ECE801
Monitoring and Estimation
Linear Multi-Parameter Systems

Instructor: Christos Panayiotou
Outline

- Least Squares
- Weighted Least Squares
- Maximum Likelihood
- Maximum A Priori (MAP) estimator
- Recursive Least Squares
Simple Example

Assume that you run an experiment and know that the model is given by

\[ y = g(x) = ax + b \]

You run \( n \) experiments with inputs \( x_1, x_2, \ldots, x_n \) and measure the outputs \( y_1, y_2, \ldots, y_n \).

Solution

- This is an overdetermined system with \( n \) equations and 2 unknowns.
- Define the mean square error (MSE) criterion

Determine the parameters \( a, b \) that minimize MSE
Simple Example

- Compute the partial derivatives with respect to the parameters $a, b$ and set them to zero.

- Let $X_1 = \sum_{i=1}^{n} x_i$, $X_2 = \sum_{i=1}^{n} x_i^2$, $Y_1 = \sum_{i=1}^{n} y_i$, $Y_2 = \sum_{i=1}^{n} x_i y_i$, we get

\[
\begin{align*}
Y_2 - aX_2 - bX_1 &= 0 \\
Y_1 - aX_1 - nb &= 0
\end{align*}
\]

- Line_LS
In vector form...

- Assume that $Y = [y_1, \ldots, y_n]^T$, $A = [a, b]^T$, $W$ is the $n$-dimensional noise (error) vector and
  \[ H = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \]

- Thus, in vector form the model is given by $Y = HA + W$

- The Least Squares objective is given by

- To minimize $J(A)$ we need to find the gradient with respect to $A$ and make it equal to 0; i.e., $\nabla J(A) = 0$
Vector gradients

\[ \nabla_x \left\{ Y^T BX \right\} = B^T Y \]

\[ \nabla_x \left\{ \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\} = B^T Y \]

\[ \nabla_x \left\{ X^T BY \right\} = B^T Y \]

\[ \nabla_x \left\{ X^T BX \right\} = (B + B^T)X \]

So, we obtain the gradient (with respect to \( A \)),

\[ J(A) = Y^T Y - Y^T HA - A^T H^T Y + A^T H^T HA \]

- assuming \( (H^T H) \) is invertible
- \textbf{Line LS2}
Weighted Least Squares

- Let the Weighted Least Squares objective given by
  
  \[
  n \quad \frac{w}{Q}\quad \text{where } Q \text{ is a positive definite symmetric weight matrix}
  \]

- Again to minimize \( J(A) \) we need to find the gradient with respect to \( A \) and make it equal to 0; i.e., \( \nabla J(A) = 0 \)
Weighted Least Squares

The gradient (with respect to $A$),

assuming $(H^T Q H)$ is invertible
Assume that the system is a multi parameter ($K$ parameters) linear system.

We run $n$ experiments, $i = 1, ..., n$ where for the $i$th-experiment the input, $x_i = [x_1, ..., x_K]^T$ and measure the outputs $Y^T = [y_1, y_2, ..., y_n]$. In vector form, all outputs are given by:

$$x_i$$ is the $i$th row of $H$.

Define the mean square error (MSE) criterion
Multi-Parameter System

So, we obtain the gradient (with respect to $A$),

- assuming $(H^TH)$ is invertible
- Line_LS2
Maximum Likelihood

- Again assume a linear model where the noise $W$ is Gaussian $W \sim N(0, R_n)$.

$$Y = HA + W$$

- Thus, the conditional pdf $f(Y|A)$ (likelihood) is Gaussian

$$f(Y|A) = \frac{1}{\sqrt{(2\pi)^n |R_n|}} \exp \left( -\frac{1}{2} (Y - HA)^T R_n^{-1} (Y - HA) \right)$$

The logarithm is given by

$$\ln(f(Y|A)) = -\frac{1}{2} \left( Y^T R_n^{-1} Y - Y^T R_n^{-1} HA - A^T H^T R_n^{-1} Y + A^T H^T R_n^{-1} HA + \ln((2\pi)^n |R_n|) \right)$$

$$\nabla \ln(f(Y|A)) =$$
Maximum Likelihood

- Unbiasedness
  \[ E[Y] = E[HA + W] = HA \]

- Therefore
  \[ E[\hat{A}_{ML}] = \]

- Define the error \( A_e = A - \hat{A}_{ML} \), and note that \( E[A_e] = 0 \)
  \( A_e = \)

- And error covariance
Again assume a linear model where the noise $W$ is Gaussian $W \sim N(0, R_n)$.

$$Y = HA + W$$

Thus, the conditional pdf $f(Y|A)$ (likelihood) is Gaussian

$$f(Y|A) = \frac{1}{\sqrt{(2\pi)^n|R_n|}} \exp \left( -\frac{1}{2}(Y - HA)^T R_n^{-1}(Y - HA) \right)$$

And the prior distribution of $A$ is given by

Where $k$ is the dimension of $A$ which is normally distributed with mean $\mu_A$ and variance $\Sigma_A$. 
MAP Estimator

Recall that

\[ \hat{A}_{\text{MAP}} = \max_A \left\{ f(y \mid A) f(A) \right\} \]

Thus the overall function to be maximized is

The logarithm is given by

\[ \nabla \ln \left( f(Y \mid A) f(A) \right) = \]
MAP Estimator

- Collecting together all terms proportional to $A$.

- Therefore

- Next we investigate the unbiasedness of the estimator
Example

- Let a system with 2 inputs and 2 outputs,
  \[ y_1 = 2x_1 + x_2 + n_1, \quad y_2 = -2x_1 + x_2 + n_2 \]

- The inputs are Gaussian
  \[ X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim N \left( \mu_X = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \Sigma_X = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix} \right) \]

- and the outputs are corrupted by Gaussian noise \( n \).
  \[ W = \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} \sim N \left( \mu_n = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, R_n = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \right) \]
Example

- In vector form

\[
Y = \begin{bmatrix}
y_1 \\
y_2
\end{bmatrix} = \begin{bmatrix}
2 & 1 \\
-2 & 1
\end{bmatrix} \begin{bmatrix}
x_1 \\
x_2
\end{bmatrix} + \begin{bmatrix}
n_1 \\
n_2
\end{bmatrix} = HX + W
\]

- Next we need to compute the inverse matrices

- So
Example

For the MAP estimator we need

\[ \hat{A}_{MAP} = \left( H^T R_n^{-1} H + \Sigma_X^{-1} \right)^{-1} \left( H^T R_n^{-1} Y + \Sigma_X^{-1} \mu_X \right) \]