MATH 320 NOTES 4

5.4 Invariant Subspaces and Cayley-Hamilton theorem

The goal of this section is to prove the Cayley-Hamilton theorem:

Theorem 1. *Let $T : V \to V$ be a linear operator, V finite dimensional, and let $f(t)$ be the characteristic polynomial of T. Then $f(T) = T_0$ i.e. the zero linear transformation. In other words T is a root of its own characteristic polynomial.*

Here, if $f(t) = a_n t^n + a_{n-1}t^{n-1} + ... + a_1 t + a_0$, plugging in T means the transformation

$$f(T) = a_n T^n + a_{n-1}T^{n-1} + ... + a_1 T + a_0 I$$

Let us give some simple examples:

Example 1 The identity $I : F^3 \to F^3$ has characteristic polynomial $f(t) = (1-t)^3$. Then $f(I) = (I - I)^3 = T_0$.

Example 2 Let $A = \begin{pmatrix} 1 & 0 & 5 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. Then the characteristic polynomial is $f(t) = (1-t)^2(2-t)$, and $f(A) = (A-I)^2(2I_3 - A) = \begin{pmatrix} 0 & 0 & 5 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}^2 \begin{pmatrix} 1 & 0 & 5 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = O$.

We will prove the main theorem by using invariant subspaces and showing that if W is T-invariant, then the characteristic polynomial of $T \restriction W$ divides the characteristic polynomial of T. So, let us recall the definition of a T-invariant space:

Definition 2. *Given a linear transformation $T : V \to V$, a subspace $W \subset V$ is called* T**-invariant** *if for all $x \in W$, $T(x) \in W$. For such a W, let $T_W : W \to W$ denote the linear transformation obtained by restricting T to W i.e. for all $x \in W$, $T_W(x) = T(x) \in W$.*

Examples:

(1) V, $\{\vec{0}\}$,

(2) $\ker(T)$, $\text{ran}(T)$,

(3) E_λ for any eigenvalue λ for T.

Let us prove the last item: suppose that $v \in E_\lambda$. We have to show that $T(v) \in E_\lambda$. Denote $y = T(v)$ and compute

$$T(y) = T(T(v)) = T(\lambda v) = \lambda T(v) = \lambda y.$$

So, y is also an eigenvector for λ. Then $y = T(v) \in E_\lambda$ as desired.

Next we give another important example of an invariant subspace.

Lemma 3. *Suppose that $T : V \to V$ is a linear transformation, and let $x \in V$. Then*

$$W := Span(\{x, T(x), T^2(x), ...\})$$

is a T-invariant subspace. Moreover, if Z is any other T-invariant subspace that contains x, then $W \subset Z$.

Proof. First we show that W is T-invariant: let $y \in W$. We have to show that $T(y) \in W$. Since $y \in W$, by definition, for some natural number n, $y = T^n(x)$. Then $T(y) = T^{n+1}(x) \in W$.

Now suppose that Z is another T-invariant subspace with $x \in Z$.

Claim 4. *For every $n \geq 1$, $T^n(x) \in Z$.*

Proof. For the base case $n = 1$, since $x \in Z$ and Z is T-invariant, it follows that $T(x) \in Z$.

For the inductive case, suppose that $T^n(x) \in Z$. Then again, by T-invariance, we have that $T^{n+1}(x) \in Z$. □

By the claim, we get that $W \subset Z$.

□

W as above is called the **T-cyclic subspace of V generated by x.**

Example. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $T(\langle a, b, c\rangle) = \langle 2a, a + b, 0\rangle$. Find the T-cyclic subspace of V generated by e_1.

Solution:

- $T(e_1) = \langle 2, 1, 0\rangle$,
- $T^2(e_1) = T(\langle 2, 1, 0\rangle) = \langle 4, 3, 0\rangle$, and so on

Note that $T^2(e_1)$ is a linear combination of $e_1, T(e_1)$. Similarly, for any n, $T^n(e_1) = \langle a_1, a_2, 0\rangle$ for some a_1, a_2, and so it is a linear combination of e_1 and $T(e_1)$. It follows, that the T-cyclic subspace of V generated by e_1 is $Span(\{e_1, T(e_1)\}) = \{\langle a_1, a_2, 0\rangle \mid a_1, a_2 \in \mathbb{R}\} = Span(\{e_1, e_2\})$.

Our next lemma generalizes the above example:

Lemma 5. *Suppose that $T : V \to V$ is linear, let W be the T-invariant cyclic subspace generated by x (nonzero vector) with* $\dim(W) = k$. *Then $\{x, T(x), ..., T^{k-1}(x)\}$ is a basis for W*

Proof. Let m be the largest such that $\alpha = \{x, T(x), ..., T^{m-1}(x)\}$ is a linearly independent. Such m has to exists because W is finite dimensional. Then we have:

- $m \leq k$, since $\alpha \subset W$ and $\dim(W) = k$, and
- $T^m(x) \in Span(\alpha)$, by definition of m.

Let $Z = Span(\alpha)$. We claim that $Z = W$. We know that $Z \subset W$ because $\alpha \subset W$. For the other direction, by the second part of Lemma 3, it is enough to show that Z is T-invariant.

To that end, let $y \in Z$; write is a linear combination of the vectors in α,

$$y = a_1 x + a_2 T(x) + ... + a_m T^{m-1}(x).$$

Compute

$$T(y) = T(a_1 x + a_2 T(x) + ... + a_m T^{m-1}(x)) = a_1 T(x) + a_2 T^2(x) + ... + a_m T^m(x).$$

This is a linear combination of vectors in α and $T^m(x)$. Since $T^m(x) \in Span(\alpha)$, we get $T(y) \in Span(\alpha) = Z$.

Then α is a basis for W, and so $m = |\alpha| = k$. □

Before we prove that the characteristic polynomial of T_W divides the characteristic polynomial of T where W is T-invariant, we need the following fact.

Fact 6. *Suppose we have an $n \times n$ matrix B of the form*

$$B = \begin{pmatrix} A & C \\ 0 & D \end{pmatrix},$$

Where A is a $k \times k$ matrix. Then $\det(A) \cdot \det(D)$

Proof. The proof is by induction on k, expanding along the first column. □

Lemma 7. *Suppose that $T : V \to V$ is linear, V finite dimensions, and W is a T-invariant subspace. Let $T_W : W \to W$ be the linear transformation obtained by T restricted to W. Then the characteristic polynomial of T_W divides the characteristic polynomial of T.*

Proof. Let $\alpha = \{v_1, ..., v_k\}$ be a basis for W, and extend α to a basis $\beta = \{v_1, ...v_k, ...v_n\}$ for V. Let $A = [T_W]_\alpha$ and $B = [T]_\beta$. Then

$$B = \begin{pmatrix} A & C \\ 0 & D \end{pmatrix}$$

So,

$$(B - tI_n) = \begin{pmatrix} A - tI_k & C \\ 0 & D - tI_{n-k} \end{pmatrix}$$

Then $\det(B - tI_n) = \det(A - tI_n)g(t)$. Since the characteristic polynomial of T_W is $\det(A - tI_n)$ and the characteristic polynomial of T is $\det(B - tI_n)$, the result follows. □

Lemma 8. *Suppose that $T : V \to V$ is linear, V finite dimensional, and let W be the T-invariant cyclic subspace generated by x (nonzero vector) with* $\dim(W) = k$. *Then there are unique scalars $a_0, ..., a_{k-1}$, such that*

- $a_0 x + a_1 T(x) + ...a_{k-1} T^{k-1}(x) + T^k(x) = \vec{0}$, *and*

- *the characteristic polynomial of* T_W *is*

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

Proof. Recall that we proved $\alpha = \{x, T(x), ..., T^{k-1}(x)\}$ is a basis for W. That means that $T^k(x)$ is the span of these vectors and the linear combination is unique. So for some unique scalars $a_0, ..., a_{k-1}$, we get:

$$a_0 x + a_1 T(x) + ... + a_{k-1}T^{k-1}(x) + T^k(x) = \vec{0}.$$

Next we compute $[T_W]_\alpha$. We have:

- $T_W(x) = 0x + T(x) + 0... + 0$, so $[T_W(x)]_\alpha = \langle 0, 1, 0, ...0\rangle$,
- $T_W(T(x)) = T^2(x) = T_W^2(x)$, so $[T_W(x)]_\alpha = \langle 0, 0, 1, ...0\rangle$,
- ...
- $T_W(T^{k-1}(x)) = T^k(x) = -1(a_0 x + a_1 T(x) + ...a_{k-1}T^{k-1}(x))$, and so $[T_W(T^{k-1}(x))]_\alpha = \langle -a_0, -a_1, ... - a_{k-1}\rangle$,

Then

$$[T_W]_\alpha = \begin{pmatrix} 0 & 0 & ... & 0 & -a_1 \\ 1 & 0 & ... & 0 & -a_2 \\ \vdots & & & & \\ 0 & 0 & ... & 1 & -a_{k-1} \end{pmatrix}$$

One can verify that this matrix has characteristic polynomial

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

. □

We are finally ready to prove the Cayley-Hamilton theorem:

Proof of Theorem 1. We have to show that T "satisfies" its characteristic polynomial. So let $f(t)$ be the characteristic polynomial of T. Then we have to show that $f(T)$ is the zero linear transformation. So let $x \in V$. We have to show that $f(T)(x) = \vec{0}$.

Assume that x is nonzero (otherwise it's clear). Let W be the T-cyclic invariant subspace generated by x; $\dim(W) = k$. By the above lemma, fix coefficients $a_0, ..., a_k$, such that

- $a_0 x + a_1 T(x) + ...a_{k-1}T^{k-1}(x) + T^k(x) = \vec{0}$, and
- the characteristic polynomial of T_W is

$$g(t) = (-1)^k(a_0 + a_1 t + ...a_{k-1}t^{k-1} + t^k).$$

Then,

$$g(T)(x) = (-1)^k(a_0 I + a_1 T + ...a_{k-1}T^{k-1} + T^k)(x) =$$

$$(-1)^k(a_0 x + a_1 T(x) + ...a_{k-1}T^{k-1}(x) + T^k(x)) = \vec{0}.$$

And since g divides f, we get that $f(T)(x) = 0$.

□

6.1, 6.2 Inner Product Spaces and ONB

Let $c \in F$, we will use $\bar{c}$ to denote complex conjugation. I.e. if $c = a + bi$, then $\bar{c} = a - bi$. If $c \in \mathbb{R}$, then $\bar{c} = c$.

Definition 9. *Let V be a vector space over F. V is an* **inner product space***, if we can define an inner product function on pairs of vectors to values in F,*

$$(x, y) \mapsto \langle x, y\rangle \in F$$

with the following properties:

1. $\langle x + z, y\rangle = \langle x, y\rangle + \langle z, y\rangle$
2. $\langle cx, y\rangle = c\langle x, y\rangle$,
3. $\langle x, y\rangle = \overline{\langle y, x\rangle}$,
4. *if* $x \neq \vec{0}$, $\langle x, x\rangle > 0$,

Definition 10. *Let V be an inner product space. For $x \in V$, define the* **norm** *of x,*

$$||x|| = \sqrt{\langle x, x\rangle}.$$

Lemma 11. *(Properties of inner product spaces) Let V be an inner product space, and $x, y, z \in V$, $c \in F$. Then*

1. $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$
2. $\langle x, cy\rangle = \bar{c}\langle x, y\rangle$,
3. $\langle x, 0\rangle = \langle 0, x\rangle = 0$,
4. $\langle x, x\rangle = 0$ *iff if* $x = \vec{0}$,
5. *if* $\langle x, y\rangle = \langle x, z\rangle$ *for all x, then* $y = z$.

Lemma 12. *(Properties of norms) Let V be an inner product space over F, and $x, y, z \in V$, $c \in F$. Then*

1. $||cx|| = |c| \cdot ||x||$,
2. $||x|| \geq 0$, $||x|| = 0$ *iff if* $x = \vec{0}$,
3. *(Cauchy-Schwarz inequality)* $|\langle x, y\rangle| \leq ||x|| \cdot ||y||$,
4. *(Triangle inequality)* $||x + y|| \leq ||x|| + ||y||$.

Examples:

1. The usual dot product for $\mathbb{R}^n$ over $\mathbb{R}$: $\langle x, y\rangle = \Sigma_{1\leq i\leq n} x_i y_i$;
2. The conjugate dot product for $\mathbb{C}^n$ over $\mathbb{C}$: $\langle x, y\rangle = \Sigma_{1\leq i\leq n} x_i \bar{y}_i$;

Definition 13. *Let V be an inner product space. Two vectors x, y in V are called* **orthogonal** *(or perpendicular) if $\langle x, y\rangle = 0$.*

A vector x in V is called **normal** *if $||x|| = 0$.*

A set S is called **orthonormal** *if any two vectors from S are orthogonal and each $x \in S$ is normal.*

Lemma 14. *If S is a set of pairwise orthogonal nonzero vectors, then S is linearly independent. So, orthonormal sets are linearly independent.*

Proof. Suppose
$$a_1v_1 + ... + a_kv_k = \vec{0}$$
for vectors $v_1, ...v_k$ in S. Suppose for contradiction some of the a_i's are non zero. By rearranging if necessary, we may assume that $a_1 \neq 0$. Then

$$0 = \langle a_1v_1 + ... + a_kv_k, v_1\rangle = a_1\langle v_1, v_1\rangle + a_2\langle v_2, v_1\rangle + ...a_k\langle v_k, v_1\rangle = a_1\langle v_1, v_1\rangle.$$

By assumption $v_1 \neq \vec{0}$, and so by the properties of inner products $\langle v_1, v_1\rangle > 0$, so $a_1 = 0$. Contradiction □

Using a process called Gram-Schmidt orthogonalization, we can obtain an orthonormal set from any basis, and get the following theorem:

Theorem 15. *Let V be an inner product space of dimension $n > 0$. Then there is an orthonormal basis for V (ONB).*

Example: the standard basis is an ONB for F^n.

The next lemma illustrates the usefulness of ONBs, in the sense that we can determine in advance the coefficients of linear combinations and matrix representations.

Lemma 16. *If $\beta := \{v_1, ..., v_n\}$ is an ONB for V and $T : V \to V$ is linear, then*

(1) *For every $x \in V$,*
$$x = \langle x, v_1\rangle v_1 + \langle x, v_2\rangle v_2 + ... + \langle x, v_n\rangle v_n$$
(2) *If $A = [T]_\beta$, then the (i, j)-th entry of the matrix is $A_{ij} = \langle T(v_j), v_i\rangle$.*

Proof. We show the first part. The second is left as an exercise. Fix $x \in V$. Say $x = a_1v_1 + ... + a_nv_n$. Then for each $1 \leq i \leq n$,
$$\langle x, v_i\rangle = \langle a_1v_1 + ... + a_nv_n, v_1\rangle = \Sigma_k a_k\langle v_k, v_i\rangle = a_i\langle v_i, v_i\rangle = a_i.$$
The above is since for $k \neq i$, $\langle v_k, v_i\rangle = 0$ and $\langle v_i, v_i\rangle = 1$.

□

Definition 17. *Let $S \subset V$ be nonempty, V a inner product space. The* **orthogonal complement of** *S is*
$$S^\perp = \{x \in V \mid \langle x, y\rangle = 0 \text{ for all } y \in S\}$$

Fact 18. *For any nonempty set S, $S^\perp$ is a subspace.*

A couple of examples:

- $\{\vec{0}\}^\perp = V$; $V^\perp = \{\vec{0}\}$,
- in F^3, $\{e_1\}^\perp = Span(\{e_2, e_3\})$.

6.3, 6.4 Adjoint operators and the Spectral theorem

Let us first look at matrices:

Definition 19. *Let $A \in M_{m\times n}(F)$. The* **conjugate transpose** *of A is the $n \times m$ matrix A^* given by setting $A^*_{ij} = \bar{A}_{ji}$ i.e. we take the conjugate of each entry and transpose it. We also write $A^* = \bar{A}^t$.*

We can also define the adjoint for a linear transformations in general.

Theorem 20. *Let $T : V \to V$ be a linear operator on a finite dimensional inner product space. Then there is a linear transformation $T^* : V \to V$, such that for all $x, y \in V$,*

$$\langle T(x), y\rangle = \langle x, T^*(y)\rangle.$$

T^* is called the **adjoint** *of T.*

Moreover, if β is an ONB, then $[T^*]_\beta = [T]^*_\beta$.

The following is a key lemma:

Theorem 21. *(Schur) Suppose that the characteristic polynomial of T splits. Then there is an ONB β such that $[T]_\beta$ is upper triangular.*

Proof. (Outline) By induction on $\dim(V)$. The idea is to pick one eigenvalue (it exists since the characteristic polynomial splits), then a corresponding eigenvector z. Then prove that $W := Span(z)^\perp$ is T-invariant and apply the inductive hypothesis to T_W.

□

Next we give a condition for a diagonalizability over $\mathbb{R}$.

Definition 22. *A linear transformation $T : V \to V$ is* **self-adjoint (Hermitian)** *if $T = T^*$. Similarly, a matrix A is* **self-adjoint (Hermitian)** *if $A = A^*$.*

Lemma 23. *If T is self adjoint, then every eigenvalue is real and the characteristic polynomial splits over $\mathbb{R}$.*

Proof. Suppose that λ is an eigenvalue with eigenvector x. Then $\lambda\langle x, x\rangle = \langle\lambda x, x\rangle = \langle T(x), x\rangle = \langle x, T^*(x)\rangle = \langle x, T(x)\rangle = \langle x, \lambda x\rangle = \bar{\lambda}\langle x, x\rangle$. Then $\lambda = \bar{\lambda}$, and so it is real.

Let $f(t)$ be the characteristic polynomial. Then $f(t)$ splits over $\mathbb{C}$. But since every eigenvalue is real, every root of $f(t)$ is real, and so it must split over $\mathbb{R}$.

□

Theorem 24. *(Spectral theorems for real spaces) Suppose that $T : V \to V$ is linear, V is a finite dimensional real inner product space. Then T is self adjoint iff it has an ONB of eigenvectors.*

Proof. By the above lemma the characteristic polynomial of T splits. So by theorem 21, let β be ONB such that $A := [T]_\beta$ is upper triangular. Then $[T^*]_\beta = A^*$ has to be lower triangular. But since $A = A^*$, it follows that A is diagonal. So β is a basis of eigenvectors and T is diagonalizable.

□

In other words, symmetric real matrices are diagonalizable.

Below we briefly state (without proofs), the case of complex spaces.

Definition 25. *$T : V \to V$ is* **normal** *if $TT^* = T^*T$.*

Theorem 26. *(Spectral theorem for complex spaces) Suppose that $T : V \to V$ is linear, V is a finite dimensional complex inner product space. Then T is normal iff it has an ONB of eigenvectors.*