Matrix Computation Homework 1

Yifan Zhang 2025251018 zhangyf52025@shanghaitech.edu.cn

November 2, 2025

Problem 1.(Subspaces and Decompositions)

1)

If a set can be determined as a subspace, it will satisfy these three condition:

1. Contain the zero vector.
2. Closure under addition.
3. Closure under scaler multiplication.

question a)

Proof.

- For the **condition 1**, it is obvious that (0,0) belong to S_1.
- For the **condition 2**. **Define.** $v_1 = (x_1, y_1)$, $v_2 = (x_2, y_2)$.
  Then $v_1 + v_2 = (x_1 + x_2, y_1 + y_2)$.
  Since v_1 and v_2 belong to set S_1,
  it is clear that $x_1 - 2y_1 = x_2 - 2y_2 = 0$.
  So $x_1 + x_2 - 2(y_1 + y_2) = 0$.
  So it satisfy condition 2.
- For the **condition 3**
  Let $a \in \mathbb{R}$, $v = (x, y)$,
  then $a * v = (ax, ay)$.
  It is clear that $x - 2y = 0$.
  So $ax - 2 * ay = a * (x - 2y) = a * 0 = 0$.
  So $a * v \in \mathbb{R}$.
  So S_1 satisfy the condition 3.

In conclusion, S_1 is a subspace of $\mathcal{V}$

question b)

Proof.

- For the **condition 1**, it is obvious that (0,0) belong to S_1.
- For the **condition 2**. **Define.** $v_1 = (x_1, y_1)$, $v_2 = (x_2, y_2)$.

  Then $v_1 + v_2 = (x_1 + x_2, y_1 + y_2)$.

  Since v_1 and v_2 belong to set S_1,

  it is clear that $x_1 + y_1 = x_2 + y_2 = 0$.

  So $x_1 + x_2 + y_1 + y_2 = 0$.

  So it satisfy condition 2.
- For the **condition 3**

  Let $a \in \mathbb{R}$, $v = (x, y)$,

  then $a * v = (ax, ay)$.

  If a is irrational number and $x + y \neq 0$, then $ax + ay = a(x + y)$ is clearly irrational number, i.e., $ax + ay \notin Q$

  So S_1 not satisfy the condition 3.

In conclusion, S_1 is not a subspace of $\mathcal{V}$

2)

question a)

Proof.
To be a subspace, the set must satisfy the three conditions mentioned above.
For $\mathcal{V}_+$:

- For **condition 1**:

  Let $\alpha = 0, \beta = 0$, it is clear that $\Phi(\mathbf{0}) = \mathbf{0}$. So the zero vector is in $\mathcal{V}_+$, i.e., it satisfy the condition 1.
- For **condition 2**:

  Let $\alpha = \beta = 1$, then $A+B = \Phi(A)+\Phi(B) = \Phi(A+B)$, i.e., $(A+B) \in \mathcal{V}_+$. So it satisfy the condition 2.
- For **condition 3**:

  It is clear that $\Phi(A) = A$, then **Let** $\beta = 0$, we can get $\Phi(\alpha A) = \alpha\Phi(A) = \alpha A$. So $\alpha A \in \mathcal{V}_+$. It satisfy the condition 3.

Above all, $\mathcal{V}_+$ is a subspace of $\mathcal{V}$ over $\mathbb{R}$
For $\mathcal{V}_-$:

- For **condition 1**:

  Let $\alpha = 0, \beta = 0$, it is clear that $\Phi(\mathbf{0}) = \mathbf{0}$. So the zero vector is in $\mathcal{V}_-$, i.e., it satisfy the condition 1.

- For **condition 2**:

  Let $\alpha = \beta = 1$, then $A + B = -\Phi(A) - \Phi(B) = -\Phi(A + B)$, i.e., $(A + B) \in \mathcal{V}_-$. So it satisfy the condition 2.

- For **condition 3**:

  It is clear that $\Phi(A) = -A$, then **Let** $\beta = 0$, we can get $\Phi(\alpha A) = \alpha\Phi(A) = -\alpha A$.So $\alpha A \in \mathcal{V}_-$. It satisfy the condition 3.

Above all, $\mathcal{V}_-$ is a subspace of $\mathcal{V}$ over $\mathbb{R}$

question b)

Proof.

If we want to prove that every matrix $A \in \mathcal{V}$ can be written in exactly one way as

$$A = A_+ + A_-,$$

where $A_+ \in V_+$ and $A_- \in V$, we need to prove that:

- A can be represented as $A_+ + A_-$, where $A_+ \in V_+$ and $A_- \in V$.
- A_+andA_- are unique.

First, we prove the **existence**:
Let

$$A_+ = \frac{1}{2}(A + \Phi(A)), A_- = \frac{1}{2}(A - \Phi(A)).$$

Then it is clear that $A = A_+ + A_-$.
We just need to prove that $A_+ \in \mathcal{V}_+$ and $A_- \in \mathcal{V}_-$.

$$\begin{aligned}
\Phi(A_+) &= \Phi\left(\frac{1}{2}(A + \Phi(A))\right) \\
&= \frac{1}{2}\Phi(A + \Phi(A)) \\
&= \frac{1}{2}(\Phi(A) + \Phi(\Phi(A))) \\
&= \frac{1}{2}(\Phi(A) + A) \\
&= A_+
\end{aligned}$$

i.e., $A_+ \in \mathcal{V}_+$.
Similarly, we can also get that $A_- \in \mathcal{V}_-$.

So the existence can be proved and we have shown that $\mathcal{V}$ can be represented as $V_+ + V_-$. Since $\mathcal{V}_+, \mathcal{V}_-$ are the subspace of $\mathcal{V}$, $\mathcal{V} = \mathcal{V}_+ + \mathcal{V}_-$ **can be shown**

Next, we prove the **uniqueness**:

Use proof by contradiction here. We assume that there also exist $B_+ \neq A_+$ and $B_- \neq A_-$ satisfy the condition.

Then $\mathcal{V} = A_+ + A_- = B_+ + B_-$ and we can get that $A_+ - B_+ = B_- - A_- = C$

Since

$$A_+, B_+ \in \mathcal{V}_+, \quad A_-, B_- \in \mathcal{V}_-,$$

we can get that

$$A_+ - B_+ \in \mathcal{V}_+, \quad B_- - A_- \in \mathcal{V}_-,$$

i.e.,

$$C \in \mathcal{V}_+, \quad C \in \mathcal{V}_-.$$

Then $\Phi(C) = C = -\Phi(C)$.

So we can get that $C = \{0\}$.

Then $\mathcal{V}_+ \cap \mathcal{V}_+ = \{0\}$ can be shown.

Since $C = \{0\}$, $A_+ - B_+ = B_- - A_- = \{0\}$, we can get that $A_+ = B_+$ and $A_- = B_-$.

This contradicts the hypothesis.

So the uniqueness can be proved.

Problem 2.(Span and Rank)

1)

question a)

Proof. If we want to proof $\mathcal{R}(L) = span\{L(v_1), \dots, L(V_n)\}$, we must show two set containments:

- $\mathcal{R}(L) \subseteq span\{L(v_1), \dots, L(v_n)\}$
- $span\{L(v_1), \dots, L(v_n)\} \subseteq \mathcal{R}(L)$

Part 1. $\mathcal{R}(L) \subseteq span\{L(v_1), \dots, L(v_n)\}$:

Let w be an arbitrary vector in $\mathcal{V}(L)$. Then there exists vector v and $L(v) = w$

v can be represented as $\alpha_1 v_1 + \cdots + \alpha_n v_n$

Then we can get:

$$\begin{aligned} w &= L(v) \\ &= L(\alpha_1 v_1 + \cdots + \alpha_n v_n) \\ &= \alpha_1 L(v_1) + \cdots + \alpha_n L(v_n) \end{aligned}$$

Since w is expressed as a linear combination of the vectors $\{L(v_1), L(v_2), \ldots, L(v_n)\}$, it follows by definition that w is in the span of these vectors.

$$w \in span\{L(v_1), L(v_2), \ldots, L(v_n)\}$$

Thus, $\mathcal{R}(L) \subseteq span\{L(v_1), \ldots, L(v_n)\}$.

Part 2.$span\{L(v_1), \ldots, L(v_n)\} \subseteq \mathcal{R}(L)$:

Let w be an arbitrary vector in $span\{L(v_1), \ldots, L(v_n)\}$, Then

$$w = \sum_{i=1}^{n} \alpha_i L(v_i) \qquad = \sum_{i=1}^{n} L(\alpha_i v_i) = L(\sum_{i=1}^{n} \alpha_i v_i)$$

Let

$$u = \sum_{i=1}^{n} \alpha_i v_i$$

It is clear that $u \in \mathbb{R}^n$

So any w in $span\{L(v_1), \ldots, L(v_n)\}$ can be represented as $L(u), u \in \mathbb{R}^n$.

Since

$$\mathcal{R}(L) := L(v) \in \mathbb{R}^m | v \in \mathbb{R}^n,$$

$L(u) \in \mathcal{R}(L)$, i.e., $w \in \mathcal{R}(L)$

Thus, $span\{L(v_1), \ldots, L(v_n)\} \subseteq \mathcal{R}(L)$.

Above all, we can prove that $\mathcal{R}(L) = span\{L(v_1), \ldots, L(V_n)\}$.

question b)

Proof.

If we want to prove that all the outputs through this network are $\mathcal{R}(W) + b$, we just need to prove that

- $y \in \mathcal{R}(w) + b$ for any input x.
- The arbitrary vector in $\mathcal{R}(w) + b$ can be represented as y, i.e., $Wx + b$.

Part 1. $y \in \mathcal{R}(w) + b$ for any input x:

Let x be an arbitrary vector in $\mathbb{R}^n$ and

$$x = (x_1, \ldots, x_n)^T, \quad W = (w_1, \ldots, w_n).$$

Then

$$\begin{aligned} y &= Wx + b \\ &= \sum_{i=1}^{n} x_i W_i + b \end{aligned}$$

Define. $\{v_1, \dots, v_k\}$ are the **Maximal linearly independent set** of W.
Thus $x_i w_i$ can be represented as $\sum_{j=1}^{k} \alpha_{ij} v_j$.
Thus we can easily get

$$\begin{aligned} y &= \sum_{i=1}^{n} \sum_{j=1}^{k} \alpha_{ij} v_j + b \\ &= \sum_{\beta_j}^{k} \beta_j v_j + b \end{aligned}$$

Thus we can prove that $y \in \mathcal{R}(w) + b$ for any input x.

Part 2. The arbitrary vector in $\mathcal{R}(w) + b$ can be represented as y, i.e., $Wx + b$.

We can easily get $\mathcal{R}(w) = \{u \in U | u = W(x), x \in \mathbb{R}^n\}$.
Thus any vector in $\mathcal{R}(w) + b$ can be represented as $Wx + b$.

Above all, all the outputs through this network are $\mathcal{R}(W) + b$.

2)

Proof.

Define. $\{x_1, \dots, x_k\}$ are the **Maximal linearly independent set** of A
Let

$$X = (x_1, \dots, x_k), \qquad A = (a_1, \dots, a_n)$$

Thus a_1 can be represented as $\sum_{j=1}^{k} \alpha_{1j} x_j$.
Let $a_i = \sum_{j=1}^{k} \alpha_{ij} x_j$, Then

$$A = \{\sum_{j=1}^{k} \alpha_{ij} x_j, \dots, \sum_{j=1}^{k} \alpha_{nj} x_j\}$$

Thus we can **Let**

$$Y^T = \begin{bmatrix} \alpha_{11} & \alpha_{21} & \cdots & \alpha_{n1} \\ \alpha_{12} & \alpha_{22} & \cdots & \alpha_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1k} & \alpha_{2k} & \cdots & \alpha_{nk} \end{bmatrix}$$

Then there exits matrices X and Y to satisfy:

$$A = XY^T$$

Since $A = XY^T$, we can get that $min\{rank(X), rank(Y)\} \geq rank(A) = p$. Since $X \in \mathbb{R}^{m \times p}$ and $Y \in \mathbb{R}^{n \times p}$, we can get that $rank(X) \leq p, rank(Y) \leq p$.
Thus $rank(X) = rank(Y) = p$

Problem 3. (Flop Counting and Algorithm Complexity)

1)

Operation	Dimensions	Flops
$\alpha = \mathbf{u}^\top \mathbf{v}$	$\mathbf{u}, \mathbf{v} \in \mathbb{R}^p$	$2p$
$\mathbf{w} = \mathbf{w} + \beta\mathbf{u}$	$\beta \in \mathbb{R}, \mathbf{u}, \mathbf{w} \in \mathbb{R}^p$	$2p$
$\mathbf{z} = \mathbf{z} + \mathbf{Mu}$	$\mathbf{M} \in \mathbb{R}^{q \times p}, \mathbf{u} \in \mathbb{R}^p, \mathbf{z} \in \mathbb{R}^q$	$2pq + q$
$\mathbf{N} = \mathbf{N} + \mathbf{vu}^\top$	$\mathbf{N} \in \mathbb{R}^{q \times p}, \mathbf{u} \in \mathbb{R}^p, \mathbf{v} \in \mathbb{R}^q$	$2pq$
$\mathbf{D} = \mathbf{D} + \mathbf{PQ}$	$\mathbf{P} \in \mathbb{R}^{q \times r}, \mathbf{Q} \in \mathbb{R}^{r \times p}, \mathbf{D} \in \mathbb{R}^{q \times p}$	$2pqr + pq$

2)

Solution.
It is clear that the matrix H can be represented as $ABC \circ D$.
Based on the matrix expression, we can design the following algorithm:

- Step 1: Compute the intermediate matrix $P = AB$.
- Step 2: Compute the matrix $M = PC$. (This gives $M = ABC$).
- Step 3: Compute the final matrix $H = M \circ D$.

Complexity Analysis:

- Step 1 ($P = AB$): This is a standard matrix-matrix multiplication of two $n \times n$ matrices. This operation takes $O(n^3)$ flops.
- Step 2 ($M = PC$): This is another $n \times n$ matrix-matrix multiplication. This operation also takes $O(n^3)$ flops.
- Step 3 ($H = M \circ D$): This is a Hadamard product (element-wise multiplication) of two $n \times n$ matrices. This operation takes $O(n^2)$ flops.

Thus,

$$\textbf{Total Complexity} = O(n^3) + O(n^3) + O(n^2) = O(n^3)$$

Problem 4. (Norms)

1)

Proof.
First, we can easily get that

$$||w||_1 = |w_1|+|w_2|+\cdots+|w_n|, \quad ||w||_\infty = max|w_1|, |w_2|, \ldots, |w_n|, \quad ||w||_2^2 = w_1^2+w_2^2+\cdots+w_n^2$$

Let $w_k = max\{|w_1|, |w_2|, \ldots, |w_n|\} = ||w||_\infty$.
Then

$$||w||_1||w||_\infty = (w_k + \sum_{i\neq k}^{n} w_i)w_k$$

$$\frac{1+\sqrt{n}}{2}||w||_2^2 = \frac{1+\sqrt{n}}{2}(w_k^2 + \sum_{i\neq k}^{n} w_i^2)$$

Thus, we just need to show that

$$(w_k + \sum_{i\neq k}^{n} w_i)w_k \leq \frac{1+\sqrt{n}}{2}(w_k^2 + \sum_{i\neq k}^{n} w_i^2)$$

$$\textbf{i.e.} \quad w_k^2 + (\sum_{i\neq k}^{n} w_i)w_k \leq \frac{1+\sqrt{n}}{2}w_k^2 + \frac{1+\sqrt{n}}{2}\sum_{i\neq k}^{n} w_i^2$$

Applying the **Cauchy-Schwarz Inequality**, we can easily get

$$\sum_{i\neq k}^{n} w_i \leq \sqrt{\sum_{i\neq k}^{n} w_i^2}\sqrt{n-1}$$

Thus

$$w_k^2 + (\sum_{i\neq k}^{n} w_i)w_k \leq w_k^2 + w_k\sqrt{n-1}\sqrt{\sum_{i\neq k}^{n} w_i^2}$$

Let $x = w_k, \quad S = \sum_{i\neq k}^{n} w_i^2$

We just need to prove that

$$x^2 + x\sqrt{n-1}\sqrt{S} \leq \frac{1+\sqrt{n}}{2}x^2 + \frac{1+\sqrt{n}}{2}S$$

$$\textbf{i.e.} \quad \frac{\sqrt{n}-1}{2}x^2 - \sqrt{n-1}\sqrt{S}x + \frac{1+\sqrt{n}}{2}S \geq 0$$

It is clear that $S \geq 0, \quad x \geq 0, \quad \frac{\sqrt{n}-1}{2} \geq 0.$

We just focus on x. Since $-\frac{b}{2a} = \frac{\sqrt{n-1}}{\sqrt{n}-1}\sqrt{S} \geq 0$, if we want the inequality to hold, we need

$$\Delta = b^2 - 4ac = (n-1)S - 4 \cdot \frac{\sqrt{n}-1}{2} \cdot \frac{1+\sqrt{n}}{2} S \leq 0.$$

Since $S \geq 0$, we just need to prove

$$\begin{aligned} n - 1 - 4 \cdot \frac{\sqrt{n}-1}{2} \cdot \frac{1+\sqrt{n}}{2} &\leq 0 \\ \textbf{i.e.} \quad n - 1 - (n-1) &\leq 0 \end{aligned}$$

That is clearly true.

2)

Proof.
Part 1. We first prove $||E||_F = ||u||_2||v||_2$

Since $E = uv^T$, we can get that $E_{ij} = u_i \cdot v_j$.
Thus

$$\begin{aligned} ||E||_F &= \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}(u_i v_j)^2} \\ &= \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} u_i^2 v_j^2} \\ &= \sqrt{\sum_{i=1}^{m} u_i^2 \sum_{j=1}^{n} v_j^2} \\ &= \sqrt{\sum_{i=1}^{m} u_i^2}\sqrt{\sum_{j=1}^{n} v_j^2} \\ &= ||u||_2||v||_2 \end{aligned}$$

Part 2. We next prove $||E||_2 = ||u||_2||v||_2$

$$||E||_2 = \sqrt{\lambda_{\max}(E^T E)}$$

Next, we focus on the eigenvalue of $E^T E$.

It is clear that

$$\begin{aligned} E^T E &= (uv^T)^T uv^T \\ &= vu^T uv^T \\ &= v(u^T u)v^T \\ &= (u^T u)vv^T \qquad (u^T u \ \textit{is a scalar}) \end{aligned}$$

Let $u^T u = ||u||_2^2 = k, \quad vv^T = A$, then we can easily get $\lambda(EE^T) = k\lambda(A)$. Next, we will find the eigenvalues of matrix A.

$$Ax = \lambda x \implies (vv^T)x = \lambda x$$

Since $v^T x$ is a scaler, we just need to solve $(v^T x)v = \lambda x$. It is clear that there are two case can satisfy this condition:

- **Case 1.** $x = v$, i.e., x is parallel to v.
- **Case 2.** $\lambda = v^T x = 0$, i.e., x is orthogonal to v.

Case 1:

$$(vv^T)v = \lambda v \implies \lambda = v^T v = ||v||_2^2$$

Case 2:

$$\lambda = 0$$

Thus

$$||E||_2 = \sqrt{k||v||_2^2} = \sqrt{||u||_2^2||v||_2^2} = ||u||_2||v||_2$$

3)

Proof.

$$\begin{aligned} \|A\|_1 &= \max_{1 \le j \le n} \sum_{i=1}^{m} |a_{ij}| && \text{(Maximum absolute column sum)} \\ \|A\|_\infty &= \max_{1 \le i \le m} \sum_{j=1}^{n} |a_{ij}| && \text{(Maximum absolute row sum)} \\ \|A\|_2^2 &= \sup_{\|x\|_2 = 1} \|Ax\|_2^2 = \sup_{\|x\|_2 \neq 0} \frac{\|Ax\|_2^2}{\|x\|_2^2} && \text{(Spectral norm squared)} \end{aligned}$$

Let $x \in \mathbb{R}^n$ be an arbitrary vector, $(Ax)_i$ be the i-th component of the vector Ax.

$$\|Ax\|_2^2 = \sum_{i=1}^{m} ((Ax)_i)^2 = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j \right)^2$$

We can write $a_{ij}x_j$ as $\sqrt{|a_{ij}|} \cdot \sqrt{|a_{ij}|}x_j$. Then we apply the Cauchy-Schwarz inequality and get:

$$\left(\sum_{j=1}^{n} a_{ij}x_j\right)^2 \leq \left(\sum_{j=1}^{n}(\sqrt{|a_{ij}|})^2\right)\left(\sum_{j=1}^{n}(\sqrt{|a_{ij}|}x_j)^2\right) = \left(\sum_{j=1}^{n}|a_{ij}|\right)\left(\sum_{j=1}^{n}|a_{ij}|x_j^2\right)$$

Since $\|A\|_\infty = \max_{1\leq i\leq m}\sum_{j=1}^{n}|a_{ij}|$, then

$$\|Ax\|_2^2 \leq \sum_{i=1}^{m}\left(\sum_{j=1}^{n}|a_{ij}|\right)\left(\sum_{j=1}^{n}|a_{ij}|x_j^2\right) \leq \|A\|_\infty \sum_{i=1}^{m}\left(\sum_{j=1}^{n}|a_{ij}|x_j^2\right)$$

Next change the sum sequence can get that

$$\sum_{i=1}^{m}\left(\sum_{j=1}^{n}|a_{ij}|x_j^2\right) = \sum_{j=1}^{n}x_j^2\sum_{i=1}^{m}|a_{ij}|$$

Since $\|A\|_\infty = \max_{1\leq i\leq m}\sum_{j=1}^{n}|a_{ij}|$, then

$$\|Ax\|_2^2 \leq \|A\|_\infty\|A\|_1\sum_{j=1}^{n}x_j^2 = \|A\|_\infty\|A\|_1\|x\|_2^2$$

Thus,

$$\|A\|_2^2 = \sup_{\|x\|_2\neq 0}\frac{\|Ax\|_2^2}{\|x\|_2^2} \leq \sup_{\|x\|_2\neq 0}\frac{\|A\|_1\|A\|_\infty\|x\|_2^2}{\|x\|_2^2} = \|A\|_1\|A\|_\infty$$

i.e.

$$\|A\|_2 \leq \sqrt{\|A\|_1\|A\|_\infty}$$

Problem 5. (LU Decomposition)

1)

Solution.

Whether the A has an LU decomposition relies on the fact that the pivots (diagonal elements of U) are all non-zero.

Based on the $A^{(0)}$, we take

$$\tau^{(1)} = \begin{bmatrix} 1 \\ 2 \\ -6 \\ 4 \end{bmatrix}$$

Since $A^{(1)} = M_1 A^{(0)}$ and $M_1 = 1 - \tau^{(1)} e_k^T$,we can get that

$$A^{(1)} = \begin{bmatrix} 1 & 3 & 1 & -2 \\ 0 & -4 & -2 & 9 \\ 0 & 21 & 10 & -4 \\ 0 & -10 & -5 & 15 \end{bmatrix}$$

Similarly, we can take

$$\tau^{(2)} = \begin{bmatrix} 0 \\ 1 \\ -21/4 \\ 5/2 \end{bmatrix}$$

$$A^{(2)} = \begin{bmatrix} 1 & 3 & 1 & -2 \\ 0 & -4 & -2 & 9 \\ 0 & 0 & -\frac{1}{2} & \frac{173}{4} \\ 0 & 0 & 0 & -\frac{15}{2} \end{bmatrix}$$

Since the entry under the pivot of the third column $A^{(2)}$ is zero, we can take that $A^{(3)} = A^{(2)}$ and $\tau^{(3)} = e_3$

Finally, we can get

$$L = M_1^{-1} M_2^{-1} M_3^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ -6 & -21/4 & 1 & 0 \\ 4 & 5/2 & 0 & 1 \end{bmatrix}$$

$$U = A^{(3)} = \begin{bmatrix} 1 & 3 & 1 & -2 \\ 0 & -4 & -2 & 9 \\ 0 & 0 & -1/2 & 173/4 \\ 0 & 0 & 0 & -15/2 \end{bmatrix}$$

2)

Solution.

We solve $LUx = b$ by solving two triangular systems:

- **step 1.** Solve $Ly = b$ for y (forward substitution).
- **step 2.** Solve $Ux = y$ for x (backward substitution).

Step 1: Solve $Ly = b$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ -6 & -21/4 & 1 & 0 \\ 4 & 5/2 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 2 \\ 5 \end{bmatrix}$$

- $y_1 = 2$

- $2y_1 + y_2 = 0 \implies y_2 = -4$
- $-6y_1 - \frac{21}{4}y_2 + y_3 = 2 \implies y_3 = -7$
- $4y_1 + \frac{5}{2}y_2 + y_4 = 5 \implies y_4 = 7$

Thus,

$$y = \begin{bmatrix} 2 \\ -4 \\ -7 \\ 7 \end{bmatrix}$$

.

Step 2: Solve $Ux = y$

$$\begin{bmatrix} 1 & 3 & 1 & -2 \\ 0 & -4 & -2 & 9 \\ 0 & 0 & -1/2 & 173/4 \\ 0 & 0 & 0 & -15/2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \\ -7 \\ 7 \end{bmatrix}$$

- $-\frac{15}{2}x_4 = 7 \implies x_4 = -\frac{14}{15}$
- $-\frac{1}{2}x_3 + \frac{173}{4}x_4 = -7 \implies x_3 = -\frac{1001}{15}$
- $-4x_2 - 2x_3 + 9x_4 = -4 \implies x_2 = \frac{484}{15}$
- $x_1 + 3x_2 + x_3 - 2x_4 = 2 \implies x_1 = -\frac{449}{15}$

Above all, we can get that

$$x = \begin{bmatrix} -449/15 \\ 484/15 \\ -1001/15 \\ -14/15 \end{bmatrix} \approx \begin{bmatrix} -29.9 \\ 32.3 \\ -66.7 \\ -0.9 \end{bmatrix}$$

3)

Solution.

We perform Gaussian change to swapping rows and ensure the largest absolute value is the pivot.

Step 1. Col 1: $[1, 2, -6, 4]^T$. Max absolute value is $|-6|$ in R_3. Swap $R_1 \leftrightarrow R_3$.

$$P_1 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad A^{(0)'} = P_1 A = \begin{bmatrix} -6 & 3 & 4 & 8 \\ 2 & 2 & 0 & 5 \\ 1 & 3 & 1 & -2 \\ 4 & 2 & -1 & 7 \end{bmatrix}$$

Based on the $A^{(0)'}$, we take

$$\tau^{(1)} = \begin{bmatrix} 1 \\ -1/3 \\ -1/6 \\ -2/3 \end{bmatrix}$$

Since $A^{(1)} = M_1 A^{(0)'}$ and $M_1 = 1 - \tau^{(1)} e_1^T$,we can get that

$$A^{(1)} = \begin{bmatrix} -6 & 3 & 4 & 8 \\ 0 & 3 & 4/3 & 23/3 \\ 0 & 7/2 & 5/3 & -2/3 \\ 0 & 4 & 5/3 & 37/3 \end{bmatrix}$$

Step 2. Similarly, Col 2 (sub-diagonal): $[3, 7/2, 4]^T$. Max absolute value is $|4|$ in R_4. Swap $R_2 \leftrightarrow R_4$.

$$P_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad A^{(1)'} = P_2 A^{(1)} = \begin{bmatrix} -6 & 3 & 4 & 8 \\ 0 & 4 & 5/3 & 37/3 \\ 0 & 7/2 & 5/3 & -2/3 \\ 0 & 3 & 4/3 & 23/3 \end{bmatrix}$$

We must also swap the computed τ

$$\tau^{(1)'} = \begin{bmatrix} 1 \\ -2/3 \\ -1/6 \\ -1/3 \end{bmatrix}$$

Based on the $A^{(1)'}$, we take

$$\tau^{(2)} = \begin{bmatrix} 0 \\ 1 \\ 7/8 \\ 3/4 \end{bmatrix}$$

Since $A^{(2)} = M_2 A^{(1)'}$ and $M_2 = 1 - \tau^{(2)} e_2^T$,we can get that

$$A^{(2)} = \begin{bmatrix} -6 & 3 & 4 & 8 \\ 0 & 4 & 5/3 & 37/3 \\ 0 & 0 & 5/24 & -275/24 \\ 0 & 0 & 1/12 & -19/12 \end{bmatrix}$$

Step 3. Col 3 (sub-diagonal): $[5/24, 1/12]^T$. Max is $5/24$ in R_3. No swap.
i.e. $P_3 = I, \quad A^{(2)'} = A^{(2)}, \quad \tau^{(2)} = \tau^{(2)'}, \quad \tau^{(1)''} = \tau^{(1)'}$

Based on the $A^{(2)'}$, we take

$$\tau^{(3)} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 2/5 \end{bmatrix}$$

Since $A^{(3)} = M_3 A^{(2)'}$ and $M_3 = 1 - \tau^{(3)} e_3^T$,we can get that

$$A^{(3)} = \begin{bmatrix} -6 & 3 & 4 & 8 \\ 0 & 4 & 5/3 & 37/3 \\ 0 & 0 & 5/24 & -275/24 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

Finally,

$$M_1 = 1 - \tau^{(1)''} e_1^T$$
$$M_2 = 1 - \tau^{(2)'} e_2^T$$
$$M_3 = 1 - \tau^{(3)} e_3^T$$

Thus, the final matrices are:

$$P = P_3 P_2 P_1 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$
$$L = M_1^{-1} M_2^{-1} M_3^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2/3 & 1 & 0 & 0 \\ -1/6 & 7/8 & 1 & 0 \\ -1/3 & 3/4 & 2/5 & 1 \end{bmatrix}$$
$$U = A^{(3)} = \begin{bmatrix} -6 & 3 & 4 & 8 \\ 0 & 4 & 5/3 & 37/3 \\ 0 & 0 & 5/24 & -275/24 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

4)

Solution.

From Problem 1: $A = L_1 U_1$

$$L_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ -6 & -21/4 & 1 & 0 \\ 4 & 5/2 & 0 & 1 \end{bmatrix} \quad \text{and} \quad U_1 = \begin{bmatrix} 1 & 3 & 1 & -2 \\ 0 & -4 & -2 & 9 \\ 0 & 0 & -1/2 & 173/4 \\ 0 & 0 & 0 & -15/2 \end{bmatrix}$$

We set $L = L_1$. We then factor U_1 into DM^T, where D is the diagonal of U_1 and M^T is a unit upper triangular matrix.

$$D = \text{Diag}(U_1) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -4 & 0 & 0 \\ 0 & 0 & -1/2 & 0 \\ 0 & 0 & 0 & -15/2 \end{bmatrix}$$

$$M^T = D^{-1}U_1 = \begin{bmatrix} 1/1 & 0 & 0 & 0 \\ 0 & 1/(-4) & 0 & 0 \\ 0 & 0 & 1/(-1/2) & 0 \\ 0 & 0 & 0 & 1/(-15/2) \end{bmatrix} \begin{bmatrix} 1 & 3 & 1 & -2 \\ 0 & -4 & -2 & 9 \\ 0 & 0 & -1/2 & 173/4 \\ 0 & 0 & 0 & -15/2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 3 & 1 & -2 \\ 0 & 1 & 1/2 & -9/4 \\ 0 & 0 & 1 & -173/2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

This M^T is unit upper triangular, which means its transpose, M, is unit lower triangular, as required by the problem definition.

The final decomposition is:

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ -6 & -21/4 & 1 & 0 \\ 4 & 5/2 & 0 & 1 \end{bmatrix}$$

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -4 & 0 & 0 \\ 0 & 0 & -1/2 & 0 \\ 0 & 0 & 0 & -15/2 \end{bmatrix}$$

$$M^T = \begin{bmatrix} 1 & 3 & 1 & -2 \\ 0 & 1 & 1/2 & -9/4 \\ 0 & 0 & 1 & -173/2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Problem 6. (Cholesky Decomposition)

1)

Solution.

Define.

$$L = \begin{pmatrix} l_{11} & 0 & 0 & 0 \\ l_{21} & l_{22} & 0 & 0 \\ l_{31} & l_{32} & l_{33} & 0 \\ l_{41} & l_{42} & l_{43} & l_{44} \end{pmatrix}$$

We can find the elements of L by equating the elements of A with the corresponding elements of LL^T column by column.

The general formulas are:

$$l_{jj} = \sqrt{a_{jj} - \sum_{k=1}^{j-1} l_{jk}^2}$$

$$l_{ij} = \frac{1}{l_{jj}} \left(a_{ij} - \sum_{k=1}^{j-1} l_{ik} l_{jk} \right) \quad \text{for } i > j$$

Column 1 $(j = 1)$

- $l_{11} = \sqrt{a_{11}} = 2$
- $l_{21} = \frac{a_{21}}{l_{11}} = 1$
- $l_{31} = \frac{a_{31}}{l_{11}} = 1$
- $l_{41} = \frac{a_{41}}{l_{11}} = 2$

Column 2 $(j = 2)$

- $l_{22} = \sqrt{a_{22} - l_{21}^2} = 2$
- $l_{32} = \frac{a_{32} - l_{31} l_{21}}{l_{22}} = 0$
- $l_{42} = \frac{a_{42} - l_{41} l_{21}}{l_{22}} = 0$

Column 3 $(j = 3)$

- $l_{33} = \sqrt{a_{33} - (l_{31}^2 + l_{32}^2)} = 2$
- $l_{43} = \frac{a_{43} - (l_{41} l_{31} + l_{42} l_{32})}{l_{33}} = 1$

Column 4 $(j = 4)$

- $l_{44} = \sqrt{a_{44} - (l_{41}^2 + l_{42}^2 + l_{43}^2)} = 1$

Final Matrix L Combining these results, the lower triangular matrix L is:

$$L = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 0 & 2 & 0 \\ 2 & 0 & 1 & 1 \end{pmatrix}$$

2)

Proof.
We decompose k in two cases:

- **Case 1.** k = 1.
- **Case 2.** k > 1.

Case 1. k=1:

It is obviously that $a_{11} = g_{11}^2, \quad \Delta_1 = a_{11}$.
Thus
$$g_{11}^2 = a_{11} = \frac{a_{11}}{1} = \frac{\Delta_1}{\Delta_0}$$

Case 1. k>1:
Since A is symmetric, A_k is symmetric. We partition it as:

$$A_k = \begin{bmatrix} A_{k-1} & \mathbf{v} \\ \mathbf{v}^T & a_{kk} \end{bmatrix}$$

where A_{k-1} is the $(k-1) \times (k-1)$ leading principal submatrix, $\mathbf{v}$ is a $(k-1) \times 1$ column vector, and a_{kk} is a scalar.

Since G is lower triangular, its leading principal submatrix G_k is also lower triangular. We partition it in a corresponding block form:

$$G_k = \begin{bmatrix} G_{k-1} & \mathbf{0} \\ \mathbf{w}^T & g_{kk} \end{bmatrix}$$

where G_{k-1} is the $(k-1) \times (k-1)$ lower triangular Cholesky factor of A_{k-1}, $\mathbf{0}$ is a $(k-1) \times 1$ zero vector, $\mathbf{w}$ is a $(k-1) \times 1$ column vector, and g_{kk} is a scalar.

The Cholesky decomposition for A_k is $A_k = G_k G_k^T$ (the property of Cholesky decomposition). We compute the right-hand side using our partitions:

$$\begin{aligned} G_k G_k^T &= \begin{bmatrix} G_{k-1} & \mathbf{0} \\ \mathbf{w}^T & g_{kk} \end{bmatrix} \begin{bmatrix} G_{k-1}^T & \mathbf{w} \\ \mathbf{0}^T & g_{kk} \end{bmatrix} \\ &= \begin{bmatrix} G_{k-1} G_{k-1}^T & G_{k-1}\mathbf{w} \\ (G_{k-1}\mathbf{w})^T & \mathbf{w}^T\mathbf{w} + g_{kk}^2 \end{bmatrix} U \end{aligned}$$

Use the property of determinants.

$$\Delta_k = \det(A_k) = \det(G_k G_k^T) = \det(G_k)\det(G_k^T) = (\det(G_k))^2$$

Since G_k is a lower triangular matrix, its determinant is the product of its diagonal elements:

$$\det(G_k) = (g_{11}g_{22}\cdots g_{k-1,k-1}) \cdot g_{kk} = (\det(G_{k-1})) \cdot g_{kk}$$

Then,
$$\Delta_k = ((\det(G_{k-1})) \cdot g_{kk})^2 = (\det(G_{k-1}))^2 \cdot g_{kk}^2$$

Similarly,

$$\Delta_{k-1} = \det(A_{k-1}) = \det(G_{k-1}G_{k-1}^T) = (\det(G_{k-1}))^2$$

We now have a system of two equations:

$$\Delta_k = (\det(G_{k-1}))^2 \cdot g_{kk}^2$$
$$\Delta_{k-1} = (\det(G_{k-1}))^2$$

It is obviously that

$$\Delta_k = \Delta_{k-1} \cdot g_{kk}^2$$

i.e.

$$g_{kk}^2 = \frac{\Delta_k}{\Delta_{k-1}}$$

Above all, we can get that for any $k \in \{1, 2, \ldots, n\}$, $g_{kk}^2 = \frac{\Delta_k}{\Delta_{k-1}}$.