Lecture 37: Quick review from previous lecture

- Let A be an $m \times n$ matrix with $\textsf{rank}(A) = r$ and nonzero singular values $\sigma_1 \geq \ldots \geq \sigma_r$ and SVD
$$A = U\Sigma V^T = U \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \; 0 \end{bmatrix}_{m\times n} V^T$$
The **pseudoinverse** of A is the $n \times m$ matrix
$$A^+ = V\Sigma^+U^T = V \begin{bmatrix} 1/\sigma_1 & & & & \\ & \ddots & & & \\ & & 1/\sigma_r & & \\ & & & 0 & \\ & & & & \ddots \; 0 \end{bmatrix}_{n\times m} U^T$$
Here
$$\Sigma^+_{ij} = \begin{cases} \frac{1}{\sigma_i} & \text{if } i = j \leq r \\ 0 & \text{otherwise} \end{cases}$$
- Suppose that $A \in M_{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$. Then
$$\mathbf{x}^* = A^+\mathbf{b}$$
is the least squares solution to the linear system $A\mathbf{x} = \mathbf{b}$.
- If $A \in M_{m\times n}$ has n linearly independent columns (rank $A = n$), then
$$\mathbf{x}^* = A^+\mathbf{b} = (A^TA)^{-1}A^T\mathbf{b},$$
where
$$A^+ = (A^TA)^{-1}A^T.$$

Today we will

- Continue Section 8.7 Singular Values

- Lecture will be recorded -

- **Exam 3:** 5/3 (Monday) in lecture. * 2 review sections on Wed. and Fri.
- Practice Exam is on Canvas now.

$D = \begin{pmatrix} 1 & 0 \\ 0 & -3 \end{pmatrix}$, $\|D\|_2 = |-3| = 3$

§ Revisit Matrix norm.

Frobenius norm and Natural Matrix norm.

Let's consider $\|\mathbf{x}\|_2 = \sqrt{x_1^2 + \ldots + x_n^2}$.

$tr(B) = b_{11} + \cdots + b_{nn}$ if $B = (b_{ij})$.

The **natural matrix norm** of A is

$$\|A\|_2 = \max\{\|A\mathbf{u}\|_2 : \|\mathbf{u}\|_2 = 1\}.$$

Ex: $A = \begin{array}{cc} \vec{a}_1 & \vec{a}_2 \\ \end{array}$ $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$.

The **Frobenius norm** of a matrix $A = (a_{ij})$ is defined by $\|A\|_F = \sqrt{1^2 + 3^2 + 2^2 + 4^2}$

$= \sqrt{\|\vec{a}_1\|_2^2 + \|\vec{a}_2\|_2^2}$

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2} = \sqrt{tr(A^T A)}$$

Fact 8: Let $A \in M_{m\times n}$ and Q is an orthogonal matrix. Then

$$① \ \|QA\|_2 = \|A\|_2, \quad ② \ \|QA\|_F = \|A\|_F. = \|AQ\|_F$$

[To see this] Old HW: Q orthogonal. Then $\|Qx\|_2 = \|x\|_2$.

① $\|A\|_2 = \max\{\|Au\|_2 : \|u\|_2 = 1\}$

$= \max\{\|QAu\|_2 : \|u\|_2 = 1\}$

$= \|QA\|_2$.

② If $A = [\vec{a}_1 \ \cdots \ \vec{a}_n]_{m\times n}$, then

$\|QA\|_F = \sqrt{\|Q\vec{a}_1\|_2^2 + \cdots + \|Q\vec{a}_n\|_2^2}$

$= \sqrt{\|\vec{a}_1\|_2^2 + \cdots + \|\vec{a}_n\|_2^2}$

$= \|A\|_F$.

since $\|Q\vec{a}_j\|_2 = \|\vec{a}_j\|_2$

Then we have

Fact 9: (1) Let $A \in M_{m\times n}$ ($m \times n$ real matrices) with rank$(A) = r$ and has positive singular values $\sigma_1 \geq \ldots \geq \sigma_r$. Then

$$\|A\|_F = \sqrt{\sigma_1^2 + \ldots + \sigma_r^2}.$$

(2) In particular, if A is a real, symmetric matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$. Then

$A = A^T$

$$\|A\|_F = \sqrt{\lambda_1^2 + \ldots + \lambda_n^2}$$

[To see this]

(1) SVD : $A = U \Sigma V^T$, where $\Sigma = \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}_{m\times n}$

Fact 8

$\|A\|_F = \|U\Sigma V^T\|_F = \|\Sigma V^T\|_F = \|\Sigma\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$. #

(2) $A = A^T$. By Fact 4, $\sigma_i = |\lambda_i|$.

By (1), $\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_n^2} = \sqrt{\lambda_1^2 + \cdots + \lambda_n^2}$

Example 3. Let $A = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$. Find $\|A\|_F$.

Ans. Previously, we have found the eigenvalues of A to be 0, 2 and 3.

$A = A^T$. By Fact 9 (2), $\|A\|_F = \sqrt{0^2 + 2^2 + 3^2}$

$= \sqrt{13}$.

Fact 10: (1) Let $A \in M_{m\times n}$ ($m \times n$ real matrices) with rank$(A) = r$ and has positive singular values $\sigma_1 \geq \ldots \geq \sigma_r$. Then

$$\|A\|_2 = \sigma_1 \qquad \text{(largest singular value).}$$

$\sigma_i = \sqrt{\lambda_i}$, λ_i eigenval. of A^TA

(2) In particular, if A is a real, symmetric matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$.

$$\|A\|_2 = \max_{1\leq i\leq n} |\lambda_i|$$

$A = A^T$, $\sigma_i = |\lambda_i|$.

[To see this] Like Fact 9.

$$\|A\|_2 = \|U\Sigma V^T\|_2 \overset{\text{Fact 8}}{=} \|\Sigma\|_2 = \max_{1\leq i\leq r} \sigma_i = \sigma_1 \text{ // (the largest one)}$$

Example 4.

(1) Consider the same matrix as Example 3: $A = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$ has eigenvalues 0, 2 and 3. Then $\|A\|_2 =$ 3

$A = A^T$, $\sigma_i = 0, 2, 3$.

(2) Also in **Example 1:**

$A = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{bmatrix}$ has singular value $\sqrt{6}, 0, 0$. Then $\|A\|_2 =$ $\sqrt{6}$

$\|A\|_F = \sqrt{6}$. (By Fact 9), or $\|A\|_F = \sqrt{2(1^2+1^2+(-1)^2)} = \sqrt{6}$.

§ Low rank approximations to a matrix.

Suppose we want to approximate a matrix $A = A_{m\times n}$ with rank r by a matrix $B = B_{m\times n}$ with rank $k < r$.

We want to find such B of rank k

"to minimize $\|A - \tilde{B}\|$ among all $m \times n$ matrix $\tilde{B}$ with rank k"

Recall a matrix A with rank r has full SVD as follows:

$$A = U\Sigma V^T = U \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix}_{m\times n} V^T, \quad \sigma_1 \geq \cdots \geq \sigma_r > 0.$$

$$\text{let } B = U \underbrace{\begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_k & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix}_{m\times n}}_{\Sigma_k} V^T, \quad k < r.$$

Then $\operatorname{rank}(B) = k$.

The best rank k approximation to A is

$$B = U\Sigma_k V^T.$$

Fact 8: This matrix B minimizes the distance to A as measured by Frobenius norm and operator norm:

$$\|A - B\|_2 = \sigma_{k+1} \qquad \|A - B\|_F = \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_r^2}$$

$$A - B = U \begin{bmatrix} \overbrace{0 & & }^{k \text{ zeros}} & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & & \sigma_{k+1} & & \\ & & & & \ddots & \\ & & & & & \sigma_r \\ & & & & & & 0 \\ & & & & & & & \ddots \\ & & & & & & & & 0 \end{bmatrix} V^T$$

has positive singular values $\underbrace{\sigma_{k+1}}_{\text{largest one}}, \ldots, \sigma_r$

Then $\|A - B\|_2 = \sigma_{k+1}$

$\|A - B\|_F = \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_r^2}$. #

$A_{3\times 4} : \mathbb{R}^4 \to \mathbb{R}^3$.

Example 5. Suppose A is a 3×4 matrix and has positive singular values ~~2, 3, 8~~ 8, 3, 2 and corresponding right singular vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ and left singular vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$, respectively. Then

$$A = 8\mathbf{u}_1\mathbf{v}_1^T + 3\mathbf{u}_2\mathbf{v}_2^T + 2\mathbf{u}_3\mathbf{v}_3^T.$$

Find the best rank 1 approximation to A and then find $\|A - B\|_2$ and $\|A - B\|_F$.

$$A = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \begin{bmatrix} 8 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ v_3^T \end{bmatrix}.$$ (Reduced SVD)

The best rank 1 appro. is

$$B = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \begin{bmatrix} 8 & & \\ & 0 & \\ & & 0 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ v_3^T \end{bmatrix}$$

$$= \underline{8\, u_1 v_1^T}$$

Then
$$A - B = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \begin{bmatrix} 0 & & \\ & 3 & \\ & & 2 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ v_3^T \end{bmatrix}$$

$\|A - B\|_2 = 3$, largest singular value of $A-B$.

$\|A - B\|_F = \sqrt{3^2 + 2^2} = \sqrt{13}$

§ Condition number

A very useful quantity for understanding the behavior of a matrix is its *condition number*.

Definition: The **condition number** of a nonsingular $n \times n$ matrix A (rank $A = n$) is the ratio between its largest σ_1 and smallest singular values σ_n, namely,

$$\kappa(A) = \frac{\sigma_1}{\sigma_n}.$$

The condition number $\kappa(A)$ measures the "sensitivity of operations" we perform with A to changes in the input data.

In particular, if $\kappa(A)$ is very large, then small changes in $\mathbf{x}$ can result in large changes in $A\mathbf{x}$.

Fact 13: If A is $n \times n$ nonsingular matrix, then A and A^{-1} have the same condition number.

$A_{n\times n} = U \Sigma V^T$, $\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}_{n\times n}$, $\sigma_1 \geq \cdots \geq \sigma_n > 0$.

$A^{-1} = V \Sigma^{-1} U^T$, $\Sigma^{-1} = \begin{bmatrix} 1/\sigma_1 & & \\ & \ddots & \\ & & 1/\sigma_n \end{bmatrix}$.

A has singular values $\sigma_1 \geq \cdots \geq \sigma_n$.

A^{-1} " $1/\sigma_1 \leq \cdots \leq 1/\sigma_n$

$K(A) = \frac{\sigma_1}{\sigma_n}$; $K(A^{-1}) = \frac{\frac{1}{\sigma_n}}{\frac{1}{\sigma_1}} = \frac{\sigma_1}{\sigma_n}$. #

Fact 14:

$$\kappa(A) = \|A\|_2 \|A^{-1}\|_2.$$

$\|A\|_2 = \sigma_1$ (largest singular value)

$\|A^{-1}\|_2 = \frac{1}{\sigma_n}$ (" of A^{-1})

Poll Question 1: Suppose $n \times n$ matrix A has full SVD

$A = U\underset{\Sigma}{D}V^T$, where U, V are orthogonal matrices and D is diagonal matrix.

Then which one of the following is the full SVD of A^T

A) UDV^T
B) VDU^T ✓

$$A^T = (UDV^T)^T = V\underline{D^T}U^T$$
$$= V\,\underline{D}\,U^T.$$