# Lecture 19

SVD decomposition intro

Q. Consider the unit circle in $\mathbb{R}^2$:

$$S = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x^2 + y^2 = 1 \right\}$$

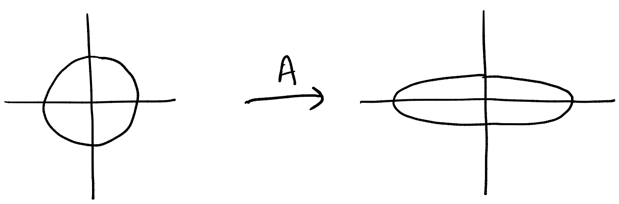
What happens to $S$ under the linear transformations

(a) $A = \begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix}$ (b) $B = \begin{bmatrix} \sqrt{3}/2 & 1/2 \\ -1/2 & \sqrt{3}/2 \end{bmatrix}$ ?

Try to draw a picture of the image of $S$ under these transformations.

(a) We can easily see the eigenvalues $\lambda_1 = 2$, $\lambda_2 = 1/2$

and corresponding eigenvectors $v_1 = e_1$, $v_2 = e_2$.




$Ae_1 = 2e_1$

$Ae_2 = \frac{1}{2} e_2$

(b) This is a planar rotation; the circle just rotates, so the image looks the same (but is actually rotated).

<u>Recall</u> $A = QR \rightsquigarrow Q = [q_1 \dots q_n]$, and $\{q_1, \dots, q_m\}$ is an ONB for the codomain of $A$.

<u>Today:</u> If we give an ONB for both the domain and codomain, we can further simplify the core information in the matrix.

"Core information": What directions are being stretched / squished, and by how much?

<u>SVD = Singular-Value Decomposition</u>

<u>thm 6.1</u> (SVD) $A \in \mathbb{R}^{m \times n}$ $(m \geq n)$ can be decomposed as

$$A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T \qquad \text{where}$$

- $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$ are orthogonal
- $\Sigma \in \mathbb{R}^{n \times n}$ is diagonal $\Sigma = \operatorname{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$

  and moreover, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$

=

As linear trans: $\quad U^T A V = \left(\begin{array}{c} \sigma_1 \ddots \sigma_n \\ \hline 0 \end{array}\right)$

$$\begin{tikzcd} \mathbb{R}^m \arrow[r, "\text{pre-}"', "\text{transformation}"'] \arrow[rrr, bend left, "\begin{pmatrix} \Sigma \\ 0 \end{pmatrix}"] & \mathbb{R}^m \arrow[r, "A"] & \mathbb{R}^n \arrow[r, "\text{post-}"', "\text{transformation}"'] & \mathbb{R}^n \end{tikzcd}$$

We understand how $\begin{pmatrix} \Sigma \\ 0 \end{pmatrix}$ transforms $\mathbb{R}^m \longrightarrow \mathbb{R}^n$ much better!

=

Note that by reordering bases (pre/post transform using permutation matrices, which are orthogonal), we achieve this ordering of $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$.

We made them all positive by doing basis changes $e_i \leftrightarrow -e_i$ (for example) as part of $U, V^T$.

## Terminology

eg. $A \in \mathbb{R}^{3\times 2}$

$$A = U\begin{pmatrix}\Sigma \\ 0\end{pmatrix}V^T$$

$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} = \begin{bmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{bmatrix}\begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ \hline 0 & 0 \end{bmatrix}\begin{bmatrix} -v_1^T- \\ -v_2^T- \end{bmatrix}$$

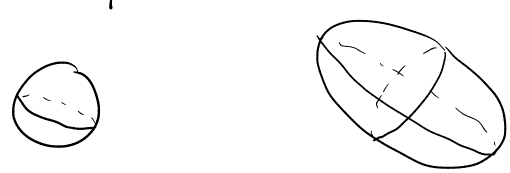
- $\{u_j\}$, $\{v_j\}$ are called (left / right) <u>singular vectors</u>
- $\{\sigma_j\}$ are the <u>singular values</u> (generalization of eigenvalues) (we'll say more later)

⚠ often, we don't consider "0" a singular value.

## What are the singular values?

As in the geometric examples at the beginning of class, the $\sigma_i$ tell you how much the unit ball is stretched or shrunk in a set of orthogonal directions

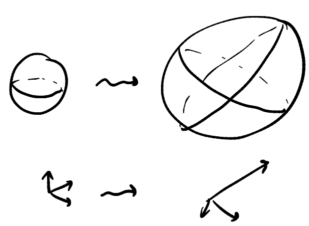
eg. cantaloupe ⇝ watermelon.



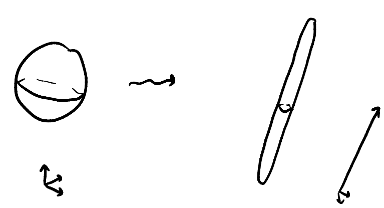
maybe: $\sigma_1 = 2$, $\sigma_2 = \sigma_3 = 1$.

Our system is well-behaved when the $\sigma_i$ are all comparable in size. Compare:

1



2



## Relation to Eigenvalues

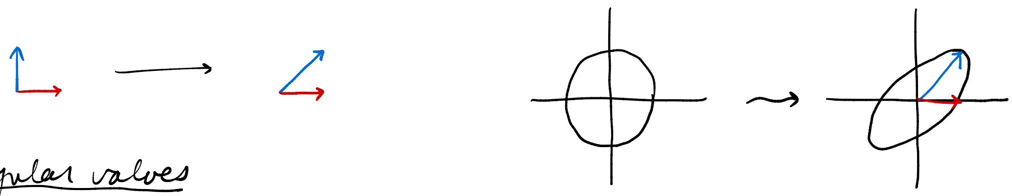
Will study more carefully later, but lets get a sense of the diff via some examples

eg. $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$

Eigenvalues

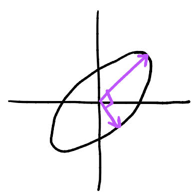
$\det(A - \lambda I) = \det\begin{pmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{pmatrix} = (1-\lambda)^2 \rightsquigarrow \lambda = 1$ (multiplicity 2)

Eigenvector $v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$



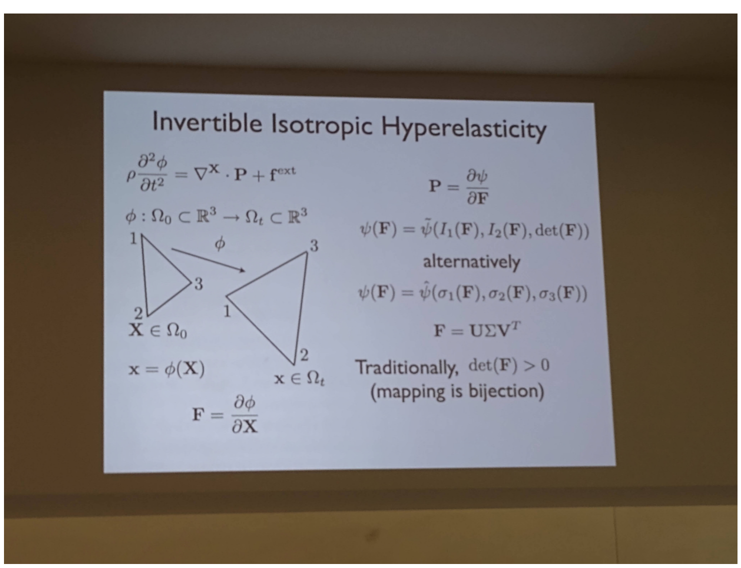
singular values

Fact $\Sigma = \begin{pmatrix} \sqrt{\frac{3+\sqrt{5}}{2}} & 0 \\ 0 & \sqrt{\frac{3-\sqrt{5}}{2}} \end{pmatrix}$



$\Rightarrow$ In applications, we often care about these geometric aspects more.

eg. CGI (at a talk by Prof. Joseph Teran @ conf last wknd)



## Remarks

① We of course also have thin SVD: $\boxed{A = \hat{U}\,\Sigma V^T}$

eg.

$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} = \begin{bmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ \hline 0 & 0 \end{bmatrix} \begin{bmatrix} - v_1^T - \\ - v_2^T - \end{bmatrix}$$

$u_3$ only gets multiplied by 0s!

In general:

$$\begin{bmatrix} A \end{bmatrix} = \left[\begin{array}{c|c} \hat{U} & U_0 \end{array}\right] \begin{bmatrix} \Sigma \\ \hline 0 \end{bmatrix} \begin{bmatrix} V^T \end{bmatrix} \Rightarrow \boxed{A = \hat{U}\,\Sigma V^T}$$

② Note also if $A \in \mathbb{R}^{m \times n}$ and $m \leq n$, we can still do SVD!

Treats cols/rows the same.

(Transpose A first, then apply the SVD theorem)

$$\begin{bmatrix} A \end{bmatrix} = \begin{bmatrix} U \end{bmatrix} \left[\begin{array}{c|c} \Sigma & 0 \end{array}\right] \begin{bmatrix} V^T \end{bmatrix}$$

---

③ Rank 1 decomposition $\rightsquigarrow$ toward low rank approx (later)

Outer product form:

$$\underbrace{(u_1 \cdots\cdots u_n)}_{\hat{U}} \underbrace{\begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix}}_{\Sigma} \underbrace{\begin{pmatrix} v_1^T \\ \vdots \\ v_n^T \end{pmatrix}}_{V^T} = \sum_{i=1}^{n} \overset{\#!}{\sigma_i}\, u_i v_i^T$$