Lecture 37: Quick review from previous lecture

- Suppose that $A \in M_{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$. Let $\mathbf{x}^* = A^+\mathbf{b}$. Then $\mathbf{x}^*$ is the least squares solution to the linear system $A\mathbf{x} = \mathbf{b}$.

  If A has n linearly independent columns($\ker A = \{\mathbf{0}\}$), then

  $$\mathbf{x}^* = A^+\mathbf{b} = (A^TA)^{-1}A^T\mathbf{b},$$

  where

  $$A^+ = (A^TA)^{-1}A^T.$$

- If $A = \sum_{j=1}^{r} \sigma_j u_j v_j^T$, then the matrix $A_k = \sum_{j=1}^{k} \sigma_j u_j v_j^T$, $k < r$ minimizes the distance to A, that is,

  $$\min_{rank(B)=k} \|A - B\|_2 = \|A - A_k\|_2.$$

$A = A_{m\times n}$ with $\text{rank}(A) = r$.

(full SVD)

$$A = U\Sigma V^T = [u_1 \cdots u_m]\begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}\begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix}$$

$$= [u_1 \cdots u_m]\begin{bmatrix} \sigma_1 v_1^T \\ \vdots \\ \sigma_r v_r^T \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$= \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T.$$

$$= \sum_{j=1}^{r} \sigma_j u_j v_j^T$$

(Reduced SVD)

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

$$= [u_1 \cdots u_r]\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}_{r\times r}\begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix}$$

$$= \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T. \ \#$$

Today we will discuss condition number and incomplete matrix.

- Lecture will be recorded -

- Information for Final Exam and Course Grade has been posted on Canvas, see "Announcements".

§ Condition number

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

For simplicity, let's take a square matrix $A = A_{n\times n}$, and suppose the rank is $r = n$ (so the matrix is nonsingular). Call the singular values

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

> **Definition:** The **condition number** of a nonsingular $n \times n$ matrix is the ratio between its largest and smallest singular values, 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.

Example. Suppose we apply A to a vector $\mathbf{x}$. Write the SVD $A = U\Sigma V^T$; then we can write

$= \sum_{j=1}^{n} \sigma_j u_j v_j^T$

δv_1, $\tilde{x} = x + \delta v_1$, x $\xrightarrow{A}$ $\delta u_1 \sigma_1$, $A\tilde{x} = Ax + \delta\sigma_1 u_1$, Ax

$$A\mathbf{x} = \sum_{i=1}^{n} \sigma_i \langle \mathbf{x}, \mathbf{v}_i \rangle \mathbf{u}_i$$

$Ax = \sum_{j=1}^{n} \sigma_j u_j v_j^T x = \sum_{j=1}^{n} \sigma_j \langle x, v_j \rangle u_j$

What if we perturb $\mathbf{x}$ a little bit, along the direction of $\mathbf{v}_1$? How much would the product $A\mathbf{x}$ change?

Pick a small number δ, define $\tilde{x} = x + \delta v_1$.

Then $A\tilde{x} = A(x + \delta v_1)$

$= Ax + \delta A(v_1)$ $\qquad Av_1 = \sum_{1}^{n} \sigma_i \langle v_1, v_i \rangle u_i = \sigma_1 u_1$

$= Ax + \delta \sigma_1 u_1$

So, $\|A\tilde{x} - Ax\|_2 = \|\delta \sigma_1 u_1\|_2 = |\delta| \sigma_1$ since $\|u_1\|_2 = 1$.

The "relative difference" between Ax and $A\tilde{x}$ is the ratio

$$\frac{\|A\tilde{x} - Ax\|_2}{\|Ax\|_2} = \frac{|\delta|\sigma_1}{(\sigma_1^2\langle x, v_1\rangle^2 + \cdots + \sigma_n^2 \langle x, v_n\rangle^2)^{1/2}}.$$

[Continue]

For example, we take $x = v_n$, then $\langle x, v_i \rangle = 0$ if $i \neq n$.

So, the relative difference is

$$\frac{\| A\tilde{x} - Ax \|_2}{\| Ax \|_2} = \frac{|\delta|\,\sigma_1}{\sigma_n} = |\delta|\,\kappa(A).$$

On the other hand,

$$\frac{\| \tilde{x} - x \|_2}{\| x \|_2} = \frac{\| \delta v_1 \|_2}{1} = |\delta|. \quad \text{since } \|v_1\|_2 = 1$$

We have shown the following:

Fact: For certain vectors $\mathbf{x}$, perturbing $\mathbf{x}$ by relative amount δ along $\mathbf{v}_1$ can result in a change in $A\mathbf{x}$ by relative amount $|\delta|\kappa(A)$.

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

Matrices with large condition numbers are said to be **ill-conditioned**. One must be careful when working with ill-conditioned matrices, since small errors in inputs can become inflated.

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

[To see this] Since A is $n\times n$ nonsingular matrix,
there are singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$.
Then singular values of A^{-1} are

$$0 < \frac{1}{\sigma_1} \leq \frac{1}{\sigma_2} \leq \cdots \leq \frac{1}{\sigma_n}.$$

So, $K(A) = \dfrac{\sigma_1}{\sigma_n}$.

$K(A^{-1}) = \dfrac{1}{\sigma_n} \Big/ \dfrac{1}{\sigma_1} = \dfrac{\sigma_1}{\sigma_n}$. Thus, $K(A) = K(A^{-1})$ #

Fact: If A is $n \times n$ nonsingular matrix, then

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

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

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

$$\|A\|_2 \|A^{-1}\|_2 = \frac{\sigma_1}{\sigma_n} = \underline{K(A)} \quad \#.$$

8.6 Incomplete Matrices

If a matrix is not complete, then it is not similar to a diagonal matrix.
For example,

$$A = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$$

, $\lambda = 2, 2$.
$A - 2I = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$

is not complete and so is not diagonalizable.

Recall: A diagonalizable matrix A can be written as $A = VDV^{-1}$.
where $D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$. A is similar to D.

However, it can be shown that a not complete matrix is similar to a matrix of the following form, called a **Jordan matrix**:

$$\begin{pmatrix} J_{\lambda_1,n_1} & & & \\ & J_{\lambda_2,n_2} & & \\ & & \ddots & \\ & & & J_{\lambda_k,n_k} \end{pmatrix}$$

Here, each matrix J_{λ_i,n_i} is a **Jordan block matrix**, which is the following n_i-by-n_i matrix:

$$J_{\lambda_i,n_i} = \begin{pmatrix} \boxed{\lambda_i} & 1 & & & & \\ & \lambda_i & 1 & & & \\ & & \lambda_i & 1 & & \\ & & & \ddots & \ddots & \\ & & & & \lambda_i & 1 \\ & & & & & \lambda_i \end{pmatrix}$$

The numbers $\lambda_1, \ldots, \lambda_k$ are the eigenvalues of A. The eigenvalue λ_i has multiplicity n_i, the size of its Jordan block J_{λ_i,n_i}.

Note that if each Jordan block were 1-by-1 (i.e. $n_i = 1$ for all i), then the Jordan matrix would be diagonal, and A would be complete.

§ **The Jordan Canonical Form.** Let's see how we get this Jordan matrix. Throughout this section, we suppose that A is an $n \times n$ matrix. Let

$$\lambda_1, \ldots, \lambda_k$$

be the eigenvalues of A.

Definition: A **Jordan chain** of length j for a square matrix A is a sequence of nonzero vectors $\mathbf{w}_1, \ldots, \mathbf{w}_j$ that satisfies

$$\underline{A\mathbf{w}_1 = \lambda\mathbf{w}_1,} \quad A\mathbf{w}_i = \lambda\mathbf{w}_i + \mathbf{w}_{i-1}, \qquad i = 2, \ldots, j,$$

where λ is an eigenvalue of A.

EX: $j = 3$, $\quad (A - \lambda I) w_1 = 0.$

$$(A - \lambda I) w_2 = w_1 \Rightarrow (A - \lambda I)^2 w_2 = (A - \lambda I) w_1 = 0$$

$$(A - \lambda I) w_3 = w_2 \Rightarrow (A - \lambda I)^3 w_3 = 0.$$

$$A \begin{bmatrix} w_1 & \underline{\underline{w_2}} & w_3 \end{bmatrix} = \begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}$$

$$Aw_2 = \lambda w_2 + 1 \underline{\underline{w_1}}$$

The initial vector $\mathbf{w}_1$ in a Jordan chain is a "genuine" eigenvector. The rest $\mathbf{w}_2, \ldots, \mathbf{w}_j$ are **generalized eigenvectors**, in the following definition:

Definition: A nonzero vector $\mathbf{w} \neq \mathbf{0}$ such that

$$(A - \lambda I)^k \mathbf{w} = \mathbf{0} \qquad \text{for some } k > 0$$

and λ is a **generalized eigenvector** of the matrix A.

Example. Find the Jordan form of $A = \begin{pmatrix} 1 & 1 \\ -1 & 3 \end{pmatrix}$

$0 = \det(A - \lambda I)$, $\lambda = \underline{2}, \underline{2}$.

$A - 2I = \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix}$. $\dim(N(A-2I)) = 1$.

Jordan form $\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$. #

Example. Find the Jordan form of $A = \begin{pmatrix} -1 & -1 & 0 \\ 0 & -1 & -2 \\ 0 & 0 & -1 \end{pmatrix}$

$\lambda = -1, -1, -1$.

$\dim(N(A+I)) = 1$.

Jordan form is $\begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}$. #