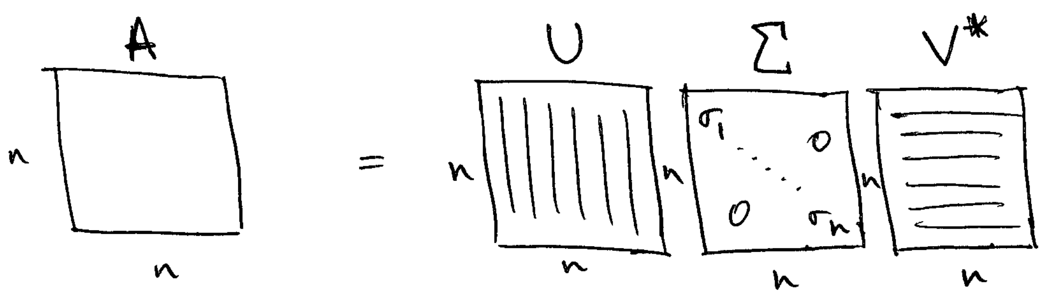
LECTURE 33
04/13/12

## SINGULAR VALUE DECOMPOSITION (SVD):

FOR ANY $A \in M_{n,n}(\mathbb{C})$,




$\Sigma = \text{DIAG}(\sigma_1, \ldots, \sigma_n)$, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$
SINGULAR VALUES
(E-VALUES OF $\sqrt{A^*A}$ OR $\sqrt{AA^*}$)

$V$ UNITARY (ORTHONORMAL E-VECTORS OF $\sqrt{A^*A}$)

$U$ UNITARY (ORTHONORMAL E-VECTORS OF $\sqrt{AA^*}$)

NOTE: $A^*A$ AND $AA^*$ ARE POSITIVE OPERATORS SINCE THEY ARE SELF-ADJOINT AND HAVE ALL EIGENVALUES POSITIVE $(\geq 0)$. TO SEE THIS, NOTE THAT $(A^*A)^* = A^*A^{**} = A^*A$. AND IF $\lambda$ IS AN E-VALUE W/ E-VECTOR $\underline{v}$ FOR $A^*A$, THEN

$$\lambda \|\underline{v}\|^2 = \lambda \langle \underline{v} | \underline{v} \rangle = \langle \underline{v} | \lambda \underline{v} \rangle = \langle \underline{v} | A^*A\underline{v} \rangle$$
$$= \langle A\underline{v} | A\underline{v} \rangle \quad \text{(BY PROPERTIES OF ADJOINT.)}$$
$$= \|A\underline{v}\|^2 \Rightarrow \lambda \geq 0.$$

Q: WHY SVD?

A: WE WILL SEE LATER THAT SVD IS USEFUL IN APPLICATIONS INVOLVING LARGE, HIGH-DIMENSIONAL DATA SETS. IT IS NOT AS USEFUL FOR DYNAMICAL APPLICATIONS SINCE

$$A^k = (U\Sigma V^*)^k = U\Sigma V^* U\Sigma V^* \cdots U\Sigma V^*$$
$$\neq U\Sigma^k V^*.$$

(UNLESS $U=V$, IN WHICH CASE THIS IS THE SAME AS JORDAN FORM.)

REMARK:

1) $A^*A$ AND $AA^*$ HAVE SAME E-VALUES!

PF. $A^*A\,\underline{v} = \lambda \underline{v} \Rightarrow AA^*(\underbrace{A\underline{v}}_{\text{CALL } \underline{u}}) = \lambda(\underbrace{A\underline{v}}_{\underline{u}})$

$AA^*\underline{u} = \lambda \underline{u} \Rightarrow A^*A(\underbrace{A^*\underline{u}}_{\text{CALL } \underline{v}}) = \lambda(\underbrace{A^*\underline{u}}_{\underline{v}}).$

2) $r \triangleq \text{RANK}(A) = \#$ OF NONZERO $\sigma_i$'s.

PF. $\text{RANK}(A) = \text{RANK}([A]_{\mathcal{E}\mathcal{F}}) = \text{RANK}(\Sigma)$
$= \#$ NONZERO $\sigma_i$'s

WHERE $\mathcal{E}, \mathcal{F}$ ARE COLUMNS OF $U, V$.

PROCEDURE TO FIND SVD:

① FIND NONZERO E-VALUES $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$, $r \leq n$. OF $A^*A$ AND CORRESPONDING E-VECTORS $\{v_i\}_{i=1}^r$.

② LET $\sigma_i = \sqrt{\lambda_i}$, $i=1,..,r$. THESE ARE THE E-VALUES OF $\sqrt{A^*A}$ (OR OF $\sqrt{AA^*}$) — THAT IS, THE SINGULAR VALUES OF $A$.

③ LET $\underline{u}_i = \frac{1}{\sigma_i} A\underline{v}_i$, $i=1,..,r$. THESE ARE THE ORTHONORMAL E-VECTORS OF $AA^*$.

(TO SEE THIS, NOTE THAT $\langle A\underline{v}_i | A\underline{v}_j \rangle = \langle A^*A\underline{v}_i | v_j \rangle$

$= \begin{cases} \sigma_i^2 & \text{IF } i=j \\ 0 & \text{ELSE} \end{cases} \Rightarrow \{\underline{u}_i\}$ ORTHONORMAL.)

④ IF NECESSARY: FOR REMAINING COLUMNS

$\underline{v}_{r+1}, \ldots, \underline{v}_n$ (E-VECTORS OF $A^*A$ W/ E-VALUE 0, I.E, IN $\mathrm{Ker}(A^*A)$)

$\underline{u}_{r+1}, \ldots, \underline{u}_n$ (E-VECTORS OF $AA^*$ W/ E-VALUE 0, I.E, IN $\mathrm{Ker}(AA^*)$)

USE GRAM-SCHMIDT. THAT IS, FIND A BASIS FOR $\mathrm{Ker}(A^*A)$ AND ORTHONORMALIZE TO GET $\{v_i\}_{i=r+1}^n$, AND SIMILARLY FIND A BASIS FOR $\mathrm{Ker}(AA^*)$ AND ORTHONORMALIZE TO GET $\{\underline{u}_i\}_{i=r+1}^n$.

EX. FIND SVD OF $A = \begin{pmatrix} 2 & 2 \\ -1 & 1 \end{pmatrix}$.

① $A^*A = \begin{pmatrix} 2 & -1 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 2 & 2 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix} \Rightarrow \lambda_1 = 8,\ \lambda_2 = 2$

(TRACE = 10, DET = 16)

$\Rightarrow \sqrt{\lambda_1} = \sqrt{8},\ \sqrt{\lambda_2} = \sqrt{2}$

② $\Rightarrow \sigma_1 = \sqrt{8},\ \sigma_2 = \sqrt{2}$.

CORRESPONDING EVECTORS OF $A^*A$ ARE $\underline{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix},\ \underline{v}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \end{pmatrix}$

ORTHONORMAL.

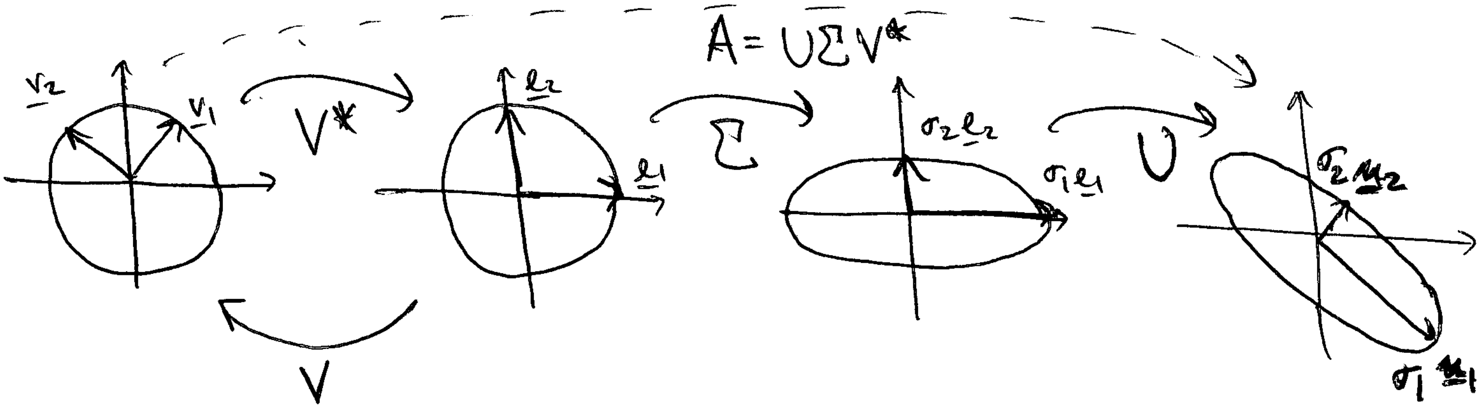
③ $\frac{1}{\sigma_1} A \underline{v}_1 \doteq \underline{u}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

$\frac{1}{\sigma_2} A \underline{v}_2 \doteq \underline{u}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

ORTHONORMAL.

$$\Rightarrow A = \underset{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}{\underset{\parallel}{U}}\ \underset{\begin{pmatrix} \sqrt{8} & 0 \\ 0 & \sqrt{2} \end{pmatrix}}{\underset{\parallel}{\Sigma}}\ \underset{\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}}{\underset{\parallel}{V^*}}$$

GEOMETRY OF SVD: EVERY LINEAR TRANSFORMATION IS GIVEN BY A ROTATION, STRETCH, AND ANOTHER ROTATION. FOR EX, IN $\mathbb{R}^2$,



**Q:** WHY DOES $A^*A$ PLAY A ROLE TO BEGIN WITH?

FROM THE GEOMETRIC PICTURE, WE SEE THAT ANY LINEAR TRANSFORMATION IS A MAPPING FROM THE UNIT SPHERE TO AN ELLIPSE (A HYPERELLIPSE IN HIGHER DIMENSIONS).

WE SEEK $\underline{x}$'s SUCH THAT $\|\underline{x}\|=1$ AND $A\underline{x}$ HAS MAXIMAL (OR MINIMAL) LENGTH.

Let $Q(\underline{x}) = \|A\underline{x}\|^2$. THEN, WE SEEK SOLUTIONS $\underline{x}$ THAT

$$\begin{cases} \text{EXTREMIZE } Q(\underline{x}) = \|A\underline{x}\|^2 = \langle A\underline{x} | A\underline{x} \rangle = \langle \underline{x} | A^*A\underline{x} \rangle \\ \text{w/ CONSTRAINT } \|\underline{x}\| = 1. \end{cases}$$

USING LAGRANGE MULTIPLIERS TO SOLVE THIS CONSTRAINED MAXIMIZATION/MINIMIZATION PROBLEM, WE GET THAT SOLUTIONS SATISFY

$$A^*A\,\underline{x} = \lambda\,\underline{x} \quad \text{FOR SOME } \lambda$$

LAGRANGE MULTIPLIER

$\Rightarrow$ I.E., LOOK AT E-VECTORS OF $A^*A$!

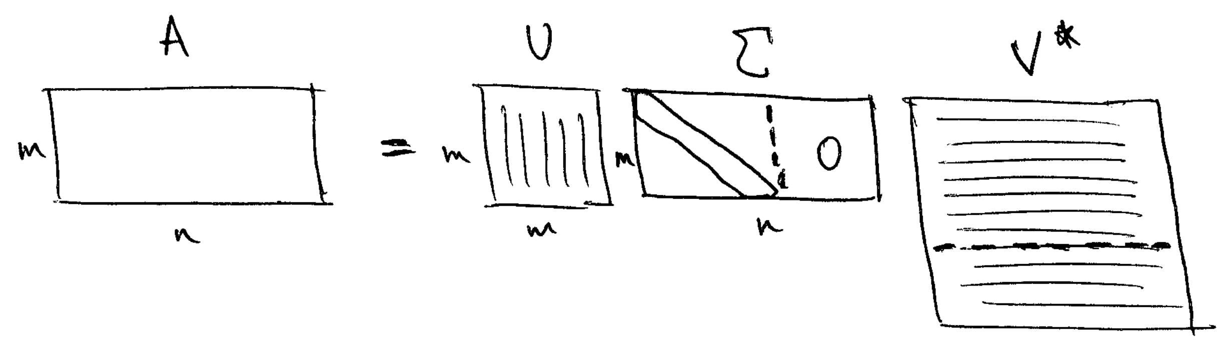
# Lecture 34
04/16/12

## Singular value decomposition (SVD) (cont'd):

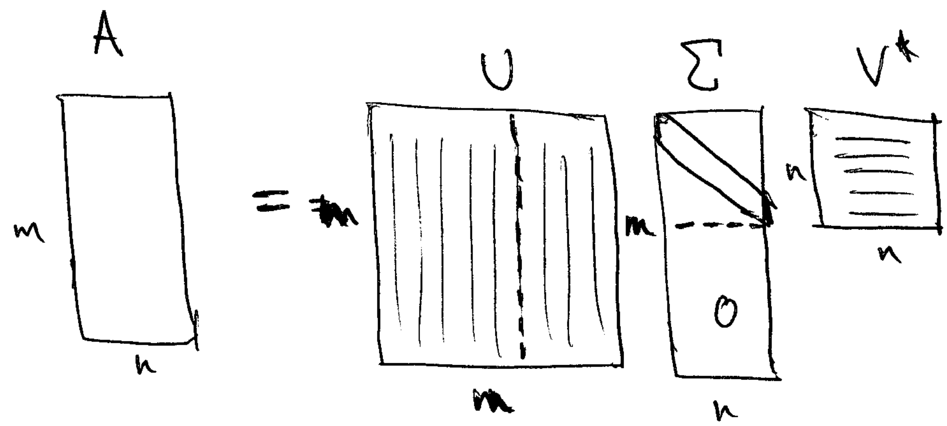
Last time, we introduced SVD for any square matrices — in fact, can find SVD of any $A \in M_{m,n}(\mathbb{C})$

$m,n$ may be different!

(m<n)




(n<m)




$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$ (rectangular diagonal matrix w/ singular values on diagonal)

$V \in M_{n,n}(\mathbb{C})$ unitary (e-vectors of $A^*A$)

$U \in M_{m,m}(\mathbb{C})$ unitary (e-vectors of $AA^*$)

PROCEDURE TO FIND FIND SVD EXACTLY SAME AS BEFORE, EXCEPT

(4') IF NECESSARY: $\underline{v}_{r+1}, \dots, \underline{v}_{n} \in \mathrm{Ker}(A^*A)$

$\underline{u}_{r+1}, \dots, \underline{u}_{m} \in \mathrm{Ker}(AA^*)$

FIND USING GRAM-SCHMIDT.

Ex. FIND SVD OF $A = \begin{pmatrix} 1 & 2 \\ 2 & 2 \\ 2 & 1 \end{pmatrix}$

(1) $A^*A = \begin{pmatrix} 9 & 9 \\ 8 & 8 \end{pmatrix} \Rightarrow \lambda_1 = 17, \lambda_2 = 1$

W/ CORRESPONDING E-VECTORS

$\underline{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \underline{v}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$

ORTHONORMAL.

(2) $\sigma_1 = \sqrt{\lambda_1} = \sqrt{17}, \quad \sigma_2 = \sqrt{\lambda_2} = 1.$

(3) $\underline{u}_1 = \frac{1}{\sigma_1} A\underline{v}_1 = \frac{1}{\sqrt{17}\sqrt{2}}\begin{pmatrix} 1 & 2 \\ 2 & 2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{34}}\begin{pmatrix} 3 \\ 4 \\ 3 \end{pmatrix}$

$\underline{u}_2 = \frac{1}{\sigma_2} A\underline{v}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 2 \\ 2 & 2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$

(4) STILL NEED $\underline{u}_3$. NOTE THAT

$AA^* = \begin{pmatrix} 5 & 6 & 4 \\ 6 & 8 & 6 \\ 4 & 6 & 5 \end{pmatrix} \xrightarrow{\text{ROW REDUCE}} \begin{pmatrix} \boxed{1} & 0 & -1 \\ 0 & \boxed{1} & \frac{3}{2} \\ 0 & 0 & 0 \end{pmatrix}$

$$\Rightarrow \operatorname{Ker}(AA^*) = \text{SPAN}\left\{\frac{1}{\sqrt{17}}(2,-3,2)^T\right\}$$

$$\Rightarrow \underline{u}_3 = \frac{1}{\sqrt{17}}(2,-3,2)^T.$$

(EASY TO CHECK THAT $\underline{u}_3 \perp \underline{u}_1$ AND $\underline{u}_3 \perp \underline{u}_2$).

SO,

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

$$U = \begin{pmatrix} \frac{3}{\sqrt{34}} & -\frac{1}{\sqrt{2}} & \frac{2}{\sqrt{17}} \\ \frac{4}{\sqrt{34}} & 0 & \frac{-3}{\sqrt{17}} \\ \frac{3}{\sqrt{34}} & \frac{1}{\sqrt{2}} & \frac{2}{\sqrt{17}} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sqrt{17} & 0 \\ 0 & 1 \end{pmatrix}, \quad V^* = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

## APPLICATIONS OF SVD:

IT IS EASY TO SEE THAT $A = U\Sigma V^*$

$$\Rightarrow A = \sum_{i=1}^{r} \sigma_i \underline{u}_i \underline{v}_i^* \qquad (r = \text{RANK}(A)).$$

NOTE THAT $\underline{u}_i \underline{v}_i^*$ IS A RANK 1 MATRIX (WHY?),

SO $A = \sum_{i=1}^{p} \sigma_i \underline{u}_i \underline{v}_i^*$ IS A RANK $p$ MATRIX.

AS WE NOW SEE, THIS IS THE "BEST" $p$-RANK APPROXIMATION TO $A$ IN SOME SENSE.

WE NEED A NOTION OF DISTANCE FOR MATRICES.

CONSIDER THE FROBENIUS INNER PRODUCT ON $V = M_{m,n}(\mathbb{C})$:

$$\langle A | B \rangle = \mathrm{Tr}(A^* B).$$

$$\Rightarrow \|A\|^2 = \langle A | A \rangle = \mathrm{Tr}(A^* A) = \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 .$$

CAN CHECK THAT FROBENIUS NORM IS UNITARILY INVARIANT — I.E.,

$W$ ISOMETRY $\Rightarrow$ $\|WA\| = \|A\|$ FOR ANY $A$.
$\|AW\| = \|A\|$

(IN SQUARED FROBENIUS NORM!)

THM. IF $A = U \Sigma V^*$, THE CLOSEST $p$-RANK APPROXIMATION TO $A$ IS $B = U \Sigma_p V^*$, WHERE $\Sigma_p = \mathrm{DIAG}(\sigma_1, \ldots, \sigma_p)$

PF. WE SEEK TO SOLVE

$$\begin{cases} \text{MINIMIZE } \|A - B\|^2 \text{ OVER ALL } B \text{ SUCH THAT} \\ \mathrm{RANK}(B) \leq p \leq r. \end{cases}$$

IF $A$ HAS SVD $A = U \Sigma V^*$, DEFINE A NEW MATRIX $S = U^* B V$, SO THAT $B = U S V^*$.

THEN,

$$\|A - B\|^2 = \|U \Sigma V^* - U S V^*\|^2$$
$$= \|U(\Sigma - S)V^*\|^2$$

$$= \|\Sigma - S\|^2 \quad (\text{SINCE } U, V^* \text{ ISOMETRIES})$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} |\Sigma_{ij} - S_{ij}|^2$$

$$= \sum_{i=1}^{r} |\sigma_i - S_{ii}|^2 + \sum_{\substack{i,j \text{ S.T.} \\ \Sigma_{ij}=0}} \sum |S_{ij}|^2 .$$

- TO MAKE THE $2^{\text{ND}}$ TERM AS SMALL POSSIBLE, WE TAKE $S_{ij} = 0$ EVERYWHERE $\Sigma_{ij} = 0$. SO, AT THIS POINT WE KNOW THAT $S = \text{DIAG}(S_{11}, S_{22}, \ldots, S_{rr})$.

- TO MAKE THE $1^{\text{ST}}$ TERM AS SMALL AS POSSIBLE, WE TAKE $S_{ii} = \sigma_i$ FOR $i = 1, \ldots, p$, AND ZERO OTHERWISE. THIS IS THE BEST $S$ OF RANK $\leq p$ WHICH MINIMIZES $\|\Sigma - S\|^2$.

SO, $S = \text{DIAG}(\sigma_1, \ldots, \sigma_p) \implies B = U \Sigma_p V^*$ IS THE BEST $p$-RANK APPROX!

THIS PROPERTY HAS WIDE APPLICABILITY TO DATA COMPRESSION, PRINCIPAL COMPONENT ANALYSIS, FACTOR ANALYSIS, PATTERN RECOGNITION, ...

EX. $A = \begin{pmatrix} 1.01 & 1 & 1 \\ 1 & 1.01 & 1 \\ 1 & 1 & 1 \end{pmatrix}$

RANK$(A) = 3$, BUT IT SHOULD BE VERY CLOSELY APPROX. BY A RANK 1 MATRIX SINCE $1.01 \approx 1$. WE FIND

$$\Sigma = \begin{pmatrix} \overset{\sigma_1}{\overset{\shortparallel}{3.01}} & & \\ & \overset{\sigma_2}{\overset{\shortparallel}{0.01}} & \\ & & \overset{\sigma_3}{\overset{\shortparallel}{0.01}} \end{pmatrix}$$

SO $\sigma_1 \gg \sigma_2, \sigma_3$.

$\rightsquigarrow$ RANK$(A) = 3$ BUT IS VERY "CLOSE" TO RANK 1 MATRIX

$$\sigma_1 \underline{u}_1 \underline{v}_1^* = \begin{pmatrix} 1.01 & 1.01 & 1.01 \\ 1.01 & 1.01 & 1.01 \\ 1.01 & 1.01 & 1.01 \end{pmatrix}$$