18.700 - LINEAR ALGEBRA, DAY 13
UPPER TRIANGULAR MATRICES, DIAGONALIZATION

SAM SCHIAVONE

CONTENTS

I. PRE-CLASS PLANNING

I.1. **Goals for lesson.**

(1) Students will learn the definition of eigenspace.
(2) Students will learn criteria for diagonalizability.
(3) Students will learn that a linear operator is diagonalizable iff its minimal polynomial splits and has no repeated roots.
(4) Students will compute powers of a linear operator using diagonalization.

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

II.1. Last time.

- Proved that the roots of the minimal polynomial are exactly the eigenvalues.
- Saw how to compute the eigenvalues and eigenvectors of a linear operator by upper triangularizing.
- Proved that the eigenvalues of an upper triangular matrix are the diagonal entries.

II.2. 5C Upper triangular matrices, cont.

Remark 1. Let $T \in \mathcal{L}(V)$. Last time we talked about choosing a basis so that $[T]_{\mathcal{B}}$ is upper triangular. This is NOT the same as the "upper triangulation" step in row reduction. Row reducing the matrix $[T]$ is equivalent to multiplying $P[T]$ by some invertible matrix P. This doesn't change the kernel of T, but it does change the outputs of T!

Proposition 2. *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 3. *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 $[T]_{\mathcal{B}}$ is upper triangular.*

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.3. 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 4. 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 5 (Give example, one where 0 is one of the diagonal entries.)**.**

Proposition 6. *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 7. 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 8. Diagonalizable $\neq$ diagonal!

Example 9. 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 10. 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 11. 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 12. λ is an eigenvalue of T iff $E(\lambda, T) \neq \{0\}$.

Theorem 13 (Sum of eigenspaces is direct)**.** *Let $T \in \mathcal{L}(V)$ and suppose $\lambda_1, \ldots, \lambda_m$ 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 $k = 1, \ldots, m$. 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 14 (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 $v_1, \dots, v_n$ 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 $v_1, \dots, v_n$. Claim: These vectors are linearly independent. (Exercise.) Since this list has length $\dim(V)$, then it also spans, hence is a basis. ☐

Remark 15.

- 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 16. 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 17. *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. □

Remark 18. The converse is NOT true. For instance, the identity operator has 1 as a repeated eigenvalue, but is clearly diagonalizable. Indeed, its matrix is diagonal with respect to any basis.

Theorem 19. *Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then T is diagonalizable iff* $\text{minpoly}(T)$ *splits into degree* 1 *factors and has no repeated roots, i.e., there exist distinct $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$ such that* $\text{minpoly}(T) = (z - \lambda_1) \cdots (z - \lambda_m)$.

Proof. ($\Rightarrow$): Assume T is diagonalizable. Then there is a basis of V consisting of eigenvectors of T. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. Then for each j, there exists λ_k such that $(T - \lambda_k I)v_j = 0$. Thus

$$(T - \lambda_1 I) \cdots (T - \lambda_m I)v_j = 0$$

for each j, so $\text{minpoly}(T) = (z - \lambda_1) \cdots (z - \lambda_m)$.

($\Leftarrow$): Assume the $\text{minpoly}(T) = (z - \lambda_1) \cdots (z - \lambda_m)$ where $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$ are distinct. By strong induction on m.

Base case: $m = 1$. Then $T - \lambda_1 I = 0$, so $T = \lambda_1 I$, which is diagonalizable.

Inductive step: Assume $m \geq 2$ and the result holds for all $k < m$. By a previous result, $U := \text{img}(T - \lambda_m I)$ is T-invariant, so we can restrict T to this subspace. Given $u \in \text{img}(T - \lambda_m I)$, then [ask students] $u = (T - \lambda_m I)(v)$ for some $v \in V$. Then

$$(T - \lambda_1 I) \cdots (T - \lambda_{m-1} I)u = (T - \lambda_1 I) \cdots (T - \lambda_{m-1} I)(T - \lambda_m I)(v) = 0\,, \qquad (\ddagger)$$

so this polynomial kills $T|_U$. Thus $(z - \lambda_1) \cdots (z - \lambda_{m-1})$ is a polynomial multiple of $\text{minpoly}(T|_U)$. By the inductive hypothesis, then $\text{img}(T - \lambda_m I)$ has a basis consisting of eigenvectors of $T|_U$, and hence of T.

Claim: $\text{img}(T - \lambda_m I) + \ker(T - \lambda_m I)$ is direct. Given $u \in \text{img}(T - \lambda_m I) \cap \ker(T - \lambda_m I)$, then $(T - \lambda_m I)(u) = 0 \iff T(u) = \lambda_m u$. Since $u \in \text{img}(T - \lambda_m I)$, then

$$0 = (T - \lambda_1 I) \cdots (T - \lambda_{m-1} I)u = (\lambda_m - \lambda_1) \cdots (\lambda_m - \lambda_{m-1})u$$

by ($\ddagger$). Since the λ_i are distinct, this implies that $u = 0$. Thus $\text{img}(T - \lambda_m I) \cap \ker(T - \lambda_m I) = \{0\}$.

Since the sum is direct, then

$$\dim(\text{img}(T - \lambda_m I) \oplus \ker(T - \lambda_m I)) = \dim(\text{img}(T - \lambda_m I)) + \dim(\ker(T - \lambda_m I)) = \dim(V)$$

by Rank-Nullity. Thus $V = \text{img}(T - \lambda_m I) \oplus \ker(T - \lambda_m I)$.

We already saw that $\text{img}(T - \lambda_m I)$ has a basis of eigenvectors of T. Since $\ker(T - \lambda_m I)$ is exatly the λ_m-eigenspace of T, taking a basis of $\ker(T - \lambda_m I)$ and concatenating it with the basis of $\text{img}(T - \lambda_m I)$ yields a basis of V of eigenvectors of T. Thus T is diagonalizable. □

Corollary 20. *Suppose $T \in \mathcal{L}(V)$ is diagonalizable and U is a T-invariant subspace of V. Then $T|_U$ is diagonalizable.*

Proof. Since T is diagonalizable, then minpoly(T) splits and has no repeated roots. By a previous results, minpoly(T) is a polynomial multiple of minpoly$(T|_U)$, so minpoly$(T|_U)$ also splits and has no repeated roots. $\square$

II.3.1. *Worksheet.*

II.3.2. *Gershgorin disc theorem.*

Definition 21. Suppose $T \in \mathcal{L}(V)$ and $\mathcal{B} := (v_1, \dots, v_n)$ is a basis of V. Let $A = [T]_{\mathcal{B}}$. A *Gershgorin disc* of T with respect to $\mathcal{B}$ is a set of the form

$$\left\{ z \in \mathbb{F} : |z - A_{j,j}| \leq \sum_{\substack{k=1 \\ k \neq j}}^{n} |A_{j,k}| \right\}$$

where $j \in \{1, \dots, n\}$.

Remark 22.

- T has n Gershgorin discs, one for each $j = 1, \dots, n$.
- For $\mathbb{F} = \mathbb{R}$, the j^{th} disc is a closed interval centered at $A_{j,j}$, with radius the sum of the absolute values of all the entries in the j^{th} row.
- For $\mathbb{F} = \mathbb{R}$, the j^{th} disc is a closed disc centered at $A_{j,j}$. [Draw picture.]

Theorem 23 (Gershgorin Disc Theorem)**.** *With notation as above, each eigenvalue of T is contained in some Gershgorin disc of T with respect to $\mathcal{B}$.*

Proof. Suppose $\lambda \in \mathbb{F}$ is an eigenvalue of T with corresponding eigenvector w. Then we can uniquely write

$$w = c_1 v_1 + \cdots + c_n v_n$$

for some $c_1, \dots, c_n \in \mathbb{F}$. Let $A = [T]_{\mathcal{B}}$. Applying T to both sides, then

$$\lambda c_1 v_1 + \cdots + \lambda c_n v_n = \lambda w = T(w) = T\left(\sum_{k=1}^{n} c_k v_k\right) = \sum_{k=1}^{n} c_k T(v_k)\,. \tag{24}$$

Now

$$[T(v_k)]_{\mathcal{B}} = [T]_{\mathcal{B}}[v_k]_{\mathcal{B}} = Ae_k = \begin{pmatrix} A_{1,k} \\ \vdots \\ A_{n,k} \end{pmatrix},$$

so

$$T(v_k) = A_{1,k} v_1 + \cdots + A_{n,k} v_n = \sum_{j=1}^{n} A_{j,k} v_j\,.$$

Substituting this into (24), we have

$$\sum_{k=1}^{n} c_k T(v_k) = \sum_{k=1}^{n} c_k \sum_{j=1}^{n} A_{j,k} v_j = \sum_{j=1}^{n} \left(\sum_{k=1}^{n} A_{j,k} c_k\right) v_j\,.$$

Equating the coefficients of v_j from the above and (24), then

$$\lambda_j = \sum_{k=1}^{n} A_{j,k} c_k \tag{†}$$

for all $j = 1, \ldots, n$. Now let $m \in \{1, \ldots, n\}$ be such that

$$|c_m| = \max\{|c_1|, \ldots, |c_n|\}\,.$$

Taking $j = m$ in (†) and subtracting $A_{m,m}c_m$ from both sides, we have [start in middle]

$$(\lambda - A_{m,m})c_m = \lambda c_m - A_{m,m}c_m = \left(\sum_{k=1}^{n} A_{m,k}c_k\right) - A_{m,m}c_m = \sum_{\substack{k=1 \\ k\neq m}}^{n} A_{m,k}c_k\,.$$

Then

$$|\lambda - A_{m,m}| = \left|\sum_{\substack{k=1 \\ k\neq m}}^{n} A_{m,k}\frac{c_k}{c_m}\right| \leq \sum_{\substack{k=1 \\ k\neq m}}^{n} |A_{m,k}| \left|\frac{c_k}{c_m}\right| \leq \sum_{\substack{k=1 \\ k\neq m}}^{n} |A_{m,k}|$$

by the triangle inequality. Thus λ is in the m^{th} Gershgorin disc. □