Lecture 22

Q. Let $A \in \mathbb{R}^{m\times n}$ (say, $m \geq n$), $r = \operatorname{rank} A$, $A = USV^T$ (SVD).
What is $AA^T\vec{u}_j$ for $1 \leq j \leq r$?

Recall Last time:

We computed SVD of a matrix A

1. Compute the singular values $\{\sigma_i\}$: $\sigma_i = \sqrt{\lambda_i(A^TA)} = \sqrt{\lambda_i(AA^T)}$
   This gives us S.

2. Compute either $\vec{u}_j$ = eigenvector of AA^T corresponding to λ_j
   or $\vec{v}_j$ = eigenvector of A^TA, corresponding to λ_j

   (You may have to expand the ONB $u_1, \dots, u_m$ or $v_1, \dots, v_n$.)

3. Compute the other ONB.

To justify Step 2:

$$AA^T = USV^TVS^TU^T = USS^TU^T = U \cdot \operatorname{diag}(\overbrace{\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2}^{\lambda_i \text{ of } AA^T}, 0, \dots, 0)\, U^T$$

Consider the case where $m \geq n$, ie A is tall.

$$AA^T\vec{u}_j = \begin{bmatrix} U \end{bmatrix} \left[\begin{array}{c|c} \Sigma^2 & 0 \\ \hline 0 & 0 \end{array}\right] \underbrace{\begin{bmatrix} U^T \end{bmatrix} \begin{bmatrix} \vec{u}_j \end{bmatrix}}_{\begin{bmatrix} 0 \\ \vdots \\ 0 \\ u_j^Tu_j \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = e_j}$$

$$= \begin{bmatrix} U \end{bmatrix} \left[\begin{array}{c|c} \Sigma^2 & 0 \\ \hline 0 & 0 \end{array}\right] \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} U \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ \sigma_j^2 \\ \vdots \\ 0 \end{bmatrix}$$

$$= (\sigma_j^2) \cdot [\vec{u}_1 \dots \vec{u}_j \dots \vec{u}_m] \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \sigma_j^2 \vec{u}_j$$

ie. $\vec{u}_j$ is indeed an eigenvector of AA^T.

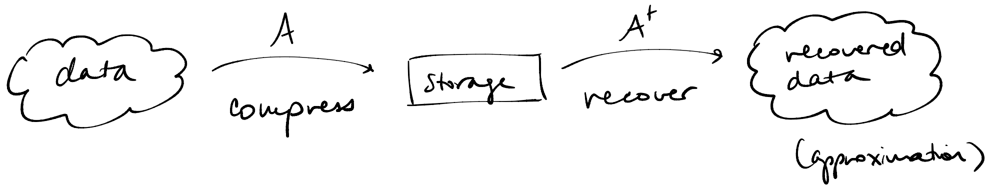
We also said the <u>pseudoinverse</u> of A is given by

$$A^+ = (A^TA)^{-1}A = VS^+U^T \quad \text{where}$$

$$\text{If } S = \left[\begin{array}{c} \begin{matrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \; 0 \end{matrix} \\ \hline 0 \end{array}\right] \quad \text{then} \quad S^+ = \left[\begin{array}{c|c} \begin{matrix} 1/\sigma_1 & & & & \\ & \ddots & & & \\ & & 1/\sigma_r & & \\ & & & 0 & \\ & & & & \ddots \; 0 \end{matrix} & 0 \end{array}\right]$$

(And similar for the other case where $n \geq m$.)

<u>Cartoon</u>:



Low-rank Approximation

Setting: $A \in \mathbb{R}^{m\times n}$ (lets say $m \geq n$), with $r = \operatorname{rank} A$.

Recall Outer product of $u \in \mathbb{R}^m, v \in \mathbb{R}^n$ is $uv^T = [v_1\vec{u} \quad v_2\vec{u} \quad \cdots \quad v_n\vec{u}]$, a rank 1 matrix.

By SVD, $A = \sum_{j=1}^{r} \sigma_j u_j v_j^T$. (Sum of r rank-1 matrices.)

Let $A_k = \sum_{j=1}^{k} \sigma_j u_j v_j^T$. This is a rank k approximation of A.

Among all $m\times n$ matrices of rank $\leq k$, A_k is the best approximation for A:

Thm For any $1 \leq k \leq r$, $\|A - A_k\|_F = \inf\limits_{\substack{B \in \mathbb{R}^{m\times n} \\ \operatorname{rank} B \leq k}} \|A - B\|_F$

ie. A_k minimizes the Frobenius norm of the residual $A - B$.

Why? $\|A - A_k\|_F = \|USV^T - US_kV^T\|_F$ where $\Sigma_k = \operatorname{diag}(\sigma_1, .., \sigma_k, 0, .., 0)$ and S_k is Σ_k extended by 0s

$= \|U(S - S_k)V^T\|_F$

$= \|U\|_F \|V\|_F \cdot \sqrt{\sigma_{k+1}^2 + \sigma_{k+2}^2 + \cdots + \sigma_r^2}$

and this is as small as possible since for other matrices B,

$$\|A - B\|_F = \|USV^T - U(U^TBV)V^T\|_F = \|U\|_F \|V\|_F \underbrace{\|S - (U^TBV)\|_F}_{\text{would have many more nonzero terms}}$$

⚠ In class I think I claimed $\|U\|_F = 1$; this is not true! It should be $\sqrt{m}$.

Similarly, $\|V\|_F = \sqrt{n}$.

How to choose k

One thing to consider: you want the approximation A_k to give rise to a well-conditioned system. (We'll see examples of these approximations in action next class.)

This brings us back to perturbation theory and condition numbers.

Recall For square, nonsingular matrix $A \in \mathbb{R}^{r\times r}$,

$$K(A) = \|A\|_2 \|A^{-1}\|_2 = \sqrt{\max \lambda(A^TA)} \cdot \sqrt{\max \lambda\left(\underbrace{(A^{-1})^T(A^{-1})}_{\substack{=(A^T)^{-1}A^{-1} \\ =(A^TA)^{-1}}}\right)}$$

$$= \sqrt{\lambda_1} \cdot \sqrt{1/\lambda_r} = \sigma_1/\sigma_r$$

In general: A any matrix.

$$K(A) = \|A\|_2 \|A^+\|_2 = \sqrt{\max \lambda(A^TA)} \cdot \sqrt{\max \lambda((A^+)^TA^+)}$$

$$= \sigma_1 \cdot \sqrt{\min \lambda(A^TA)} \quad = \sigma_1/\sigma_r.$$

How small is small enough for K(A)?

In addition to all the other errors (eg. floating point arithmetic...) perturbation theory tells us that if $K(A) \approx 10^k$, we will lose $\approx k$ digits of precision:

eg. Suppose we know some parameter should theoretically be π.

We could input to the machine $3.14\underline{15926}$.

If we apply some ($\approx$ identity) transformation and get $3.14\underline{25210}$,

Then we have lost 5 digits of precision.

ie. expect to lose $\log_{10} K(A)$ digits of precision when computing $Ax \sim b$.