Review

Last time: we did the outer product form SVD

A: $m \times n$ of rank r

$$A = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T$$

- $\sigma_1 \geq \cdots \geq \sigma_r > 0$ are the singular values
- $\{v_1, \ldots, v_r\}$ is an orthonormal set in $\mathbb{R}^n$
  - called the right singular vectors
  - forms a basis for $\mathrm{Row}(A)$
  - orthonormal eigenvectors of A^TA:

    $$A^TAv_i = \sigma_i^2 v_i$$
- $\{u_1, \ldots, u_r\}$ is an orthonormal set in $\mathbb{R}^m$
  - called the left singular vectors
  - forms a basis for $\mathrm{Col}(A)$
  - orthonormal eigenvectors of AA^T:

    $$AA^Tu_i = \sigma_i^2 u_i$$

The singular vectors are related by

$$Av_i = \sigma_i u_i \qquad A^Tu_i = \sigma_i v_i$$

SVD of A^T is

$$A^T = \sigma_1 v_1 u_1^T + \cdots + \sigma_r v_r u_r^T$$

NB: If A is a wide matrix $(m<n)$ then

$A^TA: n\times n \qquad AA^T: m\times m \leftarrow$ smaller

So it's easier to compute eigenvalues & eigenvectors of AA^T!

$$\boxed{\text{If } A \text{ is wide, compute the SVD of } A^T.}$$

Eg: $A=\begin{pmatrix} -10 & 10 & -10 & 10 \\ 10 & 5 & 10 & 5 \end{pmatrix}$

$$A^TA=\begin{pmatrix} 200 & -50 & 200 & -50 \\ -50 & 125 & -50 & 125 \\ 200 & -50 & 200 & -50 \\ -50 & 125 & -50 & 125 \end{pmatrix}$$ yikes!

Let's compute the SVD of A^T instead.

$AA^T=\begin{pmatrix} 400 & -100 \\ -100 & 250 \end{pmatrix} \qquad p(\lambda)=(\lambda-450)(\lambda-200)$

$\lambda_1=450 \Rightarrow \sigma_1=\sqrt{450}=15\sqrt{2} \quad u_1=\frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ -1 \end{pmatrix} \quad v_1=\frac{1}{\sigma_1}A^Tu_1=\frac{1}{\sqrt{10}}\begin{pmatrix} -2 \\ 1 \\ -2 \\ 1 \end{pmatrix}$

$\lambda_2=200 \Rightarrow \sigma_2=\sqrt{200}=10\sqrt{2} \quad u_2=\frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad v_2=\frac{1}{\sigma_2}A^Tu_2=\frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ 2 \\ 1 \\ 2 \end{pmatrix}$

$\Rightarrow A^T=15\sqrt{2}\, v_1u_1^T+10\sqrt{2}\, v_2u_2^T$

$\Rightarrow A=15\sqrt{2}\, u_1v_1^T+10\sqrt{2}\, u_2v_2^T$

u_i are right-singular vectors of $A^T \rightarrow$ left-singular vectors of A

SVD: Matrix Form

Let A be an $m \times n$ matrix of rank r.

Then $A = U \Sigma V^T$ where:

(square with orthonormal columns)

- $U = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{pmatrix}$ is an $m \times m$ orthogonal matrix
- $V = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}$ is an $n \times n$ orthogonal matrix
- $\Sigma = \begin{pmatrix} \sigma_1 & & & 0 & \\ & \ddots & & & \\ & & \sigma_r & & \\ 0 & & & 0 & \\ & & & & \ddots \ 0 \end{pmatrix}$ is an $m \times n$ diagonal matrix.

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the singular values

Where did $u_{r+1}, \ldots, u_m$ and $v_{r+1}, \ldots, v_n$ come from??

They're orthonormal bases for the other two fundamental subspaces!

Col(A): $\{u_1, \ldots, u_r\}$ Nul(A^T): $\{u_{r+1}, \ldots, u_m\}$

Row(A): $\{v_1, \ldots, v_r\}$ Nul(A): $\{v_{r+1}, \ldots, v_n\}$

Procedure to Compute $A = U\Sigma V^T$:

(1) Compute the singular values and singular vectors

$\{v_1, \ldots, v_r\}$ $\{u_1, \ldots, u_r\}$ $\sigma_1, \ldots, \sigma_r$

as before (eigenvectors for nonzero eigenspaces of A^TA, AA^T)

(2) Find orthonormal bases

$\{u_{r+1}, \ldots, u_m\}$ for $\mathrm{Nul}(A^T) = \mathrm{Nul}(AA^T) =$ 0-eigenspace AA^T

$\{v_{r+1}, \ldots, v_n\}$ for $\mathrm{Nul}(A) = \mathrm{Nul}(A^TA) =$ 0-eigenspace A^TA

using Gram-Schmidt.

(3) $U = \begin{pmatrix} | & & | & | & & | \\ u_1 & \cdots & u_r & u_{r+1} & \cdots & u_m \\ | & & | & | & & | \end{pmatrix}$ $V = \begin{pmatrix} | & & | & | & & | \\ v_1 & \cdots & v_r & v_{r+1} & \cdots & v_n \\ | & & | & | & & | \end{pmatrix}$

$$\Sigma = \begin{pmatrix} \sigma_1 & & & 0 & \\ & \ddots & & & \\ & & \sigma_r & & \\ & 0 & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix}$$ (same size as A)

Proof: Use the outer product version of matrix mult:

$$U\Sigma V^T = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{pmatrix} \begin{pmatrix} - & v_1 & - \\ & \vdots & \\ - & v_n & - \end{pmatrix}$$

$$= \begin{pmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{pmatrix} \begin{pmatrix} - & \sigma_1 v_1 & - \\ & \vdots & \\ - & \sigma_r v_r & - \\ - & 0 & - \\ & \vdots & \\ - & 0 & - \end{pmatrix}$$

$$= \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T + 0 + \cdots + 0$$ ✓

Eg: $A = \begin{pmatrix} -10 & 10 & -10 & 10 \\ 10 & 5 & 10 & 5 \end{pmatrix}$

(1) $A = 15\sqrt{2}\, u_1 v_1^T + 10\sqrt{2}\, u_2 v_2^T$ for

$$u_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ -1 \end{pmatrix} \qquad v_1 = \frac{1}{\sqrt{10}}\begin{pmatrix} -2 \\ 1 \\ -2 \\ 1 \end{pmatrix}$$

$$u_2 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix} \qquad v_2 = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ 2 \\ 1 \\ 2 \end{pmatrix}$$

(2) $Nul(A^T) = \{0\}$ because $r = m$

$Nul(A)$: $\begin{pmatrix} -10 & 10 & -10 & 10 \\ 10 & 5 & 10 & 5 \end{pmatrix} \xrightarrow{\text{rref}} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$

$$\xrightarrow{\text{PVF}} \text{Span}\left\{ \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

[already orthogonal — usually need Gram-Schmidt] $\xrightarrow{\text{normalize}}$ $v_3 = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad v_4 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}$

(3) So $A = U\Sigma V^T$ for

$$U = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} 15\sqrt{2} & 0 & 0 & 0 \\ 0 & 10\sqrt{2} & 0 & 0 \end{pmatrix}$$

$$V = \begin{pmatrix} -2/\sqrt{10} & 1/\sqrt{10} & -1/\sqrt{2} & 0 \\ 1/\sqrt{10} & 2/\sqrt{10} & 0 & -1/\sqrt{2} \\ -2/\sqrt{10} & 1/\sqrt{10} & 1/\sqrt{2} & 0 \\ 1/\sqrt{10} & 2/\sqrt{10} & 0 & 1/\sqrt{2} \end{pmatrix}$$

NB: $A = U\Sigma V^T$ contains full orthogonal diagonalizations of A^TA and of AA^T:

$$\boxed{A^TA = V\begin{pmatrix}\sigma_1^2 & & & & 0\\ & \ddots & & & \\ & & \sigma_r^2 & & \\ & & & 0 & \\ 0 & & & & \ddots\; 0\end{pmatrix}V^T \qquad AA^T = U\begin{pmatrix}\sigma_1^2 & & & & 0\\ & \ddots & & & \\ & & \sigma_r^2 & & \\ & & & 0 & \\ 0 & & & & \ddots\; 0\end{pmatrix}U^T}$$

It also contains orthonormal bases for all four subspaces:

$$U = \begin{pmatrix} \overbrace{\begin{matrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{matrix}}^{\substack{\text{o.n. basis} \\ \text{for } \mathrm{Col}(A)}} & \overbrace{\begin{matrix} | & & | \\ u_{r+1} & \cdots & u_m \\ | & & | \end{matrix}}^{\substack{\text{o.n. basis} \\ \text{for } \mathrm{Nul}(A^T)}} \end{pmatrix}$$

$i \le r$ $\boxed{Av_i = \sigma_i u_i \uparrow \; \downarrow A^Tu_i = \sigma_i v_i}$ $\boxed{Av_i = 0 \uparrow \; \downarrow A^Tu_i = 0}$ $i > r$

$$V = \begin{pmatrix} \underbrace{\begin{matrix} | & & | \\ v_1 & \cdots & v_r \\ | & & | \end{matrix}}_{\substack{\text{o.n. basis} \\ \text{for } \mathrm{Row}(A)}} & \underbrace{\begin{matrix} | & & | \\ v_{r+1} & \cdots & v_n \\ | & & | \end{matrix}}_{\substack{\text{o.n. basis} \\ \text{for } \mathrm{Nul}(A)}} \end{pmatrix}$$

NB: SVD of A^T is

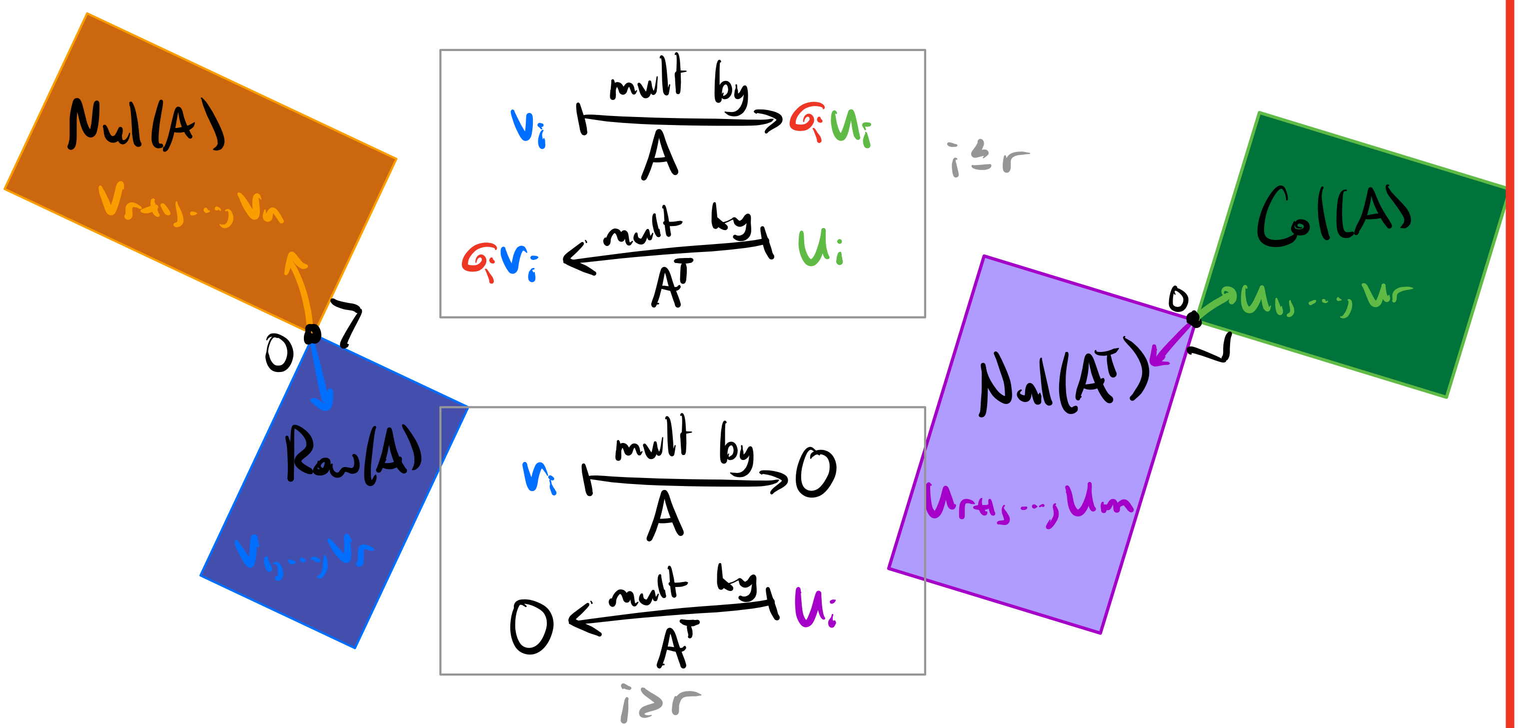
$$A^T = V\Sigma^T U^T$$

The Big Picture Revisited

for an $m \times n$ matrix A of rank r

Row Picture: $\mathbb{R}^n$

Column Picture: $\mathbb{R}^m$



Geometry of the SVD: Matrix Form

We have drawn pictures of a triple product decomposition before.

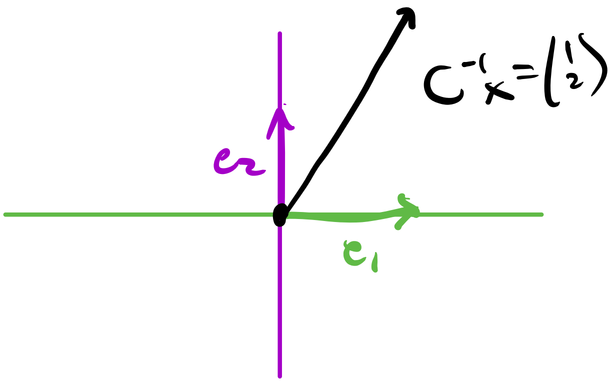
Diagonalization:

$$A = \frac{1}{10}\begin{pmatrix} 11 & 6 \\ 9 & 14 \end{pmatrix} = CDC^{-1}$$

$$\text{for } C = \begin{array}{cc} w_1 & w_2 \\ \end{array}\begin{pmatrix} 2 & -1 \\ 3 & 1 \end{pmatrix} \qquad D = \begin{array}{cc} \lambda_1 & \lambda_2 \\ \end{array}\begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

To evaluate $Ax = CDC^{-1}x$:

(1) multiply by C^{-1} (2) multiply by D (3) multiply by C



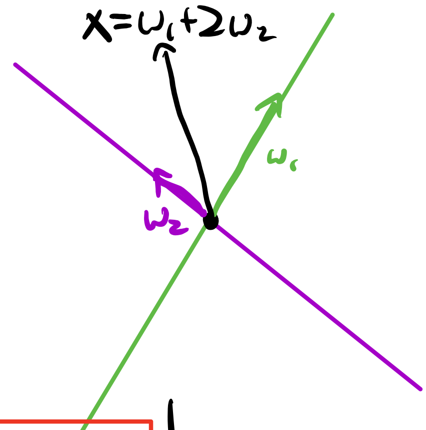
(1)

$\xleftarrow{C^{-1}}$

$C^{-1}w_1 = e_1$

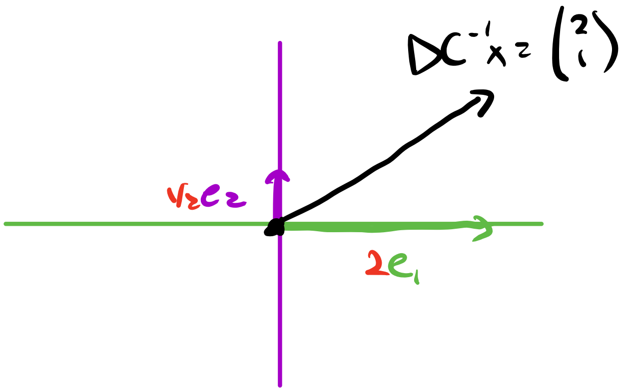
$C^{-1}w_2 = e_2$

$C^{-1}x = \begin{pmatrix}1\\2\end{pmatrix}$



$$\boxed{\begin{aligned} Aw_1 &= 2w_1 \\ Aw_2 &= \tfrac{1}{2}w_2 \end{aligned}}$$

(2) $\big\downarrow D$



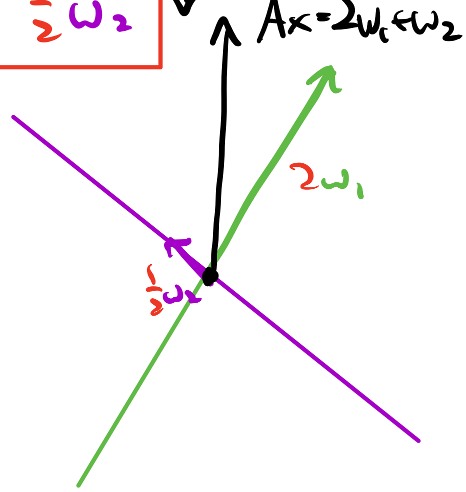
(3)

$\xrightarrow{C}$

$Ce_1 = w_1$

$Ce_2 = w_2$

$C\begin{pmatrix}2\\1\end{pmatrix} = 2w_1 + w_2$



SVD: $A = \begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix} = U\Sigma V^T$ for

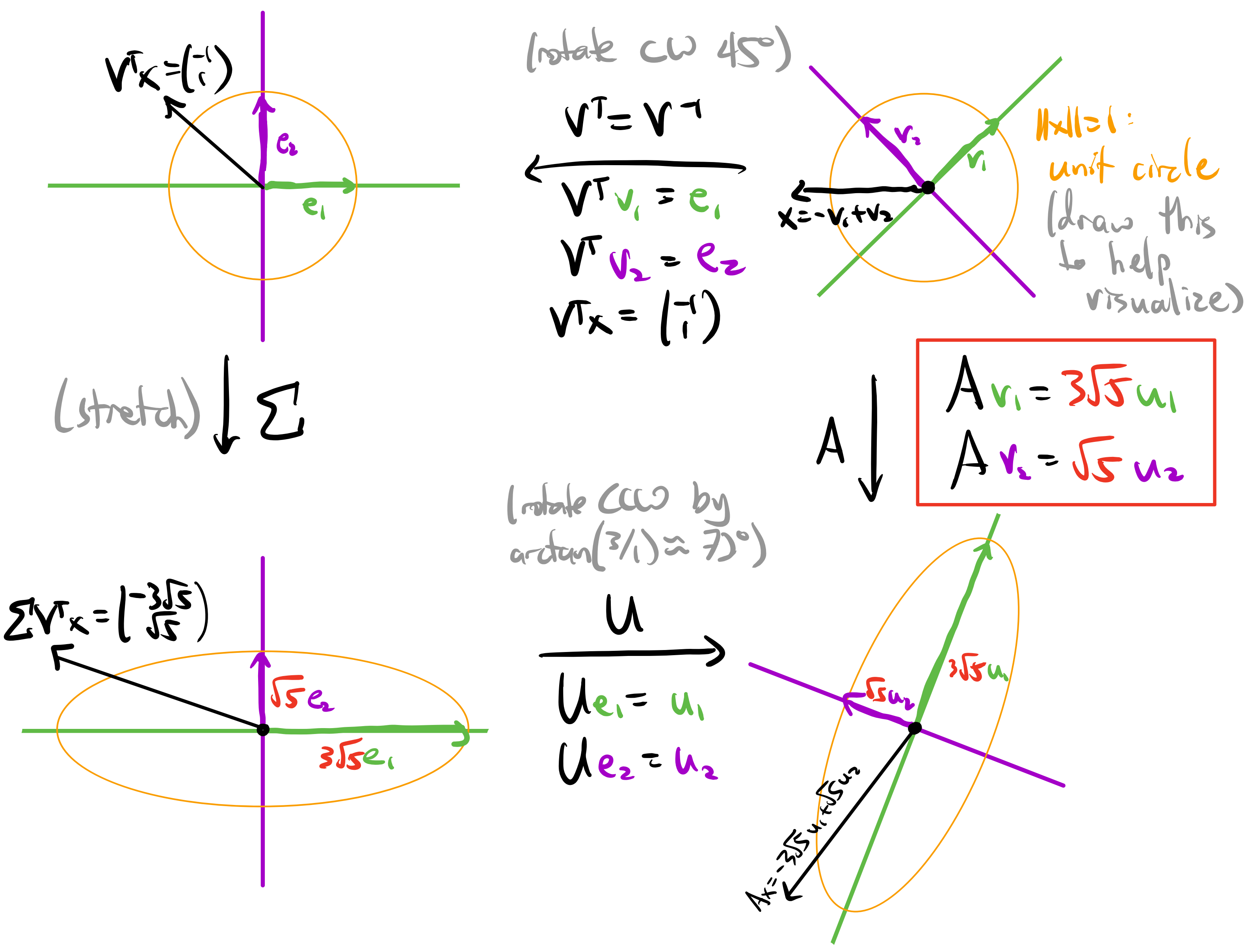
$$U = \frac{1}{\sqrt{10}} \begin{array}{c} \begin{array}{cc} u_1 & u_2 \end{array} \\ \begin{pmatrix} 1 & -3 \\ 3 & 1 \end{pmatrix} \end{array} \quad V = \frac{1}{\sqrt{2}} \begin{array}{c} \begin{array}{cc} v_1 & v_2 \end{array} \\ \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \end{array} \quad \Sigma = \begin{array}{c} \begin{array}{cc} \sigma_1 & \sigma_2 \end{array} \\ \begin{pmatrix} 3\sqrt{5} & 0 \\ 0 & \sqrt{5} \end{pmatrix} \end{array}$$

To evaluate $Ax = U\Sigma V^T x$:

(1) multiply by V^T (2) multiply by Σ (3) multiply by U

But U and V^T are orthogonal, so these just rotate/flip.

$Ax =$ (1) rotate/flip (2) stretch (3) rotate/flip

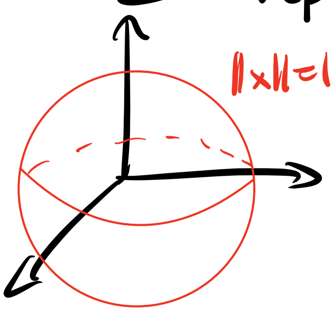


Notes / caveats:

- Diagonalization: start & end in $\{w_1, w_2\}$ basis
  SVD: start with $\{v_1, v_2\}$ & end with $\{u_1, u_2\}$ basis
  → Different bases!

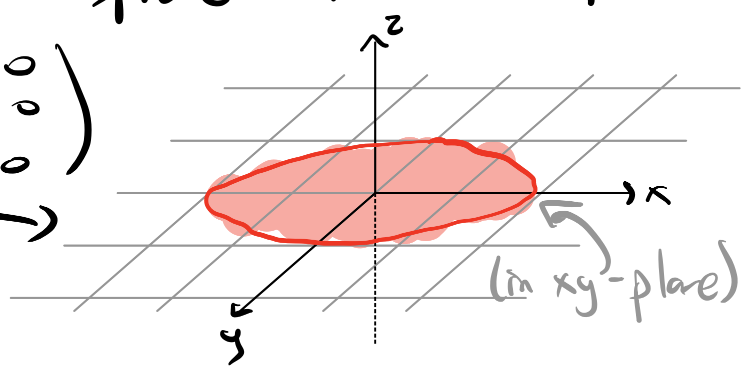
- The V^T & U steps preserve lengths & angles (rotations / flips) ⇝ easier to visualize.

- The Σ step can flatten a sphere in the same $\mathbb{R}^n$:



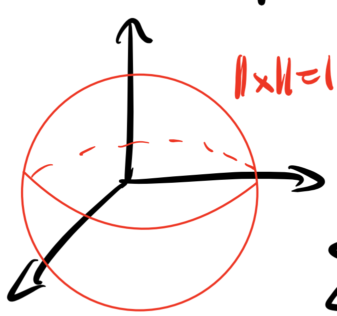
$$\Sigma = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\Sigma \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2x \\ y \\ 0 \end{pmatrix}$$



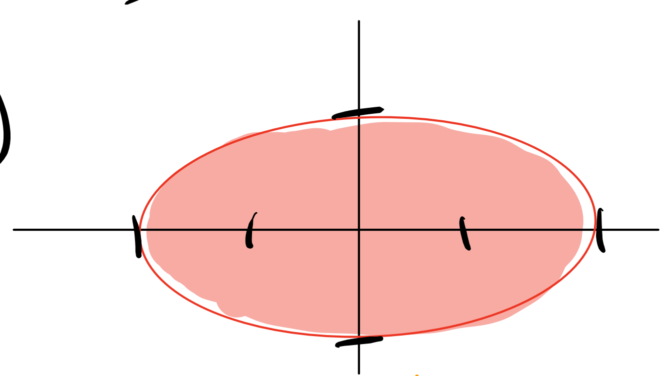
"project onto the xy-plane, then stretch"

- The Σ step can change dimensions:



$$\Sigma = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

$$\Sigma \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2x \\ y \end{pmatrix}$$



"project onto the xy-plane, forget the z-coordinate, then stretch"

Geometry of the SVD: Outer Product Form

Here is a geometric interpretation of the SVD that will be useful for the PCA. Let

$$A = \begin{pmatrix} | & & | \\ d_1 & \cdots & d_n \\ | & & | \end{pmatrix} \qquad \text{SVD} \qquad A = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T$$

$$\Rightarrow \quad A v_i = \sigma_i u_i \qquad A^T u_i = \sigma_i v_i$$

Expand out $A^T u_i = \sigma_i v_i$:

$$\sigma_i v_i = A^T u_i = \begin{pmatrix} - & d_1^T & - \\ & \vdots & \\ - & d_n^T & - \end{pmatrix} u_i = \begin{pmatrix} d_1 \cdot u_i \\ \vdots \\ d_n \cdot u_i \end{pmatrix}$$

$$\Rightarrow \sigma_i u_i v_i^T = u_i (\sigma_i v_i)^T = u_i \begin{pmatrix} d_1 \cdot u_i & \cdots & d_n \cdot u_i \end{pmatrix}$$

$$= \begin{pmatrix} | & & | \\ (d_1 \cdot u_i) u_i & \cdots & (d_n \cdot u_i) u_i \\ | & & | \end{pmatrix}$$

NB: $(d \cdot u_i) u_i =$ orthogonal projection of d onto $\mathrm{Span}\{u_i\}$ (since $u_i \cdot u_i = \|u_i\|^2 = 1$).

> The columns of $\sigma_i u_i v_i^T$ are the orthogonal projections of the columns of A onto $\mathrm{Span}\{u_i\}$.

Now look at the sum:

$$A = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T$$

The i^{th} column of this sum is:

$$\underset{i^{th}\text{ col of }A}{} \longrightarrow d_i = (d_i\cdot u_1)u_1 + \cdots + (d_i\cdot u_r)u_r$$

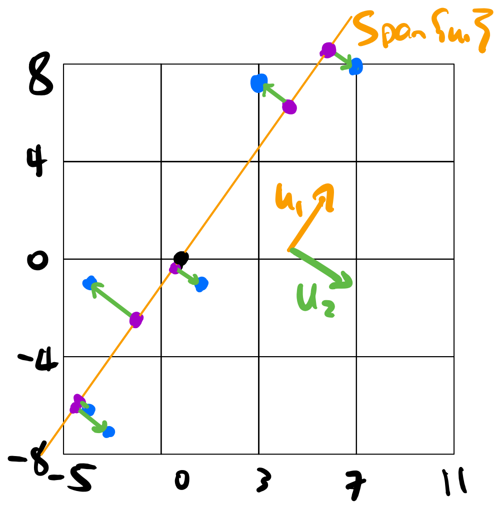
Since $\{u_1,\dots,u_r\}$ is an orthonormal basis of $\mathrm{Col}(A)$, this is just the projection formula as applied to d_i: the projection of d_i onto $\mathrm{Col}(A)$ is just d_i since $d_i \in \mathrm{Col}(A)$ (it is the i^{th} column of A).

Eg: $A = \begin{pmatrix} 3 & -4 & 7 & 1 & -4 & -3 \\ 7 & -6 & 8 & -1 & -1 & -7 \end{pmatrix}$ $\quad r=2$

$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T$

$\sigma_1 \approx 16.9 \qquad \sigma_2 \approx 3.92$

$u_1 \approx \begin{pmatrix} 0.561 \\ 0.828 \end{pmatrix} \qquad u_2 \approx \begin{pmatrix} 0.828 \\ -0.561 \end{pmatrix}$



• $= d_i = \begin{pmatrix} 3 \\ 7 \end{pmatrix}, \begin{pmatrix} -4 \\ -6 \end{pmatrix}, \dots$ (columns)

• $=$ columns of $\sigma_1 u_1 v_1^T$

$=$ projections of • onto / $= \mathrm{Span}\{u_1\}$

$\nwarrow$ $=$ columns of $\sigma_2 u_2 v_2^T$

$=$ projections of • onto \ $= \mathrm{Span}\{u_2\}$

NB: • $=$ • $+$ $\nwarrow$

So SVD "pulls apart" the columns of A in $u_1,\dots,u_r$-components