Review: the SVD, So Far

Last time, we covered the outer product form of SVD.

A: $m \times n$ matrix of rank r

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$$

- $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the singular values
- $\{v_1, v_2, \ldots, v_r\}$ are orthonormal vectors in $\mathbb{R}^n$.
    - These are the right singular vectors.
    - They form a basis for Row(A).
    - They are eigenvectors of A^TA:

$$A^TAv_i = \sigma_i^2 v_i$$

- $\{u_1, u_2, \ldots, u_r\}$ are orthonormal vectors in $\mathbb{R}^m$.
    - These are the left singular vectors.
    - They form a basis for Col(A).
    - They are eigenvectors of AA^T:

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

The singular vectors are related by

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

The SVD of A^T is

$$A^T = \sigma_1 v_1 u_1^T + \sigma_2 v_2 u_2^T + \cdots + \sigma_r u_r v_r^T$$

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

$$\underset{n\times n}{A^TA} \quad \text{vs} \quad \underset{m\times m}{AA^T} \leftarrow \text{smaller!}$$

It's much easier to compute eigenvalues & eigenvectors of AA^T in this case.

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

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

$$A^TA = \begin{pmatrix} 100 & -50 & 100 & -50 \\ -50 & 125 & -50 & 125 \\ 100 & -50 & 100 & -50 \\ -50 & 125 & -50 & 125 \end{pmatrix} \leftarrow \text{yikes!}$$

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

$$\lambda_1 = 450 \rightsquigarrow \sigma_1 = \sqrt{\lambda_1} = 15\sqrt{2} \qquad u_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ -1 \end{pmatrix} \qquad v_1 = \frac{1}{\sqrt{10}}\begin{pmatrix} -2 \\ 1 \\ -2 \\ 1 \end{pmatrix}$$

$$\lambda_2 = 200 \rightsquigarrow \sigma_2 = \sqrt{\lambda_2} = 10\sqrt{2} \qquad u_2 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix} \qquad v_2 = \frac{1}{\sqrt{10}}\begin{pmatrix} 2 \\ 1 \\ 2 \\ 1 \end{pmatrix}$$

These are right-singular vectors of A^T $\Rightarrow$ left-singular vectors of A.

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

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

SVD: Matrix Form

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

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

- U is an $m \times m$ orthogonal matrix.
- V is an $n \times n$ orthogonal matrix.
- Σ is an $m \times n$ diagonal matrix.

orthogonal: square with orthonormal columns

These matrices contain:

$$\Sigma = \begin{pmatrix} \sigma_1 & \ddots & & & 0 \\ & & \sigma_r & & \\ & & & 0 & \\ 0 & & & & \ddots & 0 \end{pmatrix}$$

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

$$U = \begin{pmatrix} | & & | & | & & | \\ \underbrace{u_1 \cdots u_r}_{\substack{\text{left singular} \\ \text{vectors:} \\ \text{Col}(A)}} & & & \underbrace{u_{r+1} \cdots u_m}_{\substack{\text{orthonormal} \\ \text{basis for} \\ \text{Nul}(A^T)}} \\ | & & | & | & & | \end{pmatrix} \qquad V = \begin{pmatrix} | & & | & | & & | \\ \underbrace{v_1 \cdots v_r}_{\substack{\text{right singular} \\ \text{vectors:} \\ \text{Row}(A)}} & & & \underbrace{v_{r+1} \cdots v_m}_{\substack{\text{orthonormal} \\ \text{basis for} \\ \text{Nul}(A)}} \\ | & & | & | & & | \end{pmatrix}$$

Recall: To compute the SVD of A, you compute orthonormal eigenvectors $v_1, v_2, \ldots, v_r$ for the nonzero eigenvalues of A^TA. The 0-eigenspace of A^TA is $\text{Nul}(A^TA) = \text{Nul}(A)$, so

$v_1, \ldots, v_r, v_{r+1}, \ldots, v_n$ is an orthonormal eigenbasis of A^TA.

Likewise,

$u_1, \ldots, u_r, u_{r+1}, \ldots, u_m$ is an orthonormal eigenbasis of AA^T.

In fact, if $A = U\Sigma V^T$ then

$$\overset{U^TU = I_m}{}$$

$$A^TA = (U\Sigma V^T)^T(U\Sigma V^T) = V\Sigma^T U^T U \Sigma V^T = V\Sigma^T\Sigma V^T$$

$$\Sigma^T\Sigma = \underset{n\times m}{\begin{pmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ 0 & & & & \ddots & 0 \end{pmatrix}} \underset{m\times n}{\begin{pmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ 0 & & & & \ddots & 0 \end{pmatrix}} = \underset{m\times m}{\begin{pmatrix} \sigma_1^2 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r^2 & & \\ & & & 0 & \\ 0 & & & & \ddots & 0 \end{pmatrix}}$$

Likewise, $AA^T = U\Sigma\Sigma^T U^T$.

same nonzero eigenvalues

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

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

(1) Compute the singular values & singular vectors

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 \qquad \{v_1, v_2, \ldots, v_r\} \qquad \{u_1, u_2, \ldots, u_r\}$$

as before.

(2) Compute orthonormal bases

$\{v_{r+1}, v_{r+2}, \ldots, v_n\}$ for $\mathrm{Nul}(A) = 0$-eigenspace of A^TA

$\{u_{r+1}, u_{r+2}, \ldots, u_m\}$ for $\mathrm{Nul}(A^T) = 0$-eigenspace of AA^T

probably using PVF and Gram–Schmidt.

(3) $$U = \begin{pmatrix} | & & | & | & & | \\ u_1 & \cdots & u_r & u_{r+1} & \cdots & u_m \\ | & & | & | & & | \end{pmatrix} \quad V = \begin{pmatrix} | & & | & | & & | \\ v_1 & \cdots & v_r & v_{r+1} & \cdots & v_m \\ | & & | & | & & | \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ 0 & & & & \ddots & 0 \end{pmatrix}$$

Proof:

$$U\Sigma V^T = \begin{pmatrix} | & & | & | & & | \\ u_1 & \cdots & u_r & u_{r+1} & \cdots & u_m \\ | & & | & | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & & & & & \\ & \ddots & & & 0 & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & 0 & & & \ddots & \\ & & & & & 0 \end{pmatrix} \begin{pmatrix} - & v_1^T & - \\ & \vdots & \\ - & v_r^T & - \\ - & v_{r+1}^T & - \\ & \vdots & \\ - & v_n^T & - \end{pmatrix}$$

$$= \begin{pmatrix} | & & | & | & & | \\ u_1 & \cdots & u_r & u_{r+1} & \cdots & u_m \\ | & & | & | & & | \end{pmatrix} \begin{pmatrix} - & \sigma_1 v_1^T & - \\ & \vdots & \\ - & \sigma_r v_r^T & - \\ - & 0 & - \\ & \vdots & \\ - & 0 & - \end{pmatrix}$$

outer product form

$$= \sigma_1 u_1 v_1^T + \sigma_2 u_2^T + \cdots + \sigma_r u_r v_r^T + 0 + \cdots + 0$$

$$= A$$

//

NB: We could have put any vectors we want in the last $m-r$ columns of U and the last $n-r$ columns of V and the product is still $A = U\Sigma V$ (since these vectors are multiplied by 0). But we really want U and V to have orthonormal columns. So $u_{r+1}, \ldots, u_m$ must be $\perp \operatorname{Span}\{u_1, \ldots, u_r\} = \operatorname{Col}(A)$, so they must be a basis for $\operatorname{Col}(A)^\perp = \operatorname{Nul}(A^T)$. Likewise for V.

NB: In matrix form,

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

is the SVD of A^T.

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

(1) $\sigma_1 = 15\sqrt{2} \qquad u_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ -1 \end{pmatrix} \qquad v_1 = \frac{1}{\sqrt{10}}\begin{pmatrix} -2 \\ 1 \\ -2 \\ 1 \end{pmatrix}$

$\sigma_2 = 10\sqrt{2} \qquad u_2 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix} \qquad v_2 = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ 2 \\ 1 \\ 2 \end{pmatrix}$

(2) In this case, $m = r = 2$ (full row rank), so $\mathrm{Nul}(A^T) = \{0\} \Rightarrow$ no more u_i's to compute.

But $n = 4$ and $r = 2 \Rightarrow \dim \mathrm{Nul}(A) = 2$, so we need to compute v_3 and v_4 (orthonormal basis for $\mathrm{Nul}(A)$).

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

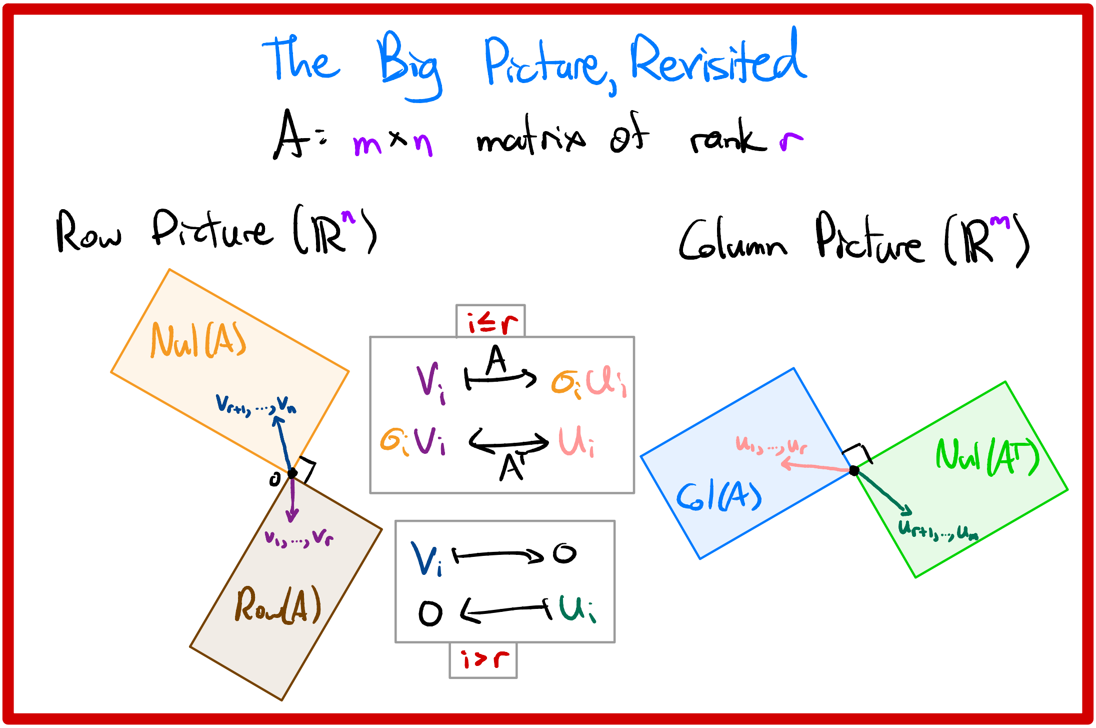
$$\xrightarrow{\text{PVF}} \text{basis } \left\{ \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

We got lucky that these vectors are already orthogonal — usually we'd have to do Gram-Schmidt. So just divide by the lengths:

$$v_3 = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \qquad v_4 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

(3) $U = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix} \quad V = \begin{pmatrix} -2/\sqrt{10} & 1/\sqrt{10} & -1/\sqrt{2} & 0 \\ 1/\sqrt{10} & 2/\sqrt{10} & 0 & -1/\sqrt{2} \\ -2/\sqrt{10} & 1/\sqrt{10} & 1/\sqrt{2} & 0 \\ 1/\sqrt{10} & 2/\sqrt{10} & 0 & 1/\sqrt{2} \end{pmatrix} \quad \Sigma = \begin{pmatrix} 15\sqrt{2} & 0 & 0 & 0 \\ 0 & 10\sqrt{2} & 0 & 0 \end{pmatrix}$

We can draw all this in the Big Picture:

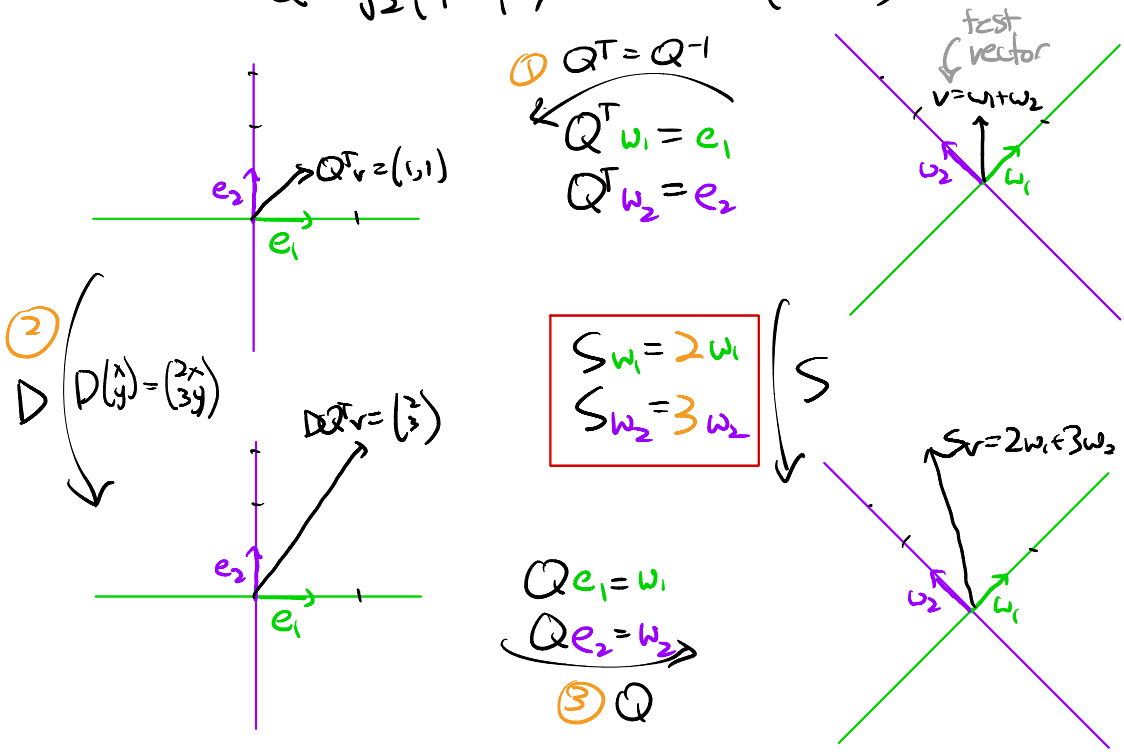


Geometry of the SVD, Matrix Form

We've drawn a picture of a triple product decomposition before.

Eg (L21): $S = \frac{1}{2}\begin{pmatrix} 5 & -1 \\ -1 & 5 \end{pmatrix} = QDQ^T$ for

$$Q = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \qquad D = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$



In the case of the SVD, $A = U\Sigma V^T$ means:

1. Multiply by V^T: orthogonal (rotate/flip)
2. Multiple by Σ: diagonal (stretch coordinates)
3. Multiply by U: orthogonal (rotate/flip)

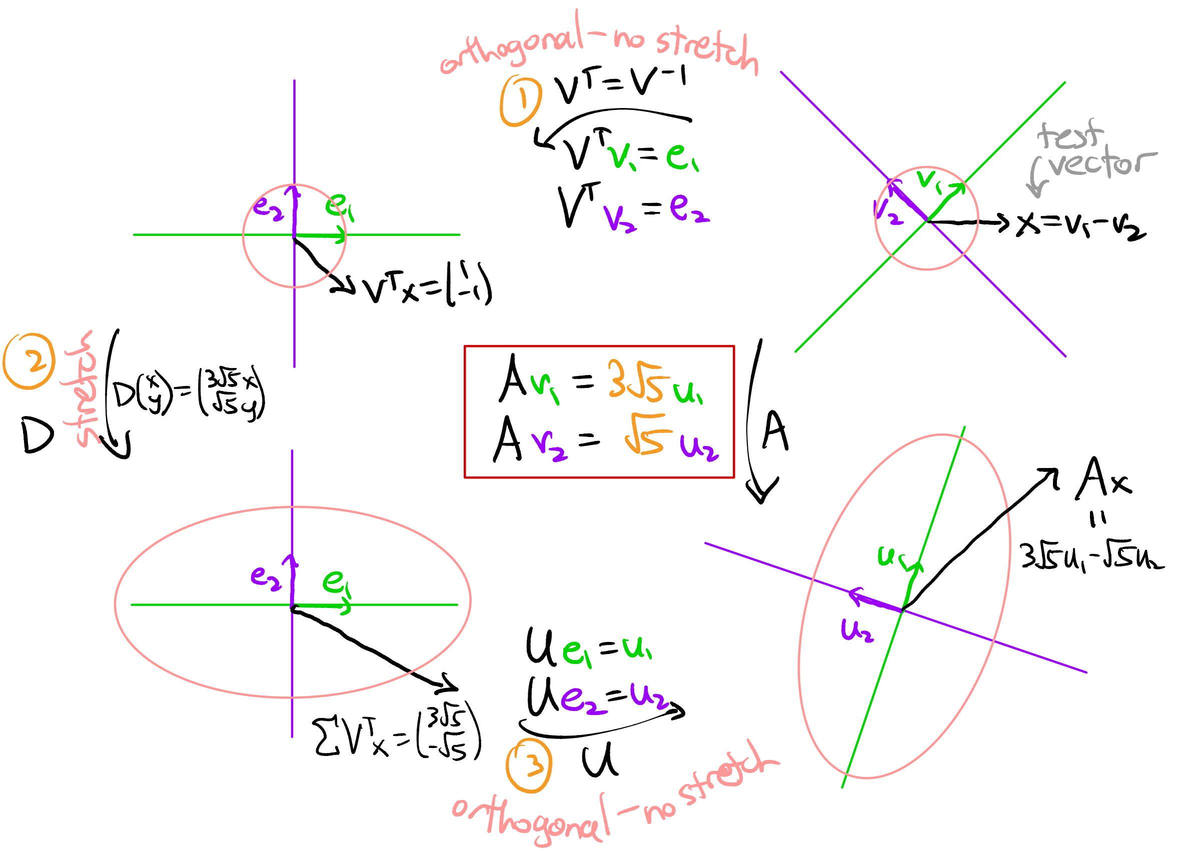
Eg: $A = \begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix} \overset{(L24)}{=} 3\sqrt{5}\, u_1 v_1^T + \sqrt{5}\, u_2 v_2^T$ where

$$u_1 = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ 3 \end{pmatrix} \quad v_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad u_2 = \frac{1}{\sqrt{10}}\begin{pmatrix} -3 \\ 1 \end{pmatrix} \quad v_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

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

$$U = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 & -3 \\ 3 & 1 \end{pmatrix} \quad V = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 3\sqrt{5} & 0 \\ 0 & \sqrt{5} \end{pmatrix}$$

Let's also draw the unit circle ○ to visualize stretching.



Reiterate: Any matrix can be expressed as:

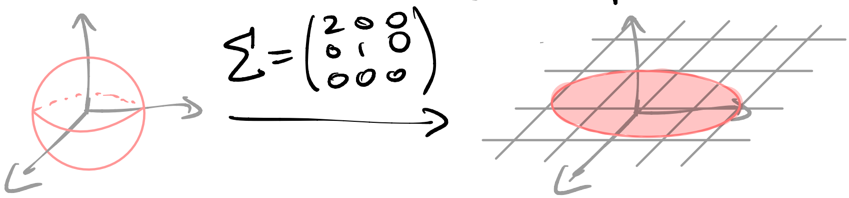
(rotate/flip) then (stretch) then (rotate/flip)

Notes/Caveats:

- In diagonalization, you start & end with the same basis $\{w_1, w_2, \ldots, w_n\}$ (eigenbasis).

  In the SVD, you start with the right singular vectors $\{v_1, v_2, \ldots, v_n\}$ and you end with the left singular vectors $\{u_1, u_2, \ldots, u_m\}$.

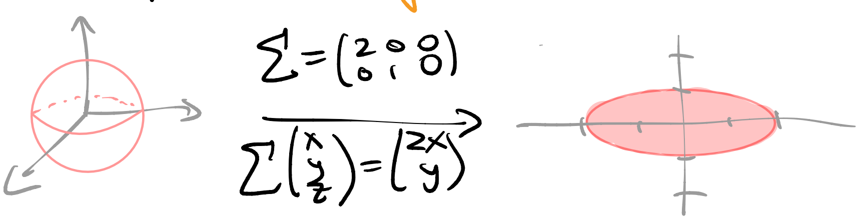
- The Σ step can flatten your sphere $\|x\|=1$:



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

(This can happen with a $S = QDQ^T$ decomposition too — that means 0 is an eigenvalue of S.)

- 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, then forget the z-coordinate."

Geometry of the SVD: Outer Product Form

This is the geometric interpretation that we will use throughout L26 & L27 when we do the PCA.

Give the columns of A a name:

$$A = \begin{pmatrix} | & & | \\ d_1 & \cdots & d_n \\ | & & | \end{pmatrix}$$

d_i = data points

SVD of A: $A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$

Recall: $Av_i = \sigma_i u_i \quad A^T u_i = \sigma_i v_i$

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

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

$$\Rightarrow \sigma_i u_i v_i^T = u_i (\sigma_i v_i)^T \overset{A^T u_i = \sigma_i v_i}{=} u_i (A^T u_i)^T$$

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

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

NB: $(d \cdot u_i) u_i$ is the orthogonal projection of d onto Span$\{u_i\}$: $\frac{d \cdot u_i}{u_i \cdot u_i} u_i = (d \cdot u_i) u_i$ because $\|u_i\| = 1$.

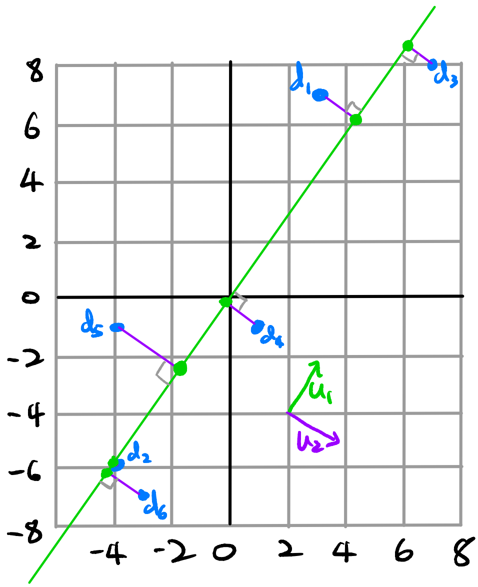
Upshot: The columns of $\sigma_i u_i v_i^T$ are the orthogonal projections of the columns of A onto $\mathrm{Span}\{u_i\}$.

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

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

$\sigma_1 \approx 16.9 \quad \sigma_2 \approx 3.92$

$u_1 \approx \begin{pmatrix} .561 \\ .828 \end{pmatrix} \quad u_2 \approx \begin{pmatrix} .828 \\ -.561 \end{pmatrix}$



LEGEND:

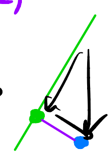
• $= d_i = \begin{pmatrix} 3 \\ 7 \end{pmatrix}, \begin{pmatrix} -4 \\ -6 \end{pmatrix}, \ldots$

• = columns of $\sigma_1 u_1 v_1^T$

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

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

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

NB: $A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T$ means • = • + \ vs 

The SVD "pulls apart" the columns of A into the u_1, u_2-directions: the principal components.

In fact, the i^{th} column of

$$\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$$

is $(d_i \cdot u_1)u_1 + (d_i \cdot u_2)u_2 + \cdots + (d_i \cdot u_r)u_r$

(projection formula) = orthogonal projection of d_i onto $Col(A) = Span\{u_1, u_2, \ldots, u_r\}$

$= d_i$ because $d_i \in Col(A)$ (it's the i^{th} column!)

This is just the i^{th} column of the SVD:

$$A = \begin{pmatrix} | & & | \\ d_1 & \cdots & d_n \\ | & & | \end{pmatrix} = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$$

More generally, for $k \leq n$, the i^{th} column of

$$\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_k u_k v_k^T$$

is equal to

$(d_i \cdot u_1)u_1 + (d_i \cdot u_2)u_2 + \cdots + (d_i \cdot u_k)u_k$

= orthogonal projection of d_i onto $Span\{u_1, u_2, \ldots, u_k\}$

Partial Sums are Projections

For $k \leq n$, the i^{th} column of

$$\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_k u_k v_k^T$$

is the orthogonal projection of d_i onto $Span\{u_1, u_2, \ldots, u_k\}$