Math 110: Linear Algebra
UC Berkeley

Kelvin Lee

October 15, 2020

These are course notes for UC Berkeley's Spring 2020 Math 110 Linear Algebra, instructed by Professor Kenneth Ribet. These notes were created with main reference to Sheldon Axler's Linear Algebra Done Right(3rd Edition) and Professor Ribet's Math 110 lecture. Every section number mentioned refers to the LADR text.

Contents

§1 Vector Spaces

§1.1 Definition of Vector Space

A vector space is defined to be a set V with an addition and a scalar multiplication on V that satisfy the properties that will be introduced below.

Definition 1.1 (addition, scalar multiplication)

1. An addition on a set V is a function that assigns an element $u + v \in V$ to each pair of elements $u, v \in V$.
2. A scalar multiplication on a set V is a function that assigns an element $\lambda v \in V$ to each $\lambda \in \mathbf{F}$ and each $v \in V$.

The formal definition of a vector space is as follows:

Definition 1.2 (vector space)

A ***vector space*** is a set V along with an addition on V and a scalar multiplication on V such that the following properties hold:

commutativity
$u + v = v + u$ for all $u, v \in V$;
associativity
$(u + v) + w = u + (v + w)$ and $(ab)v = a(bv)$ for all $u, v, w \in V$ and all $a, b \in \mathbf{F}$;
additive identity
there exists an element $0 \in V$ such that $v + 0 = v$ for all $v \in V$;
additive inverse
for every $v \in V$, there exists $w \in V$ such that $v + w = 0$;
multiplicative identity
$1v = v$ for all $v \in V$;
distributive properties
$a(u + v) = au + av$ and $(a + b)v = av + bv$ for all $a, b \in \mathbf{F}$ and all $u, v \in V$.

Lemma 1.3 (Unique additive identity)

A vector space has a unique additive identity.

Proof. Suppose 0 and 0' are both additive identities for some vector space V. Then

$$0' = 0' + 0 = 0 + 0' = 0,$$

where the first equality holds because 0 is an additive identity, the second equality comes from commutativity, and the third equality holds because $0'$ is an additive identity. Thus $0' = 0$, proving that V has only one additive identity. □

Lemma 1.4 (Unique additive inverse)

Every element in a vector space has a unique additive inverse.

Proof. Suppose V is a vector space. Let $v \in V$. Suppose w and w' are additive inverses of v. Then

$$w = w + 0 = w + (v + w') = (w + v) + w' = 0 + w' = w'.$$

Thus $w = w'$, as desired. ∎

Notation 1.5.

§2 Linear Maps

Notation 2.1 ($\mathbf{F}, V, W$)**.** F denotes $\mathbf{R}$ or $\mathbf{C}$; V and W denote vector spaces over $\mathbf{F}$.

§2.1 The Vector Space of Linear Maps

§2.1.1 Definition and Examples of Linear Maps

Now we are ready for one of the key definitions in linear algebra.

Definition 2.2 (linear map)
A ***linear map*** from V to W is a function $T : V \to W$ with the following properties:

additivity
$T(u+v) = Tu + Tv$ for all $u, v \in V$;
homogeneity
$T(\lambda v) = \lambda(Tv)$ for all $\lambda \in \mathbf{F}$ and all $v \in V$.

Note: ***Linear transformation*** *and* ***linear map*** *are identically the same concept.*

Notation 2.3 ($\mathcal{L}(V, W)$)**.** The set of all linear maps from V to W is denoted $\mathcal{L}(V, W)$.

Example 2.4 (linear maps)
zero
In addition to its other uses, we let the symbol 0 denote the function that takes each element of some vector space to the additive identity of another vector space. To be specific, $0 \in \mathcal{L}(V, W)$ is defined by

$$0v = 0.$$

The 0 on the left side of the equation above is a function from V to W, whereas the 0 on the right side is the additive identity in W. As usual, the context should allow you to distinguish between the many uses of the symbol 0.
identity
The identity *map*, denoted I, is the function on some vector space that takes each element to itself. To be specific, $I \in \mathcal{L}(V, V)$ is defined by

$$Iv = v.$$

Example 2.5 (continued)

differentiation
Define $D \in \mathcal{L}(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R}))$ by $Dp = p'$ The assertion that this function is a linear map is another way of stating a basic result about differentiation: $(f+g)' = f' + g'$ and $(\lambda f)' = \lambda f'$ whenever f, g are differentiable and λ is a constant.

integration
Define $T \in \mathcal{L}(\mathcal{P}(\mathbf{R}), \mathbf{R})$ by

$$Tp = \int_0^1 p(x)dx.$$

The assertion that this function is linear is another way of stating a basic result about integration: the integral of the sum of two functions equals the sum of the integrals, and the integral of a constant times a function equals the constant times the integral of the function. multiplication by x^2 Define $T \in \mathcal{L}(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R}))$ by

$$(Tp)(x) = x^2 p(x)$$

for $x \in \mathbf{R}$.

backward shift
Recall that $\mathbf{F}^\infty$ denotes the vector space of all sequences of elements of $\mathbf{F}$. Define $T \in \mathcal{L}(\mathbf{F}^\infty, \mathbf{F}^\infty)$ by

$$T(x_1, x_2, x_3, \ldots) = (x_2, x_3, \ldots)$$

from $\mathbf{R}^3$ to $\mathbf{R}^2$
Define $T \in \mathcal{L}(\mathbf{R}^3, \mathbf{R}^2)$ by

$$T(x, y, z) = (2x - y + 3z, 7x + 5y - 6z).$$

from $\mathbf{F}^n$ to $\mathbf{F}^m$
Generalizing the previous example, let m and n be positive integers, let $A_{j,k} \in \mathbf{F}$ for $j = 1, \ldots, m$ and $k = 1, \ldots, n$, and define $T \in \mathcal{L}(\mathbf{F}^n, \mathbf{F}^m)$ by

$$T(x_1, \ldots, x_n) = (A_{1,1}x_1 + \cdots + A_{1,n}x_n, \ldots, A_{m,1}x_1 + \cdots + A_{m,n}x_n)$$

Actually every linear map from $\mathbf{F}^n$ to $\mathbf{F}^m$ is of this form.

The following result illustrates that a linear map is completely determined by its values on a basis, i.e. we can find a linear map that takes on whatever values we wish on the vectors in a basis.

Lemma 2.6 (Linear maps and basis of domain)
Suppose $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_n \in W$. Then there exists a unique linear map $T : V \to W$ such that

$$Tv_j = w_j$$

for each $j = 1, \ldots, n$.

Proof. First we show the existence of a linear map T with the desired property. Define $T : V \to W$ by

$$T(c_1 v_1 + \cdots + c_n v_n) = c_1 w_1 + \cdots + c_n w_n$$

where $c_1, \ldots, c_n$ are arbitrary elements of $\mathbf{F}$. The list $v_1, \ldots, v_n$ is a basis of V, and thus the equation above does indeed define a function T from V to W (because each element of V can be uniquely written in the form $c_1 v_1 + \cdots + c_n v_n$) For each j, taking $c_j = 1$ and the other c 's equal to 0 in the equation above shows that $Tv_j = w_j$ If $u, v \in V$ with $u = a_1 v_1 + \cdots + a_n v_n$ and $v = c_1 v_1 + \cdots + c_n v_n$, then

$$\begin{aligned} T(u+v) &= T((a_1 + c_1) v_1 + \cdots + (a_n + c_n) v_n) \\ &= (a_1 + c_1) w_1 + \cdots + (a_n + c_n) w_n \\ &= (a_1 w_1 + \cdots + a_n w_n) + (c_1 w_1 + \cdots + c_n w_n) \\ &= Tu + Tv. \end{aligned}$$

Similarly, if $\lambda \in \mathbf{F}$ and $v = c_1 v_1 + \cdots + c_n v_n$, then

$$\begin{aligned} T(\lambda v) &= T(\lambda c_1 v_1 + \cdots + \lambda c_n v_n) \\ &= \lambda c_1 w_1 + \cdots + \lambda c_n w_n \\ &= \lambda (c_1 w_1 + \cdots + c_n w_n) \\ &= \lambda T v. \end{aligned}$$

Thus T is a linear map from V to W. To prove uniqueness, now suppose that $T \in \mathcal{L}(V, W)$ and that $Tv_j = w_j$ for $j = 1, \ldots, n$. Let $c_1, \ldots, c_n \in \mathbf{F}$. The homogeneity of T implies that $T(c_j v_j) = c_j w_j$ for $j = 1, \ldots, n$. The additivity of T now implies that

$$T(c_1 v_1 + \cdots + c_n v_n) = c_1 w_1 + \cdots + c_n w_n.$$

Thus T is uniquely determined on $\operatorname{span}(v_1, \ldots, v_n)$ by the equation above. Because $v_1, \ldots, v_n$ is a basis of V, this implies that T is uniquely determined on V. $\square$

§2.1.2 Algebraic Operations on $\mathcal{L}(V, W)$

We begin by definition addition and scalar multiplication on $\mathcal{L}(V, W)$.

Definition 2.7 (Definition addition and scalar multiplication on $\mathcal{L}(V, W)$)
Suppose $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbf{F}$. The sum $S + T$ and the product λT are the linear maps from V to W defined by

$$(S+T)(v) = Sv + Tv \quad \text{and} \quad (\lambda T)(v) = \lambda(Tv)$$

for all $v \in V$.

You should verify that the sum and the product above are indeed linear.

Lemma 2.8 ($\mathcal{L}(V, W)$ is a vector space)

With the operations of addition and scalar multiplication as defined above, $\mathcal{L}(V, W)$ is a vector space.

The routine proof of the result above is left to the reader. Note that the additive identity of $\mathcal{L}(V, W)$ is the zero linear map defined earlier in this section.

Usually it makes no sense to multiply together two elements of a vector space, but for some pairs of linear maps a useful product exists. We will need a third vector space, so for the rest of this section suppose U is a vector space over $\mathbf{F}$.

Definition 2.9 (Definition Product of Linear Maps)

If $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then the product $ST \in \mathcal{L}(U, W)$ is defined by

$$(ST)(u) = S(Tu)$$

for $u \in U$.

Lemma 2.10 (Algebraic properties of products of linear maps)

associativity

$$(T_1T_2)\,T_3 = T_1\,(T_2T_3)$$

whenever T_1, T_2, and T_3 are linear maps such that the products make sense (meaning that T_3 maps into the domain of T_2, and T_2 maps into the domain of T_1).

identity

$$TI = IT = T$$

whenever $T \in \mathcal{L}(V, W)$ (the first I is the identity map on V, and the second I is the identity map on W).

distributive properties

$$(S_1 + S_2)\,T = S_1T + S_2T \quad \text{and} \quad S\,(T_1 + T_2) = ST_1 + ST_2$$

whenever $T, T_1, T_2 \in \mathcal{L}(U, V)$ and $S, S_1, S_2 \in \mathcal{L}(V, W)$.

The routine proof of the result above is left to the reader. Multiplication of linear maps is not commutative. In other words, it is not necessarily true that $ST = TS$, even if both sides of the equation make sense.

Example 2.11

Suppose $D \in \mathcal{L}(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R}))$ is the differentiation map defined in Example 3.4 and $T \in \mathcal{L}(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R}))$ is the multiplication by x^2 map defined earlier in this section. Show that $TD \neq DT$.

Solution:

We have

$$((TD)p)(x) = x^2 p'(x) \quad \text{but} \quad ((DT)p)(x) = x^2 p'(x) + 2xp(x).$$

In other words, differentiating and then multiplying by x^2 is not the same as multiplying by x^2 and then differentiating.

§3 Operators on Inner Product Spaces

§3.1 Self-Adjoint and Normal Operators

Definition 3.1 (Adjoint, $T*$)

7.2 Definition adjoint, T^* Suppose $T \in \mathcal{L}(V, W)$. The adjoint of T is the function $T^* : W \to V$ such that

$$\langle Tv, w \rangle = \langle v, T^* w \rangle$$

for every $v \in V$ and every $w \in W$

§3.2 Polar Decomposition and Singular Value Decomposition

§3.2.1 Polar Decomposition

Notation 3.2 ($\sqrt{T}$). If T is a positive operator, then $\sqrt{T}$ denotes the unique postive square root of T.

Lemma 3.3 (Polar Decomposition)

Suppose $T \in \mathcal{L}(V)$. Then there exists an isometry $S \in \mathcal{L}(V)$ such that

$$T = S\sqrt{T^*T}.$$

Here, $\sqrt{T^*T}$ is the unique positive square root of the positive operator T^*T.

Proof. □

§3.2.2 Singular Value Decomposition

Definition 3.4 (*singular values*)

Suppose $T \in \mathcal{L}(V)$. The ***singular values*** of T are the eigenvalues of $\sqrt{T^*T}$, with each eigenvalue λ repeated $\dim E(\lambda, \sqrt{T^*T})$ times.

Note: The singular values are nonegative real numbers.
Alternatively, the singular values of T are the square roots of the diagonal entries when T^*T is put in diagonal form in an orthonormal basis.

Example 3.5

Let $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ be an operator on $\mathbf{R}^2$, then $T^*T = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. You should be able to check that the eigenvalues of T^*T are the positive numbers

$$\frac{3+\sqrt{5}}{2}, \frac{3-\sqrt{5}}{2}$$

and that the square roots of these eigenvalues are

$$\frac{\sqrt{5}+1}{2}, \frac{\sqrt{5}-1}{2}.$$

These are the singular values of T.

The next result shows that every operator on V has a clean description in terms of its singular values and two orthonormal bases of V.

Theorem 3.6 (Singular Value Decomposition)

Suppose $T \in \mathcal{L}(V)$ has singular values $s_1, \ldots, s_n$. Then there exist orthonormal abses $e_1, \ldots, e_n$ and $f_1, \ldots, f_n$ of V such that

$$Tv = s_1\langle v, e_1\rangle f_1 + \ldots + s_n\langle v, e_n\rangle f_n$$

for every $v \in V$.

Proof. By the Spectral lemma (real or complex), there is an orthonormal basis $e_1 \ldots, e_n$ of V such that $\sqrt{T^*T}(e_j) = s_j e_j$ for each j. Choose a polar decomposition

$$T = S\sqrt{T^*T}$$

for T (i.e., choose S). Then for each j,

$$Te_j = S(s_j e_j) = s_j f_j, \ f_j := Se_j.$$

Because S is an isometry, the list $f_1, \ldots, f_n$ is an orthonormal basis for V. (For more details, please refer to LADR p.237.) □

The following result has been mentioned above and it provides an alternative perspective on singular values.

Lemma 3.7 (Singular values without taking square root of an operator)

Suppose $T \in \mathcal{L}(V)$. Then the singular values of T are the nonegative square roots of the eigenvalues of T^*T, with each eigenvalue λ reapeated $\dim E(\lambda, T^*T)$ times.

Proof. The Spectral lemma implies that there are an orthonormal basis $e_1, \ldots, e_n$ and nonnegative numbers $\lambda_1, \ldots, \lambda_n$ such that $T^*Te_j = \lambda_j e_j$ for $j = 1, \ldots, n$. It is easy to see that $\sqrt{T^*T}e_j = \sqrt{\lambda_j}e_j$ for $j = 1, \ldots, n$, which implies the desired result. □

This concludes the section on polar decomposition and singular values.

§4 Operators on Complex Vector Spaces

§4.1 Gereralized Eigenvectors and Nilpotent Operators

§4.1.1 Null Spaces of Powers of an Operator

We begin this chapter with a study of null spaces of powers of an operator.

Lemma 4.1 (Sequence of increasing null spaces and decreasing range)
Suppose $T \in \mathcal{L}(V)$ and V is finite-dimensional. Then

$$\{0\} = \text{null}T^0 \subseteq \text{null}T^1 \subseteq \cdots \subseteq \text{null}T^k \subseteq \cdots \subseteq V$$

and

$$V = \text{range}T^0 \supseteq \text{range}T \supseteq \cdots \supseteq \text{range}T^k \supseteq \cdots \supseteq \{0\}.$$

Proof. Suppose k is a nonnegative integer and $v \in \text{null}T^k$. Then $T^k v = 0$, and hence

$$T^{k+1}v = T(T^k v) = T(0) = 0.$$

Thus $v \in \text{null}T^{k+1}$. Hence $\text{null}T^k \subseteq \text{null}T^{k+1}$, as desired and the remaining is left as an exercise for the reader. □

Note that the "links" between successive spaces cannot all be proper inclusions because V has a finite dimension. Thus, for instance, there is a j such that $\text{null}T^j = \text{null}T^{j+1}$, and clearly $j \leq \dim V$.

The following result states that if two consecutive terms in this sequence of subspaces are equal, then all later terms in the sequence are equal.

Lemma 4.2 (Equality in the sequence of null spaces)
Suppose $T \in \mathcal{L}(V)$. Suppose m is a nonnegative integer such that $\text{null}T^m = \text{null}T^{m+1}$. Then

$$\text{null}T^m = \text{null}T^{m+1} = \text{null}T^{m+2} = \text{null}T^{m+3} = \cdots.$$

Proof. Suppose $v \in \text{null}T^{j+2}$. Then

$$T^{j+1}(Tv) = 0,\ Tv \in \text{null}T^{j+1} = \text{null}T^j$$

and thus $T^j(Tv) = 0$, so $v \in \text{null}T^{j+1}$. □

One says that the spaces $\text{null}T^k$ *stabilize*, and in fact the stable subspace that we get is $\text{null}T^n$, where $n = \dim V$. Similarly, the spaces $\text{range}T^j$ coincide with $\text{range}T^n$ for $j \geq n$.

Lemma 4.3 (Null spaces stop growing)
Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then

$$\text{null}T^n = \text{null}T^{n+1} = \text{null}T^{n+2} = \ldots.$$

Proof. We can show this using the results from above. □

Lemma 4.4 (V is the direct sum of null$T^{\dim V}$ and range$T^{\dim V}$)
Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then

$$V = \text{null}T^n \oplus \text{range}T^n.$$

Proof. It's not too hard to see that the two subspaces have $\{0\}$ intersection. An element of rangeT^n is $T^n v$. If this element is annihilated by T^n, then $T^{2n}v = 0$ and then $T^n v = 0$ by the previous result(s).

Since the subspaces have complementary dimensions (by rank-nullity), their direct sum is the whole space V. □

Example 4.5

An example to keep in mind is the operator T on $\mathbf{F}^2$ defined by $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. The null space and range of T are $1-$dimensional; in fact, they're the *same* $1-$dimensional subspace of $\mathbf{F}^2$. The operator T^2 is 0, so its null space is all of $\mathbf{F}^2$ and its range is 0. The operator T itself is sort of dicey, but when we stabilize by taking $T^{\dim \mathbf{F}^2}$, we're in good shape.

§4.1.2 Generalized Eigenvectors

Unfortunately, some operators do not have enough eigenvectors to lead to a good description. Thus we introduce the concept of generalized eigenvectors, which will play a major role in our description of the structure of an operator.

To understand why we need more than eigenvectors, let's examine the question of describing an operator by decomposing its domain into invariant subspaces. Fix $T \in \mathcal{L}(V)$. We seek to describe T by finding a "nice" direct sum decomposition

$$V = U_1 \oplus \cdots \oplus U_m$$

where each U_j is a subspace of V invariant under T. The simplest possible nonzero invariant subspaces are 1-dimensional. A decomposition as above where each U_j is a 1 -dimensional subspace of V invariant under T is possible if and only if V has a basis consisting of eigenvectors of T. This happens if and only if V has an eigenspace decomposition

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$$

where $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of T. The Spectral lemma shows that if V is an inner product space, then a decomposition of the form holds for every normal operator if $\mathbf{F} = \mathbf{C}$ and for every self-adjoint operator if $\mathbf{F} = \mathbf{R}$ because operators of those types have enough eigenvectors to form a basis of V.

Then what about operators of other types? Sadly, the decomposition might not hold for more general operators, even on a complex vector space. Hence, we will need generalized eigenvectors and generalized eigenspaces, which we now introduce, to remedy this situation.

Definition 4.6 (generalized eigenvector)

A vector $v \in V$ is called a ***generalized eigenvector*** of T corrresponding to λ if $v \neq 0$ and

$$(T - \lambda I)^j v = 0$$

for some positive integer j.

Note: According to Axler, we don't define the concept of a generalized eigenvalue, because this would not lead to anything new. The reason would be if $(T - \lambda I)^j$ is not injective for some positive integer j, then $T - \lambda I$ is not injective, and hence λ is an eigenvalue of T.
Here, although j is allowed to be an arbitrary integer, we will soon prove that every generalized eigenvector satisifes this equation with $j = \dim V$.

Definition 4.7 (generalized eigenspace, $G(\lambda, T)$)

The ***generalized eigenspace*** of T corresponding to λ, denoted $G(\lambda, T)$, is defined to be the set of all generalized eigenvectors of T corresponding to λ, along with the 0 vector.

Because every eigenvector of T is a generalized eigenvector of T (take $j = 1$ in the definition of generalized eigenvector), each eigenspace is contained in the corresponding generalized eigenspace. In other words, if $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$, then

$$E(\lambda, T) \subset G(\lambda, T).$$

The next result implies that if $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$, then $G(\lambda, T)$ is a subspace of V (because the null space of each linear map on V is a subspace of V).

Lemma 4.8 (Description of generalized eigenspaces)

Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. Then

$$G(\lambda, T) = \operatorname{null}(T - \lambda I)^{\dim V}.$$

Proof. Suppose $v \in \operatorname{null}(T - \lambda I)^{\dim V}$. The definitions imply $v \in G(\lambda, T)$. Thus $G(\lambda, T) \supset \operatorname{null}\,(T - \lambda I)^{\dim V}$.
Conversely, suppose $v \in G(\lambda, T)$. Thus there is a positive integer j such that

$$v \in \operatorname{null}(T - \lambda I)^j.$$

Using the lemmas above, we get $v \in \operatorname{null}(T - \lambda I)^{\dim V}$. Thus $G(\lambda, T) \subset \operatorname{null}(T - \lambda I)^{\dim} V$, as desired. □

The takeaway here is that we now see that if λ is an eigenvalue for $T \in \mathcal{L}(V)$, then

$$V = G(\lambda, T) \oplus \operatorname{range}(T - \lambda I)^n.$$

Further, the subspace $\operatorname{range}(T - \lambda I)^n$ of V is visibly T-invariant. The displayed decomposition helps us to achieve our aims of analyzing and understanding T on V (whatever that means). The space $\operatorname{range}(T - \lambda I)^n$ is smaller than V, while the generalized eigenspace is specific to λ and might be tractable.

Example 4.9

From the examples $T = 0$ and $T = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ on $\mathbf{F}^2$, we see that both have $V = G(0, T)$. In the first case, T is 0(duh!); in the second case, T is nonzero but ***nilpotent***(a concept that will be covered soon).

We learned earlier that eigenvectors corresponding to distinct eigenvalues are linearly independent. Now we prove a similar result for generalized eigenvectors.

Lemma 4.10 (Linear independence of generalized eigenvectors)

Let $T \in \mathcal{L}(V)$. Suppose that $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T and $v_1, \ldots, v_m$ are corresponding generalized eigenvectors. Then $v_1, \ldots, v_m$ is linearly independent.

Proof. Suppose $a_1, \ldots, a_m$ are complex numbers such that

$$0 = a_1 v_1 + \cdots + a_m v_m \tag{1}$$

Let k be the largest nonnegative integer such that $(T - \lambda_1 I)^k v_1 \neq 0$. Let

$$w = (T - \lambda_1 I)^k v_1$$

Thus

$$(T - \lambda_1 I) w = (T - \lambda_1 I)^{k+1} w = 0$$

and hence $Tw = \lambda_1 w$. Thus $(T - \lambda I)w = (\lambda_1 - \lambda) w$ for every $\lambda \in \mathbf{F}$ and hence

$$(T - \lambda I)^n w = (\lambda_1 - \lambda)^n w \tag{2}$$

for every $\lambda \in \mathbf{F}$, where $n = \dim V$. Apply the operator

$$(T - \lambda_1 I)^k (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n$$

to both sides of (1) getting

$$\begin{aligned} 0 &= a_1 (T - \lambda_1 I)^k (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n v_1 \\ &= a_1 (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n w \\ &= a_1 (\lambda_1 - \lambda_2)^n \cdots (\lambda_1 - \lambda_m)^n w. \end{aligned}$$

The equation above implies that $a_1 = 0$. In a similar fashion, $a_j = 0$ for each j, which implies that $v_1, \ldots, v_m$ is linearly independent. □

§4.1.3 Nilpotent Operator

Now we introduce the concept of nilpotent operator, which we have mentioned above.

Definition 4.11

An operator is called ***nilpotent*** if some power of it equals 0.

*The Latin word **nil** means nothing or zero; the Latin word **potent** means power. Thus **nilpotent** literally means zero power!*

Keep in mind that we would never need to use a power higher than the dimension of the sapce and we will show why.

Lemma 4.12 (Nilpotent operator raised to dimension of domain is 0)
Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then

$$N^{\dim V} = 0.$$

Proof. The proof is simple. Because N is nilpotent, $G(0, N) = V$, which implies null $N^{\dim V} = V$, as desired. □

Given an operator T on V, we want to find a basis of V such that the matrix of T with respect to this basis is as simple as possible, meaning that the matrix contains many 0's.

Lemma 4.13 (Matrix of a nilpotent operator)
If N is ***nilpotent***, it is strictly upper-triangular in some basis of V.

The phrase "strictly upper-triangular" means upper-triangular with 0s along the diagonal. We can also rephrase being upper-triangular as follows:

There is a basis $v_1, \ldots, v_n$ of V so that $Nv_j \in span(v_1, \ldots, v_{j-1})$ for $j = 1, \ldots, n$. For $j = 1$, this means that Nv_1 is in the span of the empty list, which is $\{0\}$, and thus $Nv_1 = 0$.

Proof. The proof might be kind of wordy, but bear with it. First choose a basis of null N. Then extend this to a basis of null N^2. Then extend to a basis of null N^3. Continue in this fashion, eventually getting a basis of V (because 8.18 states that null $N^{\dim V} = V$).

Now let's think about the matrix of N with respect to this basis. The first column, and perhaps additional columns at the beginning, consists of all 0's, because the corresponding basis vectors are in null N. The next set of columns comes from basis vectors in null N^2. Applying N to any such vector, we get a vector in null N; in other words, we get a vector that is a linear combination of the previous basis vectors. Thus all nonzero entries in these columns lie above the diagonal. The next set of columns comes from basis vectors in null N^3. Applying N to any such vector, we get a vector in null N^2; in other words, we get a vector that is a linear combination of the previous basis vectors. Thus once again, all nonzero entries in these columns lie above the diagonal. Continue in this fashion to complete the proof. □

§4.2 Decomposition of an Operator

§4.2.1 Description of Operators on Complex Vector Spaces

We saw earlier that the domain of an operator might not decompose into eigenspaces, even on a finite-dimensional complex vector space. In this section we will see that every operator on a finite-dimensional complex vector space has enough generalized eigenvectors to provide a decomposition.

We observed earlier that if $T \in \mathcal{L}(V)$, then null T and range T are invariant under T (see LADR 5.3, parts (c) and (d)). Now we show that the null space and the range of each polynomial of T is also invariant under T.

Lemma 4.14 (The null space and range of $p(T)$ are invariant under T)
Suppose $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbf{F})$. Then null$p(T)$ and range$p(T)$ are invariant under T.

Proof. Suppose $v \in$ null $p(T)$. Then $p(T)v = 0$. Thus

$$((p(T))(Tv) = T(p(T)v) = T(0) = 0$$

Hence $Tv \in$ null $p(T)$. Thus null $p(T)$ is invariant under T, as desired. Suppose $v \in$ range $p(T)$. Then there exists $u \in V$ such that $v = p(T)u$. Thus

$$Tv = T(p(T)u) = p(T)(Tu).$$

Hence $Tv \in$ range $p(T)$. Thus range $p(T)$ is invariant under T, as desired. □

The following major result shows that every operator on a complex vector space can be thought of as composed of pieces, each of which is a nilpotent operator plus a scalar multiple of the identity.

Lemma 4.15 (Description of operators on complex vector spaces)
Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. Then

(a) $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$;

(b) each $G(\lambda_j, T)$ is invariant under T;

(c) each $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent.

(The following statements assume that the reader has read the results below, if you haven't, please do so before proceeding.)
We saw previously that

$$V = G(\lambda, T) \oplus U$$

where $U :=$ range$(T - \lambda I)^n$ of V is a T -invariant subspace of V We have proved that

$$G(\mu, T) \subseteq U$$

if μ is another eigenvalue of T.

Proof. Again, we will be using some results below, it would be better to read them first before looking at this proof. For part(a), the proof is by induction on dim V. If $V = \{0\}$, there are no eigenvalues and V is the empty direct sum (which is $\{0\}$ by convention). If V is 1 -dimensional, it's $E(\lambda, T) = G(\lambda, T)$ for the unique eigenvalue λ. Assume that V has dimension > 1 and that the lemma is known for spaces of dimension less than that of V. Then V has an eigenvalue λ, and $V = G(\lambda, T) \oplus U$ where $U = \operatorname{range}(T - \lambda l)^n$. The space U is T-invariant and has dimension $< \dim V$. The statement to be proved is known for U.

So far, we have

$$V = G(\lambda, T) \oplus U, \quad U = \bigoplus_{\mu} G(\mu, T_{|U}).$$

where $T_{|U}$ is the restriction of T to U. The second decomposition is true because dim $U < \dim V$ and because of the inductive assumption. Because $G(\lambda, T) \cap U = \{0\}, \lambda$ is not an eigenvalue of $T_{|U}$ Further, since $G(\mu, T) \subseteq U$, for each eigenvalue μ of T other than λ, the eigenvalues of $T_{|U}$ on U are the eigenvalues $\mu \neq \lambda$ of T. For each such μ, the inclusion $G(\mu, T) \subseteq U$ implies that $G(\mu, T_{|U}) = G(\mu, T)$ Hence we have the desired decomposition of V. Please refer to LADR for the remaining portion. □

If $v \in V$, then there is a unique way to write

$$v = v_1 + \ldots + v_m$$

with $v_j = G(\lambda_j, T)$ for each j. Each summand is either a generalized eigenvector or is 0. After removing those term that are 0, we've written v as a sum(possibly an empty sum!) of generalized eigenvectors.

Hence, if we knew that

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$$

we would learn that V is annihilated by the operator

$$(T - \lambda_1 l)^n (T - \lambda_2 l)^n \cdots (T - \lambda_m l)^n$$

because the j th factor of this product annihilates the j th summand in the direct sum decomposition.

Proposition 4.16

We do in fact know that the operator

$$(T - \lambda_1 I)^n (T - \lambda_2 I)^n \ldots (T - \lambda_m I)^n = 0$$

on V.

Here's why:

We poved long ago that T is upper-triangular in some basis(because $\mathbf{F} = \mathbf{C}$). If the numbers on the diagonal are $a_1, \ldots, a_n$, then the operator

$$(T - a_1 I) \ldots (T - a_n I) = 0$$

on V as we can see by direct computation. Further, we know that complex numbers that appear as diagonal entries are precisely the eigenvalues of T(5.32 on page 152). The number of times a give λ_j appears on the digaonal is between 1 and n. The polynomial $(z - a_1) \ldots (z - a_n)$ is certainly a divisor of $(z - \lambda_1)^n \ldots (z - \lambda_m)^n$. It follows that the operator $(T - \lambda_1 I)^n \ldots (T - \lambda_m I)^n = 0$ on V.

Proposition 4.17

Suppose that λ and μ are distinct eigenvalues of T. Then

$$G(\mu, T) \subseteq \operatorname{range}(T - \lambda I)^n.$$

Proof. Professor Ribet's proof on this might be slightly annoying, but here it is. Write $(\lambda - \mu)I = (T - \mu l) - (T - \lambda l)$ and use the formula to compute $((\lambda - \mu)l)^{2n}$ by the binomial lemma. The result can be expressed in the form

$$(\lambda - \mu)^{2n} I = f(T)(T - \mu I)^n + g(T)(T - \lambda I)^n;$$

here f and g are polynomials. □

Example 4.18

Suppose $n = 3$. Then $(\lambda - \mu)^6 I$ is a sum of seven terms

$$(T - \mu I)^6 - 6(T - \mu I)^5(T - \lambda l) + 10(T - \mu I)^4(T - \lambda I)^2 + \cdots$$

each having either a $(T - \mu I)^3$ or a $(T - \mu I)^3$ that can be factored out.

Example 4.19

Suppose now that $v \in G(\mu, T) = \operatorname{null}(T - \mu I)^n$. Then

$$(\lambda - \mu)^{2n} v = p(T)(T - \mu I)^n v + q(T)(T - \lambda I)^n v = (T - \lambda I)^n q(T) v$$

Thus

$$v = (T - \lambda I)^n \left(\frac{1}{(\lambda - \mu)^{2n}} q(T) v \right)$$

is in the range of $(T - \lambda I)^n$.

As we know, an operator on a complex vector space may not have enough eigenvectors to form a basis of the domain. The next result shows that on a complex vector space there are enough generalized eigenvectors to do this.

Lemma 4.20 (A basis of generalized eigenvectors)

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then there is a basis of V consisting of generalized eigenvectors of T.

Proof. Choose a basis of each $G(\lambda_j, T)$ in 8.21($LADR$). Put all these bases together to form a basis of V consisting of generalized eigenvectors of T. □

§4.2.2 Multiplicity of an Eigenvalue

We will now study the dimensions of the subspaces involved in the decomposition of V.

Definition 4.21 (multiplicity)

- Suppose $T \in \mathcal{L}(V)$. The multiplicity of an eigenvalue λ of T is defined to be the dimension of the corresponding generalized eigenspace $G(\lambda, T)$.
- In other words, the multiplicity of an eigenvalue λ of T equals $\dim \operatorname{null}(T - \lambda I)^{\dim V}$.

Lemma 4.22 (Sum of the multiplicities equals $\dim$ V)
Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then

$$\dim V = \sum_{\lambda} \text{multiplicity of } \lambda.$$

with the sum over the set of eigenvalues of T.

Proof. The desired result follows from 8.21 and the obvious formula for the dimension of a direct sum (see 3.78 or Exercise 16 in Section 2.C). □

In case of confusions, note that the terms ***algebraic multiplicity*** and ***geometric multiplicity*** are used differently. If $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T, then

$$\text{algebraic multiplicity of } \lambda = \dim \operatorname{null}(T - \lambda I)^{\dim V} = \dim G(\lambda, T),$$

$$\text{geometric multiplicity of } \lambda = \operatorname{dimnull}(T - \lambda I) = \dim E(\lambda, T).$$

Now here are some challenges:

Question 4.23. *If U is a T-invariant subspace of V, is it true that the multiplicity of λ for T on V is the sum of the multiplicity of λ for $T_{|U}$ (an operator on U) and the multiplicity of λ for T/U (an operator on V/U)?*

If T is upper-triangular in some basis, with diagonal entries $a_1, \dots, a_n$, is the multiplicity of λ the number of entries a_j that are equal to λ? (Hey, this statement is the last exercise of 8.C.!)

For those who are not interested, you may skip this section and the following propositions. Recall that $U \subset V$ is assumed to be T-invariant, and we'd like to consider a difference $T - \lambda I$. Let's call this difference N, so that (for example) $G(\lambda, T) = \text{null } N^n$.

Proposition 4.24

$$\operatorname{null}(N_U)^n = (\operatorname{null} N^n) \cap U$$

This is clear: a vector in U is sent to 0 by N^n if and only if it is sent to 0 by the restriction of N to U. Here's another observation:

Proposition 4.25

$$\operatorname{range}(N_U)^n = (\operatorname{range} N^n) \cap U.$$

Note

$$V = \operatorname{null} N^n \oplus \operatorname{range} N^n$$

and

$$U = \operatorname{null}(N_U)^n \oplus \operatorname{range}(N_U)^n .$$

If $u \in U$ is $0 + x$ relative to the first direct sum decomposition and $w + z$ relative to the second one, then $w = 0$ and $z = x$ by the uniqueness of the writing of u as a sum of elements of null N^n and range N^n.

Remark: In the second decomposition, we would normally replace n by the dimension of U, which in general is smaller than n. However, null spaces stop growing (8.4) and ranges corresponding stop shrinking (e.g., by the rank-nullity formula).

§4.2.3 Block Diagonal Matrices

Definition 4.26 (block diagonal matrix)

A ***block diagonal matrix*** is a square matrix of the form

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

where $A_1, \ldots, A_m$ are square matrices lying along the diagonal and all the other entries of the matrix equal 0.

Lemma 4.27 (Block diagonal matrix with upper-triangular blocks)

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T, with multiplicities $d_1, \ldots, d_m$. Then there is a basis of V with respect to which T has a block diagonal matrix of the form

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

where each A_j is a d_j -by- d_j upper-triangular matrix of the form

$$A_j = \begin{pmatrix} \lambda_j & & * \\ & \ddots & \\ 0 & & \lambda_j \end{pmatrix}.$$

Proof. Each $(T - \lambda_j I)_{G(\lambda_j, T)}$ is nilpotent. For each j, choose a basis of $G(\lambda_j, T)$, which is a vector space with dimension d_j, such that the matrix of $(T - \lambda_j I)|_{G(\lambda_j, T)}$ with respect to this basis is as in 8.19. Thus the matrix of $T|_{G(\lambda_j, T)}$, which equals $(T - \lambda_j I)|_{G(\lambda_j, T)} + \lambda_j I|_{G(\lambda_j, T)}$, with respect to this basis will look like the desired form shown above for A_j. Putting the bases of the $G(\lambda_j, T)$'s together gives a basis of $V$by 8.21(a) The matrix of T with respect to this basis has the desired form. □

§4.2.4 Square Roots

Recall that a square root of an operator $T \in \mathcal{L}(V)$ is an operator $R \in \mathcal{L}(V)$ such that $R^2 = T$ (see 7.33). Every complex number has a square root, but not every operator on a complex vector space has a square root. The noninvertibility of that operator is no accident, as we will soon see. We begin by showing that the identity plus any nilpotent operator has a square root.

> **Lemma 4.28** (Identity plus nilpotent has a square root)
> Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then $I + N$ has a square root.

Proof. Consider the Taylor series for the function $\sqrt{1+x}$:

$$\sqrt{1+x} = 1 + a_1 x + a_2 x^2 + \cdots$$

We will not find an explicit formula for the coefficients or worry about whether the infinite sum converges because we will use this equation only as motivation.

Because N is nilpotent, $N^m = 0$ for some positive integer m. In 8.32 suppose we replace x with N and 1 with I. Then the infinite sum on the right side becomes a finite sum (because $N^j = 0$ for all $j \geq m$). In other words, we guess that there is a square root of $I + N$ of the form

$$I + a_1 N + a_2 N^2 + \cdots + a_{m-1} N^{m-1}$$

Having made this guess, we can try to choose $a_1, a_2, \ldots, a_{m-1}$ such that the operator above has its square equal to $I + N$. Now

$$\begin{aligned}
&\left(I + a_1 N + a_2 N^2 + a_3 N^3 + \cdots + a_{m-1} N^{m-1}\right)^2 \\
&\quad = I + 2a_1 N + \left(2a_2 + a_1^2\right) N^2 + \left(2a_3 + 2a_1 a_2\right) N^3 + \cdots \\
&\quad + \left(2a_{m-1} + \text{ terms involving } a_1, \ldots, a_{m-2}\right) N^{m-1}
\end{aligned}$$

We want the right side of the equation above to equal $I + N$. Hence choose a_1 such that $2a_1 = 1$ (thus $a_1 = 1/2$). Next, choose a_2 such that $2a_2 + a_1^2 = 0$ (thus $a_2 = -1/8$). Then choose a_3 such that the coefficient of N^3 on the right side of the equation above equals 0 (thus $a_3 = 1/16$). Continue in this fashion for $j = 4, \ldots, m-1$, at each step solving for a_j so that the coefficient of N^j on the right side of the equation above equals 0. Actually we do not care about the explicit formula for the a_j 's. We need only know that some choice of the a_j 's gives a square root of $I + N$ □

Note: This lemma is valid on real and complex vector spaces.

Lemma 4.29 (Over $\mathbf{C}$, invertible operators have square roots)
Suppose V is a complex vector space and $T \in \mathcal{L}(V)$ is invertible. Then T has a square root.

Proof. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. For each j, there exists a nilpotent operator $N_j \in \mathcal{L}\left(G\left(\lambda_j, T\right)\right)$ such that $\left.T\right|_{G\left(\lambda_j, T\right)} = \lambda_j I + N_j$ (see 8.21(c)). Because T is invertible, none of the λ_j 's equals 0, so we can write

$$\left.T\right|_{G\left(\lambda_j, T\right)} = \lambda_j\left(I + \frac{N_j}{\lambda_j}\right)$$

for each j. Clearly N_j / λ_j is nilpotent, and so $I + N_j / \lambda_j$ has a square root (by previous lemma). Multiplying a square root of the complex number λ_j by a square root of $I + N_j / \lambda_j$, we obtain a square root R_j of $\left.T\right|_{G\left(\lambda_j, T\right)}$ A typical vector $v \in V$ can be written uniquely in the form

$$v = u_1 + \cdots + u_m$$

where each u_j is in $G\left(\lambda_j, T\right)$ (see 8.21). Using this decomposition, define an operator $R \in \mathcal{L}(V)$ by

$$R v = R_1 u_1 + \cdots + R_m u_m$$

The reader should verify that this operator R is a square root of T, completing the proof. □

By imitating the techniques in this section, the reader should be able to prove that if V is a complex vector space and $T \in \mathcal{L}(V)$ is invertible, then T has a k^{th} root for every positive integer k.

§4.3 Characteristic and Minimal Polynomials

§4.3.1 The Cayley-Hamilton Theorem

Definition 4.30 (characteristic polnomial)
Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T, with multiplicities $d_1, \ldots, d_m$. The polynomial

$$(z - \lambda_1)^{d_1} \cdots (z - \lambda_m)^{d_m}$$

is called the ***characteristic polynomial*** of T.

Lemma 4.31 (Degree and zeros of characteristic polynomial)
Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then

(a) the characteristic polynomial of T has degree $\dim V$;

(b) the zeros of the characteristic polynomial of T are the eigenvalues of T.

Proof. □

Theorem 4.32 (Cayley-Hamilton Theorem)
Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let q denote the characteristic polynomial of T. Then $q(T) = 0$

Proof. □

§5 References