Undergraduate Research Opportunity
Programme in Science

UNITARY SIMILARITIES AND
SCHUR’S THEOREM

Wang Fei

Supervisor: Dr. Victor Tan

Department of Mathematics
National University of Singapore
2001/2002
Index of Symbols

$k$, $\alpha$, complex number

$\Re(z)$, real part of a complex number

$\Im(z)$, imaginary part of a complex number

$\mathbb{C}^n$, space of $n$-tuples of complex numbers

$v$, column vector

$v^T$, row vector

$0$, zero vector

$A$, matrix

$m \times n$, size of a matrix with $m$ rows and $n$ columns

$O$, zero matrix

$I_n$, identity matrix of size $n$

$J_n$, unit matrix of size $n$

$\overline{A}$, complex conjugate of a matrix

$A^T$, transpose of a matrix

$A^H$, hermitian transpose of a matrix

$A^{-1}$, inverse of a square matrix

$\|u\|$, hermitian norm of a vector

$\langle u, v \rangle$, hermitian inner product of two vectors

$M(F)$, matrix space over the field $F$
CONTENTS

CHAPTER 1: UNITARY MATRICES AND SCHUR’S THEOREM 1
  1.1 The Hermitian Inner Product 1
  1.2 Unitary Matrices 5
  1.3 Schur’s Theorem 9
  1.4 Cayley-Hamilton Theorem 16

CHAPTER 2: APPLICATIONS OF SCHUR’S THEOREM 19
  2.1 Hermitian Matrices 19
  2.2 Normal Matrices 24
  2.3 Invariant Subspaces 29
  2.4 Simultaneous Diagonalization 35
  2.5 The Frobenius Norm 37

CHAPTER 3: THE LIMIT OF $A^m$ 46
  3.1 Gerschgorin’s Theorem 46
  3.2 The Limit of $A^m$ For $\rho(A)\neq 1$ 51
  3.3 The Limit of $A^m$ For $\rho(A)=1$ 60
  3.4 Probability Vectors and Transition Matrices 65
CHAPTER 1

UNITARY MATRICES AND SCHUR’S THEOREM

1.1. The Hermitian Inner Product

Recall that in the real space, the Euclidean inner product is defined by \( u \cdot v = \sum_{i=1}^{n} u_i v_i = u^T v \), and the norm of vector is \( \| u \| = \sqrt{u \cdot u} \). However, we cannot use these definitions directly for complex space, because they are not well defined. For example, take \( u = [1 \ i]^T \), we will obtain \( \| u \| = \sqrt{1^2 + i^2} = 0 \). Then \( u \) would be a non-zero vector with zero length, and \( \mathbb{C}^n \) would be no longer an inner product space. In this section, we will talk about how the inner product is well defined and develop some properties and theorems of vector space over the complex field.

Refer to the complex set; the norm of \( z \in \mathbb{C} \) is given by \( |z| = \sqrt{z \cdot \overline{z}} \). So we may have the similar definition of vector \( u \in \mathbb{C}^n \). Now, we introduce the “complex conjugate” for vector and matrix with complex entries first, and then give the definition of norm for vectors in \( \mathbb{C}^n \).
Definition 1.1.1 Suppose $A$ is a matrix with complex entries. The \textit{hermitian transpose} of $U$ is defined by $A^H = \overline{(A)^T}$, where $\overline{A}$ is the matrix whose entries are the complex conjugates of the corresponding entries in $A$.

\textbf{Remark.} If $A$ is a real matrix, then $A^H = (\overline{A})^T = A^T$. Its hermitian transpose is actually the real transpose. In fact, they do have some similar properties. Suppose $A$ and $B$ are matrices of the same size. Then

a) $(A^H)^H = A$

b) $(A+B)^H = A^H + B^H$

c) $(kA)^H = \overline{k}A^H$

d) $(AB)^H = B^H A^H$.

Definition 1.1.2 Let $\mathbf{u} = [u_1 \quad u_2 \quad \cdots \quad u_n]^T$ and $\mathbf{v} = [v_1 \quad v_2 \quad \cdots \quad v_n]^T$ be vectors in $\mathbb{C}^n$. The \textit{hermitian inner product} of $\mathbf{u}$ and $\mathbf{v}$ is defined by

$$\langle \mathbf{u}, \mathbf{v} \rangle = \overline{u}_1 v_1 + \overline{u}_2 v_2 + \cdots + \overline{u}_n v_n = \mathbf{u}^H \mathbf{v}.$$ 

The \textit{hermitian norm} (or \textit{hermitian length}) of a vector $\mathbf{u} = [u_1 \quad u_2 \quad \cdots \quad u_n]^T$ in $\mathbb{C}^n$ is defined by

$$\| \mathbf{u} \| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle} = \sqrt{|u_1|^2 + |u_2|^2 + \cdots + |u_n|^2}.$$ 

According to the definitions, the inner product $\langle \mathbf{u}, \mathbf{v} \rangle$ can be a complex number, but the norm $\| \mathbf{u} \|$ must be real and nonnegative. The following are some properties of the hermitian inner product. They are quite similar to those in the real case.
Theorem 1.1.3 Let $u, v$ and $w$ be vectors in $\mathbb{C}^n$, and $k$ a complex number. Then

a) $\langle u, v \rangle = \overline{\langle v, u \rangle}$

b) $\langle u, kv \rangle = k \langle u, v \rangle$

c) $\langle ku, v \rangle = \overline{k} \langle u, v \rangle$

d) $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$.

e) $\langle u, u \rangle \geq 0$, and $\langle u, u \rangle = 0$ if and only if $u = 0$.

Proof.

a) Let $u = [u_1, u_2, \ldots, u_n]^T$, $v = [v_1, v_2, \ldots, v_n]^T$ and $w = [w_1, w_2, \ldots, w_n]^T$, then

\[
\overline{\langle v, u \rangle} = \overline{\sum_{i=1}^{n} v_i u_i} = \sum_{i=1}^{n} \overline{v_i} u_i = \sum_{i=1}^{n} v_i \overline{u_i} = \langle v, u \rangle
\]

b) $\langle u, kv \rangle = \sum_{i=1}^{n} v_i (ku_i) = k \sum_{i=1}^{n} v_i (u_i) = k \langle u, v \rangle$

c) $\langle ku, v \rangle = \overline{\langle v, ku \rangle} = \overline{k} \langle v, u \rangle = \overline{k} \overline{\langle u, v \rangle} = k \langle u, v \rangle$

d) $\langle u + v, w \rangle = \sum_{i=1}^{n} (u_i + v_i) w_i = \sum_{i=1}^{n} u_i w_i + \sum_{i=1}^{n} v_i w_i = \langle u, w \rangle + \langle v, w \rangle$. 


Theorem 1.1.4 (Cauchy-Schwarz Inequality) For any vectors \( u \) and \( v \) in \( \mathbb{C}^n \)

\[
|\langle u, v \rangle| \leq \| u \| \cdot \| v \|.
\]

**Proof.** If \( v = 0 \), then both sides become 0, it holds. Assume that \( v \neq 0 \), let \( k \) be a complex number, consider the following inequality:

\[
0 \leq \| u - kv \|^2 = \langle u - kv, u - kv \rangle = \langle u, u \rangle - \langle u, kv \rangle - \langle kv, u \rangle + \langle kv, kv \rangle = \| u \|^2 - k\langle u, v \rangle - \bar{k}\langle u, v \rangle + k\bar{k} \| v \|^2 = \| u \|^2 - 2\text{Re}(k\langle u, v \rangle) + |k|^2 \| v \|^2.
\]

It holds for every \( k \). In particular, take \( k = \frac{\langle u, v \rangle}{\| v \|^2} \), we have

\[
0 \leq \| u \|^2 - 2\text{Re}\left( \frac{\langle u, v \rangle}{\| v \|^2} \langle u, v \rangle \right) + \frac{\langle u, v \rangle^2}{\| v \|^2} = \| u \|^2 - 2\left( \frac{\langle u, v \rangle}{\| v \|^2} \right)^2 + \frac{\langle u, v \rangle^2}{\| v \|^2} = \| u \|^2 - \left( \frac{\langle u, v \rangle}{\| v \|^2} \right)^2.
\]

It follows that \( \| u \|^2 \| v \|^2 \geq |\langle u, v \rangle|^2 \). Moreover, the equality holds if and only if \( \| u - kv \| = 0 \), if and only if \( u \) is a scalar multiple of \( v \). \( \Box \)

Theorem 1.1.5 (Triangle Inequality) For any vectors \( u \) and \( v \) in \( \mathbb{C}^n \)

\[
\| u + v \| \leq \| u \| + \| v \|.
\]
**Proof.** By the Cauchy-Schwarz inequality (Theorem 1.1.4),

\[
\| u + v \|^2 = \| u \|^2 + \| v \|^2 + \langle u, v \rangle + \langle v, u \rangle = \| u \|^2 + \| v \|^2 + 2 \text{Re} \langle u, v \rangle \\
\leq \| u \|^2 + \| v \|^2 + 2 \| u \| \cdot \| v \| \\
= (\| u \| + \| v \|)^2.
\]

We are done, and the quality holds if and only if \( v = 0 \) or \( u = kv \) for some real \( k \). □

### 1.2. Unitary Matrices

For matrix with real entries, the orthogonal matrix \( A^{-1} = A^T \) plays an important role in the diagonalization problem. In the complex matrix space, a new type of matrix superseding the orthogonal matrix, called unitary matrix, will be discussed in this section.

**Definition 1.2.1** Let \( u, v \) be vectors in \( \mathbb{C}^n \). They are called **orthogonal** if \( \langle u, v \rangle = 0 \), and written as \( u \perp v \). The set of vectors \( \{u_1, u_2, \cdots, u_k\} \) is an **orthonormal** set if \( \langle u_i, u_j \rangle = 0 \) for each \( i \neq j \) and \( \| u_i \| = 1 \) for all \( i \).

**Definition 1.2.2** A square matrix \( U \) with complex entries is a **unitary matrix** if

\[
U^H U = U U^H = I.
\]

**Remark.** Let \( U \) be an \( n \times n \) matrix. Then \( U \) is unitary if and only if the column (row) vectors of \( U \) form an orthonormal basis of \( \mathbb{C}^n \) with respect to the hermitian inner product. That is, \( U \) is
invertible and \( U^{-1} = U^H \). Suppose that \( U \) and \( V \) are unitary. Then \( U \), \( U^H \), \( U^T \) and \( UV \) are also unitary.

**Definition 1.2.3** If \( A \) and \( B \) are square matrices of the same size, we say that \( B \) is **unitarily similar** to \( A \) if there exists a unitary matrix \( U \), such that \( B = U^H AU \). The relationship from \( A \) to \( U^H AU \) is called a **unitary transformation**.

**Remark.** Unitary transformations are special cases of similarity transformations, so all the results which hold for similarity transformations hold for unitary transformation as well. Note that the unitary transformation is also an equivalence relation. **Reflexivity:** by choosing \( U = I \), which is a unitary matrix, \( A \) is unitarily similar to \( A \) itself. **Symmetry:** \( B = U^H AU \) implies that \( A = UBU^H = (U^H)^H BU^H \), so \( A \) is unitarily similar to \( B \) if and only if \( B \) is unitarily similar to \( A \). **Transitivity:** suppose that \( B \) is unitarily similar to \( A \) and \( C \) is to \( B \), say \( B = U^H AU \) and \( C = V^H AV \) for some unitary matrices \( U \) and \( V \). Then \( C = (UV)^H A(UV) \) is unitary similar to \( A \). Therefore, the unitary transformation is an equivalence relation.

**Definition 1.2.4** A square matrix \( A \) is called **unitarily diagonalizable** if there exists a diagonal matrix which is unitarily similar to \( A \).

**Remark.** More theorems about unitary diagonalization will be introduced in chapter 2.

As \( U^T = U^H \) for real matrix, any orthonormal matrix in \( M(\mathbb{R}) \) with respect the real inner product is also a unitary matrix respect to the hermitian inner product. Given a basis for \( \mathbb{R}^n \), we can use the Gram-Schmidt process to find an orthonormal basis. Actually, the process can also be applied in the complex case, but it requires the replacement of the real inner product with the hermitian inner product.
Theorem 1.2.5 (QR decomposition) Let \( A \) be an invertible matrix. Then there exists a unitary matrix \( Q \) and an upper triangular \( R \), such that \( A = QR \).

**Proof.** \( A \) is invertible, so the columns of \( A \) form a linearly independent set \( \{u_1, u_2, \cdots, u_n\} \). Now apply the Gram-Schmidt process. Let \( v_1 = u_1 \), and \( v_j = u_j - \sum_{i=1}^{j-1} \langle u_j, q_i \rangle q_i \) for \( j = 2, 3, \cdots, n \), where \( q_j = \frac{v_j}{\|v_j\|} \). Then \( \{q_1, q_2, \cdots, q_n\} \) is an orthonormal basis and thus \( Q = [q_1 \ q_2 \ \cdots \ q_n] \) is a unitary matrix. Denote \( r_{ij} = \langle u_j, q_i \rangle \). Then \( u_j = \|v_j\|q_j + \sum_{i=1}^{j-1} r_{ij}q_i \).

Set

\[
R = \begin{bmatrix}
\|v_1\| & r_{12} & \cdots & r_{1n} \\
0 & \|v_2\| & \cdots & r_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \|v_n\| \\
\end{bmatrix}
\]

Then \( A = QR \). \( \Box \)

Theorem 1.2.6 A square matrix \( U \) is unitary if and only if \( \langle Ux, Uy \rangle = \langle x, y \rangle \) for all \( x \) and \( y \).

**Proof.** Let \( U \) be a unitary matrix. Then for any \( x \) and \( y \),

\[
\langle Ux, Uy \rangle = (Uy)^\dagger (Ux) = x^\dagger (U^\dagger U)y = x^\dagger y = \langle x, y \rangle.
\]
Conversely, suppose that $\langle \mathbf{U} \mathbf{x}, \mathbf{U} \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$ holds for all $\mathbf{x}$ and $\mathbf{y}$. In particular, choose $\mathbf{x} = \mathbf{e}_i$ and $\mathbf{y} = \mathbf{e}_j$, where $\mathbf{e}_i$ and $\mathbf{e}_j$ are the $i$th and $j$th column vectors of the identity matrix $\mathbf{I}$ respectively. Then for every $1 \leq i, j \leq n$,

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \langle \mathbf{U} \mathbf{e}_i, \mathbf{U} \mathbf{e}_j \rangle.$$

Note that $\langle \mathbf{e}_i, \mathbf{e}_j \rangle$ is the $(i, j)$ entry of the identity matrix $\mathbf{I}$, and $\langle \mathbf{U} \mathbf{e}_i, \mathbf{U} \mathbf{e}_j \rangle$ is the hermitian inner product of the $i$th and $j$th column of $\mathbf{U}$. So the column vectors of $\mathbf{U}$ form an orthonormal basis of $\mathbb{C}^n$. By the remark of Definition 1.2.2, $\mathbf{U}$ is unitary. 

\begin{flushright}
$\square$
\end{flushright}

**Theorem 1.2.7** A square matrix $\mathbf{U}$ is unitary if and only if $\| \mathbf{U} \mathbf{x} \| = \| \mathbf{x} \|$ for every $\mathbf{x}$ in $\mathbb{C}^n$.

**Proof.** If $\mathbf{U}$ is unitary, then $\| \mathbf{U} \mathbf{x} \|^2 = (\mathbf{U} \mathbf{x})^H (\mathbf{U} \mathbf{x}) = \mathbf{x}^H (\mathbf{U}^H \mathbf{U}) \mathbf{x} = \mathbf{x}^H \mathbf{x} = \| \mathbf{x} \|^2$.

Conversely, if $\| \mathbf{U} \mathbf{x} \| = \| \mathbf{x} \|$ always holds, then for every $\mathbf{y}$ and $\mathbf{z}$, $\| \mathbf{U} (\mathbf{y} + \mathbf{z}) \| = \| \mathbf{y} + \mathbf{z} \|,

$$\| \mathbf{U} (\mathbf{y} + \mathbf{z}) \|^2 = \langle \mathbf{U} \mathbf{y}, \mathbf{U} \mathbf{y} \rangle + \langle \mathbf{U} \mathbf{z}, \mathbf{U} \mathbf{z} \rangle + \langle \mathbf{U} \mathbf{y}, \mathbf{U} \mathbf{z} \rangle + \langle \mathbf{U} \mathbf{z}, \mathbf{U} \mathbf{y} \rangle$$

$$= \| \mathbf{y} \|^2 + \| \mathbf{z} \|^2 + 2 \text{Re} \langle \mathbf{U} \mathbf{y}, \mathbf{U} \mathbf{z} \rangle,$$

$$\| \mathbf{y} + \mathbf{z} \|^2 = \langle \mathbf{y}, \mathbf{y} \rangle + \langle \mathbf{z}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle + \langle \mathbf{z}, \mathbf{y} \rangle$$

$$= \| \mathbf{y} \|^2 + \| \mathbf{z} \|^2 + 2 \text{Re} \langle \mathbf{y}, \mathbf{z} \rangle.$$

Since $\| \mathbf{U} \mathbf{y} \| = \| \mathbf{y} \|$, $\| \mathbf{U} \mathbf{z} \| = \| \mathbf{z} \|$ and $\| \mathbf{U} (\mathbf{y} + \mathbf{z}) \| = \| \mathbf{y} + \mathbf{z} \|$, we have

$$\text{Re} \langle \mathbf{U} \mathbf{y}, \mathbf{U} \mathbf{z} \rangle = \text{Re} \langle \mathbf{y}, \mathbf{z} \rangle$$

Similarly, we can express $\| \mathbf{U} (\mathbf{y} - \mathbf{z}) \| = \| \mathbf{y} - \mathbf{z} \|$ and deduce that
\[ \text{Im}(Uy, Uz) = \text{Im}(y, z) \]

Therefore, \( \langle Uy, Uz \rangle = \langle y, z \rangle \) for all \( x \) and \( y \). By Theorem 1.2.6, \( U \) is unitary. \( \square \)

**Theorem 1.2.8** Let \( U \) be a unitary matrix and \( \lambda \) an eigenvalue of \( U \). Then \( |\lambda| = 1 \), and
\[ |\det(U)| = 1. \]

**Proof.** Let \( x \) be an eigenvector of \( U \) corresponding to \( \lambda \). Then \( Ux = \lambda x \) and
\[ ||Ux|| = ||\lambda x|| = |\lambda| \cdot ||x||. \]

On the other hand, \( ||Ux|| = ||x|| \) since \( U \) is unitary. We must have \( |\lambda| = 1 \). Note that \( \det(U) \) is the product of all eigenvalues of \( U \), so \( |\det(U)| = |\lambda_1| \cdot |\lambda_2| \cdots |\lambda_n| = 1. \) \( \square \)

### 1.3. Schur’s Theorem

Given a square matrix \( A \), if it has full rank of linearly independent eigenvectors, then there exists an invertible \( P \) such that \( P^{-1}AP = \Lambda \), where \( \Lambda \) is diagonal. This is a very beautiful theorem, since the diagonal entries of \( \Lambda \) are exactly the eigenvalues of \( A \). However, not all matrices have this property, that is, some matrices are not diagonalizable. But we may reduce the restriction, and replace the diagonal matrix by an upper triangular matrix; then the new statement will hold for every square matrix, which is known as Schur’s theorem. Schur’s theorem is very useful in solving a variety of problems related to eigenvalues. In this section, we will prove this important theorem and its generalized form.
Lemma 1.3.1 Let \((\lambda, u)\) be an eigenpair of a matrix \(A\), and \(U\) a unitary matrix whose first column is \(u\). Then the first column of \(U^H AU\) is \(\lambda e_1\).

Proof. Write \(U = [u \ y_2 \ y_3 \ \cdots \ y_n]\), where \(u\) is an eigenvector corresponding to the eigenvalue \(\lambda\) of \(A\). Then \(Au = \lambda u\). Since \(U\) is a unitary matrix,

\[
I = U^H U = U^H [u \ y_2 \ y_3 \ \cdots \ y_n] = [U^H u \ U^H y_2 \ U^H y_3 \ \cdots \ U^H y_n].
\]

We deduce that

\[
U^H u = e_1.
\]

So,

\[
U^H AU = U^H A [u \ y_2 \ y_3 \ \cdots \ y_n] = [U^H Au \ U^H Ay_2 \ U^H Ay_3 \ \cdots \ U^H Ay_n] = [\lambda u \ U^H y_2 \ U^H y_3 \ \cdots \ U^H y_n] = [\lambda e_1 \ U^H y_2 \ U^H y_3 \ \cdots \ U^H y_n].
\]

Therefore, the first column of \(U^H AU\) is given by \(\lambda e_1 = [\lambda \ 0 \ \cdots \ 0]^T\).

Remark. By using a similar argument, we can prove the general case: Let \(u\) be the eigenvector corresponding to the eigenvalue \(\lambda\) of \(A\), and \(P\) a invertible matrix with the ith column \(u\). Then the ith column of \(P^{-1}AP\) is \(\lambda e_i\).

Theorem 1.3.2 (Schur's Theorem) Any square matrix is unitarily similar to an upper triangular matrix.
**Proof.** The theorem will be proved by induction. When $A$ is $1 \times 1$, it is trivial. Assume that $A$ is $n \times n$, and the statement is true for every $(n-1) \times (n-1)$ matrix. Let $(\lambda, u)$ be an eigenpair of $A$ with $\|u\|=1$. Now, find a basis of $\mathbb{C}^n$ which contains $u$, say $\{u, x_2, x_3, \ldots, x_n\}$, and then use the Gram-Schmidt process to construct an orthonormal basis $\{u, y_2, y_3, \ldots, y_n\}$. Set $U=[u \ y_2 \ y_3 \ \cdots \ y_n]$. Then $U$ is a unitary matrix. By Lemma 1.3.1, the first column of $U^H A U$ is $\lambda e_1$, that is:

$$U^H A U = \begin{bmatrix} \lambda & c^T \\ 0 & B \end{bmatrix},$$

where $B$ is an $(n-1) \times (n-1)$ matrix, $0$ and $c$ are $(n-1) \times 1$ vectors. By the induction hypothesis, $B$ is unitarily similar to an upper triangular matrix. We can write

$$V^H B V = T$$

for some unitary $V$ and upper triangular $T$. Define $W = \begin{bmatrix} 1 & 0^T \\ 0 & V \end{bmatrix}$, which is a unitary matrix, because

$$W^H W = \begin{bmatrix} 1 & 0^T \\ 0 & V^H \end{bmatrix} \begin{bmatrix} 1 & 0^T \\ 0 & V \end{bmatrix} = \begin{bmatrix} 1 & 0^T \\ 0 & V^H V \end{bmatrix} = \begin{bmatrix} 1 & 0^T \\ 0 & I \end{bmatrix} = I_n$$

Then product $UW$ is also unitary. Therefore, the following unitary transformation:

$$(UW)^H A (UW) = W^H (U^H A U) W = \begin{bmatrix} 1 & 0^T \\ 0 & V^H \end{bmatrix} \begin{bmatrix} \lambda & c^T \\ 0 & B \end{bmatrix} \begin{bmatrix} 1 & 0^T \\ 0 & V \end{bmatrix}
= \begin{bmatrix} \lambda & c^T V \\ 0 & V^H B V \end{bmatrix} = \begin{bmatrix} \lambda & c^T V \\ 0 & T \end{bmatrix},$$

gives an upper triangular matrix. \( \square \)
Theorem 1.3.3 (Generalized Schur’s Theorem) For any square matrices $A$ and $B$ of the same size, there exist unitary matrices $U$ and $V^H$, and upper triangular matrices $S$ and $T$, such that

$$V^H A U = T \quad \text{and} \quad V^H B U = S.$$ 

Proof. This theorem is also proved by induction. If $A$ and $B$ are of size $1 \times 1$, it is trivial. Assume that the theorem is true for each pair of $(n-1) \times (n-1)$ matrices. Let $A$ and $B$ be $n \times n$ matrices, and the scalar $\lambda$ is a solution to

$$\det(A - \lambda B) = 0.$$ 

Then, there exist a unit vector $u$ in $\mathbb{C}^n$, such that $(A - \lambda B)u = 0$. That is $Au = \lambda Bu$ with $\|u\|=1$. Using the same argument in the proof of Theorem 1.3.2, we can construct an orthonormal basis \{u, y_2, y_3, \ldots, y_n\} of $\mathbb{C}^n$. Set

$$Y = [u \ y_2 \ y_3 \ \cdots \ y_n].$$

Then $Y$ is unitary. Similarly, there exists a unit vector $v$ such that

$$v^T A = \lambda v^T B.$$ 

with $\|v\|=1$, and thus we can also construct an orthonormal basis \{v^T, x_2^T, x_3^T, \ldots, x_n^T\}. Set

$$X^H = \begin{bmatrix} v^T \\ x_2^T \\ x_3^T \\ \vdots \\ x_n^T \end{bmatrix}.$$ 

Then $X^H$ is also unitary, and
\[ X^H A Y = \begin{bmatrix} v^T \\ x_2^T \\ x_3^T \\ \vdots \\ x_n^T \end{bmatrix} A \begin{bmatrix} u \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} v^T A u \\ v^T A y_2 \\ v^T A y_3 \\ \vdots \\ v^T A y_n \end{bmatrix} \begin{bmatrix} x_2^T A u \\ x_2^T A y_2 \\ x_2^T A y_3 \\ \vdots \\ x_2^T A y_n \end{bmatrix} + \begin{bmatrix} x_3^T A u \\ x_3^T A y_2 \\ x_3^T A y_3 \\ \vdots \\ x_3^T A y_n \end{bmatrix} = \begin{bmatrix} \lambda v^T Bu + v^T A y_2 \\ \lambda x_2^T Bu + x_2^T A y_2 \\ \lambda x_3^T Bu + x_3^T A y_2 \\ \vdots \\ \lambda x_n^T Bu + x_n^T A y_2 \end{bmatrix} \begin{bmatrix} v^T A y_3 \\ v^T A y_3 \\ v^T A y_3 \\ \vdots \\ v^T A y_n \end{bmatrix} \]

Set \( k = v^T Bu \), \( p = \begin{bmatrix} x_2^T Bu \\ x_3^T Bu \\ \vdots \\ x_n^T Bu \end{bmatrix} \), \( c_1^T = [v^T A y_2 \ v^T A y_3 \ \cdots \ v^T A y_n] \), and \( C_1 \) be the minor matrix of the \((1,1)\) entry of \( X^H A Y \). Then the equation can be simplified to:

\[ X^H A Y = \begin{bmatrix} \lambda k & c_1^T \\ \lambda p & C_1 \end{bmatrix}. \]

Similarly,

\[ X^H B Y = \begin{bmatrix} k & c_2^T \\ p & C_2 \end{bmatrix}, \]

where \( c_2^T = [v^T B y_2 \ v^T B y_3 \ \cdots \ v^T B y_n] \), and \( C_2 \) is the minor matrix of the \((1,1)\) entry of \( X^H B Y \).

Now, we reduce the first column of \( X^H B Y \) into \([\alpha \ 0 \ \cdots \ 0]^T\) for some scalar \( \alpha \) by applying row operations, which have the same effect as left multiplying by a sequence of elementary matrices \( E_m, E_{m-1}, \ldots, E_1 \). Denote \( E \) by \( E_m E_{m-1} \cdots E_1 \). Then \( E \) is invertible such that
\[ E(X^HBY) = E \begin{bmatrix} k & c_2^T \\ p & C_2 \end{bmatrix} = \begin{bmatrix} \alpha & d_2^T \\ 0 & D_2 \end{bmatrix} \]

and

\[ E(X^HAY) = E \begin{bmatrix} \lambda k & c_1^T \\ \lambda p & C_1 \end{bmatrix} = \begin{bmatrix} \lambda \alpha & d_1^T \\ 0 & D_1 \end{bmatrix}. \]

By the QR decomposition (Theorem 1.2.5), there exists a unitary matrix \( Q \) and an upper triangular matrix \( R \) such that

\[ E^{-1} = QR, \]

implies

\[ Q^H = RE. \]

Write \( R = \begin{bmatrix} r & W^T \\ 0 & W \end{bmatrix} \). Then

\[ (Q^H X^H)BY = (RE)X^HBY = \begin{bmatrix} r & W^T \\ 0 & W \end{bmatrix} \begin{bmatrix} \alpha & d_2^T \\ 0 & D_2 \end{bmatrix} = \begin{bmatrix} r \alpha & r d_2^T + W^T D_2 \\ 0 & W D_2 \end{bmatrix} = \begin{bmatrix} \alpha & g_2^T \\ 0 & G_2 \end{bmatrix} \]

and

\[ (Q^H X^H)AY = (RE)X^HAY = \begin{bmatrix} r & W^T \\ 0 & W \end{bmatrix} \begin{bmatrix} \lambda \alpha & d_1^T \\ 0 & D_1 \end{bmatrix} = \begin{bmatrix} \lambda \alpha & r d_1^T + W^T D_1 \\ 0 & W D_1 \end{bmatrix} = \begin{bmatrix} \alpha & g_1^T \\ 0 & G_1 \end{bmatrix} \]

where \( G_1 \) and \( G_2 \) are \((n - 1) \times (n - 1)\) matrices. By the induction hypothesis, there exist unitary matrices \( H^H \) and \( K \) and upper triangular matrices \( L_1 \) and \( L_2 \) such that

\[ H^H G_1 K = L_1 \text{ and } H^H G_2 K = L_2. \]
Define

\[ M^H = \begin{bmatrix} 1 & 0^T \\ 0 & H^H \end{bmatrix} \text{ and } N^H = \begin{bmatrix} 1 & 0^T \\ 0 & K \end{bmatrix}. \]

It is easy to check that \( M \) and \( N \) are unitary by definition. The products

\[ (M^H Q^H X^H) A(YN) = M^H (Q^H X^H AY) N = \begin{bmatrix} 1 & 0^T \\ 0 & H^H \end{bmatrix} \begin{bmatrix} \lambda \alpha & g_1^T K \\ 0 & G_1 \end{bmatrix} \begin{bmatrix} 1 & 0^T \\ 0 & K \end{bmatrix} = \begin{bmatrix} \lambda \alpha & g_1^T K \\ 0 & H^H G_1 K \end{bmatrix} = \begin{bmatrix} \lambda \alpha & g_1^T K \\ 0 & L_1 \end{bmatrix} \]

and

\[ (M^H Q^H X^H) B(YN) = M^H (Q^H X^H B Y) N = \begin{bmatrix} 1 & 0^T \\ 0 & H^H \end{bmatrix} \begin{bmatrix} \alpha & g_2^T K \\ 0 & G_2 \end{bmatrix} \begin{bmatrix} 1 & 0^T \\ 0 & K \end{bmatrix} = \begin{bmatrix} \alpha & g_2^T K \\ 0 & H^H G_2 K \end{bmatrix} = \begin{bmatrix} \alpha & g_2^T K \\ 0 & L_2 \end{bmatrix} \]

are both upper triangular matrices.

Let \( U = YN \) and \( V^H = (XQM)^H \). Then \( T = \begin{bmatrix} \lambda \alpha & g_1^T K \\ 0 & L_1 \end{bmatrix} \) and \( S = \begin{bmatrix} \alpha & g_2^T K \\ 0 & L_2 \end{bmatrix} \). \( \square \)

**Remark.** \( \lambda \) is called the *generalized eigenvalue* of \( A \) and \( B \) if \( \det(A - \lambda B) = 0 \). More precisely, a non-zero vector \( u \) is called a *right eigenvector* of \( A \) and \( B \) if \( Au = \lambda Bu \), and a non-zero \( v^T \) is a *left eigenvector* of \( A \) and \( B \) if \( v^T A = \lambda v^T B \).

We can easily deduce the Schur’s Theorem from theorem 1.3.3. By choosing \( B = I \) the identity matrix, then \( S = V^H U \) is invertible. So \( V = US^{-1} \). Note that the inverse of an invertible upper
triangular matrix is an upper triangular matrix, and the product of two upper triangular matrices is also an upper triangular matrix. Therefore,

\[ V^H A V = (V^H A U)^H V = T U^H (U S^{-1}) = T S^{-1} \]

is an upper triangular matrix.

1.4. Cayley-Hamilton Theorem

Carley-Hamilton theorem tells us, every square matrix is a zero to its characteristic polynomial. However, it is not easy to prove directly. In this section, we will offer a nice proof based on the Schur’s theorem.

**Lemma 1.4.1** Suppose \( A \) is an \( n \times n \) matrix whose first \( k - 1 \) columns are zeros and \( S \) is an upper triangular matrix with a zero in the \( k \)th diagonal entry. Then the product \( AS \) is a matrix whose first \( k \) columns are all zeros.

**Proof.** Let \( B_1 \) be the \( n \times k \) matrix formed by the first \( k \) columns of \( S \). We can write

\[
A = \begin{bmatrix} O_{n \times (k-1)} & a & A_1 \end{bmatrix} \text{ where } a \text{ is the } k \text{th column vector of } A, \text{ and } B_1 = \begin{bmatrix} B_2 \\ 0^T \\ O_{(n-k) \times k} \end{bmatrix} \text{ where the zero vector } 0^T \text{ is the } k \text{th row vector of } B_1. \]

Then

\[
AB_1 = \begin{bmatrix} O_{n \times (k-1)} & a & A_1 \end{bmatrix} \begin{bmatrix} B_2 \\ 0^T \\ O_{(n-k) \times k} \end{bmatrix} = \begin{bmatrix} O_{n \times (k-1)}B_2 + a0^T + A_1O_{(n-k) \times k} = O_{n \times k} \end{bmatrix},
\]

That is, the first \( k \) columns of \( AS \) are all zeros. \( \square \)
Lemma 1.4.2 Let \( \{T_i\}_{i=1}^n \) be a sequence of \( n \times n \) upper triangular matrices, where the \( i \)th diagonal entry of \( T_i \) is zero. Then the product of the \( n \) matrices \( T_1 T_2 \cdots T_n = O \).

**Proof.** Obviously, the first column of \( T_1 \) is zero. By Lemma 1.4.1, the first two columns of \( T_1 T_2 \) are zeros. Using this lemma repeatedly, we can show that the \( n \) columns of \( T_1 T_2 \cdots T_n \) are all zeros. That is,

\[
T_1 T_2 \cdots T_n = O .
\]

**Remark.** If \( T \) is a strictly upper triangular matrix of size \( n \), then every diagonal entry of \( T \) is zero. We can immediately have \( T^n = T \cdot T \cdots T = O \).

Lemma 1.4.3 Suppose \( B = P^{-1}AP \) for some invertible \( P \). Then for any polynomial \( p(x) \),

\[
p(B) = P^{-1} p(A) P .
\]

**Proof.** Let \( p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \). For each \( 1 \leq k \leq n \),

\[
B^k = (P^{-1}AP)(P^{-1}AP) \cdots (P^{-1}AP) = P^{-1}A(PP^{-1})A \cdots AP = P^{-1}A^k P .
\]

Therefore,

\[
p(B) = p_0 I + p_1 A + p_2 A^2 + \cdots + p_n A^n
\]

\[
= a_0 P^{-1}I P + a_1 P^{-1}AP + a_2 P^{-1}A^2 P + \cdots + a_n P^{-1} A^n P
\]

\[
= P^{-1}(a_0 I + a_1 A + a_2 A^2 + \cdots + a_n A^n) P
\]

\[
= P^{-1} p(A) P .
\]

We are done. \( \square \)
**Theorem 1.4.4 (Cayley-Hamilton Theorem)** Every square matrix is a zero of its characteristic polynomial. i.e. Suppose a square matrix $A$ has the characteristic polynomial $c(\lambda)$. Then 
\[ c(A) = O. \]

**Proof.** Using Schur’s theorem, $A$ can be written as $A = U^H T U$ for some unitary $U$ and upper triangular $T$. By Lemma 1.4.3,
\[ c(A) = U^H c(T) U. \]

Note that $T$ has the same eigenvalues as $A$, say $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then
\[ c(T) = (\lambda_1 I - T)(\lambda_2 I - T) \cdots (\lambda_n I - T). \]

We can always assume that $\lambda_i$ is the $i$th diagonal entry of $T$, and then $\lambda_i I - T$ has a zero in the $i$th diagonal. By Lemma 1.4.2,
\[ c(T) = (\lambda_1 I - T)(\lambda_2 I - T) \cdots (\lambda_n I - T) = O. \]

Therefore,
\[ c(A) = U^H c(T) U = U^H O U = O. \]
CHAPTER 2

APPLICATIONS OF SCHUR’S THEOREM

2.1 Hermitian Matrices

In this part, we will introduce two special types of matrices, called hermitian matrices and skew-hermitian matrices. They are analogs of real symmetric matrices and skew-symmetric matrices respectively in the complex matrix space.

Definition 2.1.1 A square matrix $A$ is called hermitian if $A^H = A$, and $A$ is called skew-hermitian if $A^H = -A$.

Remark. According to the definition, every real symmetric matrix is hermitian, and every real skew-symmetric matrix is skew-hermitian. However, the converse may not hold in general.

The diagonal entries of a hermitian matrix must be real since $\bar{a}_{ii} = a_{ii}$, and the diagonal entries of a hermitian matrix must be purely imaginary or 0 since $\bar{a}_{ii} = -a_{ii}$. In fact, it is easy to show that $A$ is hermitian if and only if $iA$ is skew-hermitian, and $A$ is skew-hermitian if and only if $iA$ is hermitian. So every problem about skew-hermitian matrix can be rephrased as a problem about hermitian matrix.

Theorem 2.1.2 Every square matrix can be uniquely written as the sum of a hermitian matrix and a skew-hermitian matrix.
**Proof.** Given $A$ be a square matrix, suppose $A = A_H + A_S$ for some hermitian $A_H$ and skew-hermitian $A_S$. Then $A^H = A_H^H + A_S^H = A_H - A_S$. We can solve that $A_H = \frac{1}{2}(A + A^H)$ is hermitian and $A_S = \frac{1}{2}(A - A^H)$ is skew-hermitian. By the process of finding $A_H$ and $A_S$, such matrices must be unique. \hfill \Box

**Theorem 2.1.3** Every hermitian matrix is unitarily diagonalizable. i.e. Let $A$ be a hermitian matrix. Then there exits a unitary matrix $U$ and a diagonal matrix $\Lambda$, such that

$$U^H AU = \Lambda.$$

Moreover, all the eigenvalues of $A$ are real.

**Proof.** By Schur’s theorem (Theorem 1.3.2), there exist a unitary matrix $U$ and an upper triangular matrix $T$, such that

$$U^H AU = T.$$

Note that $A$ is hermitian. Then the upper triangular matrix $T$ satisfies

$$T^H = (U^H AU)^H = U^H A^H U = U^H AU = T.$$

So $T$ is also a hermitian matrix. Therefore, $T$ is diagonal and real, and the diagonal entries of $T$ are exactly all eigenvalues of $A$. \hfill \Box

**Remark.** Recall that $A$ is hermitian if and only $iA$ is skew-hermitian, and $U^H AU = T$ implies $U^H (iA) U = iT$. Hence, every skew-hermitian matrix is also unitarily diagonalizable, and all the eigenvalues of $A$ are purely imaginary or 0.
Corollary 2.1.4 Let $A$ be a hermitian matrix with only non-negative eigenvalues. Then there exists a square matrix $W$ such that

$$A = W^H W.$$ 

Proof. By Theorem 2.1.3, there exists a unitary matrix $U$ such that $U^H AU = \Lambda$, where $\Lambda = \text{diag}(\lambda_i)$. Set $\Lambda^{1/2} = \text{diag}(\sqrt{\lambda_i})$ and $W = \Lambda^{1/2} U^H$. Then

$$A = UA U^H = U(\Lambda^{1/2} \Lambda^{1/2}) U^H = (\Lambda^{1/2} U^H)^H (\Lambda^{1/2} U^H) = W^H W. \quad \Box$$

Theorem 2.1.5 A square matrix $A$ is hermitian if and only if $u^H Au$ is real for every vector $u$.

Proof. First, Suppose $A$ is a hermitian matrix. Note that $u^H Au$ is a scalar. Then

$$\overline{u^H Au} = (u^H Au)^H = u^H Au.$$

Hence $u^H Au$ is real.

Conversely, assume that $u^H Au$ is real for every $u$ in $\mathbb{C}^n$. Let $u = e_i$. Then every diagonal entry

$$a_{ii} = e_i^H Ae_i$$

is real. Now consider $u = e_i + e_j$ for each pair of $(i, j)$. Then

$$u^H Au = (e_i + e_j)^H A(e_i + e_j) = e_i^H Ae_i + e_j^H Ae_j + e_i^H Ae_j + e_j^H Ae_i = (a_{ii} + a_{jj}) + (a_{ij} + a_{ji})$$

is real, so $a_{ij} + a_{ji}$ must be real. That is

$$\text{Im} a_{ij} = -\text{Im} a_{ji}.$$ 

Let $u = (e_i + ie_j)$. Then
\[ u^H A u = (e_i + i e_j)^H A (e_i + i e_j) = e_i^H A e_i - i e_j^H A e_j + i e_j^H A e_i - i e_i^H A e_j = (a_{ii} + a_{jj}) + i(a_{ij} - a_{ji}) \]

is also real, and then \( a_{ij} - a_{ji} \) is purely imaginary or zero. We can write

\[ \Re a_{ij} = \Re a_{ji}. \]

Therefore, \( a_{ij} = \overline{a_{ji}} \) for all \( 1 \leq i, j \leq n \). It follows that \( A \) is hermitian by definition. \( \square \)

**Remark.** Recall that \( A \) is skew-hermitian if and only if \( iA \) is hermitian. So \( A \) is skew-hermitian if and only if \( u^H A u \) is purely imaginary or 0.

**Theorem 2.1.6** Let \( A \) be a hermitian matrix. Arrange its eigenvalues in non-decreasing order.

Let \( \lambda_{\min} = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n = \lambda_{\max} \). Denote the set \( \{ x^H A x \mid x \in \mathbb{C}^n, x \neq 0 \} \) by \( W(A) \). Then \( W(A) = [\lambda_{\min}, \lambda_{\max}] \).

**Proof.** Note that \( W(A) \) and \( [\lambda_{\min}, \lambda_{\max}] \) are subsets of \( \mathbb{R} \). We may always assume that \( \| u \| = 1 \).

Otherwise, use \( u = \frac{x}{\| x \|} \) to replace \( x \), we will get the same value. Therefore, \( W(A) \) can be simplified as

\[ W(A) = \{ x^H A x \mid x \in \mathbb{C}^n, \| x \| = 1 \}. \]

\( A \) is unitary diagonalizable, say \( A = U^H \Lambda U \), where \( U \) is unitary and \( \Lambda = \text{diag}(\lambda_i) \). Denote \( y = U x \). Then

\[ x^H A x = x^H (U^H \Lambda U) x = (U x)^H \Lambda (U x) = y^H A y. \]

From Theorem 1.2.7, \( \| y \| = \| U x \| = \| x \| = 1 \), we deduce that
\[ u^H A u = \begin{bmatrix} \bar{y}_1 & \bar{y}_2 & \cdots & \bar{y}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \]

\[ = \lambda_1 | y_1 |^2 + \lambda_2 | y_2 |^2 + \cdots + \lambda_n | y_n |^2. \]

with

\[ | y_1 |^2 + | y_2 |^2 + \cdots + | y_n |^2 = 1. \]

Thus, the value of \( u^H A u \) varies continuously over \([\lambda_{\min}, \lambda_{\max}]\). \(\square\)

**Corollary 2.1.7** If \( A \) is a hermitian matrix, then every diagonal entry of \( A \) satisfies:

\[ \lambda_{\min} \leq a_{ii} \leq \lambda_{\max}. \]

**Proof.** Let \( u = e_i \). Then

\[ a_{ii} = e_i^H A e_i \in W(A). \]

Therefore,

\[ \lambda_{\min} = W(A)_{\min} \leq a_{ii} \leq W(A)_{\max} = \lambda_{\max}. \] \(\square\)

**Remark.** By this corollary, we know that \( a_{ii} \leq \lambda_{\max} \leq |\lambda|_{\max} \) and \( a_{ii} \geq \lambda_{\min} \geq -|\lambda|_{\max} \), then

\[ |a_{ii}| \leq |\lambda|_{\max} \text{ for every } i = 1, 2, \cdots, n. \]

Therefore, \( |a_{ii}|_{\max} \leq |\lambda|_{\max}. \)
2.2 Normal Matrices

Theorem 2.1.3 tells that a matrix is unitarily diagonalizable if it is a hermitian matrix or a skew-hermitian matrix. However, the converse may not be true. In this section, we will define a new type of matrix, and then give the sufficient and necessary conditions for unitary diagonalization.

Definition 2.2.1 A square matrix $A$ is a \textit{normal matrix} if

$$AA^H = A^H A.$$ 

Remark. Clearly every unitary matrix, hermitian matrix or skew-hermitian matrix is normal. However, the converse may not hold in general. For example, $A = \begin{pmatrix} 1+i & 0 \\ 0 & 1+i \end{pmatrix}$ is normal but not unitary or hermitian or skew-hermitian.

Lemma 2.2.2 Let $T$ be an upper triangular matrix. If $T$ is commutative with its hermitian transpose, then $T$ is a diagonal matrix.

Proof. We show it by mathematical induction. When $T$ is $1 \times 1$, it is trivial. Suppose that the statement holds for every upper triangular matrix of size $n$ which commutative with its hermitian transpose. Let $S$ be an upper triangular matrix of size $n+1$, say

$$S = \begin{pmatrix} \lambda & c^T \\ 0 & T \end{pmatrix},$$

where $T$ is an upper triangular matrix of size $n$, and vectors $0$ and $c$ are vectors in $\mathbb{C}^n$. If $SS^H = S^H S$, then
\[
\begin{pmatrix}
\lambda & c^T \\
0 & T
\end{pmatrix}
\begin{pmatrix}
\lambda & 0^T \\
\overline{c} & T^H
\end{pmatrix}
=
\begin{pmatrix}
\lambda & 0^T \\
\overline{c} & T^H
\end{pmatrix}
\begin{pmatrix}
\lambda & c^T \\
0 & T
\end{pmatrix}.
\]

That is
\[
\begin{pmatrix}
|\lambda|^2 + \|c\|^2 & c^T T^H \\
T\overline{c} & TT^H
\end{pmatrix}
\begin{pmatrix}
|\lambda|^2 & \overline{\lambda} c^T \\
\lambda \overline{c} & T^H T
\end{pmatrix}.
\]

Then \(TT^H = T^H T\). By the induction hypothesis, \(T\) is diagonal. From \(|\lambda|^2 + \|c\|^2 = |\lambda|^2\), we have \(\|c\|^2 = 0\), so \(c = 0\) the zero vector. Therefore, \(S\) must be diagonal. \(\square\)

**Theorem 2.2.3** A square matrix \(A\) is normal if and only if it is unitary diagonalizable.

**Proof.** First, suppose that \(U^HAU = \Lambda\) for some unitary matrix \(U\) and diagonal matrix \(\Lambda\). By taking the hermitian transpose both sides, we have \(U^HA^HU = \Lambda^H\). Then
\[
U^H(AA^H)U = (U^HAU)(U^HA^HU) = \Lambda\Lambda^H
\]
and
\[
U^H(A^HA)U = (U^HA^HU)(UAU) = A^HA\Lambda.
\]

Note that \(\Lambda\) is diagonal, which is always commutative with its hermitian transpose. Then
\[
AA^H = U(\Lambda\Lambda^H)U^H = U(\Lambda^HA)U^H = A^HA.
\]

By definition, \(A\) is normal.

Conversely, assume that \(AA^H = A^HA\). By Schur’s Theorem, there exists a unitary matrix \(U\) and an upper triangular matrix \(T\) such that
\[
U^HAU = T.
\]
So \( U^H A^H U = T^H \). Therefore

\[
T T^H - T^H T = (U^H A U)(U^H A^H U) - (U^H A^H U)(U^H A U) \\
= U^H A A^H U - U^H A A^H A U \\
= U^H (A A^H - A^H A) U \\
= U^H O U \\
= O.
\]

By Lemma 2.2.2, \( T \) is diagonal. Hence \( A \) is unitarily diagonalizable. \( \square \)

**Remark.** The theorem gives a sufficient and necessary condition of unitary diagonalization.

**Corollary 2.2.4** A square matrix \( A \) is normal if and only if there is an orthonormal set of \( n \) eigenvectors of \( A \).

**Proof.** Suppose \( A \) is normal. By Theorem 2.2.3, \( U^H A U = \Lambda \) for some unitary \( U \) and diagonal \( \Lambda \). Note that the column vectors of \( U \) are the eigenvectors of \( A \). Since \( U \) is unitary, these eigenvectors must form an orthonormal set.

Conversely, assume that \( A \) has an orthonormal set of \( n \) eigenvectors, say \( \{u_1, u_2, \ldots, u_n\} \), set \( U = [u_1 \ u_2 \ \cdots \ u_n]^T \). Then \( U^H A U = \text{diag}(\lambda_i) = \Lambda \), and \( A \) is normal. \( \square \)

**Theorem 2.2.5** A square matrix \( A \) is normal if and only if \( A = H_1 + i H_2 \) for some hermitian matrices \( H_1 \) and \( H_2 \) such that \( H_1 H_2 = H_2 H_1 \).
Proof. Suppose $A$ is normal. By Theorem 2.2.3, $U^H A U = \Lambda$ for some unitary $U$ and diagonal $A$, we can write $A = \Lambda_1 + i \Lambda_2$, where $\Lambda_1$ and $\Lambda_2$ are real. Then

$$A = U \Lambda U^H = U(\Lambda_1 + i \Lambda_2) U^H = U \Lambda_1 U^H + i U \Lambda_2 U^H.$$ 

Set $H_1 = U \Lambda_1 U^H$ and $H_2 = U \Lambda_2 U^H$. Then $H_1$ and $H_2$ are hermitian, and

$$H_1 H_2 = (U \Lambda_1 U^H)(U \Lambda_2 U^H) = U \Lambda_1 \Lambda_2 U^H = U \Lambda_2 \Lambda_1 U^H = (U \Lambda_2 U^H)(U \Lambda_1 U^H) = H_2 H_1.$$ 

Conversely, if such hermitian matrices $H_1$ and $H_2$ exist, we have

$$A^H = (H_1 + i H_2)^H = H_1^H - i H_2^H = H_1 - i H_2.$$ 

Then

$$AA^H = (H_1 + i H_2)(H_1 - i H_2) = H_1^2 + H_2^2 + i(H_1 H_2 - H_2 H_1) = H_1^2 + H_2^2.$$ 

Similarly,

$$A^H A = H_1^2 + H_2^2.$$ 

So $AA^H = A^H A$, and then $A$ is normal. $\square$

Remark. We are using Theorem 2.2.3 to show this statement. Another proof is given by Theorem 2.1.2. Note that $H_2$ is hermitian if and only if $i H_2$ is skew-hermitian. By Theorem 2.1.2, $H_1 = A_1 = \frac{1}{2} (A + A^H)$ and $i H_2 = A_2 = \frac{1}{2} (A - A^H)$. So $H_1 H_2 = H_2 H_1$ if and only if $(A + A^H)(A - A^H) = (A - A^H)(A + A^H)$, if and only if $AA^H = A^H A$.

The following theorems give the condition in which a normal matrix is unitary, hermitian or skew-hermitian, respectively.
**Theorem 2.2.6** A normal matrix $A$ is unitary if and only if all its eigenvalues are of modulus $1$.

**Proof.** The *only if* part is from Theorem 1.2.8.

Consider the *if* part. Assume that all the eigenvalues of $A$ are of modulus $1$. By Schur’s Theorem, we can write $U^H A U = \Lambda$ for some unitary $U$ and diagonal $\Lambda$. Note that the diagonal entries of $\Lambda$ are exactly the eigenvalues of $A$. It is easy to check that $\Lambda$ is a unitary matrix. Then

$$A = U \Lambda U^H$$

is also a unitary matrix. \qed

**Theorem 2.2.7** A normal matrix $A$ is hermitian if and only if all its eigenvalues are real.

**Proof.** The *only if* part is the Theorem 2.1.3.

Now consider the *if* part. By Schur’s Theorem, we can write $U^H A U = \Lambda$ for some unitary $U$ and diagonal $\Lambda$. Suppose all the eigenvalues of $A$ are real. Then $\Lambda$ is real, so $\Lambda^H = \Lambda$. Therefore,

$$A^H = (U \Lambda U^H)^H = U \Lambda^H U^H = U \Lambda U^H = A.$$ 

$A$ is a hermitian matrix. \qed

**Corollary 2.2.8** A normal matrix $A$ is skew-hermitian if and only if its eigenvalues are all purely imaginary or $0$.

**Proof.** Suppose that $A$ is skew-hermitian and $\lambda$ is an eigenvalue of $A$. Then $i\lambda$ is an eigenvalue of $iA$. Note that $iA$ is a hermitian matrix. By Theorem 2.2.7, $i\lambda$ is real. Thus, $\lambda$ is purely imaginary or $0$.  


Conversely, if \( \lambda \) has zero real part, then \( i\lambda \) is be real. Therefore, \( iA \) is a hermitian matrix by Theorem 2.2.7. So \( A \) must be skew-hermitian.

### 2.3 Invariant Subspaces

Let \( V \) be a vector space of dimension \( n \). We can define a linear operation \( T: V \to V \). In particular, for \( V = \mathbb{C}^n \), every \( T \) can be represented by \( T(u) = Au \) for some square matrix \( A \) of size \( n \). Sometimes, we want to ask that if the domain is replaced by a subspace \( W \) of \( \mathbb{C}^n \), will \( T \) maps \( W \) to itself? The answer depends on the choice of \( W \). In this section, we will talk about the self-mapped subspace, and give some interesting results.

**Definition 2.3.1** A vector subspace \( \varphi \) of \( \mathbb{C}^n \) is an invariant subspace of a square matrix \( A \), if 

\[ u \in \varphi \implies Au \in \varphi. \]

**Remark.** Suppose \( \varphi \) is spanned by a set of vectors \( \{x_1, x_2, \cdots, x_n\} \). To check whether \( \varphi \) is an invariant subspace of \( A \), we need only to show that whether each \( Ax_i \) is in \( \varphi \).

**Theorem 2.3.2** Every eigenspace \( \mathcal{E}(\lambda) \) of \( A \) corresponding to eigenvalue \( \lambda \) is an invariant subspace of \( A \). The column space \( \text{col}(A) \) and the null space \( \text{null}(A) \) are also invariant subspaces of \( A \).

**Proof.** Let \( u \in \mathcal{E}(\lambda) \). Then \( Au = \lambda u \in \mathcal{E}(\lambda) \). So \( \mathcal{E}(\lambda) \) is an invariant subspace of \( A \). For every \( x \in \mathbb{C}^n \), \( Ax \in \text{col}(A) \). Then for those \( x \in \text{col}(A) \), \( Ax \) is also in \( \text{col}(A) \). If \( y \in \text{null}(A) \), then \( Ay = 0 \) is in \( \text{null}(A) \). \( \square \)
Consider the unitary transformation in Schur’s Theorem: $U^H A U = T$. Such unitary $U$ depends on the order of the eigenvalues along the diagonal of $T$. However, no matter how $U$ changes, it always has the property given below.

**Theorem 2.3.3** Suppose $U^H A U = T$ for some unitary $U$ and upper triangular $T$. Then the first $k$ columns of $U$ form a basis of an invariant subspace of $A$.

**Proof.** Write $A U = U T$. Replace $U = [u_1 \ u_2 \ \cdots \ u_n]$ into the equality. Then

$$A [u_1 \ u_2 \ \cdots \ u_n] = [u_1 \ u_2 \ \cdots \ u_n] T.$$

That is,

$$[A u_1 \ A u_2 \ \cdots \ A u_n] = [u_1 \ u_2 \ \cdots \ u_n] \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ 0 & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & t_{nn} \end{bmatrix}.$$

Comparing the corresponding columns, we have

$$Au_k = t_{kk} u_k + b_{2k} u_2 + \cdots + b_{nk} u_n, \quad k = 1, 2, \ldots, n.$$

By the remark of Definition 2.3.1, we can verify that $\text{span}\{u_1, u_2, \cdots, u_k\}$ is an invariant subspace of $A$ for every $k$. □

**Theorem 2.3.4** Let $\varphi \neq \{0\}$ be an invariant subspace of $A$. Then there exists an eigenvector of $A$ in $\varphi$. 

30
Proof. Suppose \( \{y_1, y_2, \cdots, y_k\} \) is a basis for \( \varphi \). By definition, each \( Ay_i \) \((1 \leq i \leq k)\) also belongs to \( \varphi \). Hence, \( Ay_i \) is a linear combination of \( y_1, y_2, \cdots, y_r \). Write

\[
Ay_i = g_{1i}y_1 + g_{2i}y_2 + \cdots + g_{ki}y_k.
\]

Denote vector \( g_i = \begin{bmatrix} g_{1i} & g_{2i} & \cdots & g_{ki} \end{bmatrix}^T \), and matrix \( Y = \begin{bmatrix} y_1 & y_2 & \cdots & y_k \end{bmatrix} \). That becomes

\[
Ay_i = Yg_i.
\]

Set \( G = \begin{bmatrix} g_1 & g_2 & \cdots & g_k \end{bmatrix} \). Then these \( k \) equations can be written as the matrix form:

\[
AY = YG.
\]

Note that \( G \) is a square matrix of size \( k \). Let \((\mu, v)\) be an eigenpair of \( G \). That is \( Gv = \mu v \). Premultiplying \( Y \) on both sides, we have

\[
YGv = \mu Yv.
\]

It follows that

\[
A(Yv) = (AY)v = (YG)v = \mu(Yv).
\]

Recall that \( Yv \in \varphi \) but \( Yv \neq 0 \), because \( Y \) has linear independent column vectors. Hence, \((\mu, Yv)\) is an eigenpair of \( A \). \( \square \)

Theorem 2.3.4 shows that a non-trivial invariant subspace \( \varphi \) of \( A \) contains at least one eigenvector of \( A \). In particular, if \( \dim(\varphi) = 1 \), then \( \varphi \) is spanned by that eigenvector. We may ask
that if $\varphi$ is of dimension more than 1, can $\varphi$ still be spanned by some eigenvectors of $A$? The following beautiful theorem tells that it is true if and only if $A$ is diagonalizable.

**Lemma. 2.3.5** Let $A$ be a square matrix of size $n$ and $x$ a vector in $\mathbb{C}^n$. Then $\varphi = \text{span}\{x, Ax, A^2x, \cdots, A^{n-1}x\}$ is an invariant subspace of $A$.

**Proof.** For any vector $y \in \varphi$, it is a linear combination of $x, Ax, A^2x, \cdots, A^{n-1}x$. That is,

$$y = p_0x + p_1Ax + p_2A^2x + \cdots + p_{n-1}A^{n-1}x.$$  

Pre-multiplying $A$ on both sides of the above equation, that becomes

$$Ay = p_0Ax + p_1A^2x + p_2A^3x + \cdots + p_{n-1}A^n x.$$  

Now, we only need to show $A^nx$ is in $\varphi$. Note that the set $\{x, Ax, A^2x, \cdots, A^{n-1}x, A^n x\}$ must be linearly dependent, because it contains $n + 1$ elements. Thus, there exist some scalars $q_0, q_1, q_2, \cdots, q_n$ such that

$$q_0x + q_1Ax + q_2A^2x + \cdots + q_nA^n x = 0$$

where not all $q_0, q_1, q_2, \cdots, q_n$ are zero. Suppose $m$ is the greatest integer among $q_i$’s such that

$$q_m \neq 0 \quad (0 \leq m \leq n).$$

Then

$$A^nx = -\left(\frac{q_0}{q_m}\right)x - \left(\frac{q_1}{q_m}\right)Ax - \left(\frac{q_1}{q_m}\right)A^2x - \cdots - \left(\frac{q_{m-1}}{q_m}\right)A^{m-1}x.$$  

Hence,
\[ A^n x = A^{n-m} \left[ - \left( \frac{q_0}{q_m} \right)x - \left( \frac{q_1}{q_m} \right)A x - \left( \frac{q_1}{q_m} \right)A^2 x \ldots - \left( \frac{q_{m-1}}{q_m} \right)A^{m-1} x \right] \]

\[ = - \left( \frac{q_0}{q_m} \right) A^{n-m} x - \left( \frac{q_1}{q_m} \right) A^{n-m+1} x - \left( \frac{q_1}{q_m} \right) A^{n-m+2} x \ldots - \left( \frac{q_{m-1}}{q_m} \right) A^{n-1} x \]

is in \( \mathcal{V} \). We conclude that \( \mathcal{V} \) is an invariant subspace of \( A \). \( \square \)

**Theorem 2.3.6** Let \( A \) be a square matrix. Then every invariant subspace \( \mathcal{V} \neq \{0\} \) of \( A \) can be spanned by a set of linearly independent eigenvectors of \( A \) if and only if \( A \) is diagonalizable.

**Proof.** If \( A \) is not diagonalizable, there cannot be \( n \) linear independent eigenvectors. Hence \( \mathbb{C}^n \) is not spanned by some eigenvectors of \( A \). We can always find an \( x \notin \mathcal{E}(\lambda_1) \oplus \mathcal{E}(\lambda_2) \oplus \cdots \oplus \mathcal{E}(\lambda_r) \), where \( \mathcal{E}(\lambda_i) \) is the eigenspace corresponding to the eigenvalue \( \lambda_i \) \((1 \leq i \leq r)\) of \( A \). By Lemma 2.3.6,

\[ \mathcal{V} = \text{span}\{x, Ax, A^2 x, \ldots, A^{n-1} x\} \]

is an invariant subspace of \( A \), but cannot be spanned by only eigenvectors since \( x \in \mathcal{V} \).

Now, we will show the if part by induction. Suppose \( A \) is diagonalizable. Let \( u_1, u_2, \ldots, u_n \) be the \( n \) linearly independent eigenvectors of \( A \) corresponding to the eigenvalues \( \lambda_1, \lambda_2, \ldots, \lambda_n \), respectively. When \( \text{dim}(\mathcal{V}) = 1 \), Theorem 2.3.2 shows that there exists an eigenvector \( u \) of \( A \) in \( \mathcal{V} \). It is easy to show that \( \mathcal{V} = \text{span}\{u\} \). Assume that the theorem is true for every invariant subspace of dimension \( k \), then consider \( \text{dim}(\mathcal{V}) = k+1 \). By Theorem 2.3.2, there exists an
eigenvector $u_1 \in \varphi$, and we can always extend $u_1$ to $\{u_1, y_1, y_2, \cdots, y_k\}$, a basis of $\varphi$. Note that every vector in $\mathbb{C}^n$ is a linear combination of $u_1, u_2, \cdots, u_n$. In particular

$$y_i = a_{i1}u_1 + a_{i2}u_2 + \cdots + a_{in}u_n$$

for each $i = 1, 2, \cdots, k$. Set $v_i = y_i - a_{i1}u_1 = a_{i2}u_2 + a_{i3}u_3 + \cdots + a_{in}u_n$. Then for every $x \in \varphi$

$$x = p_0u_1 + p_1 y_1 + p_2 y_2 + \cdots + p_k y_k$$
$$= p_0u_1 + p_1(a_{11}u_1 + v_1) + p_2(a_{12}u_1 + v_2) + \cdots + p_k(a_{1k}u_1 + v_k)$$
$$= (p_0 + p_1 a_{11} + p_2 a_{12} + \cdots + p_k a_{1k}) u_1 + p_1 v_1 + p_2 v_2 + \cdots + p_k v_k.$$ 

Therefore, $\{u_1, v_1, v_2, \cdots, v_k\}$ is also a basis for $\varphi$. Note that $\varphi$ is an invariant subspace of $A$.

Then for each $v_i = a_{i2}u_2 + a_{i3}u_3 + \cdots + a_{in}u_n, A v_i \in \varphi$ and

$$A v_i = a_{i2} A u_2 + a_{i3} A u_3 + \cdots + a_{in} A u_n$$
$$= a_{i2} \lambda_2 u_2 + a_{i3} \lambda_3 u_3 + \cdots + a_{in} \lambda_n u_n$$

So $A v_i$ must be a linear combination of $v_i$'s, and the set $\{v_1, v_2, \cdots, v_k\}$ an invariant subspace of $A$. By the induction hypothesis, there exists a set of linearly independent eigenvectors $\{w_1, w_2, \cdots, w_k\}$ of $A$ such that

$$\text{span}\{v_1, v_2, \cdots, v_k\} = \text{span}\{w_1, w_2, \cdots, w_k\}.$$ 

Note that $u_1, w_1, w_2, \cdots, w_k$ are linear independent eigenvectors, and

$$\varphi = \text{span}\{u_1, w_1, w_2, \cdots, w_k\}.$$ 

We finish the proof. □
2.4 Simultaneous Diagonalization

Generally, the matrix multiplication is not commutative. In this section, we will introduce a class of matrices, in which the matrix multiplication is commutative.

**Definition 2.4.1** Square matrices $A$ and $B$ are said to be *simultaneously diagonalizable*, if both $P^{-1}AP$ and $P^{-1}BP$ are diagonal matrices for some invertible $P$.

**Theorem 2.4.2** If square matrices $A$ and $B$ are simultaneously diagonalizable, then every pair of matrices from $A, B, A + B, AB$ is also simultaneously diagonalizable.

**Proof.** By definition, $P^{-1}AP = \Lambda_1$ and $P^{-1}BP = \Lambda_2$ for some invertible $P$, where $\Lambda_1$ and $\Lambda_2$ are diagonal matrices. Therefore,

$$P^{-1}(A + B)P = P^{-1}AP + P^{-1}BP = \Lambda_1 + \Lambda_2$$

and

$$P^{-1}(AB)P = (P^{-1}AP)(P^{-1}BP) = \Lambda_1\Lambda_2$$

are all diagonal matrices. □

**Remark.** Denote the class of matrices which are simultaneously diagonalizable with $A$ by $\mathcal{A}$, then $\mathcal{A}$ is closed under the matrix addition and multiplication. Moreover, it is easy to check that $\mathcal{A}$ is a ring.
Theorem 2.4.3 If square matrices $A$ and $B$ are simultaneously diagonalizable, then $AB = BA$.

Proof. By definition, $P^{-1}AP = \Lambda_1$ and $P^{-1}BP = \Lambda_2$ for some invertible $P$, where $\Lambda_1$ and $\Lambda_2$ are diagonal matrices. Note that $\Lambda_1 \Lambda_2 = \Lambda_2 \Lambda_1$, then

$$AB = (P \Lambda_1 P^{-1})(P \Lambda_2 P^{-1}) = P \Lambda_1 \Lambda_2 P^{-1} = P \Lambda_2 \Lambda_1 P^{-1} = (P \Lambda_2 P^{-1})(P \Lambda_1 P^{-1}) = BA.$$ \[
\]

We have proved that $\mathcal{A}$ is a commutative ring. Conversely, $AB = BA$ may not imply that $A$ and $B$ are simultaneously diagonalizable unless $A$ and $B$ are both diagonalizable. In fact, if $A$ and $B$ are both diagonalizable, they are also simultaneously diagonalizable. This is a strong theorem, from which we can get some interesting results.

Theorem 2.4.4 Let $A$ and $B$ be square matrices such that $AB = BA$. Suppose $\lambda_0$ is an eigenvalue of $B$. Then there exists an eigenvector of $B$ corresponding to $\lambda_0$ such that it is also an eigenvector of $A$.

Proof. Let $\{u_1, u_2, \cdots, u_k\}$ be a basis of the eigenspace $\mathcal{E}_B(\lambda_0)$ of $B$ corresponding to $\lambda_0$. Now consider $A u_1, A u_2, \cdots, A u_k$. If $A u_j = 0$ for some $1 \leq j \leq k$, then $(0, u_j)$ is an eigenpair of $A$, we are already done. Assume that none of $A u_j$ is zero, $j = 1, 2, \cdots, k$. Then from $BA = AB$, we deduce that

$$B(A u_j) = (AB)u_j = A(Bu_j) = \lambda_0(A u_j)$$

for each $j = 1, 2, \cdots, k$. Hence, $A u_j$ is also an eigenvector of $B$ corresponding to the eigenvalue $\lambda_0$. That is, $A u_j \in \mathcal{E}_B(\lambda_0)$ for every $u_j$. So $\mathcal{E}_B(\lambda_0)$ is an invariant subspace of $A$. By Theorem 2.3.4, there exists an $x \in \mathcal{E}_B(\lambda_0)$ such that $x$ is also an eigenvector of $A$. \[
\]
Remark. Similarly, we can show that for each eigenspace of $A$, there exists an eigenvector in the eigenspace such that it is also an eigenvector of $B$.

**Theorem 2.4.5** If matrices $A$ and $B$ are commutative and both diagonalizable. Then $A$ and $B$ are simultaneously diagonalizable.

**Proof.** Let $\lambda_1, \lambda_2, \ldots, \lambda_r$ be the distinct eigenvalues of $B$ with algebraic (geometric) multiplicity $d_1, d_2, \ldots, d_r$, respectively. Theorem 2.4.4 shows that, each eigenspace $E_B(\lambda_i)$ of $B$ corresponding to $\lambda_i$ is an invariant subspace of $A$. By Theorem 2.3.6, each $E_B(\lambda_i)$ can be spanned by a set of linearly independent eigenvector of $A$. That is, there exist $d_i$ linearly independent eigenvectors in $E_B(\lambda_i)$, such that they are also the eigenvectors of $A$. Note that all these $d_1 + d_2 + \cdots + d_r = n$ common eigenvectors are linearly independent. Therefore, $A$ and $B$ are simultaneously diagonalizable by the matrix $P$ whose columns are the common vectors. □

**Corollary 2.4.6** If $A$ and $B$ are normal matrices and $AB = BA$, then $AB$ is also a normal matrix.

**Proof.** $A$ is normal, so we can write $U^H A U = \Lambda$ for some unitary $U$ and upper triangular $\Lambda$. Note that $A$ and $B$ are both diagonalizable. By Theorem 2.4.5, they are also simultaneously diagonalizable. Theorem 2.4.2 shows that the unitary $U$ also diagonalizes $AB$. Therefore, $AB$ is a normal matrix. □

### 2.5 The Frobenius Norm

We have defined the norm of vectors in Chapter 1, which is the Euclidean length of the line segment. However, there is no natural definition of matrix norm. Many definitions are possible.
In fact, a matrix norm is a mapping from $\mathbb{M}_{nm}(\mathbb{C})$ to $\mathbb{R}$, denoted by $\|A\|$, if it satisfies the following three conditions:

a) $\|A\| \geq 0$, and $\|A\| = 0$ if and only if $A = 0$;

b) $\|A + B\| \leq \|A\| + \|B\|$;

c) $\|kA\| = |k| \cdot \|A\|$ for any complex scalar $k$.

In this section, we will introduce the most frequently used norm in linear algebra, the **Frobenius norm**. We usually use the notation $\|A\|_F$ for the Frobenius norm. For convenience, we use $\|A\|$ instead.

**Definition 2.5.1** The **Frobenius norm** of a square matrix $A$, denoted by $\|A\|$, is the square root of the sum of the squares of the absolute values of the entries of $A$. That is

$$\|A\| = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^2}.$$ 

**Remark.** If we write $A = [a_1 \quad a_2 \quad \cdots \quad a_n]$, then $\|A\|^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^2 = \sum_{i=1}^{n} \|a_i\|^2$, is the sum of squares of the hermitian norm of its column vectors. Obviously, $\|A\| = \|\overline{A}\| = \|A^T\| = \|A^\dagger\|$. For some special matrices, we can easily compute their Frobenius norms, such as $\|I_n\| = \sqrt{n}$, $\|J_n\| = n$.

**Theorem 2.5.2** Let $A$ be a square matrix. Then $\|A\|^2 = \text{tr}(A^\dagger A)$. Here $\text{tr}(A)$ is the trace of a square matrix $A$, defined as the sum of the diagonal entries of $A$. 
Proof. $\|A\|^2 = \sum_{j=1}^{n} \|a_j\|^2 = a_{11}^{\dagger}a_1 + a_{21}^{\dagger}a_2 + \cdots + a_{n1}^{\dagger}a_n = \text{tr}(A^H A)$. \qed

Remark. By this theorem, we can find the Frobenius norm for some other special matrices. For example, let $U$ be a unitary matrix of size $n$. Then $\|U\|^2 = \text{tr}(U^H U) = \text{tr}(I_n) = n$.

Now, we want to show that the Frobenius norm is “well defined”. That is, it satisfies the properties that given at the beginning of this section.

Theorem 2.5.3 For any scalar $k \in \mathbb{C}$ and square matrices $A$ and $B$ of the same size,

a) $\|A\| \geq 0$, and $\|A\| = 0$ if and only if $A = O$;

b) $\|kA\| = |k| \cdot \|A\|$;

c) $\|A + B\| \leq \|A\| + \|B\|$;

d) $\|AB\| \leq \|A\| \cdot \|B\|$.

Proof. a) By definition, $\|A\|^2 = \sum |a_{ij}|^2 \geq 0$, $\|A\| = 0$ if and only if every $a_{ij}$ is 0. That is $A = O$.

b) $\|kA\|^2 = \sum |ka_{ij}|^2 = |k|^2 \sum |a_{ij}|^2 = |k|^2 \cdot \|A\|^2$.

c) Let $A = [a_1, a_2, \cdots, a_n]$ and $B = [b_1, b_2, \cdots, b_n]$. By the triangle inequality

$$\|u + v\| \leq \|u\| + \|v\|.$$ 

Then
\[ \| \mathbf{A} + \mathbf{B} \|^2 = \| \mathbf{a}_1 + \mathbf{b}_1 \|^2 + \| \mathbf{a}_2 + \mathbf{b}_2 \|^2 + \cdots + \| \mathbf{a}_n + \mathbf{b}_n \|^2 \]
\[ \leq (\| \mathbf{a}_1 \| + \| \mathbf{b}_1 \|)^2 + (\| \mathbf{a}_2 \| + \| \mathbf{b}_2 \|)^2 + \cdots + (\| \mathbf{a}_n \| + \| \mathbf{b}_n \|)^2 \]
\[ = \| \mathbf{a}_1 \|^2 + \| \mathbf{a}_2 \|^2 + \cdots + \| \mathbf{a}_n \|^2 + \| \mathbf{b}_1 \|^2 + \| \mathbf{b}_2 \|^2 + \cdots + \| \mathbf{b}_n \|^2 \]
\[ + 2 \| \mathbf{a}_1 \| \cdot \| \mathbf{b}_1 \| + 2 \| \mathbf{a}_2 \| \cdot \| \mathbf{b}_2 \| + \cdots + 2 \| \mathbf{a}_n \| \cdot \| \mathbf{b}_n \|. \]

By the Cauchy inequality for scalar product,

\[ \| \mathbf{A} \|^2 \| \mathbf{B} \|^2 = (\| \mathbf{A} \|^2 + \| \mathbf{B} \|^2 + \cdots + \| \mathbf{A} \|)^2 \cdot (\| \mathbf{B} \|^2 + \| \mathbf{B} \|^2 + \cdots + \| \mathbf{B} \|)^2 \]
\[ \geq (\| \mathbf{A} \| \cdot \| \mathbf{B} \| + \| \mathbf{A} \| \cdot \| \mathbf{B} \| + \cdots + \| \mathbf{A} \| \cdot \| \mathbf{B} \|)^2. \]

Then

\[ \| \mathbf{A} + \mathbf{B} \|^2 \leq \| \mathbf{A} \|^2 + \| \mathbf{B} \|^2 + 2 \| \mathbf{A} \| \cdot \| \mathbf{B} \| \]
\[ = (\| \mathbf{A} \| + \| \mathbf{B} \|)^2. \]

d) Write \( \mathbf{A} = \begin{bmatrix} \mathbf{c}_1^T & \mathbf{c}_2^T & \cdots & \mathbf{c}_n^T \end{bmatrix} \), \( \mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix} \). Then for each \( \mathbf{b}_i \),

\[ \| \mathbf{A} \mathbf{b}_i \|^2 = \| \mathbf{c}_1 \mathbf{b}_i \|^2 + \| \mathbf{c}_2 \mathbf{b}_i \|^2 + \cdots + \| \mathbf{c}_n \mathbf{b}_i \|^2 \]
\[ \leq \| \mathbf{c}_1 \|^2 \| \mathbf{b}_i \|^2 + \| \mathbf{c}_2 \|^2 \| \mathbf{b}_i \|^2 + \cdots + \| \mathbf{c}_n \|^2 \| \mathbf{b}_i \|^2 \]
\[ = (\| \mathbf{c}_1 \|^2 + \| \mathbf{c}_2 \|^2 + \cdots + \| \mathbf{c}_n \|^2)^2 \| \mathbf{b}_i \|^2 \]
\[ = \| \mathbf{A} \|^2 \| \mathbf{b}_i \|^2. \]

Therefore,

\[ \| \mathbf{A} \mathbf{B} \|^2 = \| \begin{bmatrix} \mathbf{A} \mathbf{b}_1 & \mathbf{A} \mathbf{b}_2 & \cdots & \mathbf{A} \mathbf{b}_n \end{bmatrix} \|^2 \]
\[ = \| \mathbf{A} \mathbf{b}_1 \|^2 + \| \mathbf{A} \mathbf{b}_2 \|^2 + \cdots + \| \mathbf{A} \mathbf{b}_n \|^2 \]
\[ \leq \| \mathbf{A} \|^2 (\| \mathbf{b}_1 \|^2 + \| \mathbf{b}_2 \|^2 + \cdots + \| \mathbf{b}_n \|^2) \]
\[ = \| \mathbf{A} \|^2 \| \mathbf{B} \|^2. \]
Theorem 2.5.4 Let $U$ be a unitary matrix. Then for any square matrix $A$ of the same size,

$$
\|A\| = \|UA\| = \|AU\|. 
$$

Proof. By Theorem 2.5.2,

$$
\|UA\|^2 = \text{tr}[(UA)^H (UA)] = \text{tr}[A^H (U^H U) A] = \text{tr}(A^H A) = \|A\|^2.
$$

Then

$$
\|(AU)^H\| = \|U^H A^H\| = \|A^H\|.
$$

So

$$
\|AU\| = \|A\|. 
$$
Theorem 2.5.6 If \( A \) and \( B \) are normal matrices of the same size, then \( \| AB \| = \| BA \| \).

**Proof.** Note that \( A^*A = A^*A \) and \( B^*B = B^*B \) by hypothesis. Using Theorem 2.5.2,

\[
\| AB \|^2 = \text{tr}[(AB)^*(AB)] = \text{tr}[B^*(A^*A)B] = \text{tr}[B^*(AA^*)B] = \text{tr}[(A^*B)^*(A^*B)] \\
= \| A^*B \|^2 = \| (B^*A)^* \|^2 = \| B^*A \|^2 \\
= \text{tr}[(B^*A)^*(B^*A)] = \text{tr}[A^*(BB^*)A] = \text{tr}[A^*(BB^*)A] = \text{tr}[(BA)^*(BA)] \\
= \| BA \|^2.
\]

We are done. \( \square \)

**Remark.** The statement may not hold for non-normal matrices, For example, take \( A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \) and \( B = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} \). Then \( \| AB \| = 0 \), but \( \| BA \| = 1 \).

Theorem 2.5.7 (Schur’s Inequality) Let \( \lambda_1, \lambda_2, \cdots, \lambda_n \) be the eigenvalues of square matrix \( A \), then

\[
\| A \|^2 \geq |\lambda_1|^2 + |\lambda_2|^2 + \cdots + |\lambda_n|^2.
\]

The equality holds if and only if \( A \) is a normal matrix.

**Proof.** By Schur’s theorem, there exist a unitary matrix \( U \) and an upper triangular matrix \( T \) such that \( U^*A U = T \). Therefore, by Corollary 2.5.5,

\[
\| A \|^2 = \| T \|^2 \geq |t_{11}|^2 + |t_{22}|^2 + \cdots + |t_{nn}|^2.
\]

Recall that the eigenvalues of \( A \) are exactly the diagonal entries of \( T \). So
The equality holds if and only if $T$ is a diagonal matrix, if and only if $A$ is normal. \qed

**Corollary 2.5.8** Let $\lambda_0$ be an eigenvalue of an invertible matrix $A$. Then

$$
\| A \|^{-1} \leq |\lambda_0|^{-1} \leq \| A^{-1} \|.
$$

**Proof.** By theorem 2.5.6, $\| A \|^2 \geq |\lambda_1|^2 + |\lambda_2|^2 + \cdots + |\lambda_n|^2 \geq |\lambda_0|^2$. Note that $\lambda_0^{-1}$ is an eigenvalue of $A^{-1}$. Then $\| A^{-1} \| \geq |\lambda_0^{-1}| = |\lambda_0|^{-1}$. \qed

**Corollary 2.5.9** If a square matrix $A$ is of size $n$, then $\| A \| \geq \sqrt[n]{\det(A)}^{1/n}$. The equality holds if and only if $A$ is a scalar multiple of a unitary matrix.

**Proof.** Recall the **Arithmetic - Geometric Means Inequality** is

$$
\frac{x_1 + x_2 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 x_2 \cdots x_n}, \quad x_i \geq 0.
$$

Let $x_i = |\lambda_i|^2$ for each $i = 1, 2, \cdots, n$. Then

$$
\frac{\| A \|^2}{n} \geq \frac{|\lambda_1|^2 + |\lambda_2|^2 + \cdots + |\lambda_n|^2}{n} \geq \sqrt[n]{|\lambda_1|^2 |\lambda_2|^2 \cdots |\lambda_n|^2} = |\det(A)|^{2/n}.
$$

By taking square roots for both sides, we finish the proof. Moreover, the equality holds if and only if $A$ is normal, and all the eigenvalues of $A$ are of the same modulus. Theorem 2.2.6 shows that $A$ must be a scalar multiple of a unitary matrix. \qed
Theorem 2.5.10 Let $A$ be a square matrix. Set $A_H = \frac{1}{2}(A + A^H)$ and $A_S = \frac{1}{2}(A - A^H)$. Then

$$\|A_H\|^2 + \|A_S\|^2 = \|A\|^2.$$ 

Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of $A$. Then

$$\|A_H\|^2 \geq \sum_{i=1}^{n} \Re^2(\lambda_i)$$

and

$$\|A_S\|^2 \geq \sum_{i=1}^{n} \Im^2(\lambda_i).$$

The above two equalities hold if and only if $A$ is normal.

Proof. The $(i, j)$th entry of $A_H$ is $\frac{1}{2}(a_{ij} + \overline{a}_{ji})$, and that of $A_S$ is $\frac{1}{2}(a_{ij} - \overline{a}_{ji})$. Verify that

$$|a_{ij} + \overline{a}_{ji}|^2 + |a_{ij} - \overline{a}_{ji}|^2 = (a_{ij} + \overline{a}_{ji})(\overline{a}_{ij} + a_{ji}) + (a_{ij} - \overline{a}_{ji})(\overline{a}_{ij} - a_{ji})$$

$$= |a_{ij}|^2 + |a_{ji}|^2 + a_{ij}a_{ji} + \overline{a}_{ij}\overline{a}_{ji} + |a_{ij}|^2 - a_{ij}a_{ji} - \overline{a}_{ij}\overline{a}_{ji}$$

$$= 2(|a_{ij}|^2 + |a_{ji}|^2).$$

for each $1 \leq i, j \leq n$. Then

$$\|A_H\|^2 + \|A_S\|^2 = \frac{1}{2}\sum (|a_{ij}|^2 + |a_{ji}|^2) = \sum |a_{ij}|^2 = \|A\|^2.$$ 

By Schur’s theorem, there exist a unitary matrix $U$ and an upper triangular matrix $T$, such that $U^H A U = T$. Then $A = U T U^H$, $A^H = U T^H U^H$. We have
\[ A_H = \frac{1}{2}(U T U^H + U T^H U^H) = \frac{1}{2} U (T + T^H) U^H. \]

Recall that the diagonal entries of \( T \) are exactly the eigenvalues of \( A \). So each diagonal entry of \( T + T^H \) is given by \( \lambda_i + \bar{\lambda}_i = 2 \text{Re}(\lambda_i) \). Therefore

\[ \| A_H \|^2 = \frac{1}{4} \| U (T + T^H) U^H \|^2 = \frac{1}{4} \| T + T^H \|^2 \geq \sum_{i=1}^{n} \text{Re}^2(\lambda_i). \]

The equality holds if and only if \( T + T^H \) is diagonal, is equivalent to that the upper triangular matrix \( T \) is diagonal, but \( T \) is diagonal if and only if \( A \) is normal.

Similarly, each diagonal entry of \( T - T^H \) is given by \( \lambda_i - \bar{\lambda}_i = 2i \text{Im}(\lambda_i) \), and then

\[ \| A_S \|^2 = \frac{1}{4} \| U (T - T^H) U^H \|^2 = \frac{1}{4} \| T - T^H \|^2 \geq \sum_{i=1}^{n} \text{Im}^2(\lambda_i). \]

Moreover, the equality holds if and only if \( A \) is normal. \( \square \)

**Remark.** By this theorem, we immediately have \( \| A_H \|^2 + \| A_S \|^2 \geq \sum_{i=1}^{n} \text{Re}^2(\lambda_i) + \sum_{i=1}^{n} \text{Im}^2(\lambda_i) \). Or equivalently, \( \| A \|^2 \geq \sum_{i=1}^{n} |\lambda_i|^2 \). That is the result of theorem 2.5.7.

Recall the results of Theorem 2.1.2 that \( A_H \) is hermitian and \( A_S \) is skew-hermitian. If \( A \) is hermitian, then \( A_S = 0 \). From the second inequality \( \sum_{i=1}^{n} \text{Im}^2(\lambda_i) \leq \| A_S \|^2 = 0 \), we know that the eigenvalues of \( A \) are real. If \( A \) is skew-hermitian, then \( A_H = 0 \). \( \sum_{i=1}^{n} \text{Re}^2(\lambda_i) \leq \| A_H \|^2 = 0 \) implies that the eigenvalues of \( A \) are purely imaginary or 0. These are the results of Theorem 2.2.7 and Corollary 2.2.8.
CHAPTER 3

THE LIMIT OF $A^m$

3.1 Gerschgorin’s Theorem

In many problems of linear algebra, such as diagonalization, we are asked to find the eigenvalues of the given square matrix. Note that eigenvalues are the zeros of the characteristic polynomial $\det(\lambda I - A)$. If $A$ is of size more than 4, we can only find the approximate solutions in general. The Gerschgorin’s theorem provides an estimate for the location of the eigenvalues.

The row and column sums of the matrix play the major roles in this theorem, for convenience, we denote the sum of the magnitudes of the entries of the $i$th row by $r_i$, and the sum of the magnitudes of the entries of the $j$th column by $c_j$. That is, for each $i, j = 1, 2, \ldots, n$,

$$r_i = |a_{i1}| + |a_{i2}| + \cdots + |a_{in}| = \sum_{k=1}^{n} |a_{ik}|$$

$$c_j = |a_{1j}| + |a_{2j}| + \cdots + |a_{nj}| = \sum_{k=1}^{n} |a_{kj}|.$$ 

They will give a constraint to the location of eigenvalues.

**Definition 3.1.1** Let $A$ be a square matrix of size $n$. The Gerschgorin radius of the $i$th row is the sum of the magnitudes of the off-diagonal entries in the $i$th row, denoted by $r_i$. That is,
The corresponding **Gerschgorin disk** is the set of complex numbers such that

\[ |z - a_{ii}| \leq r_i'. \]

**Remark.** If the Gerschgorin radii \( r_i' \) equals to 0, then the Gerschgorin disk deduces to one point \( z = a_{ii} \). Since theorems given below still hold for this special case, for convenience, we just consider it as a “disk” with center \( a_{ii} \) and radius 0.

**Definition 3.1.2** The set of eigenvalues of \( A \) is called the **spectrum** of \( A \). The **spectral radius** of \( A \), denoted by \( \rho(A) \), is the maximum of the absolute values of the eigenvalues. i.e.

\[ \rho(A) = \max \{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_n|\}. \]

**Theorem 3.1.3 (Gerschgorin’s Theorem)** The spectrum of \( A \) lies within the union of the Gerschgorin’s disks. i.e. If \( \lambda_0 \) is an eigenvalue of \( A \), then \( |\lambda_0 - a_{ii}| \leq r_i' \) for some \( 1 \leq i \leq n \).
Proof. Let \((\lambda_0, \mathbf{u})\) be an eigenpair of \(A\). Consider the eigenvalue equation \(A\mathbf{u} = \lambda_0 \mathbf{u}\). It follows that

\[
\sum_{j=1}^{n} a_{ij} u_j = \lambda_0 u_i, \quad i = 1, 2, \ldots, n. \quad (*)
\]

where \(u_j\) is the \(i\)th component of \(\mathbf{u}\). Suppose \(u_k\) has the greatest modulus among \(u_j\)'s. Then choose the \(k\)th equation in (*), we have

\[
\sum_{j \neq k} a_{ij} u_j = \lambda_0 u_k - a_{kk} u_k = (\lambda_0 - a_{kk}) u_k.
\]

Apply the Triangle Inequality,

\[
\left| \lambda_0 - a_{kk} \right| \left| u_k \right| \leq \sum_{j \neq k} \left| a_{ij} \right| \left| u_j \right|.
\]

Note that \(u_k \neq 0\), otherwise \(\mathbf{u}\) would be the zero-vector. Therefore,

\[
\left| \lambda_0 - a_{kk} \right| \leq \sum_{j \neq k} \left| a_{ij} \right| \left| u_j \right| / \left| u_k \right| = r_k'.
\]

Remark. Note that if \(\lambda_0\) is an eigenvalue of \(A\), then \(\mathbf{v}^T A = \lambda_0 \mathbf{v}^T\) for some non-zero row vector \(\mathbf{v}^T\). Using the similar argument, we can show that \(\left| \lambda_0 - a_{kk} \right| \leq c_k' \) for some \(k\), where \(c_k' = c_k - |a_{kk}|\).

Corollary 3.1.4 Denote \(r_{\max} = \{r_1, r_2, \ldots, r_n\}\). Let \(\lambda_0\) be an eigenvalue of \(A\). Then \(\left| \lambda_0 \right| \leq r_{\max}\).

Moreover, \(\rho(A) \leq r_{\max}\).

Proof. Let \(\lambda_0\) be the eigenvalue of \(A\) with the greatest modulus. Gerschgorin’s Theorem shows that \(\lambda_0\) satisfies \(\left| \lambda_0 - a_{ii} \right| \leq r_i'\) for some \(i\). Using the Triangle Inequality, we have
Then
\[ |\lambda| - |a_{ii}| \leq |\lambda - a_{ii}| \leq r_i'. \]

Since the above inequality holds for every eigenvalue of $A$, so
\[ \rho(A) \leq r_{\text{max}}. \quad \Box \]

**Remark.** This corollary gives an upper bound for $\rho(A)$. When $A$ is a hermitian matrix, Corollary 2.1.7 show that $|a_{ii}|_{\text{max}} \leq \rho(A)$. This gives a lower bound for $\rho(A)$.

In this chapter, we will use the notation $A \geq B$ for real matrices $A$ and $B$ if every entry of $A$ is greater than or equal to the corresponding entry of $B$, and $A > B$ if every entry of $A$ is strictly greater than the corresponding entry of $B$. In particular, we write $A \geq O$ if $A$ is nonnegative, and $A > O$ if $A$ is positive.

The stochastic matrix is a class of real matrices widely used in statistics. Gerschgorin’s Theorem gives a simple proof that its spectral radius is 1.

**Definition 3.1.5** A square matrix $S$ is a *stochastic matrix* if $S \geq O$ and each row sums to 1.
Theorem 3.1.6 If $S$ is a stochastic matrix, then $\rho(S) = 1$. Moreover, if the diagonal entries of $S$ are all positive, then $1$ is the only eigenvalue of $S$ with absolute value $1$.

Proof. It is easy to verify that $(1, \mathbf{1})$ is an eigenpair of $S$. Note that the first $1$ is a scalar, and the second $\mathbf{1}$ is a column vector. Each row sum $r_i = 1$, so $r_{\text{max}} = 1$. Corollary 3.1.4 shows that $\rho(S) = 1$.

Suppose the diagonal entries of $S$ are all positive. Let $\lambda_0$ be an eigenvalue of $S$ with magnitude $1$. Then by Gerschgorin’s Theorem,

$$|\lambda_0 - s_{ii}| \leq r_i = 1 - s_{ii}$$

for some $1 \leq i \leq n$. Then

$$1 - s_{ii} \geq |\lambda_0 - s_{ii}| \geq |\lambda_0| - s_{ii} = 1 - s_{ii}.$$  

Therefore,

$$1 - s_{ii} = |\lambda_0 - s_{ii}| = |\lambda_0| - s_{ii}.$$  

The second equality implies that the argument of $\lambda_0$ is the same as that of $s_{ii}$. Recall that $s_{ii}$ is positive, so $\lambda_0$ is also positive. Thus, we have $\lambda_0 = 1$. \qed

Remark. We can also prove this theorem by drawing the graph. The eigenvalue $\lambda_0$ with modulus $1$ lies in both the Gerschgorin’s disk and the unit circle. The graph given below shows that $\lambda_0 = 1$. 

50
However, if $s_{ii} = 0$ for some $i$, then the circle is the boundary of the disk. The intersection point is not unique. So the statement may not be true. For example, $S = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is a stochastic matrix with eigenvalues $\lambda = \pm 1$.

### 3.2 Limit of $A^m$ For $\rho(A) \neq 1$

For complex numbers, the absolute value of constant is used to determine whether $a^m$ is convergent as $m \to \infty$. If $|a| < 1$, then $a^m \to 0$; if $|a| > 1$ then $a^m \to \infty$; if $a = 1$, then $a^m \to 1$; if $|a| = 1$ but $a \neq 1$, then the limit of $a^m$ does not make sense. For matrix, the spectral radius $\rho(A)$ is the value used to determine the limit of $A^m$. In this section, we will discuss the case when $\rho(A) \neq 1$; and in next section, we will talk about the limit if $\rho(A) = 1$. 
**Definition 3.2.1** Let \( A_1, A_2, \cdots, A_m, \cdots \) be a sequence of \( n \times n \) matrices, with \( A_m = (a_{ij}^{(m)})_{n \times n} \). If 
\[ \lim_{m \to \infty} a_{ij}^{(m)} \]
exists for every \( i, j = 1, 2, \cdots, n \), then the sequence of matrices is said to be **convergent**; otherwise, it is called **divergent**. If \( A \) is convergent, set 
\[ \lim_{m \to \infty} a_{ij}^{(m)} = a_{ij}^{(\infty)} \]
and define \( A_{\infty} = (a_{ij}^{(\infty)})_{n \times n} \).
Then \( A_{\infty} \) is the **limit** of the sequence. Written as
\[ \lim_{m \to \infty} A_m = A_{\infty} \]
or
\[ A_m \to A_{\infty} \text{ as } m \to \infty. \]

**Remark.** It is easy to show that: *if* \( \lim_{m \to \infty} A_m = A_{\infty} \) *and* \( \lim_{m \to \infty} B_m = B_{\infty} \), *then*
\[ \lim_{m \to \infty} (A_m + B_m) = A_{\infty} + B_{\infty}. \]
For any complex scalar \( c \), \( \lim_{m \to \infty} (c \ A_m) = c \ A_{\infty} \).

In this section, if the \( A_1, A_2, \cdots, A_m, \cdots \) is a power sequence. i.e. \( A_m = A^m \) for each positive integer \( m \). We usually use \( A_{\infty} \) instead of \( A_m \). That is \( A^\infty = \lim_{m \to \infty} A^m \).

**Theorem 3.2.2** If \( \lim_{m \to \infty} A_m = A_{\infty} \), then for any matrix \( B \) with \( n \) rows and \( C \) with \( n \) columns,
\[ \lim_{m \to \infty} A_m \ B = A_{\infty} \ B \text{ and } \lim_{m \to \infty} C \ A_m = C \ A_{\infty}. \]

**Proof.** Note that the \((i,j)\) entry of \( A_m \ B \) is \( \sum_{k=1}^{n} (a_{ik}^{(m)} \ b_{kj}) \), then
\[ \lim_{m \to \infty} \sum_{k=1}^{n} (a_{ik}^{(m)} \ b_{kj}) = \sum_{k=1}^{n} (\lim_{m \to \infty} a_{ik}^{(m)} \ b_{kj}) = \sum_{k=1}^{n} (a_{ik}^{(\infty)} \ b_{kj}). \]
So, \( \lim_{m \to \infty} A_m \ B = A_{\infty} \ B \). Similarly, we can show that \( \lim_{m \to \infty} C \ A_m = C \ A_{\infty} \).

\[ \square \]
**Corollary 3.2.3** If $\lim_{m \to \infty} A_m$ does not exist, then for invertible $B$ and $C$, $\lim_{m \to \infty} A_mB$ and $\lim_{m \to \infty} CA_m$ also do not exist.

**Proof.** Suppose $\lim_{m \to \infty} A_mB$ exists. By Theorem 3.2.2,

$$\lim_{m \to \infty} A_m = \lim_{m \to \infty} A_mB^{-1}$$

exists. That shows a contradiction. Hence, $\lim_{m \to \infty} A_mB$ does not exist. Similarly, $\lim_{m \to \infty} CA_m$ also does not exist. □

As for complex numbers, we can also talk about series of matrices. Recall that $\|A\|$ is the Frobenius norm of a square matrix $A$, which has been defined in Section 2.5. We can use it to test the convergence of a special kind of matrix series.

**Theorem 3.2.4** Let $A$ be a square matrix, and $\{c_m\}$ be a nonnegative sequence. Suppose that $c_1\|A\| + c_2\|A\|^2 + \cdots + c_m\|A\|^m$ converges. Then $c_1A + c_2A^2 + \cdots + c_mA^m + \cdots$ also converges.

**Proof.** By theorem 2.5.3 (d), $\|A^m\| = \|AA\cdots A\| \leq \|A\| \cdot \|A\| \cdots \|A\| = \|A\|^m$. Then for each $(i,j)$th entry of $A^m$, $|a_{ij}^{(m)}| \leq \|A^m\| \leq \|A\|^m$. Since $c_1\|A\| + c_2\|A\|^2 + \cdots + c_m\|A\|^m$ converges, we can apply the Cauchy criterion. Given any $\varepsilon > 0$, there exists a positive integer $N$ such that $c_{r+1}\|A\|^{r+1} + c_{r+2}\|A\|^{r+2} + \cdots + c_s\|A\|^s < \varepsilon$, whenever $s > r > N$. Then
\[ |c_{r+1}a_{ij}^{(r+1)} + c_{r+2}a_{ij}^{(r+2)} + \cdots + c_r a_{ij}^{(r)}| \]
\[ \leq c_{r+1} |a_{ij}^{(r+1)}| + c_{r+2} |a_{ij}^{(r+2)}| + \cdots + c_r |a_{ij}^{(r)}| \]
\[ \leq c_{r+1} \|A\|^{r+1} + c_{r+2} \|A\|^{r+2} + \cdots + c_r \|A\|^r \]
\[ < \varepsilon \]

for each \( i, j = 1, 2, \cdots, n \). So \( c_1 a_{ij}^{(1)} + c_2 a_{ij}^{(2)} + \cdots + c_m a_{ij}^{(m)} \) is also convergent. By definition, \( c_1 A + c_2 A^2 + \cdots + c_m A^m \) is convergent. \( \square \)

**Theorem 3.2.5** For any square matrix \( A \), if \( \rho(A) > 1 \), then \( \lim_{m \to \infty} A^m \) does not exist.

**Proof.** Let \( \lambda_0 \) be an eigenvalue of \( A \) with the greatest modulus. Then \( |\lambda_0| = \rho(A) > 1 \). By Schur’s Theorem, \( U^H A U = T \) for some unitary matrix \( U \) and upper triangular matrix \( T \). Note that \( \lambda_0^m \) is an eigenvalue of \( A^m \). Since \( A^m = (U T U^H)^m = U T^m U^H \), \( T^m \) must have \( \lambda_0^m \) on its diagonal. However, \( \lim_{m \to \infty} \lambda_0^m \) does not exist since \( |\lambda_0| > 1 \), so \( \{T^m\} \) is divergent. By Corollary 3.2.3, \( A^m \to U T^m U^H \) is also divergent. \( \square \)

In order to test the convergence of \( \lim_{m \to \infty} A^m \) when \( \rho(A) < 1 \), we bring in the notation \( S_T(\sigma) \), which is the related matrix of an upper triangular matrix \( T \). Let

\[
T = \begin{bmatrix}
\lambda_1 & t_{12} & t_{13} & \cdots & t_{1n} \\
0 & \lambda_2 & t_{23} & \cdots & t_{2n} \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & \cdots & t_{n-1,n} \\
0 & 0 & 0 & 0 & \lambda_n
\end{bmatrix},
\]

and \( \sigma \geq 0 \). Then it is given by
$$S_T(\sigma) = \begin{bmatrix}
\sigma & t_{12} & t_{13} & \cdots & t_{1n} \\
0 & \sigma & t_{23} & \cdots & t_{2n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & t_{n-1,n} \\
0 & 0 & 0 & 0 & \sigma
\end{bmatrix}.$$

**Remark.** Obviously, $\sigma$ is the only eigenvalue of $S_T(\sigma)$. So $\rho(S_T(\sigma)) = \sigma$.

**Lemma 3.2.6** Let $T$ be an upper triangular matrix of size $n$. Then $\lim_{m \to \infty} S_T^m(\sigma) = 0$ if and only if $\sigma < 1$.

**Proof.** Since $S_T(\sigma) = \sigma I + S_T(0)$, by the Binomial Theorem, it follows that

$$S_T^m(\sigma) = (\sigma I + S_T(0))^m = \sigma^m I + \sum_{k=1}^{m} \binom{m}{k} \sigma^{m-k} S_T^k(0).$$

Note that $S_T(0)$ is a strictly upper triangular matrix. By the remark of Lemma 1.4.2, $S_T^k(0) = 0$ if $k \geq n$. Then

$$\lim_{m \to \infty} S_T^m(\sigma) = \lim_{m \to \infty} \sigma^m I + \sum_{k=1}^{m} \binom{m}{k} \sigma^{m-k} S_T^k(0)$$

$$= \lim_{m \to \infty} \sigma^m I + \sum_{k=1}^{n} S_T^k(0) \cdot \lim_{m \to \infty} \binom{m}{k} \sigma^{m-k}.$$

For each $k$,

$$0 \leq \binom{m}{k} \sigma^{m-k} = \frac{m(m-1) \cdots (m-k+1)}{k!} \sigma^{m-k} \leq \frac{m^k}{k!} \sigma^{m-k} = \frac{m^k}{\sigma^{k-m}}.$$
If $\sigma < 1$, set $f(x) = x^k \sigma^{-x} = \frac{x^k}{\sigma^{k-x}}, \ k = 1, 2, \cdots, n$. By Hôpital's Rule,

$$\lim_{x \to \infty} f(x) = \lim_{x \to \infty} \frac{x^k}{\sigma^{k-x}} = \lim_{x \to \infty} \frac{kx^{k-1}}{\ln \sigma \cdot \sigma^{k-x}} = \frac{k}{-\ln \sigma} \cdot \lim_{x \to \infty} \frac{x^{k-1}}{\sigma^{k-x}}.$$  

Repeating the Hôpital's Rule, we have

$$\lim_{x \to \infty} f(x) = \frac{k}{-\ln \sigma} \cdot \lim_{x \to \infty} \frac{x^{k-1}}{\sigma^{k-x}} = \frac{k(k-1)}{(\ln \sigma)^2} \cdot \lim_{x \to \infty} \frac{x^{k-2}}{\sigma^{k-x}} = \cdots = \frac{k(k-1) \cdots 2}{(-\ln \sigma)^k} \cdot \lim_{x \to \infty} \frac{x}{\sigma^{k-x}} = \frac{k!}{(-\ln \sigma)^k} \cdot \lim_{x \to \infty} \sigma^{x-k} = 0.$$  

Using Squeeze Theorem, $\lim_{m \to \infty} \left( \begin{array}{c} m \\ k \end{array} \right) \sigma^{m-k} = 0$, $k = 1, 2, \cdots, n$. Therefore,

$$\lim_{m \to \infty} S_T^m(\sigma) = \lim_{m \to \infty} \sigma^m I + \sum_{k=1}^{n} \left[ \lim_{m \to \infty} \left( \begin{array}{c} m \\ k \end{array} \right) \sigma^{m-k} \right] = 0 + 0 = 0.$$  

Now consider the only if part. If $\sigma > 1$, it follows from Theorem 3.2.5 that $\lim_{m \to \infty} S_T^m(\sigma)$ does not exist. If $\sigma = 1$, then every diagonal entry of $S_T^m(\sigma)$ must be 1. We have $S_T^{(\infty)} = \lim_{n \to \infty} S_T^{(m)} = 1$. Note that $\lim_{m \to \infty} A_m = A_\infty = 0$ if and only if $\lim_{n \to \infty} d_{ij}^{(m)} = 0$ for every $i, j = 1, 2, \cdots, n$. Therefore, $S_T^m(\sigma)$ cannot tend to $0$.  \qed
Lemma 3.2.7 Let $T$ be an upper triangular matrix of size $n$. \( \lim_{m \to \infty} T^m = O \) if and only if \( \rho(T) < 1 \).

Proof. Consider the related matrix $S_T(\sigma)$, with $\sigma = \rho(T)$. It is easy to verify that $|t_{ij}| \leq s_{ij}$ for every $i, j$. Denote $(|t_{ij}|)_{\alpha\alpha}$ by $|T|$. Then $|T| \leq S_T(\sigma)$. Comparing the corresponding entries of $T^2$ and $S_T^2(\sigma)$, we have

\[
|t_{ij}^{(2)}| = |\sum_{k=1}^n t_{ik} t_{kj}| \leq \sum_{k=1}^n |t_{ik}| \cdot |t_{kj}| \leq \sum_{k=1}^n s_{ik} s_{kj} = S_{ij}^{(2)}.
\]

That is $|T^2| \leq S_T^2(\sigma)$. Using mathematical induction, we can show that

\[
|T^m| \leq S_T^m(\sigma)
\]

for each positive integer $m$. By Lemma 3.2.6, \( \lim_{m \to \infty} S_T^m(\sigma) = O \) if \( \rho(T) = \sigma < 1 \). Hence,

\[
\lim_{m \to \infty} T^m = O \quad \text{by the Squeeze Theorem.}
\]

Conversely, if \( \lim_{m \to \infty} T^m = O \), then \( \lim_{n \to \infty} t_{ii}^{(m)} = 0 \). Note that \( t_{ii}^{(m)} = t_{ii}^m \), since $T$ is an upper triangular matrix. So \( \lim_{n \to \infty} t_{ii}^m = 0 \), implies $|t_{ii}| < 1$ for all $i$. Recall that upper triangular matrix has eigenvalues along its diagonal. It follows that \( \rho(T) < 1 \). \( \square \)

Theorem 3.2.8 For any square matrix $A$, \( \lim_{m \to \infty} A^m = O \) if and only if \( \rho(A) < 1 \).

Proof. By Schur’s Theorem, there exists a unitary matrix $U$ and an upper triangular matrix $T$, such that $U^H AU = T$. So $A^m = U T^m U^H$. By Theorem 3.2.2 and Corollary 3.2.3, \( \lim_{m \to \infty} A^m = O \) if and only if \( \lim_{m \to \infty} T^m = O \), if and only if \( \rho(T) < 1 \). Note that \( \rho(A) = \rho(T) \), we are done. \( \square \)
For complex number, the geometric series $1 + a + a^2 + \cdots + a^m + \cdots$ converges to $\frac{1}{1-a}$ if $|a|<1$. For matrix series, we have a similar result.

**Theorem 3.2.9** Let $A$ be a square matrix of size $n$. Then the geometric series

$I + A + A^2 + \cdots + A^m + \cdots$ converges if and only if $\rho(A) < 1$. Moreover, if $\rho(A) < 1$, then the series converges to $(I - A)^{-1}$.

**Proof.** Suppose $I + A + A^2 + \cdots + A^m + \cdots$ converges. Then

$$\lim_{m \to \infty} (I + A + A^2 + \cdots + A^m + A^{m+1}) = \lim_{m \to \infty} (I + A + A^2 + \cdots + A^m) .$$

So

$$\lim_{m \to \infty} A^m = O .$$

It follows from Theorem 3.2.8 that $\rho(A) < 1$.

Conversely, if $\rho(A) < 1$, then consider the identity

$$(I - A)(I + A + A^2 + \cdots + A^m) = I - A^{m+1} .$$

Note that $I - A$ is invertible; otherwise $\det(I - A) = 0$ will imply that 1 is an eigenvalue of $A$. We pre-multiply $(I - A)^{-1}$ to both sides of the equality above. Then

$$(I + A + A^2 + \cdots + A^m) = (I - A)^{-1}(I - A^{m+1}) .$$

So
\[
\lim_{m \to \infty} (I + A + A^2 + \cdots A^m) = \lim_{m \to \infty} (I - A)^{-1}(I - A^{m+1}) \\
= (I - A)^{-1} \lim_{m \to \infty} (I - A^{m+1}) \\
= (I - A)^{-1} (I - \lim_{m \to \infty} A^{m+1}).
\]

\[
\lim_{m \to \infty} A^{m+1} = O \text{ because } \rho(A) < 1. \text{ Therefore,}
\]

\[
\lim_{m \to \infty} (I + A + A^2 + \cdots A^m) = (I - A)^{-1}(I - O) = (I - A)^{-1}.
\]

\[\square\]

**Corollary 3.2.10** Let \(T\) be a square upper triangular matrix with \(\rho(T) < 1\). Define \(A = \begin{bmatrix} I & B \\ O & T \end{bmatrix}\).

Then \(A^m\) converges and tends to \(\begin{bmatrix} I & B(I-T)^{-1} \\ O & O \end{bmatrix}\).

**Proof.** \(A = \begin{bmatrix} I & B \\ O & T \end{bmatrix}\), then \(A^2 = \begin{bmatrix} I & B + BT \\ O & T^2 \end{bmatrix} = \begin{bmatrix} I & B(I+T) \\ O & T^2 \end{bmatrix}\). By induction, we can show that \(A^m = \begin{bmatrix} I & B(I+T+T^2 + \cdots + T^{m-1}) \\ O & T^m \end{bmatrix}\) for any positive integer \(m\). Let \(B^{(m)} = B(I+T+T^2 + \cdots + T^{m-1})\). Then \(\lim_{m \to \infty} B^{(m)} = B(I-T)^{-1}\) by Theorem 3.2.9, and \(\lim_{m \to \infty} T^m = O\) by Lemma 3.2.7. Therefore,

\[
\lim_{m \to \infty} A^m = \begin{bmatrix} I & \lim_{m \to \infty} B^{(m)} \\ O & \lim_{m \to \infty} T^m \end{bmatrix} = \begin{bmatrix} I & B(I-T)^{-1} \\ O & O \end{bmatrix}.
\]

\[\square\]
3.3 The Limit of $A^m$ For $\rho(A) = 1$

We will continue the work in Section 3.2 to find the limit of $A^m$ when $\rho(A)=1$. First, we will show that if $A$ has an eigenvalue with $|\lambda_0|=1$ but $\lambda_0 \neq 1$, then $\lim_{m \to \infty} A^m$ does not exists; and then give the sufficient and necessary condition for the convergence of $A^m$ if $\rho(A)=1$.

**Lemma 3.3.1** Suppose that $\lambda_0 \in \mathbb{C}$ is an eigenvalue of square matrix $A$ with $|\lambda_0|=1$, then $\lim_{m \to \infty} A^m$ exists only if $\lambda_0=1$.

**Proof.** By Schur’s Theorem, there exists a unitary $U$ and an upper triangular $T$ such that $U^H A U = T$. So $A^m = UT^m U^H$. Note that $\lambda_0^m$ is an eigenvalue of $T^m$. Then $T^m$ has $\lambda_0^m$ on its diagonal. If $|\lambda_0|=1$ but $\lambda_0 \neq 1$, $\lim_{m \to \infty} A^m$ does not exist. So $\lim_{m \to \infty} T^m$ does not exist by definition. It follows from Corollary 3.2.3 that $\lim_{m \to \infty} A^m$ also does not exist. \(\square\)

**Lemma 3.3.2** Suppose that $\lambda_0=1$ is the only eigenvalue of a square matrix $A$, then $\lim_{m \to \infty} A^m$ exists if and only if $A = I$.

**Proof.** If $A = I$, obviously $\lim_{m \to \infty} A^m = I$.

Assume that $A \neq I$. By Schur’s Theorem, there exists a unitary $U$ and upper triangular $T$ such that $U^H A U = T$. So $U^H A^m U = T^m$. Note that the diagonal entries of $T$ are all 1’s, we can write $T = I + S$ where $S$ is a strictly upper triangular matrix. In the proof of Lemma 3.2.6, we have shown that

$$T^m = I + \sum_{k=1}^{m-1} \binom{m}{k} S^k \quad \text{if } m \geq n \quad (*)$$

60
Note that $S \neq O$; otherwise $T = I$, and then $A = UTU^H = I$. Therefore, there exists $s^{(1)}_{ij} = t_{ij} \neq 0$ for some $i < j$. By (*), the $(i, j)$th entry of $T^m$ can be written as

$$t^{(m)}_{ij} = \binom{m}{1}s^{(1)}_{ij} + \binom{m}{2}s^{(2)}_{ij} + \cdots + \binom{m}{n-1}s^{(n-1)}_{ij}. \quad (m \geq n)$$

Since each $\binom{m}{k} = \frac{m(m-1)(m-2)\cdots(m-k+1)}{k!}$ is a polynomial of degree $k$ with variable $m$ with the $k$th coefficient $\frac{1}{k!}$, then $t^{(m)}_{ij} = a_1m + a_2m^2 + \cdots + a_{n-1}m^{n-1}$ is also a polynomial. We will show that the degree of $t^{(m)}_{ij}$ is at least 1.

Otherwise, suppose all the coefficients $a_1 = a_2 = \cdots = a_{n-1} = 0$. Note that $\binom{m}{n-1}$ is the only term of degree $n-1$ and its $(n-1)$th coefficient is $\frac{1}{(n-1)!}$. So the $(n-1)$th coefficient of $t^{(m)}_{ij}$ is given by $a_{n-1} = \frac{s^{(n-1)}_{ij}}{(n-1)!}$, implies $s^{(n-1)}_{ij} = a_{n-1}(n-1)! = 0$. Hence, $\binom{m}{n-2}$ is the only term of degree $n-2$, and its $(n-2)$th coefficient is $\frac{1}{(n-2)!}$. Then $a_{n-2} = \frac{s^{(n-2)}_{ij}}{(n-2)!}$, implies $s^{(n-2)}_{ij} = a_{n-2}(n-2)! = 0$. By repeating this argument, we arrive at $s^{(1)}_{ij} = 0$, which contradicts with our assumption.

Note that a non-scalar polynomial is unbounded. That is, $t^{(m)}_{ij} \to \infty$ as $m \to +\infty$. Therefore, $\lim_{m \to \infty} T^m$ does not exist. It follows from Corollary 3.2.3 that $\lim_{m \to \infty} A^m$ does not exist.  

**Lemma 3.3.3** If $\lambda_0$ is an eigenvalue of $A$ with geometric multiplicity $r$, then there exists a unitary matrix $U$ and an upper triangular matrix $T$ such that
\[ U^H A U = \begin{bmatrix} \lambda_0 I_r & B \\ O & T \end{bmatrix}. \]

**Proof.** We can use Gram-Schmidt method to find an orthonormal basis of the eigenspace \( \mathcal{E}(\lambda_0) \) of \( A \) corresponding to \( \lambda_0 \), say \( \{u_1, u_2, \ldots, u_r\} \). Now, extend this set into an orthonormal basis of \( \mathbb{C}^n \). Let \( X = [u_1 \quad u_2 \ldots \quad u_r] \), and \( Y \) be the matrix with the remaining \( n - r \) vector as its columns. Then \( X^H X = I_r \) and \( Y^H X = O \), so

\[
\begin{bmatrix} X^H \\ Y^H \end{bmatrix} A \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} X^H AX & X^H AY \\ Y^H AX & Y^H AY \end{bmatrix} = \begin{bmatrix} \lambda_0 X^H X & X^H AY \\ \lambda_0 Y^H X & Y^H AY \end{bmatrix} = \begin{bmatrix} \lambda_0 I_r & B \\ O & T \end{bmatrix}.
\]

By Schur’s Theorem, \( V^H M V = T \) for some unitary matrix \( U \) and upper triangular matrix \( T \).

Define \( U = \begin{bmatrix} X & Y \\ I_r & O \end{bmatrix} \). Then

\[
U^H A U = \begin{bmatrix} I_r & O \\ O & V^H \end{bmatrix} \begin{bmatrix} X^H \\ Y^H \end{bmatrix} A \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} I_r & O \\ O & V^H \end{bmatrix} \begin{bmatrix} \lambda_0 I_r & B \\ O & M \end{bmatrix} = \begin{bmatrix} I_r & B \\ O & V^H M V \end{bmatrix} = \begin{bmatrix} I_r & B \\ O & T \end{bmatrix}.
\]

It is easy to verify that \( U \) is unitary. We complete the proof. \( \square \)
Theorem 3.3.4 Let $A$ be a square matrix with $\rho(A) = 1$, and $\lambda$ be an eigenvalue of $A$. Denote the algebraic multiplicity and geometric multiplicity of $\lambda$ by $a(\lambda)$ and $g(\lambda)$ respectively. Then

$$\lim_{m \to \infty} A^m$$ exists if and only if

a) $\lambda = 1$ is the only eigenvalue with absolute value 1, and

b) $a(1) = g(1)$.

Proof. Let $\lambda_0$ be an eigenvalue of $A$ with $|\lambda_0| = 1$. Then $\lim_{m \to \infty} A^m$ exists implies $\lambda_0 = 1$ by Lemma 3.3.1. For convenience, set $r = g(1)$, and $s = a(1)$. Then by Lemma 3.3.3

$$U^H A U = \begin{bmatrix} I & B \\ O & T \end{bmatrix}$$

for some unitary matrix $U$ and upper triangular matrix $T$. Assume that $r < s$, then there are $s - r$ diagonal entries of $T$ are 1. As reminded in Section 2.3, the order of eigenvalues can be chosen optionally. We can write,

$$T = \begin{bmatrix} T_1 & C \\ O & T_2 \end{bmatrix}$$

where $T_1$ is of size $s - r$ and has only 1’s along the diagonal. Then

$$U^H A U = \begin{bmatrix} I_r & \mathbf{D} & \mathbf{E} \\ O & T_1 & C \\ O & O & T_2 \end{bmatrix}.$$

So

$$U^H A^m U = (U^H A U)^m = \begin{bmatrix} L & D & E \\ O & T_1 & C \\ O & O & T_2 \end{bmatrix}^m = \begin{bmatrix} I & F & G \\ O & T_1^m & H \\ O & O & T_2^m \end{bmatrix}.$$
By Lemma 3.3.2, \( \lim_{m \to \infty} T_m \) exists if and only if \( T_m = I_{s-r} \). Suppose \( T_m = I_{s-r} \). Then by induction, we can verify that for any positive integer \( m \),

\[
U^H A^m U = \begin{bmatrix}
I_r & D & E
\end{bmatrix}^m = \begin{bmatrix}
I_r & mD & G
\end{bmatrix}.
\]

Note that \( D \neq O \); otherwise, there would be \( r + (s - r) = s \) independent eigenvectors of \( A \) corresponding to \( \lambda_0 = 1 \). Then \( g(1) = s = a(1) \), which contradicts with our assumption. \( D \neq O \) shows that \( \lim_{m \to \infty} mD \) does not exists. Hence, \( \lim_{m \to \infty} U^H A^m U \) and so \( \lim_{m \to \infty} A^m \) does not exist. Therefore, \( A^m \) converges unless \( a(1) = g(1) \).

Conversely, suppose that \( \lambda_0 = 1 \) is the only eigenvalue with modulus 1, and \( a(1) = g(1) \). Write

\[
U^H A U = \begin{bmatrix}
I_r & B
\end{bmatrix}.
\]

By Corollary 3.2.10,

\[
\lim_{m \to \infty} U^H A^m U = \lim_{m \to \infty} (U^H A U)^m = \lim_{m \to \infty} \begin{bmatrix}
I_r & B(I + T_2 + T_2^2 + \cdots + T_2^{m-1})
\end{bmatrix} = \begin{bmatrix}
I_r & B(I - T)^{-1}
\end{bmatrix}.
\]

Then

\[
\lim_{m \to \infty} A^m = U \begin{bmatrix}
I_r & B(I - T)^{-1}
\end{bmatrix} U^H.
\]

**Remark.** The theorem shows that, if \( \rho(A) = 1 \) and \( A^\infty = \lim_{m \to \infty} A^m \) exists, then \( a(1) = g(1) = r = \text{rank}(A^\infty) \). Moreover, since \( A^\infty = \lim_{m \to \infty} A^m = \lim_{m \to \infty} A^{m+1} = A \lim_{m \to \infty} A^m = AA^\infty \), the column vectors of \( A^\infty \) are either the zero-vector or an eigenvector of \( A \) corresponding to \( \lambda = 1 \).
3.4 Probability Vectors and Transition Matrices

The probability vector is a set of values representing the likelihood of a certain class of events. The transition matrix, formed by probability vectors, can represent the probability of all such classes. Both of them are widely used in mathematical model of time series. We will use the results stated in last section to prove some attractive properties of them.

**Definition 3.4.1** A column vector \( p \geq 0 \) is a probability vector if all the entries sum to 1.

**Remark.** Suppose \( p \geq 0 \). Then \( p \) is a probability vector if and only if \( 1^T p = 1 \).

**Definition 3.4.2** A square matrix \( A \) is called a transition matrix if all its columns are probability vectors.

**Remark.** Suppose \( A \geq 0 \). Write \( A = [p_1 \quad p_2 \quad \cdots \quad p_n] \). Then \( A \) is a transition matrix if and only if each \( p_i \) is a probability vector, if and only if

\[
1^T A = 1^T [p_1 \quad p_2 \quad \cdots \quad p_n] = [1^T p_1 \quad 1^T p_2 \quad \cdots \quad 1^T p_n] = [1 \quad 1 \quad \cdots \quad 1] = 1^T.
\]

**Theorem 3.4.3** Let \( A \) be a transition matrix. Then for any positive integer \( m \), \( A^m \) is a transition matrix.

**Proof.** We show this by induction. It is trivial when \( m = 1 \). Assume that \( A^m \) is a transition matrix for some positive integer \( m \). Then \( 1^T A^{m+1} = (1^T A^m)A = 1^T A = 1^T \). It is easy to see that \( A^{m+1} = A^m A \geq 0 \). So \( A^{m+1} \) is also a transition matrix. \( \square \)
Theorem 3.4.4 Let $A$ be a transition matrix and $p$ a probability vector. Then $Ap$ is also a probability vector.

Proof. $A \geq 0$ and $p \geq 0$ imply that $Ap \geq 0$. Since $1^T (Ap) = (1^T A)p = 1^T p = 1^T$, we know that $Ap$ is also a probability vector. \qed

Comparing Definition 3.1.5 and Definition 3.4.1, $A$ is a transition matrix if and only if $A^T$ is a stochastic matrix. Note that $A$ and $A^T$ have the same eigenvalues. By Theorem 3.1.6, if the diagonal entries of $A$ are all positive, then 1 is the only eigenvalue of $A$ with absolute value 1. Moreover, if $A$ is positive, we have the following theorem.

Theorem 3.4.5 Suppose that $A > 0$ is a transition matrix. Then the eigenspace associated with $\lambda = 1$ is one-dimensional.

Proof. Note that $A$ and $A^T$ have the same eigenvalues and the corresponding geometric multiplicities. We need to show the statement holds for stochastic matrix. For convenience, set $S = A^T$. Suppose $(1, u)$ is an eigenpair of $S$ and we will show that $u$ is a scalar multiple of $S$.

Let $u = [u_1 \ u_2 \ \cdots \ u_n]^T$. Suppose $|u_k| \geq |u_i|$ for each $i = 1, 2, \cdots, n$. Then consider the $k$th equation of the system $Su = u$:

$$s_{k1}u_1 + s_{k2}u_2 + \cdots + s_{kn}u_n = u_k.$$

By the Triangle Inequality,
\[ |U_k| = |s_{k1}u_1 + s_{k2}u_2 + \cdots + s_{kn}u_n| \]
\[ \leq s_{k1}|u_1| + s_{k2}|u_2| + \cdots + s_{kn}|u_n| \]
\[ \leq (s_{k1} + s_{k2} + \cdots + s_{kn})|u_k| \]
\[ = |u_k|. \]

We deduce that
\[ |s_{k1}u_1 + s_{k2}u_2 + \cdots + s_{kn}u_n| \]
\[ = s_{k1}|u_1| + s_{k2}|u_2| + \cdots + s_{kn}|u_n| \]
\[ = (s_{k1} + s_{k2} + \cdots + s_{kn})|u_k|. \]

The second equality shows that
\[ |u_1| = |u_2| = \cdots = |u_n|. \]

Note that \(|u_1| \neq 0\), since \(u \neq 0\); and \(s_{k1}, s_{k2}, \ldots, s_{kn}\) are positive numbers. The first equality gives \(\arg(u_1) = \arg(u_2) = \cdots = \arg(u_n)\). Then, we have \(u_1 = u_2 = \cdots = u_n\), so \(u = u_11\). Therefore, the eigenspace \(E(1)\) of \(S\) is one-dimensional. So \(g_A(1) = g_S(1) = 1\). \(\square\)

**Remark.** In this theorem, the condition that \(A > O\) is necessary. For example, if \(A = I_n\), then \(A\) is a transition matrix but \(g_A(1) = n\).

**Definition 3.4.6** Let \(A\) be a transition matrix and \(p\) a probability vector. Then \(p\) is called a **stationary vector** of \(A\) if \(p\) is an eigenvector of \(A\) corresponding to the eigenvalue \(\lambda = 1\).
Remark. It is not obviously that the stationary vector exists by definition. However, it does for $A > O$. Let $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$ and $p = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix}$ with $p_1 + p_2 + \cdots + p_n = 1$, such that $Ap = p$.

Suppose that $p_1 < 0$. Then from the first equation $a_{11}p_1 + a_{12}p_2 + \cdots + a_{1n}p_n = p_1$, we have $a_{11}p_1 + a_{12}p_2 + \cdots + a_{1n}p_n = p_1(1 - a_{11}) < 0$. Since $a_{ii} > 0$ for all $i$, there must be some $p_k < 0$, $k \geq 2$.

Suppose $p_2 < 0$. Consider the second equation. There must be another $p_{k'} < 0$, $k' \geq 3$. Suppose $p_3 < 0$. Repeating this process, finally we will get $p_i < 0$ for each $i = 1, 2, \cdots, n$, contradicting with $p_1 + p_2 + \cdots + p_n = 1$.

**Theorem 3.4.7** Let $A$ be a transition matrix. Suppose that $A^m = \lim_{m \to \infty} A^m$ exists and equals to $A^\infty$.

Then $A^\infty$ is a transition matrix whose columns are stationary vectors of $A$.

**Proof.** $A \geq O$ implies $A^m \geq O$. By theorem 2.4.3, $A^m$ is a transition matrix for every positive integer $m$, so $1^T A^\infty = 1^T \lim_{m \to \infty} A^m = \lim_{m \to \infty} (1^T A^m) = \lim_{m \to \infty} (1^T) = 1^T$. Therefore, $A^\infty$ is also a transition matrix. All the columns vectors of $A^\infty$ are probability vectors. Note that $AA^\infty = A \lim_{m \to \infty} A^m = \lim_{m \to \infty} A^{m+1} = \lim_{m \to \infty} A^m = A^\infty$, so each column vector of $A^\infty$ is the eigenvector of $A$ corresponding to $\lambda = 1$. Therefore, these columns are the stationary vectors of $A$. \[\square\]

The above theorem describes the limit of $A^m$ for transition matrix $A$ if the limit exists. The following theorem will show that the limit does exist for positive transition matrices.
Theorem 3.4.8 Let \( A > O \) be a transition matrix. Then \( A^\infty = \lim_{m \to \infty} A^m \) exists. Moreover, \( A^\infty \) is a transition matrix with \( \operatorname{rank}(A^\infty) = 1 \).

Proof. Recall that \( \rho(A) = 1 \), so \( A^\infty \) exists if and only if \( A \) satisfies the two properties stated in Theorem 3.3.4. Note that \( A \) and \( A^T \) have the same eigenvalues and the corresponding algebraic and geometric multiplicities. Also, \( A^T > O \) is a stochastic matrix. Then by Theorem 3.1.6, \( \rho(A) = \rho(A^T) = 1 \), and \( \lambda = 1 \) is the only eigenvalue of \( A \) with modulus 1. Theorem 3.4.5 says that \( g(1) = 1 \), so we only need to show \( a(1) = 1 \).

Set \( S = A^T \). Write the characteristic polynomial of \( S \) as

\[
|\lambda I - S| = a_0 + a_1(\lambda - 1) + a_2(\lambda - 1)^2 + a_n(\lambda - 1)^n.
\]

Obviously, \( a_0 = 0 \) since \( \lambda = 1 \) is an eigenvalue of \( S \). Now we shall show \( a_1 > 0 \), and thus will imply that the algebraic multiplicity of \( \lambda = 1 \) is exactly 1. Write

\[
S = \begin{bmatrix}
S_{11} & S_{12} & S_{13} & \cdots & S_{1n} \\
S_{21} & S_{22} & S_{23} & \cdots & S_{2n} \\
S_{31} & S_{32} & S_{33} & \cdots & S_{3n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
S_{n1} & S_{n2} & S_{n3} & \cdots & S_{nn}
\end{bmatrix} = \begin{bmatrix}
1 - \sum_{k \neq 1} S_{1k} & S_{12} & S_{13} & \cdots & S_{1n} \\
S_{21} & 1 - \sum_{k \neq 2} S_{1k} & S_{23} & \cdots & S_{2n} \\
S_{31} & S_{32} & 1 - \sum_{k \neq 3} S_{1k} & \cdots & S_{3n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
S_{n1} & S_{n2} & S_{n3} & \cdots & 1 - \sum_{k \neq n} S_{1k}
\end{bmatrix}.
\]

Then

\[
|\lambda I - S| = \det \begin{bmatrix}
(\lambda - 1) + \sum_{k \neq 1} S_{1k} & -S_{12} & -S_{13} & \cdots & -S_{1n} \\
S_{21} & (\lambda - 1) + \sum_{k \neq 2} S_{2k} & -S_{23} & \cdots & -S_{2n} \\
-S_{31} & -S_{32} & (\lambda - 1) + \sum_{k \neq 3} S_{3k} & \cdots & -S_{3n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
-S_{n1} & -S_{n2} & -S_{n3} & \cdots & (\lambda - 1) + \sum_{k \neq n} S_{nk}
\end{bmatrix}.
\]
Note that the coefficient of $\lambda -1$ is given by

$$\sum_{k \neq 2} S_{2k} - S_{23} \cdots - S_{2n} \quad \sum_{k \neq 1} S_{1k} - S_{13} \cdots - S_{1n} \quad \sum_{k \neq 1} S_{1k} - S_{12} \cdots - S_{1,n-1} \quad \sum_{k \neq 1} S_{1k} - S_{12} \cdots - S_{1,n-1}$$

Denote the first matrix by $M_1$. Suppose $\lambda$ is an eigenvalue of $M_1$. Then by Gerschgorin’s Theorem, there exists an $i \in \{2, 3, \cdots, n\}$, such that $|\lambda - \sum_{k \neq i} S_{ik}| \leq r_i = \sum_{k \neq i} S_{ik} - \sum_{k \neq i} S_{ik}$. This is a disk with its center on $\sum_{k \neq i} S_{ik}$ and radius less than $\sum_{k \neq i} S_{ik}$.

From the graph, we can see that every eigenvalue has positive real part. Note that if $\lambda$ is an eigenvalue, then $\bar{\lambda}$ is also an eigenvalue since $M_1$ is a real matrix. From the fact that $\lambda \bar{\lambda} = |\lambda|^2 > 0$ and $\det(M_1)$ is the product of its eigenvalues, we know that $\det(M_1) > 0$. Similarly, $\det(M_2), \det(M_3), \cdots, \det(M_n)$ are all positive, so

$$a_i = \det(M_1) + \det(M_2) + \cdots + \det(M_n) > 0.$$
Therefore, $a_\lambda(l) = a_s(l) = 1$. We conclude that $A^\infty = \lim_{m \to \infty} A^m$ exists. Moreover, $A^\infty$ is a transition matrix and. By remark of Theorem 3.3.4, rank($A^\infty$) = $g(l) = 1$. □

**Remark.** $A^\infty$ is a transition matrix with rank($A^\infty$) = 1. Therefore, the columns of $A^\infty$ are identical. Write $A^\infty = [p \ p \ \cdots \ p]$, where $p$ is a stationary matrix of $A$. Theorem 7.4.4 shows that the eigenspace associated with $\lambda = 1$ is one-dimensional. Then such $p$ must be unique. Thus, we can write $A^\infty = p1^T$ where $p$ is the unique stationary vector of $A$.

However, the statement may not hold for every transition matrix. For example, $\lambda = -1$ is an eigenvalue of $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, so $A^\infty$ does not exist.

**Corollary 7.4.9** Let $A > O$ be a matrix whose rows and columns all sum to 1. Then $A^\infty = \frac{1}{n} J_n$.

**Proof.** It is easy to verify that $A \frac{1}{n} - I = \frac{1}{n} I$, so $\frac{1}{n} I$ is the unique stationary vector of $A$, and then $A^\infty = (\frac{1}{n} I)^T = \frac{1}{n} J_n$. □
REFERENCE

