SINGULAR VALUE DECOMPOSITION (SVD)

* THE SVD IS A MATRIX FACTORIZATION WHICH IS COMMON IN MANY ALGORITHMS. IT ALSO PROVIDES IMPORTANT INSIGHT INTO GEOMETRIC ASPECTS OF THE MATRIX

SETTING: LET $A \in \mathbb{R}^{m \times n}$ BE A MATRIX
$\vec{x} \in \mathbb{R}^n$ BE A VECTOR
$\vec{b} \in \mathbb{R}^m$ " " "

- WE CAN INTERPRET THE EQUATION $A\vec{x} = \vec{b}$ AS A DEFINING A MAPPING FROM $\vec{x}$ TO $\vec{b}$
- MORE GENERALLY WE CAN INTERPRET A AS DEFINING A LINEAR MAPPING FROM $\mathbb{R}^n$ TO $\mathbb{R}^m$
  - GIVEN $c \in \mathbb{R}$, $\vec{x}, \vec{y} \in \mathbb{R}^n$

$$\left.\begin{aligned} A(\vec{x}+\vec{y}) &= A\vec{x} + A\vec{y} \\ A(c\vec{x}) &= c(A\vec{x}) \end{aligned}\right\} \text{PROPERTIES OF LINEAR MAPPING}$$

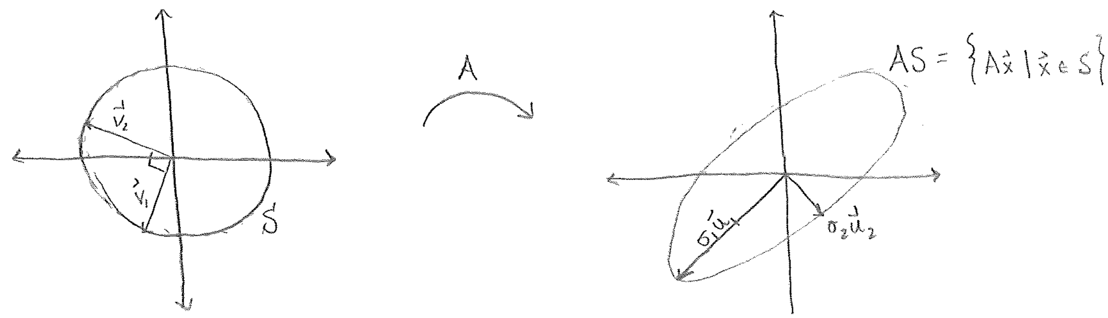
  - BECAUSE OF LINEARITY WE CAN UNDERSTAND GEOMETRIC DETAILS OF THE LINEAR TRANSFORMATION BY STUDYING THE TRANSFORMATION OF VECTORS ON UNIT SPHERE IN $\mathbb{R}^n$! FOR ANY $\vec{x} \in \mathbb{R}^n$, $\vec{x} \neq \vec{0}$

$$A\vec{x} = A\left(\|x\| \frac{\vec{x}}{\|x\|}\right) = \|x\| \, A\Big(\underbrace{\frac{\vec{x}}{\|x\|}}_{\text{ON UNIT CIRCLE}}\Big)$$

  - GEOMETRIC FACT:
    - THE IMAGE OF THE UNIT SPHERE (IN $\mathbb{R}^n$) UNDER ANY $m \times n$ MATRIX IS A HYPERELLIPSE IN $\mathbb{R}^n$!
    - A HYPERELLIPSE IN $\mathbb{R}^m$ IS THE SURFACE OBTAINED BY STRETCHING THE UNIT SPHERE IN $\mathbb{R}^m$ BY SOME FACTORS $\sigma_1, \dots, \sigma_m$ (POSSIBLY ZERO) IN SOME ORTHOGONAL DIRECTIONS $\vec{u}_1, \dots, \vec{u}_m$, $\|\vec{u}_i\| = 1$, $i = 1, \dots, m$

THE VECTORS $\{\sigma_i \vec{u}_i\}_{i=1}^m$ ARE THE PRINCIPAL SEMIAXES OF THE HYPERELLIPSE WITH LENGTHS $\sigma_1, \dots, \sigma_m$

Ex: $A \in \mathbb{R}^{2\times 2}$: LET $S = \{ \vec{x} \in \mathbb{R}^2 \cdot \|x\| = 1$



- LET $\vec{v}_j$ BE THE PREIMAGE OF $\sigma_j \vec{u}_j$, i.e. $A\vec{v}_j = \sigma_j \vec{u}_j$
  - $\vec{v}_j$, $\vec{u}_j$ MAY BE CHOSEN SO THAT $\sigma_j \geq 0$
  - $\vec{v}_j \cdot \vec{v}_k = \delta_{jk} = \begin{cases} 1 & j = k \\ 0 & \text{ELSE} \end{cases}$ * ORTHOGONALITY IS NOT OBVIOUS BUT CAN BE PROVEN!

- FOR NOW ASSUME $m \geq n$ AND $\text{RANK}(A) = n$

DEF: • THE n SINGULAR VALUES of A ARE THE LENGTHS OF THE PRINCIPAL SEMIAXES OF AS, WRITTEN $\sigma_1, \dots, \sigma_n$.
BY CONVENTION WE ORDER THEM

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n > 0$$

- THE n LEFT SINGULAR VECTORS OF A ARE THE UNIT VECTORS $\{ \hat{u}_1, \dots, \hat{u}_n \}$ ORIENTED IN THE DIRECTIONS OF THE PRINCIPAL SEMIAXES OF AS.

  $\Rightarrow$ $\sigma_j \hat{u}_j$ IS THE j^{th} LARGEST PRINCIPAL SEMIAXIS OF AS

- THE n RIGHT SINGULAR VECTORS OF A ARE THE UNIT VECTORS $\{ \vec{v}_1 \dots \vec{v}_n \} \in S$ THAT ARE PREIMAGES OF THE PRINCIPAL SEMIAXES

$$A\vec{v}_j = \sigma_j \vec{u}_j$$

REDUCED SVD

IN MATRIX NOTATION THE EQUATION $A\vec{v}_j = \sigma_j \hat{u}_j \quad j=1,\dots,n$

CAN BE EXPRESSED

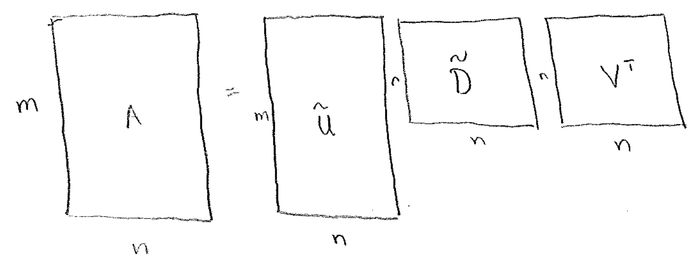
$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \underbrace{\begin{bmatrix} \vec{v}_1 \,|\, \vec{v}_2 \,|\, \dots \,|\, \vec{v}_n \end{bmatrix}}_{\tilde{V}} = \underbrace{\begin{bmatrix} \hat{u}_1 & \hat{u}_2 & \dots & \hat{u}_n \end{bmatrix}}_{\tilde{U}} \underbrace{\begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{bmatrix}}_{\tilde{D}}$$

$$AV = UD$$

SINCE COLUMNS OF V ARE ORTHONORMAL $VV^T = I_n \Rightarrow V^T = V^{-1}$

$$A = \tilde{U}\tilde{D}V^T$$

REDUCED SVD $(m \geq n)$ SCHEMATIC



FULL SVD

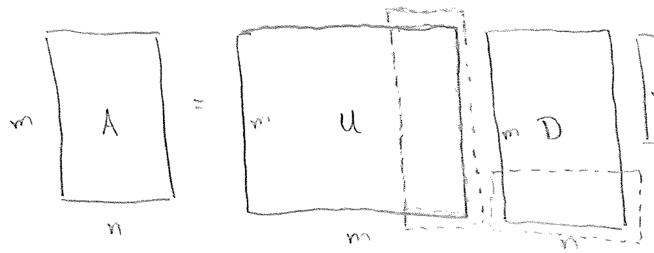
THE COLUMNS IN $\tilde{U}$ ARE n ORTHONORMAL COLUMNS IN $\mathbb{R}^m$. IF $m > n$ $\{\vec{u}_1, \dots \vec{u}_n\}$ DO NOT FORM A BASIS FOR $\mathbb{R}^m$ BUT WE CAN FIND $n-m$ ADDITIONAL ORTHONORMAL VECTORS $\vec{u}_{n+1}, \dots \vec{u}_m$ SO THAT $\{\vec{u}_1, \dots, \vec{u}_m\}$ FORM AN ORTHONORMAL BASIS FOR $\mathbb{R}^m$.

$\tilde{U}$ REPLACED WITH $U = [u_1 | \dots | \vec{u}_m]$ NOW $UU^T = I_m$

$\tilde{D}$ REPLACED WITH $D = \begin{array}{c} \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \\ & 0 & \end{bmatrix} \\ n \end{array} m$ — $(m-n) \times n$ BLOCK OF ZEROS

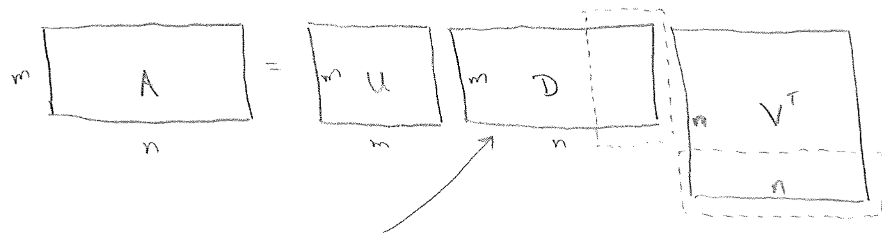
$$\Rightarrow \quad A = UDV^T$$

Full SVD ($m \geq n$)



Vectors $\vec{u}_{n+1}, \dots \vec{u}_m$ are multiplied by 0!

Full SVD ($n \geq m$)



Vector $\vec{v}_{m+1}^T \dots \vec{v}_n^T$ are multiplied by 0!

$$D = \left[\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_m & \end{array}\right]$$

$m \times (n-m)$ block of zeros

Note:

- Generically, $A \in \mathbb{R}^{m \times n}$ has $\min\{m,n\}$ singular values
- $A \in \mathbb{R}^{m \times n}$ implies $U \in \mathbb{R}^{m \times m}$, $D \in \mathbb{R}^{m \times n}$, $V \in \mathbb{R}^{n \times n}$ $\quad UU^T = I_m$, $VV^T = I_n$
- Question: What does multiplication by an orthonormal matrix represent geometrically?

Answer: Let $\vec{e}_j = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ ← j^{th} entry then $\underbrace{\left[\vec{u}_1 | \dots | \vec{u}_m\right]}_{U} \vec{e}_j = \vec{u}_j$

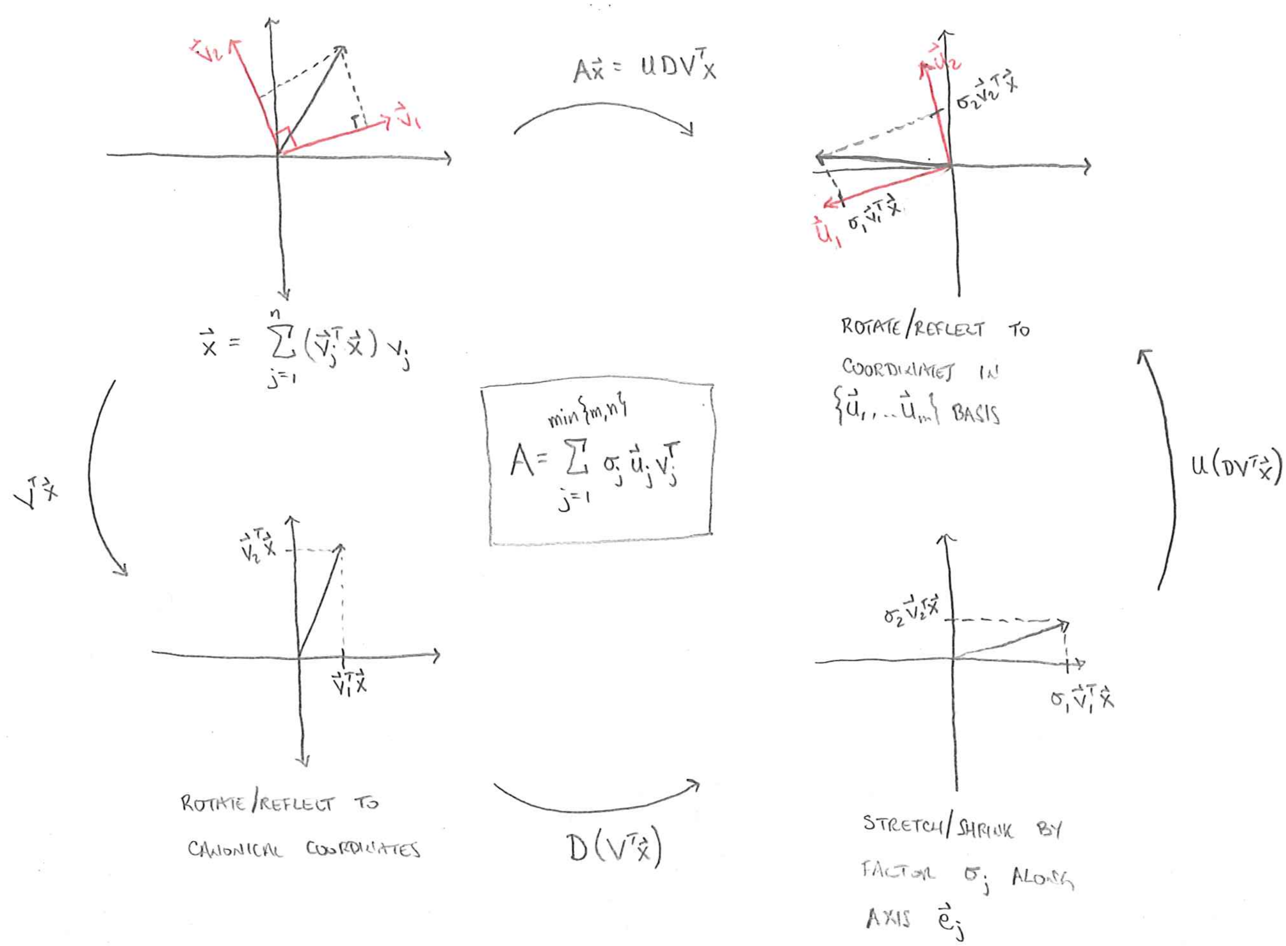
$$\underbrace{\begin{bmatrix} \vec{u}_1^T \\ \vdots \\ \vec{u}_m^T \end{bmatrix}}_{U^T} \vec{u}_j = \vec{e}_j$$

$U \vec{e}_j = \vec{u}_j$

$U^T \vec{u}_j = e_j$

- Multiplication by U rotates/reflects canonical basis vectors $\vec{e}_1, \dots \vec{e}_m$ to new basis $\vec{u}_1 \dots \vec{u}_m$
- Mult. by U^T rotates/reflects basis $\vec{u}_1, \dots, \vec{u}_m$ back to canonical basis $\vec{e}_1, \dots, \vec{e}_j$

GEOMETRY OF MATRIX MULTIPLICATION



THM: ANY $m \times n$ MATRIX HAS A SINGULAR VALUE DECOMPOSITION, $A = UDV^T$, W/ UNIQUELY DETERMINED SINGULAR VALUES σ_j, $j = 1, \dots, \min\{m,n\}$

NOTE:
- EVERY MATRIX (SQUARE OR NON-SQUARE) HAS A SVD. WE CANNOT SAY THE SAME ABOUT EIGENDECOMPOSITIONS!
- WE DO NOT NEED A FULL RANK ASSUMPTION ABOUT A!

QUESTION: ASSUME $A \in \mathbb{R}^{m \times n}$ HAS RANK $r < \min\{m,n\}$

a) WHAT CAN WE SAY ABOUT THE SINGULAR VALUES OF A?

$$A = \underbrace{U}_{\text{FULL RANK}} D \underbrace{V^T}_{\text{FULL RANK}}$$

$\Rightarrow$ RANK$(D) = r$

$\Rightarrow$ ONLY (FIRST) r SINGULAR VALUES ARE NON-ZERO!

b) HOW ARE THE SINGULAR VECTOR RELATED TO THE COL(A)?

$$A = \sum_{j=1}^{r} \sigma_j \vec{u}_j v_j^T \Rightarrow A\vec{x} = \sum_{j=1}^{r} \sigma_j u_j v_j^T \vec{x}$$

$$= \sum_{j=1}^{r} (\sigma_j v_j^T \vec{x})\, \vec{u}_j$$

$$\Rightarrow \text{COL}(A) = \text{SPAN}\{\vec{u}_1, \dots, \vec{u}_r\}$$

c) HOW ARE THE SINGULAR VECTORS RELATED TO KER (A)?

$$\underset{\mathbb{R}^n}{\underset{\cap}{\vec{x}}} = \vec{v}_1^T \vec{x} + \dots + \vec{v}_n^T x$$

$$\Rightarrow A\vec{x} = U \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & 0 & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \; 0 \\ & 0 & & & \end{bmatrix} \begin{bmatrix} v_1^T x \\ \vdots \\ \vec{v}_n^T \vec{x} \end{bmatrix} = \left[\vec{u}_1 | \dots | \vec{u}_m\right] \begin{bmatrix} \sigma_1 \vec{v}_1^T \vec{x} \\ \vdots \\ \sigma_r \vec{v}_r^T \vec{x} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\Rightarrow A\vec{x} = \vec{0} \text{ IMPLIES } \vec{v}_1^T \vec{x} = \dots = \vec{v}_r^T \vec{x}$$

$$\vec{x} \in \text{SPAN}\{\vec{v}_{r+1}, \dots, \vec{v}_n\}$$

- SUPPOSE $\underset{\mathbb{R}^{m \times n}}{\underset{\cap}{A}}$ IS SYMMETRIC AND POSITIVE SEMIDEFINITE (NON-NEG. EIGENVALUES)

$$A = W \Lambda W^T \qquad W = \left[\vec{w}_1 | \dots | \vec{w}_m\right]$$

ORTHONORMAL EIGENVECTORS

$$\Lambda = \begin{bmatrix} \lambda_1 & \ddots & 0 \\ 0 & & \lambda_n \end{bmatrix}$$

EIGENVALUES IN DECREASING ORDER

- SVD AND EIGENDECOMP ARE THE SAME IN THIS CASE!

PCA AND SVD

LET $X \in \mathbb{R}^{N\times d}$ BE A CENTERED DATA MATRIX WITH SVD

$$X = UDV^T \qquad U \in \mathbb{R}^{N\times N} \quad UU^T = I_N$$
$$V \in \mathbb{R}^{d\times d} \quad VV^T = I_d$$
$$D \in \mathbb{R}^{N\times d}$$

FIND THE PRINCIPAL COMPONENT LOADINGS, SCORES, AND VARIANCES OF X IN TERMS SINGULAR VECTORS AND VALUES

OBSERVE:

$$\hat{\Sigma} = \frac{X^T X}{N} = \frac{1}{N}\left((UDV^T)^T (UDV^T)\right)$$

$$= \frac{V D^T U^T U D V^T}{N} = V \begin{bmatrix} \sigma_1^2/N & & 0 \\ & \ddots & \\ 0 & & \sigma_d^2/N \end{bmatrix} V^T$$

- PRINCIPAL COMPONENT VARIANCES ARE $\sigma_1^2/N, \dots, \sigma_d^2/N$

  " $\lambda_j = \sigma_j^2/N$ " , $j = 1, \dots, d$

- PRINCIPAL COMPONENT LOADINS ARE RIGHT SINGULAR VECTOR

  " $\vec{w}_j = \vec{v}_j$ " $j = 1, \dots, d$

- PRINCIPAL COMPONENT SCORES ARE THE TRANSPOSE NON-ZERO PRINCIPAL AXES

  " $Y = XV$ " $= UD = \begin{bmatrix} \sigma_1 \vec{u}_1^T \\ \vdots \\ \sigma_d \vec{u}_d^T \end{bmatrix}^T$

  $Y_i^T = x_i^T [\vec{v}_1 | \dots | \vec{v}_d]$

NOTE: DIRECTLY CALCULATING THE SVD OF X IS MORE RELIABLE THAN FINDING $\frac{X^T X}{N}$ THE CALCULATING ITS EIGENDECOMPOSITION! $\Big\}$ PCA COMPUTATION RELIES ON SVD!

A NOTE ON FINDING SINGULAR VALUES AND VECTORS

LET $A \in \mathbb{R}^{m \times n}$ HAVE SVD $A = UDV^T$

$U \in \mathbb{R}^{m \times m}$ ← ORTHONORMAL
$D \in \mathbb{R}^{m \times n}$ ← DIAGONAL
$V \in \mathbb{R}^{n \times n}$ ⟷ ORTHONORMAL

THEN

$$AA^T = UDV^T V D^T U^T = U(DD^T)U^T$$

$$DD^T = \begin{cases} \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_m^2 \end{bmatrix} & m \le n \\ \begin{bmatrix} \sigma_1^2 & & & & 0 \\ & \ddots & & & \\ & & \sigma_n^2 & & \\ & & & 0 & \\ 0 & & & & \ddots \ 0 \end{bmatrix} & m > n \end{cases} \Rightarrow DD^T \text{ IS DIAGONAL}$$

* 1) THE LEFT SINGULAR VECTORS OF A ARE THE EIGENVECTORS OF AA^T W/ EIGENVALUES $\sigma_1^2, \ldots \sigma_{m \wedge n}^2$

2) THE SINGULAR VALUES OF A ARE THE SQUARE ROOTS OF THE FIRST $\min\{m, n\}$ EIGENVALUES OF AA^T

3) SIMILARLY THE RIGHT SINGULAR VECTORS OF A ARE THE EIGENVECTORS OF A^TA !