Welcome to Ma221! Lec4 Fall 24

Finish Floating Point

Last time: exception handling

$1/0 = \text{Inf}$

$\text{Inf}+3 = \text{Inf}$ etc

$0/0 = \text{NaN}$

Impact on software:

Reliable Software

Computing $\|x\|_2 = \sqrt{\sum x_i^2}$

What could go wrong with

$s = 0$, for $i = 1:n$ $s = s + s_i^2$, $s = \sqrt{s}$

if some $s_i > \sqrt{OV} = \sqrt{\text{largest finite number}}$

then $s = \text{Inf}$, result $= \sqrt{s} = \text{Inf}$

if many $s_i \le \sqrt{UN} = \sqrt{\text{smallest positive number}}$

then $s_i^2 = 0$, s too small,

but $\sqrt{s}$ looks ok, may be very small

Reliable routines for this, and many more, (not all) exist, in BLAS library

Recent worst case examples (see webpage)

Crash of Ariane 5 rocket

Robotic car crash

Current work to make BLAS and LAPACK more reliable (class projects)

Norms, SVD, condition number for $Ax=b$

Summary of how to understand accuracy despite round off:

Show algorithms are backward stable

Scalar case for evaluating $f(x)$

instead get $alg(x) = f(x+\delta)$ δ "small"

$$\approx f(x) + f'(x)\delta$$

Error bound: $\underbrace{\left|\frac{alg(x) - f(x)}{f(x)}\right|}_{\text{relative error in output}} \approx \underbrace{\left|\frac{f'(x)\,x}{f(x)}\right|}_{\substack{\text{condition number}\\ = K(x)}} \underbrace{\left|\frac{\delta}{x}\right|}_{\text{rel error in input}}$

geometric interpretation of $\kappa(x)$

when is $\kappa(\hat{x}) = \infty$? If f smooth, f' bounded

when $f(\hat{x}) = 0$

Given x with large $\kappa(x)$ how close is x to a zero $\hat{x}$?

Newton: $\hat{x} \sim x - \dfrac{f(x)}{f'(x)}$

relative dist to nearest infinitely hard problem $= \left| \dfrac{x - \hat{x}}{x} \right| = \left| \dfrac{f(x)}{x \cdot f'(x)} \right| = \dfrac{1}{\kappa(x)}$

Same approach for $Ax=b$, $Ax=\lambda x$, ...

Get $(A+\Delta)\hat{x} = b$ where

Δ "small" compared to A

What is small?

Need vector and matrix norms

$x = f(A)$ get $alg(x) = \hat{x} = f(A+\Delta)$

error $= \hat{x} - x = f(A+\Delta) - f(A)$

if Δ small enough for Taylor expansion

error $\approx J_f(A) \cdot \Delta$ $\quad J =$ Jacobian

want to bound error $|error| \le |J_f(A)| \cdot |\Delta|$

need norms, not absolute values

Matrix and Vector Norms

Def: Let B be a linear space ($\mathbb{R}^n$ or $\mathbb{C}^n$)

B is "normed" if there is

$\|\cdot\|: B \to \mathbb{R}$ s.t

(1) $\|x\| \geq 0$ and $\|x\| = 0$ iff $x = 0$
"positive definite"

(2) $\|c \cdot x\| = |c| \cdot \|x\|$ "homogeneous"

(3) $\|x + y\| \leq \|x\| + \|y\|$ "triangle inequality"

Examples: p-norm $\|x\|_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}$, $p \geq 1$

Euclidean norm = 2-norm = $\|x\|_2$

x real $\Rightarrow$ $\|x\|_2^2 = \sum_i x_i^2 = x^T x$

∞-norm $\|x\|_\infty = \max_i |x_i|$

C-norm = $\|Cx\|$ where
C has full column rank (HW Q1.5)

Lemma (1.4) All norms are equivalent:
given any $\|\cdot\|_a$ and $\|\cdot\|_b$
there are positive constants α and β

$$\alpha \|x\|_a \leq \|x\|_b \leq \beta \|x\|_a \quad \forall x$$

(proof: compactness)

Lemma is an excuse to use easiest

norm in any proof (HW Q1.14)

Def: Matrix norm: vector norm on m·n vectors
i.e. $A^{m \times n}$

(1) $\|A\| \geq 0$ and $\|A\| = 0$ iff $A = 0$

(2) $\|c \cdot A\| = |c| \cdot \|A\|$

(3) $\|A + B\| \leq \|A\| + \|B\|$

Ex: max norm = $\max_{ij} |A_{ij}|$

Frobenius norm = $\|A\|_F = \left(\sum_{i,j} |A_{ij}|^2\right)^{1/2}$

Def: Operator norm of A: given any vector norm

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

Lemma: (1.6) An operator norm is a matrix norm
pf. HW Q 1.15

Lemma (1.7) if $\|A\|$ is an operator norm
then $\exists x$ s.t. $\|x\| = 1$ and $\|Ax\| = \|A\|$

proof: $\|A\| = \max_{x \neq 0} \|Ax\| / \|x\|$

$$= \max_{x \neq 0} \| A \underbrace{\frac{x}{\|x\|}}_{\text{unit vector}} \|$$

$$= \max_{\text{unit vectors } y} \|Ay\|$$

y attaining maximum exists
since $\|Ay\|$ continuous on
closed bounded set = unit ball

Orthogonal + Unitary Matrices (needed for SVD)

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

H stands for Hermitian

(if $A = A^H$, A called Hermitian)

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

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

for simplicity, stick to real, all generalizes

Fact Q orthogonal $\Rightarrow Q^TQ = I$

$\Leftrightarrow$ $(i,j)^{th}$ entry of $Q^TQ = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$

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

Fact $QQ^T = I$ implies same for rows

Fact $\|Qx\|_2 = \|x\|_2$ (aka Pythagorean Thm)

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

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

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

Fact: if Q $m \times n$ $n < m$ and

$Q^T Q = I_n$ then can add

$\square \, \square = \square$

$m-n$ more columns to Q, get $m\begin{array}{c} \quad n \quad m-n \\ [Q, \hat{Q}] \end{array}$

which is orthogonal (infinitely many choices of $\hat{Q}$, examples later)

Lemma (most proofs in HW Q 1.16)

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

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

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

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

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

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

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 eigendecomposition

(*) $A^T A q_i = \lambda_i q_i$ where λ_i real, $\lambda_i \ge 0$

q_i unit orthogonal vectors

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

$$\|A q_i\|_2^2 = \lambda_i \|q_i\|_2^2 = \lambda_i \geq 0$$

$$Q = [q_1, \ldots, q_n], \quad \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

$$(*) \quad A^T A Q = Q \Lambda \Rightarrow A^T A = Q \Lambda Q^T$$

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

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

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

$$= \sqrt{\max_{y \neq 0} \frac{\sum_i \lambda_i y_i^2}{\sum_i y_i^2}}$$

$$\leq \sqrt{\max_{y \neq 0} \frac{\sum_i \lambda_{max} y_i^2}{\sum_i y_i^2}} =$$

$$= \sqrt{\lambda_{max}}, \text{ attainable with } y = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

assuming $\lambda_1 = \lambda_{max}$

SVD = Singular Value Decomposition

Given SVD, can
solve $Ax=b$
solve over or underdetermined LS
problems, A full rank or not
compute eval, evecs of AA^T or A^TA
" " " of A if $A=A^T$

SVD is Swiss Army Knife of NLA
more expensive than specialized algs
so may not be first resort

History: 1936: first complete version
by Eckart + Young
1965: First Backward Stable Alg
Golub + Kahan
1990: D. + Kahan: faster, much
more accurate
1995: Gu: (need singular vectors
too)
currently fastest "reliable"
alg

2010 thesis by Paul Willems
attained $O(n^2)$
(making reliable hard class project)

Thm: Suppose A $m \times m$ Then $\exists$
orthogonal $U = [u(1), \dots, u(m)]$
diagonal $\Sigma = \mathrm{diag}(\sigma_1, \dots \sigma_m)$
$\sigma_1 \geq \sigma_2 \cdots \geq \sigma_m \geq 0$
orthogonal $V = [v(1), \dots, v(m)]$
$A = U \Sigma V^T$
$v(i)$ right singular vectors
σ_i singular values
$u(i)$ left singular vectors

In general rectangular case $A^{m \times n}$, $m > n$

U $m \times m$ orthog
V $n \times n$ orthog
Σ $m \times n$, diag

$$\begin{array}{cc} & n \\ m & \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ & & \sigma_n \\ & 0 & \end{bmatrix} \end{array}$$

$A = U \Sigma V^T$
"thin SVD"

$A = [u(1) \cdots u(n)] \cdot \mathrm{diag}(\sigma_1 \dots \sigma_n) \cdot V^T$
same idea if $A_{m \times n}$ $m < n$

Geometric interpretation:

$A^{m \times n}$: linear map from $\mathbb{R}^n \to \mathbb{R}^m$
with correct orthogonal bases for
$\mathbb{R}^n$ and $\mathbb{R}^m$, A diagonal ($= \Sigma$)

"all matrices diagonal"

proof that SVD exists: induction on $n \le m$

2 base cases

$A^{m \times n} = 0$: $U = I_m$, $\Sigma = 0$, $V = I_n$

$n = 1$ (one column):

first col of $U = \frac{A}{\|A\|_2}$

other cols of U can be chosen in any way so U orthogonal

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

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

Induction step (if $A \ne 0$)

$$\|A\|_2 = \max_{x \ne 0} \|Ax\|_2 / \|x\|_2 = \max_{\|x\|_2 = 1} \|Ax\|_2$$

Let $v(1)$ be x attaining max (pick any if not unique)

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

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

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

square, orthog

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

$$= \begin{array}{c|c|c} & 1 & n-1 \\ \hline 1 & u(1)^T A v(1) & u(1)^T A \tilde{V} \\ m-1 & \tilde{U}^T A v(1) & \tilde{U}^T A \tilde{V} \end{array} = \left[\begin{array}{c|c} \sigma_1 & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right]$$

want $A_{21} = 0$, $A_{12} = 0$

$A_{21} = 0$ by def of $\tilde{U}$

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

if $\|A_{12}\| > 0$ then

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

$$= \left\| \begin{bmatrix} \sigma_1 \\ A_{12}^T \end{bmatrix} \right\|_2 = \sqrt{\sigma_1^2 + A_{12} A_{12}^T} > \sigma_1$$

if $A_{12} \neq 0$

contradiction

Induction:

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

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

$= SVD$ as desired