Linear LMS Estimation

also known as Linear MMSE estimation.

Motivation for linear LMS estimation.

- To find the LMS estimate, one needs knowledge of the full statistics $f(\theta, x)$, which is rarely the case in practice.
- In practice, we typically only have knowledge of the first and second moments of Θ, X, $\mathbb{E}[\Theta]$, $\mathbb{E}[X]$, $\mathrm{Var}(X)$, $\mathrm{Var}\,\Theta$, $\mathrm{Cov}(\Theta, X)$
- This is not, in general, enough to compute LMS estimate.
- But we will see it is sufficient to compute the estimate of the form

$$\hat{\Theta} = aX + b \quad \Leftarrow \text{ a linear function of the observation } X.$$

that minimizes the $\mathrm{MSE} = \mathbb{E}[(\Theta - \hat{\Theta})^2]$.

Linear LMS estimation.

- Unknown Θ, data X. ($p(\theta)$, $p(x|\theta)$ may not be available)
- Instead, we are given $\mathbb{E}X$, $\mathbb{E}\Theta$, $\mathrm{Var}X$, $\mathrm{Var}\Theta$, $\mathrm{Cov}(\Theta, X)$.
- Q: Find a, b such that the <u>linear</u> estimate

$$\hat{\Theta} = aX + b$$

minimizes the $\mathrm{MSE} = \mathbb{E}[(\Theta - \hat{\Theta})^2]$

- Suppose a has already been found: what is the best b?

(Recall derivation in "LMS in the absence of noise".)

$$\mathbb{E}[(\Theta-\hat{\Theta})^2] = \mathbb{E}[(\underbrace{\Theta - aX}_{Y} - b)^2] = \mathbb{E}[(Y-b)^2] = \mathbb{E}[Y^2] - 2\mathbb{E}[Y]\cdot b + b^2 \triangleq g(b).$$

$$\frac{dg(b)}{db} = 2b - 2\mathbb{E}[Y] \overset{\text{set}}{=} 0 \Rightarrow b = \mathbb{E}[Y] = \mathbb{E}[\Theta - aX].$$

Linear LMS estimation. (cont.)

- Now find the best a that minimizes $\mathbb{E}[(\Theta - aX - \overbrace{\mathbb{E}[\Theta - aX]}^{b.})^2]$

- We have $g(a) = \mathbb{E}[(\Theta - aX - \mathbb{E}[\Theta - aX])^2]$

$$= \mathrm{Var}(\Theta - aX)$$

$$= \mathrm{Var}\,\Theta - 2a\,\mathrm{Cov}(\Theta, X) + a^2\,\mathrm{Var}\,X$$

$$\boxed{\begin{aligned} &\mathrm{Var}\,Y = \mathbb{E}[(Y - \mathbb{E}Y)^2] \\ &\mathrm{Var}(a\cdot Y) = a^2\,\mathrm{Var}\,Y \\ &\mathrm{Var}(X-Y) = \mathrm{Var}\,X + \mathrm{Var}\,Y \\ &\qquad\qquad - 2\,\mathrm{Cov}(X, Y) \end{aligned}}$$

$$\frac{dg(a)}{da} = 2\,\mathrm{Var}X\; a - 2\,\mathrm{Cov}(\Theta, X) \overset{\text{set}}{=} 0$$

$$\Rightarrow a = \frac{\mathrm{Cov}(\Theta, X)}{\mathrm{Var}X}$$

$$\Rightarrow \boxed{\hat{\Theta}_{LLMS}} = aX + b = \frac{\mathrm{Cov}(\Theta, X)}{\mathrm{Var}X} X + \mathbb{E}[\Theta] - \frac{\mathrm{Cov}(\Theta, X)}{\mathrm{Var}X} \cdot \mathbb{E}X.$$

$$\boxed{= \mathbb{E}[\Theta] + \frac{\mathrm{Cov}(\Theta, X)}{\mathrm{Var}(X)}(X - \mathbb{E}X)}$$

Recall $\hat{\Theta}_{LMS} = \mathbb{E}[\Theta | X]$

Linear LMS estimation. (cont.)

$$\rho_{\Theta,X} = \frac{\mathrm{Cov}(\Theta, X)}{\sqrt{\mathrm{Var}\,\Theta \cdot \mathrm{Var}\,X}}$$

- Corresponding MSE

$$MSE_{LLMS} = \mathbb{E}[(\Theta - \hat{\Theta})^2] = \mathbb{E}\left[\left(\Theta - \mathbb{E}\Theta - \frac{\mathrm{Cov}(\Theta, X)}{\mathrm{Var}\,X}(X - \mathbb{E}X)\right)^2\right].$$

$$= \mathbb{E}[(\Theta - \mathbb{E}\Theta)^2] - 2\frac{\mathrm{Cov}(\Theta, X)}{\mathrm{Var}\,X}\mathbb{E}((\Theta - \mathbb{E}\Theta)(X - \mathbb{E}X)) + \left(\frac{\mathrm{Cov}(\Theta, X)}{\mathrm{Var}\,X}\right)^2 \mathbb{E}[(X - \mathbb{E}X)^2]$$

$$= \mathrm{Var}\,\Theta - 2\frac{[\mathrm{Cov}(\Theta, X)]^2}{\mathrm{Var}\,X} + \left(\frac{\mathrm{Cov}(\Theta, X)}{\mathrm{Var}\,X}\right)^2 \mathrm{Var}\,X$$

$$= \mathrm{Var}\,\Theta - \frac{[\mathrm{Cov}(\Theta, X)]^2}{\mathrm{Var}\,X}$$

$$= \mathrm{Var}\,\Theta\,(1 - \rho_{\Theta,X}^2).$$

Recall MSE for LMS estimator is.

$$MSE_{LMS} = \mathbb{E}[\mathrm{Var}(\Theta \mid X)].$$

$\Rightarrow$ Linear LMS only requires first & second moments $\mathbb{E}\Theta$, $\mathbb{E}X$, $\mathrm{Var}\,\Theta$, $\mathrm{Var}\,X$, $\mathrm{Cov}(\Theta, X)$.

Summary of linear LMS estimation.

- Unknown Θ, data X, ($f(\theta)$ and $f(x|\theta)$ may not be available).
- Q: Given data X, find a <u>linear</u> estimator $\hat{\Theta} = aX + b$

  (a, b constants) that minimizes MSE $\mathbb{E}[(\Theta - \hat{\Theta})^2]$.

a linear function in x.

$$X \longrightarrow \boxed{\hat{\theta}(x) = ax + b} \longrightarrow \hat{\Theta}$$

- A: $\hat{\theta}_{LLMS}(x) = \mathbb{E}\Theta + \dfrac{Cov(\Theta, X)}{Var X}(x - \mathbb{E}X)$. ← the estimate for specific data x.

  $\hat{\Theta}_{LLMS} = \mathbb{E}\Theta + \dfrac{Cov(\Theta, X)}{Var X}(X - \mathbb{E}X)$ ← estimator (a random variable)

  $$MSE_{LLMS} = Var\Theta - \frac{[Cov(\Theta, X)]^2}{Var X}$$

Example : Additive shot niose channel

$X = \Theta + W$, signal $\Theta \sim N(0,1)$, noise $W|\Theta=\theta \sim N(0,\theta^2)$.

Q: Find the linear LMS estimate of Θ given X.

Want to find $\mathbb{E}X$, $\mathbb{E}\Theta$, $\mathrm{Var}X$, $\mathrm{Var}\Theta$, and $\mathrm{Cov}(\Theta, X)$.

- $\Theta \sim N(0,1) \Rightarrow \mathbb{E}\Theta = 0$, $\mathrm{Var}\Theta = 1$. $W|\Theta=\theta \sim N(0,\theta^2) \Rightarrow \mathbb{E}[W|\Theta=\theta] = 0$, $\mathrm{Var}[W|\Theta=\theta] = \theta^2$

  $\Rightarrow \mathbb{E}[W|\Theta]=0$, $\mathrm{Var}[W|\Theta] = \Theta^2$

- $\mathbb{E}W \overset{\text{law of total expectation}}{=} \mathbb{E}_\Theta[E[W|\Theta]] = \mathbb{E}_\Theta[0] = 0$.

- $\mathrm{Var}W \overset{\text{law of total variance}}{=} \mathbb{E}_\Theta[\mathrm{Var}(W|\Theta)] + \mathrm{Var}[\mathbb{E}[W|\Theta]]$.

  $= \mathbb{E}[\Theta^2] + \mathrm{Var}[0] = \mathrm{Var}\Theta + (\mathbb{E}\Theta)^2 + 0 = 1$

- $\mathbb{E}[\Theta \cdot W] \overset{\text{law of total expectation}}{=} \mathbb{E}_\Theta[E[\Theta \cdot W|\Theta]] \overset{\text{next page}}{=} \mathbb{E}_\Theta[\Theta\, \mathbb{E}[W|\Theta]] = \mathbb{E}[\Theta \cdot 0] = 0$

Want to show $\mathbb{E}_{\Theta}[\mathbb{E}[\Theta W|\Theta]] = \mathbb{E}_{\Theta}[\Theta\, \mathbb{E}[W|\Theta]]$.

$\boxed{\begin{aligned}\mathbb{E}[g(\Theta)] &= \int g(\theta) f(\theta)\, d\theta. \\ g(\theta) &= \mathbb{E}[\Theta W | \Theta = \theta] \\ &= \theta\, \mathbb{E}[W|\Theta=\theta].\end{aligned}}$

<u>proof</u>: $\mathbb{E}_{\Theta}[\mathbb{E}[\Theta W|\Theta]] = \int \mathbb{E}[\Theta W|\Theta=\theta] f(\theta)\, d\theta$

not random given $\Theta = \theta$.

$$= \int \mathbb{E}[\theta W|\Theta=\theta] f(\theta)\, d\theta.$$

$$= \int \theta\, \mathbb{E}[W|\Theta=\theta] f(\theta)\, d\theta.$$

$$= \mathbb{E}_{\Theta}[\Theta\, \mathbb{E}[W|\Theta]].$$

- $\mathrm{Var}X = \mathrm{Var}(\Theta + W) = \mathrm{Var}\Theta + \mathrm{Var}W + 2\,\mathrm{Cov}(\Theta, W) = 1+1+2\times 0 = 2$.
- $\mathrm{Cov}(\Theta, X) = \mathbb{E}[\Theta X] - \mathbb{E}\Theta \cdot \mathbb{E}X = \mathbb{E}[\Theta(\Theta+W)] - 0 = \mathbb{E}[\Theta^2] + \mathbb{E}[\Theta W] = 1 + 0 = 1$.

$$\Rightarrow \hat{\Theta}_{LLMS} = \mathbb{E}\Theta + \frac{\mathrm{Cov}(\Theta, X)}{\mathrm{Var}X}(X - \mathbb{E}X) = \frac{1}{2}X$$

$$\mathrm{MSE}_{LLMS} = \mathrm{Var}\Theta - \frac{(\mathrm{Cov}(\Theta, X))^2}{\mathrm{Var}X} = 1 - \frac{1}{2} = \frac{1}{2}$$

Example: Romeo and Juliet.

Romeo and Juliet start dating, but Juliet will be late on any date by a random amount X uniformly distributed over the interval $[0, \theta]$. The parameter θ is unknown and is the realization of a r.v. $\Theta \sim \text{Unif}[0,1]$. Assuming that Juliet was late by an amount x on their first date, how should Romeo use this information to estimate Θ using linear LMS rule? Find the corresponding MSE.

$$\boxed{Y \sim \text{Unif}[a,b].\ \mathbb{E}Y = \frac{a+b}{2},\ \text{Var}\,Y = \frac{(b-a)^2}{12}}$$

Unknown: $\Theta \sim \text{Unif}[0,1]$. $\Rightarrow \mathbb{E}\Theta = \frac{1}{2} \quad \text{Var}\,\Theta = \frac{1}{12}$.

Observation model: $X \mid \Theta = \theta \sim \text{Unif}[0, \theta] \Rightarrow \mathbb{E}[X \mid \Theta = \theta] = \frac{\theta}{2}$, $\text{Var}[X \mid \Theta = \theta] = \frac{\theta^2}{12}$

To find the linear LMS estimator, we want to compute. $\mathbb{E}X$. $\text{Var}X$. $\text{Cov}(X, \Theta)$.

By the law of total expectation, $\mathbb{E}X = \mathbb{E}_\Theta[\mathbb{E}[X \mid \Theta]] = \mathbb{E}\left[\frac{\Theta}{2}\right] = \frac{1}{4}$

By the law of total variance,

$$\mathrm{Var}[X] = \mathbb{E}[\mathrm{Var}[X|\Theta]] + \mathrm{Var}[\mathbb{E}[X|\Theta]]$$

$$= \mathbb{E}\left[\frac{\Theta^2}{12}\right] + \mathrm{Var}\left[\frac{\Theta}{2}\right]$$

$$= \frac{1}{12}\left[\mathrm{Var}\,\Theta + (\mathbb{E}\,\Theta)^2\right] + \frac{1}{4}\mathrm{Var}\,\Theta = \frac{7}{144}$$

To compute $\mathrm{Cov}(X, \Theta) = \mathbb{E}[X\Theta] - \mathbb{E}X \cdot \mathbb{E}\Theta$, we first compute $\mathbb{E}[X \cdot \Theta]$

$$\mathbb{E}[X\Theta] = \mathbb{E}_\Theta[\mathbb{E}[X\Theta|\Theta]] = \mathbb{E}_\Theta[\Theta\, \mathbb{E}[X|\Theta]] = \mathbb{E}\left[\frac{\Theta^2}{2}\right] = \frac{1}{6}$$

$$\Rightarrow \mathrm{Cov}(X, \Theta) = \mathbb{E}[X \cdot \Theta] - \mathbb{E}X \cdot \mathbb{E}\Theta = \frac{1}{24}$$

$$\Rightarrow \quad \hat{\Theta}_{LLMS} = \mathbb{E}\Theta + \frac{\mathrm{Cov}(\Theta, X)}{\mathrm{Var}\,X}(X - \mathbb{E}X) = \frac{6}{7}X + \frac{2}{7}$$

$$MSE_{LLMS} = \mathrm{Var}\,\Theta - \frac{(\mathrm{Cov}(\Theta, X))^2}{\mathrm{Var}\,X} = \frac{1}{21}$$

Linear LMS v.s. LMS estimation.

- Q: How does linear LMS estimation ompare to LMS estimation?
- In general, $\hat{\Theta}_{LLMS} \neq \hat{\Theta}_{LMS}$, and $MSE_{LLMS} \geq MSE_{LMS}$.
- It turns out for the special case when Θ and X are jointly Gaussian

$$\hat{\Theta}_{LLMS} = \hat{\Theta}_{LMS}$$

and therefore $MSE_{LLMS} = MSE_{LMS}$.

Joint Gaussian Random Variables

- Two random variables are <u>jointly Gaussian</u> if their joint pdf is of the form

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} e^{-\frac{1}{2(1-\rho_{X,Y}^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\rho_{X,Y}\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]}$$

- The pdf is fully determined by μ_X, μ_Y, σ_X^2, σ_Y^2, $\rho_{X,Y}$

- <u>Example 1</u>: $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, X and Y are <u>independent</u>,

$\Rightarrow \rho_{X,Y} = 0$

then X and Y are jointly Gaussian with joint pdf

$$f(x,y) = f(x)f(y) = \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]}$$

Joint Gaussian Random Variables (cont.)

- If X and Y are jointly Gaussian, the contours of equal joint pdf are ellipses defined by the quadratic equation

$$\frac{(x-\mu_X)^2}{2\sigma_X^2} + \frac{(y-\mu_Y)^2}{2\sigma_Y^2} - 2\rho_{X,Y}\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} = c \geq 0 .$$

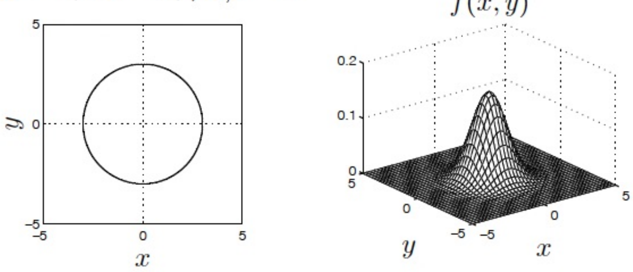
- The orientation of the major axis of the ellipse is

$$\theta = \frac{1}{2}\arctan\left(\frac{2\rho_{X,Y}\sigma_X\sigma_Y}{\sigma_X^2-\sigma_Y^2}\right)$$

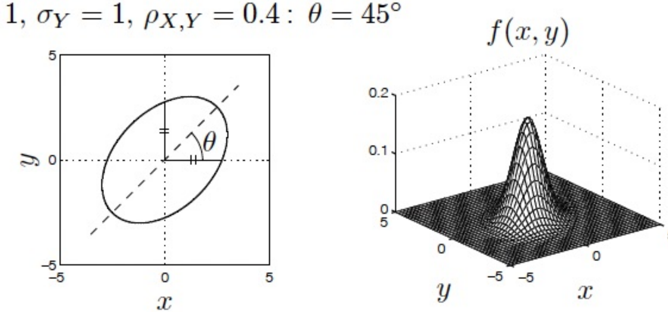
- Examples: We plots contours of equal joint pdf $f(x,y)$ for joint Gaussian X, Y with $\mu_X = \mu_Y = 0$ and different σ_X, σ_Y, and $\rho_{X,Y}$.

Contours of joint Gaussian pdf (Source : Stanford EE 278 notes).

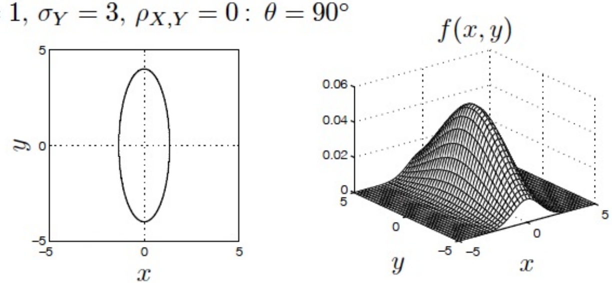
$\sigma_X = 1,\ \sigma_Y = 1,\ \rho_{X,Y} = 0$



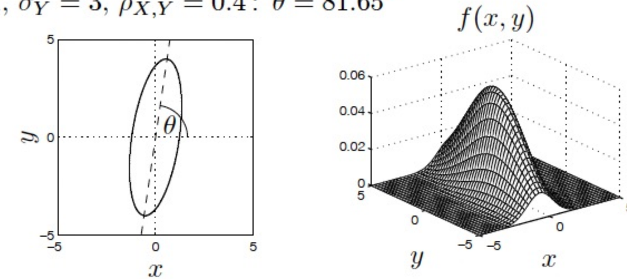
$\sigma_X = 1,\ \sigma_Y = 1,\ \rho_{X,Y} = 0.4:\ \theta = 45^\circ$



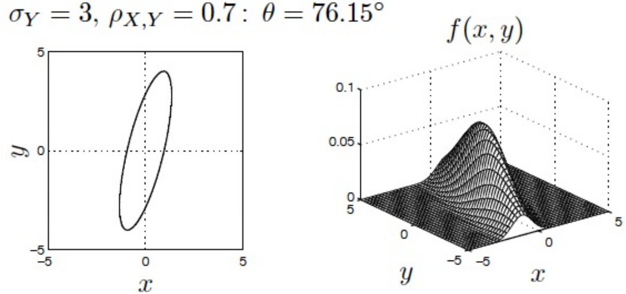
$\sigma_X = 1,\ \sigma_Y = 3,\ \rho_{X,Y} = 0:\ \theta = 90^\circ$



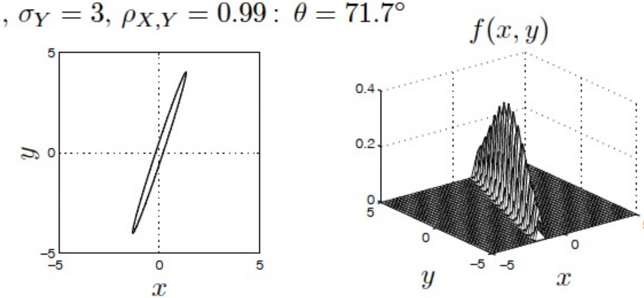
$\sigma_X = 1,\ \sigma_Y = 3,\ \rho_{X,Y} = 0.4:\ \theta = 81.65^\circ$



$\sigma_X = 1,\ \sigma_Y = 3,\ \rho_{X,Y} = 0.7:\ \theta = 76.15^\circ$



$\sigma_X = 1,\ \sigma_Y = 3,\ \rho_{X,Y} = 0.99:\ \theta = 71.7^\circ$



Jointly Gaussian random vector

- Random variables $X_1, X_2, \cdots, X_n$ are <u>jointly Gaussian</u>, or the random vector $\underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ is Gaussian $N(\underline{\mu}, \Sigma)$, if the joint pdf is of the form.

$$\boxed{f(x_1, \cdots, x_n) = f(\underline{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^T \Sigma^{-1} (\underline{x}-\underline{\mu})},}$$

where

$$\underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \underline{\mu} = \begin{pmatrix} EX_1 \\ \vdots \\ EX_n \end{pmatrix}, \quad \Sigma = \begin{bmatrix} Cov(X_1, X_1), & \cdots & Cov(X_1 X_n) \\ & \ddots \quad Cov(X_i, X_j) \quad \ddots & \\ Cov(X_n X_1) & \cdots & Cov(X_n, X_n) \end{bmatrix}.$$

Covariance matrix.

and $\det(\Sigma) > 0$.

Properties of jointly Gaussian r.v.s.

(1) Linear transforms of jointly Gaussian random variables are jointly Gaussian

i.e., given any $m \times n$ full rank matrix A with $m \le n$, let $\underline{X} \sim N(\underline{\mu}, \Sigma)$

then $\boxed{\underline{Y} = A\underline{X} \sim N(A\underline{\mu}, A\Sigma A^T).}$

(2) Marginals of jointly Gaussian r.v.s are jointly Gaussian.

i.e., if $(X_1 \cdots X_n)$ are jointly Gaussian, then for any $\{i_1, i_2, \cdots, i_k\} \subseteq \{1, \cdots, n\}$

$(X_{i_1}, X_{i_2}, \cdots, X_{i_k})$ are jointly Gaussian.

eg: If $\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N(\underline{\mu}, \Sigma)$, then $Y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_3 \end{bmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_3 \end{pmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{13} \\ \Sigma_{31} & \Sigma_{33} \end{bmatrix} \right)$

Properties of jointly Gaussian r.v.s (cont.)

(3). Conditionals of jointly Gaussian r.v.s are jointly Gaussian. If

$$\underline{X} = \begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \end{bmatrix} = N\left(\begin{bmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right),$$

where $\underline{X}$ is an n-dim vector, $\underline{X}_1$ is a k-dim vector, $\underline{X}_2$ is an (n-k)-dim vector,

then,

$$\boxed{\underline{X}_2 \mid \{\underline{X}_1 = \underline{x}_1\} \sim N\left(\Sigma_{21}\Sigma_{11}^{-1}(\underline{x}_1 - \underline{\mu}_1) + \underline{\mu}_2,\ \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right).}$$

eg: $n=2$, $k=1$.

$$X_2 \mid \{X_1 = x_1\} \sim N\Bigg(\underbrace{\frac{Cov(X_1, X_2)}{Var X_1}(x_1 - \mathbb{E}X_1) + \mathbb{E}X_2}_{=}, \ \underbrace{Var X_2 - \frac{(Cov(X_1, X_2))^2}{Var X_1}}_{=} \Bigg).$$

linear LMS estimate of X_2 given $X_1 = x_1$, MSE for linear LMS.

$\parallel$

On the other hand, $\mathbb{E}[X_2 | X_1 = x_1] =$ LMS estimate $= \arg\max_{x_2} f(x_2|x_1) =$ MAP estimate.

Linear LMS v.s. LMS v.s. MAP estimation.

> If the unknown Θ and data X (potentially can be a vector) are <u>jointly Gaussian</u>, then the linear LMS estimate of Θ given X is the same as the LMS estimate and the MAP estimate. i.e.,
>
> $$\hat{\Theta}_{LLMS} = \hat{\Theta}_{LMS} = \hat{\Theta}_{MAP}.$$

Example: Estimating Gaussian signal in Gaussian noise.

- Signal $\Theta \sim N(0,1)$, $X = \Theta + W$, $W \sim N(0,1)$ indep. of Θ
- $\begin{bmatrix} X \\ \Theta \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} W \\ \Theta \end{bmatrix} \Rightarrow X$ and Θ are jointly Gaussian
- $\Rightarrow \hat{\Theta}_{LLMS} = \hat{\Theta}_{LMS} = \hat{\Theta}_{MAP} = \frac{X}{2}$

Geometric formulation of Linear estimation

First we recall background on an inner product space.

- A <u>vector space</u> V, eg Euclidean space, consists of a set of vectors that are closed under two operations
  - <u>vector addition</u>: if $v_1, v_2 \in V$, then $v_1 + v_2 \in V$.
  - <u>scalar multiplication</u>: if $a \in \mathbb{R}$, $v \in V$, then $a \cdot v \in V$.
- A <u>inner product</u>, eg dot product $u^T v$ in Euclidean space, is a real valued operation satisfying these three conditions:
  - commutativity $u^T v = v^T u$
  - linearity: $(au+v)^T w = a u^T w + v^T w$
  - nonnegativity: $u^T u \geq 0$ and $u^T u = 0$ iff $u = 0$.

Geometric formulation of Linear estimation (cont.)

- The <u>norm</u> of u is defined as $\|u\| = \sqrt{u^T u}$
- u and v are <u>orthogonal</u> (written as $u \perp v$) if $u^T v = 0$
- A vector space with an inner product is called an <u>inner product space</u>

eg: Euclidean space with dot product

Linear LMS estimation.

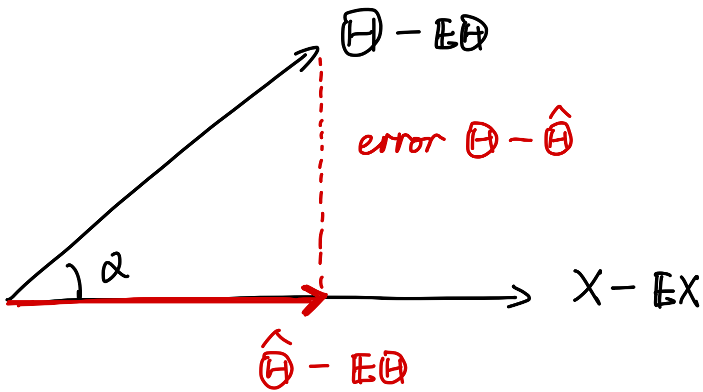
- Consider the space V of all <u>zero mean</u> random variables on the same probability space.
- View $\Theta - \mathbb{E}\Theta$ and $X - \mathbb{E}X$ as vectors in V.
  - Closed under vector addition : $V_1 . V_2 \in V \Rightarrow V_1 + V_2 \in V$.

    ($\mathbb{E}V_1 = 0$, $\mathbb{E}V_2 = 0 \Rightarrow \mathbb{E}[V_1 + V_2] = 0$.)
  - Closed under scalar multiplication : $a \in \mathbb{R}, V \in V \Rightarrow a \cdot V \in V$.

    (If $\mathbb{E}V = 0$ that $\mathbb{E}[aV] = 0$).

    $\mathbb{E}V_1 \cdot \mathbb{E}V_2 = 0$.
  - Inner product of $V_1, V_2 \in V$: $\mathbb{E}[V_1 V_2] = Cov(V_1, V_2)$.
  - norm of $V \in V$: $\|V\| = \sqrt{\mathbb{E}[V^2]} = \sqrt{Var(V)} = \sigma_V$.

Orthogonality principle for linear LMS estimation.



inner product $\Leftrightarrow$ $\mathrm{Cov}(\Theta, X)$

norm of $\Theta - \mathbb{E}\Theta$ $\Leftrightarrow$ σ_Θ

norm of $X - \mathbb{E}X$ $\Leftrightarrow$ σ_X

$\cos\alpha$ $\Leftrightarrow$ $\rho_{\Theta, X}$.

Find a vector $\hat{\Theta} - \mathbb{E}\Theta = a(X - \mathbb{E}X)$ that minimizes $\|\Theta - \hat{\Theta}\|$

- Clearly $\Theta - \hat{\Theta} \perp X - \mathbb{E}X$ minimizes $\|\Theta - \hat{\Theta}\|$, i.e.,

$$\mathbb{E}((\Theta - \hat{\Theta})(X - \mathbb{E}X)) = 0 \Rightarrow \mathbb{E}[(\Theta - \mathbb{E}\Theta)(X - \mathbb{E}X)] = \mathbb{E}[(\hat{\Theta} - \mathbb{E}\Theta)(X - \mathbb{E}X)]$$

$$\Rightarrow \mathrm{Cov}(\Theta, X) = a \,\mathrm{Var}\, X \qquad \Rightarrow a = \frac{\mathrm{Cov}(\Theta, X)}{\mathrm{Var}\, X}.$$

This arguement is called the <u>orthogonality principle</u>.