ESE 524: Detection and Estimation Theory

Recitation 3

Washington University in St. Louis

Outline

- Linear models
- Maximum-likelihood estimation

Useful Formulas

- Linear model:

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

- Minimum variance unbiased (MVU) estimator:

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

- Best linear unbiased estimator (BLUE) and minimum variance unbiased (MVU) estimator under colored noise:

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

Useful Formulas (Cont.)

- Maximum likelihood:

$$\hat{\boldsymbol{\theta}} = \underset{\theta}{\operatorname{argmax}}\, p(\boldsymbol{x}; \boldsymbol{\theta})$$

- Asymptotic distribution of the maximum likelihood estimator:

$$\lim_{N \to \infty} \sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathcal{N}(\mathbf{0}, N\mathcal{I}(\boldsymbol{\theta}))$$

System Identification

- System identification focuses on using statistical methods to build models of dynamical systems.
- Most of the time the quantitative information is usually output and/or input data.
- Sometimes have assumptions from prior information on type of model or a model from physics.
- System identification is also concerned with how to design experiments that best measure the input/outputs.

Example: Finite Impulse Response (FIR) Filter

- **Goal:** Estimate a linear model given input and output data. Linear models are completely controlled by their impulse response.
- Assume a Finite Impulse Response model, with p terms.
- $\boldsymbol{\theta} = \Big[h[0], \quad h[1], \quad ..., \quad h[p-1]\Big]^T$
- Let $u[n]$ be an input function. u can be arbitrary but in general $u[n] = 0$ for $n < 0$.
- Let $w[n] \sim \mathcal{N}(0, \sigma^2)$ be the usual white noise, and $\boldsymbol{w}$ be the vector of i.i.d. samples from the noise distribution.

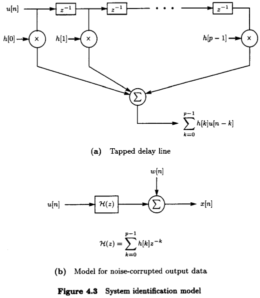

Figure 1: Problem 4.3 from Kay (a): General FIR linear model block diagram. (b): Adding noise to the FIR system in (a).

Converting This to a Linear Model

- The output signal is given by the convolution of the input u with the impulse response $\boldsymbol{\theta}$, added with white noise:

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

- Since $w[n]$ are i.i.d. $x[n]$ are all independent of each other.
- To construct a linear model, look at a couple of examples:

$$x[0] = \Sigma_{k=0}^{p-1} h[k]u[0-k] + w[0] = h[0]u[0] + h[1]u[-1] + ... + w[0]$$

$$= h[0]u[0] + w[0]$$

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

$$x[2] = \Sigma_{k=0}^{p-1} h[k]u[2-k] + w[2] = h[0]u[2] + h[1]u[1] + h[2]u[0] + w[1]$$

Converting This to a Linear Model (Cont.)

- Following this pattern we can find the linear model form:

$$\begin{bmatrix} x[0] \\ x[1] \\ . \\ . \\ . \\ x[N-1] \end{bmatrix} = \underbrace{\begin{bmatrix} u[0] & 0 & 0 & ... & 0 \\ u[1] & u[0] & 0 & ... & 0 \\ u[2] & u[1] & u[0] & ... & 0 \\ . & . & . & . & . \\ u[N-1] & u[N-2] & u[N-3] & ... & u[N-p] \end{bmatrix}}_{\boldsymbol{H}} \underbrace{\begin{bmatrix} h[0] \\ h[1] \\ h[2] \\ . \\ . \\ . \\ h[p] \end{bmatrix}}_{\boldsymbol{\theta}} + \boldsymbol{w}$$

- Using theorem 1 the MVU estimator of the impulse response is:

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

- The variances of the estimates are the diagonal entries of

$$C_{\hat{\boldsymbol{\theta}}} = \sigma^2 (\boldsymbol{H}^T \boldsymbol{H})^{-1}$$

Matlab Example

- Fix $p = 10$, $N = 100$, and $\sigma^2 = 1$
- Try several different input functions

$$
\begin{aligned}
u_1[n] &= 1 \quad \text{for} \quad n > 0 \\
u_2[n] &= \cos(2\pi n/20) \\
u_3[n] &= \delta[n] \\
u_4[n] &= e[n] \sim N(0, 2)
\end{aligned}
$$

Input Functions

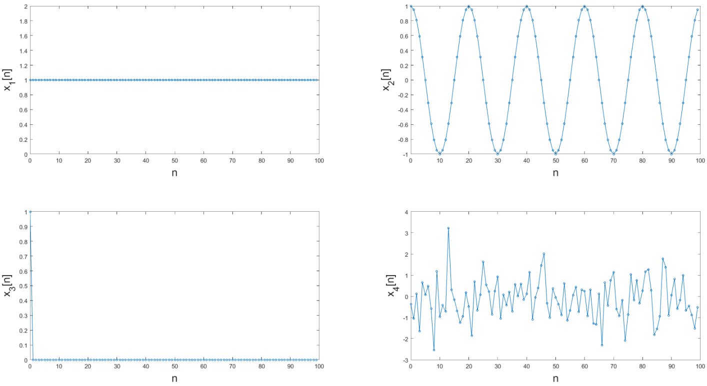

Figure 2: The four input functions.

Outputs

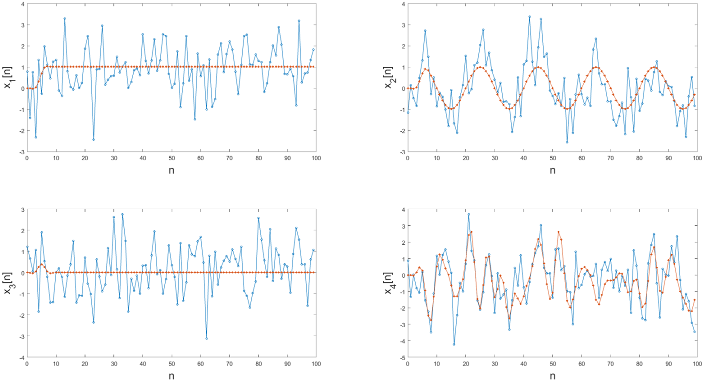

Figure 3: The four output functions, noisy signals are shown in blue, output without noise is shown in red.

Which Input Yields the Best Estimator?

- True coefficients:
  $0, -0.013, -0.025, 0.06, 0.27, 0.39, 0.27, 0.06, -0.026, -0.013, 0$
- MSE of Unit Step Function Based Estimator - 1.62
- MSE of Cosine Based Estimator - 3.27
- MSE of Dirac Delta Based Estimator - 1.34
- MSE of Random Noise Based Estimator - 0.002!

Why is Random Noise the Best Input?

- MacWilliams and Sloane showed that pseudo-random noise is the best we can do. - This is a lengthy derivation given in the Kay example.
- But examining our example we can look at the average value of the diagonals of $(\boldsymbol{H}^T\boldsymbol{H})^{-1}$ for each case.
- Unit step - 1.82
- Cosine - 1.73
- Dirac Delta - 1
- Random Noise - 0.0034.
- Random noise has the smallest entries.

The Information Matrix

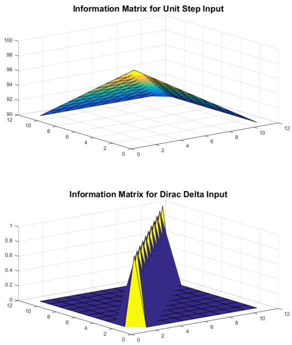

Figure 4: Visualization of information matrix resulting from each input.

Note - the Dirac Delta input also yields a diagonal information matrix. But it's peak is much lower so the information contained is less useful.

Properties of the Information Matrix $H^T H$

- The ij^th entry of $\boldsymbol{H}^T\boldsymbol{H}$ is given by

$$[\boldsymbol{H}^T\boldsymbol{H}]_{ij} = \Sigma_{n=0}^{N-1} u[n-i]u[n-j]$$

- For large N this becomes

$$[\boldsymbol{H}^T\boldsymbol{H}]_{ij} = \Sigma_{n=0}^{N-1-|i-j|} u[n]u[n+|i-j|]$$

- This represents the autocorrelation of u.
- White noise is uncorrelated with itself, this means that most of the terms in this sum will be very close to 0 except for the diagonal entries.
- From an earlier class - it is a good rule of thumb to have a diagonal Information Matrix - random noise decouples every coefficient in the impulse response from the others.

Maximum Likelihood Encoder

A more in-depth look at the example on page 36-37 in lecture 3.

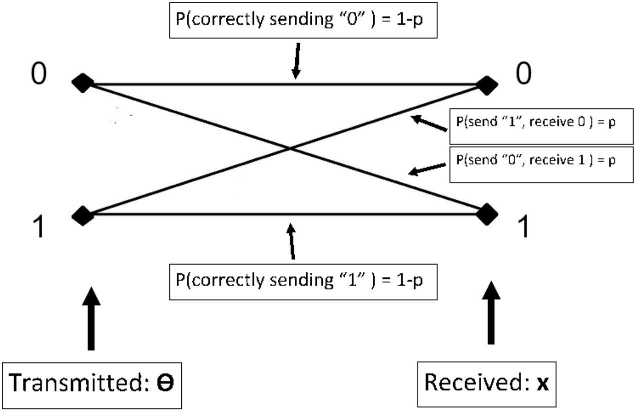

Figure 5: Diagram of the communications channel. The probability of sending the wrong message is p.

Setting Up the Model

- The transmitted signal is given by $\theta[n] \in \{0, 1\}$, and denote the vector of bits sent as $\boldsymbol{\Theta}$
- The channel noise $w[n] \in \{0, 1\}$ is represented by i.i.d. Bernoulli variables. The pmf of an individual $w[n]$ is

$$P(W = w[n]) = p^{w[n]}(1 - p)^{1 - w[n]}$$

  (Note: this pmf will appear in your homework)
- $w[n] = 1$ represents that an error has occurred in the channel transmission, because it always flips the transmitted signal.
- The received signal is

$$x[n] = \theta[n] \bigoplus w[n],$$

  where $\bigoplus$ denotes addition modulo 2. Denote the vector of independent variables as $\boldsymbol{x}$

Review of Modulo Addition

- Modulo 2 addition works by the formula $x \bigoplus y = \text{remainder}(\frac{x+y}{2})$.
- If the addition results in an even number, the modulo is 0, and the modulo is 1 when the addition is an odd number.
- In this example it is also equivalent to binary addition with no carry bit.
- **Key Formula**: $x[n] = \theta[n] \bigoplus w[n] \rightarrow w[n] = x[n] \bigoplus \theta[n]$.
- To check this, take all four possible cases.
  1. $\theta[n] = 0,\ w[n] = 0\,, x[n] = \theta[n] \bigoplus w[n] = 0 \bigoplus 0 = 0 \rightarrow x[n] \bigoplus \theta[n] = 0 \bigoplus 0 = 0 = w[n]$
  2. $\theta[n] = 1,\ w[n] = 0\,, x[n] = \theta[n] \bigoplus w[n] = 1 \bigoplus 0 = 1 \rightarrow x[n] \bigoplus \theta[n] = 1 \bigoplus 1 = 0 = w[n]$
  3. $\theta[n] = 0,\ w[n] = 1\,, x[n] = \theta[n] \bigoplus w[n] = 0 \bigoplus 1 = 1 \rightarrow x[n] \bigoplus \theta[n] = 1 \bigoplus 0 = 1 = w[n]$
  4. $\theta[n] = 1,\ w[n] = 1\,, x[n] = \theta[n] \bigoplus w[n] = 1 \bigoplus 1 = 0 \rightarrow x[n] \bigoplus \theta[n] = 0 \bigoplus 1 = 0 = w[n]$

Creating the Likelihood Function

- $p(\boldsymbol{x};\boldsymbol{\Theta}) = P(\boldsymbol{X} = \boldsymbol{x}) = P(\boldsymbol{\Theta} \bigoplus \boldsymbol{W})$

$$= P(\boldsymbol{W} = \boldsymbol{x} \bigoplus \boldsymbol{Theta})$$

- But $\boldsymbol{W}$ is a vector of i.i.d. Bernoulli random variables, so we know the pmf.
-

$$P(\boldsymbol{W} = \boldsymbol{x} \bigoplus \boldsymbol{\Theta}) = \Pi_{n=0}^{N-1} p^{w[n]}(1-p)^{1-w[n]} =$$

$$p^{\Sigma_{n=0}^{N-1} x[n] \bigoplus \theta[n]}(1-p)^{N-\Sigma_{n=0}^{N-1} x[n] \bigoplus \theta[n]}$$

- Now pull out the parts depending on $\theta[n]$:

$$P(\boldsymbol{W} = \boldsymbol{x} \bigoplus \boldsymbol{\theta}) = (1-p)^N (\frac{p}{1-p})^{\Sigma_{n=0}^{N-1} x[n] \bigoplus \theta[n]}$$

Maximizing the Likelihood Function

- In communications systems, $p < .5$.
- This implies that $(\frac{p}{1-p}) < 1$.
- To maximize this term, minimize $\Sigma_{n=0}^{N-1} x[n] \bigoplus \theta[n]$.
- This term is called the "Hamming distance".
- The next problem is to find a sequence $\theta[n]$ that minimizes the Hamming Distance.

Comments on the Hamming Distance

- Named for Richard Hamming, who invented the formula in order to construct an error correcting coding system.
- Using the Hamming distance, errors in 2-bit communications can be detected. Errors in 1 bit communications can even be corrected.
- Used to compare strings in text analysis to compare words of the same length.
- Used to compare gene codes in biology.

Fourier Transform (Kay example 4.2)

- In many applications, the signal is described as a Fourier Series. For example, cell phones estimate the Fourier coefficients from your voice for 10-20 frequencies, and then send those numbers to the cell tower and out into the world.
- In this case the signal model is given as

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

- Here $w[n] \sim \mathcal{N}(0, \sigma_2)$ are the usual i.i.d. samples of white gaussian noise.
- Denote $\boldsymbol{x} = \begin{bmatrix} x[0], & x[1], & \ldots, & x[n] \end{bmatrix}^T$.
- The vector of parameters to be estimated is $\boldsymbol{\Theta} = \begin{bmatrix} a_1, & a_2, & \ldots, & a_M, & b_1, & b_2, & \ldots, & b_M \end{bmatrix}^T$

Linear Model Formulation

- Define the model matrix $\boldsymbol{H}$ as:

$$\begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ \cos(\frac{2\pi}{N}) & \dots & \cos(\frac{2\pi M}{N}) & \sin(\frac{2\pi}{N}) & \dots & \sin(\frac{2\pi M}{N}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \cos(\frac{2\pi(N-1)}{N}) & \dots & \cos(\frac{2\pi M(N-1)}{N}) & \sin(\frac{2\pi(N_1)}{N}) & \dots & \sin(\frac{2\pi M(N-1)}{N}) \end{bmatrix}$$

- The columns of $H = \begin{bmatrix} \boldsymbol{h}_1 & \boldsymbol{h}_2 & \cdots & \boldsymbol{h}_{2M} \end{bmatrix}$ are given by either $\cos(\frac{2\pi i n}{N})$ or $\sin(\frac{2\pi i n}{N})$ for $i = 1, 2, \dots, M$ and $n = 0, 1, \dots, N-1$
- The linear model form of the Fourier series estimation problem is given by

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

The Information Matrix

- Recall from last time that the information matrix is given by $\frac{1}{\sigma^2}\boldsymbol{H}^T\boldsymbol{H}$.
- $\boldsymbol{H}^T\boldsymbol{H} = \begin{bmatrix} \boldsymbol{h}_1^T \\ \vdots \\ \boldsymbol{h}_{2M}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{h}_1 & \dots & \boldsymbol{h}_{2M} \end{bmatrix} =$

$$\begin{bmatrix} \boldsymbol{h_1^t h_1} & \boldsymbol{h_1^T h_2} & \dots & \boldsymbol{h_1^T h_{2M}} \\ \boldsymbol{h_2^t h_1} & \boldsymbol{h_2^T h_2} & \dots & \boldsymbol{h_2^T h_{2M}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{h_{2M}^t h_1} & \boldsymbol{h_{2M}^T h_2} & \dots & \boldsymbol{h_{2M}^T h_{2M}} \end{bmatrix}$$

- There are three possible cases here.

Case 1

- Case 1: $\boldsymbol{h}_i^T \boldsymbol{h}_j = \Sigma_{n=0}^{N-1} \cos(\frac{2\pi i n}{N}) \cos(\frac{2\pi j n}{N})$
- If $i = j$, then:

$$\Sigma_{n=0}^{N-1} \cos(\frac{2\pi i n}{N})^2 = \Sigma_{n=0}^{N=N-1} \frac{1}{2} + \frac{1}{2} \cos(\frac{4\pi i n}{N}) =$$
$$\frac{N}{2} + \frac{1}{2} \Sigma_{n=0}^{N-1} \cos(\frac{4\pi i n}{N}) = \frac{N}{2} + 0$$

- For a proof of the final sum, use "Lagrange's Trigonometric Identities". The proof of these identities expands the functions in terms of complex exponentials.
- If $i \neq j$ then:

$$\Sigma_{n=0}^{N-1} \cos(\frac{2\pi i n}{N}) \cos(\frac{2\pi j n}{N})$$
$$= \frac{1}{2} \Sigma_{n=0}^{N-1} \cos(\frac{2\pi (i+j) n}{N}) + \cos(\frac{2\pi (i-j) n}{N}) = 0$$

using the same identity as before for the finite sum of cosines.

Cases 2 and 3

- Case 2: $\boldsymbol{h}_i^T \boldsymbol{h}_j = \Sigma_{n=0}^{N-1} \sin(\frac{2\pi in}{N}) \sin(\frac{2\pi jn}{N})$
- This case is almost identical with the last case.
- $\boldsymbol{h}_i^T \boldsymbol{h}_j = \begin{cases} \frac{N}{2} \text{ for } i = j \\ 0 \text{ else} \end{cases}$ -$\boldsymbol{h}_i^T \boldsymbol{h}_j = \Sigma_{n=0}^{N-1} \sin(\frac{2\pi in}{N}) \sin(\frac{2\pi jn}{N})$
- Case 3 is the sum $\boldsymbol{h}_i^T \boldsymbol{h}_j = \Sigma_{n=0}^{N-1} \cos(\frac{2\pi in}{N}) \sin(\frac{2\pi jn}{N})$
- This reduces to a similar expression as Case 1.

$$\frac{1}{2}\Sigma_{n=0}^{N-1} \sin\left(\frac{2\pi(i+j)n}{N}\right) - \sin\left(\frac{2\pi(i-j)n}{N}\right) = 0$$

MVU Estimator

- Using trig identities, we can show that $\boldsymbol{H}^T\boldsymbol{H} = \frac{N}{2}I$, where I is the identity matrix.
- Then the optimal least squares estimator of the Fourier coefficients is

$$\hat{\boldsymbol{\Theta}} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{x} = \frac{2}{N}\begin{bmatrix} \boldsymbol{h}_1^T\boldsymbol{x} \\ \boldsymbol{h}_2^T\boldsymbol{x} \\ \vdots \\ \boldsymbol{h}_{2M}^T\boldsymbol{x} \end{bmatrix}$$

- The specific coefficients are

$$\hat{a}_k = \frac{2}{N}\Sigma_{n=0}^{N-1}x[n]\cos(\frac{2\pi kn}{N})$$

$$\hat{b}_k = \frac{2}{N}\Sigma_{n=0}^{N-1}x[n]\sin(\frac{2\pi kn}{N})$$

- Congratulations, we have re-invented the Discrete Fourier Transform (DFT)!