MATH 245 - Linear Algebra (Advanced)

Fantastic Theorems and How to Prove Them

Jose Luis Avilez
Faculty of Mathematics
University of Waterloo

Preface

The following is the first edition of a set of notes produced during MATH 245 in the Fall of 2017 at the University of Waterloo. They were produced by combining the proofs provided in the lectures and the supplementary textbook, *Linear Algebra* by Friedberg, Insel, and Spence.

In each chapter you will find that I have outlined the essential structure required to pass the course's final exam. The theorems you will find in the latter part of each chapter are usually exercises from the book which may be harder to prove off-the-cuff under exam conditions.

Since this is the first edition of such notes, you may find that some of the typos, errors, or jokes, may not be of your liking. I take full responsibility for the first two; however, if you don't like the jokes, then you have bigger problems than linear algebra.

Contents

Part I

MATH 146 Theorems

Here I shall list, without proof, all those little theorems from MATH 146 that you wished you remembered. I shall also include cool facts that we have proven which, on the off chance, may be useful in McKinnon's final exam.

Definition 0.1. A matrix $A \in M_{m\times n}(\mathbb{F})$ is the rectangular array:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

Yes, I know, it's embarrassing not to remember the order m and n come in. Oops.

Theorem 0.2. *Let V be a vector space and W be a subset of V. Then W is a subspace if and only if:*

1. *It contains the zero vector.*
2. *It is closed under vector addition.*
3. *It is closed under scalar multiplication*

Theorem 0.3. *The intersection of subspaces of V is a subspace of V.* **Remark***: In general, the union of subspaces is not a subspace.*

Theorem 0.4. *Let W_1 and W_2 be subspaces of V. $W_1 \cup W_2$ is a subspace of V if and only if $W_1 \subseteq W_2$ or $W_2 \subseteq W_1$.*

Definition 0.5. If S_1 and S_2 are nonempty subsets of a vector space V then the **sum** of them, denoted $S_1 + S_2 = \{x + y : x \in S_1 \text{ and } y \in S_2\}$

Definition 0.6. A vector space V is called the **direct sum** of W_1 and W_2 if W_1, W_2 are subspaces of V such that $W_1 \cap W_2 = \{0\}$ and $W_1 + W_2 = V$. We denote that V is the direct sum of these two subspaces by writing $V = W_1 \oplus W_2$. **Remark:** In class, the so called "big O plus" has been used to denote Cartesian product with trivial intersection. Mutatis mutandis, they are the same thing.

Definition 0.7. Let W be a subspace of V. Then, the **coset** of W containing v is the set $v+W = \{v+w : w \in W\}$ for some $v \in V$. The quotient space of V modulo W is then then $V/W = \{v + W : v \in V\}$.

Theorem 0.8. $\dim(V/W) = \dim V - \dim W$

Theorem 0.9. $\text{span}(S_1 \cup S_2) = \text{span}(S_1) + \text{span}(S_2)$

Theorem 0.10. $\text{span}(S_1 \cap S_2) \subseteq \text{span}(S_1) \cap \text{span}(S_2)$

Theorem 0.11. *Any superset of a linearly dependent set is linearly dependent. Equivalently, by contrapositive, any subset of a linearly independent set is linearly independent.*

Theorem 0.12. *Every finite-dimensional vector space has a basis. Additionally, if you believe in the axiom of choice, every vector space has a basis.*

Theorem 0.13. *If W_1, W_2 are finite-dimensional subspaces of a vector space V, then*

$$\dim(W_1 + W_2) = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2)$$

Moreover, V is the direct sum of W_1 and W_2 if and only if

$$\dim(W_1 + W_2) = \dim W_1 + \dim W_2$$

Definition 0.14. The null space (or kernel) of a linear transformation T is the set $N(T) = \{x \in V : T(x) = 0\}$. The range (or image) of T is the set $R(T) = \{T(x) : x \in V\}$.

Definition 0.15. Let V and W be vector spaces and let $T : V \to W$ be a linear transformation. Then $N(T)$ and $R(T)$ are subspaces of V and W, respectively.

Theorem 0.16. *Let V and W be vector spaces and let $T : V \to W$ be a linear transformation. If $\beta = \{v_1, \dots, v_n\}$ is a basis for V then,*

$$R(T) = \operatorname{span}(T(\beta))$$

Theorem 0.17. Rank-nullity theorem. *Let V and W be vector spaces and let $T : V \to W$ be a linear transformation. If V is finite-dimensional then:*

$$\operatorname{nullity}(T) + \operatorname{rank}(T) = \dim V$$

Theorem 0.18. *Let V and W be vector spaces and let $T : V \to W$ be a linear transformation. Then T is injective if and only if $N(T) = \{0\}$.*

Theorem 0.19. *Let V and W be vector spaces of* **equal dimension** *and let $T : V \to W$ be a linear transformation. Then, the following are equivalent:*

1. *T is injective*
2. *T is surjective*
3. $\operatorname{rank}(T) = \dim V$

Theorem 0.20. *If two linear transformations agree on the basis then they are equal.*

Theorem 0.21. *Let V and W be vector spaces and let $T : V \to W$ be a linear transformation. Then T is injective if and only if it carries linearly independent subsets of V onto linearly independent subsets of W.*

Definition 0.22. Let V be a vector space and let $T : V \to V$ be a linear transformation. A subspace of W is said to be T**-invariant** if $T(x) \in W$ for every $x \in W$. Namely, $T(W) \subseteq W$. If W is T-invariant, then we define the **restriction** of T on W to be the function $T_W : W \to W$ defined by $T_W(x) = T(x)$ for all $x \in W$.

Theorem 0.23. *Let V be an n-dimensional vector space, and let $T : V \to V$ be a linear transformation. Suppose that W is a T-invariant subspace of V having dimension k. Then there exists an ordered basis β such that $[T]_\beta$ has the block-matrix form:*

$$[T]_\beta = \begin{bmatrix} A & B \\ O & C \end{bmatrix}$$

where A is a $k \times k$ matrix and O is the $(n-k) \times k$ zero matrix.

Definition 0.24. Let A be an $m \times n$ matrix with entries from a field $\mathbb{F}$. We denote by L_A the mapping $L_A : \mathbb{F}^n \to \mathbb{F}^m$ defined by $L_A(x) = Ax$ (the matrix product of A and x) for each column vector $x \in \mathbb{F}^n$. We call L_A a left-multiplication transformation.

Theorem 0.25. *Let V be a vector space and $T : V \to V$ be linear. Then $T^2 = T_0$ if and only if $R(T) \subseteq N(T)$.*

Theorem 0.26. $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ *and* $\operatorname{tr}(A^t) = \operatorname{tr}(A)$

Theorem 0.27. *Let T be a linear operator on V. If $\operatorname{rank}(T) = \operatorname{rank}(T^2)$ then $R(T) \cap N(T) = \{0\}$ and $V = R(T) \oplus N(T)$.*

Theorem 0.28. *Two vector spaces over the same field are isomorphic if and only if their dimensions are the same.*

Theorem 0.29. *About matrix inverses (provided the matrices are invertible):*

1. $(AB)^{-1} = B^{-1}A^{-1}$
2. $(A^t)^{-1} = (A^{-1})^t$

Theorem 0.30. *Let V and W be finite-dimensional vector spaces and $T : V \to W$ be an isomorphism. Let V_0 be a subspace of V. Then $T(V_0)$ is a subspace of W and $\dim V_0 = \dim(T(V_0))$.*

Theorem 0.31. *Let β and β' be two ordered bases for a finite-dimensional vector space V and let $Q = [I_V]_{\beta}^{\beta'}$ (namely, Q is the change of basis matrix). Then:*

1. *Q is invertible.*
2. *For any $v \in V$, $[v]_{\beta'} = Q[v]_{\beta}$*

We say that Q changes β coordinates into β' coordinates.

Theorem 0.32. *Let T be a linear operator on a finite-dimensional vector space V, and let β and β' be ordered bases for V. Suppose that Q is the change of coordinate matrix from β to β' coordinates. Then*

$$[T]_{\beta} = Q^{-1}[T]_{\beta'}Q$$

Definition 0.33. Let $A, B \in M_{n\times}(\mathbb{F})$. Then we say that B is similar to A if there exists and invertible matrix Q such that $B = Q^{-1}AQ$.

Definition 0.34. Let V be a vector space over a field $\mathbb{F}$. A **linear functional** is a linear transformation $f : V \to F$.

Definition 0.35. The dual space of V is the vector space of all linear functionals on V and is denoted V^*.

Theorem 0.36. $\dim V = \dim(V^*)$ *and* $\dim V = \dim(V^{**})$

Theorem 0.37. *Multiplication by an invertible matrix preserves the rank of the original matrix.*

Theorem 0.38. $\operatorname{rank}(A) = \operatorname{rank}(A^t)$

Theorem 0.39. *Let $T : V \to W$ and $U : W \to Z$ be linear transformations on finite dimensional vector spaces and let A and B be matrices such that the product AB is defined. Then:*

1. $\operatorname{rank}(UT) \leq \operatorname{rank}(U)$
2. $\operatorname{rank}(UT) \leq \operatorname{rank}(T)$
3. $\operatorname{rank}(AB) \leq \operatorname{rank}(A)$
4. $\operatorname{rank}(AB) \leq \operatorname{rank}(B)$

Theorem 0.40. *Let $T, U : V \to W$ be linear. Then $R(T+U) \subseteq R(T) + R(U)$ and* $\operatorname{rank}(T+U) \leq \operatorname{rank}T + \operatorname{rank}U$

Definition 0.41. Let $A \in M_{n\times n}(\mathbb{F})$. If $n = 1$ so that $A = (A_{11})$ we define the determinant of A to be $\det(A) = A_{11}$. For $n \geq 2$, we define the determinant recursively as:

$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} A_{1j} \cdot \det(\widetilde{A}_{1j})$$

In fact, the determinant can be computed by cofactor expansion along any row or column.

Theorem 0.42. *The determinant is an n-linear function.*

Theorem 0.43. $\det(AB) = \det(A) \cdot \det(B)$

Theorem 0.44. $\det(A) = \det(A^t)$

Theorem 0.45. *The determinant of an upper triangular matrix is the product of its diagonal entries.*

Theorem 0.46. *If A and B are similar then* $\det(A) = \det(B)$

Theorem 0.47. *Let* $M = \begin{bmatrix} A & B \\ O & C \end{bmatrix}$ *be a block matrix with* A *and* C *being square matrices. Then* $\det(M) = \det(A) \cdot \det(C)$.

Theorem 0.48. *The Vandermonde matrix*

$$V = \begin{bmatrix} 1 & c_0 & c_0^2 & \dots & c_0^n \\ 1 & c_1 & c_1^2 & \dots & c_1^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & c_n & c_n^2 & \dots & c_n^n \end{bmatrix}$$

has determinant

$$\det(V) = \prod_{0 \leq i < j \leq n} (c_j - c_i)$$

Definition 0.49. A linear operator T on a finite-dimensional vector space V is said to be diagonalisable if there exists an ordered basis such that $[T]_\beta$ is diagonal.

Theorem 0.50. *A linear operator* T *on* V *is diagonalisable if and only if there exists a basis for* V *consisting of eigenvectors of* T.

Definition 0.51. The characteristic polynomial of a matrix A is $f(t) = \det(A - tI_n)$.

Theorem 0.52. *Let* T *be a linear operator on* V *such that the characteristic polynomial of* T *splits over the field. Then* T *is diagonalisable if and only if the algebraic multiplicity of each eigenvalue equals the geometric multiplicity of its corresponding eigenspace. Additionally, if* T *is diagonalisable, the union of eigenbases is a basis for* V.

Theorem 0.53. $\mathrm{tr}(A) = \sum_i m_i \lambda_i$ *and* $\det(A) = \lambda_1^{m_1} \dots \lambda_k^{m_k}$ *where* m_i *is the multiplicity of each eigenvalue.*

Theorem 0.54. *If* T *is diagonalisable,* T^{-1} *is diagonalisable.*

Part II

Inner Product Spaces

Chapter 1

Inner Products and Norms

Definition 1.1. Let V be a vector space over $\mathbb{F} = \mathbb{C}$ or $\mathbb{R}$. An **inner product** on V is a function that assigns to every ordered pair of vectors $x, y \in V$ a scalar, denoted $\langle x, y\rangle$, such that the following hold:

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. $\overline{\langle x, y\rangle} = \langle y, x\rangle$
4. $\langle x, x\rangle > 0$ if $x > 0$

Definition 1.2. The **norm** of a vector $x \in V$ is the scalar $\|x\| = \sqrt{\langle x, x\rangle}$

Theorem 1.3. *Let β be a basis for a finite-dimensional inner product space. If $\langle x, z\rangle = 0$ for all $z \in \beta$, then $x = 0$*

Proof. Write $x = a_1 z_1 + \ldots + a_n z_n$. Then,

$$\begin{aligned}\langle x, x\rangle &= \langle x, a_1 z_1 + \ldots + a_n z_n\rangle \\ &= \sum_{i=1}^{n} \overline{a_i}\langle x, z_i\rangle \\ &= 0\end{aligned}$$

Hence, $x = 0$. ∎

Theorem 1.4. *Let β be a basis for a finite-dimensional inner product space. If $\langle x, z\rangle = \langle y, z\rangle$ for all $z \in \beta$ then $x = y$.*

Proof. Since $\langle x, z\rangle = \langle y, z\rangle$, it follows that $\langle x - y, z\rangle = 0$. By Theorem 1.3, $x - y = 0$ and $x = y$. ∎

Theorem 1.5. *If V is an inner product space, then $|\langle x, y\rangle| = \|x\|\,\|y\|$ if and only if one of the vectors x, y is a multiple of the other.*

Proof. The result follows trivially if we consider $z = x - ay$, and let

$$a = \frac{\langle x, y\rangle}{\|y\|^2}$$

∎

Theorem 1.6. *Let V be an inner product space and suppose x, y are orthogonal vectors in V. Then $\|x + y\|^2 = \|x\|^2 + \|y\|^2$. (Note, this is the Pythagorean Theorem in $\mathbb{R}^2$ using the Euclidean Norm.)*

Proof. Since x and y are orthogonal, $\langle x, y\rangle = \langle y, x\rangle = 0$. We have the following,

$$\begin{aligned}\|x + y\|^2 &= \langle x + y, x + y\rangle \\ &= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\ &= \|x\|^2 + \|y\|^2\end{aligned}$$

∎

Theorem 1.7. *If V is an inner product space, then for all $x, y \in V$,*

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$$

This is called the parallelogram law.

Proof. We proceed by performing some Mickey Mouse algebra on the norm,

$$\begin{aligned}
\|x+y\|^2 + \|x-y\|^2 &= \langle x+y, x+y\rangle + \langle x-y, x-y\rangle \\
&= \|x\|^2 + \langle x,y\rangle + \langle y,x\rangle + \|y\|^2 + \|x\|^2 - \langle y,x\rangle - \langle x,y\rangle + \|y\|^2 \\
&= 2\|x\|^2 + 2\|y\|^2
\end{aligned}$$

■

Theorem 1.8. *Let $S = \{v_1, \dots, v_k\}$ be an orthogonal set in V, and let $a_1, \dots, a_k$ be scalars. Then,*

$$\left\|\sum_{i=1}^{k} a_i v_i\right\|^2 = \sum_{i=1}^{k} |a_i|^2 \|v_i\|^2$$

Proof. Since S is orthogonal, $\langle v_i, v_j\rangle = 0$ for $i \neq j$. Thus, the algebra works out neatly as follows,

$$\begin{aligned}
\left\|\sum_{i=1}^{k} a_i v_i\right\|^2 &= \langle \sum_{i=1}^{k} a_i v_i, \sum_{i=1}^{k} a_i v_i\rangle \\
&= \sum_{i=1}^{k} a_i \overline{a_i}\langle v_i, v_i\rangle + \sum_{i\neq j} a_i \overline{a_j}\langle v_i, v_j\rangle \\
&= \sum_{i=1}^{k} a_i \overline{a_i}\langle v_i, v_i\rangle \\
&= \sum_{i=1}^{k} |a_i| \|v_i\|^2
\end{aligned}$$

■

Theorem 1.9. *Suppose that $\langle\cdot,\cdot\rangle_1$ and $\langle\cdot,\cdot\rangle_2$ are two inner products on a vector space V. Then, $\langle\cdot,\cdot\rangle = \langle\cdot,\cdot\rangle_1 + \langle\cdot,\cdot\rangle_2$ is another inner product on V.*

Proof. We check the four properties of inner products.

1. *Linearity in first argument (i).* We have the following:

$$\begin{aligned}
\langle x+y, z\rangle &= \langle x+y, z\rangle_1 + \langle x+y, z\rangle_2 \\
&= \langle x, z\rangle_1 + \langle y, z\rangle_1 + \langle x, z\rangle_2 + \langle y, z\rangle_2 \\
&= \langle x, z\rangle_1 + \langle x, z\rangle_2 + \langle y, z\rangle_1 + \langle y, z\rangle_2 \\
&= \langle x, z\rangle + \langle y, z\rangle
\end{aligned}$$

2. *Linearity in the second argument (ii).*

$$\begin{aligned}
\langle cx, y\rangle &= \langle cx, y\rangle_1 + \langle cx, y\rangle_2 \\
&= c\langle x, y\rangle_1 + c\langle x, y\rangle_2 \\
&= c(\langle x, y\rangle_1 + \langle x, y\rangle_2) \\
&= c\langle x, y\rangle
\end{aligned}$$

3. *Conjugate symmetry.*

$$\begin{aligned}
\overline{\langle x, y\rangle} &= \overline{\langle x, y\rangle_1 + \langle x, y\rangle_2} \\
&= \overline{\langle x, y\rangle_1} + \overline{\langle x, y\rangle_2} \\
&= \langle y, x\rangle_1 + \langle y, x\rangle_2 \\
&= \langle y, x\rangle
\end{aligned}$$

4. *Positive definiteness.* Let $x \neq 0$. Then,

$$\langle x, x\rangle = \langle x, x\rangle_1 + \langle x, x\rangle_2$$

But, each inner product is positive definite, so, $\langle x, x\rangle_1 > 0$ and $\langle x, x\rangle_2 > 0$. Hence, $\langle x, x\rangle > 0$.

■

Theorem 1.10. *Let T be a linear operator on an inner product space V, and suppose that $\|T(x)\| = \|x\|$ for all x. Prove that T is one-to-one.*

Proof. Suppose $T(x) = 0$ Then $\|x\| = \|T(x)\| = \|0\| = 0$. Thus $x = 0$ and $N(T) = \{0\}$. Hence T is one-to-one. ■

Theorem 1.11. *Let V be a vector space over F where $F = \mathbb{C}$ or $F = \mathbb{R}$, and let W be an inner product space over F with an inner product $\langle \cdot, \cdot\rangle$. If $T : V \to W$ is linear then $\langle x, y\rangle' = \langle T(x), T(y)\rangle$ defines an inner product if and only if T is one-to-one.*

Proof. ($\Rightarrow$) If $\langle \cdot, \cdot\rangle'$ is an inner product, and $T(x) = 0$ then $\langle x, x\rangle' = \langle T(x), T(x)\rangle = 0$. Thus, $x = 0$, $N(T) = \{0\}$, and T is injective.

($\Leftarrow$) Linearity in first argument and conjugate symmetry follow trivially. For positive definiteness, $\langle x, x\rangle' = \langle T(x), T(x)\rangle > 0$, if $T(x) \neq 0$, which is guaranteed if $x \neq 0$, since T is injective.

Theorem 1.12. *Let V be an inner product space over F. (i) If $F = \mathbb{R}$ then,*

$$\langle x, y\rangle = \frac{1}{4}\|x + y\|^2 - \frac{1}{4}\|x - y\|^2$$

and (ii) if $F = \mathbb{C}$ then,

$$\langle x, y\rangle = \frac{1}{4}\sum_{k=1}^{4} i^k \left\|x + i^k y\right\|^2$$

There are called the **polar identities**.

Proof. We will prove this by arguing from right to left for the second equation. The following are statements of fact:

$$\begin{aligned}
\|x + y\|^2 &= \|x\|^2 + \|y\|^2 + 2\Re(\langle x, y\rangle) \\
\|x - y\|^2 &= \|x\|^2 + \|y\|^2 - 2\Re(\langle x, y\rangle) \\
\|x + iy\|^2 &= \|x\|^2 + \|y\|^2 + 2\Im(\langle x, y\rangle) \\
\|x - iy\|^2 &= \|x\|^2 + \|y\|^2 - 2\Im(\langle x, y\rangle)
\end{aligned}$$

We have the following,

$$\begin{aligned}
\frac{1}{4}\sum_{k=1}^{4} i^k \left\|x + i^k y\right\|^2 &= \frac{1}{4}(i\|x + iy\|^2 - \|x - y\|^2 - i\|x - iy\|^2 + \|x + y\|^2) \\
&= \frac{1}{4}(4\Re(\langle x, y\rangle) + 4i\Im(\langle x, y\rangle)) \\
&= \Re(\langle x, y\rangle) + i\Im(\langle x, y\rangle) \\
&= \langle x, y\rangle
\end{aligned}$$

■

Theorem 1.13. *Let $A \in M_{n\times n}(\mathbb{F})$. Then A can be written as the sum of two self-adjoint matrices.*

Proof. We let

$$A_1 = \frac{1}{2}(A + A^*) \qquad A_2 = \frac{1}{2i}(A - A^*)$$

It's not too hard to show that $A_1^* = A_1$ and $A_2^* = A_2$ and that $A = A_1 + iA_2$.

Moreover, this representation is unique. We suppose $A = B_1 + iB_2$ with $B_1^* = B_1$ and $B_2^* = B_2$. We have that $A^* = B_1^* - iB_2^*$ and that

$$B_1 = \frac{1}{2}(A + A^*) \qquad B_2 = \frac{1}{2i}(A - A^*)$$

which completes the proof. ■

Chapter 2

The Gram-Schmidt Orthogonalisation Process and Orthogonal Complements

Theorem 2.1. *Let V be an inner product space and $S = \{v_1, \ldots, v_k\}$ be an orthogonal subset of V consisting of non-zero vectors. If $y \in span(S)$, then,*

$$y = \sum_{i=1}^{k} \frac{\langle y, v_i \rangle}{\|v_i\|^2} v_i$$

Proof. Since $y \in span(S)$, we write $y = a_1 v_1 + \ldots + a_k v_k$ and say

$$\langle y, v_j \rangle = \langle a_1 v_1 + \ldots + a_k v_k, v_j \rangle = \sum_{i=1}^{k} a_i \langle v_j, v_i \rangle = a_j \langle v_j, v_j \rangle = a_j \|v_j\|^2$$

Hence, $a_j = \frac{\langle y, v_j \rangle}{\|v_j\|^2}$. ■

Theorem 2.2. *If, in addition to Theorem 2.1, the set S is orthonormal and $y \in span(S)$,*

$$y = \sum_{i=1}^{k} \langle y, v_i \rangle v_i$$

Proof. Each $\|v_j\|^2 = 1$ since they are unit vectors. Thus, the equation follows. What this means is that the i-th coordinate of y only depends on the i-th basis vector if the basis is orthonormal. It also follows from this that orthogonality implies linear independence. ■

Theorem 2.3. Gram-Schmidt Process. *Let V be an inner product space and $S = \{w_1, \ldots, w_n\}$ be a linearly independent subset of V. Define $S' = \{v_1, \ldots, v_n\}$, where $v_1 = w_1$ and*

$$v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j \rangle}{\|v_j\|^2} v_j \tag{2.1}$$

Then S' is an orthogonal set of nonzero vectors such that $span(S') = span(S)$.

Proof. The proof is by mathematical induction on n, the number of vectors in S. If $n = 1$, then $w_1 = v_1$ and $S = S'$. A set with a single vector is vacuously orthogonal, thus the base case holds. For the inductive step, suppose S'_{k-1} has been constructed using the process in 2.1. We show that the set S'_k is orthogonal when S'_{k-1} is extended by the algorithm. Clearly, $v_k \neq 0$, since if it were, then $w_k \in \text{span}(S_{k-1})$, contradicting the linear independence of S_k. Thus, for $1 \leq i \leq k-1$, and constructing v_k using (2.1), we obtain:

$$\langle v_k, v_i \rangle = \langle w_k, v_i \rangle - \sum_{j=1}^{k-1} \frac{\langle w_k, v_i \rangle}{\|v_j\|^2} \langle v_j, v_i \rangle = \langle w_k, v_i \rangle - \frac{\langle w_k, v_i \rangle}{\|v_i\|^2} \|v_i\|^2 = 0 \tag{2.2}$$

where the last equality holds since $\langle v_j, v_i \rangle = 0$ if $i \neq j$, by the inductive hypothesis that S'_{k-1} is orthogonal. This construction yields $\text{span}(S'_k) \subseteq \text{span}(S_k)$. Since $\dim(S'_k) = \dim(S_k)$ it follows that both sets span the same space. This completes the proof. ■

Theorem 2.4. *Let V be a nonzero finite-dimensional inner product space. Then V has an orthonormal basis β. Furthermore, if $\beta = \{v_1, \ldots, v_n\}$ and $x \in V$, then*

$$x = \sum_{i=1}^{n} \langle x, v_i \rangle v_i \tag{2.3}$$

Proof. By a corollary of the Replacement Lemma, every finite-dimensional vector space has a finite basis β'. Thus, by the Gram-Schmidt process, we can orthonormalise this basis into β. The remainder of the theorem follows from Theorem 2.2. ■

Theorem 2.5. *Let V be a finite-dimensional inner product space with an orthonormal basis $\beta = \{v_1, \ldots, v_n\}$. Let T be a linear operator on V, and let $A = [T]_\beta$. Then, for any i and j, $A_{ij} = \langle T(v_j), v_i \rangle$.*

Proof. The i,j-th entry of A is the i-th coordinate of $T(v_j)$. But we have that

$$T(v_j) = \sum_{i=1}^{n} \langle T(v_j), v_i \rangle v_i$$

Thus the i, j-th entry is $A_{ij} = \langle T(v_j), v_i \rangle$. ■

Definition 2.6. The orthogonal complement $S^\perp$ of S is defined as $S^\perp := \{x \in V : \langle x, y \rangle = 0, \forall y \in S\}$.

Theorem 2.7. *Let W be a finite-dimensional subspace of an inner product space V and let $y \in V$. Then, there exist unique vectors $u \in W$ and $z \in W^\perp$ such that $y = u + z$. Furthermore, if $\{v_1, \ldots, v_k\}$ is an orthonormal basis for W, then*

$$u = \sum_{i=1}^{k} \langle y, v_i \rangle v_i$$

Proof. We let $\{v_1, \ldots, v_k\}$ be an orthonormal basis for W and let u be a vector defined as above and let $z = y - u$. We shall prove that $z \in W^\perp$. We have,

$$\begin{aligned}
\langle z, v_j \rangle &= \langle y - \sum_{i=1}^{k} \langle y, v_i \rangle v_i, v_j \rangle \\
&= \langle y, v_j \rangle - \langle \sum_{i=1}^{k} \langle y, v_i \rangle v_i, v_j \rangle \\
&= \langle y, v_j \rangle - \sum_{i=1}^{k} \langle y, v_i \rangle \langle v_i, v_j \rangle \\
&= \langle y, v_j \rangle - \langle y, v_j \rangle \quad (*) \\
&= 0
\end{aligned}$$

Where $(*)$ follows since $\langle v_i, v_j \rangle = 0$ for $i \neq j$. Since z is orthogonal to all the basis vectors for W, it follows that $z \in W^\perp$. To show they are unique, suppose $y = u + z = u' + z'$. Then, $u' - u = z' - z \in W \cap W^\perp = \{0\}$. Thus $u' = u$ and $z' = z$. ■

Note that the definition of projection follows from here, by just taking the coordinates of some vector in V with respect to an ON basis for W.

Theorem 2.8. *Let V be an inner product space, and let W be a finite-dimensional subspace of V. If $x \notin W$, there exists $y \in V$ such that $y \in W^\perp$, but $\langle x, y \rangle \neq 0$.*

Proof. Let $\beta = \{v_1, \ldots, v_k\}$ be a basis for W. Since $x \notin W$, $x \notin span(\beta)$ and thus $\beta \cup \{x\}$ is linearly independent. Let $X = span(\beta \cup \{x\})$. Then $x = u + y$, with $u \in W$ and $y \in W^\perp \subseteq V$. Hence,

$$\langle x, y \rangle = \langle u + y, y \rangle$$

$$
\begin{aligned}
&= \langle u, y\rangle + \langle y, y\rangle \\
&= \|y\|^2 \neq 0
\end{aligned}
$$

And we are done. Trippy, eh? ■

Theorem 2.9. *Let A be an $n \times n$ matrix with complex entries. Then $AA^* = I$ if and only if the rows of A form an orthonormal basis for $\mathbb{C}^n$.*

Proof. ($\Leftarrow$) Suppose the rows of A form an orthonormal basis for $\mathbb{C}^n$. Then,

$$
\begin{aligned}
(AA^*)_{ii} &= \sum_{k=1}^{n} A_{ik}(A^*)_{ki} \\
&= \sum_{k=1}^{n} A_{ik}\overline{A}_{ik} \\
&= \langle Row_i(A), Row_i(A)\rangle \\
&= \|Row_i(A)\|^2 \\
&= 1 \quad \text{(Since the rows of } A \text{ are normal)}
\end{aligned}
$$

Likewise,

$$
\begin{aligned}
(AA^*)_{ij} &= \sum_{k=1}^{n} A_{ik}(A^*)_{kj} \\
&= \sum_{k=1}^{n} A_{ik}\overline{A}_{jk} \\
&= \langle Row_i(A), Row_j(A)\rangle \\
&= 0 \quad \text{(Since the rows of } A \text{ are orthogonal if } i \neq j)
\end{aligned}
$$

Hence, $AA^* = I$.

($\Rightarrow$) Conversely, suppose $AA^* = I$. Denote $A = (a_{ij})$ and let $v_i = (a_{i1}, \ldots, a_{in})$. Using the standard complex inner product, we have that $\langle v_i, v_j\rangle = (AA^*)_{ij} = \delta_{ij}$. ■

Note to self: You were being a sucker today. You better not continue that Mickey Mouse level algebra. As Troy Bolton would say, get your head in the game.

Theorem 2.10. *For any matrix $A \in M_{m\times n}(\mathbb{F})$, $(R(L_{A^*}))^\perp = N(L_A)$*

Proof. Let $x \in (R(L_{A^*}))^\perp$. Then, x is orthogonal to A^*y for all $y \in \mathbb{F}^n$. Thus,

$$0 = \langle x, A^*y\rangle = \langle Ax, y\rangle$$

Since this is true for all y, it must be the case that $Ax = 0$. Namely, $x \in N(A)$. Thus, $(R(L_{A^*}))^\perp \subseteq N(L_A)$. Conversely, let $x \in N(L_A)$. Then $Ax = 0$. Hence, for all $y \in \mathbb{F}^n$,

$$\langle Ax, y\rangle = \langle x, A^*y\rangle = 0$$

Thus, x is orthogonal to A^*y, for all y. Thus, $x \in (R(L_{A^*}))^\perp$. And we are done. ■

Theorem 2.11. *Let V be an inner product space, S and S_0 be subsets of V and W be a finite-dimensional subspace of V. Then, the following are true:*

1. *$S_0 \subseteq S$ implies that $S^\perp \subseteq S_0^\perp$*
2. *$S \subseteq (S^\perp)^\perp$; so $\text{span}(S) \subseteq (S^\perp)^\perp$*
3. *$W = (W^\perp)^\perp$*

4. $V = W \oplus W^{\perp}$

Proof.

1. Let $x \in S^{\perp}$ and $y \in S_0$. Thus $y \in S$ and $\langle x, y\rangle = 0$. But since $y \in S_0$, then x is orthogonal to any element in S_0. Thus, $x \in S_0^{\perp}$.

2. Let $x \in S$. Then x is orthogonal to any vector in $S^{\perp}$. But then, x is in the orthogonal complement of $S^{\perp}$; that is $x \in (S^{\perp})^{\perp}$. Since the orthogonal complement of a set is a subspace and the span of a set is the smallest subspace containing that set, it follows that $\text{span}(S) \subseteq (S^{\perp})^{\perp}$.

3. By the argument above, we have that $W \subseteq (W^{\perp})^{\perp}$. Conversely (by contrapositive), suppose $x \notin W$. By Theorem 2.8, we can find a $y \in W^{\perp}$ such that $\langle x, y\rangle \neq 0$. Thus, $x \notin (W^{\perp})^{\perp}$. It follows that $W = (W^{\perp})^{\perp}$.

4. By Theorem 2.7, we have that $V = W + W^{\perp}$. Now, we prove that $W \cap W^{\perp} = \{0\}$. Clearly, $\{0\} \subseteq W \cap W^{\perp}$ since the latter are vector spaces. Suppose $x \in W \cap W^{\perp}$. Then $\langle x, x\rangle = \|x\|^2 = 0$, since $x \in W$ implies that it is orthogonal to any element (say x) in $W^{\perp}$. Thus $x = 0$ and the set equality follows. Thus $V = W \oplus W^{\perp}$.
■

Theorem 2.12. *Let W_1 and W_2 be subspaces of a finite-dimensional inner product space. Then $(W_1 + W_2)^{\perp} = W_1^{\perp} \cap W_2^{\perp}$ and $(W_1 \cap W_2)^{\perp} = W_1^{\perp} + W_2^{\perp}$.*

Let $x \in (W_1 + W_2)^{\perp}$. Then, x is orthogonal to all y of the form $y = w_1 + w_2$ with $w_1 \in W_1$ and $w_2 \in W_2$. In particular, x is orthogonal to $y = 0 + w_2$ and $y = w_1 + 0$. Hence, $w \in W_1^{\perp} \cap W_2^{\perp}$. Conversely, if $x \in W_1^{\perp} \cap W_2^{\perp}$, then $x \in W_1^{\perp}$ and $x \in W_2^{\perp}$. Let $y = w_1 + w_2 \in (W_1 + W_2)$. Then, we have

$$\langle x, y\rangle = \langle x, w_1 + w_2\rangle = \langle x, w_1\rangle + \langle x, w_2\rangle = 0$$

Since x is orthogonal to w_1 and w_2. Thus, $x \in (W_1 + W_2)^{\perp}$.

For the second equality, we proceed by set algebra,

$$(W_1 \cap W_2)^{\perp} = ((W_1^{\perp})^{\perp} \cap (W_2^{\perp})^{\perp})^{\perp} = (W_1^{\perp} + W_2^{\perp})$$

■

Theorem 2.13. *Parseval's identity: Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis for V. For any $x, y \in V$,*

$$\langle x, y\rangle = \sum_{i=1}^{n} \langle x, v_i\rangle \overline{\langle y, v_i\rangle}$$

Proof. Since V is a finite inner product space and the gods of linear algebra have been kind enough to provide us with a basis, we know that

$$x = \sum_{i=1}^{n} \langle x, v_i\rangle v_i$$

and

$$y = \sum_{i=1}^{n} \langle y, v_i\rangle v_i$$

Then, by the beautiful powers endowed to inner products, we have,

$$\begin{aligned}
\langle x, y\rangle &= \langle \sum_{i=1}^{n} \langle x, v_i\rangle v_i, \sum_{i=1}^{n} \langle y, v_i\rangle v_i\rangle \\
&= \sum_{i=1}^{n} \langle x, v_i\rangle \overline{\langle y, v_i\rangle} \langle v_i, v_i\rangle + \sum_{i \neq j} \langle x, v_i\rangle \overline{\langle y, v_j\rangle} \langle v_i, v_j\rangle
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \langle x, v_i \rangle \overline{\langle y, v_i \rangle} \, \|v_i\|^2 \\
&= \sum_{i=1}^{n} \langle x, v_i \rangle \overline{\langle y, v_i \rangle}
\end{aligned}
$$

And we are done. I wonder if it is actually useful. ■

Theorem 2.14. *Bessel's inequality. Let V be an inner product space, and let $S = \{v_1, \ldots, v_n\}$ be an orthonormal subset of V. Then, for any $x \in V$ we have*

$$\|x\|^2 \geq \sum_{i=1}^{n} |\langle x, v_i \rangle|^2$$

Proof. Let $W = span(S)$. Then $x = u + y$, with $u \in W$ and $v \in W^{\perp}$. Denote $u = a_1 v_n + \ldots + a_n v_n$ Then,

$$
\begin{aligned}
\|x\|^2 &= \|u + y\|^2 \\
&= \|u\|^2 + \|y\|^2 \quad \text{(applying Theorem 1.4)} \\
&\geq \|u\|^2 \\
&= \left\| \sum_{i=1}^{n} a_i v_i \right\|^2 \\
&= \sum_{i=1}^{n} \|a_i v_i\|^2 \quad \text{(applying Theorem 1.4)} \\
&= \sum_{i=1}^{n} |a_i| \, \|v_i\|^2 \\
&= \sum_{i=1}^{n} |a_i| \quad \text{(these are just the } S \text{ coordinates of } x\text{)} \\
&= \sum_{i=1}^{n} |\langle x, v_i \rangle|^2
\end{aligned}
$$

Wow, that was long winded. There is a hammer proof, though, which uses exercise 6.2(1) from the textbook. Feel free to knock yourself out. ■

Theorem 2.15. *If $\{w_1, \ldots, w_n\}$ is an orthogonal set of non-zero vectors, then the vectors $v_1, \ldots, v_n$ derived from the Gram-Schmidt process satisfy $v_i = w_i$ for $i = 1, \ldots, n$.*

Proof. By mathematical induction. You all know inducing bores me. So do it yourself. ■

Chapter 3

The Adjoint of a Linear Operator

Theorem 3.1. *Let V be a finite dimensional inner product space over $\mathbb{F}$, and let $g : V \to \mathbb{F}$ be a linear transformation. Then, there exists a unique vector $y \in V$ such that $g(x) = \langle x, y\rangle$ for all $x \in V$*

Proof. Let $\beta = \{v_1, \ldots, v_n\}$ be an orthonormal basis for V. Set $h(x) = \langle x, y\rangle$ and

$$y = \sum_{i=1}^{n} \overline{g(v_i)} v_i$$

Then,

$$h(v_j) = \langle v_j, y\rangle = \langle v_j, \sum_{i=1}^{n} \overline{g(v_i)} v_i\rangle = \sum_{i=1}^{n} g(v_i)\langle v_j, v_i\rangle = g(v_j)\left\| v_j \right\|^2 = g(v_j)$$

for all $v_j \in \beta$. Since h and g agree on a basis, $h = g$. The fact that y is unique follows since $\langle x, y\rangle = \langle x, y'\rangle$ for $x \in \beta$ we have that $y = y'$. ∎

Theorem 3.2. *Let V be a finite-dimensional inner product space, and let T be a linear operator on V. Then, there exists a unique function $T^* : V \to V$ such that $\langle T(x), y\rangle = \langle x, T^*(y)\rangle$ for all $x, y \in V$. Furthermore, T^* is linear.*

Proof. This is a bit of a long-winded proof of linearity. It was proven in class. The proof strategy is as follows:

1. Define $g(x) = \langle T(x), y\rangle$ and prove it is linear.
2. Define $T^*(y) = y'$ where y' is the unique vector such that $g(x) = \langle x, y'\rangle$.

∎

Theorem 3.3. *Let V be a finite-dimensional inner product space, and let β be an orthonormal basis for V. If T is a linear operator on V. If T is a linear operator on V, then*

$$[T^*]_\beta = [T]_\beta^*$$

Proof. Let $A = [T]_\beta$, $B = [T^*]_\beta$ and $\beta = \{v_1, \ldots, v_n\}$. Then

$$B_{ij} = \langle T^*(v_j), v_i\rangle = \langle v_j, T(v_i)\rangle = \overline{\langle T(v_i), v_j\rangle} = \overline{A}_{ji} = A^*_{ij}$$

Hence, $B = A^*$. ∎

Theorem 3.4. *The following statements are cool facts of life:*

1. $(T + U)^* = T^* + U^*$
2. $(cT)^* = \overline{c}T^*$
3. $(TU)^* = U^*T^*$

4. $T^{**} = T$

5. $I^* = I$

Proof. The same facts are true if you change T, a linear operator, for A, a matrix. We will show them for the linear operator analogues.

1. We exploit the properties of adjoints.

$$\begin{aligned}\langle x, (T+U)^*(x)\rangle &= \langle (T+U)(x), x\rangle \\ &= \langle T(x), x\rangle + \langle U(x), x\rangle \\ &= \langle x, T^*(x)\rangle + \langle x, U^*(x)\rangle \\ &= \langle x, T^*(x) + U^*(x)\rangle\end{aligned}$$

   Since this is true for all x, it follows that $(T+U)^* = T^* + U^*$

2. The proofs for this one and the rest are similar. ■

■

Theorem 3.5. *Let V be a finite-dimensional inner product space, and let $T : V \to V$ be a linear transformation. If T is an isomorphism, prove that the adjoint T^* is also an isomorphism, and that $(T^{-1})^* = (T^*)^{-1}$.*

Proof. Let us prove some of the artillery that goes into the proof (yeah, that stuff from Theorem 3.4 that I said I would not prove).
Lemma. $(TU)^* = U^*T^*$.
Proof of lemma. We know that T and T^* satisfy

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

Thus, it follows that for any $x, y \in V$

$$\begin{aligned}\langle (TU)^*(x), y\rangle &= \langle x, (TU)(y)\rangle \\ &= \langle x, T(U(y))\rangle \\ &= \langle T^*(x), U(y)\rangle \\ &= \langle U^*(T^*(x)), y\rangle \\ &= \langle U^*T^*(x), y\rangle\end{aligned}$$

Since this holds for any y, then $(TU)^*(x) = U^*T^*(x)$. But this holds for all x, and from MATH 146 we know this implies that $(TU)^* = U^*T^*$. ■

Now, on to the problem. Since T is an isomorphism, it is invertible. Denote its inverse as T^{-1}, as sane people do. Then, we have

$$T^*(T^{-1})^* = (T^{-1}T)^* = I_V^* = I_V$$

Where the first equality uses our lemma. Thus, $(T^{-1})^*$ is the right inverse of T^*. Note that by a similar argument, we can prove that it is also its left inverse. Hence, $(T^*)^{-1} = (T^{-1})^*$. This shows that T^* is invertible, which means it is an isomorphism. ■

Theorem 3.6. *Let T be a linear operator on an inner product space V. Then, $\|T(x)\| = \|x\|$ for all $x \in V$ if and only if $\langle T(x), T(y)\rangle = \langle x, y\rangle$.*

Proof. ($\Leftarrow$) Suppose $\langle T(x), T(y)\rangle = \langle x, y\rangle$. Then,

$$\|T(x)\|^2 = \langle T(x), T(x)\rangle = \langle x, x\rangle = \|x\|^2$$

($\Rightarrow$) Conversely, suppose $\|T(x)\| = \|x\|$. We will use the polar identities proven in Theorem 1.9.

$$
\begin{aligned}
\langle x, y\rangle &= \frac{1}{4}\sum_{k=1}^{4} i^k \left\|x + i^k y\right\|^2 \\
&= \frac{1}{4}\sum_{k=1}^{4} i^k \left\|T(x + i^k y)\right\|^2 \quad (*) \\
&= \frac{1}{4}\sum_{k=1}^{4} i^k \left\|T(x) + i^k T(y)\right\|^2 \\
&= \langle T(x), T(y)\rangle
\end{aligned}
$$

Where $(*)$ follows since $\|T(x)\| = \|x\|$. At this point, you may be wondering how I did that. I just pulled an identity out of my ass. You should practice doing that yourself. ■

Theorem 3.7. *Let V be an inner product space, and let T be a linear operator on V. Then,*

1. $R(T^*)^\perp = N(T)$
2. *If V is finite-dimensional, then* $R(T^*) = N(T)^\perp$

Proof. For the first one, we argue by set inclusion. Let $x \in R(T^*)^\perp$. Then, x is orthogonal to any element in the range of T^*, say $T^*(y)$. That is

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

Since this is true for all y, $T(x) = 0$. Hence, $x \in N(T)$ and $R(T^*) \subseteq N(T)^\perp$. Conversely, let $x \in N(T)^\perp$. Then $T(x) = 0$. Hence,

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

That is, x is orthogonal to any $T^*(y)$. Namely, $x \in R(T^*)^\perp$. Therefore, $N(T) \subseteq R(T^*)^\perp$ and $R(T^*)^\perp = N(T)$. ■

For the second part, we use the fact that for a finite dimensional vector space V, any subspace W satisfies $(W^\perp)^\perp = W$. Equipping part (1) with the previous lemma, we have

$$R(T^*) = (R(T^*)^\perp)^\perp = N(T)^\perp$$

.

Theorem 3.8. *Let T be a linear operator on a finite-dimensional inner product space V. Then, the following are things you need to know to be successful in your dating life:*

1. $N(T^*T) = N(T)$ *and* $rank(T^*T) = rank(T)$.
2. $rank(T) = rank(T^*)$ *and* $rank(TT^*) = rank(T)$
3. *For any $n \times n$ matrix A,* $rank(A^*A) = rank(AA^*) = rank(A)$

Proof.

1. Let $x \in N(T^*T)$. Then $T^*T(x) = 0$ and, for all $y \in V$

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

   Since y is arbitrary, $T(x) \in R(T)^\perp$. Since $T(x) \in R(T)$, $\langle T(x), T(x)\rangle = 0$, which implies $T(x) = 0$ and $x \in N(T)$. Thus, $N(T^*T) \subseteq N(T)$. Conversely, suppose $x \in N(T)$. Then $T(x) = 0$. Which means,

   $$T^*T(x) = T^*(T(x)) = T^*(0) = 0$$

   Hence $x \in N(T^*T)$ and $N(T) \subseteq N(T^*T)$. A bit of set algebra hocus pocus leads to $N(T^*T) = N(T)$. The second part of the statement follows from the rank-nullity theorem. ■

2. Since V is finite-dimensional, we may partition V as $V = N(T) \oplus N(T)^\perp$. Thus, $\dim V = \dim(N(T)) + \dim(N(T)^\perp)$. By the rank-nullity theorem, $\dim V = \dim(N(T)) + \dim(R(T))$. Thus, $\dim(R(T)) = \dim(N(T)^\perp)$. But, by Theorem 3.7(2), since V is finite dimensional, $R(T^*) = N(T)^\perp$ and thus, we have

$$\dim(R(T)) = \dim(N(T)^\perp) = \dim(R(T^*))$$

   That is, $\operatorname{rank}(T) = \operatorname{rank}(T^*)$.

   The second part follows since

$$\operatorname{rank}(TT^*) = \operatorname{rank}(T^{**}T^*) = \operatorname{rank}(T^*) = \operatorname{rank}(T)$$

   where the first equality follows since $T = T^{**}$ and the second equality follows by part (1).

3. The result follows from parts 1 and 2 if we let $T = L_A$ and note that $L_{A^*} = (L_A)^*$. ■

Theorem 3.9. *Let A be an $n \times n$ matrix. Then* $\det(A^*) = \overline{\det(A)}$.

From MATH 146, we know that $\det(A) = \det(A^t)$. Hence, it remains to show that $\det(\overline{A}) = \overline{\det(A)}$. We shall proceed by induction.

Base case. For an 1×1 matrix we have that $A = (a)$ that $\det(\overline{A}) = \overline{a} = \det(\overline{A})$. Thus, our proposition holds for the base case.

Inductive step. Suppose the result is true for any $k \times k$ matrix. Now, consider the case of $A \in M_{(k+1)\times(k+1)}(\mathbb{C})$. Then, we make a phone call to our old friend, expansion by minors,

$$\begin{aligned}
\det(\overline{A}) &= \sum_{i=1}^{k+1} (-1)^{k+1} \overline{a}_{i1} \det(\widetilde{\overline{A}_{i1}}) \\
&= \sum_{i=1}^{k+1} (-1)^{k+1} \overline{a}_{i1} \overline{\det(\widetilde{A_{i1}})} \quad \text{(by inductive hypothesis)} \\
&= \sum_{i=1}^{k+1} \overline{(-1)^{k+1} a_{i1} \det(\widetilde{A_{i1}})} \quad \text{(by distributing the conjugate)} \\
&= \overline{\det(A)}
\end{aligned}$$

Maybe induction is not too bad after all. ■

Chapter 4

Normal and Self-adjoint Operators

Theorem 4.1. *Let T be a linear operator on a finite-dimensional inner product space V. If T has an eigenvector, then so does T^**

Proof. Let $T(v) = \lambda v$. Then, we have,

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

Thus, v is orthogonal to $R(T^* - \overline{\lambda} I)$. Thus $T^* - \overline{\lambda}$ is not onto, thus not one-to-one (since it is an operator) and thus has a non-zero nullspace. But vectors in its nullspace are eigenvectors of T^* with eigenvalue $\overline{\lambda}$, which completes the lemma. ■

Theorem 4.2. *Here's a bigger fish to fry. We call it Schur's Lemma, whoever he may be. It goes as follows: Let T be a linear operator on a finite-dimensional inner product space V. Suppose that the characteristic polynomial of T splits. Then, there exists an orthonormal basis β for V such that the matrix $[T]_\beta$ is upper triangular.*

Proof. Wow, that was a mouthful. We do it by induction. For the base case, where $n = 1$, the result is trivial, as all 1×1 matrices are upper triangular.

Inductive step. Suppose the result is true for linear operators on $(n - 1)$-dimensional inner product spaces whose characteristic polynomial splits. Since its characteristic polynomial splits, T has eigenvectors and, by Theorem 4.1, so does T^*. Since we can play around with norms as scaling factors, let z be a unit eigenvector of T^*; that is $T^*(z) = \lambda z$. Construct $W = span(z)$. At this point Schur decided it would be a good idea to show that $W^\perp$ is T-invariant. We show it in the following way: Let $y \in W^\perp$ and $x = cz \in W$. Then,

$$\begin{aligned}
\langle T(y), x\rangle &= \langle T(y), cz\rangle \\
&= \langle y, T^*(cz)\rangle \\
&= \langle y, cT^*(z)\rangle \\
&= \langle y, c\lambda z\rangle \\
&= \overline{c\lambda}\langle y, z\rangle \\
&= 0 \quad \text{(since } y \text{ and } z \text{ are orthogonal)}
\end{aligned}$$

If you are not angry at Schur, then you would note that, since x was arbitrary in W, then $T(y) \in W^\perp$. From MATH 146, we know that the characteristic polynomial of the restriction of a linear operator to a subspace divides the characteristic polynomial of the operator on the whole space. That is, the characteristic polynomial of $T_{W^\perp}$ divides the characteristic polynomial of T. We know that $V = W \oplus W^\perp$ and since $\dim(W) = 1$ (it only has one basis vector), it follows that $\dim(W^\perp) = n - 1$.

If you do not have murderous thoughts towards Schur right now, you might recall this was a proof by induction. Then, we apply our inductive hypothesis to obtain an orthonormal basis γ for $W^\perp$ such that $[T_{W^\perp}]_\gamma$ is upper triangular. We are almost there. Since orthogonality implies linear independence, $\beta = \gamma \cup \{z\}$ is an orthonormal basis for V. Thus, $[T]_\beta$ is upper triangular. ■

Definition 4.3. Let V be an inner product space, and let T be a linear operator on V. We say that T is **normal** if $TT^* = T^*T$. An $n \times n$ matrix is normal if $AA^* = A^*A$

Theorem 4.4. *Let V be an inner product space, and let T be a normal operator on V. Then the following are things Marcoux might say to get on "Stuff UWaterloo Profs Say":*

1. $\|T(x)\| = \|T^*(x)\|$ *for all* $x \in V$
2. $T - cI$ *is normal for every* $c \in \mathbb{F}$
3. *If x is an eigenvector of T, then x is also an eigenvector of T^*. In fact, if $T(x) = \lambda x$, then $T^*(x) = \overline{\lambda} x$.*
4. *If λ_1 and λ_2 are distinct eigenvalues of T with corresponding eigenvectors x_1 and x_2, then the latter are orthogonal.*

Proof.

1. We have the following statement of truth,

$$\begin{aligned} \|T(x)\|^2 &= \langle T(x), T(x) \rangle = \langle x, T^*T(x) \rangle = \langle x, TT^*(x) \rangle \\ &= \langle T^*(x), T^*(x) \rangle = \|T^*(x)\|^2 \end{aligned}$$

2. Note that $(T - cI)^* = T^* - \overline{c}I$. We argue by function composition. Note that,

$$\begin{aligned} (T - cI)(T^* - \overline{c}) &= TT^* - \overline{c}T - cT^* + |c|^2 I \\ &= T^*T - cT^* - \overline{c}T + |c|^2 I \\ &= (T^* - \overline{c})(T - cI) \end{aligned}$$

Hence, $T - cI$ is normal.

3. We apply part (1) of this theorem as follows. Let $T(x) = \lambda x$ and $U = T - \lambda I$. Then,

$$0 = \|U(x)\| = \|U^*(x)\| = \left\|(T^* - \overline{\lambda} I)(x)\right\| = \left\|T^*(x) - \overline{\lambda} x\right\|$$

Hence, $T^*(x) = \overline{\lambda} x$. That is, x is an eigenvector of T^*, with a corresponding eigenvalue $\overline{\lambda}$.

4. Let the conditions in the statement hold. Then,

$$\begin{aligned} \lambda_1 \langle x_1, x_2 \rangle &= \langle \lambda_1 x_1, x_2 \rangle = \langle T(x_1), x_2 \rangle \\ &= \langle x_1, T^*(x_2) \rangle = \langle x_1, \overline{\lambda_2} x_2 \rangle \\ &= \lambda_2 \langle x_1, x_2 \rangle \end{aligned}$$

Since $\lambda_1 \neq \lambda_2$, it must be the case that $\langle x_1, x_2 \rangle = 0$. That is, x_1 and x_2 are orthogonal. ■

Theorem 4.5. *Let T be a linear operator on a finite-dimensional complex inner product space V. Then T is normal if and only if there exists an orthonormal basis for V consisting of eigenvectors of T.*

Proof. Suppose that T is normal. We assault this problem with induction.

By the Fundamental Theorem of Algebra, the characteristic polynomial splits over $\mathbb{C}$. Let $\beta = \{v_1, \ldots, v_n\}$ be a basis for V onto which we will apply Schur's Lemma to make $A = [T]_\beta$ upper triangular. Since A is upper-triangular, v_1 is an eigenvector. Now, for the inductive step, suppose $v_1, \ldots, v_{k-1}$ are eigenvectors of T, for some k. We want to show that v_k is also an eigenvector (this would show that all the vectors are eigenvectors). By our inductive hypothesis, $T(v_j) = \lambda_j v_j$ for $j < k$. Hence, by Theorem 4.4.3, $T^*(v_j) = \overline{\lambda_j} v_j$. Now, since A is upper triangular, we have

$$T(v_k) = A_{1k}v_1 + A_{2k}v_2 + \ldots + A_{jk}v_j + \ldots + A_{kk}v_k \quad (*)$$

Additionally, we have, for $j < k$, the following:

$$A_{jk} = \langle T(v_k), v_j \rangle = \langle v_k, T^*(v_j) \rangle = \langle v_k, \overline{\lambda} v_j \rangle = \lambda \langle v_k, v_j \rangle = 0$$

By applying this fact to $(*)$, we can get rid of all terms except for the last one. That is $T(v_k) = A_{kk}v_k$. Namely, v_k is an eigenvector of T, and we are done.

The converse follows from the fact that diagonal matrices commute. ■

Definition 4.6. A linear operator T is said to be self-adjoint on an inner product space V if $T = T^*$. A matrix A is called self-adjoint if $A = A^*$

Theorem 4.7. *Let T be a self-adjoint operator on a finite-dimensional inner product space V. Then*

1. *Every eigenvalue of T is real.*
2. *Suppose V is a real finite-dimensional inner product space. Then the characteristic polynomial of T splits.*

Proof. (1) Let x be an eigenvector of T. Then,

$$\lambda x = T(x) = T^*(x) = \overline{\lambda} x$$

Since $x \neq 0$, it must be the case that $\lambda = \overline{\lambda}$. Namely, λ is real.

(2). Let $\dim V = n$ and β be an orthonormal basis for V. Let $A = [T]_\beta$. Then A is self-adjoint. Define $T_A : \mathbb{C}^n \to \mathbb{C}^n$ with $T_A(x) = Ax$ for all $x \in \mathbb{C}^n$. We note that T_A is also self-adjoint, since $[T_A]_\sigma = A$, where σ is the standard ordered basis for $\mathbb{C}^n$.

By the fundamental theorem of algebra, the characteristic polynomial of T_A splits over the complex numbers into linear factors $t - \lambda_i$. But by part (1), $\lambda_i \in \mathbb{R}$. Hence the characteristic polynomial splits over the real numbers. Since the characteristic polynomial of T_A is the same as A's, and A's is the same as T's, the characteristic polynomial of T splits. ■

Theorem 4.8. *Let T be a linear operator on a finite-dimensional real inner product space V. Then T is self-adjoint if and only if there exists an orthonormal basis β for V consisting of eigenvectors of T.*

Proof. ($\Rightarrow$) Suppose T is self-adjoint. Then, $[T^*]_\beta = [T]^*_\beta = [T]_\beta$. Since T is self-adjoint, its characteristic polynomial splits, thus, by Schur's Lemma there exists an orthonormal basis (say β) such that $[T]_\beta$ is upper triangular. But the same would apply for $[T^*]_beta$. Hence $[T]_\beta$ and $[T]^*_\beta$ are upper triangular, thus they are diagonal matrices. Thus, β must consist of eigenvectors of T.

($\Leftarrow$) Conversely, suppose there is an orthonormal basis β for V consisting of eigenvectors of T. Then $[T]_\beta$ is a diagonal matrix consisting of eigenvalues. $[T^*]_\beta$ is also diagonal, consisting of the conjugates of the eigenvalues of T. But since the inner product space is real, these are the same, thus $[T^*]_\beta = [T]_\beta$. ■

Theorem 4.9. *Let T and U be self-adjoint operators on an inner product space V. Then TU is self-adjoint if and only if $TU = UT$.*

Proof. ($\Rightarrow$) Suppose TU is self-adjoint. Then,

$$TU = (TU)^* = U^*T^* = UT$$

Where the second equality is an application of Theorem 3.4 and the third equality follows from the fact that T and U are self-adjoint.

($\Leftarrow$) Suppose $TU = UT$. Then,

$$(TU)^* = U^*T^* = UT = TU$$

■

Theorem 4.10. *Let T be a normal operator on a finite-dimensional complex inner product space V and let W be a subspace of V. If W is T-invariant, then W is also T^*-invariant.*

Proof. By Theorem 4.5, since T is normal on a complex inner product space, there exists an orthonormal basis β consisting of eigenvectors of T. That is, T is diagonalisable. From MATH 146, we have that T_W is also diagonalisable to, say, $[T_W]_\gamma$. Then $[T^*_W]_\gamma = [T_W]^*_\gamma$, is also diagonal. Then the eigenvectors for W only get scaled by T^*. Thus, W is T^*-invariant. ■

Theorem 4.11. *Let T be a normal operator on a finite-dimensional inner product space V. Prove that $N(T) = N(T^*)$ and $R(T) = R(T^*)$.*

Proof. Since T is normal, $T(x) = 0$ if and only if $T^*(x) = 0$ by Theorem 4.4(1). Then, $N(T) = N(T^*)$. For the second part we have,

$$R(T^*) = (R(T^*)^{\perp})^{\perp} = N(T)^{\perp} = N(T^*)^{\perp} = R(T)$$

Where the equalities follow by Theorem 3.7 and the first part of this theorem. ∎

Theorem 4.12. *Let T be a self-adjoint operator on a finite-dimensional inner product space V. Then, for all $x \in V$,*

$$\|T(x) \pm ix\|^2 = \|T(x)\|^2 + \|x\|^2$$

Additionally, $T - iI$ is invertible and $[(T - iI)^{-1}]^* = (T + iI)^{-1}$

Proof. We will show the positive case, the negative one is similar.

$$\begin{aligned}
\|T(x) + ix\|^2 &= \langle T(x) + ix, T(x) + ix\rangle \\
&= \|T(x)\|^2 + \langle T(x), ix\rangle + \langle ix, T(x)\rangle + |i|^2\langle x, x\rangle \\
&= \|T(x)\|^2 + \|x\| - i\langle T(x), x\rangle + i\langle T^*(x), x\rangle \\
&= \|T(x)\|^2 + \|x\| - i\langle T(x), x\rangle + i\langle T(x), x\rangle \\
&= \|T(x)\|^2 + \|x\|
\end{aligned}$$

Suppose $0 = \|T(x) - ix\| = \|(T - iI)(x)\|$. Thus, either x is an eigenvector of T with eigenvalue i or $x = 0$. But since T is self-adjoint, its eigenvalues are real, so $x = 0$. Thus, $N(T - iI) = \{0\}$. Finally, we compute the inverse,

$$\begin{aligned}
\langle x, [(T - iI)^{-1}]^*((T + iI))(y)\rangle &= \langle [(T - iI)^{-1}]x, (T + iI)(y)\rangle \\
&= \langle [(T - iI)^{-1}]x, (T^* + iI)(y)\rangle \\
&= \langle [(T - iI)^{-1}]x, (T - iI)^*(y)\rangle \\
&= \langle [(T - iI)(T - iI)^{-1}]x, (y)\rangle \\
&= \langle x, y\rangle
\end{aligned}$$

And the result holds. ∎

Theorem 4.13. *Let T be a normal operator on a finite-dimensional real inner product space V whose characteristic polynomial splits. Then V has an orthonormal basis of eigenvectors of T and T is self-adjoint.*

By Schur's Lemma, there exists an orthonormal basis $\beta = \{v_1, \ldots, v_n\}$ for V such that $A = [T]_\beta$ is upper triangular. We thus have that v_1 is an eigenvector (with eigenvalue A_{11}). Let this be the base case. Now, arguing by induction, suppose $\{v_1, \ldots, v_{k-1}\}$ is a set of orthonormal eigenvectors. Consider

$$T(v_k) = A_{1k}v_1 + A_{2k}v_2 + \ldots + A_{jk}v_j + \ldots + A_{kk}v_k$$

But for $j < k$, $A_{jk} = \langle T(v_k, v_j\rangle = \langle v_k, T^*(v_j)\rangle = \langle v_k, \overline{\lambda}v_j\rangle = \lambda\langle v_k, v_j\rangle = 0$, where the second equality follows since T is normal and the fourth because v_j and v_k are orthogonal. Hence $T(v_k) = A_{kk}v_k$; namely v_k is an eigenvector. That is β is an orthogonal basis of eigenvectors; thus, by Theorem 4.8, T is self-adjoint. ∎

Theorem 4.14. *Assume that T is a linear operator on a complex (not necessarily finite-dimensional) inner product space V with an adjoint T^*. Then:*

1. *If T is self-adjoint, then $\langle T(x), x\rangle$ is real for all $x \in V$.*
2. *If T satisfies $\langle T(x), x\rangle = 0$ for all $x \in V$, then $T = T_0$.*
3. *If $\langle T(x), x\rangle$ is real for all $x \in V$, then $T = T^*$.*

Proof.

1. Since T is self-adjoint, its characteristic polynomial splits, so it has eigenvectors. Let x be an eigenvector of T, and we have

$$\lambda\langle x,x\rangle = \langle \lambda x,x\rangle = \langle T(x),x\rangle = \langle x,T^*(x)\rangle = \langle x,T(x)\rangle = \langle x,\lambda x\rangle = \overline{\lambda}\langle x,x\rangle$$

   Thus $\lambda = \overline{\lambda}$, thus the eigenvalues are real. Since $\langle T(x),x\rangle$ is sneaked into that equation and $\langle x,x\rangle \in \mathbb{R}$, the result follows.

2. The result follows if we replace x with $x+y$ and explore what happens when we expand the inner product.

3. If $\langle T(x),x\rangle$ is real for all $x \in V$, then $\langle T(x),x\rangle = \langle x,T(x)\rangle = \langle T^*(x),x\rangle$. Since T and T^* agree everywhere, $T = T^*$. ■

Theorem 4.15. *An $n \times n$. real matrix A is said to be a* **Gramian** *matrix if there exists a real (square) matrix B such that $A = B^tB$. Prove that A is a Gramian matrix if and only if A is symmetric and all of its eigenvalues are nonnegative.*

Proof. ($\Rightarrow$) Suppose A is Gramian. Then $A^t = (B^tB)^t = B^tB = A$, so A is symmetric. Additionally, if x is a unit eigenvector of A, we have

$$\lambda = \lambda\langle x,x\rangle = \langle \lambda x,x\rangle = \langle Ax,x\rangle = \langle B^tBx,x\rangle = \langle Bx,Bx\rangle = \|Bx\|^2 \geq 0$$

So the eigenvalues of A are non-negative.

($\Leftarrow$) Conversely, if a matrix A is symmetric and its eigenvalues are non-negative, it is diagonalisable. Additionally, $\sqrt{\lambda_i}$ exist, since each $\lambda_i \geq 0$. Thus, if we let D be the matrix whose diagonal entries are $\sqrt{\lambda_i}$, we can write $A = [I]_\alpha^\beta D^2 [I]_\beta^\alpha$ with the right bases and $B = D[I]_\beta^\alpha$. ■

Theorem 4.16. *Let T_1 and T_2 be linear transformations from a finite-dimensional vector space V to itself, and assume that both T_1 and T_2 are diagonalisable. That is, assume that there are bases β_1 and β_2 such that the matrices $[T_i]_{\beta_i}$ are diagonal. Prove that $T_1T_2 = T_2T_1$ if and only if there is a basis β such that the matrices $[T_i]_\beta$ are both diagonal.*

($\Leftarrow$) Suppose there exists a basis β such that $[T_1]_\beta$ and $[T_2]_\beta$ are diagonal. Since diagonal matrices commute, we have that $[T_1]_\beta[T_2]_\beta = [T_2]_\beta[T_1]_\beta$. Thus, $T_1T_2 = T_2T_1$.

($\Rightarrow$) Suppose $T_1T_2 = T_2T_1$. Let $\beta_1 = \{v_1, \ldots, v_n\}$ and $\beta_2 = \{w_1, \ldots, w_n\}$ be bases of eigenvectors of T_1 and T_2, respectively, for V. Let, $T(v_i) = \lambda_i v_i$ and $T(w_i) = \mu_i w_i$. Since $T_1T_2 = T_2T_1$, we have:

$$T_1(T_2(v_i)) = T_1T_2(v_i) = T_2T_1(v_i) = T_2(\lambda_i v_i) = \lambda_i T_2(v_i)$$

Thus, $T_2(v_i)$ is a λ_i-eigenvector of T_1. Thus, $T_2(v_i) \in E_{\lambda_i}$; namely, each eigenspace E_{λ_i} is T_2 invariant. From MATH 146, we have that since T_2 is diagonalisable, its restriction to an invariant subspace is also diagonalisable: the reason for this is because the characteristic polynomial of $T|_{E_{\lambda_i}}$ divides the characteristic polynomial of T; since T is diagonalisable, its characteristic polynomial splits, thus it must be that the characteristic polynomial of $T|_{E_{\lambda_i}}$ also splits and the algebraic and geometric multiplicities of said eigenspace remain invariant. Thus $T_2|_{E_{\lambda_i}}$ is diagonalisable. Let β_{λ_i} be a basis for E_{λ_i} such that $[T_2|_{E_{\lambda_i}}]_{\beta_{\lambda_i}}$ is diagonal. Note that this is a basis for E_{λ_i} consisting of eigenvectors for both T_1 and T_2. Construct

$$\beta = \bigcup_i \beta_{\lambda_i}$$

to be a basis for V. By construction, be have that $[T_2]_\beta$ is diagonal. Since each E_{λ_i} is also T_1-invariant (since they are spaces of eigenvectors of T_1), it follows that each $[T_1|_{E_{\lambda_i}}]_{\beta_{\lambda_i}}$ is diagonal, and thus $[T_1]_\beta$ is diagonal. Thus T_1 and T_2 are simultaneously diagonalisable. ■

Theorem 4.17. *Let T be a linear operator on an inner product space V, and let W be a T-invariant subspace of V. Then,*

1. *If T is self-adjoint, then T_W is self-adjoint*
2. *$W^{\perp}$ is T^*-invariant*
3. *If W is both T and T^* invariant, then $(T_W)^* = (T^*)_W$*
4. *If W is both T- and T^*- invariant and T is normal, then T_W is normal.*

Proof.

1. For all $x, y \in W$,

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

2. For all $x \in W$, $y \in W^{\perp}$, we have,

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

since $T(x) \in W$, as W is T-invariant. Thus, $T^*(y)$ is orthogonal to x. That is, $T^*(y) \in W^{\perp}$.

3. Similar to (1), for $x, y \in W$ we have

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

where the last equality follows since W is T^*-invariant. Thus $(T^*)_W = (T_W)^*$

4. Similar to (1) and (3). ■

Theorem 4.18. *Let T be a normal operator on a finite-dimensional complex inner product, space V and let W be a subspace of V. If W is T-invariant, then W is also T^*-invariant.*

Proof. Since T is normal over a complex field, it is diagonalisable. Thus, T_W is diagonalisable, with a basis of eigenvectors of T, say β. Since T is normal, if x is an eigenvector of T, then it is also an eigenvector of T^*. That is, β is a basis for W of eigenvectors of T^*. Thus, W is T^*-invariant. ■

Chapter 5

Unitary and Orthogonal Operators and Their Matrices

Definition 5.1. Let T be a linear operator on a finite-dimensional inner product space V (over $\mathbb{F}$). Suppose $\|T(x)\| = \|x\|$ for all $x \in V$. Then,

1. If $\mathbb{F} = \mathbb{C}$ then T is called a **unitary operator**.

2. If $\mathbb{F} = \mathbb{R}$ then T is called an **orthogonal operator**.

Definition 5.2. Let T be a linear operator on an **infinite-dimensional** inner product space V. Then if $\|T(x)\| = \|x\|$, then the operator is called an **isometry**. If, in addition, T is onto, the operator is called unitary or orthogonal. Note that if it is onto and it satisfies the norm requirement, it is also one-to-one (because it must mean that the null space is just the zero vector).

Theorem 5.3. *Let T be a linear operator on a finite-dimensional inner product space V. Then the following statements are equivalent:*

1. $TT^* = T^*T = I$.

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

3. *If β is an orthonormal basis for V, then $T(\beta)$ is an orthonormal basis for V.*

4. *There exists an orthonormal basis β for V such that $T(\beta)$ is an orthonormal basis for V.*

5. $\|T(x)\| = \|x\|$ *for all* $x \in V$

Proof. (1) $\Rightarrow$ (2). Suppose $TT^* = T^*T = I$. Then, $\langle T(x), T(y)\rangle = \langle x, T^*T(y)\rangle = \langle x, y\rangle$

(2) $\Rightarrow$ (3) Suppose $\langle T(x), T(y)\rangle = \langle x, y\rangle$. Say $\beta = \{v_1, \ldots, v_n\}$ is a basis for V. Then, $\langle T(v_i), T(v_j)\rangle = \langle v_i, v_j\rangle = \delta_{ij}$. Hence, $T(\beta)$ is a set of orthonormal vectors. Since it has the same cardinality as β, it follows that $T(\beta)$ is a basis for V.

(3) $\Rightarrow$ (4) Duh.

(4) $\Rightarrow$ (5) Let $x \in V$. Then,

$$\|x\|^2 = \langle a_1v_1 + \ldots + a_nv_n, a_1v_1 + \ldots + a_nv_n\rangle = |a_1|^2 + \ldots + |a_n|^2$$

Computing $\|T(x)\|$ should yield the same. Or you are wrong.

(5) $\Rightarrow$ (1) This one is a bit harder. We prove a lemma first.

Lemma. Let U be a self-adjoint operator on a finite-dimensional inner-product space V. If $\langle x, U(x)\rangle = 0$ for all $x \in V$, then $U = T_0$.

Proof of lemma. Since U is self-adjoint, then we can choose an orthonormal basis for V consisting of eigenvectors of U. Let x be an eigenvector in the basis. Then,

$$0 = \langle x, U(x)\rangle = \langle x, \lambda x\rangle = \overline{\lambda}\langle x, x\rangle$$

But $x \neq 0$, hence $\overline{\lambda} = 0$. Hence, $U(x) = 0$ for all basis vectors x. Hence $U = T_0$. (QED Lemma)

Back to the proof. For any $x \in V$ we have

$$\langle x, x\rangle = \|T(x)\|^2 = \langle T(x), T(x)\rangle = \langle x, T^*T(x)\rangle$$

Thus, $\langle x, (I - T^*T)(x)\rangle = 0$. By our lemma, this implies that $(I - T^*T) = T_0$. Thus $T^*T = I$. We may argue similarly to show that it commutes. Victory! ■

Theorem 5.4. *Let T be a linear operator on a finite-dimensional real inner product space V. Then V has an orthonormal basis of eigenvectors of T with corresponding eigenvalues of absolute value 1 if and only if T is both self-adjoint and orthogonal.*

Proof. ($\Rightarrow$) By Theorem 4.8, T is self-adjoint. Likewise, $TT^*(v_i) = T(\lambda_i v_i) = (\lambda_i)^2 v_i = v_i$. Hence $TT^* = I$ and by Theorem 5.2 T is orthogonal.

($\Leftarrow$) If T is self-adjoint, there exists an orthonormal basis of eigenvectors. Additionally, if T is orthogonal, $\|v_i\| = \|T(v_i)\| = |\lambda_i|\,\|v_i\|$. Thus $|\lambda_i| = 1$. ■

Theorem 5.5. *Let T be a linear operator on a finite-dimensional complex inner product space V. Then V has an orthonormal basis of eigenvectors of T with corresponding eigenvalues of absolute value 1 if and only if T is unitary.*

Proof. ($\Rightarrow$) Since T has an orthonormal basis of eigenvectors, it is normal. Additionally, since it is normal, if x is an eigenvector, then $T(x) = \lambda x$ and $T^*(x) = \overline{\lambda}x$. Thus,

$$TT^*(x) = T(\overline{\lambda}x) = \lambda\overline{\lambda}x = |\lambda|^2 x = x$$

since $|\lambda| = 1$. Thus $TT^* = T^*T = I$ and T is unitary.

($\Leftarrow$) Since T is unitary, is is normal. Thus, we find a basis of unit eigenvectors of T. Since T is unitary,

$$|\lambda|\,\|x\| = \|\lambda x\| = \|T(x)\| = \|x\|$$

thus $|\lambda| = 1$. ■

Definition 5.6. Let L be a one-dimensional subspace of $\mathbb{R}^2$. We may view L as a line in the plane through the origin. A linear operator T on $\mathbb{R}^2$ is called a reflection of $\mathbb{R}^2$ about L if $T(x) = x$ for all $x \in L$ and $T(x) = -x$ for all $x \in L^\perp$.

Definition 5.7. A square matrix A is called an orthogonal matrix if $A^tA = AA^t = I$ and unitary if $A^*A = AA^* = I$.

Definition 5.8. We say that a matrix A is **unitarily equivalent** to a matrix B if and only if there exists a unitary matrix P such that $A = P^*BP$.

Theorem 5.9. *Let A be a complex $n \times n$ matrix. Then A is normal if and only if it is unitarily equivalent to a diagonal matrix.*

Proof. ($\Rightarrow$) If A is normal it is orthogonally diagonalisable. Thus $A = P^{-1}DP$ for some invertible matrix P. But since the columns of P are orthonormal vectors, P is unitary.

($\Leftarrow$) Suppose that A is unitarily equivalent to a diagonal matrix. Then $A = P^*DP$ and $A^* = P^*D^*P$. Thus $AA^* = P^*DD^*P$ and $A^*A = P^*D^*DP$. Since D and D^* are diagonal, they commute. So $AA^* = A^*A$, that is, A is normal. ■

Theorem 5.10. *Let A be a real $n \times n$ matrix. Then A is symmetric if and only if it is orthogonally equivalent to a real diagonal matrix.*

Proof. Exactly as above. ■

Definition 5.11. Let V be a real inner product space. A function $f : V \to V$ is called a rigid motion if

$$\|f(x) - f(y)\| = \|x - y\|$$

Theorem 5.12. *Let $f : V \to V$ be a rigid motion on a finite-dimcnsional real inner product space V. Then there exists a unique orthogonal operator T on V and a unique translation g on V such that $f = g \circ T$.*

Proof. Define $T : V \to V$ as $T(x) = f(x) - f(0)$. Our strategy will hinge on showing that T is an orthogonal linear operator. If we do this, then we can say that $f = g \circ T$, where g is just the translation by $f(0)$. We first note that T is a composite operation of f on x followed by a translation by $f(0)$. Since f is already a rigid motion, a further translation would not affect the "rigidity" of it. Hence T is a rigid motion. Additionally, for any $x \in V$ we have

$$\|T(x)\| = \|f(x) - f(0)\| = \|x - 0\| = \|x\|$$

where the second equality follows since f is a rigid motion. Thus, for any $x, y \in V$ we have:

$$\|T(x) - T(y)\|^2 = \langle T(x) - T(y), T(x) - T(y)\rangle = \|T(x)\|^2 + \|T(y)\|^2 - 2\langle T(x), T(y)\rangle = \|x\|^2 + \|y\|^2 - 2\langle T(x), T(y)\rangle$$

Similarly,

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2\langle x, y\rangle$$

But since T is a rigid motion, it must be the case that $\|T(x) - T(y)\| = \|x - y\|$ Hence, $\langle T(x), T(y)\rangle = \langle x, y\rangle$, namely, T is orthogonal (we still need to show it is linear to call it an operator, properly). So, we'll do that exactly; namely, we show T is linear. First we prove that T is additive.

$$\begin{aligned}
\|T(x+y) - T(x) - T(y)\|^2 &= \|[T(x+y) - T(x)] - T(y)\|^2 \\
&= \|T(x+y) - T(x)\|^2 + \|T(y)\|^2 - 2\langle T(x+y) - T(x), T(y)\rangle \\
&= \|x + y - x\|^2 + \|y\|^2 - 2(\langle T(x+y), T(y)\rangle - \langle T(x), T(y)\rangle) \\
&= 2\|y\|^2 - 2((\langle x+y, y\rangle - \langle x, y\rangle) \\
&= 2\|y\|^2 - 2((\langle x+y-x, y\rangle) \\
&= 2\|y\|^2 - 2((\langle y, y\rangle) \\
&= 2\|y\|^2 - 2\|y\|^2 \\
&= 0
\end{aligned}$$

Thus, $T(x+y) - T(x) - T(y) = 0$ and $T(x+y) = T(x) + T(y)$. Likewise, we argue for linearity of scalar multiplication as follows (luckily, it's easier):

$$\begin{aligned}
\|T(ax) - aT(x)\|^2 &= \|T(ax)\|^2 + \|aT(x)\|^2 - 2\langle T(ax), aT(x)\rangle \\
&= \|ax\|^2 + |a|^2 \|T(x)\| - 2a\langle T(ax), T(x)\rangle \\
&= 2a^2 \|x\| - 2a\langle ax, x\rangle \\
&= 2a^2 \|x\| - 2a^2\langle x, x\rangle \\
&= 0
\end{aligned}$$

Thus, $T(ax) - aT(x) = 0$ and $T(ax) = aT(x)$. Thus, T is linear. For uniqueness, suppose there exist two such operators, namely T and U, such that

$$f(x) = T(x) + u_0 = U(x) + v_0$$

But, if we inflict this transformation on $x = 0$, we obtain $u_0 = v_0$. Thus, we reduce this to $T(x) = U(x)$, but this holds for all the vectors in V, thus $T = U$. It was indeed a long-drawn battle, but we won! ■

Definition 5.13. Let $ax^2 + 2bxy + cy^2 + dx + ey + f = 0$ be a quadratic equation. We call the quadratic $ax^2 + 2bxy + cy^2$ the **associated quadratic form**. Note that if we let $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ and $X = \begin{bmatrix} x \\ y \end{bmatrix}$, then the associated quadratic form can be expressed as $\langle AX, X \rangle = X^t AX$.

Theorem 5.14. *The conic section $ax^2+2bxy+cy^2+f = 0$ may be written in terms of coordinates for $\beta = \{x', y'\}$, where β is a rotation of the standard ordered basis for $\mathbb{R}^2$. Namely, the xy term can be dropped.*

Proof. Denote A as the in Definition 5.9. Since A is symmetric (thus, self-adjoint since $\mathbb{F} = \mathbb{R}$) it is orthogonally equivalent to a diagonal matrix D. In particular, $D = P^t AP$ where D is a diagonal matrix with its diagonal entries corresponding to the eigenvalues (λ_1, λ_2) of A and P is an orthogonal matrix. We define $X' = \begin{bmatrix} x' \\ y' \end{bmatrix}$ by $X' = P^t X$, or, equivalently, by $PX' = PP^t X = X$. Then, we have the following set of equalities:

$$X^t AX = (PX')^t A(PX') = (X')^t (P^t AP)X' = (X')^t DX' = \lambda_1 (x')^2 + \lambda_2 (y')^2$$

Since P is orthogonal, we have that $\det(P) = \pm 1$. If $\det(P) = -1$, we can perform a column swap to obtain a matrix Q such that $\det(Q) = 1$. Either way, our orthogonal transformation represents a rotation of a conic section. ■

Theorem 5.15. *If T is a unitary operator on a finite-dimensional inner product space V, then T has a unitary* **square root**. *That is, there exists an operator U such that $U^2 = T$.*

Since T is a unitary linear operator on a finite dimensional, applying Corollary 2 to Theorem 6.18 from the textbook we let $\beta = \{v_1, \ldots, v_n\}$ be an orthonormal basis of eigenvectors such that:

$$[T]_\beta = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

where each λ_i is the eigenvalue corresponding to v_i and $|\lambda_i| = 1$. Now, let $\lambda_i = \sigma_i^2$ and denote

$$A = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \end{bmatrix}$$

Note that $|\sigma_i| = 1$. Construct the linear transformation U such that $[U]_\beta = A$. Note U is unitary by the converse of Corollary 2 of Theorem 6.18 from the textbook. Since, $A^2 = [T]_\beta$, it follows that $U \circ U = T$. ■

Theorem 5.16. *Let T be a self-adjoint operator on a finite-dimensional inner product space. Then, $(T + iI)(T - iI)^{-1}$ is unitary.*

Let $U = (T + iI)(T - iI)^{-1}$ and $U^* = [(T - iI)^{-1}]^*(T + iI)^*$. By Theorem 4.12, and the fact that the factors in UU^* commute, since they are polynomials in T, it follows that $UU^* = U^*U = I$. ■

Theorem 5.17. *Let A be an $n \times n$ real symmetric or complex normal matrix. Then*

$$tr(A) = \sum_{i=1}^{n} \lambda_i \qquad \text{and} \qquad tr(A^*A) = \sum_{i=1}^{n} |\lambda_i|^2$$

where λ_i represent the eigenvalues of A.

Proof. Since A is real symmetric or complex normal, it is either orthogonally (in the real case) or unitarily (in the complex case) equivalent to a diagonal matrix of eigenvalues. That is, $A = P^*DP$ where P is orthogonal (or unitary). Then,

$$tr(A) = tr(P^*DP) = tr(DP^*P) = tr(D) = \sum_{i=1}^{n} \lambda_i$$

and,

$$tr(A^*A) = tr(P^*D^*DP) = tr(D^*D) = \sum_{i=1}^{n} |\lambda_i|^2$$

■

Theorem 5.18. *Let A be an $n \times n$ real symmetric or complex normal matrix. Use this to prove that*

$$\det(A) = \prod_{i=1}^{n} \lambda_i$$

where λ_i are eigenvalues of A.

The theorem is true in general, but this one is trivial to prove.

$$\det(A) = \det(P^*DP) = \det(P^*)\det(D)\det(P) = \det(D) = \prod_{i=1}^{n} \lambda_i$$

since D is diagonal and its determinant is the product of the diagonal entries, which are simply the eigenvalues of A. ■

Theorem 5.19. *Let U be a unitary operator on an inner product space V, and let W be a finite-dimensional U-invariant subspace of V. Then,*

1. *$U(W) = W$*
2. *$W^{\perp}$ is U-invariant*

Proof.

1. Let $w \in W$. If $U(w) = 0$, then $0 = \|U(w)\|^2 = \|w\|^2$, thus $w = 0$. Thus U is an isomorphism on W and $U(W) = W$.
2. See the assignment question related to this. Note that the result is not true if W is not finite-dimensional. ■

Theorem 5.20. *A matrix that is both unitary and upper triangular is diagonal.*

Proof. Since A is unitary, $AA^* = I$. Thus $A^* = A^{-1}$. But since A is upper triangular, A^{-1} is also upper-triangular. Thus A and A^* are both upper triangular. It follows that A is diagonal. ■

Theorem 5.21. QR Factorisation. *Let $w_1, \ldots, w_n$ be linearly independent vectors in $\mathbb{F}^n$ and let $v_1, \ldots, v_n$ be the orthogonal vectors obtained by applying a Gram-Schmidt process on $w_1, \ldots, w_n$. Let $u_1, \ldots, u_n$ be the vectors obtained from normalising the v_i's. Then,*

1. *An expression for* w_k *is*

$$w_k = \|v_k\| \, u_k + \sum_{j=1}^{k-1} \langle w_k, u_j \rangle u_j$$

2. *Let* A *and* Q *denote the* $n \times n$ *matrices in which the* k*th columns are* w_k *and* u_k*, respectively. Define* $R \in M_{n\times n}(\mathbb{F})$ *by:*

$$R_{jk} = \begin{cases} \|v_j\| & j = k \\ \langle w_k, u_j \rangle & j < k \\ 0 & j > k \end{cases}$$

   Then, $A = QR$

Proof.

1. We write $v_k = \|v_k\| \, u_k$. Substituting this into the expression for the Gram-Schmidt process yields the desired result.

2. Some computation leads to this solution. ■

From the result above it follows that every invertible matrix is the product of a unitary matrix and an upper triangular matrix.

Chapter 6

Orthogonal Projections and the Spectral Theorem

Definition 6.1. T is a projection if and only if $T = T^2$.

Definition 6.2. We can formulate an alternative definition for a projection T. Suppose $V = W_1 \oplus W_2$, and $x = x_1 + x_2$ with $x_1 \in W_1$ and $x_2 \in W_2$. Then T is the projection on W_1 along W_2 if $T(x) = x_1$.

Definition 6.3. Let V be an inner product space and let $T : V \to V$ be a projection. We say that T is an orthogonal projection if $R(T)^\perp = N(T)$ and $N(T)^\perp = R(T)$. Note that if V is finite-dimensional, then having one of these two is sufficient, since we are allowed to take the orthogonal complement of both sides.

Theorem 6.4. *Let V be an inner product space, and let T be a linear operator on V. Then T is an orthogonal projection if and only if T has an adjoint T^* and $T^2 = T = T^*$.*

Proof. ($\Rightarrow$) Suppose T is an orthogonal projection. Then $T^2 = T$. Now we show that the adjoint exists and it equals T (note that V need not be finite-dimensional). We have that $V = R(T) \oplus N(T)$ since V is a linear operator, and $R(T)^\perp = N(T)$ since it is orthogonal. Let $x = x_1 + x_2$ and $y = y_1 + y_2$ with $x_1, y_1 \in R(T)$ and $x_2, y_2 \in N(T)$. We have:

$$\langle x, T(y)\rangle = \langle x_1 + x_2, T(y_1) + T(y_2)\rangle = \langle x_1 + x_2, T(y_1)\rangle = \langle x_1 + x_2, y_1\rangle = \langle x_1, y_1\rangle + \langle x_2, y_1\rangle = \langle x_1, y_1\rangle$$

Likewise,

$$\langle T(x), y\rangle = \langle x_1, y_1 + y_2\rangle = \langle x_1, y_1\rangle + \langle x_1, y_2\rangle = \langle x_1, y_1\rangle$$

Since $\langle T(x), y\rangle = \langle x, T(y)\rangle$ for all $x, y \in V$ it follows that T^* exists and $T = T^*$.

($\Leftarrow$) Conversely, suppose $T^2 = T = T^*$. Since $T^2 = T$, it is a projection. We now show it is orthogonal by showing it satisfies Definition 6.3. We argue first by set inclusion. Let $x \in R(T)$ and $y \in N(T)$. Since $x \in R(T)$, $x = T(x) = T^*(x)$, thus

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

Hence $x \in N(T)^\perp$. Now, let $y \in N(T)^\perp$. To show $y \in R(T)$ we must show that $T(y) = y$. We have:

$$\|y - T(y)\|^2 = \langle y - T(y), y - T(y)\rangle = \langle y, y - T(y)\rangle - \langle T(y), y - T(y)\rangle$$

Notice that $T(y - T(y)) = T(y) - T^2(y) = T(y) - T(y) = 0$. Hence, $y - T(y) \in N(T)$ and our first term is 0. But,

$$\langle T(y), y - T(y)\rangle = \langle y, T^*(y - T(y))\rangle = \langle y, T^*(y) - T^*T(y)\rangle = \langle y, T(y) - T(y)\rangle = 0$$

Thus, it follows that $y - T(y) = 0$ and $T(y) = y$. Thus $y \in R(T)$ and $N(T)^\perp = R(T)$.

From the above result it follows that $R(T)^\perp = (N(T)^\perp)^\perp \supseteq N(T)$ (since $(W^\perp)^\perp \subseteq W$ for any subspace W of V; note, if V is finite-dimensional set equality would hold directly and we would be done). What a bummer,

V might be infinite-dimensional. Note a problem, though; set inclusion saves us. Let $x \in R(T)^{\perp}$. Then, for any $y \in V$ we have

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

Since $y \neq 0$ in general, it must be the case that $T(x) = 0$. Hence, $x \in N(T)$ and $R(T)^{\perp} = N(T)$. ■

Theorem 6.5. *Let V be a finite dimensional vector space, let W be a subspace of V, and T be the orthogonal projection of V on W. Then, there exists an orthonormal basis β such that $[T]_\beta = \begin{bmatrix} I_k & O_1 \\ O_2 & O_3 \end{bmatrix}$*

Proof. Pick a basis $\gamma = \{v_1, \dots, v_k\}$ for W. Extend γ to a basis $\beta = \gamma \cup \{v_{k+1}, \dots, v_n\}$ for V. Since T is the orthogonal projection onto W, $T(v_i) = v_i$ for $1 \leq i \leq k$ and $T(v_j) = 0$ for $k+1 \leq j \leq n$. Hence, $[T]_\beta$ is of the required form. ■

Theorem 6.6. *A linear operator T on a finite-dimensional vector space V is diagonalisable if and only if V is the direct sum of the eigenspaces of T. (See section 3 of the lecture RDWillard lecture "Lies, half-truths, omissions").*

Proof. Let $\lambda_1, \dots, \lambda_k$ be the distinct eigenvalues of T.

($\Rightarrow$) Suppose T is diagonalisable. For each i, we choose an ordered basis γ_i for each E_{λ_i} (each λ_i eigenspace). By Theorem 5.8 on the Ross Willard Winter 2017 MATH146 notes (good times, eh?), we have that $\gamma_1 \cup \dots \cup \gamma_k$ is linearly independent, with $\dim V$ vectors, and is thus a basis for V. Thus, it follows that

$$V = E_{\lambda_1} \oplus \dots \oplus E_{\lambda_k}$$

as required.

($\Leftarrow$) Conversely, suppose that V is the direct sum of the eigenspaces of T. For each eigenspace E_{λ_i} choose an ordered basis γ_i (we can do this, as each eigenspace is finite-dimensional). Since V is the direct sum of the eigenspaces, $\gamma_1 \cup \dots \cup \gamma_k$ is a basis for V. But this is a basis of eigenvectors for V. Then, by Theorem 5.1 on the RDWillard notes, T is diagonalisable. ■

Theorem 6.7. *The Spectral Theorem.* *Suppose that T is a linear operator on a finite-dimensional inner product space V over F with the distinct eigenvalues $\lambda_1, \dots, \lambda_k$. Assume that T is normal if $\mathbb{F} = \mathbb{C}$ and that T is self-adjoint if $\mathbb{F} = \mathbb{R}$. For each $i (1 \leq i \leq k)$, let W_i be the eigenspace of T corresponding to the eigenvalue λ_i, and let T_i be the orthogonal projection of V on W. Wow, long set up. Drumroll. Then the following statements are true:*

1. *$V = W_1 \oplus \dots \oplus W_k$.*
2. *If W_i' denotes the direct sum of eigenspaces W_j with $j \neq i$, then $W_i^{\perp} = W_i'$.*
3. *$T_iT_j = \delta_i jT_i$*
4. *$I = T_1 + \dots + T_k$.*
5. *$T = \lambda_1 T_1 + \dots + \lambda_k T_k$*

Proof.

1. Since T is either normal or self-adjoint (depending on the field) then T is diagonalisable. By Theorem 6.6, V is the direct sum of its eigenspaces, namely each W_i.
2. Let $x \in W_i$ and $y \in W_j$ with $i \neq j$. Since x, y are linear combinations of orthogonal eigenvectors, $\langle x, y\rangle = 0$. Thus, $W_i' \subseteq W_i^{\perp}$. From the result above we have that $\dim V = \dim(W_1) + \dots + \dim(W_k)$. Since V is finite dimensional, we also have that $\dim(V) = \dim(W_i) + \dim((W_i^{\perp})$. Collecting and eliminating terms, it follows that $\dim(W_i') = \dim(W_i^{\perp})$. This, added to the set inclusion yields $W_i' = W_i^{\perp}$.

3. Suppose $i = j$. Since each T_i is an orthogonal projection $T_i^2 = T_i$. Now, suppose $i \neq j$. Any $v \in V$ can be expressed uniquely as $v = w_1 + \ldots + w_i + \ldots + w_j + \ldots + w_k$ with each $w_m \in W_m$. Then $T_iT_j(v) = T_i(T_j(w_1 + \ldots + w_i + \ldots + w_j + \ldots + w_k)) = T_i(w_j) = 0$. Hence $T_iT_j = 0$. It follows that $T_iT_j = \delta_{ij}T_i$.

4. Express $v = w_1 + \ldots + w_k$. Then, for each $T_i(v) = w_i$. Thus, $T_1(v) + \ldots + T_k(v) = w_1 + \ldots + w_k = v$. This is true for any v, hence $T_1 + \ldots + T_k = I$.

5. Write $v = w_1 + \ldots + w_k$. Then,

$$T(v) = T(w_1 + \ldots + w_k) = T(w_1) + \ldots + T(w_k) = \lambda_1 w_1 + \ldots + \lambda_k w_k$$

But $w_i = T_i(w_i)$. Hence, our expression becomes

$$T(v) = \lambda_1 T_1(w_1) + \ldots + \lambda_k T_k(w_k) = (\lambda_1 T_1 + \ldots + \lambda_k T_k)(v)$$

■

Definition 6.8. The set of eigenvalues of T is called the **spectrum** of T. The sum in part 4 of Theorem 6.7 is called the **resolution of the identity operator** induced by T. The sum in part 5 of Theorem 6.7 is called the **spectral decomposition of** T.

Theorem 6.9. ***Lagrange's Interpolation Theorem***. *Let $\mathbb{F}$ be a field (which is possibly infinite). Let $\{a_1, \ldots, a_n\} \subseteq \mathbb{F}$ be a set of distinct scalars from the field and $\{b_1, \ldots, b_n\} \subseteq \mathbb{F}$ also be scalars from the field (not necessarily distinct). Then, there exists a polynomial $p \in P_n(\mathbb{F})$ such that $p(a_i) = b_i$ for $1 \leq i \leq k$.*

Proof. We can construct a linear system in a^m and apply the Vandermonde determinant to show that it has solutions. ■

Theorem 6.10. *If $\mathbb{F} = \mathbb{C}$, then T is normal if and only if $T^* = g(T)$ for some polynomial g.*

Proof. ($\Leftarrow$) Suppose $T^* = g(T)$. Since T commutes with any polynomial in T, we have

$$TT^* = Tg(T) = g(T)T = T^*T$$

Thus, T is normal.

($\Rightarrow$) Suppose T is normal. Since $\mathbb{F} = \mathbb{C}$, T is diagonalisable and by the spectral theorem we have that $T = \lambda_1 T_1 + \ldots \lambda_k T_k$, where T_i is the orthogonal projection to E_{λ_i}. Taking the adjoint of both sides, we obtain $T^* = (\lambda_1 T_1 + \ldots \lambda_k T_k)^* = \overline{\lambda_1} T_1^* + \ldots + \overline{\lambda_k} T^* k = \overline{\lambda_1} T_1 + \ldots + \overline{\lambda_k} T$, where the last equality follows since orthogonal projections are self-adjoint. Using Lagrange Interpolation, we construct a polynomial $g(\lambda_i) = \overline{\lambda_i}$ for $1 \leq i \leq k$. Then

$$g(T) = g(\lambda_1)T_1 + \ldots + g(\lambda_k)T_k = \overline{\lambda_1} T_1 + \ldots + \overline{\lambda_k} T_k = T^*$$

■

Theorem 6.11. *If $\mathbb{F} = \mathbb{C}$, then T is unitary if and only if T is normal and $|\lambda| = 1$ for every eigenvalue λ of T.*

Proof. ($\Rightarrow$) Suppose T is unitary. Then T is normal and each of its eigenvalues has absolute value 1 by Theorem 5.5.

($\Leftarrow$) Let T be normal and have unit length eigenvalues. We write the spectral decomposition of T as $T = \lambda_1 T_1 + \ldots + \lambda_k T_k$. Thus we have,

$$\begin{aligned} TT^* &= (\lambda_1 T_1 + \ldots + \lambda_k T_k)(\lambda_1 T_1 + \ldots + \lambda_k T_k)^* \\ &= (\lambda_1 T_1 + \ldots + \lambda_k T_k)(\overline{\lambda_1} T_1 + \ldots + \overline{\lambda_k} T_k) \\ &= |\lambda_1|^2 T_1 + \ldots + |\lambda_k|^2 T_k \\ &= T_1 + \ldots + T_k \\ &= I \end{aligned}$$

Where the final equality follows from part 4 of the spectral theorem. Thus T is unitary. ■

Theorem 6.12. *If $\mathbb{F} = \mathbb{C}$ and T is normal, then T is self-adjoint if and only if each eigenvalue is real.*

($\Leftarrow$) Using the spectral theorem

$$T^* = \overline{\lambda_1}T_1 + \ldots + \overline{\lambda_k}T_k = \lambda_1 T_1 + \ldots + \lambda_k T_k = T$$

($\Rightarrow$) The eigenvalues are real since $\overline{\lambda}x = T^*(x) = T(x) = \lambda x$ ■

Theorem 6.13. *Let T satisfy the hypothesis for the spectral theorem and let its spectral decomposition be $T = \lambda_1 T_1 + \ldots + \lambda_k T_k$. Then each T_j is a polynomial in T.*

Proof. Choose a polynomial g_j such that $g_j(\lambda_i) = \delta_{ij}$, the Kroenecker delta. Then,

$$g_j(T) = g_j(\lambda_1)T_1 + \ldots + g_j(\lambda_k)T_k = \delta_{1j}T_1 + \ldots + \delta_{jj}T_j + \ldots + \delta kjT_k = T_j$$

■

Theorem 6.14. *Let W be a finite-dimensional subspace of an inner product space V. Then, if T is the orthogonal projection of V on W then $I - T$ is the orthogonal projection of V on $W^\perp$.*

First we show that $R(I-T) = W^\perp$. We write $V = W \oplus W^\perp$; let $v \in V$ and write $v = w + w'$ where $w \in W$ and $w' \in W^\perp$. Then,

$$(I-T)(v) = I(v) - T(v) = I(w+w') - T(w+w') = (w+w') - w = w' \in W^\perp$$

Thus $R(I-T) \subseteq W^\perp$. To show that $W^\perp \subseteq R(I-T)$, we note that if $w' \in W^\perp$, then for some arbitrary element $v = w + w'$, $w' = (I-T)(v)$, thus, $w' \in R(I-T)$.

It is now sufficient to show that $N(I-T) = W$. Let $v = w + w'$ as above and $v \in N(I-T)$. Then $(I-T)(v) = I(v) - T(v) = v - w = w' = 0$, that is, the $W^\perp$ component of v is 0 and $v \in W$. Now, take $w \in W$. Then $(I-T)(w) = I(w) - T(w) = w - w = 0$. Thus $w \in N(I-T)$. This completes the equality. Thus, we have that $N(I-T)^\perp = R(I-T)$. But since our spaces are finite dimensional, we can take the orthogonal complements of both sides and conclude that $N(I-T)^\perp = (R(I-T)^\perp)^\perp = R(I-T)$. Thus, $I-T$ is the orthogonal projection of V on $W^\perp$. ■

Theorem 6.15. *Let T be a linear operator on a finite-dimensional inner product space V. If T is a projection, then T is an orthogonal projection if and only if $\|T(x)\| \leq \|x\|$ for all $x \in V$.*

Proof. ($\Rightarrow$) Suppose T is an orthogonal projection on some subspace W of V. Let $v = w + w' \in V$ with $w \in W$ and $w \in W^\perp$. Then

$$\|T(v)\| = \left\|T(w+w')\right\| = \|w\| \leq \left\|w+w'\right\| = \|v\|$$

($\Leftarrow$) I'll prove the converse at a later date. ■

Theorem 6.16. *Let T be a normal operator on a finite-dimensional inner product space V. If T is a projection, then T is an orthogonal projection.*

Since V is finite-dimensional, it is sufficient to prove that $R(T)^\perp = N(T)$. We do so by set-inclusion. Suppose $x \in R(T)^\perp$. Then,

$$\begin{aligned} \langle T(x), T(x)\rangle &= \langle x, T^*T(x)\rangle \\ &= \langle x, T(T^*(x))\rangle \qquad \text{(since } T \text{ is normal)} \\ &= 0 \qquad \text{(since } T(T^*(x)) \in R(T)\text{)} \end{aligned}$$

Thus, $T(x) = 0$ by positive-definiteness of the inner product and $x \in N(T)$. Conversely, if $x \in N(T)$, then since T is a projection we have that

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

thus, $x \in R(T)^\perp$. Hence, $R(T)^\perp = N(T)$. Since V is finite-dimensional, we take the orthogonal complements of both sides and arrive to $R(T) = N(T)^\perp$. Thus, T is an orthogonal projection. ■

Theorem 6.17. *This one is long, but it has some nice properties you may want to use at some point in life. Let T be a normal operator on a finite-dimensional complex inner product space V (thus $T = \lambda_1 T_1 + \ldots + \lambda_k T_k$ by the spectral theorem). Then:*

1. *If g is a polynomial, then*
$$g(T) = \sum_{i=1}^{k} g(\lambda_i) T_i$$
2. *If $T^n = T_0$ for some n, then $T = T_0$*
3. *If U is a linear operator on V, then U commutes with T if and only if U commutes with all T_i*
4. *There exists a normal linear operator U such that $U^2 = T$*
5. *T is invertible if and only if $\lambda_i \neq 0$ for $1 \leq i \leq k$*
6. *T is a projection if and only if every eigenvalue of T is 0 or 1*
7. *$T = -T^*$ if and only if λ_i is a purely imaginary number.*

Proof.

1. Let $g(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$. Then, we have,
$$\begin{aligned} g(T) &= g(\lambda_1 T_1 + \ldots + \lambda_k T_k) \\ &= a_0 I + a_1(\lambda_1 T_1 + \ldots + \lambda_k T_k) + a_2(\lambda_1 T_1 + \ldots + \lambda_k T_k)^2 + \ldots + a_n(\lambda_1 T_1 + \ldots + \lambda_k T_k)^n \\ &= a_0(T_1 + \ldots T_k) + a_1(\lambda_1 T_1 + \ldots + \lambda_k T_k) + a_2(\lambda_1 T_1 + \ldots + \lambda_k T_k)^2 + \ldots + a_n(\lambda_1 T_1 + \ldots + \lambda_k T_k)^n \\ &= (T_1 + \ldots T_k) + a_1(\lambda_1 T_1 + \ldots + \lambda_k T_k) + a_2(\lambda_1^2 T_1 + \ldots + \lambda_k^2 T_k)^2 + \ldots + a_n(\lambda_1^n T_1 + \ldots + \lambda_k^n T_k)^n \\ &= (a_0 + a_1\lambda_1 + a_2\lambda_1^2 + \ldots + a_n\lambda_k^n)T_1 + \ldots + (a_0 + a_1\lambda_k + a_2\lambda_k^2 + \ldots + a_n\lambda_k^n)T_k \\ &= g(\lambda_1)T_1 + \ldots + g(\lambda_k)T_k \end{aligned}$$
Where the third equality follows by the spectral resolution to the identity and the fourth equality follows from part (3) of the spectral theorem.
2. $T^n = \lambda_1^n T_1 + \ldots \lambda_k^n T_k$. Let v_i be a λ_i-eigenvector. Then,
$$0 = T^n(v_i) = \lambda_i^n v_i$$
That is, $\lambda_i^n = 0$ and $\lambda_i = 0$ for all $1 \leq i \leq k$. Thus, $T = 0T_1 + \ldots + 0T_k = T_0$.
3. ($\Leftarrow$) Suppose U commutes with all T_i. Then,
$$UT = U(\lambda_1 T_1 + \ldots + \lambda_k T_k) = \lambda_1 U T_1 + \ldots + \lambda_k U T_k = \lambda_1 T_1 U + \ldots + \lambda_k T_k U = (\lambda_1 T_1 + \ldots + \lambda_k T_k)U = TU$$
Conversely, suppose U commutes with T. By Theorem 6.13, $T_j = g_j(T)$ for some polynomial g_j. Then, since U commutes with any polynomial of T, we have
$$UT_j = Ug_j(T) = g_j(T)U = T_j U$$
4. Choose $U = \lambda_1^{\frac{1}{2}} T_1 + \ldots + \lambda_k^{\frac{1}{2}} T_k$, where $\lambda_i^{\frac{1}{2}}$ is any square-root of the i-th eigenvalue.
5. Let $A = [T]_\beta$ for some basis β. Then, $\det A = \lambda_1^{m_1}\lambda_2^{m_2}\ldots\lambda_k^{m_k}$, which is non-zero if and only if $\lambda_i \neq 0$ for $1 \leq i \leq k$. Thus T is invertible if and only if each eigenvalue is non-zero.
6. If every eigenvalue is 0 or 1, then it follows trivially that $T^2 = T$. If T is a projection then $T^2 = T$ and $\lambda_i^2 = \lambda_i$. Solving for λ_i we obtain that it is either 0 or 1.

7. Suppose the eigenvalues are purely imaginary. Then $\overline{\lambda_i} = -\lambda_i$. Then,

$$-T^* = -(\lambda_i T_1 + \ldots + \lambda_k T_k)^* = -(\overline{\lambda_i} T_1 + \ldots + \overline{\lambda_k} T_k) = \lambda_1 T_1 + \ldots + \lambda_k T_k = T$$

Conversely, if $T = -T^*$, then we have that $\lambda_i = -\overline{\lambda_i}$. Thus, λ_i is purely imaginary. ∎

Theorem 6.18. *If T is a normal operator on a complex inner product space and U is a linear operator that commutes with T, then U commutes with T^*.*

Proof. By Theorem 6.10, $T^* = g(T)$ for some polynomial g. Since U commutes with T, it commutes with any polynomial of T. Thus,

$$UT^* = Ug(T) = g(T)U = T^*U$$

∎

Theorem 6.19. *Let U and T be normal operators on a finite-dimensional complex inner product space V such that $TU = UT$. Then, there exists an orthonormal basis β for V consisting of vectors which are eigenvectors of both T and U.*

We use induction on the dimension n of V. If $n = 1$, U and T will be diagonalised simultaneously by any orthonormal basis. Suppose the statement is true for $n \leq k1$. Consider the case $n = k$. Now pick one arbitrary eigenspace $W = E_\lambda$ of T for some eigenvalue λ. Note that W is T-invariant, naturally, and U-invariant since

$$TU(w) = UT(w) = \lambda U(w)$$

for all $w \in W$. If $W = V$, then we may apply Theorem 4.5 to the operator U and get an orthonormal basis consisting of eigenvectors of U. Those vectors will also be eigenvectors of T.

On the other hand, if W is a proper subspace of V, we may apply the induction hypothesis to T_W and U_W, which are normal by Theorem 4.18, and get an orthonormal basis β_1 for W consisting of eigenvectors of T_W and U_W. So those vectors are also eigenvectors of T and U.

Additionally, we know that $W^\perp$ is also T- and U-invariant by Theorem 4.17. They are also normal operators (Theorem 4.18). Again, by applying the induction hypothesis we get an orthonormal basis β_2 for $W^\perp$ consisting of eigenvectors of T and U. Since V is finite-dimensional, we know that $\beta = \beta_1 \cup \beta_2$ is an orthonormal basis for V consisting of eigenvectors of T and U[1]. ∎

[1] This proof was retrieved almost *verbatim* from Lin J., Su S., Lai Z. (2011) Solutions to Linear Algebra, Fourth Edition, Stephen H. Friedberg, Arnold J. Insel, Lawrence E. Spence.

Chapter 7

The Singular Value Decomposition and Pseudoinverse

We begin by extending the definition of the adjoint transformation (now not necessarily an operator) to any real or complex finite-dimensional inner product space.

Definition 7.1. Let $T : V \to W$ where V and W are finite-dimensional inner product spaces, endowed with the inner products $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$, respectively (note, these can be different inner products). A function $T : W \to V$ is called an adjoint of T if $\langle T(x), y \rangle_2 = \langle x, T^*(y) \rangle_1$ for all $x \in V$ and $y \in W$

We present Theorem 7.2 without proof. The proof is similar to the proofs for the analogue theorem for the adjoint operator presented in Chapter 3.

Theorem 7.2. *Let $T : V \to W$ be a linear transformation, where V and W are finite-dimensional inner product spaces with inner products $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$, respectively. Then the following statements are true:*

1. *There is a unique adjoint T^* of T, and T^* is linear.*
2. *If β and γ are orthonormal bases for V and W, respectively, then $([T]_\beta^\gamma)^* = [T^*]_\gamma^\beta$*
3. $\operatorname{rank}(T) = \operatorname{rank}(T^*)$
4. $\langle T^*(x), y \rangle_1 = \langle x, T(y) \rangle_2$ *for all $x \in W$ and $y \in V$*
5. *For all $x \in V$, $T^*T(x) = 0$ if and only if $T(x) = 0$.*

Equipped with the theorems above, we proceed with the introduction of the singular value decomposition.

Theorem 7.3. *Let V and W be finite-dimensional inner product spaces, and let $T : V \to W$ be a linear transformation of rank r. Then, there exist orthonormal bases $\beta = \{v_1, \dots, v_n\}$ and $\gamma = \{w_1, \dots w_m\}$ for V and W, respectively, and scalars $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$ such that*

$$T(v_i) = \begin{cases} \sigma_i u_i, & \text{for } 1 \leq i \leq r \\ 0, & \text{otherwise} \end{cases}$$

In particular, v_i is a σ_i^2-eigenvector of T^*T for $1 \leq i \leq r$; for $i > r$, v_i is a 0-eigenvector of T^*T.

Proof. We begin by showing that T^*T is positive semi-definite. Let x be a unit λ-eigenvector of T^*T. Then

$$\lambda = \lambda \langle x, x \rangle = \langle \lambda x, x \rangle = \langle T^*T(x), x \rangle = \langle T(x), T(x) \rangle \geq 0$$

Since $\operatorname{rank}(T) = r$ it follows by Theorem 3.8 that $\operatorname{rank}(T^*T) = r$. Clearly T^*T is self-adjoint, thus by Theorem 4.8 there exists an ordered orthonormal basis consisting of λ_i-eigenvectors of T^*T, say $\beta = \{v_1, \dots, v_n\}$. We can endow this basis with an ordering where $\lambda_1 \geq \lambda_2 \geq \dots \lambda_r > 0$ and $\lambda_{r+1} = \dots = \lambda_n = 0$. For $1 \leq i \leq r$, denote $\sigma_i = \sqrt{\lambda_i}$ and $w_i = \frac{1}{\sigma_i}T(v_i)$. We claim that the set $\{w_1, \ldots, w_r\}$ is an orthonormal subset of W. Suppose $1 \leq i, j \leq r$, then,

$$\begin{aligned}
\langle w_i, w_j \rangle &= \langle \frac{1}{\sigma_i}T(v_i), \frac{1}{\sigma_j}T(v_j) \rangle \\
&= \frac{1}{\sigma_i \sigma_j}\langle T^*T(v_i), v_j \rangle \\
&= \frac{1}{\sigma_i \sigma_j}\langle \lambda_i v_i, v_j \rangle \\
&= \frac{\sigma_i^2}{\sigma_i \sigma_j}\langle v_i, v_j \rangle \\
&= \delta_{ij} \quad \text{(since } \beta \text{ is an orthonormal set)}
\end{aligned}$$

Thus $\{w_1, \ldots, w_r\}$ is an orthonormal subset of W. We can extend this to an orthonormal basis for W, $\gamma = \{w_1, \ldots, w_r, \ldots, w_m\}$. By this construction $T(v_i) = \sigma_i w_i$ for $1 \leq i \leq r$ and $T(v_i) = 0$ for $i > r$. ■

Definition 7.4. The unique scalars $\sigma_1, \ldots, \sigma_r$ are called the singular values of T. If $r \leq m, n$ then we include $\sigma_{r+1} = \ldots = \sigma_k$ as singular values, where $k = \min(m, n)$.

We now develop a singular value analogue for matrices.

Theorem 7.5. *Let A be an $m \times n$ matrix of rank r with the positive singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r$, and let Σ be the $m \times n$ matrix defined by:*

$$\Sigma = \begin{cases} \sigma_i, & \textit{if } i = j \leq r \\ 0, & \textit{otherwise} \end{cases}$$

Then there exists an $m \times m$ unitary matrix U and an $n \times n$ unitary matrix V, such that

$$A = U\Sigma V^*$$

Definition 7.6. Let A be an $m \times n$ matrix of rank r with **positive singular values** $\sigma_1 \geq \ldots \geq \sigma_r$. A factorization $A = U\Sigma V^*$ where U and V are unitary matrices and E is the $m \times n$ matrix defined as above is called a **singular value decomposition** of A.

Definition 7.7. Pseudoinverse. Let V and W be finite-dimensional inner product spaces over the same field $\mathbb{F}$ and let $T : V \to W$ be a linear transformation. Let $L : N(T)^\perp \to R(T)$ be the restriction of L to the orthogonal complement of the null space of T; namely $L = T\Big|_{N(T)^\perp}$. (Note that L is an invertible linear transformation, even if T is not). The **pseudoinverse** of T, denoted $T^\dagger$, is defined as the unique linear transformation $T^\dagger : W \to V$ such that:

$$T^\dagger(y) = \begin{cases} L^{-1}(y), & \text{for } y \in R(T) \\ 0 & \text{for } y \in R(T)^\perp \end{cases}$$

Remark: If T is invertible, then $T^{-1} = T^\dagger$ since $N(T)^\perp = V$.

Theorem 7.8. *Let A be an $m \times n$ matrix with the singular value decomposition $A = U\Sigma V^*$. Let $\Sigma^\dagger$ be the $n \times n$ matrix defined by*

$$\Sigma^\dagger = \begin{cases} \frac{1}{\sigma_i}, & i = j \leq r \\ 0, & \textit{otherwise} \end{cases}$$

Then $A^\dagger = V\Sigma^\dagger U^*$ is the pseudoinverse of A.

Theorem 7.9. *Let V and W be finite-dimensional inner product spaces and let $T : V \to W$ be linear. Then,*

1. *$T^\dagger T$ is the orthogonal projection of V on $N(T)^\perp$.*
2. *$TT^\dagger$ is the orthogonal projection of W on $R(T)$.*

Proof. 1. We use the notation adopted in Definition 7.7. If $x \in N(T)^{\perp}$ then $T^{\dagger}T(x) = L^{-1}L(x) = I(x) = x$. If $x \in N(T)$ then $T^{\dagger}T(x) = T^{\dagger}(0) = 0$.

2. If $y \in R(T)$ then $TT^{\dagger}(y) = LL^{-1}(y) = I(y) = y$. If $y \in R(T)^{\perp}$ then $TT^{\dagger}(y) = T(0) = 0$. ■

Theorem 7.10. *Let $Ax = b$ be a system of linear equations where A is an $m \times n$ matrix and $b \in \mathbb{F}^n$. If $z = A^{\dagger}b$ then z has the following properties:*

1. *If $Ax = b$ is consistent, then z is the solution with least norm.*
2. *If $Ax = b$ is inconsistent, then z is the unique best approximation to a solution having minimum norm. That is, $\|Az - b\| \leq \|Ay - b\|$ for any $y \in \mathbb{F}^n$. Equality holds if and only if $Az = Ay$.*

Proof. Let $T = L_A$.

1. Since the system is consistent, $b \in R(T)$. Thus, $Az = AA^{\dagger}b = b$, so z is a solution. Suppose y is any solution to the system, then

$$T^{\dagger}T(y) = A^{\dagger}Ay = A^{\dagger}b = z$$

Thus z is the orthogonal projection of y on $N(T)^{\perp}$ by Theorem 7.9. Thus $\|z\| \leq \|y\|$ with equality only if $y = z$.

2. Since the system is inconsistent, $b \notin R(T)$. By Theorem 7.9, $Az = AA^{\dagger}b = TT^{\dagger}(b) = b$ is the orthogonal projection of b on $R(T)$. Thus Az is the vector on $R(T)$ closest to b. Thus $\|Az - b\| \leq \|Ay - b\|$. ■

Chapter 8

Bilinear and quadratic forms

Definition 8.1. Let V be a vector space over a field $\mathbb{F}$. A function B from the set $V \times V$ of ordered pairs of vectors to $\mathbb{F}$ is called a bilinear form on V if B is linear in each variable when the other variable is held fixed; that is, B is a bilinear form on V if:

$$B(ax + y, z) = aB(x, z) + B(y, z)$$
$$B(x, ay + z) = aB(x, y) + B(x, z)$$

Note that inner products over real fields are bilinear forms. Here are a few properties of all bilinear forms.

Theorem 8.2. *For any bilinear form B on a vector space V with field $\mathbb{F}$, the following are true:*

1. *We define the functionals $R_x, L_x : V \to \mathbb{F}$ as $R_x(y) = B(y, x)$ and $L_x(y) = B(x, y)$. Then L_x and R_x are linear.*
2. *$B(0, x) = B(x, 0) = 0$ for all $x \in V$*

Theorem 8.3. *The set of all bilinear forms $\mathcal{B}(V)$ on V forms a vector space.*

Proof. Since $\mathcal{B}(V)$ is a subset of the dual space of V, V^*, we only need to show to show that it is closed under vector addition, scalar multiplication and that it contains the zero bilinear form. These three things are indeed true. ■

Definition 8.4. Let $\beta = \{v_1, v_2, \ldots, v_n\}$ be an ordered basis for an n-dimensional vector space V and $B \in \mathcal{B}(V)$. The matrix A, where $A_{ij} = B(v_i, v_j)$ is the matrix representation of B with respect to β and is denoted $\psi_\beta(B)$.

Theorem 8.5. *For any n-dimensional vector space V over $\mathbb{F}$ and any ordered basis β for V, $\psi_\beta : \mathcal{B}(V) \to M_{n\times n}(\mathbb{F})$ is an isomorphism.*

Proof. Let $\beta = \{v_1, \ldots, v_n\}$ be an ordered basis for V. For any bilinear form $B \in \mathcal{B}(V)$, we have that $B : V \times V \to \mathbb{F}$. Thus, $\mathcal{B}(V) = \mathcal{L}(V \otimes V, \mathbb{F})$. We have that $\dim(\mathcal{L}(V \otimes V, \mathbb{F})) = \dim(V \otimes V) \cdot \dim \mathbb{F} = n^2 \cdot 1 = n^2$. Now let $B \in \mathcal{B}(V)$. Suppose $\psi_\beta(B) = 0$, where 0 denotes the zero matrix. Since B is completely determined by its action on a basis and $v_i, v_j \neq 0$ for all $i, j = 1, \ldots, n$, it follows that B is the zero bilinear form. This shows that ψ_β is injective. Since $\dim(\mathcal{B}(V)) = \dim(M_{n\times n}(\mathbb{F}))$, it follows from the rank-nullity theorem that ψ_β is an isomorphism. ■

From these facts it follows that $B(x, y) = x^t Ay$ where $A = \psi_\beta(B)$.

Definition 8.6. Let $A, B \in M_{n\times n}(\mathbb{F})$. Then B is said to be **congruent** to A if and only if there exists an invertible matrix $Q \in M_{n\times n}(\mathbb{F})$ such that $B = Q^t AQ$.

Theorem 8.7. *A bilinear form B on a vector space V is said to be symmetric if and only if $B(x, y) = B(y, x)$ for all $x, y \in V$.*

Theorem 8.8. *Let B be a bilinear form on a finite-dimensional vector space V, and let β be an ordered basis for V. Then B is symmetric if and only if $\psi_\beta(B)$ is symmetric.*

Proof. Let $\beta = \{v_1, \dots, v_n\}$ and $A = \psi_\beta(B)$.

($\Rightarrow$) Suppose B is symmetric. Then, for all $i, j = 1, \dots, n$, we have,

$$A_{ij} = B(v_i, v_j) = B(v_j, v_i) = A_{ji}$$

Thus, A is symmetric.

($\Leftarrow$) Conversely, suppose A is symmetric. Define a bilinear form J on V such that $J(x, y) = B(y, x)$. We proceed to show that the matrix representation of J and B is the same. Let $C = \psi_\beta(J)$. Then

$$C_{ij} = J(v_i, v_j) = B(v_j, v_i) = A_{ji} = A_{ij}$$

Thus, A and C agree on every entry; that is $A = C$. Since ψ_β is an isomorphism by Theorem 8.5, it follows that $B(x, y) = J(y, x) = B(y, x)$; thus, B is symmetric. ∎

Definition 8.9. A bilinear form B is said to be diagonalisable if and only if there exists an ordered basis β such that $\psi_\beta(B)$ is diagonal.

Theorem 8.10. *Let V be a finite dimensional vector space over a field $\mathbb{F}$ where the characteristic of the field is not 2. Let B be a bilinear form on V. Then, B is diagonalisable if and only if B is symmetric.*

Proof. ($\Rightarrow$) Suppose B is diagonalisable. Then, there exists a basis β such that $D = \psi_\beta(B)$ is diagonal. Since D is diagonal, D is symmetric. It follows trivially that B is a symmetric bilinear form.

($\Leftarrow$) For this part, we first prove a lemma.

Lemma. Let B be a non-zero symmetric bilinear form on a vector space V over a field $\mathbb{F}$ not of characteristic two. Then there is a vector x in V such that $B(x, x) \neq 0$.

Proof of lemma. Since B is non zero, we can choose $v_1, v_2 \in V$ such that $B(v_1, v_2) \neq 0$. If $B(v_1, v_1) \neq 0$ or $B(v_2, v_2) \neq 0$, we are done. Otherwise, let $y = v_1 + v_2$, then

$$B(y, y) = B(v_1, v_1) + B(v_1, v_2) + B(v_2, v_1) + B(v_2, v_2) = 2B(v_1, v_2) \neq 0$$

since $2 \neq 0$ and $B(v_1, v_2) \neq 0$. *(QED lemma)*

Now we proceed with the prove of the converse by using mathematical induction on the dimension of V. For $n = 1$, the result is trivial (since all 1×1 matrices are diagonal, duh), so the base case holds. Now, suppose that B is diagonalisable for $n - 1$ for some fixed integer $n > 1$. If B is the zero bilinear form, then its trivially diagonalisable. So suppose that B is a non-zero symmetric bilinear form. By our lemma, there exists some vector $x \in V$ such that $B(x, x) \neq 0$. Define $L_x : V \to \mathbb{F}$ as $L_x(y) = B(x, y)$ for all $y \in V$ (L stands for left). Since $L_x(x) = B(x, x) \neq 0$, it follows that $\operatorname{rank}(L_x) = 1$. By the rank-nullity theorem, $\operatorname{nullity}(L_x) = n - 1$.

Clearly, if B is symmetric on V, it is symmetric on subspaces of V. In particular, the restriction of B to $N(L_x)$ is symmetric. By the inductive hypothesis, there exists an ordered basis $\{v_1, \dots, v_{n-1}\}$ for $N(L_x)$ such that B is diagonalisable (that is $B(v_i, v_j) = 0$ for $i \neq j$). Set $v_n = x$. By our construction, $x \notin N(L_x)$, so $\beta = \{v_1, \dots, v_n\}$ is an ordered basis for V. In addition, since B is symmetric we have that for $i = 1, \dots, n - 1$,

$$\psi_\beta(B)_{in} = B(v_i, v_n) = B(v_n, v_i) = B(x, v_i) = L_x(v_i) = 0$$

Thus, the matrix representation of B is diagonal. That is, B is diagonalisable, and we are done! ∎

Definition 8.11. Let V be a vector space over $\mathbb{F}$. A function $K : V \times V \to \mathbb{F}$ is called a **quadratic form** if and only if there exists a bilinear form B such that

$$K(x) = B(x, x) \qquad \text{for all } x \in V$$

Theorem 8.12. *Let V be a vector space over a field $\mathbb{F}$ not of characteristic two and let B be a symmetric bilinear form on V. If $K(x) = B(x,x)$ is the quadratic form associated with B, then for all $x, y \in V$,*

$$B(x,y) = \frac{1}{2}\left[K(x+y) - K(x) - K(y)\right]$$

Proof. We show that the right-hand side is equal to the left-hand side. We have,

$$\begin{aligned}
\frac{1}{2} &= [K(x+y) - K(x) - K(y)] \\
&= \frac{1}{2}[B(x+y, x+y) - B(x,x) - B(y,y)] \\
&= \frac{1}{2}[B(x,x) + B(x,y) + B(y,x) + B(y,y) - B(x,x) - B(y,y)] \\
&= \frac{1}{2}[B(x,y) + B(y,x)] \\
&= \frac{1}{2}[2B(x,y)] \qquad \text{(since } B \text{ is symmetric)} \\
&= B(x,y)
\end{aligned}$$

■

Theorem 8.13. *Let V be a finite-dimensional real inner product space, and let B be a symmetric bilinear form on V. Then there exists an orthonormal basis β for V such that $\psi_\beta(B)$ is a diagonal matrix.*

Proof. Choose an orthonormal basis $\gamma = \{v_1, \ldots v_n\}$ for V and let $A = \psi_\gamma(B)$. Since A is symmetric, it is orthogonally diagonalisable. That is, there exists an orthogonal matrix Q such that $D = Q^t AQ$, where D is diagonal. Let $\beta = \{w_1, \ldots, w_n\}$ be defined by

$$w_j = \sum_{i=1}^{n} Q_{ij} v_i \qquad \text{for } 1 \leq j \leq n$$

Since Q is orthogonal and γ is orthonormal, β is an orthonormal basis that diagonalises B; that is $\psi_\beta(B) = D$ is diagonal. ■

Theorem 8.14. Sylvester's Law of Inertia *Let B be a symmetric bilinear form on a finite-dimensional real vector space V. Then, the number of positive diagonal entries and negative diagonal entries in any diagonal representation of B are each independent of the diagonal representation.*

Definition 8.15. The number of positive diagonal entries in a diagonal representation of a symmetric bilinear form on a real vector space is called the **index** of the form. The difference between the number of positive and the number of negative diagonal entries in a diagonal representation of a symmetric bilinear form is called the **signature** of the form. The three terms, rank, index, and signature are called the **invariants** of the bilinear form because they are invariant with respect to matrix representations. These same terms apply to the associated quadratic form. Notice that the values of any two of these invariants determine the value of the third.

Theorem 8.16. *Two real $n \times n$ matrices are congruent if and only if they have the same invariants.*

Proof. ($\Rightarrow$) If A and B are congruent, then they are both congruent to the same diagonal matrix and thus have the same invariants.

Part III

Canonical Forms

Chapter 9

The Jordan Canonical Form - Theory

One of the sad facts about life is that not all matrices are diagonalisable. That's a shame, because diagonal matrices are cool. Luckily, though, when the characteristic polynomial of a linear operator splits, we can find a matrix almost as good as a diagonal matrix, called the Jordan canonical form.

Unfortunately, I was not able to complete the class notes for this section. Instead, I have included solutions to some exercise problems.

Definition 9.1. Suppose that V is a finite-dimensional vector space over a field $\mathbb{F}$. Let T be a linear operator on V and suppose its characteristic polynomial splits and has eigenvalues $\lambda_1, \ldots, \lambda_r$. Then a matrix of the form

$$A = \begin{bmatrix} \lambda & 1 & 0 & \ldots & 0 & 0 \\ 0 & \lambda & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \lambda & 1 \\ 0 & 0 & 0 & \ldots & 0 & \lambda \end{bmatrix}$$

is called a **Jordan block**. A matrix of the form

$$[T]_\beta = \begin{bmatrix} A_1 & O & \ldots & O \\ O & A_2 & \ldots & O \\ \vdots & \vdots & & \vdots \\ O & O & \ldots & A_k \end{bmatrix}$$

is called a **Jordan canonical Form**. We say that the ordered basis β is a **Jordan canonical basis**

Definition 9.2. Let T be a linear operator on a vector space V, and let λ be an eigenvalue. A non-zero vector x in V is called a **generalized eigenvector** of T corresponding to λ if $(T - \lambda I)^p)(x) = 0$ for some positive integer p.

Definition 9.3. Let T be a linear operator on a vector space V, and let λ be an eigenvalue of T. The **generalized eigenspace** of T corresponding to λ, denoted K_λ, is the subset of V defined as:

$$K_\lambda = \{x \in V : (T - \lambda I)^p(x) = 0 \text{ for some positive interger } p\}$$

Theorem 9.4. *Let T be a linear operator on a vector space V and let λ be an eigenvalue of T. Then K_λ is a T-invariant subspace of V containing E_λ (the λ-eigenspace).*

Proof. First, we prove that K_λ is a subspace of V. Clearly $0 \in K_\lambda$ since $(T - \lambda I)(0) = 0$. Let $x \in K_\lambda$ and let c be a scalar. Since $x \in K_\lambda$, $(T - \lambda I)^p(x) = 0$. But since $(T - \lambda I)^p$ is a linear transformation, $(T - \lambda I)^p(cx) = c(T - \lambda I)^p(x) = c \cdot 0 = 0$. Thus $cx \in K_\lambda$. Finally, we show closure under vector addition. Let $x, y \in K_\lambda$. Then, for some positive integers p, q,

$$(T - \lambda I)^p(x) = 0 \qquad (T - \lambda I)^q(y) = 0$$

Notice that $(T-\lambda I)^p$ and $(T-\lambda I)^q$ are polynomials in T, so they commute; thus we have,

$$\begin{aligned}(T-\lambda I)^{p+q}(x+y) &= (T-\lambda I)^{p+q}(x)+(T-\lambda I)^{p+q}(y) \\ &= (T-\lambda I)^q(T-\lambda I)^p(x)+(T-\lambda I)^p(T-\lambda I)^q(y) \\ &= (T-\lambda I)^q(0)+(T-\lambda I)^p(0) \\ &= 0\end{aligned}$$

Thus, $(x+y)\in K_\lambda$. Since 0 is contained in the set, and it is closed under scalar multiplication and vector addition, K_λ is a subspace of V.

We are left to show that K_λ is T-invariant. Let $x\in K_\lambda$, then for some p, $(T-\lambda I)^p(x)=0$. But,

$$(T-\lambda I)^pT(x)=T(T-\lambda I)^p(x)=T(0)=0$$

Thus, $T(x)\in K_\lambda$; that is K_λ is T-invariant. ■

Theorem 9.5. *Let T be a linear operator on a vector space V and let λ be an eigenvalue of T. Then, if $\mu\neq\lambda$, the $(T-\mu I)\Big|_{K_\lambda}$ is an isomorphism.*

Proof. Let $x\in K_\lambda$ and $(T-\mu I)(x)=0$. Arguing by contradiction, suppose that $(T-\mu I)|_{K_\lambda}$ is not injective and say $x\neq 0$. Since $x\in K_\lambda$, let p be the smallest integer such that $(T-\lambda I)^p(x)=0$. Let $y\in(T-\lambda I)^{p-1}(x)$. Clearly, y is a λ-eigenvector, so $y\in E_\lambda$. Additionally, we have,

$$(T-\mu I)(y)=(T-\mu I)(T-\lambda I)^{p-1}(x)=(T-\lambda I)^{p-1}(T-\mu I)(x)=(T-\lambda I)^{p-1}(0)=0$$

Thus, $y\in E_\mu$. Since $y\in E_\mu\cap E_\lambda=\{0\}$, $y=0$. Then, $x=0$, a contradiction, thus $(T-\mu I)\Big|_{K_\lambda}$ is injective. But since $(T-\mu I)$ is an operator, by the rank-nullity theorem we have that it is full rank, and thus it is an isomorphism. ■

Theorem 9.6. *Let T be a linear operator on a finite-dimensional vector space V such that the characteristic polynomial of T splits. Suppose that λ is an eigenvalue of T with multiplicity m. Then:*

1. $\dim(K_\lambda)=m$
2. $K_\lambda=N((T-\lambda I)^m)$

Theorem 9.7. *Let T be a linear operator on a finite-dimensional vector space V such that the characteristic polynomial of T splits, and let $\lambda_1,\ldots,\lambda_k$ be the distinct eigenvalues of T with corresponding multiplicities $m_1,m_2,\ldots,m_k$. For $1\le i\le k$, let β_i be an ordered basis for K_{λ_i}. Then the following statements are true.*

1. *$\beta_i\cap\beta_j=\emptyset$ for $i\neq j$.*
2. *$\beta=\beta_1\cup\ldots\cup\beta_k$ is an ordered basis for V.*

Theorem 9.8. *Let T be a linear operator on a finite-dimensional vector space V such that the characteristic polynomial of T splits. Then T is diagonalisable if and only if $E_\lambda=K_\lambda$ for every eigenvalue λ of T.*

Definition 9.9. Let T be a linear operator on a vector space V and let x be a generalized eigenvector of T corresponding to the eigenvalue λ. Suppose that p is the smallest positive integer for which $(T-\lambda I)^p(x)=0$. Then the ordered set

$$\{(T-\lambda I)^{p-1}(x),\ldots(T-\lambda I)(x),x\}$$

is called a **cycle of generalised eigenvectors** (or, as the cool kids say, **chains**) corresponding to λ. The vector $(T-\lambda I)^{p-1}$ is called the initial vector of the cycle. The length of the cycle is p.

Theorem 9.10. *Let T be a linear operator on a finite-dimensional vector space V whose characteristic polynomial splits, and suppose that β is a basis for V such that β is a disjoint union of cycles of generalized eigenvectors of T. Then the following statements are true.*

1. *For each cycle γ of generalised eigenvectors contained in β, $W = \mathrm{span}(\gamma)$ is T-invariant, and $[T_W]_\gamma$ is a Jordan block.*

2. *β is a Jordan Canonical basis for V*

Theorem 9.11. *Let T be a linear operator on a vector space V, and let λ be an eigenvalue of T. Suppose that $\gamma_1, \ldots, \gamma_q$ are cycles of generalised eigenvectors of T corresponding to λ such that the initial vectors of the γ_i's are distinct and form a linearly independent set. Then the γ_i's are disjoint and their union*

$$\gamma = \bigcup_{i=1}^{q} \gamma_i$$

is linearly independent.

Proof. We prove that the cycles are disjoint. Say $x \in \gamma_1$ and $x \in \gamma_2$. Since they are in cycles, there exists a smallest integer q such that $(T - \lambda I)^q(x) = 0$. Thus $(T - \lambda I)^{q-1}(x)$ is an eigenvector of T contained in two cycles. This is a contradiction, thus cycles with distinct initial vectors are disjoint. ■

Theorem 9.12. *Every cycle of generalised eigenvectors of a linear operator T is linearly independent.*

Theorem 9.13. *Let T be a linear operator on a finite-dimensional vector space V, and let λ be an eigenvalue of T. Then K_λ has an ordered basis consisting of a union of disjoint cycles of generalised eigenvectors corresponding to λ.*

Theorem 9.14. *Let T be a linear operator on a finite-dimensional vector space V whose characteristic polynomial splits. Then V is the direct sum of the generalised eigenspaces of T.*

Theorem 9.15. *Let T be a linear operator on a finite-dimensional vector space V and let γ be a cycle of generalised eigenvectors that corresponds to the eigenvalue λ. Then $\mathrm{span}(\gamma)$ is a T-invariant subspace of V.*

Proof. Naturally, $W = \mathrm{span}(\gamma)$ is $T - \lambda I$ invariant. Thus,

$$T(w) = (T - \lambda I)(w) + \lambda I(w) = (T - \lambda I)(w) + \lambda w \in W$$

■

Chapter 10

Jordan Canonical Form - Computation

Theorem 10.1. *For any positive integer r, the vectors in β_i that are associated with the dots in the first r rows of the dot diagram of T_i constitute a basis for $N((T - \lambda_i I)^r)$. Hence the number of dots in the first r rows of the dot diagram equals* $\text{nullity}((T - \lambda I)^r)$.

Theorem 10.2. *Let r_j denote he number of dots in the j-th row of the dot diagram of T_i, the restriction of T to K_{λ_i}. Then, the following statements are true.*

1. $r_1 = \dim(V) - \text{rank}(T - \lambda_i I)$
2. $r_j = \text{rank}((T - \lambda_i I)^{j-1}) - \text{rank}((T - \lambda_i I)^j) \quad j > 1$

Theorem 10.3. *For any eigenvalue λ_i of T, the dot diagram of T_i is unique. Thus, subject to the convention that the cycles of generalised eigenvectors for the bases of each generalized eigenspace are listed in order of decreasing length, the Jordan canonical form of a linear operator or a matrix is unique up to the ordering of the eigenvalues.*

Theorem 10.4. *Let A and B be $n \times n$ matrices, each having Jordan canonical forms computed according to the conventions of this section. Then A and B are similar if and only if they have (up to an ordering of their eigenvalues) the same Jordan canonical form.*

Theorem 10.5. *Let A be an $n \times n$ matrix whose characteristic polynomial splits. Then A and A^t have the same Jordan Canonical Form and they are similar.*

Proof. Since $(A - \lambda I)^t = (A^t - \lambda I)$ and rank is invariant under transposes, we have that $\text{rank}((A - \lambda I)^r) = \text{rank}((A^t - \lambda I)^r)$. Since the JCF is completely determined by its dot diagram and the dot diagram is completely determined by the difference between the ranks of the subsequent powers of $A - \lambda I$ (see Theorem 10.2), it follows that they have the same JCF. Since they have the same JCF, they are similar. ■

Definition 10.6. A linear operator T on a vector space V is called nilpotent if $T^p = T_0$ for some positive integer p. An $n \times n$ matrix A is called nilpotent if $A^p = O$ for some positive integer p.

Theorem 10.7. *Any strictly upper triangular matrix is nilpotent.*

A strictly upper triangular matrix is a matrix which is upper triangular and all its diagonal entries are zero. The characteristic polynomial of A is $p_A(t) = (-1)^n t^n$. Thus, by the Cayley-Hamilton theorem, $A^n = 0$. ■

Theorem 10.8. *Let T be a linear operator on a finite-dimensional vector space V, and let J be the Jordan canonical form of T. Let D be the diagonal matrix whose diagonal entries are the diagonal entries of J, and let $M = J - D$. Then, the following are true.*

1. *M is nilpotent.*
2. $MD = DM$

3. *If p is the smallest integer for which $M^p = O$, then, for any positive integer $r < p$:*

$$J^r = D^r + rD^{r-1}M + \frac{r(r-1)}{2!}D^{r-2}M^2 + \ldots + rDM^{r-1} + M^r$$

and for any positive integer $r \geq p$,

$$J^r = D^r + rD^{r-1}M + \frac{r(r-1)}{2!}D^{r-2}M^2 + \ldots + \frac{r!}{(r-p+1)!(p-1)!}D^{r-p+1}M^{p-1}$$

Proof.

1. M is strictly upper triangular. Thus, it is nilpotent.

2. Since M only has nonzero entries in its supradiagonal, we obtain,

$$(MD)_{ij} = \sum_{k=1}^{n} M_{ik}D_{kj} = M_{i(i+1)}D_{(i+1)j} = D_{(i+1)j}M_{i(i+1)} = \sum_{k=1}^{n} D_{ik}M_{kj} = (DM)_{ij}$$

3. For the first part, by the binomial theorem we obtain,

$$J^r = (M+D)^r = \sum_{i=0}^{r} \binom{r}{i} M^{r-i}D^i = D^r + rD^{r-1}M + \frac{r(r-1)}{2!}D^{r-2}M^2 + \ldots + rDM^{r-1} + M^r$$

The second equality follows from the fact that $M^r = 0$, since M is nilpotent. ■

Theorem 10.9. *Let*

$$J = \begin{bmatrix} \lambda & 1 & 0 & \ldots & 0 \\ 0 & \lambda & 1 & \ldots & 0 \\ 0 & 0 & \lambda & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ 0 & 0 & 0 & \ldots & \lambda \end{bmatrix}$$

be the $m \times m$ Jordan block corresponding to λ, and let $N = J - \lambda I_m$. Then, the following are true:

1. *$N^m = O$, and for $1 \leq r < m$,*

$$N_{ij}^r = \begin{cases} 1 & \text{if } j = i + r \\ 0 & \text{otherwise} \end{cases}$$

2. *For any integer $r \geq m$:*

$$J^r = \begin{bmatrix} \lambda^r & r\lambda^{r-1} & \frac{r(r-1)}{2!}\lambda^{r-1} & \ldots & \frac{r(r-1)\ldots(r-m+2)}{(m-1)!}\lambda^{r-m+1} \\ 0 & \lambda^r & r\lambda^{r-1} & \ldots & \frac{r(r-1)\ldots(r-m+3)}{(m-2)!}\lambda^{r-m+2} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & \lambda^r \end{bmatrix}$$

3. *The limit*

$$\lim_{r\to\infty} J^r$$

exists if and only if $|\lambda| < 1$ or $\lambda = 1$ and $m = 1$.

Proof.

1. Clearly N is nilpotent. It remains to show that iterating powers kills N after m iterations. Naturally this happens since right-multiplication by N shifts vectors to the left, since N is strictly upper triangular. All vectors will be shifted after m iterations, thus yielding $N^m = O$.

2. Follows from Theorem 10.8.

3. If $|\lambda| < 1$, then the limit tends to the zero matrix. If $\lambda = 1$ and $m = 1$, then the limit tends to the one-dimensional identity. Conversely, if $|\lambda| > 1$, then the diagonal entries will not converge. The result follows. ■

Chapter 11

Minimal Polynomials

Definition 11.1. Let T be a linear operator on a finite-dimensional vector space. A polynomial $p(t)$ is called a **minimal polynomial** of T if $p(t)$ is a monic polynomial of least positive degree for which $p(T) = T_0$

Theorem 11.2. *Let $p(t)$ be a minimal polynomial of a linear operator T on a finite-dimensional vector space V.*

1. *For any polynomial $g(t)$, if $g(T) = T_0$ then $p(t)$ divides $g(t)$. In particular $p(t)$ divides the characteristic polynomial of T.*

2. *The minimal polynomial of T is unique.*

Theorem 11.3. *Let T be a linear operator on a finite-dimensional vector space V, and let $p(t)$ be the minimal polynomial of T. A scalar λ is an eigenvalue of T if and only if $p(\lambda) = 0$ Hence the characteristic polynomial and the minimal polynomial of T have the same zeroes.*

Theorem 11.4. *Let T be a linear operator on an n-dimensional vector space V such that V is a T-cyclic subspace of itself. Then, the characteristic polynomial $f(t)$ and the minimal polynomial $p(t)$ have the same degree and $f(t) = (-1)^n p(t)$.*

Theorem 11.5. *Let T be a linear operator on a finite-dimensional vector space V. Then T is diagonalisable if and only if the minimal polynomial of T is of the form*

$$m(t) = (t - \lambda_1) \dots (t - \lambda_k)$$

where $\lambda_1, \dots, \lambda_k$ are the distinct eigenvalues of T.

Theorem 11.6. *Let T be a diagonalisable linear operator on a finite-dimensional vector space V. Then V is a T-cyclic subspace if and only if each of the eigenspaces of T is one-dimensional.*

We solve this by a sequence of if and only if statements.

$$\begin{aligned} V \text{ is a } T\text{-cyclic subspace} &\Leftrightarrow m(t) = (-1)^n p(t) \\ &\Leftrightarrow p(t) = \prod_i (t - \lambda_i) \qquad \text{(since } T \text{ is diagonalisable)} \\ &\Leftrightarrow \text{Each eigenspace of } T \text{ is one-dimensional} \end{aligned}$$

■

Theorem 11.7. *Let T be a linear operator on a finite-dimensional vector space V, and suppose that W is a T-invariant subspace of V. Then the minimal polynomial of T_W divides the minimal polynomial of T.*

Let $p(t)$ be the minimal polynomial of T. Then $p(T_W) = p(T) = T_0$. Thus since $p(t)$ is a polynomial that kills T_W, it follows that the minimal polynomial of T_W divides $p(t)$. ■

Theorem 11.8. *Let T be a linear operator on a finite-dimensional vector space, and suppose that the characteristic polynomial of T splits. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T, and for each i let p_i be the order of the largest Jordan block corresponding to λ_i in a Jordan canonical form of T. Then the minimal polynomial of T is*

$$\prod_{i=1}^{k}(t-\lambda_i)^{p_i}$$

Definition 11.9. Let T be a linear operator on a finite-dimensional vector space V, and let x be a nonzero vector in V. The polynomial $p(t)$ is called a T-**annihilator** of x if $p(t)$ is a monic polynomial of least degree for which $p(T)(x) = 0$.

Theorem 11.10. *Let T be a linear operator on a finite dimensional vector space V, and let x be a nonzero vector in V. Then the following are true:*

1. *The vector x has a unique T-annihilator.*
2. *The T-annihilator of x divides any polynomial $g(t)$ for which $g(T) = T_0$*
3. *If $p(t)$ is the T-annihilator of x and W is the T-cyclic subspace generated by x, then $p(t)$ is the minimal polynomial of T_W and $\dim W$ equals the degree of $p(t)$.*
4. *The degree of the T-annihilator of x is 1 if and only if x is an eigenvector of T.*

Proof. We prove these in one go. First, suppose $p(t)$ and $q(t)$ are T-annihilators of x. Then $p|q$ and $q|p$, thus $p(t) = q(t)$ and the annihilator is unique.

Now let $x \in W$, $\dim W = m$ and let C_x be the T-cyclic subspace generated by x and T. Then the characteristic polynomial of T_W is

$$p_{T_W}(t) = (-1)^m(a_0 + a_1 t + \ldots + a_m t^m)$$

where a_i are the negative coefficients of the rightmost column of T with respect to the basis $\beta = \{x, T(x), \ldots, T^{m-1}(x)\}$. Now, arguing by contradiction, suppose the minimal polynomial has degree k where $k < m$. By the Wu quick-eye method of proof, we see that this would lead to the basis for W being linearly dependent, a contradiction. Thus, the T-annihilator of x is simply the minimal polynomial of T_W.

This makes the other statements trivial. ■

Part IV

Computations and other Distractions

11.1 Some interesting inner products, norms, and orthogonal complements

Example 11.11. The **Frobenius Inner Product** is the standard inner product for $V = M_{n\times n}(\mathbb{F})$. It is defined as $\langle A, B\rangle = tr(B^*A)$.

Example 11.12. Let H be the vector space of all complex-values functions on $[0, 2\pi]$ whose inner product is

$$\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(t)\overline{g(t)}dt$$

Denote $f_n(t) = e^{int}$ and $S = \{f_n : n \text{ is an integer}\}$. Then S is an orthonormal subset of H.

Example 11.13. Let $V = C([-1, 1])$, that is the set of all continuous functions on $[-1, 1]$. Denote W_e and W_o as the subspaces of even and odd functions, respectively. Then, $W_e^\perp = W_o$ provided that

$$\langle f, g\rangle = \int_{-1}^{1} f(t)g(t)dt$$

Proof. Fix $f \in W_o$. Then, for any even function $g \in W_e$, we have $f(-t)g(-t) = -f(t)g(t)$. That is, $f(t)g(t)$ is odd. Thus, $\langle f, g\rangle = 0$ since this is a symmetric integral of an odd function. Thus, $W_o \subseteq W_e^\perp$.

Conversely, let $h \in W_e^\perp$. Since h is a continuous function, we can write it as the sum of an even function f and an odd function g. That is, $h(t) = f(t) + g(t)$ where

$$f(t) = \frac{1}{2}(h(t) + h(-t)) \qquad g(t) = \frac{1}{2}(h(t) - h(-t))$$

Since $h \in W_e^\perp$ and $f \in W_e$, we have

$$0 = \langle h, f\rangle = \langle f + g, f\rangle = \langle f, f\rangle + \langle f, g\rangle = \langle f, f\rangle = \|f\|^2$$

where $\langle f, g\rangle = 0$ by the forward direction. Thus $f = 0$ and h is odd. That is $W_e^\perp \subseteq W_o$. Set equality follows. ■

Example 11.14. Define the inner product

$$\langle f, g\rangle = \int_{-1}^{1} f(t)g(t)dt$$

If we apply an Gram-Schmidt process on the standard ordered basis for $P_n(\mathbb{F})$ we obtain the **Legendre polynomials**. The first few (normalised) Legendre polynomials are

1. $P_0(x) = \frac{1}{\sqrt{2}}$
2. $P_1(x) = \sqrt{\frac{3}{2}}x$
3. $P_2(x) = \sqrt{\frac{5}{8}}(3x^2 - 1)$
4. $P_3(x) = \frac{\sqrt{7}}{2\sqrt{2}}(5x^3 - 3x)$
5. $P_4(x) = \frac{3}{8\sqrt{2}}(35x^4 - 30x^2 + 3)$
6. $P_5(x) = \frac{\sqrt{11}}{8\sqrt{2}}(63x^5 - 70x^3 + 15x)$

If the inner product is modified to obtain the integral in the interval $(0, 1)$ then we obtain the shifted Legendre polynomials, as in A01Q6.

11.2 Adjoints, self-adjoints, normal, and unitary operators

Example 11.15. Let V be an inner product space, and let $y, z \in V$. Define $T : V \to V$ by $T(x) = \langle x, y\rangle z$ for all $x \in V$. First prove that T is linear. Then show that T^* exists, and find an explicit expression for it.

Proof. Linearity is trivial. Note that for all $x \in V$

$$\langle x, T^*(w)\rangle = \langle T(x), w\rangle = \langle\langle x, y\rangle z, w\rangle = \langle x, y\rangle\langle z, w\rangle = \langle x, \overline{\langle z, w\rangle} y\rangle = \langle x, \langle w, z\rangle y\rangle$$

Since this is true for all x, is follows that $T^*(w) = \langle w, z\rangle y$. ∎

Example 11.16. Give an example of a linear operator T and an ordered basis β such that T is normal, but $[T]_\beta$ is not normal.

One such example is $T(a, b) = (a, 2b)$ and $\beta = \{(1, 1), (0, 1)\}$.

Example 11.17. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a rotation by $0 < \theta < \pi$. Clearly, $\|T(x)\| = \|x\|$. Let σ be the standard basis for our space, then

$$[T]_\sigma = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

Clearly, T is normal, but not self-adjoint (we notice it is not self adjoint since it does not have any eigenvectors). Also, it is clear that,

$$[T^*]_\sigma = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

that is, the rotation by $-\theta$. This fully characterises a rotation as an orthogonal transformation.

Example 11.18. We fully characterise reflections in $\mathbb{R}^2$ with respect to the standard ordered basis σ. Let T be a reflection across the line L. Let α be the angle between the positive x-axis and L. Let $\gamma = \{v_1, v_2\}$ be an orthogonal basis for $\mathbb{R}^2$ where $T(v_1) = v_1$ and $T(v_2) = -v_2$. Thus, we can express $v_1 = (\cos\alpha, \sin\alpha) \in L$ and $v_2 = (-\sin\alpha, \cos\alpha) \in L^\perp$, by our definition of reflections. Thus

$$[T]_\gamma = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Now, using a change of basis matrix (note that this is the orthogonal matrix whose columns are simply v_1 and v_2, we let

$$Q = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$$

Then, expressing $[T]_\sigma$ using a change of basis, we obtain:

$$[T]_\sigma = Q[T]_\gamma Q^{-1} = \begin{bmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{bmatrix}$$

Example 11.19. Suppose that T and U are reflections of $\mathbb{R}^2$ about the respective lines L and L' through the origin and that ϕ and ψ are the angles from the positive x-axis to L and L'. respectively. Then UT is a rotation. Find its angle of rotation.

Let σ be the standard ordered basis for $\mathbb{R}^2$. Using the characterisation above, we have,

$$[T]_\sigma = \begin{bmatrix} \cos 2\phi & \sin 2\phi \\ \sin 2\phi & -\cos 2\phi \end{bmatrix} \qquad \text{and} \qquad [U]_\sigma = \begin{bmatrix} \cos 2\psi & \sin 2\psi \\ \sin 2\psi & -\cos 2\psi \end{bmatrix}$$

We multiply these two matrices to obtain,

$$[UT]_\sigma = [U]_\sigma[T]_\sigma = \begin{bmatrix} \cos 2(\psi-\phi) & -\sin 2(\psi-\phi) \\ \sin 2(\psi-\phi) & \cos 2(\psi-\phi) \end{bmatrix}$$

11.3 Jordan and Rational Canonical Thingies

Example 11.20. Let $V = M_{n\times n}(\mathbb{R})$ and $T(A) = A^t$. Find the minimal polynomial of T

We spot that $T^2 = I$ and $T^2 - I = T_0$. Thus, the minimal polynomial must divide the polynomial $p(t) = t^2 - 1 = (t-1)(t+1)$. Clearly $(T - I) \neq T_0$ and $T + I \neq T_0$, thus the minimal polynomial is $m(t) = t^2 - 1$. ■

Example 11.21. Describe all linear operators T on $\mathbb{R}^2$ such that T is diagonalisable and $T^3 - 2T^2 + T = T_0$.

The polynomial we seek to satisfy is

$$t^3 - 2t^2 + t = t(t-1)^2$$

Clearly T_0 and I satisfy it. Now, since T is diagonalisable, we seek for the minimal polynomial to be $m(t) = t(t-1)$. Thus, we complete our characterisation by letting

$$[T]_\beta = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

for some basis β. ■

Example 11.22. Let T be a linear operator on a finite-dimensional vector space V, and let W_1 and W_2 be T-invariant subspaces of V such that $V = W_1 \oplus W_2$. Suppose that $p_1(t)$ and $p_2(t)$ are the minimal polynomials of T_{W_1} and T_{W_2}, respectively. Prove or disprove that $p_1(t)p_2(t)$ is the minimal polynomial of T.

It is false. Let $T = I$, W_1 the x-axis and W_2 the y-axis. The minimal polynomial of T_{W_i} is $t - 1$, and so is the minimal polynomial of I on V. ■