JORDAN CANONICAL FORM

We have seen that not every linear transformation $T: V \to V$ is diagonalizable

<u>example</u> : $T: \mathbb{C}^2 \to \mathbb{C}^2$
$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x+y \\ y \end{pmatrix}$

Let β be the standard basis for $\mathbb{C}^2$
Then $[T]_\beta^\beta = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, which we have seen is not diagonalizable, so T is not diagonalizable

But we will show over the next couple of classes that even though we cannot always find a basis β for V such that $[T]_\beta^\beta$ is diagonal, we can (provided the characteristic polynomial of T splits) always find a basis β for V such that $[T]_\beta^\beta$ is in JORDAN CANONICAL FORM.

This means that

$$[T]_\beta^\beta = \left(\begin{array}{c|c|c|c} A_1 & 0 & 0 & 0 \\ \hline 0 & A_2 & 0 & 0 \\ \hline 0 & 0 & \ddots & 0 \\ \hline 0 & 0 & 0 & A_k \end{array}\right)$$

where each matrix A_i is either a 1×1 matrix (λ) or takes the form

$$\begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & & \vdots \\ 0 & 0 & \lambda & \ddots & 0 \\ \vdots & \vdots & 0 & \ddots & 1 \\ 0 & 0 & 0 & & \lambda \end{pmatrix}$$

The matrices A_i are called Jordan blocks.

Better still, the Jordan form of a transformation $T : V \to V$ is unique apart from the fact that one could re-order the Jordan blocks $A_1, \dots, A_k$.

Note that the basis β is not unique, but any Jordan canonical basis β gives the same Jordan canonical form (up to re-ordering) example : Let V be a finite-dim. vector space over $\mathbb{C}$.

Then $T: V \to V$ is diagonalizable if and only if all the Jordan blocks in the Jordan canonical form for T are 1×1 matrices.

The Jordan canonical form theorem ~~Theorem~~, which is contained in Chapter 7 of the textbook, gives a complete answer to one of our major questions from the beginning of the course: "given $T: V \to V$ a linear map, if I am allowed to choose a basis β for V then how simple can I make the matrix $[T]_\beta^\beta$ look?"

It will also allow us to fill in the missing part of our work on linear differential equations: if the auxiliary polynomial

$p(t)$ has a repeated root $t = c$ (of multiplicity k, say) then the corresponding part of a basis to the solution space of the differential equation

$$p(D)(y) = 0$$

is $\{e^{ct}, te^{ct}, \ldots, t^{k-1}e^{ct}\}$

We will spend the next few classes working on the proof and the applications of the Jordan canonical form theorem.

~~[illegible]~~

INVARIANT SUBSPACES

In analysing a linear transformation $T: V \to V$, it will help to break up the space V into a direct sum of simpler (smaller) pieces V. But we should not choose these pieces at random: they should be mapped to themselves by T.

<u>Def^n</u>: A subspace W of a vector space V is called T-invariant if $T(W) \subset W$. In other words,

here $T: V \to V$ is a linear map

$$W \text{ is } T\text{-invariant} \iff T(w) \in W \text{ for all } w \in W$$

This notion should be familiar to you from the homework.

<u>example</u>: given $T: V \to V$ linear, $R(T)$ is a T-invariant subspace of V

<u>Proof</u>: If $x \in R(T)$ then $T(x) \in R(T)$. □

example Suppose that $\beta = \{v_1, \ldots, v_a, w_1, \ldots, w_b, \ldots, z_1, \ldots, z_c\}$ is ~~the~~ a basis with respect to which $[T]_\beta^\beta$ is in Jordan canonical form:

$$[T]_\beta^\beta = \begin{array}{c} \\ a \\ b \\ \vdots \\ c \end{array} \begin{array}{c} \begin{array}{cccc} a & b & \cdots & c \end{array} \\ \begin{pmatrix} A_1 & 0 & 0 & 0 \\ 0 & A_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & A_k \end{pmatrix} \end{array}$$

Then the "first block" span$\{v_1, \ldots, v_a\}$ is T-invariant : if $v = \alpha_1 v_1 + \cdots + \alpha_a v_a$ then

$$[T(v)]_\beta = [T]_\beta^\beta [v]_\beta$$

$$= \begin{array}{c} a \\ \\ \\ \\ \end{array} \begin{pmatrix} A_1 & 0 & 0 & 0 \\ 0 & A_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & A_k \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_a \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} A_1 v \\ \hline 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{array}{l} \leftarrow \text{vector of size } a \\ \\ \leftarrow \text{rest are zeroes} \end{array}$$

and so $T(v)$ is in the "first block" span $\{v_1, \ldots, v_a\}$.

Similarly the "second block" ~~span~~ span $\{w_1, \ldots, w_b\}$ is T-invariant, and so on.

So in the process of putting a transformation T into Jordan canonical form, we will need to break T up ~~into~~ as a direct sum of invariant subspaces.

This example suggests that we should figure out how to find T-invariant subspaces of V. The next section describes a simple method for doing so.

CYCLIC SUBSPACES

Let $T: V \to V$ be a linear map and let $v \in V$. The subspace

$$W = \text{span}\{v, T(v), T^2(v), \dots\}$$

is called the *T-cyclic subspace of V generated by v*

Claim: W is T-invariant

Proof: If $w \in W$ then

$$w = a_1 T^{i_1}(v) + \dots + a_n T^{i_n}(v)$$

for some $i_1, \dots, i_n$ and scalars $a_1, \dots, a_n$.

Then

$$T(w) = a_1 T^{i_1+1}(v) + \dots + a_n T^{i_n+1}(v) \quad \text{(by linearity of } T\text{)}$$

$$\in W$$

$\square$

We now figure out:

- how to find a basis for a T-cyclic subspace of V
- what the matrix of T looks like in this basis

Theorem (cf 5.22(a))

Let V be a vector space, $T: V \to V$ be a linear map and W be a T-cyclic subspace of V generated by $v \in V$. Suppose that $\dim W = k$. Then

$$\beta = \{v, T(v), \dots, T^{k-1}(v)\}$$

is a basis for W

Proof: Idea: a largest-possible LI set is a basis

so:

let j be the largest integer such that

$\{v, T(v), \dots, T^{j-1}(v)\}$ is LI.

If $j=0$ then $v=0$ and $W=\{0\}$, so the Theorem is true.

Otherwise $j \geq 1$; we know that $j \leq k$ as $\dim W = k$, and we want to show that $j=k$.

<u>Claim</u>: $\gamma = \{v, T(v), \ldots, T^{j-1}(v)\}$ is a basis for W

<u>Proof</u>: It is LI by definition. We need to show it spans W.

Since $W = \operatorname{span}\{v, T(v), T^2(v), \ldots\}$ it suffices to show that $T^N(v) \in \operatorname{span}(\gamma)$ for all N. This is because any $w \in W$ can be written as a LC of $v, T(v), \ldots$ so if each $T^N(v)$ can be written as a LC of elements of γ then each $w \in W$ can also be written as a LC of elements of γ.

Subclaim: $T^N(v) \in \text{span } \gamma$ for all N

Proof of Subclaim:

This is obvious for $N = 0, 1, 2, \dots, j-1$.
We prove it for $N \geq j$ by induction on N

Base case $N = j$

The set $\{v, T(v), \dots, T^{j-1}(v), T^j(v)\}$ is linearly dependent (look at the definition of j!) so we can find scalars $a_0, \dots, a_j$ with

$$a_0 v + a_1 T(v) + \dots + a_j T^j(v) = 0$$

and not all the $a_i = 0$.
But a_j cannot be zero, as otherwise

$$a_0 v + a_1 T(v) + \dots + a_{j-1} T^{j-1}(v) = 0$$

$\Rightarrow$ $a_0 = a_1 = \dots = a_{j-1} = 0$ too (as γ is LI)

$\times\!\!\!\times$

So $a_j \neq 0$, and

$$T^j(v) = -\frac{a_0}{a_j} v - \frac{a_1}{a_j} T(v) - \dots - \frac{a_{j-1}}{a_j} T^{j-1}(v)$$

Induction step: assume $T^j(v), T^{j+1}(v), \ldots, T^{N-1}(v)$ are all in span γ. We need to show $T^N(v) \in \text{span}(\gamma)$

But $T^{N-1}(v) = b_0 v + b_1 T(v) + \cdots + b_{j-1} T^{j-1}(v)$
~~Pla~~
by the induction hypothesis, so

$$T^N(v) = b_0 T(v) + b_1 T^2(v) + \cdots + b_{j-1} T^j(v)$$

$$= -\frac{b_{j-1} a_0}{a_j} v + \left(b_0 - \frac{b_{j-1} a_1}{a_j}\right) T(v) + \cdots$$

$$+ \left(b_{j-2} - \frac{b_{j-1} a_{j-1}}{a_j}\right) T^{j-1}(v)$$

here we used the fact that $T^j(v) = -a_0/a_j\, v - \cdots - \frac{a_{j-1}}{a_j} T^{j-1}(v)$

and so $T^N(v) \in \text{span}(\gamma)$

By induction, we are done. $\square$

This proves the Subclaim, which shows that γ spans W. So γ is a basis for W. But $\dim W = k$ and γ has ~~k~~ j elements, so $j = k$ (which is what we wanted). $\square$

Since W is T-invariant, the restriction T_W of T to W makes sense:

$$T_W : W \longrightarrow W$$
$$w \longmapsto T(w)$$

Let us compute the matrix $[T_W]_\beta^\beta$ of T_W with respect to the basis

$$\beta = \{v, T(v), \ldots, T^{k-1}(v)\}$$

Since β is a basis for W and $T^k(v) \in W$ we can find scalars $a_0, a_1, \ldots, a_{k-1}$ with

$$a_0 v + a_1 T(v) + \cdots + a_{k-1} T^{k-1}(v) + T^k(v) = 0$$

[warning: these aren't the same a_i as on pp 11–12]

The matrix $[T_W]_\beta^\beta$ is

$$[T_W]_\beta^\beta = \begin{array}{cc} & \overbrace{\hphantom{0 \quad 0 \quad \cdots \quad 0}}^{k-1} \\ k & \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & & 0 & -a_1 \\ 0 & 1 & & 0 & \cdot \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{k-1} \end{pmatrix} \end{array}$$

because e.g. for the first column:

$$T(v) = 0.v + 1.T(v) + 0.T^2(v) + \cdots$$

and for the last column

$$T(T^{n-1}(v)) = -a_0 v - a_1 T(v) - \cdots - a_{k-1} T^{k-1}(v)$$

This lets us compute the characteristic polynomial $f(t)$ of T_W:

<u>Claim</u> (cf Th$^{\underline{m}}$ 5.22 (b))

$$f(t) = (-)^k (a_0 + a_1 t + \cdots + a_{k-1} t^{k-1} + t^k)$$

Proof By induction on k

Base case: $k=1$

$$f(t) = |-t| = -t \quad \checkmark$$

Induction step: assume it is true for $k-1$.

Then:

$$f(t) = \begin{array}{c} \overbrace{\hphantom{-t \quad 0 \quad \cdots \quad 0}}^{k-1} \hphantom{\quad -a_{k-1}-t} \\ \begin{vmatrix} -t & 0 & \cdots & 0 & -a_0 \\ 1 & -t & & 0 & -a_1 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & 0 & \ddots & -t & \vdots \\ 0 & 0 & & 1 & -a_{k-1}-t \end{vmatrix} \end{array}$$

$$= (-t) \begin{array}{c} \overbrace{\hphantom{-t \quad 0 \quad \cdots \quad 0}}^{k-2} \hphantom{\quad -a_{k-1}-t} \\ \begin{vmatrix} -t & 0 & \cdots & 0 & -a_1 \\ 1 & -t & & 0 & -a_2 \\ 0 & 1 & & \vdots & \vdots \\ \vdots & 0 & & \vdots & \vdots \\ 0 & 0 & & -t & \\ & & & 1 & -a_{k-1}-t \end{vmatrix} \end{array}$$

this looks like exactly the same sort of determinant but now k is smaller by 1 and $a_0 \rightsquigarrow a_1$, $a_1 \rightsquigarrow a_2$ etc.

$$+ (-1)^k a_0 \begin{vmatrix} 1 & -t & 0 & \cdots & & 0 \\ 0 & 1 & -t & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & & & \vdots \\ \vdots & & & \ddots & & \vdots \\ & & & & \ddots & -t \\ 0 & \cdots & & & 0 & 1 \end{vmatrix}$$

expanding along the first row

$$= (-t)(-1)^{k-1}\underbrace{(a_1 + a_2 t + \dots + a_{k-1}t^{k-2} + t^{k-1})}_{\text{using the induction hypothesis}}$$

$$+ \quad (-1)^k a_0 \cdot 1$$

$$= (-1)^k (a_0 + a_1 t + a_2 t^2 + \dots + a_{k-1}t^{k-1} + t^k).$$

By induction, we are done. $\square$

Thinking about T-invariant subspaces can help in computing the characteristic polynomial of T:

<u>Theorem</u> (5.21) Let V be a finite-dim. vector space, $T: V \to V$ a linear map and W a T-invariant subspace of V. Let $f(t)$ be the characteristic polynomial of T and $g(t)$ be the characteristic polynomial of the restriction T_W of

T to W. Then $g(t)$ divides $f(t)$.

<u>Proof</u>: Pick a basis $\gamma = \{w_1, \dots, w_k\}$ for W and extend it to a basis $\beta = \{w_1, \dots, w_k, v_1, \dots, v_l\}$ for V. Then

$$[T]_\beta^\beta = \begin{array}{c} \\ k \\ l \end{array} \begin{array}{c} \begin{array}{cc} k & l \end{array} \\ \left(\begin{array}{c|c} A & B \\ \hline O & C \end{array}\right) \end{array}$$

where $A = [T_W]_\gamma^\gamma$.

Thus
$$f(t) = \det \left(\begin{array}{c|c} A - tI & B \\ \hline O & C - tI \end{array}\right)$$

$$= \det(A - tI)\,\det(C - tI)$$

$$= g(t) \cdot \det(C - tI)$$

and so $g(t)$ divides $f(t)$. $\square$

this step follows from the fact that $\det\left(\begin{array}{c|c} \tilde{A} & \tilde{B} \\ \hline 0 & \tilde{C} \end{array}\right) = (\det \tilde{A})(\det \tilde{C})$. One can see this by, for example, observing that

(a) if $\det \tilde{A} = 0$ then $N(\tilde{A}) \neq \{0\}$ so

$\exists\ v \in N(\tilde{A}),\ v \neq 0.$

But then
$$\left(\begin{array}{c|c} \tilde{A} & \tilde{B} \\ \hline 0 & \tilde{C} \end{array}\right)\begin{pmatrix} v \\ \hline 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow \quad N\left(\begin{array}{c|c} \tilde{A} & \tilde{B} \\ \hline 0 & \tilde{C} \end{array}\right) \neq \{0\}$$

$$\Rightarrow \quad \det\left(\begin{array}{c|c} \tilde{A} & \tilde{B} \\ \hline 0 & \tilde{C} \end{array}\right) = 0$$

and (b) if $\det \tilde{A} \neq 0$ then we can ~~pick our basis for ... such that A~~ change basis in W such that $\tilde{A}$ becomes upper-triangular (by e.g. midterm 2 Q4), and since none of the entries on the diagonal of $\tilde{A}$ are zero we can use column operations (adding scalar multiples of the first k columns to other columns) to get rid of $\tilde{B}$ and the ~~uppe~~ non-diagonal parts of $\tilde{A}$.

clear rows using column operations

$$\det\left(\begin{array}{c|c} \tilde{A} & \tilde{B} \\ \hline 0 & \tilde{C} \end{array}\right) = \det\left(\begin{array}{c|c} \begin{matrix} \lambda_1 & & \ast \\ & \ddots & \ast \\ 0 & & \lambda_k \end{matrix} & \tilde{B} \\ \hline 0 & \tilde{C} \end{array}\right)$$

$$= \det\left(\begin{array}{c|c} \begin{matrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_k \end{matrix} & 0 \\ \hline 0 & \tilde{C} \end{array}\right)$$

$$= \lambda_1 \lambda_2 \cdots \lambda_k \det(C)$$

$$= (\det \tilde{A})(\det \tilde{C}). \qquad \square$$

The textbook uses this result to prove the Cayley-Hamilton theorem:

<u>Thm</u> (5.23) Let $T: V \to V$ be a linear map, where V is a finite-dimensional vector space.

Let $f(t) = a_0 + a_1 t + \cdots + (-1)^n t^n$ be the characteristic polynomial of T. Then $f(T) = 0$, i.e.

$$a_0 I + a_1 T + \cdots + a_{n-1} T^{n-1} + (-)^n T^n$$

is the zero transformation.

Read their proof of this now; we will prove it a little later in the course, once we know about Jordan Canonical Form.