ECE 275A Homework # 2 Solutions

1. ECE 174 Midterm solutions are in a separate file located on the class website.

2. Let $\mathcal{V}$ and $\mathcal{W}$ be independent (disjoint) subspaces of a vector space $\mathcal{X}$, such that $\mathcal{X} = \mathcal{V} + \mathcal{W}$.[1] Consider any two vectors $x = x_1 + x_2$ and $y = y_1 + y_2$ in $\mathcal{X}$, where $x_1, y_1 \in \mathcal{V}$ and $x_2, y_2 \in \mathcal{W}$ give the unique decomposition of x and y along the companion subspaces $\mathcal{V}$ and $\mathcal{W}$ respectively. Let $P = P_{\mathcal{V}|\mathcal{W}}$ denote the projection operator which projects $\mathcal{X}$ onto $\mathcal{V}$ along $\mathcal{W}$. Then $P\,x = x_1$ and $P\,y = y_1$. For any two scalars α and β we have

$$\alpha x + \beta y = \alpha(x_1 + x_2) + \beta(y_1 + y_2) = (\alpha x_1 + \beta y_1) + (\alpha x_2 + \beta y_2) \in \mathcal{X}\,,$$

with $(\alpha x_1 + \beta y_1) \in \mathcal{V}$ and $(\alpha x_2 + \beta y_2) \in \mathcal{W}$ since $\mathcal{V}$ and $\mathcal{W}$ are vector subspaces. Therefore

$$P(\alpha x + \beta y) = \alpha x_1 + \beta y_1 = \alpha P x + \beta P y\,,$$

showing that a (possibly non-orthogonal) projection operator is linear.

3. Recall the the columns of $[V\,W]$ are a basis for $\mathcal{X}$ iff they form a spanning set of vectors for $\mathcal{X}$ which are also linearly independent. Also recall that a linear operator P is a projection operator iff it is idempotent, $P = P^2$.

   (a) The complementary subspace condition $\mathcal{X} = \mathcal{V} \oplus \mathcal{W}$ and the fact that the columns of V and and the columns of W each respectively forms a basis for $\mathcal{V}$ and $\mathcal{W}$ implies that for each $x \in \mathcal{X}$ there exists a $v = V\alpha \in \mathcal{V}$ and a $w = W\beta \in \mathcal{W}$ such that

   $$x = v + w = V\alpha + W\beta = [V\,W]\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$$

   showing that the columns of $[V\,W]$ span $\mathcal{X}$.

   The complementary subspace condition $\mathcal{X} = \mathcal{V} \oplus \mathcal{W}$ implies that $\mathcal{V}$ and $\mathcal{W}$ are disjoint, which is true iff $\mathcal{V} \cap \mathcal{W} = \{0\}$. Now suppose that the columns of V and W *taken together* are *not* linearly independent. Then there exists $\alpha \neq 0$ and $\beta \neq 0$ such that $0 = V\alpha + W\beta$. (Recall that the columns of V are a linearly independent set, as are the columns of W.) This yields $V\alpha = -W\beta \neq 0$. But $V\alpha \in \mathcal{V}$ and $-W\beta \in \mathcal{W}$, implying that

   $$0 \neq V\alpha \in \mathcal{V} \cap \mathcal{W}$$

   which contradicts the assumption that $\mathcal{V}$ and $\mathcal{W}$ are disjoint, $\mathcal{V} \cap \mathcal{W} = \{0\}$. Therefore the assumption that the columns of V and W taken together are not linearly independent must be incorrect.

[1] I.e., let $\mathcal{V}$ and $\mathcal{W}$ be *companion* subspaces of $\mathcal{X}$.

(b) Note that to show that a linear operator is a projection operator onto $\mathcal{V}$ we need to show both that it is *idempotent* (i.e., that it is indeed a *projector*) and that *its range is* $\mathcal{V}$ (i.e., that it is a projector *onto* $\mathcal{V}$). Let n be the dimension of $\mathcal{X}$. With the columns $M = [V\,W]$ linearly independent, M is an $n \times n$ invertible matrix.

Consider the matrix

$$P = M \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} M^{-1} = [V\,0]\,M^{-1}.$$

It is easily shown that $P^2 = P$, so that P is a projection operator.

Because $P = [V\,0]\,M^{-1}$, we have that $\mathcal{R}(P) = \mathcal{R}(V) = \mathcal{V}$. Thus P is a projector from $\mathcal{X}$ onto the subspace $\mathcal{V} \subset \mathcal{X}$.

We now need to show that P projects "along" $\mathcal{W}$, which is equivalent to the requirement that $\mathcal{N}(P) = \mathcal{R}(W)$, i.e. that $0 = Px$ for all $x \in \mathcal{W}$. This is readily shown:

$$x \in \mathcal{R}(W) \iff x = W\beta = [V\,W]\begin{pmatrix} 0 \\ \beta \end{pmatrix} = M\begin{pmatrix} 0 \\ \beta \end{pmatrix} \implies Px = [V\,0]\,M^{-1}M\begin{pmatrix} 0 \\ \beta \end{pmatrix} = 0.$$

4. Proof of Fact 1:

$$\begin{aligned} \langle x_1, \alpha x_2 \rangle &= \alpha \langle x_1, x_2 \rangle \qquad \text{(assuming linearity in the second argument)} \\ &= \overline{\overline{\alpha} \langle x_2, x_1 \rangle} \\ &= \overline{\langle x_2, \overline{\alpha} x_1 \rangle} \\ &= \langle \overline{\alpha} x_1, x_2 \rangle \ . \end{aligned}$$

Proof of Fact 2:

$$\begin{aligned} \langle x_1 + x_2, x \rangle &= \overline{\langle x, x_1 + x_2 \rangle} \\ &= \overline{\langle x, x_1 \rangle + \langle x, x_2 \rangle} \qquad \text{(assuming linearity in the second argument)} \\ &= \langle x_1, x \rangle + \langle x_2, x \rangle \ . \end{aligned}$$

Proof of Fact 3:

$$\begin{aligned} \langle \alpha_1 x_1 + \alpha_2 x_2, x \rangle &= \langle \alpha_1 x_1, x \rangle + \langle \alpha_2 x_2, x \rangle \qquad \text{(from Fact 2)} \\ &= \overline{\alpha_1} \langle x_1, x \rangle + \overline{\alpha_2} \langle x_2, x \rangle \ . \qquad \text{(from Fact 1)} \end{aligned}$$

Proof of Fact 4: For all vectors $x \in \mathcal{X}$, for all vectors $y_1, y_2 \in \mathcal{Y}$, and for all scalars $\alpha_1, \alpha_2 \in \mathbb{C}$,

$$\begin{aligned} \langle A^* (\alpha_1 y_1 + \alpha_2 y_2), x \rangle &= \langle \alpha_1 y_1 + \alpha_2 y_2, Ax \rangle \\ &= \overline{\alpha_1} \langle y_1, Ax \rangle + \overline{\alpha_2} \langle y_2, Ax \rangle \qquad \text{(from Fact 3)} \\ &= \overline{\alpha_1} \langle A^* y_1, x \rangle + \overline{\alpha_2} \langle A^* y_2, x \rangle \\ &= \langle \alpha_1 A^* y_1 + \alpha_2 A^* y_2, x \rangle , \qquad \text{(from Fact 3)} \end{aligned}$$

and therefore $A^*(\alpha_1 y_1 + \alpha_2 y_2) = \alpha_1 A^* y_1 + \alpha_2 A^* y_2$. (Why does this last step follow?[2])

Proof of Fact 5: For all x and y,

$$\begin{aligned}\langle (\alpha A)^* y, x\rangle &= \langle y, \alpha A x\rangle \\ &= \alpha \langle y, Ax\rangle \\ &= \alpha \langle A^* y, x\rangle \\ &= \langle \overline{\alpha} A^* y, x\rangle \,, \qquad \text{(from Fact 1)}\end{aligned}$$

and therefore $(\alpha A)^* = \overline{\alpha} A^*$. (Why?)

Proof of Fact 6: For all x and y,

$$\begin{aligned}\langle (A+B)^* y, x\rangle &= \langle y, (A+B)x\rangle \\ &= \langle y, Ax\rangle + \langle y, Bx\rangle \\ &= \langle A^* y, x\rangle + \langle B^* y, x\rangle \\ &= \langle A^* y + B^* y, x\rangle \qquad \text{(from Fact 2)} \\ &= \langle (A^* + B^*)\, y, x\rangle \qquad \text{(from definition of addition of operators)}\end{aligned}$$

and therefore $(A+B)^* = A^* + B^*$. (Why?)

Proof of Fact 7:

$$\begin{aligned}(\alpha A + \beta B)^* &= (\alpha A)^* + (\beta B)^* \qquad \text{(from Fact 5)} \\ &= \overline{\alpha} A^* + \overline{\beta} B^* \,. \qquad \text{(from Fact 4)}\end{aligned}$$

Proof of Fact 8: For all x and y,

$$\langle A^* y, x\rangle = \langle y, Ax\rangle \Leftrightarrow \overline{\langle x, A^* y\rangle} = \overline{\langle Ax, y\rangle} \Leftrightarrow \langle x, A^* y\rangle = \langle Ax, y\rangle \,.$$

Proof of Fact 9: For all x and z,

$$\langle (CA)^* z, x\rangle = \langle z, CAx\rangle = \langle C^* z, Ax\rangle = \langle A^* C^* z, x\rangle \,,$$

and therefore $(CA)^* = A^* C^*$. (Why?)

5. Moon 3.8.10. Let a linear mapping, $\mathcal{A} : \mathbb{C}^m \to \mathbb{C}^{p\times q}$, between the space of complex m-vectors and complex $p \times q$-matrices be given by,

$$\mathcal{A}\, c = \sum_{i=1}^{m} c_i X_i \,.$$

[2] On an exam, I can ask you to fill in *every* step of the proof, *including* the ones not explicitly given here.

The inner product on $\mathbb{C}^m$ is taken to be the standard inner product, $\langle x, y\rangle = x^H y$, while the inner product on $\mathbb{C}^{p\times q}$ is taken to be the Frobenius inner product,

$$\langle X, Y\rangle = \operatorname{tr} X^H Y \,.$$

Assume that X_i are linearly independent so that the mapping, $\mathcal{A}$, is one-to-one. We wish to solve the (possibly inconsistent) inverse problem,

$$Y = \mathcal{A}\, c\,,$$

The normal equations are,

$$\mathcal{A}^*\mathcal{A}\, c = \mathcal{A}^* Y\,, \tag{1}$$

which, since $\mathcal{A}$ is one-to-one, yields the least squares solution,

$$\widehat{c} = \mathcal{A}^+ Y = (\mathcal{A}^*\mathcal{A})^{-1}\mathcal{A}^* Y\,. \tag{2}$$

To obtain the normal equations (1) and the least squares solution (2) it is evident that we need to compute the so–called *Grammian operator* $\mathcal{A}^*\mathcal{A}$ and the *cross-correlation vector* $\mathcal{A}^*Y$. (These terms are defined in Section 3.1 of Moon.) The adjoint operator is determined from,

$$\begin{aligned}
\langle Y, \mathcal{A}\, c\rangle &= \left\langle Y, \sum_{i=1}^m c_i X_i \right\rangle = \sum_{i=1}^m c_i \langle Y, X_i\rangle \\
&= \sum_{i=1}^m \overline{\langle X_i, Y\rangle}\, c_i = \langle \mathcal{A}^*(Y),\, c\rangle \,.
\end{aligned}$$

This yields,

$$\mathcal{A}^*(Y) = \begin{pmatrix} \langle X_1, Y\rangle \\ \vdots \\ \langle X_m, Y\rangle \end{pmatrix} = \begin{pmatrix} \operatorname{tr} X_1^H Y \\ \vdots \\ \operatorname{tr} X_m^H Y \end{pmatrix} \in \mathbb{C}^m\,, \tag{3}$$

which is precisely the cross-correlation vector given as Equation (3.5) of Moon. (Because we define linearity in the *second* argument, our inner product arguments are reversed compared to Moon, who defines linearity to be in the first argument.) Taking $Y = \mathcal{A}\, c$ in the above enables us to determine $\mathcal{A}^*\mathcal{A}$. In particular, note that the i^{th} component of the m-vector $\mathcal{A}^*\mathcal{A}\, c$ is given by,

$$\begin{aligned}
(\mathcal{A}^*\mathcal{A}\, c)_i &= \langle X_i,\, \mathcal{A}\, c\rangle = \left\langle X_i,\, \sum_{j=1}^m c_j X_j \right\rangle \\
&= \sum_{j=1}^m c_j \langle X_i,\, X_j\rangle = \left(\langle X_i,\, X_1\rangle, \cdots, \langle X_i,\, X_m\rangle\right) c\,.
\end{aligned}$$

This shows that the full Grammian matrix is given by (cf. Equation (3.7) of Moon),

$$\mathcal{A}^*\mathcal{A} = \begin{pmatrix} \langle X_1,\, X_1\rangle & \cdots & \langle X_1,\, X_m\rangle \\ \vdots & \ddots & \vdots \\ \langle X_m,\, X_1\rangle & \cdots & \langle X_m,\, X_m\rangle \end{pmatrix} = \begin{pmatrix} \mathrm{tr}\, X_1^H X_1 & \cdots & \mathrm{tr}\, X_1^H X_m \\ \vdots & \ddots & \vdots \\ \mathrm{tr} X_m^H X_1 & \cdots & \mathrm{tr}\, X_m^H X_m \end{pmatrix} . \tag{4}$$

Equations (1)–(4) taken together yield the least-squares solution.

6. From Equation (1.2) of Moon we have (with $a_0 = 1$),

$$y[t] = -\overline{a}_1\, y[t-1] - \cdots - \overline{a}_p\, y[t-p] + \overline{b}_0\, f[t] + \cdots + \overline{b}_q\, f[t-q] = x^H\, r[t]\,, \tag{5}$$

where,

$$x = (a_1, \cdots, a_p, \cdots, b_0, \cdots, b_q)^T \in \mathbb{C}^{p+q+1}\,,$$

and

$$r[t] = (-y[t-1], \cdots, -y[t-p],\, f[t], \cdots, f[t-q])^T \in \mathbb{C}^{p+q+1}\,.$$

Note that we can rewrite (5) as

$$\overline{y}[t] = r^H[t] x\,. \tag{6}$$

Now suppose we have collected data sufficient to fill in the values of $y[t]$ and $r[t]$ for $t = 1, \cdots, m$. Then (6) enables us to fill in the m rows of the following vector-matrix equation,

$$\begin{pmatrix} \overline{y}[1] \\ \vdots \\ \overline{y}[m] \end{pmatrix} = \begin{pmatrix} r^H[1] \\ \vdots \\ r^H[m] \end{pmatrix} x\,,$$

which we can write as,

$$\eta = Ax\,, \tag{7}$$

where,

$$\eta = \begin{pmatrix} \overline{y}[1] \\ \vdots \\ \overline{y}[m] \end{pmatrix} \in \mathbb{C}^m \quad \text{and} \quad A = \begin{pmatrix} r^H[1] \\ \vdots \\ r^H[m] \end{pmatrix} \in \mathbb{C}^{m\times(p+q+1)}\,. \tag{8}$$

The system (7)–(8) can be solved in the least-squares sense to yield estimates of the unknown parameter vector, x. Note that this is a purely *data-driven* approach and (other than the putative validity of the ARMA model assumption) no statistical information about the data (such as knowledge of correlations) is assumed. The data $y[0], \cdots, y[p-1], \cdots, f[0], \cdots, f[-q]$ are known as the *initial conditions*. If they are not available, their values are often (suboptimally) set to zero. This is usually a reasonable approximation when $m \gg (p+q+1)$.

7. Proof of Moon Equation (4.34):

$$\begin{aligned}
(A+XRY)^{-1}XR &= A^{-1}X\left(R^{-1}+YA^{-1}X\right)^{-1}; \qquad \text{(Moon Eq. (4.34))}\\
XR &= (A+XRY)\,A^{-1}X\left(R^{-1}+YA^{-1}X\right)^{-1};\\
XR\left(R^{-1}+YA^{-1}X\right) &= (A+XRY)\,A^{-1}X;\\
X+XRYA^{-1}X &= X+XRYA^{-1}X;\\
X &= X\,.
\end{aligned}$$

Proof of Moon Equation (4.33):

$$\begin{aligned}
(A+XRY)^{-1} &= A^{-1}-A^{-1}X\left(R^{-1}+YA^{-1}X\right)^{-1}YA^{-1} \qquad \text{(Moon Eq. (4.33))}\\
&= A^{-1}-(A+XRY)^{-1}XRYA^{-1}; \qquad \text{(Using Moon Eq. (4.34))}\\
I &= (A+XRY)\,A^{-1}-XRYA^{-1}\\
&= I+XRYA^{-1}-XRYA^{-1}=I\,.
\end{aligned}$$

To show the validity of Moon Equation (4.32), take $R = I$, $X = x$, and $Y = y^H$ in Moon Equation (4.33).

8. Kay 8.10. Write the vector x as,

$$x = x+\widehat{s}-\widehat{s} = \widehat{s}+(x-\widehat{s}) = \widehat{s}+e\,,$$

where $\widehat{s}$ is the orthogonal projection of x onto the subspace. Note that $e = x-\widehat{s}$ is orthogonal to $\widehat{s}$, $\langle e,\widehat{s}\rangle = \langle \widehat{s},e\rangle = 0$, as a consequence of the Orthogonality Principle. Exploiting this fact results in,

$$\|x\|^2 = \|\widehat{s}+e\|^2 = \langle \widehat{s}+e,\widehat{s}+e\rangle = \langle \widehat{s},\widehat{s}\rangle + \langle e,e\rangle = \|\widehat{s}\|^2+\|e\|^2\,.$$

9. Let $\mathbb{C}^n$ be the space of n-dimensional vectors over the field of complex numbers with the standard inner product,

$$\langle x_1,x_2\rangle = x_1^H x_2\,,$$

and corresponding induced 2-norm,

$$\|x\|_2 = \sqrt{x^H x}\,.$$

Let RV^m be the Hilbert space of zero mean, finite second-order moment m-dimensional random vectors over the field of complex numbers with inner product,

$$\langle y_1,y_2\rangle = \mathrm{E}\left\{y_1^H(\omega)y_2(\omega)\right\}\,,$$

with corresponding induced norm,

$$\|y\|^2 = \mathrm{E}\left\{\|y(\omega)\|_2^2\right\}\,,$$

where $\|y(\omega)\|_2$ here denotes the standard 2-norm on $\mathbb{C}^m$. (As is standard, random vectors are identified as equivalent if they are identical almost surely.) Let

$$\mathcal{A} : \mathbb{C}^n \to \mathsf{RV}^m$$

be defined by

$$\mathcal{A}(x) = A(\omega)\, x,$$

where A is a random $m \times n$ matrix whose columns belong to, and are linearly independent in, the space RV^m,

$$A(\omega) = [a_1(\omega), \cdots, a_n(\omega)] \,, \qquad a_i \in \mathsf{RV}^m, \quad i = 1, \cdots, n\,.$$

Because the n columns of $A(\omega)$ belong to RV^m, they each have zero mean. Furthermore, since they are assumed to be linearly independent in the space RV^m, they each must be nonzero with probability one.

Given an arbitray random vector, $y \in \mathsf{RV}^m$, we wish to find the minimum mean-square error approximation to y in the range of $\mathcal{A}$. This is equivalent to determining a vector $\widehat{x}$ such that,

$$\widehat{x} \in \operatorname{argmin} \mathrm{E}\left\{\|y(\omega) - \mathcal{A}(x)\|_2^2\right\} = \operatorname{argmin} \mathrm{E}\left\{\|y(\omega) - A(\omega)x\|_2^2\right\}\,.$$

Equivalently, we desire to solve $y \approx \mathcal{A}(x)$ in the least-squares sense. The standard application of the projection theorem in the codomain of $\mathcal{A}$ yields the Normal Equations,

$$\mathcal{A}^*\mathcal{A}\,\widehat{x} = \mathcal{A}^* y\,,$$

which we write as

$$R_{AA}\,\widehat{x} = R_{Ay}\,,$$

where,

$$R_{AA} = \mathcal{A}^*\mathcal{A} \quad \text{and} \quad R_{Ay} = \mathcal{A}^* y.$$

With the assumption that the columns of $A(\omega)$ are linearly independent as random vectors in RV^m, we will see below that $R_{AA} = \mathcal{A}^*\mathcal{A}$ is invertible (one-to-one) and hence $\mathcal{A}$ is one-to-one. *Therefore the least squares solution is unique and is the unique solution to the Normal Equations.*

To determine the adjoint operator, $\mathcal{A}^*$, note that,

$$\langle y, \mathcal{A}(x)x\rangle = \mathrm{E}\left\{y^H A x\right\} = \left(\mathrm{E}\left\{A^H y\right\}\right)^H x = \left\langle \mathrm{E}\left\{A^H y\right\}, x\right\rangle\,, \tag{9}$$

and therefore,

$$\mathcal{A}^* y = R_{Ay} = \mathrm{E}\left\{A^H y\right\}\,.$$

Now note that,

$$\mathcal{A}^*(\mathcal{A}x) = \mathrm{E}\left\{A^H(Ax)\right\} = \mathrm{E}\left\{A^H A\right\} x$$

Since this is true for all x, the operator $R_{AA} = \mathcal{A}^*\mathcal{A}$ is seen to be,

$$\mathcal{A}^*\mathcal{A} = R_{AA} = \mathrm{E}\left\{A^H A\right\} = \mathrm{diag}\left(\|a_1\|^2, \cdots, \|a_m\|^2\right),$$

which is invertible since $\|a_i\| \neq 0,\ i = 1, \cdots, m$.[3] Thus we have determined that the minimum mean-square estimate of y is given by $\widehat{y} = \mathcal{A}\widehat{x}$ where $\widehat{x}$ is the unique least-squares (pseudoinverse) solution

$$\widehat{x} = R_{AA}^{-1} R_{Ay} = \left(\mathrm{E}\left\{A^H A\right\}\right)^{-1} \mathrm{E}\left\{A^H y\right\}.$$

Note that we have shown that,

$$\mathcal{A}^+ y = \left(\mathrm{E}\left\{A^H A\right\}\right)^{-1} \mathrm{E}\left\{A^H y\right\},$$

and

$$\hat{y} = P_{\mathcal{R}(\mathcal{A})} y = \mathcal{A}\widehat{x} = \mathcal{A}\mathcal{A}^+ y = A\left(\left(\mathrm{E}\left\{A^H A\right\}\right)^{-1} \mathrm{E}\left\{A^H y\right\}\right).$$

Since this is true for all $y \in \mathsf{RV}^m$, we have shown that the orthogonal projector onto the range of $\mathcal{A}$ is given by

$$P_{\mathcal{R}(\mathcal{A})}(\,\cdot\,) = A\left(\left(\mathrm{E}\left\{A^H A\right\}\right)^{-1} \mathrm{E}\left\{A^H(\,\cdot\,)\right\}\right).$$

10. Fredholm's Alternative. Condition (a) says that b is in the range of A. Condition (b) says that b is not in the range of A. Obviously these conditions are mutually exclusive and one and only one of them is true.

[3]Recall that $a_i \neq 0$ by assumption that the columns of A are linearly independent in RV^m.