- Missed 1 lec -

(Rhombus example).

11/02

$$|J(x,y)| = \left| J\left(\frac{z,w}{x,y}\right) \right|$$

$$= \begin{vmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{vmatrix}$$

$$f_{zw}(z,w) = \frac{f_{XY}(x,y)}{J(x,y)} + \dots$$ for all roots possible.

eg

$f_X(x) = \frac{1}{2}$; $|x|<1$

$f_Y(y) = \frac{1}{2}$; $|y|<1$.

$Z = X+Y$

$W = X-Y$.

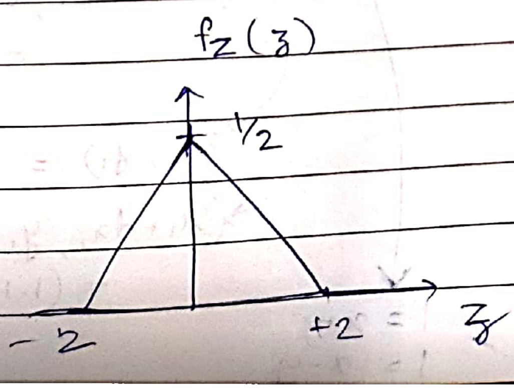
Answer : $f_{zw}(z,w) = \frac{1}{8}$.

$$f_Z(z) = \int_{z-2}^{2-z} \frac{1}{8}\, dw = \frac{2-z-z+2}{8}$$

$$= \frac{4-2z}{8}$$

$$= \frac{1}{2} - \frac{z}{4}$$

~~3/10~~.



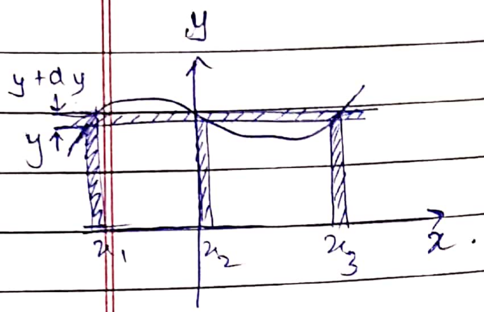
$-1 < x < 1$

$-1 < y < 1$

Earlier, we saw :

* $f_Y(y)\,dy = P(y \le Y \le y+dy)$

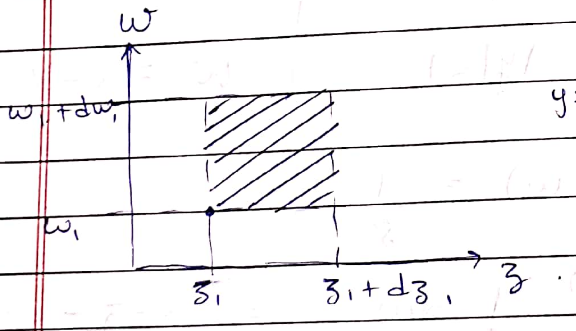
$= P(x_1 \le X \le x_1 + dx_1)$
$+ P(x_2 \le X \le x_2 + dx_2)$
$+ \dots$
(all roots).

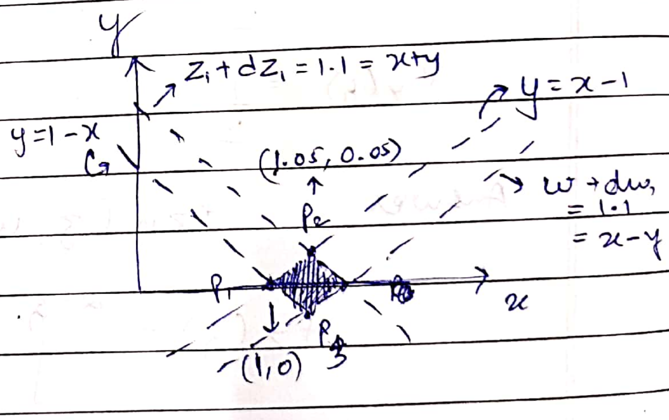


* Now, for 2 variables :

$f_{ZW}(z, w)\,dz\,dw = P(z < Z \le z+dz,\; w < W \le w + dw)$

$= f_{XY}(x,y)\,dx\,dy.$





$Z = X + Y$

$W = X - Y$

$X = \dfrac{Z+W}{2}$

$Y = \dfrac{Z-W}{2}$

$\dfrac{\partial x}{\partial z} = \dfrac{1}{2}$

$\dfrac{\partial y}{\partial w} = \dfrac{-1}{2}$

$\dfrac{\partial x}{\partial w} = \dfrac{1}{2}$

$\dfrac{\partial y}{\partial z} = \dfrac{1}{2}$

Let $(z_1, w_1) = (1,1)$

$(z_1+dz_1,\; w_1+dw_1) = (1.1, 1.1)$

$(x_1, y_1) = (1,0)$

$(x_1+dx_1,\; y_1+dy_1) = (1.1, 0)$

$1 = x+y$
$1 = x-y.$

Date
Page

$$P_1 = \begin{pmatrix} x \\ y \end{pmatrix}, \quad P_2 = \begin{pmatrix} x + \frac{\partial x}{\partial z} dz \\ y + \frac{\partial y}{\partial z} dz \end{pmatrix}.$$

~~P_3~~ and similarly, $P_3 = \begin{pmatrix} x + \frac{\partial x}{\partial w} dw \\ y + \frac{\partial y}{\partial w} dw \end{pmatrix}$

Area of $\parallel$gm (in XY plane) $= |(P_2 - P_1) \times (P_3 - P_1)|$

$$= \left| \begin{pmatrix} \frac{\partial x}{\partial z} dz \\ \frac{\partial y}{\partial z} dz \end{pmatrix} \times \begin{pmatrix} \frac{\partial x}{\partial w} dw \\ \frac{\partial y}{\partial w} dw \end{pmatrix} \right|$$

$$= \begin{vmatrix} i & j \\ \frac{\partial x}{\partial z} dz & \frac{\partial y}{dz} dz \\ \frac{\partial x}{\partial w} dw & \frac{\partial y}{\partial w} dw \end{vmatrix}$$

$$= \frac{\partial x}{\partial z} dz \frac{\partial y}{\partial w} dw - \frac{\partial x}{\partial w} dw \frac{\partial y}{dz} dz.$$

$$= \left(\frac{\partial x}{\partial z} \frac{\partial y}{\partial w} - \frac{\partial x}{\partial w} \frac{\partial y}{\partial z} \right) (dz\, dw)$$

$\therefore$ $f_{ZW}(z, w)\, dz\, dw = f_{XY}(x, y)\, dx\, dy.$

$f_{ZW}(z, w)$ ~~$dz\, dw$~~ $= f_{XY}(x, y) \left| \frac{\partial x}{\partial z} \frac{\partial y}{\partial w} - \frac{\partial x}{\partial w} \frac{\partial y}{\partial z} \right|$

$= f_{XY}(x, y) \cdot |(-1/4 - 1/4)|$ $\quad = f_{XY}(x, y)(1/2)$

Method can't be used when Z, W are linearly dependent because $J = 0$.

12/02.

★ RANDOM VECTOR:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}$$

If we have 2 r.v. X_1 and X_2, they can be characterised by using joint Pdf $f_{X_1,X_2}(x_1,x_2)$. Hence we have N numbers of r.v.s, these can be characterized using multivariate Pdf.

$$f_{X_1,X_2,\ldots X_n}(x_1,x_2,\ldots,x_n).$$

$$F_{X_1,X_2,\ldots X_n}(x_1,x_2,\ldots,x_n)$$

$$= P(X_1 \leq x_1, X_2 \leq x_2 \ldots, X_n \leq x_n)$$

* COVARIANCE

If the outcome of one random variable is a given value, what can we say about the outcome of ~~one~~ another random variable? Answered by using the concept of covariance.

$$Var(X) = E[X^2] - (E[X])^2 = E[(X - m_x)^2]$$

Covariance between X & Y is given by

$$Cov(X, Y) = E[(X - m_x)(Y - m_y)]$$

When $Y = X$, $Cov(X, Y) = Var(X)$.

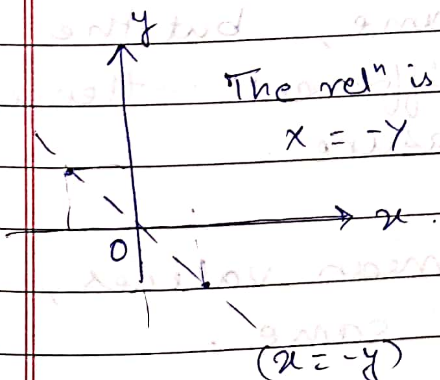
eg ① X & Y

$P(X = 1, Y = 1) = P(X = -1, Y = -1) = \frac{1}{2}$ Rel^n is $X = Y$.

$$E(XY) = \sum x_i y_i \, P(X = x_i, Y = y_i)$$

$$E(XY) = 1 \cdot 1 \cdot \frac{1}{2} + (-1)(-1) \cdot \frac{1}{2} = 1.$$

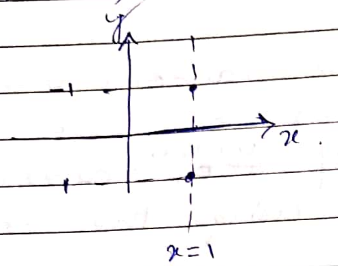
eg ② $P(X = 1, Y = -1) = P(X = -1, Y = 1) = \frac{1}{2}$



The rel^n is $X = -Y$

$$E(XY) = 1 \cdot (-1) \cdot \frac{1}{2} + (-1) \cdot (1) \cdot \frac{1}{2} = -1$$

eg. $P(X=1, Y=1) = P(X=1, Y=-1) = \frac{1}{2}$



$$E(XY) = 1.1.\frac{1}{2} + 1.(-1)\frac{1}{2}$$

$$E(XY) = 0.$$

When X takes 1, Y takes 1 or -1.
There doesn't seem to be any relationship

In eg ① $E(X) = 0 = E(Y)$
eg ② $E(X) = 0 = E(Y)$

Let us consider $P(X=0, Y=0) = P(X=2, Y=2) = \frac{1}{2}$

$$E(X) = 0.\frac{1}{2} + 2.\frac{1}{2} = 1$$

Still, we have $\underline{X=Y}$. (like in eg ①)

$$E(XY) = 0.0.\frac{1}{2} + 2.2.\frac{1}{2} = 2$$

The relationship is the same, but the ~~mean~~ expected value is different. Hence $E(XY)$ is not a good measure.

But if we subtract the mean values, the measure remains the same.

$$\text{cov}(XY) = E\left[(X - m_x)(Y - m_y)\right]$$

Date
Page

- $cov(X,Y)$ takes values +ve, −ve and 0.

- Another measure that takes values between −1 and +1 can also be used to measure how two r.v.s X, Y covary.

That measure is called correlation coefficient ρ_{XY}.

$$\rho_{XY} = \frac{E[(X-m_X)(Y-m_Y)]}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}.$$

$$-1 \leq \rho_{XY} \leq 1$$

↳ Let $u = X - m_X$

$E(X) = m_X$

$E(u)$

$= E(X) - E(m_X)$

$= 0.$

u is a 0 mean random variable.

- Consider r.v.s X and Y.

$E[(X+\alpha Y)^2] = f(\alpha) \Rightarrow$ always +ve.

$$f(\alpha) = E[X^2 + \alpha^2 Y^2 + 2\alpha XY]$$

$$f(\alpha) = E[X^2] + \alpha^2 E[Y^2] + 2\alpha E[XY]$$

$$f(\alpha) = A\alpha^2 + B\alpha + C$$

$A = E(Y^2)$

$B = 2E(XY)$

$C = E(X^2).$

Here, the discriminant is $\Delta \leq 0$

$$B^2 \leq 4AC$$

$$4(E(XY))^2 \leq 4E(Y^2)E(X^2)$$

$$|E(XY)| \leq \sqrt{E(Y^2)E(X^2)}$$

Replacing X, Y with mean subtracted U, W.

$$|E(UW)| \leq \sqrt{E[(Y-m_Y)^2]\,E[(X-m_X)^2]}$$

$$|E[(X-m_X)(Y-m_Y)]| \leq \sigma_Y \sigma_X$$

$$|\rho_{XY}| \leq 1$$

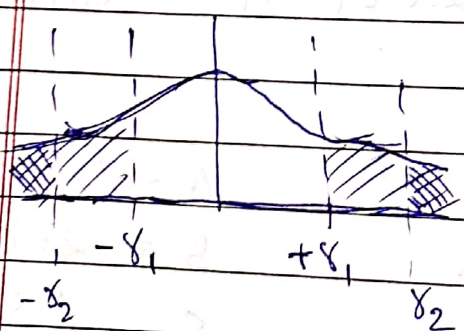
$$\therefore \quad -1 \leq \rho_{XY} \leq 1.$$

* CHEBYSHEV INEQUALITY

Know about

The bounds on the probability of occurrence of a random variable, based on mean and variance, i.e., without using PDF.

$$P(\underbrace{|X - m_X|}_{\text{r.v. with zero mean}} > \gamma) \leq B$$



$B_2 < B_1$

This inequality does not require PDF. Only mean, variance.

classmate

Date

Page

18/02

$$var(x) = \sigma_x^2 = \int_{-\infty}^{\infty} (x - m_x)^2 f_x(x)\, dx$$

$$= \int_{|x-m_x| > \gamma} (x-m_x)^2 f_x(x)\, dx + \int_{|x-m_x| \le \gamma} (x-m_x)^2 f_x(x)\, dx.$$

$$var(x) \geq \int_{|(x-m_x)| > \gamma} \underbrace{(x-m_x)^2}_{\gamma} f_x(x)\, dx$$

$$var(x) \geq \gamma^2 \underbrace{\int_{|x-m_x| > \gamma} f_x(x)\, dx}_{P(|x-m_x| > \gamma)}$$

$$\therefore \quad P(|x-m_x| > \gamma) \leq \frac{var(x)}{\gamma^2}$$

let

$$\gamma = K\sigma_x.$$

$$P(|(x-m_x)| > K\sigma_x) \leq \frac{\sigma_x^2}{K^2\sigma_x^2}$$

$$\therefore \quad P(|x-m_x| > K\sigma_x) \leq \frac{1}{K}$$

*see applications to Gaussians

Scanned by CamScanner

$\hookrightarrow E[(X-\sigma_x)(Y-\sigma_x)] \le \sigma_Y \sigma_X$.

MARKOFF INEQUALITY :

For $f_X(x) = 0$ for $x < 0$

$$P[X \geq \alpha] \leq \frac{m_X}{\alpha}$$

* Continue with covariance.

$$Z = X + Y$$

$$Var(X+Y) = Var(Z) = Var(X) + Var(Y) + 2\,cov(X,Y)$$

Since, $var(Z) = E[(Z - M_Z)^2]$ or

$$= E(Z^2) - [E(Z)]^2$$

$$= E(X^2 + Y^2 + 2XY) - [E(X+Y)]^2$$

$$= E[X^2] + E[Y^2] + E[2XY] - (E[X] + E[Y])^2$$

$$= E[X^2] - (E[X])^2 + E[Y^2] - (E[Y])^2 + E[2XY] - 2E[X]E[Y].$$

$$= var(X) + var(Y) + 2\,Cov(X,Y)$$

Uncorrelated

$$var(Z) = var(X) + var(Y).$$

$\hookrightarrow$ and $var(X-Y) = var(X) + var(Y) - 2\,cov(X,Y)$

Date ______
Page ______

$\hookrightarrow$ $var(x+y) = var(x) + var(y) + cov(x,y) + cov(y,x)$

$var(x_1 + x_2) = var(x_1) + var(x_2) + cov(x_1, x_2) + cov(x_2 x_1)$

$$\begin{aligned} var(x_1 + x_2 + x_3) = {} & var(x_1) + var(x_2) + var(x_3) \\ & + cov(x_1, x_2) + cov(x_1, x_3) \\ & + cov(x_3 x_2) + cov(x_2, x_1) \\ & + cov(x_3, x_1) + cov(x_2, x_3). \end{aligned}$$

$\star$ Consider

$$var(x_1 + x_2) = \underset{1\times 2}{\begin{bmatrix} 1 & 1 \end{bmatrix}} \underset{2\times 2}{\begin{bmatrix} \sigma_{x_1}^2 & cov(x_1, x_2) \\ cov(x_2 x_1) & \sigma_{x_2}^2 \end{bmatrix}} \underset{2\times 1}{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}$$

COVARIANCE MATRIX.

In general if there are n random variables

$$C_x = \begin{bmatrix} var(x_1) & cov(x_1 x_2) & \cdots & cov(x_1 x_n) \\ cov(x_2 x_1) & var(x_2) & \cdots & cov(x_2, x_n) \\ \vdots & \vdots & & \vdots \\ cov(x_n x_1) & \cdots & \cdots & var(x_n) \end{bmatrix}$$

19/02

Properties of C_x (covariance matrix)

1) It is a real symmetric matrix.

$C_x = C_x^T$

2) Covariance matrix is positive semidefinite

- If a matrix $A_{(M\times n)}$ is semidefinite then,

$$\underline{x}^T A \underline{x} \geq 0 \qquad \text{for } \underline{x} \in \mathbb{R}^n$$

$u_1 = x_1 - m_{x_1}$ and $u_2 = x_2 - m_{x_2}$

u_1 and u_2 are mean subtracted random variables.

then, $var(a_1u_1 + a_2u_2) = E\left[(a_1u_1 + a_2u_2)^2\right] - \left[E(a_1u_1 + a_2u_2)\right]^2$

$\rightarrow 0$ since u_1, u_2 have zero mean.

$$var = E(a_1^2u_1^2 + a_2^2u_2^2 + 2a_1a_2u_1u_2)$$

$$= a_1^2 E(u_1^2) + a_2^2 E(u_2^2) + 2a_1a_2 E(u_1u_2)$$

$$= a_1 E\left[(x_1 - m_{x_1})^2\right] + a_2^2\left[(x_2 - m_{x_2})^2\right] + 2a_1a_2 E\left[(x_1 - m_{x_1})(x_2 - m_{x_2})\right]$$

$$= (a_1 \ a_2)\begin{bmatrix} E(x_1 - m_{x_1})^2 & E(x_1 - m_{x_1})(x_2 - m_{x_2}) \\ E(x_2 - m_{x_2})(x_1 - m_{x_1}) & E(x_2 - m_{x_2})^2 \end{bmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \geq 0$$

Since var has to be ≥ 0. P

$\therefore P \geq 0$.

$\downarrow$

$x^T A x$

square matrix

$$v^T C_x v \geq 0 \qquad\qquad C v = \lambda v$$

eigenvalue.

$$v^T \lambda v \geq 0$$

$$\lambda \|v\|^2 \geq 0.$$

Since C_x is real, symmetric matrix, what can you say about eigen vectors and values of C_x?

- ↳ Eigen values are all real.
- ↳ Eigen vectors of distinct eigenvalues are all orthogonal.

* PRINCIPAL COMPONENT ANALYSIS (PCA): or KL transform.

DFT
$$\begin{bmatrix} X[0] \\ X[1] \\ X[2] \\ X[3] \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & j & -1 & -j \\ 1 & -1 & 1 & -1 \\ 1 & -j & -1 & j \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \\ x[2] \\ x[3] \end{bmatrix}$$

$$X = Bx.$$

DFT is a linear transform.

$$x = B^{-1} X = \frac{1}{4} (B^*)^T X.$$

B is orthogonal matrix.

Consider $Y = AX$.

$$\begin{bmatrix} \\ \\ \end{bmatrix} = \begin{bmatrix} & & \\ & & \end{bmatrix}\begin{bmatrix} \\ \\ \end{bmatrix}$$

~~Set~~ highly decorrelated.

highly correlated.

(*). All X_1 to X_n will packed in fewer Y_1 to Y_m $(m << n)$.
⇒ highest energy ⇒ optimum transform

PCA is a data dependent transformation

21/02

Eigenvalues of a symmetric matrix are real and eigenvectors of distinct eigenvalues are orthogonal.

$A \rightarrow$ Symmetric matrix.

$Ax = \lambda x$.

$(x^*)^T A x = (x^*)^T \lambda x$

and $\quad Ax^* = \lambda^* x^* \qquad (\lambda = \lambda^*$ when λ real).

$\therefore \quad (x^*)^T A x =$

Scanned by CamScanner

-Missed 1 lec 21/02-

25/02

$Y = AX$

X → correlated input

Y → uncorrelated output.

Hence, covariance matrix of Y will be diagonal.

- prev lec -

$A = U^T$ → choosing this makes C_Y diagonal

$$U = \begin{bmatrix} U_{11} & U_{21} \\ U_{12} & U_{22} \end{bmatrix}$$

with columns u_1, u_2.

u_1: 1st column of eigenvector matrix of X.

This will be useful in compression, feature extraction, etc.

$$\underset{(N\times 1)}{\underline{Y}} = \underset{(N\times N)}{U^T}\ \underset{(N\times 1)}{\underline{X}}$$

↓

Transformation is decided by $x_1 \dots x_N$ unlike DFT, DCT, etc. (Hence, data dependent).

which leads to an optimum linear transform in ~~a sense~~ MSE between the original input and reconstructed output when M number of Y's are retained where $M < N$.

$\underline{X} = U\underline{Y}$

Since $U^T = U^{-1}$ (orthogonal).

↓

Here, the $MSE = 0$ since all Ys are used to get back X.

Suppose $N = 8$, $M = 3$.

$$\underset{(8\times1)}{\hat{X}} = \underset{(8\times8)}{U}\ \underset{(8\times1)}{\underline{Y}}$$

$$\begin{bmatrix} \text{———} \\ \text{———} \\ \text{———} \\ \\ \\ \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ 0 \\ 0 \\ \vdots \end{bmatrix}_{8\times1}$$

Only $(M=3)$ non zero values in Y.

$\therefore$ only the first 3 columns of U are used.

$\therefore$ U becomes 8×3, taking Y as 3×1

$$\therefore \underset{(8\times1)}{\hat{X}} = \underset{(8\times3)}{U}\ \underset{(3\times1)}{\underline{Y}}.$$

$E\left[(X-\hat{X})^2\right]$ will give the MSE.

This MSE is less than other methods ~~that~~ we ~~using~~ fixed transforms.

$\therefore$ 8 RVs can be stored using 3 RVs.

$$C_Y = \Sigma = \begin{bmatrix} \sigma_{Y_1}^2 & cov(Y_1 Y_2) & \cdots & cov(Y_1, Y_n) \\ cov(Y_2 Y_1) & \sigma_{Y_2}^2 & \cdots & cov(Y_2 Y_n) \\ \vdots & & & \vdots \\ cov(Y_n Y_1) & cov(Y_n Y_2) & \cdots & \sigma_{Y_n}^2 \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \lambda_N \end{bmatrix}$$

Date
Page

σ^2 : contains total power content.

↳ Total power content can be found by adding all elements of covariance matrix. (NxN)

After transformation, the total power is sum of diagonal elements only (N elements)

Since by Parseval's th., input power = output power, the ~~input~~ data's ~~distributed~~ input power is concentrated in the output's Y values.

↳ Most of the power contained in Y_1.

↳ $C_Y = A C_X A^T$

$$\begin{bmatrix} \sigma_{Y_1}^2 & cov(Y_1Y_2) \\ cov(Y_2Y_1) & \sigma_{Y_2}^2 \end{bmatrix} = \underset{2\times2}{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}} \underset{2\times2}{\begin{bmatrix} \sigma_{X_1}^2 & cov(X_1X_2) \\ cov(X_2X_1) & \sigma_{X_2}^2 \end{bmatrix}} \underset{2\times2}{\begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}}$$

$$\cancel{\sigma_{Y_1}^2} = \begin{bmatrix} a_{11}\sigma_{x_1}^2 + a_{12}cov(x_2x_1) & a_{11}cov(x_1x_2) + \sigma_{x_2}^2 a_{12} \\ a_{21}\sigma_{x_1}^2 + a_{22}cov(x_2x_1) & a_{21}cov(\,) + a_{22}\sigma_{x_2}^2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}$$

$$\sigma_{Y_1}^2 = a_{11}^2\sigma_{x_1}^2 + a_{11}a_{12}cov(x_2x_1) + a_{11}a_{12}cov(x_1x_2) + a_{12}^2\sigma_{x_2}^2$$

$$\sigma_{Y_1}^2 = \begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix}^T \begin{bmatrix} \sigma_{x_1}^2 & cov \\ cov & \sigma_{x_2}^2 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix}$$

$$\sigma_{Y_1}^2 = a_1^T \underline{C_X}\, a_1$$

$\sigma_{y_1}^2 = a_1^T C_x a_1$

Eigenvectors of C_x are orthogonal since C_x is symmetric and real.

Now consider $Y = AX$

$L(a_1, \lambda) = a_1^T C_x a_1 - \lambda(\|a_1\|^2 - 1)$

Find a_1 such that this maximized.

Diff. wrt a_1 and equate to zero.

$\frac{\partial(a_1, \lambda)}{\partial a_{11}}$ and $\frac{\partial(a_1, \lambda)}{\partial a_{12}}$

$\downarrow$

$2a_{11}\sigma_{x_1}^2 + 2a_{12}\,cov(x_1, x_2) - 2\lambda a_{11} = 0.$

||ly for a_{12}.

$\therefore\ 2C_x a_1 - 2\lambda a_1 = 0.$

$\therefore\ C_x a_1 = \lambda a_1$

a_1: eigenvector corresponding to max. eigen value.