References: ① Robert G Gallager : Stochastic Processes, theory for applications (Chapter 3)
② Chapter 5.6, 5.7 : Stanley Chan: Introduction to Probability for data science

Lec 37

① Random Vectors

A collection of n R.Vs $X = (X_1, X_2, \ldots X_n)^T$ is refered to as a random vector

ⓐ Expectation

$$E[X] = \underbrace{\begin{pmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{pmatrix}}_{n \times 1}$$

can assume similar defn for expectation of a random matrix containing random variable.

ⓑ Covariance matrix

Correlation matrix : $\underbrace{E[X X^T]}_{\substack{n\times n \\ \text{matrix}}} = \begin{bmatrix} E[X_1^2] & E[X_1X_2] & \cdots & E[X_1X_n] \\ \vdots & & & \\ E[X_nX_1] & - & - & E[X_n^2] \end{bmatrix}$

↓

It is symmetric.

Covariance matrix : $\underbrace{K_X}_{n\times n \text{ matrix}} = E\left[(X - E[X])(X - E[X])^T\right]$.

$\bar{X}_i = X_i - E[X_i]$

$$= E\left[\begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_n \end{pmatrix} (\bar{X}_1 \ \cdots \ \bar{X}_n)\right] = \begin{bmatrix} Var(X_1), & Cov(X_1, X_2) & \cdots & Cov(X_1, X_n) \\ & & & \\ Cov(X_n, X_1) & \cdots & & Cov(X_n, X_n) \end{bmatrix}$$

Lec 38

(i) Properties of the Covariance matrix.

ⓐ K_X is symmetric

(b) K_x is positive-semi-definite matrix.

Any symmetric matrix K ($n \times n$ matrix) is said to be p.s.d if it satisfies the following inequality:

$$\underbrace{\underline{a}^T}_{1\times n} \underbrace{K}_{n\times n} \underbrace{\underline{a}}_{n\times 1} \geq 0 \quad \forall\ \underline{a} \in \mathbb{R}^n$$

Consider the Covariance matrix K_x is p.s.d

$$K_x = E[\bar{x}\,\bar{x}^T] \quad \text{where } \bar{x} = x - E[x]$$

Let us consider a random variable

$$Y = \underline{a}^T \bar{x}, \qquad Y^2 = (a^T\bar{x})(a^T\bar{x})$$

$$E[Y^2] \geq 0.$$

as $a^T\bar{x}$ is a scalar $\leftarrow$ $= (\underline{a}^T\bar{x})(a^T\bar{x})^T$

$= a^T\bar{x}\,\bar{x}^T a$

$$\Rightarrow E[a^T\bar{x}\,\bar{x}^T a] \geq 0.$$

$$\Rightarrow a^T E[\bar{x}\,\bar{x}^T] a \geq 0. \Bigg\}$$

$E[a^T M] = a^T E[M]$ and similarly $E[Ma] = E[M]a$

$$\Rightarrow \underline{a}^T K_x \underline{a} \geq 0$$

Can check that correlation matrix $E[xx^T]$ is also p.s.d.

(ii) Spectral decomposition

For Any symmetric matrix K, there exists an $\underbrace{\text{orthonormal matrix } Q}$ st

$QQ^T = I$ and each column of Q has norm 1.

$K = QDQ^T$ where D is a diagonal matrix containing eigen values of K.

(iii) Equivalent definitions of Positive semi-definite matrices

K is p.s.d matrix iff

(1) $\forall a \in \mathbb{R}^n, \quad a^T K a \geq 0.$

(2) if K has non-negative eigen values.

(3) if K can be decomposed as $K = RR^T = R^2$

(2) $\Rightarrow$ (3):

We assume K is symmetric & has non-negat eigen values

Can write $K = QDQ^T$ (as K is symmetric).

$$= Q \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} Q^T \quad \lambda_n \rightarrow \text{eigen values of } K$$

$$= Q D^{1/2} D^{1/2} Q^T.$$

$$= \underbrace{Q D^{1/2} Q^T}\, Q D^{1/2} Q^T.$$

Let $R = Q D^{1/2} Q^T.$

$R^T = R.$

$\Rightarrow \quad K = RR^T.$

(1) $\Rightarrow$ (2). From (1) $a^T K a \geq 0 \quad \forall a \in \mathbb{R}^n.$

Since K is symmetric $K = QDQ^T$ and let $Q = [\underline{q}_1 \ \underline{q}_2 \ .. \ \underline{q}_n]$

$\underline{q}_i$'s are eigen vectors.

Set $\underline{a} = \underline{q}_i$

$\Rightarrow \quad \underline{q}_i^T K \underline{q}_i = \underline{q}_i^T \lambda_i \underline{q}_i$

≥ 0 from (1)

$\Rightarrow \lambda_i \ \underline{q}_i^T \underline{q}_i \geq 0.$

$\Rightarrow \lambda_i \geq 0$ as $\underline{q}_i^T \underline{q}_i \geq 0$

$3 \Rightarrow 1$.

$$a^T K a = a^T R R^T a.$$

Let $b = R^T a \Rightarrow b^T = a^T R$

$$\Rightarrow a^T R R^T a = b^T b \geq 0.$$

We have shown that

$$\begin{tikzcd} 1 \arrow[r, Rightarrow] & 2 \arrow[r, Rightarrow] & 3 \arrow[ll, Rightarrow, bend left] \end{tikzcd}$$

$\therefore$ All the conditions are equivalent.

② Joint density of Independent Gaussian R.Vs:

Let $Z = (z_1, \dots z_n)$ where z_i's are iid standard normal random variables.

$$f_Z(z) = f_{Z_1, Z_2 \dots Z_n}(z_1, z_2, \dots z_n) = \prod_{i=1}^{n} f_{Z_i}(z_i)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-z_i^2/2}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^{n} z_i^2/2} = \frac{1}{(2\pi)^{n/2}} e^{-\frac{z^T z}{2}}$$

③ Jointly Gaussian R.Vs / Gaussian Random Vector.

$X = (X_1, X_2, \dots X_n)$ is a Gaussian Random Vector with mean 0. if $\exists$ a $n \times m$ matrix A st.

$X = AZ$ where Z is a Gaussian vector comprising of m iid standard normal R.Vs.

U is a Gaussian random vector if U can be written as $U = AZ + \underline{\mu}$

ⓐ Linear Transformation of Gaussian R.V is a Gaussian R.V.

Let X be a n-Gaussian R.V with zero mean. Then $Y = MX$ where M is a $m \times n$ matrix is an m-Gaussian R.V.

Since X is Gaussian R.V $\exists$ a matrix A st $X = AZ$.

Then $Y = \underbrace{MA}_{B} Z = BZ$.

$\Rightarrow Y$ is m-Gaussian R.V.

Examples :

$$X = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}.$$

Z_1, Z_2 Standard normal iid.

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} Z_1 + 2Z_2 \\ Z_2 \end{bmatrix}.$$

④ Special Gaussian vectors. (A is non-singular $n \times n$ matrix).

Assume A is non-singular and X is a Gaussian R.V st $X = AZ$.
and $Z = (Z_1, \dots Z_n)$ where Z_i's are iid standard normal.

ⓐ Covariance matrix

$$K_X = E[X X^T] \quad \text{as } X \text{ is mean } 0 \text{ vector.}$$

$$= E[A Z Z^T A^T]$$

$$= A \underbrace{E[Z Z^T]}_{\text{Identity matrix}} A^T = A A^T.$$

A is assumed to be invertible $\Rightarrow$ K_X is invertible as $\det(K_X) = \det(A) * \det(A^T) = \det(A)^2 \neq 0$

ⓑ Joint density.

$$(Z_1, Z_2, \dots Z_n) \Longleftrightarrow (X_1, X_2, \dots X_n)$$

$$X_1 = g_1(z_1, \dots z_n) = \sum_{j=1}^{n} a_{1j} z_j$$

$$X_i = g_i(z_1, \dots z_n) = \sum_{j=1}^{n} a_{ij} z_j$$

$$f_{X_1, \dots X_n}(x_1, \dots x_n) = \frac{f_Z(z)}{J(z_1, \dots z_n)} = \frac{f_Z(A^{-1}x)}{J(z_1, \dots z_n)}$$

z st $x = Az$

$$J(z_1, \dots z_n) = \begin{vmatrix} \frac{\partial g_1(z_1, \dots z_n)}{\partial z_1} \; \text{"}a_{11} & \dots & \frac{\partial g_1(z_1, \dots z_n)}{\partial z_n} \; \text{"}a_{1n} \\ \vdots & & \vdots \\ \frac{\partial g_n(z_1, \dots z_n)}{\partial z_1} = a_{n,1} & \dots & \frac{\partial g_n(z_1, \dots z_n)}{\partial z_n} \; \text{"}a_{n,n} \end{vmatrix}$$

$$= |A| \longrightarrow \text{determinant of } A$$

$$f_X(x) = \frac{f_Z(A^{-1}x)}{|A|}. \qquad f_Z(z) = \frac{e^{-\frac{z^T z}{2}}}{(2\pi)^{n/2}}$$

$$= \frac{e^{-\frac{x^T (A^{-1})^T A^{-1} x}{2}}}{(2\pi)^{n/2} |A|}.$$

We know that

$$K_X = A A^T$$
$$K_X^{-1} = (A^T)^{-1} A^{-1}$$

$$\boxed{f_X(x) = \frac{e^{-\frac{x^T K_x^{-1} x}{2}}}{(2\pi)^{n/2} |K_x|^{1/2}}, \; x \in \mathbb{R}^n} \rightarrow (1)$$

$$\det(K_X) = \det(A)\det(A^T).$$
$$= \det(A)^2$$

In this case K has eigen values > 0.

$$K = Q D Q^T$$
$$\det(K) = \det(Q)\det(D)\det(Q^T)$$
$$= 1 \cdot \det(D)$$

$$1 = \det(QQ^T) = \det(Q) * \det(Q^T) = 1$$

$\Rightarrow$ D can't have 0's. as K is invertible.

$\Rightarrow$ eigen values > 0.

Made with Goodnotes

(c) A Random vector x with joint density defined by K_x with equation (1) is a Gaussian random vector.
We assume X is a random vector with p.d.f given by (1).
Given a covariance matrix K_x it should be p.s.d $\Rightarrow$ eigen values of K_x are ≥ 0.
$\Rightarrow \exists$ a matrix R st $K_x = RR^T$.

$$f_x(x) = \frac{e^{-\frac{x^T (RR^T)^{-1} x}{2}}}{(2\pi)^{n/2} \det(R)}$$

as $\det(K_x) = (\det(R))^2$

$$= \frac{e^{-\frac{(xR^{-1})^T (R^{-1}x)}{2}}}{(2\pi)^{n/2} \det(R)}.$$

define $Z = R^{-1} X$.

Clear to see that

$$f_Z(z) = \frac{e^{-\frac{z^T z}{2}}}{(2\pi)^{n/2}}.$$

$\Rightarrow$ Z is standard normal iid Gaussian Random Vector.

(d) Examples. (To find A from K_x).

(i) $K_x = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\qquad A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $AA^T = K_x$.

(ii) $K_x = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$ $\qquad A = \begin{bmatrix} \sqrt{5} & 0 \\ 0 & \sqrt{3} \end{bmatrix}$; $AA^T = K_x$

(iii) $K_x = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$

$$K_x = QDQ^T = \underbrace{QD^{1/2}Q^T}_{A} \underbrace{QD^{1/2}Q^T}_{A^T}.$$

$$|K_x - \lambda I| = 0$$

$$\left| \begin{bmatrix} 1-\lambda & -0.5 \\ -0.5 & 1-\lambda \end{bmatrix} \right| = 0.$$

$$(1-\lambda)^2 - \frac{1}{4} = 0.$$

$$\lambda^2 + 1 - \frac{1}{4} - 2\lambda = 0$$

$$\lambda^2 - 2\lambda + \frac{3}{4} = 0.$$

$$\Rightarrow \lambda = 3/2, \ 1/2.$$

$$\begin{bmatrix} 1 & -0.5 \\ -0.5 & 1. \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \frac{3}{2} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

$$e_1 - \frac{1}{2} e_2 = \frac{3}{2} e_1.$$

$$-\frac{e_1}{2} - \frac{e_2}{2} = 0 \qquad e_1 = -e_2.$$

$$Q = \begin{bmatrix} \underline{q}_1 & \underline{q}_2 \end{bmatrix}, \quad D = \begin{bmatrix} 3/2 & \\ & 1/2 \end{bmatrix}.$$

$$\underline{q}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \underline{q}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

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

$$A = QD^{1/2}Q^T = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{3}/\sqrt{2} & \\ & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

5. When A is singular (square)

$$X = \begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}.$$

Let $K := \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$, $A = \frac{1}{\sqrt{5}} \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$.

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

$$\Rightarrow \quad X_1 = \frac{1}{\sqrt{5}} [4z_1 + 2z_2]$$

$$X_2 = \frac{1}{\sqrt{5}} [2z_1 + z_2]$$

$$= \frac{X_1}{2}.$$

Can write:

$X_1 = 2Z_1$
and $X_2 = Z_1$

$$f_{X_1, X_2}(x_1, x_2) = f_{X_1}(x_1)\, \delta\left(x_2 - \frac{x_1}{2}\right)$$

Equiprobability points. 2-Gaussian R.V, X with mean 0

We want to find points such that their pdf evaluates to 80% of the maximum possible value.

$$f_X(x) = \frac{e^{-\frac{x^T K^{-1} x}{2}}}{(2\pi)(\det(K))^{1/2}}.$$

$x \in \mathbb{R}^2$ st.

$$f_X(x) = \frac{0.8}{2\pi (\det(K))^{1/2}}.$$

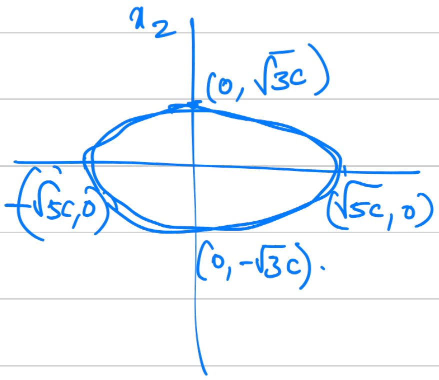
$$\Rightarrow \frac{e^{-\frac{x^T K^{-1} x}{2}}}{\cancel{2\pi (\det(K))^{1/2}}} = \frac{0.8}{\cancel{2\pi (\det(K))^{1/2}}}$$

$$\frac{x^T K^{-1} x}{2} = \log\left(\frac{1}{0.8}\right) = c$$

Let Example! :

$$K = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$$

$$(x_1 \ \ x_2) \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = c.$$



$$\frac{x_1^2}{5} + \frac{x_2^2}{3} = c.$$

eigen values.

Continuing with the general 2x2 covariance matrix

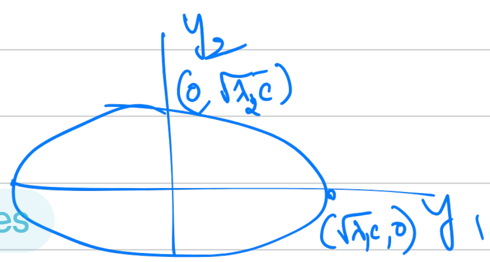
To find all x st

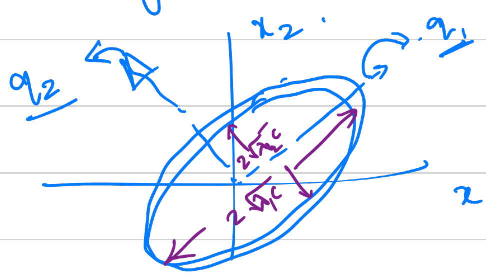
$$x^T K^{-1} x = c$$

$$x^T [Q D Q^T]^{-1} x = c$$

$$x^T Q D^{-1} Q^T x. = c.$$

Let $y = Q^T x$.
$y^T = x^T Q$ then $y^T D^{-1} y = c.$





Example 2:

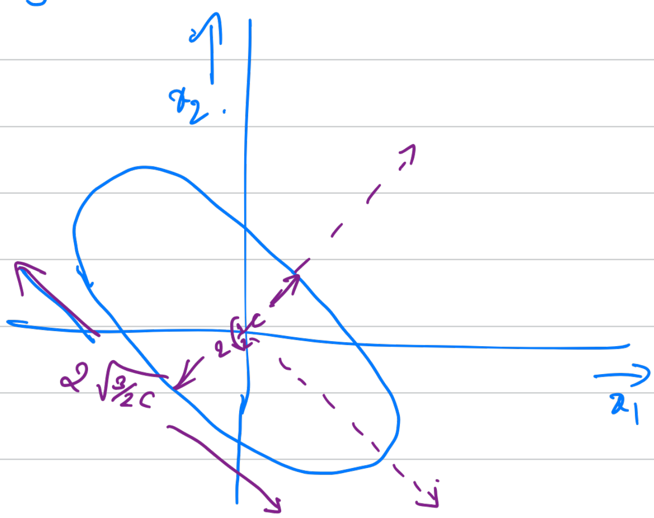
$$K = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}.$$

$$Q = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 3/2 & \\ & 1/2 \end{bmatrix}$$

$$x^T Q D^{-1} Q^T x = c.$$

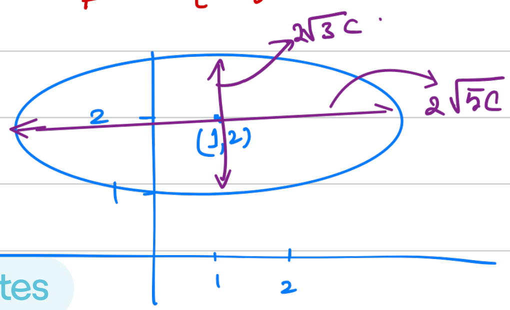
$$y = Q^T x = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

Think of $Q^T = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$. Then $\theta = -45°$.

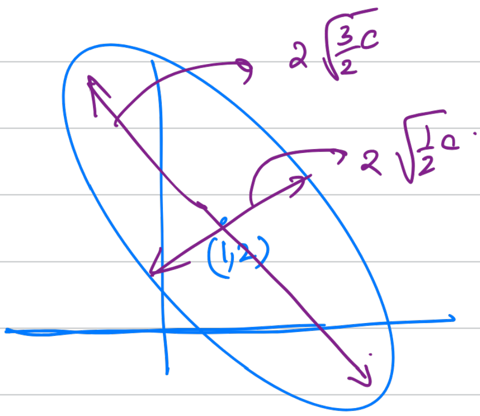


If the n-Gaussian R.V vector has mean $\underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ and Covariance matrix, to find points where p.d.f evaluates to a certain % of maximum. Follow the same steps but shift the center of the ellipse.

Example 1: $\quad \underline{\mu} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad K = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}.$



Example 2: $\mu = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ $K = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$



Finish !!

Multi Parameter estimation from multiple Observations
Appendix

$X = (X_1, \dots X_n) \longrightarrow$ $\boxed{Y|X}$ $\longrightarrow Y = Y_1 Y_2 \dots Y_m \longrightarrow$ $\boxed{\text{estimator}}$

$$\begin{pmatrix} \hat{X}_1 \\ \hat{X}_2 \\ \vdots \\ \hat{X}_n \end{pmatrix} = \hat{X} = \hat{x}(Y) = \begin{bmatrix} \hat{x}_1(Y) \\ \hat{x}_2(Y) \\ \hat{x}_n(Y) \end{bmatrix}.$$

Sum of
Mean Squared Errors.

$$\sum_{i=1}^{n} E\left[(X_i - \hat{X}_i)^2\right] = E\left[(X - \hat{X})^T (X - \hat{X})\right].$$

$$E\left[(X_i - \hat{X}_i)^2 \mid Y = y\right] \geq E\left[(X_i - E[X_i \mid Y = y])^2 \mid Y = y\right].$$

$$E\left[(X_i - \hat{X}_i)^2\right] \geq E\left[(X_i - E[X_i \mid Y])^2\right].$$

for each RV MMSE estimator minimizes individual error.

$$\hat{x}_{i,MMSE}(y) = E[X_i \mid Y = y].$$

$$\hat{x}_{MMSE}(y) = \begin{bmatrix} E[X_1 \mid Y = y] \\ \vdots \\ E[X_n \mid Y = y] \end{bmatrix} = E[X \mid Y = y].$$

Linear MMSE estimator.

$$\hat{x}_{LMMSE}(y) = \underset{n\times m}{A}\,\underset{m\times 1}{y} + b \quad \text{st.}$$

$$(A, b) = \arg\min_{(A,b)} E\left[[X-(Ay+b)]^T [X-(Ay+b)] \right].$$

Let $A = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix}$.

We have earlier shown that

$E[(X_i - a_i^T y - b_i)^2]$ is

minimized by $a_i^T = R_{X_i,Y} K_Y^{-1}$.

$b_i = E[X_i] - a_i^T E[Y]$.

$\Rightarrow$ If individual sums are minimized, the entire sum get minimized by.

$$\hat{x}_{LMMSE}(y) = \begin{bmatrix} R_{X_1,Y} \\ R_{X_2,Y} \\ \vdots \\ R_{X_n,Y} \end{bmatrix} K_Y^{-1}(y - E[Y]) + E[X]$$

$$= R_{X,Y} K_Y^{-1}(y - E[Y]) + E[X].$$

MMSE Estimator is same as linear MMSE estimator if (X, Y) jointly Gaussian

Finding $E[X \mid Y = y]$. Assume X is $n \times 1$ vector and Y is $m \times 1$ vector

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Let $K = \begin{bmatrix} \overbrace{K_X}^{n\times n} & \overbrace{R_{XY}}^{n\times m} \\ \underbrace{R_{YX}}_{m\times n} & \underbrace{K_Y}_{m\times m} \end{bmatrix}$ and Let $K^{-1} = \begin{bmatrix} \overbrace{A}^{n\times n} & \overbrace{B}^{n\times m} \\ \underbrace{C}_{m\times n} & \underbrace{D}_{m\times m} \end{bmatrix}$.

$$KK^{-1} = K^{-1}K = I_{m+n}.$$

$$f_Y(y) = \frac{1}{(2\pi)^{m/2}\det(K_Y)} e^{-\frac{y^T K_y^{-1} y}{2}}$$

$$f_{X,Y}(x,y) = \frac{1}{(2\pi)^{\frac{m+n}{2}}\det(K)} e^{-(x^T\ y^T)\overbrace{\begin{bmatrix} A & B \\ C & D \end{bmatrix}}^{K^{-1}}\begin{bmatrix} x \\ y \end{bmatrix}}.$$

$$= \frac{1}{(2\pi)^{\frac{m+n}{2}}\det(K)} e^{-x^TAx - x^TBy - y^TCx - y^TDy}$$

$$f_{X|Y}(x|y) = \frac{1}{(2\pi)^{n/2}\left\{\frac{\det(K)}{\det(K_y)}\right\}} e^{-x^TAx - 2x^TBy - y^TDy + y^TK_Y^{-1}y}.$$

as K is symmetric K^{-1} should be symmetric as well

$\Rightarrow$ A, D are symmetric and $B^T = C$

$$\begin{bmatrix} K_X & R_{XY} \\ R_{YX} & K_Y \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ 0 & I_m \end{bmatrix}$$

$\Rightarrow$ $K_XB + R_{XY}D = 0.$

$R_{YX}A + K_YC = 0.$

$K_XA + R_{XY}C = I_n$

$R_{YX}B + K_YD = I_n.$

$\Rightarrow A = (K_X - R_{XY}K_Y^{-1}R_{YX})^{-1}$

$\Rightarrow D = (K_Y - R_{YX}K_X^{-1}R_{XY})^{-1}$

$B = -K_X^{-1}R_{XY}D.$

$C = -K_Y^{-1}R_{YX}A.$

Can simplify

$$f_{X|Y}(x|y) = \frac{1}{(2\pi)^{n/2}\det(A^{-1})} e^{-\frac{[x + A^{-1}By]^T A [x + A^{-1}By]}{2}}.$$

$$E[X|Y=y] = -A^{-1}By.$$

$$= A^{-1} \underline{K_X^{-1} R_{X,Y} D} y.$$

$B = C^T$

$K_X^{-1} R_{X,Y} D$

$= A R_{X,Y} K_Y^{-1}$

$$= A^{-1} A R_{X,Y} K_Y^{-1} y.$$

$$\boxed{= R_{X,Y} K_Y^{-1} y}$$

exactly same as the linear estimator.

Needed for simplification

① $\det\begin{pmatrix} \overbrace{A}^{n\times n} & 0 \\ C & \underbrace{D}_{m\times m} \end{pmatrix} = \det(A)\det(D) = \det\begin{pmatrix} A & B \\ 0 & D \end{pmatrix}$

② $\det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det\begin{pmatrix} A & B \\ 0 & D - CA^{-1}B \end{pmatrix}$

$= \det(A) \det(D - CA^{-1}B)$

row operations can be be represented by multiplication by a matrix on the left

$$\underbrace{\begin{pmatrix} I & 0 \\ -CA^{-1} & I \end{pmatrix}}_{\text{determinant is 1 from part ①}}\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A & B \\ 0 & D - CA^{-1}B \end{pmatrix}$$

③ $$\det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det\begin{pmatrix} A - BD^{-1}C & 0 \\ C & D \end{pmatrix}$$

$\downarrow$

$$\underbrace{\begin{pmatrix} I & -BD^{-1} \\ 0 & I \end{pmatrix}}_{\text{det is } 1.}\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A - BD^{-1}C & 0 \\ C & D \end{pmatrix}.$$

$$\det(K) = \det\begin{pmatrix} K_X & R_{XY} \\ R_{YX} & K_Y \end{pmatrix}.$$

$$= \det(K_Y)\,\det\left(K_X - R_{XY}K_Y^{-1}R_{YX}\right).$$

$$\frac{\det(K)}{\det(K_Y)} = \det(A).$$ where A is the top left matrix of K^{-1}.

$$K^{-1} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

$$y^T B^T (A^{-1})^T \cancel{A}\,\cancel{A^{-1}} B y = y^T B^T A^{-1} B y$$

A^{1} is symmetric
as A is symmetric
(due to K^{-1} being symmetric).

To show $y^T B^T A^{-1} B y = y^T D y - y^T K_Y^{-1} y$

i.e., $B^T A^{-1} B = D - K_y^{-1}$.

We showed earlier $B = -K_X^{-1} R_{XY} D$

$$B^T = C = -K_y^{-1} R_{yx} A.$$

$$B^T A^{-1} B = K_y^{-1} R_{yx} K_x^{-1} R_{xY} D.$$

$$D^{-1} D = 1.$$

$$(K_y - R_{yx} K_x^{-1} R_{xY}) D = 1.$$

$$\Rightarrow (1 - K_y^{-1} R_{Yx} K_x^{-1} R_{xY}) D = K_y^{-1}.$$

$$\Rightarrow \boxed{B^T A^{-1} B = D - K_y^{-1}}$$