We can also project onto a plane that does not go through the origin (i.e. a plane that is not a subspace of $\mathbb{R}^3$).

Such a plane has formula $\vec{n}^T(\vec{v}-\vec{v}_0)=0$

- Translate $\vec{v}_0$ to the origin so the plane goes through 0
- Project
- Translate the origin back to $\vec{v}_0$

$$\left[\begin{array}{c|c} I & \begin{matrix}0\\0\\0\end{matrix} \\ \hline -v_0 & 1\end{array}\right]\left[\begin{array}{c|c} I-\vec{n}\vec{n}^T & \begin{matrix}0\\0\\0\end{matrix} \\ \hline 0\ 0\ 0 & 1\end{array}\right]\left[\begin{array}{c|c} I & \begin{matrix}0\\0\\0\end{matrix} \\ \hline v_0 & 1\end{array}\right] = \text{Projection onto } \vec{n}^T(\vec{v}-\vec{v}_0)=0$$

Eg. Projecting onto $x+y+2z=1$.

$$\begin{bmatrix}1\\1\\2\end{bmatrix}^T\left(\begin{bmatrix}x\\y\\z\end{bmatrix}\right)=1 \qquad \vec{n}^T(\vec{v}-\vec{v}_0)$$

Choose $\vec{v}_0=\begin{bmatrix}1\\0\\0\end{bmatrix}$, get $\begin{bmatrix}1\\1\\2\end{bmatrix}^T\left(\begin{bmatrix}x\\y\\z\end{bmatrix}-\begin{bmatrix}1\\0\\0\end{bmatrix}\right)=0$

$$T_-PT_+ = \left[\begin{array}{ccc|c}1&0&0&0\\0&1&0&0\\0&0&1&0\\ \hline -1&0&0&1\end{array}\right]\left[\begin{array}{ccc|c}0&-1&-2&0\\-1&0&-2&0\\-2&-2&-3&0\\ \hline 0&0&0&1\end{array}\right]\left[\begin{array}{ccc|c}1&0&0&0\\0&1&0&0\\0&0&1&0\\ \hline 1&0&0&1\end{array}\right]$$

$$\vec{n}\vec{n}^T=\begin{bmatrix}1\\1\\2\end{bmatrix}\begin{bmatrix}1&1&2\end{bmatrix}=\begin{bmatrix}1&1&2\\1&1&2\\2&2&4\end{bmatrix}$$

Class 22: 11/27

Today: SVD (Singular Value Decomposition)

<u>Question</u>: To what extent can we diagonalize a non-square matrix?

When $A=S\Lambda S^{-1}$, we had $AS=S\Lambda$

so $A\vec{x}_1=\lambda_1\vec{x}_1$

$A\vec{x}_2=\lambda_2\vec{x}_2$ etc.

But that required A to be square.

For SVD, equations are

$$\begin{aligned}A\vec{v}_1&=\sigma_1\vec{u}_1\\ A\vec{v}_2&=\sigma_2\vec{u}_2\\ &\vdots\\ A\vec{v}_r&=\sigma_r\vec{u}_r\end{aligned}$$

(u's and v's are called <u>singular vectors</u> and σ's are <u>singular values</u>)

SVD: $\boxed{A = U\Sigma V^T}$ where U, V are orthogonal, Σ diagonal.

(So $AV = U\Sigma$) We make Σ have $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$.

Note: this says $\boxed{A = \vec{u}_1 \sigma_1 \vec{v}_1^T + \dots + \vec{u}_r \sigma_r \vec{v}_r^T}$

Two issues:

1. How do we find U, Σ, V?
2. Why do we care?

1. To find U, Σ, V, use that if $A = U\Sigma V^T$, then

$$A^TA = V\Sigma U^T U\Sigma V^T = V\Sigma^2 V^T$$

and $$AA^T = U\Sigma V^T V\Sigma U^T = U\Sigma^2 U^T$$

So diagonalize A^TA to find V and Σ^2

Then we can diagonalize AA^T to find U

<u>or</u> use $AV = U\Sigma$ to find U.

(Note: A^TA is symmetric, so V is orthogonal)

<u>Example</u>:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

$$A^TA = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

A^TA's eigenvalues come from $\det(A^TA - \lambda I) = \det \begin{bmatrix} 1-\lambda & 0 & 0 \\ 0 & 1-\lambda & 1 \\ 0 & 1 & 1-\lambda \end{bmatrix}$

$$= (1-\lambda) \begin{vmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{vmatrix} = (1-\lambda)\left[(1-\lambda)^2 - 1\right]$$

$$= (1-\lambda)(-2\lambda + \lambda^2) = (1-\lambda)\lambda(\lambda - 2)$$

So $A^TA = V \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} V^T$ and $\Sigma = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

Find V: $\lambda_1 = 2$: $\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{bmatrix} \vec{x} = 0 \rightsquigarrow \vec{x} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

$\lambda_2 = 1$: $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \vec{x} = 0 \rightsquigarrow \vec{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

$\lambda_3 = 0$: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \vec{x} = 0 \rightsquigarrow \vec{x} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$

So $V = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix}$

Now $AV = U\Sigma$, so $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} = U \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

$$\begin{bmatrix} 0 & 1 & 0 \\ \frac{2}{\sqrt{2}} & 0 & 0 \end{bmatrix} = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{2}\,\vec{u}_1 & \vec{u}_2 & 0 \end{bmatrix}$$

So $\vec{u}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\vec{u}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

2. Why do we care?

Two reasons.

1) Easily read off bases for 4 fundamental subspaces
2) Data compression.

1) <u>Fact:</u> First r columns of V are a basis for the row space of A
Last $n-r$ columns of V are a basis for the nullspace of A
First r columns of U are a basis for the column space of A
Last $m-r$ columns of U are a basis for the left nullspace of A

<u>Why?</u> Use $AV = U\Sigma$.

If A is $m \times n$, this says that A takes an orthonormal basis of $\mathbb{R}^n$ to an <u>orthogonal</u> basis of $\mathbb{R}^m$.

$$A \begin{bmatrix} | & | & & | \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_m \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots\,\sigma_r & \\ & & & \end{bmatrix}$$

$n-r$ columns of 0's.

$m-r$ rows of 0's

So $A\vec{v}_1, A\vec{v}_2, \dots A\vec{v}_r$ (which are $\vec{u}_1\sigma_1, \vec{u}_2\sigma_2, \dots \vec{u}_r\sigma_r$) form a basis for $C(A)$. $A\vec{v}_{r+1}, \dots A\vec{v}_n = 0$, so they don't contribute.

* First r columns of U form an orthonormal basis for $C(A)$.
But the left nullspace of A is $C(A)^\perp$, so the last $m-r$ columns of U form an orthonormal basis for the left nullspace.
(Repeat argument with A^T to get results for V)

2) Data compression.

An image is made of pixels (say 2^9 by 2^{10} pixels). Each pixel has a grayscale value. You can keep track of this with a $2^9 \times 2^{10}$ matrix. This has 2^{19} entries. No big deal!

Until you start making movies with 30 images per second, or doing a CT scan which has tons of cross section images. Now there are lots of $2^9 \times 2^{10}$ matrices to keep track of.

Options: 1) Bigger pixels. Double pixel height + width and reduce the number of pixels by a factor of 4.
Downside: $\frac{1}{4}$ of 2^{19} (times many images per second) is still huge, and picture quality decreases.

2) Use SVD! If A is your pixel matrix, approximate A using the largest singular values
$A \approx \sigma_1 \vec{u}_1 \vec{v}_1^T$ ($\vec{u}_1$ has $\underset{2^9}{m}$ entries, $\vec{v}_1$ has $\underset{2^{10}}{n}$ entries)

So this means we only need to keep track of $2^9 + 2^{10}$ numbers!

In reality, use the first few singular values (until the approx. is pretty good).

Say: $A \approx \sigma_1 \vec{u}_1 \vec{v}_1^T + \sigma_2 \vec{u}_2 \vec{v}_2^T + \sigma_3 \vec{u}_3 \vec{v}_3^T$

Keep track of $3(2^9 + 2^{10})$ numbers.
Way less than 2^{19}!