HINTS

PROBLEM 1:
Prove that $\|Q^*A\|^2_F = \langle AA^T, QQ^T \rangle$ and use Von Neumann's trace inequality. What are the sing. values of $QQ^T$? To get $Q$ that attains the max, consider the SVD of $A$.

PROBLEM 2:
1) W.l.o.g. $\|A\|_q = 1$; in this case all singular values $\sigma_i \leq 1$ so $\sigma_i \leq \sigma_i^q$ ...
2) To prove $\langle A, C \rangle \leq \|A\|_p$ use Von Neumann's trace inequality + Hölder's inequality. To get $C$ that attains equality consider the SVD $A = U \Sigma V^T$ and search for $C$ of the form $U \Sigma C V^T$ for a suitable $\Sigma$.
3) For the triangular inequality use point (2).

PROBLEM 3:
1) Note that if $x \in \mathbb{R}^n$ and $P$ is an orthogonal projection onto $S$ then $Px \in S$ and $(I - P)x \in S^\perp$
2) Show that $X_0 = X(X^*X)^{-1}X^T$ satisfies the properties in the definition.
3) For $\| \cdot \|_F$: recall that the best rank-1 approx is unique.
   For $\| \cdot \|_2$: $2 \times 2$ matrices are too small to find.
   Counterexamples, but $3 \times 3$ matrices will do the job.

PROBLEM 4:
Construct the matrices $A_k$ by modifying the $\Sigma$ in the SVD of $A$.

PROBLEM 5:
$U$ and $V$ will be permutation matrices.
To get best rank-1 approx, use the SVD (truncated).

PROBLEM 6:
Let $A = \begin{bmatrix} B \\ c \end{bmatrix}$ and consider the characterization of the largest sing. val. as $\sigma_1 = \max_{x \neq 0} \frac{x^T A^T A x}{x^T x}$ (prove that $\sigma_1(A) \geq \sigma_1(B)$).
**SOLUTIONS**

**PROBLEM 1**

For \( Q \in \mathbb{R}^{m \times k} \) such that \( Q^T Q = I_k \), we have that

\[
\| Q F \|_F^2 = < Q A, Q^T A > = \text{trace} \left( A^T Q Q^T A\right) = \text{trace} \left( Q Q^T A A^T\right) = < Q Q^T, A A^T > \leq \sigma_1(A A^T) \sigma_1(Q Q^T) + \ldots + \sigma_m(A A^T) \sigma_m(Q Q^T)
\]

\[\downarrow\]

**Von Neumann’s trace inequality**

Now we have that \( \sigma_i(A A^T) = \sigma_i(A)^2 \) and as \( Q Q^T = Q \cdot I_k \cdot Q^T \) is an SVD of \( Q Q^T \) the singular values of \( Q Q^T \) are \( \sigma_i(Q Q^T) = \left\{ \begin{array}{ll} 1 & \text{for } i \leq k \\ 0 & \text{for } i > k \end{array} \right. \)

Therefore \( \| Q F \|_F^2 \leq \| A \|_F^2 = \sigma_1(A)^2 + \ldots + \sigma_k(A)^2 \).

The equality is attained for \( Q = U_k \) (first \( k \) left singular vectors of \( A \)) indeed,

\[
\| Q F \|_F^2 = \| U_k \Sigma U \Sigma U \|_F^2 = \| \left[ I_k \begin{array}{c} 0 \end{array} \right] \left[ \begin{array}{c} Z_k \\
\Sigma_k \
\end{array} \right] \|_F^2 = \| \left[ \begin{array}{c} \Sigma_k \end{array} \right] \|_F^2 = \sigma_1^2 + \ldots + \sigma_k^2.
\]

**PROBLEM 2**

1. \( \| A \|_{(\infty)} = \sigma_1 \leq (\sigma_1^p + \text{positive terms})^{1/p} = (\| A \|_p)^{1/p} \) for any \( 1 \leq p < \infty \), where \( \sigma_1 = \| A \|_{\infty} \).

   Now we prove that for any \( 1 \leq p \leq q \), we have \( \| A \|_{\infty} \leq \| A \|_p \leq \| A \|_q \).

   As \( \| A \|_p = \left( \frac{1}{n} \sum_{i=1}^{n} |a_i|^p \right)^{1/p} \) and \( \| A \|_q = \left( \frac{1}{n} \sum_{i=1}^{n} |a_i|^q \right)^{1/q} \), we have

   \[\| A \|_p \leq \| A \|_q \] for any \( 1 \leq p \leq q \).

   Therefore, \( \| A \|_p \leq \| A \|_q \).

2. \( \| A \|_p \) is a Hölder inequality.

   This proves that \( \max \langle A, C \rangle \leq \| A \|_p \cdot \| C \|_q \).

   To prove equality, consider the SVD \( A = U \Sigma V^T \), \( \Sigma = \Sigma / \| \Sigma \|_q \).

   Then we have

   \[
   \langle A, C \rangle = \text{trace} \left( (C^T A) \right) = \text{trace} \left( V \Sigma U \Sigma V^T \right) = \text{trace} \left( \Sigma \Sigma V^T \right) = \left( \sum_{i=1}^{n} \sigma_i \right)^{1-\frac{1}{q}} \leq \left( \sum_{i=1}^{n} \sigma_i \right)^{1-\frac{1}{p}} = \| A \|_p.
   \]

3. \( \| A \|_p = 0 \) trivial; \( \| A \|_p = 1 \) for \( A \) such that \( \| A \|_p = 1 \) in point 1.

   The triangle inequality: \( \| A + B \|_p \leq \| A \|_p + \| B \|_p \).

   \[\begin{align*}
   \| A + B \|_p & \leq \max \langle A, C \rangle + \max \langle B, C \rangle = \| A \|_p + \| B \|_p \\leq 2 \| C \|_p = 2.
   \end{align*}\]

   with \( p = q = 1 \).
Let \( z \in \mathbb{R}^n \), then
\[
\| (P_1 - P_2) z \|_2^2 = (P_1 - P_2)^T (P_1 - P_2) z =
\]
\[
= z^T P_1^T P_1 z - z^T P_1^T P_2 z - z^T P_2^T P_1 z + z^T P_2^T P_2 z
= z^T P_1^T z - z^T P_1^T P_2 z - z^T P_2^T P_1 z + z^T P_2^T P_2 z \ldots
\]

\[
L = P_1, \quad P_1 = P_1 - P_1^T P_1
\]

\[
+ z^T P_2^T z = (P_1 z)^T (I - P_2) z + (P_2 z)^T (I - P_1) z.
\]

Now let \( P_1 \) and \( P_2 \) be orthogonal projections onto \( S \). To prove that \( P_1 = P_2 \), it is sufficient to prove that the RHS of (1) is 0 for all \( z \in \mathbb{R}^n \). Note that \( P_2 z \in S \) so there exists \( u \in \mathbb{R}^n \) such that
\[
P_2 z = P_1 z \quad \text{(because range}(P_2) = S) ; \quad \text{then} \quad (P_1 z)^T (I - P_2) z = w^T P_2^T (I - P_2) z =
\]
\[
w^T (P_2^T - P_1^T) z = w^T (P_2 - P_1) z = w^T (P_2 - P_2) z = 0.
\]

Analogously,
\[
(P_2 z)^T (I - P_1) z = 0 \quad \forall z \in \mathbb{R}^n.
\]

Then we conclude that \( P_1 = P_2 \).

If \( X \) is a basis of \( S \), then \( X^T X \) is invertible so \( X (X^T)^{-1} X^T \) exists, as the matrix \( (X^T X)^{-1} X^T \) has full row rank, we have that
\[
\text{range} \left( X \cdot (X^T X)^{-1} X^T \right) = \text{range} (X) = S.
\]

Moreover, \( X (X^T X)^{-1} X^T = X \cdot (X^T)^{-1} X^T \) and
\[
\left[ X \cdot (X^T)^{-1} X^T \right]^2 = X \cdot (X^T)^{-1} X^T \cdot (X^T)^{-1} X^T
\]

so all the properties of the definition are satisfied.

If \( P \) is an orthogonal projection onto a \( k \)-dim. subspace, \( P A \) has rank at most \( k \). As the minimizer of \( \| A - B \|_F \) for \( B \in \{ \text{matrices of rank} \leq k \} \) is unique when all the singular values of \( A \) are different, and the minimizer corresponds to \( P = U_k U_k^T \) (projection onto the span of the first \( k \) left singular vectors of \( A \)), we conclude that only \( \min \| (I - P) A \|_F \) is unique.

In general the solution is not unique for the spectral norm; consider for example \( A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \) and \( P = q q^T \) with \( q = \begin{bmatrix} 1 \\ -\varepsilon \\ \varepsilon \end{bmatrix} \) for \( \varepsilon > 0 \) small enough.

\[
\| A - q q^T A \|_2 = \| \begin{bmatrix} 3 \varepsilon^2 & 0 & -3\sqrt{1 - \varepsilon^2} \\ 0 & 2 & 0 \\ -3\sqrt{1 - \varepsilon^2} & 0 & 1 - \varepsilon^2 \end{bmatrix} \|_2
\]

\[
\leq \max \left\{ \| \begin{bmatrix} 3 \varepsilon^2 & -3\sqrt{1 - \varepsilon^2} \\ -3\sqrt{1 - \varepsilon^2} & 1 - \varepsilon^2 \end{bmatrix} \|_F, 2 \right\}
\]

use that \( \| \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \|_2 \leq \max \| A_1 \|_2, \| A_2 \|_2 \) [prove this] and \( \| A_1 \|_2 \leq \| A \|_F \)

\[\Box\]
PROBLEM 4

Consider an SVD of $A = U \Sigma V^T$ with $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n)$ and define $A_k := U \cdot \text{diag}(\sigma_1 + \frac{1}{k}, \ldots, \sigma_n + \frac{1}{k}) \cdot V^T$.

All matrices $A_k$ have full rank because the min. singular value is bounded by $\frac{1}{k}$ from below. Moreover,

$$||A - A_k||_2 = ||U \cdot \left( \text{diag}(\sigma_1, \ldots, \sigma_n) - \text{diag}(\sigma_1 + \frac{1}{k}, \ldots, \sigma_n + \frac{1}{k}) \right) \cdot V^T||_2 = ||\text{diag}(\frac{1}{k}, \ldots, \frac{1}{k})||_2 = \frac{1}{k} \rightarrow 0 \quad \text{for} \quad k \rightarrow +\infty.$$ □

**Spectral norm is unitarily invariant.**

---

PROBLEM 5

$$A = U \Sigma V^T \quad U = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \quad V = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$$

Select first column of $U$, a $1 \times 1$ leading submatrix of $\Sigma$, first column of $V$ to get best rank-1 approximation in the spectral & Frobenius norm: $\begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$ □

---

**PROBLEM 6**

Let $A = \begin{bmatrix} B \\ c^T \end{bmatrix}$ for a vector $c \in \mathbb{R}^m$. We prove that $\sigma_1(A) > \sigma_1(B)$.

We have that $\sigma_1(A)^2 = \lambda_{\max}(AA^T) = \lambda_{\max}\begin{bmatrix} BB^T & Bc \\ c^T B & c^T c \end{bmatrix} = \max_{||x||_2 = 1} x^T \begin{bmatrix} BB^T & Bc \\ c^T B & c^T c \end{bmatrix} x = \max_{||x||_2 = 1} x^T \begin{bmatrix} BB^T & Bc \\ c^T B & c^T c \end{bmatrix} x$.

Because we take the max of the same quantity on a smaller set, we have $\max_{||x||_2 = 1} x^T \begin{bmatrix} BB^T & Bc \\ c^T B & c^T c \end{bmatrix} x = \max_{||y||_2 = 1} y^T BB^T y = \lambda_{\max}(BB^T) = \sigma_1(B)^2$. □

**Bonus question:** What happens to the smallest singular value?

(solution in the following page)
Let $B \in \mathbb{R}^{m \times n}$ with $m \geq n$ and consider $A = \begin{bmatrix} B & c^T \end{bmatrix}^{m+1}$. Then $\sigma_{\min}(A) \geq \sigma_{\min}(B)$.

Proof:
$$\sigma_{\min}^2(A) = \lambda_{\min}(A^TA) = \lambda_{\min}(B^TB + cc^T) = \min_{\|x\|_2^2 = 1} x^T(B^TB + cc^T)x = \min_{\|x\|_2^2 = 1} (x^T B^TBx + x^Tcc^Tx) \geq 0$$
$$= \min_{\|x\|_2^2 = 1} x^T B^TBx = \lambda_{\min}(B^TB) = \sigma_{\min}^2(B).$$

Let $B \in \mathbb{R}^{m \times n}$ with $m \leq n-1$ and consider $A = \begin{bmatrix} B & c^T \end{bmatrix}^{m+1}$. Then $\sigma_{\min}(A) \leq \sigma_{\min}(B)$.

Proof:
$$\sigma_{\min}^2(A) = \lambda_{\min}(A^TA) = \lambda_{\min}(\begin{bmatrix} BB^T & Bc \\ c^TB & cc^T \end{bmatrix})$$
$$= \min_{\|x\|_2^2 = 1} x^T \begin{bmatrix} BB^T & Bc \\ c^TB & cc^T \end{bmatrix}x \leq \min_{\|y\|_2^2 = 1} [y^T, 0] \begin{bmatrix} BB^T & Bc \\ c^TB & cc^T \end{bmatrix} [y^T, 0]^T$$
We are taking the min. of the same quantity on a smaller set.
$$= \min_{\|y\|_2^2 = 1} y^TBB^Ty = \lambda_{\min}(BB^T) = \sigma_{\min}^2(B).$$