Welcome Back to Ma 221! Lecture 6
Sep 6

Recall: $\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$

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

Orthogonal (+ Unitary) Matrices

Notation: $Q^* = (\overline{Q^T})$

Sometimes write Q^H

H stands for Hermitian

(if $A = A^*$, A is Hermitian)

Def: orthogonal: Q square, real. $Q^{-1} = Q^T$

unitary: Q square, complex $Q^{-1} = Q^*$

for simplicity, real case.

Fact Q orthog $\Leftrightarrow Q^T Q = I$

$\Leftrightarrow (i,j)^{th}$ of $Q^T Q$ = dot product of columns i,j of Q

$\Leftrightarrow$ all columns of Q pairwise orthogonal, and unit vectors

$Q Q^T = I$ same story for rows of Q

Fact: $\|Qx\|_2 = \|x\|_2$ Pythagorean Thm

proof: $\|Qx\|_2^2 = (Qx)^T(Qx) = x^T \underbrace{Q^T Q}_{I} x = x^T x$
$= \|x\|_2^2$

Fact: Q, Z orthog $\Rightarrow Q \cdot Z$ orthog

proof: $(QZ)^T(QZ) = Z^T \underbrace{Q^T Q}_{I} Z = \underbrace{Z^T Z}_{I} = I$

Fact: if Q $m \times n$ $m > n$ $Q^T Q = I_n$

then can add $m-n$ columns to Q

get $m[\overset{n}{Q}, \overset{m-n}{\tilde{Q}}]$ orthogonal

(proof later, infinitely many choices of $\tilde{Q}$)

Lemma (most proofs in HW Q1.16)

(1) $\|Ax\| \le \|A\| \cdot \|x\|$ for any vector norm and its operator norm

(2) $\|A \cdot B\| \le \|A\| \cdot \|B\|$ for any operator norm

(3) $\|QAZ\|_2 = \|A\|_2$ Q, Z orthog

(4) $\|Q\|_2 = 1$

(5) $\|A\|_2 = \sqrt{\lambda_{max}(A^T A)}$ ←

(6) $\|A^T\|_2 = \|A\|_2$

proof of (5)

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

$$= \max_{x \neq 0} \frac{\sqrt{(Ax)^T (Ax)}}{\sqrt{x^T x}}$$

$$= \max_{x \neq 0} \frac{\sqrt{x^T (A^T A) x}}{\sqrt{x^T x}}$$

$A^T A$ symmetric $\Rightarrow$ has eigendecomp.

(*) $A^T A q_i = \lambda_i q_i$ where λ_i all real

q_i unit orthogonal vectors

$Q = [q_1, \dots, q_n]$, orthog.

$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$

$A^T A Q = Q \Lambda \Leftrightarrow \underline{A^T A = Q \Lambda Q^T}$

$q_i^T A^T A q_i = q_i^T \lambda_i q_i = \lambda_i \geq 0$

$= \|A q_i\|_2^2$

$$\|A\|_2 = \max_{x \neq 0} \sqrt{\frac{x^T A^T A x}{x^T x}}$$

$$= \max_{x \neq 0} \sqrt{\frac{x^T Q \Lambda Q^T x}{x^T x}}$$

$$= \max_{x \neq 0} \sqrt{\frac{(x^T Q) \Lambda (Q^T x)}{(x^T Q)(Q^T x)}} \qquad y = Q^T x$$

$$= \max_{y \neq 0} \sqrt{\frac{y^T \Lambda y}{y^T y}}$$

$$= \max_{y \neq 0} \sqrt{\frac{\sum_{i=1}^n \lambda_i y_i^2}{\sum_{i=1}^n y_i^2}}$$

$$\leq \max_{y \neq 0} \sqrt{\frac{\sum_{i=1}^n \lambda_{max} y_i^2}{}}$$

$$\sum_{i=1}^{n} y_i^2$$

$= \lambda_{max}$, attainable by $y = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ if $\lambda_1 = \lambda_{max}$

SVD = Singular Value Decomposition

Given SVD of A

Solve $Ax = b$

Solve over or underdetermined LS problems, full rank A or not

compute evals, evecs of AA^T, A^TA

of A if $A = A^T$

+ error bounds (for free)

SVD = Swiss Army Knife of NLA

but more expensive than specialized alg eg Gauss. Elim for $Ax = b$

History: 1936: first complete version by Eckart + Young

1965: first backward stable alg Golub + Kahan

1990: D. + Kahan, arbitrarily more accurate for some matrices

1995: Faster, standard alg
Eisenstat + Gu

2010: Fastest alg so far
Willems: optimal complexity but not reliable enough

Thm: Suppose A $m \times m$, then $\exists$

orthogonal $V = [v(1), \ldots, v(m)]$

" $U = [u(1), \ldots, u(m)]$

diagonal $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_m)$

$\sigma_i \geq \sigma_{i+1} \geq 0$

$v(i)$ right singular vectors
$u(i)$ left " "
σ_i " values

$$A = U \cdot \Sigma \cdot V^T$$

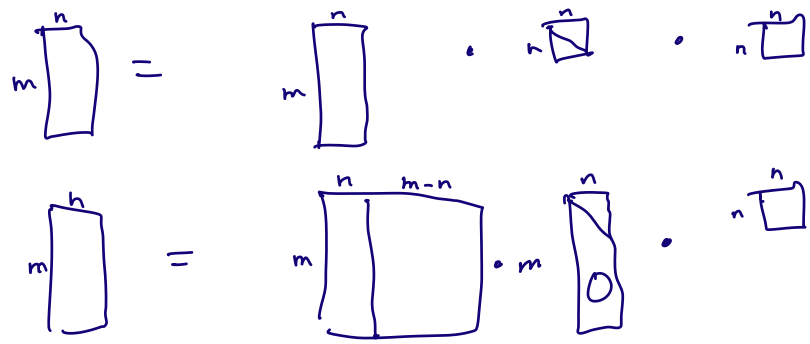
More Generally A $m \times n$, $m > n$

U $m \times m$ orthog
V $n \times n$ orthog
Σ $m \times n$, $= \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & 0 & \end{bmatrix}$

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

if $m > n$, sometimes use "thin SVD"

$$A = [u(1), \ldots, u(n)] \cdot \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \cdot [v(1) \cdots v(n)]$$



Geometric Interpretation:

$A^{m\times n}$ linear map from $\mathbb{R}^n \to \mathbb{R}^m$

with right orthogonal bases for

$\mathbb{R}^m$ and $\mathbb{R}^n$ (cols of U and V)

then A diagonal

i.e. all matrices diagonal with

right choice of bases

Proof that SVD exists, induction on n, $m \geq n$

2 base cases 1) $A^{m\times n} = 0$, $U = I_m$, $\Sigma = 0$, $V = I_n$

2) $n = 1$ (one column)

U's first col $= \dfrac{A}{\|A\|_2}$

other cols of U can be chosen in any way to make U square, orthog.

$\sigma_1 = \|A\|_2$, $V = 1$

$$\Sigma = \begin{bmatrix} \sigma_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Induction (if $A \neq 0$)

$$\|A\|_2 = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2 = 1} \|Ax\|_2$$

let $v(1)$ be x attaining max

$$\sigma_1 = \|A\|_2 = \|A\, v(1)\|_2$$

$$u(1) = \frac{A\, v(1)}{\|A\, v(1)\|_2} = \frac{A\, v(1)}{\sigma_1}$$

$$V = [v(1), \tilde{V}]$$

$$U = [u(1), \tilde{U}]$$

both square, orthog

$$\tilde{A} = U^T A V = \begin{bmatrix} u^T(1) \\ \tilde{U}^T \end{bmatrix} A \begin{bmatrix} v(1), \tilde{V} \end{bmatrix}$$

$$= \begin{array}{c c} & \begin{array}{c c} 1 & n-1 \end{array} \\ \begin{array}{c} 1 \\ m-1 \end{array} & \left[\begin{array}{c|c} u^T(1) A v(1) & u^T(1) A \tilde{V} \\ \hline \tilde{U}^T A\, v(1) & \tilde{U}^T A \tilde{V} \end{array}\right] \end{array}$$

$$= \begin{array}{c c} & \begin{array}{c c} 1 & n-1 \end{array} \\ \begin{array}{c} 1 \\ m-1 \end{array} & \left[\begin{array}{c|c} \sigma_1 & A_{12} = 0 \\ \hline A_{21} = 0 & A_{22} \end{array}\right] \end{array}$$

$$A_{21} = 0 = \tilde{U}^T A v(1) = \tilde{U}^T u(1) \cdot \sigma_1 = 0$$

$A_{12} = 0$ by def of $\sigma_1 = \|A\|_2$

if $\|A_{12}\| > 0$ get contradiction:

$$\|A\|_2 = \|A^T\|_2 = \|\tilde{A}^T\|_2 \geq \left\| \tilde{A}^T \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right\|_2 = \left\| \begin{bmatrix} \sigma_1 \\ A_{12}^T \end{bmatrix} \right\|_2$$

$$= \sqrt{\sigma_1^2 + A_{12} A_{12}^T} > \sigma_1 \text{ if } A_{12} \neq 0$$

would contradict def of σ_1 as $\max \frac{\|Ax\|_2}{\|x\|_2}$

By induction $A_{22} = U_2 \Sigma_2 \cdot V_2^T$ SVD

$$A = U \tilde{A} V^T = U \left[\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & U_2 \Sigma_2 V_2^T \end{array} \right] V^T$$

$$= U \underbrace{\left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & U_2 \end{array} \right]}_{\text{orthog}} \underset{\text{diag}}{\left[\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & \Sigma_2 \end{array} \right]} \underbrace{\left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & V_2^T \end{array} \right]}_{\text{orthog}} V^T$$

$= \text{SVD}$

most algorithms start by finding U, V so $U^T A V = \begin{bmatrix} \ddots & \ddots & & 0 \\ & \ddots & \ddots & \\ 0 & & \ddots & \ddots \end{bmatrix}$ "bidiagonal"

and then finding SVD of bidiagonal