HINTS

PROBLEM 1: Prove that $\|Q^T A\|_F^2 = \langle AA^T, QQ^T \rangle$ 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 $\sigma_i^p \leq \sigma_i^q$ ...
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 $U\tilde{\Sigma}V^T$ for a suitable $\tilde{\Sigma}$
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 $X(X^T X)^{-1} X^T$ satisfies the properties in the definition
3) For $\|\cdot\|_F$: recall that the best rank-k approx is unique
   For $\|\cdot\|_2$: 2x2 matrices are too small to find counterexamples, but 3x3 matrices will do the job ;-)

PROBLEM 4: Construct the matrices A_k by modifying the Σ 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 $\sigma_1^2 = \max_{x \neq 0} \frac{x^T A^T A x}{x^T x}$
(prove that $\sigma_1(A) \geq \sigma_1(B)$).

SOLUTIONS

PROBLEM 1

For $Q \in \mathbb{R}^{m \times k}$ such that $Q^TQ = I_k$ we have that

$\|Q^TA\|_F^2 = \langle Q^TA, Q^TA\rangle = \operatorname{trace}(A^TQQ^TA) = \operatorname{trace}(QQ^TAA^T) =$

$= \langle QQ^T, AA^T\rangle \underset{\substack{\uparrow \\ \text{Von Neumann's trace inequality}}}{\leq} \sigma_1(AA^T)\sigma_1(QQ^T) + \dots + \sigma_m(AA^T)\sigma_m(QQ^T)$

Now we have that $\sigma_i(AA^T) = \sigma_i(A)^2$ and as $QQ^T = Q \cdot I_k \cdot Q^T$ is an SVD of QQ^T the sing. values of QQ^T are $\sigma_i(QQ^T) = \begin{cases} 1 & \text{for } i \leq k \\ 0 & \text{for } i > k \end{cases}$

Therefore $\|Q^TA\|_F^2 \leq \sigma_1(A)^2 + \dots + \sigma_k(A)^2$.

The equality is attained for $Q = U_k$ (first k left sing. vectors of A): indeed,

$\|Q^TA\|_F^2 = \|U_k^TU\Sigma V\|_F^2 = \|U_k^TU\Sigma\|_F^2 = \left\|\left[I_k \ \underbrace{0}_{m-k}\right]\cdot\begin{bmatrix}\Sigma_k & \\ & \tilde{\Sigma}_k\end{bmatrix}\right\|_F^2 =$

$= \left\|\left[\Sigma_k \ 0\right]\right\|_F^2 = \sigma_1^2 + \dots + \sigma_k^2.$ $\square$

PROBLEM 2

① $\|A\|_{(\infty)} = \sigma_1 \leq (\sigma_1^p + \text{positive terms})^{1/p} = (\sigma_1^p + \dots + \sigma_n^p)^{1/p} = \|A\|_{(p)}$

Now we prove that for $1 \leq p < q$ we have $\|A\|_{(p)} \geq \|A\|_{(q)}$.

As $\|\alpha A\|_{(q)} = (|\alpha\sigma_1|^q + \dots + |\alpha\sigma_n|^q)^{1/q} = |\alpha| \cdot \|A\|_{(q)}$, wlog we can assume that $\|A\|_{(q)} = 1$. Then $\sigma_i \leq 1$ for $i = 1, \dots, n$, so $\sigma_i^p \geq \sigma_i^q$ $\forall i = 1, \dots, n$. Therefore,

$\|A\|_{(p)} = (\sigma_1^p + \dots + \sigma_n^p)^{1/p} \geq (\sigma_1^q + \dots + \sigma_n^q)^{1/p} = (\|A\|_{(q)})^{q/p} = 1$ $\square$

② $\langle A, C\rangle \underset{\substack{\uparrow \\ \text{Von Neumann's trace inequality}}}{\leq} \sigma_1(A)\sigma_1(C) + \dots + \sigma_n(A)\sigma_n(C) \underset{\substack{\uparrow \\ \text{Hölder inequality}}}{\leq} \underbrace{\left(\sum_1^n \sigma_i(A)^p\right)^{1/p} \cdot \left(\sum_1^n \sigma_i(C)^q\right)^{1/q}}_{= \|A\|_{(p)} \cdot \underbrace{\|C\|_{(q)}}_{=1}}$

This proves that $\max_{\|C\|_{(q)}=1} \langle A, C\rangle \leq \|A\|_{(p)}$. To prove equality, consider the SVD $A = U\Sigma V^T$, $\tilde{C} := U \cdot \underbrace{\begin{bmatrix}\sigma_1^{p/q} & & \\ & \ddots & \\ & & \sigma_n^{p/q}\end{bmatrix}}_{=:\tilde{\Sigma}} V^T$,

$C := \tilde{C} / \|\tilde{C}\|_{(q)}$. Then we have

$\langle A, C\rangle = \operatorname{trace}(C^TA) = \operatorname{trace}(V\tilde{\Sigma}U^TU\Sigma V^T) = \operatorname{trace}(\tilde{\Sigma}\Sigma) =$

$= \left(\sum_{i=1}^n \sigma_i^{1+p/q}\right) / \|\tilde{C}\|_{(q)} = \left(\sum_{i=1}^n \sigma_i^p\right) / \left(\sum_{i=1}^n \sigma_i^p\right)^{1/q} =$

$= \left(\sum_{i=1}^n \sigma_i^p\right)^{1-\frac{1}{q}} = \left(\sum_{i=1}^n \sigma_i^p\right)^{1/p} = \|A\|_{(p)}.$ $\square$

③ $\|A\|_{(p)} \geq 0$ trivial; $\|\alpha A\|_{(p)} = |\alpha| \cdot \|A\|_{(p)}$ proved in point 1.

Triangular inequality: $\|A+B\|_{(p)} \overset{(\text{point } 2)}{=} \max_{\substack{\|C\|_{(q)}=1 \\ \downarrow \\ \text{with } \frac{1}{p}+\frac{1}{q}=1}} \langle A+B, C\rangle$

$\leq \max_{\|C\|_{(q)}=1} \langle A, C\rangle + \max_{\|C\|_{(q)}=1} \langle B, C\rangle \overset{\swarrow \text{again point 2}}{=}$

$= \|A\|_{(p)} + \|B\|_{(p)}.$ $\square$

[SOLUTIONS]

PROBLEM 3

① Let $z \in \mathbb{R}^n$, then $\|(P_1-P_2)z\|_2^2 = z^T(P_1-P_2)^T(P_1-P_2)z =$

$= z^T \underbrace{P_1^T P_1}_{= P_1 \cdot P_1 = P_1 = P_1^T} z - z^T P_1^T P_2 z - z^T P_2^T P_1 z + z^T P_2^T P_2 z = z^T P_1^T z - z^T P_1^T P_2 z - z^T P_2^T P_1 z \cdots$

$+ z^T P_2^T z = (P_1 z)^T (I-P_2) z + (P_2 z)^T (I-P_1) z.$ $(*)$

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 $(*)$ is 0 for all $z \in \mathbb{R}^n$. Note that $P_1 z \in S$ so there exists $w \in \mathbb{R}^n$ such that $P_2 w = P_1 z$ (because range$(P_2) = S$); then $(P_1 z)^T (I-P_2) z = w^T P_2^T (I-P_2) z =$
$= w^T (P_2^T - P_2^T P_2) z = w^T (P_2 - P_2^2) z = w^T (P_2 - P_2) z = 0$. Analogously, $(P_2 z)^T (I-P_1) z = 0 \quad \forall z \in \mathbb{R}^n$. We then conclude that $P_1 = P_2$. □

② If X is a basis of S, then $X^T X$ is invertible so $X(X^TX)^{-1}X^T$ exists; as the matrix $(X^TX)^{-1}X^T$ has full row-rank, we have that range$(X \cdot (X^TX)^{-1}X^T) =$ range$(X) = S$. Moreover, $(X(X^TX)^{-1}X^T)^T = X \overbrace{(X^TX)}^{\text{symmetric}}{}^{-T} X^T$
$= X \cdot (X^TX)^{-1} X^T$ and $[X \cdot (X^TX)^{-1}X^T]^2 = X(X^TX)^{-1} \boxed{\underbrace{X^TX(X^TX)^{-1}}_{=I}} X^T = X(X^TX)^{-1}X^T$
so all the properties of the definition are satisfied. □

③ If P is an orthogonal projection onto a k-dim. subspace, $P \cdot A$ has rank at most k. As the minimizer of $\|A-B\|_F$ for $B \in \{$matrices of rank $\le k\}$ is unique when all the sing. 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 also $\min_{P \ldots} \|(I-P)A\|_F$ 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 $P = qq^T$ with $q = \begin{bmatrix} \sqrt{1-\varepsilon^2} \\ 0 \\ \varepsilon \end{bmatrix}$ for $\varepsilon > 0$ small enough.

$$\hookrightarrow \|A - qq^TA\|_2 = \left\| \begin{bmatrix} 3\varepsilon^2 & 0 & -\varepsilon\sqrt{1-\varepsilon^2} \\ 0 & 2 & 0 \\ -3\varepsilon\sqrt{1-\varepsilon^2} & 0 & 1-\varepsilon^2 \end{bmatrix} \right\|_2 =$$

$$= \left\| \left[\begin{array}{cc:c} 3\varepsilon^2 & -\varepsilon\sqrt{1-\varepsilon^2} & 0 \\ -3\varepsilon\sqrt{1-\varepsilon^2} & 1-\varepsilon^2 & 0 \\ \hdashline 0 & 0 & 2 \end{array}\right] \right\|_2 \le \max\left\{ \left\| \begin{bmatrix} 3\varepsilon^2 & -\varepsilon\sqrt{1-\varepsilon^2} \\ -3\varepsilon\sqrt{1-\varepsilon^2} & 1-\varepsilon^2 \end{bmatrix} \right\|_F, 2 \right\}$$

use that $\left\| \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} \right\|_2 \le \max\{\|A_1\|_2, \|A_2\|_2\}$ [PROVE THIS!]
and $\|A_1\|_2 \le \|A_1\|_F$ □

[SOLUTIONS]

PROBLEM 4

Consider an SVD of $A = U\Sigma V^T$ with $\Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_n)$ and define

$A_k := U \cdot \operatorname{diag}\left(\sigma_1 + \frac{1}{k}, \dots, \sigma_n + \frac{1}{k}\right) \cdot V^T$.

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

$$\| A - A_k \|_2 = \| U \cdot \left(\operatorname{diag}(\sigma_1, \dots, \sigma_n) - \operatorname{diag}(\sigma_1 + \tfrac{1}{k}, \dots, \sigma_n + \tfrac{1}{k}) \right) \cdot V^T \|_2 =$$

$$= \| \operatorname{diag}(1/k, \dots, 1/k) \|_2 = 1/k \longrightarrow 0 \quad \text{for } k \to +\infty. \qquad \square$$

Spectral norm is unitarily invariant

PROBLEM 5

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

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

PROBLEM 6

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

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

$$= \max_{\substack{\|x\|_2 = 1 \\ x \in \mathbb{R}^n}} x^T \cdot \begin{bmatrix} BB^T & Bc \\ c^T B & c^T c \end{bmatrix} x \geq \max_{\substack{\|x\|_2 = 1 \text{ with} \\ \text{last component} = 0 \\ x \in \mathbb{R}^n}} x^T \cdot \begin{bmatrix} BB^T & Bc \\ c^T B & c^T c \end{bmatrix} x =$$

because we take the max of the same quantity on a smaller set

$$= \max_{\substack{\|y\|_2 = 1 \\ y \in \mathbb{R}^{n-1}}} y^T BB^T y = \lambda_{max}(BB^T) = \sigma_1(B)^2. \qquad \square$$

Bonus question: what happens to the SMALLEST singular value?

(solution in the following page)

[SOLUTIONS]

* Let $B \in \mathbb{R}^{m\times n}$ with $m \geq n$ and consider $A = \overbrace{\begin{bmatrix} B \\ c^T \end{bmatrix}}^{n} \Big\}\, m+1$

Then $\sigma_{\min}(A) \geq \sigma_{\min}(B)$

Proof: $\sigma^2_{\min}(A) = \lambda_{\min}(A^TA) = \lambda_{\min}(B^TB + cc^T) =$

$$= \min_{\substack{\|x\|_2 = 1 \\ x\in\mathbb{R}^n}} x^T(B^TB+cc^T)x = \min_{\substack{\|x\|_2 = 1 \\ x\in\mathbb{R}^n}} (x^TB^TBx + \underbrace{x^Tcc^Tx}_{\geq 0}) \geq$$

$$\geq \min_{\substack{\|x\|_2 = 1 \\ x\in\mathbb{R}^n}} x^TB^TBx = \lambda_{\min}(B^TB) = \sigma^2_{\min}(B). \quad \blacksquare$$

* Let $B \in \mathbb{R}^{m\times n}$ with $m \leq n-1$ and consider $A = \overbrace{\begin{bmatrix} B \\ c^T \end{bmatrix}}^{n} \Big\}\, m+1$

Then $\sigma_{\min}(A) \leq \sigma_{\min}(B)$.

Proof: $\sigma^2_{\min}(A) = \lambda_{\min}(AA^T) = \lambda_{\min}\left(\begin{bmatrix} BB^T & Bc \\ c^TB^T & c^Tc \end{bmatrix}\right) =$

$$= \min_{\substack{\|x\|_2 = 1 \\ x\in\mathbb{R}^{m+1}}} x^T \cdot \begin{bmatrix} BB^T & Bc \\ c^TB^T & c^Tc \end{bmatrix} x \leq \min_{\substack{x = (y,0)^T \\ y\in\mathbb{R}^m \\ \|y\|_2 = 1}} [y^T, 0] \cdot \begin{bmatrix} BB^T & Bc \\ c^TB^T & c^Tc \end{bmatrix} \begin{bmatrix} y \\ 0 \end{bmatrix} =$$

we are taking the min. of the same quantity on a smaller set

$$= \min_{\substack{y\in\mathbb{R}^m \\ \|y\|_2 = 1}} y^TBB^Ty = \lambda_{\min}(BB^T) = \sigma^2_{\min}(B). \quad \blacksquare$$