Math. 115A H. B. Enderton

The Spectral Theorem

For the final topic in this course, we combine our work in Chapter 5 (on diagonalization) with our work in Chapter 6 (on inner product spaces).

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.

For example, in Problem Set VIII, a certain symmetric matrix A has the property that L_A is represented by a diagonal matrix, relative to an orthonormal basis for $\mathbf{R}^3$. And a certain Hermitian matrix B, over $\mathbf{C}$, has the property that L_B is represented by a *real* diagonal matrix, relative to an orthonormal basis for $\mathbf{C}^3$. We want to investigate this phenomenon in general.

Certainly orthonormal bases are the "nicest" bases. Suppose that β is an orthonormal ordered basis $\{b_1, b_2, \ldots, b_n\}$ for an n-dimensional inner product space V. Then coordinate vectors can be found by using the Fourier coefficients:

$$[v]_\beta = \begin{bmatrix} \langle v, b_1 \rangle \\ \vdots \\ \langle v, b_n \rangle \end{bmatrix}$$

so that

$$v = \langle v, b_1 \rangle b_1 + \cdots + \langle v, b_n \rangle b_n.$$

Relative to β, a linear operator T is represented by the matrix $A = [T]_\beta^\beta$ whose entries are given by the equation:

$$(A)_{ij} = \langle T(b_j), b_i \rangle$$

(As a special case, the standard ordered basis $\{e_1, \ldots, e_n\}$ for $\mathbf{C}^n$ or $\mathbf{R}^n$ is orthonormal, and $(A)_{ij} = \langle Ae_j, e_i \rangle = e_i^t A e_j$.) And the coordinate map $[\ \ ]_\beta$ relative to an orthonormal basis is an isometry, i.e., it preserves the inner product:

$$\langle u, v \rangle = \langle [u]_\beta, [v]_\beta \rangle$$

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

The following fact gives us cause for optimism regarding symmetric matrices.

Lemma. Consider the space $\mathbf{R}^n$ with the usual inner product, and assume that A is an $n \times n$ (real) symmetric matrix. Then any two eigenvectors for *different* eigenvalues are orthogonal.

Proof. We make use of the fact that for any real matrix B, symmetric or not,

$$\langle Bx, y \rangle = \langle x, B^t y \rangle.$$

So for a symmetric matrix A,

$$\langle Ax, y \rangle = \langle x, Ay \rangle.$$

Now suppose that x and y are eigenvectors for the distinct eigenvalues λ and μ, respectively. Then $\lambda\langle x, y \rangle = \langle \lambda x, y \rangle = \langle Ax, y \rangle = \langle x, Ay \rangle = \langle x, \mu y \rangle = \mu\langle x, y \rangle$. Since $\lambda \neq \mu$, we must have $\langle x, y \rangle = 0$ and hence $x \perp y$. $\dashv$

Thus for a real symmetric matrix, 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 the lemma).

On the other hand, if A is a real $n \times n$ matrix that is *not* symmetric, then as we will see, there is no orthonormal basis for $\mathbf{R}^n$ that diagonalizes A. So it is natural for us to focus our attention on symmetric matrices.

As mentioned above, for a symmetric matrix,

$$\langle Ax, y\rangle = \langle x, Ay\rangle$$

for x and y in $\mathbf{R}^n$. In fact this property characterizes exactly the symmetric matrices: If the displayed equation is always true of A, then $(A)_{ij} = \langle Ae_j, e_i\rangle = \langle e_j, Ae_i\rangle = \langle Ae_i, e_j\rangle = (A)_{ji}$. This is the clue as to which linear operators have the best chance of being represented in the way we seek: those operators T with the property that

$$\langle T(u), v\rangle = \langle u, T(v)\rangle$$

for all vectors u and v. Call such operators *self-adjoint.*

(Digression: Section 6.3 defines the *adjoint* T^* of T to be the unique operator such that

$$\langle T(u), v\rangle = \langle u, T^*(v)\rangle$$

always holds. Then the self-adjoint operators can be defined by requiring that $T = T^*$. This definition is equivalent to the definition given above.)

In the complex inner product space $\mathbf{C}^n$, the usual inner product is defined by the equation $\langle x, y\rangle = y^*x$. A complex $n \times n$ matrix A is called *Hermitian* if $A = A^*$. In particular, a real matrix is Hermitian if and only if it is symmetric. Note that in a Hermitian matrix, all the entries on the main diagonal must be real. Recall that for any matrix A, we have $\langle Ax, y\rangle = \langle x, A^*y\rangle$. Thus for a Hermitian matrix A,

$$\langle Ax, y\rangle = \langle x, Ay\rangle$$

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

Thus multiplication L_A by a Hermitian $n \times n$ matrix A is a self-adjoint linear operator on $\mathbf{C}^n$, by the above. Multiplication by a real symmetric matrix is a self-adjoint linear operator on $\mathbf{R}^n$.

Theorem 0. Let T be a linear operator on an n-dimensional real or complex inner product space V, and let β be an orthonormal ordered basis for V. Then T is self-adjoint if and only if its representing matrix A relative to β is a symmetric matrix (in the real case), or a Hermitian matrix (in the complex case).

Proof. Suppose that β is an orthonormal ordered basis $\langle b_1, b_2, \dots, b_n\rangle$ and let $A = [T]_\beta^\beta$. On the one hand, if T is self-adjoint then

$$(A)_{ij} = \langle T(b_j), b_i\rangle = \langle b_j, T(b_i)\rangle = \overline{\langle T(b_i), b_j\rangle} = \overline{(A)_{ji}}$$

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

On the other hand, if A is a Hermitian matrix then

$$\langle T(u), v\rangle = \langle [T(u)]_\beta, [v]_\beta\rangle = \langle A[u]_\beta, [v]_\beta\rangle = \langle [u]_\beta, A[v]_\beta\rangle = \langle [u]_\beta, [T(v)]_\beta\rangle = \langle u, T(v)\rangle$$

so that T is a self-adjoint operator. $\dashv$

In particular, any real diagonal matrix D is automatically symmetric (and Hermitian). So Theorem 0 tells us that if we hope to have T represented by D relative to an orthonormal basis, then T must be self-adjoint. In other words, *at most* the self-adjoint operators have the property we want. What is left is the other direction: to show that any self-adjoint operator is indeed represented by a real diagonal matrix relative to an orthonormal basis.

In Chapter 5, 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. Finally, Theorem 2 will remove the second.

Our first objective is to prove that all eigenvalues of a self-adjoint linear operator are real numbers.

Theorem 1. (a) Any eigenvalue of a self-adjoint 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. (a) Assume that $T(v) = \lambda v$ for nonzero v. Then on the one hand

$$\langle T(v), v \rangle = \langle \lambda v, v \rangle = \lambda \langle v, v \rangle$$

and on the other hand

$$\langle v, T(v) \rangle = \langle v, \lambda v \rangle = \overline{\lambda} \langle v, v \rangle.$$

For a self-adjoint operator T, these are equal:

$$\lambda \langle v, v \rangle = \overline{\lambda} \langle v, v \rangle$$

Since v is nonzero, we conclude that $\lambda = \overline{\lambda}$.

For part (b), suppose that A is an $n \times n$ real symmetric matrix. Let L_A be the self-adjoint linear operator defined by $L_A(x) = Ax$ on the complex space $\mathbf{C}^n$. Then any root λ of the characteristic polynomial for A is an eigenvalue of L_A. By part (a), λ must be real.

Part (c) is the same. ⊣

We can now upgrade the earlier lemma: For a self-adjoint linear operator T, any two eigenvectors for *different* eigenvalues are orthogonal. The proof is unchanged. Suppose that u and v are eigenvectors for the distinct eigenvalues λ and μ, respectively. Then μ is *real* and $\lambda\langle u, v \rangle = \langle \lambda u, v \rangle = \langle T(u), v \rangle = \langle u, T(v) \rangle = \langle u, \mu v \rangle = \mu \langle u, v \rangle$. Since $\lambda \neq \mu$, we must have $\langle u, v \rangle = 0$ and hence $u \perp v$. As a special case, where T is multiplication by a Hermitian matrix, we can say that any two eigenvectors of a Hermitian matrix for different eigenvalues must be orthogonal.

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

Theorem 2. Assume that $T : V \to V$ is a self-adjoint linear operator on a finite-dimensional inner product space V, real or complex. Then V has an orthonormal basis of eigenvectors for T.

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 u be a corresponding eigenvector. By normalizing, we may suppose that $||u|| = 1$.

Let $W = \{u\}^{\perp}$. 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 $\dim W = n - 1$. By the projection theorem, $V = (\operatorname{span} u) \oplus W$; any vector is the sum of a vector p in $\operatorname{span} u$ and a vector e in W. So to any basis for W, we need add only the one additional vector u to get a spanning set for all of V.)

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

$$\langle T(w), u \rangle = \langle w, T(u) \rangle = \langle w, \lambda u \rangle = \lambda \langle w, u \rangle = 0$$

which verifies the claim. The significance of the claim is that the restriction of T to W is a self-adjoint linear operator on W. So by the inductive hypothesis, W has an orthonormal basis, $b_1, \ldots, b_{n-1}$ of eigenvectors for T. Adjoin the new basis vector $b_n = u$. The result is still orthonormal. And now it is a basis (of eigenvectors) for V. ⊣

Theorem 2 gives us the final result we seek.

Spectral theorem. Assume that T is a linear operator on a finite-dimensional inner product space V (real or complex). Then there exists an orthonormal basis of V relative to which T is represented by a real diagonal matrix if and only if T is self-adjoint.

Proof. For the easy direction, suppose we have such a basis; let D be the real diagonal matrix. Then D is Hermitian, so by Theorem 0, T is self-adjoint.

For the hard direction, suppose that T is self-adjoint. Then we apply Theorem 2 to obtain an orthonormal basis of eigenvectors; the representing matrix is real (by Theorem 1) and diagonal. ⊣

Corollary. (a) For a real $n \times n$ matrix A, there is an orthonormal basis for $\mathbf{R}^n$ relative to which L_A is represented by a (real) diagonal matrix if and only if A is symmetric.

(b) For a complex $n \times n$ matrix B, there is an orthonormal basis for $\mathbf{C}^n$ relative to which L_B is represented by a real diagonal matrix if and only if B is Hermitian.

Examples. See Problem Set VIII. For a real symmetric matrix A, we can make a (real) change-of-basis matrix Q with orthonormal columns (such matrices are called *orthogonal*) such that $Q^{-1}AQ$ is a real diagonal matrix. For a Hermitian matrix B, we can make a (complex) change-of-basis matrix U with orthonormal columns (such matrices are called *unitary*) such that $U^{-1}BU$ is a real diagonal matrix.

The spectral theorem is Theorem 6.17 on page 374. See also page 401. The *spectrum* of 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.

(Digression: We have focused here on *real* diagonal matrices. If you are willing to accept *complex* diagonal matrices, then the condition of self-adjointness can be weakened to a condition called *normality*, discussed in the book.)

One question not yet addressed is this: What do we do if T is *not* diagonalizable? Give up? If T has *some* eigenvectors, then we can form a basis with as many eigenvectors as we can. The resulting matrix may not be diagonal, but it may be better than whatever we started with. Beyond that, there is something called *Jordan normal form.* In 115B, that form will be studied.