ECE 7251: Signal Detection and Estimation

Spring 2002

Prof. Aaron Lanterman
Georgia Institute of Technology

Lecture 6, 1/16/02:
The Orthogonality Principle
in MMSE Estimation

The Setup

- Let X be a random variable with $E[|X|^2] < \infty$
- Can think of such random variables as elements of a Hilbert space with inner product

$$\langle X, Y\rangle = E[XY^*]$$

- Let $\mathcal{V}$ be a collection of random variables on the same space such that:
  - $E[|Z|^2] < \infty$ for all $Z \in \mathcal{V}$
  - $\mathcal{V}$ is a linear class, i.e., if $Z_1, Z_2 \in \mathcal{V}$, then $Z_1+Z_1 \in \mathcal{V}$
  - $\mathcal{V}$ is closed in the mean-square sense:

(slides based on notes by Bruce Hajek of UIUC)

The Orthogonality Principle

- Given setup on previous slide, there exists a unique $\tilde{Z} \in \mathcal{V}$ such that

$$E[|X - \tilde{Z}|^2] \le E[|X - Z|^2]$$

- For all $Z \in \mathcal{V}$,

$$X - \tilde{Z} \perp Z = 0$$

$$E[(X - \tilde{Z})Z^*] = 0$$

- "If and only if" type of statement

General MMSE Estimation

- To keep things simple, let's now focus on real-valued parameters and data
- Let Θ be a random variable with $E[\Theta^2] < \infty$
- Let Y be a random variable (or m-vector)
- Let $\mathcal{V}$ be the set of all possible squared-integrable estimators based on Y:

$$\mathcal{V} = \{g(Y) : \Re^m \to \Re, E[g^2(Y)] < \infty\}$$

- What $g(Y)$ minimizes $E[(\Theta - g(Y))^2]$?
- Previously found:

$$g(Y) = E[\Theta | Y]$$

Double-Check Using O.P.

- Does $E[(\Theta - E[\Theta | Y])h(Y)] = 0$?

$$\overbrace{\iint (q - E[\Theta | y])h(y)\underbrace{p(q, y)}\, dq\, dy}$$

$$= \iint \{(q - E[\Theta | y])\overbrace{p(q | y)}h(y)\, dq\}\overbrace{p(y)}\, dy$$

$$= \int \underbrace{(E[\Theta | y] - E[\Theta | y])}_{= 0} h(y) p(y)\, dy$$

So all is well.

Multiparameter Estimation

- Extends immediately to $\Theta = [\Theta_1, \ldots, \Theta_n]^T$
- Just do one coordinate at a time:

$$E[\Theta | Y] = \begin{bmatrix} E[\Theta_1 | Y] \\ E[\Theta_2 | Y] \\ \vdots \\ E[\Theta_m | Y] \end{bmatrix}$$

- Aside: MAP estimation extends naturally the same way

Error of MMSE Estimator

- Define the est. error $e = \underbrace{\Theta - E[\Theta \mid Y]}_{\text{(a function of } Y)}$
- Expected error:

$$E[e] = E[\Theta] - E[E[\Theta \mid Y]]$$
$$= E[\Theta] - E[\Theta] = 0$$

- Error covariance:

$$\mathrm{Cov}[\Theta] = \mathrm{Cov}[e + E[\Theta \mid Y]]$$
$$= \mathrm{Cov}[e] + \mathrm{Cov}[E[\Theta \mid Y]] \text{ (by O.P.)}$$
$$MSE = E[ee^T] = \mathrm{Cov}[\Theta] - \mathrm{Cov}[E[\Theta \mid Y]]$$

"Linear" MMSE Estimation

- Let Θ and Y be a random vectors
- Let $\mathcal{V}$ be the set of all possible <u>affine</u> estimators based on Y:

$$\mathcal{V} = \{g(Y) = AY + b,\ A \in \Re^{m \times n}, b \in \Re^{m \times 1}\}$$

- What $g(Y)$ minimizes $E[(q - g(Y))^2]$?
- Could set it up Bayes risk, take derivatives and set equal to zero, etc., as before; but can also use O.P. directly!

(slides based on notes by Stefano Soatto of UCLA)

Step 1

- Assume Θ scalar, Θ and Y both zero-mean
- O.P. says: $E[\Theta - AY, Y_1] = 0$

$$E[\Theta - AY, Y_2] = 0$$
$$\vdots$$
$$E[\Theta - AY, Y_n] = 0$$

- Hold iff $E[\Theta Y_i] = AE[YY_i], \quad i = 1 \ldots n$

$$E[\Theta Y^T] = AE[YY^T],$$
$$K_{\Theta Y} = AK_Y \quad \text{(since zero mean)}$$
$$K_{\Theta Y} K_Y^{-1} = A$$
$$g(Y) = AY$$

Only need 2[nd] order statistics!

Step 2

- Now Θ *vector*, Θ and Y both zero-mean
- Just repeat previous for $\Theta_1, \Theta_2, \ldots, \Theta_m$

$$A_1 = K_{\Theta_1 Y} K_Y^{-1}$$
$$A_2 = K_{\Theta_2 Y} K_Y^{-1}$$
$$\vdots$$
$$A_m = K_{\Theta_m Y} K_Y^{-1}$$

row index

Combining: $A = K_{\Theta Y} K_Y^{-1}$

Step 3

- Step 1: Θ and Y *not* zero mean
- Do transformation $\tilde{\Theta} = \Theta - m_\Theta, \quad \tilde{Y} = Y - m_Y$
- For transformed problem, have:

$$A = K_{\tilde{\Theta}\tilde{Y}} K_{\tilde{Y}}^{-1} = K_{\Theta Y} K_Y^{-1}$$

- Substitution yields:

$$g(Y) = m_\Theta + K_{\Theta Y} K_Y^{-1}(Y - m_Y) \equiv \underbrace{\hat{E}[\Theta \mid Y]}$$

Called "projection of Θ over span of Y" by inhabitants of Hilbert space

Error of LMMSE Estimator

- Define the est. error $e = \underbrace{\Theta - \hat{E}[\Theta \mid Y]}_{\text{(a func. of } Y)}$
- Expec. of estimator:

$$E[\hat{E}[\Theta \mid Y]] = E[m_\Theta + K_{\Theta Y} K_Y^{-1}(Y - m_Y)]$$
$$= m_\Theta + K_{\Theta Y} K_Y^{-1} \underbrace{E[(Y - m_Y)]}_{=0}$$
$$= m_\Theta$$

- Expected value of error:

$$E[e] = E[\Theta - \hat{E}[\Theta \mid Y]] = 0$$

MSE of LMMSE Estimator

$$\mathrm{Cov}[\Theta] = \mathrm{Cov}[e + \hat{E}[\Theta \mid Y]]$$

$$= \mathrm{Cov}[e] + \mathrm{Cov}[\hat{E}[\Theta \mid Y]] \quad \text{(by O.P.)}$$

$$MSE = E[ee^T] = \mathrm{Cov}[\Theta] - \mathrm{Cov}[E[\Theta \mid Y]]$$

$$= \mathrm{Cov}[\Theta] - \mathrm{Cov}[AY]$$

$$= K_\Theta - AK_Y A^T$$

$$= K_\Theta - K_{\Theta Y} K_Y^{-1} K_Y (K_{\Theta Y} K_Y^{-1})^T$$

$$= K_\Theta - K_{\Theta Y} K_Y^{-1} K_{Y\Theta} = K_e$$

Gaussian Case

- In general,

$$E[(\Theta - E[\Theta \mid Y])]^2 \leq E[(\Theta - \hat{E}[\Theta \mid Y])^2]$$

(take a performance hit for using linear estimator)

- But, if Θ and Y are *jointly* Gaussian, then

$$E[\Theta \mid Y] = \hat{E}[\Theta \mid Y]$$

- Also have easy way to compute conditionals:

$$q \mid Y = y \sim \mathcal{N}(\hat{E}[X \mid Y = y], K_e)$$

$$\sim \mathcal{N}(m_\Theta + K_{\Theta Y} K_Y^{-1}(Y - m_Y), K_\Theta - K_{\Theta Y} K_Y^{-1} K_{Y\Theta})$$