Math 510 **Fall 2022**
Homework 3 **Drew Armstrong**

---

1. The Important Matrices A^TA and AA^T. Let A be any $m\times n$ matrix. Consider the $n\times n$ (symmetric) matrix A^TA and the $m\times m$ (symmetric) matrix AA^T.

(a) Show that $N(A^TA)=N(A)$. [Hint: Use the trick formula $\mathbf{x}^TA^TA\mathbf{x}=\|A\mathbf{x}\|^2$.]

(b) Use part (a) to show that

$$\text{rank}(A^TA)=\text{rank}(A)=\text{rank}(AA^T).$$

[Hint: The Fundamental Theorem says that $\text{rank}(A)=\text{rank}(A^T)$.]

(c) If A has independent columns, prove that $(A^TA)^{-1}$ exists. If A has independent rows, prove that $(AA^T)^{-1}$ exists. [Hint: Use part (b).]

(a): Suppose that $\mathbf{x}\in\mathcal{N}(A)$, so that $A\mathbf{x}=\mathbf{0}$. Then we also have

$$(A^TA)\mathbf{x}=A^T(A\mathbf{x})=A^T\mathbf{x}=\mathbf{0},$$

so that $\mathbf{x}\in\mathcal{N}(A^TA)$. Conversely, suppose that $\mathbf{x}\in\mathcal{N}(A^TA)$, so that $(A^TA)\mathbf{x}$. Then we also have

$$\|A\mathbf{x}\|^2=(A\mathbf{x})^T(A\mathbf{x})=\mathbf{x}^TA^TA\mathbf{x}=\mathbf{x}^T\mathbf{0}=\mathbf{0}.$$

By properties of the norm $\|-\|$, this implies that $A\mathbf{x}=\mathbf{0}$ and hence $\mathbf{x}\in\mathcal{N}(A)$.

(b): By the Rank-Nullity Theorem applied to A^TA and A we have

$$\begin{aligned}
\text{rank}(A^TA)&=(\#\text{ columns of }A^TA)-\dim\mathcal{N}(A^TA) && \text{Rank-Nullity}\\
&=n-\dim\mathcal{N}(A^TA)\\
&=n-\dim\mathcal{N}(A) && \text{part (a)}\\
&=(\#\text{ columns of }A)+\dim\mathcal{N}(A)\\
&=\text{rank}(A). && \text{Rank-Nullity}
\end{aligned}$$

The other equality follows by taking $B=A^T$ and applying the Fundamental Theorem:

$$\begin{aligned}
\text{rank}(AA^T)&=\text{rank}(B^TB)\\
&=\text{rank}(B) && \text{previous result}\\
&=\text{rank}(A^T)\\
&=\text{rank}(A). && \text{Fundamental Theorem}
\end{aligned}$$

(c): Suppose that the columns of A are independent, so that $\text{rank}(A)$ equals the number of columns of A.[1] Recall that a square matrix is invertible if and only if its rank equals the number of columns. From part (b) we have

$$\begin{aligned}
\text{rank}(A^TA)&=\text{rank}(A)\\
&=(\#\text{ columns of }A),\\
&=(\#\text{ columns of }A^TA).
\end{aligned}$$

Since A^TA is square this implies that $(A^TA)^{-1}$ exists.

[1]By definition, the rank is the dimension of the column space.

2. Projection Matrices. A square matrix P is called a *projection matrix* when

$$P^2 = P \quad \text{and} \quad P^T = P.$$

(a) If P is a projection matrix, show that $P + (I - P) = I$ and $P(I - P) = O$.
(b) If P is a projection matrix, show that $I - P$ is also a projection matrix.
(c) Let P be a projection matrix. For any vector $\mathbf{x}$ we define $\mathbf{x}_1 = P\mathbf{x}$ and $\mathbf{x}_2 = (I - P)\mathbf{x}$. Show that $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ and $\mathbf{x}_1^T\mathbf{x}_2 = 0$. This is the geometric meaning of projection.
(d) Let A be any matrix with independent columns, so that $(A^TA)^{-1}$ exists. Check that the following matrix is a projection matrix:

$$P = A(A^TA)^{-1}A^T.$$

[Remark: This matrix projects onto the column space of A.]
(e) Use part (d) to find the 3×3 matrix that projects onto the plane $x - 2y + 1z = 0$. [Hint: It is easier to find the matrix P that projects onto the orthogonal complement, which is the line spanned by $(1, -2, 1)$. Then the projection onto the plane is $I - P$.]

(a): For any square[2] matrix P we have $P + (I - P) = I$. If $P^2 = P$ then we also have $P(I - P) = P - P^2 = P - P = O$.

(b): Suppose that $P^2 = P$ and $P^T = P$. Then we have

$$(I - P)^2 = I^2 - 2P + P^2 = I - 2P + P = I - P \quad \text{and} \quad (I - P)^T = I^T - P^T = I - P.$$

(c): Suppose that $P^2 = P$ and $P^T = P$. For any $\mathbf{x}$, let $\mathbf{x}_1 = P\mathbf{x}$ and $\mathbf{x}_2 = (I - P)\mathbf{x}$. Then we have

$$\mathbf{x}_1 + \mathbf{x}_2 = P\mathbf{x} + (I - P)\mathbf{x} = P\mathbf{x} + \mathbf{x} - P\mathbf{x} = \mathbf{x}$$

and

$$\mathbf{x}_1^T\mathbf{x}_2 = (P\mathbf{x})^T(I - P)\mathbf{x} = \mathbf{x}^TP^T(I - P)\mathbf{x} = \mathbf{x}^TP(I - P)\mathbf{x} = \mathbf{x}^TO\mathbf{x} = 0.$$

(d): Let A be any matrix with independent columns, so that $(A^TA)^{-1}$ exists, and define

$$P = A(A^TA)^{-1}A^T.$$

Then we have

$$\begin{aligned} P^2 &= [A(A^TA)^{-1}A^T][A(A^TA)^{-1}A^T] \\ &= A\cancel{(A^TA)^{-1}(A^TA)}(A^TA)^{-1}A^T \\ &= AI(A^TA)^{-1}A^T \\ &= P \end{aligned}$$

and

$$\begin{aligned} P^T &= [A(A^TA)^{-1}A^T]^T \\ &= (A^T)^T[(A^TA)^{-1}]^T(A)^T \\ &= A[(A^TA)^T]^{-1}A^T \\ &= A[A^T(A^T)^T]^{-1}A^T \\ &= A(A^TA)^{-1}A^T \\ &= P. \end{aligned}$$

[2]If P is not square then $I - P$ makes no sense.

(e): Let P be the 3×3 matrix that projects onto the plane $x - 2y + 1z = 0$ in $\mathbb{R}^3$. Then $Q = I - P$ is the matrix that projects onto the line spanned by $(1, -2, 1)$. The matrix Q is easier to compute:

$$
\begin{aligned}
Q &= \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \left(\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & 1 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & -2 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} (6)^{-1} \begin{pmatrix} 1 & -2 & 1 \end{pmatrix} \\
&= \frac{1}{6} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & 1 \end{pmatrix} \\
&= \frac{1}{6} \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
P &= I - Q \\
&= \frac{1}{6} \begin{pmatrix} 6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{pmatrix} - \frac{1}{6} \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix} \\
&= \frac{1}{6} \begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}.
\end{aligned}
$$

That was the quick method. On the other hand, we can choose any basis for the plane $x - 2y + 1z = 0$, such as $(1, 0, -1)$ and $(0, 1, 2)$, and form the matrix

$$
A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 2 \end{pmatrix}.
$$

Then the projection onto the plane is

$$
\begin{aligned}
P &= A(A^T A)^{-1} A^T \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 2 \end{pmatrix} \left[\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 2 \end{pmatrix} \right]^{-1} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 2 \end{pmatrix} \frac{1}{6} \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{pmatrix} \\
&= \frac{1}{6} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \end{pmatrix}
\end{aligned}
$$

$$= \frac{1}{6}\begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}.$$

3. CMR Factorization. Let A be any $m \times n$ matrix of rank r. Pick any r columns of A that form a basis for the column space and call them $\mathbf{c}_1, \ldots, \mathbf{c}_r \in \mathbb{R}^m$. Pick any r rows of A that form a basis for the row space and call them $\mathbf{r}_1, \ldots, \mathbf{r}_r \in \mathbb{R}^n$. Define the matrices

$$C = \begin{pmatrix} | & & | \\ \mathbf{c}_1 & \cdots & \mathbf{c}_r \\ | & & | \end{pmatrix} \quad \text{and} \quad R = \begin{pmatrix} - & \mathbf{r}_1^T & - \\ & \vdots & \\ - & \mathbf{r}_r^T & - \end{pmatrix}.$$

(a) Prove that there exists an invertible $r \times r$ matrix M such that $A = CMR$. [Hint: We know from Problem 1 that $(C^TC)^{-1}$ and $(RR^T)^{-1}$ exist. Check that $M = (CC^T)^{-1}(C^TAR^T)(RR^T)^{-1}$ works.]

(b) Compute a CMR factorization for the rank 1 matrix

$$A = \begin{pmatrix} 2 & 4 \\ 3 & 6 \end{pmatrix}.$$

(c) Compute a CMR factorization for the rank 2 matrix

$$A = \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix}.$$

[Hint: Use the first two columns and the first two rows.]

(a): **Assuming** that that there exists a matrix M satisfying $A = CMR$, it is easy to find a formula for M. Since C has independent columns and R has independent rows by definition, the matrices $(C^TC)^{-1}$ and $(RR^T)^{-1}$ exist, and hence

$$\begin{aligned} CMR &= A \\ C^T(CMR)R^T &= C^TAR^T \\ (C^TC)M(RR^T) &= C^TAR^T \\ M &= (C^TC)^{-1}C^TAR^T(RR^T)^{-1}. \end{aligned}$$

However, it is harder than I realized to prove that this formula for M actually satisfies $A = CMR$. Let M be defined by the previous formula and consider the matrix

$$A' = CMR = C(C^TC)^{-1}C^TAR^T(RR^T)^{-1}R.$$

Our goal is to prove that $A' = A$. Notice that

$$A' = PAQ,$$

where $P = C(C^TC)^{-1}C^T$ is the projection onto the column space $\mathcal{C}(A)$ and $Q = R^T(RR^T)^{-1}R^T$ is the projection onto the row space $\mathcal{R}(A)$. Any $\mathbf{x} \in \mathbb{R}^n$ can be expressed as $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ with $\mathbf{x}_1 \in \mathcal{R}(A)$ and $\mathbf{x}_2 \in \mathcal{R}(A)^\perp = \mathcal{N}(A)$, in which case we have $A\mathbf{x} = A\mathbf{x}_1 + A\mathbf{x}_2 = A\mathbf{x}_1 + \mathbf{0} = A\mathbf{x}_1$. On the other hand, since Q is the projection onto $\mathcal{R}(A)$ we have $Q\mathbf{x} = \mathbf{x}_1$. Then since $AQ\mathbf{x}_1 = A\mathbf{x}_1$ is in the column space $\mathcal{C}(A)$, the projection P onto the column space does nothing:

$$A'\mathbf{x} = PAQ\mathbf{x} = PA\mathbf{x}_1 = A\mathbf{x}_1 = A\mathbf{x}.$$

Since $A'\mathbf{x} = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$, we conclude that $A' = A$ as desired.

It is also true that the (unique) matrix M is invertible, but at the moment I don't see a slick proof of this. See the course notes for an ugly proof. It is even true that M^{-1} consists of the $r \times r$ submatrix of A that is the intersection of the columns of C with the rows of R.

(b): Here we can choose the first column and the first row, so that

$$A = \begin{pmatrix} 2 & 4 \\ 3 & 6 \end{pmatrix}, \quad C = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad R = \begin{pmatrix} 2 & 4 \end{pmatrix}.$$

Then we must have

$$\begin{aligned} M &= (C^T C)^{-1} C^T A R^T (R R^T)^{-1} \\ &= (13)^{-1} \begin{pmatrix} 2 & 3 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ 3 & 6 \end{pmatrix} \begin{pmatrix} 3 \\ 6 \end{pmatrix} (20)^{-1} \\ &= \frac{1}{260} \begin{pmatrix} 2 & 3 \end{pmatrix} \begin{pmatrix} 20 \\ 30 \end{pmatrix} \\ &= \frac{1}{260} \cdot 130 \\ &= \frac{1}{2}. \end{aligned}$$

Indeed, we observe that

$$CMR = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \frac{1}{2} \begin{pmatrix} 2 & 4 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 4 & 8 \\ 6 & 12 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 3 & 6 \end{pmatrix} = A.$$

(b): Here we can choose the first two columns and the first two rows, so that

$$A = \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 3 \\ 1 & 2 \\ 0 & 1 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \end{pmatrix}.$$

Then we must have

$$\begin{aligned} M &= (C^T C)^{-1} C^T A R^T (R R^T)^{-1} \\ &= \begin{pmatrix} 2 & 5 \\ 5 & 14 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 3 & 2 \\ 8 & 6 \end{pmatrix} \begin{pmatrix} 74 & 55 \\ 55 & 41 \end{pmatrix}^{-1} \\ &= \frac{1}{3} \begin{pmatrix} 14 & -5 \\ -5 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 3 & 2 \\ 8 & 6 \end{pmatrix} \frac{1}{9} \begin{pmatrix} 41 & -55 \\ -55 & 74 \end{pmatrix} \\ &= \frac{1}{27} \begin{pmatrix} 14 & -5 \\ -5 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 74 & 55 \\ 55 & 41 \\ 19 & 14 \end{pmatrix} \begin{pmatrix} 41 & -55 \\ -55 & 74 \end{pmatrix} \\ &= \frac{1}{27} \begin{pmatrix} 14 & -5 \\ -5 & 2 \end{pmatrix} \begin{pmatrix} 129 & 96 \\ 351 & 261 \end{pmatrix} \begin{pmatrix} 41 & -55 \\ -55 & 74 \end{pmatrix} \\ &= \frac{1}{27} \begin{pmatrix} 14 & -5 \\ -5 & 2 \end{pmatrix} \begin{pmatrix} 9 & 9 \\ 36 & 9 \end{pmatrix} \\ &= \frac{1}{27} \begin{pmatrix} -54 & 81 \\ 27 & -27 \end{pmatrix} \\ &= \begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix}. \end{aligned}$$

Indeed, we observe that

$$\begin{aligned} CMR &= \begin{pmatrix} 1 & 3 \\ 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 3 \\ 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix} \\ &= A. \end{aligned}$$

4. Distance Between Skew Lines. Consider two lines $(1,0,0) + s(1,2,1)$ and $(1,1,1) + t(1,1,1)$ living in $\mathbb{R}^3$.

(a) Suppose that the lines intersect, so that $(1,0,0) + s(1,2,1) = (1,1,1) + t(1,1,1)$ for some values of s and t. Express this as a single matrix equation:

$$A \begin{pmatrix} s \\ t \end{pmatrix} = \mathbf{b},$$

(b) If the lines **don't** intersect then the matrix equation in part (a) has no solution. In this case, find the least squares solution for (s,t).

(a): Let $\mathbf{x}_1 = (1,0,0) + s(1,2,1)$ and $\mathbf{x}_2 = (1,1,1) + t(1,1,1)$ be general points on the two lines. Assuming that the lines intersect, we have

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{x}_2 \\ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + t \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ s \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} - t \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \\ \begin{pmatrix} 1 & -1 \\ 2 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} &= \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}. \end{aligned}$$

(b): Since the lines don't intersect, the equation in part (a) has no solution. In this case we multiply both sides on the left by A^T to get

$$\begin{aligned} \begin{pmatrix} 1 & -1 \\ 2 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} &= \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \\ \begin{pmatrix} 1 & 2 & 1 \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 2 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} &= \begin{pmatrix} 1 & 2 & 1 \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \\ \begin{pmatrix} 6 & -4 \\ -4 & 3 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} &= \begin{pmatrix} 3 \\ -2 \end{pmatrix} \end{aligned}$$

$$\begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} 6 & -4 \\ -4 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$
$$\begin{pmatrix} s \\ t \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 3 & 4 \\ 4 & 6 \end{pmatrix} \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$
$$= \frac{1}{2}\begin{pmatrix} 1 \\ 0 \end{pmatrix}$$
$$= \begin{pmatrix} 1/2 \\ 0 \end{pmatrix}.$$

The least squares solution $(s,t) = (1/2, 0)$ corresponds to the points

$$\mathbf{x}_1 = (1,0,0) + \frac{1}{2}(1,2,1) = (3/2, 1, 1/2) \quad \text{and} \quad \mathbf{x}_2 = (1,1,1) + 0(1,1,1) = (1,1,1).$$

5. Bilinear Forms. Given a matrix $B \in \mathbb{R}^n$ we define a function $\langle -, - \rangle_B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by

$$\langle \mathbf{x}, \mathbf{y} \rangle_B := \mathbf{x}^T B \mathbf{y}.$$

Remark: If I is the identity matrix then $\langle \mathbf{x}, \mathbf{y} \rangle_I$ is just the dot product on $\mathbb{R}^n$.

(a) Show that this function is bilinear.

(b) Let $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathbb{R}^n$ be the standard basis. In this case show that

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle_B = (ij \text{ entry of } B).$$

(c) For any two $n \times n$ matrices B and C, show that

$$B = C \iff \langle \mathbf{x}, \mathbf{y} \rangle_B = \langle \mathbf{x}, \mathbf{y} \rangle_C \text{ for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

[Hint: One direction uses part (b).]

(d) **Symmetric.** Show that $B = B^T$ if and only if $\langle \mathbf{x}, \mathbf{y} \rangle_B = \langle \mathbf{y}, \mathbf{x} \rangle_B$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

(e) **Semi-Definite.** If $B = A^T A$ for some rectangular $A \in \mathbb{R}^{m \times n}$, show that

$$\langle \mathbf{x}, \mathbf{x} \rangle_B \geq 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n.$$

(f) **Definite.** If $B = A^T A$ for some A with **independent columns**, show that

$$\langle \mathbf{x}, \mathbf{x} \rangle_B = 0 \iff \mathbf{x} = \mathbf{0}.$$

[Hint for parts (e) and (f): Use the trick formula $\langle \mathbf{x}, \mathbf{x} \rangle_B = \|A\mathbf{x}\|^2$.]

(a): For all linear combinations $\sum a_i \mathbf{x}_i$ we have

$$\left\langle \sum a_i \mathbf{x}_i, \mathbf{y} \right\rangle_B = \left(\sum a_i \mathbf{x}_i\right)^T B\mathbf{y} = \left(\sum a_i \mathbf{x}_i^T\right) B\mathbf{y} = \sum a_i \mathbf{x}_i^T B\mathbf{y} = \sum a_i \langle \mathbf{x}_i, \mathbf{y} \rangle_B.$$

And for all linear combinations $\sum b_i \mathbf{y}_i$ we have

$$\left\langle \mathbf{x}, \sum a_i \mathbf{b}_i \right\rangle_B = \mathbf{x}^T B \left(\sum a_i \mathbf{y}_i\right) = \sum a_i \mathbf{x}^T B \mathbf{y}_i = \sum a_i \langle \mathbf{x}, \mathbf{y}_i \rangle_B.$$

(b): For any basis vectors $\mathbf{e}_i, \mathbf{e}_j \in \mathbb{R}^n$ we have

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle_B = \mathbf{e}_i^T B \mathbf{e}_j$$
$$= \begin{pmatrix} 0 & \cdots & 1 & \cdots & 0 \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nn} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} b_{1j} \\ \vdots \\ b_{nj} \end{pmatrix}$$
$$= b_{ij}.$$

(c): If $B = C$ then we have $\langle \mathbf{x}, \mathbf{y} \rangle_B = \mathbf{x}^T C \mathbf{y} = \mathbf{x}^T B \mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle_C$ for all $\mathbf{x}, \mathbf{y} \rangle_C$. Conversely, suppose that $\langle \mathbf{x}, \mathbf{y} \rangle_B = \langle \mathbf{x}, \mathbf{y} \rangle_C$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. In particular, we can take $\mathbf{x} = \mathbf{e}_i$ and $\mathbf{y} = \mathbf{e}_j$. Then part (b) gives

$$b_{ij} = \langle \mathbf{e}_i, \mathbf{e}_j \rangle_B = \langle \mathbf{e}_i, \mathbf{e}_j \rangle_C = c_{ij},$$

so that $B = C$.

(d): First suppose that $B^T = B$. Then for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ we have

$$\langle \mathbf{x}, \mathbf{y} \rangle_B = \mathbf{x}^T B \mathbf{y} = \mathbf{x}^T B^T \mathbf{y} = (B\mathbf{x})^T \mathbf{y} = \mathbf{y}^T (B\mathbf{x}) = \mathbf{y}^T B \mathbf{x} = \langle \mathbf{y}, \mathbf{x} \rangle_B.$$

Conversely, suppose that $\langle \mathbf{x}, \mathbf{y} \rangle_B = \langle \mathbf{y}, \mathbf{x} \rangle_B$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. In particular, we may choose $\mathbf{x} = \mathbf{e}_i$ and $\mathbf{y} = \mathbf{e}_j$. Then part (b) gives

$$b_{ij} = \langle \mathbf{e}_i, \mathbf{e}_j \rangle_B = \langle \mathbf{e}_j, \mathbf{e}_i \rangle_B = b_{ji},$$

so that $B^T = B$.

(e): If $B = A^T A$ then for all $\mathbf{x} \in \mathbb{R}^n$ we have

$$\langle \mathbf{x}, \mathbf{x} \rangle_B = \mathbf{x}^T B \mathbf{x} = \mathbf{x}^T A^T A \mathbf{x} = (A\mathbf{x})^T (A\mathbf{x}) = \|A\mathbf{x}\|^2 \geq 0.$$

(f): We have $\langle \mathbf{0}, \mathbf{0} \rangle_B = \mathbf{0}^T B \mathbf{0} = 0$ for any matrix B. Now suppose that $B = A^T A$ where A has independent columns. From part (e) we have $\langle \mathbf{x}, \mathbf{x} \rangle_B = \|A\mathbf{x}\|^2$. If $\langle \mathbf{x}, \mathbf{x} \rangle_B = 0$ then this implies that $\|A\mathbf{x}\| = 0$ and hence $A\mathbf{x} = \mathbf{0}$. If A has independent columns, then $A\mathbf{x} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$ because $\mathcal{N}(A) = \{\mathbf{0}\}$. Alternatively, we can use the fact that $(A^T A)^{-1}$ exists to get

$$\begin{aligned} A\mathbf{x} &= \mathbf{0} \\ A^T A \mathbf{x} &= A^T \mathbf{0} \\ \mathbf{x} &= (A^T A)^{-1} A^T \mathbf{0} \\ \mathbf{x} &= \mathbf{0}. \end{aligned}$$

6. Orthogonal Subspaces (Optional). Let V be a Hermitian inner product space. For any subspace $U \subseteq V$ we define its *orthogonal complement*:

$$U^\perp := \{\mathbf{v} \in V : \langle \mathbf{u}, \mathbf{v} \rangle = 0 \text{ for all } \mathbf{u} \in U\}.$$

(a) Prove that $U^\perp$ is also a subspace of V.
(b) Prove that $U \cap U^\perp = \{\mathbf{0}\}$.
(c) If U is finite dimensional with basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$, show that

$$\mathbf{v} \in U^\perp \iff \langle \mathbf{u}_i, \mathbf{v} \rangle = 0 \text{ for all } i.$$

(d) If U is finite dimensional, prove that $U + U^\perp = V$, which means that any vector $\mathbf{v} \in V$ can be expressed as $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ for some $\mathbf{v}_1 \in U$ and $\mathbf{v}_2 \in U^\perp$. [Hint: Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be an orthonormal basis for U and define the *projection function* $\pi : V \to V$ by

$$\pi(\mathbf{v}) = \sum_{i=1}^{m} \langle \mathbf{u}_i, \mathbf{v} \rangle \mathbf{u}_i.$$

For any $\mathbf{v} \in V$, use part (c) to show that $\mathbf{v} - \pi(\mathbf{v}) \in U^\perp$.]

(e) Combine (b) and (d) to prove that $U \oplus U^\perp = V$, which means that any vector $\mathbf{v} \in V$ can be expressed as $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ for some **unique** $\mathbf{v}_1 \in U$ and $\mathbf{v}_2 \in U^\perp$.

(f) If V is finite dimensional, prove that $\dim U + \dim U^\perp = \dim V$. [Hint: Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be an orthonormal basis for U and let $\mathbf{w}_1, \ldots, \mathbf{w}_n$ be an orthonormal basis for $U^\perp$. Use part (e) to prove that $\{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{w}_1, \ldots, \mathbf{w}_n\}$ is a basis for V.]

(a): For any $\mathbf{v}_i \in U^\perp$ and for any $\mathbf{u} \in U$ we have $\langle \mathbf{u}, \mathbf{v}_i \rangle = 0$. Then for any scalars a_i we have $\langle \mathbf{u}, \sum a_i \mathbf{v}_i \rangle = \sum a_i \langle \mathbf{u}, \mathbf{v}_i \rangle = \sum a_i \cdot 0 = 0$, and hence $\sum a_i \mathbf{v}_i \in U^\perp$.

(b): Suppose that $\mathbf{u} \in U$ and $\mathbf{u} \in U^\perp$. By definition, this means that $\langle \mathbf{u}, \mathbf{u} \rangle = 0$. Since $\langle -, - \rangle$ is an innder product, this implies that $\mathbf{u} = \mathbf{0}$.

(c): Suppose that U is finite dimensional with basis $\mathbf{u}_1, \ldots, \mathbf{u}_m \in U$. If $\mathbf{v} \in U^\perp$ then for all i we have $\langle \mathbf{u}_i, \mathbf{v} \rangle = 0$ because $\mathbf{u}_i \in U$. Conversely, suppose that $\langle \mathbf{u}_i, \mathbf{v} \rangle = 0$ for all i. In this case we will show that $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ for all $\mathbf{u} \in U$, and hence $\mathbf{v} \in U^\perp$. Indeed, any element $\mathbf{u} \in U$ can be expressed as $\mathbf{u} = \sum a_i \mathbf{u}_i$, which implies that

$$\langle \mathbf{u}, \mathbf{v} \rangle = \left\langle \sum a_i \mathbf{u}_i, \mathbf{v} \right\rangle = \sum a_i \langle \mathbf{u}_i, \mathbf{v} \rangle = \sum a_i 0 = 0.$$

(d): Let U be finite dimensional with orthonormal basis $\mathbf{u}_1, \ldots, \mathbf{u}_m \in U$.[3] Our goal is to prove that every $\mathbf{v} \in V$ can be expressed as $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ with $\mathbf{v}_1 \in U$ and $\mathbf{v}_2 \in U^\perp$. To do this, we first consider the projection function $\pi : V \to V$ defined by[4]

$$\pi(\mathbf{v}) = \sum \langle \mathbf{u}_i, \mathbf{v} \rangle \mathbf{u}_i.$$

Note that $\pi(\mathbf{v})$ is a linear combination of the basis vectors $\mathbf{u}_i \in U$, and hence $\pi(\mathbf{v}) \in U$. Furthermore, for any $\mathbf{v} \in V$ and for any basis element $\mathbf{u}_j \in U$ we have

$$\begin{aligned}
\langle \mathbf{u}_j, \mathbf{v} - \pi(\mathbf{v}) \rangle &= \langle \mathbf{u}_j, \mathbf{v} \rangle - \langle \mathbf{u}_j, \pi(\mathbf{v}) \rangle \\
&= \langle \mathbf{u}_j, \mathbf{v} \rangle - \left\langle \mathbf{u}_j, \sum_i \langle \mathbf{u}_i, \mathbf{v} \rangle \mathbf{u}_i \right\rangle \\
&= \langle \mathbf{u}_j, \mathbf{v} \rangle - \sum_i \langle \mathbf{u}_i, \mathbf{v} \rangle \langle \mathbf{u}_j, \mathbf{u}_i \rangle \\
&= \langle \mathbf{u}_j, \mathbf{v} \rangle - \sum_i \langle \mathbf{u}_i, \mathbf{v} \rangle \delta_{ij} \\
&= \langle \mathbf{u}_j, \mathbf{v} \rangle - \langle \mathbf{u}_j, \mathbf{v} \rangle \\
&= 0.
\end{aligned}$$

It follows from part (c) that $\mathbf{v} - \pi(\mathbf{v}) \in U^\perp$ for all $\mathbf{v}$. Hence for all $\mathbf{v} \in V$ we can write $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ with $\mathbf{v}_1 := \pi(\mathbf{v}) \in U$ and $\mathbf{v}_2 := \mathbf{v} - \pi(\mathbf{v}) \in U^\perp$.

(e): Consider any $\mathbf{v} \in V$. From part (d) we can write $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ for some $\mathbf{v}_1 \in U$ and $\mathbf{v}_2 \in U^\perp$. Suppose we also have $\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_2$ with $\mathbf{w}_1 \in U$ and $\mathbf{w}_2 \in U^\perp$, so that

$$\begin{aligned}
\mathbf{v}_1 + \mathbf{v}_2 &= \mathbf{w}_1 + \mathbf{w}_2 \\
\mathbf{v}_1 - \mathbf{w}_1 &= \mathbf{w}_2 - \mathbf{v}_2.
\end{aligned}$$

---

[3] Any basis can be turned into an orthonormal basis via the Gram-Schmidt process.

[4] One can show that π is a linear function with image U and kernel $U^\perp$, but we don't need these facts.

Since subspaces are closed under subtraction, we have $\mathbf{v}_1 - \mathbf{w}_1 \in U$ and $\mathbf{w}_2 - \mathbf{v}_2 \in U^\perp$. Hence the vector $\mathbf{v}_1 - \mathbf{w}_1 = \mathbf{w}_2 - \mathbf{v}_2$ is in $U \cap U^\perp$. But from part (b) we know that $U \cap U^\perp = \{\mathbf{0}\}$, so we must have $\mathbf{v}_1 - \mathbf{w}_1 = \mathbf{0}$ and $\mathbf{w}_2 - \mathbf{v}_2 = \mathbf{0}$, i.e., $\mathbf{v}_1 = \mathbf{w}_1$ and $\mathbf{v}_2 = \mathbf{w}_2$.

(f): Suppose that V is finite dimensional, so that U and $U^\perp$ are finite dimensional. Let $\mathbf{u}_1, \ldots, \mathbf{u}_m$ be a basis for U and let $\mathbf{w}_1, \ldots, \mathbf{w}_n$ be a basis for $U^\perp$, so that $\dim U = m$ and $\dim U^\perp = n$. In this case I claim that $\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{w}_1, \ldots, \mathbf{w}_n$ is a basis for V, so that

$$\dim V = m + n = \dim U + \dim U^\perp.$$

Spanning. Consider any $\mathbf{v} \in V$. From part (d) we can write $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ with $\mathbf{v}_1 \in U$ and $\mathbf{v}_2 \in U^\perp$. But then we can write $\mathbf{v}_1 = \sum a_i \mathbf{u}_i$ and $\mathbf{v}_2 = \sum b_j \mathbf{w}_j$, and hence

$$\mathbf{v} = \sum a_i \mathbf{u}_i + \sum b_j \mathbf{w}_j,$$

for some scalars $a_1, \ldots, a_m$ and $b_1, \ldots, b_n$.

Independence. Suppose that $\sum a_i \mathbf{u}_i + \sum b_j \mathbf{w}_j = \mathbf{0}$ for some scalars $a_1, \ldots, a_m$ and $b_1, \ldots, b_n$. Here we have written $\mathbf{0} = \mathbf{u} + \mathbf{w}$ with $\mathbf{u} = \sum a_i \mathbf{u}_i \in U$ and $\mathbf{w} = \sum b_j \mathbf{w}_j \in U^\perp$. On the other hand, we can write $\mathbf{0} = \mathbf{0} + \mathbf{0}$ with $\mathbf{0} \in U$ and $\mathbf{0} \in U^\perp$. Hence from part (e)[5] we must have $\mathbf{u} = \mathbf{0}$ and $\mathbf{w} = \mathbf{0}$, i.e., we must have $\sum a_i \mathbf{u}_i = \mathbf{0}$ and $\sum b_j \mathbf{w}_j = \mathbf{0}$. Then the independence of the $\mathbf{u}_i$ gives $a_i = 0$ for all i and the independence of the $\mathbf{w}_j$ gives $b_j = 0$ for all j.

[5]Alternatively, we just prove part (e) from scratch in this case. If $\mathbf{u} + \mathbf{w} = \mathbf{0}$ then $\mathbf{u} = -\mathbf{w}$. Since $\mathbf{u} \in U$ and $\mathbf{w} \in U^\perp$ this implies that $\mathbf{u}$ and $\mathbf{w}$ are in $U \cap U^\perp$, hence $\mathbf{u} = \mathbf{w} = \mathbf{0}$.