Last time:

<u>Thm</u> (Backward error of GE)

Let $A \in \mathbb{R}^{n \times n}$ and suppose GE produces computed LU

factor $\hat{L}$, $\hat{U}$ and a computed solution $\hat{x}$ to $Ax = b$

then $\qquad (A + \Delta A)\hat{x} = b$, $\qquad$ with $\qquad |\Delta A| \leq \gamma_{3n} |\hat{L}| |\hat{U}|$ $\qquad$ (⭐)

___

There are two issues regarding GE

Problem 1 : $A = LU$ doesn't always exist

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{but} \quad \text{no} \quad LU \text{ fact.} \longleftarrow \text{exercise.}$$

Problem 2 : (⭐) doesn't imply the stability of GE / Doolittle

- (⭐) provides a bound $|\Delta A| \leq \gamma_{3n} |\hat{L}| |\hat{U}|$.

  Ideally, we would like $|\Delta A| \leq C_n \varepsilon_{mach} |A|$

  If $\underline{|\hat{L}| |\hat{U}| = |\hat{L} \hat{U}|}$, we obtain that

  $|\hat{L}| |\hat{U}| = |\hat{L} \hat{U}| = |A + \Delta A|$

  $\qquad \leq |A| + \gamma_n |\hat{L}| |\hat{U}|$

  $\Rightarrow |\hat{L}| |\hat{U}| \leq \dfrac{1}{1 - \gamma_n} |A|$

  $\overset{(⭐)}{\Rightarrow} |\Delta A| \leq \dfrac{\gamma_{3n}}{1 - \gamma_n} |A|$

- $|\hat{L}| |\hat{U}| = |\hat{L} \hat{U}|$ does not always hold true. One condition

is that A being totally nonnegative, that is, if the determinant of every square submatrix is nonnegative and ε_{mach} is small enough (to ensure that $\hat{L} \approx L$ and $\hat{U} \approx U$)

- In general, we don't have this condition, and backward error can be quite large.

ex. $A = \begin{bmatrix} \varepsilon & 1 \\ 1 & 1 \end{bmatrix}$, $K_\infty(A) = \frac{4}{1-\varepsilon}$ ($\varepsilon < 1$)

Exact LU fact. $L = \begin{bmatrix} 1 & 0 \\ \frac{1}{\varepsilon} & 1 \end{bmatrix}$, $U = \begin{bmatrix} \varepsilon & 1 \\ 0 & 1-\frac{1}{\varepsilon} \end{bmatrix}$

If $fl(1-\frac{1}{\varepsilon}) = -\frac{1}{\varepsilon}$

$$\hat{L} = \begin{bmatrix} 1 & 0 \\ \frac{1}{\varepsilon} & 1 \end{bmatrix}, \quad \hat{U} = \begin{bmatrix} \varepsilon & 1 \\ 0 & -\frac{1}{\varepsilon} \end{bmatrix}$$

$$\hat{L}\hat{U} = \begin{bmatrix} \varepsilon & 1 \\ 1 & 0 \end{bmatrix}, \quad |\hat{L}||\hat{U}| = \begin{bmatrix} \varepsilon & 1 \\ 1 & \frac{2}{\varepsilon} \end{bmatrix}$$

$$A - \hat{L}\hat{U} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad \||\hat{L}||\hat{U}|\|_\infty = O(\frac{1}{\varepsilon})$$

$$\gg \|A\|_\infty = O(1)$$

Def: growth factor

$$\rho_n = \frac{\max\{|L|_\infty, |U|_\infty\}}{|A|_\infty}$$

We should use $\leftarrow |\hat{L}| \cdot |\hat{U}|$ but then the definition will involve ε_{mach}, which is cumbersome.

- Goal: Need to control $|L|$ and $|U|$

- Solution: pivoting

  At k^{th} stage of GE

  $$A^{(k)} = \begin{bmatrix} A_{11}^{(k)} & A_{12}^{(k)} \\ 0 & A_{22}^{(k)} \end{bmatrix} \qquad \begin{array}{l} A_{11}^{(k)} \in \mathbb{R}^{(k-1)\times(k-1)} \\ \text{upper triangular} \end{array}$$

  $$L = \begin{bmatrix} 1 & & & \\ l_{21} & 1 & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & 1 \end{bmatrix}, \qquad l_{jk} = \frac{a_{jk}^{(k)}}{a_{kk}^{(k)}}, \quad j = k, \cdots, n$$

- Idea: use row permutation to avoid small pivots
  and control growth in $|L|$

Option 1: Partial pivoting: Select $a_{rk}^{(k)} = \arg\max_{k \le i \le n} |a_{ik}^{(k)}|$

and permute rows

$$\underset{A}{\begin{bmatrix} x & x & x \\ x & x & x \\ \triangle & x & x \end{bmatrix}} \xrightarrow{P_1} \underset{P_1 A}{\begin{bmatrix} \triangle & x & x \\ x & x & x \\ x & x & x \end{bmatrix}} \xrightarrow{E_1} \underset{E_1 P_1 A}{\begin{bmatrix} \triangle & x & x \\ 0 & x & x \\ 0 & x & x \end{bmatrix}}$$

$\uparrow$

Select pivot $\qquad \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$

The GE with partial pivoting is equivalent to
GE without pivoting applied to row-permuted matrix

ex. $U = E_2 P_2 E_1 P_1 A$

$\qquad = E_2 \cdot \underbrace{P_2 E_1 P_2} \cdot \underbrace{(P_2 P_1 A)}_{P} \qquad$ since $P_i^2 = I$

$$= E_2 E_1' \, PA \quad \Rightarrow \quad PA = LU$$

$$\underbrace{}_{L^{-1}}$$

where $E_i' = P_n \cdots P_{i+1} E_i P_{i+1} \cdots P_n$

$$= I - \underbrace{(P_n \cdots P_{i+1}) l_i}_{=: l_i'} e_i^T$$

Let
$$E_i = I - l_i e_i^T$$
$$l_i = (0, \dots, 0, l_{i+1,i}, \dots, l_{ni})^T$$
$$e_i = (0, \dots, 0, 1, 0, \dots, 0)^T$$
$$\underset{\uparrow}{}$$
$$\text{ith entry}$$

From how we choose P_i

$$\max_{k+1 \le j \le n} |l_{jk}'| = \max_{k+1 \le j \le n} |l_{jk}| = \max_{k+1 \le j \le n} \frac{|a_{jk}^{(k)}|}{|a_{kk}^{(k)}|} \le 1.$$

hence $\quad \rho_n = \dfrac{\max\{1, |U|_\infty\}}{|A|_\infty}$

<u>Thm</u> (Wilkinson) Let $A \in \mathbb{R}^{n \times n}$, suppose GE with partial pivoting,

and $\hat{x}$ be computed solution $\hat{x}$ to $Ax = b$

then $\quad (A + \Delta A)\hat{x} = b$,

with $\|\Delta A\|_\infty \le n^2 \gamma_{3n} \rho_n \|A\|_\infty$

Some constant that polynomially depend on n

Remark: Forward error $\dfrac{\|x - \hat{x}\|}{\|x\|} \lesssim C_n \, \kappa(A) \, \rho_n \, \varepsilon_{mach}$

dependence on n?

• How large is ρ_n?

By induction $|u_{ij}| \leq 2^{i-1} \max_{k \leq i} |a_{kj}|$ $\longleftarrow$ exercise

$\Rightarrow$ $|U|_\infty \leq 2^{n-1} |A|_\infty$ $\Rightarrow$ $\rho_n \leq 2^{n-1}$ $\longleftarrow$ exponentially large in n

This bound is achievable:

$$A = \begin{bmatrix} 1 & & & 1 \\ -1 & 1 & 0 & 1 \\ -1 & -1 & 1 & & 1 \\ \vdots & \vdots & & \ddots & \vdots \\ -1 & -1 & \cdots & -1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ -1 & -1 & 1 & & \\ \vdots & \vdots & & \ddots & \\ -1 & -1 & \cdots & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & & 1 \\ & 1 & & & 2 \\ & & 1 & & \vdots \\ 0 & & & \ddots & 2^{n-2} \\ & & & & 2^{n-1} \end{bmatrix} , \qquad \rho_n = 2^{n-1}$$

- $K_\infty(A) = O(n)$

  when $\varepsilon_{mach} = 2^{-53}$, $n = 53$, $\dfrac{\|x - \hat{x}\|}{\|x\|} = O(1)$

  But this not observed in the IJulia notebook on Canvas.

  <u>why?</u> $\longleftarrow$ worst-case error

- GE with partial pivoting is NOT backward stable.

  But in practice exponential growth $\rho_n(A)$ seems "rare"

- growth factor for random matrix is small:

  $$\rho_n \approx O(\sqrt{n}) \quad \text{for} \quad a_{ij} \overset{iid}{\sim} N(0,1)$$

Option 2: Complete pivoting: Select $a_{rs}^{(k)} = \underset{k \leq i, j \leq n}{argmax} |a_{ij}^{(k)}|$

and permute rows and/or columns

- equivant to apply GE to $PAQ = LU$

permute rows ↗ permute columns ↖

- For complete pivoting, $\rho_n \leq \frac{1}{2}\sqrt{n}\; n^{\log(n)/4}$

↑ tight bound is open problem

GE is good for dense matrix without structure.

For matrix with structure, GE can be improved.

- Cholesky factorization :

$$A = LL^T \quad, \quad L \text{ low triangular}$$

Thm : If A is symmetric positive definite, then

A has a unique Cholesky factorization

with positive diagonal elements.

Pf: Since A is positive definite, we know

$\det(A(1:i, 1:i)) > 0, \quad \forall\; i = 1, \cdots, n$

This implies $A = LU$ exists ← exercise

Let $U = D\tilde{U}$ where $\tilde{U}$ has unit diagonal element

upper triangular

From symmetricity of A,

$$A^T = A \implies \tilde{U}^T D^T L^T = L D \tilde{U}$$

unit lower triangular

From uniqueness of LU,

$$\tilde{U}^T = L, \quad D^T L^T = D\tilde{U}$$

So we have $A = LD\tilde{U} = LDL^T$

Clearly $D_{ii} > 0$. Let $\tilde{L} = LD^{\frac{1}{2}}$ then $A = \tilde{L}\tilde{L}^T$

- An algorithm to compute $A = LL^T$ can be derived following how we derive Doolittle's method.

  But operation count $\approx \frac{1}{3}n^3$ ← half of GE $(\frac{2}{3}n^3)$

- Similar backward error bound for Cholesky

  $$(A + \Delta A)\hat{x} = b, \quad \text{with} \quad |\Delta A| \leq \gamma_{3n+1} |\hat{R}^T||\hat{R}|$$

  But doesn't suffer from growth factor because

  $$\| |R|^T|R| \|_2 = \| |R| \|_2^2 \leq n \|R\|_2^2 = n \|A\|_2$$

  ← exercise $\left(\begin{array}{l}\text{show that} \\ \| |A| \|_2^2 \leq rank(A) \|A\|_2^2\end{array}\right)$

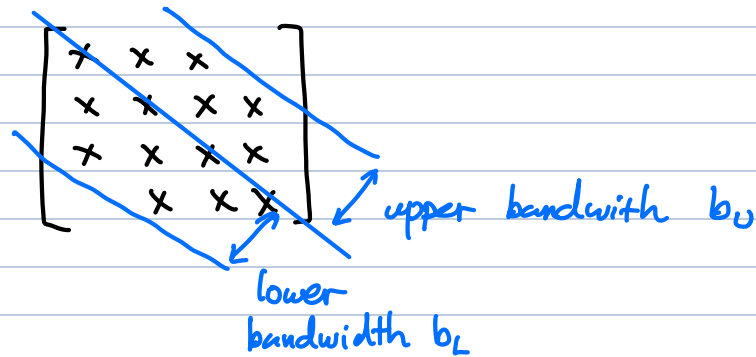- Nonpositive definite symmetric

  $$A = LDL^T$$

- Tridiagonal matrix (Thomas algorithm)

  $$A = L\ U$$

  bidiagonal

  $\begin{bmatrix} x & & & \\ x & x & & \\ & x & x & \\ & & x & x \end{bmatrix}$

Solve in $O(n)$ FLOPs.

- Banded matrix

$$\begin{bmatrix} \times & \times & \times & \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ & \times & \times & \times \end{bmatrix}$$

upper bandwith b_U

lower bandwidth b_L

$A = L\ U$ ← upper triangular with upper bandwidth b_U

↑ lower triangular with lower bandwidth b_L

With partial pivoting.

$PA = L\ U$

Solve in $O(n(b_U + b_L)^2)$

↑ lower bandwidth b_L

upper bandwidth $b_U + b_L$

- Sparse matrix: mostly zero entries ⟹ sparse direct solvers