Problem Set 7 — Not Fully Graded
For the Exercise Sessions on Dec 13 and Dec 20

<table>
<thead>
<tr>
<th>Last name</th>
<th>First name</th>
<th>SCIPER Nr</th>
<th>Points</th>
</tr>
</thead>
</table>

Problem 1: Canonical Correlation Analysis

Let $X$ and $Y$ be zero-mean real-valued random vectors with covariance matrices $R_X$ and $R_Y$, respectively. Moreover, let $R_{XY} = E[XY^T]$. Our goal is to find vectors $u$ and $v$ such as to maximize the correlation between $u^T X$ and $v^T Y$, that is,

$$
\max_{u,v} \frac{E[u^T X Y^T v]}{\sqrt{E[|u^T X|^2]} \sqrt{E[|v^T Y|^2]}}.
$$

(1)

Show how we can find the optimizing choices of the vectors $u$ and $v$ from the problem parameters $R_X, R_Y$, and $R_{XY}$.

**Hint:** Recall for the singular value decomposition that

$$
\max_v \frac{\|Av\|}{\|v\|} = \max_{\|v\|=1} \|Av\| = \sigma_1(A),
$$

(2)

where $\sigma_1(A)$ denotes the maximum singular value of the matrix $A$. The corresponding maximizer is the right singular vector $v_1$ (i.e., eigenvector of $A^T A$) corresponding to $\sigma_1(A)$.

**Solution**

Adapting the hint to this scenario, we prove that

$$
\max_{u,v} \frac{E[u^T X Y^T v]}{\sqrt{E[|u^T X|^2]} \sqrt{E[|v^T Y|^2]}} = \max_{u,v} \frac{\|u^T X Y^T v\|}{\sqrt{E[|u^T X|^2]} \sqrt{E[|v^T Y|^2]}}
$$

$$
= \max_{u,v} \frac{\|u^T X Y^T v\|}{\sqrt{E[|u^T X|^2]} \sqrt{E[|v^T Y|^2]}}
$$

$$
= \max_{u,v} \frac{u^T R_{XY} v}{\sqrt{u^T R_X u} \sqrt{v^T R_Y v}}
$$

$$
= \max_{\|u\|=\sqrt{1}} u^T R_{X}^{-1/2} R_{XY} R_{Y}^{-1/2} v
$$

$$
= \sigma_1(R_{X}^{-1/2} R_{XY} R_{Y}^{-1/2})
$$

(3)
Problem 2: Some review problems on linear algebra

(a) (Frobenius norm) Prove that \( \|A\|_F^2 = \text{trace}(A^H A) \).

(b) (Singular Value Decomposition) Let \( \sigma_i(A) \) denote the \( i \)th singular value of an \( m \times n \) matrix \( A \). Prove that \( \|A\|_F^2 = \sum_{i=1}^{\min\{m,n\}} \sigma_i^2(A) \)

(c) (Projection Matrices) Consider a set of \( k \) orthonormal vectors in \( \mathbb{C}^n \), denoted by \( \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k \).

The projection matrix (that projects an arbitrary vector into the subspace spanned by these orthonormal vectors) is given by

\[
P = \sum_{i=1}^{k} \mathbf{u}_i \mathbf{u}_i^H.
\]  

- Prove that this matrix is Hermitian, i.e., \( P^H = P \).
- Prove that this matrix is idempotent, i.e., \( P^2 = P \). (In words, projecting twice into the same subspace is the same as projecting only once.)
- Prove that \( \text{trace}(P) = k \), i.e., equal to the dimension of the subspace.
- Prove that the diagonal entries of \( P \) must be real-valued and non-negative. Then, prove that the diagonal entries of \( P \) cannot be larger than 1 (this is a little more tricky).

Solution

(a) Let \( A \) be an \( m \times n \) matrix and denote by \( a_{ij} \) the entry of \( A \) at row \( i \) and column \( j \). Hence, we have

\[
A^H A = \begin{bmatrix}
\sum_{i=1}^{m} |a_{i1}|^2 & \sum_{i=1}^{m} a_{i1}^* a_{i2} & \cdots & \sum_{i=1}^{m} a_{i1}^* a_{in} \\
\sum_{i=1}^{m} a_{i2}^* a_{i1} & \sum_{i=1}^{m} |a_{i2}|^2 & \cdots & \sum_{i=1}^{m} a_{i2}^* a_{in} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{i=1}^{m} a_{in}^* a_{i1} & \sum_{i=1}^{m} a_{in}^* a_{i2} & \cdots & \sum_{i=1}^{m} |a_{in}|^2
\end{bmatrix}
\]  

\[
\text{trace}(A^H A) = \sum_{j=1}^{n} \sum_{i=1}^{m} |a_{ij}|^2 = \left( \sum_{j=1}^{n} \sum_{i=1}^{m} |a_{ij}|^2 \right)^2 = \|A\|_F^2.
\]  

(b) Let \( U \Sigma V^H \) be the singular value decomposition of \( A \). Then \( U \) is an \( m \times m \) unitary matrix (i.e. \( U^H U = I \)), \( \Sigma \) is a diagonal \( m \times n \) matrix (\( \Sigma_{ii} = \sigma_i(A) \)), and \( V \) is an \( n \times n \) unitary matrix (i.e. \( V^H V = I \)). using part (a) we have

\[
\|A\|_F^2 = \text{trace}(A^H A) = \text{trace}\left((U \Sigma V^H)^H U \Sigma V^H\right) = \text{trace}(V \Sigma H U^H U \Sigma V^H) = \text{trace}(\Sigma H \Sigma) = \sum_{i=1}^{\min\{m,n\}} \sigma_i^2(A)
\]  

where we use the property of unitary matrix and the cyclic property of trace.
(c) For the first bullet item, \( P^H = \left( \sum_{i=1}^{k} u_i u_i^H \right)^H = \sum_{i=1}^{k} u_i u_i^H = P. \)

For the second bullet item, \( P^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} u_i u_i^H u_j u_j^H. \) Since \( \{u_1, \ldots, u_k\} \) are orthonormal vectors, we have

\[
u_i^H u_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}
\] (12)

Hence, all the terms with \( i = j \) survive. \( P^2 = \sum_{i=1}^{k} u_i u_i^H = P. \)

For the third bullet item, \( \text{trace}(P) = \text{trace}(\sum_{i=1}^{k} u_i u_i^H) = \sum_{i=1}^{k} \text{trace}(u_i u_i^H) = \sum_{i=1}^{k} \text{trace}(u_i u_i^H) = k. \)

For the last bullet item, note that we can express the diagonal elements in the following form:

\[
P_{11} = e_1^H P e_1,
\] (13)

where \( e_1 \) is the vector \((1, 0, 0, \ldots, 0)^T.\) Moreover, we know that \( P = P^H = P^2 = P^H P. \) Therefore,

\[
P_{11} = e_1^H P e_1 = e_1^H P^H P e_1 = \|P e_1\|^2,
\] (14)

which establishes that \( P_{11} \) is real-valued and non-negative. The tricky part is now to show that it is also upper bounded by 1.

An elegant proof of this fact follows by considering the matrix \( Q = I - P. \) Clearly, if we can show that the diagonal entries of \( Q \) are non-negative, then we have established that the diagonal entries of \( P \) are upper bounded by 1 (since we know that these entries are non-negative). But the matrix \( Q \) is also a projection matrix. Specifically, we have \( Q^H = (I - P)^H = I - P^H = I - P = Q \) and \( Q^2 = (I - P)(I - P) = I - 2P + P^2 = I - 2P + P = I - P = Q. \) Hence, proceeding exactly as above,

\[
Q_{11} = e_1^H Q e_1 = e_1^H Q^H Q e_1 = \|Q e_1\|^2.
\] (15)

An alternative, less elegant proof starts by observing that \( u_1, \ldots, u_k \) are an orthonormal basis for a \( k \)-dimensional subspace. Complete this into an orthonormal basis for \( \mathbb{C}^n \) by adding \( u_{k+1}, \ldots, u_n. \) Then, we can express

\[
e_1 = \sum_{i=1}^{n} \mu_i u_i,
\] (16)

where \( \mu_i \) are appropriate coefficients satsifying \( \sum_{i=1}^{n} |\mu_i|^2 = 1 \) (since \( \|e_1\|^2 = 1 \)). Using this representation, we find

\[
P_{11} = \|P e_1\|^2 = \left\| P \sum_{i=1}^{n} \mu_i u_i \right\|^2 = \left\| \sum_{i=1}^{n} \mu_i P u_i \right\|^2 = \left\| \sum_{i=1}^{k} \mu_i P u_i \right\|^2,
\] (17)

where the last step is because the vectors \( u_{k+1}, \ldots, u_n \) are orthogonal to all the vectors \( u_1, \ldots, u_k. \) Moreover, for any vector \( x \) inside the subspace spanned by \( u_1, \ldots, u_k, \) we have \( P x = x, \) hence,

\[
P_{11} = \left\| \sum_{i=1}^{k} \mu_i u_i \right\|^2 = \sum_{i=1}^{k} |\mu_i|^2 \leq 1,
\] (18)

where the second step is because \( u_1, \ldots, u_k, \) are orthonormal and the last step is because we know that \( \sum_{i=1}^{n} |\mu_i|^2 = 1. \) This establishes the claim.
Problem 3: Eckart–Young Theorem

In class, we proved the converse part of the Eckart–Young theorem for the spectral norm. Here, you do the same for the case of the Frobenius norm.

(a) For any matrix $A$ of dimension $m \times n$ and an arbitrary orthonormal basis $\{x_1, \cdots, x_n\}$ of $\mathbb{C}^n$, prove that
$$\|A\|_F^2 = \sum_{k=1}^{n} \|Ax_k\|^2.$$ (19)

(b) Consider any $m \times n$ matrix $B$ with rank$(B) \leq p$. Clearly, its null space has dimension no smaller than $n - p$. Therefore, we can find an orthonormal set $\{x_1, \cdots, x_{n-p}\}$ in the null space of $B$. Prove that for such vectors, we have
$$\|A - B\|_F^2 \geq \sum_{k=1}^{n-p} \|Ax_k\|^2.$$ (20)

(c) (This requires slightly more subtle manipulations.) For any matrix $A$ of dimension $m \times n$ and any orthonormal set of $n - p$ vectors in $\mathbb{C}^n$, denoted by $\{x_1, \cdots, x_{n-p}\}$, prove that
$$\sum_{k=1}^{n-p} \|Ax_k\|^2 \geq \sum_{j=p+1}^{r} \sigma_j^2.$$ (21)

Hint: Consider the case $m \geq n$ and the set of vectors $\{z_1, \cdots, z_{n-p}\}$, where $z_k = V^H x_k$. Express your formulas in terms of these and the SVD representation $A = U\Sigma V^H$.

(d) Briefly explain how (a)-(c) imply the desired statement.

Solution

(a) Let us collect the vectors $\{x_1, \cdots, x_n\}$ (as columns) into an $n \times n$ matrix $X$. With this, we can express
$$\sum_{k=1}^{n} \|Ax_k\|^2 = \|AX\|_F^2.$$ (22)

Using the result that $\|A\|_F^2 = \text{trace}(A^HA)$, we find
$$\|AX\|_F^2 = \text{trace}((AX)^HAX) = \text{trace}(X^HA^HXAX) = \text{trace}(A^HAXX^H),$$ (23)

where the last step is the property that $\text{trace}(AB) = \text{trace}(BA)$. But since by construction, $X$ is a unitary matrix, we have that $XX^H$ is simply the identity matrix. Hence, $\text{trace}(A^HAXX^H) = \text{trace}(A^HA)$, which completes the proof.

(b) Let us first expand our orthonormal set $\{x_1, \cdots, x_{n-p}\}$ to a full basis for $\mathbb{C}^n$ by including orthonormal vectors $\{x_{n-p+1}, \cdots, x_n\}$. Then, from Part (a), we have
$$\|A - B\|_F^2 = \sum_{k=1}^{n} \|(A - B)x_k\|^2 \geq \sum_{k=1}^{n-p} \|(A - B)x_k\|^2,$$ (24)

where the last step is simply because all terms in the sum are non-negative. But by construction, $\{x_1, \cdots, x_{n-p}\}$ are in the null space of $B$, thus for them, $Bx_k = 0$, which implies $(A - B)x_k = Ax_k$. This completes the proof.
(c) The first point of this exercise was to recall the often surprisingly useful “trick” that \( \|y\|^2 = \text{trace}(y^H y) \), where of course the trace-operator does not do anything (yet). Applying this, we can express:

\[
\begin{align*}
\|Ax_k\|^2 &= \text{trace}(x_k^H A^H Ax_k) = \text{trace}(x_k^H V \Sigma^H U^H U \Sigma V^H x_k) \\
&= \text{trace}(z_k^H \Sigma \Sigma^H z_k^H),
\end{align*}
\]

(25)

where in the last step, we have used the property \( \text{trace}(AB) = \text{trace}(BA) \). Hence,

\[
\sum_{k=1}^{n-p} \|Ax_k\|^2 = \sum_{k=1}^{n-p} \text{trace}(\Sigma^H \Sigma z_k z_k^H)
\]

(26)

where the last step is due to the fact that \( \Sigma^H \Sigma \) is a diagonal matrix with entries \( \sigma_i^2 \), and where \( G_{ij} \) denote the entries of the matrix \( G = \sum_{k=1}^{n-p} z_k z_k^H \). The matrix \( G \) is a projection matrix. As we have seen in an earlier homework problem, its trace is \( n-p \) and its diagonal entries are non-negative and no larger than one. Under these constraints, it should be clear that the last expression is minimized if we select \( G_{11} = G_{22} = \ldots = G_{pp} = 0 \) and \( G_{p+1,p+1} = \ldots = G_{nn} = 1 \). Hence,

\[
\sum_{k=1}^{n-p} \|Ax_k\|^2 = \sum_{k=1}^{n} \sigma_k^2 G_{kk}
\]

(30)

\[
\geq \sum_{k=p+1}^{n} \sigma_k^2
\]

(31)

\[
\geq \sum_{k=p+1}^{r} \sigma_k^2,
\]

(32)

which completes the proof.

(d) Combining Parts (b) and (c):

\[
\|A - B\|_F^2 \geq \sum_{k=1}^{n-p} \|Ax_k\|^2 \geq \sum_{j=p+1}^{r} \sigma_j^2
\]

(33)

shows that for any matrix \( B \) of rank \( p \), we have the above lower bound. This is precisely the statement needed to complete the proof of the Eckart-Young theorem for the Frobenius norm.

Additional remark: Another proof of the Eckart-Young theorem (which works both for the Frobenius and the spectral norm) leverages the Weyl theorem, which states that for any two matrices \( C \) and \( D \) of the same dimension \( (m \times n) \), and assume w.l.o.g. \( m \geq n \), we have that

\[
\sigma_{i+j-1}(C + D) \leq \sigma_i(C) + \sigma_j(D), \quad \text{for } 1 \leq i, j \leq n, \text{ and } i + j - 1 \leq n.
\]

(34)

I am not aware of a simple proof of this theorem (the standard proof uses the variational characterization of eigenvalues). But suppose that \( B \) is of rank no larger than \( k \), meaning that \( \sigma_i(B) = 0 \) for \( i > k \). Then, setting \( C = A - B \) and \( D = B \), Weyl’s theorem says that

\[
\sigma_{i+k}(A) \leq \sigma_i(A - B) \quad \text{for } 1 \leq i \leq n - k,
\]

(35)

and thus,

\[
\|A - B\|_F^2 \geq \sum_{i=1}^{n-k} \sigma_i^2(A - B) \geq \sum_{i=k+1}^{n} \sigma_i^2(A).
\]

(36)
Problem 4: Inner Products

Consider the standard \( n \)-dimensional vector space \( \mathbb{R}^n \).

1. Characterize the set of matrices \( W \) for which \( y^T W x \) is a valid inner product for any \( x, y \in \mathbb{R}^n \).

2. Prove that every inner product \( \langle x, y \rangle \) on \( \mathbb{R}^n \) can be expressed as \( y^T W x \) for an appropriately chosen matrix \( W \).

3. For a subspace of dimension \( k < n \), spanned by the basis \( b_1, b_2, \ldots, b_k \in \mathbb{R}^n \), express the orthogonal projection operator (matrix) with respect to the general inner product \( \langle x, y \rangle = y^T W x \). Hint: For any vector \( x \in \mathbb{R}^n \), express its projection as \( \hat{x} = \sum_{j=1}^k \alpha_j b_j \).

Solution

1. Looking at the lecture notes, Section 7.3, an inner product must satisfy linearity properties, which clearly hold for all matrices \( W \). The symmetry property \( \langle x, y \rangle = \langle y, x \rangle \) only holds if the matrix \( W \) is symmetric, i.e., \( W^T = W \). The crucial requirement is the last property, namely, \( \langle x, x \rangle \geq 0 \), with equality if and only if \( x = 0 \). To tackle this, note that \( W \) has to be symmetric, so it has a spectral decomposition \( W = U \Lambda U^H \). Hence, it is a clever idea to express the vectors \( x \) and \( y \) in terms of the eigenvectors of \( W \).

   Then, clearly, if all eigenvalues of \( W \) are strictly positive, then the property is satisfied. Conversely, if there is a eigenvalue equal to zero, or a negative eigenvalue, then there exists a choice \( x \neq 0 \) for which \( \langle x, x \rangle = 0 \). In conclusion, \( y^T W x \) is a valid inner product if and only if \( W \) is a symmetric and positive definite.

2. To prove this, use the standard basis vectors to express \( x = x_1 e_1 + \ldots + x_n e_n \), and likewise for \( y \).

3. As we have seen in class, the error \( x - \hat{x} \) must be orthogonal to the estimate \( \hat{x} \), or, equivalently, orthogonal to all of the basis vectors \( b_i \). That is,

\[
\langle x - \hat{x}, b_i \rangle = 0.
\]  

Plugging in the hint \( \hat{x} = \sum_{j=1}^k \alpha_j b_j \), we get

\[
\langle x - \sum_{j=1}^k \alpha_j b_j, b_i \rangle = 0,
\]  

and using the standard properties of the inner product,

\[
\langle x, b_i \rangle - \sum_{j=1}^k \alpha_j \langle b_j, b_i \rangle = 0.
\]  

Defining the \( n \times k \) matrix

\[
B = (b_1, b_2, \ldots, b_k),
\]

we can collect all \( k \) conditions (for \( i = 1, 2, \ldots, k \)) into

\[
B^H W x - B^H WB \alpha = 0,
\]  

where \( \alpha \) denotes the column vector of all the coefficients \( \alpha_i \). Hence,

\[
\alpha = (B^H WB)^{-1} B^H W x,
\]
where we note that $B^H W B$ is invertible since the vectors $b_j$ constitute a basis. Finally, we observe that we can write

$$\hat{\mathbf{x}} = B\alpha = B (B^H W B)^{-1} B^H W x,$$

which is thus the desired projection matrix.