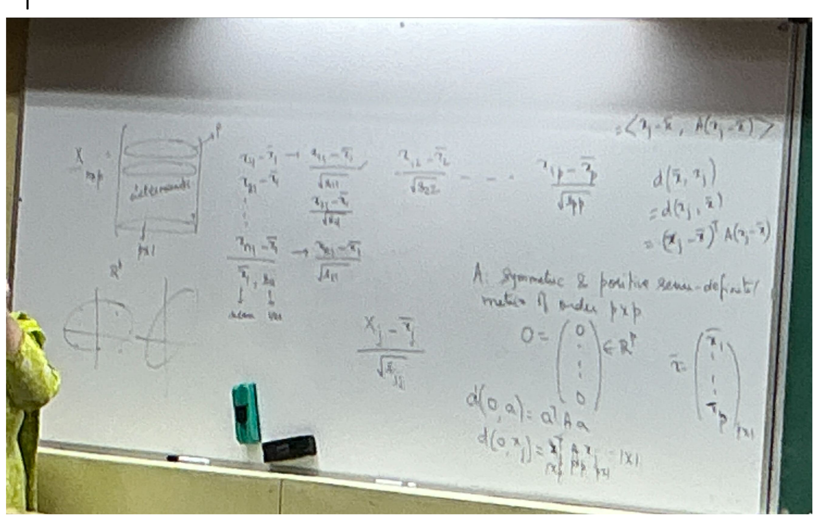


Take $A \leftarrow S^{-1}$
$\hookrightarrow$ variance – covariance matrix $p \times p$

$$S = \begin{bmatrix} s_{11} & 0 & \cdots & 0 \\ 0 & s_{22} & \cdots & 0 \\ \vdots & \vdots & & \\ 0 & 0 & \cdots & s_{pp} \end{bmatrix}_{p\times p}$$

All features are independent

$$S^{-1} = \begin{bmatrix} \frac{1}{s_{11}} & 0 & \cdots & 0 \\ & \frac{1}{s_{22}} & \ddots & \\ & & & \frac{1}{s_{pp}} \end{bmatrix}$$

$s_k^{-1} = \begin{bmatrix} 0 \\ \vdots \\ \frac{1}{s_k} \\ \vdots \end{bmatrix}$ each row

$d(x_i, x_j) = x_i^T x_j$

- distance is just an inner product
- distance is not unique we can define distance b/w vectors, matrix, etc;

A matrix can also defined such as diagonal matrix, etc.
special cases: A can be identity or A can be S^{-1}!

Since we take A is invertible, none of the eigen values can be zero. So we need S is positive definite.
$A = S^{-1}$ only if S is pd.
Mahalanobis distance.

we don't specifically need a rotation matrix, only ~~spd~~ psd matrix.

Linear Algebra

$A_{p\times p}$: symmetric & p.d.
$\hookrightarrow \underbrace{\lambda_1, \cdots \lambda_p}_{\text{real} +\text{ve}}$ eigenvalues of A

Any symmetric matrix are diagonalisable
All its E.V. $\to$ real, +ve

Matrix decomposition, symm. matrix are diagonalisable

$A = P^T \Lambda P$ $\to$ diagonal matrix with λ_i at diagonal
$\hookrightarrow$ P an eigenvector $p \times p$ matrix
$\to$ all eigenvectors are L.I. for diagonalisability
$\Rightarrow$ P is invertible & its columns are basis of $\mathbb{R}^p$
$\hookrightarrow$ orthonormalise this basis by Gram Schmidt orthonormalisation
P is orthonormal matrix $P^TP = PP^T = I \Leftrightarrow P^{-1} = P^T$

geo mul = Alg mul for these; eigen vectors L.I. [even if the eigen values are same. $A = I$ $\lambda_1 = \cdots = \lambda_p = 1$]
$\Rightarrow$ they form a basis
$\Rightarrow$ we can orthonormalise by Gram Schmidt process?

Suppose $\vec{a} \in \mathbb{R}^p$ &

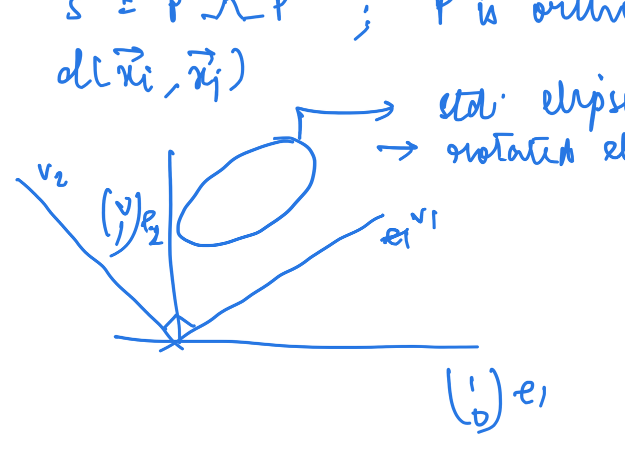
$$\begin{aligned} d(0, \vec{a}) &= \underset{1\times p}{\vec{a}^T} \underset{p\times p}{A} \underset{p\times 1}{\vec{a}} \\ &= \vec{a}^T P^T \Lambda P \vec{a} \\ &= (P^T\vec{a})^T \Lambda (P^T \vec{a}) \\ &= b^T \Lambda b \end{aligned}$$

No cross diag term.
rotation matrix.

$(\underset{p\times 1}{\vec{b}} = \underset{p\times p}{\vec{P}} \underset{p\times 1}{\vec{a}})$

P $\to$ orthonormal matrix $P^T \to$ orthonormal.
orig vector a rotated to b.
b becomes a standard vector.

In particular if $A = S^{-1}$
$S^{-1} = P \Lambda P^T$; P is orthonormal
$d(\vec{x}_i, \vec{x}_j)$
$\to$ std. ellipse w.r.t. (v_1, v_2) $\to$ orthonormal
$\to$ rotated ellipse w.r.t. (e_1, e_2)



$$\begin{aligned} d(0, \vec{x}_j) &= x_j^T S^{-1} x_j \\ &= x_j^T P \Lambda P^T x_j \\ &= x_j^{*T} \Lambda x_j^* \end{aligned}$$

$x_j^* - \vec{x} = P^T(x_j - \vec{x})$

$d(\vec{x}, x_j) = (x_j^* - \vec{x}^*)^T \Lambda (x_j^* - \vec{x}^*)$

$x_j^{*T} \Lambda x_j^* = \sum_{i=1}^{p} \lambda_i (x_{ij}^*)^2 = c^2 \to$ constant

$\sum_{i=1}^{p} \frac{(x_{ji})^2}{\left(\frac{c}{\sqrt{\lambda}}\right)^2} = 1$ $\to \lambda_i > 0$ (a S^{-1} is p.d.)

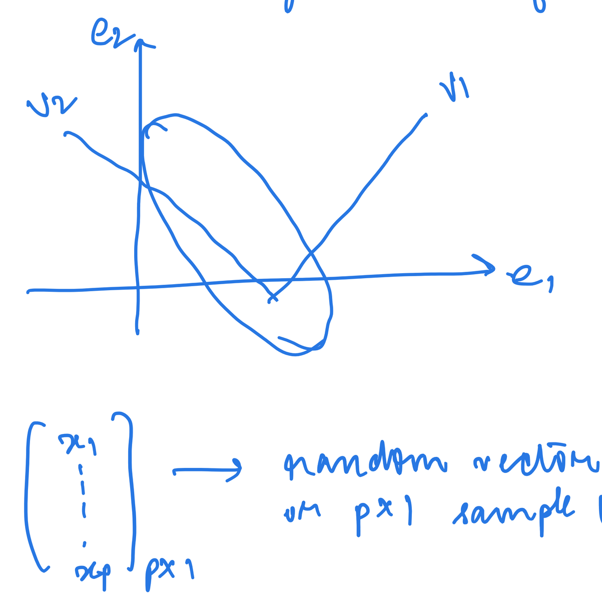
hyper ellipse in $\mathbb{R}^p$ with length of semi axis $\frac{c|c|}{\sqrt{\lambda_i}}$

x_j $(x_{j1}\ x_{j2}\ \cdots\ x_{jp})$ $\to j^{th}$ sample

under root & reciprocal of the eigen value represents the semi-axes.

The p-axes equations is $x_j^* = 0$
$\Rightarrow (\underset{p\times p}{P^T} \underset{p\times 1}{x_j}) = 0$

d.c's of normal (columns) to the planes.
eigenvectors of S^{-1} are giving new frame of reference



tilted in the original system.
in v_1 v_2 it is standard.

$X = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}_{p\times 1}$ $\to$ random vector of $p \times 1$
or $p \times 1$ sample (after experiment)

Let $\vec{a} \in \mathbb{R}^p$
$Y = \vec{a}^T \cdot X = a_1 \cdot x_1 + \cdots + a_p \cdot \vec{x}_p$
$\downarrow$
random variable

$E(Y) = a_1 E(x_1) + \cdots + a_p E(x_p) = a^T E(X)$

$$\begin{aligned} var(Y) = var(a^T X) &= E(Y - E(Y))^2 \\ &= E(a^T \underset{\sim}{X} - E(a^T X))^2 \\ &= E[a^T X - a^T \bar{X}]^2 \\ &= E[\underset{1\times p}{a^T} \underset{p\times 1}{(X - \bar{X})}]^2 \\ &= E(\underset{1\times p}{\vec{a}^T} \underset{p\times 1}{(X - \bar{X})} \underbrace{\underset{1\times p}{(X - \bar{X})^T} \underset{p\times 1}{\vec{a}}}_{1\times 1}) \\ &= \underset{1\times p}{\vec{a}^T} E(X - \bar{X})(X - \bar{X})^T \underset{p\times 1}{\vec{a}} \end{aligned}$$

These are all sample, we can replace w population & it becomes population matrix.

$\vec{a}^T S \vec{a}$ (if we are in sample).
$\vec{a}^T \Sigma \vec{a}$ (if we take population w $\bar{x} = \mu$).