Welcome to Ma221! Lecture 8, Sep 11

Start using SVD and norms to analyze condition numbers for A^{-1} and solving $Ax=b$:
If A (or A and b) change a little, how much can A^{-1} (or $A^{-1}b$) change?

Scalar case: if $|x|<1$, $\frac{1}{1-x} = \sum_{i=0}^{\infty} x^i = 1+x+x^2+\cdots$

Matrix case: If for any operator norm $\|X\|<1$
then $I-X$ invertible, $(I-X)^{-1} = \sum_{i=0}^{\infty} X^i = I+X+X^2+\cdots$
$\|(I-X)^{-1}\| \le \frac{1}{1-\|X\|}$

proof: claim: $I+X+X^2+\cdots$ converges
$\|X^i\| \le \|X\|^i \to 0$ as $i\to\infty$
$\Rightarrow$ each entry of $(I-X)^{-1} = I+X+X^2$
bounded by convergent geometric series

$$\underbrace{(I-X)(I+X+X^2+\cdots+X^i)}_{\to (I-X)^{-1} \text{ as } i\to\infty} = I - \underline{X^{i+1}} \to I \text{ as } i\to\infty$$

$$\begin{aligned}
\|(I-X)^{-1}\| &= \|I+X+X^2+\cdots\| \\
&\le \|I\| + \|X\| + \|X^2\| + \cdots \\
&\le \|I\| + \|X\| + \|X\|^2 + \cdots \\
&= 1 + \|X\| + \|X\|^2 + \cdots \\
&= 1/(1-\|X\|)
\end{aligned}$$

Generalizes to other matrix functions

$$\|e^X\| = \left\|\sum_{i=0}^{\infty} \frac{X^i}{i!}\right\| \le e^{\|X\|}$$

Lemma: if A invertible

Then $A-E$ invertible if $\|E\| < \frac{1}{\|A^{-1}\|}$

in which case

$$(A-E)^{-1} = A^{-1} + A^{-1}(EA^{-1}) + A^{-1}(EA^{-1})^2 + \cdots$$

$$\|(A-E)^{-1}\| \le \frac{\|A^{-1}\|}{1 - \underbrace{\|E\|\cdot\|A^{-1}\|}_{<1}}$$

proof: $(A-E)^{-1} = ((I - EA^{-1})A)^{-1}$

$$= A^{-1}(I - \underbrace{EA^{-1}}_{X})^{-1} \qquad \|X\| \le \|E\|\cdot\|A^{-1}\| < 1$$

$$(A-E)^{-1} = A^{-1}(I + X + X^2 + \cdots)$$

$$= A^{-1}(I + EA^{-1} + (EA^{-1})^2 + \cdots)$$

$$\|(A-E)^{-1}\| \le \|A^{-1}\| \cdot \frac{1}{1 - \|E\|\cdot\|A^{-1}\|}$$

QED

How much can A^{-1} and $(A-E)^{-1}$ differ?

Lemma: $\|(A-E)^{-1} - A^{-1}\| \le \frac{\|A^{-1}\|^2 \cdot \|E\|}{1 - \|E\|\cdot\|A^{-1}\|}$

proof: $(A-E)^{-1} - A^{-1} =$

$$A^{-1}(EA^{-1}) + A^{-1}(EA^{-1})^2 + \cdots$$

$$= A^{-1}EA^{-1}(I + EA^{-1} + (EA^{-1})^2 + \cdots)$$

take norms

$$\|(A-E)^{-1} - A^{-1}\| \le \frac{\|A^{-1}\|^2 \cdot \|E\|}{1 - \|E\|\cdot\|A^{-1}\|}$$

$$\underbrace{\frac{\|(A-E)^{-1}-A^{-1}\|}{\|A^{-1}\|}}_{\text{relative error in output}} \le \frac{\|A^{-1}\|\cdot\|A\|}{1-\frac{\|E\|}{\|A\|}\cdot\|A^{-1}\|\cdot\|A\|}\ \underbrace{\frac{\|E\|}{\|A\|}}_{\text{relative error in input}}$$

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

= condition number

Fact: $\kappa(A) \ge 1$

proof: $1 = \|I\| = \|A\cdot A^{-1}\| \le \|A\|\cdot\|A^{-1}\| = \kappa(A)$

Thm: $\underbrace{\min\left\{\frac{\|E\|}{\|A\|} : A-E \text{ singular}\right\}}_{\text{distance to singularity}} = \frac{1}{\kappa(A)}$

proof: for $\|\cdot\|_2$ using SVD:

$\min\{\|E\|_2 : A-E \text{ singular}\} = \sigma_{min}(A)$

relative dist to singularity $= \frac{\sigma_{min}(A)}{\|A\|_2}$

$= \frac{\sigma_{min}(A)}{\sigma_{max}(A)} = \frac{1}{\sigma_{max}(A)/\sigma_{min}(A)} = \frac{1}{\|A\|\cdot\|A^{-1}\|} = \kappa(A)$

Extend analysis solving $Ax=b$ vs $(A-E)\hat{x} = b+f$

$\hat{x} = x+\delta x$, solve for δx

Subtract, solve for δx:

$A\cdot\delta x - Ex - E\cdot\delta x = f$

$(A-E)\delta x = f + Ex$

$\delta x = (A-E)^{-1}(f+Ex)$

$\|\delta x\| \le \|(A-E)^{-1}\|\cdot(\|f\| + \|Ex\|)$

$$\le \frac{\|A^{-1}\|\cdot\|A\|}{1-\|E\|\cdot\|A^{-1}\|}\left(\frac{\|e\|}{\|A\|}+\frac{\|E\|\,\|x\|}{\|A\|}\right)$$

$$\underbrace{\frac{\|\delta x\|}{\|x\|}}_{\text{rel change in } x} \le \frac{\|A^{-1}\|\cdot\|A\|}{1-\|E\|\cdot\|A^{-1}\|}\left(\frac{\|e\|}{\|A\|\cdot\|x\|}+\underbrace{\frac{\|E\|}{\|A\|}}_{\text{rel change in } A}\right)$$

$$\le \underbrace{\frac{\|A^{-1}\|\cdot\|A\|}{1-\|E\|\cdot\|A^{-1}\|}}_{\text{cond \#}}\left(\underbrace{\frac{\|e\|}{\|b\|}}_{\text{rel change in } b}+\underbrace{\frac{\|E\|}{\|A\|}}_{\text{rel change in } A}\right)$$

Goal of Backward Stability:

$$\frac{\|e\|}{\|b\|},\ \frac{\|E\|}{\|A\|} = O(\varepsilon) = O(\text{machine epsilon})$$

Practical Questions

given $\hat{x}$, what is backward error?

how big is $\|E\|$?

Compute residual $r = A\hat{x} - b$

$r = A\hat{x} - Ax = A(\hat{x} - x) = A\cdot$ error

$\|\text{error}\| \le \|A^{-1}\|\cdot\|r\|$

Thm: Smallest E in norm such that $(A+E)\hat{x} = b$ has norm $\dfrac{\|r\|}{\|\hat{x}\|}$

attainable: (2-norm) $E = \dfrac{-r\cdot\hat{x}^T}{\|\hat{x}\|_2^2}$

"proof": $r = A\hat{x} - b = -E\hat{x}$

$\|r\| = \|E\hat{x}\| \le \|E\|\cdot\|\hat{x}\| \Rightarrow \|E\| \ge \dfrac{\|r\|}{\|\hat{x}\|}$

What if error bound too big?
take a few steps of Newton
but be careful about roundoff

Practical error bounds

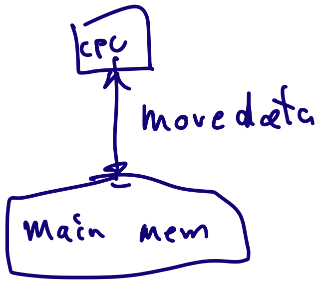
$$\|A^{-1}\| = \max_{\|x\|=1} \|A^{-1}x\| = \max_{\|x\|\le 1} \|A^{-1}x\|$$

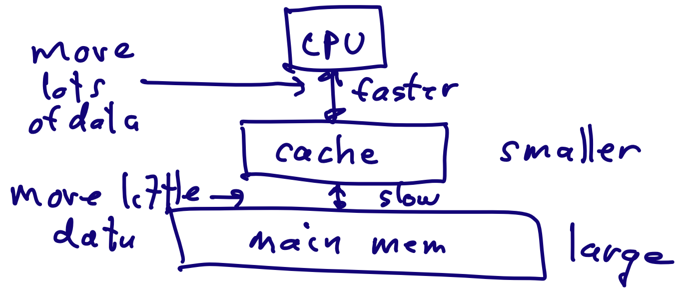
use gradient ascent "go uphill"
to maximize $\|A^{-1}x\|$ costs $O(n^2)$ per step, $O(5)$ steps works in practice

Thm (D., Diament, Malajovich, 2000)
to estimate $\|A^{-1}\|$ with any constant factor guaranteed, costs as much as mat mul.

Goal: understanding real cost in time (and energy) of running an alg.

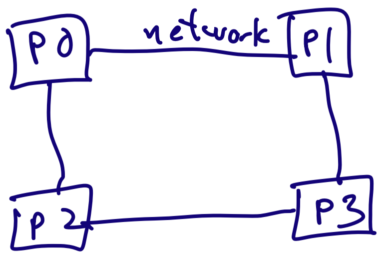
Traditionally: count flops,
but this is cheapest op.
Costs orders of magnitude more to move data from where it is stored to where you do arithmetic





Goal: minimize data movement between cache and main mem

Same idea in parallel



moving data over network very slow

Notation: "minimize communication"

Matmul: Theorm: gives a lower bound on how much data needs to move between cache and main mem to do matmul: $n \times n$ matmul
cache size M

(Hong, Kung 1981) # words moved $= \Omega\left(\frac{n^3}{\sqrt{M}}\right)$

Widely used optimal algorithm attains bound

2004: extended to parallel case

2011: extended to all dense
linear alg: GE, LS, ...