§7.3 THE SPECTRAL THEOREM

[The following improves on pages 405–410.]

We have seen that a linear operator T on a finite-dimensional vector space V is diagonalizable if and only if V has a basis of eigenvectors of T. Now suppose that V is actually an inner product space. Then we might hope to obtain an *orthonormal* basis of eigenvectors, so that T, relative to this orthonormal basis, is represented by a diagonal matrix, or even a real diagonal matrix.

We know from §6.2 that orthonormal bases are the "nicest" bases. Suppose that $\mathcal{B}$ is a list $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n$ of orthonormal basis vectors for an n-dimensional inner product space V. Then coordinate vectors can be found by using the "Fourier coefficients" (as on page 339):

$$[\![\mathbf{v}]\!]_{\mathcal{B}} = \begin{bmatrix} (\mathbf{v}, \mathbf{b}_1) \\ \vdots \\ (\mathbf{v}, \mathbf{b}_n) \end{bmatrix}$$

Relative to $\mathcal{B}$, a linear operator T is represented by the matrix A whose entries are given explicitly by the equation from page 340:

$$(A)_{ij} = (T(\mathbf{b}_j), \mathbf{b}_i)$$

(As a special case, the standard basis for $\mathbb{R}^n$ is orthonormal, and for any $n \times n$ matrix A we have $(A)_{ij} = (A\mathbf{e}_j, \mathbf{e}_i) = \mathbf{e}_i^{\mathrm{T}} A \mathbf{e}_j$.) And the coordinate map $[\![\ \ ]\!]_{\mathcal{B}}$ relative to an orthonormal basis is an isometry, i.e., it preserves the inner product:

$$(\mathbf{u}, \mathbf{v}) = ([\![\mathbf{u}]\!]_{\mathcal{B}}, [\![\mathbf{v}]\!]_{\mathcal{B}})$$

where the inner product on the left is in V, and the inner product on the right is in $\mathbb{R}^n$ (or $\mathbb{C}^n$).

Let us say that a linear operator T on a finite-dimensional inner product space V is *orthogonally diagonalizable* if there exists an orthonormal basis $\mathcal{B}$ relative to which T's representing matrix, $[\![T]\!]_{\mathcal{B}}^{\mathcal{B}}$, is a real diagonal matrix D. The problem, then, is to determine which linear operators are orthogonally diagonalizable. (It might seem that the phrase "orthonormally diagonalizable" would be more suitable. So it would, but we will stick to the phrase commonly used.)

For a real $n \times n$ matrix A, we will say that A is *orthogonally diagonalizable* if the linear operator L_A on $\mathbb{R}^n$ is orthogonally diagonalizable. This is equivalent to saying that there exists a transition matrix Q whose columns form an orthonormal basis for $\mathbb{R}^n$, such that $Q^{-1}AQ$ is a diagonal matrix. Such a matrix Q is said to be an *orthogonal* matrix; see §6.5. Similarly, a complex matrix U whose columns form an orthonormal basis for $\mathbb{C}^n$ is called a *unitary* matrix; see §6.5.

A valuable clue is that a real diagonal matrix D is always a *symmetric* matrix (and is a Hermitian matrix). We might start by determining conditions under which T is represented, relative to an orthonormal basis, by a symmetric (or Hermitian) matrix.

Recall from §6.2 that in $\mathbb{R}^n$ we have the equation

$$(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A^{\mathrm{T}}\mathbf{y})$$

and in $\mathbb{C}^n$ we have the equation

$$(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A^*\mathbf{y}).$$

Consequently, in $\mathbb{R}^n$ the equation

$$(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A\mathbf{y})$$

holds whenever A is a symmetric matrix, and the equation holds in $\mathbb{C}^n$ whenever A is Hermitian. The converse is also true:

Lemma. (a) A real $n \times n$ matrix A is symmetric if and only if the equation

$$(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A\mathbf{y})$$

holds for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$.

(b) A complex $n \times n$ matrix A is Hermitian if and only if the equation

$$(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A\mathbf{y})$$

holds for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{C}^n$.

Proof. (a) To prove the new direction, suppose that A is a matrix with the property that $(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A\mathbf{y})$ always holds. Then $(A)_{ij} = (A\mathbf{e}_j, \mathbf{e}_i) = (\mathbf{e}_j, A\mathbf{e}_i) = (A\mathbf{e}_i, \mathbf{e}_j) = (A)_{ji}$.

(b) This is the same, the equation now being $(A)_{ij} = (A\mathbf{e}_j, \mathbf{e}_i) = (\mathbf{e}_j, A\mathbf{e}_i) = \overline{(A\mathbf{e}_i, \mathbf{e}_j)} = \overline{(A)_{ji}}$. ⊣

Let us say that the linear operator T is *symmetric* if the equation

$$(T(\mathbf{u}), \mathbf{v}) = (\mathbf{u}, T(\mathbf{v}))$$

holds for all $\mathbf{u}$ and $\mathbf{v}$ in V. (The term *self-adjoint* is also used for this property, particularly in the context of complex inner product spaces. The terminology stems from the fact that one can define the *adjoint* T^* of T to be the unique operator for which the equation $(T(\mathbf{u}), \mathbf{v}) = (\mathbf{u}, T^*(\mathbf{v}))$ always holds. We will not go into this topic.)

Theorem 0. Assume that T is a linear operator on an n-dimensional real or complex inner product space V. Let $\mathcal{B}$ be a list of orthonormal basis vectors for V. Then T is a symmetric operator if and only if its representing matrix A relative to $\mathcal{B}$ is a symmetric matrix (in the real case) or a Hermitian matrix (in the complex case).

Proof. On the one hand, if the given $\mathcal{B}$ yields a Hermitian matrix A, then

$$(T(\mathbf{u}), \mathbf{v}) = ([\![T(\mathbf{u})]\!]_{\mathcal{B}}, [\![\mathbf{v}]\!]_{\mathcal{B}}) = (A[\![\mathbf{u}]\!]_{\mathcal{B}}, [\![\mathbf{v}]\!]_{\mathcal{B}}) = ([\![\mathbf{u}]\!]_{\mathcal{B}}, A[\![\mathbf{v}]\!]_{\mathcal{B}}) = ([\![\mathbf{u}]\!]_{\mathcal{B}}, [\![T(\mathbf{v})]\!]_{\mathcal{B}}) = (\mathbf{u}, T(\mathbf{v}))$$

so that T is a symmetric operator. (This covers both the real case and the complex case. Over the reals, the Hermitian matrices are exactly the symmetric matrices, and the complex conjugate operation is the identity operation.)

One the other hand, suppose T is a symmetric operator. Then given any $\mathcal{B}$ consisting of orthonormal basis vectors $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n$, we have

$$(A)_{ij} = (T(\mathbf{b}_j), \mathbf{b}_i) = (\mathbf{b}_j, T(\mathbf{b}_i)) = \overline{(T(\mathbf{b}_i), \mathbf{b}_j)} = \overline{(A)_{ji}}$$

so that A is a Hermitian matrix (and if real, is symmetric). ⊣

Corollary. (a) If T is a linear operator that is orthogonally diagonalizable, then T must be a symmetric operator.

(b) If a real $n \times n$ matrix A is orthogonally diagonalizable, then A must be a symmetric matrix.

(c) If a complex $n \times n$ matrix A is unitarily similar to a real diagonal matrix D (i.e., $D = U^{-1}AU$ for a unitary matrix U), then A must be a Hermitian matrix.

Proof. (a) A real diagonal matrix is symmetric (and Hermitian); apply Theorem 0 to the orthonormal basis that orthogonally diagonalizes T.

(b) and (c) We are given that L_A is an orthogonally diagonalizable operators, so by part (a), it is a symmetric operator. So its representing matrix relative to the standard basis—which is simply A—is a symmetric matrix (in the real case) or a Hermitian matrix (in the complex case). ⊣

Thus we have an upper bound: *At most* the symmetric linear operators can be orthogonally diagonalizable. The good news is that the converse also holds: Any symmetric linear operator is indeed orthogonally diagonalizable.

In Sections 1–2, we saw that there were two potential barriers to diagonalization. One is that the characteristic polynomial might have non-real roots (in a real vector space). The other barrier is that the dimension of some eigenspace E_λ might fall short of the algebraic multiplicity of λ in the characteristic polynomial. Theorem 1 below will remove the first barrier. And the spectral theorem will remove the second.

Theorem 1. (a) Any eigenvalue of a symmetric linear operator is real.

(b) For a real symmetric matrix, any root of its characteristic polynomial is real.

(c) For a Hermitian matrix, any root of its characteristic polynomial is real.

Proof. It suffices to prove part (c). Then part (b) follows because any real symmetric matrix is also Hermitian. And part (a) follows, where we calculate the eigenvalues using an orthonormal basis to find the representing matrix.

Assume, then, that A is a Hermitian matrix and look at the linear operator L_A on $\mathbb{C}^n$. Any root λ of its characteristic polynomial is an eigenvalue; there is some nonzero eigenvector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$. Then on the one hand

$$(A\mathbf{x}, \mathbf{x}) = (\lambda\mathbf{x}, \mathbf{x}) = \lambda(\mathbf{x}, \mathbf{x})$$

and on the other hand

$$(\mathbf{x}, A\mathbf{x}) = (\mathbf{x}, \lambda\mathbf{x}) = \overline{\lambda}(\mathbf{x}, \mathbf{x}).$$

For a Hermitian matrix, these are equal:

$$\lambda(\mathbf{x}, \mathbf{x}) = \overline{\lambda}(\mathbf{x}, \mathbf{x})$$

Since $\mathbf{x}$ is nonzero, we conclude that $\lambda = \overline{\lambda}$, so this is a real number. ⊣

The main work of this material is done in the proof of the next theorem.

Spectral Theorem. Assume that $T : V \to V$ is a symmetric linear operator on a finite-dimensional inner product space V, real or complex. Then V has an orthonormal basis of eigenvectors for T. Relative to this basis, T is represented by a real diagonal matrix D.

Proof. Use induction on $\dim V$.

Basis. The result is clear if $\dim V = 1$. (Why?) It is even clearer if $\dim V = 0$.

Inductive step. The characteristic polynomial of T has some root λ; by Theorem 1 we know that λ is real. Thus λ is an eigenvalue of T (even if V is a real inner product space); let $\mathbf{u}$ be a corresponding eigenvector. We may suppose that $|\mathbf{u}| = 1$.

Let $W = (\mathrm{sp}\{\mathbf{u}\})^\perp$. Then W is a proper subspace of V, so $\dim W < \dim V$. (In fact it is not hard to see that if $\dim V = n$, then $\dim W = n - 1$. From Chapter 6 we know that $\dim U + \dim U^\perp = \dim V$ for any subspace U; we apply this to the one-dimensional subspace $U = \mathrm{sp}\{\mathbf{u}\}$.)

Claim: Whenever $\mathbf{w} \in W$ then also $T(\mathbf{w}) \in W$. (That is, W is a "T-invariant" subspace.) The reason is that

$$\begin{aligned}(T(\mathbf{w}), \mathbf{u}) &= (\mathbf{w}, T(\mathbf{u})) && \text{(by symmetry of } T\text{)}\\ &= (\mathbf{w}, \lambda\mathbf{u})\\ &= \lambda(\mathbf{w}, \mathbf{u})\\ &= 0 && \text{(since } \mathbf{w} \text{ belongs to } W\text{)}\end{aligned}$$

which verifies the claim. The significance of the claim is that the restriction of T to W is a symmetric linear operator on W. So by the inductive hypothesis, W has an orthonormal basis $\mathbf{b}_1, \ldots, \mathbf{b}_{n-1}$ of eigenvectors for T. Adjoin the new basis vector $\mathbf{b}_n = \mathbf{u}$. The result is still orthonormal. And now it is a basis (of eigenvectors) for V. ⊣

Together with Theorem 0, the spectral theorem show that a linear operator on a finite-dimensional inner product space V is orthogonally diagonalizable if and only if T is a symmetric operator.

Corollary. (a) A real $n \times n$ matrix is orthogonally similar to a (real) diagonal matrix if and only if it is a symmetric matrix.

(b) A complex $n \times n$ matrix is unitarily similar to a real diagonal matrix if and only if it is Hermitian.

The spectral theorem assures us that a symmetric operator *has* an orthonormal basis of eigenvectors. Now suppose we want to calculate such a basis. The following fact is helpful.

Theorem 3. Assume that T is a symmetric linear operator. Then any two eigenvectors for *different* eigenvalues are orthogonal.

Proof. Assume that T is a symmetric linear operator, $T(\mathbf{u}) = \lambda\mathbf{u}$, and $T(\mathbf{v}) = \mu\mathbf{v}$ for nonzero $\mathbf{u}$ and $\mathbf{v}$, where $\lambda \neq \mu$. We will simplify $(T(\mathbf{u}), \mathbf{v})$ in two ways. One the one hand

$$(T(\mathbf{u}), \mathbf{v}) = (\lambda\mathbf{u}, \mathbf{v}) = \lambda(\mathbf{u}, \mathbf{v})$$

and on the other hand, using the symmetry of T and the fact that μ is real, we have

$$(T(\mathbf{u}), \mathbf{v}) = (\mathbf{u}, T(\mathbf{v})) = (\mathbf{u}, \mu\mathbf{v}) = \overline{\mu}(\mathbf{u}, \mathbf{v}) = \mu(\mathbf{u}, \mathbf{v}).$$

Thus $\lambda(\mathbf{u}, \mathbf{v}) = \mu(\mathbf{u}, \mathbf{v})$. Since $\lambda \neq \mu$, we conclude that $(\mathbf{u}, \mathbf{v}) = 0$; these two vectors are orthogonal. ⊣

Corollary. For a real symmetric $n \times n$ matrix, any two eigenvectors for *different* eigenvalues are orthogonal in $\mathbb{R}^n$. For a complex Hermitian $n \times n$ matrix, any two eigenvectors for *different* eigenvalues are orthogonal in $\mathbb{C}^n$

Thus for a symmetric operator, its various eigenspaces will all be orthogonal to each other. For each individual eigenspace E_λ, we can make an orthonormal basis, using the Gram–Schmidt process. When we combine all these bases for all the different E_λ's, the resulting set will still be orthonormal (by Theorem 3). Thus we have a computational procedure that, given a symmetric operator T will produce an orthonormal basis of eigenvectors. Given a real symmetric matrix A, the procedure will produce an orthogonal matrix Q for which $Q^{-1}AQ$ is diagonal. Given a Hermitian matrix A, the procedure will produce a unitary matrix U for which $U^{-1}AU$ is a real diagonal matrix.

The *spectrum* of a symmetric operator T is its list $\lambda_1, \ldots, \lambda_k$ of eigenvalues. The idea is that properly viewed (i.e., relative to just the right orthonormal basis), T can be taken apart into a linear combination of projections. For example, in terms of the representing matrix,

$$\begin{bmatrix} \lambda & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & \nu \end{bmatrix} = \lambda \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \mu \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \nu \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

In terms of linear operators, in this example,

$$T = \lambda P_\lambda + \mu P_\mu + \nu P_\nu$$

where P_λ the operation of orthogonal projection onto the eigenspace E_λ, and similarly for μ and ν. The sum $P_\lambda + P_\mu + P_\nu$ equals the identity operator (one says we have a "resolution of the identity"). The equation

$$T = \lambda P_\lambda + \mu P_\mu + \nu P_\nu$$

is referred to as the "spectral decomposition" of T.

A final comment: We have considered here only *real* diagonal matrices. One can also examine conditions under which an operator T is representable by a complex diagonal matrix. This leads to the less important concept of a *normal* operator. We will not go into this topic here.

[Continue with Examples 1 and 2 from pages 408–409.]

Example 3 For the following matrix A, find an orthogonal matrix U such that $U^{-1}AU$ is diagonal.

$$A = \begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix}$$

First, we compute the characteristic polynomial

$$\begin{aligned} C_A(x) &= \begin{vmatrix} x+1 & -2 & -2 \\ -2 & x+1 & -2 \\ -2 & -2 & x+1 \end{vmatrix} \\ &= x^3 + 3x^2 - 9x - 27 \\ &= (x-3)(x+3)^2 \end{aligned}$$

Thus the eigenvalues are -3, -3, 3.

First take $\lambda = 3$. E_3 is the nullspace of $3I - A = \begin{bmatrix} 4 & -2 & -2 \\ -2 & 4 & -2 \\ -2 & -2 & 4 \end{bmatrix}$, which is row-equivalent to $\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$. So E_3 is the line spanned by $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$; a unit vector on this line is $\frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Secondly, take $\lambda = -3$. E_{-3} is the nullspace of $-3I - A = \begin{bmatrix} -2 & -2 & -2 \\ -2 & -2 & -2 \\ -2 & -2 & -2 \end{bmatrix}$, which is row-equivalent to $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. So E_{-3} is the plane spanned by the vectors $\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$. (Note that this plane is perpendicular to the line E_3, as Theorem 3 promised.) But these two vectors are not orthogonal. So we apply the Gram–Schmidt algorithm: First, $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ which when normalized gives us $\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$. Secondly,

$$\mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1/2 \\ -1/2 \\ 1 \end{bmatrix}$$

which when normalized gives us $\mathbf{u}_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}$.

Combining the two eigenspaces, we obtain the orthonormal basis of eigenvectors:

$$\begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \quad \begin{bmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}, \quad \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

Relative to this basis, L_A is represented by the diagonal matrix $\begin{bmatrix} -3 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$. The orthogonal change-of-basis matrix is given by the equation $U = \frac{1}{\sqrt{6}} \begin{bmatrix} -\sqrt{3} & -1 & \sqrt{2} \\ \sqrt{3} & -1 & \sqrt{2} \\ 0 & 2 & \sqrt{2} \end{bmatrix}$.