

Name:

Instructions: You may not use notes or books on this exam. Don't spend too much time on any one problem. Show your work!

NAME:

1	/21	2	/15	3	/9	4	/20
5	/22	6	/13	T		/100	

[21 pts] (1a) Suppose you are using 3-digit decimal **chopped** arithmetic to solve the system  $Ax = b$  where

$$A = \begin{bmatrix} .001 & 1.00 \\ 1.00 & 2.00 \end{bmatrix}$$

and

$$b = \begin{bmatrix} 1.00 \\ 3.00 \end{bmatrix}.$$

Apply **naive** Gaussian elimination to determine the factors  $L$  and  $U$  in this case. (Hint: remember to chop your answers to three digits!)

$$\text{multiplier } M_{21} = \frac{1.00}{.001} = 1000 \quad \text{so}$$

$$L = \begin{bmatrix} 1 & 0 \\ 1000 & 1 \end{bmatrix} \quad U = \begin{bmatrix} .001 & 1.00 \\ 0 & -1000 \end{bmatrix}$$

$$\begin{aligned}
 \text{chopped arithmetic for } U_{22} &= 2.00 - (1000)(1.00) = 2 - 1000 \\
 &= -998 \quad (\text{real arithmetic}) \\
 &\approx -1000 \quad \text{in 2 digit chopped arithmetic}
 \end{aligned}$$

(b) Solve the system  $Ax = b$  using your factors  $L$  and  $U$ . Note: to get credit you should solve the two triangular systems, not the original system  $Ax = b$ . What is your computed solution in this case? (Hint: use three digit chopped decimal arithmetic.)

$$LUx = b \Rightarrow \begin{cases} Lc = b \\ Ux = c \end{cases} \quad Lc = b: \begin{bmatrix} 1 & 0 \\ 1000 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

$$\boxed{\hat{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}} \quad c_1 = 1, \quad 1000c_1 + c_2 = 3 \Rightarrow 1000 + c_2 = 3 \Rightarrow c_2 = 3 - 1000 = -997$$

In 2 digit arithmetic,  $c_2 = -1000$

$$Ux = c: \begin{bmatrix} .001 & 1.00 \\ 0 & -1000 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1000 \end{bmatrix} \Rightarrow x_2 = 1 \text{ and} \\ .001x_1 + x_2 = 1 \\ \Rightarrow .001x_1 = 1 - 1 = 0 \Rightarrow x_1 = 0$$

(c) Contrast the solution you obtained in Part (c) above with the exact solution of the original system  $Ax = b$ . Explain this difference.

$Ax = b$  in exact arithmetic is

$$\begin{bmatrix} .001 & 1.00 \\ 1.00 & 2.00 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1.00 \\ 3.00 \end{bmatrix}$$

$$\left\{ \begin{array}{l} .001x_1 + 1.00x_2 = 1.00 \\ 1.00x_1 + 2.00x_2 = 3.00 \end{array} \right.$$

$$\boxed{x = \begin{bmatrix} 1.002\dots \\ .9989\dots \end{bmatrix}}$$

$$\curvearrowleft .002x_1 + 2.00x_2 = 2.00$$

$$- \quad \underline{1.00x_1 + 2.00x_2 = 3.00}$$

$$\underline{.998x_1 = 1.00} \Rightarrow x_1 = 1.002004$$

$$\text{and } .001(1.002004) + 1.00x_2 = 1.00$$

$$\Rightarrow x_2 = 1 - .001002 = .998998$$

(d) What algorithm modification could you use to correct for the error between exact and computed solutions above (assuming you are stuck with 3-digit decimal arithmetic)? Go ahead and apply this new algorithm and see if it fixes the error.

2

Pivoting would help.

$$A = \begin{bmatrix} .001 & 1.00 \\ 1.00 & 2.00 \end{bmatrix} \quad b = \begin{bmatrix} 1.00 \\ 3.00 \end{bmatrix} \quad P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$PA = \begin{bmatrix} 1.00 & 2.00 \\ .001 & 1.00 \end{bmatrix} \quad Pb = \begin{bmatrix} 3.00 \\ 1.00 \end{bmatrix}$$

$PAx = Pb \Rightarrow$  new LU factorization

$$L = \begin{bmatrix} 1 & 0 \\ .001 & 1 \end{bmatrix}$$

$$M_{21} = \frac{.001}{1.00} = .001$$

$$U = \begin{bmatrix} 1.00 & 2.00 \\ 0 & 1.00 \end{bmatrix}$$

$$U_{22} = 1.00 - (.001)(1.00) \\ = 1.00 - .001 \approx 1.00$$

$$\begin{cases} Lc = b \\ Ux = c \end{cases}$$

Now  $Lc = b$ :  $\begin{bmatrix} 1 & 0 \\ .001 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 3.00 \\ 1.00 \end{bmatrix} \Rightarrow c_1 = 3.00$

$$(.001)(3.00) + (z = 1.00) \Rightarrow$$

$$c_2 = 1 - .003 \approx 1.00$$

$Ux = c$ :  $\begin{bmatrix} 1.00 & 2.00 \\ 0 & 1.00 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} = x_2 = 1$

$$1.00x_1 + z(1) = 3 \Rightarrow x_1 = 1$$

$$\boxed{\hat{x} = \begin{bmatrix} 1.00 \\ 1.00 \end{bmatrix}}$$

which is much closer to answer than exact solution

[15 pts](2a) Show that it takes approximately  $n^2/2$  multiplication-subtraction steps to solve each of the two triangular systems  $Lc = b$  and  $Ux = c$  which result from Gaussian Elimination.

Last unknown found in one operation (division by pivot). 2<sup>nd</sup> to last unknown requires 2 operations and so forth. (Here an "operation" includes a multiply-subtract combination.) So the total cost for back substitution is  $1+2+3+\dots+n$

$$= \frac{n(n+1)}{2} = \frac{n^2+n}{2} \text{ (which for } n \text{ large)}$$

$$\approx \frac{n^2}{2}$$

(b) What is the total cost of using elimination to solve 10 systems with the same 60x60 coefficient matrix A?

Forward elimination takes  $\approx \frac{n^3}{3}$  operations.

$n=60$  so forward elimination takes about

$\frac{160}{3}^3 \approx 72,000$  ops. But you only do this once. Back substitution takes  $\approx \frac{n^2}{2}$  ops.

$$= \frac{(60)^2}{2} = 1800 \text{ ops. So total for 10 systems}$$

$$n = 72,000 + 10(1800)(\frac{1}{2}) = \boxed{108,000 \text{ ops}}$$

$\downarrow$        $\downarrow$        $\downarrow$        $\downarrow$   
 forward      rhs's      back       $\Delta$  systems  
 elimination                substitution

[9 pts] (3) Assume  $A$  is a real symmetric matrix with an  $LDU$  decomposition. Show that if all the pivots are positive then  $A$  is positive definite. (Note: assume we are not permuting rows when we apply Gaussian elimination to get the pivots.)

Proof: In GT for symmetric matrices

$$U = L^t \quad \text{so } A = LDU = LDL^t.$$

(Note: we assumed  $A$  had an  $LDU$  decomposition.)

So positive pivots in  $D$  multiplying

perfect squares makes  $x^T A x = x^T L D L^T x$

$$= x^T L \sqrt{D} \sqrt{D} L^T x > 0.$$

( $L$  is unit lower  $\Delta$  so nonsingular.)

Specifically, we can let  $R = \sqrt{D} L^T$  which

is upper  $\Delta$ . Then  $A = R^T R$  which is

The Cholesky decomposition of  $A$

with positive diagonal elements.  $\square$

[20 pts] (4) Let  $\epsilon > 0$  be a small number and let

$$A = \begin{bmatrix} 1 & 1+\epsilon \\ 1-\epsilon & 1 \end{bmatrix}.$$

(a) Find  $\|A\|_\infty$ .

infinity norm is max row sum so

$$\|A\|_\infty = 2+\epsilon.$$

(b) Find the condition number  $\kappa_\infty(A)$ .

$$A^{-1} = \frac{1}{\epsilon^2} \begin{bmatrix} 1 & -1-\epsilon \\ -1+\epsilon & 1 \end{bmatrix}$$

$$\begin{aligned} \det(A) &= 1(1) - (1+\epsilon)(1-\epsilon) \\ &= 1 - 1 + \epsilon^2 = \epsilon^2 \end{aligned}$$

$$\|A^{-1}\|_\infty = \frac{1}{\epsilon^2} + \frac{\epsilon+1}{\epsilon^2} = \frac{2+\epsilon}{\epsilon^2}$$

$$\text{So } \kappa_\infty(A) = \|A\|_\infty \|A^{-1}\|_\infty = (2+\epsilon) \left( \frac{2+\epsilon}{\epsilon^2} \right)$$

$$\boxed{\kappa_\infty(A) = \left( \frac{2+\epsilon}{\epsilon} \right)^2}$$

(c) Now let  $\epsilon = 0.01$  and calculate  $\kappa(A)$ .

$$\kappa(A) = \left(\frac{z+\epsilon}{\epsilon}\right)^2$$

$$\text{If } \epsilon = .01 \text{ we get } \left(\frac{z+.01}{.01}\right)^2 = \left(\frac{z+.01}{.01}\right)^2 \\ = 40,401$$

(d) Imagine you are solving the system  $Ax = b$  and know that the relative error in your data vector  $b$  is  $\approx 10^{-6}$ . How many digits of the solution  $x$  can you trust?

$$\frac{\|f_x\|}{\|x\|} \leq \kappa(A) \frac{\|f_b\|}{\|b\|}$$

$$\kappa(A) \approx 40,401, \quad \frac{\|f_b\|}{\|b\|} \approx 10^{-6}$$

$$\text{So } \frac{\|f_x\|}{\|x\|} \leq (40,401)(10^{-6}) \approx .04 \cancel{\times 10^{-6}}$$

So computed solution is accurate to  
1-2 decimal places.

[22 pts] (5a) Apply the Gram-Schmidt process to the two vectors  $(1, 2, 2)$  and  $(1, 3, 1)$  in that order.

$$q_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \quad q_2 = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$$

$$\|q_1\| = \sqrt{1^2 + 2^2 + 2^2} = \sqrt{9} = 3$$

$$\text{so } q_1 = \boxed{\begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}}$$

$$\begin{aligned} q_2' &= \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} - (q_1^T q_2) q_1 = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} - \left( \begin{bmatrix} 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \right) \begin{bmatrix} 1 \\ 2/3 \\ 2/3 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} - 3 \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix} \end{aligned}$$

(b) Using the result of Part (a) above, write down the QR decomposition of the matrix

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 2 & 1 \end{bmatrix}.$$

$$\text{so } \|q_2'\| = \sqrt{2}$$

$$q_1^T q_1 = \left[ \begin{bmatrix} 1/3 & 2/3 & 2/3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \right] = \frac{1}{3} + 4/3 + 4/3 = 9/3 = 3$$

$$\boxed{q_2 = \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}} \rightarrow$$

$$q_1^T q_2 = \left[ \begin{bmatrix} 1/3 & 2/3 & 2/3 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \right] = \frac{1}{3} + 2 + 2/3 = 3$$

$$q_2^T q_2 = \left[ \begin{bmatrix} 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \right] = 3/\sqrt{2} - 1/\sqrt{2} = 2/\sqrt{2} = \sqrt{2}$$

$$\boxed{\text{so } A = QR = \begin{bmatrix} 1/3 & 0 \\ 2/3 & 1/\sqrt{2} \\ 2/3 & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 3 & 3 \\ 0 & \sqrt{2} \end{bmatrix}}$$

(c) Using the QR decomposition of  $A$ , solve the least squares problem  $Ax = b$  where  $A$  is the matrix in Part (b) above and

$$b = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}.$$

$$A = QR$$

$$\text{normal eqns: } A^t A x = A^t b$$

$$(QR)^t (QR)x = (QR)^t b$$

$$\Rightarrow R^t Q^t Q R x = R^t Q^t b$$

$$\Rightarrow R^t R x = R^t Q^t b$$

$$\Rightarrow R x = Q^t b$$

$$Q^t b = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} + \frac{4}{3} + \frac{2}{3} \\ \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{7}{3} \\ 0 \end{bmatrix}$$

and we solve

$$\begin{bmatrix} 3 & 3 \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{7}{3} \\ 0 \end{bmatrix}$$

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

so  $\bar{x} = \begin{bmatrix} 3x_1 + 3(1/2) \\ 5/18 \\ 1/2 \end{bmatrix}$

$$3x_1 + 3(1/2) = \frac{7}{3} \Rightarrow 3x_1 = \frac{7}{3} - \frac{3}{2} = \frac{14-9}{6}$$

$$\Rightarrow 3x_1 = \frac{5}{6} \Rightarrow x_1 = \frac{5}{18}$$

[13 pts](6) Show that the following operation in the back substitution algorithm for solving  $Rx = b$  is backwards stable:

$$\tilde{x}_1 = b_1 \oslash r_{11}.$$

Specifically you should show that the floating point algorithm gives exactly the right answer to a nearby problem  $(R + \delta R)x = b$ . (Here  $\oslash$  is a floating point divide.)

By the Fundamental Axiom of Floating Point Arithmetic, every floating point arithmetic operation is exact up to a relative error of at most  $\epsilon_{mach}$ . So  $\tilde{x}_1 = \frac{b_1}{r_{11}}(1+\epsilon_1)$  where  $|\epsilon_1| \leq \epsilon_{mach}$ . We want all the "data perturbations" to occur on the matrix entries of  $R$ . So we use a truncated Taylor series approximation to define  $\epsilon_1' = \frac{-\epsilon_1}{1+\epsilon_1}$  [Aside:  $\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n \quad -1 < x < 1$ . If  $x = -\epsilon_1$ , then  $\sum_{n=0}^{\infty} (-\epsilon_1)^n = \frac{1}{1+\epsilon_1}$  or  $-\epsilon_1 \sum_{n=0}^{\infty} (-\epsilon_1)^n = \frac{-\epsilon_1}{1+\epsilon_1} \Rightarrow -\epsilon_1 \approx \frac{-\epsilon_1}{1+\epsilon_1} = \epsilon_1'$ .] Then  $\frac{b_1}{r_{11}}(1+\epsilon_1) = \frac{b_1}{r_{11}(1+\epsilon_1')}$

Please sign the following honor statement: *On my honor, I pledge that I have neither given nor received any aid on this exam.*

$$= \frac{b_1}{r_{11}\left(\frac{1+\epsilon_1-\epsilon_1}{1+\epsilon_1}\right)} = \frac{b_1}{r_{11}\left(1 - \frac{\epsilon_1}{1+\epsilon_1}\right)} = \frac{b_1}{r_{11}(1+\epsilon_1')}$$

So  $\tilde{x}_1$  is exactly the correct solution to the Perturbed Problem  $(r_{11} + \delta r_{11})\tilde{x}_1 = b_1$