Week 3 : Session 1

4. Vector Parameter Estimation

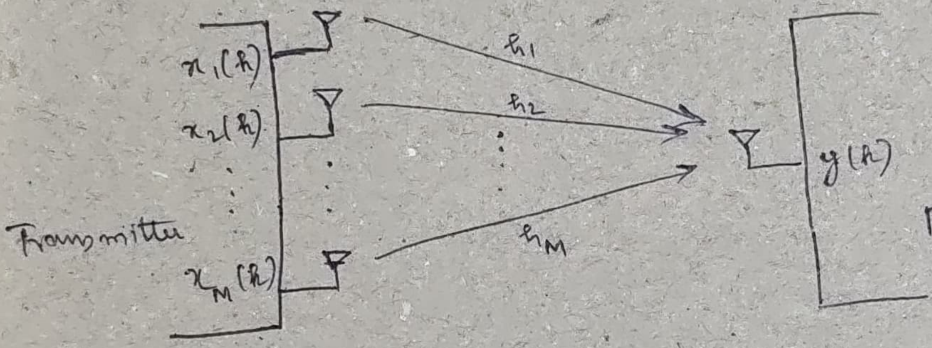
Vector Parameter Estimation is nothing but, simultaneous estimation of a vector containing multiple individual parameters as its components.

So far, we have considered the unknown parameter as Scalar (h). Now, let us consider the parameter as vector ($\bar{h}$).

$$\text{Vector Parameter } (\bar{h}) = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_M \end{bmatrix}_{M \times 1}$$

∴ The Parameter has M components ($h_1, h_2, \ldots, h_M$).

Consider a MISO (Multiple Input Single Output) system with M transmit antennas and single receive antenna.



Tx. Antennas = M

Rx. Antenna = 1

Fig. MISO Channel Schematic

$x_i(k) \rightarrow$ Symbol transmitted from Tx antenna i at time k.

$y(k) \rightarrow$ Output symbol at the receiver at time k.

$h_i \rightarrow$ Channel coefficients.

The MISO system is given as

$$y(k) = x_1(k)\, h_1 + x_2(k)\, h_2 + \cdots + x_M(k)\, h_M + v(k)$$

$$= \underbrace{\begin{bmatrix} x_1(k) & x_2(k) & \cdots & x_M(k) \end{bmatrix}}_{\bar{x}^T(k)} \underbrace{\begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_M \end{bmatrix}}_{\bar{h}} + v(k)$$

$$\boxed{y(k) = \bar{x}^T(k)\, \bar{h} + v(k)}$$

The MISO system is given as $\bar{y}(k) = \bar{x}^T(k)\,\bar{h} + v(k)$,

where $\bar{x}(k) = \begin{bmatrix} x_1(k) \\ x_2(k) \\ \vdots \\ x_M(k) \end{bmatrix}$ is the PILOT vector at time k.

Consider now, the transmission of N pilot vectors, for N time instances.

$$y(1) = \bar{x}^T(1)\,\bar{h} + v(1)$$
$$y(2) = \bar{x}^T(2)\,\bar{h} + v(2)$$
$$\vdots$$
$$y(N) = \bar{x}^T(N)\,\bar{h} + v(N)$$

This can be written in the Matrix form vectors.

$$\underbrace{\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix}}_{\substack{N\times 1 \\ \text{output} \\ \text{vector}}} = \underbrace{\begin{bmatrix} \bar{x}^T(1) \\ \bar{x}^T(2) \\ \vdots \\ \bar{x}^T(N) \end{bmatrix}}_{\substack{N\times M \\ \text{Pilot} \\ \text{Matrix}}} \underbrace{\begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_M \end{bmatrix}}_{\substack{M\times 1 \\ \text{Parameter} \\ \text{vector}}} + \underbrace{\begin{bmatrix} v(1) \\ v(2) \\ \vdots \\ v(N) \end{bmatrix}}_{\substack{N\times 1 \\ \text{Noise} \\ \text{vector}}}$$

Each $\bar{x}(k)$ is $M\times 1$ pilot vector

$$\Rightarrow \boxed{\bar{y} = X\,\bar{h} + \bar{v}}$$ is the compact model.

Note that, X is of size $N\times M$, $N \geq M$.

This is known as a Tall Matrix.

$$\underbrace{\begin{bmatrix} \bar{x}^T(1) \\ \bar{x}^T(2) \\ \vdots \\ \bar{x}^T(N) \end{bmatrix}}_{X}$$

\# rows $\geq$ \# columns

$\Rightarrow$ "Tall Matrix"

Now, how do we estimate the vector parameter $\bar{h}$? First we have to construct the Likelihood and then maximize the likelihood to obtain the estimate of the parameter vector $\bar{h}$.

The Likelihood for the estimation of the vector parameter $\bar{h}$ can be obtained / constructed as follows.

Note that $v(k)$ is Gaussian with mean = 0, var = σ^2.

(i.e) $v(k) \sim \mathcal{N}(0, \sigma^2)$.

The PDF of the noise sample $v(k)$ is

$$f_{V(k)}(v(k)) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} v^2(k)}$$

The noise samples $v(1), v(2), \ldots, v(N)$ are iid. The Joint PDF of the noise samples is given by the product of the individual PDFs.

$$\text{(i.e)}\quad f_{\bar{V}}(\bar{v}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} v^2(1)} \times \ldots \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} v^2(N)}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N}{2}} e^{-\frac{1}{2\sigma^2} \sum_{k=1}^{N} v^2(k)}$$

$$\sum_{k=1}^{N} v^2(k) = \|\bar{v}\|^2$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N}{2}} e^{-\frac{\|\bar{v}\|^2}{2\sigma^2}}$$

We have,

$$\bar{y} = X\bar{h} + \bar{v} \Rightarrow \bar{v} = \bar{y} - X\bar{h}.$$

Since $\bar{v}$ is Gaussian, $\bar{y}$ is also Gaussian with mean $= X\bar{h}$.

Therefore, the Joint PDF of $\bar{y}$ is obtained by substituting $\bar{v} = \bar{y} - X\bar{h}$ in $f_{\bar{v}}(\bar{v})$.

$$\Rightarrow f_Y(\bar{y}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N}{2}} e^{-\frac{1}{2\sigma^2}\|\bar{y} - X\bar{h}\|^2}$$

The Joint PDF of $\bar{y}$, when viewed as a function of the unknown parameter vector $\bar{h}$, becomes the likelihood of $\bar{h}$.

$$p(\bar{y}; \bar{h}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N}{2}} e^{-\frac{1}{2\sigma^2}\|\bar{y} - X\bar{h}\|^2}$$

where, $\bar{y}$ – $N \times 1$ output vector

X – $N \times M$ Pilot Matrix

$\bar{h}$ – $M \times 1$ Parameter vector

Once again, maximize the likelihood to estimate $\bar{h}$.

$$\Rightarrow \max \; p(\bar{y}\,;\,\bar{h})$$

$$\Rightarrow \max \; \underbrace{\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N}{2}}}_{\text{Constant}} e^{-\frac{1}{2\sigma^2}\|\bar{y} - X\bar{h}\|^2}$$

due to '$-$' in the exponent

$$\Rightarrow \underline{\min} \; \|\bar{y} - X\bar{h}\|^2$$

"Minimize Square of Norm". This is termed as Least Squares (LS) Problem. The Solution to this problem is given as

$$\underbrace{\hat{h} = (X^T X)^{-1} X^T \cdot \bar{y}}_{\text{Least Squares (LS) Solution}}$$

The quantity $(X^T X)^{-1} X^T$ is termed as "Pseudo-inverse of X".

$$\text{(ie)} \; \underbrace{(X^T X)^{-1} X^T}_{X^{-1}} * X = I$$

But, recall X is Tall Matrix

$\Rightarrow X$ is NOT invertible unless # Rows $\neq$ # Columns.

Thus, $(X^T X)^{-1} X^T$ acts as a left inverse of X.

Hence the name "Pseudo-inverse of X".

The LS Estimate is

$$\bar{y} = X\bar{h} + \bar{v}$$

$$\begin{aligned}
\hat{h} &= (X^T X)^{-1} X^T \cdot \bar{y} \\
&= (X^T X)^{-1} X^T \cdot (X\bar{h} + \bar{v}) \\
&= \underbrace{(X^T X)^{-1} X^T X}_{I} \bar{h} + (X^T X)^{-1} X^T \bar{v}
\end{aligned}$$

$$\boxed{\hat{h} = \bar{h} + (X^T X)^{-1} X^T \bar{v}}$$

Week 3 : Session 2

Properties of LS Estimate

Mean

$$E\{\hat{h}\} = E\left\{\bar{h} + (X^TX)^{-1}X^T\bar{v}\right\}$$

$$= \bar{h} + (X^TX)^{-1}X^T \cancel{E\{\bar{v}\}}$$

$\xrightarrow{} 0$, since $\bar{v}$ is zero mean Gaussian noise.

$$\boxed{E\{\hat{h}\} = \bar{h}}$$

Mean of the estimate coincides with the True parameter. Such an estimator is termed as UNBIASED ESTIMATOR.

Covariance Matrix

WKT, $\hat{h} = \bar{h} + (X^TX)^{-1}X^T\bar{v}$

$$\Rightarrow (\hat{h} - \bar{h}) = (X^TX)^{-1}X^T\bar{v}.$$

The covariance matrix of the estimate is given as

$$E\{(\hat{h}-\bar{h})(\hat{h}-\bar{h})^T\} = E\left\{\underbrace{(X^TX)^{-1}X^T\bar{v}}_{(\hat{h}-\bar{h})} \cdot \underbrace{\bar{v}^TX(X^TX)^{-1}}_{(\hat{h}-\bar{h})^T}\right\}$$

$$= (X^TX)^{-1}X^T\, \underbrace{E[\bar{v}\bar{v}^T]}\, X(X^TX)^{-1}$$

$$E\{\bar{v}\bar{v}^T\} = E\left\{\begin{bmatrix} v(1) \\ v(2) \\ \vdots \\ v(N)\end{bmatrix}\begin{bmatrix} v(1) & v(2) & \cdots & v(N)\end{bmatrix}\right\}$$

$$= E\left\{\begin{bmatrix} v^2(1) & \cancel{v(1)v(2)} & \cdots \\ \cancel{v(2)v(1)} & v^2(2) & \\ \vdots & & \ddots \end{bmatrix}\right\}$$

Since $v(1), v(2), \ldots, v(N)$ are independent

$$= \begin{bmatrix} \sigma^2 & 0 & 0 & \cdots \\ 0 & \sigma^2 & & \\ \vdots & & \ddots & \\ & & & \sigma^2 \end{bmatrix}$$

$$= \sigma^2 I$$

$$\Rightarrow E\left\{(\hat{h}-\bar{h})(\hat{h}-\bar{h})^T\right\} = (X^TX)^{-1}X^T \underline{\sigma^2 I}\, X(X^TX)^{-1}$$

$$= \sigma^2 \underbrace{(X^TX)^{-1}X^TX}_{I} \cdot (X^TX)^{-1}$$

$$\boxed{E\{(\hat{h}-\bar{h})(\hat{h}-\bar{h})^T\} = \sigma^2 (X^TX)^{-1}}$$

This is known as Error Covariance.

Mean Square Error (MSE)

The MSE of the estimate is given as "Trace of Error Covariance Matrix".

Trace $\rightarrow$ Sum of Diagonal Elements of Square matrix.

$$MSE = Tr\left\{\sigma^2 (X^TX)^{-1}\right\}$$

$$\boxed{MSE = \sigma^2\, Tr\left\{(X^TX)^{-1}\right\}}$$

The MSE can also be represented as follows.

$$MSE = Tr\left\{ E\left\{(\hat{h}-\bar{h})(\hat{h}-\bar{h})^T\right\}\right\}$$

$$= E\left\{\|\bar{h}-\hat{h}\|^2\right\}$$

Sum of MSEs of individual parameters.

$$\boxed{MSE = \sum_{i=1}^{M} E\left\{(\hat{h}_i - h_i)^2\right\}}$$

Example: Consider the Pilot matrix $X = \begin{bmatrix} 1 & 4 \\ 1 & 2 \\ 1 & 1 \\ 1 & 3 \end{bmatrix}_{N \times M}$,

Output vector $\bar{y} = \begin{bmatrix} -1 \\ 3 \\ -2 \\ -1 \end{bmatrix}_{N\times 1}$. What is the ML estimate of the unknown vector parameter $(\bar{h})$.

No. of Pilot vectors, $N = 4$

No. of Antennas, $M = 2$.

The ML Estimate, $\hat{h} = (X^T X)^{-1} X^T \bar{y}$.

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 4 & 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 1 & 2 \\ 1 & 1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$(X^T X)^{-1} = \frac{1}{|X^T X|} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$= \frac{1}{20} \begin{bmatrix} 30 & -10 \\ -10 & 4 \end{bmatrix}$$

$$\hat{h} = \frac{1}{20} \begin{bmatrix} 30 & -10 \\ -10 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 4 & 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} -1 \\ 3 \\ -2 \\ -1 \end{bmatrix}$$

$$= \frac{1}{20} \begin{bmatrix} -10 & -10 & 20 & 0 \\ 6 & -2 & -6 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 3 \\ -2 \\ -1 \end{bmatrix}$$

$$= \frac{1}{20} \begin{bmatrix} 0 \\ -2 \end{bmatrix}$$

$$= \frac{1}{10} \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

$$\boxed{\hat{h} = \begin{bmatrix} 0 \\ -1/10 \end{bmatrix}}.$$

Given $\sigma^2 = \frac{1}{2}$, what is the Error Covariance Matrix and MSE ?

$$\text{Error Covariance} = \sigma^2 (X^T X)^{-1}$$

$$= \frac{1}{2} \times \frac{1}{20} \begin{bmatrix} 30 & -10 \\ -10 & 4 \end{bmatrix}$$

$$= \frac{1}{40} \begin{bmatrix} 30 & -10 \\ -10 & 4 \end{bmatrix}$$

$$\text{MSE} = \text{Tr}(\text{Error covariance})$$

$$= \text{Tr}\left(\frac{1}{40} \begin{bmatrix} 30 & -10 \\ -10 & 4 \end{bmatrix} \right)$$

$$= \frac{1}{40} (34)$$

$$= \frac{17}{20}.$$

Week 5 : Session 1.

Least Squares Solution

Let us understand how to derive the Least Squares Estimate formula. We saw that, the Least squares Solution as

$$\boxed{\hat{h} = (X^T X)^{-1} X^T \bar{y}}$$ How to derive this?

WKT, The LS cost function,

$$\| \bar{y} - X \bar{h} \|^2$$

$$= (\bar{y} - X\bar{h})^T (\bar{y} - X\bar{h})$$

$$= (\bar{y}^T - \bar{h}^T X^T)(\bar{y} - X\bar{h})$$

$$= \bar{y}^T \bar{y} - \bar{h}^T X^T \bar{y} - \bar{y}^T X \bar{h} + \bar{h}^T X^T X \bar{h}$$

These are Scalar quantities and Transpose. Hence they are EQUAL!
eg: $[4]^T = [4]$
$(\bar{h}^T X^T \bar{y})^T = \bar{y}^T X \bar{h}$

$$= \bar{y}^T \bar{y} - 2 \bar{h}^T X^T \bar{y} + \bar{h}^T X^T X \bar{h}$$

Now, minimize this cost function w.r.t $\bar{h}$.

Note: When h is Scalar, we find the derivative and set to 0.
Here $\bar{h}$ is not Scalar, rather it is vector.
So, in order to minimize, we now Calculate the Gradient and Set to 0.

What is Gradient?

$$\text{Gradient of } \bar{h} ; \nabla f(\bar{h}) = \begin{bmatrix} \frac{\partial f}{\partial h_1} \\ \frac{\partial f}{\partial h_2} \\ \vdots \\ \frac{\partial f}{\partial h_M} \end{bmatrix}_{M \times 1}$$

Partial derivatives w.r.t. each component of $\bar{h}$.

Let $\bar{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_M \end{bmatrix}$. We use the following principles.

① $\bar{c}^T \bar{h} = c_1 h_1 + \cdots + c_M h_M = \bar{h}^T \bar{c}$.

$$\nabla \bar{c}^T \bar{h} = \nabla \bar{h}^T \bar{c} = \begin{bmatrix} \frac{\partial}{\partial h_1} c_1 h_1 \\ \frac{\partial}{\partial h_2} c_2 h_2 \\ \vdots \\ \frac{\partial}{\partial h_M} c_M h_M \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_M \end{bmatrix} = \bar{c}$$

Therefore,

$$\boxed{\nabla \bar{h}^T \bar{c} = \nabla \bar{c}^T \bar{h} = \bar{c}} \text{ — (1)}$$

② For any symmetric matrix $P = P^T$

$$\boxed{\nabla (\bar{h}^T P \bar{h}) = 2 P \bar{h}} \text{ — (2)}$$

Now, the Gradient of the LS cost function is

$$\nabla \| \bar{y} - X \bar{h} \|^2$$

$$= \underbrace{\nabla \bar{y}^T \bar{y}}_{\substack{0,\text{ since } \bar{y} \text{ is} \\ \text{a given fixed} \\ \text{vector!}}} - \underbrace{\nabla 2 \bar{h}^T X^T \bar{y}}_{\substack{\nabla \bar{h}^T \bar{y} = \bar{y} \\ \text{(principle ①)}}} + \underbrace{\nabla \bar{h}^T X^T X \bar{h}}_{\substack{\nabla (\bar{h}^T X \bar{h}) = 2X\bar{h} \\ \text{(Principle ②)}}}$$

$$= 0 - 2 X^T \bar{y} + 2 (X^T X) \bar{h}$$

To minimize, set the Gradient to zero.

$$\Rightarrow -2 X^T \bar{y} + 2 (X^T X) \bar{h} = 0$$

$$\Rightarrow \cancel{2} X^T X \bar{h} = \cancel{2} X^T \bar{y}$$

$$\Rightarrow \boxed{\hat{h} = (X^T X)^{-1} X^T \bar{y}}$$

This is the Maximum Likelihood Estimate of the vector parameter ($\bar{h}$) using the Least Squares (LS) Solution.