Problem 1 (Question 1.3)

let U be an orthogonal, upper triangular matrix with column vectors $u_{*1}, u_{*2}, ..., u_{*n}$. For each vector u_{*j}, we have

$$u_{ij} = 0 \quad \text{if } i > j$$

Since U is orthogonal, we see that

$$u_{*1}^T u_{*j} = \begin{bmatrix} u_{11} & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} u_{1j} \\ u_{2j} \\ \vdots \\ u_{nj} \end{bmatrix} = 0 \implies u_{11}u_{1j} = 0 \implies u_{1j} = 0 \quad \text{for all } j$$

The last implication follows because $u_{*1}^T u_{*1} = u_{11}^2 = 1 \neq 0$.

To prove the rest, we proceed by induction.
We claim that $u_{jj} \neq 0$ for all j, and $u_{ij} = 0$ for all $i < j$.
We perform induction on i.

Base case: True as shown above
Inductive hypothesis: suppose the claim holds for all $i < k < j$
Inductive step: for $i = k$, note that

$$u_{*k}^T u_{*j} = \begin{bmatrix} 0 & \cdots & u_{kk} & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ u_{kj} \\ u_{k+1j} \\ \vdots \\ 0 \end{bmatrix} = 0$$

By the inductive hypothesis. Multiplying out, we see

$$u_{kk}u_{kj} = 0$$

Since $u_{*k}^T u_{*k} = u_{kk}^2 = 1 \neq 0$, it follows that $u_{kj} = 0$ for all j and $u_{kk} \neq 0$.

Hence, by induction, $u_{ij} = 0$ when $i \neq j$, and $u_{ii} \neq 0$ for all i. This proves that U is a diagonal matrix with diagonal elements being 1.

Problem 2 (Question 1.5)

1. $||x||_C \geq 0$:
Since $||Cx||$ is a vector norm, it follows that $||x||_C = ||Cx|| \geq 0$ for all x.
Suppose $||x||_C = ||Cx|| = 0$, then $Cx = 0$.
Since C has full rank, by rank nullity theorem, its null space has dimension 0.
In other words, $Cx = 0 \implies x = 0$.
Hence, $||x||_C = 0$ iff $x = 0$.

2. $||\alpha x||_C = \alpha||x||_C$:

$$\begin{aligned}||\alpha x||_C &= ||\alpha Cx|| \\ &= \alpha||Cx|| \\ &= \alpha||x||_C\end{aligned}$$

3. $||x + y||_C \leq ||x||_C + ||y||_C$:

$$\begin{aligned}||x + y||_C &= ||C(x + y)|| \\ &= ||Cx + Cy|| \\ &\leq \alpha||Cx|| + ||Cy|| \\ &= ||x||_C + ||y||_C\end{aligned}$$

Problem 3 (Question 1.7)

Note that for Frobenius norm:

$$||xy^*||_F^2 = \mathrm{Tr}((xy^*)^*(xy^*)) = \mathrm{Tr}(yx^*xy^*) = (x^*x)\mathrm{Tr}(yy^*) = ||x||_2^2||y||_2^2$$

On the other hand, for operator two norms

$$||xy^*||_2^2 = \lambda_{\max}(yx^*xy^*) = x^*x\lambda_{\max}(yy^*) = ||x||_2^2\lambda_{\max}(yy^*)$$

Since yy^* is Hermitian, we have

$$\lambda_{\max}(yy^*) = \max_{v^*v=1} v^*yy^*v = \max_{v^*v=1}(v^*y)^2 \leq \max_{v^*v=1}||v||_2^2||y||_2^2 \leq ||y||_2^2$$

Inequality is attained when v is in the same direction as y. Hence, we have

$$||xy^*||_2^2 = ||x||_2^2\lambda_{\max}(yy^*) = ||x||_2^2||y||_2^2$$

This proves the equality

$$||xy^*||_F = ||xy^*||_2 = ||x||_2||y||_2$$

Problem 4 (Question 1.13)

We will prove the case where A is Hermitian positive definite. The case for symmetric positive definition follows immediately.
Let A be Hermitian positive definite (i.e. $\overline{A}^T = A$)
Define $\langle\ ,\ \rangle : \mathcal{B} \times \mathcal{B} \to \mathbb{C}$ given by

$$\langle x, y\rangle = \overline{x}^T Ay$$

We claim that $\langle\ ,\ \rangle$ is an inner product.

1. Conjugate symmetry

$$\langle x, y\rangle = \overline{x}^T Ay = (\overline{x}^T Ay)^T = y^T A^T \overline{x} = \overline{\overline{y^T A^T \overline{x}}} = \overline{y^T \overline{A}^T x} = \overline{\overline{y}^T Ax} = \overline{\langle y, x\rangle}$$

2. Linearity

$$\langle x, y+z\rangle = \overline{x}^T A(y+z) = \overline{x}^T Ay + \overline{x}^T Az = \langle x, y\rangle + \langle x, z\rangle$$

3. Homogeneity

$$\langle \alpha x, y\rangle = \overline{\alpha x}^T Ay = \overline{\alpha x}^T Ay = \overline{\alpha}\langle x, y\rangle$$

4. Positive definiteness

By definition of Hermitian positive definite, we have

$$\langle x, y\rangle = \overline{x}^T Ay \geq 0$$

And $\langle x, y\rangle = 0$ if and only if $x = 0$, since otherwise $\overline{x}^T Ay$ is strictly positive.

Suppose, on the other hand $\langle\ ,\ \rangle$ is an inner product.
Let $b_1, b_2, ...b_n$ be a basis of $\mathcal{B}$. Define matrix A by

$$A_{i,j} = \langle b_i, b_j\rangle$$

Then A is conjugate symmetric since

$$A_{i,j} = \langle b_i, b_j\rangle = \overline{\langle b_j, b_i\rangle} = \overline{A_{j,i}}$$

And furthermore positive definite
To prove this, let $x = \sum_{i=1}^n \lambda_i b_i \in \mathcal{B}$, then

$$\langle x, x\rangle = \langle\sum_{i=1}^n \lambda_i b_i, \sum_{i=1}^n \xi_i b_i\rangle = \sum_{i=1}^n \sum_{j=1}^n \overline{\lambda_i}\langle b_i, b_j\rangle\lambda_j = \overline{\lambda}^T A\lambda > 0$$

For all $\lambda \neq 0$. Hence, A is Hermitian positive definite.

Problem 5 (Question 1.14)

1. $||x||_2 \leq ||x||_1 \leq \sqrt{n}||x||_2$:
First note that

$$\sum_{i=1} |x_i|^2 \leq \left(\sum_{i=1}^{n} |x_i|\right)^2 \implies \sqrt{\sum_{i=1} |x_i|^2} \leq \sum_{i=1}^{n} |x_i| \implies ||x||_2 \leq ||x||_1$$

Next, by Cauchy Schwartz inequality, we have

$$||x||_1 = \langle |x_1|, \cdots, |x_n| \rangle \cdot \langle 1, \cdots, 1 \rangle \leq \sqrt{\sum_{i=1} |x_i|^2} \sqrt{\sum_{i=1}^{n} 1} = \sqrt{n}||x||_2$$

2. $||x||_\infty \leq ||x||_2 \leq \sqrt{n}||x||_\infty$:
Note that

$$||x||_\infty = \max_i |x_i| = \sqrt{\max_i |x_i|^2} \leq \sqrt{\sum_{i=1}^{n} |x_i|^2} = ||x||_2$$

Next,

$$||x||_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} \leq \sqrt{\sum_{i=1}^{n} \max_i |x_i|^2} = \sqrt{n} \max_i |x_i| = \sqrt{n}||x||_\infty$$

3. $||x||_\infty \leq ||x||_1 \leq n||x||_\infty$:
Note that

$$||x||_\infty = \max_i |x_i| \leq \sum_{i=1}^{n} |x_i| = ||x||_1$$

Next,

$$||x||_1 = \sum_{i=1}^{n} |x_i| \leq \sum_{i=1}^{n} \max_i |x_i| = n \max_i |x_i| = n||x||_\infty$$

Problem 6 (Question 1.15)

Note that we can rewrite the operator norm as

$$||A||_{mn} = \max_{||x||_n=1} ||Ax||_m$$

1. Positive definiteness:
Since $||Ax||_m \geq 0$, it follows that $||A||_{mn} \geq 0$.
Suppose $||A||_{mn} = 0$, then $||Ax||_m = 0$.
This implies that $Ax = 0$ for all $||x||_n = 1$, it follows that $A = 0$.

2. Homogeneity:

$$||\alpha A||_{mn} = \max_{||x||_n=1} ||\alpha Ax||_m = \alpha \max_{||x||_n=1} ||Ax||_m = \alpha ||A||_{mn}$$

3. Triangle inequality

$$\begin{aligned} ||A+B||_{mn} &= \max_{||x||_n=1} ||(A+B)x||_m \\ &\leq \max_{||x||_n=1} ||Ax||_m + ||Bx||_m \\ &\leq \max_{||x||_n=1} ||Ax||_m + \max_{||x||_n=1} ||Bx||_m \\ &= ||A||_{mn} + ||B||_{mn} \end{aligned}$$

Hence, the operator norm is indeed a norm.

Problem 7 (Question 1.16)

1. Lemma 1.7.1:
The case for operator norm is trivial, since

$$||A||_{mn} = \max_{x\neq 0} \frac{||Ax||_m}{||x||_n} \implies \frac{||Ax||_m}{||x||_n} \leq ||A||_{mn} \implies ||Ax||_m \leq ||A||_{mn}||x||_n$$

Suppose the norm is the Frobenius norm, let

$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \dots & a_n \\ | & | & & | \end{bmatrix}$$

Then

$$||Ax||_2^2 = \sum_{i=1}^{m} (a_i^T x)^2 \leq \sum_{i=1}^{m} ||a_i||^2||x||^2 = ||x||_2^2||A||_F^2 \implies ||Ax||_2 \leq ||A||_F||x||_2$$

This proves the claim.

2. Lemma 1.7.2:
Let $A : \mathbb{R}^n \to \mathbb{R}^k, B : \mathbb{R}^m \to \mathbb{R}^n$, consider $AB : \mathbb{R}^m \to \mathbb{R}^k$
Consider the operator norm:

$$||AB||_{mk} = \max_{x\neq 0} \frac{||ABx||_k}{||x||_m} \leq \max_{x\neq 0} \frac{||A||_{nk}||Bx||_n}{||x||_m} \leq ||A||_{nk}||B||_{mn}$$

Let a_i, b_i be the columns of A, B, respectively. Consider the Frobenius norm:

$$||AB||_F^2 = \sum_{i,j} (a_i^T b_j)^2 \leq \sum_{i,j} ||a_i||^2||b_j||^2 = \sum_i ||a_i||^2 \sum_j ||b_j||^2 = ||A||_F||B||_F$$

This proves the claim.

3. Lemma 1.7.4:
Note that it suffice showing that for any utinary matrix Q, we have

$$||QA|| = ||A||$$

First consider the case for operator norm.

$$||QA||_2 = \sqrt{\lambda_{\max}(QA)^*(QA)} = \sqrt{\lambda_{\max} A^*Q^*QA} = \sqrt{\lambda_{\max} A^*A} = ||A||_2$$

Consider the Frobenius norm.

$$||QA||_F = \sqrt{\mathrm{Tr}((QA)^*(QA))} = \sqrt{\mathrm{Tr}(A^*(QQ)A) = \mathrm{Tr}(A^*A)} = ||A||_F$$

4. Lemma 1.7.5:
Let $\mathbf{a}_i$ be the row of A, then

$$||A||_\infty = \max_{\substack{||\mathbf{x}||_\infty=1 \\ i}} |\mathbf{a}_i\mathbf{x}| \le \max_{\substack{||\mathbf{x}||_\infty=1 \\ i}} ||\mathbf{a}_i||_\infty ||\mathbf{x}||_\infty = \max_i ||\mathbf{a}_i||_\infty = \max_i \sum_{j=1}^n |a_{ij}|$$

Note that the bound is attainable by choosing $\mathbf{x}$ such that its kth entry is $\text{sign}(a_{ik})$. This proves the claim.

5. Lemma 1.7.6:
Let $\mathbf{a}_i$ be the columns of A, then

$$||Ax||_1 = \left|\left| \sum_{i=1}^n x_i \mathbf{a}_i \right|\right|_1 \le \sum_{i=1}^n |x_i| ||\mathbf{a}_i||_1$$

By triangle inequality. Let $\mathbf{a}_j$ be the column with maximum column sum, then

$$\begin{aligned}
||A||_1 = \max_{||x||_1=1} ||Ax||_1 &= \max_{||x||_1=1} \sum_{i=1}^n |x_i| ||\mathbf{a}_i||_1 \\
&= \max_{||x||_1=1} \left(|x_j| ||\mathbf{a}_j||_1 + \sum_{i\neq j} |x_i| ||\mathbf{a}_i||_1 \right) \\
&= \max_{||x||_1=1} \left((1 - \sum_{i\neq j} |x_i|) ||\mathbf{a}_j||_1 + \sum_{i\neq j} |x_i| ||\mathbf{a}_i||_1 \right) \\
&= \max_{||x||_1=1} \left(||\mathbf{a}_j||_1 + \sum_{i\neq j} (||\mathbf{a}_j|| - ||\mathbf{a}_i||) |x_i| \right) \\
&\le \max_{||x||_1=1} ||\mathbf{a}_j||_1
\end{aligned}$$

Equality can be attained by choosing $x = \mathbf{e}_j$. This proves the claim.

6. Lemma 1.7.8:
Since

$$||A||_2 = \sqrt{\lambda_{\max}(A^T A)} \qquad ||A^T||_2 = \sqrt{\lambda_{\max}(AA^T)}$$

It suffice proving that

$$\lambda_{\max}(A^T A) = \lambda_{\max}(AA^T)$$

Let $\lambda \neq 0$ be an eigenvalue of $A^T A$, and x be its corresponding eigenvector, then

$$(AA^T)Ax = A(A^T A)x = \lambda Ax$$

This means that λ is also an eigenvalue of AA^T.
The other way around can be shown by the same argument.
Hence, AA^T and $A^T A$ has the same set of eigenvalues.

This implies that

$$\lambda_{\max}(A^T A) = \lambda_{\max}(AA^T)$$

Thus proving the claim.

7. Lemma 1.7.10:
Recall that

$$||x||_2 \leq ||x||_1 \leq \sqrt{n}||x||_2 \implies \frac{1}{\sqrt{n}||x||_2} \leq \frac{1}{||x||_1} \leq \frac{1}{||x||_2}$$

Likewise

$$||Ax||_2 \leq ||Ax||_1 \leq \sqrt{n}||Ax||_2$$

This implies that

$$\frac{||Ax||_2}{\sqrt{n}||x||_2} \leq \frac{||Ax||_1}{||x||_1} \leq \frac{\sqrt{n}||Ax||_2}{||x||_2}$$

Taking maximum on both sides, we see that

$$n^{-\frac{1}{2}}||A||_2 \leq ||A||_1 \leq n^{\frac{1}{2}}||A||_2$$

Problem 7 (Question 3.9)

1. $$(A^TA)^{-1} = (V\Sigma^TU^TU\Sigma V^T)^{-1} = V(\Sigma^T\Sigma)^{-1}V^T$$

2. $$(A^TA)^{-1}A^T = V(\Sigma^T\Sigma)^{-1}V^T(V\Sigma^TU^T) = V\Sigma^{-1}U^T$$

3. $$A(A^TA)^{-1} = (U\Sigma V^T)V(\Sigma^T\Sigma)^{-1}V^T = U(\Sigma^T)^{-1}V^T$$

4. $$A(A^TA)^{-1}A^T = U\Sigma V^TV\Sigma^{-1}U^T = I_n$$