Singular Value Decomposition

Note Title

LECTURE 13

- SVD is a matrix factorization that is useful for many applications, e.g., search engines, LS problems, tomographic image reconstruction, ...

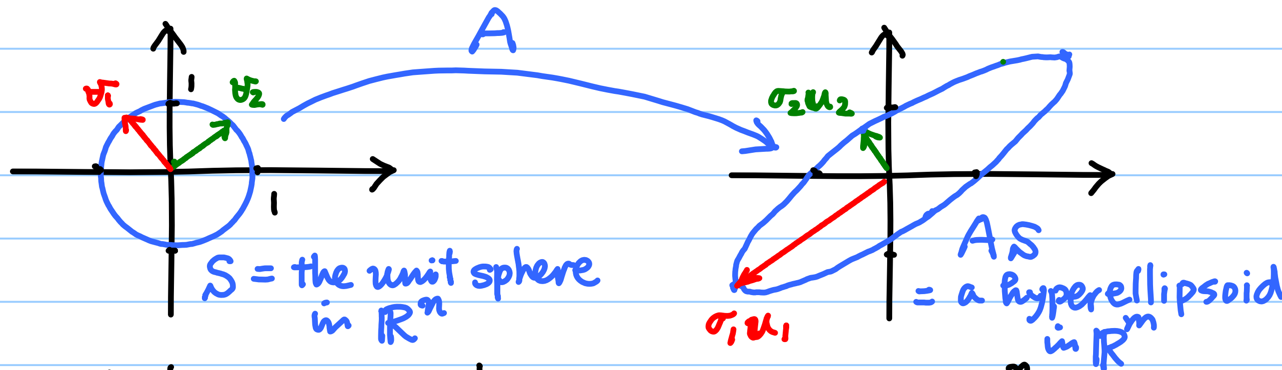
- SVD can be a concepual tool in linear algebra
  $\Rightarrow$ via SVD, we can check:
  - a given matrix is near singular
  - rank of the matrix
  - etc.

- $\exists$ a numerically stable algorithm to compute the SVD of a given matrix (it's expensive though...)
  In fact, one of the hottest topics in numerical linear algebra is how to compute a good approximation to the SVD of a huge matrix fast!

☆ A Geometric Observation

Let $A \in \mathbb{R}^{m \times n}$, and consider how A maps an input vector in $\mathbb{R}^n$ to an output vector in $\mathbb{R}^m$.

"The image of the unit sphere under any $m \times n$ matrix is a hyperellipsoid"



ONB = orthonormal basis

Let $\{v_1, \cdots, v_n\}$ be an ONB of $\mathbb{R}^n$
Let $\{u_1, \cdots, u_m\}$ be an ONB of $\mathbb{R}^m$
Let $\{\sigma_1, \cdots, \sigma_m\}$ be a set of m scalars with $\sigma_i \geq 0$, $i = 1, \cdots m$.
Then, $\sigma_i u_i$ is the ith principal semiaxis with length σ_i in $\mathbb{R}^m$.

Now, if $\text{rank}(A) = r$, then exactly r of $\{\sigma_1, \cdots, \sigma_m\}$ are nonzero, and exactly $m-r$ of σ_i's are zero.
So, if $\underline{m \geq n}$, then $\underline{\text{rank}(A) \leq n}$. full rank if $= n$
i.e., at most n of σ_i's are nonzero.

For simplicity, let's assume $m \geq n$ and $\text{rank}(A) = n$ for the time being.

<u>Def.</u> The singular values of A
$\overset{\text{def}}{\Longleftrightarrow}$ The lengths of the n principal semiaxes of the hyperellipsoid AS

Our convention: $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \underset{\neq}{\geq} 0$

Def. The n **left singular vectors** of A
$\overset{\text{def}}{\Longleftrightarrow}$ $\{u_1, \cdots, u_n\}$: the unit vectors in $\mathbb{R}^m$ along the principal semiaxes of AS.
So, $\sigma_i u_i$ is the ith largest principal semiaxis of AS.

Def. The n **right singular vectors** of A
$\overset{\text{def}}{\Longleftrightarrow}$ $\{v_1, \cdots, v_n\} \in S$: the preimages of the principal semiaxes of AS, i.e.,

$$A v_i = \sigma_i u_i \qquad i = 1, \cdots, n.$$

☆ Reduced SVD

$$m\underset{n}{\begin{bmatrix} A \end{bmatrix}} \underbrace{\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}^{n}}_{n,\ = V} = m\underbrace{\begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix}}_{n,\ = \hat{U}} \underbrace{\begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{bmatrix}^{n}}_{n,\ = \hat{\Sigma}}$$

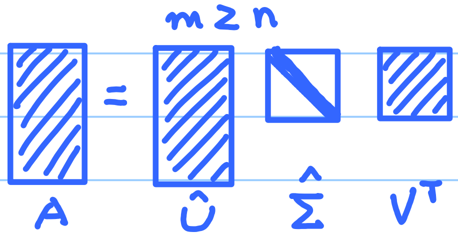
$$\Rightarrow \quad \underset{m\times n}{A}\ \underset{n\times n}{V} = \underset{m\times n}{\hat{U}}\ \underset{n\times n}{\hat{\Sigma}}$$

Since V is an orthogonal matrix,

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

The **reduced** SVD of A.

$m \geq n$



$m < n$

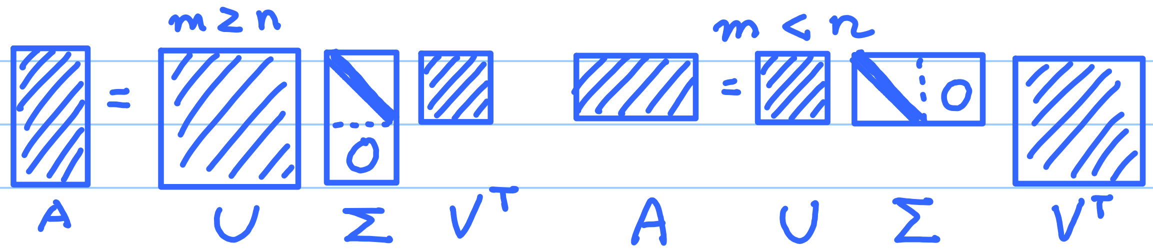
☆ Full SVD

Note $\hat{U} \in \mathbb{R}^{m \times n}$ in the reduced SVD with $m \geq n$.

$\Rightarrow$ The column vectors of $\hat{U}$ do not form an ONB of $\mathbb{R}^m$ unless $m = n$.

$\Rightarrow$ Remedy: Adjoin $m-n$ ON vectors to $\hat{U}$ to form an orthogonal matrix U. Then $\hat{\Sigma}$ must be changed to $\Sigma \in \mathbb{R}^{m \times n}$

$$\boxed{A = U \Sigma V^T}$$ The full SVD of A



For non-full rank matrices, i.e., $\operatorname{rank}(A) = r < \min(m, n)$, $\exists$ only r positive singular values.

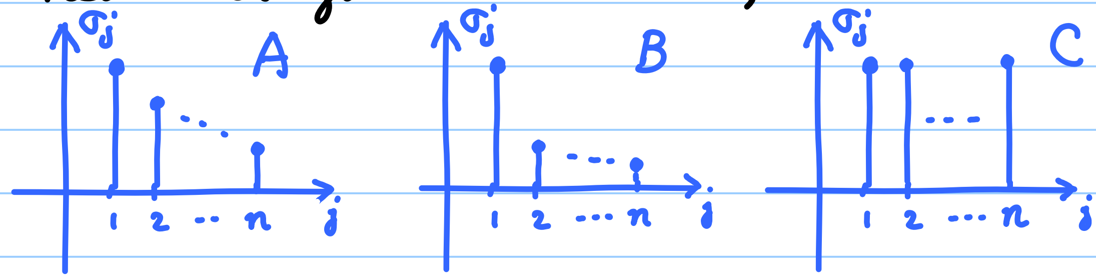
So,

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \\ \hline & & & O & & \end{bmatrix} \text{ or } \left[\begin{array}{cccccc|c} \sigma_1 & & & & & & \\ & \ddots & & & & & \\ & & \sigma_r & & & & O \\ & & & 0 & & & \\ & & & & \ddots & & \\ & & & & & 0 & \end{array}\right]$$

$m \geq n$ $\qquad\qquad$ $m \leq n$

Let's consider $m = n$ and full rank case. Theoretically, it's invertible, nonsingular.

However, we can gain more info by checking the distribution of the singular values of A $\Rightarrow$ We can see whether A is near singular or not, etc.



Out of these three scenarios, which matrix do you think behaves best numerically?
$\Rightarrow$ C.

★ Pseudoinverse via SVD

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

where

$$\Sigma^{+} := \begin{bmatrix} \frac{1}{\sigma_1} & & & & \\ & \ddots & & & \\ & & \frac{1}{\sigma_r} & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \\ \hline & & \text{O} & & \end{bmatrix} \quad \text{or} \quad \left[\begin{array}{cccccc|c} \frac{1}{\sigma_1} & & & & & & \\ & \ddots & & & & & \\ & & \frac{1}{\sigma_r} & & & & \text{O} \\ & & & 0 & & & \\ & & & & \ddots & & \\ & & & & & 0 & \end{array}\right]$$

$$m \geq n \qquad\qquad\qquad m \leq n$$

Check:

$$AA^+ = U\Sigma V^T V \Sigma^+ U^T$$
$$= U\Sigma\Sigma^+ U^T$$
$$= U \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots \\ & & O & & & 0 \end{bmatrix} U^T$$
$$= [\mathbf{u}_1 \cdots \mathbf{u}_r \; \mathbf{0} \cdots \mathbf{0}] \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_m^T \end{bmatrix}$$
$$= \hat{U}\hat{U}^T$$

$\longrightarrow$ reduced version.

Similarly, $A^+A = \hat{V}\hat{V}^T$

The Moore-Penrose Conditions

For a given matrix $A \in \mathbb{R}^{m\times n}$, if $X \in \mathbb{R}^{n\times m}$ satisfies the following:

$$\begin{cases} (1) & AXA = A \\ (2) & XAX = X \\ (3) & (AX)^T = AX \\ (4) & (XA)^T = XA \end{cases}$$

then X is called the pseudoinverse (or the Moore-Penrose inverse) of A and written as A^+

$\exists$ many applications using A^+!

Note: If $\| AX - I_m \|_F \to \min$ then $X = A^+$.