Problem 1

(a): Using the definition of matrix-matrix multiplication, show that

$$\operatorname{tr}(AB) = \operatorname{tr}(BA)$$

for square matrices $A, B \in \mathbb{R}^{n\times n}$.

Proof. Let $C = AB$, using the definition of matrix multiplication, the (i, j)-th entry of C is:

$$C_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj}.$$

The diagonal entries of C are given by C_{ii}, which gives us:

$$C_{ii} = \sum_{k=1}^{n} A_{ik}B_{ki}.$$

Recall that the trace of a square matrix is the sum of its diagonal elements. Taking the trace of C, we sum over all diagonal elements:

$$\operatorname{tr}(C = AB) = \sum_{i=1}^{n} C_{ii} = \sum_{i=1}^{n}\sum_{k=1}^{n} A_{ik}B_{ki}.$$

Now, let $D = BA$ and consider $\operatorname{tr}(D = BA)$. Using the same reasoning, the (i, j)-th entry of D is:

$$D_{ij} = \sum_{k=1}^{n} B_{ik}A_{kj}.$$

Taking the trace of D, we get:

$$\operatorname{tr}(D = BA) = \sum_{i=1}^{n} D_{ii} = \sum_{i=1}^{n}\sum_{k=1}^{n} B_{ik}A_{ki}.$$

Swapping the variables i, k, we get:

$$\operatorname{tr}(BA) = \sum_{k=1}^{n}\sum_{i=1}^{n} B_{ki}A_{ik} = \sum_{i=1}^{n}\sum_{k=1}^{n} A_{ik}B_{ki} = \operatorname{tr}(AB).$$

This proves our statement. □

(b) Explain why this property implies that for any set of square matrices $\{A_i\}_{i=1}^n$, the trace satisfies the cyclic property:

$$\operatorname{tr}(A_{\sigma(1)} \cdots A_{\sigma(n)}) = \operatorname{tr}(A_1 \cdots A_n),$$

where $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ is an order-preserving permutation of the indices.

Proof. Note that since σ is an order-preserving permutation of the indices, any permuation should either be the trivial one, eg. $A_1A_2 \cdots A_n$ or be in the form $A_iA_{i+1} \cdots A_nA_1 \cdots A_{i-1}$ for some $1 \le i \le n - 1$. Note that we can split this product into two parts: $B = A_iA_{i+1} \cdots A_n$ and $C = A_1 \cdots A_{i-1}$. Observe that $CB = A_1 \cdots A_n$ and $BC = A_{\sigma(1)} \cdots A_{\sigma(n)}$. From part 1, we know that both B and C are square matrices as they are the product of square matrices. Subsequently, applying part (a), we get:

$$\operatorname{tr}(BC) = \operatorname{tr}(A_{\sigma(1)} \cdots A_{\sigma(n)}) = \operatorname{tr}(A_1 \cdots A_n) = \operatorname{tr}(CB)$$

This proves our statement. □

Problem 2

Let $A \in \mathbb{R}^{m\times n}$. Use the definition of matrix-matrix multiplication to show that

$$\|A\|_F^2 = \mathrm{tr}(A^TA).$$

Proof. For $C = A^TA$, from matrix-matrix multiplication and recalling $A^T_{ik} = A_{ki}$, we have:

$$C_{ii} = \sum_{k=1}^{m} A^T_{ik}A_{ki} = \sum_{k=1}^{m} A_{ki}A_{ki}.$$

As the trace of A^TA, $\mathrm{tr}(A^TA)$, is the sum of its diagonal entries:

$$\mathrm{tr}(A^TA) = \sum_{i=1}^{n} C_{ii}.$$

Substituting the expression for C_{ii}, we get:

$$\mathrm{tr}(A^TA) = \sum_{i=1}^{n}\sum_{k=1}^{m} A_{ki}^2.$$

Recall that the square of the Frobenius norm $\|A\|_F^2$ is the sum of the squares of all entries of A:

$$\|A\|_F^2 = \sum_{i=1}^{n}\sum_{k=1}^{m} A_{ki}^2.$$

Comparing this with the expression for $\mathrm{tr}(A^TA)$, we see that:

$$\|A\|_F^2 = \sum_{i=1}^{n}\sum_{k=1}^{m} A_{ki}^2 = \mathrm{tr}(A^TA).$$

This proves our statement.

□

Problem 3

Let $A \in \mathbb{R}^{m\times n}$. Show that the Frobenius norm satisfies

$$\|A\|_F = \sqrt{\sum_{i=1}^{r} \sigma_i^2},$$

where $\{\sigma_i\}$ are the singular values of A, and $r = \mathrm{rank}(A)$. Hint: Use the results of Problem 1 and Problem 2.

Proof. Recall that from Problem 2 we can write the the Frobenius norm of A as

$$\|A\|_F^2 = \sqrt{\mathrm{tr}(A^TA)}.$$

Let the singular value decomposition (SVD) of A be represented by

$$A = U\Sigma V^T,$$

where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices (since A is a real values matrix), and $\Sigma \in \mathbb{R}^{m\times n}$ is a diagonal matrix with the singular values $\{\sigma_i\}$ of A on the diagonal. Substituting $A = U\Sigma V^T$ into $A^T A$, we get:

$$A^T A = (U\Sigma V^T)^T (U\Sigma V^T) = V\Sigma^T U^T U\Sigma V^T$$

Recall that $U^T U = I$ due to the orthogonality of U):

$$A^T A = V\Sigma^T U^T U\Sigma V^T = V\Sigma^T \Sigma V^T$$

Now from problem 1, recall that $\mathrm{tr}((V\Sigma^T\Sigma)V^T) = \mathrm{tr}(V^T(V\Sigma^T\Sigma))$. Again, due to the orthogonality of V, this simplifies to $\mathrm{tr}(\Sigma^T\Sigma)$. Noting that $\Sigma^T\Sigma$ is a diagonal matrix with entries σ_i^2 (the squared singular values of A), we get:

$$\mathrm{tr}(A^T A) = \sum_{i=1}^{r} \sigma_i^2,$$

where $r = \mathrm{rank}(A)$ is the number of nonzero singular values of A. Taking the square root to match with our original definition, we get:

$$\|A\|_F = \sqrt{\mathrm{tr}(A^T A)} = \sqrt{\sum_{i=1}^{r} \sigma_i^2}.$$

□

Problem 4

Show that the Frobenius and induced matrix 2-norms satisfy

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{\mathrm{rank}(A)}\|A\|_2$$

for all $A \in \mathbb{R}^{m\times n}$, i.e., these norms are equivalent. Bonus: For what class of matrices does $\|A\|_2 = \|A\|_F$?

Proof. Recall that the induced matrix 2-norm of a matrix A is given by its largest singular value:

$$\|A\|_2 = \sigma_{\max}.$$

Additionally, from Problem 3, the Frobenius norm of A can be expressed as:

$$\|A\|_F = \sqrt{\sum_{i=1}^{r} \sigma_i^2},$$

where $r = \mathrm{rank}(A)$ is the number of nonzero singular values.

To prove the inequality $\|A\|_2 \leq \|A\|_F$, observe that $\|A\|_F$ includes all the singular values of A, while $\|A\|_2$ considers only the largest singular value. Since the sum of the squares of the singular values is at least as large as the square of the largest singular value, it follows that:

$$\|A\|_F = \sqrt{\sum_{i=1}^{r} \sigma_i^2} \geq \sqrt{\sigma_{\max}^2} = \|A\|_2.$$

To prove the inequality $\|A\|_F \leq \sqrt{\text{rank}(A)}\|A\|_2$, note that $\sigma_{\max} \geq \sigma_i$ for all i by definition. This gives us:

$$\sum_{i=1}^{r} \sigma_i^2 \leq \sum_{i=1}^{r} \sigma_{\max}^2 = r\sigma_{\max}^2.$$

Taking the square root of both sides, we get:

$$\|A\|_F = \sqrt{\sum_{i=1}^{r} \sigma_i^2} \leq \sqrt{r\sigma_{\max}^2} = \sqrt{\text{rank}(A)}\|A\|_2.$$

Combining these, we get:

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{\text{rank}(A)}\|A\|_2.$$

Bonus: For $\|A\|_2 = \|A\|_F$, observe that equality occurs when $\sqrt{\sum_{i=1}^{r} \sigma_i^2} = \sqrt{\sigma_{\max}^2}$. This implies that all singular values except $\sigma_{\max}$ are zero – therefore, A must have rank 1.

For matrices of rank 1, we have:

$$\|A\|_F = \sqrt{\sigma_{\max}^2} = \|A\|_2.$$

Thus, $\|A\|_2 = \|A\|_F$ if and only if A is a rank-1 matrix. □

Problem 5

Let $A \in \mathbb{R}^{n\times n}$ be an invertible matrix. Let $A = U\Sigma V^T$ be its SVD with the diagonal entries of $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n)$ given in descending order. The Eckart-Young theorem states that

$$\arg \min_{Z\in\mathbb{R}^{n\times n}:\text{rank}(Z)=k} \|A - Z\|_2 = A_k,$$

where $A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$ is the truncated SVD of rank k.

Part (a): Use this theorem to show that

$$\min_{X\in\mathbb{R}^{n\times n}:X \textbf{ singular}} \|A - X\|_2 = \sigma_n,$$

i.e., the smallest singular value measures the absolute distance from A to the nearest singular matrix.

Proof. Notice that the Eckart-Young theorem gives us the nearest matrix of A to a rank r matrix is given by A_r. Notice that we need the singular matrix to have rank $r < n$, as rank is always an integer, the largest possible value we can have is $n-1$. The Eckart-Young theorem, then, tells us the closest matrix of rank $n-1$ is A_{n-1}. Now consider the difference between these two matrices. Recall that:

$$A = \sum_{i=1}^{n} \sigma_i u_i v_i^T, \qquad A_{n-1} = \sum_{i=1}^{n-1} \sigma_i u_i v_i^T$$

The difference $A - A_{n-1}$ consists only of the last term in the SVD:

$$A - A_{n-1} = \sigma_n u_n v_n^T.$$

where σ_n is the smallest singular value of A, and u_n and v_n are the corresponding singular vectors. Since the 2-norm of a matrix is equal to its largest singular value, we have:

$$\|A - A_{n-1}\|_2 = \|\sigma_n u_n v_n^T\|_2 = \sigma_n.$$

Thus, the minimum distance from A to a singular matrix is achieved by A_{n-1}, and this distance is equal to the smallest singular value σ_n. This proves our statement. □

Part (b): Use this theorem to show that

$$\min_{X \in \mathbb{R}^{n\times n}: X \textbf{ singular}} \frac{\|A - X\|_2}{\|A\|_2} = \frac{1}{\kappa(A)},$$

i.e., the reciprocal of the condition number measures the relative distance from A to the nearest singular matrix.

Proof. Recall from part (a), we have:

$$\min_{X \in \mathbb{R}^{n\times n}: X \text{ singular}} \|A - X\|_2 = \sigma_n,$$

where $\sigma_n = \sigma_{min}$, the smallest singular value of A, as the values are in descending order. Now recall that $\|A\|_2 = \sigma_{max}$. Substituting this to our expression, we get:

$$\frac{\|A - X\|_2}{\|A\|_2} = \frac{\sigma_{min}}{\sigma_{max}}.$$

Now recall that the condition number of A with respect to the 2-norm is given by:

$$\kappa(A) = \frac{\sigma_{max}}{\sigma_{min}}$$

This gives us $\frac{1}{\kappa(A)} = \frac{\sigma_{min}}{\sigma_{max}}$. Substituting it back to our expression:

$$\min_{X \in \mathbb{R}^{n\times n}: X \text{ singular}} \frac{\|A - X\|_2}{\|A\|_2} = \frac{1}{\kappa(A)}.$$

This proves our statement. □