Problem 1 (10 points)
Let $u_1$ and $u_2$ be vectors such that $\|u_1\| = \|u_2\| = 1$, and $\langle u_1, u_2 \rangle = 0$. For any vector $x$, we define $P(x)$ as the vector $P(x) = \langle x, u_1 \rangle u_1 + \langle x, u_2 \rangle u_2$.

1. How would you geometrically interpret $P(x)$? (Hint: Think about projections)
2. Show that: $\|P(x)\|^2 = \langle x, u_1 \rangle^2 + \langle x, u_2 \rangle^2$.
3. Using parts (1) and (2), show that $\|P(x)\| \leq \|x\|$. When is $\|P(x)\| = \|x\|$?

Solutions
1. $P(x)$ is the projection of $x$ onto the subspace spanned by $u_1$ and $u_2$.
   Let $V$ be the subspace spanned by $u_1$ and $u_2$. $P(x)$ is the projection of $x$ onto subspace $V$ if $x - P(x)$ is orthogonal to $V$. We first show that $x - P(x) \perp u_1$ and $x - P(x) \perp u_2$.

   $\langle x - P(x), u_1 \rangle = \langle x - \langle x, u_1 \rangle u_1 - \langle x, u_2 \rangle u_2, u_1 \rangle$
   $= \langle x, u_1 \rangle - \langle \langle x, u_1 \rangle u_1, u_1 \rangle - \langle \langle x, u_2 \rangle u_2, u_1 \rangle$
   $= \langle x, u_1 \rangle - \langle x, u_1 \rangle \langle u_1, u_1 \rangle - \langle x, u_2 \rangle \langle u_2, u_1 \rangle$
   $= \langle x, u_1 \rangle - \langle x, u_1 \rangle - \langle x, u_2 \rangle \cdot 0$
   $= 0,$

   $\langle x - P(x), u_2 \rangle = \langle x - \langle x, u_1 \rangle u_1 - \langle x, u_2 \rangle u_2, u_2 \rangle$
   $= \langle x, u_2 \rangle - \langle \langle x, u_1 \rangle u_1, u_2 \rangle - \langle \langle x, u_2 \rangle u_2, u_2 \rangle$
   $= \langle x, u_2 \rangle - \langle x, u_1 \rangle \langle u_1, u_2 \rangle - \langle x, u_2 \rangle \langle u_2, u_2 \rangle$
   $= \langle x, u_2 \rangle - \langle x, u_1 \rangle \cdot 0 - \langle x, u_2 \rangle \cdot 1$
   $= 0.$

Since $x - P(x) \perp u_1$, $x - P(x) \perp u_2$ and $u_1, u_2$ are linearly independent, $x - P(x)$ is orthogonal to any vector in subspace $V$, which means that $x - P(x)$ is orthogonal to $V$. Therefore, $P(x)$ is the projection of $x$ onto the subspace spanned by $u_1$ and $u_2$.

![Figure 1: Visualization of $P(x)$, when $x, u_1, u_2 \in \mathbb{R}^3$](image)
2. We show \(\|P(x)\|^2 = \langle x, u_1 \rangle^2 + \langle x, u_2 \rangle^2\) by expanding \(\|P(x)\|^2\).

\[
\|P(x)\|^2 = \langle P(x), P(x) \rangle = \langle x, u_1 \rangle^2 + \langle x, u_2 \rangle^2 + \langle x, u_1 \rangle^2 + \langle x, u_2 \rangle^2 = \langle x, u_1 \rangle^2 + \langle x, u_2 \rangle^2 = \langle x, u_1 \rangle^2 + \langle x, u_2 \rangle^2.
\]

3. Since \(P(x) \perp x - P(x)\), we have \(\|x\|^2 = \|P(x)\|^2 + \|x - P(x)\|^2\). Or, from part (1), we have \(\langle u_1, x - P(x) \rangle = 0\) and \(\langle u_2, x - P(x) \rangle = 0\), thus

\[
\|x\|^2 = \langle x, x \rangle = \langle P(x) + (x - P(x)), P(x) + (x - P(x)) \rangle = \langle P(x), P(x) \rangle + \langle P(x), x - P(x) \rangle + \langle x - P(x), P(x) \rangle + \langle x - P(x), x - P(x) \rangle = \|P(x)\|^2 + 2\langle P(x), x - P(x) \rangle + \|x - P(x)\|^2.
\]

Therefore, \(\|P(x)\|^2 \leq \|x\|^2\). Since \(\|P(x)\| \geq 0\) and \(\|x\| \geq 0\), we have \(\|P(x)\| \leq \|x\|\).

When \(\|x - P(x)\|^2 = 0\), i.e. \(x = P(x)\) or \(x\) itself is in the subspace spanned by \(u_1\) and \(u_2\), we have \(\|P(x)\| = \|x\|\).

**Problem 2 (10 points)**

Given two column vectors \(x\) and \(y\) in \(d\)-dimensional space, the outer product of \(x\) and \(y\) is defined to be the \(d \times d\) matrix \(x \circ y = xy^\top\).

1. Show that for any \(x\) and \(y\), \(x^\top(x \circ y)y = \|x\|^2\|y\|^2\). When is this equal to \(x^\top \langle x, y \rangle y\)?

2. Show that for any non-zero \(x\) and \(y\), the outer product \(x \circ y\) always has rank 1.

3. Let \(x_1, \ldots, x_n\) be \(n \times d\) data vectors, and let \(X\) be the \(n \times d\) data matrix whose \(i\)-th row is the row vector \(x_i^\top\). Show that:

\[
X^\top X = \sum_{i=1}^n x_i \circ x_i
\]

**Solutions**

We know that for any vector \(x\), \(x^\top x = \|x\|^2\). Thus,

\[
x^\top(x \circ y)y = x^\top(xy^\top)y = (x^\top x)(y^\top y) = \|x\|^2\|y\|^2
\]

Also, \(x^\top \langle x, y \rangle y = \langle x, y \rangle(x^\top y) = \langle x, y \rangle(x, y) = \langle x, y \rangle^2 = \|x\|\|y\| \cos \theta \|^2 = \|x\|^2\|y\|^2 \cos^2 \theta\). This quantity is equal to \(\|x\|^2\|y\|^2\) when \(\theta = 0^\circ\) or \(180^\circ\). This means that the two quantities are equal when the vectors \(x\) and \(y\) are collinear.
Let $x_i$ be the $i$th element of vector $x$ and $y_i$ be the $i$th element of vector $y$. Thus,

$$x \circ y = xy^\top = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} [y_1, y_2, \ldots, y_d] = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_d \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_d \\ \vdots & \vdots & \ddots & \vdots \\ x_d y_1 & x_d y_2 & \cdots & x_d y_d \end{bmatrix}$$

Notice that every row is a scalar multiple of the first row of the above matrix. Therefore, when this matrix is reduced to a row echelon form, it will contain only one non-zero row. Therefore, the outer product $x \circ y$ always has rank 1.

Let $Y = X^\top X$. Therefore, $Y$ is a $d \times d$ matrix. Let $x_{ij}$ be the $j$th element of vector $x_i$. Therefore, $X$ can be written as

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{bmatrix}$$

Therefore,

$$Y = X^\top X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{12} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1d} & x_{2d} & \cdots & x_{nd} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{bmatrix} = \sum_{k=1}^{n} x_{ki} x_{kj}$$

where $i, j = 1, 2, \ldots, d$. Now we work out the right side of the equation.

$$\sum_{k=1}^{n} x_k \circ x_k = \sum_{k=1}^{n} x_k x_k^\top = \sum_{k=1}^{n} \begin{bmatrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{kd} \end{bmatrix} [x_{k1}, x_{k2}, \ldots, x_{kd}] = \sum_{k=1}^{n} \begin{bmatrix} x_{k1}^2 & x_{k1} x_{k2} & \cdots & x_{k1} x_{kd} \\ x_{k2} x_{k1} & x_{k2}^2 & \cdots & x_{k2} x_{kd} \\ \vdots & \vdots & \ddots & \vdots \\ x_{kd} x_{k1} & x_{kd} x_{k2} & \cdots & x_{kd}^2 \end{bmatrix}$$

Thus, the right side of the equation equals $Y$.

**Problem 3 (10 points)**

Suppose $A$ and $B$ are $d \times d$ matrices which are symmetric (in the sense that $A_{ij} = A_{ji}$ and $B_{ij} = B_{ji}$ for all $i$ and $j$) and positive semi-definite. Also suppose that $u$ is a $d \times 1$ vector such that $\|u\| = 1$. Which of the following matrices are always positive semi-definite, no matter what $A$, $B$ and $u$ are? Justify your answer.

1. $10A$
2. $A + B$
3. $uu^\top$
4. $A - B$
5. $I - uu^\top$ (Hint: Write down $x^\top (I - uu^\top) x$ in terms of some dot-products, and try using Cauchy-Schwartz.)
Solutions

A general strategy for solving this problem is to first try to prove that the matrix \( M \) is positive semi-definite; if you fail, then try to find a counter-example to disprove the claim. For the latter, you need find out a specific vector \( x \) for which \( x^\top Mx < 0 \).

By the definition of positive semi-definite matrices, for all \( d \times 1 \) vector \( x \),

\[
x^\top Ax \geq 0, x^\top Bx \geq 0
\]

1. For the matrix \( 10A \), for all \( d \times 1 \) vector \( x \),

\[
x^\top(10A)x = 10(x^\top Ax) \geq 0
\]

thus it is positive semidefinite.

2. For the matrix \( A + B \), for all \( d \times 1 \) vector \( x \),

\[
x^\top(A + B)x = (x^\top Ax) + (x^\top Bx) \geq 0,
\]

as both \( x^\top Ax \) and \( x^\top Bx \) are \( \geq 0 \). Thus it is positive semidefinite.

3. For the matrix \( uu^\top \), for all \( d \times 1 \) vector \( x \),

\[
x^\top(uu^\top)x = (x^\top u)(u^\top x) = (\langle x, u \rangle)(\langle u, x \rangle) = (\langle x, u \rangle)^2 \geq 0
\]

thus it is positive semidefinite.

4. The matrix \( A - B \) is not always positive semi-definite. As a concrete counter-example, take \( d = 2 \), \( A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \), and \( B = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \). Then \( A - B = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \). There exists a \( 2 \times 1 \) vector \( x = [1, 0]^\top \) such that

\[
x^\top(A - B)x = -1
\]

which proves that \( A - B \) is in fact not positive semi-definite.

5. For the matrix \( I - uu^\top \), for all \( d \times 1 \) vector \( x \),

\[
x^\top(I - uu^\top)x = x^\top x - (\langle x, u \rangle)^2
\]

Now applying Cauchy-Schwarz to \( (\langle x, u \rangle) \) and using the fact that \( \|u\| = 1 \), we find that

\[
(\langle x, u \rangle)^2 \leq \|x\|^2\|u\|^2 = \|x\|^2 = x^\top x
\]

Thus, we conclude

\[
x^\top(I - uu^\top)x \geq 0
\]

This establishes the fact that \( (I - uu^\top) \) is positive semi-definite.

**Problem 4 (10 points)**

In class, we discussed how to define a norm or a length for a vector. It turns out that one can also define a norm or a length for a matrix. Two popular matrix norms are the Frobenius norm and the spectral norm.

The Frobenius norm of a \( m \times n \) matrix \( A \), denoted by \( \|A\|_F \) is defined as:

\[
\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2}
\]

The spectral norm of a \( m \times n \) matrix \( A \), denoted by \( \|A\| \) is defined as:

\[
\|A\| = \max_x \frac{\|Ax\|}{\|x\|}
\]

where \( x \) is a \( n \times 1 \) vector.
1. Let \( I \) be the \( n \times n \) identity matrix. What is its Frobenius norm? What is its spectral norm? Justify your answer.

2. Suppose \( A = uv^T \) where \( u \) is a \( m \times 1 \) vector and \( v \) is a \( n \times 1 \) vector. Write down the Frobenius norm of \( A \) as function of \( \|u\| \) and \( \|v\| \). Justify your answer.

3. Write down the spectral norm of \( A \) in terms of \( \|u\| \) and \( \|v\| \). Justify your answer.

Solutions

Since \( I \) is an \( n \times n \) identity matrix, therefore it has \( n \) elements along the diagonal which are 1 and all the remaining elements are 0. Therefore, the Frobenius norm of \( I \) is given by

\[
\|I\|_F = \sqrt{n}
\]

The spectral norm of \( I \) is given by

\[
\|I\| = \max_x \frac{\|Ix\|}{\|x\|} = \max_x \frac{\|x\|}{\|x\|} = 1
\]

Let \( u = [u_1, u_2 \ldots u_m]^T \) and \( v = [v_1, v_2 \ldots v_n]^T \). Since \( A = uv^T \), therefore

\[
A = \begin{bmatrix}
  u_1v_1 & u_1v_2 & \cdots & u_1v_n \\
  u_2v_1 & u_2v_2 & \cdots & u_2v_n \\
  \vdots & \vdots & \ddots & \vdots \\
  u_mv_1 & u_mv_2 & \cdots & u_mv_n
\end{bmatrix}
\]

The Frobenius norm of \( A \) is given by

\[
\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} u_i^2 v_j^2} = \sqrt{\sum_{i=1}^{m} u_i^2 \sum_{j=1}^{n} v_j^2} = \sqrt{\|u\|^2 \|v\|^2} = \|u\| \|v\|
\]

In order to find the spectral norm of \( A \), observe that for any \( n \times 1 \) \( x \),

\[
\|Ax\| = \|u \langle v, x \rangle\| = \|u\| \|\langle v, x \rangle\| = \|u\| \|v\| \|x\| |\cos \theta|
\]

where \( \theta \) is the angle between \( v \) and \( x \).

\(|\cos \theta|\) attains a maximum value of 1 at \( \theta = 0 \) or 180. Therefore, \( \|A\| = \|u\| \|v\| \).