ECE4110: Random Processes

MMSE and Linear MMSE Estimation

Qing Zhao
School of Electrical and Computer Engineering
Cornell University, Ithaca, NY 14850
qz16@cornell.edu

Outline

- Correlation and covariance.
- Orthogonality and uncorrelatedness.
- The Hilbert space of random variables.
- MMSE estimation:
  - The Orthogonality principle.
  - Unconstrained MMSE estimator.
  - Linear MMSE estimator.
  - MMSE estimator for jointly Gaussian random variables.
  - MMSE estimator for random vectors.

Correlation and Covariance

Correlation and Covariance:

- Correlation: $\mathbb{E}[XY]$
- Covariance:

$$\begin{aligned}\mathsf{Cov}(X,Y) &\stackrel{\Delta}{=} \mathbb{E}\left[(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right] \\ &= \mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

- Correlation coefficient:

$$\rho_{XY} = \frac{\mathsf{Cov}(X,Y)}{\sqrt{\mathsf{Var}(X)\mathsf{Var}(Y)}}$$

Orthogonality and Uncorrelatedness:

- X and Y are orthogonal if $\mathbb{E}[XY]=0$.
- X and Y are uncorrelated if $\mathsf{Cov}(X,Y)=0$, or equivalently, $\mathbb{E}[XY]=\mathbb{E}[X]\mathbb{E}[Y]$.

Correlation and Covariance

Properties:

- Independence implies uncorrelatedness, but not vise versa.

- If at least one of X and Y has zero mean, then $\mathbb{E}[XY] = \mathsf{Cov}(X, Y)$. In this case, X and Y are orthogonal iff they are uncorrelated.

- $\mathsf{Cov}(X, aY + bZ) = a\mathsf{Cov}(X, Y) + b\mathsf{Cov}(X, Z)$.

- $\mathsf{Cov}(X - \mathbb{E}[X], Y - \mathbb{E}[Y]) = \mathsf{Cov}(X, Y)$.

- Cauchy-Schwarz inequality:
$$(\mathbb{E}[XY])^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2],$$
with equality iff $\Pr(X = cY) = 1$ for some constant c (assuming $\mathbb{E}[Y^2] \neq 0$).

- $|\rho_{XY}| \leq 1$ with equality iff $X = aY + b$ for some constants a and b.

- If $X_1, \ldots, X_m$ are pairwise uncorrelated, then
$$\mathsf{Var}(\sum_{i=1}^{m} X_i) = \sum_{i=1}^{m} \mathsf{Var}(X_i).$$

Geometric Intepretation: Hilbert Space

Inner Product:

An inner product on a vector space $\mathcal{V}$ is a map

$$< \cdot, \cdot >: \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{R}$$

such that for all $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ and $\alpha, \beta \in \mathcal{R}$,

- $< \mathbf{x}, \mathbf{y} > = < \mathbf{y}, \mathbf{x} >$
- $< \mathbf{x}, \mathbf{x} > \geq 0$ with equality iff $\mathbf{x} = 0$.
- $< \mathbf{x}, \alpha\mathbf{y} + \beta\mathbf{z} > = \alpha < \mathbf{x}, \mathbf{y} > + \beta < \mathbf{x}, \mathbf{z} >$.

Hilbert Space:

A Hilbert space is a vector space that

- has an inner product $< \cdot, \cdot >$ defined
- and is complete with respect to the norm $||\mathbf{x}|| \triangleq \sqrt{< \mathbf{x}, \mathbf{x} >}$ induced by the inner product.

Example: 2-dimensional Euclidean plane with inner product $< \mathbf{x}, \mathbf{y} >$ (also referred to as dot product $\mathbf{x} \cdot \mathbf{y}$) defined as

$$< \mathbf{x}, \mathbf{y} > = x_1 y_1 + x_2 y_2$$

The Hilbert Space of Random Variables

The Hilbert Space of Random Variables:

All random variables with finite second moments form a Hilbert space with inner product

$$< X, Y > \stackrel{\Delta}{=} \mathbb{E}[XY]$$

Geometry:

1. The length (norm) of X:
$$||X|| = \sqrt{< X, X >} = \sqrt{\mathbb{E}[X^2]}$$
2. The angle θ_{XY} between X and Y is given by
$$\cos\theta_{XY} = \frac{< X, Y >}{||X||\,||Y||} = \frac{\mathbb{E}[XY]}{\sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}}$$
3. CauchySchwarz inequality follows directly from $|\cos\theta_{XY}| \leq 1$.
4. The projection $\sqcap_Y(X)$ of X onto Y is
$$\sqcap_Y(X) = ||X|| \cos\theta_{XY} \frac{Y}{||Y||} = \frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]} Y$$
5. X and Y are orthogonal if $\mathbb{E}[XY] = 0$ $\big(\sqcap_Y(X) = 0\big)$.
6. Pythagorean theorem: if $X_1, \ldots, X_m$ are pairwise orthogonal, then
$$\mathbb{E}\left[(\sum_{i=1}^{m} X_i)^2\right] = \sum_{i=1}^{m} \mathbb{E}(X_i^2).$$

MMSE Estimation

Mean Square Error (MSE):

Let $\hat{X}$ denote an estimate of X.

- The estimation error W is

$$W = X - \hat{X}$$

- The MSE of the estimate $\hat{X}$ is

$$\mathbb{E}\left[(X-\hat{X})^2\right] = \mathbb{E}[W^2] = ||W||^2 =< W, W >,$$

  which is the length of the error W squared, or equivalently, the distance $d(X, \hat{X}) = ||X - \hat{X}||$ between X and $\hat{X}$ squared.

- The MMSE estimator of X is the one that minimizes the MSE:

$$\min_{\hat{X}} \mathbb{E}\left[(X-\hat{X})^2\right]$$

- The MMSE estimator is the one that produces an error W that is the shortest in length $||W||$. In other words, $\hat{X}_{\text{MMSE}}$ is the closest to X.

MMSE Estimator of X using a Constant

MMSE Estimator of X using a constant:

- Choose a constant a to minimize MSE:

$$\hat{X} = a^* \quad \text{where } a^* = \arg\min_{a \in \mathcal{R}} \mathbb{E}\left[(X-a)^2\right]$$

- The estimator and the MSE:

$$a^* = \mathbb{E}[X], \quad \text{MSE} = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \text{Var}(X)$$

Proof:

$$\begin{aligned}
\text{MSE} &= \mathbb{E}[(X-a)^2] \\
&= \mathbb{E}\left[(X - \mathbb{E}[X] + \mathbb{E}[X] - a)^2\right] \\
&= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] + 2\mathbb{E}\left[(X - \mathbb{E}[X])(\mathbb{E}[X] - a)\right] + \mathbb{E}\left[(\mathbb{E}[X] - a)^2\right] \\
&= \text{Var}(X) + (\mathbb{E}[X] - a)^2
\end{aligned}$$

which is minimized at $a = \mathbb{E}[X]$, resulting in an MSE $= \text{Var}(X)$.

MMSE Estimator of X using $g(Y)$

MMSE Estimator of X using $g(Y)$:

- The MMSE estimator $g^*(Y)$:

$$g^*(Y) = \arg\min_g \mathbb{E}\left[(X - g(Y))^2\right]$$

- For each given $Y = y$, the problem is reduced to estimating X using a constant $g(y)$. The MMSE estimator is

$$g^*(y) = \mathbb{E}\left[X|Y = y\right]$$

- Given $Y = y$, the MSE is the conditional variance:

$$\text{MSE}_{Y=y} = \mathbb{E}\left[(X - \mathbb{E}[X|Y = y])^2|Y = y\right] = \text{Var}(X|Y = y)$$

- The MMSE estimator of X using a function g of Y is the conditional mean (a random variable):

$$\hat{X}_{\text{MMSE}} = g^*(Y) = \mathbb{E}[X|Y]$$

- The error W has zero mean ($\mathbb{E}[W] = 0$), i.e., $\hat{X}_{\text{MMSE}}$ is an unbiased estimator.

- The overall MSE achieved by $g^*(Y) = \mathbb{E}(X|Y)$ is the expectation (over Y) of the conditional variance:

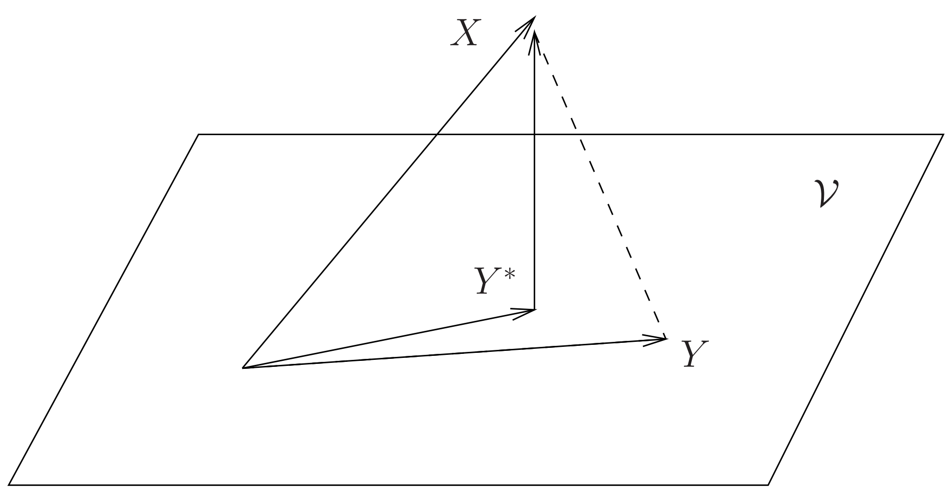
$$\begin{aligned}
\text{MSE} &= \mathbb{E}_Y\left[\text{Var}(X|Y)\right] = \int_y \text{Var}(X|Y = y) f_Y(y) dy \\
&= \mathbb{E}[X^2] - \mathbb{E}[(\hat{X}_{\text{MMSE}})^2] \quad \left(\text{Pythagorean: } W \perp \hat{X}_{\text{MMSE}}\right) \\
&= \text{Var}(X) - \text{Var}(\hat{X}_{\text{MMSE}}) \quad (\because \mathbb{E}[X] = \mathbb{E}[\hat{X}_{\text{MMSE}}])
\end{aligned}$$

The Orthogonality Principle

The Orthogonality Principle:

Let $\mathcal{L}^2(\Omega, \mathcal{F}, \Pr)$ denote the set of all random variables on $(\Omega, \mathcal{F}, \Pr)$ with finite second moments. Let $\mathcal{V}$ be a closed (in the mean square sense) linear subspace of $\mathcal{L}^2(\Omega, \mathcal{F}, \Pr)$, and let $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \Pr)$.

- *Existence and uniqueness:* There exists a unique element $Y^* \in \mathcal{V}$ such that, for all $Y \in \mathcal{V}$,
$$\mathbb{E}\left[(X - Y^*)^2\right] \leq \mathbb{E}\left[(X - Y)^2\right].$$
- *Characterization:* Let Z be a random variable in $\mathcal{V}$. Then $Z = Y^*$ if and only if $\mathbb{E}\left[(X - Z)Y\right] = 0$ for all $Y \in \mathcal{V}$, i.e., $X - Z \perp \mathcal{V}$.
- *Error expression:* The MMSE for estimating X using $\mathcal{V}$ is given by
$$\text{MSE} = \mathbb{E}\left[(X - Y^*)^2\right] = \mathbb{E}[X^2] - \mathbb{E}[(Y^*)^2]$$



Remarks:

- The MMSE estimator of X using $\mathcal{V}$ is the random variable Y^* in $\mathcal{V}$ that results in an error $X - Y^*$ being orthogonal to $\mathcal{V}$ (i.e., orthogonal to all random variables in $\mathcal{V}$).
- It is easy to check that the error $W = X - \mathbb{E}[X|Y]$ of the MMSE estimator $g^*(Y) = \mathbb{E}[X|Y]$ is orthogonal to $g(Y)$ for all functions $g(\cdot)$.

MMSE Estimator of X using aY

MMSE Estimator of X using aY:

- Choose a to minimize MSE:

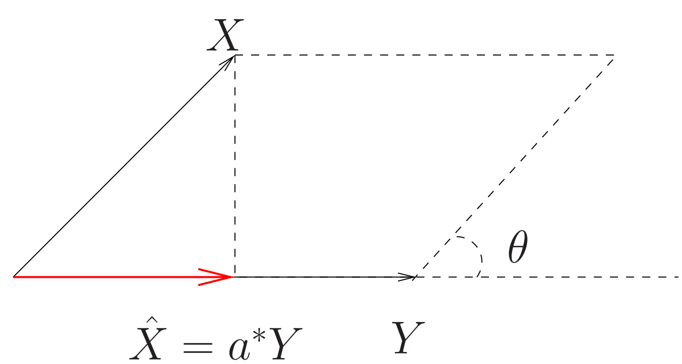
$$\hat{X} = a^*Y \quad \text{where } a^* = \arg\min_{a \in \mathcal{R}} \mathbb{E}\left[(X - aY)^2\right]$$

- From the geometric interpretation, the best estimate is given by the projection of X onto Y:

$$a^* = \frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}$$

- The error $W = X - a^*Y$ is orthogonal to Y:

$$\mathbb{E}[WY] = \mathbb{E}\left[(X - a^*Y)Y\right] = 0$$



When Y is a constant ($Y = c$):

$$a^* = \frac{\mathbb{E}[X]}{c}, \qquad \hat{X} = a^*c = \mathbb{E}[X]$$

$$\text{MSE} = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \text{Var}(X)$$

Linear MMSE Estimators

Linear MMSE Estimator:

- Estimate X using affine functions of Y:

$$\hat{X} = aY + b$$

- Choose a and b to minimize MSE:

$$\{a^*, b^*\} = \arg\min_{a,b} \mathbb{E}\left[(X - (aY + b))^2\right]$$

- By the orthogonality principle, the error $W \triangleq X - (a^*Y + b^*)$ is orthogonal to all affine functions of Y:

$$W \perp aY + b, \quad \forall a, b$$

- It suffices to have

$$W \perp 1 \; (\mathbb{E}[W] = 0) \quad \text{and} \quad W \perp Y \; (\mathbb{E}[WY] = 0)$$

- The above leads to

$$b^* = \mathbb{E}[X] - a^*\mathbb{E}[Y], \quad a^* = \frac{\mathsf{Cov}(X, Y)}{\mathsf{Var}(Y)}$$

- The best linear estimator and its MSE:

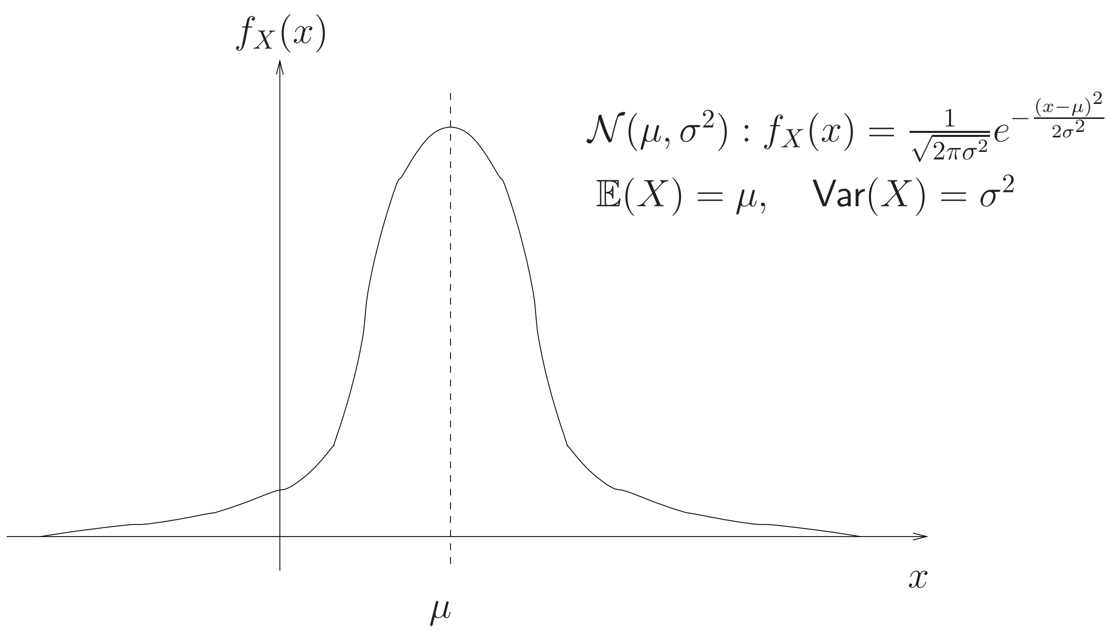
$$\begin{aligned}
\hat{X}_{\mathsf{LMMSE}} &= \mathbb{E}[X] + \frac{\mathsf{Cov}(X, Y)}{\mathsf{Var}(Y)}[Y - \mathbb{E}(Y)] \\
\mathsf{MSE} = \mathbb{E}\left[W^2\right] &= \mathbb{E}[X^2] - \mathbb{E}[\hat{X}^2_{\mathsf{LMMSE}}] \quad \left(\text{Pythagorean: } W \perp \hat{X}_{\mathsf{LMMSE}}\right) \\
&= \mathsf{Var}(X) - \mathsf{Var}(\hat{X}_{\mathsf{LMMSE}}) \quad (\because \mathbb{E}[X] = \mathbb{E}[\hat{X}_{\mathsf{LMMSE}}]) \\
&= \mathsf{Var}(X) - \frac{\mathsf{Cov}^2(X, Y)}{\mathsf{Var}(Y)} \\
&= \mathsf{Var}(X)(1 - \rho^2_{XY})
\end{aligned}$$

Jointly Gaussian Random Variables

Gaussian Random Variable:

A random variable X is Gaussian with mean μ and variance $\sigma^2 > 0$ if X has PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Jointly Gaussian Random Variables:

X and Y are jointly Gaussian if every linear combination of them is Gaussian, i.e., $a_1 X + a_2 Y$ is Gaussian for all a_1, a_2.

Properties

Properties of Jointly Gaussian Random Variables:

- Uncorrelated jointly Gaussian random variables are independent.
- Jointly Gaussian implies marginally Gaussian, but not vise versa.
- Independent Gaussian random variables are jointly Gaussian.
- If X and Y are jointly Gaussian, then $a_1X + b_1Y + c_1$ and $a_2X + b_2Y + c_2$ are jointly Gaussian for all a_i, b_i, c_i.
- For jointly Gaussian X and Y with $|\rho_{XY}| < 1$, their joint PDF $f_{XY}(x, y)$ is given by

$$\frac{1}{(2\pi)\sqrt{|\mathbf{K}|}} exp\left\{-\frac{1}{2}\left(\begin{pmatrix}X\\Y\end{pmatrix} - \begin{pmatrix}\mu_X\\\mu_Y\end{pmatrix}\right)^T \mathbf{K}^{-1}\left(\begin{pmatrix}X\\Y\end{pmatrix} - \begin{pmatrix}\mu_X\\\mu_Y\end{pmatrix}\right)\right\}$$

where $\mathbf{K}$ is the covariance matrix of $\begin{pmatrix}X\\Y\end{pmatrix}$:

$$\mathbf{K} = \begin{pmatrix}\mathsf{Cov}(X,X) & \mathsf{Cov}(X,Y)\\ \mathsf{Cov}(Y,X) & \mathsf{Cov}(Y,Y)\end{pmatrix} = \begin{pmatrix}\sigma_X^2 & \rho_{XY}\sigma_X\sigma_Y\\ \rho_{XY}\sigma_X\sigma_Y & \sigma_Y^2\end{pmatrix}$$

- For jointly Gaussian X and Y, we have $\hat{X}_{\mathsf{MMSE}} = \hat{X}_{\mathsf{LMMSE}}$, i.e.,

$$\begin{aligned}\mathbb{E}(X|Y) &= \hat{X}_{\mathsf{LMMSE}} = \frac{\mathsf{Cov}(X,Y)}{\mathsf{Var}(Y)}[Y - \mathbb{E}(Y)] + \mathbb{E}(X)\\ \mathsf{MMSE} &= \mathsf{Var}(X) - \frac{\mathsf{Cov}^2(X,Y)}{\mathsf{Var}(Y)} = \mathsf{Var}(X)(1-\rho_{XY}^2)\end{aligned}$$

- For jointly Gaussian X and Y, the conditional distribution of X given $Y = y$ is Gaussian with mean and variance given by

$$\begin{aligned}\mathbb{E}(X|Y=y) &= \frac{\mathsf{Cov}(X,Y)}{\mathsf{Var}(Y)}[y - \mathbb{E}(Y)] + \mathbb{E}(X)\\ \mathsf{Var}(X|Y=y) &= \mathsf{Var}(X) - \frac{\mathsf{Cov}^2(X,Y)}{\mathsf{Var}(Y)} = \mathsf{Var}(X)(1-\rho_{XY}^2)\end{aligned}$$

Random Vectors

Random Vector:

A random vector $\mathbf{X}$ of dimension m consists of m random variables defined on the same probability space:

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix}$$

Expectation, Correlation and Covariance Matrixes:

$$\mathbb{E}(\mathbf{X}) = \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_m) \end{pmatrix} \quad \mathbb{E}(\mathbf{X}\mathbf{X}^T) = \begin{pmatrix} \mathbb{E}(X_1^2) & \cdots & \mathbb{E}(X_1 X_m) \\ \vdots & \vdots & \vdots \\ \mathbb{E}(X_m X_1) & \cdots & \mathbb{E}(X_m^2) \end{pmatrix}$$

$$\text{Cov}(\mathbf{X}) = \begin{pmatrix} \text{Var}(X_1) & \cdots & \text{Cov}(X_1, X_m) \\ \vdots & \vdots & \vdots \\ \text{Cov}(X_m, X_1) & \cdots & \text{Var}(X_m) \end{pmatrix}$$

Cross Correlation and Covariance Matrixes:

Let $\mathbf{X}$ (dimension m) and $\mathbf{Y}$ (dimension n) be two random vectors defined on the same probability space.

$$\mathbb{E}(\mathbf{X}\mathbf{Y}^T) = \Big\{\mathbb{E}(X_i Y_j)\Big\}_{m\times n} \qquad \text{Cov}(\mathbf{X}, \mathbf{Y}) = \Big\{\text{Cov}(X_i, Y_j)\Big\}_{m\times n}$$

Gaussian Random Vectors

Gaussian Random Vector:

- $\mathbf{X}$ is a Gaussian random vector if its coordinate random variables are jointly Gaussian.
- $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K})$ denotes a Gaussian random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{K}$.

Properties: Suppose that $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K})$.

- Jointly Gaussian implies marginally Gaussian: any sub-vector of $\mathbf{X}$ is Gaussian. In particular,

$$X_i \sim \mathcal{N}(\mu_i, \mathbf{K}_{i,i}).$$

- For any matrix $\mathbf{A}$ and vector $\mathbf{b}$, $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ is Gaussian and

$$\mathbf{Y} \sim \mathcal{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\mathbf{K}\mathbf{A}^T).$$

- If $\mathbf{K}$ is nonsingular, then $\mathbf{X}$ has a PDF given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^m |\mathbf{K}|}} exp\{-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{X} - \boldsymbol{\mu})\}$$

- If $\mathbf{K}$ is a diagonal matrix, then the coordinates $X_1, \ldots, X_m$ are independent.
- If $\mathbf{X}$ and $\mathbf{Y}$ are jointly Gaussian vectors, then they are independent iff $\mathsf{Cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{0}$.

MMSE Estimation of Random Vectors

MMSE Estimation of $\mathbf{X}$ using $\mathbf{g}(\mathbf{Y})$:

- MSE for estimation of $\mathbf{X}$ is the sum of the MSEs of the coordinates:

$$\mathbb{E}(||\mathbf{X}-\mathbf{g}(\mathbf{Y})||^2) = \sum_{i=1}^{m} \mathbb{E}[(X_i - g_i(\mathbf{Y}))^2]$$

  Thus finding the MMSE estimator of $\mathbf{X}$ decomposes into finding the MMSE estimators of each X_i separately.

- The MMSE estimator is given by the conditional mean:

$$\mathbf{g}^*(\mathbf{Y}) = \mathbb{E}(\mathbf{X}|\mathbf{Y}) = \begin{pmatrix} \mathbb{E}(X_1|\mathbf{Y}) \\ \vdots \\ \mathbb{E}(X_m|\mathbf{Y}) \end{pmatrix}$$

Linear MMSE Estimation of $\mathbf{X}$:

- Estimate X using a linear transform of Y:

$$\hat{\mathbf{X}} = \mathbf{A}\mathbf{Y} + \mathbf{b}$$

- By the orthogonality principle,

$$\begin{aligned} \hat{\mathbf{X}}_{\text{LMMSE}} &= \mathbb{E}(\mathbf{X}) + \text{Cov}(\mathbf{X},\mathbf{Y})\text{Cov}^{-1}(\mathbf{Y})(\mathbf{Y}-\mathbb{E}(\mathbf{Y})) \\ \text{Cov}(\mathbf{X}-\hat{\mathbf{X}}_{\text{LMMSE}}) &= \text{Cov}(\mathbf{X}) - \text{Cov}(\mathbf{X},\mathbf{Y})\text{Cov}^{-1}(\mathbf{Y})\text{Cov}(\mathbf{Y},\mathbf{X}) \end{aligned}$$

When $\mathbf{X}$ and $\mathbf{Y}$ are jointly Gaussian:

$$\hat{\mathbf{X}}_{\text{LMMSE}} = \mathbf{g}^*(\mathbf{Y}) = \mathbb{E}(\mathbf{X}|\mathbf{Y})$$

Example: Gaussian Signal in Gaussian Noise

$$Y_i = X + W_i, \quad i = 1, \dots, n,$$

where $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ is the signal, $W_i \sim \mathcal{N}(0, \sigma_w^2)$ is noise, and $X, W_1, \dots, W_n$ are independent.

- X and $\mathbf{Y}$ are jointly Gaussian. In particular,

$$\mathbf{Y} \sim \mathcal{N}(\mu_x \mathbf{1}, \sigma_w^2 \mathbf{I} + \sigma_x^2 \mathbf{1}\mathbf{1}^T), \quad \text{where } \mathbf{1} = [1, \dots, 1]^T.$$

- Compute all required statistics:

$$\begin{aligned}
\mathbb{E}(X) &= \mu_x, \quad \mathbb{E}(\mathbf{Y}) = \mu_x \mathbf{1} \\
\text{Cov}(X, \mathbf{Y}) &= \sigma_x^2 \mathbf{1}^T, \quad \text{Cov}(\mathbf{Y}) = \sigma_w^2 \mathbf{I} + \sigma_x^2 \mathbf{1}\mathbf{1}^T \\
\text{Cov}^{-1}(\mathbf{Y}) &= \frac{1}{\sigma_w^2}[\mathbf{I} - \frac{\sigma_x^2}{\sigma_w^2 + n\sigma_x^2}\mathbf{1}\mathbf{1}^T]
\end{aligned}$$

where the Matrix Inversion Lemma was used to obtain $\text{Cov}^{-1}(\mathbf{Y})$:

$$(\mathbf{A} + \mathbf{b}\mathbf{b}^T)^{-1} = \mathbf{A}^{-1} - \frac{1}{1 + \mathbf{b}^T \mathbf{A}^{-1} \mathbf{b}} \mathbf{A}^{-1} \mathbf{b}\mathbf{b}^T \mathbf{A}^{-1}$$

- The MMSE estimator:

$$\begin{aligned}
\hat{X} &= \mathbb{E}(X|\mathbf{Y}) = \mathbb{E}(X) + \text{Cov}(X, \mathbf{Y})\text{Cov}^{-1}(\mathbf{Y})(\mathbf{Y} - \mathbb{E}(\mathbf{Y})) \\
&= \mu_x + \frac{\sigma_x^2}{\sigma_x^2 + \frac{\sigma_w^2}{n}}(\overline{\mathbf{Y}} - \mu_x) = \mu_x + \frac{\text{SNR}}{1 + \text{SNR}}(\overline{\mathbf{Y}} - \mu_x) \\
&\approx \begin{cases} \mu_x, & \text{SNR} << 1 \\ \overline{\mathbf{Y}}, & \text{SNR} >> 1 \end{cases}
\end{aligned}$$

where

$$\overline{\mathbf{Y}} \triangleq \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad \text{SNR} \triangleq \frac{n\sigma_x^2}{\sigma_w^2}$$