Problem 1 (Question 1.2)

Suppose $A = ab^T$ for some vector $a, b \in \mathbb{R}^n$. Note that for every $v \in \mathbb{R}^n$, we have

$$Av = ab^T v = \langle b, v \rangle a$$

This implies that the column space of A is spanned by a. In other words,

$$\text{rank}(A) = \dim \text{Col}(A) = 1$$

Suppose, on the other hand, $\text{rank}(A) = 1$. Let $\text{Col}(A)$ be spanned by some vector $a \in \mathbb{R}^n$, then

$$Av = \lambda a$$

For all $v \in \mathbb{R}^n$. Let $e_1, e_2, ...e_n$ be a basis of $\mathbb{R}^n$, then under this basis, A has a matrix representation of

$$A = \begin{bmatrix} | & | & & | \\ \lambda_1 a & \lambda_2 a & \dots & \lambda_n a \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | \\ a \\ | \end{bmatrix} \begin{bmatrix} \lambda_1 & \lambda_2 & \dots & \lambda_n \end{bmatrix} = ab^T$$

This concludes that A has rank 1 iff $A = ab^T$ for some a, b

Problem 2 (Question 1.4)

Let A be a strictly upper triangular matrix. Let $a_{i,j}^{(k)}$ denote that i, j entry of A^k. We claim that

$$a_{i,j}^{(k)} = 0 \quad \text{when } i - j \leq k - 1$$

We proceed by induction.

Base case: The case for $n = 1$ is true by the definition of A.
Inductive hypothesis: Suppose the statement holds for all $n \leq k$.
Inductive step: For $n = k + 1$, note that

$$a_{i,j}^{(k+1)} = \sum_{m=1}^{n} a_{i,m}^{(k)} a_{m,j}^{(1)}$$

By inductive hypothesis, we have

$$a_{i,m}^{(k)} = 0 \quad \text{when } i - m \leq k - 1$$

$$a_{m,j}^{(1)} = 0 \quad \text{when } m - j \leq 0$$

In other words

$$a_{i,m}^{(k)} a_{m,j}^{(1)} = 0 \quad \text{if } m \geq i - k + 1 \text{ or } m \leq j$$

This holds for all m for $i - j \leq k$. This is because if $m > j$, then

$$m > j \implies m > i - k \implies m \geq i - k + 1$$

This means that, for all m, either one of the conditions $m \geq i - k + 1$ or $m \leq j$ must be true. Hence

$$a_{i,j}^{(k+1)} = 0 \quad \text{when } i - j \leq k$$

Thus proving the claim.

With this result, taking $k = n$, we see that

$$a_{i,j}^{(n)} = 0 \quad \text{when } i - j \leq n - 1$$

But this is true for all i, j. This implies that $a_{i,j}^{(n)} = 0$ for all i, j, or

$$A^n = 0$$

Problem 3 (Question 1.10)

Note that

$$\mathrm{fl}\left(\sum_{i=1}^{d} x_k y_k\right) = x_1 y_1 (1+\delta_1) \prod_{i=1}^{d-1} (1+\hat{\delta}_i) + \sum_{k=2}^{d} x_k y_k (1+\delta_k) \prod_{i=k-1}^{d-1} (1+\hat{\delta}_i)$$

Since

$$1 - d\epsilon \leq (1+\delta_1) \prod_{i=1}^{d-1} (1+\hat{\delta}_i) \leq 1 + d\epsilon$$

$$1 - d\epsilon \leq 1 - (d-k+2)\epsilon \leq (1+\delta_k) \prod_{i=k-1}^{d-1} (1+\hat{\delta}_i) \leq 1 + (d-k+2)\epsilon \leq 1 + d\epsilon$$

Therefore,

$$(1-d\epsilon) \sum_{i=1}^{d} x_i y_i \leq \mathrm{fl}\left(\sum_{i=1}^{d} x_k y_k\right) \leq (1+d\epsilon) \sum_{i=1}^{d} x_i y_i$$

Hence,

$$\mathrm{fl}\left(\sum_{i=1}^{d} x_k y_k\right) = \sum_{i=1}^{d} (1+\tilde{\delta}_i) x_i y_i$$

With $|\tilde{\delta}_i| \leq d\epsilon$

Now let

$$c_{i,j} = \sum_{k=1}^{n} a_{i,k} b_{k,j}$$

Then by what we've just shown

$$\begin{aligned} |\mathrm{fl}(c_{i,j}) - c_{i,j}| &= \left| \mathrm{fl}\left(\sum_{k=1}^{n} a_{i,k} b_{k,j}\right) - \sum_{k=1}^{n} a_{i,k} b_{k,j} \right| \\ &= \left| \sum_{k=1}^{n} \tilde{\delta}_k a_{i,k} b_{k,j} \right| \\ &\leq n\epsilon \left| \sum_{k=1}^{n} a_{i,k} b_{k,j} \right| \\ &\leq n\epsilon \sum_{k=1}^{n} |a_{i,k}||b_{k,j}| \\ &= n\epsilon (|A||B|)_{i,j} \end{aligned}$$

Hence,

$$|\mathrm{fl}(A \cdot B) - A \cdot B| \leq n\epsilon |A||B|$$

Problem 4 (Question 1.11)

Let $l_{i,j}$ be the i, j entry of matrix L, then the solution, by forward substitution, is given by

$$x_k = \frac{b_k - \sum_{i=1}^{k-1} l_{i,k} x_k}{l_{k,k}}$$

Let $x_k^* = \text{fl}(x_k)$, then

$$x_k^* = \frac{b_k - \text{fl}(\sum_{i=1}^{k-1} l_{i,k} x_i^*)}{l_{k,k}}(1+\delta_0)(1+\delta_1)$$

From the previous problem, we have

$$x_k^* = \frac{b_k - \sum_{i=1}^{k-1} l_{i,k} x_i^*(1+\delta_{i,k})}{l_{k,k}}(1+\delta_0)(1+\delta_1)$$

Where $|\delta_{i,k}| \leq (i-1)\epsilon$. Rewriting, we have

$$\frac{1}{(1+\delta_0)(1+\delta_1)} l_{k,k} x_k^* + \sum_{i=1}^{k-1} l_{i,k} x_i^*(1+\delta_{i,k}) = b_k$$

Since

$$(1+\delta_0)(1+\delta_1) \geq (1-\epsilon)^2 \geq 1 - 2\epsilon \implies \frac{1}{(1+\delta_0)(1+\delta_1)} \leq 1 + 2\epsilon$$

So

$$\frac{1}{(1+\delta_0)(1+\delta_1)} = 1 + \delta_{i,1}$$

Where $|\delta_{i,1}| \leq 2\epsilon$. Hence, we have

$$\sum_{i=1}^{k} l_{i,k} x_i^*(1+\delta_{i,k}) = b_k$$

Letting $[\delta L]_{i,j} = l_{i,j}\delta_{i,j}$, we see that x^* is the solution of

$$(L + \delta L)x^* = b$$

Where $|[\delta L]_{i,j}| = |l_{i,j}\delta_{i,j}| \leq n\epsilon |l_{i,j}|$.

In the case of backward substitution, the solution is given by

$$x_k^* = \frac{c_k - \sum_{i=k+1}^{n} u_{k,i} x_i^*(1+\delta_{k,i})}{u_{k,k}}(1+\delta_0)(1+\delta_1)$$

By the same analysis,

$$\frac{1}{(1+\delta_0)(1+\delta_1)}u_{k,k}x^* + \sum_{i=k+1}^{n} u_{k,i}x_i^*(1+\delta_{k,i}) = c_k$$

Hence, we have

$$\sum_{i=k}^{n} u_{k,i}x_i^*(1+\delta_{k,i}) = c_k$$

Letting $[\delta U]_{i,j} = u_{i,j}\delta_{i,j}$, we see that x^* is the solution of

$$(U + \delta U)x^* = c$$

Where $|[\delta U]_{i,j}| = |u_{i,j}\delta_{i,j}| \leq n\epsilon |u_{i,j}|$