Chapter 4
Linear Models

General Linear Model

Recall signal + WGN case: $x[n] = s[n;\theta] + w[n]$

$$\mathbf{x} = \mathbf{s}(\theta) + \mathbf{w}$$

Here, dependence on θ is general

Now we consider a special case: **Linear "Observations"**:

$$\underbrace{\mathbf{s}(\theta)}_{N\times 1} = \underbrace{\mathbf{H}}_{\underline{\text{known}}\text{ “observation matrix” } (N\times p)}\underbrace{\theta}_{p\times 1} + \underbrace{\mathbf{b}}_{\underline{\text{known}}\text{ “offset”}(p\times 1)}$$

The General Linear Model:

$$\boxed{\underbrace{\mathbf{x}}_{\text{Data Vector}} = \underbrace{\mathbf{H}}_{\text{Known \& Full Rank}}\underbrace{\theta}_{\text{To Be Estimated}} + \underbrace{\mathbf{b}}_{\text{Known}} + \underbrace{\mathbf{w}}_{\sim N(\mathbf{0},\mathbf{C})}}$$

$\mathbf{w}$: zero-mean, Gaussian, $\mathbf{C}$ is pos. def.

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** : $\exists\ \theta_1, \theta_2$ such that $\mathbf{s} = \mathbf{H}\theta_1 = \mathbf{H}\theta_2$

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

$$\boxed{\hat{\boldsymbol{\theta}}_{MVU} = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}^{-1}\left(\mathbf{x}-\mathbf{b}\right)} \quad \text{… as we'll see!!!}$$

MVUE for Linear Model

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

$$\boxed{\hat{\boldsymbol{\theta}}_{MVU} = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}^{-1}\left(\mathbf{x}-\mathbf{b}\right)}$$

$$\boxed{\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}}$$ and achieves the CRLB.

Proof: We'll do this for the $\mathbf{b} = \mathbf{0}$ case but it can easily be done for the more general case.

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

$$E\{\mathbf{x}\} = E\{\mathbf{H}\theta + \mathbf{w}\} = \mathbf{H}\theta + E\{\mathbf{w}\} = \mathbf{H}\theta$$

$$\text{cov}\{\mathbf{x}\} = E\{(\mathbf{x} - \mathbf{H}\theta)\,(\mathbf{x} - \mathbf{H}\theta)^{\mathrm{T}}\} = E\{\mathbf{w}\,\mathbf{w}^{\mathrm{T}}\} = \mathbf{C}$$

Recalling CRLB Theorem… Look at the partial of LLF:

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

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

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

Constant w.r.t. θ

Linear w.r.t. θ

Quadratic w.r.t. θ
(Note: $\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}$ is symmetric)

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

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

Pull out
$\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}$

$$= \underbrace{\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}}_{\mathbf{I}(\boldsymbol{\theta})}\left[\underbrace{\left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}}_{g(\boldsymbol{\theta})=\hat{\boldsymbol{\theta}}} - \boldsymbol{\theta}\right]$$

The “CRLB Theorem” says that if we have this form we have found the MVU <u>and</u> it achieves the CRLB of $\mathbf{I}^{-1}(\theta)$!!

Whitening Filter Viewpoint

For simplicity… assume $\mathbf{b} = \mathbf{0}$

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

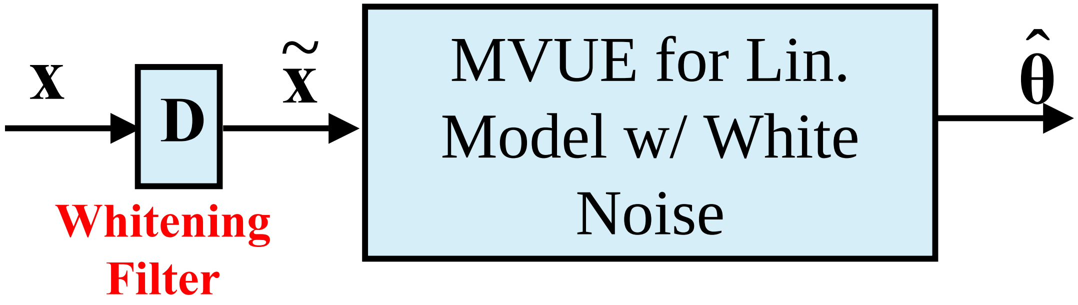
Thus, from (A1.2): for pos. def. $\mathbf{C}$ $\exists$ $N\times N$ invertible matrix $\mathbf{D}$, s.t.

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

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

$$\boxed{\begin{aligned} E\{\tilde{\mathbf{w}}\tilde{\mathbf{w}}^T\} &= E\{(\mathbf{Dw})(\mathbf{Dw})^T\} = E\{\mathbf{Dww}^T\mathbf{D}^T\} \\ &= \mathbf{DCD}^T = \mathbf{D}\left(\mathbf{D}^{-1}\left(\mathbf{D}^T\right)^{-1}\right)\mathbf{D}^T = \mathbf{I} \end{aligned}}$$

Claim: White!!



$\mathbf{x}$

D

$\tilde{\mathbf{x}}$

MVUE for Lin. Model w/ White Noise

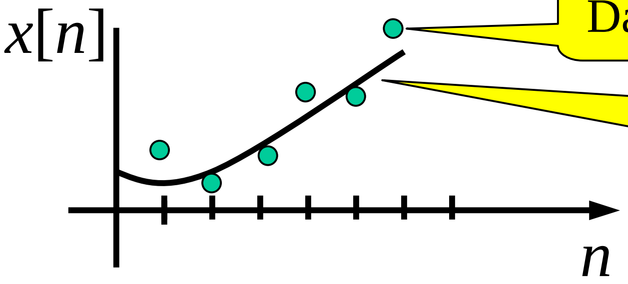
$\hat{\boldsymbol{\theta}}$

Whitening Filter

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 !!



Data

Model is Quadratic in Index n…
But Model is Linear in Parameters

$$\boxed{x[n] = \underbrace{\theta_1 + \theta_2 n + \theta_3 n^2}_{\text{Linear in } \theta\text{'s}} + w[n]} \Rightarrow \boxed{\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} \qquad \mathbf{H} = \begin{bmatrix} 1 & 1 & 1^2 \\ 1 & 2 & 2^2 \\ \vdots & \vdots & \vdots \\ 1 & N & N^2 \end{bmatrix}$$

Ex. 4.2: Fourier Analysis (not most general)

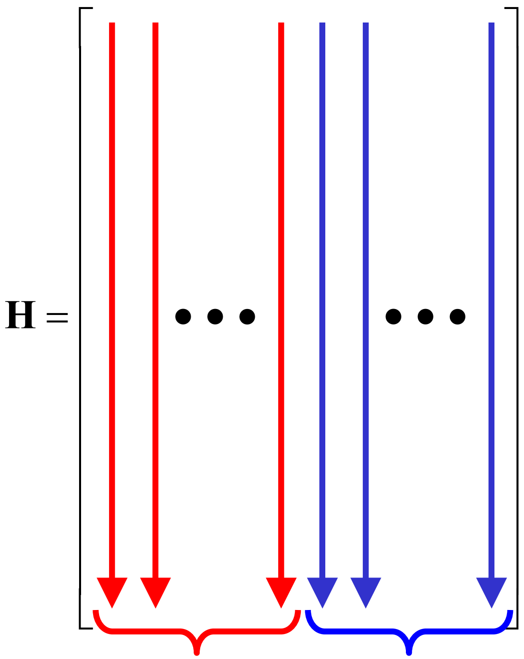
Data Model: $$x[n] = \sum_{k=1}^{M} a_k \cos\left(\frac{2\pi kn}{N}\right) + \sum_{k=1}^{M} b_k \sin\left(\frac{2\pi kn}{N}\right) + w[n]$$

Parameters to Estimate

AWGN

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

Observation Matrix:



$$\cos\left(\frac{2\pi kn}{N}\right) \qquad \sin\left(\frac{2\pi kn}{N}\right) \qquad k = 1, 2, \ldots, M$$

$n = 0, 1, 2, \ldots, N$
Down each column

Now apply MVUE Theorem for Linear Model:

$$\hat{\theta}_{MVU} = \big(\underbrace{\mathbf{H}^T\mathbf{H}}_{=\frac{N}{2}\mathbf{I}}\big)^{-1}\mathbf{H}^T\mathbf{x} \quad \Longrightarrow \quad \boxed{\hat{\theta}_{MVU} = \frac{N}{2}\mathbf{H}^T\mathbf{x}}$$

Using standard orthogonality of sinusoids (see book)

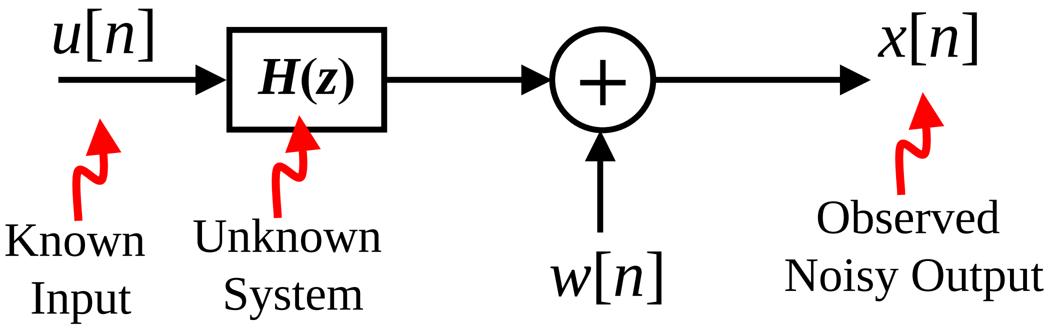
Each Fourier coefficient estimate is found by the inner product of a column of **H** with the data vector **x**

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



$u[n]$ $H(z)$ $x[n]$

Known Input

Unknown System

$w[n]$

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*)

unknown, but here we’ll assume known

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

Measured

Estimation Parameters

Known Input
Assume $u[n] = 0,\ n < 0$

AWGN

Write FIR convolution in matrix form:

$$\mathbf{x} = \underbrace{\begin{bmatrix} u[0] & 0 & 0 & \cdots & \cdots & 0 \\ u[1] & u[0] & 0 & \cdots & \cdots & 0 \\ u[2] & u[1] & u[0] & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ \vdots & & & & \ddots & u[0] \\ \vdots & & & & & u[1] \\ \vdots & & & & & \vdots \\ u[N-1] & \cdots & \cdots & \cdots & \cdots & u[N-p] \end{bmatrix}}_{\mathbf{H}\,(Nxp)} \underbrace{\begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[p-1] \end{bmatrix}}_{\boldsymbol{\theta}} + \mathbf{w}$$

Measured Data

Known Input Signal Matrix

Estimate This

The Theorem for the Linear Model says:

$$\boxed{\hat{\boldsymbol{\theta}}_{MVU} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}}$$

$$\boxed{\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \sigma^2\left(\mathbf{H}^T\mathbf{H}\right)^{-1}}$$

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. $\mathbf{H}^T\mathbf{H}$ is diagonal will minimize variance

$\Rightarrow$ Choose *u*[*n*] to be **pseudo-random noise (PRN)**
u[*n*] is $\perp$ to all its shifts $u[n-m]$

Proof uses: $\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \sigma^2 \left(\mathbf{H}^T\mathbf{H}\right)^{-1}$

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