^T = \mathbf{I}$

As a consequence, if we transform our generalized model

$\boldsymbol{x} = \boldsymbol{H\theta} + \boldsymbol{w}$

to

$$\begin{aligned}\mathbf{x}' &= \mathbf{Dx} \\ &= \mathbf{DH}\boldsymbol{\theta} + \mathbf{Dw} \\ &= \mathbf{H}'\boldsymbol{\theta} + \mathbf{w}'\end{aligned}$$

the noise will be whitened since $\mathbf{w}' = \mathbf{Dw} \sim N(0,\mathbf{I})$, and the usual linear model will result. The MVU estimator of $\boldsymbol{\theta}$ is,

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= (\mathbf{H}'^T\mathbf{H}')^{-1}\mathbf{H}'^T\mathbf{x}' \\ &= (\mathbf{H}^T\mathbf{D}^T\mathbf{D}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{D}^T\mathbf{D}\mathbf{x}\end{aligned}$$

so that

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x} \quad \text{.........................................(15)}$$

In a similar fashion we find that

$$\mathbf{C}_{\hat{\theta}} = (\mathbf{H}'^T\mathbf{H}')^{-1}$$

or finally,

$$\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1} \quad \text{..............................................(16)}$$

If $\mathbf{C} = \sigma^2\mathbf{I}$, we have the same previous result.

## a. Whitening of Colored Noise

We now extend the example of DC level in noise to the colored noise case. If x[n] = A + w[n] for n = 0,1,...,N-1, where w[n] is colored Gaussian noise with N x N covariance matrix C, it immediately follows from (15) that with $\mathbf{H} = \mathbf{1} = [1\ 1\ \ldots\ 1]^T$, the MVU estimator of the DC level is

$$\begin{aligned}\hat{A} &= (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x} \\ &= \frac{\mathbf{1}^T\mathbf{C}^{-1}\mathbf{x}}{\mathbf{1}^T\mathbf{C}^{-1}\mathbf{1}}\end{aligned}$$

and the variance is

$$\begin{aligned}\operatorname{var}(\hat{A}) &= (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1} \\ &= \frac{1}{\mathbf{1}^T\mathbf{C}^{-1}\mathbf{1}}.\end{aligned}$$

If $\mathbf{C} = \sigma^2\mathbf{I}$, we have as our MVU estimator the sample mean with a variance of $\sigma^2/N$. An interesting interpretation of the MVU estimator follows by considering the factorization of $\mathbf{C}^{-1}$ as $\mathbf{D}^T\mathbf{D}$. We noted previously that **D** is a whitening matrix. The MVU estimator is expressed as

$$\begin{aligned}\hat{A} &= \frac{\mathbf{1}^T\mathbf{D}^T\mathbf{D}\mathbf{x}}{\mathbf{1}^T\mathbf{D}^T\mathbf{D}\mathbf{1}} \\ &= \frac{(\mathbf{D}\mathbf{1})^T\mathbf{x}'}{\mathbf{1}^T\mathbf{D}^T\mathbf{D}\mathbf{1}} \\ &= \sum_{n=0}^{N-1} d_n x'[n]\end{aligned} \quad \text{.......................................................(17)}$$

where $\boldsymbol{d}_n = [\mathbf{Dl}]_n/\mathbf{1}^T\mathbf{D}^T\mathbf{D1}$. According to this result, the data are first prewhitened to form $\mathbf{x}'[n]$ and then "averaged" using prewhitened averaging weights $\boldsymbol{d}_n$. The prewhitening has the effect of decorrelating and equalizing the variances of the noises at each observation time before averaging.

Another extension to the linear model allows for signal components that are known. Assume that **s** represents a known signal contained in the data. Then, a linear model that incorporates this signal is

$$\boldsymbol{x} = \boldsymbol{H\theta} + \boldsymbol{s} + \boldsymbol{w}$$

To determine the MVU estimator let $\mathbf{x}' = \mathbf{x}-\mathbf{s}$, so that

$$\boldsymbol{x}' = \boldsymbol{H\theta} + \boldsymbol{w}$$

which is now in the form of the linear model. The MVU estimator follows as

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T(\mathbf{x}-\mathbf{s}) \quad \text{.........................................(18)}$$

with covariance

$$\mathbf{C}_{\hat{\theta}} = \sigma^2(\mathbf{H}^T\mathbf{H})^{-1} \quad \text{...................................................(19)}$$

## b. Known Signal Components

If $x[n] = A + r^n + w[n]$ for $n = 0,1,..., N-1$, where r is known, A is to be estimated, and w[n] is WGN the model is $x = A[1\ 1\ \ldots 1]^T + s + w$, where $s = [1\ r\ \ldots\ r^{N-1}]^T$. The MVU estimator is

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

with variance,

$$\mathrm{var}(\hat{A}) = \frac{\sigma^2}{N}.$$

It should be clear that the two extensions described can be combined to produce the general linear model summarized by the following theorem.

# 5. MVUE FOR GENERAL LINEAR MODEL

**Theorem 2: (Minimum Variance Unbiased Estimator for General Linear Model)**

If the data can be modeled as

$$\boldsymbol{x} = \boldsymbol{H\theta} + \boldsymbol{s} + \boldsymbol{w} \qquad (20)$$

where **x** is an N x 1 vector of observations, **H** is a known N x p observation matrix (N > p) of rank p, 0 is a p x 1 vector of parameters to be estimated, s is an N x 1 vector of known signal samples. and w is an N x 1 noise vector with PDF *N*(0, **C**), then the MVU estimator is

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}(\mathbf{x} - \mathbf{s}) \qquad (21)$$

and the covariance matrix is

$$\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1} \qquad (22)$$

For the general Linear model the MVU estimator is efficient in that it attains the CRLB.

# 6. REFERENCES

**[Kay]** S.M. Kay's Fundamentals of Statistical Signal Processing: Estimation Theory (Volume 1), Prentice Hall Signal Processing Series.

**[Mono]** D. G. Monolakis, V. K. Ingle, and S. M. Kogon, Statistical and. Adaptive Signal Processing, McGraw-Hill, USA, 2000.