1 Overview

In the last lecture, we talked about Chernoff bound and defined the characteristic function of a RV. Then we gave some examples and concluded by proving the Central Limit Theorem with examples.

In this lecture, we will introduce random vectors, define Positive Semi-Definite (P.S.D.) matrices, give some examples theorem and proofs, then use them to prove some properties in covariance matrices.

2 Random Vector

**Definition 1.** A random vector $X = (X_1, X_2, \ldots, X_n)^T$, is a vector of random variables $X_i$, $i = 1, \ldots, n$.

**Definition 2.** The mean vector of $X$, denoted by $\mu$, is $\mu = (\mu_1, \mu_2, \ldots, \mu_n)^T$ where $\mu_i = E[X_i], i = 1, \ldots, n$.

**Definition 3.** The covariance matrix $K_{XX}$ or $K$, of $X$ is an $n \times n$ matrix defined as

$$K_{XX} \triangleq E\left[ (X - \mu)(X - \mu)^T \right].$$

$$K_{XX} = E \begin{bmatrix} (X_1 - \mu_1) & (X_2 - \mu_2) & \cdots & (X_n - \mu_n) \\ (X_1 - \mu_1)(X_2 - \mu_2) & (X_2 - \mu_2)^2 & \cdots & (X_2 - \mu_2)(X_n - \mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ (X_n - \mu_n)(X_1 - \mu_1) & (X_n - \mu_n)(X_2 - \mu_2) & \cdots & (X_n - \mu_n)^2 \end{bmatrix},$$

$$K_{XX} = \begin{bmatrix} \sigma_1^2 & K_{12} & \cdots & K_{1n} \\ K_{21} & \sigma_2^2 & \cdots & K_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ K_{n1} & K_{n2} & \cdots & \sigma_n^2 \end{bmatrix}.$$

Remark: The matrix $K_{XX}$ is real symmetric and $K_{ij} = K_{ji} = cov(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] = K_{ij}$, and $\sigma_i^2 = V(X_i)$. 

1
Definition 4. The correlation matrix $R_{XX}$, or $R$, is defined as $R = E[XX^T]$. 

Corollary 1. $K = R - \mu \mu^T$. 

Example 1. $X = (X_1, X_2)$, 

$$\text{Cov}(X_1, X_2) = E[X_1, X_2] - \mu_1 \mu_2,$$

$$K_{XX} = \begin{bmatrix} \sigma^2_{X_1} & \text{cov}(X_1, X_2) \\ \text{cov}(X_1, X_2) & \sigma^2_{X_2} \end{bmatrix} = \begin{bmatrix} E[X_1^2] & E[X_1X_2] \\ E[X_1X_2] & E[X_2^2] \end{bmatrix} - \begin{bmatrix} \mu_1^2 & \mu_1 \mu_2 \\ \mu_1 \mu_2 & \mu_2^2 \end{bmatrix}.$$

Definition 5. For any random vectors $X$ and $Y$ of same length.

1. If the cross-covariance matrix $K_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E[XY^T] - \mu_X \mu^T_Y = 0 \Rightarrow$ we say that $X$ and $Y$ are uncorrelated.

2. If $E[XY^T] = 0 \Rightarrow$ we say that $X$ and $Y$ are orthogonal.

3 Properties of Covariance Matrices

Can any $n \times n$ real symmetric matrix be a covariance matrix? Answer: No.

Example 2. $M = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$, can it be covariance matrix of a vector $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$?

No. Because $V[X_2] = -2 < 0$.

Example 3. Consider matrix $M = \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix}$, can it be a covariance matrix?

Take $Y + X_1 - X_2$,

$$V(Y) = V(X_1 - X_2) = V(X_1) + V(X_2) - 2 \text{cov}(X_1, X_2) = 2 + 2 - 2 \times 3 = -2$$

So $M$ cannot be covariance matrix.

Therefore we want for any linear combination of $X = (X_1, \ldots, X_n)$, say $Y = a_1 X_1 + \ldots + a_n X_n$, to have $V(Y) \geq 0$.

$$V(Y) = E(Y^2) - (E(Y))^2$$

$$E(Y) = E[a^T X] = a^T \mu_X$$

$$E[Y^2] = E[(a^T X)(a^T X)] = E[a^T X \cdot X^T a] = a^T E[X \cdot X^T] a$$

$$\Rightarrow V(Y) = a^T E[X \cdot X^T] a - a^T \mu_X \mu_X^T a$$

$$= a^T K_{XX} a \quad \text{should be} \quad \geq 0$$

So we want $M$ to satisfy $a^T M a \geq 0$, for any $a$. 

2
Definition 6. A matrix $M$ is positive semi-definite (P.S.D) if

$$X^T M X \geq 0 \quad \forall X \in \mathbb{R}^n \text{ (we say } M \succeq 0).$$

Example 4. The identity matrix $I$ is P.S.D. because for any $X = (X_1, X_2)^T$,

$$X^T I X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix},$$

$$= \|X\|^2 \geq 0.$$

Similarly, any diagonal matrix with all non-negative diagonal entries is psd.

Example 5. Consider the same matrix $M$ of example 3,

$$\begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = (1) \begin{bmatrix} -1 & 1 \end{bmatrix} = -2 < 0.$$

Thus, this matrix is not P.S.D.

Theorem 1. Any covariance matrix $K$ is P.S.D.

Proof. Let $X = (X_1, X_2, \ldots, X_n)^T$ be a a zero-mean random vector, i.e., $E[X] = (0, 0, \ldots, 0)^T$, and let

$$K = E[XX^T].$$

Our goal is to prove that $K \succeq 0$, which means that if we pick $Z = (Z_1, Z_2, \ldots, Z_n)^T$ we need to show that $Z^T K Z \geq 0$.

$$Z^T K Z = Z^T E[XX^T] Z, \quad (1)$$

$$= E[Z^T X X^T Z], \quad (2)$$

$$= E[(Z^T X) (Z^T X)^T], \quad (3)$$

$$= E[Y^2] \geq 0. \quad (4)$$

Where equation (2) is a result of the linearity of expectations and equation (3) results from

$$(AB^T) = B^T A^T,$$

and in equation (4) $Y = Z^T X$ is a single random variable.

Definition 7. The eigenvalues of a matrix $M$ are the scalars $\lambda$ such that

$$\exists \Phi \neq 0, M \Phi = \lambda \Phi. \quad (6)$$

The vectors $\Phi$ are called eigenvectors. Typically we choose $\phi_i$ such that $||\phi_i|| = 1$. 

3
**Theorem 2.** A real symmetric matrix $M$ is P.S.D if and only if all its eigenvalues are non-negative.

**Theorem 3.** Let $M$ be a real symmetric matrix then $M$ has $n$ mutually orthogonal unit eigenvectors $\phi_1, \ldots, \phi_n$.

**Proof.** From linear Algebra or in the textbook. \hfill \square

**Example 6.** Find the eigenvalues and eigenvectors of the matrix $M = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$.

1. **Eigenvalues**:

   $\det \left( \begin{bmatrix} 4 - \lambda & 2 \\ 2 & 4 - \lambda \end{bmatrix} \right) = 16 + \lambda^2 - 8\lambda - 4 = 0,$

   $\lambda_1 = 6$ and $\lambda_2 = 2$ therefore $M \succ 0$.

2. **Eigenvectors**:

   For $\lambda_1 = 2$ set $\Phi_1 = \begin{bmatrix} \Phi_{11} \\ \Phi_{12} \end{bmatrix}^T$ such that

   $\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} \Phi_{11} \\ \Phi_{12} \end{bmatrix} = 2 \begin{bmatrix} \Phi_{11} \\ \Phi_{12} \end{bmatrix}$.

   $4\Phi_{11} + 2\Phi_{12} = 2\Phi_{11}$

   $2\Phi_{11} + 4\Phi_{12} = 2\Phi_{12}$

   $\Rightarrow \Phi_{11} = -\Phi_{12} \Rightarrow \Phi_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}^T$.

   For $\lambda_2 = 6$: we repeat the same steps and get

   $\Phi_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}^T$.

**Claim 1.** *(Eigenvalue Decomposition)* The matrix $M$ having $\Phi_1, \Phi_2$ as eigenvectors can be expressed as

$M = U\Lambda U^T,$

Where

$U = \begin{bmatrix} \Phi_1 \\ \Phi_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix},$

$\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}$.

Check:

$U\Lambda U^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix},$

$= \frac{1}{2} \begin{bmatrix} 2 & 6 \\ -2 & 6 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix},$

$= \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix},$

$= M.$
**Theorem 4.** (Eigenvalue Decomposition Theorem) Let $M$ be a real symmetric matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and corresponding eigenvectors $\Phi_1, \Phi_2, \ldots, \Phi_n$ then

$$U^T M U = \Lambda,$$

With :

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

**Proof.** We can write from equation (6) :

$$MU = U\Lambda$$

and

$$U^{-1}MU = \Lambda,$$

Since $U$ is a real symmetric matrix :

$$U^T = U^{-1} \Rightarrow \Lambda = U^T MU,$$

and

$$M = (U^T)^{-1} \Lambda U^{-1},$$

$$= U\Lambda U^T.$$

$$\Box$$

**Example 7.** Let $X = (X_1, X_2)^T$ and $K = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$.

Suppose $X_1$ and $X_2$ are correlated with $\text{cov}(X_1, X_2) = 2$.

**Question:** Find $A$ such that $Y = AX$, $Y = (Y_1, Y_2)^T$ and $Y_1 \& Y_2$ are uncorrelated.

**Solution:** Let

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \ \Rightarrow \ \begin{cases} Y_1 = a_{11}X_1 + a_{12}X_2, \\ Y_2 = a_{21}X_1 + a_{22}X_2. \end{cases}$$

We know that $X \sim N(0, 1)$ and $Y \sim N(0, 1)$, we need $K_{YY}$ to be

$$K_{YY} = \begin{bmatrix} \sigma^2_{Y_1} & 0 \\ 0 & \sigma^2_{Y_2} \end{bmatrix}.$$
Recall that $Y = AX$. Hence,

$$\mu_Y = E[Y],$$
$$= E[AX],$$
$$= AE[X],$$
$$= A\mu_X.$$

By definition, the covariance matrix $K_{YY}$ is

$$K_{YY} = E\left[(Y - \mu_Y)(Y - \mu_Y)^T\right],$$
$$= E\left[A(X - \mu_X)(A(X - \mu_X)^T)\right],$$
$$= AE\left[(X - \mu_X)(A(X - \mu_X)^T)\right],$$
$$= AK_{XX}A^T.$$

By theorem 4 (Eigenvalue Decomposition Theorem) we have:

$$\Lambda = U^T MU.$$ 

Therefore, we need to pick the matrix $A$ such that $A = U^T$ for $K_{YY}$ to be a diagonal matrix.

$$A = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

This leads to the final result

$$Y_1 = \frac{1}{\sqrt{2}}(X_1 - X_2),$$
$$Y_2 = \frac{1}{\sqrt{2}}(X_1 + X_2).$$

4 Multidimensional Jointly Gaussian Distribution

Recall that if two random variables are jointly Gaussian, then the marginal distributions are also Gaussian, but the converse is not necessarily true.

**Definition 8.** A vector $X = (X_1, X_2, \ldots, X_n)^T$ with $E(X) = \mu = (\mu_1, \mu_2, \ldots, \mu_n)^T$ is called jointly Gaussian if

$$f_X(x) = \frac{1}{(2\pi)^{n/2}\sqrt{|K_{XX}|}} \exp\left[\frac{-1}{2}(X - \mu)^T K_X^{-1}(X - \mu)\right],$$

where, $|K_{XX}| = \det(K_{XX}).$
Example 8. For \( n = 1 \),
\[
f_X(x) = \frac{1}{(2\pi)^{1/2}\sigma} \exp \left[ -\frac{1}{2} \left( \frac{X - \mu}{\sigma} \right)^T \frac{1}{\sigma^2} (X - \mu) \right].
\]

Example 9. For \( n = 2 \), \( X = (X_1, X_2)^T \) and the covariance matrix \( K_{XX} \) is defined by
\[
K_{XX} = \begin{bmatrix}
\sigma_{X_1}^2 & Cov(X_1, X_2) \\
Cov(X_1, X_2) & \sigma_{X_2}^2
\end{bmatrix},
\]
\[
= \begin{bmatrix}
\sigma_{X_1}^2 & \rho \sigma_{X_1} \sigma_{X_2} \\
\rho \sigma_{X_1} \sigma_{X_2} & \sigma_{X_2}^2
\end{bmatrix}.
\]
And,
\[
det(K_{XX}) = \sigma_{X_1}^2 \sigma_{X_2}^2 - \rho^2 \sigma_{X_1}^2 \sigma_{X_2}^2,
\]
\[
= (1 - \rho^2) \sigma_{X_1}^2 \sigma_{X_2}^2.
\]
Hence,
\[
f_{X_1, X_2}(x_1, x_2) = \frac{1}{(2\pi)^{1/2}\sigma_{X_1}\sigma_{X_2}\sqrt{1 - \rho^2}} \exp \left[ -\frac{1}{2(1 - \rho^2)} \beta \right],
\]
Where,
\[
\beta = \left( \frac{(x_1 - \mu_{X_1})^2}{\sigma_{X_1}^2} - 2\rho \left( \frac{x_1 - \mu_{X_1}}{\sigma_{X_1}} \right) \left( \frac{x_2 - \mu_{X_2}}{\sigma_{X_2}} \right) + \left( \frac{x_2 - \mu_{X_2}}{\sigma_{X_2}} \right)^2 \right).
\]

Example 10. Let \( X, Y, Z \) be three jointly Gaussian random variables with \( \mu_X = \mu_Y = \mu_Z = 0 \).
\[
K = \begin{bmatrix}
1 & 0.2 & 0.3 \\
0.2 & 1 & 0.3 \\
0.3 & 0.2 & 1
\end{bmatrix},
\]

Question: Find the pdf \( f_{X,Z}(x, z) \).

Answer: From the given information, \( X \) and \( Z \) are jointly Gaussian and
\[
K_{XZ} = \begin{bmatrix}
1 & 0.3 \\
0.3 & 1
\end{bmatrix}.
\]
From \( K_{XZ} \) we know that:
\[
\sigma_X = \sigma_Z = 1, \quad Cov[XZ] = 0.3 \Rightarrow \rho = \frac{0.3}{1} = 0.3.
\]
Therefore,
\[
f_{XZ}(x, z) = \frac{1}{(2\pi)^{1/2}\sqrt{0.91}} \exp \left[ -\frac{1}{2(0.91)} \left( x^2 - 0.6xz + z^2 \right) \right].
\]

7
**Theorem 5.** Let $X$ be jointly Gaussian, $A$ be an invertible matrix and,\[ Y = AX. \]

Then, $Y$ is jointly Gaussian.

**Proof.** From Chapter 3, \( f_Y(y) = \frac{f_X(x)}{|A|} \) but, \[ X = A^{-1}Y; \]

Therefore, \[ f_Y(Y) = \frac{1}{|A|} f_X(A^{-1}Y), \]

\[ f_Y(Y) = \frac{1}{(2\pi)^{n/2} \sqrt{|K_{XX}| |A|}} \exp \left[ -\frac{1}{2} \left( (A^{-1}Y - \mu_X)^T K_{XY}^{-1} (A^{-1}Y - \mu_X) \right) \right]. \]

Recall that \[ \mu_Y = E[Y], \quad \mu_X = A\mu_Y. \] \[ = AE[X], \quad = A\mu_X, \] \[ \Rightarrow \mu_X = A^{-1}\mu_Y. \] \[ \Rightarrow \mu_X = A^{-1}\mu_Y. \] \[ \Rightarrow \mu_X = A^{-1}\mu_Y. \]

In addition, from last lecture we have, \[ K_{YY} = E[YY^T] - \mu_Y\mu_Y^T, \]

\[ = AK_{XX}A^T. \]

Hence, \[ \alpha = -\frac{1}{2} (A^{-1}Y - \mu_X)^T K_{XY}^{-1} (A^{-1}Y - \mu_X), \]

\[ = -\frac{1}{2} A^{-1}(Y - \mu_Y)^T K_{XY}^{-1} A^{-1}(Y - \mu_Y), \]

\[ = -\frac{1}{2} (Y - \mu_Y)^T A^{-1T} K_{XY}^{-1} A^{-1}(Y - \mu_Y). \]

Where, equation (12) result by substituting $\mu_X$ by $A^{-1}\mu_Y$ (from equation (10)). We still need to show that $\beta = \sqrt{|K_{YY}|}$. \[ \text{det}(K_{YY}) = \text{det}(AK_{XX}A^T), \]

\[ = \text{det}(A) \text{det}(K_{XX}) \text{det}(A^T), \]

\[ = \text{det}^2(A) \text{det}(K_{XX}), \]

\[ \Rightarrow \sqrt{|K_{YY}|} = |A| \sqrt{|K_{XX}|}. \]

Hence, $Y$ is jointly Gaussian with $\mu_Y = A\mu_X$ and $K_{YY} = AK_{XX}A^T$. \[ \Box \]
Example 11. Transform $X$ (jointly Gaussian) into $Y = (Y_1, \ldots, Y_n)$ where $Y_i$ are iid.

Since for $Y$ to be iid,
\[ K_{YY} = \begin{bmatrix} \sigma_{Y_1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{Y_1}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{Y_n}^2 \end{bmatrix}, \]

where the covariance is zero and uncorrelated jointly Gaussian random variables are independent. Pick random vector $Y = AX$, where $A$ is to be chosen such that:
\[ K_{YY} = AK_{XX}A^T. \]

Since $K_{XX}$ is symmetric, from the Eigenvalue Decomposition Theorem (see previous lecture) we have,
\[ U^T K_{XX} U = \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}, \]

where $\lambda_n$ are the eigenvalues of $K_{XX}$ and $U = [\Phi_1, \Phi_2, \ldots, \Phi_n]$ is the eigenvector matrix. Hence, $A = U^T$ (Hint: Use the “eig” function in Matlab to generate the matrices).

Lemma 1. If $X_1, X_2, \ldots, X_n$ are jointly Gaussian random variables, then
\[ Z_1 = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n, \]
is a Gaussian random variable $\forall a_i$ such that $\exists i$ for which $a_i \neq 0$.

Remark 1. When asked to find the pdf $f_{Z_1}(Z_1)$, all we have to do is find $E[Z_1]$ and $V(Z_1)$.

Let $a = (a_1, \ldots, a_n)^T$, $Z_1$ can be written as $Z_1 = a^T X$ and
\[ E[Z_1] = a^T \mu_X. \]

However, since $X_1, X_2, \ldots, X_n$ might be dependent,
\[ V(Z_1) \neq a_1^2 V(X_1) + \cdots + a_n^2 V(X_n). \]

For example for $n = 2$ and $\mu_X = 0$,
\[ V(Z_1) = E \left[ (a_1 X_1 + a_2 X_2)^2 \right], \]
\[ = E \left[ a_1^2 X_1^2 + a_2^2 X_2^2 + 2a_1 a_2 X_1 X_2 \right], \]
\[ = a_1^2 \sigma_X^2 + a_2^2 \sigma_X^2 + 2a_1 a_2 \text{Cov}(X_1, X_2). \]
In general:

\[
\text{Var}(Z_1) = E[Z_1^2] - \mu^2 Z_1, \\
= E[Z_1 Z_1^T] - \mu Z_1 \mu^T Z_1, \\
= E[a^T X X^T a] - a^T \mu X \mu^T X^T a, \\
= a^T (E[X X^T] - \mu X \mu^T) a, \\
= a^T K_{XX} a \in \mathbb{R}.
\]

Proof. (of lemma 1) Let,

\[
\begin{bmatrix}
Y_1 \\
Y_2
\end{bmatrix}
= \begin{bmatrix} 1 & 1 \\
3 & 2 \end{bmatrix}
\begin{bmatrix}
X_1 \\
X_2
\end{bmatrix}
= \begin{bmatrix}
X_1 + X_2 \\
3X_1 + 2X_2
\end{bmatrix}.
\]

\(Y_1 = X_1 + X_2 \& Y_2 = 3X_1 + 2X_2\) are Gaussian (theorem 5). We can think of \(Z_1\) being a component of \(Z = (Z_1, Z_2, \ldots, Z_n)^T\) where,

\[
\begin{bmatrix}
Z_1 \\
Z_2 \\
\vdots \\
Z_n
\end{bmatrix}
= \begin{bmatrix}
a_1 & a_2 & \cdots & a_n \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1
\end{bmatrix}
\begin{bmatrix}
X_1 \\
X_2 \\
\vdots \\
X_n
\end{bmatrix}
= \begin{bmatrix}
a_1 X_1 + a_2 X_2 + \cdots + a_n X_n \\
X_2 \\
\vdots \\
X_n
\end{bmatrix}.
\]

We know that \(A\) is invertible (full rank) which means that \(Z\) is jointly Gaussian (theorem 5). Thus, each component of \(Z\) is Gaussian, in particular \(Z_1\).

Remark 2. Any linear combination of the components of a jointly Gaussian random vector is a Gaussian random variable.

5 Overview on Estimation

Recall:

1. Tossing a die \(X \in \{0, 1, 2, 3, 4, 5, 6\}\), we want to estimate \(X\) by \(\hat{X}\).

   What is the best estimate?

   \(MSE = E[(X - \hat{X})^2]\).

   We want to minimize \(E[(X - \hat{X})^2]\)

   Take \(\hat{X}_{\text{min}} = E[X]\)

   (check previous notes)

2. Find the Minimum Mean Square Error (MMSE) of \(X\) given \(Y\).

   \(\hat{X}_{\text{MMSE}} = E[X|Y]\).
3. Linear MMSE (LMMSE)

Here $\hat{X}_{M\text{MSE}} = aY + b$.

$$\min_{a,b} E[(X - \hat{X})^2] \iff (X - \hat{X}) \perp Y.$$ 

Recall that we say $X$ is orthogonal to $Y (X \perp Y)$ if and only if $E[XY] = 0$.

By the orthogonality principle, we know that if $X_1 \perp X_2 \Rightarrow E[X_1,X_2] = 0$.

Thus, $E[(X - \hat{X})Y] = 0$.

$\hat{X}_{L\text{MSE}} = \frac{\rho \sigma_X}{\sigma_Y} (Y - \mu_Y) + \mu_X$.

Where $\rho = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$.

So,

$$\hat{X}_{L\text{MSE}} = \frac{\text{Cov}(X,Y)}{\sigma_Y^2} (Y - \mu_Y) + \mu_X.$$ 

$L\text{MSE} = E[(X - \hat{X}_{L\text{MSE}})^2] = E(X^2) - E(\hat{X}^2) = ||X||^2 - ||\hat{X}||^2$.

Recall that $E[X^2] = ||X||^2$.

Example 12.

$$f_{XY} = \begin{cases} 
2e^{-x}e^{-y} & \text{if } 0 \leq y \leq x < \infty, \\
0 & \text{otherwise.}
\end{cases}$$

1. Find MMSE and LMMSE of $X$ given $Y$

$\hat{X}_{M\text{MSE}} = E[X|Y] = Y + 1$. (Check exam solution for a detailed proof.)

Since $\hat{X}_{M\text{MSE}}$ is linear then,

$\hat{X}_{L\text{MSE}} = Y + 1$.

Straight calculations give $\mu_X = 3/2, \mu_y = 1/2, \text{Var}(X) = 5/4, \text{Var}(Y) = 1/4, \text{and Cov}(X,Y) = 1/4$.

2. Find the MMSE & LMMSE of $Y$ given $X$.

First, we will find the MMSE; but to do this we need to calculate the covariance of $X$ and $Y$.

$\text{Cov}(XY) = E[XY] - \mu_x \mu_y$.

$E[XY] = \int \int xy f(x,y) dx dy = \int_0^{+\infty} \int_0^x 2xye^{-x}e^{-y}dy dx = 1.$

$\text{Cov}(XY) = 1 - 3/2 \times 1/2 = 1/4.$
Usually, finding the LMMSE is much easier than finding the MMSE because you simply apply to formula.

\[ \hat{Y}_{\text{LMMSE}} = \frac{\text{Cov}(XY)}{\sigma^2_x} (X - \mu_x) + \mu_y. \]

Thus, if you restrict yourself to linear functions of the form \( aX + b \), then the best choices are \( a = 1/5 \) and \( b = 1/5 \).

Next, we will find the best MMSE estimator. Recall the definition of the best MMSE estimator.

\[ \hat{Y}_{\text{MMSE}} = E[Y|X]. \]

\[ \hat{Y}_{\text{MMSE}} = \int y f_{Y|X}(y|x) \, dy. \]

\[ \hat{Y}_{\text{MMSE}} = \int_0^x y \frac{e^{-y}}{1 - e^{-x}} \, dy = \frac{-e^{-y}(y + 1)}{1 - e^{-x}} \bigg|_0^x = 1 - \frac{xe^{-x}}{1 - e^{-x}}. \]

As homework, find the error associated with each estimate.

### 6 The Orthogonality Principle

**Theorem 6** (The Orthogonality Principle). The MMSE of \( \hat{X} \) of \( X \) given \( Y \), where \( \hat{X} = g(Y) \), where \( g(*) \in \Gamma \) and (\( \Gamma^* \) is all functions, linear functions, constants), is found when \( \hat{X} = \min E[(X - g(Y))^2] \) where the minimization is over \( g(*) \in \Gamma \). The MMSE = \( E[X^2] - E[X]^2 \). In this case, \( \hat{X} \) is unique and the error is orthogonal to the observation \( (X - \hat{X}) \perp Y \). The * indicates there are some technical conditions on gamma but they are not discussed here.

**Proof.** Proof is omitted. \( \square \)

**Example 13.** \( X = (X_1, X_2, X_3) \) are jointly Gaussian and, \( \mu_x = (0, 0, 0) \),

\[ K_{XX} = R_{XX} = \begin{bmatrix} 1 & 0.2 & 0.1 \\ 0.2 & 2 & 0.3 \\ 0.1 & 0.3 & 4 \end{bmatrix}. \]

Find the LMMSE of \( X_3 \) Given \( X_1 \) and \( X_2 \).
Because all $\mu_x = 0$,

$$K_{X_3 Y} = \text{Cov}(X_3X_1) \text{ Cov}(X_3X_2) = [0.1 \ 0.3].$$

$\hat{X}_3 \text{ LMMSE} = [0.1 \ 0.3].$

$[K_{YY}^{-1}] = a_1X_1 + a_2X_2,$ \quad $a_1 = 0.0714,$ \quad $a_2 = 0.1429.$

Find the MMSE of the $X_3$.

$$\hat{X}_3 \text{ MMSE} = E[(X_3 - \hat{X})^2] = E[X_3^2] - E[\hat{X}^2]$$

$$= 4 - E[(a_1X_1 + a_2X_2)^2]$$

$$= 4 - a_1^2E[X_1^2] - a_2^2E[X_2^2] - 2a_1a_2E[X_1X_2]$$

$$= 3.95.$$ 

7 MMSE Based on Vector Observation

**Theorem 7.** The Linear Minimum Mean-Square Estimate LMMSE $\hat{X}_{LMMSE}$ of $X$ given an observed random vector $Y = (Y_1, \ldots, Y_n)^T$ is given by

$$\hat{X}_{LMMSE} = K_{XY}^T K_{YY}^{-1} (Y - \mu_Y) + \mu_X,$$

where,

$$\mu_X = E[X],$$

$$\mu_Y = (E[Y_1], E[Y_2], \ldots, E[Y_n]),$$

$$K_{YY} = E[YY^T] - \mu_Y \mu_Y^T,$$

and $K_{XY} = (\text{Cov}[XY_1], \text{Cov}[XY_2], \ldots, \text{Cov}[XY_n])^T$,

where $K_{YY}$ is the covariance matrix of $Y$.

And, the MMSE is given by

$$\text{MMSE} = \min E[(X - \hat{X}_{LMMSE})^2]$$

$$= E[X^2] - E[\hat{X}_{LMMSE}^2].$$
Proof. First, let us assume that $\mu_X = 0$ and $\mu_Y = 0$. Then, we can write

$$\hat{X}_{LMMSE} = a_1Y_1 + a_2Y_2 + \cdots + a_nY_n = a^tY.$$  

By the orthogonality principle: $(X - \hat{X}_{LMMSE}) \perp Y_i \quad i = 1, 2, \ldots, n$ ,

$$E[a^t Y \cdot Y_i] = E[XY_i] \quad i = 1, 2, \ldots, n,$$

$$E[(a_1Y_1 + a_2Y_2 + \cdots + a_nY_n)Y_i] = E[XY_i] \quad i = 1, 2, \ldots, n.$$  

So, we get the following $n \times n$ linear system with $n$ unknowns, $a_1, \ldots, a_n$:

$$a_1E[Y_1] + a_2E[Y_1Y_2] + \cdots + a_nE[Y_1Y_n] = E[XY_1],$$

$$a_1E[Y_2] + a_2E[Y_2Y_2] + \cdots + a_nE[Y_2Y_n] = E[XY_2],$$

$$\vdots$$

$$a_1E[Y_n] + a_2E[Y_nY_2] + \cdots + a_nE[Y_nY_n] = E[XY_n].$$

In matrix form, this can be written as

$$a^t R_{YY} = R_{XY}^t,$$

$$a^t = R_{XY}^t R_{YY}^{-1}.$$  

Where,

$$K_{YY} = \begin{bmatrix}
E[Y_1^2] & E[Y_1Y_2] & \cdots & E[Y_1Y_n]
E[Y_2Y_1] & E[Y_2^2] & \cdots & E[Y_2Y_n]
\vdots & \vdots & \ddots & \vdots
E[Y_nY_1] & E[Y_nY_2] & \cdots & E[Y_n^2]
\end{bmatrix},$$

and,

$$K_{XY} \overset{\text{def}}{=} \begin{bmatrix}
\text{Cov}[XY_1]
\text{Cov}[XY_2]
\vdots
\text{Cov}[XY_n]
\end{bmatrix} = \begin{bmatrix}
E[XY_1]
E[XY_2]
\vdots
E[XY_n]
\end{bmatrix}.$$  

So,

$$\hat{X}_{LMMSE} = K_{XY}^T K_{YY}^{-1} Y.$$  

In general, if $\mu_X \neq 0$ and $\mu_Y \neq 0$, apply the same method above to $X' = X - \mu_X$ and $Y' = Y - \mu_Y$, then we get

$$\hat{X}_{LMMSE} = K_{XY}^T K_{YY}^{-1} (Y - \mu_Y) + \mu_X.$$  

Example 14. Multiple Antenna Receiver

Assume 2 antennas receive signals independently. $Y_1 = X + N_1$, $Y_2 = X + N_2$, $X \sim N(0, 2)$, $N_1, N_2 \sim N(0, 1)$. Assume they are all independent.
1. Find the LMMSE of $X$ given $Y_1$.

$$
\hat{X}_{LMMSE} = \frac{\text{Cov}(XY_1)}{V(Y_1)} Y_1.
$$

$$
\text{Cov}(XY_1) = E[XY_1] - E[X]E[Y_1] \quad \text{Note that } E[X]E[Y_1] = 0
$$

$$
= E[X(X + N_2)]
$$

$$
= E[X^2] + E[XN_2] = 2 + 0 = 2.
$$

$$
V(Y_1) = V(X) + V(N_1) = 2 + 1 = 3.
$$

So that, $\hat{X}_{LMMSE} = \frac{2}{3} Y_1$

$$
X_{MMSE} = E[X^2] - E[\hat{X}^2]
$$

$$
= 2 - E[(\frac{2}{3}Y_1)^2]
$$

$$
= 2 - \frac{4}{9}E[Y_1^2] = \frac{2}{3}.
$$

2. Find the LMMSE of $X$ given $Y_1$ and $Y_2$.

Usually, we want to find that $\hat{X} = a_1 Y_1 + a_2 Y_2 + C$.

In this case, $C = 0$.

While $X - \hat{X} \perp Y_1$, and $X - \hat{X} \perp Y_2$, we can obtain,

$$
E[(X - a_1 Y_1 - a_2 Y_2) Y_1] = 0.
$$

$$
E[(X - a_1 Y_1 - a_2 Y_2) Y_2] = 0.
$$

$$
a_1 E[Y_1^2] + a_2 E[Y_1 Y_2] = E[XY_1].
$$

$$
a_1 E[Y_1 Y_2] + a_2 E[Y_2^2] = E[XY_2].
$$

$$
K_{Y_1 Y_2} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = K_{XY}.
$$

Therefore,

$$
\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = K_{Y_1 Y_2}^{-1} K_{XY} = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 2 \end{bmatrix}.
$$
And,

\[ MMSE = E[X^2] - E[\hat{X}_{LMMSE}^2] \]
\[ = 2 - E[0.4(Y_1 + Y_2)^2] \]
\[ = 0.4 < MMSE \text{ with only } Y_1. \]