LECTURE 4

LAST TIME

- ☐ NORMS AND INDUCED MATRIX NORMS
- ☐ SVD

$A = U_r \Sigma_r V_r^*$ $\nearrow$ REDUCED OR FULL SVD. ; $A \in \mathbb{C}^{m \times n}$

$r = \text{rank}(A)$

$U_r \in \mathbb{C}^{m\times r}$; $V_r \in \mathbb{C}^{n \times r}$; $\Sigma_r = \text{diag}(\sigma_i)$
$\sigma_i > 0$; $i = 1, \dots, r$

TODAY

- ☐ EXISTENCE AND UNIQUENESS
- ☐ LOW RANK APPROXIMATIONS (THM 5.7/ 5.8) Book

SOME ADDITIONAL PROPERTIES OF SVD

$\{u_1, \dots, u_r\}$: ORTHONORMAL BASIS FOR RANGE(A)

$\{v_1, \dots, v_r\}$: ORTHONORMAL BASIS FOR RANGE(A^*)

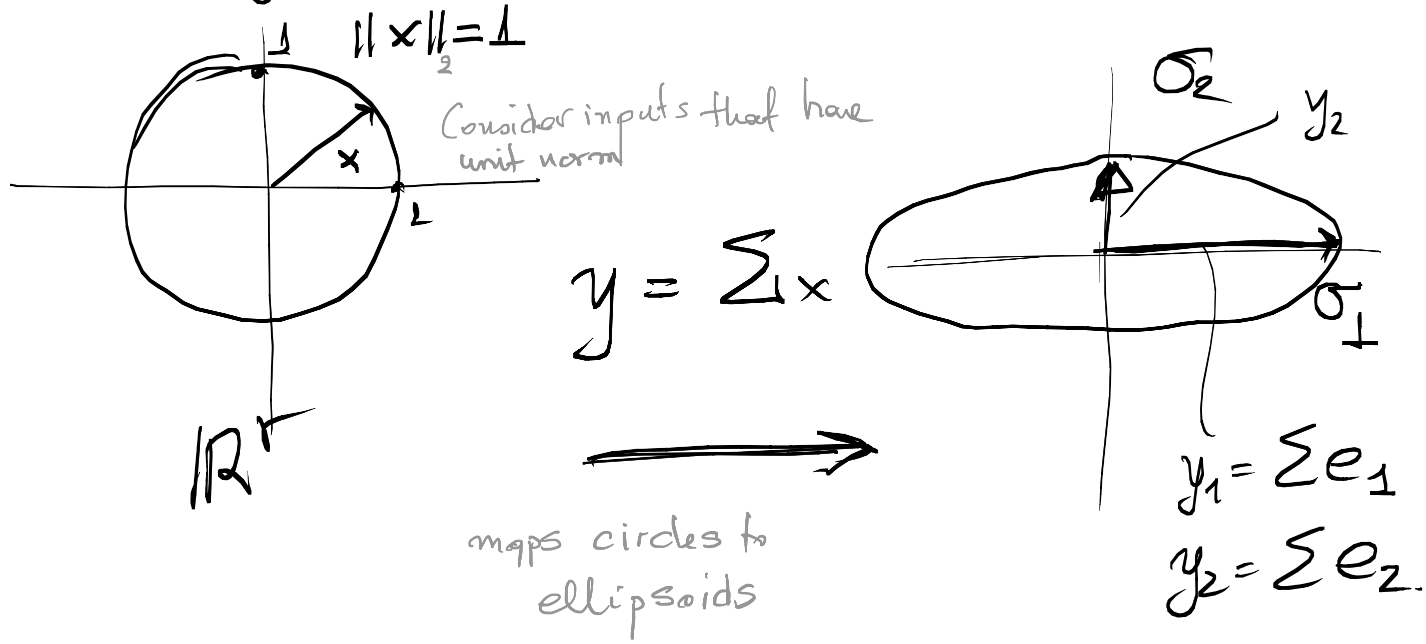
$\{v_{r+1}, \dots, v_n\}$: ORTHONORMAL BASIS FOR NULL(A)

LET $A \in \mathbb{C}^{m\times m}$, A FULL RANK : $A^{-1} = V \Sigma^{-1} U^*$

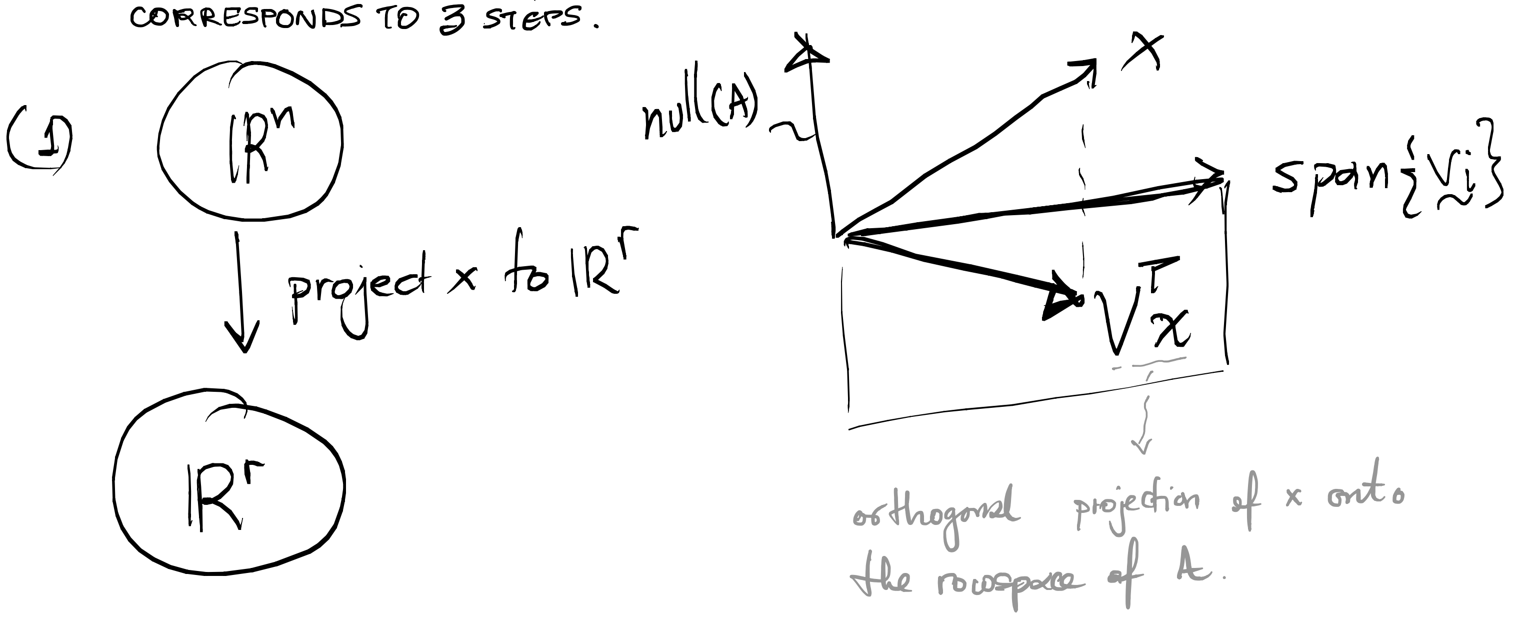
RECALL BASIC RESULT BASED ON SVD:

All matrices $A \in \mathbb{C}^{m \times n}$ are diagonal in the appropriate coordinate systems.

· Diagonal matrix $\Sigma \in \mathbb{R}^{r \times r}$

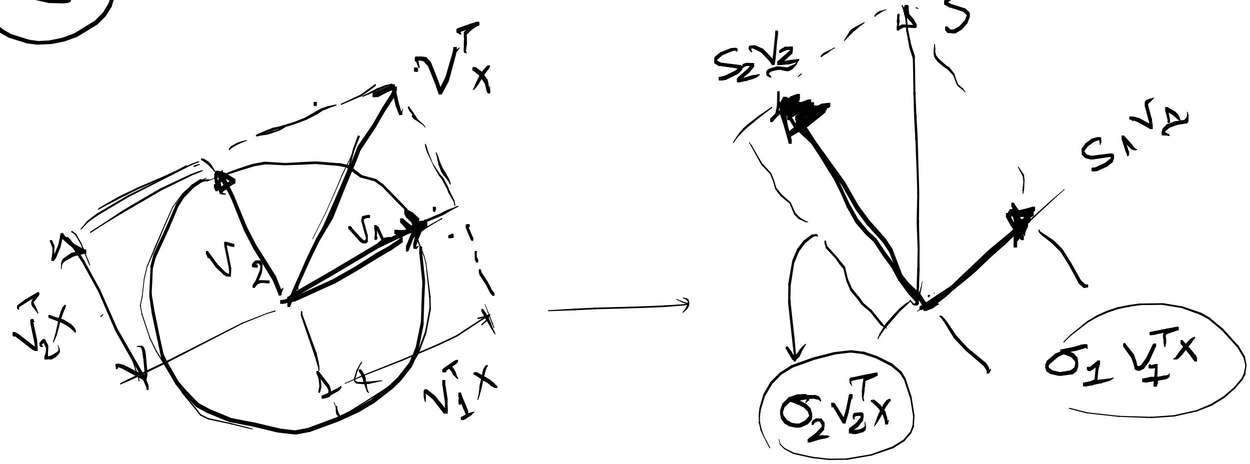


For arbitrary matrix $A \in \mathbb{R}^{m \times n}$, Ax CORRESPONDS TO 3 STEPS.



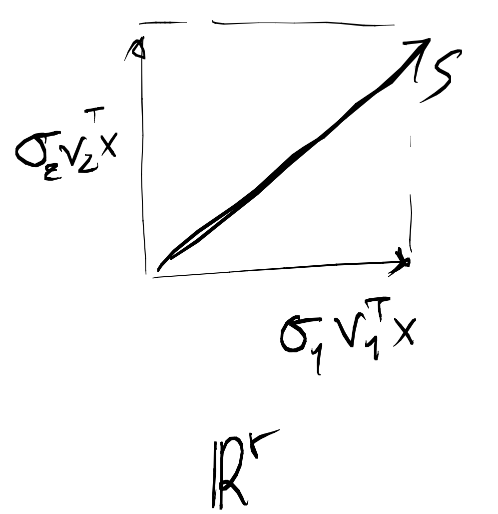
2

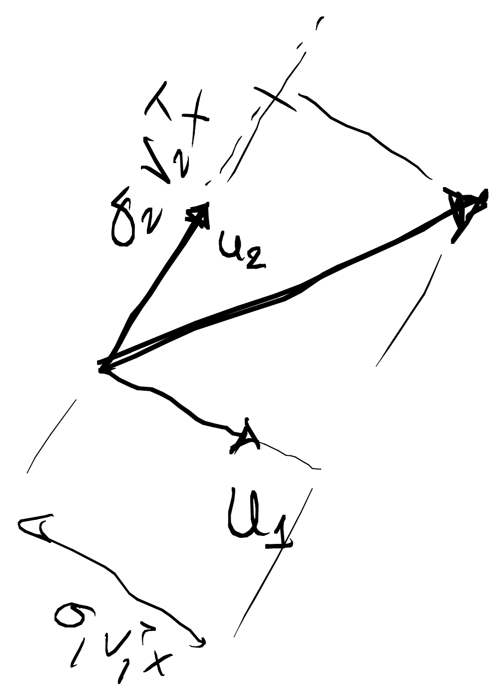
$$\mathbb{R}^r \xrightarrow{s=\Sigma V^T x} \mathbb{R}^r$$



3

$$\mathbb{R}^r \xrightarrow{y=Us} \mathbb{R}^m$$





$\underline{\underline{y=Ax}}$

In summary

A can be viewed as

a diagonal mapping if the input coord system

---

e.g. $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$; $U = V = I$; $\Sigma = \begin{bmatrix} 2 & \\ & 1 \end{bmatrix}$

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

more examples:

```
A = rand (m,n);
[U,S,V] = svd (A,'econ');
norm( A - U*S*V');
```

2 - NORM OF A MATRIX

LET $A \in \mathbb{C}^{m \times n}$

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

HERE WE USE THE FULL SVD SO, U, V ARE UNITARY.

$$\|Ax\|_2^2 = (Ax)^*(Ax) = (U\Sigma V^* x)^*(U\Sigma V^* x)$$

$$= (\Sigma V^* x)^* \underbrace{U^* U}_{\text{UNITARY}} \Sigma V^* x$$

$$= (\Sigma V^* x)^* (\Sigma V^* x) \quad (\text{LET } p = V^* x)$$

$$= (\Sigma p)^*(\Sigma p). \quad \text{LET } q = \Sigma p$$

THEN. $q(i) = \sigma_i \, p(i) \quad ; \; i = 1, \dots, r$

$q(i) = 0 \qquad i = r+1, \dots, m$

THUS: $\|Ax\|_2^2 = \sum_{i=1}^{r} q^2(i) = \sum_{i=1}^{r} \sigma_i^2 p^2(i) \qquad (1)$

But since $p = V^* x$; $\|p\|_2 = 1$

≤ 1 (2)

(1) BECOMES: $\|Ax\|_2^2 \le \sigma_1^2 \sum_{i=1}^{} p^2(i) \le \sigma_1^2$, for all x

CHOOSE $x = v_1$. THEN $V^* x = e_1 \Rightarrow \|A v_1\|_2^2 = \sigma_1^2 \Rightarrow \|A\|_2 = \sigma_1$

- LEMMA:

Let $A \in \mathbb{C}^{m \times n}$

Let $Q \in \mathbb{C}^{m \times m}$, orthonormal

Let $P \in \mathbb{C}^{n \times n}$, orthonormal

Then $\|QAP\|_2 = \|A\|_2$.

Pf:

(1) Show $\|QA\|_2 = \|A\|_2$

$\|QA\|_2 = \|QAx_*\|_2$ for some $x_* \in \mathbb{C}^n$, $\|x_*\| = 1$.

$\|QAx_*\|_2^2 = QAx_* \cdot QAx_* = Ax_* \cdot Q^*QAx_*$

$= Ax_* \cdot Ax_* \leq \|A\|_2^2$. (a)

Let $\|A\|_2 = \|Ay_*\|_2$ for some $y_* \in \mathbb{C}^n$

$\|A\|_2^2 = Ay_* \cdot Ay_* = Ay_* \cdot Q^*QAy_* = QAy_* \cdot QAy_*$

$\leq \|QA\|_2$. (b) $\overset{(a)+(b)}{\Rightarrow}$ $\underline{\|QA\|_2 = \|A\|_2}$

(2) Show $\|AP\|_2 = \|A\|_2$.

$\boxed{a}$ $\|AP\|_2^2 = APx_* \cdot APx_*$, for some $x_* \in \mathbb{C}^n$, $\|x\|_2 = 1$.

$= Ay \cdot Ay$ where $y = Px_*$ and $\|y\|_2 = 1$

$\leq \|A\|_2^2$.

$\boxed{b}$ $\|A\|_2^2 = Ay_* \cdot Ay_* = APP^* y_* \cdot APP^* y_*$, Since $PP^* = I$

$= APx \cdot APx$, where $x = P^* y$ and $\|x\|_2 = 1$

$\leq \|AP\|_2^2$

$a \,\&\, b \Rightarrow \|AP\|_2^2 = \|A\|_2^2$. $\blacksquare$

That means orthonormal, full rank tranformations preserve ^matrix & vector Euclidian norms.

Notice the importance of the 2-norm in this proof. It is the only ✓INDUCED NORM that is invariant to rotations. (FROBENIUS) ASLO IS INVARIANT

THM 4.1 (Existence/Uniqueness of SVD)

(Text Book).

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

[The proof extends to complex matrices with with minor changes]

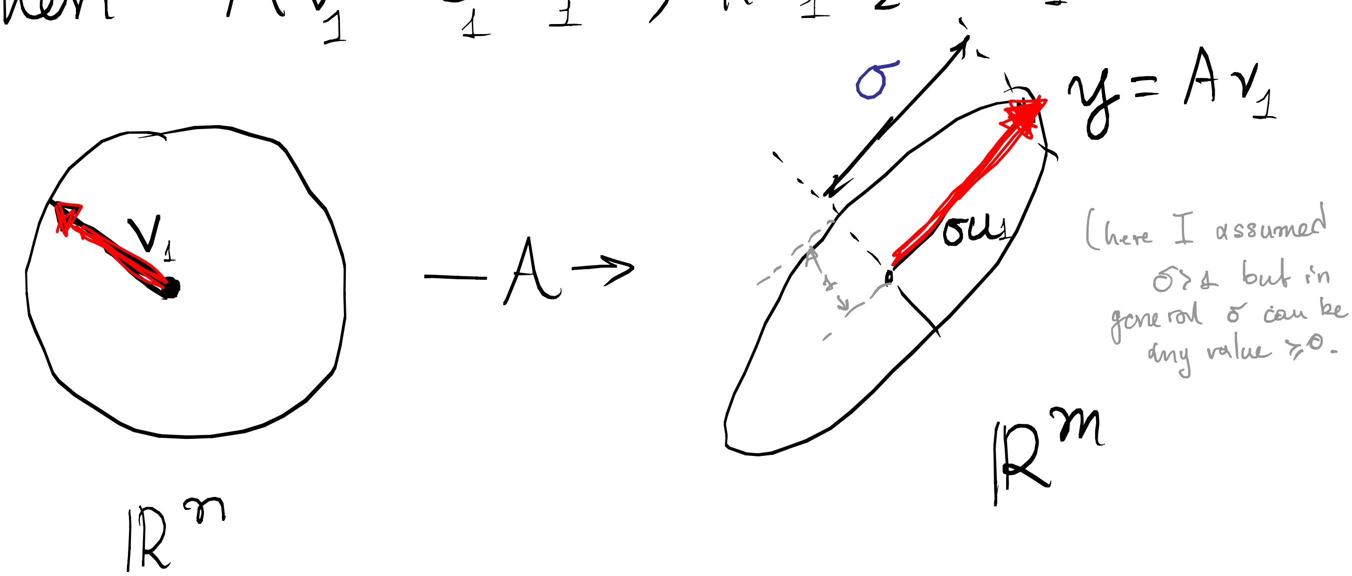
- Let $\sigma_1 := \|A\|_2$

- Then there exists (by definition of induced norms). $\underline{v_1 \in \mathbb{R}^n}$ such that $\|Av_1\|_2 = \sigma_1$

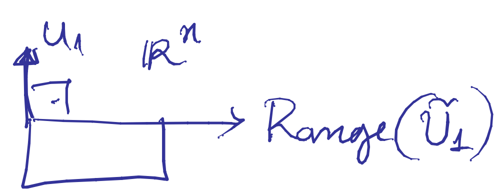
  [we will discuss uniqueness later]

- Let $y = Av_1$. ; Let $u_1 = \dfrac{y}{\|y\|_2} = \dfrac{y}{\sigma_1}$

Then $Av_1 = \sigma_1 u_1$, $\|v_1\|_2 = \|u_1\|_2 = 1$.



- Define $U_1 = \begin{bmatrix} | & \boxed{\tilde{U}_1} \\ u_1 & \\ | & \end{bmatrix} \in \mathbb{R}^{m\times m}$ orthonormal



i.e. construct an orthonormal basis so that u_1 is its first vector. (It is always possible to do that, we will see how to do it next week).

- Define $V_1 = \begin{bmatrix} | & \\ v_1 & \tilde{V}_1 \\ | & \end{bmatrix}$; orthonormal in $\mathbb{R}^{n\times n}$.

Notice: $\tilde{U}_1^T u_1 = 0$ and $\tilde{V}_1^T v_1 = 0$ by construction.

- Let $S_1 = U_1^T A V_1$.

$$S_1 = \begin{bmatrix} \text{---}\ u_1^T\ \text{---} \\ \tilde{U}_1^T \end{bmatrix} A \begin{bmatrix} | & \\ v_1 & \tilde{V}_1 \\ | & \end{bmatrix}$$

$$= \begin{bmatrix} \text{---}\ u_1^T\ \text{---} \\ \tilde{U}_1^T \end{bmatrix} \begin{bmatrix} | & \\ Av_1 & A\tilde{V}_1 \\ | & \end{bmatrix}$$

make sure you understand this block matrix multiplication

$$
= \left[\begin{array}{c|c} u_1^T A v_1 & -u_1^T A \tilde{V}_1 - \\ \hline | & \\ \tilde{U}_1^T A v_1 & \tilde{U}_1^T A \tilde{V}_1 \\ | & \end{array}\right] = \begin{array}{c} \\ \left[\begin{array}{c|c} \sigma & w^T \\ \hline q & A_1 \end{array}\right] \end{array}
$$

(block sizes: 1, $n-1$ columns; 1, $m-1$ rows)

. But since $Av_1 = u_1\sigma$ and $\tilde{U}_1^T u_1 = 0$,

$$
S_1 = \left[\begin{array}{c|c} \sigma_1 & w^T \\ \hline 0 & A_1 \end{array}\right].
$$

- $w = 0$. To show this use the norm of S_1.

Pick $z \in \mathbb{R}^n$, $\|z\| = 1$. and

compute $\|S_1 z\|_2^2 \leq \|S_1\|_2^2$ (Since norm is a maximum)

Pick $z = \begin{Bmatrix} \sigma_1 \\ w \end{Bmatrix} \dfrac{1}{(\sigma_1^2 + w^T w)^{1/2}}$ this factor is required to ensure that $\|z\|_2 = 1$

Then $S_1 z = \dfrac{1}{(\sigma_1^2 + w^T w)^{1/2}} \begin{bmatrix} \sigma_1^2 + w^T w \\ A_1 w \end{bmatrix}$ $\begin{array}{l} \in \mathbb{R} \\ \in \mathbb{R}^{m-1} \end{array}$

$$\Rightarrow S_1 z \cdot S_1 z = \frac{1}{\sigma_1^2 + w^T w}\left[(\sigma_1^2 + w^T w)^2 + A_1 w \cdot A_1 w \right]$$

$\boxed{A}$ $$\Rightarrow \underline{(\sigma_1^2 + w^T w) + \|A_1 w\|_2^2 / \sigma_1^2 + w^T w \leq \|S_1\|_2^2}$$

But since $S_1 = U_1^T A V_1$ with U_1, V_1 orthonormal, we will have

$\boxed{B}$ $$\underline{\|S_1\|_2 = \|A\|_2 = \sigma_1.}$$

Combining $\boxed{A}$ and $\boxed{B}$ we obtain

$$(\sigma_1^2 + w^T w) + \frac{\|A_1 w\|^2}{(\sigma_1^2 + w^T w)} \leq \sigma_1^2.$$

$$\boxed{\Rightarrow \quad w = 0}$$

since all terms are ≥ 0.

So far we have the following

$$\sigma_1 := \|A\|_2 = \|Av_1\|_2 = \sigma_1 \|u_1\|_2$$

$$Av_1 = \sigma_1 u_1 , \quad \|v_1\|_2 = \|u_1\|_2 = 1.$$

$\bar{U}_1, \bar{V}_1$ are orthonormal matrices

$$\bar{U}_1^T A \bar{V}_1^T = S_1 = \left[\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & A_1 \end{array}\right] ; \quad A_1 = \tilde{U}_1^T A \tilde{V}_1$$

$$A_1 \in \mathbb{R}^{m-1 \times n-1}.$$

repeat for A_1

$$\sigma_2 = \|A_1\|_2 ; \quad A_1 v_2 = \sigma_2 u_2 ; \quad \|v_2\|_2 = \|u_2\|_2 = 1.$$

$$S_2 = U_2^T A_1 V_2 = \left[\begin{array}{c|c} \sigma_2 & 0 \\ \hline 0 & A_3 \end{array}\right] , \quad A_3 \in \mathbb{R}^{m-2 \times n-2}$$

Let $\bar{U}_2 = \left[\begin{array}{c|c} 1 & 0\ 0\ 0 \\ \hline 0 & \\ 0 & U_2 \\ 0 & \end{array}\right] ; \quad V_2 = \begin{bmatrix} 0 \\ | \\ v_2 \end{bmatrix} , \quad u_2 = \begin{bmatrix} 0 \\ | \\ u_2 \\ | \end{bmatrix}$

$\bar{V}_2 = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & V_2 \end{array}\right] ;$

That is, we extend U_2 and V_2 from $\mathbb{R}^{(m-1) \times (n-1)}$ to $\mathbb{R}^{m \times n}$ (to match our original matrix).

Since $S_2 = U_2^T A_1 V_2 \Rightarrow$

$$\Rightarrow A_1 = U_2 S_2 V_2^T$$

$$\Rightarrow S_1 = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \boxed{U_2 S_2 V_2^T} \end{bmatrix} = \bar{U}_2 \begin{bmatrix} \sigma_1 & 0 \\ 0 & \boxed{S_2} \end{bmatrix} \bar{V}_2^T$$

$$= \bar{U}_2 \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \boxed{A_2} \end{bmatrix} \bar{V}_2^T$$

[drop bar-notation above U_2 and V_2]

$$\Rightarrow A = U_1 U_2 \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \boxed{A_2} \end{bmatrix} V_2^T V_1^T$$

By INDUCTION

$$A = \underbrace{\left(\prod_{i=1}^{r} U_i\right)}_{U} \begin{array}{c|cc} & r & n-r \\ \hline r & \begin{matrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r\end{matrix} & 0 \\ m-r & 0 & \boxed{A_r} \end{array} \underbrace{\left(\prod_{i=r}^{1} V_i^T\right)}_{V}$$

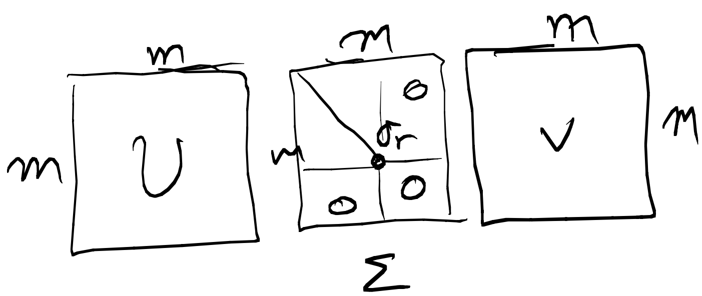
Exercise: Show that U, V ARE UNITARY MATRICES.

When $r = \operatorname{rank}(A)$ then $A_r = 0$

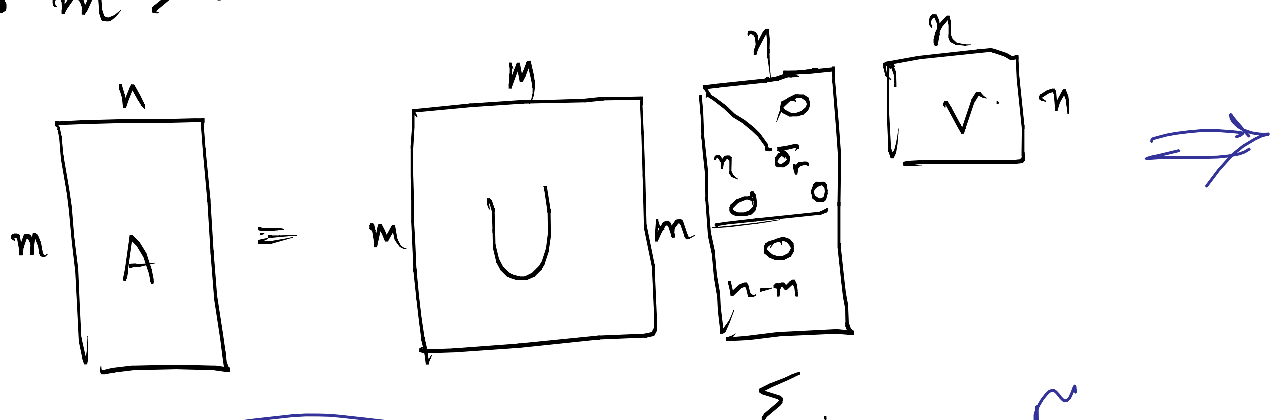
(we find that : $\sigma_{r+1} = 0$) and we terminate.

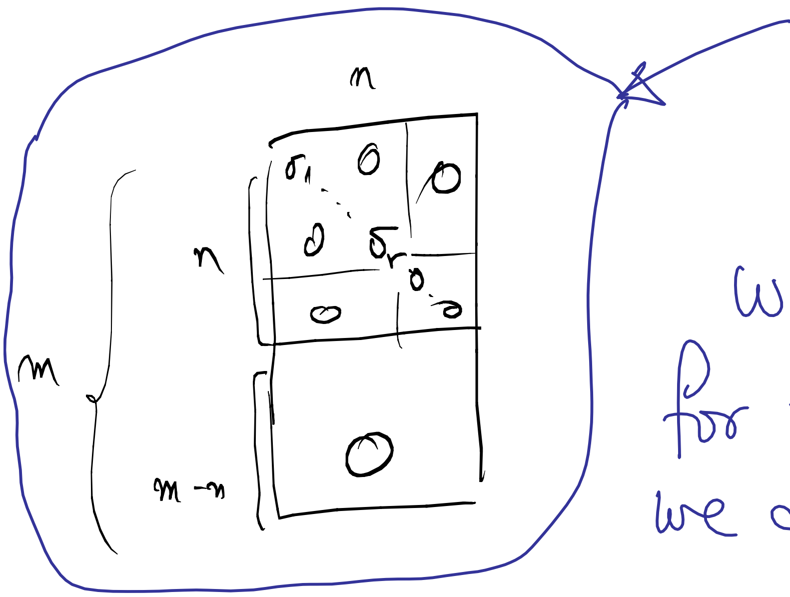
This decomposition is the FULL SVD.

- $m = n$



- $m > n$

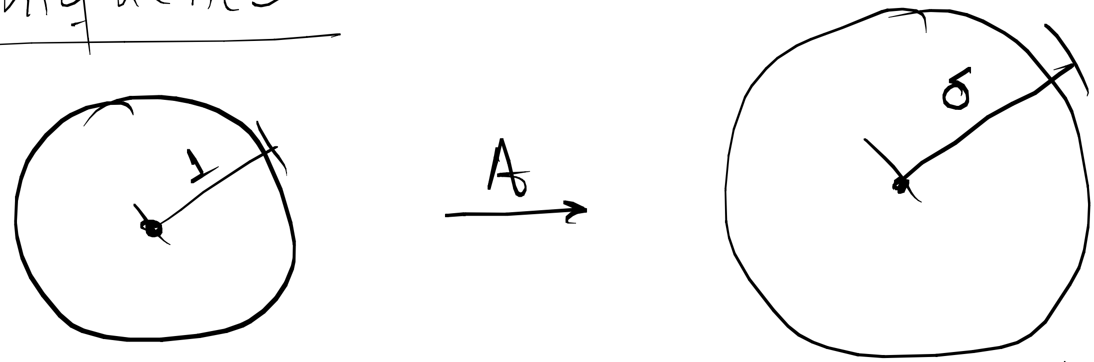




$$\Rightarrow A = \sum_{i=1}^{r} \sigma_i \underset{\sim}{u}_i \underset{\sim}{v}_i^*$$

which is the formula for the reduced SVD we discussed in lecture 4.

- Uniqueness



Consider A that maps spheres to spheres.

then $\|A\|_2 = \sigma$.

e.g. $A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$; $\|A\|_2 = 2$.

In this case, there exist more than one vector v such that $\|Av\|_2 = \sigma$.

For the example above,

$$\|Ae_1\|_2 = \|Ae_2\|_2 = \|A(a_1 e_1 + a_2 e_2)\|_2 = \sigma$$

with. $(\alpha_1^2 + a_2^2)^{1/2} = 1$

in fact, there are infinite number of vectors so that $\|Av_1\|_2 = \sigma$, $\|v_1\| = 1$.

If however there is only a unique vector (up to a sign) so that $\|Av_1\|_2 = \sigma$

Then σ is called <u>simple singular value</u>.

Examples (full SVD)

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} ; \quad U = V = \begin{bmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{bmatrix} ; \quad \Sigma = \mathrm{diag}([2,3])$$

$$A = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix} ; \quad U = \begin{bmatrix} 1 & 0 \\ 0 & \pm 1 \end{bmatrix} ; \quad V = \begin{bmatrix} 1 & 0 \\ 0 & \mp 1 \end{bmatrix}$$

$$\Sigma = \mathrm{diag}([3,2])$$

$$A = \begin{bmatrix} 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad V = \begin{bmatrix} 0 & \pm 1 \\ \pm 1 & 0 \end{bmatrix} ; \quad U = \begin{bmatrix} \pm 1 & 0 \quad 0 \\ \begin{matrix} 0 \\ 0 \end{matrix} & \boxed{V} \end{bmatrix}$$

any orthogonal basis in $\mathbb{R}^2$

$$\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} ; \quad V = \begin{bmatrix} \pm \sqrt{2}/2 & \mp \sqrt{2}/2 \\ \pm \sqrt{2}/2 & \pm \sqrt{2}/2 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 0 \end{bmatrix} ; \quad U = \begin{bmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{bmatrix}$$

THM 5.8 (TEXTBOOK)

Let $A \in \mathbb{C}^{m \times n}$

Let $r = \operatorname{rank}(A)$

Let $0 < \nu \leq r$

Let $A_\nu = \sum_{j=1}^{\nu} \sigma_j u_j v_j^*$

Then

(1) $$\min_{\substack{B \in \mathbb{C}^{m \times n} \\ \operatorname{rank}(B) \leq \nu}} \|A - B\|_2 = \sigma_{\nu+1}$$

(2) $$\|A - A_\nu\|_2 = \sigma_{\nu+1}$$

In other words, A_ν is the best ν-rank approximation to A.

(in the 2-norm or forbenius norm)

Notice that there exist many rank-1 decompositions of A. For example, if c_i are the columns of A and r_i are the rows,

$$A = \sum_{i=1}^{n} c_i \underset{\sim}{e}_i^{*} = \sum_{i=1}^{m} e_i \underline{r}_i^{*}$$

However, these expansions don't have best approximation properties.