Lecture 35: Quick review from previous lecture

• Suppose that a symmetric matrix $A$ has real eigenvalues

$$\lambda_1 \geq \cdots \geq \lambda_n.$$  

Then

$$\lambda_1 = \max\{\langle Ax, x \rangle : \|x\|_2 = 1\}, \quad \lambda_n = \min\{\langle Ax, x \rangle : \|x\|_2 = 1\}.$$  

The maximal value is achieved when $x = \pm u_1$, the unit eigenvector associated with the largest eigenvalue $\lambda_1$.

The minimal value is achieved when $x = \pm u_n$, the unit eigenvector associated with the smallest eigenvalue $\lambda_n$.

• Information for Final Exam and Course Grade has been posted on Canvas, see "Announcements".
**Definition:** The square roots of the eigenvalues of $A^T A$ are called the **singular values** $\sigma_1, \sigma_2, \ldots, \sigma_n$ of an $m \times n$ matrix $A$.

Thus, we have shown that

**Full SVD for a matrix:**
Let $A$ be an $m \times n$ matrix of rank $r$ with the positive singular values

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r,$$

and let $\Sigma$ be the $m \times n$ matrix defined by

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 0 & \sigma_r & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & \cdots & 0 \end{bmatrix}_{m \times n}$$

Then there exist an $m \times m$ orthogonal matrix $U$ and an $n \times n$ orthogonal matrix $V$ such that

$$A = U \Sigma V^T.$$
Example. Find full and reduced SVD for $A = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{bmatrix}$.

Solution: $A^t A = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 2 & 2 & -2 \\ 2 & 2 & -2 \\ -2 & -2 & 2 \end{pmatrix}$ has eigenvalues $\lambda = 6, 0, 0$.

- $\lambda = 6$, $v_1 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$;
- $\lambda = 0$, $v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $v_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

$u_1 = \frac{Av_1}{\sqrt{6}} = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$, $u_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $u_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

Finding $u_2 \perp u_1$, $(u_2 \in \text{coker } A)$, we have $u_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. 

(Full SVD)

(Reduced SVD)
Fact: Let \( A \in M_{m \times n} \) (\( m \times n \) real matrices). Then the following are true.

1. The nonzero eigenvalues of \( A^T A \) and \( A A^T \) are the same.

2. \( A \) and \( A^T \) have the same nonzero singular values.

\[ A = \begin{bmatrix} \frac{1}{\sigma_1} \\ \frac{1}{\sigma_2} \\ \vdots \\ \frac{1}{\sigma_r} \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} \begin{bmatrix} \hat{v}_1 \\ \hat{v}_2 \\ \vdots \\ \hat{v}_r \end{bmatrix} \]

### Reduced SVD:

\[ A = \begin{bmatrix} \frac{1}{\sigma_1} \\ \frac{1}{\sigma_2} \\ \vdots \\ \frac{1}{\sigma_r} \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} \begin{bmatrix} \hat{v}_1 \\ \hat{v}_2 \\ \vdots \\ \hat{v}_r \end{bmatrix} \]
Fact: Let $A \in M_{m \times n}$ ($m \times n$ real matrices). Then the following is true.

1. If $A = A^T$, then the singular values of $A$ are the absolute values of the eigenvalues of $A$.

\[ A v = \lambda v, \quad v \neq 0. \]
\[ (A^T A) v = A^T (A v) = A^T (\lambda v) = \lambda A v = (\lambda^2 v) . \]

So, $\lambda^2$ is eigenvalue of $A^T A$, it leads to $\sqrt{\lambda^2} = |\lambda|$ is the singular value of $A$. 

Likewise, we also have

Fact: Let $A \in M_{m \times n}$ ($m \times n$ real matrices). Then the following are true.

1. $\| A \|_2 = \sigma_1$
2. $\| A \|_F = \sqrt{\sigma_1^2 + \ldots + \sigma_r^2}$.

The proof is similar to the one for symmetric matrix we showed in Lecture 33. Thus, we skip the proof here.
§ Least square solutions - Pseudo inverse

How do we “almost” solve a system?

For instance, we consider an experimenter collects data by taking measurements

\[ b_1, b_2, \ldots, b_m \]

at times \( t_1, t_2, \ldots, t_m \) respectively.

Suppose that the data

\[ (b_1, t_1), (b_2, t_2), \ldots, (b_m, t_m) \]

are plotted in the plane.

Suppose there exists a “linear relationship” between \( b \) and \( t \), say \( b = \alpha t + \beta \).

We want to find the constants \( \alpha, \beta \) so that the line \( b = \alpha t + \beta \) represents the best possible fit to the data collected. One way is to minimize the error

\[
E \overset{\text{def}}{=} \sum_{i=1}^{m} \left( \overbrace{b_i}^{\text{collected data}} - \overbrace{(\alpha t_i + \beta)}^{\text{linear relationship}} \right)^2,
\]

which can be written as

\[
\|Ax - b\|^2,
\]

where

\[
A = \begin{bmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \quad x = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}.
\]

Since we cannot solve \( Ax = b \) exactly in many cases, our goal here is to find a \( x \) that minimizes \( \|Ax - b\| \). We will develop a general method for finding a vector \( x^* \) that minimizes the error \( E \), that is,

\[
\|Ax^* - b\| \leq \|Ax - b\| \quad \text{for all } x \in \mathbb{R}^n.
\]

**Definition:** Suppose that \( A \in M_{m \times n}, b \in \mathbb{R}^m \). The least squares problem is to find \( x \in \mathbb{R}^n \) for which that \( \|Ax - b\| \) is minimized.

A vector \( x \) that minimizes \( \|Ax - b\| \) is called the least squares solution.
**Definition:** The pseudoinverse of a nonzero $m \times n$ matrix $A$ with SVD $A = U\Sigma V^T$ is the $n \times m$ matrix

$$A^+ = V\Sigma^+ U^T.$$ 

$A$ has rank $r$ with singular values $\sigma_1 \geq \ldots \geq \sigma_r > 0$.

$$\Sigma = \begin{bmatrix} \sigma_1 & \sigma_2 & \cdots & 0 \\ & \sigma_3 & \cdots & 0 \\ & & \ddots & \vdots \\ & & & \sigma_r \end{bmatrix}_{m \times n}$$

$$\Sigma^+ = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_r} \end{bmatrix}_{n \times m}; \quad \Sigma_{ij}^+ = \begin{cases} \frac{1}{\sigma_i}, & \text{if } i = j \leq r, \\ 0, & \text{otherwise}. \end{cases}$$

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

\[ b = b_r + b_n \]

$b_r$ and $b_n$ are orthogonal projections of $b$ onto $\text{img} A$ and $\text{coker} A$, respectively.

\[ \|Ax - b\|_2^2 = \|Ax - b_r - b_n\|_2^2 = \|Ax - b_r\|_2^2 + \|b_n\|_2^2 \]

$\|Ax - b\|_2 \leq \|Ax - b_r\|_2 + \|b_n\|_2$. 

The least squares solution satisfies $Ax = br$.

**Check:** $x^* = A^+b$ satisfies $Ax = br$. 

\[ MATH 4242-Week 14-3 \quad Spring 2020 \]
Remark: The equation \( A^T Ax - A^T b = 0 \) is called the normal equation.

Fact: Suppose that \( A \in M_{m \times n}, \ b \in \mathbb{R}^m \). Let \( x^* = A^+ b \). If \( A \) has \( n \) linearly independent columns(\( \ker A = \{0\} \)), then

\[
x^* = A^+ b = (A^T A)^{-1} A^T b,
\]

where

\[ A^+ = (A^T A)^{-1} A^T. \]

Thus, \((A^T A)^{-1}\) exists. \( x^* \) satisfies \( A^T Ax = A^T b \).

So,

\[ x^* = (A^T A)^{-1} A^T b. \]

Example. Consider the linear system

\[
\begin{align*}
x + y - z &= 1, \\
x + y - z &= 1.
\end{align*}
\]

Find the best approximation to a solution having minimum norm.

To be continued!