Problem 1

Let

$$N = \begin{bmatrix} 1 & . & . & . & . \\ . & 1 & . & . & . \\ . & . & 1 & . & . \\ . & . & \ell_{43} & 1 & . \\ . & . & \ell_{53} & . & 1 \end{bmatrix}, \quad M = \begin{bmatrix} 1 & . & . & . & . \\ . & 1 & . & . & . \\ . & . & 1 & . & . \\ . & . & -\ell_{43} & 1 & . \\ . & . & -\ell_{53} & . & 1 \end{bmatrix}$$

where ℓ_{43} and ℓ_{53} are arbitrary real numbers. Show that $NM = I$ and $MN = I$, i.e., that M is the inverse of N.

Proof. First, consider NM. Observe that the product the first three rows remain unchanged because their corresponding rows in both N and M are standard basis vectors. The only nontrivial row operations occur in the fourth and fifth rows of N. The fourth row of N is $(0, 0, \ell_{43}, 1, 0)$, and we compute its dot product with each column of M:

- Dot product with the third column of M:

$$(0, 0, \ell_{43}, 1, 0) \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \\ -\ell_{43} \\ 0 \end{bmatrix} = \ell_{43} - \ell_{43} = 0.$$

- Dot product with the fourth column of M:

$$(0, 0, \ell_{43}, 1, 0) \cdot \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = 1.$$

- The remaining columns give the dot product of zero.

Thus, the fourth row of NM is $(0, 0, 0, 1, 0)$. The fifth row of N is $(0, 0, \ell_{53}, 0, 1)$, and we compute its dot product with each column of M:

- Dot product with the third column of M:

$$(0, 0, \ell_{53}, 0, 1) \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ -\ell_{53} \end{bmatrix} = \ell_{53} - \ell_{53} = 0.$$

- Dot product with the fifth column of M:

$$(0, 0, \ell_{53}, 0, 1) \cdot \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = 1.$$

- The remaining columns give the dot product of zero.

Thus, the fifth row of NM is $(0,0,0,0,1)$. Since all rows match the identity matrix, we conclude:

$$NM = I.$$

Now, consider the product MN:

$$MN = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -\ell_{43} & 1 & 0 \\ 0 & 0 & -\ell_{53} & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \ell_{43} & 1 & 0 \\ 0 & 0 & \ell_{53} & 0 & 1 \end{bmatrix}$$

Observe that this is exactly expression NM; where we set $-\ell_{43}$ to ℓ_{43} and $-\ell_{53}$ to ℓ_{53}. Since the above equation holds for any value of $\ell_{53}, -\ell_{43}$, the proof follows directly, giving us $NM = I = MN$. ☐

Problem 2

Let

$$M_1^{-1} = \begin{bmatrix} 1 & . & . & . & . \\ \ell_{21} & 1 & . & . & . \\ \ell_{31} & . & 1 & . & . \\ \ell_{31} & . & . & 1 & . \\ \ell_{21} & . & . & . & 1 \end{bmatrix}, \quad M_2^{-1} = \begin{bmatrix} 1 & . & . & . & . \\ . & 1 & . & . & . \\ . & \ell_{32} & 1 & . & . \\ . & \ell_{42} & . & 1 & . \\ . & \ell_{52} & . & . & 1 \end{bmatrix} \quad M_3^{-1} = \begin{bmatrix} 1 & . & . & . & . \\ . & 1 & . & . & . \\ . & . & 1 & . & . \\ . & . & \ell_{43} & 1 & . \\ . & . & \ell_{53} & . & 1 \end{bmatrix} \quad M_4^{-1} = \begin{bmatrix} 1 & . & . & . \\ . & 1 & . & . \\ . & . & 1 & . \\ . & . & . & 1 \\ . & . & . & \ell_{54} \end{bmatrix}$$

By successively computing $M_1^{-1}M_2^{-1}$, $(M_1^{-1}M_2^{-1})M_3^{-1}$, and $((M_1^{-1}M_2^{-1})M_3^{-1})M_4^{-1}$, show that

$$M_1^{-1}M_2^{-1}M_3^{-1}M_4^{-1} = \begin{bmatrix} 1 & & & & \\ \ell_{21} & 1 & & & \\ \ell_{31} & \ell_{32} & 1 & & \\ \ell_{41} & \ell_{42} & \ell_{43} & 1 & \\ \ell_{51} & \ell_{52} & \ell_{53} & \ell_{54} & 1 \end{bmatrix}$$

Proof. First consider the product $M_1^{-1}M_2^{-1}$:

$$M_1^{-1}M_2^{-1} = \begin{bmatrix} 1 & . & . & . & . \\ \ell_{21} & 1 & . & . & . \\ \ell_{31} & . & 1 & . & . \\ \ell_{41} & . & . & 1 & . \\ \ell_{51} & . & . & . & 1 \end{bmatrix} \begin{bmatrix} 1 & . & . & . & . \\ . & 1 & . & . & . \\ . & \ell_{32} & 1 & . & . \\ . & \ell_{42} & . & 1 & . \\ . & \ell_{52} & . & . & 1 \end{bmatrix} = \begin{bmatrix} 1 & . & . & . & . \\ \ell_{21} & 1 & . & . & . \\ \ell_{31} & \ell_{32} & 1 & . & . \\ \ell_{41} & \ell_{42} & . & 1 & . \\ \ell_{51} & \ell_{52} & . & . & 1 \end{bmatrix}$$

Now consider $(M_1^{-1}M_2^{-1})M_3^{-1}$:

$$(M_1^{-1}M_2^{-1})M_3^{-1} = \begin{bmatrix} 1 & . & . & . & . \\ \ell_{21} & 1 & . & . & . \\ \ell_{31} & \ell_{32} & 1 & . & . \\ \ell_{41} & \ell_{42} & . & 1 & . \\ \ell_{51} & \ell_{52} & . & . & 1 \end{bmatrix} \begin{bmatrix} 1 & . & . & . & . \\ . & 1 & . & . & . \\ . & . & 1 & . & . \\ . & . & \ell_{43} & 1 & . \\ . & . & \ell_{53} & . & 1 \end{bmatrix} = \begin{bmatrix} 1 & . & . & . & . \\ \ell_{21} & 1 & . & . & . \\ \ell_{31} & \ell_{32} & 1 & . & . \\ \ell_{41} & \ell_{42} & \ell_{43} & 1 & . \\ \ell_{51} & \ell_{52} & \ell_{53} & . & 1 \end{bmatrix}$$

Finally, consider $((M_1^{-1}M_2^{-1})M_3^{-1})M_4^{-1}$:

$$((M_1^{-1}M_2^{-1})M_3^{-1})M_4^{-1} = \begin{bmatrix} 1 & . & . & . & . \\ \ell_{21} & 1 & . & . & . \\ \ell_{31} & \ell_{32} & 1 & . & . \\ \ell_{41} & \ell_{42} & \ell_{43} & 1 & . \\ \ell_{51} & \ell_{52} & \ell_{53} & . & 1 \end{bmatrix} \begin{bmatrix} 1 & . & . & . & . \\ . & 1 & . & . & . \\ . & . & 1 & . & . \\ . & . & . & 1 & . \\ . & . & . & \ell_{54} & 1 \end{bmatrix} = \begin{bmatrix} 1 & . & . & . & . \\ \ell_{21} & 1 & . & . & . \\ \ell_{31} & \ell_{32} & 1 & . & . \\ \ell_{41} & \ell_{42} & \ell_{43} & 1 & . \\ \ell_{51} & \ell_{52} & \ell_{53} & \ell_{54} & 1 \end{bmatrix}$$

This proves our statement. □

Problem 3*

GEPP applied to an invertible matrix $A \in \mathbb{R}^{3\times 3}$ produces a factorization

$$M_2P_2M_1P_1A = U$$

where L is unit lower triangular, U is upper triangular, and P_1 and P_2 are permutation matrices. Recall that M_1 and M_2 are elementary lower triangular matrices of the form

$$M_1 = \begin{bmatrix} 1 & 0 & 0 \\ -m_{21} & 1 & 0 \\ -m_{31} & 0 & 1 \end{bmatrix}, \quad M_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -m_{32} & 1 \end{bmatrix},$$

and that P_1 may permute any rows while P_2 may only permute the second and third rows.

(a) Explain why the factorization can be written as

$$\tilde{M}_2\tilde{M}_1PA = U$$

where $\tilde{M}_2 = M_2$, $\tilde{M}_1 = P_2M_1P_2$, and $P = P_2P_1$.

Proof. Expanding the equation, we get:

$$\tilde{M}_2\tilde{M}_1PA = M_2P_2M_1P_2P_2P_1A$$

Observe that as P permutes only the first and the third rows, it could either swap them or leave them as they are. If the rows are not swapped $P_2 = I$ – it trivially follows $P_2^2 = I$. If the rows are swapped, multiplying by P_2 again swaps row 3 (formerly 2) and 2 (formerly 3) again; giving us $P_2^2 = I$. Plugging this back, we get:

$$M_2P_2M_1P_2P_2P_1 = M_2P_2M_1P_1A = U$$

This proves our statement. □

(b) Note that there are two possibilities for the permutation matrix P_2. For both cases, explicitly compute the matrix $\tilde{M}_1$.

- If $P_2 = I$ (no row swap), then $\tilde{M}_1 = M_1$:

$$\tilde{M}_1 = \begin{bmatrix} 1 & 0 & 0 \\ -m_{21} & 1 & 0 \\ -m_{31} & 0 & 1 \end{bmatrix}$$

- If P_2 swaps the second and third rows:

$$P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Then, applying P_2 to M_1 gives:

$$\tilde{M}_1 = P_2 M_1 P_2 = \begin{bmatrix} 1 & 0 & 0 \\ -m_{31} & 0 & 1 \\ -m_{21} & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -m_{31} & 1 & 0 \\ -m_{21} & 0 & 1 \end{bmatrix}$$

(c) For both possibilities of P_2, find the unit lower triangular matrix L such that $PA = LU$.

Proof. Note that we have $\tilde{M}_2 \tilde{M}_1 PA = U$, subsequently we need $L = \tilde{M}_1^{-1} \tilde{M}_2^{-1}$. Recall that these are lower triangular matrices; subsequently, we get:

$$\tilde{M}_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & m_{32} & 1 \end{bmatrix}$$

- If $P_2 = I$, then:

$$\tilde{M}_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ m_{21} & 1 & 0 \\ m_{31} & 0 & 1 \end{bmatrix}$$

Which gives us (following the proof from problem 2):

$$L = M_1^{-1} M_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ m_{21} & 1 & 0 \\ m_{31} & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & m_{32} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ m_{21} & 1 & 0 \\ m_{31} & m_{32} & 1 \end{bmatrix}$$

- If P_2 swaps rows 2 and 3:

$$\tilde{M}_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ m_{31} & 1 & 0 \\ m_{21} & 0 & 1 \end{bmatrix}$$

$$L = \tilde{M}_1^{-1} M_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ m_{31} & 1 & 0 \\ m_{21} & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & m_{32} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ m_{31} & 1 & 0 \\ m_{21} & m_{32} & 1 \end{bmatrix}$$

□

Problem 4*

Analyze the computational cost of obtaining the LU decomposition via GE, assuming no pivoting is required.

(a) In the kth step of GE, compute the number of flops required for the multipliers in matrix M_k.

Proof. At the kth step of GE, we need to compute the multipliers $m_{ik} = \frac{a_{ik}}{a_{kk}}$ for $k+1 \leq i \leq n$. This requires $n-k$ divisions (also flops). Summing over all steps $k=1$ to $n-1$:

$$\sum_{k=1}^{n-1}(n-k) = \frac{n(n-1)}{2} = O(n^2).$$

□

(b) Compute the number of flops for matrix-vector product $M_k x$.

Proof. Recall from Pset 1, problem 3 that the number of flops for a matrix (size $n \times n$) vector (size n) product is $O(n^2)$. □

(c) Compute the number of flops for the matrix-matrix product $M_k A_{k-1} = A_k$.

Proof. Again recall from Pset 1, problem 3, the flops required for the matrix-matrix product $M_k A_{k-1} = A_k$.is $(2n-1)n^2 = O(n^3)$. □

(d) Use these results to show that GE to produce the factorization $A = LU$ costs $O(n^3)$ flops.

Proof. Recall that GE uses matrix multiplications to apply elimination steps, producing the lower triangular matrix L from multipliers M_k and the upper triangular matrix U from the transformed A. Each elimination step updates A using $M_k A_{k-1}$ (matrix-matrix multiplication) and computes the effect on partial updates using $M_k x$ (matrix-vector multiplication). Combining all steps gives us $O(n^2) + O(n^2) + O(n^3) = O(n^3)$ as the cost.

□

Problem 5* (Datta 5.13)

Apply GEPP and GECP to the following matrices:

$$A_1 = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 1 \\ -1 & -1 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 100 & 99 & 98 \\ 98 & 55 & 11 \\ 0 & 1 & 1 \end{bmatrix}.$$

Show all work and state all matrices in intermediate factorizations for both GEPP and GECP. Compute the growth factor in all cases.

Proof. We consider this problem in four parts:

1. Gaussian Elimination with Partial Pivoting (GEPP) for A_1

Recall that in GEPP, we select the largest absolute value in the current column as the pivot.

- The largest entry in column 1 is already in position (1,1), so no row swaps are needed.
- Elimination proceeds with finding multipliers:

$$m_{21} = \frac{-1}{1} = 1, \quad m_{31} = \frac{-1}{1} = 1.$$

- Updating the second and third rows:

$$R_2 \leftarrow R_2 - m_{21}R_1, \quad R_3 \leftarrow R_3 - m_{31}R_1.$$

  This gives us:

$$U = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & -1 & 2 \end{bmatrix}.$$

- Updating the third row ($m_{32} = \frac{-1}{1} = -1.$):

$$R_3 \leftarrow R_3 - m_{32}R_2.$$

  This results in:

$$U = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 4 \end{bmatrix}.$$

Gaussian Elimination with Complete Pivoting (GECP) for A_1

Recall that in GECP, we select the largest absolute value in the remaining submatrix.

- The largest absolute value is 1 at (1,1), so no swaps are needed. The elimination steps are identical to GEPP, giving us

$$U = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & -1 & 2 \end{bmatrix}.$$

- The largest absolute value in the remaining submatrix is 1 at position (3,2), so we swap column 2 and 3.

$$U = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 2 & -1 \end{bmatrix}.$$

- Updating the third row ($m_{32} = \frac{2}{2} = -1.$):

$$R_3 \leftarrow R_3 - m_{32}R_2.$$

  This results in:

$$U = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & -2 \end{bmatrix}.$$

Growth Factor

- The element with the largest absolute value of U at any step for GEPP is 4, the element with the largest absolute value of A is 1; thus the growth factor is $4/1 = 4$.

- The element with the largest absolute value of U at any step for GECP is 2, the element with the largest absolute value of A is 1; thus the growth factor is $2/1 = 2$.

3. Gaussian Elimination with Partial Pivoting (GEPP) for A_2

Recall that $A_2 = \begin{bmatrix} 100 & 99 & 98 \\ 98 & 55 & 11 \\ 0 & 1 & 1 \end{bmatrix}$

- The largest entry in column 1 is already in position (1,1), so no row swaps are needed.
- Elimination proceeds with multipliers:

$$m_{21} = \frac{98}{100} = 0.98$$

- Updating the second row:

$$R_2 \leftarrow R_2 - 0.98R_1$$

  This gives us:

$$\begin{bmatrix} 100 & 99 & 98 \\ 0 & 55 - 0.98(99) & 11 - 0.98(98) \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 100 & 99 & 98 \\ 0 & -42.02 & -85.04 \\ 0 & 1 & 1 \end{bmatrix}$$

- Updating the third row $m_{32} = \frac{1}{-42.02} = -0.0238$:

$$R_3 \leftarrow R_3 - (-0.0238)R_2$$

  This gives us:

$$U = \begin{bmatrix} 100 & 99 & 98 \\ 0 & -42.02 & -85.04 \\ 0 & 0 & -1.0238 \end{bmatrix}.$$

4. Gaussian Elimination with Complete Pivoting (GECP) for A_2

- The largest absolute value is 1 at (1,1), so no swaps are needed. The elimination steps are identical to GEPP, giving us:

$$U = \begin{bmatrix} 100 & 99 & 98 \\ 0 & -42.02 & -85.04 \\ 0 & 1 & 1 \end{bmatrix}$$

- The largest absolute value in the remaining submatrix is 1 at position (3,2), so we swap columns 2 and 3.

$$U = \begin{bmatrix} 100 & 98 & 99 \\ 0 & -85.04 & -42.02 \\ 0 & 1 & 1 \end{bmatrix}$$

- Updating the third row $m_{32} = \frac{1}{-85.04} = -0.0118$:

$$R_3 \leftarrow R_3 + 0.0118R_2$$

  This gives us:

$$U = \begin{bmatrix} 100 & 98 & 99 \\ 0 & -85.04 & -42.02 \\ 0 & 0 & 0.5058 \end{bmatrix}$$

Growth Factor

- The element with the largest absolute value of U at any step for GEPP is 100, the element with the largest absolute value of A is 100; the growth factor is $100/100 = 1$.
- The element with the largest absolute value of U at any step for GECP is 100, the element with the largest absolute value of A is 100; thus the growth factor is $100/100 = 1$.

□

Problem 6

Explain how the linear system $Ax = b$ may be solved using the $PAQ = LU$ decomposition produced by GECP.

Proof. Recall that for $A = LU$ we can solve $Ly = b$ by forward substitution and $Ux = y$ by back substitution. Now observe that we know $LU = PAQ \implies P^{-1}LUQ^{-1} = A$. We need to solve:

$$P^{-1}LUQ^{-1}x = b.$$

To proceed, define $y_i = LUQ^{-1}x$ and $x_i = Q^{-1}x$. First, note that we can solve $P^{-1}y_i = b \implies y_i = Pb$ by multiplication. Then, note that we can find $Ly = y_i$ by forward substitution. This gives us $Ux_i = y$ which is solved by backward substitution. Lastly, we get $Q^{-1}x = x_i \implies x = Qx_i$ which is solved by multiplication. □

Problem 7

Explain how $\log(|\det(A)|)$ can be computed using each of the factorizations $A = LU$ (as in GE), $PA = LU$ (as in GEPP), and $PAQ = LU$ (as in GECP).

Proof. Consider this proof in three parts:

A. $A = LU$ **(Gaussian Elimination without Pivoting)**:

$$\det(A) = \det(L)\det(U).$$

Since L is a unit lower triangular matrix, we know $\det(L) = 1$, so:

$$\det(A) = \det(U).$$

Since U is an upper triangular matrix, its determinant is the product of its diagonal entries:

$$\det(A) = \prod_i U_{ii}.$$

Taking the logarithm gives us:

$$\log(|\det(A)|) = \sum_i \log(|U_{ii}|).$$

B. $PA = LU$ **(GEPP)** : Since P is a permutation matrix, we know $\det(P) = \pm 1$. This gives us:

$$\det(A) = \det(P^{-1})\det(L)\det(U) = \det(U).$$

This gives us the same expression as part (A) :

$$\log(|\det(A)|) = \sum_i \log(|U_{ii}|).$$

C. $PAQ = LU$ **(GECP)** Since P and Q are both permutation matrices, and permutations only change the sign of the determinant (i.e., $\det(P), \det(Q) = \pm 1$), we again get:

$$|\det(A)| = |\det(P^{-1})\det(L)\det(U)\det(Q^{-1})| = |\det(U)|.$$

Thus, the expression remains the same:

$$\log(|\det(A)|) = \sum_i \log(|U_{ii}|).$$

□

Problem 8 (AG 5.23)

Let $b + \delta b$ be a perturbation of a vector b ($b \neq 0$), and let x and δx be such that

$$Ax = b \quad \text{and} \quad A(x + \delta x) = b + \delta b,$$

where A is a given nonsingular matrix. Show that

$$\frac{\|\delta x\|_2}{\|x\|_2} \leq \kappa(A)\frac{\|\delta b\|_2}{\|b\|_2}.$$

Proof. Observe that subtracting the two equations gives us:

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

which simplifies to:

$$A\delta x = \delta b \implies \delta x = A^{-1}\delta b$$

Taking the norm of both sides, we get:

$$\|\delta x\|_2 = \|A^{-1}\delta b\|_2.$$

Using the sub-multiplicative property of the 2-norm, we can bound the right-hand side:

$$\|\delta x\|_2 \leq \|A^{-1}\|_2\|\delta b\|_2.$$

Now consider the first equation $Ax = b$. Taking the 2-norm of both sides gives us;

$$\|Ax\|_2 = \|b\|_2.$$

Using the sub-multiplicative property of the 2-norm, we get:

$$\|b\|_2 = \|Ax\|_2 \leq \|A\|_2\|x\|_2.$$

which gives us:

$$\frac{1}{\|A\|_2\|x\|_2} \leq \frac{1}{\|b\|_2} \implies \frac{1}{\|x\|_2} \leq \frac{\|A\|_2}{\|b\|_2}$$

Multiplying our two expressions together, we get:

$$\|\delta x\|_2\frac{1}{\|x\|_2} \leq \|A^{-1}\|_2\|\delta b\|_2\frac{\|A\|_2}{\|b\|_2}$$

Reorganizing our expression:

$$\frac{|\delta x\|_2}{\|x\|_2} \leq \|A\|_2\|A^{-1}\|_2\frac{\|\delta b\|_2}{\|b\|_2}$$

Recall that the condition number of A is defined as:

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

This gives us:

$$\frac{|\delta x\|_2}{\|x\|_2} \leq \kappa(A)\frac{\|\delta b\|_2}{\|b\|_2}$$

$\square$

Problem 9

This problem builds on Problem 4. A matrix $T \in \mathbb{R}^{n\times n}$ is said to be *tridiagonal* if $t_{ij} = 0$ whenever $|i-j| > 1$ for $i, j = 1, \dots, n$. A matrix $H \in \mathbb{R}^{n\times n}$ is said to be *upper Hessenberg* if $h_{ij} = 0$ whenever $i > j + 1$.

Part (a)

Give a detailed description of how the factorization $H = LU$ (assuming that no pivoting is required) can be obtained in only $O(n^2)$ flops when H is upper Hessenberg.

Part (b)

Give a detailed description of how the factorization $T = LU$ (assuming that no pivoting is required) can be obtained in only $O(n)$ flops when T is triangular.

Problem 10

Compute (by hand, although you may use code to check your work) the Cholesky decomposition of the matrix

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 5 & 5 \\ 1 & 5 & 14 \end{bmatrix}.$$

Show all of your work.

Proof. The hand-written computation is as follows:

we need to find:

$$\begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix}$$

st. $l_{11}^2 = 1$, $\quad l_{11} \cdot l_{21} = 1 \quad l_{11} \cdot l_{31} = 1$

since it needs to be positive $\to l_{11} = 1$ $\quad l_{21} = 1 \quad l_{31} = 1$

$l_{21} \cdot l_{21} + l_{22} l_{22} = 5$

$1 \cdot 1 + l_{22}^2 = 5$

$\hookrightarrow l_{22} = \sqrt{5-1}$

$= 2$

$l_{21} \cdot l_{31} + l_{22} \cdot l_{32} = 5$

$1 \cdot 1 + 2 \cdot l_{32} = 5$

$\hookrightarrow l_{32} = 2$

$l_{31}^2 + l_{32}^2 + l_{33}^2 = 14$

$1 + 4 + l_{33}^2 = 14 \quad \to l_{33} = \sqrt{14-5} = \sqrt{9} = 3$

$$A = LL^T = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 3 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix}}_{L^T} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 5 & 5 \\ 1 & 5 & 14 \end{bmatrix}$$

□

Problem 11

Let $A = LL^\top$ be the Cholesky factorization of a symmetric positive definite matrix $A \in \mathbb{R}^{n\times n}$.

Part (a)

Explain why the diagonal of A can contain no negative (or zero) entries.

Proof. Note that L is a lower triangular matrix whose diagonal entries are strictly positive. Then, $L^\top$ is a upper triangular matrix whose diagonal entries are strictly positive; with the same entries on the diagonal. Observe that, as all other entries are zero, the diagonal entries of A are given by:

$$a_{ii} = \sum_{k=1}^{i} l_{ik}^2.$$

Since each term in the sum is nonnegative and at least one term l_{ii}^2 is strictly positive (because L has positive diagonal entries), it follows that $a_{ii} > 0$ for all i. Thus, the diagonal entries of A are strictly positive and cannot contain negative or zero entries. □

Part (b)

Show that

$$l_{ij}^2 \leq a_{ii}, \quad i, j = 1, \ldots, n,$$

i.e., the squares of the entries in any row of L are bounded above by the corresponding diagonal entry of A.

Proof. From the Cholesky factorization, we have:

$$A = LL^\top.$$

Recall from part (a), expanding the diagonal elements gives us:

$$a_{ii} = \sum_{k=1}^{i} l_{ik}^2.$$

Since each term in this sum is nonnegative, we immediately get:

$$l_{ij}^2 \leq \sum_{k=1}^{i} l_{ik}^2 = a_{ii}.$$

This proves our statement.

□

Acknowledgments

Mathstackexchange for the sub-multiplicative property of the

Problems solved: (need 8)

- ☑ Problem 1
- ☑ Problem 2
- ☑ Problem 3*
- ☑ Problem 4*
- ☑ Problem 5*
- ☑ Problem 6
- ☑ Problem 7
- ☑ Problem 8
- ☐ Problem 9
- ☑ Problem 10
- ☑ Problem 11