Day 4: Linear Filtering and Stochastic Processes

Zach Neveu

September 26, 2019

1 Recap

- Last week: general systems + noise
- General system $y = \theta^T * [1x] + \eta$ where η is noise
- y is linear w.r.t θ, affine w.r.t x

2 This week

- η included in input signal, rather than added after the system
- η doesn't have to be IID anymore
- Idea: filter design as a learning problem to minimize a cost function
- Specify cost function in time domain
- $e_n = d_n - \hat{d}_n$ is error function
- If we minimize the MSE, we get a **Wiener Filter**
- $\hat{d}_n = \sum_{i=0}^{l-1} w_i u_{n-i}$ where u_n is a realization of a random variable and w_i is a weight

3 Applications

System Identification

- If we have access to a system but don't know what it is we might want to model it
- Using Wiener filter, possible to learn the impulse response of the system
- Also useful if output is a mixture of two signals and we want to remove one
- $d_n = s_n + y_n$, where $y[n] = H(u[n])$. We can learn $H()$ to subtract s_n from d_n
- Example: echo cancellation
- Benefit of learned IR is that it can change over time for example if someone moves their phone etc.
- Example 2: noise cancellation
- Want to cancel noise in the cabin, can learn IR between engine and cabin

4 Stochastic Processes

- Stationary: RP is stationary if the distribution of the sample at all times is equal
- Useful Statistics about RPs
  - Mean
  - Autocovariance at time instants: covariance of process at one time with process at different time
  - Autocorrelation at time instants: Product of two different time instances
  - Cross-correlation at time instants: measure similarity of two different signals at two time instances
- WSS: wide sense stationary (weaker form of stationary): mean value is constant, and autocorrelation of samples depends only on their difference
- Mean and Covariance Ergotic - Mean and covariance over time are equal to mean and covariance across realizations. Ergotic processe are also WSS

5 Autoregressive Models

- Assume that the next sample depends on the last l samples
- $\sum_{k=1}^{l} a_k u_{n-k} + \eta_n$
- This simple model is quite useful
- AR(1) process - $l = 1$, depends only on previous value
- Notion of causality embedded in this model
- Large $a \rightarrow$ more low frequencies, small $a \rightarrow$ closer to white

6 Normal Equations

- Looking to estimate θ for linear model
- Loss function: $J(\theta) = E[(y - \hat{y})^2]$ (MSE)
- $\theta_* := argmin_\theta J(\theta)$ is MSE optimal θ (not $\hat{\theta}$ because if properties of signal are known, this is known to be optimal)
- To analytically minimize $J(\theta)$, plug in $\hat{y}$, take gradient w.r.t. θ, set equal to 0
- Result: $E[xx^T]\theta = E[xy]$
- Left side of this is auto correlation of x, right side is cross-correlation of x,y
- $\theta_* = E[xy] * E[xx^T]^{-1}$

- In practice, these values would take infinite time to calculate
- Makes the assumption that auto correlation matrix invertible
- can estimate both matrices from finite amount of data
- Wiener filter: $\theta_* = \Sigma_x^{-1} r_{xy}$ where r_{xy} is cross correlation
- To use Wiener filter we must:
  - Specify desired signal
  - Compute autocorrelation of input signal
  - Compute cross-correlation between input signal and desired signal