H and K notes on eigenstuff

Jonathan Lin

July 5, 2019

So the goal of eigen analysis and all that of linear operators is that we would like to analyze a linear operator and "decompose it" into some sort of simpler structure. The canonical example are diagonalizable matrices: if we can diagonalize a matrix we have definitely simplified the problem greatly.

1 All Preliminary Definitions and basic results

Definition 1. Let V be a vector space over the field F and let T be a linear operator on V. An *eigenvalue* of T is a scalar c of F such that there is a non-zero vector v in V such that $T(v) = cv$. If c is an eigenvalue of T then any such vector where $T(v) = cv$ is called an *eigenvector* of T associated with c and the collection of all v such that $T(v) = cv$ is called the *eigenspace* associated with c.

It is clear that the eigenspace associated with c is a subspace of V.

The eigenspace associated with c is the null space of the linear transformation $T - cI$ (Forward and converse directions are very easy). Then an equivalent definition would be that c is an eigenvalue of T if $T - cI$ is not the zero subspace. In summary:

Theorem 1. *Let T be a linear operator on a finite dimensional vector space V and let c be a scalar. The following are equivalent:*

1. *c is a characteristic value of T*

2. *$T - cI$ is not an invertible linear transformation*

3. *$\det(T - cI) = 0$.*

We define the characteristic polynomial of T to be $\det(T - cI)$. In the next proposition we show that similar matrices have the same characteristic polynomial, so that the characteristic polynomial of a linear operator can be defined unambiguously.

Proposition 1. *Similar matrices have the same characteristic polyomial.*

Proof. If $B = PAP^{-1}$, then

$$\det(B - cI) = \det(PAP^{-1} - cI)$$
$$= \det(P(A - cI)P^{-1})$$
$$= \det(A - cI)$$

so we are done. $\qquad\square$

If V is an n dimensional vector space and $T \in \mathcal{L}(V, V)$ there are 4 things that could happen:

- T has no characteristic values.

- T has exactly n characteristic values.

- T has less than n characteristic values, but a collection of eigenvalues spans V.

- T has less than n characteristic values and no collection of eigenvalues spans V.

Definition 2. A linear operator T is diagonalizable if there is a basis for V where each vector in the basis is an eigenvector of T.

Suppose that T is diagonalizable and $c_1, \ldots, c_n$ the distinct eigenvalues of T. Then there is an ordered basis B where T is represented by a diagonal matrix where the scalars c_i are on the main diagonal repeated a certain number of times. We deduce that the characteristic polynomial for T is the product of linear factors $(\lambda - c_i)$ and hence the characteristic polynomial will have the form

$$(x - c_1)^{d_1} \cdots (x - c_k)^{d_k}.$$

Now we show that eigenvectors with different eigenvalues are mutually linearly independent.

Theorem 2. *Let T be a linear operator on the finite-dimensional vector space V. Let $c_1, \ldots, c_k$ be the distinct characteristic values of T and let W_i be the eigenspace of c_i. If $W = W_1 + \cdots + W_k$, then*

$$\dim W = \dim W_1 + \cdots + \dim W_k.$$

In fact, if we let B_i be a basis for W_i (for each i) then we have that $B = (B_1, \ldots, B_k)$ is basis for W.

Proof. It suffices to show that W is a direct sum of the W_i, that is, if $w_i \in W_i$ and $w_1 + \cdots + w_k = 0$ then all the w_i are 0.

Suppose $w_i \in W_i$ and that $w_1 + \cdots + w_k = 0$. For any polynomial f we have that

$$0 = f(T)0 = f(T)w_1 + \cdots + f(T)w_k$$
$$= f(c_1)w_1 + \cdots = f(c_k)w_k.$$

If we choose a polynomial f_i such that $f_i(c_j) = \delta_{ij}$ then it follows that $0 = f_i(c_i)w_i = w_i$. So we are done. For the last statement this just follows from properties of linear independence and the invariance of dimension. $\square$

Corollary 1. *The following statements are all equivalent:*

1. T is diagonalizable.

2. The characteristic polynomial for T is

$$(x - c_1)^{d_1} \cdots (x - c_k)^{d_k}.$$

and $dim W_i = d_i$ for i from 1 to k.

3.

$$\dim W = \dim W_1 + \cdots + \dim W_k.$$

Proof. This is pretty straightforward once you consider the previous lemma. □

Here is a matrix analogue of the previous theorem. Let A be an $n \times n$ matrix with entries in a field F and let $c_1, \ldots, c_k$ be the distinct eigenvalues of A in F. For each i let W_i be the null space os $A - c_i I$ and let B_i be an ordered basis for W_i. Then we can string the bases to form the columns of a matrix P. Then the previous theorem implies that A is similar to a diagonal matrix if and only if P is a square matrix. When P is square, P is invertible and $P^{-1}AP$ is diagonal.

2 Polynomials which annihilate a linear operator T

Suppose T is a linear operator on V. If p is a polynomial in $F[x]$, then $p(T)$ is a linear operator on V. We say that p annihilates T if we have that $p(T) = 0$. It is easy to see that the collection of polynomials p which annihilate T are an ideal in $F[x]$.

If V is finite-dimensional, then it is not true that this ideal is the zero ideal. To see why, note that $\dim(\mathcal{L}(V, V)) = n^2$, so that the $n^2 + 1$ vectors $I, T, \ldots, T^{n^2}$ are linearly dependent. So there are non-zero scalars such that

$$a_0 I + \cdots + a_{n^2} T^{n^2} = 0.$$

We say the minimal polynomial for T is the unique monic generator of the annihilators of T.

Suppose that $A \in M_{n \times n}(F)$ and suppose that $F_1 \supset F$ is an extension field (or whatever correct term to use) of F. Then it is true that $A \in M_{n \times n}(F_1)$. We claim that the minimal polynomial is the same in both cases. To see why, we note that over F the minimal polynomial is

$$f(x) = x^k + \sum_{j=0}^{k-1} a_j x^j.$$

Expanding any polynomial at A we obtain a system of n^2 homogeneous linear equations for $a_0, a_1, \ldots, a_{k-1}$. The coefficients for these lie in F and it is clear that if a solution exists in F_1 a solution must also exist for F. So the two minimal polynomials are actually the same.

Here is the important trick used here: Suppose A and b are a matrix and vector with entries of a field F and suppose $Ax = b$ has some non-trivial solutions in some extension field F_1. Then it is clear that row-reduction will find a solution, all of whose entries are in F.

Theorem 3. *Suppose T is a linear operator on a finite dimensional vector space V. Then the characteristic polynomial and the minimal polynomial for T have the same roots disregarding multiplicities.*

Proof. It suffices to show that $p(c) = 0$ if and only if c is an eigenvalue of T.

First, suppose $p(c) = 0$. Then there is a unique polynomial q such that

$$p(x) = (x - c)q(x).$$

Choose any vector v such that $q(T)v \neq 0$ (this is possible because p is the minimal polyonmial). Then we have that

$$0 = p(T)v = (T - cI)q(T)v$$

which implies that $q(T)v$ is an eigenvector of T.

Conversely, suppose that c is an eigenvalue of T (that is, $T(v) = cv$ for some vector v). Then we have that $0 = p(T)v = p(c)v$. Since we assume v to be non-zero this implies that $p(c) = 0$, as desired. $\square$

We observe that p must be the product of the linear polynomials which have eigenvalues as their roots.

Here is an important theorem about the relationship between the minimal polynomial and the characteristic polynomial.

Theorem 4. *The minimal polynomial divides the characteristic polynomial. That is, if p is the characteristic polynomial of some linear operator T then $p(T) = 0$.*

There are a couple of proofs of this theorem.

Proof. The idea of the proof is that we move the perspective to the ring of polynomial operators of T. Using a determinant which we can see is $p(T)$, where p is the characteristic polynomial, then we can more easily see that T must be 0.

Let $K = F[T]$, that is, the set of polynomials in T with coefficients in F. Choose a basis $v_1, \ldots, v_n$ for V, and let A be the matrix which represents T in the given basis.

Then we have that by the definition of matrix multiplication, that

$$T(v_i) = \sum_{j=1}^{n} A_{ji}v_j \implies \sum_{j=1}^{n}(\delta_{ij}T - A_{ji}I)v_j = 0.$$

Let B denote the element of $K^{n \times n}$ with entries

$$B_{ij} = \delta_{ij}T - A_{ij}I.$$

As an example, when $n = 2$, we have

$$B = \begin{bmatrix} T - A_{11}I & -A_{21}I \\ -A_{12}I & T - A_{22}I \end{bmatrix}$$

It is clear through both commutativity of $F[T]$ and the ring isomorphism $T \mapsto x$ to the ring $F[x]$ that $\det B = f(T)$ where f is the characteristic polynomial of T.

4

We would like to show now that $(\det B)v_j = 0$ for all $1 \leq j \leq n$, this implies that $\det B = 0$ as desired.

Let $\tilde{B}$ be the classical adjoint of B. By the definition of B, we have that

$$\sum_{j=1}^{n} B_{ij}v_j = 0.$$

Doing some double sum shenanigans we have that

$$
\begin{aligned}
0 &= \sum_{j=1}^{n} B_{ij}v_j \\
&= \sum_{j=1}^{n} \tilde{B}_{ki} B_{ij} v_j \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \tilde{B}_{ki} B_{ij} v_j \\
&= \sum_{j=1}^{n} \sum_{i=1}^{n} \tilde{B}_{ki} B_{ij} v_j \\
&= \sum_{j=1}^{n} \delta_{kj} \det B v_j \\
&= \det B v_k,
\end{aligned}
$$

hence we are done. $\qquad\square$

3   Invariant Subspaces

Definition 3. Let $T \in \mathcal{L}(V, V)$. Then W is invariant under T if for each vector w in W, $T(w)$ is in W.

Let $T \in \mathcal{L}(V, V)$, and let U be any linear operator which commutes with T. Let $W = \operatorname{im} U$ and $N = \ker U$. Then both U and N are invariant under T. Suppose $w \in \operatorname{im} U$. Then $w = U(v)$ for some $v \in V$ so that $T(w) = T(U(v)) = U(T(v)) \in \operatorname{im} U$. Similarly, if $w \in N$, we have that $U(T(w)) = T(U(w)) = T(0) = 0$.

For example, suppose $U = p(T)$ for some polynomial p. Then $UT = TU$. It follows that letting p be the characteristic polynomial that the eigenspace of all eigenvectors with some eigenvalue c is invariant under T, since explicitly if a vector is in the null space of $T - cI$ by the previous remarks it is invariant under T.

Suppose W is invariant under T. Then T induces a linear functional $T_W \in \mathcal{L}(W, W)$ which is the same as T but restricted to W. Suppose we choose an ordered basis $B = (v_1, \ldots, v_n)$ where $B' = (v_1, \ldots, v_r)$ is a basis for W. Letting $A = [T]_B$, we see that A has the block form

$$
\begin{bmatrix} B & C \\ 0 & D \end{bmatrix},
$$

where B is the matrix $[T]_{B'}$. It follows from the block form that $\det A - \lambda I = \det B - \lambda I \det D - \lambda I$. Hence the characteristic polynomial of T_W divides the characteristic polynomial of T. Also, since

$$A^k = \begin{bmatrix} B^k & C_k \\ 0 & D^k \end{bmatrix},$$

for some $r \times n - r$ matrix C_k, it follows that any annihilator of A will also annihilate B. Hence the minimal polynomial for B divides the minimal polynomial for A.

Instead of considering all polynomials which annihilate the whole space, or even a subspace, we can consider an even more general definition of polynomials whose evaluations at T will map vector v into some subspace W. Here is the rigorous definition.

Definition 4. Let W be an invariant subspace for T and let v be a vector in V. Then the T-conductor of v, $S_T(v; W)$, is the set which consists of all polynomials $g \in F[x]$ such that $g(T)v \in W$.

When $W = \{0\}$ the T-conductor for v is instead called the T-annihilator for v.

Proposition 2. *If W is invariant under T then W is invariant under $p(T)$ for any polynomial $p \in F[x]$. Hence the conductor $S_T(v; W)$ is an ideal in $F[x]$.*

Proof. For the first statement, it suffices to show that W is invariant under T^k for all T. This is done using induction.

Now suppose $f, g \in S_T(v; W)$. Then $cf(T) + g(T) \in W$. Now suppose $g \in S_T(v; W)$. Then $f(T)(g(T)w) = fg(T)w \in W$. We conclude that $S_T(v; W)$ is an ideal. $\square$

We call the unique monic generator for $S_T(v; W)$ the T-conductor of v into W. This is used interchangeably with the set (since the set can be associated with its respective monic generator).

Proposition 3. *Let V be a finite-dimensional vector space over F. Let T be a linear operator on V such that the minimal polynomial for T is a product of linear factors $p(x) = (x - c_1)^{r_1} \cdots (x - c_k)^{r_k}$, with $c_i \in F$.*

Let W be a proper subspace of V which is invariant under T. There exists a vector $v \in V$ such that $v \notin W$ and $(T - cI)v \in W$ for some eigenvalue c of T.

Proof. The key idea in the proof is that the T-conductor for any vector divides the minimal polynomial.

Let x be any vector not in W, ad let g be the T-conductor of x into W. Then $g \mid p$, where p is the minimal polynomial of T. By assumption, g cannot be constant, hence

$$g(\lambda) = (\lambda - c_1)^{e_1} \cdots (\lambda - c_n)^{e_n},$$

where at least one of e_1 through e_n is non-zero. Then $g(\lambda) = (\lambda - c_j)h(\lambda)$ for some c_j and a unique h (determined by c_j). By the definition of g, $h(T)x \notin W$, but

$$(T - c_j I)(h(T)v) = g(T)v \in W$$

so we are done. $\square$

Using this proposition we can characterize the linear operators which admit an upper/lower triangular form with respect to some basis B.

Theorem 5. *Let V be a finite dimensional vector space over the field F and let T be a linear operator on V. Then T is triangulable if and only if the minimal polynmoial for T splits over F.*

Proof. Suppose the minimal polynomial factors $p(\lambda) = (\lambda - c_1)^{r_1} \cdots (\lambda - c_n)^{r_n}$.

Apply the lemma above to 0 to obtain a vector v_1 such that $(T - c_1 I)v = 0$. Then, apply the lemma to v_1 and we obtain v_2 such that $(T - c_2 I)v_2 = kv_1 \implies T(v_2) = c_2 v_2 + k_1 v_1$.

At the $(k+1)$th step of the algorithm we apply the lemma to $\{v_1, \ldots, v_k\}$ to obtain v_{k+1} such that

$$(T - c_{j+1}I)v_{k+1} \in \text{span}(v_1, \ldots, v_k).$$

Note that this process generates subspaces

$$0 = W \subset W_1 \subset \cdots \subset W_n = V$$

which are all invariant under T.

Conversely, if T is triangulable it must be that the characteristic polynomial is a product of linear subspaces (the determinant of a triangular matrix is the product of the diagonals). $\square$

Here is a similar theorem, found in Axler. In Axler the only fields he works with are $\mathbb{R}$ and $\mathbb{C}$, so the theorem is framed in terms of a vector space over $\mathbb{C}$.

Theorem 6. *Suppose V is a complex vector space and $T \in \mathcal{L}(V, V)$. Then T has an upper-triangular matrix with respect to some basis of V.*

Proof. The result is clear if $\dim V = 1$.

Suppose $\dim V = n$ and the result holds for all vector spaces with dimension less than n. Let λ be any eigenvalue of T, and let $U = \text{im}(T - \lambda I)$. By rank-nullity it is clear that $T - \lambda I$ is not surjective. So $\dim U < \dim V$.

It is also straightforward to see that U is invariant under T. If $u \in U$ then $T(u) = (T - \lambda I)u + \lambda u \in U$. It follows that the restriction operator T_U is well defined. By the induction hypothesis there is a basis $(u_1, \ldots, u_m)$ of U for which the striction is an upper triangular matrix.

Extending $(u_1, \ldots, u_m)$ to a basis $(u_1, \ldots, u_m, v_1, \ldots, v_n)$ of V, for k we have $T(v_k) = (T - \lambda I)v_k + \lambda v_k$. Then

$$T(v_k) \in \text{span}(u_1, \ldots, u_m, v_1, \ldots, v_k),$$

as desired. $\square$

Note that the second proof requires the result that all complex linear operators have at least one eigenvalue. This is covered in the proof in H and K because the splitting of the characteristic polynomial implies that the operator has at least one eigenvalue. Hence the first proof is actually more general than the second one.

Theorem 7. *A linear operator T is diagonalizable if and only if the minimal polynomial splits into distinct linear factors.*

Proof. Suppose T is diagonalizable. Then it has a diagonal matrix with respect to some basis, where its diagonal entries represent all the eigenvalues of λ. It is clear that the polynomial $p(\lambda) = (\lambda - c_1) \cdots (\lambda - c_n)$ is a polynomial which annihilates T, because it is a matrix multiplication of diagonal matrices, with each entry having a product which is zero somewhere. Hence the total matrix product is just the zero matrix. It is clear that if q is any polynomial which divides p, then some entries on the matrix will remain non-zero.

Conversely, let W be the subaspace spanned by all the eigenvectors, and suppose $W \neq V$. There is some $v \notin W$ and some λ_j which is an eigenvalue of T such that

$$(T - \lambda_j I)(v) = w \in W.$$

We have $w = w_1 + \cdots + w_k$ where $T(w_j) = \lambda_j w_j$. Hence we have

$$h(T)w = h(T)w_1 + \cdots + h(T)w_k$$
$$= h(\lambda_1)w_1 + \cdots + h(\lambda_k)w_k.$$

By the divisor and remainder theorems, respectively, there exist polynomials $p(x) = (x - \lambda_j)q(x)$ and $q(x) - q(\lambda_j) = (x - c_j)h(x)$. It follows that

$$q(T) - q(c_j)I = (T - c_j I)h(T)$$
$$\implies q(T)v - q(c_j)v = (T - c_j I)(h(T)v) = h(T)w.$$

we have that $h(T)w \in W$ and $0 = p(T)v = (T - \lambda_j I)(q(T)v)$, so that $q(T)v \in W$. This and the above equation implies that $q(c_j)v \in W$. But this implies that $q(c_j) = 0$ since $v \notin W$. So p has multiple roots. $\square$

this apparently gives a proof of Cayley Hamilton.