Question 1

[10 marks]

Recall that a floating-point operation, or *flop*, is an operation of the form $mx + b$. Show how to convert a $(k+1)$-digit base b ($b \neq 10$) positive integer

$$d_k\, d_{k-1} \ldots d_1\, d_0$$

into its base 10 equivalent in k flops or less.

$d_k\, d_{k-1} \ldots d_1\, d_0$

You first want to know its

multiplier product remainder

then you'd convert

$m\,(d_n d_{n-1} \ldots d_1 d_0) + b$

md_n

2

not really
go ez plz

Question 2

[15 marks]

Consider the linear system $Ax = b$ where

13 $\qquad PA = LU$

$$A = \begin{bmatrix} 2 & 5 & 10 \\ 8 & 32 & 8 \\ 1 & 8 & 13 \end{bmatrix}, \quad b = \begin{bmatrix} 7 \\ -16 \\ 6 \end{bmatrix}.$$

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

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow P_1A = \begin{bmatrix} 8 & 32 & 8 \\ 2 & 5 & 10 \\ 1 & 8 & 13 \end{bmatrix} \Rightarrow L_1 = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{4} & 1 & 0 \\ -\frac{1}{8} & 0 & 1 \end{bmatrix} \Rightarrow L_1P_1A = \begin{bmatrix} 2 & 5 & 10 \\ 0 & -3 & 8 \\ 0 & 4 & 12 \end{bmatrix}$$

$(r_1 \Leftrightarrow r_2)$

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \Rightarrow P_2L_1P_1A = \begin{bmatrix} 2 & 5 & 10 \\ 0 & 4 & 12 \\ 0 & -3 & 8 \end{bmatrix} \Rightarrow L_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{3}{4} & 1 \end{bmatrix} \Rightarrow L_2P_2L_1P_1A = \begin{bmatrix} 2 & 5 \\ 0 & 4 \\ 0 & 0 \end{bmatrix}$$

$L_2P_2L_1P_1A = U$

$L_2L_1(P_2P_2)P_1A = U$

$(P_2L_1P_2)P_2P_1A = U$

$L_2\tilde{L}_1P_2P_1A = U$

$$\underbrace{P_2P_1}_{P}A = \underbrace{\tilde{L}_1^{-1}L_2^{-1}}_{L}U$$

$$L = \tilde{L}_1^{-1} \cdot L_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{8} & 1 & 0 \\ \frac{1}{4} & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{3}{4} & 1 \end{bmatrix}$$

$$L = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{8} & 1 & 0 \\ \frac{1}{4} & -\frac{3}{4} & 1 \end{bmatrix} \checkmark \quad 5$$

$$P_2P_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \checkmark \quad 2$$

CONTINUED . . .

P, A, L, U

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

$Pb = B$

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 7 \\ -16 \\ 6 \end{bmatrix} = \begin{bmatrix} -16 \\ 6 \\ 7 \end{bmatrix} = \tilde{B}$$

by forward solve:

$Ld = \tilde{B}$

$$\begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{8} & 1 & 0 \\ \frac{1}{4} & -\frac{3}{4} & 1 \end{bmatrix}\begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} -16 \\ 6 \\ 7 \end{bmatrix}$$

$\boxed{1x = -16}$

$\frac{1}{8}x + y = 6 \Rightarrow \frac{1}{8}(-16) + y = 6 \Rightarrow \boxed{y = 8}$

$\frac{1}{4}x - \frac{3}{4}y + z = 7 \Rightarrow \frac{1}{4}(-16) + -\frac{3}{4}(8) + z = 7$

$(-4) + (-6) + z = 7 \Rightarrow \boxed{z = 17}$

$$d = \begin{bmatrix} -16 \\ 8 \\ 17 \end{bmatrix}$$

backward solve:

$Ux = d$

3

$$x = \begin{bmatrix} -21/2 \\ -1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} \cdots & \cdots & 10 \\ & \cdots & 12 \\ & & 17 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -16 \\ 8 \\ 17 \end{bmatrix}$$

$17z = 17 \Rightarrow \boxed{z = 1}$

$4y + 12z = 8 \Rightarrow 4y + 12(1) = 8 \Rightarrow \boxed{y = -1}$

$2x + 5y + 10z = -16 \Rightarrow 2x + 5(-1) + 10(1) = -$

$2x - 5 + 10 = -$

$\boxed{x = -}$

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 $Ax = b$)?

This is usually done to optimize computation the scenario RHS is too unbalanced. An example of this occurrence is iterative improv

CONTINUED

Question 3

[5 marks] 5

Consider the iterative improvement algorithm discussed in tutorial:

Solve $Ax = b$ for initial approximation $\hat{x}_0$.
for $i = 0, 1, \ldots$ until convergence
  compute $r_i = b - A\hat{x}_i$
  solve $Az_i = r_i$
  update $\hat{x}_{i+1} = \hat{x}_i + z_i$
end for

After the first iteration of this algorithm,

$$
\begin{aligned}
\hat{x}_1 &= \hat{x}_0 + z_0 \\
&= \hat{x}_0 + A^{-1}r_0 \\
&= \hat{x}_0 + A^{-1}(b - A\hat{x}_0) \\
&= \hat{x}_0 + A^{-1}b - A^{-1}A\hat{x}_0 \\
&= \hat{x}_0 + x - \hat{x}_0 \\
&= x
\end{aligned}
$$

Apparently the algorithm converges to the true solution x in just one iteration! What is the fallacy in this argument?

The fallacy in this algorithm is that z_i (which represents the approximation of the round-off error must be updated into $\hat{x}_{i+1}$ to be accounted for. This means then that there is no way it will take just 1 iteration.

Question 4

[10 marks] 10

Let $\hat{x}$ be a computed solution to $Ax = b$, $A \in \mathcal{R}^{n\times n}$. The following bound for the relative error in $\hat{x}$ was derived in class:

$$\frac{\|x - \hat{x}\|}{\|x\|} \le \text{cond}(A)\frac{\|r\|}{\|b\|},$$

where $r = b - A\hat{x}$. Starting with the equations $Ax = b$ and $A\hat{x} = b - r$, derive a *lower* bound for $\|x - \hat{x}\|/\|x\|$. What do these bounds tell us about the reliability of $\hat{x}$?

$$Ax = b \quad (1)$$
$$A\hat{x} = b - r \quad (2)$$

First subtract (2) from (1)

$$Ax - A\hat{x} = b - (b - r)$$
$$A(x - \hat{x}) = b - b + r$$
$$\|A\|\,\|x - \hat{x}\| \ge \|r\| \quad (3)$$

From $Ax = b$, we know $x = A^{-1}b$

$$\|x\| \le \|A^{-1}\|\,\|b\| \quad (4)$$

let us combine (3) & (4)

$$\frac{\|r\|}{\|A\|\,\|x - \hat{x}\|} \le \frac{\|A^{-1}\|\,\|b\|}{\|x\|}$$

$$\frac{\|r\|}{\|A\|\,\|A^{-1}\|\,\|b\|} \le \frac{\|x - \hat{x}\|}{\|x\|}$$

$$\frac{1}{\|A\|\,\|A^{-1}\|}\cdot\frac{\|r\|}{\|b\|} \le \frac{\|x - \hat{x}\|}{\|x\|}$$

$$\boxed{\frac{1}{\text{cond}(A)}\cdot\frac{\|r\|}{\|b\|} \le \frac{\|x - \hat{x}\|}{\|x\|}}$$

∴ The bounds tell us that if it is ill conditioned then we cannot gaurantee a small relative error.

CONTINUED ...

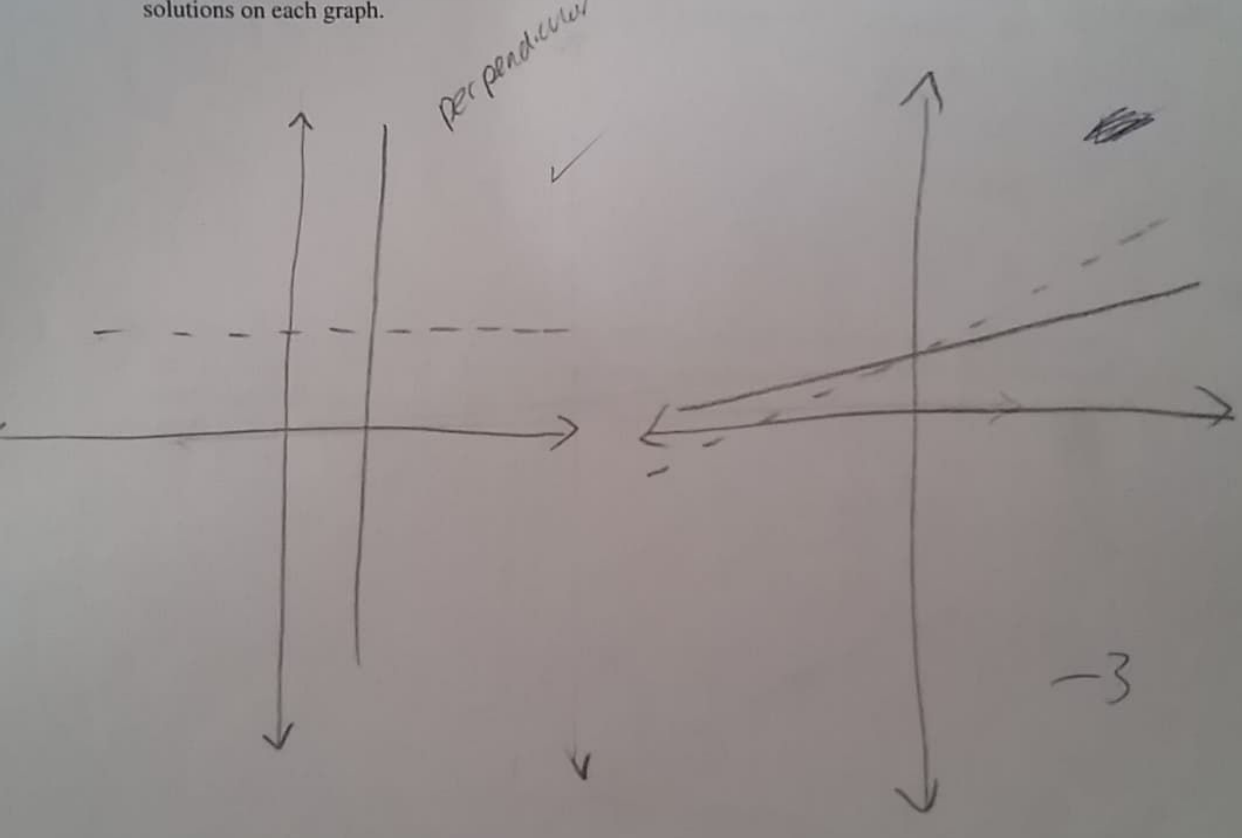
If $\hat{x}$ has a large con[...] then it is unreliable otherwise it is relia[...]

Question 5

[10 marks] 5

Recall in lecture we discussed the geometric interpretation of the manifestation of round-off error during the Gaussian Elimination/LU factorization process. We drew two graphs depicting the intersection of lines which represented, respectively, the solution of a poorly conditioned and a perfectly conditioned linear system $Ax = b, A \in \mathcal{R}^{2\times 2}, x, b \in \mathcal{R}^2$.

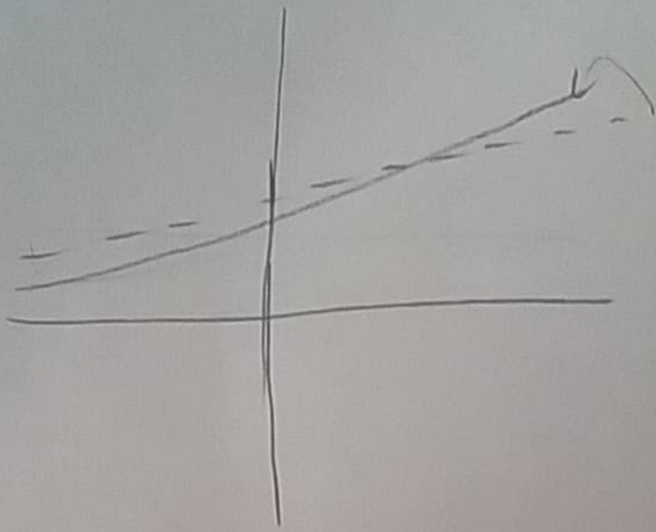
a. Reproduce the graphs below. As in lecture, draw the true systems with solid lines and the systems resulting from roundoff error with dashed lines. Clearly label the true solution and the approximate solutions on each graph.



not complete graph

CONTINUED ...

b. Copy the graph representing the poorly conditioned system to the space below. Show how the residual vector $r = b - A\hat{x}$ manifests on the graph. (**Note:** This was not discussed in lecture.)

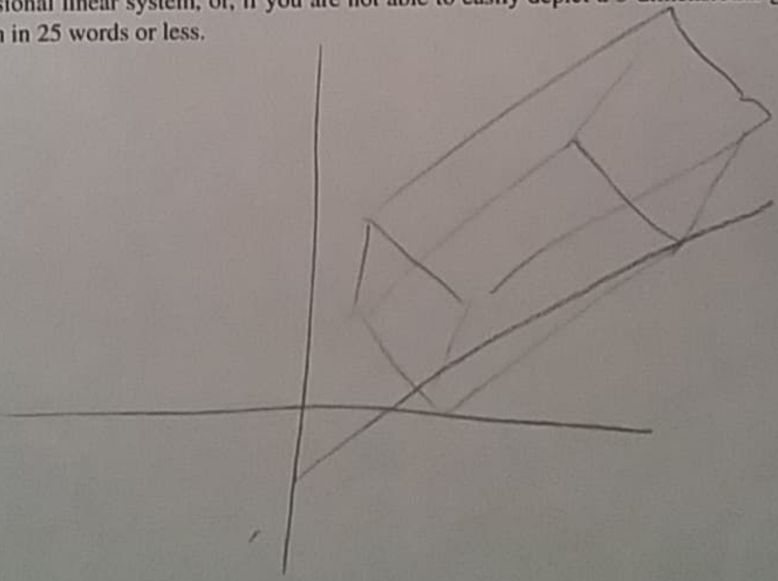


$(x + (5+\epsilon)y = b_1)$

As $\epsilon \to \infty$, $b_1 \to \infty$, showing an ill condition graph

X

c. The solution of a linear system $Ax = b, A \in \mathcal{R}^{3\times 3}, x, b \in \mathcal{R}^3$ is the line or point of intersection of three planes. In the space below, either draw a graph representing a *perfectly* conditioned 3-dimensional linear system, or, if you are not able to easily depict a 3-dimensional graph, describe the graph in 25 words or less.



$\frac{2}{3}$

END OF EXAM