Welcome to Ma221! Lecture 12, Sep 20

Gauss. Elim (GE)

for $i = 1 : n-1$

if $A(i,i) = 0$ but $A(j,i) \neq 0, j > i$

swap rows i and j, record swap P

if all $A(j,i) = 0$, signal error

$A(i+1:n, i) = A(i+1:n, i) / A(i,i)$ ... BLAS1

$A(i+1:n, i+1:n) = A(i+1:n, i+1:n) - A(i+1:n, i) \cdot A(i, i+1:n)$ ... BLAS2

when done $A = \begin{bmatrix} U \\ L \end{bmatrix}$ $P \cdot L \cdot U =$ original A

How to "pivot", choose $A(j,i)$, or $A(i,i)$ to put on diagonal

Goal: Backward Stability:

$P \cdot L \cdot U = A + E$ $\quad \|E\| = O(\varepsilon) \cdot \|A\|$

Not guaranteed just by $A(i,i) \neq 0$

Ex: single precision $\varepsilon \approx 10^{-7}$

$$A = \begin{bmatrix} 10^{-8} & 1 \\ 1 & 1 \end{bmatrix} \quad A^{-1} = \begin{bmatrix} -1 & 1 \\ 1 & -10^{-8} \end{bmatrix}$$

$\kappa(A) \sim 2.6$ very well conditioned

$$L = \begin{bmatrix} 1 & 0 \\ 10^{8} & 1 \end{bmatrix} \quad U = \begin{bmatrix} 10^{-8} & 1 \\ 0 & fl(1 - 10^{8} \cdot 1) \end{bmatrix} = \begin{bmatrix} 10^{-8} & 1 \\ 0 & -10^{8} \end{bmatrix}$$

"forgotten"

$$L \cdot U = \begin{bmatrix} 1 & 0 \\ 10^8 & 1 \end{bmatrix} \cdot \begin{bmatrix} 10^{-8} & 1 \\ 0 & -10^8 \end{bmatrix} = \begin{bmatrix} 10^{-8} & 1 \\ 1 & 0 \end{bmatrix}$$

Get same L, U if $A(2,2)$ were $.5, -1, \ldots$
any $O(1)$ number

$\Rightarrow$ Solving $Ax = b$ gives same answer
for all these different A's, all wrong

Instead: swap rows 1 & 2, get $\begin{bmatrix} 1 & 1 \\ 10^{-8} & 1 \end{bmatrix}$
proceed with GE, all good to 8 digits

Intuition: want large entry of A on diagonal
pick largest entry: "partial pivoting"

Recall: HW 1.10 $\quad fl(A \cdot B) = A \cdot B + E$

$$|E| \le n \cdot \varepsilon \, |A| \cdot |B|$$

since $A = P \cdot L \cdot U$, get similar bound:

Thm: Backward Error for GE:
If P, L, U from GE:

$$A - E = P \cdot L \cdot U, \quad |E| \le n \cdot \varepsilon \cdot |P| \cdot |L| \cdot |U|$$

Cor: Solve $Ax = b$ by 1) GE, 2) forward sub with L
3) back sub. with U
Computed $\hat{x}$ satisfies $(A - F)\hat{x} = b$

$$|F| \le 3 \cdot n \cdot \varepsilon \cdot |P| \cdot |L| \cdot |U|$$

Proof of Cor: Use HW 1.11 (assume $P = I$)
Solve $Ly = b$, get $(L + \delta L)\hat{y} = b$, $|\delta L| \le n\varepsilon \cdot |L|$

Solve $Ux = \hat{y}$, get $(U+\delta U)\hat{x} = \hat{y}$, $|\delta U| \le n\cdot\varepsilon\cdot|U|$

$$b = (L+\delta L)\hat{y} = (L+\delta L)(U+\delta U)\hat{x}$$
$$= (LU + \delta L\cdot U + L\cdot\delta U + \delta L\cdot\delta U)\hat{x}$$
$$= (A - E + \delta L\cdot U + L\cdot\delta U + \delta L\cdot\delta U)\hat{x} \quad \text{by Thm}$$
$$= (A - F)\hat{x}$$

$$|F| \le |E| + |\delta L\cdot U| + |L\cdot\delta U| + |\delta L\cdot\delta U|$$
$$\le |E| + |\delta L|\cdot|U| + |L|\cdot|\delta U| + |\delta L|\cdot|\delta U|$$
$$\le n\varepsilon|L|\cdot|U| + n\cdot\varepsilon|L|\cdot|U| + n\cdot\varepsilon(L)\cdot|U| + n^2\varepsilon^2|L|\cdot|U|$$
$$\approx 3n\varepsilon|L|\cdot|U| \qquad \text{QED of Cor}$$

for backward stability need $\|F\| = O(\varepsilon)\|A\|$

$\|F\| \le n\varepsilon\,\|\,|L|\cdot|U|\,\|$, want

$\|\,|L|\cdot|U|\,\| = O(\|A\|)$ depends on P

Proof of Thm: Trace through Alg

$U(i,j)$ computed as

$$U(i,j) = A(i,j) - L(i,1)\cdot U(1,j) - L(i,2)\cdot U(2,j) - \cdots$$
$$= A(i,j) - \sum_{k=1}^{i-1} L(i,k)\cdot U(k,j)$$

also follows from $A(i,j)$ = dot product of row i of L & col j of U

use previous analysis of dot products

when $i > j$ same idea

$$L(i,j) = \left(A(i,j) - \sum_{k=1}^{j-1} L(i,k)\cdot U(k,j)\right)/U(i,i)$$

again dot product analysis applies QED of Thm

⇒ intuition: choose P so $|L(i,j)|$ and $|U(j,k)|$ not too large

(1) Standard: Partial pivoting (GEPP)

At each step choose largest entry (in abs. value) in each column to put on diagonal

$|L(k,i)| = |A(k,i)/A(i,i)| \le 1$

Thm (easy) with GE, $|L_{ij}| \le 1$

and $\max_{i,j}(|U(i,j)|) \le 2^{n-1} \max_{i,j}(|A(i,j)|)$

Bad news: attainable, lose all precision in single, at $n = 24$

Good news: hardly ever happens, so GEPP standard alg

Empirical observation: $\frac{\| |L| \cdot |U| \|}{\|A\|} = g \lesssim n^{2/3}$

If entries of A were i.i.d $N(0,1)$ true with high probability

(2) Complete Pivoting: Permute rows and columns so each $A(i,i)$ is largest in all remaining rows and cols

$A = P_r \cdot L \cdot U \cdot P_c$

more stable: Thm: $g < n^{(\log n + 4)}$ vs 2^{n-1}

Empirically: $g < n^{1/2}$ vs $n^{2/3}$

rarely used, cost $= O(n^3)$ additional

(3) Can't hit lower bound on latency in parallel case, instead use Tournament Pivoting

(4) Threshold Pivoting (sparse case) tradeoff stability vs. sparsity of L, U

Final Error bound for $Ax=b$, get $(A-f)\hat{x}=b$

$$\frac{\|x-\hat{x}\|}{\|x\|} \leq \text{condition number} \cdot \text{backward error}$$

$$\leq \kappa(A) \cdot 3 \cdot n \cdot \underbrace{g \cdot \|A\|}_{\| |L| \cdot |U| \|}$$

(g: growth factor)

we can estimate $\kappa(A)$, g in $O(n^2)$ extra work

What if error bound too large, or you want a guarantee?

: Try iterative refinement, aka Newton's Method

$$f(x) = Ax - b$$
$$f'(x) = A$$
$$x_{i+1} = x_i - (f'(x_i))^{-1} f(x_i)$$
$$= x_i - A^{-1}(Ax_i - b)$$
$$= A^{-1}b$$

($Ax_i - b$: residual)

why not immediate convergence?

round off

residual mostly noise

Use: mixed precision:

most work ($O(n^3)$ part) in low prec.

rest ($O(n^2)$ part) in high prec:

computing $r = Ax_i - b$

Low/high could mean

single/double

half/single

or

bfloat16

double/quad (software)

even 3 precisions in use

Do GEPP to solve $Ax=b$ in low prec

get initial solution $x(1)$

$i=1$

repeat $r = A \cdot x(i) - b$ .. in high prec

$O(n^2)$ cost

solve $Ad = r$ in low prec

$O(n^2)$ cost

update $x(i+1) = x(i) - d$ in low prec

$O(n)$ cost

until "convergence", depends on goals

(1) Getting a small backward error in high precision

straightforward stopping criterion:

$\|r\| \leq \varepsilon_{high} \cdot (\|A\| \cdot \|x_{computed}\| + \|b\|)$

(2) Getting small forward error

$\|x_{comp} - x_{exact}\| / \|x_{true}\| = O(\varepsilon)$

complicated stopping criterion:
gesvxx in LAPACK

Is it worth doing itref in one precision?

get $|E| \leq n \cdot \varepsilon \, |A|$