ECE511: Analysis of Random Signals Fall 2016

Chapter 5

Dr. Salim El Rouayheb

Scribe: H. Xie, L. Gan, W. Yi, Jaijo, R. Tajeddine, S. Jones, Y. Yang, L. Liu

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

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

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

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

$$\begin{aligned}
K_{XX} &= E\left[\begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_n - \mu_n \end{pmatrix} \begin{pmatrix} X_1 - \mu_1 & X_2 - \mu_2 & \ldots & X_n - \mu_n \end{pmatrix}^{\mathrm{T}}\right], \\
&= E\begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) & \cdots & (X_1 - \mu_1)(X_n - \mu_n) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 & \cdots & (X_2 - \mu_2)(X_n - \mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ (X_n - \mu_n)(X_1 - \mu_1) & (X_n - \mu_n)(X_2 - \mu_2) & \cdots & (X_n - \mu_n)^2 \end{bmatrix}, \\
&= \begin{bmatrix} \sigma_1^2 & K_{12} & \cdots & K_{1n} \\ K_{21} & \sigma_2^2 & \cdots & K_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ K_{n1} & K_{n2} & \cdots & \sigma_n^2 \end{bmatrix}.
\end{aligned}$$

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

Definition 4. *The correlation matrix R_{XX}, or R, is defined as $R = E\left[\underset{\sim}{X}\underset{\sim}{X}^T\right]$.*

Corollary 1. $K = R - \underset{\sim}{\mu}\underset{\sim}{\mu}^T$.

Example 1. $\underset{\sim}{X} = (X_1, X_2)$,

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

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

Definition 5. *For any random vectors $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ of same length.*

1. If the cross-covariance matrix $K_{XY} = E\left[\left(\underset{\sim}{X} - \underset{\sim}{\mu}_X\right)\left(\underset{\sim}{Y} - \underset{\sim}{\mu}_Y\right)\right] = E\left[\underset{\sim}{X}\underset{\sim}{Y}^T\right] - \underset{\sim}{\mu}_X\underset{\sim}{\mu}_Y^T = \mathbf{0} \Rightarrow$ we say that $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ are uncorrelated.
2. If $E\left[\underset{\sim}{X}\underset{\sim}{Y}^T\right] = 0 \Rightarrow$ we say that $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ 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* $\underset{\sim}{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$*?*
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 $Y + X_1 - X_2$,

$$\begin{aligned} V(Y) &= V(X_1 - X_2) \\ &= V(X_1) + V(X_2) - 2cov(X_1, X_2) \\ &= 2 + 2 - 2 \times 3 \\ &= -2 \end{aligned}$$

So M cannot be covariance matrix.

Therefore we want for any linear combination of $\underset{\sim}{X} = (X_1, \ldots, X_n)$, say $\underset{\sim}{Y} = a_1X_1 + \ldots, +a_nX_n$, to have $V(Y) \geq 0$.

$$\begin{aligned} V(Y) &= E(Y^2) - (E(Y))^2 \\ E(Y) &= E[\underset{\sim}{a}^T\underset{\sim}{X}] = \underset{\sim}{a}^T\underset{\sim}{\mu}_X \\ E[Y^2] &= E[(\underset{\sim}{a}^T\underset{\sim}{X})(\underset{\sim}{a}^T\underset{\sim}{X})] = E[\underset{\sim}{a}^T\underset{\sim}{X}\cdot\underset{\sim}{X}^T\underset{\sim}{a}] \\ &= \underset{\sim}{a}^T E[\underset{\sim}{X}\cdot\underset{\sim}{X}^T]\underset{\sim}{a} \\ \Longrightarrow \quad V(Y) &= \underset{\sim}{a}^T E[\underset{\sim}{X}\cdot\underset{\sim}{X}^T]\underset{\sim}{a} - \underset{\sim}{a}^T\underset{\sim}{\mu}_X\underset{\sim}{\mu}_X^T\underset{\sim}{a} \\ &= \underset{\sim}{a}^T K_{XX}\underset{\sim}{a} \quad \text{should be } \geq 0 \end{aligned}$$

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

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

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

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

$$\begin{aligned} \underset{\sim}{X}^T I \underset{\sim}{X} &= \begin{pmatrix} X_1 & X_2 \end{pmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \\ &= ||\underset{\sim}{X}||^2 \geq 0. \end{aligned}$$

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

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

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

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

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

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

$$K = E\left[\underset{\sim}{X}\underset{\sim}{X}^T\right].$$

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

$$\begin{aligned} \underset{\sim}{Z}^T K \underset{\sim}{Z} &= \underset{\sim}{Z}^T E\left[\underset{\sim}{X}\underset{\sim}{X}^T\right] \underset{\sim}{Z}, && (1) \\ &= E\left[\underset{\sim}{Z}^T \underset{\sim}{X}\underset{\sim}{X}^T \underset{\sim}{Z}\right], && (2) \\ &= E\left[\left(\underset{\sim}{Z}^T \underset{\sim}{X}\right)\left(\underset{\sim}{Z}^T \underset{\sim}{X}\right)^T\right], && (3) \\ &= E\left[Y^2\right] \geq 0. && (4) \\ & && (5) \end{aligned}$$

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) $Y = \underset{\sim}{Z}^T \underset{\sim}{X}$ is a single random variable. □

Definition 7. *The eigenvalues of a matrix M are the scalars λ such that*

$$\exists \underset{\sim}{\Phi} \neq 0, M\underset{\sim}{\Phi} = \lambda \underset{\sim}{\Phi}. \quad (6)$$

The vectors $\underset{\sim}{\Phi}$ are called eigenvectors. Typically we choose ϕ_i such that $||\phi_i|| = 1$.

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. □

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 $\underline{\Phi}_1 = \begin{bmatrix} \Phi_{11} & \Phi_{21} \end{bmatrix}^T$ such that
$$\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} \Phi_{11} \\ \Phi_{12} \end{bmatrix} = 2 \begin{bmatrix} \Phi_{11} \\ \Phi_{12} \end{bmatrix}.$$
$$\left.\begin{aligned} 4\Phi_{11} + 2\Phi_{12} = 2\Phi_{11} \\ 2\Phi_{11} + 4\Phi_{12} = 2\Phi_{12} \end{aligned}\right\} \Rightarrow \Phi_{11} = -\Phi_{21} \Rightarrow \underline{\Phi}_1 = \begin{bmatrix} 1 & -1 \end{bmatrix}^T.$$
For $\lambda_2 = 6$: we repeat the same steps and get
$$\underline{\Phi}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^T.$$

Claim 1. *(Eigenvalue Decomposition) The matrix M having $\underline{\Phi}_1$, $\underline{\Phi}_2$ as eigenvectors can be expressed as*
$$M = U \Lambda U^{\mathrm{T}},$$
Where
$$U = \begin{bmatrix} \underline{\Phi}_1 & \underline{\Phi}_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix},$$
$$\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}.$$

Check:
$$\begin{aligned} U\Lambda U^{\mathrm{T}} &= \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \\ &= \frac{1}{2} \begin{bmatrix} 2 & 6 \\ -2 & 6 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \\ &= \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}, \\ &= M. \end{aligned}$$

Theorem 4. *(Eigenvalue Decomposition Theorem) Let M be a real symmetric matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and corresponding eigenvectors $\underline{\Phi}_1, \underline{\Phi}_2, \ldots, \underline{\Phi}_n$ then*

$$U^{\mathrm{T}} M U = \Lambda,$$

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) :

$$MU = U\Lambda \text{ and } U = \begin{bmatrix} | & & | \\ \phi_1 & \cdots & \phi_n \\ | & & | \end{bmatrix},$$

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

Since U is a real symmetric matrix :

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

and

$$\begin{aligned} M &= (U^{\mathrm{T}})^{-1} \Lambda U^{-1}, \\ &= U \Lambda U^{\mathrm{T}}. \end{aligned}$$

□

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

Suppose X_1 and X_2 are correlated with cov $(X_1, X_2) = 2$.

Question: Find A such that $\underline{Y} = A\underline{X}$, $\underline{Y} = (Y_1,\ Y_2)^T$ and Y_1 & Y_2 are uncorrelated.

Solution: Let

$$\left.\begin{aligned} A &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\ \underline{Y} &= \begin{pmatrix} Y_1 & Y_2 \end{pmatrix}^T \end{aligned}\right\} \Rightarrow \begin{aligned} Y_1 &= a_{11} X_1 + a_{12} X_2, \\ Y_2 &= a_{21} X_1 + a_{22} X_2. \end{aligned}$$

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

$$K_{YY} = \begin{bmatrix} \sigma_{Y_1}^2 & 0 \\ 0 & \sigma_{Y_2}^2 \end{bmatrix}.$$

Recall that $\underline{Y} = A\underline{X}$. Hence,

$$\begin{aligned}
\underline{\mu}_Y &= E\left[\underline{Y}\right], \\
&= E\left[A\underline{X}\right], \\
&= AE\left[\underline{X}\right], \\
&= A\underline{\mu}_X.
\end{aligned}$$

By definition, the covariance matrix K_{YY} is

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

By theorem 4 (Eigenvalue Decomposition Theorem) we have:

$$\Lambda = U^{\mathrm{T}}MU.$$

Therefore, we need to pick the matrix A such that $A = U^{\mathrm{T}}$ 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

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

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* $\underline{X} = (X_1, X_2, \ldots, X_n)^T$ *with* $E(\underline{X}) = \underline{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^T$ *is called jointly Gaussian if*

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

where, $|K_{XX}| = \det(K_{XX})$.

Example 8. *For $n = 1$,*

$$f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left[\frac{-1}{2}(\underline{X} - \underline{\mu})^T \frac{1}{\sigma^2}(\underline{X} - \underline{\mu})\right].$$

Example 9. *For $n = 2$, $\underline{X} = (X_1, X_2)^T$ and the covariance matrix K_{XX} is defined by*

$$\begin{aligned} K_{XX} &= \begin{bmatrix} \sigma_{X_1}^2 & Cov(X_1, X_2) \\ Cov(X_1, X_2) & \sigma_{X_2}^2 \end{bmatrix}, \\ &= \begin{bmatrix} \sigma_{X_1}^2 & \rho\sigma_{X_1}\sigma_{X_2} \\ \rho\sigma_{X_1}\sigma_{X_2} & \sigma_{X_2}^2 \end{bmatrix}. \end{aligned}$$

And,

$$\begin{aligned} \det(K_{XX}) &= \sigma_{X_1}^2\sigma_{X_2}^2 - \rho^2\sigma_{X_1}^2\sigma_{X_2}^2, \\ &= (1 - \rho^2)\sigma_{X_1}^2\sigma_{X_2}^2. \end{aligned}$$

Hence,

$$f_{X_1X_2}(x_1, x_2) = \frac{1}{(2\pi)\sigma_{X_1}\sigma_{X_2}\sqrt{1 - \rho^2}} \exp\left[\frac{-1}{2(1 - \rho^2)}\beta\right],$$

Where,

$$\beta = \left(\frac{(x_{X_1} - \mu_{X_1})^2}{\sigma_{X_1}} - 2\rho\left(\frac{x_{X_1} - \mu_{X_1}}{\sigma_{X_1}}\right)\left(\frac{x_{X_2} - \mu_{\mu_{X_2}}}{\sigma_{X_2}}\right) + \frac{(x_{X_2} - \mu_{X_2})^2}{\sigma_{X_2}}\right).$$

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:

$$\left.\begin{aligned} \sigma_X = \sigma_Z = 1 \\ Cov[XZ] = 0.3 \end{aligned}\right\} \Rightarrow \rho = \frac{0.3}{1} = 0.3.$$

Therefore,

$$f_{XZ}(x, z) = \frac{1}{(2\pi)\sqrt{0.91}} \exp\left[\frac{-1}{2(0.91)}\left(x^2 - 0.6xz + z^2\right)\right].$$

Theorem 5. *Let $\underline{X}$ be jointly Gaussian, A be an invertible matrix and,*

$$\underline{Y} = A\underline{X}.$$

Then, $\underline{Y}$ is jointly Gaussian.

Proof. From Chapter 3, $f_Y(y) = \frac{f_X(x)}{|A|}$ but,

$$\underline{X} = A^{-1}\underline{Y},$$

Therefore,

$$f_{\underline{Y}}(Y) = \frac{1}{|A|} f_{\underline{X}}\left(A^{-1}Y\right),$$
$$f_{\underline{Y}}(Y) = \frac{1}{(2\pi)^{n/2}\underbrace{\sqrt{|K_{XX}|}|A|}_{\beta}} \exp\underbrace{\left[-\frac{1}{2}\left(\left(A^{-1}\underline{Y} - \underline{\mu}_X\right)^T K_{XY}^{-1}(A^{-1}\underline{Y} - \underline{\mu}_X)\right)\right]}_{\alpha}.$$

Recall that

$$\underline{\mu}_Y = E[\underline{Y}], \tag{7}$$
$$= AE[\underline{X}], \tag{8}$$
$$= A\underline{\mu}_X, \tag{9}$$
$$\Rightarrow \underline{\mu}_X = A^{-1}\underline{\mu}_Y. \tag{10}$$

In addition, from last lecture we have,

$$K_{YY} = E[\underline{Y}\underline{Y}^T] - \underline{\mu}_Y\underline{\mu}_Y^T,$$
$$= AK_{XX}A^T.$$

Hence,

$$\alpha = \frac{-1}{2}(A^{-1}\underline{Y} - \underline{\mu}_X)^T K_{XY}^{-1}(A^{-1}\underline{Y} - \underline{\mu}_X), \tag{11}$$
$$= \frac{-1}{2}A^{-1}(\underline{Y} - \underline{\mu}_Y)^T K_{XY}^{-1}A^{-1}(\underline{Y} - \underline{\mu}_Y), \tag{12}$$
$$= \frac{-1}{2}(\underline{Y} - \underline{\mu}_Y)^T \underbrace{{A^{-1}}^T K_{XY}^{-1} A^{-1}}_{K_{YY}}(\underline{Y} - \underline{\mu}_Y). \tag{13}$$

Where, equation (12) result by substituting $\underline{\mu}_X$ by $A^{-1}\underline{\mu}_Y$ (from equation (10)). We still need to show that $\beta = \sqrt{|K_{YY}|}$.

$$\det(K_{YY}) = \det(AK_{XX}A^T),$$
$$= \det(A)\det(K_{XX})\det(A^T),$$
$$= \det^2(A)\det(K_{XX}),$$
$$\Rightarrow \sqrt{|K_{YY}|} = |A|\sqrt{|K_{XX}|}.$$

Hence, $\underline{Y}$ is jointly Gaussian with $\underline{\mu}_Y = A\underline{\mu}_X$ and $K_{YY} = AK_{XX}A^T$. □

Example 11. *Transform $\underline{X}$ (jointly Gaussian) into $\underline{Y} = (Y_1, \dots, Y_n)$ where Y_i are iid.*

Since for $\underline{Y}$ 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 $\underline{Y} = A\underline{X}$, 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,

$$U^T K_{XX} U = \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix},$$

where λ_n are the eigenvalues of K_{XX} and $U = [\underline{\Phi}_1,\ \underline{\Phi}_2,\ \dots, \underline{\Phi}_n]$ is the eigenvector matrix. Hence, $A = U^T$ (Hint: Use the "eig" function in Matlab to generate the matrices).

Lemma 1. *If $X_1,\ X_2,\ \dots,\ X_n$ are jointly Gaussian random variables, then*

$$Z_1 = a_1X_1 + a_2X_2 + \cdots + a_nX_n,$$

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 $\underline{a} = (a_1, \dots, a_n)^T$, Z_1 can be written as $Z_1 = \underline{a}^T\underline{X}$ and

$$E[Z_1] = \underline{a}^T\underline{\mu}_X.$$

However, since $X_1, X_2, \dots, X_n$ 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 $\underline{\mu}_X = \underline{0}$,

$$\begin{aligned} V(Z_1) &= E\left[(a_1X_1 + a_2X_2)^2\right], \\ &= E\left[a_1^2X_1^2 + a_2^2X_2^2 + 2a_1a_2X_1X_2\right], \\ &= a_1^2\sigma_{X_1}^2 + a_2^2\sigma_{X_1}^2 + 2a_1a_2Cov(X_1, X_2). \end{aligned}$$

In general:

$$\begin{aligned}
Var\left(Z_1\right) &= E\left[Z_1\right]^2 - \mu_{Z_1}^2, \\
&= E\left[Z_1 Z_1^T\right] - \mu_{Z_1}\mu_{Z_1}^T, \\
&= E\left[\underline{a}^T \underline{X}\,\underline{X}^T \underline{a}\right] - \underline{a}^T \underline{\mu}_X \underline{\mu}_X^T \underline{a}, \\
&= \underline{a}^T \left(E\left[\underline{X}\,\underline{X}^T\right] - \mu_X \mu_X^T\right) \underline{a}, \\
&= \underline{a}^T K_{XX} \underline{a} \in \mathbb{R}.
\end{aligned}$$

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 $\underline{Z} = \left(Z_1,\ Z_2,\ \ldots,\ Z_n\right)^T$ where,

$$\begin{bmatrix} Z_1 \\ Z_2 \\ \cdots \\ Z_n \end{bmatrix} = \underbrace{\begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}}_{A} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} a_1X_1 + a_2X_2 + \cdots + a_nX_n \\ X_2 \\ \vdots \\ X_n \end{bmatrix}.$$

We know that A is invertible (full rank) which means that $\underline{Z}$ is jointly Gaussian (theorem 5). Thus, each component of $\underline{Z}$ 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 $\hat{X}_{min} = E[X]$
   (check previous notes)

2. Find the Minimum Mean Square Error (MMSE) of X given Y.
   $\hat{X}_{MMSE} = E[X|Y]$.

3. Linear MMSE (LMMSE)
   Here $\hat{X}_{MMSE} = aY + b$.

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

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

$$\hat{X}_{LMMSE} = \frac{\rho\sigma_X}{\sigma_Y}(Y - \mu_Y) + \mu_X,$$

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

$$\hat{X}_{LMMSE} = \frac{Cov(X,Y)}{\sigma_Y^2}(Y - \mu_Y) + \mu_X.$$
$$LMMSE = E[(X - \hat{X}_{LMMSE})^2]$$
$$= E(X^2) - E(\hat{X}^2) = ||X||^2 - ||\hat{X}||^2.$$

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

Example 12.

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

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

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

   Since $\hat{X}_{MMSE}$ *is linear then,*

   $\hat{X}_{LMMSE} = Y + 1$.
   Straight calculations give $\mu_X = 3/2, \mu_y = 1/2, Var(X) = 5/4, Var(Y) = 1/4,$ *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.

$$Cov(XY) = E[XY] - \mu_x\mu_y.$$
$$E[XY] = \iint xy\ f(x,y)dx\ dy = \int_0^{+\infty}\int_0^x 2xye^{-x}e^{-y}dy\ dx = 1.$$
$$Cov(XY) = 1 - 3/2 \times 1/2 = 1/4.$$

Usually, finding the LMMSE is much easier than finding the MMSE because you simply apply to formula.

$$\hat{Y}_{LMMSE} = \frac{Cov(XY)}{\sigma_x^2}(X - \mu_x) + \mu_y.$$
$$\hat{Y}_{LMMSE} = \frac{1/4}{5/4}(X - 3/2) + 1/2 = \; X/5 - 1/5.$$

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.

$$\hat{Y}_{MMSE} = E\left[Y|X\right].$$

$$\hat{Y}_{MMSE} = \int y f_{Y|X}\left(y|x\right) dy.$$

$$\hat{Y}_{MMSE} = \int_0^x y\frac{e^{-y}}{1-e^{-x}}\; dy = \left.\frac{-e^{-y}(y+1)}{1-e^{-x}}\right|_0^x = 1 - \frac{xe^{-x}}{1-e^{-x}}.$$

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 (Γ^* is all functions, linear functions, constants), is found when $\hat{X} = \min E[(X - g\left(Y\right))^2]$ where the minimization is over $g\left(*\right) \in \Gamma$. The MMSE $= E[X^2] - E[\hat{X}^2]$. 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. □

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.

$$K_{YY} = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 2 \end{bmatrix},$$

$$\Rightarrow K_{YY}^{-1} = \begin{bmatrix} 1.0204 & -0.102 \\ -0.102 & 0.5102 \end{bmatrix}.$$

Because all $\mu_x = 0$*,*

$$K_{X_3Y}^T = [Cov(X_3X_1) \;\; Cov(X_3X_2)] = [0.1 \;\; 0.3].$$

$\hat{X}_{3\ LMMSE} = [0.1 \;\; 0.3].$

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

Find the MMSE of the X_3*.*

$$\begin{aligned} \hat{X}_{3\ MMSE} &= E[(X_3 - \hat{X})^2] = E[X_3^2] - E[\hat{X}^2] \\ &= 4 - E[(a_1X_1 + a_2X_2)^2] \\ &= 4 - a_1^2E[X_1^2] - a_2^2E[X_2^2] - 2a_1a_2E[X_1X_2] \\ &= 3.95. \end{aligned}$$

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* $\underline{Y} = (Y_1, \ldots, Y_n)^T$ *is given by*

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

where,

$$\begin{aligned} \mu_X &= E[X], \\ \underline{\mu}_Y &= (E[Y_1], E[Y_2], \ldots, E[Y_n]), \\ K_{YY} &= E[\underline{Y}\underline{Y}^T] - \mu_Y\mu_Y^T, \\ \text{and } K_{XY} &= (Cov[XY_1], Cov[XY_2], \ldots, Cov[XY_n])^T, \end{aligned}$$

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

And, the MMSE is given by

$$\begin{aligned} MMSE &= \min E[(X - \hat{X}_{LMMSE})^2] \\ &= E[X^2] - E[\hat{X}_{LMMSE}{}^2]. \end{aligned}$$

Proof. First, let us assume that $\mu_X = 0$ and $\underline{\mu}_Y = \underline{0}$. Then, we can write

$$\begin{aligned}\hat{X}_{LMMSE} &= a_1 Y_1 + a_2 Y_2 + \cdots + a_n Y_n \\ &= \underline{a}^t \underline{Y}.\end{aligned}$$

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

$$E[\underline{a}^t \underline{Y} \cdot Y_i] = E[XY_i] \;\; i = 1, 2, \ldots, n,$$

$$E[(a_1 Y_1 + a_2 Y_2 + \cdots + a_n Y_n) Y_i] = E[XY_i] \;\; i = 1, 2, \ldots, n.$$

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

$$\begin{aligned} a_1 E[Y_1^2] + a_2 E[Y_1 Y_2] + \cdots + a_n E[Y_1 Y_n] &= E[XY_1], \\ a_1 E[Y_2 Y_1] + a_2 E[Y_2^2] + \cdots + a_n E[Y_2 Y_n] &= E[XY_2], \\ &\vdots \\ a_1 E[Y_n Y_1] + a_2 E[Y_n Y_2] + \cdots + a_n E[Y_n^2] &= E[XY_n]. \end{aligned}$$

In matrix form, this can be written as

$$\begin{aligned}\underline{a}^t R_{YY} &= R_{XY}^t, \\ \underline{a}^t &= R_{XY}^t R_{YY}^{-1}.\end{aligned}$$

Where,

$$K_{YY} = \begin{bmatrix} E[Y_1^2] & E[Y_1 Y_2] & \ldots & E[Y_1 Y_n] \\ E[Y_2 Y_1] & E[Y_2^2] & \ldots & E[Y_2 Y_n] \\ \vdots & \vdots & & \vdots \\ E[Y_n Y_1] & E[Y_n Y_2] & \ldots & E[Y_n^2] \end{bmatrix},$$

and,

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

So,

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

In general, if $\mu_X \neq 0$ and $\underline{\mu}_Y \neq \underline{0}$,
Apply the same method above to $X' = X - \mu_X$ and $\underline{Y}' = \underline{Y} - \underline{\mu}_Y$, then we get

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

□

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.*

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

$$\begin{aligned} Cov(XY_1) &= E[XY_1] - E[X]E[Y_1] \qquad \textit{Note that } E[X]E[Y_1] = 0 \\ &= E[X(X + N_2)] \\ &= E[X^2] + E[XN_2] = 2 + 0 = 2. \end{aligned}$$

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

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

$$\begin{aligned} X_{MMSE} &= E[X^2] - E[\hat{X}^2] \\ &= 2 - E[(\frac{2}{3}Y_1)^2] \\ &= 2 - \frac{4}{9}E[Y_1^2] = \frac{2}{3}. \end{aligned}$$

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

Usually, we want to find that $\hat{X} = a_1Y_1 + a_2Y_2 + C$.
In this case, $C = 0$.
While $\quad X - \hat{X} \perp Y_1, \quad$ *and* $\quad X - \hat{X} \perp Y_2,$
we can obtain,

$$\begin{aligned} &E[(X - aY_1 - a_2Y_2)Y_1] = 0. \\ &E[(X - aY_1 - a_2Y_2)Y_2] = 0. \\ &\quad a_1E[Y_1^2] + a_2E[Y_1Y_2] = E[XY_1]. \\ &\quad a_1E[Y_1Y_2] + a_2E[Y_2^2] = E[XY_2]. \end{aligned}$$

$$K_{Y_1Y_2} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = K_{XY}.$$

Therefore,

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = K_{Y_1Y_2}^{-1} K_{XY} = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 2 \end{bmatrix}.$$

And,

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