18.700 - LINEAR ALGEBRA, DAY 12
UPPER TRIANGULAR MATRICES AND DIAGONALIZATION

SAM SCHIAVONE

CONTENTS

I. PRE-CLASS PLANNING

I.1. Goals for lesson.

(1) Students will learn that the roots of the minimal polynomial are exactly the eigenvalues.
(2) Students will learn how to compute the eigenvalues and eigenvectors of a linear operator.
(3) Students will learn that the eigenvalues of an upper triangular matrix are the diagonal entries.
(4) Students will learn the definition of eigenspace.
(5) Students will learn criteria for diagonalizability.

I.2. Methods of assessment.

(1) Student responses to questions posed during lecture
(2) Student responses to worksheet

I.3. Materials to bring.
(1) Laptop + adapter (2) Worksheets

(0:00)

II. Lesson Plan

Announcements: • Exam grades posted. Median 77, Mean 74. [Show distribution, if possible.]

II.1. **Last time.**

- Proved some properties of eigenvalues.
- Defined $p(T)$ where p is a polynomial and T is a linear operator.
- Showed the existence of the minimal polynomial.
- Proved that the roots of $\operatorname{minpoly}(T)$ are eigenvalues of T.

II.2. **The Minimal Polynomial, continued.**

Corollary 1. *Let V be a nonzero finite-dimensional $\mathbb{C}$-vector space and $T \in \mathcal{L}(V)$. Then T has an eigenvalue.*

Proof. Let $m := \operatorname{minpoly}(T)$. Note that m is nonconstant: if $m = c$ were constant, then we would have $cI = 0$, contradicting the fact that $V \neq 0$.

By the Fundamental Theorem of Algebra, there exists $\lambda \in \mathbb{C}$ such that $m(\lambda) = 0$. Then

$$m(z) = (z - \lambda)q(z)$$

for some monic $q \in \mathcal{P}(\mathbb{C})$. Then

$$0 = m(T) = (T - \lambda I)q(T)\,.$$

Since $\deg(q) < \deg(m)$ and m is the minpoly, then $q(T) \neq 0$. Then there is some vector $v \in V$ such that $q(T)(v) \neq 0$. Then

$$0 = m(T)(v) = (T - \lambda I)(q(T)(v))$$

so $q(T)(v)$ is an eigenvector of T with eigenvalue λ. □

Remark 2. Here we used the fact that $\mathbb{C}$ is algebraically closed in an important way. The result is not true over $\mathbb{R}$!

Corollary 3. *With notation as above, the eigenvalues of T are exactly the roots of* $\operatorname{minpoly}(T)$.

Proof. We have seen that all the roots of $m := \operatorname{minpoly}(T)$ are eigenvalues of T. Conversely, suppose $\lambda \in \mathbb{F}$ is an eigenvalue of T. Then there exists $0 \neq v \in V$ such that $T(v) = \lambda v$. Applying T to both sides repeatedly, we see that $T^k(v) = \lambda^k v$ for all $k \in \mathbb{Z}_{\geq 0}$. Taking linear combinations of these monomials, we have [write "0 = ..." last]

$$0 = m(T)v = m(\lambda)v\,.$$

Since $v \neq 0$, then $m(\lambda) = 0$. □

Q: Given a linear operator T, how can we compute its eigenvalues and eigenvectors?
A:

(1) To compute $\operatorname{minpoly}(T)$, we need to find the smallest d such that

$$c_0 I + c_1 T + \cdots + c_{d-1}T^{d-1} = -T^d$$

has a solution for $c_0, \ldots, c_{d-1} \in \mathbb{F}$. We can choose a basis $\mathcal{B}$ for V and apply $[\cdot]_{\mathcal{B}}$ to the above equation. This produces a matrix equation which can be thought of as a linear system of $(\dim(V))^2$ equations in d unknowns.

This yields the following algorithm: for each $d = 1, 2, \ldots$, see if the above system of equations has a solution. By the theorem, this algorithm terminates at the latest when $d = \dim(V)$.

(2) Usually faster, but not guaranteed to always work: choose $v \in V, v \neq 0$ and consider the equation

$$c_0 v + c_1 T(v) + \cdots + c_{n-1} T^{n-1}(v) = -T^n(v)$$

where $n := \dim(V)$. Again, by choosing a basis for V and applying $[\cdot]_{\mathcal{B}}$, we obtain a system of n equations in the n unknowns $c_0, \ldots, c_{n-1}$. If the solution to this system is unique, this yields the coefficients of $\operatorname{minpoly}(T)$.

(3) Choose a basis $\mathcal{B}$ and let $A = [T]_{\mathcal{B}}$. Compute $\ker(A - \lambda I)$, treating λ as a variable.

Proposition 4. *Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$ and $q \in \mathcal{P}(\mathbb{F})$. Then $q(T) = 0$ iff* $\operatorname{minpoly}(T)$ *divides q, i.e., $q = \operatorname{minpoly}(T) f$ for some $f \in \mathcal{P}(\mathbb{F})$.*

Proof idea. Use the division algorithm to divide q by $\operatorname{minpoly}(T)$ and consider the remainder. □

Corollary 5. *With the same assumptions, suppose U is a T-invariant subspace of V. Then* $\operatorname{minpoly}(T|_U)$ *divides* $\operatorname{minpoly}(T)$.

Corollary 6. *With the same assumptions, T is not invertible iff the constant term of* $\operatorname{minpoly}(T)$ *is* 0.

Proof. Let $m := \operatorname{minpoly}(T)$. Then

$$\begin{aligned} T \text{ is not invertible} &\iff 0 \text{ is an eigenvalue of } T \\ &\iff 0 \text{ is a zero of } m \\ &\iff m(0) = 0 \\ &\iff \text{the constant term of } m \text{ is } 0. \end{aligned}$$

□

II.3. **Worksheet.**

II.4. **5C Upper triangular matrices.** Let $T : V \to W$ be a linear map and $\mathcal{B} = (v_1, \ldots, v_n)$ and $\mathcal{C} = (w_1, \ldots, w_m)$ be bases of V and W, respectively. Recall that

$$_{\mathcal{C}}[T]_{\mathcal{B}} = \begin{pmatrix} | & & | \\ [T(v_1)]_{\mathcal{C}} & \cdots & [T(v_n)]_{\mathcal{C}} \\ | & & | \end{pmatrix}.$$

When $V = W$, so T is a linear operator, then its matrix is [ask students] square.

Goal: Find a basis of V such that $[T]_{\mathcal{B}} = {}_{\mathcal{B}}[T]_{\mathcal{B}}$ is particularly simple. Suppose V is a $\mathbb{C}$-vector space. As we saw, then T has an eigenvalue λ; let $v \neq 0$ be a corresponding eigenvector, so $T(v) = \lambda v$. If we take v to be the first element in a basis for $\mathcal{B}$ of V, then

$$[T]_{\mathcal{B}} = \begin{pmatrix} \lambda & & \\ 0 & * & \\ \vdots & & \\ 0 & & \end{pmatrix}.$$

(Here the $*$ indicate that we don't know the other entries.)

Definition 7. A square matrix is *upper triangular* if all entries below the diagonal are 0.

Example 8. [Give examples of diagonal and not diagonal matrices.]

Proposition 9. *Suppose $T \in \mathcal{L}(V)$ and $\mathcal{B} := (v_1, \dots, v_n)$ is a basis of V. TFAE.*

(i) $[T]_{\mathcal{B}}$ *is upper triangular.*
(ii) $T(v_k) \in \operatorname{span}(v_1, \dots, v_k)$ *for each* $k = 1, \dots, n$.
(iii) $\operatorname{span}(v_1, \dots, v_k)$ *is* T*-invariant for each* $k = 1, \dots, n$.

Proof sketch. (i) $\implies$ (ii): Recall that the k^{th} column of $[T]_{\mathcal{B}}$ is the coordinate vector $[T(v_k)]_{\mathcal{B}}$. Write

$$[T(v_k)]_{\mathcal{B}} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

Since $[T]_{\mathcal{B}}$ is upper triangular, then $0 = a_{k+1} = \cdots = a_n$, so

$$T(v_k) = a_1 v_1 + \cdots + a_k v_k \in \operatorname{span}(v_1, \dots, v_k).$$

(ii) $\implies$ (i): Similar.
(ii) $\iff$ (iii): Exercise. □

Proposition 10. *Suppose $T \in \mathcal{L}(V)$ and V has a basis $\mathcal{B}$ with respect to which $[T]_{\mathcal{B}}$ is upper triangular with diagonal entries $\lambda_1, \dots, \lambda_n$. Then*

(i)

$$(T - \lambda_1 I) \cdots (T - \lambda_n I) = 0.$$

(ii) The eigenvalues of T are exactly $\lambda_1, \dots, \lambda_n$.

Proof. (i) Claim: $(T - \lambda_1 I) \cdots (T - \lambda_k I) v_j = 0$ for all $j \leq k$, for all $k = 1, \dots, n$. By induction on k.

Base case: $k = 1$. Since $T(v_1) = \lambda_1 v_1$, then $(T - \lambda_1 I) v_1 = 0$.

Inductive step: Let $k \geq 2$ and assume the result holds for $k - 1$. Fix $j \in \{1, \dots, k\}$.

Case 1: $j < k$. Then $j \leq k - 1$. Note that the $T - \lambda_i I$ commute with each other since they are polynomials in T. Then

$$\begin{aligned}(T - \lambda_1 I) \cdots (T - \lambda_k I) v_j &= (T - \lambda_k) \overbrace{(T - \lambda_1 I) \cdots (T - \lambda_{k-1} I) v_j}^{0 \text{ by inductive hypothesis}} \\ &= (T - \lambda_k)(0) = 0\end{aligned}$$

Case 2: $j = k$. Since $[T]_{\mathcal{B}}$ is upper triangular, then

$$T(v_k) = a_1 v_1 + \cdots + a_{k-1} v_{k-1} + \lambda_k v_k$$

for some $a_1, \dots, a_{k-1} \in \mathbb{F}$. Then

$$(T - \lambda_k I)(v_k) = a_1 v_1 + \cdots + a_{k-1} v_{k-1} + \underbrace{\lambda_k v_k - \lambda_k v_k}_{\to 0}.$$

Then

$$\begin{aligned}(T - \lambda_1 I) \cdots (T - \lambda_{k-1} I)(T - \lambda_k I)(v_k) &= (T - \lambda_1 I) \cdots (T - \lambda_{k-1} I)(a_1 v_1 + \cdots + a_{k-1} v_{k-1}) \\ &= a_1 \cdot 0 + \cdots + a_{k-1} \cdot 0 = 0\end{aligned}$$

by the inductive hypothesis.

(ii) Recall that λ is an eigenvalue of T iff $T - \lambda I$ is not invertible. Since $T(v_1) = \lambda_1 v_1$, then λ_1 is an eigenvalue.

Given $k \in \{2, \ldots, n\}$, then $(T - \lambda_k I)v_k \in \operatorname{span}(v_1, \ldots, v_{k-1})$ by the above. Thus $T - \lambda_k I$ maps $\operatorname{span}(v_1, \ldots, v_k)$ into $\operatorname{span}(v_1, \ldots, v_{k-1})$. Since

$$\dim(\operatorname{span}(v_1, \ldots, v_k)) = k \quad \text{and} \quad \dim(\operatorname{span}(v_1, \ldots, v_{k-1})) = k - 1$$

then $T - \lambda_k I$ is not injective by a previous result (consequence of Rank-Nullity). Thus λ_k is an eigenvalue of T.

□

Proposition 11. *Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then V has a basis $\mathcal{B}$ with respect to which $[T]_{\mathcal{B}}$ is upper triangular iff* minpoly(T) *splits into degree 1 factors, i.e.,*

$$\operatorname{minpoly}(T)(z) = (z - \lambda_1) \cdots (z - \lambda_m)$$

for some $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$.

Proof sketch. ($\Rightarrow$): Suppose T has an upper triangular matrix with respect to some basis of V. Denote the diagonal entries of this matrix by $\alpha_1, \ldots, \alpha_n$. Letting

$$q(z) = (z - \alpha_1) \cdots (z - \alpha_n),$$

then $q(T)$ by a previous result. Then minpoly(T) divides q, so minpoly also splits into degree 1 factors.

($\Leftarrow$): Suppose $\operatorname{minpoly}(T)(z) = (z - \lambda_1) \cdots (z - \lambda_m)$ for some $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$. If $m = 1$, done. Otherwise, let $U = \operatorname{img}(T - \lambda_m I)$. Then U is T-invariant, so we can consider the restriction $T|_U$. Apply the inductive hypothesis, extend the basis, and prove that this results in an upper triangular matrix. □

Corollary 12. *Let V be a finite dimensional $\mathbb{C}$-vector space and $T \in \mathcal{L}(V)$. Then there exists a basis $\mathcal{B}$ of V such that T has an upper triangular matrix with respect to $\mathcal{B}$.*

Proof. By the Fundamental Theorem of Algebra, nonconstant polynomial over $\mathbb{C}$ splits into degree 1 factors, so this is true of minpoly(T) in particular. □

II.5. **Diagonalizable Operators.** Say we have linear operators $S, T \in \mathcal{L}(V)$ and we want to compute their composition ST with respect to some choice of basis. In general, matrix multiplication is an expensive operation: naively, it requires n^3 operations, where $n = \dim(V)$. But if we can cleverly choose a basis of V that makes it so many of the entries of $[S]$ and $[T]$ are 0, then this will make this computation faster.

Definition 13. A *diagonal matrix* is a square matrix all of whose off-diagonal entries are 0. That is, A is diagonal if $A_{ij} = 0$ when $i \neq j$.

Example 14 (Give example, one where 0 is one of the diagonal entries.)**.**

Proposition 15. *Let V be a finite-dimensional vector space and $T \in \mathcal{L}(V)$. If there exists a basis $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}$ is diagonal, then the eigenvalues of T are precisely the diagonal entries of $[T]_{\mathcal{B}}$.*

Proof. Diagonal matrices are upper triangular, so follows from a previous result. □

Definition 16. An operator $T \in \mathcal{L}(V)$ is *diagonalizable* if there exists a basis $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}$ is diagonal. Similarly, we say that a square matrix A is diagonalizable if the linear map $L_A : \mathbb{F}^n \to \mathbb{F}^n$ is diagonalizable.

Remark 17. Diagonalizable $\neq$ diagonal!

Example 18. Define

$$\begin{aligned} T : \mathbb{F}^2 &\to \mathbb{F}^2 \\ v &\mapsto Av \end{aligned}$$

where

$$A = \begin{pmatrix} -14 & 9 \\ -30 & 19 \end{pmatrix}.$$

Then A is not diagonal. However, with respect to the basis $\mathcal{B}$

$$v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \; v_2 = \begin{pmatrix} 3 \\ 5 \end{pmatrix},$$

we have

$$[T]_{\mathcal{B}} = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}.$$

Thus T (and A) is diagonalizable.

Remark 19. Note that if v is an eigenvector of T, then so is cv for all $0 \neq c \in \mathbb{F}$:

$$T(cv) = cT(v) = c\lambda v = \lambda(cv).$$

Definition 20. Let $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. The *eigenspace* of T corresponding to λ is the subspace

$$E(\lambda) := E(\lambda, T) := \ker(T - \lambda I) = \{v \in V : T(v) = \lambda v\}.$$

So $E(\lambda, T)$ is the set of all eigenvectors of T corresponding to λ, along with the 0 vector.

Remark 21. λ is an eigenvalue of T iff $E(\lambda, T) \neq \{0\}$.

Theorem 22 (Sum of eigenspaces is direct)**.** *Let $T \in \mathcal{L}(V)$ and suppose $\lambda_1, \dots, \lambda_m$ are distinct eigenvaues of T. Then*

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

is a direct sum. Moreover, if V is finite-dimensional, then

$$\dim(E(\lambda_1, T)) + \cdots + \dim(E(\lambda_m, T)) \leq \dim(V).$$

Proof. Suppose $v_1 + \cdots + v_m = 0$ where $v_k \in E(\lambda_k)$ for all $k = 1, \dots, m$. Since eigenvectors corresponding to distinct eigenvalues are linearly independent, then $v_k = 0$ for all k. (Otherwise, this would be a nontrivial linear relation.) Thus the sum is direct.

If V is finite-dimensional, then

$$\dim(E(\lambda_1, T)) + \cdots + \dim(E(\lambda_m, T)) = \dim(E(\lambda_1) \oplus \cdots \oplus E(\lambda_m)) \leq \dim(V).$$

□

Q: How can we tell when a linear operator is diagonalizable?

Theorem 23 (Criteria for diagonalizability)**.** *Let V be finite-dimensional and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. TFAE.*

(i) *T is diagonalizable.*
(ii) *V has a basis consisting of eigenvectors of T.*
(iii) $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$.
(iv) $\dim(V) = \dim(E(\lambda_1, T)) + \cdots + \dim(E(\lambda_m, T))$.

Proof. (a) $\iff$ (b): T has a diagonal matrix

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

with respect to a basis $v_1, \ldots, v_n$ iff $T(v_k) = \lambda_k$ for each k.

(b) $\implies$ (c): Assume V has a basis of eigenvectors of T. Then every $v \in V$ can be written as a linear combination of eigenvectors, so

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

and we know the sum is direct from the previous result.

(c) $\implies$ (d): Dimension of direct sum is sum of dimensions of the summands.

(d) $\implies$ (b): Assume

$$\dim(V) = \dim(E(\lambda_1, T)) + \cdots + \dim(E(\lambda_m, T))\,.$$

Choose a basis for each $E(\lambda_k, T)$, and concatenate these to form a list $v_1, \ldots, v_n$. Claim: These vectors are linearly independent. (Exercise.) Since this list has length $\dim(V)$, then it also spans, hence is a basis. □

Remark 24.

- Every linear operator on a finite-dimensional $\mathbb{C}$-vector space has an eigenvalue.
- Every linear operator on a finite-dimensional $\mathbb{C}$-vector space can be upper triangularized.
- Not every linear operator on a finite-dimensional $\mathbb{C}$-vector space can be diagonalized.

Example 25. Define

$$\begin{aligned} T : \mathbb{F}^3 &\to \mathbb{F}^3 \\ v &\mapsto Av \end{aligned}$$

where

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

Note that A is upper triangular, but not diagonal. [Compute A^2, A^3.] Thus $T^3 = 0$, so 0 is the only possible eigenvalue of T. Since $E(0, T) = \ker(T)$, we see that

$$E(0, T) = \{(a, 0, 0) \in \mathbb{F}^3 : a \in \mathbb{F}\}\,,$$

which is 1-dimensional. Thus (d) fails, so T is not diagonalizable.

Proposition 26. *Let V be finite-dimensional and suppose* $T \in \mathcal{L}(V)$ *has* $\dim(V)$ *distinct eigenvalues. Then T is diagonalizable.*

Proof. Let $n := \dim(V)$ and suppose T has distinct eigenvalues $\lambda_1, \ldots, \lambda_n$. Then $\dim(E(\lambda_i)) \geq 1$ for each i, so

$$\dim(E(\lambda_1)) + \cdots + \dim(E(\lambda_n)) \geq 1 + \cdots + 1 = n = \dim(V) .$$

The reverse inequality is always true so we have equality, hence T is diagonalizable. $\square$