8.7 Complex Matrices

If A is an $n \times n$ matrix, the characteristic polynomial $c_A(x)$ is a polynomial of degree n and the eigenvalues of A are just the roots of $c_A(x)$. In most of our examples these roots have been *real* numbers (in fact, the examples have been carefully chosen so this will be the case!); but it need not happen, even when the characteristic polynomial has real coefficients. For example, if $A = \left[\begin{array}{rr} 0 & 1 \\ -1 & 0 \end{array} \right]$ then $c_A(x) = x^2 + 1$ has roots i and $-i$, where i is a complex number satisfying $i^2 = -1$. Therefore, we have to deal with the possibility that the eigenvalues of a (real) square matrix might be complex numbers.

In fact, nearly everything in this book would remain true if the phrase *real number* were replaced by *complex number* wherever it occurs. Then we would deal with matrices with complex entries, systems of linear equations with complex coefficients (and complex solutions), determinants of complex matrices, and vector spaces with scalar multiplication by any complex number allowed. Moreover, the proofs of most theorems about (the real version of) these concepts extend easily to the complex case. It is not our intention here to give a full treatment of complex linear algebra. However, we will carry the theory far enough to give another proof that the eigenvalues of a real symmetric matrix A are real (Theorem 5.5.7) and to prove the spectral theorem, an extension of the principal axes theorem (Theorem 8.2.2).

The set of complex numbers is denoted $\mathbb{C}$. We will use only the most basic properties of these numbers (mainly conjugation and absolute values), and the reader can find this material in Appendix A.

If $n \geq 1$, we denote the set of all n-tuples of complex numbers by $\mathbb{C}^n$. As with $\mathbb{R}^n$, these n-tuples will be written either as row or column matrices and will be referred to as **vectors**. We define vector operations on $\mathbb{C}^n$ as follows:

$$\begin{aligned} (v_1, v_2, \dots, v_n) + (w_1, w_2, \dots, w_n) &= (v_1 + w_1, v_2 + w_2, \dots, v_n + w_n) \\ u(v_1, v_2, \dots, v_n) &= (uv_1, uv_2, \dots, uv_n) \quad \text{for } u \text{ in } \mathbb{C} \end{aligned}$$

With these definitions, $\mathbb{C}^n$ satisfies the axioms for a vector space (with complex scalars) given in Chapter 6. Thus we can speak of spanning sets for $\mathbb{C}^n$, of linearly independent subsets, and of bases. In all cases, the definitions are identical to the real case, except that the scalars are allowed to be complex numbers. In particular, the standard basis of $\mathbb{R}^n$ remains a basis of $\mathbb{C}^n$, called the **standard basis** of $\mathbb{C}^n$.

A matrix $A = \left[a_{ij} \right]$ is called a **complex matrix** if every entry a_{ij} is a complex number. The notion of conjugation for complex numbers extends to matrices as follows: Define the **conjugate** of $A = \left[a_{ij} \right]$ to be the matrix

$$\overline{A} = \left[\begin{array}{c} \overline{a}_{ij} \end{array} \right]$$

obtained from A by conjugating every entry. Then (using Appendix A)

$$\overline{A + B} = \overline{A} + \overline{B} \quad \text{and} \quad \overline{AB} = \overline{A}\,\overline{B}$$

holds for all (complex) matrices of appropriate size.

The Standard Inner Product

There is a natural generalization to $\mathbb{C}^n$ of the dot product in $\mathbb{R}^n$.

Definition 8.15 Standard Inner Product in $\mathbb{R}^n$

Given $\mathbf{z} = (z_1, z_2, \dots, z_n)$ *and* $\mathbf{w} = (w_1, w_2, \dots, w_n)$ *in* $\mathbb{C}^n$, *define their* ***standard inner product*** $\langle \mathbf{z}, \mathbf{w} \rangle$ *by*

$$\langle \mathbf{z}, \mathbf{w} \rangle = z_1\overline{w}_1 + z_2\overline{w}_2 + \cdots + z_n\overline{w}_n = \mathbf{z} \cdot \overline{\mathbf{w}}$$

where $\overline{w}$ *is the conjugate of the complex number* w.

Clearly, if $\mathbf{z}$ and $\mathbf{w}$ actually lie in $\mathbb{R}^n$, then $\langle \mathbf{z}, \mathbf{w} \rangle = \mathbf{z} \cdot \mathbf{w}$ is the usual dot product.

Example 8.7.1

If $\mathbf{z} = (2, 1-i, 2i, 3-i)$ and $\mathbf{w} = (1-i, -1, -i, 3+2i)$, then

$$\begin{aligned}\langle \mathbf{z}, \mathbf{w} \rangle &= 2(1+i) + (1-i)(-1) + (2i)(i) + (3-i)(3-2i) = 6-6i \\ \langle \mathbf{z}, \mathbf{z} \rangle &= 2 \cdot 2 + (1-i)(1+i) + (2i)(-2i) + (3-i)(3+i) = 20\end{aligned}$$

Note that $\langle \mathbf{z}, \mathbf{w} \rangle$ is a complex number in general. However, if $\mathbf{w} = \mathbf{z} = (z_1, z_2, \dots, z_n)$, the definition gives $\langle \mathbf{z}, \mathbf{z} \rangle = |z_1|^2 + \cdots + |z_n|^2$ which is a nonnegative real number, equal to 0 if and only if $\mathbf{z} = \mathbf{0}$. This explains the conjugation in the definition of $\langle \mathbf{z}, \mathbf{w} \rangle$, and it gives (4) of the following theorem.

Theorem 8.7.1

Let $\mathbf{z}$, $\mathbf{z}_1$, $\mathbf{w}$, *and* $\mathbf{w}_1$ *denote vectors in* $\mathbb{C}^n$, *and let* λ *denote a complex number.*

1. $\langle \mathbf{z} + \mathbf{z}_1, \mathbf{w} \rangle = \langle \mathbf{z}, \mathbf{w} \rangle + \langle \mathbf{z}_1, \mathbf{w} \rangle$ *and* $\langle \mathbf{z}, \mathbf{w} + \mathbf{w}_1 \rangle = \langle \mathbf{z}, \mathbf{w} \rangle + \langle \mathbf{z}, \mathbf{w}_1 \rangle$.
2. $\langle \lambda\mathbf{z}, \mathbf{w} \rangle = \lambda \langle \mathbf{z}, \mathbf{w} \rangle$ *and* $\langle \mathbf{z}, \lambda\mathbf{w} \rangle = \overline{\lambda} \langle \mathbf{z}, \mathbf{w} \rangle$.
3. $\langle \mathbf{z}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{z} \rangle}$.
4. $\langle \mathbf{z}, \mathbf{z} \rangle \geq 0$, *and* $\langle \mathbf{z}, \mathbf{z} \rangle = 0$ *if and only if* $\mathbf{z} = \mathbf{0}$.

Proof. We leave (1) and (2) to the reader (Exercise 8.7.10), and (4) has already been proved. To prove (3), write $\mathbf{z} = (z_1, z_2, \dots, z_n)$ and $\mathbf{w} = (w_1, w_2, \dots, w_n)$. Then

$$\begin{aligned}\overline{\langle \mathbf{w}, \mathbf{z} \rangle} = (\overline{w_1\overline{z}_1 + \cdots + w_n\overline{z}_n}) &= \overline{w}_1\overline{\overline{z}}_1 + \cdots + \overline{w}_n\overline{\overline{z}}_n \\ &= z_1\overline{w}_1 + \cdots + z_n\overline{w}_n = \langle \mathbf{z}, \mathbf{w} \rangle\end{aligned}$$

□

Definition 8.16 Norm and Length in $\mathbb{C}^n$

*As for the dot product on $\mathbb{R}^n$, property (4) enables us to define the **norm** or **length** $\|\mathbf{z}\|$ of a vector $\mathbf{z} = (z_1, z_2, \dots, z_n)$ in $\mathbb{C}^n$:*

$$\|\mathbf{z}\| = \sqrt{\langle \mathbf{z}, \mathbf{z} \rangle} = \sqrt{|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2}$$

The only properties of the norm function we will need are the following (the proofs are left to the reader):

Theorem 8.7.2

If $\mathbf{z}$ is any vector in $\mathbb{C}^n$, then

1. $\|\mathbf{z}\| \geq 0$ *and* $\|\mathbf{z}\| = 0$ *if and only if* $\mathbf{z} = \mathbf{0}$.
2. $\|\lambda \mathbf{z}\| = |\lambda| \|\mathbf{z}\|$ *for all complex numbers* λ.

A vector $\mathbf{u}$ in $\mathbb{C}^n$ is called a **unit vector** if $\|\mathbf{u}\| = 1$. Property (2) in Theorem 8.7.2 then shows that if $\mathbf{z} \neq \mathbf{0}$ is any nonzero vector in $\mathbb{C}^n$, then $\mathbf{u} = \frac{1}{\|\mathbf{z}\|}\mathbf{z}$ is a unit vector.

Example 8.7.2

In $\mathbb{C}^4$, find a unit vector $\mathbf{u}$ that is a positive real multiple of $\mathbf{z} = (1 - i, i, 2, 3 + 4i)$.

Solution. $\|\mathbf{z}\| = \sqrt{2 + 1 + 4 + 25} = \sqrt{32} = 4\sqrt{2}$, so take $\mathbf{u} = \frac{1}{4\sqrt{2}}\mathbf{z}$.

Transposition of complex matrices is defined just as in the real case, and the following notion is fundamental.

Definition 8.17 Conjugate Transpose in $\mathbb{C}^n$

*The **conjugate transpose** A^H of a complex matrix A is defined by*

$$A^H = (\overline{A})^T = \overline{(A^T)}$$

Observe that $A^H = A^T$ when A is real.[14]

Example 8.7.3

$$\begin{bmatrix} 3 & 1 - i & 2 + i \\ 2i & 5 + 2i & -i \end{bmatrix}^H = \begin{bmatrix} 3 & -2i \\ 1 + i & 5 - 2i \\ 2 - i & i \end{bmatrix}$$

[14]Other notations for A^H are A^* and $A^\dagger$.

The following properties of A^H follow easily from the rules for transposition of real matrices and extend these rules to complex matrices. Note the conjugate in property (3).

Theorem 8.7.3

Let A and B denote complex matrices, and let λ be a complex number.

1. $(A^H)^H = A$.
2. $(A+B)^H = A^H + B^H$.
3. $(\lambda A)^H = \overline{\lambda} A^H$.
4. $(AB)^H = B^H A^H$.

Hermitian and Unitary Matrices

If A is a real symmetric matrix, it is clear that $A^H = A$. The complex matrices that satisfy this condition turn out to be the most natural generalization of the real symmetric matrices:

Definition 8.18 Hermitian Matrices

A square complex matrix A is called **hermitian**[15] *if $A^H = A$, equivalently if $\overline{A} = A^T$.*

Hermitian matrices are easy to recognize because the entries on the main diagonal must be real, and the "reflection" of each nondiagonal entry in the main diagonal must be the conjugate of that entry.

Example 8.7.4

$\left[\begin{array}{ccc} 3 & i & 2+i \\ -i & -2 & -7 \\ 2-i & -7 & 1 \end{array}\right]$ is hermitian, whereas $\left[\begin{array}{cc} 1 & i \\ i & -2 \end{array}\right]$ and $\left[\begin{array}{cc} 1 & i \\ -i & i \end{array}\right]$ are not.

The following Theorem extends Theorem 8.2.3, and gives a very useful characterization of hermitian matrices in terms of the standard inner product in $\mathbb{C}^n$.

Theorem 8.7.4

An $n \times n$ complex matrix A is hermitian if and only if

$$\langle A\mathbf{z},\ \mathbf{w}\rangle = \langle \mathbf{z},\ A\mathbf{w}\rangle$$

for all n-tuples $\mathbf{z}$ and $\mathbf{w}$ in $\mathbb{C}^n$.

[15] The name hermitian honours Charles Hermite (1822–1901), a French mathematician who worked primarily in analysis and is remembered as the first to show that the number e from calculus is transcendental—that is, e is not a root of any polynomial with integer coefficients.

Proof. If A is hermitian, we have $A^T = \overline{A}$. If $\mathbf{z}$ and $\mathbf{w}$ are columns in $\mathbb{C}^n$, then $\langle \mathbf{z}, \mathbf{w} \rangle = \mathbf{z}^T \overline{\mathbf{w}}$, so

$$\langle A\mathbf{z}, \mathbf{w} \rangle = (A\mathbf{z})^T \overline{\mathbf{w}} = \mathbf{z}^T A^T \overline{\mathbf{w}} = \mathbf{z}^T \overline{A} \overline{\mathbf{w}} = \mathbf{z}^T (\overline{A\mathbf{w}}) = \langle \mathbf{z}, A\mathbf{w} \rangle$$

To prove the converse, let $\mathbf{e}_j$ denote column j of the identity matrix. If $A = \left[a_{ij}\right]$, the condition gives

$$\overline{a}_{ij} = \langle \mathbf{e}_i, A\mathbf{e}_j \rangle = \langle A\mathbf{e}_i, \mathbf{e}_j \rangle = a_{ij}$$

Hence $\overline{A} = A^T$, so A is hermitian. □

Let A be an $n \times n$ complex matrix. As in the real case, a complex number λ is called an **eigenvalue** of A if $A\mathbf{x} = \lambda \mathbf{x}$ holds for some column $\mathbf{x} \neq \mathbf{0}$ in $\mathbb{C}^n$. In this case $\mathbf{x}$ is called an **eigenvector** of A corresponding to λ. The **characteristic polynomial** $c_A(x)$ is defined by

$$c_A(x) = \det(xI - A)$$

This polynomial has complex coefficients (possibly nonreal). However, the proof of Theorem 3.3.2 goes through to show that the eigenvalues of A are the roots (possibly complex) of $c_A(x)$.

It is at this point that the advantage of working with complex numbers becomes apparent. The real numbers are incomplete in the sense that the characteristic polynomial of a real matrix may fail to have all its roots real. However, this difficulty does not occur for the complex numbers. The so-called fundamental theorem of algebra ensures that *every* polynomial of positive degree with complex coefficients has a complex root. Hence every square complex matrix A has a (complex) eigenvalue. Indeed (Appendix A), $c_A(x)$ factors completely as follows:

$$c_A(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n)$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A (with possible repetitions due to multiple roots).

The next result shows that, for hermitian matrices, the eigenvalues are actually real. Because symmetric real matrices are hermitian, this re-proves Theorem 5.5.7. It also extends Theorem 8.2.4, which asserts that eigenvectors of a symmetric real matrix corresponding to distinct eigenvalues are actually orthogonal. In the complex context, two n-tuples $\mathbf{z}$ and $\mathbf{w}$ in $\mathbb{C}^n$ are said to be **orthogonal** if $\langle \mathbf{z}, \mathbf{w} \rangle = 0$.

Theorem 8.7.5

Let A denote a hermitian matrix.

1. *The eigenvalues of A are real.*
2. *Eigenvectors of A corresponding to distinct eigenvalues are orthogonal.*

Proof. Let λ and μ be eigenvalues of A with (nonzero) eigenvectors $\mathbf{z}$ and $\mathbf{w}$. Then $A\mathbf{z} = \lambda \mathbf{z}$ and $A\mathbf{w} = \mu \mathbf{w}$, so Theorem 8.7.4 gives

$$\lambda \langle \mathbf{z}, \mathbf{w} \rangle = \langle \lambda \mathbf{z}, \mathbf{w} \rangle = \langle A\mathbf{z}, \mathbf{w} \rangle = \langle \mathbf{z}, A\mathbf{w} \rangle = \langle \mathbf{z}, \mu \mathbf{w} \rangle = \overline{\mu} \langle \mathbf{z}, \mathbf{w} \rangle \tag{8.6}$$

If $\mu = \lambda$ and $\mathbf{w} = \mathbf{z}$, this becomes $\lambda \langle \mathbf{z}, \mathbf{z} \rangle = \overline{\lambda} \langle \mathbf{z}, \mathbf{z} \rangle$. Because $\langle \mathbf{z}, \mathbf{z} \rangle = \|\mathbf{z}\|^2 \neq 0$, this implies $\lambda = \overline{\lambda}$. Thus λ is real, proving (1). Similarly, μ is real, so equation (8.6) gives $\lambda \langle \mathbf{z}, \mathbf{w} \rangle = \mu \langle \mathbf{z}, \mathbf{w} \rangle$. If $\lambda \neq \mu$, this implies $\langle \mathbf{z}, \mathbf{w} \rangle = 0$, proving (2). □

The principal axes theorem (Theorem 8.2.2) asserts that every real symmetric matrix A is orthogonally diagonalizable—that is $P^T AP$ is diagonal where P is an orthogonal matrix $(P^{-1} = P^T)$. The next theorem identifies the complex analogs of these orthogonal real matrices.

Definition 8.19 Orthogonal and Orthonormal Vectors in $\mathbb{C}^n$

*As in the real case, a set of nonzero vectors $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m\}$ in $\mathbb{C}^n$ is called **orthogonal** if $\langle \mathbf{z}_i, \mathbf{z}_j \rangle = 0$ whenever $i \neq j$, and it is **orthonormal** if, in addition, $\|\mathbf{z}_i\| = 1$ for each i.*

Theorem 8.7.6

The following are equivalent for an $n \times n$ complex matrix A.

1. *A is invertible and $A^{-1} = A^H$.*
2. *The rows of A are an orthonormal set in $\mathbb{C}^n$.*
3. *The columns of A are an orthonormal set in $\mathbb{C}^n$.*

Proof. If $A = \left[\begin{array}{cccc} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{array}\right]$ is a complex matrix with jth column $\mathbf{c}_j$, then $A^T \overline{A} = \left[\langle \mathbf{c}_i, \mathbf{c}_j \rangle\right]$, as in Theorem 8.2.1. Now (1) $\Leftrightarrow$ (2) follows, and (1) $\Leftrightarrow$ (3) is proved in the same way. □

Definition 8.20 Unitary Matrices

*A square complex matrix U is called **unitary** if $U^{-1} = U^H$.*

Thus a real matrix is unitary if and only if it is orthogonal.

Example 8.7.5

The matrix $A = \left[\begin{array}{cc} 1+i & 1 \\ 1-i & i \end{array}\right]$ has orthogonal columns, but the rows are not orthogonal.

Normalizing the columns gives the unitary matrix $\frac{1}{2}\left[\begin{array}{cc} 1+i & \sqrt{2} \\ 1-i & \sqrt{2}i \end{array}\right]$.

Given a real symmetric matrix A, the diagonalization algorithm in Section 3.3 leads to a procedure for finding an orthogonal matrix P such that $P^T AP$ is diagonal (see Example 8.2.4). The following example illustrates Theorem 8.7.5 and shows that the technique works for complex matrices.

Example 8.7.6

Consider the hermitian matrix $A = \left[\begin{array}{cc} 3 & 2+i \\ 2-i & 7 \end{array}\right]$. Find the eigenvalues of A, find two orthonormal eigenvectors, and so find a unitary matrix U such that $U^H AU$ is diagonal.

Solution. The characteristic polynomial of A is

$$c_A(x) = \det(xI - A) = \det\left[\begin{array}{cc} x-3 & -2-i \\ -2+i & x-7 \end{array}\right] = (x-2)(x-8)$$

Hence the eigenvalues are 2 and 8 (both real as expected), and corresponding eigenvectors are $\begin{bmatrix} 2+i \\ -1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 2-i \end{bmatrix}$ (orthogonal as expected). Each has length $\sqrt{6}$ so, as in the (real) diagonalization algorithm, let $U = \frac{1}{\sqrt{6}} \begin{bmatrix} 2+i & 1 \\ -1 & 2-i \end{bmatrix}$ be the unitary matrix with the normalized eigenvectors as columns.

Then $U^H AU = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$ is diagonal.

Unitary Diagonalization

An $n \times n$ complex matrix A is called **unitarily diagonalizable** if $U^H AU$ is diagonal for some unitary matrix U. As Example 8.7.6 suggests, we are going to prove that every hermitian matrix is unitarily diagonalizable. However, with only a little extra effort, we can get a very important theorem that has this result as an easy consequence.

A complex matrix is called **upper triangular** if every entry below the main diagonal is zero. We owe the following theorem to Issai Schur.[16]

Theorem 8.7.7: Schur's Theorem

If A is any $n \times n$ complex matrix, there exists a unitary matrix U such that

$$U^H AU = T$$

is upper triangular. Moreover, the entries on the main diagonal of T are the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ of A (including multiplicities).

Proof. We use induction on n. If $n = 1$, A is already upper triangular. If $n > 1$, assume the theorem is valid for $(n-1) \times (n-1)$ complex matrices. Let λ_1 be an eigenvalue of A, and let $\mathbf{y}_1$ be an eigenvector with $\|\mathbf{y}_1\| = 1$. Then $\mathbf{y}_1$ is part of a basis of $\mathbb{C}^n$ (by the analog of Theorem 6.4.1), so the (complex analog of the) Gram-Schmidt process provides $\mathbf{y}_2, \dots, \mathbf{y}_n$ such that $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$ is an orthonormal basis of $\mathbb{C}^n$. If $U_1 = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_n \end{bmatrix}$ is the matrix with these vectors as its columns, then (see Lemma 5.4.3)

$$U_1^H AU_1 = \begin{bmatrix} \lambda_1 & X_1 \\ 0 & A_1 \end{bmatrix}$$

in block form. Now apply induction to find a unitary $(n-1) \times (n-1)$ matrix W_1 such that $W_1^H A_1 W_1 = T_1$ is upper triangular. Then $U_2 = \begin{bmatrix} 1 & 0 \\ 0 & W_1 \end{bmatrix}$ is a unitary $n \times n$ matrix. Hence $U = U_1 U_2$ is unitary (using Theorem 8.7.6), and

$$U^H AU = U_2^H (U_1^H AU_1) U_2$$

[16]Issai Schur (1875–1941) was a German mathematician who did fundamental work in the theory of representations of groups as matrices.

$$= \begin{bmatrix} 1 & 0 \\ 0 & W_1^H \end{bmatrix} \begin{bmatrix} \lambda_1 & X_1 \\ 0 & A_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & W_1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & X_1W_1 \\ 0 & T_1 \end{bmatrix}$$

is upper triangular. Finally, A and $U^H AU = T$ have the same eigenvalues by (the complex version of) Theorem 5.5.1, and they are the diagonal entries of T because T is upper triangular. □

The fact that similar matrices have the same traces and determinants gives the following consequence of Schur's theorem.

Corollary 8.7.1

Let A be an $n \times n$ complex matrix, and let $\lambda_1, \lambda_2, \dots, \lambda_n$ denote the eigenvalues of A, including multiplicities. Then

$$\det A = \lambda_1\lambda_2\cdots\lambda_n \quad \textit{and} \quad \operatorname{tr} A = \lambda_1 + \lambda_2 + \cdots + \lambda_n$$

Schur's theorem asserts that every complex matrix can be "unitarily triangularized." However, we cannot substitute "unitarily diagonalized" here. In fact, if $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, there is no invertible complex matrix U at all such that $U^{-1}AU$ is diagonal. However, the situation is much better for hermitian matrices.

Theorem 8.7.8: Spectral Theorem

If A is hermitian, there is a unitary matrix U such that $U^H AU$ is diagonal.

Proof. By Schur's theorem, let $U^H AU = T$ be upper triangular where U is unitary. Since A is hermitian, this gives

$$T^H = (U^H AU)^H = U^H A^H U^{HH} = U^H AU = T$$

This means that T is both upper and lower triangular. Hence T is actually diagonal. □

The principal axes theorem asserts that a real matrix A is symmetric if and only if it is orthogonally diagonalizable (that is, $P^T AP$ is diagonal for some real orthogonal matrix P). Theorem 8.7.8 is the complex analog of half of this result. However, the converse is false for complex matrices: There exist unitarily diagonalizable matrices that are not hermitian.

Example 8.7.7

Show that the non-hermitian matrix $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ is unitarily diagonalizable.

Solution. The characteristic polynomial is $c_A(x) = x^2 + 1$. Hence the eigenvalues are i and $-i$, and it is easy to verify that $\begin{bmatrix} i \\ -1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ i \end{bmatrix}$ are corresponding eigenvectors. Moreover, these eigenvectors are orthogonal and both have length $\sqrt{2}$, so $U = \frac{1}{\sqrt{2}} \begin{bmatrix} i & -1 \\ -1 & i \end{bmatrix}$ is a unitary matrix such that $U^HAU = \left[\begin{array}{cc} i & 0 \\ 0 & -i \end{array} \right]$ is diagonal.

There is a very simple way to characterize those complex matrices that are unitarily diagonalizable. To this end, an $n \times n$ complex matrix N is called **normal** if $NN^H = N^HN$. It is clear that every hermitian or unitary matrix is normal, as is the matrix $\left[\begin{array}{rr} 0 & 1 \\ -1 & 0 \end{array} \right]$ in Example 8.7.7. In fact we have the following result.

Theorem 8.7.9

An $n \times n$ complex matrix A is unitarily diagonalizable if and only if A is normal.

Proof. Assume first that $U^HAU = D$, where U is unitary and D is diagonal. Then $DD^H = D^HD$ as is easily verified. Because $DD^H = U^H(AA^H)U$ and $D^HD = U^H(A^HA)U$, it follows by cancellation that $AA^H = A^HA$.

Conversely, assume A is normal—that is, $AA^H = A^HA$. By Schur's theorem, let $U^HAU = T$, where T is upper triangular and U is unitary. Then T is normal too:

$$TT^H = U^H(AA^H)U = U^H(A^HA)U = T^HT$$

Hence it suffices to show that a normal $n \times n$ upper triangular matrix T must be diagonal. We induct on n; it is clear if $n = 1$. If $n > 1$ and $T = \left[t_{ij} \right]$, then equating $(1, 1)$-entries in TT^H and T^HT gives

$$|t_{11}|^2 + |t_{12}|^2 + \cdots + |t_{1n}|^2 = |t_{11}|^2$$

This implies $t_{12} = t_{13} = \cdots = t_{1n} = 0$, so $T = \left[\begin{array}{cc} t_{11} & 0 \\ 0 & T_1 \end{array} \right]$ in block form. Hence $T = \left[\begin{array}{cc} \overline{t}_{11} & 0 \\ 0 & T_1^H \end{array} \right]$ so $TT^H = T^HT$ implies $T_1T_1^H = T_1T_1^H$. Thus T_1 is diagonal by induction, and the proof is complete. □

We conclude this section by using Schur's theorem (Theorem 8.7.7) to prove a famous theorem about matrices. Recall that the characteristic polynomial of a square matrix A is defined by $c_A(x) = \det(xI - A)$, and that the eigenvalues of A are just the roots of $c_A(x)$.

Theorem 8.7.10: Cayley-Hamilton Theorem[17]

If A is an $n \times n$ complex matrix, then $c_A(A) = 0$; that is, A is a root of its characteristic polynomial.

Proof. If $p(x)$ is any polynomial with complex coefficients, then $p(P^{-1}AP) = P^{-1}p(A)P$ for any invertible complex matrix P. Hence, by Schur's theorem, we may assume that A is upper triangular. Then the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ of A appear along the main diagonal, so

$$c_A(x) = (x - \lambda_1)(x - \lambda_2)(x - \lambda_3) \cdots (x - \lambda_n)$$

[17]Named after the English mathematician Arthur Cayley (1821–1895) and William Rowan Hamilton (1805–1865), an Irish mathematician famous for his work on physical dynamics.

Thus

$$c_A(A) = (A - \lambda_1 I)(A - \lambda_2 I)(A - \lambda_3 I)\cdots(A - \lambda_n I)$$

Note that each matrix $A - \lambda_i I$ is upper triangular. Now observe:

1. $A - \lambda_1 I$ has zero first column because column 1 of A is $(\lambda_1, 0, 0, \dots, 0)^T$.
2. Then $(A - \lambda_1 I)(A - \lambda_2 I)$ has the first two columns zero because the second column of $(A - \lambda_2 I)$ is $(b, 0, 0, \dots, 0)^T$ for some constant b.
3. Next $(A - \lambda_1 I)(A - \lambda_2 I)(A - \lambda_3 I)$ has the first three columns zero because column 3 of $(A - \lambda_3 I)$ is $(c, d, 0, \dots, 0)^T$ for some constants c and d.

Continuing in this way we see that $(A - \lambda_1 I)(A - \lambda_2 I)(A - \lambda_3 I)\cdots(A - \lambda_n I)$ has all n columns zero; that is, $c_A(A) = 0$. □

Exercises for 8.7

Exercise 8.7.1 In each case, compute the norm of the complex vector.

a. $(1, 1-i, -2, i)$

b. $(1-i, 1+i, 1, -1)$

c. $(2+i, 1-i, 2, 0, -i)$

d. $(-2, -i, 1+i, 1-i, 2i)$

Exercise 8.7.2 In each case, determine whether the two vectors are orthogonal.

a. $(4, -3i, 2+i), (i, 2, 2-4i)$

b. $(i, -i, 2+i), (i, i, 2-i)$

c. $(1, 1, i, i), (1, i, -i, 1)$

d. $(4+4i, 2+i, 2i), (-1+i, 2, 3-2i)$

Exercise 8.7.3 A subset U of $\mathbb{C}^n$ is called a **complex subspace** of $\mathbb{C}^n$ if it contains 0 and if, given $\mathbf{v}$ and $\mathbf{w}$ in U, both $\mathbf{v}+\mathbf{w}$ and $z\mathbf{v}$ lie in U (z any complex number). In each case, determine whether U is a complex subspace of $\mathbb{C}^3$.

a. $U = \{(w, \overline{w}, 0) \mid w \text{ in } \mathbb{C}\}$

b. $U = \{(w, 2w, a) \mid w \text{ in } \mathbb{C}, a \text{ in } \mathbb{R}\}$

c. $U = \mathbb{R}^3$

d. $U = \{(v+w, v-2w, v) \mid v, w \text{ in } \mathbb{C}\}$

Exercise 8.7.4 In each case, find a basis over $\mathbb{C}$, and determine the dimension of the complex subspace U of $\mathbb{C}^3$ (see the previous exercise).

a. $U = \{(w, v+w, v-iw) \mid v, w \text{ in } \mathbb{C}\}$

b. $U = \{(iv+w, 0, 2v-w) \mid v, w \text{ in } \mathbb{C}\}$

c. $U = \{(u, v, w) \mid iu - 3v + (1-i)w = 0;\ u, v, w \text{ in } \mathbb{C}\}$

d. $U = \{(u, v, w) \mid 2u + (1+i)v - iw = 0;\ u, v, w \text{ in } \mathbb{C}\}$

Exercise 8.7.5 In each case, determine whether the given matrix is hermitian, unitary, or normal.

a. $\left[\begin{array}{rr} 1 & -i \\ i & i \end{array}\right]$

b. $\left[\begin{array}{rr} 2 & 3 \\ -3 & 2 \end{array}\right]$

c. $\left[\begin{array}{rr} 1 & i \\ -i & 2 \end{array}\right]$

d. $\left[\begin{array}{rr} 1 & -i \\ i & -1 \end{array}\right]$

e. $\frac{1}{\sqrt{2}}\left[\begin{array}{rr} 1 & -1 \\ 1 & 1 \end{array}\right]$

f. $\left[\begin{array}{cc} 1 & 1+i \\ 1+i & i \end{array}\right]$

g. $\left[\begin{array}{cc} 1+i & 1 \\ -i & -1+i \end{array}\right]$

h. $\frac{1}{\sqrt{2}|z|}\left[\begin{array}{cc} z & z \\ \overline{z} & -\overline{z} \end{array}\right]$, $z \neq 0$

Exercise 8.7.6 Show that a matrix N is normal if and only if $\overline{N}N^T = N^T\overline{N}$.

Exercise 8.7.7 Let $A = \left[\begin{array}{cc} z & \overline{v} \\ v & w \end{array}\right]$ where v, w, and z are complex numbers. Characterize in terms of v, w, and z when A is

a. hermitian

b. unitary

c. normal.

Exercise 8.7.8 In each case, find a unitary matrix U such that U^HAU is diagonal.

a. $A = \left[\begin{array}{cc} 1 & i \\ -i & 1 \end{array}\right]$

b. $A = \left[\begin{array}{cc} 4 & 3-i \\ 3+i & 1 \end{array}\right]$

c. $A = \left[\begin{array}{cc} a & b \\ -b & a \end{array}\right]$; a, b, real

d. $A = \left[\begin{array}{cc} 2 & 1+i \\ 1-i & 3 \end{array}\right]$

e. $A = \left[\begin{array}{ccc} 1 & 0 & 1+i \\ 0 & 2 & 0 \\ 1-i & 0 & 0 \end{array}\right]$

f. $A = \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 1+i \\ 0 & 1-i & 2 \end{array}\right]$

Exercise 8.7.9 Show that $\langle A\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, A^H\mathbf{y}\rangle$ holds for all $n \times n$ matrices A and for all n-tuples $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{C}^n$.

Exercise 8.7.10

a. Prove (1) and (2) of Theorem 8.7.1.

b. Prove Theorem 8.7.2.

c. Prove Theorem 8.7.3.

Exercise 8.7.11

a. Show that A is hermitian if and only if $\overline{A} = A^T$.

b. Show that the diagonal entries of any hermitian matrix are real.

Exercise 8.7.12

a. Show that every complex matrix Z can be written uniquely in the form $Z = A + iB$, where A and B are real matrices.

b. If $Z = A + iB$ as in (a), show that Z is hermitian if and only if A is symmetric, and B is skew-symmetric (that is, $B^T = -B$).

Exercise 8.7.13 If Z is any complex $n \times n$ matrix, show that ZZ^H and $Z + Z^H$ are hermitian.

Exercise 8.7.14 A complex matrix B is called **skew-hermitian** if $B^H = -B$.

a. Show that $Z - Z^H$ is skew-hermitian for any square complex matrix Z.

b. If B is skew-hermitian, show that B^2 and iB are hermitian.

c. If B is skew-hermitian, show that the eigenvalues of B are pure imaginary ($i\lambda$ for real λ).

d. Show that every $n \times n$ complex matrix Z can be written uniquely as $Z = A + B$, where A is hermitian and B is skew-hermitian.

Exercise 8.7.15 Let U be a unitary matrix. Show that:

a. $\|U\mathbf{x}\| = \|\mathbf{x}\|$ for all columns $\mathbf{x}$ in $\mathbb{C}^n$.

b. $|\lambda| = 1$ for every eigenvalue λ of U.

Exercise 8.7.16

a. If Z is an invertible complex matrix, show that Z^H is invertible and that $(Z^H)^{-1} = (Z^{-1})^H$.

b. Show that the inverse of a unitary matrix is again unitary.

c. If U is unitary, show that U^H is unitary.

Exercise 8.7.17 Let Z be an $m \times n$ matrix such that $Z^H Z = I_n$ (for example, Z is a unit column in $\mathbb{C}^n$).

a. Show that $V = ZZ^H$ is hermitian and satisfies $V^2 = V$.

b. Show that $U = I - 2ZZ^H$ is both unitary and hermitian (so $U^{-1} = U^H = U$).

Exercise 8.7.18

a. If N is normal, show that zN is also normal for all complex numbers z.

b. Show that (a) fails if *normal* is replaced by *hermitian*.

Exercise 8.7.19 Show that a real 2×2 normal matrix is either symmetric or has the form $\left[\begin{array}{rr} a & b \\ -b & a \end{array} \right]$.

Exercise 8.7.20 If A is hermitian, show that all the coefficients of $c_A(x)$ are real numbers.

Exercise 8.7.21

a. If $A = \left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array} \right]$, show that $U^{-1}AU$ is not diagonal for any invertible complex matrix U.

b. If $A = \left[\begin{array}{rr} 0 & 1 \\ -1 & 0 \end{array} \right]$, show that $U^{-1}AU$ is not upper triangular for any *real* invertible matrix U.

Exercise 8.7.22 If A is any $n \times n$ matrix, show that $U^H AU$ is lower triangular for some unitary matrix U.

Exercise 8.7.23 If A is a 3×3 matrix, show that $A^2 = 0$ if and only if there exists a unitary matrix U such that $U^H AU$ has the form $\left[\begin{array}{ccc} 0 & 0 & u \\ 0 & 0 & v \\ 0 & 0 & 0 \end{array} \right]$ or the form $\left[\begin{array}{ccc} 0 & u & v \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array} \right]$.

Exercise 8.7.24 If $A^2 = A$, show that rank $A = $ tr A. [*Hint*: Use Schur's theorem.]

Exercise 8.7.25 Let A be any $n \times n$ complex matrix with eigenvalues $\lambda_1, \dots, \lambda_n$. Show that $A = P + N$ where $N^n = 0$ and $P = UDU^T$ where U is unitary and $D = \text{diag}(\lambda_1, \dots, \lambda_n)$. [Hint: Schur's theorem]

8.8 An Application to Linear Codes over Finite Fields

For centuries mankind has been using codes to transmit messages. In many cases, for example transmitting financial, medical, or military information, the message is disguised in such a way that it cannot be understood by an intruder who intercepts it, but can be easily "decoded" by the intended receiver. This subject is called *cryptography* and, while intriguing, is not our focus here. Instead, we investigate methods for detecting and correcting errors in the transmission of the message.

The stunning photos of the planet Saturn sent by the space probe are a very good example of how successful these methods can be. These messages are subject to "noise" such as solar interference which causes errors in the message. The signal is received on Earth with errors that must be detected and corrected before the high-quality pictures can be printed. This is done using error-correcting codes. To see how, we first discuss a system of adding and multiplying integers while ignoring multiples of a fixed integer.