1. Almost orthonormal basis. Let \( u_1, u_2, \ldots, u_n \) form an orthonormal basis for an inner product space \( V \) and let \( v_1, v_2, \ldots, v_n \) be a set of vectors in \( V \) such that
\[
\|u_j - v_j\| < \frac{1}{\sqrt{n}}, \quad j = 1, 2, \ldots, n.
\]
Show that \( v_1, v_2, \ldots, v_n \) form a basis for \( V \).

Solution:
Since \( \dim(V) = n \), we are done if we can show that \( (v_1, v_2, \ldots, v_n) \) is independent. Assume to the contrary that they are not. Then, \( W := \text{span}(v_1, \ldots, v_n) \) has dimension < \( n \). Therefore, there exists \( w \in W^\perp \) such that \( w \neq 0 \), and without loss of generality, we can assume \( \|w\| = 1 \). We have
\[
\sum_{i=1}^{n} \| u_i - v_i \|^2 = \sum_{i=1}^{n} \| u_i - v_i \|^2 \| w \|^2 \\
\geq \sum_{i=1}^{n} \langle u_i - v_i, w \rangle^2 \\
= \sum_{i=1}^{n} (\langle u_i, w \rangle - \langle v_i, w \rangle)^2 \\
\overset{(b)}{=} \sum_{i=1}^{n} (\langle u_i, w \rangle)^2 \\
= \sum_{i=1}^{n} (u_i^T w)^2 \\
= \sum_{i=1}^{n} w^T u_i u_i^T w \\
= w^T \left( \sum_{i=1}^{n} u_i u_i^T \right) w \\
\overset{(c)}{=} w^T U U^T w \\
= w^T w \\
= 1,
\]
where (a) follows by the Cauchy–Schwarz inequality (positive-definiteness of the inner product), (b) follows since \( \langle v_i, w \rangle = 0 \) for every \( i \), and in (c), \( U \) is the orthogonal matrix \([u_1 \ \cdots \ \ u_n]\) whose columns are the \( u_i \) vectors. Since \( \|u_1 - v_1\|^2, \ldots, \|u_n - v_n\|^2 \) are non-negative, (1) implies that \( \|u_i - v_i\|^2 \geq 1/n \) for at least one \( i \), contradicting the assumption.
2. **Matrix inversion lemmas.** Let $A \in \mathbb{F}^{n \times n}$, $B \in \mathbb{F}^{n \times k}$, $C \in \mathbb{F}^{k \times n}$, and $D \in \mathbb{F}^{k \times k}$. Suppose that $A$, $D$, and $D - CA^{-1}B$ are invertible.

(a) Show that $A^{-1}B(D - CA^{-1}B)^{-1} = (A - BD^{-1}C)^{-1}BD^{-1}$.

(b) Show that

$$(A - BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}.$$ 

(c) Similarly, suppose that $A$, $D$, and $D + CA^{-1}B$ are invertible. Show that

$$(A + BD^{-1}C)^{-1} = A^{-1} - A^{-1}B(D + CA^{-1}B)^{-1}CA^{-1},$$

which is sometimes referred to as the Woodbury matrix identity.

(d) Let $\tilde{A} \in \mathbb{F}^{n \times n}$ be identical to $A$ except that the $(i, j)$-th entry differs by $\delta$, i.e., $\tilde{A} = A + \delta e_i e_j^T$. Show that

$$\tilde{A}^{-1} = A^{-1} - \frac{1}{1/\delta + (A^{-1})_{ji}} f_i g_j^T,$$

where $f_i$ is the $i$-th column of $A^{-1}$ and $g_j^T$ is its $j$-th row.

(Hint: Consider the block matrix

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

and its inverse.)

**Solution:**

(a) Consider

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ 0 & I_k \end{bmatrix} = I_{n+k}.$$ 

Hence,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{bmatrix}.$$ 

Comparing the two block matrices, we have

$$A^{-1}B(D - CA^{-1}B)^{-1} = (A - BD^{-1}C)^{-1}BD^{-1},$$

where the invertibility of $A - BD^{-1}C$ follows from part (b). More directly, we have

$$(A - BD^{-1}C)A^{-1}B = BD^{-1}(D - CA^{-1}B),$$

which implies that

$$A^{-1}B(D - CA^{-1}B)^{-1} = (A - BD^{-1}C)^{-1}BD^{-1}.$$
(b) We will show that

\[
(A - BD^{-1}C) (A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}) = (A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}) (A - BD^{-1}C) = I_n.
\]

We have

\[
(A - BD^{-1}C) (A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1})
\]
\[
= I_n + B(D - CA^{-1}B)^{-1}CA^{-1} - BD^{-1}CA^{-1} - BD^{-1}CA^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}
\]
\[
= I_n - BD^{-1}CA^{-1} + (B - BD^{-1}CA^{-1}B)(D - CA^{-1}B)^{-1}CA^{-1}
\]
\[
= I_n - BD^{-1}CA^{-1} + BD^{-1}(D - CA^{-1}B)(D - CA^{-1}B)^{-1}CA^{-1}
\]
\[
= I_n - BD^{-1}CA^{-1} + BD^{-1}CA^{-1}
\]
\[
= I_n.
\]

Also,

\[
(A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}) (A - BD^{-1}C)
\]
\[
= I_n - A^{-1}BD^{-1}C + A^{-1}B(D - CA^{-1}B)^{-1}C - A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}BD^{-1}C
\]
\[
= I_n - A^{-1}BD^{-1}C + A^{-1}B(D - CA^{-1}B)^{-1}(C - CA^{-1}BD^{-1}C)
\]
\[
= I_n - A^{-1}BD^{-1}C + A^{-1}B(D - CA^{-1}B)^{-1}(D - CA^{-1}B)D^{-1}C
\]
\[
= I_n - A^{-1}BD^{-1}C + A^{-1}BD^{-1}C
\]
\[
= I_n.
\]

Therefore, \((A - BD^{-1}C)\) is invertible and

\[
(A - BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}.
\]

(c) Replacing \(D\) by \(-D\) in (b), we have

\[
(A + BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(-D - CA^{-1}B)^{-1}CA^{-1}
\]
\[
= A^{-1} - A^{-1}B(D + CA^{-1}B)^{-1}CA^{-1}.
\]

(d) Let \(k = 1, B = e_i, C = e_j^T,\) and \(D = -1/\delta.\) Then we have \(A = A - BD^{-1}C,\) therefore, from part (c), if \(D - CA^{-1}B \equiv -(1/\delta + (A^{-1})_{ji}) \neq 0,\)

\[
\tilde{A}^{-1} = A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}
\]
\[
= A^{-1} - A^{-1}e_i (1/\delta + (A^{-1})_{ji})^{-1} e_j^T A^{-1}
\]
\[
= A^{-1} - f_i (1/\delta + (A^{-1})_{ji})^{-1} g_j^T
\]
\[
= A^{-1} - \frac{1}{1/\delta + (A^{-1})_{ji}} f_i g_j^T,
\]

where \(f_i\) and \(g_j^T\) are, respectively, the \(i^{th}\) column and \(j^{th}\) row of \(A^{-1}.\)
3. **Moore–Penrose pseudoinverse.** A pseudoinverse of \( A \in \mathbb{R}^{m \times n} \) is defined as a matrix \( A^+ \in \mathbb{R}^{n \times m} \) that satisfies

\[
AA^+ A = A,
A^+ A A^+ = A^+,
\]

and \( AA^+ \) and \( A^+ A \) are symmetric.

(a) Show that \( A^+ \) is unique.

(b) Show that \((A^T A)^{-1} A^T\) is the pseudoinverse and a left inverse of a full-rank tall matrix \(A\).

(c) Show that \(A^T (A A^T)^{-1}\) is the pseudoinverse and a right inverse of a full-rank fat matrix \(A\).

(d) Show that \(A^{-1}\) is the pseudoinverse of a full-rank square matrix \(A\).

(e) Show that \(A\) is the pseudoinverse of itself for a projection matrix \(A\) (cf. Question 4 in Homework Set #3).

(f) Show that \((A^T)^+ = (A^+)^T\).

(g) Show that \((A A^T)^+ = (A^+)^T A^+\) and \((A^T A)^+ = A^+ (A^+)^T\).

(h) Suppose that \(A\) has a rank decomposition \(A = BC\), for example, \(B = Q \in \mathbb{R}^{m \times r}\) and \(C = R \in \mathbb{R}^{r \times n}\) as in the QR decomposition. Find \(A^+\) in terms of \(B\) and \(C\).

(i) Show that \(\mathcal{R}(A^+) = \mathcal{R}(A^T)\) and \(\mathcal{N}(A^+) = \mathcal{N}(A^T)\).

(j) Show that \(P = AA^+\) and \(Q = A^+ A\) are projection matrices.

(k) Show that \(y = Px\) and \(z = Qx\) are the projections of \(x\) onto \(\mathcal{R}(A)\) and \(\mathcal{R}(A^T)\), respectively, where \(P\) and \(Q\) are defined as in \([3]\).

(l) Show that

\[
A^+ = \lim_{\delta \to 0} (A^T A + \delta I)^{-1} A^T = \lim_{\delta \to 0} A^T (A A^T + \delta I)^{-1}.
\]

(m) Show that \(x^* = A^+ b\) is a least-squares solution to the linear equation \(Ax = b\), i.e.,

\[
\|Ax^* - b\| \leq \|Ax - b\|
\]

for every other \(x\).

(n) Show that \(x^* = A^+ b\) is the least-norm solution to the linear equation \(Ax = b\), i.e.,

\[
\|x^*\| \leq \|x\|
\]

for every other solution \(x\), provided that a solution exists.

**Solution:** We have the following relations that define \(A^+\), which we summarize here for easy reference.

\[
AA^+ A = A, \quad (2)
A^+ A A^+ = A^+, \quad (3)
A^T (A^+)^T = A^+ A, \quad (4)
(A^+)^T A^T = AA^+. \quad (5)
\]

(a) If possible, let \(A\) have two pseudo-inverses \(B\) and \(C\). Then, we have

\[
AB = ACAB = C^T A^T B^T A^T = C^T (ABA)^T = C^T A^T AC.
\]

Similarly, we can show that \(BA = CA\). We therefore have

\[
B = BAB = CAB = CAC = C.
\]
(b) Since $A$ is full-rank and tall, $A^T A$ is non-singular. Defining $B := (A^T A)^{-1} A^T$, we see that $B \in \mathbb{R}^{n \times m}$, and
\[
ABA = A (A^T A)^{-1} A^T A = A, \\
BAB = (A^T A)^{-1} A^T A (A^T A)^{-1} A^T = (A^T A)^{-1} A^T = B, \\
A^T B^T = A^T A (A^T A)^{-1} = I_n = (A^T A)^{-1} A^T A = BA,
\]
and
\[
B^T A^T = A (A^T A)^{-1} A^T = A ((A^T A)^{-1} A^T) = AB.
\]
Thus, $(A^T A)^{-1} A^T$ is the pseudoinverse of $A$.

(c) Since $A$ is full-rank and fat, $AA^T$ is non-singular. Defining $B := A^T (AA^T)^{-1}$, we see that $B \in \mathbb{R}^{n \times m}$, and
\[
ABA = AA^T (AA^T)^{-1} A = A, \\
BAB = A^T (AA^T)^{-1} AA^T (AA^T)^{-1} = A^T (AA^T)^{-1} = B, \\
A^T B^T = A^T (AA^T)^{-1} A = (A^T (AA^T)^{-1}) A = BA,
\]
and
\[
B^T A^T = (AA^T)^{-1} AA^T = I_m = (AA^T) (AA^T)^{-1} = AB.
\]
Thus, $A^T (AA^T)^{-1}$ is the pseudoinverse of $A$.

(d) If $A$ is full-rank and square, then so is $A^T$, and using part (c) and the uniqueness of the pseudoinverse (proved in part (a)), we can compute the pseudoinverse of $A$ as
\[
A^+ = A^T (AA^T)^{-1} = A^T (A^T)^{-1} A^T = A^{-1}.
\]

(e) Since $A$ is a projection matrix, it is symmetric and satisfies $A^2 = A$. We then have
\[
A \cdot A \cdot A = A^2 \cdot A = A \cdot A = A^2 = A,
\]
and $A \cdot A = A^2 = A$ is symmetric, which show that $A$ is its own pseudoinverse.

(f) Let $B = (A^+)^T$. Then $B \in \mathbb{R}^{m \times n}$, and we have
\[
A^T B A^T = A^T (A^+)^T A^T = (AA^+)^T \quad \text{(6)} \quad A^T, \\
B A^T B = (A^+)^T A^T (A^+)^T = (A^+ AA^+)^T \quad \text{(7)} \quad (A^+)^T = B, \\
(A^T)^T B^T = AA^+ \quad \text{(8)} \quad (A^+)^T A^T = BA^T,
\]
and
\[
B^T (A^T)^T = A^+ A \quad \text{(9)} \quad A^T (A^+)^T = A^T B.
\]
This shows that $(A^T)^+ = (A^+)^T$. 

5
(g) Let \( B := (A^+)^T A^+ \). Then we have
\[
(AA^T)B(AA^T) = AA^T (A^+)^T A^+ AA^T + AA^+ AA^T = AA^T + AA^+ AA^T = AA^T,
\]
and
\[
B(AA^T)B = (A^+)^T A^+ AA^T (A^+)^T A^+ = (A^+)^T A^+ AA^+ (A^+)^T A^+ = B.
\]
Now, since \( B \) and \( AA^T \) are both symmetric, their products will be symmetric if and only if they commute. We have
\[
BAA^T = (A^+)^T A^+ AA^T = (A^+)^T A^T (A^+)^T A^T = (A^+ AA^+)^T A^T = (A^+)^T A^T = AA^+, \]
and
\[
AA^T B = AA^T (A^+)^T A^+ AA^+ AA^+ = AA^+. \]
Thus, \((AA^T)^+ = (A^+)^T A^+\). Now, using this relation and part (f), we have
\[
(A^T A)^+ = (A^T (A^+)^T) + = ((A^T)^+ (A^T)^+) + = ((A^+)^T (A^+)^T) = A^+(A^+)^T. \]

(h) Since \( B \) and \( C \) are respectively tall, full-rank and fat, full-rank, \( B^T B \) and \( CC^T \) are both square and non-singular.

Let us define \( D := C^T (CC^T)^{-1} (B^T B)^{-1} B^T \). Then, \( D \in \mathbb{R}^{n \times m} \), and
\[
ADA = BCC^T (CC^T)^{-1} (B^T B)^{-1} B^T BC = BC = A, \]
\[
DAD = C^T (CC^T)^{-1} (B^T B)^{-1} B^T BCC^T (CC^T)^{-1} (B^T B)^{-1} B^T = D, \]
\[
D^T A^T = B(B^T B)^{-1} (CC^T)^{-1} CC^T B^T = B(B^T B)^{-1} B^T = BCC^T (CC^T)^{-1} (B^T B)^{-1} B^T = AD, \]
and
\[
A^T D^T = C^T B^T B(B^T B)^{-1} (CC^T)^{-1} C = C^T (CC^T)^{-1} C = C^T (CC^T)^{-1} (B^T B)^{-1} B^T BC = DA. \]

Therefore, \( A^+ = C^T (CC^T)^{-1} (B^T B)^{-1} B^T = C^T (B^T A C)^{-1} B^T \).

Remark: Note that this can also be written as \( A^+ = C^+ B^+ \). In parts (g) and (h), we have seen two situations where \((BC)^+ = C^+ B^+\). This is in general not true, however.

Consider, for example,
\[
A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \]

Then, \((AB)^+ = [1/2 \ 1/2] \), but \( B^+ A^+ = [1/4 \ 1/4] \).

(i) We will show that \( \mathcal{R}(A^+) \subseteq \mathcal{R}(A^T) \) and \( \mathcal{R}(A^+) \supseteq \mathcal{R}(A^T) \). For showing the first part, let \( y \in \mathcal{R}(A^+) \), and let \( y = A^+ x \). Then, we have
\[
A^T (A^+)^T y = A^+ Ay = A^+ AA^+ x = A^+ x = y,
\]
so defining \( \tilde{x} := (A^+)^T y \), we see that \( y \) can be written as \( A^T \tilde{x} \), which shows that \( y \in \mathcal{R}(A^T) \).
For the opposite direction, if \( y \in \mathcal{R}(A^T) \) is written as \( y = A^T x \), then we can similarly show that \( y = A^+ \tilde{x} \), where \( \tilde{x} := Ay \). Thus, \( y \in \mathcal{R}(A^+) \). Therefore, we have shown that \( \mathcal{R}(A^+) = \mathcal{R}(A^T) \).

Now, let \( x \in \mathcal{N}(A^+) \). We then have

\[
A^+ x = 0 \implies AA^+ x = 0 \implies (A^+)^T A^T x = 0 \implies A^T (A^+)^T A^T x = 0 \implies (AA^+)^T x = 0 \implies A^T x = 0.
\]

Thus, \( x \in \mathcal{N}(A^T) \).

Similarly, if \( x \in \mathcal{N}(A^T) \), we have

\[
A^T x = 0 \implies (A^+)^T A^T x = 0 \implies AA^+ x = 0 \implies A^+ AA^+ x = 0 \implies A^+ x = 0.
\]

Thus, \( x \in \mathcal{N}(A^+) \).

We have therefore shown that \( \mathcal{N}(A^+) = \mathcal{N}(A^T) \).

(j) \( P \) and \( Q \) are symmetric by the properties of \( A^+ \), and

\[
P^2 = AA^+ AA^+ \implies AA^+ = P.
\]

Similarly,

\[
Q^2 = A^+ AA^+ A = A^+ A = Q.
\]

Therefore, \( P \) and \( Q \) are projection matrices.

(k) Clearly, for every \( x \in \mathbb{R}^m \), \( y = Px = AA^+ x \in \mathcal{R}(A) \). Thus, we are done if we can show that \( x - Px \in \mathcal{R}(A)^\perp = \mathcal{N}(A^T) \). We have

\[
A^T(x - Px) = A^T x - A^T AA^+ x \implies A^T x - A^T (A^+)^T A^T x = A^T x - (AA^+)^T x \implies A^T x - A^T x = 0,
\]

therefore \( P \) is indeed the projection onto \( \mathcal{R}(A) \).

Similarly, for every \( x \in \mathbb{R}^n \), \( z = Qx = A^+ Ax \in \mathcal{R}(A^+) = \mathcal{R}(A^T) \), where the last equality follows from part (i). Thus, we are done if we can show that \( x - Qx \in \mathcal{N}(A) \). We have

\[
A(x - Qx) = Ax - AA^+ Ax \implies Ax - Ax = 0,
\]

therefore \( Q \) is indeed the projection onto \( \mathcal{R}(A^T) \).

(l) We have \( A^T (AA^+ + \delta I) = (A^T A + \delta I) A^T \), therefore

\[
(A^T A + \delta I)^{-1} A^T = A^T (AA^+ + \delta I)^{-1} =: B_\delta.
\]

From problem 2, we have

\[
A^{-1} B(D + CA^{-1}B)^{-1} CA^{-1} = A^{-1} - (A + BD^{-1}C)^{-1}, \tag{6}
\]

and

\[
A^{-1} B(D + CA^{-1}B)^{-1} = (A + BD^{-1}C)^{-1} BD^{-1}. \tag{7}
\]
Now, let \( \text{rank}(A) = r \) and \( A = QR \), where \( Q \in \mathbb{R}^{m \times r} \) and has orthonormal columns (i.e. \( Q^T Q = I_r \)), and \( R \in \mathbb{R}^{r \times n} \) with \( \text{rank}(R) = r \). Then,

\[
AB_\delta = A(A^T A + \delta I)^{-1} A^T = QR(R^T R + \delta I_n)^{-1} R^T Q^T.
\]

Substituting \( I_r \) in place of \( A \), \( \delta I_n \) in place of \( D \), and \( R^T \) and \( R \) in place of \( C \) and \( B \) respectively, in (6), we have

\[
R(R^T R + \delta I_n)^{-1} R^T = I - \left( I + \frac{R R^T}{\delta} \right)^{-1} = I - \delta(RR^T + \delta I)^{-1},
\]

therefore

\[
AB_\delta = QQ^T - \delta Q(RR^T + \delta I)^{-1} Q^T. \tag{8}
\]

Similarly, we have

\[
B_\delta A = (R^T R + \delta I_n)^{-1} R^T R,
\]

and substituting \( \delta I_n \) in place of \( A \), \( I_r \) in place of \( D \), and \( R^T \) and \( R \) in place of \( B \) and \( C \) respectively, in (7), we have

\[
(\delta I_n + R^T R)^{-1} R^T = \frac{1}{\delta} R^T \left( I + \frac{R R^T}{\delta} \right)^{-1} = R^T(RR^T + \delta I)^{-1}, \tag{9}
\]

therefore

\[
B_\delta A = R^T(RR^T + \delta I)^{-1} R. \tag{10}
\]

Since \( RR^T \), being square and full-rank, is invertible, taking limits of (8) gives

\[
\lim_{\delta \to 0} AB_\delta = QQ^T, \tag{11}
\]

which is symmetric. Taking limits of (10) gives

\[
\lim_{\delta \to 0} B_\delta A = R^T(RR^T)^{-1} R, \tag{12}
\]

which is symmetric.

Right-multiplying (8) by \( A \) or left-multiplying (10) by \( A \) and taking limits gives

\[
\lim_{\delta \to 0} AB_\delta A = QR = A. \tag{13}
\]

Multiplying \( B_\delta = (A^T A + \delta I)^{-1} A^T = (R^T R + \delta I)^{-1} R^T Q^T \) to the left of (8) gives

\[
B_\delta AB_\delta = (R^T R + \delta I)^{-1} R^T Q^T - \delta(R^T R + \delta I)^{-1} R^T(RR^T + \delta I)^{-1} Q^T
\]

\[
= (R^T R + \delta I)^{-1} R^T(I - \delta(RR^T + \delta I)^{-1})Q^T. \tag{14}
\]
Therefore,
\[ B_δAB_δ − B_δ = −δ(R^TR + δI)^{-1}R^T(RR^T + δI)^{-1}Q^T. \] (15)

Using (9) in (15) gives
\[ B_δAB_δ − B_δ = −δRR^T(δI)^{-1}(RR^T + δI)^{-1}Q^T. \] (16)

Finally, since \( RR^T \) is invertible, taking limits of (16) gives
\[ \lim_{δ \to 0} (B_δAB_δ − B_δ) = 0. \] (17)

Letting \( B := \lim_{δ \to 0} B_δ \), we see from (11), (12), (13), and (17), that \( B = A^+ \).

**Alternative proof.** For a proof that does not assume a priori that the limit exists, we need the following lemma.

**Lemma 1** For a symmetric matrix \( M \in \mathbb{R}^{k \times k} \) and for any \( y \in \mathbb{R}^k \),
\[ \lim_{δ \to 0} (M + δI)^{-1}My = \text{Proj}_{\mathcal{R}(M)}(y). \]

Using the lemma with \( ATA = : M \), we have, for every \( y \in \mathbb{R}^n \),
\[ \lim_{δ \to 0} (AT^T A + δI)^{-1}AT^TAy = \text{Proj}_{\mathcal{R}(AT^TA)}(y) = \text{Proj}_{\mathcal{R}(AT^T)}(y) \text{ part (k)} = A^+ Ay. \] (18)

Now, consider an arbitrary \( x \in \mathbb{R}^m \). \( x \) can be written as \( x = \bar{x} + x^\perp \), where \( \bar{x} \in \mathcal{R}(A) \) and \( x^\perp \in \mathcal{N}(AT) \). Let \( \bar{x} = A\bar{z} \). We then have
\[ \lim_{δ \to 0} (AT^T A + δI)^{-1}AT^TAx = \lim_{δ \to 0} (AT^T A + δI)^{-1}AT^T(A\bar{z} + x^\perp) \]
\[ \stackrel{(a)}{=} \lim_{δ \to 0} (AT^T A + δI)^{-1}AT^TA\bar{z} \]
\[ \stackrel{(18)}{=} A^+ A\bar{z} \]
\[ \stackrel{(b)}{=} A^+(A\bar{z} + x^\perp) \]
\[ = A^+ x, \]
which proves the result since \( x \) is arbitrary. Here, (a) follows since \( x^\perp \in \mathcal{N}(AT^T) \) and (b) follows since from part (i), \( \mathcal{N}(AT) = \mathcal{N}(A^+) \) and \( x^\perp \in \mathcal{N}(AT) \), therefore \( A^+ x^\perp = 0 \).

(m) We have
\[ AT^T(Ax^* − b) = AT^TAA^+b − AT^Tb \overset{(a)}{=} AT^T(A^+)^TAT^Tb − AT^Tb = (AA^+)AT^Tb − AT^Tb \overset{(b)}{=} AT^Tb − AT^Tb = 0. \]

Therefore, for any \( x \in \mathbb{R}^m \),
\[ ||Ax − b||^2 = ||Ax^* − b + A(x − x^*)||^2 \]
\[ = ||Ax^* − b||^2 + ||A(x − x^*)||^2 + 2(x − x^*)^T AT^T(Ax^* − b) \]
\[ = ||Ax^* − b||^2 + ||A(x − x^*)||^2 \]
\[ ≥ ||Ax^* − b||^2, \]

demonstrating that \( x^* \) is indeed a least-squares solution to \( Ax = b \).
(n) Suppose that the linear equation \( b = Ax \) has a solution. Then, by part (m), \( x^* = A^+b \) is a solution to \( b = Ax \). Now, let \( z \) be any other solution to \( b = Ax \), i.e., we have \( Az = b \). Then,

\[
(x^*)^T (z - x^*) = b^T (A^+)^T (z - A^+b) \\
= b^T (A^+)^T z - b^T (A^+)^T A^+b \\
\equiv b^T (A^+ A A^+)^T z - b^T (A^+)^T A^+b \\
= b^T (A^+)^T A^T (A^+)^T z - b^T (A^+)^T A^+b \\
\equiv b^T (A^+)^T A^+ A z - b^T (A^+)^T A^+b \\
(a) b^T (A^+)^T A^+ b - b^T (A^+)^T A^+b \\
= 0.
\]

Here, (a) follows since \( Az = b \). We then have

\[
\|z\|^2 = \|x^* + (z - x^*)\|^2 = \|x^*\|^2 + \|z - x^*\|^2 \geq \|x^*\|^2.
\]

4. Least squares for a new inner product.

(a) Let \( A \in \mathbb{R}^{m \times n} \) be full-rank and tall. Show that \( \langle x, y \rangle_A = \langle Ax, Ay \rangle = x^T A^T Ay \) is a valid inner product.

(b) Let \( B \in \mathbb{R}^{n \times k} \) be full-rank and tall. Find the unique solution to the following least-squares problem

\[
\begin{aligned}
\text{minimize} & \quad \|y - Bx\|_A, \\
\text{where} & \quad \|v\|_A = \sqrt{\langle v, v \rangle_A}. \\
\end{aligned}
\]

Your answer should be in terms of \( y \) and \( B \).

Solution:

(a) Let \( \alpha, \beta \in \mathbb{R} \) and \( x, y, z \in \mathbb{R}^n \). Then, we have

\[
\langle y, x \rangle_A = y^T A^T Ax = (y^T A^T Ax)^T = x^T A^T Ay = \langle x, y \rangle_A. \tag{19}
\]

Also,

\[
\langle \alpha x + \beta z, y \rangle = (\alpha x + \beta z)^T A^T Ay = \alpha (x^T A^T Ay) + \beta (z^T A^T Ay) = \alpha \langle x, y \rangle_A + \beta \langle z, y \rangle_A. \tag{20}
\]

Finally,

\[
\langle x, x \rangle_A = \|Ax\|^2 \geq 0, \tag{21}
\]

and

\[
\langle x, x \rangle_A = 0 \implies \|Ax\|^2 = 0 \implies Ax = 0 \implies x = 0, \tag{22}
\]

since \( A \) is tall and full-rank (independent columns). \( \tag{19} \) through \( \tag{22} \) show that \( \langle x, y \rangle_A \) is a valid inner product.
5. Recursive least squares. The least-squares problem for \( y = Ax \) can be viewed as finding the best fit for noisy observations \( y_1, y_2, \ldots, y_m \) from linear measurements \( \tilde{a}_1^T x, \tilde{a}_2^T x, \ldots, \tilde{a}_m^T x \), where \( \tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_m \) are rows of the measurement matrix \( A \). We know that if \( A \) is full-rank and tall, \[ x(m) = (A^T A)^{-1} A^T y = \left( \sum_{i=1}^{m} \tilde{a}_i \tilde{a}_i^T \right)^{-1} \sum_{i=1}^{n} y_i \tilde{a}_i \] is the least-squares solution. Now suppose that there is an additional measurement \( \tilde{a}_{m+1}^T x \), which results in a new observation \( y_{m+1} \). The new least-squares solution can be found from scratch as \[ x(m + 1) = \left( \sum_{i=1}^{m+1} \tilde{a}_i \tilde{a}_i^T \right)^{-1} \sum_{i=1}^{n} y_i \tilde{a}_i, \] which is computational expensive. This problem explores a low-complexity alternative that can compute the new solution \( x^{(m+1)} \) based on the old one \( x^{(m)} \), and can incorporate subsequent measurement outcomes recursively.

(a) Let \( P(m) = \left( \sum_{i=1}^{m} \tilde{a}_i \tilde{a}_i^T \right)^{-1} \). Using Problem 2, show that \[ P(m + 1) = \frac{P(m) \tilde{a}_{m+1} \tilde{a}_{m+1}^T P(m)}{1 + \tilde{a}_{m+1}^T P(m) \tilde{a}_{m+1}}. \]

(b) Show that the solution \( x(m + 1) \) at the \( (m + 1) \)-st iteration can be obtained as \[ x(m + 1) = x(m) + \epsilon(m + 1) q(m + 1), \] where \[ q(m + 1) = P(m + 1) \tilde{a}_{m+1} \] \[ \epsilon(m + 1) = y_{m+1} - \tilde{a}_{m+1}^T x(m). \]
Solution:

(a) We have

\[ P(m + 1) = \left( \sum_{i=1}^{m+1} \tilde{a}_i \tilde{a}_i^T \right)^{-1} = \left( P(m)^{-1} + \tilde{a}_{m+1} \tilde{a}_{m+1}^T \right)^{-1}. \]

From part 2(c), we have

\[ (A + B D^{-1} C)^{-1} = A^{-1} - A^{-1} B (D + C A^{-1} B)^{-1} C A^{-1}. \]  \hspace{1cm} (26)

Substituting \( P(m)^{-1} \) for \( A \), \( \tilde{a}_{m+1} \) for \( B \), \( \tilde{a}_{m+1}^T \) for \( C \), and 1 for \( D \), we have

\[ (P(m)^{-1} + \tilde{a}_{m+1} \tilde{a}_{m+1}^T)^{-1} = P(m) - P(m) \tilde{a}_{m+1} (1 + \tilde{a}_{m+1}^T P(m) \tilde{a}_{m+1})^{-1} \tilde{a}_{m+1}^T P(m) \]
\[ = P(m) - \frac{P(m) \tilde{a}_{m+1} \tilde{a}_{m+1}^T P(m)}{1 + \tilde{a}_{m+1}^T P(m) \tilde{a}_{m+1}}. \]  \hspace{1cm} (27)

(b) We have

\[ x(m + 1) = P(m + 1) \left( y_{m+1} \tilde{a}_{m+1} + \sum_{i=1}^{m} y_i \tilde{a}_i \right) \]
\[ = \left( P(m) - \frac{P(m) \tilde{a}_{m+1} \tilde{a}_{m+1}^T P(m)}{1 + \tilde{a}_{m+1}^T P(m) \tilde{a}_{m+1}} \right) \left( \sum_{i=1}^{m} y_i \tilde{a}_i \right) + P(m + 1) y_{m+1} \tilde{a}_{m+1} \]
\[ = x(m) + y_{m+1} P(m + 1) \tilde{a}_{m+1} - \frac{P(m) \tilde{a}_{m+1} \tilde{a}_{m+1}^T P(m)}{1 + \tilde{a}_{m+1}^T P(m) \tilde{a}_{m+1}} \sum_{i=1}^{m} y_i \tilde{a}_i \]
\[ = x(m) + y_{m+1} P(m + 1) \tilde{a}_{m+1} - \frac{P(m) \tilde{a}_{m+1}}{1 + \tilde{a}_{m+1}^T P(m) \tilde{a}_{m+1}} \tilde{a}_{m+1}^T x(m). \]  \hspace{1cm} (28)

From (27), we have

\[ P(m + 1) \tilde{a}_{m+1} = P(m) \tilde{a}_{m+1} - \frac{P(m) \tilde{a}_{m+1} \tilde{a}_{m+1}^T P(m) \tilde{a}_{m+1}}{1 + \tilde{a}_{m+1}^T P(m) \tilde{a}_{m+1}} \]
\[ = P(m) \tilde{a}_{m+1} \left( 1 - \frac{\tilde{a}_{m+1}^T P(m) \tilde{a}_{m+1}}{1 + \tilde{a}_{m+1}^T P(m) \tilde{a}_{m+1}} \right) = \frac{P(m) \tilde{a}_{m+1}}{1 + \tilde{a}_{m+1}^T P(m) \tilde{a}_{m+1}}. \]  \hspace{1cm} (29)

Substituting (29) into (28) gives

\[ x(m + 1) = x(m) + y_{m+1} P(m + 1) \tilde{a}_{m+1} - P(m + 1) \tilde{a}_{m+1} \tilde{a}_{m+1}^T x(m) \]
\[ = x(m) + P(m + 1) \tilde{a}_{m+1} \left( y_{m+1} - \tilde{a}_{m+1}^T x(m) \right), \]

as required.