18.700 - LINEAR ALGEBRA, DAY 23
JORDAN CANONICAL FORM, TRACE
BILINEAR AND MULTILINEAR FORMS

SAM SCHIAVONE

CONTENTS

I. PRE-CLASS PLANNING

I.1. Goals for lesson.

(1) Students will learn that every linear map has Jordan basis.
(2) Students will learn the definition of trace.
(3) Students will learn the definition of a multilinear form.
(4) Students will learn the definitions of an inversion and the sign of a permutation.
(5) Students will learn how permuting the entries affects the value of an alternating multilinear form.

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.**

- Proved the Cayley-Hamilton theorem: Let $q(z) = \text{charpoly}(T)$. Then $q(T) = 0$.
- Defined Jordan basis and Jordan canonical form:

  Definition 1. 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*.

- Proved the matrix form of the Generalized Eigenspace Decomposition Theorem: there exists a basis $\mathcal{B}$ such that $[T]_{\mathcal{B}}$ is block diagonal with upper triangular blocks.

II.2. **8C: Jordan form, cont.**

Proposition 2. *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

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

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 3. *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.3. **8D: Trace.**

Definition 4. Let A be a square matrix with entries in $\mathbb{F}$. The *trace of A*, denoted $\operatorname{tr}(A)$, 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

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

Proposition 5. *Suppose $A \in M_{m\times n}(\mathbb{F})$ and $B \in M_{n\times m}(\mathbb{F})$. Then*

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

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

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

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]

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

by the previous result. □

Definition 7. Let $T \in \mathcal{L}(V)$. The *trace of T*, denoted $\operatorname{tr}(T)$, is defined to be

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

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

Remark 8. By the previous result, $\operatorname{tr}(T)$ is well-defined.

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

Proposition 9. *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*

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

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 10. *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. [Skip, if necessary.] 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 11. *The function* $\mathrm{tr} : \mathcal{L}(V) \to \mathbb{F}$ *is linear. I.e.,* tr *is a linear functional on* $\mathcal{L}(V)$.

Proof. Exercise. □

II.4. **9A, 9B: Bilinear and multilinear forms.**

Definition 12. A *bilinear form* on V is a function $\beta : V \times V \to \mathbb{F}$ that is linear in each component: for each $w \in V$, the maps

$$\begin{aligned} V &\to \mathbb{F} \\ v &\mapsto \beta(v, w) \end{aligned}$$

and

$$\begin{aligned} V &\to \mathbb{F} \\ v &\mapsto \beta(w, v) \end{aligned}$$

are both linear. Denote the set of bilinear forms on V by $V^{(2)}$.

More concretely,

$$\beta(cu + v, w) = c\beta(u, w) + \beta(v, w)$$

and

$$\beta(w, cu + v) = c\beta(w, u) + \beta(w, v)$$

for all $u, v, w \in V$ and all $c \in \mathbb{F}$.

Lemma 13. $V^{(2)}$ *is a vector space under pointwise addition and scalar multiplication of functions.*

Proof. Exercise. □

Example 14.

- Let $\mathbb{F} = \mathbb{R}$ and V be an $\mathbb{R}$-vector space. Then every inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is a bilinear form.
- Given $A \in M_{n\times n}(\mathbb{R})$, define

  $$\begin{aligned} \beta : \mathbb{F}^n \times F^n &\to \mathbb{F} \\ (x, y) &\mapsto x^t A y \,. \end{aligned}$$

  Then β is bilinear by properties of matrix multiplication.

Definition 15. Fix $m \in \mathbb{Z}_{>0}$. Denote $V^m = \overbrace{V \times \cdots \times V}^{m \text{ times}}$.

- An *m-linear form* on V is a function $\beta : V^m \to \mathbb{F}$ that is linear in each component when the others are held fixed. I.e., for each $k \in \{1, \ldots, m\}$ and $u_1, \ldots, u_m \in V$, the map

  $$\begin{aligned} V &\to \mathbb{F} \\ v &\mapsto \beta(u_1, \ldots, u_{k-1}, v, u_{k+1}, \ldots, u_m) \end{aligned}$$

  is linear.
- Denote by $V^{(m)}$ the set of all m-linear forms on V.

- A *multilinear form* on V is an m-linear form on V for some $m \in \mathbb{Z}_{>0}$.

Definition 16. Let $m \in \mathbb{Z}_{>0}$.

- An m-linear form $\alpha \in V^{(m)}$ is *alternating* if $\alpha(v_1, \ldots, v_m) = 0$ whenever $v_j = v_k$ for some $j, k \in \{1, \ldots, m\}$ with $j \neq k$.
- Let $V^{(m)}_{\text{alt}}$ be the set of all alternating m-linear forms on V.

Lemma 17. *$V^{(m)}$ is a vector space, and $V^{(m)}_{\text{alt}}$ is a subspace.*

Proof. Exercise. □

Lemma 18. *Let $\alpha \in V^{(m)}_{\text{alt}}$. If $v_1, \ldots, v_m \in V$ is linearly dependent, then*

$$\alpha(v_1, \ldots, v_m) = 0\,.$$

Proof idea. Use the Linear Dependence Lemma to express v_k as a linear combination of the others. Then use multilinearity and alternating property. Details left as an exercise. □

Proposition 19. *Let $\alpha \in V^{(m)}_{\text{alt}}$ and $v_1, \ldots, v_m \in V$. Swapping the vectors in any two slots of $\alpha(v_1, \ldots, v_m)$ changes the value by a factor of -1.*

Proof idea. For simplicity, suppose $m = 2$. Then

$$0 = \alpha(v+w, v+w) = \cancelto{0}{\alpha(v,v)} + \alpha(v,w) + \alpha(w,v) + \cancelto{0}{\alpha(w,w)}\,.$$

The proof is virtually the same for $m \geq 2$. □

<u>Q</u>: What if we perform multiple swaps? For example, suppose that $\alpha \in V^{(3)}_{\text{alt}}$ and $v_1, v_2, v_3 \in V$. Then

$$\alpha(v_3, v_1, v_2) = -\alpha(v_1, v_3, v_2) = \alpha(v_1, v_2, v_3)\,.$$

This leads us to investigate more general permutations.

Definition 20. Let $m \in \mathbb{Z}_{>0}$.

- A *permutation* of $(1, \ldots, m)$ is a rearrangement, i.e., a list $(j_1, \ldots, j_m)$ that contains each of $1, \ldots, m$ exactly once.
- Denote the set of all permutations of $(1, \ldots, m)$ by $\operatorname{perm}(m)$.

Example 21. $(2, 1, 4, 3) \in \operatorname{perm}(4)$.

Definition 22. Suppose $(j_1, \ldots, j_m) \in \operatorname{perm}(m)$.

- An *inversion* of $(j_1, \ldots, j_m)$ is a pair of integers (k, ℓ) with $k, \ell \in \{1, \ldots, m\}$ such that $k < \ell$ and k appears *after* ℓ in the list $(j_1, \ldots, j_m)$.
- Let N be the number of inversions of $(j_1, \ldots, j_m)$. The *sign* of $(j_1, \ldots, j_m)$ is

$$\operatorname{sgn}(j_1, \ldots, j_m) := (-1)^N\,.$$

Example 23.

- Consider $(2, 1, 3, 4) \in \operatorname{perm}(4)$. It has exactly one inversion, namely $(1, 2)$, so it has sign $(-1)^1 = -1$.
- The permutation $(1, \ldots, m)$ has no inversions (the numbers are all in increasing order), so it has sign $(-1)^0 = 1$.

- Consider the permutation $(2, 3, \ldots, m, 1)$. Its inversions are

$$(1, 2), (1, 3), \ldots, (1, m)$$

so it has sign $(-1)^{m-1}$.

Proposition 24. *Swapping two entries in a permutation multiplies the sign of the permutaiton by -1.*

Proof. Let π be the original position, and π' be the permutation obtained from swapping the i^{th} and j^{th} entries of π. Denote the i^{th} entry of π by $\pi(i)$. Then $\pi(i) < \pi(j)$ iff $\pi'(i) > \pi'(j)$, so we have either added or subtracted exactly 1 inversion so far.

Consider the entries not in between the i^{th} and j^{th} spots. For these entries, there is no change in whether they were in order or not. [Draw picture.]

Now consider $\pi(k)$ with $i < k < j$.

Case 1: $\pi(k)$ was in order with respect to both $\pi(i)$ and $\pi(j)$, i.e., $\pi(i) < \pi(k) < \pi(j)$. Then

$$\pi'(i) > \pi'(k) > \pi'(j)$$

so we have 2 more inversions, multiplying the sign by $(-1)^2 = 1$.

Case 2: $\pi(i) > \pi(k) > \pi(j)$. Similar.

Case 3: $\pi(i) < \pi(k)$ and $\pi(k) > \pi(j)$. Then

$$\pi'(i) = \pi(j) < \pi(k) = \pi'(k)$$
$$\pi'(k) = \pi(k) > \pi(i) = \pi'(j)$$

so we have the same number of inversions that we started with, and the sign is unchanged.

Case 4: $\pi(i) > \pi(k)$ and $\pi(k) < \pi(j)$. Similar.

Thus in all cases we have an odd number of inversions, so $\text{sign}(\pi') = -\text{sign}(\pi)$. □

Proposition 25. *Suppose $m \in \mathbb{Z}_{>0}$ and $\alpha \in V_{\text{alt}}^{(m)}$. Then*

$$\alpha(v_{j_1}, \ldots, v_{j_m}) = \text{sign}(j_1, \ldots, j_m)\alpha(v_1, \ldots, v_m)$$

Proof idea. We can get from $(j_1, \ldots, j_m)$ to $(1, \ldots, m)$ by a series of swaps. Each swap changes the sign of α by a factor of -1, and also changes the sign of the remaining permutation by a factor of -1. □

Theorem 26. *Let $n := \dim(V)$. Suppose $e_1, \ldots, e_n$ is a basis of V. Suppose $v_1, \ldots, v_n \in V$. For each k, write*

$$v_k = \sum_{j=1}^{n} b_{j,k} e_j$$

for some $b_{1,k}, \ldots, b_{n,k} \in \mathbb{F}$. Then

$$\alpha(v_1, \ldots, v_n) = \alpha(e_1, \ldots, e_n) \sum_{(j_1,\ldots,j_n) \in \text{perm}(n)} \text{sign}(j_1, \ldots, j_n) b_{j_1,1} \cdots b_{j_n,n}$$

for all $\alpha \in V_{\text{alt}}^{(m)}$.

Proof.

$$\begin{aligned}
\alpha(v_1,\ldots,v_n) &= \alpha\left(\sum_{j_1=1}^{n} b_{j_1,1}e_{j_1},\ldots \sum_{j_n=1}^{n} b_{j_n,1}e_{j_n}\right) = \sum_{j_1=1}^{n}\cdots\sum_{j_n=1}^{n} b_{j_1,1}\cdots b_{j_n,n}\alpha(e_{j_1},\ldots,e_{j_n}) \\
&= \sum_{(j_1,\ldots,j_n)\in\mathrm{perm}(n)} b_{j_1,1}\cdots b_{j_n,n}\alpha(e_{j_1},\ldots,e_{j_n}) \\
&= \sum_{(j_1,\ldots,j_n)\in\mathrm{perm}(n)} b_{j_1,1}\cdots b_{j_n,n}\,\mathrm{sign}(j_1,\ldots,j_n)\alpha(e_1,\ldots,e_n) \\
&= \alpha(e_1,\ldots,e_n)\sum_{(j_1,\ldots,j_n)\in\mathrm{perm}(n)} \mathrm{sign}(j_1,\ldots,j_n)b_{j_1,1}\cdots b_{j_n,n}\,,
\end{aligned}$$

where the third equality holds because $\alpha(e_{j_1},\ldots,e_{j_n}) = 0$ if $j_1,\ldots,j_n$ are not distinct. □

Corollary 27. $\dim(V_{\mathrm{alt}}^{(n)}) = 1$.

Proof. Let $n := \dim(V)$. Suppose $\alpha,\alpha' \in V_{\mathrm{alt}}^{(n)}$ with $\alpha \neq 0$. Then $\alpha(e_1,\ldots,e_n) \neq 0$ for some $e_1,\ldots,e_n \in V$. Then $e_1,\ldots,e_n$ is linearly independent (contrapositive of earlier result). Let

$$c := \frac{\alpha'(e_1,\ldots,e_n)}{\alpha(e_1,\ldots,e_n)}\,.$$

Letting $b_{j,k}$ be as above, then

$$\begin{aligned}
\alpha'(v_1,\ldots,v_n) &= \alpha'(e_1,\ldots,e_n)\sum_{(j_1,\ldots,j_n)\in\mathrm{perm}(n)} \mathrm{sign}(j_1,\ldots,j_n)b_{j_1,1}\cdots b_{j_n,n} \\
&= c\alpha(e_1,\ldots,e_n)\sum_{(j_1,\ldots,j_n)\in\mathrm{perm}(n)} \mathrm{sign}(j_1,\ldots,j_n)b_{j_1,1}\cdots b_{j_n,n} \\
&= c\alpha(v_1,\ldots,v_n)\,.
\end{aligned}$$

Thus $\alpha' = c\alpha$. Thus $\dim(V_{\mathrm{alt}}^{(n)}) \leq 1$.

It remains to show that $\dim(V_{\mathrm{alt}}^{(m)}) = 1$. For details, see 9.37 in the text book. □

[Skip if necessary.] To prove the next result, we will need some more results on linear functionals. Recall that $V^{\vee}$, the dual space, is

$$V^{\vee} = \mathcal{L}(V,\mathbb{F}) = \{\varphi : V \to \mathbb{F} \mid \varphi \text{ is linear}\}\,.$$

Fix $j \in \{1,\ldots,n\}$. Define

$$\begin{aligned}
\varphi_j : \mathbb{F}^n &\to \mathbb{F} \\
(x_1,\ldots,x_n) &\mapsto x_j\,,
\end{aligned}$$

i.e., projection onto the j^{th} coordinate. Then φ_j is linear and

$$\varphi_j(e_k) = \begin{cases} 1 & \text{if } k = j; \\ 0 & \text{otherwise.} \end{cases}$$

We can define a similar notion in general.

Definition 28. Let $\mathcal{B} := (v_1, \ldots, v_n)$ be a basis of V. The *dual basis of* $\mathcal{B}$ is the list $\mathcal{B}^\vee := (\varphi_1, \ldots, \varphi_n)$ in $V^\vee$, where φ_j is defined by

$$\varphi_j(v_k) = \begin{cases} 1 & \text{if } k = j; \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 29. *Suppose V is finite-dimensional. Then $\mathcal{B}$ is a basis of $V^\vee$.*

Remark 30. φ_j is sometimes denoted $v_j^\vee$.