1. **Proof of (a).** Obviously $Ax = 0 \Rightarrow \langle y, Ax \rangle = 0$ for all $y$. To show sufficiency, note that if $\langle y, Ax \rangle = 0$ for all $y$, then it must certainly be true for the particular value of $y = Ax$. This yields $\|Ax\|^2 = 0 \Rightarrow Ax = 0$.

**Proof of (b).** Let $A$ be $m \times n$ (i.e., be a mapping from an $n$-dimensional space to an $m$-dimensional space) and $r(A) = \dim \mathcal{R}(A)$. Recall that $r(A) = r(A^*)$. We have

$$A \text{ is onto} \iff r(A) = r(A^*) = m \iff \dim \mathcal{N}(A) = m - r(A^*) = 0 \iff A^* \text{ is one-to-one}.$$ 

Applying this result, we have that $A^*$ is onto if and only if $A^{**} = A$ is one-to-one.

**Proof of (c).** We have,

$$y \in \mathcal{N}(A^*) \iff A^*y = 0 \iff \langle A^*y, x \rangle = 0, \ \forall x \quad \text{(from part (a))} \iff \langle y, Ax \rangle = 0, \ \forall x \iff y \in \mathcal{R}(A)^\perp.$$ 

**Proof of (d).** Let $A$ be $m \times n$. Then $A^*$ is $n \times m$, $A^*A$ is $n \times n$, and $AA^*$ is $m \times m$.

First note that it is obvious that $\mathcal{N}(A) \subset \mathcal{N}(A^*A)$ as

$$x \in \mathcal{N}(A) \Rightarrow Ax = 0 \Rightarrow A^*Ax = 0 \Rightarrow x \in \mathcal{N}(A^*A).$$

On the other hand,

$$x \in \mathcal{N}(A^*A) \iff A^*Ax = 0 \iff \langle \xi, A^*Ax \rangle = 0, \ \forall \xi \quad \text{(from part (a))} \iff \langle A\xi, Ax \rangle = 0, \ \forall \xi \Rightarrow \langle Ax, Ax \rangle = \|Ax\|^2 = 0 \quad \text{(in particular take } \xi = x) \Rightarrow Ax = 0 \Rightarrow x \in \mathcal{N}(A)$$

showing that $\mathcal{N}(A^*A) \subset \mathcal{N}(A)$. It is therefore the case that indeed $\mathcal{N}(A^*A) = \mathcal{N}(A)$ whereby ensuring that $\nu(A) = \nu(A^*A)$. Thus $A^*A$ is invertible (has rank $n$) if and only
if \( A \) is one-to-one (has rank \( n \)). (And which is true if and only if \( A^* \) is onto from part (b) above.)

If we apply the above result to the \( n \times m \) operator \( M = A^* \), we obtain \( \mathcal{N}(AA^*) = \mathcal{N}(A^*) \). Thus \( AA^* \) is invertible if and only if \( A^* \) is one-to-one which is true if and only if \( A \) is onto, from part (b) above.

**Proof of (e).** Most generally, let us assume that the space, \( X \), is a vector space (not necessarily a normed or inner product space). Consider two (complete) disjoint subspaces, \( V \cap W = \{0\} \) such that \( X = V + W \). We call two such subspaces \( V \) and \( W \) *complementary subspaces*. It is the case that every vector, \( x \), in \( X \) can be uniquely represented as \( x = v + w \) where \( v \in V \subset X \) and \( w \in W \subset X \).\(^1\) Furthermore, the ‘component’ vectors \( v \) and \( w \) are linearly independent. Note that \( v \) and \( w \) remain well defined vectors in \( X \). The projection of \( X \) onto \( V \) along a complementary space \( W \), is defined by the mapping,

\[
x = v + w \mapsto v, \quad \text{for all } x \in X.
\]

Because the decomposition of \( x \) into components along \( V \) and \( W \) is unique for every \( x \) in \( X \), this mapping is obviously unique. Regardless of what we call this mapping, say \( P \) or \( Q \), the mapping always has the same domain, codomain, and value \( v = Px = Qx \) for all \( x \) in the domain \( X \). By the definition of a function, operator, or mapping, if any two mappings \( P \) and \( Q \) have the same domain and codomain and \( Px = Qx \) for all \( x \) then it must be the case that \( P = Q \).

Let \( P \) be the projection operator defined, as above, by the mapping \( x = v + w \mapsto v \) for arbitrary \( x \in X \) with unique decomposition \( v \) and \( w \) along a pair of specified complementary subspaces \( V \) and \( W \). We have that for all vectors \( x \in X \),

\[
P^2x = P(Px) = Pv = v = Px
\]

and therefore \( P^2 = P \), showing that a projection operator \( P \) is idempotent. By induction, we have more generally that \( P^n = P \) for all integers \( n \geq 1 \). Essentially this says that \( P \) acts like the identity operator when acting on its range space (i.e., on the subspace, \( V \), that it projects onto.) Thus, the first time \( P \) is applied, it “annihilates” any component, \( w \in W \), of \( x \) that may exist along the complementary space \( W \), and thereafter acts like the identity operator on the “remnant” component \( v \in V \).

**Proof of (f).** Let \( P \) be an orthogonal projection operator on a Hilbert space \( X \).\(^2\)\(^3\)

Recall that \( P \) and \( I - P \) project onto complementary subspaces \( V = \mathcal{R}(P) \) and

---

\(^1\)Can you prove this?

\(^2\)Recall that this means that the complementary subspaces \( V \) and \( W \) associated with \( P \) are orthogonal complements with respect to the inner product on \( X \).

\(^3\)As a projection operator \( P \) must be idempotent.
\[ \mathcal{W} = \mathcal{R}(I - P). \] Because these complementary spaces are assumed to be orthogonal complements, \( \mathcal{W} = \mathcal{V}^\perp \), we have that for any pair of arbitrary vectors \( x, \xi \in \mathcal{X} \)

\[ 0 = \langle (I - P)x, P\xi \rangle = \langle P^* (I - P)x, \xi \rangle \]

and therefore\(^4\)

\[ 0 = P^*(I - P) = P^* - P^*P \Rightarrow P = P^*P \]

showing that \( P \) must be self-adjoint as well as idempotent. On the other hand, if \( P \) is idempotent and self-adjoint we have

\[ P = P^2 = PP = P^*P \]

and we can reverse the above steps to show that the projections \( (I - P)x \) and \( P\xi \) are orthogonal for all \( x \) and \( \xi \) so that the complementary subspaces associated with \( P \) and \( I - P, \mathcal{V} = \mathcal{R}(P) \) and \( \mathcal{W} = \mathcal{R}(I - P) \), are orthogonal complements.

**Proof of (g).** This is actually geometrically obvious from the fact that the nullspaces of \( A \) and \( A^* \) are both trivial. However, we will give a more mundane proof.

The matrix \( A \) is square iff \( A^* \) is square. Furthermore \( r(A^*) = r(A) \). Therefore if \( A \) is square and full rank (and hence invertible) then \( A^* \) is square and full rank (and hence invertible). Since \( A \) (being square) has full column rank and full row rank, we have

\[ A^+ = A^*(AA^*)^{-1} = (A^*A)^{-1}A^*. \]

With \( A \) and \( A^* \) both invertible, it is easily shown using either of the above expressions for \( A^+ \) that \( A^+ = A^{-1} \).

**Proof of (h).** From the definition of the adjoint operator it can readily be shown that \( A^* = \Omega^{-1}A^HW \).

(a) Suppose \( A \) is onto. Then (in the manner used to show invertibility of \( AA^* \) in problem 1a above) it can be shown that \( A\Omega^{-1}A^H \) is invertible (or just temporarily pretend it is the case that \( W = I \), so that \( A^* = \Omega^{-1}A^H \), and apply the result of problem 1a). This results in,

\[ A^+ = A^*(AA^*)^{-1} = \Omega^{-1}A^HW (A\Omega^{-1}A^HW)^{-1} = \Omega^{-1}A^HWW^{-1}(A\Omega^{-1}A^H)^{-1} = \Omega^{-1}A^H (A\Omega^{-1}A^H)^{-1}, \]

which is independent of the weighting matrix \( W \).

\(^4\)Why? On an exam I could ask you to justify this step.
(b) Suppose $A$ is one-to-one. Then it can be shown that $A^HWA$ is invertible. This yields,

\[
A^+ = (A^* A)^{-1} A^* = (\Omega^{-1} A^HWA)^{-1} \Omega^{-1} A^H W
\]

\[
= (A^HWA)^{-1} \Omega \Omega^{-1} A^H W = (A^HWA)^{-1} A^H W,
\]

which is independent of $\Omega$.

**Proof of (i).** Part (a). Since $x_p$ and $x'_p$ are both assumed to any two particular least-squares solutions to the inverse problem $y = Ax$, we have

\[
\hat{y} \triangleq P_{R(A)} y = Ax_p = Ax'_p
\]

so that

\[
A x_p = A x'_p
\]

Define $\hat{x} = P_{R(A^*)} x_p$ and $\hat{x}' = P_{R(A^*)} x'_p$ and note that

\[
\hat{x} - \hat{x}' \in \mathcal{R}(A^*)
\]

Also note that

\[
x_p - \hat{x} \in \mathcal{N}(A) = \mathcal{R}(A^*)^\perp \quad \text{and} \quad x'_p - \hat{x}' \in \mathcal{N}(A) = \mathcal{R}(A^*)^\perp
\]

or

\[
A(x_p - \hat{x}) = 0 \quad \text{and} \quad A(x'_p - \hat{x}') = 0
\]

We have

\[
A x_p = A x'_p
\]  

\[
\Leftrightarrow A (\hat{x} + (x_p - \hat{x})) = A \left(\hat{x}' + (x'_p - \hat{x}')\right)
\]

\[
\Leftrightarrow A \hat{x} = A \hat{x}'
\]

\[
\Leftrightarrow A(\hat{x} - \hat{x}') = 0
\]

\[
\Leftrightarrow \hat{x} - \hat{x}' = 0
\]

where the last step follows because $\hat{x} - \hat{x}' \in \mathcal{R}(A^*) \cap \mathcal{N}(A) = \{0\}$. Thus it must be the case that $\hat{x} = \hat{x}'$ as claimed.

Part (b). By assumption $\hat{x} = A^+ y$. Let $\hat{y} = P_{R(A)}$ be the least squares estimate of $y$. Then

\[
A \hat{x} = \hat{y}
\]

Suppose $x_p \in \{\hat{x}\} + \mathcal{N}(A)$. Then $x_p$ must be of the form $x_p = \hat{x} + x_0$ with $x_0 \in \mathcal{N}(A)$. Obviously then

\[
A x_p = A(\hat{x} + x_0) = A \hat{x} = \hat{y}
\]
showing that \( x_p \) is a particular least-squares solution. This shows that \( \{\hat{x}\} + \mathcal{N}(A) \) is a subset of the set of all least-squares solutions.

On the other hand, suppose \( x_p \) is any particular least-squares solution, \( \hat{y} = A x_p \). Then we can write

\[
x_p = \hat{x} + (x_p - \hat{x})
\]

\[
= \hat{x} + (x_p - P_{\mathcal{R}(A^*)} x_p) \quad \text{(from the } \hat{x} \text{ uniqueness result proven above)}
\]

\[
= \hat{x} + x_0 \quad \text{where} \quad x_0 = x_p - P_{\mathcal{R}(A^*)} x_p \in \mathcal{N}(A)
\]

This shows that the set of all least-squares solution must be a subset of \( \{\hat{x}\} + \mathcal{N}(A) \)

Therefore we have shown that the set of all particular least-squares solutions must be precisely the set \( \{\hat{x}\} + \mathcal{N}(A) \).

2. **Solution to Part (a).** The matrices in \( \mathcal{V} \) inherit the properties of vector addition and scalar multiplication from the parent space \( \mathcal{X} \). It is easily shown from linearity of the transpose operator that for all scalars \( \alpha, \beta \in \mathbb{C} \) and symmetric \( n \times n \) complex matrices \( A \) and \( B \),

\[
(\alpha A + \beta B)^T = \alpha A^T + \beta B^T = \alpha A + \beta B
\]

showing closure. Hence that \( \mathcal{V} \) is a subspace. Because to uniquely and unambiguously specify any symmetric complex matrix one must specify the \( n \) elements on the diagonal and the \( \frac{n^2 - n}{2} \) elements above (or, equivalently, below) the diagonal, the dimension of the space \( \mathcal{V} \) is

\[
n + \frac{n^2 - n}{2} = \frac{n(n + 1)}{2}
\]

The set of hermitian matrices is not a subspace as

\[
(\alpha A)^H = \overline{\alpha} A^H = \overline{\alpha} A \neq \alpha A
\]

for arbitrary scalar \( \alpha \) and hermitian matrix \( A \).

**Solution to Part (b).** First we show that \( P \) is a linear operator,\(^5\)

\[
P(\alpha A + \beta B) = \frac{(\alpha A + \beta B) + (\alpha A + \beta B)^T}{2} = \alpha \frac{A + A^T}{2} + \beta \frac{B + B^T}{2} = \alpha P(A) + \beta P(B) .
\]

\(^5\)An operator is a projection operator onto its range (which is necessarily a subspace of its domain) if and only if it is linear and idempotent. Therefore to prove that \( P \) is a projection operator onto \( \mathcal{V} \) we have to show that \( P \) is a) linear; b) idempotent; and c) has range equal to \( \mathcal{V} \).

Two nonlinear symmetrizing operators are

\[
P'(A) \triangleq \frac{A + A^H}{2} \quad \text{and} \quad P''(A) \triangleq \frac{A^2 + (A^2)^T}{2} .
\]

The first definition is idempotent and produces Hermitian symmetric matrices, but is nonlinear and is therefore not a projection operator. The second definition is an obviously nonlinear symmetrizing operator which is not idempotent.
for all scalars $\alpha, \beta \in \mathbb{C}$ and $A, B \in \mathbb{C}^{n \times n}$.

Now we determine the range of $P$. Obviously (by the definition of $P$)

$$P(\mathcal{X}) = \mathcal{R}(P) \subset \mathcal{V}$$

On the other hand

$$P(\mathcal{V}) = \mathcal{V}$$

since $Pv = v$ for all $v \in \mathcal{V}$. Thus, since

$$\mathcal{V} \subset \mathcal{X} \Rightarrow P(\mathcal{V}) \subset P(\mathcal{X}),$$

we have

$$\mathcal{V} = P(\mathcal{V}) \subset P(\mathcal{X}) = \mathcal{R}(P) \subset \mathcal{V}$$

showing that the operator $P$ is onto $\mathcal{V}$, $\mathcal{R}(P) = \mathcal{V}$.

Now we prove that $P$ is idempotent, showing that $P$ is a projection operator onto $\mathcal{R}(P) = \mathcal{V}$. For any $n \times n$ complex matrix $A$, we have

$$P(P(A)) = P\left(\frac{A + A^T}{2}\right) = \frac{A + A^T}{2} + \frac{A + A^T}{2} = A + A^T = P(A)$$

demonstrating that $P$ is indeed idempotent.

Finally, to prove that $P : \mathcal{X} \to \mathcal{X}$ is an orthogonal projection operator, we show that it is self-adjoint, $P = P^*$. To prove that $P$ is self-adjoint we must show that

$$\langle A, P(B) \rangle = \langle P(A), B \rangle$$

for any square complex matrices $A, B \in \mathcal{X}$. From the definition of the Frobenius inner product on the domain and range, this condition is equivalent to

**Identity:**

$$\text{tr} A^H P(B) = \text{tr} P(A)^H B. \quad (1)$$

The proof of the useful identity (1) is straightforward,

$$\text{tr} A^H P(B) = \text{tr} A^H \left(\frac{B + B^T}{2}\right)$$

$$\quad = \frac{1}{2} \left(\text{tr} A^H B + \text{tr} A^H B^T\right)$$

$$\quad = \frac{1}{2} \left(\text{tr} A^H B + \text{tr} B^T A^H\right)$$

$$\quad = \frac{1}{2} \left(\text{tr} A^H B + \text{tr} A B\right)$$

$$\quad = \text{tr} \left(\frac{A + A^T}{2}\right)^H B = \text{tr} P(A)^H B.$$

---

6Note that we cannot claim the existence of an adjoint, much less inquire into its properties, until we have first proved that $P$ is a linear operator.
Comments on the Solution to Part (b).

I. The property that $P$ is self-adjoint, $P = P^*$, is equivalent to the orthogonality condition that

$$\langle (I - P)A, PB \rangle = 0$$  \hfill (2)

for all matrices $A, B \in X$.\footnote{Cf. the proof of self-adjointness of an orthogonal projection operator given in lecture.} Therefore one can show that $P$ is self-adjoint by proving the equivalent orthogonality condition (2).

Note that it is not enough to prove the weaker condition

$$\langle (I - P)A, PA \rangle = 0,$$

which is a necessary, \emph{but not sufficient}, condition for the orthogonality condition (2) to hold.\footnote{A simple drawing should convince you of this fact.}

II. One might be tempted to treat the range space $V$ as the codomain of $P$, and then determine the adjoint of this codomain-restricted operator. However, to do this is to treat $V$ as a separate Hilbert space which is \emph{not} a subspace of $X$ (so that the resulting operator is not a projector). Remember, every projection operator has codomain equal to its domain and projects onto its range which is a subspace of the domain/codomain.

Solution to Part (c). The orthogonal projector onto the orthogonally complementary subspace $V^\perp$ is given by $I - P$. If we call $P$ a “symmetrizing operator,” then the complementary orthogonal projection operator $I - P$ can be considered an “skew-symmetrizing operator.” This is because it is readily determined that

$$(I - P)A = \frac{A - A^T}{2}$$

so that the resulting matrix $B = (I - P)A$ is \emph{skew-symmetric}

$$B^T = -B.$$ 

Thus the subspace of symmetric matrices and the subspace of skew-symmetric matrices are orthogonal complements with respect to the Frobenius inner product.

That symmetric and skew-symmetric matrices are orthogonal to each other with respect to the Frobenius inner product can be directly shown. Let $A$ be symmetric, $A^T = A$, let $B$ be skew symmetric, $B^T = -B$, and note that therefore $A^H = \overline{A}$. We have

$$\text{tr} A^H B = \text{tr} \overline{A}B = \text{tr} B\overline{A} = -\text{tr} A^H B$$

showing that

$$\langle A, B \rangle = \text{tr} A^H B = 0.$$
3. **Solution to Part (a).** It is evident that $\mathcal{A}(\cdot)$ is a linear operator. Because the matrix $A$ is onto, so that the matrix $AA^H$ is invertible, we have

\[
0 = \mathcal{A}(X) \iff 0 = X A \\
\iff 0 = X AA^H \\
\iff 0 = X
\]

showing that $\mathcal{A}(\cdot)$ is one-to-one. Thus a unique least squares solution exists and is the unique solution to the normal equations

\[
\mathcal{A}^*(\mathcal{A}(X)) = \mathcal{A}^*(Y) .
\]  

(3)

It remains to determine the adjoint $\mathcal{A}^*(\cdot)$ in order to complete the determination of the normal equations.\(^9\)

The adjoint is determined from the condition that

\[
\langle Y, \mathcal{A}(X) \rangle = \langle \mathcal{A}^*(Y), X \rangle
\]

for all $Y \in \mathcal{Y}$ and $X \in \mathcal{X}$. Note that $X$ is necessarily symmetric, so that

\[
X = P(X)
\]

where

\[
P(A) = \frac{A + A^T}{2}
\]

was considered in the previous homework problem. We have

\[
\langle Y, \mathcal{A}(X) \rangle = \text{tr} \ Y^H \mathcal{A}(X) \\
= \text{tr} \ Y^H X A \\
= \text{tr} \ A Y^H X \\
= \text{tr} \ A Y^H P(X) \\
= \text{tr} \ (YA^H)^H P(X) \\
= \text{tr} \ [P(YA^H)]^H X \quad \text{(from Identity (1) of the previous problem)} \\
= \langle \mathcal{A}^*(Y), X \rangle
\]

showing that

\[
\mathcal{A}^*(Y) = P(YA^H) .
\]  

(4)

\(^9\)Recall that the action of the adjoint must produce an output in $\mathcal{Y}$, $\mathcal{A}^*(\cdot) : \mathcal{Y} \rightarrow \mathcal{X}$. Therefore $\mathcal{A}^*(Y)$ must be a symmetric matrix. If you did not get a symmetric form for $\mathcal{A}^*(Y)$ in your solution, you should have known that you had done something wrong.
The normal equations then become
\[ \begin{align*}
A^*(A(X)) &= A^*(Y) \\
A^*(XA) &= A^*(Y) \\
P(XAA^H) &= P(YA^H) \\
X(AA^H) + (AA^H)^T X &= YA^H + (YA^H)^T 
\end{align*} \]
which is the desired solution.

**Solution to Part (b).**

(i) Term-by-term manipulation of the power series expansion
\[ e^{At} = I + At + \frac{1}{2!} A^2 t^2 + \frac{1}{3!} A^3 t^3 + \cdots \]
readily shows the validity of the claimed properties.

(ii) The theoretical solution
\[ X = \int_0^\infty e^{-Mt} \Lambda e^{-Mt} dt \]  \hspace{1cm} (5)
to the Lyapunov equation
\[ XM + M^T X = \Lambda \]
exists because the eigenvalues of \( M = AA^H \) are all real (since \( M \) is hermitian) and strictly greater than zero (because \( A \) onto implies that \( M \) is positive definite, not just positive semidefinite). Note that because \( \Lambda \) is symmetric, the theoretical solution \( X \) is symmetric.\(^{10}\)

To show that the theoretical solution satisfies the Lyapunov equation is straightforward. For \( X \) given by (5) we have
\[ \begin{align*}
XM + M^T X &= \int_0^\infty \left( e^{-Mt} \Lambda e^{-Mt} M + M^T e^{-Mt} \Lambda e^{-Mt} \right) dt \\
&= -\int_0^\infty d \left( e^{-Mt} \Lambda e^{-Mt} \right) \\
&= \Lambda - \lim_{t \to \infty} e^{-Mt} \Lambda e^{-Mt} \\
&= \Lambda 
\end{align*} \]
where the last step follows because the eigenvalues of \( M \) (and therefore \( M^T \)) are strictly positive.

\(^{10}\)The integral of a matrix function is defined component wise. Thus the integral of a transpose is the transpose of the integral.
4. For the two linear operators \( B \) and \( C \) we are given that \( \mathcal{R}(C) \cap \mathcal{N}(B) = \{0\} \). We have

\[
BC = 0 \iff (BC)(x) = 0 \ \forall x \\
\iff B(C(x)) = 0 \ \forall x \\
\iff C(x) = 0 \ \forall x \quad \text{[Because } C(x) \in \mathcal{R}(C) \cap \mathcal{N}(B) \text{]} \\
\iff C = 0.
\]

5. **Proof of (a).** The goal is to show that the candidate matrix \( M = (A^+)^* \) is the pseudoinverse of \( A^* \), which is accomplished by showing that \( M \) satisfies the four Moore-Penrose pseudoinverse conditions:

I. \((MA^*)^* = AM^* = AA^+ = (AA^*)^* = (A^+)^*A^* = MA^*\).
III. \(A^*MA^* = A^*(A^+)^*A^* = (AA^+A)^* = (A)^* = A^*\).
IV. \(MA^*M = (A^+)^*A^*(A^*)^* = (A^+AA^+)^* = (A^*)^* = M\).

Note that in the proof we exploit the fact that \( A^+ \) is assumed *a priori* to be a pseudoinverse and therefore itself satisfies the four Moore-Penrose conditions.

**Proof of (b).** Let \( M = A^*(AA^*)^+ \) and \( N = (A^*A)^+A^* \). To proved that these are both pseudoinverses will require some work.

Using the identity \((A^+)^* = (A^*)^+ \) proved in Part (a), it is easy to show that,

\[
((A^*A)^+)^* = (A^*A)^+ \quad \text{and} \quad ((AA^*)^+)^* = (AA^*)^+,
\]

two facts that we will use repeatedly below.

In addition, we will also use the extremely useful ‘Theorem A’, which was proved in the previous problem.

**Theorem A** Let the nullspace of \( B \) and the range of \( C \) be disjoint. Then \( C = 0 \iff BC = 0 \).

Finally, note that \((AA^*)^+\) and \((A^*A)^+\) used in the definition of the candidate pseudoinverses \( M \) and \( M \) are both pseudoinverses by assumption and therefore both satisfy the four Moore-Penrose pseudoinverse conditions (MPCs).

To prove that \( M \) is a pseudoinverse, we show that it satisfies the MPC’s.

\[
(AM)^* = (AA^*(AA^*)^+)^* \\
= AA^*(AA^*)^+ \quad \text{(Property of } (AA^*)^+) \\
= AM.
\]
\[(MA)^* = (A^*(AA^*)^+ A)^* = A^* ((AA^*)^+)^* A \]
\[= A^*(AA^*)^+ A = MA. \]

\[MAM = A^*(AA^*)^+ AA^*(AA^*)^+ \]
\[= A^*(AA^*)^+ \text{ (Property of } (AA^*)^+) \]
\[= M. \]

Because \((AA^*)^+\) is a pseudoinverse it satisfies,
\[AA^*(AA^*)^+ AA^* = AA^* \]
which we rewrite as,
\[A (A^*(AA^*)^+ AA^* - A^*) = 0. \]
Applying Theorem A results in,
\[A^*(AA^*)^+ AA^* = A^*, \]
which (by taking the adjoint of both sides) is readily shown to be equivalent to the final MPC,
\[AMA = A. \]

We have now proved that \(M\) is a pseudoinverse.

To prove that \(N\) is a pseudoinverse, we show that it also satisfies the MPC’s.

\[(NA)^* = ((A^* A)\^+ A^* A)^* \]
\[= (A^* A)^+ A^* A \text{ (Property of } (A^*)^+ A) \]
\[= NA. \]

\[(AN)^* = (A (A^* A)^+ A^*)^* = A ((A^* A)^+ A^* \]
\[= A (A^* A)^+ A^* = AN. \]

\[= (A^* A)^+ A^* \text{ (Property of } (A^* A)^+) \]
\[= N. \]

Because \((A^* A)^+\) is a pseudoinverse it satisfies,
\[A^* A (AA^*)^+ A^* A = A^* A \]
which we rewrite as,
\[A^* (A (A^* A)^+ A^* A - A) = 0. \]
Applying Theorem A results in,

\[ A(A^*A)^+A^*A = A, \]

which is just the final MPC,

\[ ANA = N. \]

**Proof of (c).** One can readily show that,

\[ \alpha^+ = \begin{cases} \frac{1}{\alpha} & \text{for } \alpha \neq 0 \\ 0 & \text{for } \alpha = 0 \end{cases}, \]

satisfies all four MPC's.

**Proof of (d).** One can readily show that,

\[ D^+ = \text{diag}(d_1^+, \ldots, d_n^+) \]

satisfies all four MPC's.

**Proof of (e).** One can readily show that,

\[ \Sigma^+ = \begin{pmatrix} S^+ & 0 \\ 0 & 0 \end{pmatrix}, \]

satisfies all four MPC's, where the shown block zero matrices of \( \Sigma^+ \) are identical to the block zero matrices of \( \Sigma^T \).

**Proof of (f).** Let \( M = C^*B^+A^* \) be a candidate pseudoinverse. One can readily ascertain that \( M \) satisfies all four MPC's and therefore indeed \( M = (ABC)^+ \).

**Proof of (g).** We have,

\[ A^+ = (U\Sigma V^H)^+ = V\Sigma U^H \quad \text{(From (f) above)} \]

\[ = V \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix}^+ U^H \]

\[ = V \begin{pmatrix} S^+ & 0 \\ 0 & 0 \end{pmatrix} U^H \quad \text{(From (e) above)} \]

\[ = (V_1 V_2) \begin{pmatrix} S^+ & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} U^H_1 \\ U^H_2 \end{pmatrix} = V_1 S^+ U^H_1 \]

\[ = V_1 S^{-1} U^H_1 \]

\[ = \frac{1}{\sigma_1} v_1 u_1^H + \cdots + \frac{1}{\sigma_r} v_r u_r^H. \]
6. Note that although the SVD is highly constrained, it is *not* necessarily unique. We exploit the fact that the SVD of any real matrix, $M$, can be partitioned as,

$$M = U \Sigma V^T = \left( U_1 : U_2 \right) \begin{pmatrix} S & \vdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \vdots & 0 \end{pmatrix} \begin{pmatrix} V_1^H \\ \vdots \\ V_2^H \end{pmatrix},$$

where the real matrices $U$ and $V$ are orthogonal and the $r = \text{rank}(M)$ nonzero elements of the real $r \times r$ diagonal matrix $S$ are the (unique) nonzero singular values of $M$. Obtaining any factorization of a real matrix $M$ such that the factorization has these properties is equivalent to providing an SVD of $M$. For the matrices shown, we can perform this factorization by inspection and geometric reasoning alone. We do *not* have to perform an eigenvector and eigenvalue analysis of $MM^T$ and $M^T M$, although this is often done for general, arbitrary matrices. (Remember, ours are hand crafted for ease of factorization.) Although the SVD itself may not be unique, the four orthogonal projectors onto the fundamental subspaces of $A$, the pseudoinverse of $A$, and the singular values of $A$ all are unique (why?). Having the SVD factorization at hand, we know the rank of $A$, the dimensions of the four fundamental subspaces, and orthonormal bases for each of the four subspaces. And we can readily construct the pseudoinverse and the orthogonal projectors onto each of the four fundamental subspaces. (See the first problem below for a detailed discussion. The subsequent solutions only give an SVD factorization and assume that you can perform the required matrix multiplications to obtain the remaining requested quantities.)

Matrix $A$ is obviously full rank, $r = n = 1$, and can be decomposed as,

$$A = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & \vdots & 0 \\ 0 & \vdots & 1 \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix} \quad (1).$$

We immediately see that (nonzero) $V_2$ is nonexistent, showing that the nullspace of $A$, $\mathcal{N}(A)$, is trivial ($\nu = n - r = 0$), which is consistent with the obvious fact that the matrix is full rank, $r = n = 1$. The column $U_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is an orthonormal basis for the $r = 1$ dimension range of $A$, $\mathcal{R}(A)$. The column $U_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ is a orthonormal basis for the $m - r = 2 - 1 = 1$ dimensional nullspace of $A^T$, $\mathcal{N}(A^T)$. The (scalar) column $V_1 = (1)$ is an orthonormal basis for the $r = 1$ dimensional range of $A^T$, $\mathcal{R}(A^T)$. Nonunqueness of the SVD shows up in the choice of signs used when defining the orthonormalized columns of $U$ and $V$ (the signs just have to be chosen in a consistent manner). We obtain,

$$P_{\mathcal{R}(A)} = I - P_{\mathcal{N}(A^T)} = U_1 U_1^T = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad P_{\mathcal{N}(A^T)} = I - P_{\mathcal{R}(A)} = U_2 U_2^T = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

$$P_{\mathcal{R}(A^T)} = I - P_{\mathcal{N}(A)} = V_1 V_1^T = (1), \quad P_{\mathcal{N}(A)} = I - P_{\mathcal{R}(A^T)} = (0).$$
Note that more than one way exists to construct the projection operators. This fact can be used to check for errors. Finally, the pseudoinverse is obtained as,

\[ A^+ = V^T \Sigma^+ U = V_1^T S^{-1} U_1 = (1) \cdot \frac{1}{1} \cdot (1 0) = (1 0). \]

Matrix \( B \) is obviously full rank, \( r = n = 2 \), and can be decomposed as,

\[
B = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} \\ 1 & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & \sqrt{2} \\ \cdots & \cdots \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.
\]

Nonuniqueness here shows up in the fact that the sign of the last column of \( U \) can be reversed. From this factorization, one can determine the subspace dimensions.

Matrix \( C \) is obviously rank deficient, \( r = 2 < \min(m, n) = 3 \). One possible SVD is,

\[
C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} \\ 1 & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & \sqrt{2} \\ \cdots & \cdots \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ \cdots & \cdots & \cdots \end{pmatrix}.
\]

Matrix \( D \) is obviously full rank, \( r = m = 1 \). An SVD is given by,

\[
D = (0 3 0) = (1) \begin{pmatrix} 3 & \cdots & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ \cdots & \cdots & \cdots \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.
\]

Nonuniqueness shows up here not just in the choice of sign. For instance, equally good SVD’s are,

\[
D = (-1) \begin{pmatrix} 3 & \cdots & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 & 0 \\ \cdots & \cdots & \cdots \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},
\]

and

\[
D = (1) \begin{pmatrix} 3 & \cdots & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ \cdots & \cdots & \cdots \\ \sqrt{\frac{1}{2}} & 0 & \sqrt{\frac{1}{2}} \\ -\sqrt{\frac{1}{2}} & 0 & -\sqrt{\frac{1}{2}} \end{pmatrix}.
\]

There are, in fact, an infinity of possibilities. (Can you categorize all of them?)
7. Kay 8.13. Note that Kay interchangeably uses \( x[n] \) and \( s[n] \) to mean the same sampled signal (this is likely a typo). Assume the signal model \( s[n] = A + Bn \), where \( A \) and \( B \) are real scalars, and determine a least squares fit of this model to empirical data \( \{s[n], n = 0, \ldots, N-1\} \). We have

\[
\begin{pmatrix}
  s[0] \\
  s[1] \\
  s[2] \\
  \vdots \\
  s[N-1]
\end{pmatrix}
= 
\begin{pmatrix}
  1 & 0 \\
  1 & 1 \\
  1 & 2 \\
  \vdots \\
  1 & N-1
\end{pmatrix}
\begin{pmatrix}
  A \\
  B
\end{pmatrix},
\]

which we write in vector-matrix form as,

\[
s = H\theta.
\]

Note that \( H \) is full-rank, so we can determine the least-squares estimates as,

\[
\hat{\theta} = (H^TH)^{-1}H^Ts
\]

One determines that (using some standard integer summation formulas),

\[
H^Ts = \begin{pmatrix}
\sum_{n=0}^{N-1} s[n] \\
\sum_{n=0}^{N-1} n s[n]
\end{pmatrix},
\]

\[
H^TH = \begin{pmatrix}
N & \frac{N(N-1)}{2} \\
\frac{N(N-1)}{2} & \frac{N(N-1)(2N-1)}{6}
\end{pmatrix},
\]

and (using the standard formula for inverting \( 2 \times 2 \) matrices),

\[
(H^TH)^{-1} = \begin{pmatrix}
\frac{2(2N-1)}{N(N+1)} & \frac{-6}{N(N+1)} \\
\frac{-6}{N(N+1)} & \frac{12}{N(N^2-1)}
\end{pmatrix}.
\]

This yields equation (8.24) of Moon (remember that \( s[n] \) and \( x[n] \) are the same quantity).