Problem Set #2, Inner Product Spaces

OSM Lab: Math
Rebekah Dix

Exercise 3.1

Part (i)

$$\begin{aligned}
\frac{1}{4}(||x+y||^2 - ||x-y||^2) &= \frac{1}{4}\left(\langle x+y, x+y\rangle - \langle x-y, x-y\rangle\right) \\
&= \frac{1}{4}(\langle x,x\rangle + 2\langle x,y\rangle + \langle y,y\rangle - (\langle x,x\rangle - 2\langle x,y\rangle + \langle y,y\rangle)) \\
&= \frac{1}{4}(4\langle x,y\rangle) \\
&= \langle x,y\rangle
\end{aligned}$$

where we used that fact that $\langle x,y\rangle = \langle y,x\rangle$ because we're in a real inner product space.

Part (ii)

$$\begin{aligned}
\frac{1}{2}(||x+y||^2 - ||x-y||^2) &= \frac{1}{4}\left(\langle x+y, x+y\rangle + \langle x-y, x-y\rangle\right) \\
&= \frac{1}{2}(\langle x,x\rangle + 2\langle x,y\rangle + \langle y,y\rangle + \langle x,x\rangle - 2\langle x,y\rangle + \langle y,y\rangle) \\
&= \frac{1}{2}(2\langle x,x\rangle + 2\langle y,y\rangle) \\
&= \langle x,x\rangle + \langle y,y\rangle) \\
&= ||x||^2 + ||y||^2
\end{aligned}$$

Exercise 3.2

Using the properties of inner products outlined in Definition 3.1.1 and Proposition 3.1.3, observe that,

$$\begin{aligned}
||x-iy||^2 &= \langle x,x\rangle + i\langle y,x\rangle - i\langle x,y\rangle + \langle y,y\rangle \\
||x+iy||^2 &= \langle x,x\rangle - i\langle y,x\rangle + i\langle x,y\rangle + \langle y,y\rangle
\end{aligned}$$

Therefore,

$$\begin{aligned}
i||x-iy||^2 - i||x+iy||^2 &= i\langle x,x\rangle - \langle y,x\rangle + \langle x,y\rangle + i\langle y,y\rangle - i\langle x,x\rangle + \langle y,x\rangle - \langle x,y\rangle - i\langle y,y\rangle \\
&= 0
\end{aligned}$$

Then, using the calculation from Exercise 3.1.1, we have that,

$$\begin{aligned}
\frac{1}{4}(||x+y||^2 - ||x-y||^2 + i||x-iy||^2 - i||x+iy||^2) &= \frac{1}{4}\left(4\langle x,y\rangle + 0\right) \\
&= \langle x,y\rangle
\end{aligned}$$

Exercise 3.3

Part (i) First observe that,

$$\langle x, x^5 \rangle = \int_0^1 x^6 dx = \frac{1}{7}$$
$$\langle x, x \rangle = \int_0^1 x^2 dx = \frac{1}{3}$$
$$\langle x^5, x^5 \rangle = \int_0^1 x^{10} dx = \frac{1}{11}$$

Therefore by Equation (3.8),

$$\cos\theta = \frac{\langle x, x^5 \rangle}{||x||||x^5||} = \frac{\sqrt{33}}{7} \tag{1}$$

We find that $\theta = .608$ radians or 34.85 degrees.
Part (ii) First observe that,

$$\langle x^2, x^4 \rangle = \int_0^1 x^6 dx = \frac{1}{7}$$
$$\langle x^2, x^2 \rangle = \int_0^1 x^4 dx = \frac{1}{5}$$
$$\langle x^4, x^4 \rangle = \int_0^1 x^8 dx = \frac{1}{9}$$

Therefore by Equation (3.8),

$$\cos\theta = \frac{\langle x^2, x^4 \rangle}{||x^2||||x^4||} = \frac{\sqrt{45}}{7} \tag{2}$$

We find that $\theta = .29$ radians or 16.6 degrees.

Exercise 3.8

Part (i)

Claim 1. $S = \{\cos(t), \sin(t), \cos(2t), \sin(2t\}$ is an orthonormal set.

Proof. Let δ_{ij} be the Kronecker delta. We show that $\langle x_i, x_j \rangle = \delta_{ij}$ for all $x_i, x_j \in S$.

To this end,

$$\begin{aligned}
\frac{1}{\pi}\int_{-\pi}^{\pi}\cos(t)\sin(t)dt &= \left(-\frac{\cos^2(t)}{2\pi}\right)\Bigg|_{-\pi}^{\pi} = 0\\
\frac{1}{\pi}\int_{-\pi}^{\pi}\cos(t)\cos(2t)dt &= \left(\frac{\sin(3x)+3\sin(x)}{6\pi}\right)\Bigg|_{-\pi}^{\pi} = 0\\
\frac{1}{\pi}\int_{-\pi}^{\pi}\cos(t)\sin(2t)dt &= \left(-\frac{\cos(3x)+3\cos(x)}{6\pi}\right)\Bigg|_{-\pi}^{\pi} = 0\\
\frac{1}{\pi}\int_{-\pi}^{\pi}\sin(t)\cos(2t)dt &= \left(\frac{3\cos(x)-\cos(3x)}{6\pi}\right)\Bigg|_{-\pi}^{\pi} = 0\\
\frac{1}{\pi}\int_{-\pi}^{\pi}\sin(t)\sin(2t)dt &= \left(\frac{3\sin(x)-\sin(3x)}{6\pi}\right)\Bigg|_{-\pi}^{\pi} = 0\\
\frac{1}{\pi}\int_{-\pi}^{\pi}\cos(2t)\sin(2t)dt &= \left(-\frac{\cos(4x)}{8\pi}\right)\Bigg|_{-\pi}^{\pi} = 0
\end{aligned}$$

$$\begin{aligned}
\frac{1}{\pi}\int_{-\pi}^{\pi}\cos(t)\cos(t)dt &= \left(\frac{1}{\pi}\left(\frac{t}{2}+\frac{1}{4}\sin(2t)\right)\right)\Bigg|_{-\pi}^{\pi} = 1\\
\frac{1}{\pi}\int_{-\pi}^{\pi}\sin(t)\sin(t)dt &= \left(\frac{1}{\pi}\left(\frac{t}{2}-\frac{1}{4}\sin(2t)\right)\right)\Bigg|_{-\pi}^{\pi} = 1\\
\frac{1}{\pi}\int_{-\pi}^{\pi}\cos(2t)\cos(2t)dt &= \left(\frac{1}{\pi}\left(\frac{t}{2}+\frac{1}{8}\sin(2t)\right)\right)\Bigg|_{-\pi}^{\pi} = 1\\
\frac{1}{\pi}\int_{-\pi}^{\pi}\sin(2t)\sin(2t)dt &= \left(\frac{1}{\pi}\left(\frac{t}{2}-\frac{1}{8}\sin(2t)\right)\right)\Bigg|_{-\pi}^{\pi} = 1
\end{aligned}$$

This calculations show that $\langle x_i, x_j\rangle = \delta_{ij}$ for all $x_i, x_j \in S$; therefore, S is an orthonormal set. $\square$

Part (ii)

$$\langle t,t\rangle = \frac{1}{\pi}\int_{\pi}^{-\pi} t^2 dt = \frac{1}{\pi}\left(\frac{t^3}{3}\right)\Bigg|_{-\pi}^{\pi} = \frac{2\pi^2}{3}$$

Therefore, $||t|| = \pi\sqrt{\frac{2}{3}}$.

Part (iii) In Part (i) we proved that S is an orthonormal set. Let $X = \text{span}(S)$. Therefore, by Definition 3.2.6, the orthogonal projection of $\cos(3t)$ onto X is given by,

$$\text{proj}_X(\cos(3t)) = \sum_{x_i\in S}\langle x_i, \cos(3t)\rangle x_i \tag{3}$$

We need to calculate the following inner products:

$$\begin{aligned}
\langle \cos(t), \cos(3t) \rangle &= \frac{1}{\pi}\int_{-\pi}^{\pi} \cos(t)\cos(3t)dt = \left(\frac{1}{\pi}\left(\frac{\sin(4x)}{8} + \frac{\sin(2x)}{4}\right)\right)\Bigg|_{-\pi}^{\pi} = 0 \\
\langle \sin(t), \cos(3t) \rangle &= \frac{1}{\pi}\int_{-\pi}^{\pi} \sin(t)\cos(3t)dt = \left(\frac{1}{\pi}\left(\frac{\cos(2x)}{4} - \frac{\cos(4x)}{8}\right)\right)\Bigg|_{-\pi}^{\pi} = 0 \\
\langle \cos(2t), \cos(3t) \rangle &= \frac{1}{\pi}\int_{-\pi}^{\pi} \cos(2t)\cos(3t)dt = \left(\frac{1}{\pi}\left(\frac{\sin(5x)}{10} + \frac{\sin(x)}{2}\right)\right)\Bigg|_{-\pi}^{\pi} = 0 \\
\langle \sin(2t), \cos(3t) \rangle &= \frac{1}{\pi}\int_{-\pi}^{\pi} \sin(2t)\cos(3t)dt = \left(\frac{1}{\pi}\left(\frac{\cos(x)}{2} - \frac{\cos(5x)}{10}\right)\right)\Bigg|_{-\pi}^{\pi} = 0
\end{aligned}$$

Therefore, we find that $\operatorname{proj}_X(\cos(3t)) = \mathbf{0}$.

Part (iv) In Part (i) we proved that S is an orthonormal set. Let $X = \operatorname{span}(S)$. Therefore, by Definition 3.2.6, the orthogonal projection of t onto X is given by,

$$\operatorname{proj}_X(t) = \sum_{x_i \in S} \langle x_i, t \rangle x_i \tag{4}$$

We need to calculate the following inner products:

$$\begin{aligned}
\langle \cos(t), t \rangle &= \frac{1}{\pi}\int_{-\pi}^{\pi} \cos(t)t dt = \left(\frac{1}{\pi}\left(\cos(t) + t\sin(t)\right)\right)\Bigg|_{-\pi}^{\pi} = 0 \\
\langle \sin(t), t \rangle &= \frac{1}{\pi}\int_{-\pi}^{\pi} \sin(t)t dt = \left(\frac{1}{\pi}\left(\sin(t) - t\cos(t)\right)\right)\Bigg|_{-\pi}^{\pi} = 2 \\
\langle \cos(2t), t \rangle &= \frac{1}{\pi}\int_{-\pi}^{\pi} \cos(2t)t dt = \left(\frac{1}{\pi}\left(\frac{\cos(2t)}{4} + \frac{t\sin(2t)}{2}\right)\right)\Bigg|_{-\pi}^{\pi} = 0 \\
\langle \sin(2t), t \rangle &= \frac{1}{\pi}\int_{-\pi}^{\pi} \sin(2t)t dt = \left(\frac{1}{\pi}\left(-\frac{t\cos(2t)}{2} + \frac{\sin(2t)}{4}\right)\right)\Bigg|_{-\pi}^{\pi} = -1
\end{aligned}$$

Therefore,

$$\operatorname{proj}_X(t) = 2\sin(t) - \sin(2t) \tag{5}$$

Exercise 3.9

The rotation matrix in $\mathbb{R}^2$ is given by,

$$R_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \tag{6}$$

Fix $x = (x_1, x_2), y = (y_1, y_2) \in \mathbb{R}^2$. Then,

$$L(x) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} = \begin{bmatrix} x_1\cos(\theta) + x_2\sin(\theta) & -x_1\sin(\theta) + x_2\cos(\theta) \end{bmatrix} \tag{7}$$

Similarly,

$$L(y) = \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} = \begin{bmatrix} y_1\cos(\theta) + y_2\sin(\theta) & -y_1\sin(\theta) + y_2\cos(\theta) \end{bmatrix} \tag{8}$$

Therefore,

$$\begin{aligned} \langle L(x), L(y)\rangle &= x_1y_1\cos^2(\theta) + x_1y_2\cos(\theta)\sin(\theta) + x_2y_1\cos(\theta)\sin(\theta) + x_2y_2\sin^2(\theta) \\ &\quad + x_1y_1\sin^2(\theta) - x_1y_2\cos(\theta)\sin(\theta) - x_2y_1\cos(\theta)\sin(\theta) + x_2y_2\cos^2(\theta) \\ &= x_1y_1 + x_2y_2 \\ &= \langle x, y\rangle \end{aligned}$$

Then by Definition 3.2.11, a rotation in $\mathbb{R}^2$ is an orthonormal transformation.

Exercise 3.10

Claim 2. The matrix $Q \in M_n(\mathbb{F})$ is an orthonormal matrix if and only if $Q^HQ = I = QQ^H$.

Proof. Suppose $Q \in M_n(\mathbb{F})$ is an orthonormal matrix and fix $x \in \mathbb{F}^n$ Then,

$$\langle x, x\rangle = \langle Qx, Qx\rangle = x^HQ^HQx = \langle x, Q^HQx\rangle \tag{9}$$

This holds for all x, which implies $x = Q^HQx$. Thus, it must be that $Q^HQ = I$. Since Q is square, this equation implies that $Q^{-1} = Q^H$. Therefore, it must also be the case that $QQ^H = I$.

Conversely, suppose $Q^HQ = I$. Fix $x, y \in \mathbb{F}^n$. Then,

$$\langle Qx, Qy\rangle = x^HQ^HQy = x^Hy = \langle x, y\rangle \tag{10}$$

Therefore, Q is an orthonormal matrix. □

Claim 3. If $Q \in M_n(\mathbb{F})$ is an orthonormal matrix, then $||Qx|| = ||x||$ for all $x \in \mathbb{F}^n$.

Proof. Let $Q \in M_n(\mathbb{F})$ be an orthonormal matrix and fix $x \in \mathbb{F}^n$. By Definition 3.2.14, Q is the matrix representation of an orthonormal operator on $\mathbb{F}^n$. Therefore, by Definition 3.2.11, Q satisfies,

$$\langle x, x\rangle = \langle Qx, Qx\rangle \tag{11}$$

We can simplify this equation as follows,

$$\begin{aligned} &\langle x, x\rangle = \langle Qx, Qx\rangle \\ &\implies ||x||^2 = ||Qx||^2 \\ &\implies ||x|| = ||Qx|| \end{aligned}$$

Our choice of x was arbitrary, therefore $||Qx|| = ||x||$ for all $x \in \mathbb{F}^n$. □

Claim 4. If $Q \in M_n(\mathbb{F})$ is an orthonormal matrix, then so is Q^{-1}.

Proof. Let $Q \in M_n(\mathbb{F})$ be an orthonormal matrix. By part (i), we know that $QQ^H = Q^H Q = I$, so that $Q^{-1} = Q^H$. Now, observe that $I = QQ^H = (Q^H)^H Q^H = (Q^{-1})^H Q^H = (Q^{-1})^H Q^{-1}$. Therefore, again by (i), we have that $Q^{-1} = Q^H$ is an orthonormal matrix. □

Claim 5. The columns of an orthonormal matrix $Q \in M_n(\mathbb{F})$ are orthonormal.

Proof. Let $Q \in M_n(\mathbb{F})$ be an orthonormal matrix. We write $Q = [q_1 q_2 \dots q_n]$ where q_i is the ith column of Q. By part (i), we have that $Q^H Q = I$. Therefore,

$$\begin{aligned} I &= [q_1 q_2 \dots q_n]^H [q_1 q_2 \dots q_n] \\ &= \begin{bmatrix} q_1^H \\ q_2^H \\ \vdots \\ q_n^H \end{bmatrix} [q_1 q_2 \dots q_n] \\ &= \begin{bmatrix} q_1^H q_1 & q_1^H q_2 & \dots & q_1^H q_n \\ q_2^H q_1 & q_2^H q_2 & \dots & q_2^H q_n \\ \vdots & \vdots & \ddots & \vdots \\ q_n^H q_1 & q_n^H q_2 & \dots & q_n^H q_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \end{aligned}$$

Observe that in the final line, the corresponding elements of the two matrices are equal. Therefore, we have that $q_i^H q_i = 1$ and $q_i^H q_j =$ for all $i \neq j$. Therefore, the columns of Q are by definition (Definition 3.2.1) orthonormal. □

Claim 6. If $Q \in M_n(\mathbb{F})$ is an orthonormal matrix, then $|\det(Q)| = 1$. The converse is false.

Proof. Let $Q \in M_n(\mathbb{F})$ be an orthonormal matrix. By part (i), we know that $Q^H Q = I$. Also note that a basic property of determinants is that $det(Q^H) = det(Q)$ and $detQ^H Q) = det(Q^H)det(Q)$. Therefore,

$$1 = det(I) = det(Q^H Q) = det(Q^H)det(Q) = (det(Q))^1. \tag{12}$$

Therefore, $|\det(Q)| = 1$. Conversely, consider the matrix,

$$A = \begin{bmatrix} 4 & 1 \\ 3 & 1 \end{bmatrix} \tag{13}$$

The $det(A) = 1$, but clearly A is not orthonormal. Therefore, the converse is not true. □

Claim 7. If $Q_1, Q_2 \in M_n(\mathbb{F})$ are orthonormal matrices, then the product $Q_1 Q_2$ is also an orthonormal matrix.

Proof. Let $Q_1, Q_2 \in M_n(\mathbb{F})$ be orthonormal matrices. We will show that $(Q_1Q_2)^H(Q_1Q_2) = I$. To that end,

$$\begin{aligned}
(Q_1Q_2)^H(Q_1Q_2) &= Q_2^H Q_1^H Q_1 Q_2 \\
&= Q_2^H (Q_1^H Q_1) Q_2 \\
&= Q_2^H I Q_2 && (Q_1^H Q_1 = I \text{ because } Q_1 \text{ is orthonormal}) \\
&= Q_2^H Q_2 \\
&= I && (Q_2^H Q_2 = I \text{ because } Q_2 \text{ is orthonormal})
\end{aligned}$$

Therefore, by part (i), because $(Q_1Q_2)^H(Q_1Q_2) = I$, we have that Q_1Q_2 is an orthonormal matrix. □

Exercise 3.11

Let $\{x_i\}_{i=1}^n$ be a collection of linearly dependent vectors. Without loss of generality, suppose that vector k is the first vector that makes the subset of vectors linearly dependent i.e. k is the index of first vector that can be written in terms of vectors 1 to $k-1$. Observe that the projection of x_k onto the span of $\{x_1, x_2, \ldots, x_{k-1}\}$ is simply x_k. Therefore, following the notation from the Gram-Schmidt Orthonormalization Process in the textbook, $p_{k-1} = x_k$. Therefore, x_k does not extend the dimension of $\{x_1, x_2, \ldots, x_{k-1}\}$ and the denominator of q_k is not well-defined. However, to make the process more useful, when this occurs, we could simply set $q_k = 0$. Therefore, when the Gram-Schmidt Orthonormalization Process is conducted in this manner on a set of linearly dependent vectors, it will return a set of vectors, some of which form an orthonormal set, and some of which are 0. To recover an orthonormal set we can simply discard the vectors that were set to 0.

Exercise 3.16

Claim 8. The QR decomposition is not unique.

Proof. Consider a diagonal matrix D (of dimension $n \times n$) where all entries on the diagonal are 1, except the entry in the last row and column (i.e. d_{nn}). Observe that $D^{-1} = D$. Then, let $A = QR$ be the QR decomposition of a matrix A. Then, $A = QR = QIR = QDD^{-1}R = \tilde{Q}\tilde{R}$. Now, $\tilde{Q} = QD$ is still an orthonormal matrix. This follows by part (vi) of Exercise 3.10. Clearly D is an orthonormal matrix, so that QD is also an orthonormal matrix. Furthermore, $\tilde{R} = D^{-1}R$ is still an upper diagonal matrix. Therefore, $A = \tilde{Q}\tilde{R}$ is another QR decomposition for A so that the QR decomposition is not unique. □

Claim 9. If A is invertible, then there is a unique QR decomposition of A such that R has only positive diagonal elements.

Proof. Suppose A has two QR decompositions: $A = QR = Q'R'$, where Q, Q' are orthonormal matrices and R, R' have positive diagonal elements. If A is invertible, then Q, Q', R, R' must also be invertible. Therefore, $QR = Q'R'$ implies that

$(Q')^{-1}Q = R'R^{-1}$. By Exercise 3.10, we know that $(Q')^{-1}$ is an orthonormal matrix, so that the product $(Q')^{-1}Q$ is also an orthonormal matrix. Similarly, $R'R^{-1}$ is an upper triangular matrix. Then, it must be the case that $(Q')^{-1}Q = I = R'R^{-1}$. Therefore, it follows that $R' = R$ and $Q' = Q$, which shows that the QR decomposition in this case is unique. □

Exercise 3.17

Claim 10. Let $A \in M_{m\times n}$ have rank $n \leq m$, and let $A = \hat{Q}\hat{R}$ be a reduced QR decomposition. Then, solving the system $A^H Ax = A^H b$ is equivalent to solving the system $\hat{R}x = \hat{Q}^H b$.

Proof. Let $x \in \mathbb{F}^n$ solve $A^H Ax = A^H b$. We can substitute in the reduced QR decomposition in to get that $A^H Ax = (\hat{Q}\hat{R})^H(\hat{Q}\hat{R})x = \hat{R}^H\hat{Q}^H\hat{Q}\hat{R}x$. $\hat{Q}$ is an orthonormal $m \times n$ matrix, so $\hat{Q}^H\hat{Q} = I_{n\times n}$. Therefore, $A^H Ax = \hat{R}^H\hat{R}x$. Additionally, $A^H b = \hat{R}^H\hat{Q}^H x$. Thus, $\hat{R}^H\hat{R}x = \hat{R}^H\hat{Q}^H x$. Observe that in the reduced QR decomposition, $\hat{R}$ is invertible (and hence $\hat{R}^H$ is invertible). This follows from the fact that A has full rank (linearly independent columns). Therefore, we can multiply both sides by $\left(\hat{R}^H\right)^{-1}$ on the left to find that $\hat{R}x = \hat{Q}^H x$. Therefore, by reversing this chain of equalities, solving the system $A^H Ax = A^H b$ is equivalent to solving the system $\hat{R}x = \hat{Q}^H b$. □

Exercise 3.23

Claim 11. $||\,||x|| - ||y||\,|| \leq ||x-y||$ for all $x, y \in V$.

Proof. Let $x, y \in V$. Then,

$$||x|| = ||x - y + y|| \leq ||x-y|| + ||y|| \tag{14}$$

where the inequality follows from the triangle inequality. Therefore,

$$||x|| - ||y|| \leq ||x-y|| \tag{15}$$

Similarly,

$$||y|| = ||y - x + x|| \leq ||y-x|| + ||x|| \tag{16}$$

Therefore,

$$||y|| - ||x|| \leq ||y-x|| = ||x-y|| \tag{17}$$

Combining (15) and (17), we find that $||\,||x|| - ||y||\,|| \leq ||x-y||$. □

Exercise 3.24

Claim 12. $||f||_{L^1} = \int_a^b |f(t)|dt$ is a norm on $C([a,b],\mathbb{F})$.

Proof. Let $f \in C([a,b],\mathbb{F})$.

(Positivity) It is clear that that $\int_a^b |f(t)|dt \geq 0$ for all $t \in [a,b]$, since the absolute value of $f(t)$ is weakly positive. Next, suppose $\int_a^b |f(t)|dt = 0$. This could occur in the degenerate case where $a = b$, but we disregard this. Because $|f(t)| \geq 0$ for all $t \in [a,b]$, for the integral to be 0, it must be that $f(t) = 0$ for all $t \in [a,b]$. Conversely, suppose $f(t) = 0$ for all $t \in [a,b]$. Then, clearly $\int_a^b |f(t)|dt = 0$. Thus, this norm satisfies the positivity condition.

(Scale preservation) Fix $a \in \mathbb{F}$. Then,

$$\begin{aligned} ||af||_{L^1} &= \int_a^b |af(t)|dt \\ &= \int_a^b |a||f(t)|dt \\ &= |a| \int_a^b |f(t)|dt \\ &= |a|||f||_{L^1} \end{aligned}$$

(Triangle Inequality) Let $g \in C([a,b],\mathbb{F})$. Then,

$$\begin{aligned} ||f+g||_{L^1} &= \int_a^b |f(t)+g(t)|dt \\ &\leq \int_a^b (|f(t)|+|g(t)|)\,dt \\ &= \int_a^b |f(t)|dt + \int_a^b |g(t)|dt \\ &= ||f||_{L^1} + ||g||_{L^1} \end{aligned}$$

These three properties verify that $||f||_{L^1}$ is indeed a norm. □

Claim 13. $||f||_{L^2} = \left(\int_a^b |f(t)|^2\right)^{\frac{1}{2}} dt$ is a norm on $C([a,b],\mathbb{F})$.

Proof. Let $f \in C([a,b],\mathbb{F})$.

(Positivity) This argument is analogous to that in Claim 12. $|f(t)|^2$ follows the sign of $|f(t)|$, so the argument presented there holds.

(Scale preservation) Fix $a \in \mathbb{F}$. Then,

$$\begin{aligned}
||af||_{L^2} &= \left(\int_a^b |af(t)|^2\right)^{\frac{1}{2}} dt \\
&= \left(\int_a^b |a|^2|f(t)|^2\right)^{\frac{1}{2}} dt \\
&= \left(|a|^2\int_a^b |f(t)|^2\right)^{\frac{1}{2}} dt \\
&= |a|\left(\int_a^b |f(t)|^2\right)^{\frac{1}{2}} dt \\
&= |a|||f||_{L^2}
\end{aligned}$$

(Triangle Inequality) Let $g \in C([a,b],\mathbb{F})$. Then,

$$\begin{aligned}
||f+g||_{L^2}^2 &= \int_a^b |f(t)+g(t)|^2 dt \\
&= \int_a^b (|f(t)|^2 + 2|f(t)g(t)| + |g(t)|^2)dt \\
&= \int_a^b |f(t)|^2 dt + \int_a^b |g(t)|^2 dt + 2\int_a^b (|f(t)g(t)|^2 dt \\
&\leq \int_a^b |f(t)|^2 dt + \int_a^b |g(t)|^2 dt + 2\left(\int_a^b |f(t)|^2\right)^{\frac{1}{2}}\left(\int_a^b |g(t)|^2\right)^{\frac{1}{2}} \\
&\qquad \text{(by Cauchy Schwarz)} \\
&= \left(\left(\int_a^b |f(t)|^2 dt\right)^{\frac{1}{2}} + \left(\int_a^b |g(t)|^2 dt\right)^{\frac{1}{2}}\right)^2
\end{aligned}$$

Therefore, $||f+g||_{L^2} \leq ||f||_{L^2} + ||g||_{L^2}$. These three properties verify that $||f||_{L^2}$ is indeed a norm. ☐

Claim 14. $||f||_{L^\infty} = \sup_{x\in[a,b]} |f(x)|$ is a norm on $C([a,b],\mathbb{F})$.

Proof. Let $f \in C([a,b],\mathbb{F})$.

(Positivity) It is clear that that $\sup_{x\in[a,b]} |f(x)| \geq 0$ for all $t \in [a,b]$, since the absolute value of $f(t)$ is weakly positive. Next, suppose $\sup_{x\in[a,b]} |f(x)|dt = 0$. Recall that the supremum takes the largest value of $|f(t)|$ over this interval. Therefore, if $\sup_{x\in[a,b]} |f(x)|dt = 0$, then it must be that $f(t) = 0$ for all $x \in [a,b]$. Conversely, suppose $f(t) = 0$ for all $x \in [a,b]$. Then clearly $\sup_{x\in[a,b]} |f(x)|dt = 0$.

(Scale Preservation) Fix $a \in \mathbb{F}$. Then,

$$\begin{aligned} ||af||_{L^\infty} &= \sup_{x\in[a,b]} |af(x)| \\ &= |a| \sup_{x\in[a,b]} |f(x)| \\ &= |a| ||f||_{L^\infty} \end{aligned}$$

(Triangle Inequality) Let $g \in C([a,b],\mathbb{F})$. Then,

$$\begin{aligned} ||f+g||_{L^\infty} &= \sup_{x\in[a,b]} |f(x)+g(x)| \\ &\le \sup_{x\in[a,b]} (|f(x)|+|g(x)|) \\ &= \sup_{x\in[a,b]} |f(x)| + \sup_{x\in[a,b]} |g(x)| \\ &= ||f||_{L^\infty} + ||g||_{L^\infty} \end{aligned}$$

These three properties verify that $||f||_{L^\infty}$ is indeed a norm. □

Exercise 3.26

Claim 15. Topological equivalence is an equivalence relation.

Proof. (Reflexivity) Let $||\cdot||_a$ be a norm on the vector space X. Then, set $m = M = 1$ in the definition of topological equivalence. Clearly,

$$m||x||_a \le ||x||_a \le M||x||_a \tag{18}$$

for all $x \in X$ when $m = M = 1$. Therefore, $||\cdot||_a$ is equivalent to itself, so topological equivalence is reflexive.

(Symmetry) Let $||\cdot||_b$ be a norm on the vector space X and suppose $||\cdot||_a$ is topologically equivalent to $||\cdot||_b$. Therefore, there exist constants $0 < m \le M$ such that

$$m||x||_a \le ||x||_b \le M||x||_a \tag{19}$$

for all $x \in X$. However, it also follows that,

$$\frac{1}{M}||x||_b \le ||x||_a \le \frac{1}{m}||x||_b \tag{20}$$

for all $x \in X$. Therefore, $||\cdot||_b$ is topologically equivalent to $||\cdot||_a$, so that the relation is symmetric.

(Transitivity) Let $||\cdot||_c$ be a norm on the vector space X. Suppose $||\cdot||_a$ is topologically equivalent to $||\cdot||_b$ and $||\cdot||_b$ is topologically equivalent to $||\cdot||_c$. Therefore, there exist constants $0 < m_1 \le M_1$ such that $m_1||x||_a \le ||x||_b \le M_1||x||_a$ for all $x \in X$. Similarly, there exist constants $0 < m_2 \le M_2$ such that $m_2||x||_b \le ||x||_c \le$

$M_2||x||_b$ for all $x \in X$. However, it follows from these inequalities and the fact that the constants are positive that,

$$m_1 m_2 ||x||_a \le ||x||_c \le M_1 M_2 ||x||_a \tag{21}$$

for all $x \in X$. Therefore, $||\cdot||_a$ is topologically equivalent to $||\cdot||_c$, so that topological equivalence is transitive. With these three properties, it follows that topological equivalence is an equivalence relation. $\square$

Claim 16. If $x \in \mathbb{F}^n$, then, $||x||_2 \le ||x||_1 \le \sqrt{n}||x||_2$.

Proof. Fix $x \in \mathbb{F}^n$. First observe that $(\|x\|_2)^2 = \sum_{i=1}^n |x_i|^2 \le (\sum_{i=1}^n |x_i|)^2 = (\|x\|_1)^2$. Therefore, $||x||_2 \le ||x||_1$. Secondly, observe that $\|x\|_1 = \sum_{i=1}^n |x_i| = \sum_{i=1}^n 1 \cdot |x_i| = \langle 1, |x| \rangle$. Therefore, by the Cauchy-Schwarz inequality and the fact that $\langle 1, |x| \rangle$ is positive, we have that, $\|x\|_1 = \langle 1, |x| \rangle \le (\sum_{i=1}^n 1^2)^{\frac{1}{2}} (\sum_{i=1}^n |x_i|^2)^{\frac{1}{2}} = \sqrt{n}\,\|x\|_2$. Therefore, chaining these two inequalities together, we have that $||x||_2 \le ||x||_1 \le \sqrt{n}||x||_2$. $\square$

Claim 17. If $x \in \mathbb{F}^n$, then, $||x||_\infty \le ||x||_2 \le \sqrt{n}||x||_\infty$.

Proof. Fix $x \in \mathbb{F}^n$. Suppose, without loss of generality, that $|x_k| = \|x\|_\infty$, where $1 \le k \le n$. Clearly, we have that $\|x\|_\infty = |x_k| \le (\sum_{i=1}^n |x_i|^2)^{\frac{1}{2}} = \|x\|_2$, because only including index k in the sum of the 2-norm leads to the inequality $|x_k| \le |x_k|$; therefore, including other positive terms will not reverse the inequality. Thus, $\|x\|_\infty \le \|x\|_2$. Secondly, observe that $(\|x\|_2)^2 = \sum_{i=1}^n |x_i|^2 \le n|x_k|^2$. This implies that $\|x\|_2 \le \sqrt{n}|x_k| = \sqrt{n}\,\|x\|_\infty$. Chaining these inequalities together shows that $||x||_\infty \le ||x||_2 \le \sqrt{n}||x||_\infty$. $\square$

Therefore, because topological equivalence is an equivalence relation (and hence symmetric and transitive), the above two claims imply that the p-norms for $p = 1, 2, \infty$ on $\mathbb{F}^n$ are topologically equivalent.

Exercise 3.28

Let A be an $n \times n$ matrix.

Claim 18. $\frac{1}{\sqrt{n}}\|A\|_2 \le \|A\|_1 \le \sqrt{n}\,\|A\|_2$

Proof. By the topological equivalence of the the 1-norm and 2-norm in $\mathbb{F}^n$ (Exercise 3.26), we have the following inequalities,

$$\frac{1}{\sqrt{n}\,\|x\|_2} \le \frac{1}{\|x\|_1} \le \frac{1}{\|x\|_2} \tag{22}$$

and

$$\|Ax\|_2 \le \|Ax\|_1 \le \sqrt{n}\,\|Ax\|_2 \tag{23}$$

We use the inequalities to find that,

$$\|A\|_1 \ge \frac{\|Ax\|_1}{\|x\|_1} \ge \frac{\|Ax\|_2}{\sqrt{n}\,\|x\|_2} \tag{24}$$

where the first inequality follows from the definition of the 1-norm and the second inequality follows from (22) and (23). Since (24) holds for all x, it follows that $\|A\|_1 \geq \frac{1}{\sqrt{n}} \|A\|_2$. Similarly, observe that,

$$\sqrt{n}\,\|A\|_2 \geq \frac{\sqrt{n}\,\|Ax\|_2}{\|x\|_2} \geq \frac{\|Ax\|_1}{\|x\|_1} \tag{25}$$

where again the first inequality follows from the definition of the 2-norm and the second inequality follows from (22) and (23). Since (24) holds for all x, it follows that $\sqrt{n}\,\|A\|_2 \geq \|A\|_1$. Putting these two results together shows that $\frac{1}{\sqrt{n}} \|A\|_2 \leq \|A\|_1 \leq \sqrt{n}\,\|A\|_2$. ☐

Claim 19. $\frac{1}{\sqrt{n}} \|A\|_\infty \leq \|A\|_1 \leq \sqrt{n}\,\|A\|_\infty$

Proof. This proof is analogous to that of the above claim. By the topological equivalence of the the 2-norm and ∞-norm in $\mathbb{F}^n$ (Exercise 3.26), we have the following inequalities,

$$\frac{1}{\sqrt{n}\,\|x\|_\infty} \leq \frac{1}{\|x\|_2} \leq \frac{1}{\|x\|_\infty} \tag{26}$$

and

$$\|Ax\|_\infty \leq \|Ax\|_2 \leq \sqrt{n}\,\|Ax\|_\infty \tag{27}$$

We use the inequalities to find that,

$$\|A\|_2 \geq \frac{\|Ax\|_2}{\|x\|_2} \geq \frac{\|Ax\|_\infty}{\sqrt{n}\,\|x\|_\infty} \tag{28}$$

where the first inequality follows from the definition of the 2-norm and the second inequality follows from (26) and (27). Since (28) holds for all x, it follows that $\|A\|_2 \geq \frac{1}{\sqrt{n}} \|A\|_\infty$. Similarly, observe that,

$$\sqrt{n}\,\|A\|_\infty \geq \frac{\sqrt{n}\,\|Ax\|_\infty}{\|x\|_\infty} \geq \frac{\|Ax\|_2}{\|x\|_2} \tag{29}$$

where again the first inequality follows from the definition of the ∞-norm and the second inequality follows from (26) and (27). Since (28) holds for all x, it follows that $\sqrt{n}\,\|A\|_\infty \geq \|A\|_2$. Putting these two results together shows that $\frac{1}{\sqrt{n}} \|A\|_\infty \leq \|A\|_2 \leq \sqrt{n}\,\|A\|_\infty$. ☐

Exercise 3.29

Claim 20. Any orthonormal matrix $Q \in M_n(\mathbb{F})$ has $\|Q\| = 1$, where $\|\cdot\|$ is the induced norm from $\mathbb{F}^n$ having the 2-norm.

Proof. Let $Q \in M_n(\mathbb{F})$ be an orthonormal matrix. Then,

$$\begin{aligned}
\|Q\| &= \sup_{x \neq 0} \frac{\|Qx\|_2}{\|x\|_2} \\
&= \sup_{x \neq 0} \frac{\|x\|_2}{\|x\|_2} && \text{(orthonormal transformations preserve lengths)} \\
&= 1
\end{aligned}$$

□

Claim 21. For any $x \in \mathbb{F}^n$, let $R_x : M_n(\mathbb{F}) \to \mathbb{F}^n$ be the linear transformation $A \mapsto Ax$. Then the induced norm of the transformation R_x is equal to $\|x\|_2$.

Proof. Fix $x \in \mathbb{F}^n$. Then, $\|R_x\| = \sup_{A:\|A\|\neq 0}\left(\frac{\|Ax\|_2}{\|A\|}\right)$. However, $\|A\| = \sup_{y\neq 0}\left(\frac{\|Ay\|_2}{\|y\|_2}\right) \geq \frac{\|Ax\|_2}{\|x\|_2}$. Therefore, $\|R_x\| \leq \sup_{A:\|A\|\neq 0}\left(\frac{\|Ax\|_2\|x\|_2}{\|Ax\|_2}\right) = \|x\|_2$. In sum, $\|R_x\| \leq \|x\|_2$.

Next, observe that if A is orthonormal, then equality is possible. Indeed, if A is orthonormal, then, $\|R_x\| = \|x\|_2$, because $\|Ax\|_2 = \|x\|_2$ and $\|A\| = 1$. □

Exercise 3.30

Claim 22. Let $S \in M_n(\mathbb{F})$ be an invertible matrix. Fix a matrix norm $\|\cdot\|$ on M_n. Then $\|A\|_S = \|SAS^{-1}\|$ is a matrix norm on $M_n(\mathbb{F})$.

Proof. We must show that $\|\cdot\|_S$ is indeed a norm and satisfies the submultiplicative property to prove that $\|\cdot\|_S$ is a matrix norm.

(Positivity) Fix $A \in M_n(\mathbb{F})$. Observe that $\|A\|_S = \|SAS^{-1}\| \geq 0$ because $\|\cdot\|$ is a matrix norm. Similarly, $\|A\|_S = \|SAS^{-1}\| = 0$ if and only if $SAS^{-1} = 0$ because $\|\cdot\|$ is a matrix norm. S is invertible implies that this can only occur when $A = 0$.

(Scale preservation) Let $\alpha \in \mathbb{R}$. Then, $\|\alpha A\|_S = \|\alpha SAS^{-1}\| = \alpha\|SAS^{-1}\| = \alpha\|A\|_S$, because $\|\cdot\|$ is a matrix norm.

(Triangle inequality) Let $B \in M_n(\mathbb{F})$. Then, $\|A+B\|_S = \|S(A+B)S^{-1}\| = \|SAS^{-1} + SBS^{-1}\| \leq \|SAS^{-1}\| + \|SBS^{-1}\| = \|A\|_S + \|B\|_S$, where the inequality follows because $\|\cdot\|$ is a matrix norm.

(Submultiplicative) We must show that $\|AB\|_S \leq \|A\|_S\|B\|_S$. Then,

$$\begin{aligned}\|AB\|_S &= \left\|AS^{-1}SB\right\|_S = \left\|S(AS^{-1}SB)S^{-1}\right\| = \left\|(SAS^{-1})(SBS^{-1})\right\| \\ &\leq \left\|SAS^{-1}\right\|\left\|SBS^{-1}\right\| = \|A\|_S\|B\|_S\end{aligned}$$

Therefore, $\|\cdot\|_S$ satisfies the properties of a norm and is submultiplicative. Thus, it is a matrix norm. □

Exercise 3.37

Observe that we may represent the space $V = \mathbb{R}[x;2]$ by a vector of coefficients in $\mathbb{R}^3$ on the basis $[1;x;x^2]$. Indeed, consider $p(x) \in \mathbb{R}[x;2]$. Then, $p(x) = [a_2;a_1;a_0][x^2;x;1]^T$. Then, we are looking for the unique q such that $L(p) = \langle q,p\rangle$. Observe that $L(p) = 2a_2 + a_1$. Thus, we want the q such that,

$$2a_2 + a_1 = L(p) = \langle q,p\rangle = q^T(a_2;a_1;a_0)^T \tag{30}$$

Clearly, $q^T = [2;1;0]$ satisfies this. By the discussion in the book before the Finite-Dimensional Riesz Representation Theorem, this q is unique.

Exercise 3.38

Suppose $p(x) = [a_2; a_1; a_0][x^2; x; 1]^T$. Then $p'(x) = [0; 2a_2; a_1][x^2; x; 1]^T$. Let

$$D = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{31}$$

Then,

$$D[p](x) = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} a_2 \\ a_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 2a_2 \\ a_1 \end{bmatrix} \tag{32}$$

is the derivative operator $D : V \to V$. The adjoint of D is given by the Hermitian conjugate,

$$D^* = D^H = \begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \tag{33}$$

Exercise 3.39

Claim 23. Let V and W be finite-dimensional inner product spaces. The adjoint has the following properties:

1. If $S, T \in \mathscr{L}(V; W)$, then $(S + T)^*) = S^* + T^*$ and $(\alpha T)^* = \bar{\alpha} T^*$, $\alpha \in \mathbb{F}$.
2. If $S \in \mathscr{L}(V; W)$, then $(S^*)^* = S$.
3. If $S, T \in \mathscr{L}(V)$, then $(ST)^* = T^* S^*$.
4. If $T \in \mathscr{L}(V)$ and T is invertible, then $(T^*)^{-1} = (T^{-1})^*$.

Proof. Let $v \in V$ and $w \in W$, and fix $\alpha \in \mathbb{F}$.

1. $\langle w, (S + T)v \rangle = \langle w, Sv \rangle + \langle w, Tv \rangle = \langle S^* v, w \rangle + \langle T^* v, w \rangle = \langle (S^* + T^*)v, w \rangle$. Therefore, by Definition 3.7.6, $(S + T)^* = S^* + T^*$. Similarly, $\langle v, \alpha Sw \rangle = \alpha \langle v, Sw \rangle = \alpha \langle S^* w, v \rangle = \langle \bar{\alpha} S^* w, v \rangle$. Thus, $\bar{\alpha} S^* = (\alpha S)^*$.
2. By the definition of the adjoint, we have that,

   $$\langle (S^*)^* w, v \rangle = \langle w, S^* v \rangle = \langle Sw, v \rangle \tag{34}$$

   This holds for all $v \in V$ and $w \in W$, which implies $(S^*)^* = S$.
3. By the definition of the adjoint, we have that,

   $$\langle (ST)^* w, v \rangle = \langle w, STv \rangle = \langle S^* w, Tv \rangle = \langle T^* S^* w, v \rangle \tag{35}$$

   This holds for all $v \in V$ and $w \in W$, which implies $(ST)^* = T^* S^*$.

4. A key observation is that $I = I^*$. Then, we have that,

$$\begin{aligned} T^*(T^*)^{-1} &= I \\ &= I^* \\ &= (T^{-1}T)^* \\ &= T^*(T^{-1})^* && \text{(by the previous claim)} \end{aligned}$$

Thus, $(T^*)^{-1} = (T^{-1})^*$.

□

Exercise 3.40

Let $M_n(\mathbb{F})$ be endowed with the Frobenius inner product.

Claim 24. Let $A \in M_n(\mathbb{F})$. Then, $A^* = A^H$.

Proof. Let $B, C \in M_n(\mathbb{F})$. Then,

$$\langle B, AC\rangle = tr(B^H AC) = tr((A^H B)^H C) = \langle A^H B, C\rangle \tag{36}$$

Therefore, by definition, $A^* = A^H$. □

Claim 25. For any $A_1, A_2, A_3 \in M_n(\mathbb{F})$, we have $\langle A_2, A_3A_1\rangle = \langle A_2A_1^*, A_3\rangle$.

Proof. Let $A_1, A_2, A_3 \in M_n(\mathbb{F})$. Then,

$$\langle A_2, A_3A_1\rangle = tr(A_2^H A_3A_1) = tr(A_1A_2^H A_3) = tr((A_2A_1^H)^H A_3) = \langle A_2A_1^H, A_3\rangle \tag{37}$$

Then by Claim 24, we know that $A_1^H = A_1^*$. Therefore, $\langle A_2, A_3A_1\rangle = \langle A_2A_1^*, A_3\rangle$. □

Exercise 3.44

Claim 26. Fredholm Alternative: Given $A \in M_{m\times n}(\mathbb{F})$ and $b \in \mathbb{F}^m$, either $Ax = b$ has a solution $x \in \mathbb{F}$ or there exists $y \in \mathscr{N}(A^H)$ such that $\langle y, b\rangle \neq 0$.

Proof. First note that the above claim is logically equivalent to the following claim: $Ax = b$ has a solution $x \in \mathbb{F}$ if and only if for all $y \in \mathscr{N}(A^H)$ we have that $\langle y, b\rangle = 0$. First suppose that $Ax = b$ has a solution $x \in \mathbb{F}$. Fix $y \in \mathscr{N}(A^H)$. Then,

$$\langle y, b\rangle = \langle y, Ax\rangle = \langle A^H y, x\rangle = \langle 0, x\rangle = 0 \tag{38}$$

because $y \in \mathscr{N}(A^H)$ implies that $A^H y = 0$. Conversely, suppose that for all $y \in \mathscr{N}(A^H)$ we have that $\langle y, b\rangle = 0$. For the sake of contradiction, assume that $Ax = b$ has no solution. This implies that $b \notin \mathscr{R}(A)$, which in turn implies that $b \in \mathscr{R}(A^H)$. Thus, it must be that $\langle b, b\rangle = 0$. This of course would only be true if $b = 0$ and be a contradiction otherwise. However, even in the case that $b = 0$, $Ax = b$ has a solution, namely $x = 0$. Thus, we have reached a contradiction and $Ax = b$ has a solution $x \in \mathbb{F}$. □

Exercise 3.45

Claim 27. $Sym_n(\mathbb{R})^\perp = Skew_n(\mathbb{R})$

Proof. We first show that $Skew_n(\mathbb{R}) \subset Sym_n(\mathbb{R})^\perp$. Let $A \in Skew_n(\mathbb{R})$. Fix $B \in Sym_n(\mathbb{R})$. We must show that $A \in Sym_n(\mathbb{R})^\perp$, so we must show that $\langle A, B\rangle = 0$. Indeed,

$$\langle A, B\rangle = \langle -A, B\rangle = -\langle A, B\rangle = -tr(AB) \tag{39}$$

Moreover, because we are working in the real numbers, we know that the inner product is symmetric. Therefore,

$$\langle A, B\rangle = \langle B, A\rangle = tr(B^T A) = tr(BA) = tr(AB) \tag{40}$$

where $B^T = B$ because B is symmetric. Therefore, these two chains of equalities give that $-tr(AB) = tr(AB)$, which implies that $tr(AB) = 0$. Thus, $\langle A, B\rangle = 0$, so $A \in Sym_n(\mathbb{R})^\perp$.

Next, we show that $Sym_n(\mathbb{R})^\perp \subset Skew_n(\mathbb{R})$. Let $A \in Sym_n(\mathbb{R})^\perp$ and $B \in Sym_n(\mathbb{R})$. Thus, $tr(A^T B) = 0$. We must show that $A \in Skew_n(\mathbb{R})$, or that $A^T = -A$. To that end, consider the sum $A + A^T$. Clearly, $(A + A^T) \in Sym_n(\mathbb{R})$, because $(A + A^T)^T = (A + A^T)$. Then,

$$\begin{aligned}
\langle A + A^T, B\rangle &= \langle A, B\rangle + \langle A^T, B\rangle \\
&= tr(A^T B) + tr(AB) \\
&= tr(AB) \\
&= tr(AB^T) && \text{(because } B \text{ is symmetric)} \\
&= tr(B^T A) \\
&= tr((A^T B)^T) \\
&= tr(A^T B) = 0
\end{aligned}$$

Therefore, $\langle A + A^T, B\rangle = 0$ for all $B \in Sym_n(\mathbb{R})$. Consider $B = A + A^T$. Then, by the positivity of the inner product, we have that $\langle A + A^T, A + A^T\rangle \geq 0$. However, we showed that this inner product was actually 0. This only occurs when one of the arguments itself is 0. Thus, $A + A^T = 0$, which implies that $A^T = -A$. ☐

Exercise 3.46

Claim 28. Let A be an $m \times n$ matrix.

1. If $x \in \mathscr{N}(A^H A)$, then Ax is both in $\mathscr{R}(A)$ and $\mathscr{N}(A^H)$.
2. $\mathscr{N}(A^H A) = \mathscr{N}(A)$.
3. A and $A^H A$ have the same rank.
4. If A has linearly independent columns, then $A^H A$ is nonsingular.

Proof.

1. Let $x \in \mathscr{N}(A^H A)$. Therefore, $A^H Ax = A^H(Ax) = 0$, so $Ax \in \mathscr{N}(A^H)$. Additionally, Ax is clearly in $\mathscr{R}(A)$. $\mathscr{R}(A)$ is the set of linear combinations of the columns of A, and Ax is precisely this.

2. We first show that $\mathscr{N}(A) \subset \mathscr{N}(A^H A)$. Let $x \in \mathscr{N}(A)$. Therefore, $Ax = 0$. Multiply both sides by A^H on the left to observe that $A^H Ax = 0$. Therefore, $x \in \mathscr{N}(A^H A)$. Next, we show that $\mathscr{N}(A^H A) \subset \mathscr{N}(A)$. Let $x \in \mathscr{N}(A^H A)$. Thus, $A^H Ax = 0$. We must show that $Ax = 0$. To that end, consider $\|Ax\|^2 = \langle Ax, Ax\rangle = x^H A^H Ax = 0$. Therefore, $\|Ax\| = 0$, which implies that $Ax = 0$.

3. Suppose the underlying space has dimension k and A has rank r. Therefore, $\mathscr{N}(A)$ has dimension $k - r$. But by (2), we have that $\mathscr{N}(A^H A) = \mathscr{N}(A)$. Therefore, $\mathscr{N}(A^H A)$ has dimension $k - r$, which implies that $A^H A$ has rank r. Thus, A and $A^H A$ have the same rank.

4. Suppose A has linearly independent columns. Then, A has full rank, i.e. rank n. Now, observe that $A^H A$ is an $n \times n$ matrix. By (3), we know that A and $A^H A$ have the same rank, so that $A^H A$ has rank n. Therefore, $A^H A$ is nonsingular.

□

Exercise 3.47

Claim 29. Let A be an $m \times n$ matrix of rank n. Let $P = A(A^H A)^{-1} A^H$. Then the following hold,

1. $P^2 = P$
2. $P^H = P$
3. $rank(P) = n$

Proof.

1. By direct calculation, observe that

$$\begin{aligned} P^2 &= (A(A^H A)^{-1} A^H)(A(A^H A)^{-1} A^H) \\ &= A(A^H A)^{-1}(A^H A)(A^H A)^{-1} A^H \\ &= A(A^H A)^{-1} A^H \\ &= P \end{aligned}$$

2. Again by direct calculation,

$$\begin{aligned} P^H &= (A(A^H A)^{-1} A^H)^H \\ &= A((A^H A)^{-1})^H A^H \\ &= A((A^H A)^H)^{-1} A^H \\ &= A(A^H A)^{-1} A^H \\ &= P \end{aligned}$$

3. I will use the following identity in my proof: If B is an $n \times k$ matrix of rank n, then $rank(AB) = rank(A)$. First consider the matrix product $(A^H A)^{-1} A^H$. By Exercise 3.46, if A has rank n, then $(A^H A)^{-1}$ has rank n. Similarly, the conjugate transpose of a matrix also preserves rank, so A^H has rank n. Therefore, $rank((A^H A)^{-1} A^H) = rank((A^H A)^{-1}) = n$. Now, consider $P = A(A^H A)^{-1} A^H$. Using the same argument, observe that $rank(P) = rank(A(A^H A)^{-1} A^H) = rank(A) = n$.

□

Exercise 3.48

Claim 30. Let $P(A) = \frac{A+A^T}{2}$ be the map $P : M_n(\mathbb{R}) \to M_n(\mathbb{R})$:

1. P is linear
2. $P^2 = P$
3. $P^* = P$
4. $\mathscr{N}(P) = Skew_n(\mathbb{R})$
5. $\mathscr{R}(P) = Sym_n(\mathbb{R})$
6. $\|A - P(A)\|_F = \sqrt{\frac{tr(A^T A) - tr(A^2)}{2}}$

Proof. Let $A, B \in M_n(\mathbb{R})$ and $\alpha, \beta \in \mathbb{R}$.

1. $P(\alpha A + \beta B) = \frac{(\alpha A + \beta B) + (\alpha A + \beta B)^T}{2} = \alpha \frac{A+A^T}{2} + \beta \frac{B+B^T}{2} = \alpha P(A) + \beta P(B)$. Therefore, P is linear.

2. When $P(A) = \frac{A+A^T}{2}$, we have that

$$P^2(A) = P\left(\frac{A + A^T}{2}\right) = \frac{\left(\frac{A+A^T}{2}\right) + \left(\frac{A+A^T}{2}\right)^T}{2} = \frac{2A + 2A^T}{4} = \frac{A + A^T}{2} = P(A) \tag{41}$$

   Therefore, $P^2 = P$.

3. By definition P^*, must satisfy $\langle A, P(B)\rangle = \langle P^*(A), B\rangle$ for all $A, B \in M_n(\mathbb{R})$. Consider the case where $A = B$. Then, we have that,

$$\begin{aligned}\langle A, P(A)\rangle = \langle P^* A, A\rangle &= tr((P^*(A))^T A) = tr((A^T P^*(A))^T) \\ &= tr(A^T P^*(A)) = \langle A, P^*(A)\rangle\end{aligned}$$

   Therefore, $P = P^*$.

4. We first show $\mathscr{N}(P) \subset Skew_n(\mathbb{R})$. To that end, let $A \in \mathscr{N}(P)$. Therefore, $P(A)A = 0$. Thus,
$$0 = \frac{A + A^T}{2} A = \frac{AA + A^T A}{2} \tag{42}$$
which implies $AA = -A^T A$, so it must be that $-A = A^T$, and hence $A \in Skew_n(\mathbb{R})$. We now show $\mathscr{N}(P) \supset Skew_n(\mathbb{R})$. To that end, let $A \in Skew_n(\mathbb{R})$. Therefore, $A^T = -A$. Observe that,
$$P(A)A = \frac{A + A^T}{2} A = \frac{AA + A^T A}{2} = \frac{AA - AA}{2} = 0 \tag{43}$$
Therefore, $A \in \mathscr{N}(P)$, so we have that $\mathscr{N}(P) = Skew_n(\mathbb{R})$.

5. We first show $\mathscr{R}(P) \supset Sym_n(\mathbb{R})$. Let $A \in Sym_n(\mathbb{R})$. Therefore, $A^T = A$. Then, observe that.
$$P(A) = \frac{A + A^T}{2} = \frac{2A}{A} = A \tag{44}$$
Thus, $A \in \mathscr{R}(P)$. We next show $\mathscr{R}(P) \subset Sym_n(\mathbb{R})$. Let $A \in \mathscr{R}(P)$. Therefore, there exists some $B \in M_n(\mathbb{R})$ such that $P(B) = A$, or $\frac{B^T + B}{2} = A$. Then,
$$A^T = \left(\frac{B + B^T}{2}\right)^T = \frac{B + B^T}{2} = A \tag{45}$$
Thus, $A^T = A$, so $A \in Sym_n(\mathbb{R})$. Therefore, $\mathscr{R}(P) \subset Sym_n(\mathbb{R})$, which implies that $\mathscr{R}(P) = Sym_n(\mathbb{R})$.

6.
$$\begin{aligned}
\|A - P(A)\|_F^2 &= tr((A - P(A))^T (A - P(A))) \\
&= tr((A^T - P(A))(A - P(A))) \\
&= tr(A^T A) - 2tr(P(A)A) + tr((P(A))^2) \\
&= tr(A^T A) - tr(A^T A) - tr(A^2) + \frac{1}{4}(2tr(A^2) + 2tr(A^T A)) \\
&= \frac{tr(A^T A) - tr(A^2)}{2}
\end{aligned}$$
Therefore,
$$\|A - P(A)\|_F = \sqrt{\frac{tr(A^T A) - tr(A^2)}{2}} \tag{46}$$
□

Exercise 3.50

Observe that we can write $rx^2 + sy^2 = 1$ as $y^2 = \frac{1}{s} - \frac{r}{s}x^2$. Therefore, to find the least square approximate for r and s set,

$$A = \begin{bmatrix} 1 & x_1^2 \\ 1 & x_2^2 \\ \vdots & \vdots \\ 1 & x_n^2 \end{bmatrix} \quad x = \begin{bmatrix} \frac{1}{s} \\ -\frac{r}{s} \end{bmatrix} \quad b = \begin{bmatrix} y_1^2 \\ y_2^2 \\ \vdots \\ y_n^2 \end{bmatrix} \tag{47}$$

and then solve the corresponding normal equations $A^H Ax = A^H b$ for r and s.