18.700 - LINEAR ALGEBRA, DAY 22
GENERALIZED EIGENSPACE DECOMPOSITION
JORDAN CANONICAL FORM, TRACE

SAM SCHIAVONE

CONTENTS

I. PRE-CLASS PLANNING

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

(1) Students will learn the Cayley-Hamilton theorem.
(2) Students will learn the definition of Jordan basis and Jordan canonical form.
(3) Students will learn the definition of trace.

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 (3) Chalk

(0:00)

II. LESSON PLAN

II.1. **Last time.**

- Defined generalized eigenvectors.
- Defined generalized eigenspaces: for $T \in \mathcal{L}(V)$,

$$\begin{aligned} G_\lambda(T) &= \{v \in V : (T - \lambda I)^k(v) = 0 \text{ for some } k \in \mathbb{Z}_{\geq 0}\} \\ &= \ker((T - \lambda I)^{\dim(V)}) . \end{aligned}$$

- Proved the generalized eigenspace decomposition theorem:

$$V = G_{\lambda_1}(T) \oplus \cdots \oplus G_{\lambda_m}(T)$$

  where $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of T.
- Defined geometric and algebraic (aka generalized) multiplicities of an eigenvalue λ:

$$\begin{aligned} \text{geometric multiplicity of } \lambda &= \dim(E_\lambda(T)) \\ \text{algebraic multiplicity of } \lambda &= \dim(G_\lambda(T)) . \end{aligned}$$

- Defined the characteristic polynomial:

$$\operatorname{charpoly}(T) := (z - \lambda_1)^{d_1} \cdots (z - \lambda_m)^{d_m}$$

  where $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of T, and λ_i has algebraic multiplicity d_i.

II.2. **8B: Generalized eigenspace decomposition, cont.**

Let V be a nonzero finite-dimensional vector space.

Proposition 1. *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Then*

(a) charpoly(T) *has degree* $\dim(V)$*; and*
(b) the zeroes of charpoly(T) *are exactly the eigenvalues of T.*

Proof. (a) Recall that the algebraic multiplicity of λ_k is $\dim(G_{\lambda_k}(T))$. Since

$$V = G_{\lambda_1} \oplus \cdots \oplus G_{\lambda_m} ,$$

where $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of T, then

$$\dim(V) = \dim(G_{\lambda_1}) + \cdots + \dim(G_{\lambda_m}) = \deg(\operatorname{charpoly}(T)) .$$

(b) Immediate from the definition.

□

Theorem 2 (Cayley-Hamilton)**.** *Suppose $\mathbb{F} = \mathbb{C}$. Suppose $T \in \mathcal{L}(V)$ and let $q =$* charpoly(T)*. Then $q(T) = 0$ (i.e., the zero linear map).*

Proof. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T, and let $d_k := \dim(G_{\lambda_K})$ be the algebraic multiplicity of λ_k for $k = 1, \ldots, m$. For each k, we have seen that $(T - \lambda_k I)|_{G_{\lambda_k}}$ is nilpotent, so

$$(T - \lambda_k I)^{d_k}|_{G_{\lambda_k}} .$$

By the generalized eigenspace decomposition, each vector $v \in V$ can be written as $v = v_1 + \cdots + v_m$ with $v_k \in G_{\lambda_k}$ for each k. Thus to show that $q(T) = 0$, it suffices to show $q(T)|_{G_{\lambda_k}} = 0$ for each k.

Fix $k \in \{1, \dots, m\}$. We have

$$q(T) = (T - \lambda_1 I)^{d_1} \cdots (T - \lambda_m I)^{d_m} .$$

Recall that polynomials in T commute, so we can change the order of the factors above. Thus

$$\begin{aligned} q(T)|_{G_k} &= (T - \lambda_1 I)^{d_1} \cdots (T - \lambda_{k-1})^{d_{k-1}} (T - \lambda_{k+1})^{d_{k+1}} \cdots (T - \lambda_m I)^{d_m}|_{G_k} \underbrace{(T - \lambda_k I)^{d_k}|_{G_k}}_{0} \\ &= 0 . \end{aligned}$$

□

Proposition 3. *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Then* minpoly(T) *divides* charpoly(T), *i.e.,*

$$\text{charpoly}(T) = \text{minpoly}(T)\, f(z)$$

for some $f(z) \in \mathcal{P}(\mathbb{F})$.

Proof. Letting $q := \text{charpoly}(T)$, then $q(T) = 0$. By a previous result, then minpoly(T) must divide q. □

Proposition 4. *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Let $\mathcal{B}$ be a basis of V such that $[T]_{\mathcal{B}}$ is upper triangular. For each eigenvalue λ of T, then number of times that λ appears on the diagonal of $[T]_{\mathcal{B}}$ is equal to the algebraic multiplicity of λ.*

Proof. Let $A := [T]_{\mathcal{B}}$. [Write out A with its entries in the k^{th} column. Recall that $[T]_{\mathcal{B}}$ has columns $[T(v_i)]_{\mathcal{B}}$.] Then for each k we have

$$T(v_k) = \overbrace{c_1 v_1 + \cdots + c_{k-1} v_{k-1}}^{u_k} + \lambda_k v_k ,$$

where $u_k \in \text{span}(v_1, \dots, v_{k-1})$. Thus if $\lambda_k \neq 0$, then $T(v_k)$ is not a linear combination of $T(v_1), \dots, T(v_{k-1}) \in \text{span}(v_1, \dots, v_{k-1})$. (These only involve the vectors $v_1, \dots, v_{k-1}$.) By the Linear Dependence Lemma, then the collection of $T(v_k)$ such that $\lambda_k \neq 0$ is linearly independent.

Let d be the number of indices $k \in \{1, \dots, n\}$ such that $\lambda_k = 0$. By the above, then

$$n - d \leq \dim(\text{img}(T)) = \dim(V) - \dim(\ker(T)) = n - \dim(\ker(T))$$

by Rank-Nullity. Then $\dim(\ker(T)) \leq d$.

Now, note that $[T^n]_{\mathcal{B}} = [T]_{\mathcal{B}}^n = A^n$. Moreover, the diagonal entries of A^n are $\lambda_1^n, \dots, \lambda_n^n$. Since $\lambda_k^n = 0$ iff $\lambda_k = 0$, then 0 appears on the diagonal of A^n d times, too. Thus the reasoning above applies just as well to T^n, so we have

$$\dim(\ker(T^n)) \leq d . \tag{5}$$

For each eigenvalue λ of T, let m_λ denote the algebraic multiplicity of λ, and let d_λ be the number of times λ appears on the diagonal of A. Replacing T with $T - \lambda I$ in (5), then

$$m_\lambda \leq d_\lambda \tag{6}$$

for each eigenvalue λ of T. Summing over all eigenvalues λ, we have [start in middle]

$$n = \dim(V) = \sum_{\lambda} m_{\lambda} \leq \sum_{\lambda} d_{\lambda} = n$$

where the second equality follows from the generalized eigenspace decomposition, and the last equality from the fact that the diagonal of A consists of n entries.

Thus the inequality in (6) must in fact be an equality for all eigenvalues λ. □

Definition 7. A *block diagonal matrix* is a square matrix of the form

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

where $A_1, \dots, A_m$ are square matrices (of possibly different sizes) lying on the diagonal, and all other entres are 0.

Example 8 (Give example. 2×2, 1×1 and 3×3 together.)**.**

Proposition 9. *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T, with algebraic multiplicities $d_1, \dots, d_m$. Then there is a basis $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}$ is block diagonal*

$$[T]_{\mathcal{B}} = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix}$$

where each A_k is a $d_k \times d_k$ upper triangular matrix of the form

$$A_k := \begin{pmatrix} \lambda_k & * & \\ & \ddots & \\ 0 & & \lambda_k \end{pmatrix}.$$

Proof. By a previous result, $(T - \lambda_k I)|_{G_{\lambda_k}}$ is nilpotent for each k. Thus for each k we can choose a basis $\mathcal{B}_k$ such that $[(T - \lambda_k I)|_{G_{\lambda_k}}]_{\mathcal{B}_k}$ is strictly upper triangular. [Draw picture.] Now,

$$T|_{G_{\lambda_k}} = (T - \lambda_k I)|_{G_{\lambda_k}} + \lambda_k I|_{G_{\lambda_k}}$$

so

$$[T|_{G_{\lambda_k}}]_{\mathcal{B}} = [(T - \lambda_k I)|_{G_{\lambda_k}}]_{\mathcal{B}} + [\lambda_k I|_{G_{\lambda_k}}]_{\mathcal{B}}$$

[draw picture below].

This deals with a single block. Now concatenate the bases $\mathcal{B}_1, \dots, \mathcal{B}_m$ to form a basis $\mathcal{B}$ of V. Then $[T]_{\mathcal{B}}$ is of the desired form. □

II.3. **8C: Jordan form.** We have seen that, for $\mathbb{F} = \mathbb{C}$, for every linear operator $T \in \mathcal{L}(V)$ there is a basis $\mathcal{B}$ such that $[T]_{\mathcal{B}}$ is upper triangular. And even more: we can find a basis such that $[T]_{\mathcal{B}}$ is a block diagonal matrix whose blocks are upper triangular. We'll now see that we can do even better: we can find a basis $\mathcal{B}$ such that the only nonzero entries of $[T]_{c}alB$ possibly occur on the diagonal and the super-diagonal (i.e., the line directly above the diagonal). [Draw picture.]

Example 10. Let $T \in \mathcal{L}(V)$ be defined by $T(v) = Av$ where

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

Then $T^3 = 0$ so T is nilpotent. Since A has 2 pivots, we see that $\dim(E_0(T)) = \dim(\ker(T)) = 1$. We can see that $v_1 := (0,0,1)$ is an eigenvector with eigenvalue 0. Now we want to find the generalized eigenvectors with eigenvalue 0 that are not eigenvectors. One way to do this: find v_2 such that $T(v_2) = v_1$. Then $T^2(v_2) = T(v_1) = 0$. Solving this system by row reducing the augmented matrix $(A|v_1)$, we find that

$$v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + c \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + cv_1$$

for any $c \in \mathbb{F}$. Taking $c = 0$, we have $v_2 = (0,1,0)$. We now repeat this process and search for a vector v_3 such that $T(v_3) = v_2$. Row reducing $(A|v_2)$, we find $v_3 = (1,0,0)$. Letting $\mathcal{B}$ be the basis

$$v_1, v_2, v_3 = T^2(v_3), T(v_3), v_3$$

then

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

Remark 11. In general, there may be several eigenvectors, and one will have to work backwards from each eigenvector to obtain a basis of generalized eigenvectors. Consider the matrix

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

for example.

Definition 12. Let $T \in \mathcal{L}(V)$. A *Jordan basis* for T is a basis $\mathcal{B}$ of V such that

$$[T]_{\mathcal{B}} = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix}$$

is block diagonal, and each block A_k is of the form

$$A_k = \begin{pmatrix} \lambda_k & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_k \end{pmatrix}.$$

We say that the matrix $[T]_{\mathcal{B}}$ is in *Jordan canonical form*.

Proposition 13. *Suppose $T \in \mathcal{L}(V)$ is nilpotent. Then T has a Jordan basis.*

Proof. Let $n := \dim(V)$. By strong induction on n.

Base case: $n = 1$. Then T must be the 0 operator, and any basis is a Jordan basis for T.

Inductive step: Let $n \geq 2$ and assume the result holds for all $k < n$. As we have done several times before, we will find a T-invariant subspace U and apply the inductive hypothesis to the restriction $T|_U$.

Let m be the smallest positive integer such that $T^m = 0$. Then there exists $u \in V$ such that $T^{m-1}(u) \neq 0$. Let

$$U := \operatorname{span}(u, T(u), \ldots, T^{m-1}(u)).$$

By Exercise 2 of Section 8A, $u, T(u), \ldots, T^{m-1}(u)$ is linearly independent. If $U = V$, then $T^{m-1}(u), \ldots, T(u), u$ is a Jordan basis for T.

Thus it suffices to consider the case $U \neq V$. Note that U is T-invariant: applying T to one of the basis vectors simply shifts us over one spot, and $T(T^{m-1}(u)) = T^m(u) = 0$. Since $U \neq V$, then by the inductive hypothesis there is a basis of U that is a Jordan basis for $T|_U$. Goal: Find a subspace W of V such that $V = U \oplus W$.

Let $\varphi : V \to \mathbb{F}$ be a linear functional such that $\varphi(T^{m-1}(u)) \neq 0$. (Such a linear functional exists: since $u, T(u), \ldots, T^{m-1}(u)$ is linearly independent, we can extend it to a basis for V. We can then freely choose the values of φ on these basis vectors.) Define

$$W := \{v \in V : \varphi(T^k(v)) = 0 \ \forall k = 1, \ldots, m-1\}.$$

Then W is a subspace and is moreover T-invariant (exercise). Claim: $V = U \oplus W$.

(i) Suppose $v \in U$ with $v \neq 0$. We will show that $v \notin W$, so $U \cap W = \{0\}$. Since $v \in U$, then

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

for some $c_0, \ldots, c_{m-1} \in \mathbb{F}$. Let j be the smallest index such that $c_j \neq 0$. Applying T^{m-j-1} kills all the terms after the j^{th} one on the righthand side, so

$$T^{m-j-1}(v) = c_j T^{m-1}(u).$$

Now applying φ, we have

$$\varphi(T^{m-j-1}(v)) = c_j \varphi(T^{m-1}(u)) \neq 0$$

by the definition of φ and c_j. Thus $v \notin W$, so $U \cap W = \{0\}$.

(ii) Goal: $V = U + W$. Define

$$\begin{aligned} S &\to \mathbb{F}^m \\ v &\mapsto (\varphi(v), \varphi(T(v)), \ldots, \varphi(T^{m-1}(v))). \end{aligned}$$

Then $\ker(S) = W$. [Recall definition of W.] Then

$$\begin{aligned} \dim(W) = \dim(\ker(S)) &= \dim(V) - \dim(\operatorname{img}(S)) \geq \dim(V) - \dim(\mathbb{F}^m) \\ &= \dim(V) - m \end{aligned}$$

by Rank-Nullity. Then

$$\dim(U \oplus W) = \dim(U) + \dim(W) \geq m + (\dim(V) - m) = \dim(V),$$

so we must have equality. Thus $V = U \oplus W$.

□

We can extend the previous result to all operators by using the generalized eigenspace decomposition.

Theorem 14. *Let $\mathbb{F} = \mathbb{C}$ and suppose $T \in \mathcal{L}(V)$. Then T has a Jordan basis.*

Proof. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. By the generalized eigenspace decomposition, we have

$$V = G_{\lambda_1} \oplus \cdots \oplus G_{\lambda_m}$$

and $(T - \lambda_k I)|_{G_{\lambda_k}}$ is nilpotent. By the previous result, then for each k there is a basis $\mathcal{B}_k$ of G_{λ_k} that is a Jordan basis for $(T - \lambda_k I)|_{G_{\lambda_k}}$. Concatenating these bases produces a basis $\mathcal{B}$ of V that is a Jordan basis for T. □

II.4. **8D: Trace.**

Definition 15. Let A be a square matrix with entries in $\mathbb{F}$. The *trace of A*, denoted $\mathrm{tr}(A)$, is the sum of the diagonal entries of A. In other words, if $A \in \mathrm{M}_{\mathrm{n}\times\mathrm{n}}(\mathbb{F})$, then

$$\mathrm{tr}(A) = \sum_{i=1}^{n} A_{ii} = A_{11} + \cdots + A_{nn}\,.$$

Proposition 16. *Suppose $A \in M_{m\times n}(\mathbb{F})$ and $B \in M_{n\times m}(\mathbb{F})$. Then*

$$\mathrm{tr}(AB) = \mathrm{tr}(BA)\,.$$

Proof. Exercise. (See worksheet.) □

This fact will allow us to define the trace of a linear operator, one that is independent of the choice of basis.

Proposition 17. *Suppose $T \in \mathcal{L}(V)$. Let $\mathcal{B}$ and $\mathcal{C}$ be bases of V. Then*

$$\mathrm{tr}([T]_{\mathcal{B}}) = \mathrm{tr}([T]_{\mathcal{C}})\,.$$

Proof. Let $A := [T]_{\mathcal{B}}$, $B := [T]_{\mathcal{C}}$, and $P = {}_{\mathcal{C}}[I]_{\mathcal{B}}$. Then

$$A = [T]_{\mathcal{B}} = {}_{\mathcal{B}}[I]_{\mathcal{C}}\ [T]_{\mathcal{C}}\ {}_{\mathcal{C}}[I]_{\mathcal{B}} = P^{-1}BP\,,$$

so [ask students]

$$\mathrm{tr}(A) = \mathrm{tr}(P^{-1}BP) = \mathrm{tr}((P^{-1}B)P) = \mathrm{tr}(P(P^{-1}B) = \mathrm{tr}(B)$$

by the previous result. □

Definition 18. Let $T \in \mathcal{L}(V)$. The *trace of T*, denoted $\mathrm{tr}(T)$, is defined to be

$$\mathrm{tr}(T) := \mathrm{tr}([T]_{\mathcal{B}})$$

where $\mathcal{B}$ is any basis of V.

Remark 19. By the previous result, $\mathrm{tr}(T)$ is well-defined.

The trace has an interesting relationship with eigenvalues: it is their sum.

Proposition 20. *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of T, with each repeated as many times as its algebraic multiplicity. Then*

$$\mathrm{tr}(T) = \lambda_1 + \cdots + \lambda_n\,.$$

Proof. By a previous result, there exists a basis $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}$ is upper triangular with diagonal entries $\lambda_1, \ldots, \lambda_n$ (again, repeated with algebraic multiplicity). Then

$$\operatorname{tr}(T) = \operatorname{tr}([T]_{\mathcal{B}}) = \lambda_1 + \cdots + \lambda_n \,.$$

□

The trace also has an interpretation in terms of the characteristic polynomial.

Proposition 21. *Suppose* $\mathbb{F} = \mathbb{C}$ *and* $T \in \mathcal{L}(V)$. *Let* $n := \dim(V)$. *Then* $\operatorname{tr}(T)$ *equals negative the coefficient of* z^{n-1} *in the characteristic polynomial of* T. *I.e., wiriting*

$$\operatorname{charpoly}(T) = z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0 \,,$$

then $\operatorname{tr}(T) = -a_{n-1}$.

Proof. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of T, with each repeated as many times as its algebraic multiplicity. Then

$$\operatorname{charpoly}(T) = (z - \lambda_1) \cdots (z - \lambda_n) \,.$$

(Instead of writing $(z - \lambda_k)^{d_k}$, we're just writing $(z - \lambda_k)$ d_k times.) Multiplying this expression out [explain about choosing $n - 1$ factors of z], we have

$$\operatorname{charpoly}(T) = z^n - (\lambda_1 + \cdots + \lambda_n) z^{n-1} + \cdots + (-1)^n (\lambda_1 \cdots \lambda_n) \,.$$

□

Proposition 22. *The function* $\operatorname{tr} : \mathcal{L}(V) \to \mathbb{F}$ *is linear. I.e.,* tr *is a linear functional on* $\mathcal{L}(V)$.

Proof. Exercise. □