Midterm

1. **Honor Code (0 pts)**

   Please copy the following statement in the space provided below and sign your name.

   As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others. I will follow the rules and do this exam on my own.

   IF YOU DO NOT COPY THE HONOR CODE AND SIGN YOUR NAME,
   YOU WILL GET A 0 ON THE EXAM.

Solution:

2. **SID (3 pts)**

   WHEN THE EXAM STARTS, WRITE YOUR SID AT THE TOP OF EVERY PAGE.
   No extra time will be given for this task.

3. **Favorites. Any answer, as long as you write it down, will be given full credit. (2 pts)**

   (a) (1 pts) **What's your favorite restaurant in Berkeley?**

   Solution: Any answer is fine.

   (b) (1 pts) **What's some music that makes you happy?**

   Solution: Any answer is fine.

4. Fun with Rank (12 pts)

Consider two matrices $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times p}$. Let $\mathcal{R}(A)$ denote the range (i.e. column) space of the matrix A.

(a) (4 pts) Prove that $\mathcal{R}(AB) \subseteq \mathcal{R}(A)$.

HINT: $\mathcal{R}(A) = \{\vec{y} : \vec{y} = A\vec{x}$ *for some* $\vec{x} \in \mathbb{R}^n\}$.

Solution: To solve this problem, it is sufficient to show that if $\vec{y} \in \mathcal{R}(AB)$ then $\vec{y} \in \mathcal{R}(A)$. Consider arbitrary $\vec{y} \in \mathcal{R}(AB)$ then there exists a $\vec{z} \in \mathbb{R}^p$ such that $\vec{y} = AB\vec{z}$. Define $\vec{w} \in \mathbb{R}^n$ to be $\vec{w} = B\vec{z} \in \mathbb{R}^n$. Consequently, it holds that $\vec{y} = A\vec{w}$. Therefore, we can conclude that $\vec{y} \in \mathcal{R}(A)$.

Alternate Solution: An alternate solution is to note that we can represent the product matrix AB as

$$AB = \begin{bmatrix} A\vec{b}_1 & A\vec{b}_2 & \dots & A\vec{b}_p \end{bmatrix},$$

where $\vec{b}_1, \vec{b}_2, ..., \vec{b}_p \in \mathbb{R}^n$ are the columns of matrix B.

Recall that the range space of any matrix is the span of its columns. Therefore, $\mathcal{R}(AB) = \text{span}\{A\vec{b}_1, A\vec{b}_2, ..., A\vec{b}_p\}$. Since $A\vec{b}_i \in \mathcal{R}(A)$ for every $i \in \{1, 2, .., p\}$, it holds that $\text{span}\{A\vec{b}_1, A\vec{b}_2, ..., A\vec{b}_p\} \subseteq \mathcal{R}(A)$.

(b) (4 pts) Prove that the following inequality holds:

$$0 \leq \text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}.$$

HINT: Recall that the rank of a matrix is the dimension of its range space.

HINT: You may use the result of part (a), and the fact that the rank of any matrix is the same as the rank of its transpose.

Solution: Recall that rank of any matrix is the dimension of the column space of that matrix. Therefore the rank has to be always non-negative. This proves the lower bound in the problem.

To show the upper bound, it is enough to show that

$$\text{rank}(AB) \leq \text{rank}(A); \tag{P1}$$

$$\text{rank}(AB) \leq \text{rank}(B). \tag{P2}$$

First, we show (P1). This is a consequence of part (a) where we showed that $\mathcal{R}(AB) \subseteq \mathcal{R}(A)$. Therefore

$$\begin{aligned}\text{rank}(AB) &= \dim(\mathcal{R}(AB)) \\ &\leq \dim(\mathcal{R}(A)) = \text{rank}(A).\end{aligned}$$

Next, we show (P2). Observe that

$$\begin{aligned}\text{rank}(AB) &\underset{\star}{=} \text{rank}((AB)^\top) \\ &= \text{rank}(B^\top A^\top) \\ &\underset{\star\star}{\leq} \text{rank}(B^\top) \\ &= \text{rank}(B),\end{aligned}$$

where $\star$ is due to the fact that the rank of any matrix is same as the rank of its transpose and $\star\star$ is due to (P1) by replacing A with $B^\top$ and B with $A^\top$.

(c) (4 pts) Give an example of matrices A, B such that $\text{rank}(A) \neq 0, \text{rank}(B) \neq 0$ but $\text{rank}(AB) = 0$, by finding suitable values of $x, y, z \in \mathbb{R}$ in the following matrices:

$$A = \begin{bmatrix} x & y \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 1 & z \end{bmatrix}.$$

Solution: In general, we can show that any solution with $x + y = 0$, $x \neq 0$, $z = 0$ will work.

To do this, we multiply AB and get:

$$AB = \begin{bmatrix} x & y \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & z \end{bmatrix} = \begin{bmatrix} x + y & yz \\ 0 & 0 \end{bmatrix}.$$

For this to have rank zero, we need $x + y = 0$ and $yz = 0$. Also, for A to have nonzero rank, we need at least one of x and y to be nonzero.

For a specific example, consider the matrix

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}. \tag{1}$$

Note that $\text{rank}(A) = \text{rank}(B) = 1$. But

$$AB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Therefore, $\text{rank}(AB) = 0$. This corresponds to $x = 1$, $y = -1$, $z = 0$.

5. Compact SVD (4 pts)

Consider two sets of orthonormal vectors $\{\vec{p}_1, \vec{p}_2\} \subset \mathbb{R}^m$ with $\vec{p}_1 \perp \vec{p}_2$, and $\{\vec{q}_1, \vec{q}_2\} \subset \mathbb{R}^n$ with $\vec{q}_1 \perp \vec{q}_2$. Let $C \in \mathbb{R}^{m\times n}$ be defined as

$$C = \vec{p}_1\vec{q}_1^\top + \frac{1}{2}\vec{p}_2\vec{q}_2^\top.$$

Write the compact SVD representation of matrix C in terms of $\vec{p}_1, \vec{p}_2, \vec{q}_1, \vec{q}_2$. That is, compute the SVD matrices $U_r \in \mathbb{R}^{m\times r}$, $\Sigma_r \in \mathbb{R}^{r\times r}$, $V_r \in \mathbb{R}^{n\times r}$, such that $C = U_r\Sigma_r V_r^\top$.

Solution: The compact SVD representation of matrix C in terms of $\vec{p}_1, \vec{p}_2, \vec{q}_1, \vec{q}_2$ with all three matrices is given by:

$$C = \begin{bmatrix} \vec{p}_1 & \vec{p}_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \vec{q}_1^\top \\ \vec{q}_2^\top \end{bmatrix}.$$

6. Exploring SVD, Least Squares, and Min-Norm (13 pts)

(a) (4 pts) Let $A \in \mathbb{R}^{m\times n}$ be a matrix with rank $r > 0$. Consider the equation $A\vec{x} = \vec{b}$ for some $\vec{b} \in \mathcal{R}(A)$. Show that $\vec{x}_0 = V_r\Sigma_r^{-1}U_r^\top\vec{b}$ is a solution to $A\vec{x} = \vec{b}$, where the compact SVD of A is $A = U_r\Sigma_rV_r^\top$. *Show your work.*

HINT: Remember that $U_rU_r^\top$ isn't necessarily the identity, but $U_rU_r^\top\vec{d} = \vec{d}$ for any $\vec{d} \in \mathcal{R}(U_r)$.

Solution: One way is to note that $A\vec{x}_0 = U_r\Sigma_rV_r^\top V_r\Sigma_r^{-1}U_r^\top\vec{b} = U_r\Sigma_r\Sigma_r^{-1}U_r^\top\vec{b} = U_rU_r^\top\vec{b} = \vec{b}$, where we used the fact that V_r has orthonormal columns (so $V_r^\top V_r = I$). The last step involves noting that the columns of U_r span $\mathcal{R}(A)$ and $\vec{b} \in \mathcal{R}(A)$, so $\vec{b} = U_r\vec{z}$ for some $\vec{z}$. Then, $U_rU_r^\top\vec{b} = U_rU_r^\top U_r\vec{z} = U_r\vec{z} = \vec{b}$.

Another way is to see that $Ax = U_r\Sigma_rV_r^\top x = b$. Left multiplying on both sides by $U_r^\top$ and using the fact that $U_r^\top U_r = I$, we have $\Sigma_rV_r^\top x = U_r^\top b$. Once again left multiplying on both sides by Σ_r^{-1}, we have $V_r^\top x = \Sigma_r^{-1}U_r^\top b$. Now, we substitute $x = x_0 = V_r\Sigma_r^{-1}U_r^\top b$ to see that $V_r^\top V_r\Sigma_r^{-1}U_r^\top b = \Sigma_r^{-1}U_r^\top b$, and using $V_r^\top V_r = I$ shows the left hand side and right hand side are equal.

(b) (4 pts) Let $A \in \mathbb{R}^{m\times n}$ be a matrix with $m < n$ and rank $r > 0$, and let $\vec{b} \in \mathcal{R}(A)$.

Also, let the compact and full SVD representations of A be, respectively,

$$\underbrace{A}_{m\times n} = \underbrace{U_r}_{m\times r}\underbrace{\Sigma_r}_{r\times r}\underbrace{V_r^\top}_{r\times n}, \qquad A = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \underbrace{\begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}}_{m\times n} \begin{bmatrix} V_r^\top \\ V_{n-r}^\top \end{bmatrix},$$

where $U_{m-r} \in \mathbb{R}^{r\times(m-r)}$, $V_{n-r} \in \mathbb{R}^{r\times(n-r)}$. Show that $V_r\Sigma_r^{-1}U_r^\top\vec{b} + V_{n-r}\vec{z}$ is a solution to $A\vec{x} = \vec{b}$, for any $\vec{z} \in \mathbb{R}^{n-r}$.

Solution: We plug in the given solution into $A\vec{x}$, yielding $A(V_r\Sigma_r^{-1}U_r^\top\vec{b} + V_{n-r}\vec{z}) = AV_r\Sigma_r^{-1}U_r^\top\vec{b} + AV_{n-r}\vec{z} = U_r\Sigma_r V_r^\top V_r\Sigma_r^{-1}U_r^\top\vec{b} + U_r\Sigma_r(V_r^\top V_{n-r})\vec{z}$. Then, we get $U_r\Sigma_r\Sigma_r^{-1}U_r^\top\vec{b} = U_rU_r^\top\vec{b} = \vec{b}$, where we used the fact that V_r has orthonormal columns (so $V_r^\top V_r = I$, $V_r^\top V_{n-r} = 0$).

(c) (5 pts) Again, let $A \in \mathbb{R}^{m\times n}$ be a matrix with $m < n$ and rank r, and let $\vec{b} \in \mathcal{R}(A)$.

Also, let the compact and full SVD representations of A be, respectively,

$$\underbrace{A}_{m\times n} = \underbrace{U_r}_{m\times r}\underbrace{\Sigma_r}_{r\times r}\underbrace{V_r^\top}_{r\times n}, \qquad A = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \underbrace{\begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}}_{m\times n} \begin{bmatrix} V_r^\top \\ V_{n-r}^\top \end{bmatrix},$$

where $U_{m-r} \in \mathbb{R}^{r\times(m-r)}$, $V_{n-r} \in \mathbb{R}^{r\times(n-r)}$.

Let $\vec{x}^\star$ be the solution to the following problem:

$$\vec{x}^\star = \operatorname*{argmin}_{\vec{x}\in\mathbb{R}^n} \|\vec{x}\|_2^2 \quad \text{s.t.} \quad A\vec{x} = \vec{b}.$$

Find $\vec{x}^\star$ and justify your answer.

HINT: You may use the fact that $\{V_r\Sigma_r^{-1}U_r^\top\vec{b} + V_{n-r}\vec{z} : \vec{z} \in \mathbb{R}^{n-r}\}$ *is the set of all solutions to* $A\vec{x} = \vec{b}$*, without proof.*

Solution: Using the fact in the hint, we reduce our problem to

$$\min_{\vec{z}\in\mathbb{R}^{n-r}} \|V_r\Sigma_r^{-1}U_r^\top\vec{b} + V_{n-r}\vec{z}\|_2^2.$$

We can expand the inner product to get:

$$\min_{\vec{z}\in\mathbb{R}^{n-r}} \left(\|V_r\Sigma_r^{-1}U_r^\top\vec{b}\|_2^2 + \|V_{n-r}\vec{z}\|_2^2 + 2(V_r\Sigma_r^{-1}U_r^\top\vec{b})^\top(V_{n-r}\vec{z}) \right).$$

Now, note that V is an orthogonal matrix (or that $\mathcal{R}(V_r) = \mathcal{R}(A^\top), \mathcal{R}(V_{n-r}) = N(A)$), so $\mathcal{R}(V_r) \perp \mathcal{R}(V_{n-r})$. Thus, the last term (an inner product between vectors in both of those subspaces) is zero, and we are left with $\min_{\vec{z}\in\mathbb{R}^{n-r}}(\|V_r\Sigma_r^{-1}U_r^\top\vec{b}\|_2^2 + \|V_{n-r}\vec{z}\|_2^2)$.

The first term here is independent of $\vec{z}$, so this becomes $\|V_r\Sigma_r^{-1}U_r^\top\vec{b}\|_2^2 + \min_{\vec{z}\in\mathbb{R}^{n-r}} \|V_{n-r}\vec{z}\|_2^2$. The second term is lower bounded by 0, which is achieved when $\vec{z} = 0$ (and this is the sole minimizer since V_{n-r} is full column rank). So, the answer is $p^\star = \|V_r\Sigma_r^{-1}U_r^\top\vec{b}\|_2^2$ and $\vec{z}^\star = 0$, which corresponds to $\vec{x}^\star = V_r\Sigma_r^{-1}U_r^\top\vec{b}$.

7. Convexity (10 pts)

(a) (6 pts) Let $f : \mathbb{R}^n \to \mathbb{R}$ be defined as $f(\vec{x}) = \|A\vec{x}\|_2^2$, where $A \in \mathbb{R}^{m \times n}$ is a matrix. Is f a convex function? Prove or disprove.

Solution:

One solution is to derive the Hessian $2A^T A$ and note that it is symmetric positive semidefinite.

(b) (4 pts) Let $g, h : \mathbb{R}^n \to \mathbb{R}$ be fixed twice-differentiable convex functions, and fix real numbers $a, b > 0$. Define $f : \mathbb{R}^n \to \mathbb{R}$ by $f(\vec{x}) = a \cdot g(\vec{x}) + b \cdot h(\vec{x})$ for each $\vec{x} \in \mathbb{R}^n$. Prove f is a convex function.

Solution: Fix $\vec{x}, \vec{y} \in \mathbb{R}^n$, $\alpha \in [0, 1]$ arbitrarily. Then, appealing to the definition of convexity:

$$\begin{aligned}
& f(\alpha\vec{x} + (1-\alpha)\vec{y}) \\
&= a \cdot g(\alpha\vec{x} + (1-\alpha)\vec{y}) + b \cdot h(\alpha\vec{x} + (1-\alpha)\vec{y}) \\
&= a\alpha \cdot g(\vec{x}) + a(1-\alpha) \cdot g(\vec{y}) + b\alpha \cdot h(\vec{x}) + b(1-\alpha) \cdot h(\vec{y}) \\
&\leq \alpha \cdot \big[a \cdot g(\vec{x}) + b \cdot h(\vec{x})\big] + (1-\alpha) \cdot \big[a \cdot g(\vec{y}) + b \cdot h(\vec{y})\big] \\
&= \alpha \cdot f(\vec{x}) + (1-\alpha) \cdot f(\vec{y}).
\end{aligned}$$

Thus, f is a convex function.

8. Vector Calculus (12 pts)

(a) (6 pts) Compute the gradient and Hessian with respect to $\vec{x}$ of the function $f\colon \mathbb{R}^n \to \mathbb{R}$

$$f(\vec{x}) = 1 - (\vec{a}^\top \vec{x})^2, \tag{2}$$

where $\vec{a} \in \mathbb{R}^n$.

Solution: By the chain rule,

$$\nabla f(\vec{x}) = -2(\vec{a}^\top \vec{x})\vec{a}.$$

Furthermore, the Hessian is seen to be

$$\nabla^2 f(\vec{x}) = -2\vec{a}\vec{a}^\top.$$

(b) (6 pts) Consider $f(\vec{x}) = \sum_{i=1}^{m} \log\left(b_i - \vec{a}_i^\top \vec{x}\right)$, where $\vec{a}_i \in \mathbb{R}^n$ for $i = 1, \ldots, m$, and $b_1, b_2, \ldots, b_m > 0$.

The domain of f is the set $\{\vec{x} \in \mathbb{R}^n \mid b_i - \vec{a}_i^\top \vec{x} > 0 \text{ for all } i = 1, \ldots, m\}$, which is assumed to be nonempty.

Compute the gradient of $f(\vec{x})$ with respect to $\vec{x}$.

HINT: Consider what happens in the special case of $f(x) = log(b - x)$ for a scalar variable $x \in \mathbb{R}$. Then, use the chain rule.

HINT: Recall that $\dfrac{d}{dx}\log(x) = \dfrac{1}{x}$.

Solution: The derivative of $\log(b - x)$ with respect to a scalar variable x is easily seen to be $-\frac{1}{b-x}$. Now, using the chain rule, the gradient of $\log\left(b - \vec{a}^\top \vec{x}\right)$ with respect to $\vec{x}$ is

$$-\frac{\nabla(\vec{a}^\top \vec{x})}{b - \vec{a}^\top \vec{x}} = -\frac{\vec{a}}{b - \vec{a}^\top \vec{x}}.$$

Hence, summing the gradients for each term, we get:

$$\nabla f(\vec{x}) = -\sum_{i=1}^{m} \frac{\vec{a}_i}{b_i - \vec{a}_i^\top \vec{x}}.$$

9. Best Approximations (24 pts)

We start by recalling the Eckart-Young Theorem: Consider any square matrix $C \in \mathbb{R}^{n\times n}$ and assume that we can write its full singular value decomposition as $C = U\Sigma V^\top$ where Σ is the $n \times n$ diagonal matrix with distinct diagonal entries $\sigma_1 > \sigma_2 > \cdots > \sigma_n > 0$. Then, for $0 \le k \le n$, the Eckart-Young Theorem for Frobenius norm states that

$$C_k = U_k \Sigma_k V_k^\top = \underset{\substack{B\in\mathbb{R}^{n\times n}\\ \text{rank}(B)\le k}}{\text{argmin}} \|C - B\|_F$$

where $U_k \in \mathbb{R}^{n\times k}$ is the matrix consisting of the first k columns of U, $V_k \in \mathbb{R}^{n\times k}$ is the matrix consisting of the first k columns of V, and Σ_k is the $k \times k$ diagonal matrix with the top-k singular values $\sigma_1 > \cdots > \sigma_k$ as its diagonal entries.

(a) (8 pts) Now, consider any square matrix $C \in \mathbb{R}^{n\times n}$ and let $C_k \in \mathbb{R}^{n\times n}$ denote its best rank-k approximation in the Frobenius norm. Then, for any orthonormal matrix $W \in \mathbb{R}^{n\times n}$, show that

$$WC_kW^\top = \underset{\substack{B\in\mathbb{R}^{n\times n}\\ \text{rank}(B)\le k}}{\text{argmin}} \|WCW^\top - B\|_F.$$

Solution:

Using the fact that the Frobenius norm is invariant under orthogonal transformations, we have $\|B - WCW^\top\|_F = \|W^\top BW - C\|_F$. Therefore, we can write

$$\min_{\substack{B\in\mathbb{R}^{n\times n}\\ \text{rank}(B)\le k}} \|B - WCW^\top\|_F = \min_{\substack{B\in\mathbb{R}^{n\times n}\\ \text{rank}(B)\le k}} \|W^\top BW - C\|_F.$$

Next, we show that $\{W^\top BW : B \in \mathbb{R}^{n\times n}, \text{rank}(B) \le k\} = \{Z \in \mathbb{R}^{n\times n} : \text{rank}(Z) \le k\}$. For any $B \in \mathbb{R}^{n\times n}$ such that $\text{rank}(B) \le k$, $Z = W^\top BW$ satisfies $\text{rank}(Z) \le k$. On the other hand, for any $Z \in \mathbb{R}^{n\times n}$ such that $\text{rank}(Z) \le k$, there exists $B = WZW^\top$ that satisfies $\text{rank}(B) \le k$ and $W^\top BW = W^\top WZW^\top W = Z$ because W is an orthonormal matrix. Therefore,

$$\begin{aligned}\min_{\substack{B\in\mathbb{R}^{n\times n}\\ \text{rank}(B)\le k}} \|W^\top BW - C\|_F &= \min_{\substack{Z\in\mathbb{R}^{n\times n}\\ \text{rank}(Z)\le k}} \|Z - C\|_F \\ &= \|C_k - C\|_F.\end{aligned}$$

Lastly, noting that $\|WC_kW^\top - WCW^\top\|_F = \|W(C_k - C)W^\top\|_F = \|C_k - C\|_F$, we can write

$$\|WC_kW^\top - WCW^\top\|_F = \min_{\substack{B\in\mathbb{R}^{n\times n}\\ \text{rank}(B)\le k}} \|B - WCW^\top\|_F.$$

Since $\text{rank}(WC_kW^\top) \le k$, we conclude that $WC_kW^\top$ is a minimizer of the given optimization problem.

For the remainder of this problem, we consider a square matrix $A \in \mathbb{R}^{n \times n}$ and assume A has full rank. Using Gram-Schmidt Orthonormalization (GSO), we can write the matrix A as

$$A = QR \tag{3}$$

where $Q \in \mathbb{R}^{n \times n}$ is an orthonormal matrix and $R \in \mathbb{R}^{n \times n}$ is an upper triangular matrix.

(b) (4 pts) Find the best rank-n approximation to $AA^\top$ in the Frobenius norm. *Justify your answer.*

Solution: $AA^\top$ is the best rank-n approximation since it already has rank n.

(c) (6 pts) Recall that $A \in \mathbb{R}^{n \times n}$ is a square matrix with full rank, and we can write the matrix A as

$$A = QR \tag{4}$$

where $Q \in \mathbb{R}^{n \times n}$ is an orthonormal matrix and $R \in \mathbb{R}^{n \times n}$ is an upper triangular matrix.

Assume that $R = \text{diag}(r_1, r_2, \ldots, r_n) \in \mathbb{R}^{n \times n}$ is a diagonal matrix with

$$|r_1| > |r_2| > \cdots > |r_n|,$$

and all $r_1, r_2, \ldots, r_n$ are real numbers. Let $k < n$. Then, show that the best rank-k approximation to $AA^\top$ in the Frobenius norm is $QSQ^\top$, where S is a diagonal matrix defined as

$$S = \text{diag}(r_1^2, \ldots, r_k^2, 0, \ldots, 0) \in \mathbb{R}^{n \times n}.$$

Solution: We can write $AA^\top = QRR^\top Q^\top$. Note that Q is an orthonormal matrix and $RR^\top = \text{diag}(r_1^2, r_2^2, \ldots, r_n^2)$ is a diagonal matrix. Therefore, $AA^\top = QRR^\top Q^\top$ forms an SVD for $AA^\top$. Then, the result follows by the Eckart-Young Theorem.

(d) (6 pts) Recall that $A \in \mathbb{R}^{n\times n}$ is a square matrix with full rank, and we can write the matrix A as

$$A = QR \tag{5}$$

where $Q \in \mathbb{R}^{n\times n}$ is an orthonormal matrix and $R \in \mathbb{R}^{n\times n}$ is an upper triangular matrix.

Now, we no longer assume that R is diagonal. Let $k < n$ and assume that the best rank-k approximation to $RR^\top \in \mathbb{R}^{n\times n}$ in the Frobenius norm is given by $G \in \mathbb{R}^{n\times n}$. Then, using the result of part (a), show that the best rank-k approximation to $AA^\top$ in the Frobenius norm is given by $QGQ^\top$.

Solution: We can write $AA^\top = QRR^\top Q^\top$. We use part (a) with $W = Q$ and $C = RR^\top$. As a result,

$$QGQ^\top = \operatorname*{argmin}_{\substack{B\in\mathbb{R}^{n\times n}\\ \text{rank}(B)\le k}} \|B - QRR^\top Q^\top\|_F.$$

Therefore, the best rank-k approximation to $AA^\top$ in the Frobenius norm is given by $QGQ^\top$.

