System Identification

- System Identification focuses on using statistical methods to build models of dynamical systems.
- Generally, the known information are the input signal $\mathbf{u} = [u[0], u[1], \cdots, u[N-1]]$ and output measurements $\mathbf{x} = [x[0], x[1], \cdots, x[N-1]]$.
- Sometimes have assumptions from prior information on the type of model or equations from physics.
- System Identification is also concerned with how to design experiments that best measure the input/outputs.

Example: FIR Filter

- Example 4.3 from Kay.
- **Goal:** Estimate a Linear, Time Invariant System given input and output data.
- The output of a linear system is entirely governed by their Impulse Response, so estimating this is our goal.
- Assume a Finite Impulse Response model, with p terms.
- $\theta = \left[h[0], \quad h[1], \quad \ldots \quad h[p-1]\right]^T$
- The input $\mathbf{u}$ is arbitrary but in general $u[n] = 0$ for $n < 0$.
- Let $w[n] \sim N(0, \sigma^2)$ be the usual white noise, and $\mathbf{w}$ be the vector of i.i.d samples from the noise distribution.

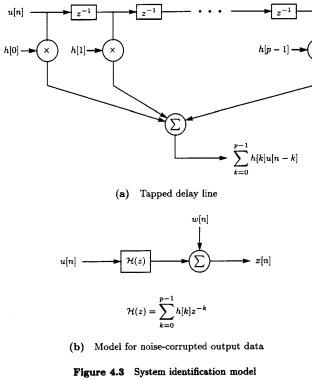

Figure 1: (a): General FIR linear model block diagram. (b): Adding noise to the FIR system in (a).

Converting this to a Linear Model

- The output signal is given by the **convolution** of the input u with the impulse response θ, added with white noise:

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

Converting this to a Linear Model

- The output signal is given by the **convolution** of the input u with the impulse response θ, 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, so we can use the linear model formulation.

Converting this to a Linear Model

- The output signal is given by the **convolution** of the input u with the impulse response θ, 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, so we can use the linear model formulation.
- To construct a linear model, look at a couple of examples:

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

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

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

Converting this to a Linear Model cont.

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

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

Converting this to a Linear Model cont.

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

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

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

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

Converting this to a Linear Model cont.

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

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

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

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

- The variances of the estimates are the diagonal entries of

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

Matlab example

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

$$u_1[n] = 1 \quad \text{for} \quad n > 0$$

$$u_2[n] = \cos(2 * \pi * n/20)$$

$$u_3[n] = \delta[n]$$

$$u_4[n] = e[n] \sim N(0, 2)$$

Input Functions

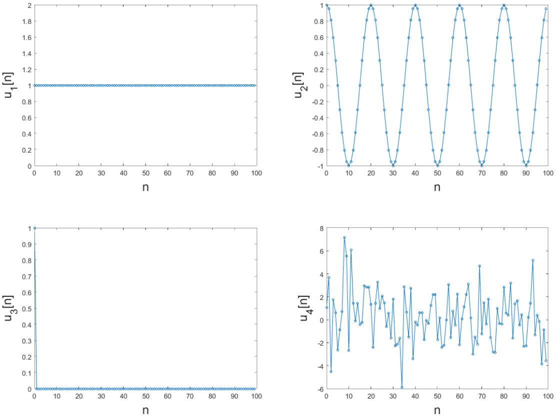

Figure 2: Four candidate input functions to determine the FIR coefficients: Unit Step, Cosine Wave, Dirac Delta, and White Noise

Outputs

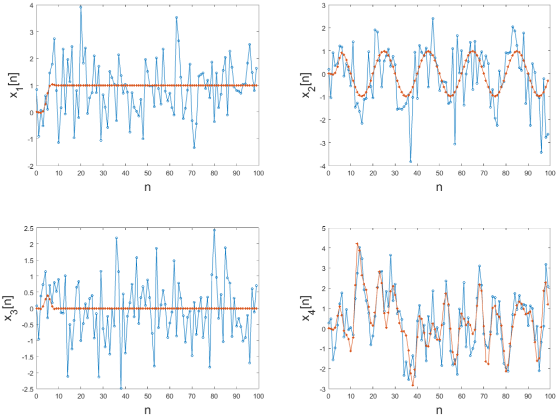

Figure 3: The four output functions, noisy signals are shown in blue, output without noise is shown in red

Which input yields the best estimator?

- True coefficient values:

$$[0, -0.013, -0.025, 0.27, 0.39, 0.27, 0.06, -0.026, -0.013, 0]$$

- 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 a random noise input the best?

- 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 $(\mathbf{H}^T\mathbf{H})^{-1}$ for each case. - This is the average CRB 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

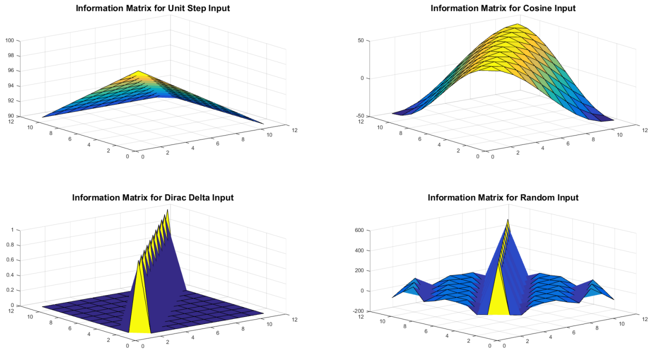

Figure 4: Visualization of information matrix resulting from each input.

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 for estimation.

Properties of the information Matrix $\mathbf{H}^T\mathbf{H}$

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

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

- For large N this becomes

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

- 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.