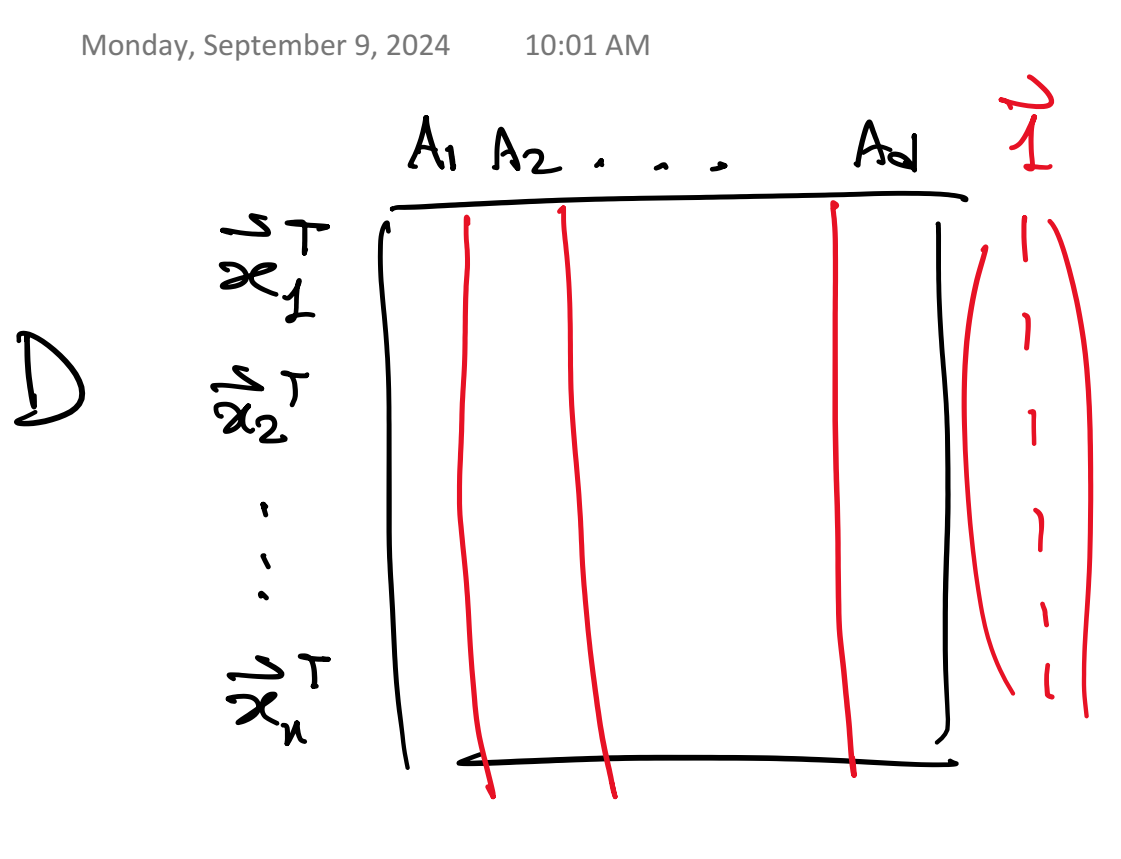
Lecture 3

Monday, September 9, 2024 10:01 AM



$D \in \mathbb{R}^{n \times d}$

is there a smaller subspace that is more relevant for (a subset) of the data

Q: $\underline{k} \ll d$?

$\vec{\mu} = \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \vdots \\ \hat{\mu}_d \end{pmatrix}$ mean of the data

$\hat{\mu}_i = \frac{A_i^T \vec{1}}{\vec{1}^T \vec{1}}$ mean along A_i

$D = \left[\vec{A}_1 \middle| \vec{A}_2 \middle| \cdots \middle| \vec{A}_d \right] \in \mathbb{R}^{n \times d}$

$D^T \in \mathbb{R}^{d \times n}$

$$\begin{bmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_d^T \end{bmatrix}$$

$$\boxed{\hat{\vec{\mu}} = \frac{D^T \vec{1}}{n}}$$

do matrix operations

$\underbrace{\underset{d \times n}{D^T} \quad \underset{n \times 1}{\vec{1}}}_{d \times 1}$

Covariance matrix (cov : $\hat{\Sigma}$)

$\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_d^2$: variance along each A_i (avg. sq. length of centered vector $\bar{A}_i$) : $\frac{\|\bar{A}_i\|^2}{n}$

$\hat{\sigma}_{ij}$: cov(A_i, A_j)

Covariance

$\frac{\bar{A}_i^T \bar{A}_j}{n}$: dot product between centered vectors (div by n)

$\hat{\rho}_{ij}$: Cor(A_i, A_j) ≡ Correlation

$\cos \theta_{ij} = \left(\frac{\bar{A}_i}{\|\bar{A}_i\|} \right)^T \left(\frac{\bar{A}_j}{\|\bar{A}_j\|} \right)$

$\hat{\Sigma}$: Covariance matrix

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1d} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \cdots & \sigma_{2d} \\ \vdots & & & & \\ \sigma_{1d} & \sigma_{2d} & & \cdots & \sigma_d^2 \end{pmatrix} = \{ \sigma_{ij}^2 \}_{\substack{i = 1 \ldots d \\ j = 1 \ldots d}}$$

$$\boxed{\sigma_{ii} \equiv \sigma_i^2}$$

real matrix $\mathbb{R}^{d \times d}$

d×d : square

$\hat{\Sigma}^T = \hat{\Sigma}$: symmetric

$\hat{\Sigma}$ is positive semi-definite (PSD)

all eigenvalues are real and non-negative

⇒ arrange them as follows

$\lambda_1 \geq \lambda_2 \geq \lambda_3 \cdots \geq \lambda_d \geq 0$ eigenvalues

$\downarrow \quad \downarrow \quad \downarrow \quad \quad \downarrow$

$\vec{u}_1 \quad \vec{u}_2 \quad \vec{u}_3 \quad \quad \vec{u}_d$ ← eigenvectors (special direction)

Σ is PSD iff $\forall \vec{x} \in \mathbb{R}^d$
(d×d)

$\underbrace{\boxed{\vec{x}^T \Sigma \vec{x}}}_{\text{quadratic form}} \geq 0$

Scalar

$\|\vec{x}\|^2 = \langle \vec{x}, \vec{x} \rangle = \boxed{\vec{x}^T \vec{x}}$

$= x_1^2 + x_2^2 + \cdots + x_d^2$

$= \sum x_i^2$

$I = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & & 1 \end{pmatrix}$

Identity matrix (PSD)

$\vec{x}^T I \vec{x}$

⇓

$\underbrace{\vec{x}^T \Sigma \vec{x}} = \sum_i \sum_j x_i \sigma_{ij} x_j$

$\underset{\vec{x}^T}{(a \;\; b)} \underbrace{\boxed{\begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix}}}_{\Sigma} \underset{\vec{x}}{\begin{pmatrix} a \\ b \end{pmatrix}}$ $x = \begin{pmatrix} a \\ b \end{pmatrix}$

$(a \;\; b) \begin{pmatrix} 5a - 2b \\ -2a + 2b \end{pmatrix}$

$a(5a - 2b) + b(-2a + 2b)$

$\boxed{5a^2 - 2ab - 2ab + 2b^2} \geq 0$

$\underline{5a^2} - \underline{4ab} + \underline{2b^2}$

$\hat{\Sigma} : \underline{\underline{\mathbb{R}^{d \times d}}}$

$M : \underline{\underline{\mathbb{R}^{m \times n}}}$ matrix

$\underset{3 \times 2}{M} \quad \underbrace{\begin{pmatrix} 2 & 1 \\ 1 & 2 \\ 1 & 1 \end{pmatrix}}_{M} \underset{2 \times 1}{\begin{pmatrix} 1 \\ 1 \end{pmatrix}} = \underset{3 \times 1}{\begin{pmatrix} 3 \\ 3 \\ 2 \end{pmatrix}}$

$M \vec{x} = \vec{y}$

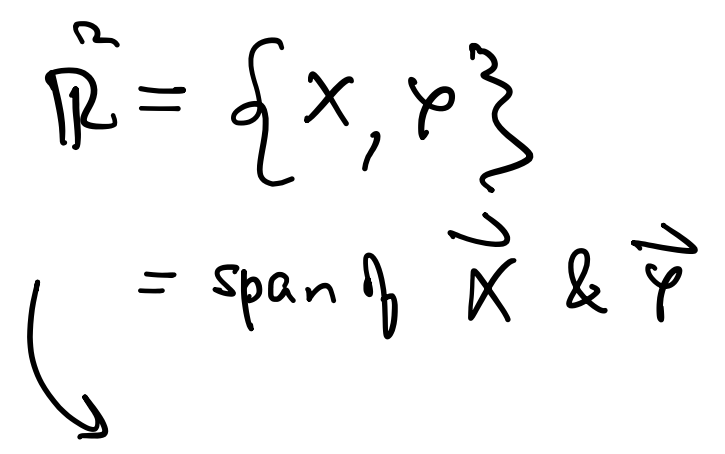
$M : \underbrace{f(\vec{x})}_{\text{operator}} : \underset{2D}{\mathbb{R}^2} \rightarrow \underset{3D}{\mathbb{R}^3}$

$\underset{m \times n}{M} \equiv f : \mathbb{R}^n \rightarrow \mathbb{R}^m$

$\hat{\Sigma} : d \rightarrow d$

$\hat{\Sigma} \vec{x} = \vec{y}$ $\vec{x}, \vec{y} \in \mathbb{R}^d$

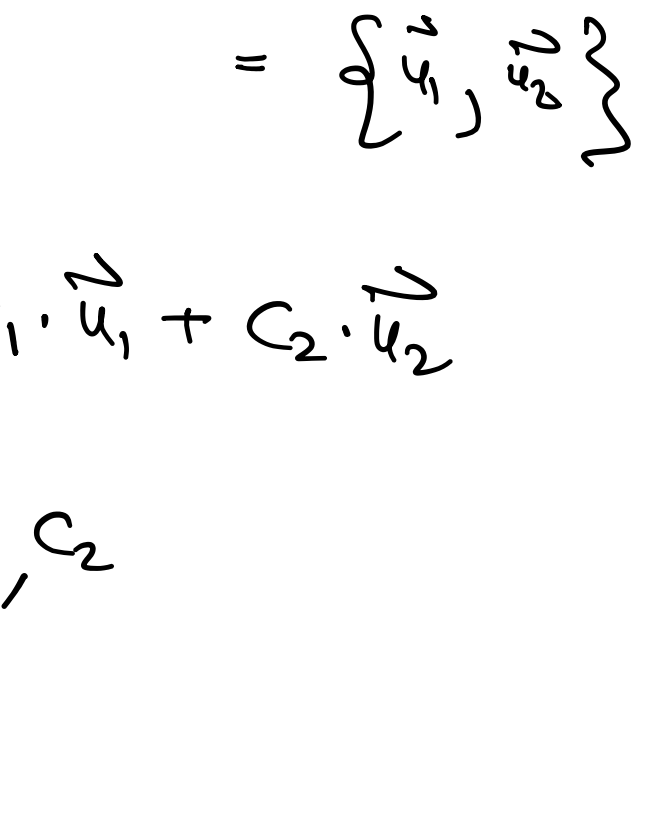
$\begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 3 \\ 0 \end{pmatrix}$



$$\boxed{\hat{\Sigma} \vec{u} = \textcircled{\lambda} \vec{u}}$$

eigenvector · Scalar eigenvalue · eigenvector

$\vec{u}$: eigenvector
invariant to $\hat{\Sigma}$ except for scaling/flipping



$\hat{\Sigma}$ is symmetric ⇒ eigenvectors are orthogonal

e.g. $\hat{\Sigma} \in \mathbb{R}^{2 \times 2}$ → $\vec{u}_1$ — λ_1

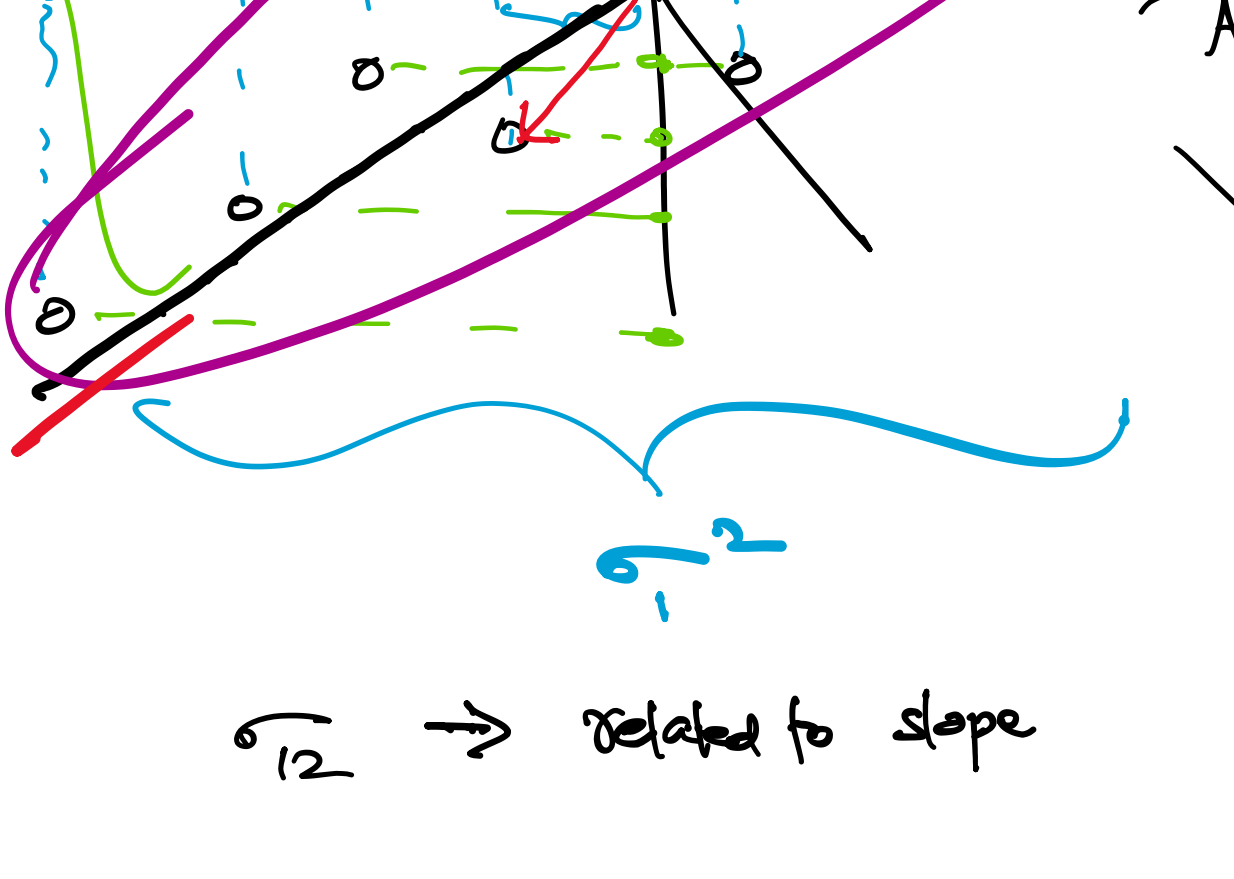
→ $\vec{u}_2$ — λ_2

→ $\vec{u}_1^T \vec{u}_2 = 0$

$\mathbb{R}^2 = \{X, Y\}$

$= \text{span of } \vec{X} \text{ \& } \vec{Y}$

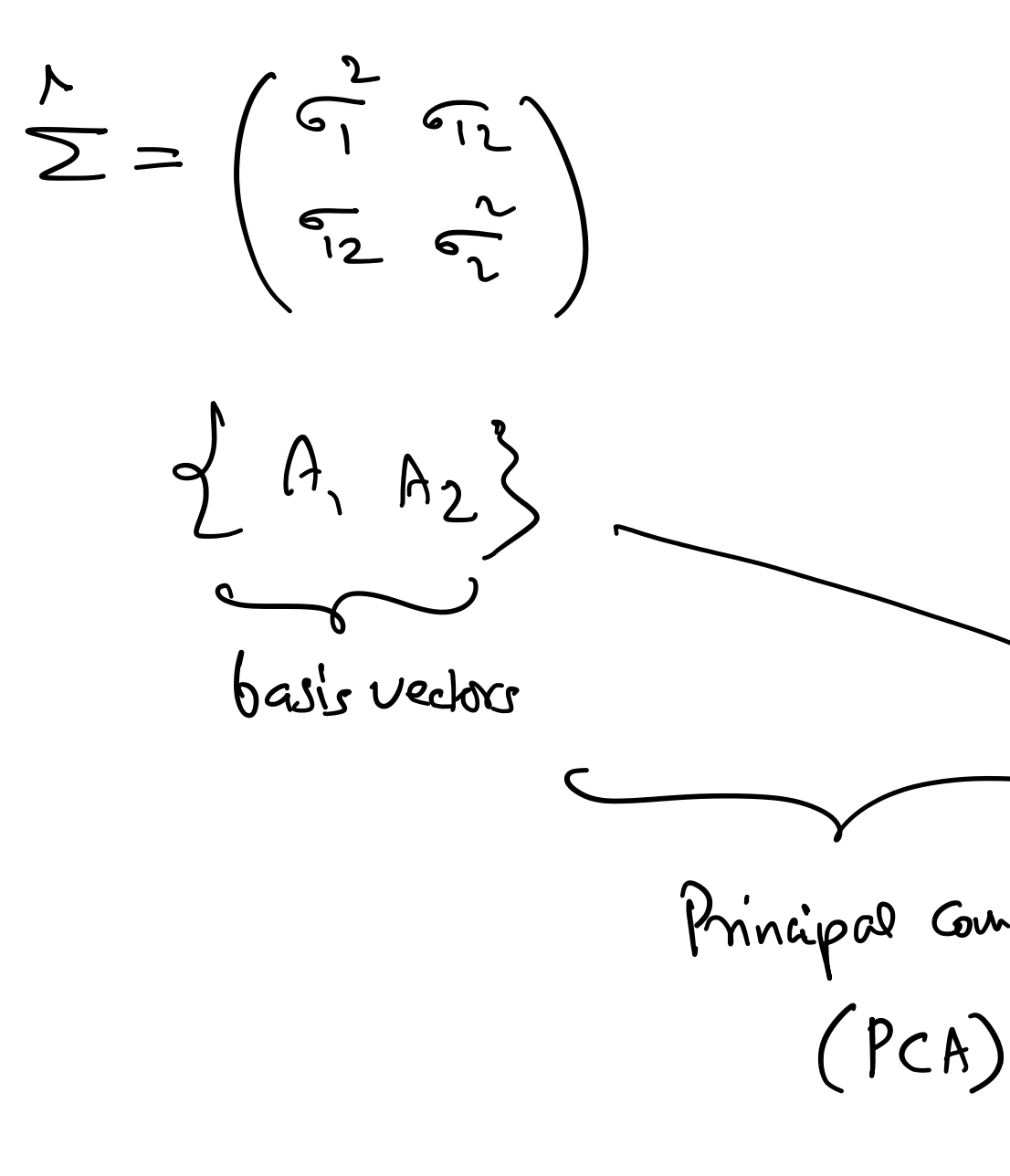
$= \{\vec{u}_1, \vec{u}_2\}$



$\vec{x} = c_1 \cdot \vec{u}_1 + c_2 \cdot \vec{u}_2$

$\forall c_1, c_2$

$\mathbb{R}^2$

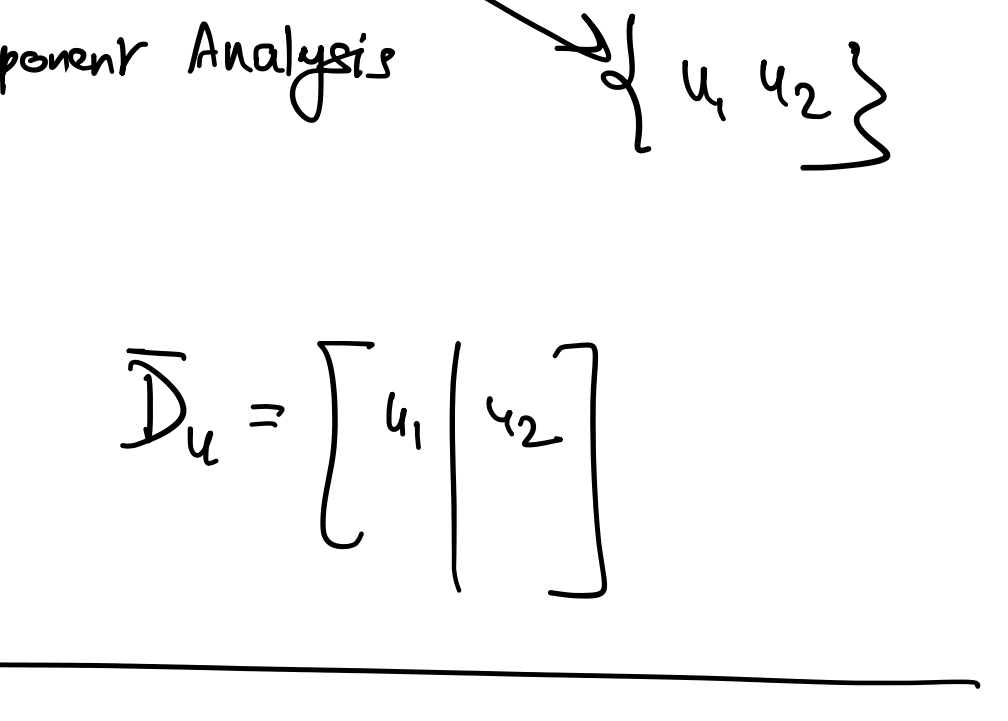


$\bar{D}$ ≡ remove $\vec{\mu}$ (mu)

Centered data

$= \left[\bar{A}_1 \;\; \bar{A}_2 \right]$

σ_{12} → related to slope



no rotation

⇒ no slope

⇒ no covariance

$\hat{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$

$\Lambda = \underbrace{\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}}$

$\underbrace{\{A_1, A_2\}}_{\text{basis vectors}}$ → $\{u_1 \; u_2\}$

Principal Component Analysis (PCA)

$\bar{D}_u = \left[u_1 \middle| u_2 \right]$

$\hat{\Sigma} = \begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix}$

1) $\hat{\Sigma} \vec{u}_1 = \lambda_1 \vec{u}_1$
2) $\hat{\Sigma} \vec{u}_2 = \lambda_2 \vec{u}_2$

$\vec{u}_1 \neq \vec{0}$
$\vec{u}_2 \neq \vec{0}$

$\hat{\Sigma} \vec{u} = \lambda \underline{\vec{u}}$

$\hat{\Sigma} \vec{u} = \lambda \vec{u}$

$$\boxed{\hat{\Sigma} \vec{u} - \lambda \vec{u} = \vec{0}}$$

2×1

$\underset{\substack{2 \times 2 \\ M}}{\underbrace{(\hat{\Sigma} - \lambda I)}} \underset{2 \times 1}{\vec{u}} = \underset{2 \times 1}{\vec{0}}$

$\underset{2 \times 2}{\hat{\Sigma}} - \underset{1}{\lambda_1}$

$M = \underbrace{\hat{\Sigma} - \lambda I}$

M must be singular (non invertible not full rank)

⇒ det(M) = 0

$\det(\hat{\Sigma} - \lambda I) = 0$

$\underbrace{\begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}}$

$M = \begin{pmatrix} 5 - \lambda & -2 \\ -2 & 2 - \lambda \end{pmatrix}$

$\det(M) = \boxed{(5 - \lambda)(2 - \lambda) - (-2 \times -2)}$ Characteristic polynomial

$= 10 - 7\lambda + \lambda^2 - 4$

$= \lambda^2 - 7\lambda + 6 = 0$

$= (\lambda - 6)(\lambda - 1) = 0$

$\det(\Sigma - \lambda I) = 0$
solve for λ

$\Rightarrow \underbrace{\underset{\lambda_1}{\lambda = 6} \quad \text{or} \quad \underset{\lambda_2}{\lambda = 1}}_{\text{eigen values}}$

$\hat{\Sigma} \vec{u}_1 = \lambda_1 \vec{u}_1$

$\vec{u}_1 = \begin{pmatrix} x \\ y \end{pmatrix}$

$\begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix} \vec{u}_1 = 6 \vec{u}_1$

$\begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 6 \begin{pmatrix} x \\ y \end{pmatrix}$

$\begin{pmatrix} 5x - 2y \\ -2x + 2y \end{pmatrix} = \begin{pmatrix} 6x \\ 6y \end{pmatrix}$

$5x - 2y = 6x$

$x = -2y$

$\vec{u}_1 = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ -0.5 \end{pmatrix}$

⇓

unit vector

$\vec{u}_1 = \frac{\vec{u}_1}{\|\vec{u}_1\|} = \frac{1}{1.118} \begin{pmatrix} 1 \\ \underline{\underline{-0.5}} \end{pmatrix}$

$\|\vec{u}_1\| = \sqrt{1^2 + 0.5^2}$

$= \sqrt{1.25}$

$= 1.118$

$\lambda_1 = 6$ $\boxed{\vec{u}_1 = \begin{pmatrix} 0.894 \\ -0.447 \end{pmatrix}}$

$\|\vec{u}_1\| = 1$

$\vec{u}_1^T \begin{pmatrix} a \\ b \end{pmatrix} = 0$

$(0.894 \;\; 0.447) \begin{pmatrix} a \\ b \end{pmatrix} = 0$

$\Rightarrow 0.894a + 0.447b = 0$ a = 1

$0.447b = -0.894$

$b = -\frac{0.894}{0.447} \approx -2$

$\begin{pmatrix} 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 0.447 \\ 0.894 \end{pmatrix}$

$\lambda_1 = 1$

Power Iteration

$\hat{\Sigma} = \begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix}$ $\vec{x}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

Start with a random vector & normalize

$\hat{\Sigma} \vec{x}_0 = \begin{pmatrix} 3 \\ 0 \end{pmatrix} \xrightarrow[\text{divide by 3}]{\lambda_1 \sim 3} \underline{\underline{\begin{pmatrix} 1 \\ 0 \end{pmatrix}}} = \vec{x}_1$

(divide by the norm)

$\vec{u}_1^T \vec{u}_1 = 1$
$\vec{u}_2^T \vec{u}_2 = 1$
$\vec{u}_1^T \vec{u}_2 = 0$

Symmetric

$\hat{\Sigma} \vec{x}_1 = \begin{pmatrix} 3 \\ -2 \end{pmatrix} \xrightarrow[\text{div by 3}]{\lambda_1 \sim 3} \begin{pmatrix} 1 \\ -0.67 \end{pmatrix} = \vec{x}_2$

$\hat{\Sigma} \vec{x}_2 = \begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ -0.67 \end{pmatrix} = \begin{pmatrix} 6.34 \\ -3.34 \end{pmatrix} = \vec{x}_3$ $\lambda_1 \sim \underline{\underline{6.34}}$

⇓

$\begin{pmatrix} 1 \\ -0.53 \end{pmatrix} \approx \begin{pmatrix} 1 \\ -0.5 \end{pmatrix}$

$\hat{\Sigma} \vec{x}_0 = \vec{x}_1$

$\vec{x}_2 = \hat{\Sigma} \vec{x}_1 = \hat{\Sigma}(\hat{\Sigma} \vec{x}_0) = \hat{\Sigma}^2 \vec{x}_0$

$\vec{x}_3 = \hat{\Sigma} \vec{x}_2 = \hat{\Sigma}^3 \vec{x}_0$

⋮

$\underline{\underline{\vec{x}_t}} = \hat{\Sigma}^t \vec{x}_0$

$\hat{\Sigma} \vec{u}_1 = \lambda_1 \vec{u}_1$

$t \rightarrow \infty$: $\vec{x}_t \approx \vec{u}_1$

ratio of the largest element after & before ≈ λ_1

$\vec{u}_1$ — λ_1 power iteration

$\vec{u}_2$ $\vec{u}_3$ . . .

λ_2 λ_3

$\vec{x}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ $\vec{y}_0 = \begin{pmatrix} 0.5 \\ 2 \end{pmatrix}$: $\vec{y}_0$ is orthogonal to $\vec{x}$

$\hat{\Sigma} \vec{x}_0$ $\hat{\Sigma} \vec{y}_0$

↓ ↓

$\vec{u}_1, \lambda_1$ $\vec{u}_2, \lambda_2$

⇓

before $\hat{\Sigma}$ operation