Singular value Decomposition (SVD)

Applications :-

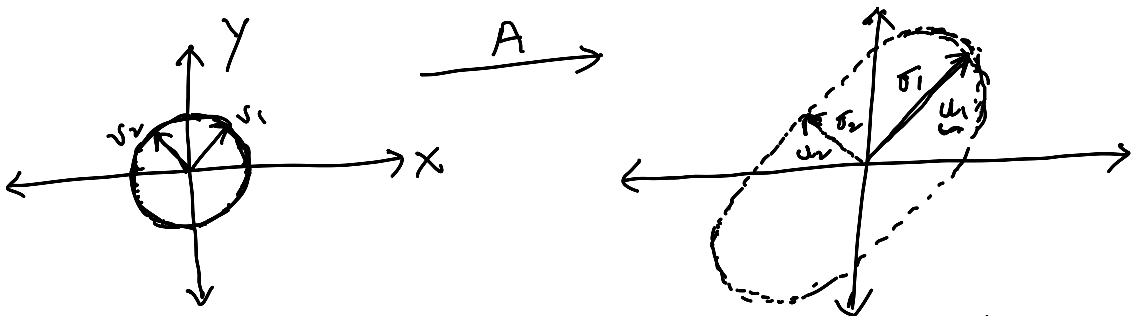
- Searching closest related images
- Image compression (Reducing image size)
- Image recovery
- Principal component analysis (Find most representative/ dominant features)
- Solution of Least squares problems
- Find null-space, column-space
- Computing Matrix norm

Geometric Intution:-

Let us consider for simplicity we are working in $\mathbb{R}^2$.

→ Action of any matrix $\underset{\sim}{A} \in \mathbb{R}^{2\times 2}$ on a unit circle → Gives an ellipse

$\underset{\sim}{A}$ is a non-singular and hence full rank



The image of unit circle under the action of matrix $\underset{\sim}{A}$ is ellipse

$$\underset{\sim}{A}\,\underset{\sim}{S} \to \text{ellipse}$$

$\underset{\sim}{u_1}, \underset{\sim}{u_2}$ are the principal semi-axes of my ellipse with lengths σ_1, σ_2

$\underset{\sim}{v_1}, \underset{\sim}{v_2}$ are pre-image vectors generating images $\underset{\sim}{u_1}, \underset{\sim}{u_2}$ the axes of the ellipse.

$$\underset{\sim}{A}\,\underset{\sim}{v_1} = \sigma_1 \underset{\sim}{u_1}$$

$$\underset{\sim}{A}\,\underset{\sim}{v_2} = \sigma_2 \underset{\sim}{u_2}$$

$$\underset{\sim}{A} \underbrace{\begin{bmatrix} \underset{\sim}{v_1} & \underset{\sim}{v_2} \end{bmatrix}}_{V} = \underbrace{\begin{bmatrix} \underset{\sim}{u_1} & \underset{\sim}{u_2} \end{bmatrix}} \underbrace{\begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}}$$

$$\underset{2\times 2}{\underset{\sim}{A}}\ \underset{2\times 2}{\underset{\sim}{V}} = \underset{2\times 2}{\underset{\sim}{U}}\ \underset{2\times 2}{\underset{\sim}{\Sigma}}$$

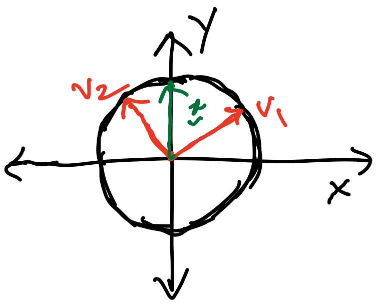
$$\underset{\sim}{A}\,\underset{\sim}{V}\,\underset{\sim}{V}^T = \underset{\sim}{U}\,\underset{\sim}{\Sigma}\,\underset{\sim}{V}^T \implies \underset{\sim}{A} = \underset{\sim}{U}\,\underset{\sim}{\Sigma}\,\underset{\sim}{V}^T$$

Consider action of any matrix on a

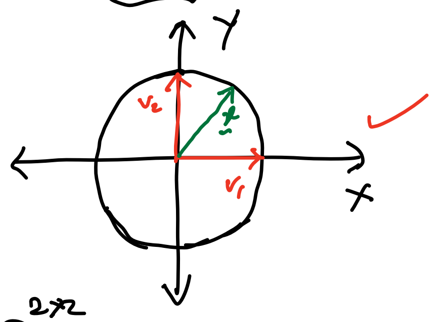
unit circle S -via- sequence of steps

$$\underset{\sim}{A} = \underset{\sim}{U}\,\underset{\sim}{\Sigma}\,\underset{\sim}{V}^T$$

$$\underset{\sim}{A}\,\underset{\sim}{S} = \boxed{\underset{\sim}{U}\,\boxed{\underset{\sim}{\Sigma}\,\boxed{\underset{\sim}{V}^T S}}}$$



Rotation V^T $\longrightarrow$



Action of any matrix $\underset{\sim}{A} \in \mathbb{R}^{2\times 2}$ on a unit circle S

$\downarrow$ Σ Stretching $\begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$ $\sigma_1 > \sigma_2$

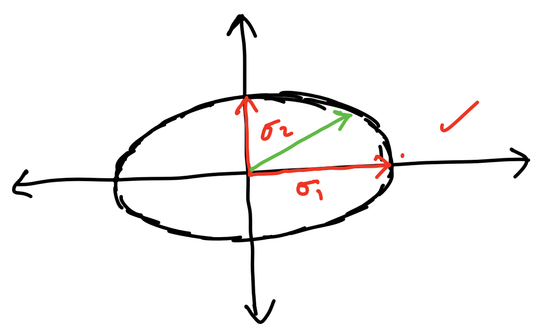
Rotation

$\downarrow$

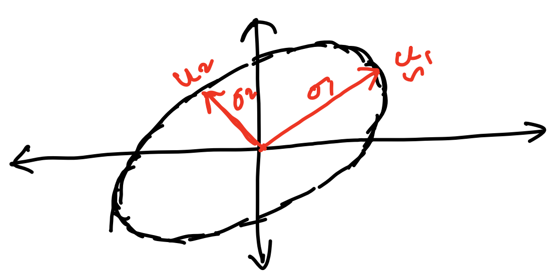
Stretching to Ellipse

$\downarrow$

Rotation of Ellipse



$\downarrow$ Rotation U



This process can be generalized to higher dimension ≥ 2

Let $\underset{\sim}{S}$ be unit hyper sphere in $\mathbb{R}^n$. The action of $\underset{\sim}{A} \in \mathbb{R}^{m \times n}$ $(m \geq n)$ and full rank on $\underset{\sim}{S}$ is a hyperellipse in $\mathbb{R}^m$ with the following properties :-

(i) $\underset{\sim}{A}\,\underset{\sim}{v}_j = \sigma_j \underset{\sim}{u}_j$

$\sigma_1, \sigma_2, \sigma_3 \ldots \sigma_n$ are lengths of the principal semi-axes of $\underset{\sim}{A}\underset{\sim}{S}$ (hyperellipse in $\mathbb{R}^m$) and these σ_j's are called singular values of $\underset{\sim}{A}$. By convention they are ordered in descending order

$$\sigma_1 \geq \sigma_2 \cdots \geq \sigma_n$$

(ii) The set of unit vectors $\{u_1, u_2, \ldots u_n\}$ are directions of principal semi axes of our hyperellipse and these are called left singular vectors and are orthonormal vectors.

$\therefore$ $\sigma_j \underset{\sim}{u}_j$ is the j^{th} largest principal semiaxes of our hyperellipse $\underset{\sim}{A}\underset{\sim}{S}$

(iii) The set of unit vectors $\{\underset{\sim}{v}_1, \underset{\sim}{v}_2, \ldots v_n\}$ which also form an orthonormal set are the pre-images of $\underset{\sim}{u}_i$ and are called right singular vectors $\underset{\sim}{A}$

$$\underset{\sim}{A}\,\underset{\sim}{v}_j = \sigma_j \underset{\sim}{u}_j \quad \text{— (1)}$$

$$[\underset{\sim}{A}]_{m\times n} \underbrace{\begin{bmatrix} | & | & | & & | \\ \underset{\sim}{v}_1 & \underset{\sim}{v}_2 & \underset{\sim}{v}_3 & \cdots & \underset{\sim}{v}_n \\ | & | & | & & | \end{bmatrix}}_{V}{}_{n\times n}$$

$$= \underbrace{\begin{bmatrix} | & | & | & & | \\ \underset{\sim}{u}_1 & \underset{\sim}{u}_2 & \underset{\sim}{u}_3 & \cdots & \underset{\sim}{u}_n \\ | & | & | & & | \end{bmatrix}_{m\times n}}_{\hat{U}} \underbrace{\begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_n \end{bmatrix}_{n\times n}}_{\hat{\Sigma}}$$

$$\underset{\sim}{A}\,\underset{\sim}{V} = \underset{\sim}{\hat{U}}\,\underset{\sim}{\Sigma}$$

$\underset{\sim}{A} \in \mathbb{R}^{m\times n}$ $\quad \underset{\sim}{\hat{U}} \in \mathbb{R}^{m\times n}$

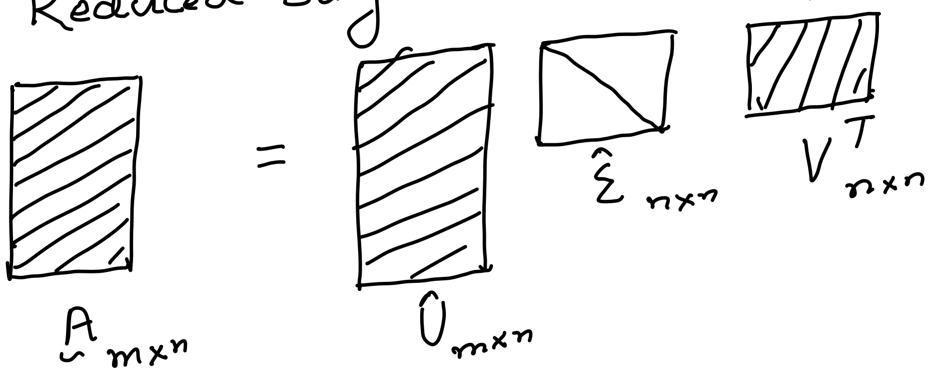
$\underset{\sim}{V} \in \mathbb{R}^{n\times n}$ $\quad \underset{\sim}{\hat{\Sigma}} \in \mathbb{R}^{n\times n}$

Since $\underset{\sim}{V}$ is orthogonal matrix

$$A V V^T = \hat{U} \hat{\Sigma} V^T$$

$$\boxed{A = \hat{U} \hat{\Sigma} V^T}$$

Reduced Singular Value Decompositi

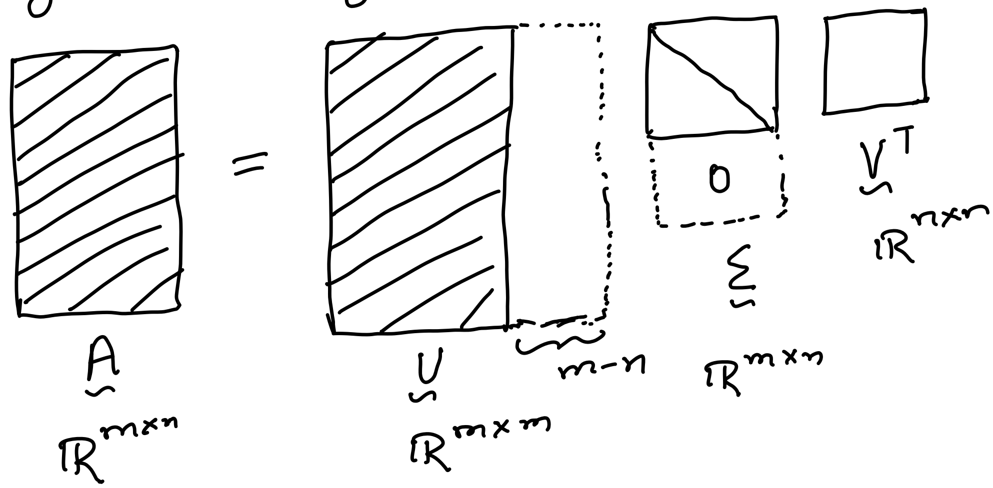


Full Singular value decomposition :-

$\hat{U}$ is made of n orthonormal vectors in $\mathbb{R}^m$ and hence this cannot be a basis for $\mathbb{R}^m$. Now to make $\hat{U}$ to be orthogonal, we add additional $m-n$ orthonormal columns to $\hat{U}$ and call this new matrix $U \in \mathbb{R}^{m\times m}$

Now we write SVD of $A \in \mathbb{R}^{m\times n}$ in terms of orthogonal matrix $U \in \mathbb{R}^{m\times m}$

but for factorization not to change
we must change Σ to $m \times n$ matrix
by introducing m additional rows of 0's



$$\boxed{A = U \Sigma V^T}$$ is called full SVD

$A \in \mathbb{R}^{m \times n}$

$U \in \mathbb{R}^{m \times m}$

$\Sigma \in \mathbb{R}^{m \times n}$

$V \in \mathbb{R}^{n \times n}$

1 We no longer need A to be full rank.

If $\text{rank}(A) = r < n$

Only r singular vectors will be determined by geometry of hyperellipse and we will add $m-r$ orthonormal columns to build U and $n-r$ arbitrary columns to extend r columns determined by the geometry

The matrix Σ will now have r positive diagonal entries with remaining $n-r$ values to be equal to zero.

Infact reduced SVD of $\underset{\sim}{A}$ less than full rank and this is given as

$$\underset{\sim}{A}_{m\times n} = \underset{\sim}{\hat{U}}_{m\times r}\,\underset{\sim}{\hat{\Sigma}}_{r\times r}\,\underset{\sim}{V}^{T}_{r\times n}$$

<u>Definition of SVD formally can be stated as follows</u>:-

The singular value decomposition of $\underset{\sim}{A} \in \mathbb{R}^{m\times n}$ is a factorization of the form $\underset{\sim}{A} = \underset{\sim}{U}\,\underset{\sim}{\Sigma}\,\underset{\sim}{V}^{T}$ where $\underset{\sim}{U} \in \mathbb{R}^{m\times m}$, $\underset{\sim}{V} \in \mathbb{R}^{n\times n}$ are orthogonal matrices and $\underset{\sim}{\Sigma}$ has $n\times n$ diagonal matrix block (Σ_1) and is of dimension $m\times n$ matrix.

(1) There is no assumption that $m \geq n$ or $\underset{\sim}{A}$ has to be full rank.

(2) All diagonal elements of $\underset{\sim}{\Sigma}^{d}$ are non-negative and increasing order

$\sigma_1 \geq \sigma_2 \geq \sigma_3 \dots \geq \sigma_p \geq 0$

where $p = \min(m, n)$

(3) Σ has the same shape as $\underset{\sim}{A}$ but $\underset{\sim}{U}, \underset{\sim}{V}$ are square matrices

(4) Every matrix $\underset{\sim}{A} \in \mathbb{R}^{m \times n}$ has a full SVD, $\underset{\sim}{A} = \underset{\sim}{U} \underset{\sim}{\Sigma} \underset{\sim}{V}^T$. The singular values $\{\sigma_j\}$ are uniquely determined for a give matrix $\underset{\sim}{A}$.

(5) If $\underset{\sim}{A}$ is a square matrix and if $(\sigma_i \neq \sigma_j)$ (for all i and j), then the left singular vectors $\{\underset{\sim}{u}_j\}$ and the right singular vectors $\{\underset{\sim}{v}_j\}$ are uniquely

determined to within a factor of ± 1.

Change of basis - via - SVD

$$\underset{\sim}{A} \in \mathbb{R}^{m\times n}, \quad \underset{\sim}{b} \in \mathbb{R}^{m}, \quad \underset{\sim}{x} \in \mathbb{R}^{n}$$

$$\underset{\sim}{A}\,\underset{\sim}{x} = \underset{\sim}{b}$$

Let $\underset{\sim}{b} \in \mathbb{R}^m$ be expanded using the left singular vectors of $\underset{\sim}{A}$ (columns of $\underset{\sim}{U}$)

and $\underset{\sim}{x} \in \mathbb{R}^n$ be expanded in basis of right singular vectors $\underset{\sim}{V}$

$$\underset{\sim}{b} = \sum_{i=1}^{m} b_i' \underset{\sim}{u}_i \Rightarrow \underset{\sim}{b} = \underset{\sim}{U}\,\underset{\sim}{b}' \qquad \underset{\sim}{b}' = \begin{bmatrix} b_1' \\ b_2' \\ \vdots \\ b_m' \end{bmatrix}$$

$$\underset{\sim}{x} = \sum_{i=1}^{n} x_i' \underset{\sim}{v}_i$$

$$\Rightarrow \underset{\sim}{x} = \underset{\sim}{V}\,\underset{\sim}{x}' \qquad \underset{\sim}{x}' = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{bmatrix}$$

$$\underset{\sim}{A}\,\underset{\sim}{x} = \underset{\sim}{b}$$

$$\underset{\sim}{U}\,\underset{\sim}{\Sigma}\,\underbrace{\underset{\sim}{V}^T\underset{\sim}{V}}_{I}\,\underset{\sim}{x}' = \underset{\sim}{U}\,\underset{\sim}{b}'$$

$$\Rightarrow \quad \underset{\sim}{U}\,\underset{\sim}{\Sigma}\,\underset{\sim}{x}' = \underset{\sim}{U}\,\underset{\sim}{b}'$$

$$\Rightarrow \quad \underset{\sim}{U}^T \underset{\sim}{U} \underset{\sim}{\Sigma} \underset{\sim}{x}' = \underset{\sim}{U}^T \underset{\sim}{U} \underset{\sim}{b}'$$

$$\Rightarrow \quad \underset{\sim}{\Sigma} \underset{\sim}{x}' = \underset{\sim}{b}'$$

Wheneve we have $\underset{\sim}{b} = \underset{\sim}{A} \underset{\sim}{x}$.
we have $\underset{\sim}{b}' = \underset{\sim}{\Sigma} \underset{\sim}{x}'$ i.e $\underset{\sim}{A}$ reduces to a diagonal matrix Σ when range is expressed in basis of columns of $\underset{\sim}{U}$ and domain is expressed in basis of columns of $\underset{\sim}{V}$

Matrix properties via SVD

Let $\underset{\sim}{A} \in \mathbb{R}^{m \times n}$ and let p be the minimum of m and n where $r \leq p$ denote the number of non-zero singular values of $\underset{\sim}{A}$

Thm 1 :- The number of non zero singular values is the rank of the matrix r.

Pf :- $\underset{\sim}{A} = \underset{\sim}{U} \underset{\sim}{\Sigma} \underset{\sim}{V}^T$ $\quad$ $\underset{\sim}{U} \in \mathbb{R}^{m \times m}$, $\underset{\sim}{\Sigma} \in \mathbb{R}^{m \times n}$, $\underset{\sim}{V} \in \mathbb{R}^{n \times n}$

$\text{rank}(\underset{\sim}{A}) = \text{rank}(\underset{\sim}{U} \underset{\sim}{\Sigma} \underset{\sim}{V}^T)$

Recall from Assgn 2, if a matrix Q is non-singular then $\text{rank}(PQ) = \text{rank}(P)$

$P \in \mathbb{R}^{m\times n}$

$Q \in \mathbb{R}^{n\times n}$

$$\text{rank}(\underbrace{\underset{\sim}{U}\underset{\sim}{\Sigma}}_{P}\underbrace{\underset{\sim}{V}^T}_{Q}) = \text{rank}(\underset{\sim}{U}\underset{\sim}{\Sigma}) \quad -(1)$$

Again if you have two matrices $\underset{\sim}{X} \in \mathbb{R}^{m\times n}$; $\underset{\sim}{Y} \in \mathbb{R}^{n\times p}$, $\text{rank}(\underset{\sim}{X}) + \text{rank}(\underset{\sim}{Y}) - n$

$$\leq \text{rank}(XY)$$

$$\leq \min(\text{rank}(X), \text{rank}(Y))$$

$\underset{\sim}{X} \to \underset{\sim}{U}$ and $\underset{\sim}{Y} \to \underset{\sim}{\Sigma}$

$\underset{\sim}{U} \in \mathbb{R}^{m\times m}$ and $\underset{\sim}{\Sigma} \in \mathbb{R}^{m\times n}$

$$\text{rank}(\underset{\sim}{U}) + \text{rank}(\Sigma) - m \leq \text{rank}(U\Sigma) \leq \min(\text{rank}(U), \text{rank}(\Sigma))$$

$$m + r - m \leq \text{rank}(U\Sigma) \leq \min(m, r)$$

$$r \leq \text{rank}(U\Sigma) \leq r$$

$$\Rightarrow \text{rank}(U\Sigma) = r$$

$$\text{rank}(\underset{\sim}{A}) = \text{rank}(\underset{\sim}{U}\underset{\sim}{\Sigma}\underset{\sim}{V}^T) = r$$

<u>Thm</u> 2:- range$(\underset{\sim}{A})$ i.e column space of $\underset{\sim}{A}$ is same as the space spanned by $\langle \underset{\sim}{u}_1, \underset{\sim}{u}_2, \dots \underset{\sim}{u}_r \rangle$

and null$(\underset{\sim}{A})$ is the space spanned by $\langle \underset{\sim}{v}_{r+1}, \underset{\sim}{v}_{r+2}, \dots \underset{\sim}{v}_n \rangle$

Pf:- $\underset{\sim}{A} = \underset{\sim}{U} \underset{\sim}{\Sigma} \underset{\sim}{V}^T$

$\underset{\sim}{A}\underset{\sim}{V} = \underset{\sim}{U}\underset{\sim}{\Sigma}$ (Multiply by $\underset{\sim}{V}$ on both sides)

$$\underset{\sim}{A} \begin{bmatrix} | & | & | & & | & | & & | \\ \underset{\sim}{v}_1 & \underset{\sim}{v}_2 & \underset{\sim}{v}_3 & \cdots & \underset{\sim}{v}_r & v_{r+1} & \cdots & v_n \\ | & | & | & & | & | & & | \end{bmatrix}_{n \times n}$$

$$= \begin{bmatrix} | & | & & | & | & & | \\ \underset{\sim}{u}_1 & \underset{\sim}{u}_2 & \cdots & \underset{\sim}{u}_r & u_{r+1} & \cdots & \underset{\sim}{u}_m \\ | & | & & | & | & & | \end{bmatrix}_{m \times m}$$

$$\begin{bmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \sigma_3 & & & & \\ & & & \ddots & & & \\ & & & & \sigma_r & & \\ & & & & & 0 & \\ & & & & & & 0 \\ - & - & - & & - & - & - \\ & & & 0 & & & \end{bmatrix}_{m \times n}$$

$\underset{\sim}{A}\underset{\sim}{v}_j = \sigma_j \underset{\sim}{u}_j$ for $j = 1$ to r

$\frac{1}{\sigma_j} \underset{\sim}{A}\underset{\sim}{v}_j = \underset{\sim}{u}_j$

clearly $\underset{\sim}{u}_j$ lies in the column-space of $\underset{\sim}{A}$ for $j = 1 .. r$

show $\langle \underset{\sim}{u}_1, .. u_r \rangle$ spans the column-space of $\underset{\sim}{A}$?

$$A v_{r+1} = 0$$
$$A v_{r+2} = 0$$
$$\vdots$$
$$A v_n = 0$$

$\langle v_{r+1}, v_{r+2}, \dots v_n \rangle$ are in the null-space of A and form a basis for $n-r$ dimensional space.

ie null(A) is spanned by $\langle v_{r+1}, v_{r+2}, \dots v_n \rangle$

Thm 3 :-

$$\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2 = \sigma_1$$

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2 \right)^{1/2}$$

$$= \sqrt{\sigma_1^2 + \sigma_2^2 + \dots \sigma_r^2}$$

Pf :-

$$\|A\|_2 = \|U \Sigma V^T\|_2 = \|\Sigma V^T\|_2$$
$$= \|\Sigma\|_2$$
$$= \max \{|\sigma_j|\}$$
$$= \sigma_1$$

$$\|A\|_F = \|U \Sigma V^T\|_F = \|\Sigma\|_F$$
$$= \sqrt{\sigma_1^2 + \sigma_2^2 + \dots \sigma_r^2}$$

SVD and eigenvalue decomposition :-

Eigendecomposition :-

If $A \in \mathbb{R}^{m \times m}$ is non-defective, then it has complete set of m linearly independent eigenvectors.

$$\left.\begin{aligned} A x_1 &= \lambda_1 x_1 \\ A x_2 &= \lambda_2 x_2 \\ &\vdots \\ A x_m &= \lambda_m x_m \end{aligned}\right\} \quad -(1)$$

$$A \begin{bmatrix} | & | & | & & | \\ x_1 & x_2 & x_3 & \cdots & x_m \\ | & | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_m \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_m \end{bmatrix}$$

$$A X = X \Lambda \quad -(2)$$

$$\Rightarrow \quad A = X \Lambda X^{-1} \quad \Rightarrow \quad \Lambda = X^{-1} A X \quad -(3)$$

For $A \in \mathbb{R}^{m \times m}$, $Ax = b$ and we expand $b, x \in \mathbb{R}^m$ in the basis of eigenvectors $b = \sum_{i=1}^{m} b_i' x_i$

$$b = X b', \quad b' = \begin{bmatrix} b_1' \\ b_2' \\ \vdots \\ b_m' \end{bmatrix}$$

and $x = \sum_{i=1}^{m} x_i' x_i$

$$\Rightarrow x = X x'$$

where $\underset{\sim}{x}' = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_m' \end{bmatrix}$

$$\underset{\sim}{A}\,\underset{\sim}{x} = \underset{\sim}{b}$$

$$\underset{\sim}{A}\,\underset{\sim}{X}\,\underset{\sim}{x}' = \underset{\sim}{X}\,\underset{\sim}{b}'$$

$$\Rightarrow \underbrace{\underset{\sim}{X}^{-1}\underset{\sim}{A}\,\underset{\sim}{X}}\,\underset{\sim}{x}' = \underset{\sim}{b}'$$

$$\underset{\sim}{\Lambda}\,\underset{\sim}{x}' = \underset{\sim}{b}'$$

Comparison of SVD and eigen decomp:-

* SVD uses two different bases (left singular vectors basis and right singular vectors basis) whereas eigendecomposition uses just one basis ie basis of eigenvectors

* SVD results in orthonormal bases whereas $\underset{\sim}{X}$ in eigendecomposition is generally not orthogonal.

* SVD exists for all matrices no matter dimension; eigendecomposition

exist for square non-defective matrices.

Thm 4:- Non zero singular values of $\underset{\sim}{A} \in \mathbb{R}^{m \times n}$ are square roots of non-zero eigenvalues of $\underset{\sim}{A}^T \underset{\sim}{A}$ or $\underset{\sim}{A}\underset{\sim}{A}^T$.

$$\underset{\sim}{A} = \underset{m\times m}{\underset{\sim}{U}} \underset{m\times n}{\underset{\sim}{\Sigma}} \underset{n \times n}{\underset{\sim}{V}^T}$$

Pf:-
$$\underset{\sim}{A}^T \underset{\sim}{A} = (\underset{\sim}{U}\underset{\sim}{\Sigma}\underset{\sim}{V}^T)^T(\underset{\sim}{U}\underset{\sim}{\Sigma}\underset{\sim}{V}^T)$$
$$= (\underset{\sim}{V}^T)^T \underset{\sim}{\Sigma}^T \underbrace{\underset{\sim}{U}^T\underset{\sim}{U}} \underset{\sim}{\Sigma}\underset{\sim}{V}^T$$
$$= \underset{n\times n}{\underset{\sim}{V}} \underset{n\times m}{\underset{\sim}{\Sigma}^T} \underset{m\times n}{\underset{\sim}{\Sigma}} \underset{n\times n}{\underset{\sim}{V}^T} = \underset{\sim}{V} \tilde{\underset{\sim}{\Sigma}} \underset{\sim}{V}^T$$

where $\tilde{\Sigma}$ is diag matrix of squares of singular values

$\underset{\sim}{V}\tilde{\underset{\sim}{\Sigma}}\underset{\sim}{V}^T$ is eigendecomp of $\underset{\sim}{A}^T\underset{\sim}{A}$ comprising $\sigma_1^2, \sigma_2^2, \ldots \sigma_p^2$ as diagonal entries of $\tilde{\Sigma}$ with $n-p$ additional zero eigenvalues if $n > p$

Thm 5:- If $\underset{\sim}{A} = \underset{\sim}{A}^T$ singular values of $\underset{\sim}{A}$ are absolute values of eigenvalues of $\underset{\sim}{A}$.

Pf:-
$$\underset{\sim}{A} = \underset{\sim}{X} \underset{\sim}{\Lambda} \underset{\sim}{X}^{-1}$$
$$= \underset{\sim}{Q} \underset{\sim}{\Lambda} \underset{\sim}{Q}^{-1}$$
(where $\underset{\sim}{Q}$ is an orthogonal matrix for matrices which are symmetric)
$$= \underset{\sim}{Q} \underset{\sim}{\Lambda} \underset{\sim}{Q}^{T}$$
$$= \underset{\sim}{Q} |\underset{\sim}{\Lambda}| \, \underline{\mathrm{sgn}(\Lambda)} \, \underset{\sim}{Q}^{T}$$

where $|\Lambda|$ and $\mathrm{sgn}(\Lambda)$ are diagonal matrices whose entries are $|\lambda_j|$ and $\mathrm{sgn}(\lambda_j)$.

$$\underset{\sim}{A} = \underset{\sim}{Q} |\Lambda| \underbrace{\mathrm{sgn}(\Lambda) \underset{\sim}{Q}^{T}}$$
$$= \underset{\sim}{Q} |\Lambda| \bar{Q}^{T}$$
$\underset{\sim}{A} = \underset{\sim}{Q} |\Lambda| \bar{Q}^{T}$ is SVD of A

Thm 6:- $\underset{\sim}{A} \in \mathbb{R}^{m \times m}$; $|\det(A)| = \prod_{i=1}^{m} \sigma_i$

$$|\det(A)| = |\det(\underset{\sim}{U} \underset{\sim}{\Sigma} \underset{\sim}{V}^{T})|$$
$$= |\det U| \, |\det \Sigma| \, |\det V^{T}|$$
$$= |\det \Sigma| = \prod_{i=1}^{m} \sigma_i$$
$[|\det V| = |\det U| = 1]$

Low rank approximations:-

Thm 7:- $A \in \mathbb{R}^{m \times n}$ of rank "r" can be written as sum of "r" rank-one matrices of the form $A = \sum_{j=1}^{r} \sigma_j u_j v_j^T$

where $\{\sigma_j\}$ are singular values and $\{u_j\}, \{v_j\}$ are the appropriate singular vectors.

Pf:- Recall $u v^T$ is a rank-one matrix

$$A = U \Sigma V^T$$

$$\Sigma = \text{Sum}_{j=1}^{r} \Sigma_j \quad \text{where}$$

$$\Sigma_j = \begin{bmatrix} 0 & & & & \\ & 0 & & & \\ & & 0 \cdots & & \\ & & & \sigma_j & \\ & & & & \cdots 0 \\ \hline & & 0 & & \end{bmatrix}_{m \times n}$$

$$A = U \Sigma V^T$$

$$= U \left\{ \text{Sum}_{j=1}^{r} \Sigma_j \right\} V^T$$

$$= \text{Sum}_{j=1}^{r} U \Sigma_j V^T = \sum_{j=1}^{r} \sigma_j u_j v_j^T$$

$$\underline{A = \sum_{j=1}^{r} \sigma_j u_j v_j^T} \qquad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$$

Then k^{th} partial sum $\sum_{j=1}^{k} \sigma_j u_j v_j^T$ $\quad k \leq r$

has as much energy (information) of A as possible

Thm 8:- For any k with $1 \leq k \leq r$

define $\boxed{A_k = \sum_{j=1}^{k} \sigma_j u_j v_j^T}$

then $\|A - A_k\|_2 = \min_{\substack{B \in \mathbb{R}^{m \times n} \\ rank(B) \leq k}} \|A - B\|_2 = \sigma_{k+1}$

Eckhart - Young Theorem

Proof:- Let there is some (B) whose $rank(B) \leq k$ such that $\|A - B\|_2 < \|A - A_k\|_2 = \sigma_{k+1}$

$\dim(N(B)) \geq n-k$

as $rank(B) \leq k$

Consider the subspaces

(i) W_1 : The null space of (B) which is of dimension of atleast $n-k$

(ii) W_2 : The space spanned by $k+1$ right singular vectors of A i.e $v_1, v_2, v_3 \cdots v_{k+1}$

These two subspaces have to intersect? why?

Dimensions of the two subspace add to $(n-k)+(k+1)$ i.e the subspaces must atleast have 1 common vector.

Let such a non-zero vector be $\underset{\sim}{x}$ i.e $\underset{\sim}{x} \in W_1 \cap W_2$

$\underset{\sim}{x} \neq 0$; $\underset{\sim}{x} \in N(B)$ i.e $\underset{\sim}{B}\underset{\sim}{x} = \underset{\sim}{0}$ and $\underset{\sim}{x} \in W_2$

$$\underset{\sim}{x} = \sum_{i=1}^{k+1} c_i \underset{\sim}{v_i}$$

$$\|\underset{\sim}{A}\underset{\sim}{x}\|_2 = \|(\underset{\sim}{A}-\underset{\sim}{B})\underset{\sim}{x}\|_2$$
$$\leq \|A-B\|_2 \|\underset{\sim}{x}\|_2 < \sigma_{k+1} \|\underset{\sim}{x}\|_2 \quad — (1)$$

$\underset{\sim}{A}\underset{\sim}{v_i} = \underbrace{\sigma_i \underset{\sim}{u_i}}$

$$\|\underset{\sim}{A}\underset{\sim}{x}\|_2^2 = \|\underset{\sim}{A} \sum_{i=1}^{k+1} c_i \underset{\sim}{v_i}\|_2^2$$

$\|\underset{\sim}{a}\|_2^2 = \underset{\sim}{a}^T\underset{\sim}{a}$

$$= \|\sum_{i=1}^{k+1} c_i \sigma_i \underset{\sim}{u_i}\|_2^2 = \left[\sum_{i=1}^{k+1} c_i \sigma_i \underset{\sim}{u_i}\right]^T \left[\sum_{i=1}^{k+1} c_i \sigma_i \underset{\sim}{u_i}\right]$$

$$= \sum_{i=1}^{k+1} c_i^2 \sigma_i^2 \quad \text{(use orthonormality of } \underset{\sim}{u_i}\text{)}$$

$$\underbrace{\sum_{i=1}^{k+1} c_i^2 \sigma_i^2} \geq \underbrace{\sum_{i=1}^{k+1} c_i^2 \sigma_{k+1}^2}$$

$$= \left(\sum_{i=1}^{k+1} c_i^2\right) \sigma_{k+1}^2$$

$$= \|\underset{\sim}{x}\|^2 \sigma_{k+1}^2$$

$\|\underset{\sim}{A}\underset{\sim}{x}\|_2^2 \geq \|\underset{\sim}{x}\|^2 \sigma_{k+1}^2$ i.e $\|\underset{\sim}{A}\underset{\sim}{x}\|_2 \geq \sigma_{k+1} \|x\|_2$ — (2)

① & ② is a contradiction which means you cannot have a matrix B with $\text{rank}(B) \leq k$ such that $\|A-B\|_2 < \|A-A_k\|_2$

Eckhart - Young in Frobenius norm :-

$\boxed{\text{Thm 9}}$: For any k, with $1 \leq k \leq r$, the matrix $A_k = \sum_{j=1}^{k} \sigma_j u_j v_j^T$ also satisfies

$$\|A - A_k\|_F = \min_{\substack{B \in \mathbb{R}^{m \times n} \\ \text{rank}(B) \leq k}} \|A - B\|_F$$

$$= \sqrt{\sigma_{k+1}^2 + \sigma_{k+2}^2 + \dots \sigma_r^2}$$

Principal components :-

u_i, v_i are called principal components!

$A_0 \in \mathbb{R}^{m \times n}$ — n features

eg: Age, height, weight, marks etc

$$m \text{ samples} \left\{ \begin{bmatrix} & & & \\ & & & \\ & & & \end{bmatrix}_{m \times n} \right.$$

$\uparrow \uparrow \uparrow \dots \uparrow$ n features

m samples can be number of students in class

→ Mean is the average of the data (in each column). Subtract these means of each of these columns of $\underset{\sim}{A_0}$ and reconstruct the data matrix which produces centered matrix $\underset{\sim}{A}$.

→ Variance as sum of squares of distance from the mean — along each column of $\underset{\sim}{A}$

$$V_i \propto \|\underset{\sim}{a_i}\|_2^2 \left(\frac{1}{m-1}\right)$$

→ Total variance in the full data is the sum of variances of individual columns.

$$\underset{\sim}{A} = \begin{bmatrix} | & | & & | \\ \underset{\sim}{a_1} & \underset{\sim}{a_2} & \cdots & \underset{\sim}{a_n} \\ | & | & & | \end{bmatrix}$$

$$T \propto \underbrace{\left(\|a_1\|_2^2 + \|a_2\|_2^2 + \dots \|a_n\|^2\right]}$$

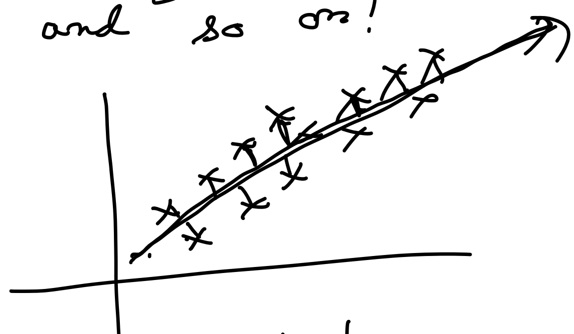
$$T \propto \|\underset{\sim}{A}\|_F^2$$

$$\propto \left(\underline{\sigma_1}^2 + \underline{\sigma_2}^2 + \dots \underline{\sigma_r}^2\right)$$

σ_1^2 accounts for maximum contribution to the total variance, σ_2^2 accounts for next largest contribution to total variance and so on!

The first component $\underset{\sim}{u_1}$ (left singular vector) is along the direction of maximum variance, $\underset{\sim}{u_2}$ is along the next largest variance and so on!

Geometrically first principal component corresponds to a line that minimizes the sum of squares of orthogonal distances of the points from the line!



The key point is that $k < n$ singular vectors explain most of the data than any other set of k vectors. So we can choose $\underset{\sim}{u}_1, \underset{\sim}{u}_2, \ldots \underset{\sim}{u}_k$ as a basis for k-dimension subspace closest to n-dimensional subspace corresponding to our m-data points.