Question 1

8

[10 marks]

Let $x,\ y \in \mathbb{R}$. Recall that $fl(x),\ fl(y) \in \mathbb{R}_b(t,s)$ denote the floating-point representations of x and y, respectively, where $fl(x) = x(1-\delta_x)$, $fl(y) = y(1-\delta_y)$, and $\delta_x,\ \delta_y$ quantify the relative roundoff errors in the respective representations.

In lecture, we showed that a typical computer estimates the product of x and y as

$$fl(fl(x)\cdot fl(y)) = (x\cdot y)(1-\delta_.)$$

where $|\delta_.| \le 3$ eps. Using similar techniques, derive a tight error bound for computer division.

$$fl[fl(x)/fl(y)] = [x(1-\delta_1)/y(1-\delta_2)](1-\delta_3)$$

$$= \frac{x(1-\delta_1)(1-\delta_3)}{y(1-\delta_2)}$$

$$= \frac{x(1-\delta_1)(1-\delta_3)(1+\delta_2)}{(1-\delta_2^2)}$$

$$\approx x(\underbrace{1-\delta_1-\delta_3+\delta_2}_{\delta_.})$$

$$= x(1-\delta_.)$$

$$|\delta_.| \le 3\cdot eps$$

Question 2

[15 marks]

Consider calculating the LU-factorization of $A \in \mathcal{R}^{5\times5}$, using Gaussian Elimination with partial pivoting. After stage 4 of the elimination we have

$$\mathcal{L}_4\,\mathcal{P}_4\,\mathcal{L}_3\,\mathcal{P}_3\,\mathcal{L}_2\,\mathcal{P}_2\,\mathcal{L}_1\,\mathcal{P}_1\,A = U \tag{1}$$

where $\mathcal{P}_i$, $\mathcal{L}_i$ are, respectively, the permutation and Gauss transform used in the i-th stage of the elimination, and U is the upper-triangular factor of the factorization.

The final form of the factorization is

$$PA = LU$$

where

$$P = \mathcal{P}_4\,\mathcal{P}_3\,\mathcal{P}_2\,\mathcal{P}_1$$

and

$$L = \tilde{\mathcal{L}}_1^{-1}\,\tilde{\mathcal{L}}_2^{-1}\,\tilde{\mathcal{L}}_3^{-1}\,\mathcal{L}_4^{-1}$$

a. Express $\tilde{\mathcal{L}}_1^{-1}$ in terms of the original $\mathcal{P}_i$ and $\mathcal{L}_i$ appearing in (1). **Show all of your work.**

$$P_4 L_3 P_3 L_2 P_2 L_1 P_1 A = L_4^{-1} U$$

$$\underbrace{P_4 L_3 P_4}_{\tilde{L}_3} P_4 P_3 L_2 P_2 L_1 P_1 A = L_4^{-1} U$$

$$P_4 P_3 L_2 P_2 L_1 P_1 A = \tilde{L}_3^{-1} L_4^{-1} U$$

$$\underbrace{P_4 P_3 L_2 P_3 P_4}_{\tilde{L}_2} P_4 P_3 P_2 L_1 P_1 A = \tilde{L}_3^{-1} L_4^{-1} U$$

$$P_4 P_3 P_2 L_1 P_1 A = \tilde{L}_2^{-1} \tilde{L}_3^{-1} L_4^{-1} U$$

$$\underbrace{P_4 P_3 P_2 L_1 P_2 P_3 P_4}_{\tilde{L}_1} P_4 P_3 P_2 P_1 A = \tilde{L}_2^{-1} \tilde{L}_3^{-1} L_4^{-1} U$$

10

$$P_4 P_3 P_2 P_1 A = \tilde{L}_1^{-1} \tilde{L}_2^{-1} \tilde{L}_3^{-1} L_4^{-1} U$$

$$\tilde{L}_1^{-1} = P_4 P_3 P_2 L_1 P_2 P_3 P_4$$

b. Given that

$$\mathcal{L}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ m_{21} & 1 & 0 & 0 & 0 \\ m_{31} & 0 & 1 & 0 & 0 \\ m_{41} & 0 & 0 & 1 & 0 \\ m_{51} & 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$\begin{aligned} \mathcal{P}_1 &\equiv \mathcal{P}_{14} \\ \mathcal{P}_2 &\equiv \mathcal{P}_{25} \\ \mathcal{P}_3 &\equiv \mathcal{P}_{34} \\ \mathcal{P}_4 &\equiv \mathcal{P}_{45} \end{aligned}$$

($\mathcal{P}_{ij}$ interchanges rows i and j for $j > i$), and considering your answer in part (**a**), write out the matrix representation of $\tilde{\mathcal{L}}_1^{-1}$ showing precisely the sign and position of the four multipliers m_{i1}.

$$P_4 P_3 P_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \qquad P_4 P_3 P_2 L_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ m_{51} & 0 & 0 & 0 & 1 \\ m_{41} & 0 & 0 & 1 & 0 \\ m_{21} & 1 & 0 & 0 & 0 \\ m_{31} & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$P_4 P_3 P_2 L_1 P_2 P_3 P_4 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ m_{51} & 1 & 0 & 0 & 0 \\ m_{41} & 0 & 0 & 1 & 0 \\ m_{21} & 0 & 0 & 0 & 1 \\ m_{31} & 0 & 1 & 0 & 0 \end{bmatrix} P_3 P_4$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ m_{51} & 1 & 0 & 0 & 0 \\ m_{41} & 0 & 1 & 0 & 0 \\ m_{21} & 0 & 0 & 0 & 1 \\ m_{31} & 0 & 0 & 1 & 0 \end{bmatrix} P_4$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ m_{51} & 1 & 0 & 0 & 0 \\ m_{41} & 0 & 1 & 0 & 0 \\ m_{21} & 0 & 0 & 1 & 0 \\ m_{31} & 0 & 0 & 0 & 1 \end{bmatrix} \qquad \tilde{L}_1^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -m_{51} & 1 & 0 & 0 & 0 \\ -m_{41} & 0 & 1 & 0 & 0 \\ -m_{21} & 0 & 0 & 1 & 0 \\ -m_{31} & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \tilde{L}_1$$

Question 3

[15 marks]

Consider the linear system $Ax = b$ where

$$A = \begin{bmatrix} 3 & 5 & 9 \\ 4 & 4 & 4 \\ 1 & 5 & 5 \end{bmatrix}, \quad b = \begin{bmatrix} 40 \\ 24 \\ 26 \end{bmatrix}.$$

8

a. Compute the $PA = LU$ factorization of A. Use exact arithmetic. Show all intermediate calculations, including Gauss transforms and permutation matrices.

$P_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ $P_1 A = \begin{bmatrix} 4 & 4 & 4 \\ 3 & 5 & 9 \\ 1 & 5 & 5 \end{bmatrix}$ $L_1 = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{3}{4} & 1 & 0 \\ -\frac{1}{4} & 0 & 1 \end{bmatrix}$ $L_1 P_1 A = \begin{bmatrix} 4 & 4 & 4 \\ 0 & 2 & 6 \\ 0 & 4 & 4 \end{bmatrix}$ $-\frac{2}{4}$

$P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ $P_2 L_1 P_1 A = \begin{bmatrix} 4 & 4 & 4 \\ 0 & 4 & 4 \\ 0 & 2 & 6 \end{bmatrix}$ $L_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{1}{2} & 1 \end{bmatrix}$ $L_2 P_2 L_1 P_1 A = \begin{bmatrix} 4 & 4 & 4 \\ 0 & 4 & 4 \\ 0 & 0 & 4 \end{bmatrix}$

U

$L_2 P_2 L_1 P_1 A = U$

$P_2 L_1 P_1 A = L_2^{-1} U$ $\quad L_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{1}{2} & 1 \end{bmatrix}$

$\underbrace{P_2 L_1 P_2}_{\tilde{L}_1} P_2 P_1 A = L_2^{-1} U$

$\tilde{L}_1 = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{4} & 1 & 0 \\ -\frac{3}{4} & 0 & 1 \end{bmatrix}$ $\tilde{L}_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{4} & 1 & 0 \\ \frac{3}{4} & 0 & 1 \end{bmatrix}$

$\underbrace{P_2 P_1}_{P} A = \underbrace{\tilde{L}_1^{-1} L_2^{-1}}_{L} U$

$L = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{4} & 1 & 0 \\ \frac{3}{4} & \frac{1}{2} & 1 \end{bmatrix}$ $U = \begin{bmatrix} 4 & 4 & 4 \\ 0 & 4 & 4 \\ 0 & 0 & 4 \end{bmatrix}$

$PA = LU$

$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$

b. Use the factorization computed in part (**a**) to solve the system.

$Ax = b$ $\quad b = \begin{bmatrix} 40 \\ 24 \\ 26 \end{bmatrix}$

$PAx = Pb$
$= \hat{b}$ $\quad \hat{b} = \begin{bmatrix} 24 \\ 26 \\ 40 \end{bmatrix}$

$LUx = \hat{b}$

$Ld = b$

$Ux = d$

Solve $Ld = \hat{b}$ first

① $\begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{4} & 1 & 0 \\ \frac{3}{4} & \frac{1}{2} & 1 \end{bmatrix} d = \begin{bmatrix} 24 \\ 26 \\ 40 \end{bmatrix}$

$d_1 = 24$

$d_2 + 6 = 26 \Rightarrow d_2 = 20$

$12 + 10 + d_3 = 40 \Rightarrow d_3 = 18$

$d = \begin{bmatrix} 24 \\ 20 \\ 18 \end{bmatrix}$ X

② $Ux = d$

$\begin{bmatrix} 4 & 4 & 4 \\ 0 & 4 & 4 \\ 0 & 0 & 4 \end{bmatrix} x = \begin{bmatrix} 24 \\ 20 \\ 18 \end{bmatrix}$

$4x_3 = 18$

$x_3 = \frac{18}{4}$

$4x_2 + 18 = 20$

$x_2 = \frac{1}{2}$

$4x_1 + 2 + 18 = 24$

$4x_1 = 4$

$x_1 = 1$

$X = \begin{bmatrix} 1 \\ \frac{1}{2} \\ \frac{9}{2} \end{bmatrix}$ X

c. Why is Gaussian Elimination usually implemented as in this question (i.e., $PA = LU$ is computed separately, and then the factorization is used to solve $Ax = b$)?

Rather than do GE to create an upper triangular matrix and then back sub. to solve x. We do LU factorization because if we wanted to solve multiple systems like $Ay = c$, $Az = d$, etc we use the LU to back and forward solve which is $O(n^2)$ whereas the GE steps take $O(n^3)$. This saves us time solving other systems.

We use pivoting to deal with scenarios where we get a 0 ~~on~~ our diagonal so we don't divide by 0 and for better roundoff properties. The pivoting cost is essentially free as well.

10

Question 4

[10 marks]

In lecture we saw that Gaussian elimination with partial pivoting usually, but not always, leads to a stable factorization of $A \in \mathcal{R}^{n\times n}$. A stable factorization is guaranteed if we use *full* pivoting, which employs both row and column interchanges before the k-th stage of the elimination to ensure that the largest element in magnitude in the $(n-k)\times(n-k)$ submatrix finds its way to the pivot position.

Full pivoting leads to a $PAQ = LU$ factorization, where P and Q are permutation matrices. Show how this factorization can be used to solve $Ax = b$.

$Ax = b$

$PAx = Pb$ $\quad$ $Pb = \hat{b}$ $\quad$ multiplying b by P is fine since it is a column vector

$PAQQx = \hat{b}$.

$L\underbrace{UQx}_{d} = \hat{b}$ $\quad$ $QQ = I$

$Ld = \hat{b}$

forward solve for d

$U\underbrace{Qx}_{f} = d$

$Uf = d$

back solve for f

$Qx = f$

$Q^{-1}Qx = Q^{-1}f$

$x = Q^{-1}f$ $\quad$ $\because Q$ is a permutation matrix $Q = Q^T$

$x = Q^T f$

Solve x with matrix-vector multiplication

CONTINUED ...

Question 5

6

[10 marks]

Explain exactly how row pivoting controls the growth of roundoff error during the LU factorization process. Use a small example and picture as was done in lecture.

If we are not using pivoting to get the largest absolute value in the column we could use a very small number that will make the values in the submatrix have bad roundoff.

Here is an example.

$$\begin{array}{c} ① \\ ② \\ ③ \end{array} \begin{bmatrix} 1 & 3 & 2 \\ 0 & 10^{-20} & 1 \\ 0 & 1 & 4 \end{bmatrix}$$

Here we use the value of 10^{-20} to eliminate the values in the column below that pivot. We do it by multiplying row ② by $\frac{a_{23}}{a_{22}}$ and subtracting from row ③.

This gets us the matrix

$$\begin{bmatrix} 1 & 3 & 2 \\ 0 & 10^{-20} & 1 \\ 0 & 0 & x \end{bmatrix}$$

So the value $x = 4 - \frac{1 \times 1}{10^{-20}}$
$= 4 - 10^{20}$

You can see here that the number x blows up in it's absolute value and we lose a lot of precision with larger absolute values as seen from assignment 1. We may not be even able to store this x in our floating point system from overflow.

Pivoting to use the largest number minimizes the error and it is more likely the number we subtract to another row will be a fraction which is more accurately represented.

SA.

- - - -