Singular value Decomposition (SVD)

Applications :-

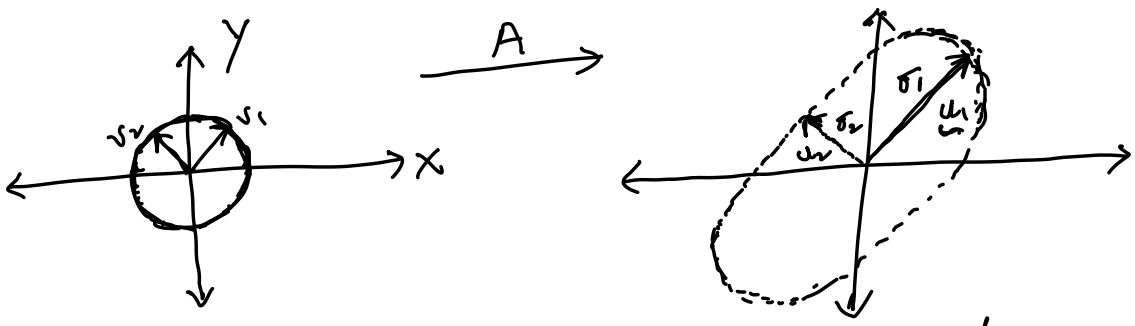
- Searching closest related images
- Image compression (Reducing image size)
- Image recovery
- Principal component analysis (Find most representative/ dominant features)
- Solution of Least squares problems
- Find null-space, column-space
- Computing Matrix norm

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

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

$\underset{\sim}{u_1}, \underset{\sim}{u_2}$ 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.

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

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

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

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

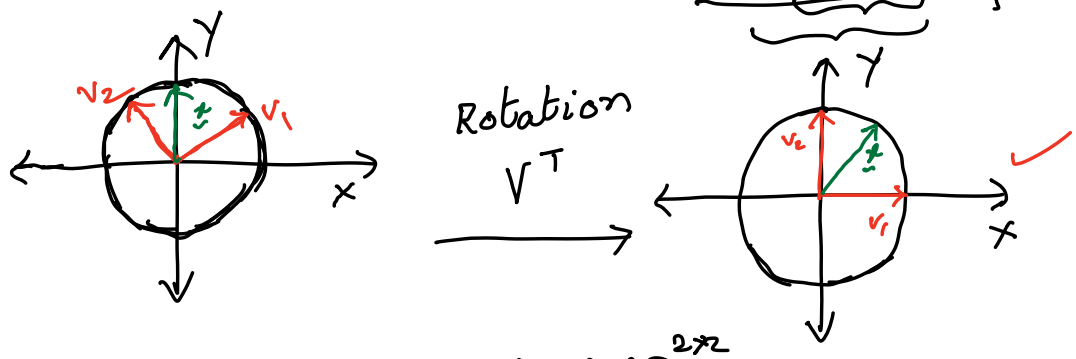
$$\underset{\sim}{A}\,\underset{\sim}{V}\,\underset{\sim}{V}^T = \underset{\sim}{U}\,\underset{\sim}{\Sigma}\,\underset{\sim}{V}^T \quad \Rightarrow \underset{\sim}{A} = \underset{\sim}{U}\,\underset{\sim}{\Sigma}\,\underset{\sim}{V}^T$$

Consider action of any matrix on a

unit circle S -via- sequence of steps

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

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



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

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

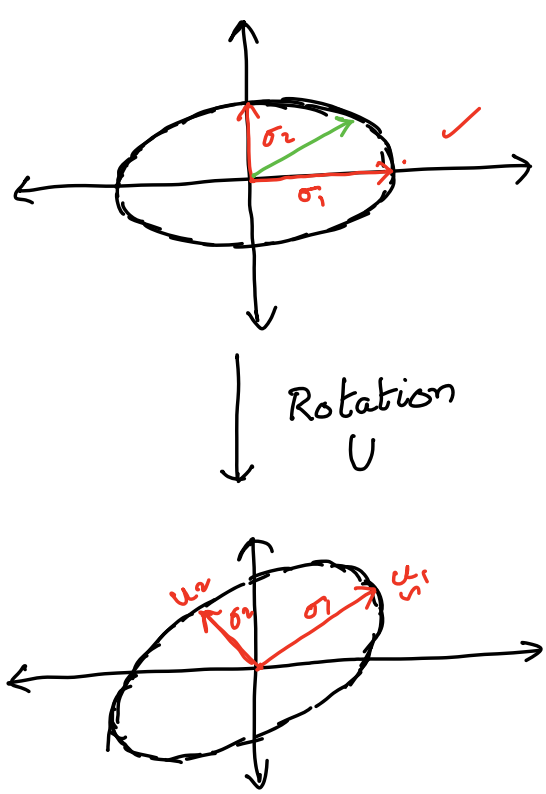
Rotation

↓

Stretching to Ellipse

↓

Rotation of Ellipse



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

$$\sigma_1 \geq \sigma_2 \ldots \geq \sigma_n$$

(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 jth largest principal semiaxes of our hyperellipse $\underset{\sim}{A}\xi$

(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 $\underset{\sim}{A}$

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

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

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

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

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

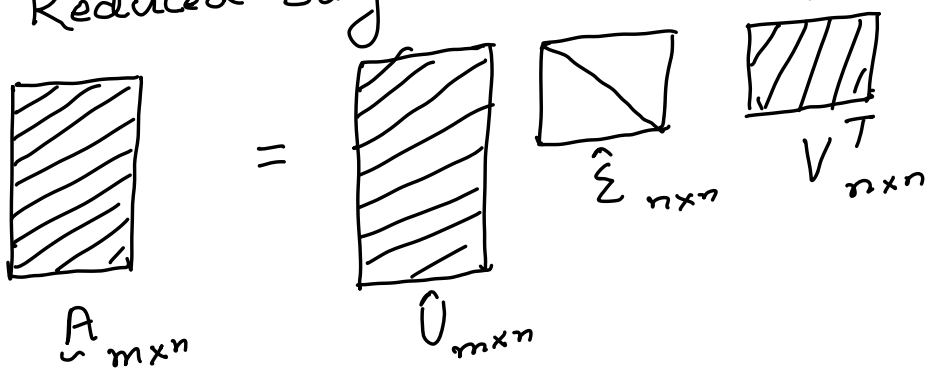
$\underset{\sim}{V} \in \mathbb{R}^{n\times n}$ $\quad \hat{\underset{\sim}{\Sigma}} \in \mathbb{R}^{n\times n}$

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

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

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

Reduced Singular Value Decomposition

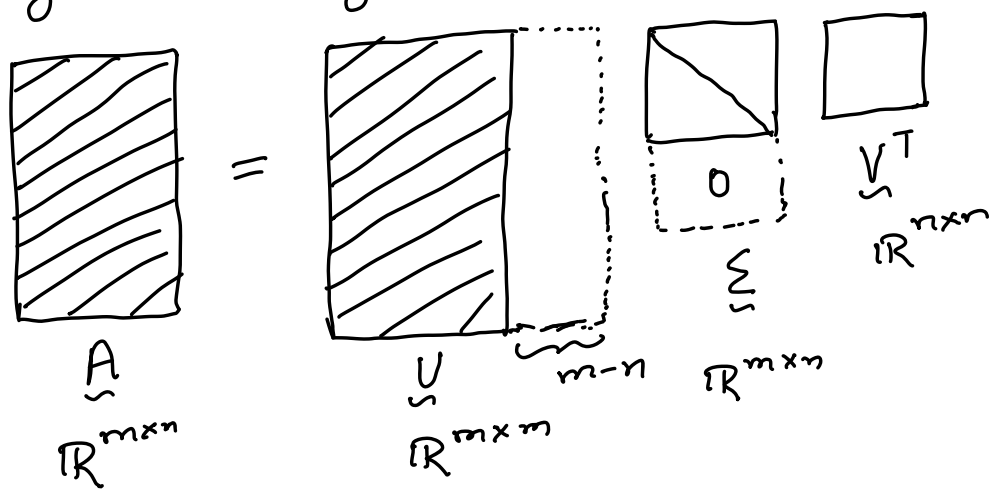


<u>Full Singular value decomposition</u> :-

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

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

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



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

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

(1) We no longer need $\underset{\sim}{A}$ to be full rank.

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

Only r singular vectors will be determined by geometry of hyperellipse and we will add $m-r$ orthonormal columns to build $\underset{\sim}{U}$ 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

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

<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

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

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

(3) $\underset{\sim}{\Sigma}$ 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 $\underset{\sim}{V}$

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

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

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

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

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

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

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

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

whenever 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 $\underset{\sim}{V}$

<u>Matrix properties via SVD</u>

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 $\underset{\sim}{A}$

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

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

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

Recall from Assgn 1, if a matrix Q is non-singular then $\text{rank}(PQ) = \text{rank}(P)$

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

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

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

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 $\langle \underset{\sim}{u}_1, \underset{\sim}{u}_2, \ldots \underset{\sim}{u}_r \rangle$

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)

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

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

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

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

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

$$\underset{\sim}{A}\,\underset{\sim}{v}_{r+1} = \underset{\sim}{0}$$
$$\underset{\sim}{A}\,\underset{\sim}{v}_{r+2} = \underset{\sim}{0}$$
$$\vdots$$
$$\underset{\sim}{A}\,\underset{\sim}{v}_{n} = \underset{\sim}{0}$$

$\langle \underset{\sim}{v}_{r+1}, \underset{\sim}{v}_{r+2}, \ldots \underset{\sim}{v}_n \rangle$ are in the null-space of $\underset{\sim}{A}$ and form a basis for $n-r$ dimensional space.

i.e. null(A) is spanned by $\langle \underset{\sim}{v}_{r+1}, \underset{\sim}{v}_{r+2}, \ldots \underset{\sim}{v}_n \rangle$

<u>Thm 3</u> :-

$$\|\underset{\sim}{A}\|_2 = \max_{\|x\|_2=1} \|\underset{\sim}{A}\,\underset{\sim}{x}\|_2 = \sigma_1$$

$$\|\underset{\sim}{A}\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2\right)^{1/2}$$
$$= \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots \sigma_r^2}$$

Pf :-

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

$$\|\underset{\sim}{A}\|_F = \|\underset{\sim}{U}\,\underset{\sim}{\Sigma}\,\underset{\sim}{V}^T\|_F = \|\underset{\sim}{\Sigma}\|_F$$
$$= \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots \sigma_r^2}$$