$\ddot{\smile}$

7.1 Diagonalization of Symmetric Matrices

symmetric matrix is $A^T = A$ or $a_{i,j} = a_{j,i}$ (must be square)

$$\begin{bmatrix} 2 & 5 & -1 \\ 5 & 3 & 0 \\ -1 & 0 & 4 \end{bmatrix}$$

main diagonal can be anything but must be symmetric about main diagonal

zero matrix, diagonal matrix → symmetric matrix

diagonalize matrix if possible: ex → $A = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{bmatrix}$ symmetric

sol'n → $\det(A - \lambda I) = \det\left(\begin{bmatrix} 6-\lambda & -2 & -1 \\ -2 & 6-\lambda & -1 \\ -1 & -1 & 5-\lambda \end{bmatrix}\right)$

find all λ's and eigenvectors: $0 = (6-\lambda)^2(5-\lambda) - 2 - 2 - (6-\lambda) - (6-\lambda) - 4(5-\lambda)$

$$0 = (6-\lambda)(\lambda-8)(\lambda-3)$$

eigenvalues: 3, 6, 8

$$[A - 3I \mid 0] = \left[\begin{array}{ccc|c} 3 & -2 & -1 & 0 \\ -2 & 3 & -1 & 0 \\ -1 & -1 & 2 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \rightarrow \begin{array}{l} x_1 = x_3 \\ x_2 = x_3 \\ x_3 = x_3 \end{array} \quad \vec{E}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$[A - 6I \mid 0] = \left[\begin{array}{ccc|c} 0 & -2 & -1 & 0 \\ -2 & 0 & -1 & 0 \\ -1 & -1 & -1 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 1/2 & 0 \\ 0 & 1 & 1/2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \quad \begin{array}{l} x_1 = -1/2\, x_3 \\ x_2 = -1/2\, x_3 \end{array} \quad \vec{E}_6 = \begin{bmatrix} -1/2 \\ -1/2 \\ 1 \end{bmatrix} \text{ or } \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}$$

$$[A - 8I \mid 0] = \left[\begin{array}{ccc|c} -2 & -2 & -1 & 0 \\ -2 & -2 & -1 & 0 \\ -1 & -1 & -3 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \quad \begin{array}{l} x_1 = -x_2 \\ x_3 = 0 \\ x_2 = x_2 \end{array} \quad \vec{E}_8 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

so A is diagonalizable. we can make our P and D but we can normalize P since it is also an orthogonal set → ($E_3 \cdot E_6 = 0$, $E_6 \cdot E_8 = 0$, $E_3 \cdot E_8 = 0$)

$$\frac{E_3}{\|E_3\|} = V_1 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}, \quad V_2 = \begin{bmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}, \quad V_3 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$$ ← so use these for P instead

$$P = \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{6} & -1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{6} & 1/\sqrt{2} \\ 1/\sqrt{3} & 2/\sqrt{6} & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 8 \end{bmatrix}$$

2/1/19

$$Av_i = \lambda v$$

$A = \lambda_1 v_1$?

not exactly sure where λ_2 comes from how does this make sense from other

7.1 Theorem

$\boxed{\text{If } A \text{ is symmetric then its eigenvectors will be orthogonal}}$

Proof: $\{\lambda_1 : v_1\}, \{\lambda_2 : v_2\}$,

$Av_2 = \lambda_2 v_2 \Rightarrow (\lambda_1 v_1)\cdot v_2 = (\lambda_1 v_1)^T v_2 = (Av_1)^T v_2$

$(Av_1)^T v_2 = (v_1^T A^T) v_2 = v_1^T (Av_2) \rightarrow A^T = A$ then

$v_1^T(Av_2) = v_1^T(\lambda_2 v_2) = \lambda_2 v_1 \cdot v_2$

$\lambda_1(v_1\cdot v_2) = \lambda_2(v_1 \cdot v_2)$, $\lambda_1 \neq \lambda_2$ so $v_1 \cdot v_2 = 0$

- A is orthogonally diagonalizable if there is an orthonormal matrix P and a diagonal matrix D such that: (saw from last page!)

$A = PDP^{-1} = PDP^T \qquad P^{-1} = P^T$ cuz orthonormal

- A is orthogonally diagonalizable iff A is symmetric, proof complicated

ex $A = \begin{bmatrix} 3 & -2 & -4 \\ -2 & 6 & 2 \\ -4 & 2 & 3 \end{bmatrix} \quad \det(A - \lambda I) = \det\left(\begin{bmatrix} 3-\lambda & -2 & -4 \\ -2 & 6-\lambda & 2 \\ -4 & 2 & 3-\lambda \end{bmatrix}\right) = -(\lambda-7)^2(\lambda+2)$

multiplicity 2

$\boxed{\lambda = -2}$ $[A + 2I \mid 0] = \left[\begin{array}{ccc|c} 5 & -2 & 4 & 0 \\ -2 & 8 & 2 & 0 \\ 4 & 2 & 5 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 1/2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$

$x_1 = -x_3$
$x_2 = -\frac{1}{2}x_3$
$x_3 = x_3$

$\Rightarrow E_{-2} = \begin{bmatrix} -1 \\ -1/2 \\ 1 \end{bmatrix} = v_1$

$\boxed{\lambda = 7}$ $[A - 7I \mid 0] = \left[\begin{array}{ccc|c} -4 & -2 & 4 & 0 \\ -2 & -1 & 2 & 0 \\ 4 & 2 & -4 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 1/2 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$ $v_2 = E_7 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ $E_{7,2} = \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix} = v_3$

We have $v_1 \perp (v_2, v_3)$ but $v_2 \not\perp v_3$:

$v_3' = v_3 - \frac{v_3 \cdot v_2}{v_2 \cdot v_2} \cdot v_2 = \begin{bmatrix} -1/4 \\ 1 \\ 1/4 \end{bmatrix}$ ← Gram-Schmidt

now normalize v_1, v_2, v_3'

$P = \begin{bmatrix} -2/3 & 1/\sqrt{2} & -1/\sqrt{18} \\ -1/3 & 0 & 4/\sqrt{18} \\ 2/3 & 1/\sqrt{2} & 1/\sqrt{18} \end{bmatrix}, \quad D = \begin{bmatrix} -2 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 7 \end{bmatrix}$

$A = PDP^{-1} = PDP^T =$

$= [u_1 \cdots u_n] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} u_1^T \\ \vdots \\ u_n^T \end{bmatrix} = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + \cdots + \lambda_n u_n u_n^T$

orthonormal eigenvectors ↑

Spectral Decomposition of A

7.4

Singular values of an $m \times n$ matrix

→ if A is an $m \times n$ matrix then $A^T A$ is symmetric ($n \times n$)

$(A^T A)^T = A^T A^{TT} = A^T A$

By spectral theorem we see $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ eigenvals

$v_1 \quad v_2 \quad \ldots \quad v_n$ eigenvecs

such that $\{v_1, \ldots, v_n\}$ is an orthonormal basis

In fact, these eigenvals pop as:

$$\|Av_i\|^2 = (Av_i)^T \cdot (Av_i) = v_i^T A^T A v_i = v_i^T \lambda_i v_i = \lambda_i (v_i \cdot v_i) = \lambda_i \|v_i\|^2$$

if orthonormal then just λ_i

The singular values of A are $\sigma_1 = \sqrt{\lambda_1}, \ldots, \sigma_n = \sqrt{\lambda_n}$

↳ Square roots of the eigenvals of $A^T A$

if eigenvalue has multiplicity write singular value that many times.

σ_i is the length of Av_i, σ_1 is the largest factor by which a vector can grow

↳ we can use these to decompose A in a way that emphasizes its priorities.

Singular value decomposition (SVD):

$$\Sigma = \begin{array}{c|cc} & \overset{\leftarrow r \rightarrow}{} & \overset{\leftarrow n-r \rightarrow}{} \\ \hline r & D & 0 \\ m-r & 0 & 0 \end{array}$$

D is the diagonal matrix and its entries are the singular values.

$$D = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix}$$

★ if we have an $m \times n$ matrix, of rank r, we can write it like:

↳ $A = U \Sigma V^T$, $\Sigma = \left[\begin{array}{cc|c} \sigma_1 & 0 & 0 \\ 0 & \sigma_n & \\ \hline & 0 & 0 \end{array}\right]$, U = $m \times n$ orthogonal (orthonormal columns)

V = $n \times n$ orthogonal

$V = [v_1 \cdots v_n]$ → $\{v_1, \ldots, v_n\}$ is an orthonormal basis for $\mathbb{R}^n$

$U = [u_1 \cdots u_m]$ → $\{u_1, \ldots, u_m\}$ is an orthonormal basis for $\mathbb{R}^m$

$\begin{bmatrix} A & I \\ I & A^{-1} \end{bmatrix}$ its like doing $A^{-1}A$ but just vector

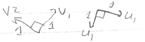
→ feeding i^{th} vector into its inverse returns the part of I, e_i

$$Av_i = U\Sigma V^T v_i = U \Sigma V^{-1} v_i = U \underbrace{\Sigma e_i}_{i^{th} \text{ column of } \Sigma} = U \sigma_i e_i = \sigma_i U e_i = \boxed{\sigma_i u_i}$$

$Av_i = \sigma_i u_i$

$\sigma_i = \sqrt{\lambda_i}$

example $\quad A = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}$



$$A = \underset{U}{\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}} \underset{\Sigma}{\begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix}} \underset{V}{\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}^T}$$

$V_1 \to$ scaler $\cdot u_1$ $\quad$ scalars of σ $\Big\}$ $Av_i = \sigma u_i$

$V_2 \to$ scaler $\cdot u_2$

By hand:

① find an ON diagonalization of A^TA $(\lambda, 1)$

② set up V & $\Sigma = \left[\begin{array}{cc|c} \sigma_1 & 0 & 0 \\ 0 & \sigma_n & \\ \hline 0 & & 0 \end{array}\right]$ $\quad$ σ are the $\sqrt{\lambda_i}$ from A^TA

V is orthonormal basis A^TA

③ construct U $\quad \sigma_i u_i = Av_i \quad u_i = \frac{1}{\sigma_i} Av_i$ for $\sigma_i \neq 0$

↳ if $\sigma_i = 0$, a little more work. (gram schmidt, include more vectors)

$\boxed{A = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}}$, find diagonalization of $A^TA = \begin{bmatrix} 2 & -1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix} = Z$

① → roots of $\det(Z - \lambda I) = \det\begin{pmatrix} 5-\lambda & 3 \\ 3 & 5-\lambda \end{pmatrix}$ $\quad \lambda = 2, 8$

→ eigenvectors: $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ gramschmidt

$\lambda = 2 \quad [Z - 2I \mid 0] \sim \left[\begin{array}{cc|c} 3 & 3 & 0 \\ 3 & 3 & 0 \end{array}\right] \to \text{Span}\left\{\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right\}$ E_2 $\quad$ turn these into ← orthonormal basis, already orthogonal so just normalize

$\lambda = 8 \quad [Z - 8I \mid 0] \sim \left[\begin{array}{cc|c} -3 & 3 & 0 \\ 3 & -3 & 0 \end{array}\right] \to \text{Span}\left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}$ E_8

$E_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}, \quad E_8 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$

② Singular values!

bigger $\to \sigma_1 = \sqrt{8} = 2\sqrt{2}, \quad \sigma_2 = \sqrt{2} \to \Sigma = \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix}, \quad V = \begin{array}{cc} v_1 & v_2 \end{array} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$

③ u_1 is normalized (unit vector) from $Av_1 = \sigma u_1 \qquad Av_1 = \begin{bmatrix} 4/\sqrt{2} \\ 0 \end{bmatrix}$

so $u_1 = \frac{1}{\sigma_1} \cdot \begin{bmatrix} 4/\sqrt{2} \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$u_2 = \frac{1}{\sigma_2} \cdot Av_2 = \frac{1}{\sigma_2}\begin{bmatrix} 0 \\ -\sqrt{2} \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}^T$$

pieces again:

corresponds to biggest σ

largest change in A

V_1 is the direction in the input space that corresponds to largest output vector

V_2 output changes the least

↳ u would by Av. cuz $Av = \sigma u$

Low rank Approximation:

$$A = U \Sigma V^T = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \left[\begin{array}{cc|c} \sigma_1 & 0 & 0 \\ 0 & \ddots\ \sigma_r & \\ \hline & 0 & 0 \end{array}\right] \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}^T$$

m×n matrix

$$\underbrace{\sigma_1 \geq \cdots \geq \sigma_b}_{\text{big}} \geq \underbrace{\sigma_{b+1} \geq \cdots \geq \sigma_r}_{\text{small}} > 0$$

get rid of small

chopped off

$$A \approx \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \left[\begin{array}{c|c} \sigma_1 \cdots \sigma_b & 0 \\ \hline 0 & 0 \end{array}\right] \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}^T$$

only keep up to b

b(m+n+1) entries

$$A \approx \begin{bmatrix} u_1 & \cdots & u_b \end{bmatrix} \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_b \end{bmatrix} \begin{bmatrix} v_1 & \cdots & v_b \end{bmatrix}^T$$

carries less info just big stuff

Image process

each entry encodes a pixel

could use low rank approx to carry less info, compress