Boston University

Department of Electrical and Computer Engineering

SC505 STOCHASTIC PROCESSES

Overall Class Summary

---

1. **Characterization and Manipulation of Random Processes**
   - Tools of stochastic processes
   - Joint pdfs
   - Marginal pdfs
   - Conditional pdfs
   - General Expectation $E[g(x)]$
   - Conditional Expectation $E[g(x)|y]$
   - Moments: Means, autocorrelations, autocovariances, power spectral density
   - Distribution-based properties: Stationarity, wide-sense stationarity, ergodicity, independence, IIP, Markov, etc
   - Moment-based properties: Uncorrelated, orthogonal, etc
   - Functions of random variables (Derived pdfs): Equivalent events have equal probability
   - Random vectors
   - Special random processes and their means and variances: e.g.
     - Gaussian
     - Poisson
     - Exponential
     - etc

2. **Random Signals and Systems**
   - I/O relationships for random processes through linear systems
   - I/O relationships for LTI systems and Stationary processes
   - Time, frequency, and Laplace/z domain expressions
   - Special case: Discrete time linear models and finding system parameters (AR, MA, ARMA)
   - LCCDE descriptions
   - Sampling
   - Noise modeling and special processes: Wiener process (Gaussian noise), Poisson process, White noise,...
   - Spectral Factorization for shaping filter design

3. **Detection Theory/Hypothesis Testing**

- $x = H_i$ Discrete-valued
- Given:
  - Observation Model $p_{Y|X}(y|x = H_i)$
  - Prior Model $p_X(x = H_i)$
  - Costs C_{ij} = Cost of deciding H_i and H_j true
- Find Optimal detection rule to minimize $E(\text{Cost})$.
- Special Case Rules: MAP Rule (MPE=$C_{ij} = 1 - \delta_{ij}$), ML Rule (MPE and $P_i = P_j$), Neyman-Person (max P_D for $P_F \leq \alpha$), etc
- Binary vs multi-valued/M-ary detection (For Binary: ROC, likelihood-ratio test, P_D, P_F, etc)
- Finding performance: $\Pr(\text{Err}) = P_e$ or $E(\text{Cost})$
- Detection of signals – role of matched filter for a signal in white noise
- KL expansions for signal detection, White vs Colored noise
- Special results: Min distance classifier, Gaussian processes, MPE rule, etc.

4. **Estimation Theory**

- General Bayes (Random Parameter) Estimation
  - Setup:
    (a) Parameter Model: $P_X(x)$, Probabilistic Prior Density
    (b) Observation Process: $P_{Y|X}(y \mid x)$, Conditional density
    (c) Costs: $J(\widehat{x}, x)$ = Cost of Estimating $\widehat{x}$ when x True.
  - Estimation Rule: Minimize Expected Cost $\Longrightarrow \widehat{x}(y) = \arg\min_x E[J(\widehat{x}, x)] = \arg\min_x E[J(\widehat{x}, x) \mid y]$
  - Performance Measures: Define error $e \equiv x - \widehat{x}(y)$
    * $E[\text{Cost}] = E[J(\widehat{x}, x)]$
    * Bias: $b \equiv E[e]$. Just a number for Bayes Estimation.
    * Error Covariance: $\Lambda_e \equiv E\left[(e-b)(e-b)^T\right] = E\left[ee^T\right] - bb^T$ Uncertainty in estimate
    * Mean Square Error: $\text{MSE} = E[e^Te] = \text{Tr}\left[[E[ee^T]\right] = \text{Tr}\left[\Lambda_e + bb^T\right]$
- Bayes Least Squares Estimation (BLSE)
  - Cost: $J(\widehat{x}, x) = ||\widehat{x} - x||^2 = ||e||^2 \Longrightarrow$ BLSE is Minimum Mean Square Error Estimate (MSEE)
  - Estimate: $\widehat{x}_B(y) = E[x \mid y]$. $\Longrightarrow$ BLSE Estimate is Conditional Mean
  - Bias: $b = E[x - \widehat{x}_B(y)] = E[x] - E\left[E[x \mid y]\right] = 0$. $\Longrightarrow$ BLSE estimates are unbiased
  - Error Covariance: $\Lambda_B = E\left[(e-0)(e-0)^T\right] = E\left[\Lambda_{x|y}(y)\right]$. Expected value of conditional covariance
  - $E[\text{Cost}] = \text{MSE} = E[e^Te] = \text{Tr}\left\{\Lambda_B\right\} = \text{Tr}\left\{E[\Lambda_{x|y}]\right\}$. Minimum value of MSE over all estimators (linear and nonlinear).
  - Alternate characterization of BLSE
    * $E\left[x - \widehat{x}_B(y)\right] = 0$. Unbiased
    * $E\left\{\left[x - \widehat{x}_B(y)\right] g(y)\right\} = 0, \forall g(\cdot)$. Error orthogonal to any function of the data

– Gaussian Vector Case:

$$\begin{bmatrix} \underline{x} \\ \underline{y} \end{bmatrix} \sim N\left(\begin{bmatrix} \underline{m}_x \\ \underline{m}_y \end{bmatrix}, \begin{bmatrix} \Lambda_x & \Lambda_{xy} \\ \Lambda_{xy}^T & \Lambda_y \end{bmatrix}\right) \implies \begin{array}{rcl} \widehat{\underline{x}}_B(y) & = & \underline{m}_x + \Lambda_{xy}\Lambda_y^{-1}\left(\underline{y} - \underline{m}_y\right) \\ \Lambda_B & = & \Lambda_{x|y} = \Lambda_x - \Lambda_{xy}\Lambda_y^{-1}\Lambda_{xy}^T \\ \text{Cost} & = & \text{MSE} = \text{Tr}\left(\Lambda_B\right) \end{array}$$

<u>Estimate is linear in Gaussian case</u> and $\Lambda_{x|y}$ is not a function of $\underline{y}$

- <u>Bayes Maximum A Posteriori Estimation (MAP)</u>
  - Cost: $J(\widehat{x}, x) = \begin{cases} 1 & |\widehat{x} - x| > \Delta \\ 0 & |\widehat{x} - x| > \Delta \end{cases} \quad \Delta \to 0.$ Uniform Cost
  - Estimate: $\widehat{x}_{MAP}(y) = \arg\max_x p_{X|Y}(x \mid y) = \arg\max_x p_{Y|X}(y \mid x)p_X(x). \implies$ <u>MAP Estimate is Conditional Mode</u>
  - MAP Equation for Estimate: $\left.\dfrac{\partial \ln\left[p_{Y|X}(y \mid x)\right]}{\partial x} + \dfrac{\partial \ln\left[p_X(x)\right]}{\partial x}\right|_{x=\widehat{x}_{MAP}(y)} = 0$
  - Bias: $b = E[x - \widehat{x}_{MAP}(y)] \neq 0$ in general. $\implies$ <u>MAP estimates can be biased</u>
  - MAP Estimation requires knowledge of details of density
- <u>Bayes Linear Least Squares Estimation (BLLSE)</u>
  - BLSE with estimator constrained to have a <u>linear</u> form: $\widehat{x}_L(y) = C\underline{y} + \underline{d}$
  - Estimate: $\widehat{x}_L(y) = \underline{m}_x + \Lambda_{xy}\Lambda_y^{-1}\left(\underline{y} - \underline{m}_y\right)$
  - <u>BLLSE Estimators only require second order properties</u>
  - Bias: $b = E[x - \widehat{x}_L(y)] = 0. \implies$ <u>LLSE estimates are unbiased</u>
  - Error Covariance: $\Lambda_L = E\left[(e-0)(e-0)^T\right] = \Lambda_x - \Lambda_{xy}\Lambda_y^{-1}\Lambda_{xy}^T$.
  - $E[\text{Cost}] = \text{MSE} = E[e^T e] = \text{Tr}\{\Lambda_L\}$. Minimum value of MSE over all <u>linear</u> estimators.
  - Alternate characterization of BLLSE. Unique linear function of y such that:
    * $E\left[x - \widehat{x}_L(y)\right] = 0$. <u>Unbiased</u>
    * $E\left\{\left[x - \widehat{x}_L(y)\right]\underline{y}^T\right\} = 0, \forall g(\cdot)$. <u>Error orthogonal to (linear functions of) the data</u>
- <u>General Nonrandom Parameter Estimation</u>
  - Setup:
    (a) Parameter Model: x is an <u>unknown deterministic</u> parameter
    (b) Observation Process: $P_{Y|X}(y \mid x)$, <u>Parameterized</u> density (aka "likelihood function").
  - Estimation Rule: No general procedure as in Bayes case.
  - Performance Measures: Define error $e(x) \equiv x - \widehat{x}(y)$
    * All are a function of x and not just numbers.
    * Bias: $b(x) = E[e \mid X = x] \equiv E[x - \widehat{x}(y) \mid X = x] = \int [x - \widehat{x}(y)]p_{Y|X}(y \mid x)\, dy$.
    * Error Covariance: $\Lambda_e(x) \equiv E\left[(e - b(x))(e - b(x))^T \mid X = x\right]$
    * Mean Square Error: $\text{MSE}(x) = E[e^T e \mid X = x] = \text{Tr}\left[\left[E[ee^T \mid X = x\right]\right] = \text{Tr}\left[\Lambda_e(x) + b(x)b(x)^T\right]$.
    * <u>Cramer-Rao Estimation Error Covariance Bound</u>

· If $\widehat{x}(y)$ is any unbiased (nonrandom parameter) estimate of x and $\Lambda_e(x)$ its associated estimation error covariance:

$$\Lambda_e(x) \quad \geq \quad \frac{1}{I_Y(x)}, \qquad I_Y(x) = E\left\{ \left[\frac{\partial}{\partial x} \ln p_{Y|X}(y \mid x)\right]^2 \middle| X = x \right\} = -E\left\{ \frac{\partial^2}{\partial x^2} \ln p_{Y|X}(y \mid x) \right.$$

· Any unbiased estimator that achieves the CRB is termed efficient.

- Maximum Likelihood Estimation (Nonrandom parameter)
  - Estimate: $\widehat{x}_{ML}(y) = \arg\max_x P_{Y|X}(y \mid x)$.
  - ML Equation for Estimate: $\left.\frac{\partial \ln \left[p_{Y|X}(y \mid x)\right]}{\partial x}\right|_{x=\widehat{x}_{ML}(y)} = 0 \qquad \Longrightarrow$ Limit of MAP as $\partial p_X(x)/\partial x \to 0$.
  - Performance:
    * If an efficient estimator does exist it is $\widehat{x}_{ML}(y)$ and in this case $\widehat{x}_{ML}(y)$ is the minimum variance, unbiased estimator.
    * If an efficient estimator does not exist, there may be unbiased estimators with lower variances.
  - ML Facts:
    * If $z = g(x) \Longrightarrow \widehat{z}_{ML}(y) = g\left(\widehat{x}_{ML}(y)\right)$
    * As number of observations $N \to \infty$ ML estimate is asymptotically unbiased, efficient, and consistant.

5. **LLSE Estimation of Random Processes based on Random Processes**

- $x(t)$, $y(\tau)$ assumed zero-mean. If not estimate $\tilde{x}(t) = (x(t) - m_x(t))$ based on $\tilde{y}(t) = (y(t) - m_y(t))$
- Linear estimator $\Longrightarrow$ Only need second order properties $K_{xx}(t,\tau)$, $K_{yx}(t,\tau)$, $K_{yy}(t,\tau)$.
- Form of estimator:

$$\text{CT}: \; \widehat{x}(t) = \int_{T_i}^{T_f} h(t,\sigma) y(\sigma)\, d\sigma$$

- Orthogonality conditions for optimal solution $\Longrightarrow$ Wiener-Hopf Equations
- General Wiener-Hopf Equations for optimal estimator:

$$\text{CT:} \quad K_{xy}(t,\tau) \;=\; \int_{T_i}^{T_f} h(t,\sigma)\, K_{yy}(\sigma,\tau)\, d\sigma \qquad \forall \tau \in [T_i,\, T_f]$$

- General Error Covariance:

$$\text{CT:} \quad \Lambda_{LSE}(t) \;=\; K_{xx}(t,t) - \int_{T_i}^{T_f} h(t,\sigma)\, K_{yx}(\sigma,t)\, d\sigma$$

- Discrete-time, finite-length: Same solution as for random vectors (i.e. normal equations) Estimate: $\Lambda_{xy} = h^T \Lambda_{yy}$. Error Variance: $\Lambda_{LSE} = \Lambda_x - h^T \Lambda_{xy^T}$
- Noncausal Wiener Filter:

- LLSE, $x(t)$, $y(t)$ zero mean
- $x(t)$, $y(t)$ Jointly wide-sense stationary
- Observation interval: $T_i = -\infty$, $T_f = +\infty$
- Optimal Estimate: WH equation just a convolution – use transform techniques

$$\text{CT:} \quad H_{nc}(s) = \frac{S_{yx}(s)}{S_{yy}(s)}$$

- Error-Covariance:

$$\text{CT:} \quad \Lambda_{nc} = K_{xx}(0) - \int_{-\infty}^{\infty} h(u)\, K_{yx}(u)\, du$$

or

$$\begin{aligned} \text{CT:} \quad S_{ee}(s) &= S_{xx}(s) - \frac{S_{yx}(s) S_{yx}(-s)}{S_{yy}(s)} \\ S_{ee}(j\omega) &= S_{xx}(j\omega) - \frac{|S_{yx}(j\omega)|^2}{S_{yy}(j\omega)} \\ \Lambda_{nc} &= R_{ee}(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{ee}(j\omega)\, d\omega \end{aligned}$$

- Causal Wiener Filter:
  - LLSE, $x(t)$, $y(t)$ zero mean
  - $x(t)$, $y(t)$ Jointly wide-sense stationary
  - Observation interval: $T_i = -\infty$, $T_f = t$
  - Optimal Estimate: Whiten data first via $W(s)$ then use CWF for white noise $G(s)$.

$$\text{CT:} \quad H_c(s) = W(s)\, G(s) = \underbrace{\frac{1}{S_{yy}^{+}(s)}}_{\text{Whitening Filter}} \; \underbrace{\left\{ \frac{S_{yx}(s)}{S_{yy}^{-}(s)} \right\}_{+}}_{\text{CWF for Innovations}}$$

  - Error-Covariance:

$$\text{CT:} \quad \Lambda_c = K_{xx}(0) - \int_{0}^{\infty} h(\tau)\, K_{yx}(\tau)\, d\tau$$

or

$$\text{CT:} \quad \Lambda_c = K_{xx}(0) - \int_{0}^{\infty} K_{\nu x}^2(\tau)\, d\tau = K_{xx}(0) - \int_{0}^{\infty} g^2(\tau)\, d\tau$$

- Similar expressions for Discrete-time.
- $\Lambda_{nc} \leq \Lambda_c$
- Know important special cases: e.g. $y(t) = x(t) + v(t)$, $x(t) \perp v(t)$.

6. **Recursive Filtering and the Discrete-Time Kalman Filter**

- Basic concept of sequential estimation of a random variable from sequential observations.

- Key idea of recursive estimation is whitening – role of orthogonality relationships
- Kalman-Filtering: LLSE for problem satisfying particular assumptions:
  - Covariance structure for $x(t)$ specified implicitly via state space model:

$$\underline{x}(t+1) = A(t)\underline{x}(t) + B(t)\underline{u}(t) + G(t)\underline{w}(t)$$

  - Observation Model:

$$\underline{y}(t) = C(t)\underline{x}(t) + \underline{v}(t)$$

- Notation:

$$\begin{aligned} \widehat{\underline{x}}(t \mid s) &= \text{LLSE of } \underline{x}(t) \text{ given } y(\tau),\ \tau \leq s \\ \underline{e}(t \mid s) &= \underline{x}(t) - \widehat{\underline{x}}(t \mid s) \\ P(t \mid s) &= E\left[\underline{e}(t \mid s)\underline{e}(t \mid s)^T\right] \end{aligned}$$

- To solve:
  (a) Set up state space model, identify $A(t)$, $B(t)$, $G(t)$, $C(t)$, $u(t)$, $w(t)$, $v(t)$, and covariances $Q(t)$, $R(t)$.
  (b) Find initial conditions: $\widehat{\underline{x}}(t_0|t_0-1)$, $P(t_0|t_0-1)$
  (c) Iterate Kalman filtering equations
- Kalman Filtering Equations:

  Initialization:

$$\begin{aligned} \widehat{\underline{x}}(t_0|t_0-1) &= \underline{m}_x(t_0) \\ P(t_0|t_0-1) &= P_x(t_0) \end{aligned}$$

  Update Step:

$$\widehat{\underline{x}}(t|t) = \widehat{\underline{x}}(t|t-1) + P(t|t-1)C^T(t)\left[C(t)P(t|t-1)C^T(t) + R(t)\right]^{-1}\left[\underline{y}(t) - C(t)\widehat{\underline{x}}(t|t-1)\right]$$

$$P(t|t) = P(t|t-1) - P(t|t-1)C^T(t)\left[C(t)P(t|t-1)C^T(t) + R(t)\right]^{-1}C(t)P(t|t-1)$$

  Prediction Step:

$$\widehat{\underline{x}}(t+1|t) = A(t)\widehat{\underline{x}}(t|t) + B(t)\underline{u}(t)$$

$$P(t+1|t) = A(t)P(t|t)A^T(t) + G(t)Q(t)G^T(t)$$

- Kalman gain: $K = P(t|t-1)C^T(t)\left[C(t)P(t|t-1)C^T(t) + R(t)\right]^{-1}$
- For stationary processes and long observation intervals, Kalman filter $\Longrightarrow$ Causal Wiener Filter as $t \longrightarrow \infty$. Steady state analysis.

7. **Advice:**

- Have basic results at your fingertips
- Know the assumptions/conditions behind formulas that you use!
- Perform sanity checks on answers – go back to basics if totally stuck (i.e. defining equation of expectation, variance etc)
- Know Fourier/Laplace transforms, partial fraction expansions, and properties
- Make sure you're clear on difference between e.g. $S_{yy}^{+}(s)$ and $\{S_{yy}\}_{+}$
- Don't forget to deal with the mean! (e.g. $R_{xx}(t)$ vs $K_{xx}(t)$, etc.)