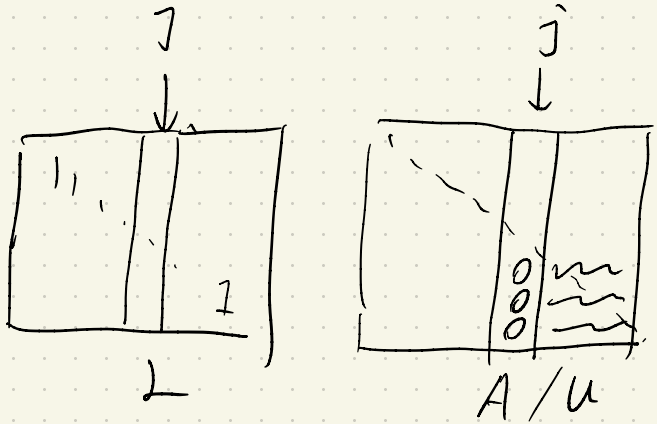
CS 4220 Feb 26th

- Recap on GEPP
- Error Analysis
- Condition Number of Solving Linear Systems



- Algo GEPP

$L = I$, $P = I$

for $j = 1 : n-1$

pick $k = \underset{l>j}{\text{argmax}} \, |A(l,j)|$

swap row k with row j in L, A, update P

$L(j+1:n, j) = A(j+1:n, j) / A(j,j)$ ← too small

$A(j+1:n, j+1:n) -= L(j+1:n, j) * A(j, j+1:n)$

$U = \text{triu}(A)$

return L, U, P

- $Ax = b$ $\quad PAx = Pb$

① $PA = LU$ ← GEPP

② $c = Pb$ ← matrix - vect

③ $Ly = c$ ← forward sub

④ $Ux = y$ ← backward sub.

← $|E| \le n \cdot \varepsilon |L||U|$

① + ③ + ④

Thm After ① ~ ④, we get vector x that solve

$(A+\delta A)x = b$

with $|\delta A| \le 3 \cdot n \varepsilon |L||U|$

- Error Analysis

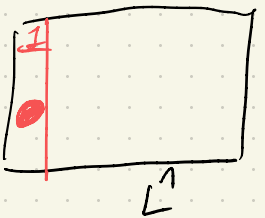
$A = LU$

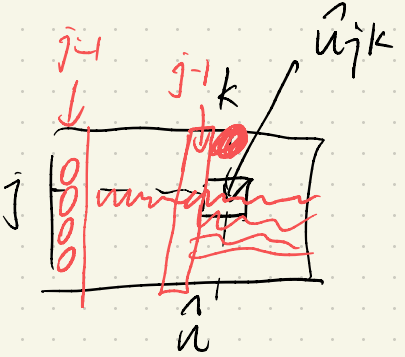
$A + E = \hat{L}\hat{U}$

entry wise.

Prop. $|E| \le n \cdot \varepsilon_{mach} \cdot |L| \cdot |U|$

proof (sketch)





if $j > k$, $\hat{u}_{jk} = 0$ at the end.

focus on the case $j \le k$

step 0: $\hat{u}_{jk} = a_{jk}$

step 1: $\hat{u}_{jk} = \hat{u}_{jk} - \hat{l}_{j2}\,\hat{u}_{1k}$

step 2: $\hat{u}_{jk} = \hat{u}_{jk} - \hat{l}_{j2}\,\hat{u}_{2k}$

$\vdots$

step $j-1$: $\hat{u}_{jk} = \hat{u}_{jk} - \hat{l}_{j\,j-1}\,\hat{u}_{j-1\,k}$

$$\hat{u}_{jk} = fl\left(a_{jk} - \sum_{i=1}^{j-1} \hat{l}_{ji}\,\hat{u}_{ik}\right) \leftarrow \begin{matrix}\text{HW1} \\ \text{Q6 (a)}\end{matrix}$$

$$= a_{jk}(1+\delta_0) - \sum_{i=1}^{j-1} \hat{l}_{ji}\,\hat{u}_{ik}(1+\delta_i) \qquad |\delta_i| \le c(j-1)\,\varepsilon_{mach}$$

$$a_{jk} = \frac{1}{1+\delta_0}\left[\hat{u}_{jk}\,\hat{l}_{jj} + \sum_{i=1}^{j-1} \hat{l}_{ji}\,\hat{u}_{ik}\,(1+\delta_i) \right]$$

Lem 1 for $|\delta| = O(\varepsilon)$

$$\frac{1}{1+\delta} = 1-\delta+\delta^2-\cdots$$
$$= 1-\delta+O(\varepsilon^2)$$

Lem 2 for $|\delta_{1,2}| = O(\varepsilon)$

$$\frac{1+\delta_1}{1+\delta_2} = 1+\delta_1-\delta_2+O(\varepsilon^2)$$

$$a_{jk} = \hat{u}_{jk}\,\hat{l}_{jj}\,(1-\delta_0) + \sum_{i=1}^{j-1} \hat{l}_{ji}\,\hat{u}_{ik}\,(1+\delta_i-\delta_0)$$

$$= (|\hat{L}||\hat{U}|)_{jk} + E_{jk}$$

where $E_{jk} = -\hat{u}_{jk}\,\hat{l}_{jj}\,\delta_0 + \sum_{i=1}^{j-1} \hat{l}_{ji}\,\hat{u}_{ik}\,(\delta_i-\delta_0) + O(\varepsilon^2)$

and can be bounded by $j\,\varepsilon_{mach}\,(|\hat{L}||\hat{U}|)_{jk} + O(\varepsilon^2)$

If we combine it together with analysis on L and on U, we prove the proposition.

- the algo is backwable if

$$\frac{\|\delta A\|}{\|A\|} = O(\varepsilon)$$

$$\|M\| = \max_{i,j} |M_{ij}|$$

$\|L\| = 1$ in GEPP

$$\|\delta A\| = O(n\varepsilon \|\hat{L}\hat{U}\|)$$

$$\leq O(n\varepsilon (\|\hat{L}\| \|\hat{U}\|))$$

$$= O(n\varepsilon \|\hat{U}\|)$$

$$\frac{\|\delta A\|}{\|A\|} = O\left(n\varepsilon \frac{\|\hat{U}\|}{\|A\|}\right)$$

define growth factor

$$g := \frac{\|\hat{U}\|}{\|A\|} = \frac{\max_{i,j} |\hat{U}_{ij}|}{\max_{i,j} |A_{ij}|}$$

— in GE

$$A = \begin{bmatrix} 10^{-20} & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 10^{20} & 1 \end{bmatrix} \begin{bmatrix} 10^{-20} & 0 \\ 0 & 1-10^{20} \end{bmatrix}$$

L U

— in GEPP

$$PA = \begin{bmatrix} 1 & 0 & 0 & 1 \\ -1 & 1 & 0 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 1 \end{bmatrix} m = \begin{bmatrix} 1 & & & \\ -1 & 1 & & 0 \\ -1 & -1 & 1 & \\ -1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & & 0 & 1 \\ & 1 & & 2 \\ & & 1 & 4 \\ 0 & & & 8 \end{bmatrix} \leftarrow$$

A L U

Thm $g \le 2^{n-1}$

by **def**. GEPP is backward stable


probability
density
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
$m = 32$
$m = 16$
$m = 8$
0
1
2
3
4
5
6
7
8
9
10
g

— Solving Linear Systems $A \in \mathbb{R}^{n \times n}$, non-singular

$$(A, b) \xrightarrow{f} x : Ax = b$$

$$(A+\delta A, b) \xrightarrow{f} x+\delta x : (A+\delta A)(x+\delta x) = b$$

$$\Rightarrow Ax + A\delta x + \delta A x + \underbrace{\delta A \delta x}_{O(\varepsilon^2)} = b$$

$$Ax = b$$

$$A\delta x + \delta A x = 0$$

$$\delta x = -A^{-1} \delta A \cdot x$$

RE in output

RE in input

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

$\kappa(A)$ condition number of A and of this problem

Relative error in the output of f, w.r.t the relative error in input

$$\frac{\|\delta x\|}{\|x\|} = \frac{\|A^{-1} \delta A \, x\|}{\|x\|} \le \|A^{-1}\| \|\delta A\|$$

triang ineq, tight

- Revisit Mat-Vect Mult

$$(A, v) \xrightarrow{f} y: \quad y = Av$$

RE $$\frac{\|y - \hat{y}\|}{\|y\|} \leq \frac{\|E\|}{\|A\|} \underbrace{\|A^{-1}\| \|A\|}_{\uparrow \text{cond}(A)}$$

where $(A + E) \cdot v = \hat{y}$

- GEPP

$$\frac{\|\delta A\|}{\|A\|} = O(n \varepsilon_g) \Rightarrow \frac{\|\delta x\|}{\|x\|} \leq O(k(A) n \varepsilon_g)$$

- $k_2(A) = \|A^{-1}\|_2 \|A\|_2$

$$= \frac{\sigma_1(A)}{\sigma_n(A)}$$