MA398 Matrix Analysis and Algorithms: Exercise Sheet 2

1. Create a Python function **p_norm** that can calculate the p-norm and the inner product of two vectors. Hence, prove that the function **p_norm** correctly implements the p-norm definition and the function inner product correctly implements the standard inner product.

2. Implement a function that verifies whether a given square matrix is unitary. Hence, prove that the function **is_unitary** correctly implements the definition of a unitary matrix. Also, give an intuitive explanation of why this condition ensures the columns of Q are orthonormal.

3. Prove the following Theorem:

   "If $A \in \mathbb{C}^{n\times n}$ is Hermitian then there is a unitary Q and a real $\Lambda \in \mathbb{R}^{n\times n}$ such that $A = Q\Lambda Q^*$."

   Hint: Remember that a Hermitian matrix is one that is equal to its own conjugate transpose, i.e., $A = A^*$. You might want to utilize the Spectral Theorem for Hermitian matrices in your proof, which states that every Hermitian matrix can be diagonalized by a unitary matrix.

   Answer: Firstly, we need to understand what Hermitian means in the context of complex matrices. A matrix $A \in \mathbb{C}^{n\times n}$ is Hermitian if it is equal to its own conjugate transpose, i.e., $A = A^*$.

   Now, let us prove the theorem:

   Proof:

   The Spectral Theorem for Hermitian matrices states that a Hermitian matrix A can be diagonalized by a unitary matrix, i.e., there exists a unitary matrix Q and a diagonal matrix Λ such that $A = Q\Lambda Q^*$.

   We need to show that the diagonal elements of Λ are real. This can be deduced from the property of Hermitian matrices that their eigenvalues are real.

   Consider the equation $Ax = \lambda x$, where x is an eigenvector corresponding to the eigenvalue λ.

   Taking the complex conjugate transpose of both sides, we get $(Ax)^* = \lambda^* x^*$.

   This becomes $x^*A^* = \lambda^* x^*$.

   Since A is Hermitian, we can replace A^* with A. This gives us $x^*A = \lambda^* x^*$.

   But we know from the original eigenvalue equation that $x^*A = \lambda x^*$. Therefore, we can equate $\lambda = \lambda^*$, which implies that λ is real.

   Therefore, all the diagonal elements of Λ, which are the eigenvalues of A, are real. Thus, we have shown that if A is Hermitian, then there exists a unitary matrix Q and a real diagonal matrix Λ such that $A = Q\Lambda Q^*$.

   This concludes the proof.

4. Implement a Python function **is_normal(A)** that checks whether a given square matrix A is normal. Recall from the lecture notes that a matrix A is normal if it satisfies $A^*A = AA^*$. The function should take as input a NumPy array A and return a Boolean value (True or False).

5. (Geometric series for matrices) Let $\|\cdot\|$ be a matrix norm on $\mathbb{C}^{n\times n}$. Assume that $\|X\| < 1$ for some $X \in \mathbb{C}^{n\times n}$. Show that,

(a) $I - X$ is invertible with $(I - X)^{-1} = \sum_{i=0}^{\infty} X^i$,

(b) $\|(I - X)^{-1}\| \le (1 - \|X\|)^{-1}$.

Answer: For every $m \in \mathbb{N}$

$$\|\sum_{i=0}^{m} X^i\| \le \sum_i \|X^i\| \le \sum_i \|X\|^i = \frac{1 - \|X\|^{m+1}}{1 - \|X\|} \tag{1}$$

where the properties of a matrix norm were used. The right hand side converges as $m \to \infty$ since $\|X\| < 1$ (geometric series). An analogous argument shows that $n \mapsto \sum_{i=0}^{n} X^i$ is a Cauchy sequence. As a consequence, $\sum_{i=0}^{\infty} X^i$ exists and $X^i \to 0$ as $i \to \infty$ in $\mathbb{C}^{n\times n}$. We infer that

$$(I - X)\sum_{i=0}^{\infty} X^i = \lim_{m\to\infty} (I - X)\sum_{i=0}^{m} X^i = \lim_{m\to\infty} (I - X^{m+1}) = I.$$

Letting $m \to \infty$ in the right hand side of (1) we obtain the second claim:

$$\|(I - X)^{-1}\| = \lim_{m\to\infty} \|\sum_{i=0}^{m} X^i\| \le \lim_{m\to\infty} \frac{1 - \|X\|^{m+1}}{1 - \|X\|} = (1 - \|X\|)^{-1}.$$

6. (Cholesky factorisation) If $A \in \mathbb{C}^{n\times n}$ is Hermitian and positive definite then there exists a unique upper triangular matrix $R \in \mathbb{C}^{n\times n}$ with (real and) positive diagonal elements such that $A = R^*R$. (Hint: Induction on n.)

   Answer: **Base case:** For $n = 1$ where A is a 1×1 matrix, we can construct R as a 1×1 matrix with $R_{11} = \sqrt{A_{11}}$. This satisfies the conditions of an upper triangular matrix with real and positive diagonal elements, and $R^*R = A$.

   Inductive step: Assume that the theorem holds for $n = k$ and let us prove it for $n = k+1$.

   Consider an $(k + 1) \times (k + 1)$ Hermitian positive definite matrix A. We can write A in block form as:

$$\begin{bmatrix} A' & b \\ b^* & a \end{bmatrix}$$

   where A' is a $k \times k$ submatrix, b is a $k \times 1$ column vector, b^* is its conjugate transpose ($1 \times k$ row vector), and a is a scalar.

   Since A is Hermitian positive definite, the Schur complement $a - b^*A'^{-1}b$ must be positive. Therefore, B is also positive definite.

   By the induction hypothesis, there exists an upper triangular matrix R' of size $k \times k$ with real and positive diagonal elements such that $A' = R'^*R'$.

   Now, define a new matrix R as:

$$R = \begin{bmatrix} R' & v \\ 0 & r \end{bmatrix}$$

   where v is a $k \times 1$ column vector and r is a scalar.

   We want to find v and r such that R is an upper triangular matrix with real and positive diagonal elements, and $A = R^*R$.

   Using the block matrix multiplication, we have:

$$R^*R = \begin{bmatrix} R'^* & 0 \\ v^* & r^* \end{bmatrix} \begin{bmatrix} R' & v \\ 0 & r \end{bmatrix} = \begin{bmatrix} R'^*R' & R'^*v \\ v^*R' & v^*v + r^*r \end{bmatrix}$$

   We want this to be equal to A. Equating corresponding entries, we have:

$$R'^*R' = A', \quad R'^*v = b, \quad v^*R' = b^*, \quad v^*v + r^*r = a$$

From the first equation, we know that $R'^*R' = A'$. From the second equation, we can solve for v as $v = (R'^*)^{-1}b$, where $(R'^*)^{-1}$ is the inverse of R'^* (since R' is invertible).

Now, let's examine the equation $v^*v + r^*r = a$, where $v = (R'^*)^{-1}b$ and R' is the upper triangular matrix that satisfies $A' = R'^*R'$.

Substituting the expression for v, we have:

$$\begin{aligned} v^*v + r^*r &= ((R'^*)^{-1}b)^*((R'^*)^{-1}b) + r^*r \\ &= b^*(R'^*)^{-T}(R'^*)^{-1}b + r^*r. \end{aligned}$$

Since R' is an upper triangular matrix, its conjugate transpose $(R'^*)^{-T}$ is a lower triangular matrix. Let's denote $(R'^*)^{-T}$ as L. Then the equation becomes:

$$b^*L^TLb + r^*r$$

Now, we need to choose r such that $b^*L^TLb + r^*r = a$. In order to do that, let's choose r as $r = \sqrt{a - b^*L^TLb}$, where we take the positive square root since $a - b^*L^TLb$ is positive (as A is positive definite). With this choice of r, the equation becomes:

$$b^*L^TLb + (\sqrt{a - b^*L^TLb})^2 = a$$

Hence, we have found a suitable choice of v and r such that $v^*v + r^*r = a$. This completes the proof.

7. (Norms)

   (a) Is $\|A\|_{\max} = \max_{i,j} |a_{ij}|$ a matrix norm?

   Answer: In order to show that the function $|A|_{\max} = \max i, j|a_{ij}|$ defines a matrix norm, we need to check if it satisfies the following four properties:

      i. $|A|\max \geq 0$, and $|A|_{\max} = 0$ if and only if A is the zero matrix.
      ii. $|\alpha A|_{\max} = |\alpha||A|_{\max}$ for all scalars α.
      iii. $|A + B|_{\max} \leq |A|_{\max} + |B|_{\max}$ (Triangle inequality).
      iv. $|AB|_{\max} \leq |A|_{\max}|B|_{\max}$ (Submultiplicativity).

   We can easily verify the first three properties:

      i. It's clear that $|A|_{\max}$ is always nonnegative, since it's the maximum of absolute values. It's also clear that $|A|_{\max} = 0$ if and only if all entries of A are zero, i.e., A is the zero matrix.
      ii. If we multiply A by a scalar α, all entries a_{ij} of A get multiplied by α, so the maximum absolute value gets multiplied by $|\alpha|$.
      iii. If we add two matrices A and B, any entry of the resulting matrix is the sum of corresponding entries in A and B. The absolute value of a sum is always less than or equal to the sum of absolute values, so the maximum absolute value in $A + B$ is less than or equal to the sum of maximum absolute values in A and B.

   However, the function $|A|_{\max}$ fails to satisfy the submultiplicativity property, i.e., $|AB|_{\max} \leq |A|_{\max}|B|_{\max}$ does not necessarily hold. For example, consider two 2×2 matrices $A = \begin{pmatrix} 1 & 1 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 1 0 & 1 \end{pmatrix}$. Then, $AB = \begin{pmatrix} 1 & 2 0 & 1 \end{pmatrix}$, so $|AB|_{\max} = 2$, while $|A|_{\max} = |B|_{\max} = 1$, and hence $|A|_{\max}|B|_{\max} = 1 < |AB|_{\max}$.

   So, the function $|A|_{\max}$ does not define a matrix norm because it doesn't satisfy all four required properties.

   (b) Show that $\|uv^*\|_2 = \|u\|_2\|v\|_2$ for all $u, v \in \mathbb{C}^n$. Does this also hold true if $\|\cdot\|_2$ is replaced by the Frobenius norm $\|\cdot\|_F$?

Answer: Let $u, v \in \mathbb{C}^n$ be arbitrary. We have:

$$\begin{aligned}
|uv^|2 &= \sqrt{\sum_{i=1}^{n} |(uv^*)_i|^2} \\
&= \sqrt{\sum_{i=1}^{n} |u_i v_i^*|^2} && \text{(because } (uv^*)i = u_i v_i^*\text{)} \\
&= \sqrt{\sum_{i=1}^{n} |u_i|^2 |v_i|^2} && \text{(because } |u_i v_i^*| = |u_i||v_i^*| = |u_i||v_i|\text{)} \\
&= \sqrt{\left(\sum_{i=1}^{n} |u_i|^2\right)\left(\sum_{i=1}^{n} |v_i|^2\right)} \\
&= |u|_2 |v|_2
\end{aligned}$$

Therefore, we have shown that $|uv^*|_2 = |u|_2|v|_2$ for all $u, v \in \mathbb{C}^n$.

We want to show that for all $u, v \in \mathbb{C}^n$, $|uv^*|_F = |u|_F|v|_F$. Since uv^* is a scalar (an 1×1 matrix), its Frobenius norm is simply the magnitude of the scalar. Therefore, we can rewrite the equation as:
(1) $|uv^*| = |u|_F|v|_F$
Now, let's compute the left-hand side of the equation:
(2) $|uv^*| = |u_1v_1^* + u_2v_2^* + \cdots + u_nv_n^*|$
To compute the right-hand side of the equation, we first find the Frobenius norms of u and v:
(3) $|u|_F = \sqrt{|u_1|^2 + |u_2|^2 + \cdots + |u_n|^2}$
(4) $|v|_F = \sqrt{|v_1|^2 + |v_2|^2 + \cdots + |v_n|^2}$
Then, we compute their product:
(5) $|u|_F|v|_F = \sqrt{(|u_1|^2 + |u_2|^2 + \cdots + |u_n|^2)(|v_1|^2 + |v_2|^2 + \cdots + |v_n|^2)}$
Now, let's square both sides of the equation (1) to simplify the expression:
(6) $|uv^*|^2 = (|u|_F|v|_F)^2$
Substituting the expressions for $|uv^*|$ and $|u|_F|v|_F$ from equations (2) and (5) into equation (6), we have:
(7) $(u_1v_1^* + u_2v_2^* + \cdots + u_nv_n^*)(u_1v_1^* + u_2v_2^* + \cdots + u_nv_n^*)^* = (|u_1|^2 + |u_2|^2 + \cdots + |u_n|^2)(|v_1|^2 + |v_2|^2 + \cdots + |v_n|^2)$
Expanding the left-hand side, we get:
(8) $\sum_{i=1}^n u_iv_i^* \sum_{i=1}^n u_i^*v_i = \sum_{i=1}^n |u_i|^2|v_i|^2$
And the right-hand side is:
(9) $\sum_{i=1}^n |u_i|^2|v_i|^2$
Since the left-hand side (equation 8) and the right-hand side (equation 9) are equal, we can conclude that:
$|uv^*|_F = |u|_F|v|_F$
for all $u, v \in \mathbb{C}^n$.

(c) Let $p \in [1, \infty)$. Prove the following statement:

$$\|x\|_\infty \le \|x\|_p \le \sqrt[p]{n}\|x\|_\infty \quad \forall x \in \mathbb{C}^n.$$

Answer: Firstly, we will show that $|x|_\infty \le |x|_p$.

The ∞-norm of a vector x is defined as the maximum absolute value of its elements, i.e. $|x|_\infty = \max_i |x_i|$. The p-norm is defined as $|x|_p = (\sum_i |x_i|^p)^{\frac{1}{p}}$. Since $|x_i|^p \leq (\max_i |x_i|)^p = |x|_\infty^p$ for all i and $p \geq 1$, summing over all elements gives $\sum_i |x_i|^p \leq n|x|_\infty^p$, and hence $|x|_p = (\sum_i |x_i|^p)^{\frac{1}{p}} \geq |x|_\infty$.
Secondly, we will show that $|x|_p \leq \sqrt[p]{n}|x|_\infty$.
By definition, $|x|_p = (\sum_i |x_i|^p)^{\frac{1}{p}}$. We have $|x_i|^p \leq |x|_\infty^p$ for all i, so $\sum_i |x_i|^p \leq \sum_i |x|_\infty^p = n|x|_\infty^p$. Hence, $|x|p = (\sum i|x_i|^p)^{\frac{1}{p}} \leq (n|x|_\infty^p)^{\frac{1}{p}} = \sqrt[p]{n}|x|_\infty$.
Therefore, we have $|x|_\infty \leq |x|_p \leq \sqrt[p]{n}|x|\infty$ for all $x \in \mathbb{C}^n$ and $p \in [1, \infty)$.