3.1. We have

$$\begin{aligned}\|x+y\|^2 - \|x-y\|^2 &= \langle x+y, x+y\rangle - \langle x-y, x-y\rangle \\ &= \langle x,x\rangle + \langle y,y\rangle + \langle x,y\rangle + \langle y,x\rangle - (\langle x,x\rangle + \langle y,y\rangle - \langle x,y\rangle - \langle y,x\rangle) \\ &= 4\langle x,y\rangle. \quad \text{(because this is a real inner product space so } \langle x,y\rangle = \langle y,x\rangle\text{)}\end{aligned}$$

Similarly,

$$\begin{aligned}\|x+y\|^2 + \|x-y\|^2 &= \langle x+y, x+y\rangle + \langle x-y, x-y\rangle \\ &= \langle x,x\rangle + \langle y,y\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle x,x\rangle + \langle y,y\rangle - \langle x,y\rangle - \langle y,x\rangle \\ &= 2(\|x\|^2 + \|y\|^2).\end{aligned}$$

3.2. First, note that:

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

Therefore,

$$i(\langle x-iy, x-iy\rangle - \langle x+iy, x+iy\rangle) = 2\langle x,y\rangle - 2\langle y,x\rangle.$$

Using the same calculation as in 3.1, we have:

$$\begin{aligned}\|x+y\|^2 - \|x-y\|^2 + i\|x+iy\|^2 - i\|x+iy\|^2 &= \langle x+y, x+y\rangle - \langle x-y, x-y\rangle + i\langle x-iy, x-iy\rangle - i\langle x+iy, x+iy\rangle \\ &= 2\langle x,y\rangle + 2\langle y,x\rangle + 2\langle x,y\rangle - 2\langle y,x\rangle \\ &= 4\langle x,y\rangle.\end{aligned}$$

3.3. For any function f, denote $x \to f(x)$ as the corresponding vector in $\mathbb{R}[x]$. That is, $x \to x$ is used to denote $f \in \mathbb{R}[x] : f(x) = x$. We have

$$\begin{aligned}\|x \to x\| &= \sqrt{\int_0^1 x^2 dx} \\ &= \frac{\sqrt{3}}{3} \\ \left\|x \to x^5\right\| &= \sqrt{\int_0^1 x^{10} dx} \\ &= \frac{\sqrt{11}}{11} \\ \left\langle x \to x, x \to x^5\right\rangle &= \int_0^1 x^6 dx \\ &= \frac{1}{7}\end{aligned}$$

Hence,

$$\cos(\theta_{(i)}) = \frac{\langle x \to x, x \to x^5\rangle}{\|x \to x\|\,\|x \to x^5\|} = \frac{\sqrt{21}}{7} \Rightarrow \theta_{(i)} = 0.857.$$

Similarly, we have

$$\begin{aligned}
\left\|x \to x^2\right\| &= \sqrt{\int_0^1 x^4 dx} \\
&= \frac{\sqrt{5}}{5} \\
\left\|x \to x^4\right\| &= \sqrt{\int_0^1 x^8 dx} \\
&= \frac{1}{3} \\
\left\langle x \to x^2, x \to x^4 \right\rangle &= \int_0^1 x^6 dx \\
&= \frac{1}{7}
\end{aligned}$$

Hence,

$$\cos(\theta_{(ii)}) = \frac{\left\langle x \to x^2, x \to x^4 \right\rangle}{\|x \to x^2\| \, \|x \to x^4\|} = \frac{3\sqrt{5}}{7} \Rightarrow \theta_{(ii)} = 0.29.$$

3.8.

(i) We have

$$\|\cos(t)\| = \sqrt{\frac{1}{\pi} \int_{-\pi}^{\pi} \cos^2(t) dt} = 1.$$

$$\|\sin(t)\| = \sqrt{\frac{1}{\pi} \int_{-\pi}^{\pi} \sin^2(t) dt} = 1.$$

$$\|\cos(2t)\| = \sqrt{\frac{1}{\pi} \int_{-\pi}^{\pi} \cos^2(2t) dt} = \sqrt{\frac{1}{2\pi} \int_{-2\pi}^{2\pi} \cos^2(u) du} = 1.$$

$$\|\sin(2t)\| = \sqrt{\frac{1}{\pi} \int_{-\pi}^{\pi} \sin^2(2t) dt} = \sqrt{\frac{1}{2\pi} \int_{-2\pi}^{2\pi} \sin^2(u) du} = 1.$$

$$\langle \cos(t), \sin(t) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \cos(t) \sin(t) dt = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sin(2t) dt = \frac{1}{4\pi} \int_{-2\pi}^{2\pi} \sin(u) du = 0.$$

$$\langle \cos(2t), \sin(2t) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \cos(2t) \sin(2t) dt = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sin(4t) dt = \frac{1}{8\pi} \int_{-4\pi}^{4\pi} \sin(u) du = 0.$$

$$\langle \cos(t), \sin(2t) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \cos(t) \sin(2t) dt = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\sin(3t) + \sin(t)) dt = 0.$$

$$\langle \cos(2t), \sin(t) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \cos(2t) \sin(t) dt = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\sin(3t) - \sin(t)) dt = 0.$$

$$\langle \cos(t), \cos(2t) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \cos(t)\cos(2t)dt = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\cos(3t) + \cos(t))dt = 0.$$

$$\langle \sin(t), \sin(2t) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \sin(t)\sin(2t)dt = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\cos(t) - \cos(3t))dt = 0.$$

Thus, S is an orthonormal set.

(ii) We have

$$\|t\| = \sqrt{\frac{1}{\pi} \int_{-\pi}^{\pi} t^2 dt} = \sqrt{\frac{2}{3}}\pi.$$

(iii) We know that S contains an orthonormal basis of X. By definition of projection, we have

$\operatorname{proj}_X(\cos(3t)) = \langle \cos(3t), \cos(t) \rangle \cos(t) + \langle \cos(3t), \sin(t) \rangle \sin(t) + \langle \cos(3t), \cos(2t) \rangle \cos(2t) + \langle \cos(3t), \sin(2t) \rangle \sin(2t)$

We have

$$\langle \cos(3t), \cos(t) \rangle = \int_{-\pi}^{\pi} \cos(3t)\cos(t)dt = 0.$$

$$\langle \cos(3t), \sin(t) \rangle = \int_{-\pi}^{\pi} \cos(3t)\sin(t)dt = 0.$$

$$\langle \cos(3t), \cos(2t) \rangle = \int_{-\pi}^{\pi} \cos(3t)\cos(2t)dt = 0.$$

$$\langle \cos(3t), \sin(2t) \rangle = \int_{-\pi}^{\pi} \cos(3t)\sin(2t)dt = 0.$$

Hence $\operatorname{proj}_X(\cos(3t)) = 0$.

(iv) We have

$$\langle t, \cos(t) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} t\cos(t)dt = 0.$$

$$\langle t, \sin(t) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} t\sin(t)dt = 2.$$

$$\langle t, \cos(2t) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} t\cos(2t)dt = 0.$$

$$\langle t, \sin(2t) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} t\sin(2t)dt = -1.$$

Hence, $\operatorname{proj}_X(t) = 2\sin(t) - \sin(2t)$.

3.9. Let R_θ be the rotation transformation about an angle θ. Then for any $x, y \in \mathbb{R}$ we have:

$$R_\theta(x, y) = (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta).$$

Therefore, for any (x,y) and (u,v) in $\mathbb{R}^2$:

$$\begin{aligned}\langle R_\theta(x,y), R_\theta(u,v)\rangle &= (x\cos\theta - y\sin\theta)(u\cos\theta - v\sin\theta) + (x\sin\theta + y\cos\theta)(u\sin\theta + v\cos\theta)\\ &= \cos^2\theta(xu+yv) + \cos\theta\sin\theta(-xv-yu+xv+yu) + \sin^2\theta(yv+xu)\\ &= xu+yv\\ &= \langle (x,y),(u,v)\rangle .\end{aligned}$$

Hence R_θ is an orthonomal transformation for any θ.

3.10.

(i) First note that if w satisfies $\langle w,v\rangle$ for all v, then choosing $v=w$ yields $\|w\rangle\| = 0$ and thus $w=0$. The same argument applies if $\langle v,w\rangle = 0$ for all v.
Now, suppose Q is an orthonormal matrix, then for all u,v we have:

$$\begin{aligned}\langle v,u\rangle &= \langle Qv, Qu\rangle\\ &= \left\langle Q^H Qv, u\right\rangle .\end{aligned}$$

This holds for all u, thus we deduce that $v = Q^H Qv$ for all v, and hence $Q^H Q = I$. It follows that $Q^{-1} = Q^H$ and hence $QQ^H = I$ also.
Conversely, suppose $QQ^H = Q^H Q = I$. Then for all u,v we have

$$\begin{aligned}\langle Qu, Qv\rangle &= \left\langle Q^H Qu, v\right\rangle\\ &= \langle u,v\rangle .\end{aligned}$$

Hence, Q is orthonormal.

(ii) For all $\mathbf{x}\in\mathbb{F}^n$, we have

$$\begin{aligned}\|Q\mathbf{x},\mathbf{x}\|^2 &= \langle Q\mathbf{x}, Q\mathbf{x}\rangle\\ &= \left\langle Q^H Q\mathbf{x}, \mathbf{x}\right\rangle\\ &= \|\mathbf{x}\|^2 . \quad \text{(because } Q^H Q = I\text{)}\end{aligned}$$

Therefore, $\|Q\mathbf{x}\| = \|\mathbf{x}\|$.

(iii) From (i) we know that $Q^H = Q^{-1}$, and therefore $(Q^{-1})^H = (Q^H)^H = Q$, so $(Q^{-1})^H Q^{-1} = QQ^{-1} = I$, so Q^{-1} is also orthonormal according to (i).

(iv) Let $\mathbf{x}_1,\ldots,\mathbf{x}_n$ be the column vectors of Q. We can deduce that $\overline{\mathbf{x}_1},\ldots,\overline{\mathbf{x}_n}$ are the corresponding row vectors of Q^H. It follows that the (i,j)-th entry of $Q^H Q$ is $\left\langle \mathbf{x}_i,\mathbf{x}_j\right\rangle$. Since $Q^H Q = I$, we deduce that $\left\langle \mathbf{x}_i,\mathbf{x}_j\right\rangle = 0$ if $i\neq j$ and $\left\langle \mathbf{x}_i,\mathbf{x}_j\right\rangle = 1$ if $i=j$. Therefore, $\mathbf{x}_1,\ldots,\mathbf{x}_n$ are orthonormal.

(v) The converse is obviously not true. For example, this matrix has determinant 1 but not orthonormal:

$$\begin{pmatrix}\frac{1}{2} & 0\\ 0 & 2\end{pmatrix}$$

.

3.11. Suppose the vectors that we apply Gram-Schmidt process to are $\mathbf{x}_1,\mathbf{x}_2,\ldots$ and suppose n is the largest integer such that $\mathbf{x}_1,\ldots,\mathbf{x}_n$ are independent. Then, by the Gram-Schmidt process for the first n vector, we will obtain an orthonrmal basis $\mathbf{q}_1,\ldots,\mathbf{q}_n$ of $X_n = \mathrm{span}(\mathbf{x}_1,\ldots,\mathbf{x}_n)$. Since $\mathbf{x}_1,\ldots,\mathbf{x}_{n+1}$ are dependent, we know that $\mathbf{x}_{n+1}\in X_n$ and thus, can be expressed uniquely as:

$$\mathbf{x}_{n+1} = \sum_{i=1}^{n} a_i\mathbf{q}_i .$$

By Gram-Schmidt process, we also know that the projection of $\mathbf{x}_{n+1}$ on X_n is

$$\begin{aligned}\mathbf{p}_{n+1} &= \sum_{i=1}^{n} \langle \mathbf{x}_{n+1}, \mathbf{q}_i \rangle \mathbf{q}_i \\ &= \sum_{i=1}^{n} \left\langle \sum_{i=1}^{n} a_i \mathbf{q}_i, \mathbf{q}_i \right\rangle \mathbf{q}_i\end{aligned}$$

Note that $\langle \mathbf{q}_i, \mathbf{q}_j \rangle = 0$ if $i \neq j$ and $\langle \mathbf{q}_i, \mathbf{q}_i \rangle = 1$ for every i. Thus, for every $i = 1, \ldots, n$:

$$\left\langle \sum_{i=1}^{n} a_i \mathbf{q}_i, \mathbf{q}_i \right\rangle = a_i.$$

We deduce that

$$\begin{aligned}\mathbf{p}_{n+1} &= \sum_{i=1}^{n} a_i \mathbf{q}_i \\ &= \mathbf{x}_{n+1}.\end{aligned}$$

Hence, $\mathbf{q}_{n+1} = 0$. The intuition is that during the Gram-Schmidt process, for any vector that can be expressed as a linear combination of the previous vectors, the process will return a zero vector. Hence, Gram-Schmidt will return a maximal orthonormal set, which is an orthonormal basis of span(given vectors).

3.16.

(i) Assume QR is a QR decomposition of A. Then $(-Q)(-R)$ is another QR decomposition, hence QR decomposition is not unique.

(ii)

3.17. We have
$$A^H A\mathbf{x} = A^H \mathbf{b} \Leftrightarrow \hat{R}^H \hat{Q}^H \hat{Q} \hat{R} \mathbf{x} = \hat{R}^H \hat{Q}^H \mathbf{b} \Leftrightarrow \hat{R}^H \hat{R} \mathbf{x} = \hat{R}^H \hat{Q}^H \mathbf{b}.$$
Since A has rank n, it follows that $\hat{R}^H$ also has rank n and thus injective (because it is a lower triangular matrix), hence $\hat{R}^H \hat{R} \mathbf{x} = \hat{R}^H \hat{Q}^H \mathbf{b}$ is equivalent to $\hat{R}\mathbf{x} = \hat{Q}^H \mathbf{b}$.

3.23. By the triangle inequality, we have

$$\|\mathbf{x} - \mathbf{y}\| + \|\mathbf{y}\| \geq \|\mathbf{x}\|$$

$$\|\mathbf{y} - \mathbf{x}\| + \|\mathbf{x}\| \geq \|\mathbf{y}\|$$

Note that $\|\mathbf{y} - \mathbf{x}\| = \|\mathbf{x} - \mathbf{y}\|$. By the two inequalities above, we deduce that

$$|\, \|\mathbf{x}\| - \|\mathbf{y}\| \,| = \max\{\|\mathbf{y} - \mathbf{x}\|, \|\mathbf{x} - \mathbf{y}\|\} \leq \|\mathbf{x} - \mathbf{y}\|.$$

3.24.

(i) Obviously $\|f\|_{L^1} \geq 0$ for all f. Since f is continuous, $\|f\|_{L^1} = \int_a^b |f(t)| dt = 0$ if and only if $f \equiv 0$. Indeed, suppose there exists $x_0 \in [a, b]$ such that $f(x_0) \neq 0$. Then because f is continuous, there exists $\varepsilon > 0$ such that $|f(x)|$ is very near to $|f(x_0)|$ (say $|f(x)| > \dfrac{|f(x_0)|}{2}$) for x in $[x_0 - \varepsilon, x_0 + \varepsilon]$. Thus, $\int_a^b |f(t)| dt > \varepsilon |f(x_0)| > 0$. The second property is obvious because $\int_a^b |af(t)|dt = |a| \int_a^b |f(t)|dt$ for every a. Finally, we have

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

(ii) The first and second properties can be proved similarly to part (i). It remains to show the triangle inequality. We will show

$$\int_a^b |f(t)+g(t)|^2 dt \leq \left(\sqrt{\int_a^b |f(t)|^2 dt} + \sqrt{\int_a^b |g(t)|^2 dt} \right)^2.$$

After expanding, the above inequality is equivalent to

$$\int_a^b |f(t)| \cdot |g(t)| dt \leq \sqrt{\int_a^b |f(t)|^2 dt} \cdot \sqrt{\int_a^b |g(t)|^2 dt},$$

which is true according to the Cauchy's Schwarz Inequality, with respect to the inner product

$$\langle f,g \rangle = \int_a^b |f(t)||g(t)|dt.$$

(iii) The first and second properties are obvious. It remains to show the triangle inequality. Note that for any f, we have $\|f\|_{L^\infty} \geq |f(x)|$ for all $x \in [a,b]$. We need to show for any $f,g \in C([a,b];\mathbb{R})$:

$$\sup_{x\in[a,b]} |f(x)+g(x)| \leq \sup_{x\in[a,b]} |f(x)| + \sup_{x\in[a,b]} |g(x)|.$$

Now, for any $x \in [a,b]$, we have $|f(x)+g(x)| \leq |f(x)|+|g(x)| \leq \|f\|_{L^\infty} + \|g\|_{L^\infty}$. From this we deduce that

$$\sup_{x\in[a,b]} |f(x)+g(x)| \leq \|f\|_{L^\infty} + \|g\|_{L^\infty}.$$

3.26. Obviously $\|\cdot\|_a \sim \|\cdot\|_a$ (just choose $m = M = 1$). Now suppose $\|\cdot\|_a \sim \|\cdot\|_b$, we will show $\|\cdot\|_b \sim \|\cdot\|_a$. We know that there exist $M \geq m > 0$ such that for all $\mathbf{x} \in X$:

$$m\|\mathbf{x}\|_a \leq \|\mathbf{x}\|_b \leq M\|\mathbf{x}\|_a$$

which can be rewritten as

$$\frac{1}{M}\|\mathbf{x}\|_b \leq \|\mathbf{x}\|_a \leq \frac{1}{m}\|\mathbf{x}\|_b.$$

Hence, $\|\cdot\|_b \sim \|\cdot\|_a$. Finally, suppose $\|\cdot\|_a \sim \|\cdot\|_b$ and $\|\cdot\|_b \sim \|\cdot\|_c$. There exist $k,K,l,L > 0$ such that:

$$\begin{aligned}
k\|\mathbf{x}\|_a &\leq \|\mathbf{x}\|_b \leq K\|\mathbf{x}\|_a \\
l\|\mathbf{x}\|_b &\leq \|\mathbf{x}\|_c \leq L\|\mathbf{x}\|_b.
\end{aligned}$$

We deduce that

$$k\|\mathbf{x}\|_a \le \|\mathbf{x}\|_b \le \frac{1}{l}\|\mathbf{x}\|_c$$

. and

$$\|\mathbf{x}\|_c \le L\|\mathbf{x}\|_b \le LK\|\mathbf{x}\|_a.$$

Therefore,

$$kl\|\mathbf{x}\|_a \le \|\mathbf{x}\|_c \le KL\|\mathbf{x}\|_a.$$

so $\|\cdot\|_a \sim \|\cdot\|_c$, and thus topological equivalence is an equivalence relation.

(i) Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. We have

$$x_1^2 + \ldots + x_n^2 \le (|x_1| + \ldots + |x_n|)^2 \Rightarrow \|\mathbf{x}\|_2 \le \|\mathbf{x}\|_1.$$

and

$$\begin{aligned} &\sum_{i\neq j}(|x_i| - |x_j|)^2 \ge 0 \\ \Leftrightarrow\ &(n-1)\sum_{i=1}^{n} x_i^2 \ge 2\sum_{i<j}|x_i x_j| \end{aligned}$$

plus $\sum_{i=1}^{n} x_i^2$ to both sides, we get

$$n\sum_{i=1}^{n} x_i^2 \ge (\sum_{i=1}^{n}|x_i|)^2.$$

or equivalently, $\|\mathbf{x}\|_1 \le \sqrt{n}\|\mathbf{x}\|_2$. Hence $\|\mathbf{x}\|_2 \le \|\mathbf{x}\|_1 \le \sqrt{n}\|\mathbf{x}\|_2$.

(ii) Obviously $|x_i|^2 \le \sum_{i=1}^{n} x_i^2$ for all $i = 1, \ldots, n$, so $\max_i |x_i| \le \sqrt{\sum_{i=1}^{n} x_i^2}$, or $\|\mathbf{x}\|_\infty \le \|\mathbf{x}\|_2$. Moreover, $\sum_{i=1}^{n} x_i^2 \le n\max_i |x_i|$, hence $\|\mathbf{x}\|_2 \le \sqrt{n}\|\mathbf{x}\|_\infty$. Thus, $\|\mathbf{x}\|_\infty \le \|\mathbf{x}\|_2 \le \sqrt{n}\|\mathbf{x}\|_\infty$.

3.28.

(i) let $A = (a_{ij})$. Note that $\mathbf{x} \to A\mathbf{x}$ is linear, so We have $\|A\|_2 = \sup_{\|\mathbf{x}\|_2=1}\|Ax\|$. For any $\mathbf{x}$ such that $\|\mathbf{x}\|_2 = 1$, we have

$$\begin{aligned} \|A\mathbf{x}\|_2^2 &= \sum_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}x_j\right)^2 \\ &\le \sum_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}^2\right) \quad \text{(by the Cauchy Schwarz Inequality and the fact that } \|\mathbf{x}\| = 1) \\ &= \sum_{i,j=1}^{n} a_{ij}^2 \\ &= \sum_{j=1}^{n}\left(\sum_{i=1}^{n} a_{ij}^2\right) \\ &\le \sum_{j=1}^{n}\left(\sum_{i=1}^{n}|a_{ij}|\right)^2 \\ &\le n\sup_{1\le j\le n}\left(\sum_{i=1}^{n}|a_{ij}|\right)^2 \\ &= n\|A\|_1^2. \end{aligned}$$

Since this holds for all $\mathbf{x}$ such that $\|\mathbf{x}\|_2 = 1$, we deduce that $\|A\|_2 = \sup_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2 \le \sqrt{n}\,\|A\|_1$. Next, WLOG, suppose $\|A\|_1 = \sum\limits_{i=1}^{n} |a_{ij_0}|$. Let $\mathbf{c}_{j_0}$ be this j_0-th column vector of A. Let $\mathbf{x}$ be the vector such that $x_{j_0} = 1$ and $x_i = 0$ everywhere else. Then it is easy to see that $\|A\mathbf{x}\|_1 = \left\|\mathbf{c}_{j_0}\right\|_1 = \|A\|_1$ and $\|A\mathbf{x}\|_2 = \left\|\mathbf{c}_{j_0}\right\|_2$. By the previous exercise and the fact that $\|x\|_2 = 1$, we know that

$$\|A\|_1 = \left\|\mathbf{c}_{j_0}\right\|_1 \le \sqrt{n}\left\|\mathbf{c}_{j_0}\right\|_2 = \sqrt{n}\,\|A\mathbf{x}\|_2 \le \sqrt{n}\,\|A\|_2 .$$

(ii) Similar to part (i), for any $\mathbf{x}$ such that $\|\mathbf{x}\|_2 = 1$, we have

$$\begin{aligned}
\|A\mathbf{x}\|_2^2 &= \sum_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}x_j\right)^2 \\
&\le \sum_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}^2\right) \quad \text{(by the Cauchy Schwarz Inequality and the fact that } \|\mathbf{x}\| = 1) \\
&= \sum_{i,j=1}^{n} a_{ij}^2 \\
&= \sum_{j=1}^{n}\left(\sum_{i=1}^{n} a_{ij}^2\right) \\
&\le \sum_{j=1}^{n}\left(\sum_{i=1}^{n} |a_{ij}|\right)^2 \\
&\le n \sup_{1\le i\le n}\left(\sum_{j=1}^{n} |a_{ij}|\right)^2 \\
&= n\,\|A\|_\infty^2 .
\end{aligned}$$

Since this holds for all $\mathbf{x}$ such that $\|\mathbf{x}\|_2 = 1$, we deduce that $\|A\|_2 = \sup_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2 \le \sqrt{n}\,\|A\|_\infty$. Next, WLOG, suppose $\|A\|_\infty = \sum\limits_{j=1}^{n} |a_{i_0 j}|$. Let $\mathbf{r}_{i_0}$ be this i_0-th row of A. Choose $\mathbf{x}$ be the vector such that $x_j \in \{1,-1\}$ and $x_j a_{i_0 j} = |a_{i_0 j}|$ for $j = 1,\dots,n$. Then it is easy to see that $\|A\mathbf{x}\|_\infty = \left\|\mathbf{r}_{i_0}\right\|_1$. Then use the previous exercise and the fact that $\|\mathbf{x}\|_2 = \sqrt{n}$, we have

$$\|A\|_\infty = \left\|\mathbf{r}_{i_0}\right\|_1 = \|A\mathbf{x}\|_\infty \le \|A\mathbf{x}\|_2 \le \|A\|_2\,\|\mathbf{x}\|_2 = \sqrt{n}\,\|A\|_2 .$$

3.29. We will denote $\|\cdot\|$ as the 2-norm throughout this problem. Since Q is orthonormal, for every $\mathbf{x}$, we have

$$\|Q\mathbf{x}\|^2 = \langle Q\mathbf{x}, Q\mathbf{x}\rangle = \langle \mathbf{x}, \mathbf{x}\rangle = \|\mathbf{x}\|^2 .$$

Hence, we have $\|Q\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$, and thus $\|Q\| = \sup_{\mathbf{x}} \dfrac{\|Q\mathbf{x}\|}{\|\mathbf{x}\|} = 1$.

Now, we have

$$\|R_{\mathbf{x}}\| = \sup_{A} \frac{\|A\mathbf{x}\|}{\|A\|}.$$

Since $\|A\mathbf{x}\| \le \|A\|\,\|\mathbf{x}\|$ for all A, we deduce that $\|R_{\mathbf{x}}\| \le \|\mathbf{x}\|$. Now since $\mathbb{F}$ is a field, it has the multiplicative identity 1, and therefore $M_n(\mathbb{F})$ has the identity matrix I. We have $R_{\mathbf{x}}(I) = I\mathbf{x} = \mathbf{x}$, thus $\dfrac{\|R(I)\|}{\|I\|} = \|\mathbf{x}\|$. From this we conclude that $\|R_{\mathbf{x}}\| \ge \dfrac{\|R(I)\|}{\|I\|} = \|\mathbf{x}\|$. Therefore, $\|R_{\mathbf{x}}\| = \|\mathbf{x}\|$.

3.30. Obviously $\|A\|_S = \left\|SAS^{-1}\right\| \ge 0$. Moreover, $\|A\|_S = 0$ if and only if $\left\|SAS^{-1}\right\| = 0$, and since $\|\cdot\|$ is a norm, it follows that $SAS^{-1} = 0$. But then $A = S^{-1}(SAS^{-1})S = 0$. Thus the first property is satisfied. Next, we have $\|aA\|_S = \left\|S(aA)S^{-1}\right\| = \left\|aSAS^{-1}\right\| = |a| \left\|SAS^{-1}\right\| = |a| \|A\|_S$. Triangle inequality is also obvious because $\|A+B\|_S = \left\|S(A+B)S^{-1}\right\| = \left\|SAS^{-1}+SBS^{-1}\right\| \le \left\|SAS^{-1}\right\| + \left\|SBS^{-1}\right\| = \|A\|_S + \|B\|_S$. Finally, we check submultiplicative property. We have $\|AB\|_S = \left\|SAS^{-1}SBS^{-1}\right\| \le \left\|SAS^{-1}\right\| \left\|SBS^{-1}\right\| = \|A\|_S \|B\|_S$ ($\|\cdot\|$ is a matrix norm so it already has submultiplicative property). Hence we conclude that $\|\cdot\|_S$ is a matrix norm for every invertible S.

3.37. Each polynomial $p \in V$ can be written uniquely as $p(x) = ax^2+bx+c$, where $a,b,c \in \mathbb{R}$. From this we have $L[ax^2+bx+c] = 2a+b$. Note that $3x^2, 2x, 1$ forms an orthonormal basis of V, and we know by the discussion before Theorem 3.7.1 that if we define $q(x) = L(3x^2)\cdot 3x^2 + L(2x)\cdot 2x + L(1)\cdot 1$ (since we are working in the real numbers, we don't need to worry about conjugates), then $L(p) = \langle q, p\rangle$ for all $p \in V$. We have

$$L(3x^2) = 6, \quad L(2x) = 2, \quad L(1) = 0.$$

Thus $q(x) = 18x^2+4x$.

3.38. We have $D[ax^2+bx+c] = 2ax+b$, where $ax^2+bx+c, 2ax+b \in V$. We can rewrite as $D[(c,b,a)] = (b,2a,0)$. Thus, we have

$$A_D = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

The adjoint of A_D is just its transpose:

$$A_D^* = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 2 & 0 \end{pmatrix}$$

3.39.

(i) Let $U = (S+T)^*$. We have

$$\langle u, (S+T)v\rangle_W = \langle Uu, v\rangle_V \Rightarrow \langle u, Sv\rangle_W + \langle u, Tv\rangle_W = \langle Uu, v\rangle_V.$$

It follows that

$$\langle Uu, v\rangle_V = \langle S^*u, v\rangle_V + \langle T^*u, v\rangle_V = \langle (S^*+T^*)u, v\rangle_V.$$

This holds for all $u \in W, v \in V$, hence we deduce that $(S+T)^* = S^*+T^*$.

(ii) For all $w \in W, v \in V$, we have

$$\langle v, S^*w\rangle_V = \langle (S^*)^*v, w\rangle_W$$

and

$$\overline{\langle w, Sv\rangle}_W = \overline{\langle S^*w, v\rangle}_V \Rightarrow \langle Sv, w\rangle_W = \langle v, S^*w\rangle_V.$$

Hence, we deduce that $\langle (S^*)^*v, w\rangle_W = \langle Sv, w\rangle_w$ for all $v \in V, w \in W$, thus $(S^*)^* = S$.

(iii) For any $w, v \in V$, we have

$$\langle w, (ST)v\rangle_V = \langle w, S(Tv)\rangle_V = \langle S^*w, Tv\rangle_V = \langle T^*S^*w, v\rangle_V.$$

Since this holds for all $w, v \in V$, we deduce that $(ST)^* = T^*S^*$.

(iv) For all $w, v \in V$, we have

$$\left\langle (T^{-1})^*T^*w, v\right\rangle_V = \left\langle T^*w, T^{-1}v\right\rangle_V = \left\langle w, TT^{-1}v\right\rangle_V = \langle w, v\rangle_V.$$

Since this holds for all $w, v \in V$, we deduce that if T is invertible then $(T^*)^{-1} = (T^{-1})^*$.

3.40. For any A, denote the operator $A : M_n(\mathbb{F}) \to M_n(\mathbb{F})$ as $A[B] = BA$.

(i) It suffices to show that for any $B, C \in M_n(\mathbb{F})$, we have

$$\langle B, A[C] \rangle = \left\langle A^H[B], C \right\rangle$$

or equivalently,

$$\langle B, CA \rangle = \left\langle BA^H, C \right\rangle.$$

We have $\langle B, CA \rangle = \operatorname{tr}(B^H CA)$ and $\left\langle BA^H, C \right\rangle = \operatorname{tr}((BA^H)^H C) = \operatorname{tr}(AB^H C)$ (because $(A^H)^H = A$). Note that $\operatorname{tr}(XY) = \operatorname{tr}(YX)$ for all $X, Y \in M_n(\mathbb{F}$, thus $\operatorname{tr}(B^H CA) = \operatorname{tr}((B^H C)A) = \operatorname{tr}(AB^H C)$, and therefore $\langle B, CA \rangle = \left\langle BA^H, C \right\rangle$. Thus, $A^* = A^H$ (in terms of a linear operator on $M_n(\mathbb{F})$.

(ii) Basically part (i).

(iii) Note that $A^* = A^H$. It suffices to show that for all $X, Y \in M_n(\mathbb{F})$, we have

$$\langle Y, T_A(X) \rangle = \langle T_{A^H}(Y), X \rangle$$

or equivalently,

$$\langle Y, AX - XA \rangle = \left\langle A^H Y - YA^H, X \right\rangle.$$

We have

$$\begin{aligned}
\langle Y, AX - XA \rangle &= \operatorname{tr}(Y^H(AX - XA)) \\
&= \operatorname{tr}(Y^H AX - Y^H XA) \\
&= \operatorname{tr}(Y^H AX) - \operatorname{tr}(Y^H XA).
\end{aligned}$$

and

$$\begin{aligned}
\left\langle A^H Y - YA^H, X \right\rangle &= \operatorname{tr}((A^H Y - YA^H)^H X) \\
&= \operatorname{tr}((A^H Y)^H X - (YA^H)^H X) \\
&= \operatorname{tr}(Y^H AX - AY^H X) \\
&= \operatorname{tr}(Y^H AX) - \operatorname{tr}(AY^H X) \\
&= \operatorname{tr}(Y^H AX) - \operatorname{tr}(Y^H XA) \quad (\text{because } \operatorname{tr}(A(Y^H X)) = \operatorname{tr}((Y^H X)A)).
\end{aligned}$$

Therefore, $\langle Y, AX - XA \rangle = \left\langle A^H Y - YA^H, X \right\rangle$ for all X, Y, and thus $(T_A)^* = T_{A^*}$.

3.44. Suppose $A\mathbf{x} = \mathbf{b}$ does not have a solution in $\mathbb{F}^n$. Then we know that $\mathbf{b} \notin \mathscr{R}(A)$. Since $\mathscr{R}(A) = \mathscr{R}((A^*)^*) = \mathscr{N}(A^*)^\perp = \mathscr{N}(A^H)^\perp$, it follows that $\mathbf{b} \notin \mathscr{N}(A^H)^\perp$. This means that there exists $\mathbf{y} \in \mathscr{N}(A^H)$ such that y and b are not orthogonal, or $\langle y, b \rangle \neq 0$, as desired.

3.45. Consider $B = (b_{ij}) \in \operatorname{Sym}_n(\mathbb{R})^\perp$. Then $\operatorname{tr}(A^T B) = \operatorname{tr}(AB) = 0$ for all $A \in \operatorname{Sym}_n(\mathbb{R})$. Consider any $i_0, j_0 \in \{1, 2, \ldots, n\}$. Let $A_{i_0 j_0} = (a_{ij}) \in M_n(\mathbb{R})$ be a matrix such that $a_{i_0 j_0} = a_{j_0 i_0} = 1$ and $a_{ij} = 0$ everywhere else. Then $A_{i_0 j_0} \in \operatorname{Sym}_n(\mathbb{R})$ and $\operatorname{tr}(A_{i_0 j_0} B) = b_{i_0 j_0} + b_{j_0 i_0} = 0$. This holds for all $i_0, j_0 \in \{1, 2 \ldots, n\}$, thus we deduce that $b_{ij} = -b_{ji}$ for any i, j, therefore $B^T = -B$ and $B \in \operatorname{Skew}_n(\mathbb{R})$. Hence we have shown that $\operatorname{Sym}_n(\mathbb{R})^\perp \subseteq \operatorname{Skew}_n(\mathbb{R})$. Now conversely, consider $B = (b_{ij}) \in \operatorname{Skew}_n(\mathbb{R})$. Let $A = (a_{ij}) \in \operatorname{Sym}_n(\mathbb{R})$ be any symmetric matrix. Note that $a_{ij} = a_{ji}$ and $b_{ij} = -b_{ji}$ for all i, j. We have

$$\begin{aligned}
\operatorname{tr}(A^T B) &= \operatorname{tr}(AB) \\
&= \sum_{i=1}^n \sum_{j=1}^n a_{ij} b_{ji} \\
&= \sum_{i<j}^n a_{ij}(b_{ij} + b_{ji}) \\
&= 0.
\end{aligned}$$

This holds for any $B \in \operatorname{Skew}_n(\mathbb{R})$ and all $A \in \operatorname{Sym}_n(\mathbb{R})$, thus $\operatorname{Skew}_n(\mathbb{R}) \subseteq \operatorname{Sym}_n(\mathbb{R})^\perp$. Thus $\operatorname{Sym}_n(\mathbb{R})^\perp = \operatorname{Skew}_n(\mathbb{R})$.

3.46.

(i) $\mathbf{x} \in \mathcal{N}(A^H A)$ implies that $A^H A\mathbf{x} = 0$, or $A^H(A\mathbf{x}) = 0$ and thus $A\mathbf{x} \in \mathcal{N}(A^H)$. $A\mathbf{x} \in \mathcal{R}(A)$ is obvious.

(ii) By the Fundamental Subspaces Theorem, we know that $\mathcal{R}(A)^\perp = \mathcal{N}(A^H)$, and since $\mathcal{R}(A) \cap \mathcal{R}(A)^\perp = \{0\}$, it follows from part (i) that for any $\mathbf{x} \in \mathcal{N}(A^H A)$, $A\mathbf{x} \in \mathcal{R}(A) \cap \mathcal{N}(A^H) = \mathcal{R}(A) \cap \mathcal{R}(A)^\perp = \{0\}$, hence $\mathbf{x} \in \mathcal{N}(A)$. Thus $\mathcal{N}(A^H A) \subseteq \mathcal{N}(A)$. On the other hand, obviously $\mathcal{N}(A) \subseteq \mathcal{N}(A^H A)$ because $A\mathbf{x} = 0$ implies $A^H(A\mathbf{x}) = 0$. Hence, $\mathcal{N}(A^H A) = \mathcal{N}(A)$.

(iii) By the rank-nullity theorem and the fact that both $A^H A$ and A are matrices with n columns, we have:

$$n = \operatorname{rank}(A) + \dim \mathcal{N}(A) = \operatorname{rank}(A^H A) + \dim \mathcal{N}(A^H A).$$

Since $\mathcal{N}(A^H A) = \mathcal{N}(A)$, we deduce that A and $A^H A$ have the same rank.

(iv) If A has linearly independent columns then its rank is n (because the rank of matrix cannot exceed its number of columns, but since the columns are linearly independent the rank is greater than or equal the number of columns, thus the rank is the number of columns). But then $A^H A$ is a $n \times n$ matrix with rank n, and thus singular.

3.47.

(i) We have
$$P^2 = A[(A^H A)^{-1}(A^H A)](A^H A)^{-1}A^H = A(A^H A)^{-1}A^H = P.$$

(ii) Let $T = A^H A$. Then $T^H = (A^H A)^H = A^H (A^H)^H = A^H A = T$. Moreover, $I = (TT^{-1})^H = (T^{-1})^H T^H = (T^{-1})^H T$, therefore $(T^{-1})^H = T^{-1}$. We have

$$P^H = [A(A^H A)^{-1}A^H]^H = (A^H)^H[(A^H A)^{-1}]^H A^H = A(T^{-1})^H A^H = AT^{-1}A^H = P.$$

(iii) We will show that $\mathcal{N}(P) = \mathcal{N}(A^H)$. Consider any $\mathbf{x} \in \mathcal{R}(P)$ such that $P\mathbf{x} = 0$. Equivalently, we have $A(A^H A)^{-1}A^H\mathbf{x} = 0$, hence $(A^H A)^{-1}A^H\mathbf{x} \in \mathcal{N}(A)$. However, since A has rank n, it has n linearly independent columns and thus, $A\mathbf{v} = 0$ if and only if $\mathbf{v} = 0$. This implies that $(A^H A)^{-1}A^H\mathbf{x} = 0$. Again, we know that $(A^H A)^{-1}$ is invertible, or nonsingular, and therefore $\mathcal{N}((A^H A)^{-1}\{0\}$. We deduce that $A^H\mathbf{x} = 0$, thus $\mathbf{x} \in \mathcal{N}(A^H)$. We have shown that any $\mathbf{x} \in \mathcal{N}(P)$ is also in $\mathcal{N}(A^H)$, hence $\mathcal{N}(P) \subseteq \mathcal{N}(A^H)$. On the other hand, any $\mathbf{y} \in \mathcal{N}(A^H)$ satisfies $\mathbf{y} \in \mathcal{N}(P)$ because $A^H\mathbf{y} = 0 \Rightarrow P\mathbf{y} = A(A^H A)^{-1}(A^H\mathbf{y}) = 0$. Consequently, we have $\mathcal{N}(P) = \mathcal{N}(A^H) = \mathcal{R}(A)^\perp$. Because A is a $m \times n$ matrix, it maps a n dimensional vector space into a m-dimensional one, and thus $\mathcal{R}(A)$ is a subset of a m-dimensional vector space. We have $m = \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$ and since $\dim \mathcal{R}(A) = \operatorname{rank}(A) = n$, we deduce that $\dim \mathcal{R}(A)^\perp = m - n$. It follows that $\dim \mathcal{N}(P) = m - n$. Note that P is a $m \times m$ matrix, hence by the rank-nullity theorem, we have

$$\operatorname{rank}(P) = m - \dim \mathcal{N}(P) = m - (m - n) = n.$$

3.48.

(i) Since $A \to A$ and $A \to A^T$ are both linear, it follows that P is linear also.

(ii) We have

$$\begin{aligned} P^2(A) &= \frac{P(A) + P(A)^T}{2} \\ &= \frac{\frac{A+A^T}{2} + \left(\frac{A+A^T}{2}\right)^T}{2} \\ &= \frac{A + A^T}{2} \quad \text{(because } (A^T)^T = A) \\ &= P(A). \end{aligned}$$

(iii) It suffices to show that for all $A, B \in M_n(\mathbb{R})$, we have

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

or equivalently,

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

Since $\operatorname{tr}(aA) = a\operatorname{tr}(A)$ for any $a \in \mathbb{R}$ and $A \in M_n(\mathbb{R})$. In short, we need to show

$$\begin{aligned} \operatorname{tr}(A^T B + A^T B^T) &= \operatorname{tr}(AB + A^T B) \\ \Leftrightarrow \operatorname{tr}(A^T B) + \operatorname{tr}(A^T B^T) &= \operatorname{tr}(AB) + \operatorname{tr}(A^T B) \\ \Leftrightarrow \operatorname{tr}(A^T B^T) &= \operatorname{tr}(AB). \end{aligned}$$

But the last equality is true because $\operatorname{tr}(A^T B^T) = \operatorname{tr}(B^T A^T) = \operatorname{tr}[(AB)^T] = \operatorname{tr}(AB)$ (taking transpose does not change elements on the diagonal, hence the trace remains the same). Thus $P^* = P$.

(iv) $A \in \mathcal{N}(P)$ if and only if $A + A^T = 0$, which is equivalent to $A \in \operatorname{Skew}_n(\mathbb{R})$.

(v) Obviously $\mathscr{R}(P) \subseteq \operatorname{Sym}_n(\mathbb{R})$, because if $A = (a_{ij})$ then $A^T = (a'_{ij})$ where $a'_{ij} = a_{ji}$, thus $(A + A^T) = (a^*_{ij})$, where $a^*_{ij} = a_{ij} + a'_{ij} = a_{ij} + a_{ji} = a^*_{ji}$, hence $\dfrac{A+A^T}{2}$ is symmetric for all A. It remains to show that any symmetric matrix X can be written as $X = \dfrac{A+A^T}{2}$ for some A. Indeed, just choose $A = X$, then apparently $X = \dfrac{2X}{2} = \dfrac{X+X^T}{2}$.

(vi) We have

$$\begin{aligned} \|A - P(A)\|_F &= \sqrt{\operatorname{tr}[(A-P(A))^T (A-P(A))]} \\ &= \sqrt{\operatorname{tr}\left(\frac{A^T - A}{2} \cdot \frac{A - A^T}{2}\right)} \\ &= \sqrt{\frac{\operatorname{tr}(A^T A) - \operatorname{tr}((A^T)^2) - \operatorname{tr}(A^2) + \operatorname{tr}(AA^T)}{4}} \end{aligned}$$

Note that $\operatorname{tr}(AA^T) = \operatorname{tr}(A^T A)$ and $\operatorname{tr}((A^T)^2) = \operatorname{tr}((A^2)^T) = \operatorname{tr}(A^2)$. Therefore,

$$\|A - P(A)\|_F = \sqrt{\frac{\operatorname{tr}(A^T A) - \operatorname{tr}(A^2)}{2}}.$$

3.50. We want to find r, s such that $rx_i^2 + sy_i^2 = 1$ for each i. That is, we want to find the best approximate solution for $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{pmatrix} x_1^2 & y_1^2 \\ x_2^2 & y_1^2 \\ \vdots & \vdots \\ x_n^2 & y_n^2 \end{pmatrix}, \mathbf{x} = \begin{bmatrix} r \\ s \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$