Generalized eigenspaces

November 20, 2019

Contents

1 Introduction

We've seen that sometimes a nice linear transformation T (from a vector space V to itself) can be *diagonalized*, and that doing this is closely related to finding *eigenvalues* of T. The eigenvalues are exactly the roots of a certain polynomial p_T, of degree equal to $\dim V$, called the *characteristic polynomial*. I explained in class how to compute p_T, and I'll recall that in these notes.

Chapter 8 of the text is devoted to almost-diagonalizing linear transformations for complex vector spaces. Diagonalizing is not quite possible in general, because the eigenspaces may be a little too small; so Chapter 8 introduces *generalized eigenspaces*, which are just enough larger to make things work. Understanding generalized eigenspaces is closely tied to *factoring* the characteristic polynomial as a product of linear factors: see the definition in the text on page 261.

When the field is not the complex numbers, polynomials need not have roots, so they need not factor into linear factors. In these notes we're going to adapt Chapter 8 so that it works over any field. In particular, this will more or less include what is done in Chapter 9 for the real numbers.

2 Polynomials

Recall from pages 30–31 of the text that $\mathcal{P}(F)$ is the (F-vector space) of all polynomials with coefficients in F. The *degree* of a nonzero polynomial is the highest power of x appearing with nonzero coefficient; the polynomial zero is defined to have degree $-\infty$ (to make the formula $\deg(pq) = \deg(p) + \deg(q)$ work when one of the factors is zero). We write

$$\mathcal{P}_m(F) = \{a_m x^m + \cdots + a_1 x + a_0 \mid a_i \in F\} \tag{2.1}$$

for the $m+1$-dimensional subspace of polynomials of degree less than or equal to m. A polynomial is called *monic* if its leading coefficient is 1; therefore

$$1, \qquad x^2 - 7x + 1/2, \qquad x - \pi$$

are monic polynomials, but $2x + 1$ is not monic.

Definition 2.2. A polynomial $p(x)$ is called *irreducible* if

1. $\deg(p) > 0$, and
2. p cannot be written as a product of two polynomials of positive degree.

If a is any nonzero scalar, then p is irreducible if and only if ap is irreducible; so we can concentrate on monic polynomials in discussing irreducibility.

There is a really important analogy in mathematics between *polynomials* and *integers*. In this analogy, irreducible polynomials correspond to (plus or minus) prime numbers. Monic polynomials correspond to positive integers. A good example of the analogy is Proposition 2.7 below. Just as with integers, addition and multiplication of polynomials is easily and rapidly computable, and even division of polynomials (with remainder) is not bad. But *factorization* is much more difficult (and interesting).

We can talk about the degree of a polynomial and about monic polynomials without paying too much attention to the field F; but the notion of irreducible depends heavily on F. For example, the polynomial $p(x) = x^2+1$ is irreducible as a real polynomial; but as complex polynomial

$$p(x) = (x + i)(x - i),$$

so p is not irreducible. In the same way $q(x) = x^2 - 2$ is irreducible in $\mathcal{P}(\mathbb{Q})$, but in $\mathcal{P}(\mathbb{R})$
$$q(x) = (x + \sqrt{2})(x - \sqrt{2}).$$

Every degree one polynomial

$$ax + b \text{ is irreducible}, \qquad (a, b \in F, \quad a \neq 0) \tag{2.3}$$

for any field F.

For degree two polynomials, the story is

$$ax^2 + bx + c \text{ irr} \iff \text{no root in } F, \qquad (a, b, c \in F, \quad a \neq 0) \tag{2.4}$$

In case $2 \neq 0$ in F, we can apply the quadratic formula and deduce

$$ax^2 + bx + c \text{ irr} \iff b^2 - 4ac \text{ has no square root in } F. \tag{2.5}$$

This statement covers the examples above: for $x^2 + 1$ over $\mathbb{R}$, $b = 0$ and $a = c = 1$; $b^2 - 4ac = -4$ has no square root in $\mathbb{R}$, so $x^2 + 1$ is irreducible. In case $F = \mathbb{F}_2$ is the field with two elements, one can check that 0 and 1 are not roots of $p(x) = x^2 + x + 1$; so p is irreducible, even though $b^2 - 4ac = 1 = 1^2$. (In fact p is the *only* irreducible polynomial of degree 2 in $\mathcal{P}(\mathbb{F}_2)$. Can you see why?)

Proposition 2.6. *Suppose F is a field. The following conditions are equivalent:*

1. *every polynomial of positive degree in $\mathcal{P}(F)$ has a root;*
2. *every nonzero polynomial in $\mathcal{P}(F)$ is a product of linear factors:*
$$p(x) = a \prod_{j=1}^{\deg p} (x - \lambda_j) \qquad (\lambda_j \in F, 0 \neq a \in F).$$
3. *the irreducible polynomials in $\mathcal{P}(F)$ are those of degree one ((2.3)).*

If these equivalent conditions are satisfied, we say that F is algebraically closed.

The fundamental theorem of algebra says that $\mathbb{C}$ is algebraically closed. The factorization of arbitrary polynomials into linear factors is at the heart of finding eigenvalues, and therefore of Chapter 8. Here is a substitute that works in any field.

Proposition 2.7. *Any monic polynomial $p \in \mathcal{P}(F)$ can be written as a product of powers of distinct monic irreducible polynomials $\{q_i \mid 1 \leq i \leq r\}$:*

$$p(x) = \prod_{i=1}^{r} q_i(x)^{m_i}, \qquad \deg p = \sum_{i=1}^{r} m_i \deg q_i.$$

Here m_i and $\deg q_i$ are positive integers, so $r \leq \deg p$. This factorization of p is unique up to rearranging the factors. The irreducible q_i that appear are precisely the irreducible factors of p.

In the analogy between polynomials and integers, this proposition corresponds to the fundamental theorem of arithmetic (which says that a positive integer has a unique factorization as a finite product of prime powers).

If F is algebraically closed, then $q_i = x - \lambda_i$, and the λ_i that appear are precisely the roots of p. So this proposition is a generalization of the fundamental theorem of algebra that applies to any field.

Example 2.8. Suppose

$$p(x) = x^3 - 1.$$

Clearly $p(1) = 1^3 - 1 = 1 - 1 = 0$, so 1 is a root of p and $x - 1$ is a factor. Doing long division shows that

$$p(x) = (x - 1)(x^2 + x + 1).$$

So far this is true over any field F. If $F = \mathbb{R}$, then $x^2 + x + 1$ is irreducible (because $1^2 - 4 = -3$ is not a square; so the factorization above is the factorization of p into irreducibles in $\mathcal{P}(\mathbb{R})$: we have

$$q_1 = x - 1, \quad m_1 = 1, \qquad q_2 = x^2 + x + 1, \quad m_2 = 1.$$

But for $\mathbb{C}$, $x^2 + x + 1$ has roots $(-1 \pm i\sqrt{3})/2$, and therefore factors as

$$x^2 + x + 1 = (x - (-1 + i\sqrt{3})/2)(x - (-1 - i\sqrt{3})/2),$$

and

$$p(x) = (x - 1)(x - (-1 + i\sqrt{3})/2)(x - (-1 - i\sqrt{3})/2)$$

is the factorization into irreducibles:

$$\begin{aligned} q_1 &= x - 1, \quad m_1 = 1, \\ q_2 &= x - (-1 + i\sqrt{3})/2, \quad m_2 = 1, \\ q_3 &= x - (-1 - i\sqrt{3})/2, \quad m_3 = 1. \end{aligned}$$

We are going to need just one slightly exotic fact about polynomials and factorization. It's based on

Definition 2.9. We say that the nonzero polynomials p_1 and p_2 are *relatively prime* if they have no common factor of positive degree; equivalently, if there is no irreducible polynomial q that divides both p_1 and p_2.

Example 2.10. The polynomials p_1 and $x-\lambda$ are relatively prime if and only if λ is not a root of p_1; that is, if and only if $p_1(\lambda) \neq 0$.

If $p_1 \neq 0$, then p_1 and 1 are always relatively prime (because 1 has no factors of positive degree).

The polynomials $\prod_{i=1}^{m}(x-\lambda_i)$ and $\prod_{j=1}^{n}(x-\mu_j)$ are relatively prime if and only if

$$\{\lambda_i\} \cap \{\mu_j\} = \emptyset;$$

that is, if and only if there are no roots in common.

Here is the exotic fact.

Proposition 2.11. *Suppose p_1 and p_2 are relatively prime (nonzero) polynomials. Then there are polynomials a_1 and a_2 with the property that*

$$a_1 p_1 + a_2 p_2 = 1.$$

I won't prove this fact; it's stated in Artin's algebra book [1] as Theorem 11.1.5 (page 390), for which the proof is similar to the one given for the corresponding fact for integers (Proposition 2.2.6 on page 46). Here are some examples meant to be supportive.

Example 2.12. Suppose $p_1 = x - \lambda$ and $p_2 = x - \mu$, with $\lambda \neq \mu$. Then

$$a_1 = 1/(\mu - \lambda), \qquad a_2 = 1/(\lambda - \mu).$$

It's easy to check that this works, but maybe not so easy to see how to generalize it.

So suppose

$$p_1(x) = (x-\lambda_1)\cdots(x-\lambda_r), \qquad p_2(x) = (x-\mu_1)\cdots(x-\mu_s),$$

with $\lambda_1, \ldots, \mu_s$ some $r+s$ distinct elements of s. This time I'll choose a_1 to be the (unique) polynomial of degree $s-1$ with the property that

$$a_1(\mu_j) = p_1(\mu_j)^{-1} \qquad (j = 1, \ldots, s).$$

It's not so difficult to write down such a polynomial a_1, using the ideas appearing in the solutions to PS2. Similarly, you can find a_2 of degree $r-1$ with the property that

$$a_2(\lambda_i) = p_2(\lambda_i)^{-1} \qquad (i = 1, \dots, r).$$

Then it follows that

$$(a_1p_1 + a_2p_2)(\lambda_i) = (a_1p_1 + a_2p_2)(\mu_j) = 1 \qquad (i = 1, \dots, r,\ j = 1, \dots, s).$$

Therefore $a_1p_1 + a_2p_2 - 1$ is a polynomial of degree at most $r+s-1$ having $r+s$ distinct zeros. It follows that $a_1p_1 + a_2p_2 - 1 = 0$, as we wished to show.

3 Calculating the characteristic polynomial

The heart of the idea is this, most of which I proved in class in October.

Proposition 3.1. *Suppose V is a finite-dimensional vector space, $T \in \mathcal{L}(V)$, and $0 \neq v_0 \in V$. Define*

$$v_j = T^j v_0.$$

Let m be the smallest positive integer with the property that

$$v_m \in \operatorname{span}(v_0, \dots v_{m-1}) =_{def} U.$$

Then $(v_0, \dots, v_{m-1})$ is linearly independent, so there is a unique expression

$$v_m = -a_0v_0 - \cdots - a_{m-1}v_{m-1}.$$

Define

$$p(x) = x^m + a_{m-1}x^{m-1} + \cdots + a_0,$$

a monic polynomial of degree $m \geq 1$ in $\mathcal{P}(F)$.

1. *The m-dimensional subspace U of V is preserved by T: $TU \subset U$.*

2. *The linear transformation*

$$p(T) = T^m + a_{m-1}T^{m-1} + \cdots + a_0I$$

acts by zero on U.

3. *If z is a polynomial and $z(T)$ acts by zero on U, then p divides z.*

4. *The eigenvalues of T on U are precisely the roots of p.*

5. *If q is an irreducible polynomial and* $\mathrm{Null}(q(T))$ *has a nonzero intersection with U, then q divides p.*

In Definition 3.3 below, we will define the characteristic polynomial of T on U to be the polynomial p described in (2). In order to relate this information about U to the rest of V, we used the next proposition.

Proposition 3.2. *Suppose V is a finite-dimensional vector space, $T \in \mathcal{L}(V)$, and $U \subset V$ is an invariant subspace: $TU \subset U$. Choose a basis $(e_1, \ldots e_m)$ for U, and extend it to a basis $(e_1, \ldots, e_m, e_{m+1}, \ldots e_n)$ for V. Recall that then*

$$(e_{m+1} + U, \ldots, e_n + U)$$

is a basis of the quotient space V/U.

1. *The matrix of T in the basis $(e_1, \ldots, e_n)$ has the form*

$$\begin{pmatrix} A & B \\ 0 & C \end{pmatrix}.$$

Here A is the $m \times m$ matrix of T on the subspace U; 0 represents the $(n-m)\times m$ zero matrix; B is $m\times(n-m)$; and C is the $(n-m)\times(n-m)$ matrix of T acting on the quotient space V/U.

2. *The set of eigenvalues of T is the union of the set of eigenvalues of A and the set of eigenvalues of C.*

3. *Suppose q is an irreducible polynomial. Then*

$$\mathrm{Null}(q(T)) \neq 0 \iff \mathrm{Null}(q(A)) \neq 0 \quad \textit{or} \quad \mathrm{Null}(q(C)) \neq 0.$$

4. *Suppose p_1 and p_2 are polynomials such that $p_1(A) = 0$ and $p_2(C) = 0$. Then $(p_1 p_2)(T) = 0$.*

Definition 3.3. The *characteristic polynomial* p_T of $T \in \mathcal{L}(V)$ is a monic polynomial of degree equal to the dimension of V, defined by induction on $\dim V$ as follows. If $V = 0$, then $p_T = 1$. If $V \neq 0$, choose any nonzero vector $v_0 \in V$, and define $U \subset V$ as in Proposition 3.1, and p as in (2) of that Proposition. Then $\dim U = m > 0$, and $\deg p = m$; so $n - m < n$. Write a matrix for T as in Proposition 3.2. Then $C \in \mathcal{L}(F^{n-m})$. By induction, the (monic) characteristic polynomial p_C of C is already defined, and $\deg p_C = n - m$. We can therefore define

$$p_T = p \cdot p_C.$$

Clearly p_T is monic, and

$$\deg p_T = \deg p + \deg p_C = m + (n - m) = n.$$

By inductive hypothesis, $p_C(C) = 0$, and by Proposition 3.1 $p(A) = 0$; so by Proposition 3.2,

$$p_T(T) = 0.$$

What is *not* obvious from this definition is that the characteristic polynomial is well-defined. The construction requires choosing a nonzero vector v_0, and different choices lead to different U and different p. I will provide some of the ingredients for proving that different choices lead to the same p_T, but I won't prove this completely. For now I'll just recall one thing I did prove in class: for *any* choice of p_T,

$$\{\text{eigenvalues of } T\} = \{\text{roots of } p_T\}. \tag{3.4}$$

The "base case" of this fact (the case of the subspace U constructed in Proposition 3.1) is Proposition 3.1(4). The induction step is Proposition 3.2(2). Once we've defined "eigenpolynomials" in Section 6, the same proof will show

$$\{\text{eigenpolynomials of } T\} = \{\text{irreducible factors of } p_T\}. \tag{3.5}$$

4 Projections

The theorems we want about generalized eigenvalues and so on (Theorem 5.6, for example) provide direct sum decompositions of vector spaces. So it is useful to say a few words about how to find direct sum decompositions.

Definition 4.1. Suppose V is a vector space. A *projection* on V is a linear map $P \in \mathcal{L}(V)$ with the property that $P^2 = P$; equivalently, a map satisfying the polynomial $x^2 - x$.

Proposition 4.2. *Suppose V is a vector space. The following things are in one-to-one correspondence:*

1. *projections $P \in \mathcal{L}(V)$;*

2. *pairs of linear transformations P and Q in $\mathcal{L}(V)$ such that*

$$P + Q = I, \qquad PQ = 0; \tag{4.2a}$$

3. *pairs of linear transformations P and Q in $\mathcal{L}(V)$ such that*

$$P^2 = P, \qquad Q^2 = Q, \qquad PQ = 0, \text{ and } P + Q = I; \tag{4.2b}$$

and

4. *direct sum decompositions $V = U \oplus W$.*

The correspondence can be described as follows. Given a projection P as in (1), define $Q = I - P$ to get a pair (P, Q) as in (2) or (3).

Given P and Q as in (2) or (3), define

$$\begin{aligned} U &= \mathrm{Range}(P) = \mathrm{Null}(Q) = \mathrm{Null}(I - P) = \mathrm{Range}(I - Q), \\ W &= \mathrm{Range}(Q) = \mathrm{Null}(P) = \mathrm{Null}(I - Q) = \mathrm{Range}(I - P). \end{aligned} \tag{4.2c}$$

Then $V = U \oplus W$ as in (4).

Given $V = U \oplus W$ as in (4), define

$$P(u + w) = u \qquad (u \in U, w \in W). \tag{4.2d}$$

Then P is a projection.

In terms of the eigenspaces of P, we have

$$U = V_{1,P}, \qquad W = V_{0,P}. \tag{4.2e}$$

In terms of the eigenspaces of Q,

$$U = V_{0,Q}, \qquad W = V_{1,Q}. \tag{4.2f}$$

The condition $P^2 = P$ is called *idempotent* (pronounced I think with a long "i") meaning more or less "self-power." The first three conditions in (3) say that P and Q are what's called *orthogonal idempotents*, because the conditions look formally a bit like the conditions for two vectors to be orthogonal.

I have tried to state the proposition in such a way that it almost proves itself. But here are some of the words.

Proof. Suppose P is a projection, so that $P^2 = P$, or equivalently

$$P - P^2 = 0, \qquad P(I - P) = 0. \tag{4.3}$$

In order to get the first requirement in (4.2a), we are forced to define $Q = I - P$. Then the formulation $P(I - P) = 0$ of the definition of projection gives the second requirement in (4.2a).

Next, suppose P and Q satisfy (4.2a). First notice that

$$P = PI = P(P + Q) = P^2 + PQ = P^2,$$

so P is a projection. In the same way we see that $Q^2 = Q$.

We want to get a direct sum decomposition of V; so we define $U = PV$, $W = QV$. Then any $v \in V$ may be written

$$v = Iv = (P + Q)v = Pv + Qv = u + w,$$

with $u = Pv \in U$ and $w = Qv \in W$. So $U + W = V$. To see that the sum is direct, suppose that $U \cap W$ contains the element $Pv_1 = Qv_2$. Applying P to this equation and using $P^2 = P$ and $PQ = 0$, we get $Pv_1 = 0$. This proves that $U \cap W = 0$, so the sum is direct.

If $w = Qv$ belongs to W, then $Pw = PQv = 0$ by the assumption that $PQ = 0$; so

$$W = \mathrm{Range}(Q) \subset \mathrm{Null}(P).$$

Conversely, if $x \in \mathrm{Null}(P)$, then $x = Ix = (P + Q)x = Qx$, so

$$\mathrm{Null}(P) \subset \mathrm{Range}(Q) = W.$$

This proves the first two equalities for W in (4.2c). The last two are just $P = I - Q$ and $Q = I - P$. The equalities for U follow by interchanging the roles of P and Q.

Finally, suppose $V = U \oplus W$ as in (3). The definition (4.2d) shows that

$$P^2(u + w) = P(u) = u = P(u + w),$$

so $P^2 = P$ as we wished to show. It is easy to see from the definitions that if we follow these constructions around the circle (for example starting with P, getting Q, then getting U and W and finally a new P') we always return to where we began (that $P' = P$). The remaining assertions about eigenvalues are just reformulations of what we have already proved. $\square$

Suppose now that V is an *inner product space*, and that P and Q are the projections corresponding to the direct sum decomposition $V = U \oplus W$. Then P^* and Q^* are the projections corresponding to the decomposition $V = W^\perp \oplus U^\perp$. This means that P and Q are self-adjoint if and only if the direct sum decomposition is *orthogonal*. That's not automatic: the term "orthogonal idempotents" makes sense without inner products, and has very little to do with the notion of "orthogonal" in inner product spaces.

5 Generalized eigenvalues

This section summarizes Chapter 8 of the text. Remember from the text (page 134) the description of the λ-eigenspace for any $T \in \mathcal{L}(V)$:

$$\begin{aligned} V_\lambda &=_{\text{def}} \{v \in V \mid Tv = \lambda v\} \\ &= \text{Null}(T - \lambda) \\ &= \text{Null}(q(T)) \qquad (q(x) = x - \lambda). \end{aligned} \tag{5.1}$$

The *multiplicity* $m(\lambda)$ of the eigenvalue λ is by definition the dimension of the eigenspace V_λ.

The text (pages 254–255) does not use this definition of $m(\lambda)$; instead the text defines the multiplicity to be what I call *generalized multiplicity* (see Definition 5.4 below). I think that the definition above is the one that most people use. The text refers to the definition above as *geometric multiplicity.*

Example 5.2. Here is one of the most important examples of eigenspaces. Suppose

$$V = \text{infinitely differentiable functions on } \mathbb{R}, \qquad D = \frac{d}{dt} \in \mathcal{L}(V).$$

For any complex number λ, the λ-eigenspace of T is

$$V_\lambda = \{f \in V \mid \frac{df}{dt} = \lambda f\}.$$

This is a first-order differential equation that you learned to solve in 18.03:

$$V_\lambda = \{Ae^{\lambda t} \mid A \in \mathbb{C}\},$$

a one-dimensional space with basis vector the function sending t to $e^{\lambda t}$. Therefore the multiplicity of the eigenvalue λ for D is $m(\lambda) = 1$.

To say that T is *diagonalizable* means that there is a basis of V consisting of eigenvectors; equivalently, that

$$V = \sum_{\lambda \in F} V_\lambda \qquad (T \text{ diagonalizable}).$$

Even for $\mathbb{C}$, not all operators are diagonalizable.

Example 5.3. Suppose $V = \mathbb{C}^2$, and

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Because T is upper triangular, its eigenvalues are the diagonal entries; that is, the only eigenvalue is 1. It's easy to calculate the eigenspace:

$$V_1 = \mathrm{Null}(T - 1 \cdot I) = \mathrm{Null} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} \mid x \in \mathbb{C} \right\}.$$

So the only eigenspace is one-dimensional, so $\mathbb{C}^2$ cannot be the direct sum of the eigenspaces. We are missing (from the eigenspace) the vectors whose second coordinate is not zero. These are not in the null space of $T - 1 \cdot I$:

$$(T - I) \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ 0 \end{pmatrix}.$$

This is not zero (if y is not zero), but now it *is* in the null space of $T - 1 \cdot I$. Therefore

$$(T - I)^2 \cdot \begin{pmatrix} x \\ y \end{pmatrix} = (T - I) \cdot \begin{pmatrix} y \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

To summarize, only the line through the first basis vector is in the 1-eigenspace (the null space of $T - 1 \cdot I$), but all of $\mathbb{C}^2$ is in the null space of $(T - 1 \cdot I)^2$.

Because eigenspaces are not big enough to decompose V, we need a good way to enlarge them. The example suggests

Definition 5.4. Suppose $T \in \mathcal{L}(V)$. The λ-*generalized eigenspace of* T is

$$V_{[\lambda]} =_{\text{def}} \{v \in V \mid (T - \lambda I)^m v = 0 \text{ (some } m > 0).\} \subset V.$$

Clearly

$$V_\lambda \subset V_{[\lambda]};$$

because if $Tv = \lambda v$, then we can take $m = 1$ in the definition of $V_{[\lambda]}$. The *generalized multiplicity* $M(\lambda)$ of the eigenvalue λ is by definition the dimension of the generalized eigenspace $V_{[\lambda]}$.

This is what the text calls "multiplicity." I don't believe there is a standard notation for generalized eigenspaces. The text uses the notation $G(\lambda, T)$ instead of $V_{[\lambda]}$, and $E(\lambda, T)$ for the eigenspace V_λ. Because of the lack of a standard, you always need to explain these notations if you use them.

Example 5.5. We return to the example of

$$V = \text{infinitely differentiable functions on } \mathbb{R}, \qquad D = \frac{d}{dt} \in \mathcal{L}(V).$$

You saw generalized eigenspaces in 18.03 as well; they appear when you try to solve higher order differential equations like

$$f'' - 2f' + f = 0,$$

which can be written as

$$(D - I)^2 f = 0.$$

One solution is e^t; and another is te^t. The reason is

$$D(te^t) = (Dt)e^t + t(De^t) = e^t + te^t,$$

so

$$(D - I)(te^t) = e^t \in \operatorname{Null}(D - I).$$

In this way you learned (in 18.03!) that the λ generalized eigenspace of D is

$$V_{[\lambda]} = \{p(t)e^{\lambda t} \mid p \in \mathcal{P}(\mathbb{C})\}.$$

This generalized eigenspace is infinite-dimensional (since the space of all polynomials is infinite-dimensional) so the generalized multiplicity $M(\lambda)$ is infinite.

The main theorem of Chapter 8 is

Theorem 5.6 (text, Theorem 8.21)**.** *Suppose V is a finite-dimensional complex vector space, $T \in \mathcal{L}(V)$, and $\lambda_1, \ldots, \lambda_r$ are all the distinct eigenvalues of T. Then*

1. *Each generalized eigenspace $V_{[\lambda_i]}$ is preserved by T.*
2. *The space V is the direct sum of the generalized eigenspaces:*

$$V = \bigoplus_{i=1}^{r} V_{[\lambda_i]}.$$

3. *The dimension of V is equal to the sum of the generalized multiplicities of all the eigenvalues:*

$$\dim V = \sum_{i=1}^{r} M(\lambda_i).$$

Proof. Part (1) of the theorem is almost obvious. I will prove (2) by induction on r (the number of *distinct* eigenvalues). Write p_T for the characteristic polynomial of T (Definition 3.3). Because $\mathbb{C}$ is algebraically closed, p_T is a product of linear factors. Because the roots of p_T are precisely the eigenvalues of T, these linear factors must each be $x - \lambda_i$ for some i. Therefore there are positive integers m_i so that

$$p_T(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_r)^{m_r} \tag{5.7a}$$

If $r = 0$, then $p_T = 1$, $V = 0$, $T = 0$, and there is nothing to prove: the zero vector space is an empty direct sum. So suppose $r \geq 1$, so that λ_1 is an eigenvector, and

$$0 \neq V_{\lambda_1} \subset V_{[\lambda_1]}.$$

We factor the polynomial p_T as

$$p_T = p_1 p_2, \quad p_1 = (x - \lambda_1)^{m_1}, \quad p_2 = (x - \lambda_2)^{m_2} \cdots (x - \lambda_r)^{m_r} \tag{5.7b}$$

The only irreducible factor of p_1 is $x - \lambda_1$, and that is not a factor of p_2; so p_1 and p_2 are relatively prime (2.9). According to Proposition 2.11, we can find polynomials a_1 and a_2 so that

$$a_1 p_1 + a_2 p_2 = 1. \tag{5.7c}$$

We now define linear transformations P and Q on V by

$$P = a_2(T) p_2(T), \qquad Q = a_1(T) p_1(T). \tag{5.7d}$$

Applying the identity (5.7c) to T gives

$$P + Q = a_2(T) p_2(T) + a_1(T) p_1(T) = 1(T) = I. \tag{5.7e}$$

Similarly, the identity $p_T(T) = 0$ (Definition 3.3), together with the factorization (5.7b), gives

$$PQ = a_2(T) p_2(T) a_1(T) p_1(T) = a_1(T) a_2(T) p_T(T) = 0. \tag{5.7f}$$

So P and Q satisfy the conditions of Proposition 4.2(2), so we get a direct sum decomposition

$$V = U \oplus W. \tag{5.7g}$$

We find

$$\begin{aligned}
U &= \mathrm{Range}(P) && \text{(Proposition 4.2)} \\
&= \mathrm{Range}(p_2(T)a_2(T)) && \text{(definition of } P\text{)} \\
&\subset \mathrm{Null}(p_1(T)) && \text{(since } p_1(T)p_2(T) = p_T(T) = 0\text{)} \\
&\subset \mathrm{Null}(a_1(T)p_1(T)) && \\
&= \mathrm{Null}(Q) = U && \text{(Proposition 4.2)}
\end{aligned} \tag{5.7h}$$

So the inclusions are equalities, and (by the middle formula)

$$U = \mathrm{Null}((T - \lambda_1)^{m_1}) = V_{[\lambda_1]}. \tag{5.7i}$$

Similarly we find

$$W = \mathrm{Null}(p_2(T)), \tag{5.7j}$$

so W is a T-invariant subspace on which the eigenvalues of T are the $r-1$ scalars $\lambda_2, \ldots, \lambda_r$. By inductive hypothesis,

$$W = \bigoplus_{i=2}^{r} W_{[\lambda_i]}, \tag{5.7k}$$

and therefore

$$V = U \oplus W = V_{[\lambda_1]} \oplus \bigoplus_{i=2}^{r} W_{[\lambda_i]}. \tag{5.7l}$$

as we wished to show.

Statement (3) is just the fact the dimension of a direct sum is the sum of the dimensions of the summands. $\square$

Example 5.8.

$$V = \mathbb{R}^3,$$

$$T = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & -1 & -1 \end{pmatrix}.$$

I defined

$$C = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix},$$

the block in the lower right of T. The characteristic polynomial of the 1×1 matrix (1) is $x - 1$, so according to Proposition 3.2, the characteristic polynomial of T is

$$p_T = (x-1)p_C.$$

We calculated $p_C = x^2$ (which is a good exercise using Proposition 3.1). So

$$p_T = (x-1)x^2,$$

and the two eigenvalues are 1 and 0. We calculated

$$V_1 = \mathbb{R} \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad V_0 = \mathbb{R} \cdot \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}.$$

The eigenspaces are not quite big enough to span $\mathbb{R}^3$. The *generalized* eigenspace for 0 has a chance to be bigger, because 0 is a double root of p_T. In fact

$$\begin{aligned} V_{[0]} &= \mathrm{Null}(T^2) \\ &= \mathrm{Null} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ &= \mathrm{span}\left(\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 3 \\ -1 \\ -1 \end{pmatrix} \right). \end{aligned}$$

6 Eigenpolynomials

In this section (corresponding to Chapter 9 of the text) we look for replacements for eigenvalues to use over fields other than $\mathbb{C}$. Recall from (5.1) the description of an eigenspace

$$V_\lambda = \mathrm{Null}(q(T)) \qquad (q(x) = x - \lambda).$$

Because $x - \lambda$ is an irreducible polynomial, this description suggests a generalization.

Definition 6.1. Suppose q is a monic irreducible polynomial in $\mathcal{P}(F)$, and $T \in \mathcal{L}(V)$. The *q-eigenspace of T* is

$$V_q =_{\mathrm{def}} \mathrm{Null}(q(T)) = \{v \in V \mid q(T)v = 0\} \subset V.$$

If $V_q \neq 0$, we say that q is an *eigenpolynomial* for T. The *multiplicity* of the eigenpolynomial q is by definition

$$m(q) = \frac{\dim V_q}{\deg q}.$$

From its definition the multiplicity looks like a rational number, but we'll see in Proposition 6.2 that it's actually an integer.

Comparing (5.1) to Definition 6.1, we see that λ is an eigenvalue of T if and only if $x - \lambda$ is an eigenpolynomial of T.

I don't know a good and widely used name for what I called "eigenpolynomial."

Proposition 6.2. *Suppose*

$$q(x) = x^d + a_{d-1}x^{d-1} + \cdots + a_0$$

is a monic irreducible polynomial, and $T \in \mathcal{L}(V)$.

1. *The subspace $V_q = \mathrm{Null}(q(T)$ is a T-invariant subspace of V.*

2. *The subspace V_q is not zero if and only if q divides the characteristic polynomial of T.*

3. *Suppose v_0 is any nonzero vector in V_q. Define $v_i = T^i v_0$. Then the list $(v_0, \ldots v_{d-1})$ is linearly independent, and*

$$v_d = -a_{d-1}v_{d-1} - \cdots - a_0 v_0.$$

The subspace $\mathrm{span}(v_0, \ldots, v_{d-1})$ is T-invariant, and the matrix of T on this subspace and in this basis is the $d \times d$ matrix

$$A_q =_{def} \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ & & \ddots & \vdots & \\ 0 & 0 & \cdots & 1 & -a_{d-1} \end{pmatrix}.$$

4. *There is a basis of V_q in which the matrix of T is block diagonal, with every block equal to A_q.*

5. *The characteristic polynomial of T acting on V_q is a power of q.*

6. *The dimension of V_q is a multiple of the degree d of q; so the multiplicity $m(q)$ is an integer.*

The general formula for the matrix A_q looks a bit peculiar, but sometimes it's quite nice. For example

$$A_{x-\lambda} = (\lambda), \qquad A_{x^2+1} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}; \tag{6.3}$$

the second formula applies whenever $x^2 + 1$ is irreducible, for example for the real numbers. The proposition is also true if A_q is replaced by any other matrix A'_q having characteristic polynomial q. Suppose for example that $F = \mathbb{R}$, and that q is the real polynomial having roots

$$re^{\pm i\theta} = r\cos\theta \pm ir\sin\theta = a \pm ib, \tag{6.4a}$$

with $r > 0$ and $0 < \theta < \pi$ (equivalently, $b > 0$). Then

$$q(x) = x^2 - 2r\cos\theta + r^2 = x^2 - 2ax + a^2 + b^2, \tag{6.4b}$$

Consequently

$$A_q = \begin{pmatrix} 0 & -r^2 \\ 1 & 2r\cos\theta \end{pmatrix} = \begin{pmatrix} 0 & -a^2 - b^2 \\ 1 & 2a \end{pmatrix}. \tag{6.4c}$$

This formula is not so nicely related to what we understand about complex numbers. So it is often convenient to use instead

$$A'_q = \begin{pmatrix} r\cos\theta & -r\sin\theta \\ r\sin\theta & r\cos\theta \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \qquad (b > 0), \tag{6.4d}$$

which also has characteristic polynomial q. The matrices A'_q appear in the text in Theorem 9.34 (concerning *normal* linear transformations on real inner product spaces). But they don't need to have anything to do with inner products.

According to Proposition 6.2, these "eigenspaces" for irreducible polynomials behave very much like eigenspaces for eigenvalues. Using exactly the same proof as for eigenvalues, one can deduce

Corollary 6.5. *Suppose V is a finite-dimensional vector space, and $T \in \mathcal{L}(V)$. Then there is a basis of T in which the matrix of T is block-upper-triangular, and every diagonal block is one of the matrices A_q, with q an irreducible factor of the characteristic polynomial of T.*

To have a complete theory, we just need to add the word "generalized."

Definition 6.6. Suppose q is a monic irreducible polynomial, and $T \in \mathcal{L}(V)$. The *q-generalized eigenspace of T* is

$$V_{[q]} =_{\text{def}} \{v \in V \mid q(T)^m v = 0 \text{ (some } m > 0).\} \subset V.$$

Clearly

$$V_q \subset V_{[q]};$$

The *generalized multiplicity $M(q)$* of the eigenpolynomial q is

$$M(q) =_{\text{def}} \frac{\dim V_{[q]}}{\deg q}.$$

Theorem 6.7 (text, Theorem 8.21)**.** *Suppose V is a finite-dimensional vector space, $T \in \mathcal{L}(V)$, p_T is the characteristic polynomial of T (a monic polynomial of degree equal to the dimension of V. Factor p_T into distinct irreducible factors in accordance with Proposition 2.7:*

$$p_T(x) = \prod_{i=1}^{r} q_i(x)^{m_i}.$$

1. *Each generalized eigenspace $V_{[q_i]}$ is preserved by T.*

2. *The space V is the direct sum of the generalized eigenspaces:*

$$V = \bigoplus_{i=1}^{r} V_{[q_i]}.$$

3. *The dimension of V is*

$$\dim V = \sum_{i=1}^{r} M(q_i) \deg q_i.$$

The proof is almost word-for-word the same as the proof of Theorem 5.6; it's a good exercise to try to write down the proof without consulting that one.

References

[1] Michael Artin, *Algebra*, Prentice Hall Inc., Englewood Cliffs, NJ, 1991.