

# ECE 532 - lecture 11 - the SVD

①

★ Matrix norms: just like vector norms (same rules).

- i)  $\|A\| = 0$  if and only if  $A = 0$
- ii)  $\|\alpha A\| = |\alpha| \|A\|$  for all  $\alpha \in \mathbb{R}$
- iii)  $\|A + B\| \leq \|A\| + \|B\|.$

Two often-used matrix norms:

★ Frobenius norm:  $\|X\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n x_{ij}^2}$

this is the same as arranging all elements of  $X$  into a long vector of length  $mn$  and taking the  $\|\cdot\|_2$  norm.

★ Induced norm:  $\|X\|$  (this is the default norm).

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} \quad (\text{does this even make sense?})$$

because of the scaling property of the norm, we also have:

$$\|A\| = \max_{\|x\|=1} \|Ax\|. \quad \text{This is clearly a bounded quantity.}$$

not the same norm! one  
is a matrix norm, the other is  
a vector norm!

(2)

## Property of induced norms:

$$(i) \|Ax\| \leq \|A\| \cdot \|x\|. \quad \text{for all } A \in \mathbb{R}^{m \times m}, x \in \mathbb{R}^n$$

$$(ii) \|AB\| \leq \|A\| \cdot \|B\| \quad \text{for all } A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times k}.$$

Proof:  $\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} \geq \frac{\|A\tilde{x}\|}{\|\tilde{x}\|}$  where  $\tilde{x}$  is any particular  $x$ .

$$\Rightarrow \|A\tilde{x}\| \leq \|A\| \|\tilde{x}\|. \quad \blacksquare$$

For part (ii),

$$\|ABx\| = \|A(Bx)\| \leq \|A\| \cdot \|B\| \cdot \|x\|.$$

$$\Rightarrow \frac{\|ABx\|}{\|x\|} \leq \|A\| \cdot \|B\|.$$

This is true for all  $x \neq 0$ . In particular, it's true if we maximize over  $x$ . The left-hand side becomes  $\|AB\|$ .

$$\Rightarrow \|AB\| \leq \|A\| \cdot \|B\|. \quad \blacksquare$$

Note: the induced norm is sometimes called just the "norm". Or the "induced 2-norm", or the "spectral norm". We can also define other norms such as the induced 1-norm:

$$\|A\|_{i1} = \max_{x \neq 0} \frac{\|Ax\|_1}{\|x\|_1}$$

sometimes, we write

$$\|A\| = \|A\|_2 \text{ or } \|A\| = \|A\|_{i2}$$

just to be clear.

or the mixed 2- $\infty$  induced norm:

$$\|A\|_{i2\infty} = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_\infty}$$

(3)

Recall orthogonal matrices :  $U \in \mathbb{R}^{m \times k}$  satisfy  $U^T U = I$ .

\* If  $U$  is square,  $UU^T = UU = I$  and  $U^{-1} = U^T$ .

\* every  $U_1 \in \mathbb{R}^{m \times k}$  can be completed by some  $U_2 \in \mathbb{R}^{m \times (m-k)}$  such that  $U = [U_1, U_2]$  is square and orthogonal.

\*  $\|Ux\| = \|x\|$  for any orthogonal  $U$ .

For matrices, a similar property holds:

$$(i) \|UAV^T\| = \|A\| \quad \text{if } U, V \text{ are orthogonal}$$

$$(ii) \|UAV^T\|_F = \|A\|_F \quad \text{if } U, V \text{ are orthogonal}$$

$$\text{prof: } \|UAV^T\| = \max_{x \neq 0} \frac{\|UAV^T x\|}{\|x\|} = \max_{x \neq 0} \frac{\|AV^T x\|}{\|x\|}.$$

note that  $V^T x$  can be anything. i.e.  $y = V^T x$  always has a solution. (simply let  $x = Vy$ ). The set of all solutions is

$$X = Vy + \underbrace{V^\perp w}_{\text{null space of } V^T, \text{ or perp space of } V}.$$

so let  $y = V^T x$ .

$$\begin{aligned} \Rightarrow \max_{y,w} \frac{\|Ay\|}{\|Vy + V^\perp w\|} &= \sqrt{\max_{y,w} \frac{\|Ay\|^2}{\|Vy + V^\perp w\|^2}} = \sqrt{\max_{y,w} \frac{\|Ay\|^2}{\|Vy\|^2 + \|V^\perp w\|^2}} \quad (\text{maximized when } w=0). \\ &= \max_y \frac{\|Ay\|}{\|y\|} = \|A\|. \end{aligned}$$

Pythagorean thm! ■

(4)

Second part of proof: we use two properties:

$$\text{trace}(AB) = \text{trace}(BA), \quad \text{and} \quad \|A\|_F^2 = \text{trace}(ATA).$$

Note:  $\text{trace}(X)$  for a square matrix is the sum of diagonal entries. i.e.  $\text{trace}(X) = \sum_{i=1}^n X_{ii}$ .

prove the properties above as an exercise!

$$\begin{aligned}
 \text{Now: } \|UAV^T\|_F^2 &= \text{trace}\left[(UAV^T)^T(UAV^T)\right] \\
 &= \text{trace}\left[V A^T \underbrace{U^T U}_I A V^T\right] \\
 &= \text{trace}\left[\underbrace{V A^T A V^T}_{\substack{\longleftarrow \\ \text{swap.}}}\right] \\
 &= \text{trace}\left[\underbrace{V^T V}_I A^T A\right] \\
 &= \text{trace}(A^T A) \\
 &= \|A\|_F^2.
 \end{aligned}$$

■

(5)

## Induced norm of a diagonal matrix

what is  $\left\| \begin{bmatrix} a_1 & & 0 & \\ & a_2 & & \\ 0 & & \ddots & a_n \end{bmatrix} \right\|$  ?

$$\|A\|^2 = \max_{x \neq 0} \frac{\left\| \begin{pmatrix} a_1 & & 0 & \\ & a_2 & & \\ 0 & & \ddots & a_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right\|_2^2}{\|x\|_2^2}$$

$$= \max_{x \neq 0} \frac{a_1^2 x_1^2 + a_2^2 x_2^2 + \dots + a_n^2 x_n^2}{x_1^2 + x_2^2 + \dots + x_n^2}$$

let  $a_l^2$  be the largest one. Clearly :

$$\max_{x \neq 0} \frac{a_1^2 x_1^2 + \dots + a_n^2 x_n^2}{x_1^2 + \dots + x_n^2} \leq \frac{a_l^2 x_1^2 + \dots + a_l^2 x_n^2}{x_1^2 + \dots + x_n^2} = a_l^2$$

but if we pick the vector  $x_i = \begin{cases} 0 & \text{if } i \neq l \\ 1 & \text{if } i = l \end{cases}$

$$\max_{x \neq 0} \frac{a_1^2 x_1^2 + \dots + a_n^2 x_n^2}{x_1^2 + \dots + x_n^2} \geq \frac{a_l^2 \cdot 1}{1} = a_l^2.$$

so  $\|A\| = |a_l|$  (largest diagonal entry).

(6)

## The SVD (Finally!).

Every matrix  $A \in \mathbb{R}^{m \times n}$  has a factorization:

$$A = U_1 \Sigma_1 V_1^T$$

(m × n)      (m × r)      (r × r)      (r × n)

where  $\boxed{r = \text{rank}(A)}$

also,  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ .

where:

$U_1 \in \mathbb{R}^{m \times r}$  is orthogonal,

$V_1 \in \mathbb{R}^{n \times r}$  is orthogonal

$$\Sigma_1 = \begin{bmatrix} \sigma_1 & & \\ & \ddots & 0 \\ 0 & & \sigma_r \end{bmatrix}$$

diagonal,  
positive

(they are sorted). These are called the singular values of A.

vectors  $V_1 = [v_1 \ v_2 \ \dots \ v_r]$  are the right singular vectors of A  
 ↪ each in  $\mathbb{R}^n$

vectors  $U_1 = [u_1 \ u_2 \ \dots \ u_r]$  are the left singular vectors of A  
 ↪ each in  $\mathbb{R}^m$

★ factorization is not unique (e.g. can flip signs of U, V)

★  $\Sigma$  is unique. i.e.  $\{\sigma_1, \dots, \sigma_r\}$  are a property of A.

(7)

$A = U, \Sigma, V_1^T$  is called the "economy" or "thin" SVD

if we complete  $U_1$  and  $V_1$ , i.e.  $U = \underbrace{[U_1, U_2]}_{\in \mathbb{R}^{m \times m}}$ ,  $V = \underbrace{[V_1, V_2]}_{\in \mathbb{R}^{n \times n}}$

$$\text{and let } \Sigma = \left\{ \begin{bmatrix} \underbrace{\Sigma_1}_{r \times r} & 0 \\ 0 & 0 \end{bmatrix} \right\}_n^m$$

$$\text{then } A = U, \Sigma, V_1^T = \underbrace{[U_1, U_2]}_{m \times r} \underbrace{\begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}}_{m \times n} \underbrace{\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}}_{n \times r}^r = U \Sigma V^T.$$

This is called the "full" SVD, or just the SVD.

If  $A$  has full column rank,

$$[A] = [U_1] \underbrace{[\Sigma_1]}_{\text{square}} [V_1^T] \quad (r=n).$$

$$= [U_1, U_2] [\Sigma_1] [V_1^T]$$

If  $A$  has full row rank,

$$[A] = \underbrace{[U_1]}_{\text{square}} [\Sigma_1] [V_1^T] \quad (r=m).$$

$$= [U_1] [\Sigma_1, 0] \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$

(8)

## Properties

★ A and  $A^T$  have the same singular values. ( $\text{rank}(A) = \text{rank}(A^T)$ ).

If  $A = U\Sigma V^T$  then  $A^T = \underbrace{V\Sigma^T U^T}_{\text{SVD for } A^T}$ .

★ norms:

$$\|A\| = \|U, \Sigma, V^T\| = \|\Sigma\| = \sigma_1 \quad (\text{max singular value})$$

$$\|A\|_F = \|U, \Sigma, V^T\|_F = \|\Sigma\|_F = \sqrt{\sigma_1^2 + \dots + \sigma_r^2}$$

★ subspaces:

$U_1$  : basis for  $R(A)$

$U_2$  : basis for  $R(A)^\perp$

$V_1$  : basis for  $N(A)^\perp$

$V_2$  : basis for  $N(A)$

"four fundamental subspaces".

★ range-nullspace orthogonality:

$$R(A)^\perp = N(A^T).$$

$$A = \underbrace{[U, U_2]}_{R(A)^\perp} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \quad \Rightarrow \quad A^T = \underbrace{[V_1, V_2]}_{N(A^T)} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$$