Chapter 1

Cayley Hamilton Theorem

We will build up the machinery needed to prove the following:

Theorem 1.0.1 (Cayley Hamilton Theorem). [thm:cayley] Given a linear operator $T : V \to V$ with $\dim(v) = n$, then

$$P_T(T) = \mathbb{0}_{n\times n}.$$

where P_T is the characteristic polynomial of T.

1.1 Invariant Subspaces

Definition 1.1.1 (Invariant Subspace). Given a linear operator $T : V \to V$ and subspace $W \subseteq V$, if $T[W] \subseteq W$ then

$$T|_W : W \to W$$

is a linear operator on W and W is T invariant.

Example 1.1.1. Consider some eigenvalue λ of T. Then there is a subspace E_λ of V associated with that eigenvalue. Taking any $v \in E_\lambda$, note that by definition $Tv = \lambda v \in E_\lambda$. Therefore E_λ is an invariant subspace for any eigenvalue λ of T.

Theorem 1.1.1. If $W \subseteq V$ is an invariant subspace under T, then for $T|_W : W \to W$ we have

$$P_{T|_W}(t) \mid P_T(t).$$

Proof. Let $\beta_w = \{w_1, \ldots, w_k\}$ be a basis of W and $\beta = \beta_w \cup \{v_{k+1}, \ldots, v_n\}$ be a basis of V where w_i form a basis of W. Then the matrix form of T is

$$[T]_\beta = \begin{pmatrix} B_1 & B_2 \\ 0 & B_3 \end{pmatrix}$$

where B_1 is $[T|_W]_{\beta_w}$. By subtracting tI from both sides and taking the determinant, we get

$$\det(T - tI) = \det(B_1 - tI)\det(B_3 - tI).$$

But this is just

$$P_T(t) = P_{T|_W}(t) \cdot q(t)$$

with $q(t) = \det(B_3 - tI)$. ∎

1.1.1 Generating Invariant Subspaces

Consider some linear operator T on a finite dimensional space V with $\dim V = n$. Then note for any $v \in V$ that

$$\{0, Tv, T^2v, \ldots\}$$

must be a linearly dependent set of vectors. If this wasn't the case, then repeated applications of T would produce infinitely many linearly independent vectors within V. Therefore there is some $k \leq n$ such that

$$\{0, Tv, T^2v, \ldots, T^{k-1}v\}$$

is linearly independent. The span of this set gives a subspace W that is T invariant, something analogous to cyclic groups in group theory. This motivates the following definition.

Definition 1.1.2 (Cyclic Subspace). Let T be a linear operator on V and $v \in V$. Then the subspace

$$W = \operatorname{span}\{0, Tv, T^2v, T^3v, \ldots\}$$

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

Example 1.1.2. Consider the operator $T : P(\mathbb{R}) \to P(\mathbb{R})$ with $T(p(x)) = p'(x)$. Starting with x^3, we see that

$$\{0, Tx^3, T^2x^3, \ldots\} = \{0, 3x^2, 6x, 6\}.$$

The span of this set then is then $P_3(\mathbb{R})$ which is invariant under T.

Theorem 1.1.2. If $a_0 + a_1Tv + a_2T^2v + \ldots + a_{k-1}T^{k-1}v + T^kv = 0$, then

$$P_{T|_W}(t) = (-1)^k\left(a_0 + a_1t + \ldots a_{k-1}t^{k-1} + t^k\right).$$

Proof. We consider the cyclic invariant subspace W spanned by the basis $\beta = \{v, Tv, T^2v, \ldots, T^{k-1}v\}$. If $w \in W$, then we know that

$$w = a_0v + a_1Tv + a_2T^2v + \ldots + a_kT^{k-1}v$$

which gives

$$Tw = a_0Tv + a_1T^2v + \ldots + a_kT^kv.$$

■

1.1.2 The Proof

thm:cayley. Let T be a linear operator on V, $v \in V \neq 0$, and W be the cyclic subspace generated by v. ■

Chapter 2

Inner Products and Norms

When working in $\mathbb{R}^n$, there is the familar idea of the scalar/dot product. Given two vectors $\mathbb{x}$ and $\mathbb{y}$ then their scalar product is

$$x \cdot y = \sum_{i=1}^{n} x_i y_i.$$

The concept of euclidean length is also captured by scalar products via

$$\sqrt{x \cdot x} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

This scalar product on $\mathbb{R}^n$ does not generalize to other vector spaces, or it may not be a useful notion of length/product of vectors even when working in $\mathbb{R}^n$. Therefore it is useful to generalize this notion of a scalar product.

Definition 2.0.1 (Inner Product). A mapping $\langle \cdot, \cdot \rangle : V \times V \to F$ is an inner product if for all $x, y \in V$ and $s \in F$

1. $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$ for all $z \in V$
2. $\langle sx, y \rangle = s\langle x, y \rangle$
3. $\overline{\langle x, y \rangle} = \langle y, x \rangle$
4. $\langle x, x \rangle > 0$ when $x \neq 0$

Example 2.0.1. The vector space $M_{n\times n}(\mathbb{R})$ of real n by n matrices can be endowed with an inner product where $\langle A, B \rangle = \operatorname{tr} B^t A$.

Example 2.0.2. The vector space $C([0, 2\pi])$ of continuous complex functions on the interval 0 to 2π can be endowed with an inner product where

$$\langle f, g \rangle = \int_0^{2\pi} f(t)\overline{g(t)} \mathrm{d}t.$$

An important concept that can be generalized from $\mathbb{R}^n$ is orthogonality. It is common to compare the scalar product of two vectors to 0 to determine if they are orthogonal or not. This motivates a generalized notion of orthogonality.

Definition 2.0.2 (Orthogonal Vectors). Let V be a vector space equipped with an inner product $\langle \cdot, \cdot \rangle$. Then $x, y \in V$ are orthogonal if $\langle x, y \rangle = 0$.

Example 2.0.3. Consider from 2.0.2 the family of functions $f_m(t) = e^{imt}$. Then for any f_m, f_n

$$\begin{aligned}\langle f_m, f_n\rangle &= \int_0^{2\pi} f_m(\tau)\overline{f_n(\tau)}\mathrm{d}\tau \\ &= \int_0^{2\pi} e^{i(m-n)\tau}\mathrm{d}\tau \\ &= \left.\frac{e^{i(m-n)\tau}}{i(m-n)}\right|_0^{2\pi} = 0.\end{aligned}$$

Hence all f_m are orthogonal to each other.

Definition 2.0.3 (Vector Norm). Let V be a vector space equipped with an inner product $\langle\cdot,\cdot\rangle$. Then the **norm or length** of x is

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

Theorem 2.0.1 (Cauchy-Schwarz Inequality). For any vector space V with an inner product $\langle\cdot,\cdot\rangle$ and $x, y \in V$,

$$|\langle x, y\rangle| \leq \|x\|\|y\|.$$

Proof. ■

The triangle inequality then follows quickly from Cauchy-Schwarz.

Theorem 2.0.2 (Triangle Inequality). For any vector space V with an inner product $\langle\cdot,\cdot\rangle$ and $x, y \in V$,

$$\|x + y\| \leq \|x\| + \|y\|.$$

This means that for any inner product on a space, it has its own version of a triangle inequality. This offloads the burden of proving directly that a norm satisfies the triangle inequality to finding some notion of an inner product that gives rise to that norm.

2.1 A General Notion of Norms

It is important to note that while every inner product gives rise to a norm, not every norm can be reverse engineered into a inner product.

Example 2.1.1. On the vector space $M_{n\times n}(\mathbb{R})$,

$$\|A\| = \sup_{\substack{x\in\mathbb{R}^n \\ \|x\|\leq 1}} Ax$$

defines a norm, but there exists no inner product that gives rise to it.

Definition 2.1.1 (Generalized Norm). Let V be a vector space. Then a norm $\|\cdot\| : V \to \mathbb{R}$ satisfies $\forall x, y \in V$ and $s \in C$

1. $\|x\| \geq 0$ and $\|x\| = 0 \Leftrightarrow x = 0$
2. $\|sx\| = |s|\|x\|$
3. $\|x + y\| \leq \|x\| + \|y\|$ $(\star)$

Example 2.1.2. The map

$$\|x\|_\infty := \max_{i \in \{1,\dots,n\}} |x_i|$$

on $\mathbb{R}^n$ is a norm. Consider the requirements to be a norm

1. Since the norm takes the maximum of the absolute value of each component, the norm will be a non negative result, meaning $\|x\|_\infty \geq 0$. If the norm is 0, then the largest term in magnitude was 0, hence $x = 0$. The reverse follows easily.

2. With $s \in C$

$$\begin{aligned} \|sx\|_\infty &= \max_{i \in \{1,\dots,n\}} |sx_i| \\ &= |s| \max_{i \in \{1,\dots,n\}} |x_i| \\ &= |s| \|x\|_\infty. \end{aligned}$$

3. The triangle inequality follows from the triangle inequality on the reals and the linearity of the maximum function.

There is a famous and important class of norms defined on euclidean space known as the p-norms. They give rise to L^p spaces which are crucial to functional analysis.

Definition 2.1.2 (L_p Norm). Given $p \in \mathbb{N}$, the map

$$L_p(x) := \sum_i \left(|x_i|^p\right)^{\frac{1}{p}}$$

is a norm for any $\mathbb{R}^n$.

Chapter 3

Orthogonality

When a vector space has an inner product, there is a notion of orthgonality as was defined in Orthogonal Vectors. Orthogonality of vectors tends to make computations and proofs simpler, hence building and working in an orthogonal basis is advantageous. Imposing normality of the basis further improves the situation.

Definition 3.0.1 (Orthogonal Basis). A basis $\beta = \{v_1, \ldots, v_n\}$ of a vector space V with inner product $\langle \cdot, \cdot \rangle$ is an **orthonormal basis** if $\|v_i\| = 1$ and $\langle v_i, v_j \rangle = 0$ for $i \neq j$.

Theorem 3.0.1. Suppose $\beta = \{v_1, \ldots, v_n\}$ is an orthonormal basis of some vector space V. Then for any $x \in V$

$$x = \sum_i \langle x, v_i \rangle v_i.$$

Proof. Since β is a basis, $x \in V$ can be written as

$$x = \sum_i a_i v_i$$

for scalars a_i. Then note

$$\begin{aligned} \langle x, v_i \rangle = \left\langle \sum_j a_j v_j, v_i \right\rangle &= \sum_j \langle a_j v_j, v_i \rangle \\ &= \sum_j a_j \langle v_j, v_i \rangle \\ &= a_i \langle v_i, v_i \rangle \\ &= a_i \end{aligned}$$

Substituting the expression for each a_i gives the desired result. ∎

Theorem 3.0.2. Any set of non-zero orthogonal vectors is linearly independent.

Proof. Let $\{v_1, \ldots, v_k\}$ be a set of orthogonal vectors with $v_i \neq 0$. Assume towards contradiction that this set is not linearly independent. Then there exists scalars a_i such that

$$\sum_i a_i v_i = 0.$$

Therefore at least one a_i is non-zero. Note that for any v_j

$$\left\langle \sum_i a_i v_i, v_j \right\rangle = a_j \|v_j\|^2$$

from the previous proof. But at the same time

$$\left\langle \sum_i a_i v_i, v_j \right\rangle = \langle 0, v_j \rangle = 0$$

meaning $a_j \|v_j\|^2 = 0$. Since v_j is non-zero, then $a_j = 0$. However, this is true for any j meaning all a_i must be zero, a contradiction. ∎

Theorem 3.0.3 (Grahm-Schmidt). Let V be an inner product space and $S = \{w_1, \ldots, w_n\}$ a set of linearly independent vectors. Then the set $\tilde{S} = \{v_1, \ldots, v_n\}$ where

$$v_1 = w_1 \qquad v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j \rangle}{\|v_j\|^2} v_j$$

is an orthogonal set with $\operatorname{span} S = \operatorname{span} \tilde{S}$.

Proof. We induct on the number of vectors in S. The base case is trivial for $n = 1$. Let $k \in \mathbb{N}$ and assume that $\tilde{S}_{k-1}$ can be constructed. If S has k linearly indpenedent vectors, then we can consider just $k - 1$ of them. By the induction hypothesis, we can produce $\tilde{S}_{k-1}$ that is an orthogonal set.

∎

Corollary 3.0.1. Every basis of an inner product space can be turned into an orthonormal basis.

Definition 3.0.2 (Orthogonal Complement). Given a set of vectors S in an inner product space V, the orthogonal complement is

$$S^\perp := \{v \in V : \langle v, s \rangle = 0 \text{ for all } s \in S\}.$$

Theorem 3.0.4 (Orthogonal Decomposition). Let $W \subseteq V$ be a subspace. Given $y \in V$, there is a unique $w \in W$ and $z \in W^\perp$ such that $y = w + z$. Equivalently, $V = W \oplus W^\perp$.

Proof. Let $\beta = \{v_1, \ldots, v_n\}$ be an orthonormal basis of W and $k = \dim W$. Note that

$$w = \sum_i \langle v_i, w \rangle v_i.$$

Let $z = y - w$. Note that

$$\begin{aligned} \langle z, v_j \rangle &= \left\langle y - \sum_i \langle v_i, w \rangle v_i, v_j \right\rangle \\ &= \langle y, v_j \rangle - \langle y, v_j \rangle \langle v_j, v_j \rangle \\ &= \langle y, v_j \rangle - \langle y, v_j \rangle = 0. \end{aligned}$$

Therefore z is orthogonal to all the vectors in β. It is orthogonal to every vector in W since taking $v \in W$ gives

$$\langle z, v \rangle = \left\langle z, \sum_i \langle v, v_i \rangle v_i \right\rangle = \sum_i \langle \langle v, v_i \rangle v_i, z \rangle = 0.$$

Chapter 4

Spectral Theorem

Lemma 4.0.1. Let T be a linear operator on a finite dimensional inner product space V. If T has an eigenvector, then T^* does as well.

Proof. Suppose v is an eigenvector T with eigenvalue λ. Then for any $x \in V$

$$0 = \langle 0, x\rangle = \langle (T-\lambda I)v, x\rangle = \langle v, (T-\lambda I)^* x\rangle = \left\langle v, (T^* - \overline{\lambda} I)x\right\rangle.$$

Therefore v is orthogonal to the range of $T^* - \lambda I$ ■

Theorem 4.0.1 (Schur's Theorem). Let T be a linear operator on V. Suppose that the characteristic polynomial of T splits. Then there exists an orthonormal basis β for V such that $[T]_\beta$ is upper triangular.

Proof. We proceed with induction on the dimension n of V. If $n = 1$, then $[T]_\beta$ is a single entry and therefore is upper triangular. Assume the statement is true for an $n-1$ dimensional space. By lemma 4.0.1, T^* has a unit eigenvector z with some eigenvalue λ. Let $W = \text{span}\{z\}$. Note that for any $y \in W^\perp$ and $x = cz \in W$

$$\begin{aligned}
\langle Ty, x\rangle &= \langle Ty, cz\rangle \\
&= \langle y, T^* cz\rangle \\
&= \langle y, cT^* z\rangle \\
&= \langle y, c\lambda z\rangle \\
&= \overline{c\lambda}\langle y, z\rangle = 0.
\end{aligned}$$

Therefore $Ty \in W^\perp$ meaning $W^\perp$ is T-invariant, meaning the characteristic polynomial of $T|_{W^\perp}$ divides the one of T. Since the characteristic polynomial of T splits, $T|_{W^\perp}$ must also split. Additionally, $\dim W^\perp = n-1$ since $\dim W = 1$. Therefore the induction hypothesis can be applied to $W^\perp$ to get an orthonormal basis γ such that $[T|_{W^\perp}]_\gamma$ is upper triangular. Choosing $\beta = \gamma \cup \{z\}$ gives an orthonormal basis for V since z is a unit vector and z is from W and hence orthogonal to the basis γ of $W^\perp$. It is clear then the matrix $[T]_\beta$ is upper triangular. ■

4.1 Normal Operators

Definition 4.1.1 (Normal Operator). Let V be an inner product space. A linear operator T (or matrix A) is **normal** if $TT^* = T^*T$ (or $AA^* = A^*A$).

Theorem 4.1.1. If T is a normal operator on a complex vector space V, then there exists an orthonormal basis of V consisting of eigenvectors of T.

Proof. Since V is a complex vector space, the characteristic polynomial of T splits by the fundamental theorem of algebra. Therefore by 4.0.1 there exists an orthonormal basis β with $[T]_\beta$ upper triangular. Since $[T]_\beta$ is upper triangular, the first basis vector v_1 is an eigenvector of T. We can then consider an induction argument over the basis vectors to show that all are eigenvectors. Assume that $\{v_1, \ldots, v_{k-1}\}$ are eigenvectors of T. Note that $T^* v_j = \overline{\lambda_j} v_j$ for any $j < k$. Since $[T]_\beta$ is upper triangular,

$$Tv_k = A_{1k}v_1 + A_{2k}v_2 + \ldots + A_{kk}v_k.$$

But note that

$$A_{jk} = \langle Tv_k, v_j \rangle = \langle v_k, T^* v_j \rangle = \left\langle v_k, \overline{\lambda_j} v_j \right\rangle = \overline{\lambda} \langle v_k, v_k \rangle = 0$$

for $j < k$ since v_k and v_j come from an orthonormal basis. Therefore

$$Tv_k = A_{kk}v_k$$

and $A_{kk} \neq 0$ meaning v_k must be an eigenvector of T. Therefore β is an orthonormal basis consisting of eigenvectors of T. ■

Definition 4.1.2 (Self Adjoint). A linear operator T on a space V (or square matrix A) is **self-adjoint** if $T = T^*$ (or $A = A^*$)

$$
\begin{aligned}
AQ = \begin{pmatrix} — & Ae_1 & — \\ & \vdots & \\ — & Ae_n & — \end{pmatrix}\begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix} &= \begin{pmatrix} \langle Ae_1, v_1\rangle & \dots & \langle Ae_1, v_n\rangle \\ \vdots & & \vdots \\ \langle Ae_n, v_1\rangle & \dots & \langle Ae_n, v_n\rangle \end{pmatrix} \\
&= \begin{pmatrix} \lambda_1 \langle v_1, e_1\rangle & \cdots & \lambda_n \langle v_n, e_1\rangle \\ \vdots & & \vdots \\ \lambda_1 \langle v_1, e_n\rangle & \cdots & \lambda_n \langle v_n, e_n\rangle \end{pmatrix} \qquad (\star) \\
&= \begin{pmatrix} | & & | \\ \lambda_1 v_1 & \cdots & \lambda_n v_n \\ | & & | \end{pmatrix} \\
&= QD
\end{aligned}
$$

Theorem 4.1.2. A linear operator $T : V \to V$ is an orthogonal projection if and only if T^* exists and $T^2 = T^* = T$.

Proof. We consider both directions.

$\Rightarrow$) Assume that T is an orthogonal projection. Then note that $V = R(T) \oplus N(T)$. Therefore for any $x, y \in V$, there are decompositions

$$
\begin{aligned}
x &= x_R + x_N \\
y &= y_R + y_N
\end{aligned}
$$

where x_R, y_R are from the range and x_N, y_N are from the null space. Note then

$$\langle x, Ty\rangle = \langle x_R + x_N, y_R\rangle = \langle x_R, y_R\rangle + \cancelto{0}{\langle x_N, y_R\rangle} = \langle x_R, y_R\rangle .$$

By similar logic, $\langle Tx, y\rangle = \langle x_R, y_R\rangle$. Therefore $\langle x, Ty\rangle = \langle Tx, y\rangle = \langle x, T^* y\rangle$ meaning $T = T^*$.

$\Leftarrow$) Suppose $T^2 = T = T^*$. Take $x \in R(T)$ and $y \in R(T)$. By definition of the range, we have $x = Tx'$. But then $T^*x = Tx = T^2x' = Tx' = x$. Therefore

$$\langle x, y\rangle = \langle T^* x, y\rangle = \langle x, Ty\rangle = \langle x, 0\rangle = 0$$

meaning $x \in R(T)^\perp$.

■

Theorem 4.1.3 (Spectral Theorem). Let $T : V \to V$ be a linear operator over a finite dimensional inner product space with underlying field $\mathbb{C}$ or $\mathbb{R}$. Assume T has eigenvalues $\lambda_1, \dots, \lambda_n$ and that T is normal/self-adjoint. Let W_i be the ith eigenspace and T_i the orthogonal projection onto W_i. Then all of the following are true

a) $V = \bigoplus_{i=1}^{n} W_i$

b) $W_i^\perp = \bigoplus_{j \neq i} W_j$

c) $T_i T_j = \delta_{ij} T_i$

d) $I = \sum_{i=1}^{n} T_i$

e) $T = \sum_{i=1}^{n} \lambda_i T_i$