Chapter 4
Linear Models
General Linear Model
Recall signal + WGN case: \( x[n] = s[n;\theta] + w[n] \)
\[
\begin{align*}
x &= s(\theta) + w \\
\text{Here, dependence on } \theta \text{ is general}
\end{align*}
\]
Now we consider a special case: **Linear “Observations”:**
\[
\begin{align*}
s(\theta) &= H\theta + b
\end{align*}
\]
Note: “Gaussian” is part of the “Linear Model”
Need For Full-Rank H Matrix

**Note:** We must assume H is full rank

**Q:** Why?

**A:** If not, the estimation problem is “ill-posed”

...given vector s there are multiple θ vectors that give s:

If H is not full rank...

Then for any s : ∃ θ₁, θ₂ such that s = Hθ₁ = Hθ₂
Importance of The Linear Model

There are several reasons:

1. Some applications admit this model
2. Nonlinear models can sometimes be linearized
3. Finding Optimal Estimator is Easy

$$\hat{\theta}_{MVU} = \left( H^T C^{-1} H \right)^{-1} H^T C^{-1} (x - b)$$ … as we’ll see!!!
**MVUE for Linear Model**

**Theorem:** The MVUE for the General Linear Model and its covariance (i.e. its accuracy performance) are given by:

\[
\hat{\theta}_{MVU} = \left( H^T C^{-1} H \right)^{-1} H^T C^{-1} (x - b)
\]

\[
C_{\hat{\theta}} = \left( H^T C^{-1} H \right)^{-1}
\]

and achieves the CRLB.

**Proof:** We’ll do this for the \( b = 0 \) case but it can easily be done for the more general case.

First we have that \( x \sim N(H\theta, C) \) because:

\[
E\{x\} = E\{H\theta + w\} = H\theta + E\{w\} = H\theta
\]

\[
\text{cov}\{x\} = E\{(x - H\theta)(x - H\theta)^T\} = E\{w w^T\} = C
\]
Recalling CRLB Theorem…  Look at the partial of LLF:

\[
\frac{\partial \ln p(x; \theta)}{\partial \theta} = -\frac{1}{2} \frac{\partial}{\partial \theta} \left[ (x - H\theta)^T C^{-1} (x - H\theta) \right]
\]

\[
= -\frac{1}{2} \frac{\partial}{\partial \theta} \left[ x^T C^{-1} x - 2x^T C^{-1} H\theta + \theta^T H^T C^{-1} H\theta \right]
\]

Constant w.r.t. \( \theta \)
Linear w.r.t. \( \theta \)
Quadratic w.r.t. \( \theta \)
(Note: \( H^T C^{-1} H \) is symmetric)

Now use results in “Gradients and Derivatives” posted on BB:

\[
\frac{\partial \ln p(x; \theta)}{\partial \theta} = -\frac{1}{2} \left[ -2H^T C^{-1} x + 2H^T C^{-1} H\theta \right] = \left[ H^T C^{-1} x - H^T C^{-1} H\theta \right]
\]

\[
= \frac{H^T C^{-1} H}{I(\theta)} \left[ (H^T C^{-1} H)^{-1} H^T C^{-1} x - \theta \right]
\]

Pull out \( H^T C^{-1} H \)

The “CRLB Theorem” says that if we have this form we have found the MVU and it achieves the CRLB of \( I^{-1}(\theta)!! \)
For simplicity… assume $b = 0$

**Whitening Filter Viewpoint**

Assume $C$ is positive definite (necessary for $C^{-1}$ to exist)

Thus, from (A1.2): for pos. def. $C \ni N \times N$ invertible matrix $D$, s.t.

$$C^{-1} = D^T D$$

$$C = D^{-1}(D^T)^{-1}$$

Transform data $x$ using matrix $D$: $	ilde{x} = Dx = DH\theta + Dw = \tilde{H}\theta + \tilde{w}$

$$E(\tilde{w}\tilde{w}^T) = E((Dw)(Dw)^T) = E(Dww^T D^T)$$

$$= DCD^T = D \left( D^{-1}(D^T)^{-1} \right) D^T = I$$

**Claim:** White!!

![Diagram](image-url)
Ex. 4.1: Curve Fitting

**Caution**: The “Linear” in “Linear Model” does **not** come from fitting **straight lines** to data.

It is more general than that!!

\[ x[n] = \theta_1 + \theta_2 n + \theta_3 n^2 + w[n] \]

Data

Model is **Quadratic in Index** \( n \)…
But Model is **Linear in Parameters**

\[ x = H\theta + w \]

\[ \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} \quad H = \begin{bmatrix} 1 & 1 & 1^2 \\ 1 & 2 & 2^2 \\ \vdots & \vdots & \vdots \\ 1 & N & N^2 \end{bmatrix} \]
Ex. 4.2: Fourier Analysis (not most general)

**Data Model:**
\[ x[n] = \sum_{k=1}^{M} a_k \cos \left( \frac{2\pi kn}{N} \right) + \sum_{k=1}^{M} b_k \sin \left( \frac{2\pi kn}{N} \right) + w[n] \]

**Parameters:**
\[ \theta = [a_1 \ldots a_M \ b_1 \ldots b_M]^T \] (Fourier Coefficients)

**Observation Matrix:**
\[ H = \begin{bmatrix} \cos \left( \frac{2\pi kn}{N} \right) & \sin \left( \frac{2\pi kn}{N} \right) \\ \vdots & \vdots \\ \cos \left( \frac{2\pi kM}{N} \right) & \sin \left( \frac{2\pi kM}{N} \right) \end{bmatrix} \]

\[ n = 0, 1, 2, \ldots, N \]
Down each column
Now apply MVUE Theorem for Linear Model:

\[ \hat{\theta}_{MVU} = \left( H^T H \right)^{-1} H^T x \]

Using standard orthogonality of sinusoids (see book)

\[ \hat{\theta}_{MVU} = \frac{N}{2} H^T x \]

Each Fourier coefficient estimate is found by the inner product of a column of \( H \) with the data vector \( x \)

Interesting!!! Fourier Coefficients for signal + AWGN are MVU estimates of the Fourier Coefficients of the noise-free signal

**COMMENT: Modeling and Estimation (are Intertwined)**
- Sometimes the parameters have some physical significance (e.g. delay of a radar signal).
- But sometimes parameters are part of non-physical assumed model (e.g. Fourier)
- Fourier Coefficients for signal + AGWN are MVU estimates of the Fourier Coefficients of the noise-free signal
Ex. 4.3: System Identification

\[ u[n] \rightarrow H(z) \rightarrow + \rightarrow x[n] \]

Known Input \quad Unknown System \quad w[n] \quad Observed Noisy Output

Goal: Determine a model for the system

Some Application Areas:
- Wireless Communications (identify & equalize multipath)
- Geophysical Sensing (oil exploration)
- Speakerphone (echo cancellation)

In many applications: assume that the system is FIR (length \( p \))

\[ x[n] = \sum_{k=0}^{p-1} h[k] u[n-k] + w[n] \]

Measured \quad Estimation Parameters \quad Known Input Assume \( u[n] = 0, n < 0 \) \quad unknown, but here we’ll assume known

AWGN
Write FIR convolution in matrix form:

\[
x = \begin{bmatrix}
    u[0] & 0 & 0 & \ldots & \ldots & 0 \\
    u[1] & u[0] & 0 & \ldots & \ldots & 0 \\
    u[2] & u[1] & u[0] & 0 & \ldots & 0 \\
    \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
    \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\
    \vdots & \vdots & \vdots & \vdots & \vdots & u[0] \\
    \vdots & \vdots & \vdots & \vdots & u[1] & \vdots \\
    u[N-1] & \ldots & \ldots & \ldots & \ldots & u[N-p]
\end{bmatrix}
\begin{bmatrix}
    h[0] \\
    h[1] \\
    \vdots \\
    h[p-1]
\end{bmatrix} + w
\]

The Theorem for the Linear Model says:

\[
\hat{\theta}_{MVU} = \left( H^T H \right)^{-1} H^T x
\]

\[
\sigma^2 \left( H^T H \right)^{-1}
\]

and achieves the CRLB.
Q: What signal $u[n]$ is best to use?

A: The $u[n]$ that gives the smallest estimated variances!!

Book shows: Choosing $u[n]$ s.t. $H^TH$ is diagonal will minimize variance

⇒ Choose $u[n]$ to be **pseudo-random noise (PRN)**

$u[n]$ is $\perp$ to all its shifts $u[n - m]$

Proof uses: $C_\theta = \sigma^2 (H^T H)^{-1}$

And Cauchy-Schwarz Inequality (same as Schwarz Ineq.)

**Note:** PRN has approximately flat spectrum

So from a frequency-domain view a PRN signal equally probes at all frequencies