Last time

- Orthogonal diagonalization
- Spectral theorem

Our next application is based on the following innocent seeming observations.

Observation 1 For *any* matrix A, $A^T A$ is symmetric.

Indeed, $(A^T A)^T = A^T (A^T)^T = A^T A$.

<u>Observation 2</u> Given orthogonal eigenvectors $\vec{v}_1$ and $\vec{v}_2$ of A^TA, $A\vec{v}_1$ and $A\vec{v}_2$ are <u>still</u> orthogonal.

Indeed,
$$
\begin{aligned}
A\vec{v}_1 \cdot A\vec{v}_2 &= (A\vec{v}_1)^T A\vec{v}_2 \\
&= \vec{v}_1^T A^T A\vec{v}_2 \\
&= \vec{v}_1 \cdot A^T A\vec{v}_2 \\
&= \vec{v}_1 \cdot \lambda_2 \vec{v}_2 \\
&= \lambda_2 (\vec{v}_1 \cdot \vec{v}_2) \\
&= 0.
\end{aligned}
$$

$\underline{\text{Ex}}$ $A = \begin{bmatrix} 6 & 2 \\ -7 & 6 \end{bmatrix}$, $A^TA = \begin{bmatrix} 6 & -7 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 6 & 2 \\ -7 & 6 \end{bmatrix} = \begin{bmatrix} 85 & -30 \\ -30 & 40 \end{bmatrix}$

$$f_{A^TA}(\lambda) = \lambda^2 - 125\lambda + 2500$$

$$= (\lambda - 100)(\lambda - 25) \qquad \lambda = 25, 100$$

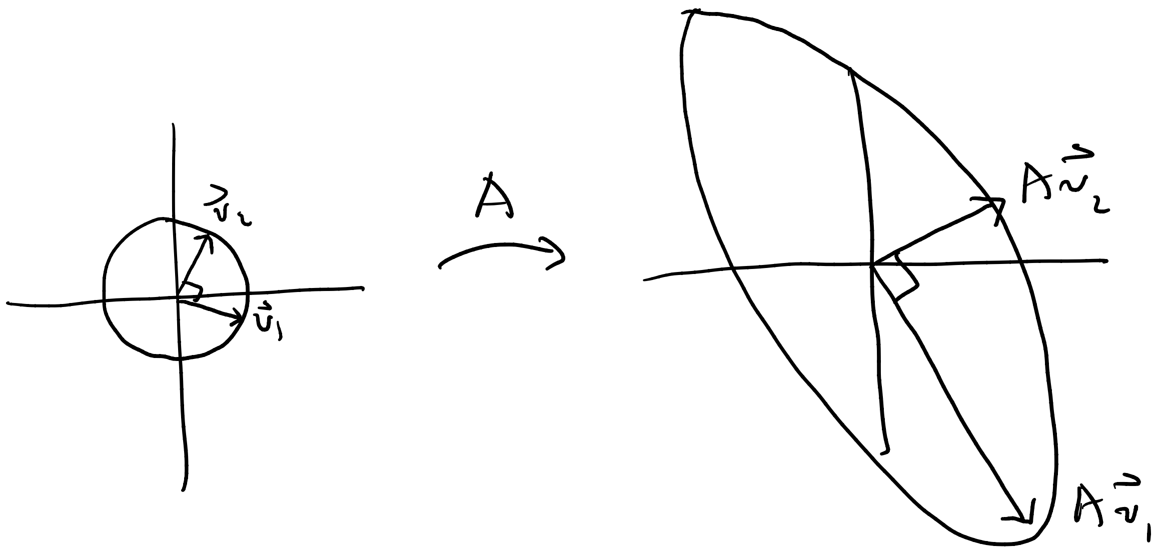
$$E_{100} = \ker \begin{bmatrix} -15 & -30 \\ -30 & -60 \end{bmatrix} = \ker \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} = \operatorname{Span} \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

$$E_{25} = \ker \begin{bmatrix} 60 & -30 \\ -30 & 15 \end{bmatrix} = \ker \begin{bmatrix} 2 & -1 \\ 0 & 0 \end{bmatrix} = \operatorname{Span} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\vec{v}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ -1 \end{bmatrix} \qquad \vec{v}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$A\vec{v}_1 = \begin{bmatrix} 6 & 2 \\ -7 & 6 \end{bmatrix} \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 10 \\ -20 \end{bmatrix}$$

$$A\vec{v}_2 = \begin{bmatrix} 6 & 2 \\ -7 & 6 \end{bmatrix} \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 10 \\ 5 \end{bmatrix}$$



The effect of A is to transform the unit circle into an ellipse. The lengths of the axes of the ellipse are

$$|A\vec{v}_1| = \sqrt{\frac{100}{5} + \frac{400}{5}} = 10 = \sqrt{\lambda_1}$$

$$|A\vec{v}_2| = \sqrt{\frac{100}{5} + \frac{25}{5}} = 5 = \sqrt{\lambda_2}.$$

<u>Def</u> The <u>singular values</u> of the $m \times n$ matrix A are the square roots of the eigenvalues of A^TA. We denote them

$\sigma_1, \ldots, \sigma_n$ and list them in descending order:

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0.$$

For any $m \times n$ matrix A, there is a basis $\{\vec{v}_1, \ldots, \vec{v}_n\}$ such that

(1) $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is orthonormal

(2) $\{A\vec{v}_1, \ldots, A\vec{v}_n\}$ is orthogonal

(3) $|A\vec{v}_i| = \sigma_i$.

Ex $A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$, $A^TA = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$

$$f_{A^TA}(\lambda) = \det \begin{bmatrix} 1-\lambda & 1 & 0 \\ 1 & 2-\lambda & 1 \\ 0 & 1 & 1-\lambda \end{bmatrix}$$

$$= (1-\lambda)(\lambda^2 - 3\lambda + 1) - (1-\lambda)$$

$$= -\lambda^3 + 3\lambda^2 + \lambda^2 - \lambda - 3\lambda + 1 - 1 + \lambda$$

$$= -\lambda^3 + 4\lambda^2 - 3\lambda$$

$$= -\lambda(\lambda^2 - 4\lambda + 3)$$

$$= -\lambda(\lambda - 3)(\lambda - 1) \qquad \lambda = 0, 1, 3$$

So the singular values of A are

$$\sigma_1 = \sqrt{3}$$

$$\sigma_2 = 1$$

$$\sigma_3 = 0$$

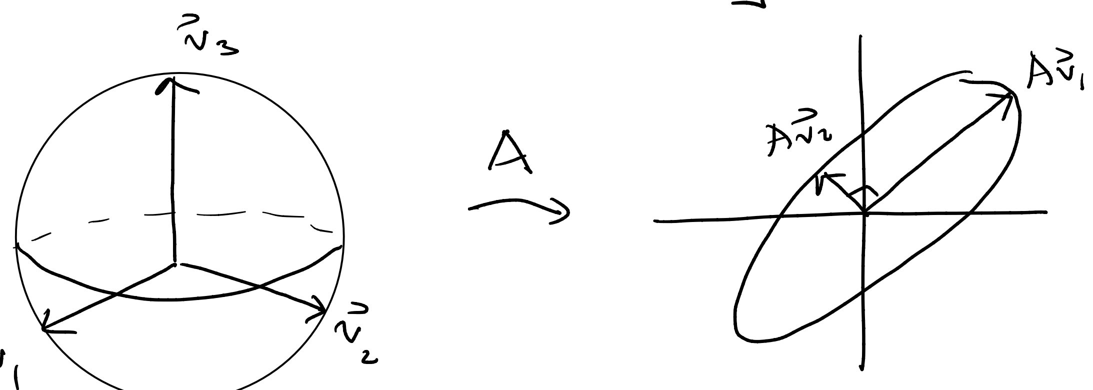
$$E_3 = \ker \begin{bmatrix} -2 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 1 & -2 \end{bmatrix} = \ker \begin{bmatrix} 1 & -1 & 1 \\ 0 & -1 & 2 \\ 0 & 1 & -2 \end{bmatrix}$$

$$= \ker \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix} = \operatorname{Span} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

$$E_1 = \ker \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \ker \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \operatorname{Span} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

$$E_1 = \ker \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \ker \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \operatorname{Span} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

$$\vec{v}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \quad \vec{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad \vec{v}_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}.$$



The example shows that, in general, the image of A has dimension equal to the number of nonzero singular values.

$$\boxed{\begin{aligned} \operatorname{rank}(A) &= \#\text{ of } \sigma_i \neq 0 \\ &= \#\text{ of } \lambda_i \neq 0 \\ &= \operatorname{rank}(A^T A) \end{aligned}}$$

As we've seen, $\{A\vec{v}_1, \dots, A\vec{v}_n\}$ is orthogonal, and $|A\vec{v}_i| = \sigma_i$. If A has rank r, then $A\vec{v}_i \neq 0$, so $\{\vec{u}_1, \dots, \vec{u}_r\}$ is <u>orthonormal</u>, where

$$\vec{u}_i = \frac{1}{\sigma_i} \vec{v}_i$$

Expanding this to an ONB for $\mathbb{R}^m$, we have the equations

$$A\vec{v}_i = \begin{cases} \sigma_i \vec{u}_i & i = 1, \dots, r \\ 0 & i = r+1, \dots, m \end{cases}$$

or $AV = U\Sigma$, where

$$V = [\vec{v}_1 \cdots \vec{v}_n] \qquad U = [\vec{u}_1 \cdots \vec{u}_m]$$

$$\Sigma = \left.\begin{bmatrix} \begin{matrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{matrix} & \Big| & 0 \\ \hline 0 & & 0 \end{bmatrix}\right\} m$$

$$\underbrace{\qquad\qquad\qquad}_{n}$$

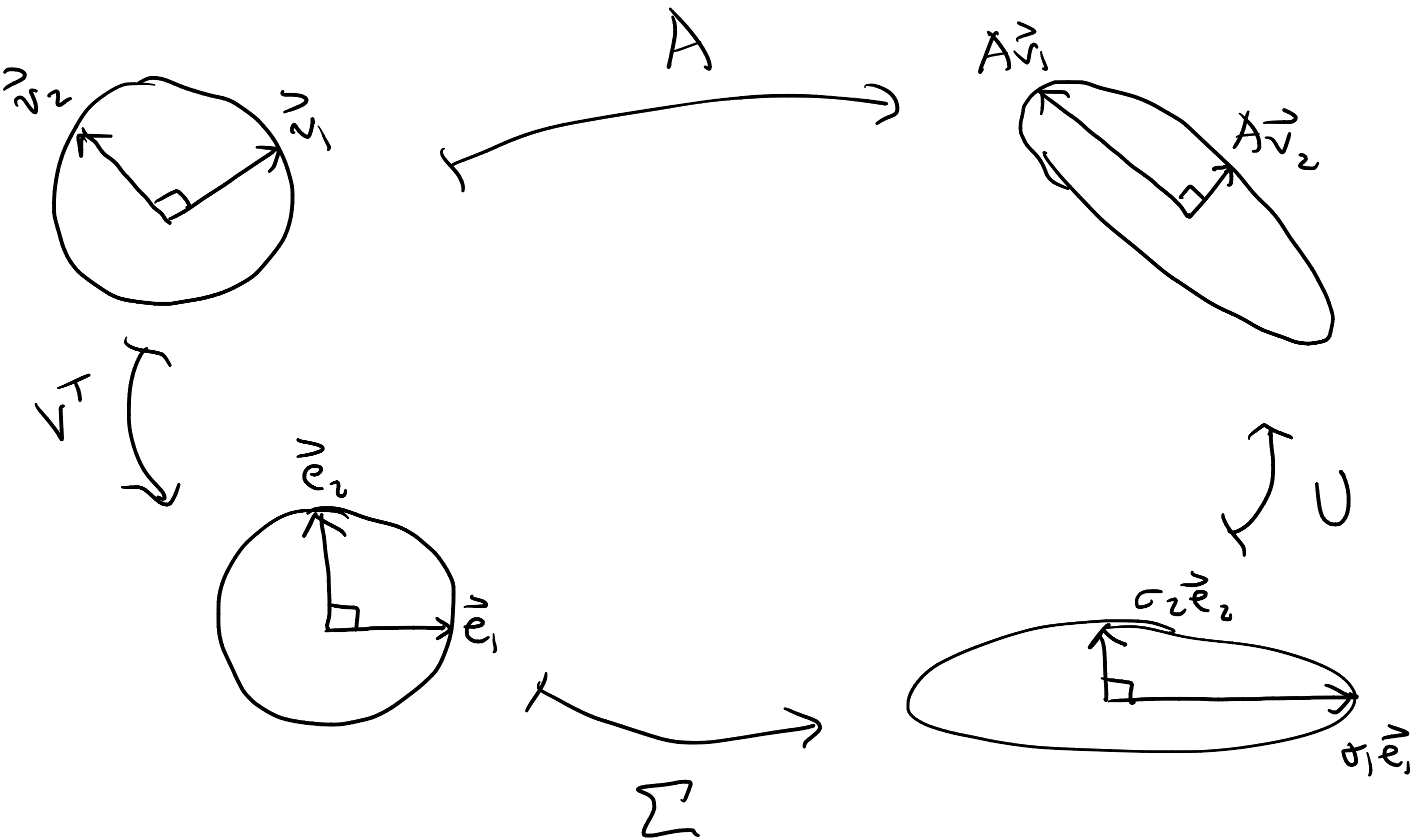
The matrices U and V are both orthogonal, so this equation becomes

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

<u>Singular value decomposition</u> An $m \times n$ matrix A of rank r can be written as

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

where U is $m \times m$ orthogonal, V is $n \times n$ orthogonal, and Σ is $m \times n$ with the first r diagonal entries positive and all other entries 0.


A
$\vec{v}_2$
$\vec{v}_1$
$A\vec{v}_1$
$A\vec{v}_2$
V^T
U
$\vec{e}_2$
$\vec{e}_1$
$\sigma_2\vec{e}_2$
$\sigma_1\vec{e}_1$
Σ

To find SVD for A:

(1) Diagonalize A^TA to obtain ONEB $\{\vec{v}_1, \ldots, \vec{v}_n\}$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$.

(2) Set $\sigma_i = \sqrt{\lambda_i}$ and $\vec{u}_i = \frac{1}{\sigma_i} A\vec{v}_i$ for $i = 1, \ldots, r$.

(3) Expand $\{\vec{u}_1, \ldots, \vec{u}_r\}$ to an ONB $\{\vec{u}_1, \ldots, \vec{u}_m\}$.

(4) $V = [\vec{v}_1 \cdots \vec{v}_n]$ $\quad \Sigma = \left[\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right]$

$U = [\vec{u}_1 \cdots \vec{u}_m]$