Elijah Thompson

Linear Algebra: A Complete Guide

– In Pursuit of Abstract Nonsense –

Sunday 24th April, 2022

Springer Nature

Preface

This text consists of a collection of linear algebra notes taken at the University of Calgary.

<table><tr><td>Place(s),
month year</td><td align="right">Firstname Surname
Firstname Surname</td></tr></table>

Contents

Part I

Vector Spaces

Chapter 1

Vector Spaces

In this chapter we shall build up the basic formulation of abstract linear algebra in terms of vector spaces and their basic formulations.

1.1 Vector Space Basics

Before defining a vector space we need the notion of a field.

Definition 1.1 A triple $(F, +, \cdot)$ of a set F with two binary operations $+ : F\times F \to F$ and $\cdot : F\times F \to F$ satisfying the following properties is known as a field.

- The pair $(F, +)$ is an abelian group with identity 0_F
- The pair $(F^*, \cdot)$ is an abelian group with identity 1_F, with $F^* = F\backslash\{0_F\}$
- For any $a, b, c \in F$, the distributive laws hold

$$a \cdot (b + c) = a \cdot b + a \cdot c, \quad (b + c) \cdot a = b \cdot a + c \cdot a$$

Moving forward we shall denote a general field by k. Now we have the notion of a vector space.

Definition 1.2 A vector space is a quadruple $(V, +, k, \cdot)$ where $(V, +)$ is an abelian group with identity 0_V, and k is a field wtih a left field action $\cdot : k \times V \to V$ known as **scalar multiplication** satisfying the following properties. For all $a, b \in F$ and all $u, v \in V$ we have distributivity

$$a \cdot (u + v) = a \cdot u + a \cdot v, \quad (a + b) \cdot u = a \cdot u + b \cdot u$$

associativity

$$(ab) \cdot u = a \cdot (b \cdot u)$$

and identity

$$1_k \cdot u = u$$

The collection of all (small) vector spaces form a category $\mathbf{Vect}_k$ with maps to be defined in Chapter 2. As with all algebraic structures we have the notion of a **subspace** which consists of a subset $U \subseteq V$ which is a vector space under the restrictions of the operations on V. We shall write $U \leq V$ to denote a subspace, and $U < V$ to denote a proper subspace.

Example 1.1 A few common examples of vector spaces and subspaces are as follows.

- The field k is a vector space over itself with multiplication in place of scalar multiplication.
- For any set X and any vector space $V \in \mathbf{Vect}_k$, the set of functions V^X from X to V inherits a vector space structure from V with pointwise addition and scalar multiplication of functions. A notable subspace is the space $(V^X)_0$ of such functions with finite support.
- A special case is the space $V^{\mathbb{N}}$ of all sequences in V and the subspace $(V^{\mathbb{N}})_0$ of all sequences with finite support. For example in $\mathbb{C}^{\mathbb{N}}$ we have the subspace ℓ^∞ of all bounded complex sequences, or for any $p > 1$ we have the subspace ℓ^p of all complex sequences $(z_n)_{n\in\mathbb{N}}$ for which $\left(\sum_{n=1}^\infty |z_n|^p\right)^{1/p} < \infty$.
- The set $M_{m,n}(k)$ of all $m \times n$ matrices with entries in k is a vector space over k with element wise addition and scalar multiplication.
- A special case is the space $k^n = M_{n,1}(k)$ of n-tuples whose components lie in k.
- The polynomial ring $k[X]$ in one indeterminate is a vector space over k with standard addition and multiplication by scalars. It is often denoted $\mathbb{P}(k)$. Additionally, for any $n \in \mathbb{N}$ the space $\mathbb{P}_n(k)$ of polynomials of degree less than or equal to n is a subspace.

Subspaces hold a special structure in vector spaces. If $V \in \mathbf{Vect}_k$ let $\mathcal{S}(V)$ denote the collection of all subspaces of V. Then $\mathcal{S}(V)$ is partially ordered by inclusion and contains both a minimal element $\{0\}$ and a maximal element V. In addition, for $S, T \in \mathcal{S}(V)$ the intersection $S \cap T$ forms the largest subspace of V containing both S and T, or in other words the greatest lower bound of S and T with respect to inclusion. Similarly, for any family $\{S_\alpha\}_{\alpha\in I}$ we can form the intersection $\bigcap_{\alpha\in I} S_\alpha$ which is the greatest lower bound of the family. Similarly, we also have a notion of least upper bound formed by **sums**.

Definition 1.3 Let $S, T \leq V$. The **sum** of S and T is defined to be

$$S + T := \{u + v | u \in S, v \in T\}$$

In general, for a family $\{S_\alpha\}_{\alpha\in I}$ the sum is the set of all finite sums of vectors from $\bigcup_{\alpha\in I} S_\alpha$:

$$\sum_{\alpha\in I} S_\alpha := \left\{ s_{\alpha_1} + \cdots + s_{\alpha_n} \middle| n \in \mathbb{N}, s_{\alpha_j} \in \bigcup_{\alpha\in I} S_\alpha \right\}$$

This construction in fact defines the least upper bound of S and T in $\mathcal{S}(V)$, or of $\{S_\alpha\}_{\alpha\in I}$ in $\mathcal{S}(V)$ with respect to inclusion. This shows that the collection $\mathcal{S}(V)$ forms a **lattice** with respect to the partial ordering of inclusion.

1.2 Direct Sums

We start our journey into vector space constructions with the notion of **direct sums**.

Definition 1.4 Let $V_1, ..., V_n \in \mathbf{Vect}_k$. The **external direct sum** of $V_1, ..., V_n$ is the vector space

$$V_1 \boxplus \cdots \boxplus V_n := \{(v_1, ..., v_n)|v_i \in V_i, i = 1, ..., n\}$$

with operations of componentwise addition and scalar multiplication. In general for an index set I and a family of vector space $\{V_\alpha\}_{\alpha\in I}$, the direct sum is the space subspace of $(\amalg_{\alpha\in I} V_\alpha)_0^I$ consisting of functions such that $f(\alpha) \in V_\alpha$. We denote this space by

$$\bigoplus_{\alpha\in I}^{ext} V_\alpha = \left\{ f : I \to \amalg_{\alpha\in I} V_\alpha \middle| f(i) \in V_i, f \text{ has finite support} \right\}$$

We also call the subspace of $(\amalg_{\alpha\in I} V_\alpha)^I$ consisting of all functions such that $f(\alpha) \in V_\alpha$ to be the **direct product** of the family, and denote it by

$$\prod_{\alpha\in I} V_\alpha := \left\{ f : I \to \amalg_{\alpha\in I} V_\alpha \middle| f(i) \in V_i \right\}$$

In the case that $V_\alpha = V \in \mathbf{Vect}_k$ for all α, these spaces are just $(V^I)_0$ and V^I respectively. In the case of I finite we have that these constructions coincide.

Identically to the external direct sum we define the notion of an internal direct sum.

Definition 1.5 A vector space $V \in \mathbf{Vect}_k$ is the **internal direct sum** of a family $\mathcal{F} = \{S_\alpha\}_{\alpha\in I}$ of subspaces of V, written

$$V = \bigoplus \mathcal{F} = \bigoplus_{\alpha\in I} S_\alpha$$

if the following hold:

- V is the sum, or join, of the family $\mathcal{F}$:

$$V = \sum_{\alpha\in I} S_\alpha$$

- For each $\alpha \in I$,

$$S_\alpha \cap \left(\sum_{\alpha \neq \beta \in I} S_\beta \right) = \{0\}$$

In this case each S_α is called a direct summand of V.

In the case that $V = S \oplus T$ we say that the subspace T is a **complement** of S in V. One should note that a subspace generally has many distinct complements. For example consider lines in $\mathbb{R}^2$ (one-dimensional subspaces). There are also equivalent ways to characterize the independence of a family in a direct sum, as we now show.

Theorem 1.1 *Let $\mathcal{F} = \{S_\alpha\}_{\alpha \in I}$ be a family of distinct subspaces of V. The following are equivalent:*

- *For each $\alpha \in I$,*

$$S_\alpha \cap \left(\sum_{\alpha \neq \beta \in I} S_\beta \right) = \{0\}$$

- *The zero vector* 0 *cannot be written as a sum of nonzero vectors from distinct subspaces of $\mathcal{F}$*
- *Every nonzer $v \in V$ has a unique, except for order of terms, expression as a sum*

$$v = s_1 + \cdots + s_n$$

of nonzero vectors from distinct subspaces in $\mathcal{F}$.

Proof First assume the first claim and that the second fails. Then there exist $v_1, ..., v_n$ in distinct families which are non-zero and $v_1 + \cdots + v_n = 0$. However this implies $-v_1 = v_2 + \cdots + v_n$, or $v_1 \in S_{\alpha_1} \cap \left(\sum_{\alpha_1 \neq \beta} S_\beta \right)$ for $v_1 \neq 0$, contradicting the first claim.

Next if the second claim holds and $\sum_{\alpha \in I} s_\alpha = \sum_{\alpha \in I} r_\alpha$, then $\sum_{\alpha \in I} (s_\alpha - r_\alpha) = 0$ is an expression of zero in terms of elements of the summands, so $s_\alpha = r_\alpha$ for each $\alpha \in I$. Finally, if the third claim holds, any non-zero element of S_α cannot be written as a sum of elements in $\{S_\beta\}_{\beta \neq \alpha}$, so we have that the first statement holds. □

1.3 Spans and Linear Independence

We now investigate possibly one of the most central concepts in vector space theory: the notion of spanning and linearly independent sets.

Definition 1.6 Let $S \subseteq V$ be a nonempty set. We define the span of S to be the set of all linear combinations of vectors from S:

$$\langle S \rangle = \text{span}(S) = \{r_1 v_1 + \cdots r_n v_n | n \in \mathbb{N}, r_i \in k, v_i \in S\}$$

The set S is said to **span** or generate V if $V = \mathrm{span}(S)$.

By convention we take span $\emptyset = \{0\}$. The span gives us the smallest subspace of V which contains the subset S. Next we have the notion of linear independence, relating to minimal generating sets.

Definition 1.7 Let $V \in \mathbf{Vect}_k$. A nonempty set $S \subseteq V$ is **linearly independent** if for any distinct vectors $s_1, ..., s_n \in S$,

$$a_1 s_1 + \cdots + a_n s_n = 0 \implies a_i = 0, \ \forall i \in \{1, ..., n\}$$

If S is not linearly independent we say that it is **linearly dependent**.

For convenience we consider the emptyset $\emptyset$ to be linearly independent. It is an immediate consequence that a set S is linearly independent if and only if every vector $v \in \mathrm{span}(S)$ can be expressed uniquely as a linear combination of vectors in S, and no vector in S is a linear combination of other vectors in S.

Theorem 1.2 *Let $V \in \mathbf{Vect}_k$ and let $S \subseteq V$ be linearly independent. If $v \in V$ and $v \notin S$, then $S \cup \{v\}$ is linearly independent if and only if $v \notin \mathrm{span}(S)$.*

Proof First if $v \in \mathrm{span}(S)$ then we could express $v = a_1 v_1 + \cdots + a_n v_n$ for $v_i \in S$, or in other words $v - a_1 v_1 - \cdots - a_n v_n = 0$, implying that $S \cup \{v\}$ is not linearly independent. Conversely, if $S \cup \{v\}$ is linearly independent, we have $av + a_1 v_1 + \cdots + a_n v_n = 0$ for some $a_i \in S$, where $a \neq 0$ since S is linearly independent. Hence $v = -\frac{1}{a}(a_1 v_1 + \cdots + a_n v_n) \in \mathrm{span}(S)$. □

Combining linear independence and spanning we arrive at the notion of a basis.

Definition 1.8 A subset $\mathcal{B}$ of $V \in \mathbf{Vect}_k$ is said to be a **basis** if it is linearly independent and spans V.

Combining with our previous we find equivalently that a every vector in V can be expressed uniquely as a linear combination of vectors in $\mathcal{B}$, $\mathcal{B}$ is a minimal spanning set, and $\mathcal{B}$ is a maximal linearly independent set. Additionally, if $\mathcal{B} = \{v_1, ..., v_n\}$ is finite, we can write $V = \langle v_1 \rangle \oplus \cdots \oplus \langle v_n \rangle$.

Theorem 1.3 *Let $V \in \mathbf{Vect}_k$ be a non-zero vector space. Let I be a linearly independent set in V and S a spanning set containing I. Then there is a basis $\mathcal{B}$ of V for which $I \subseteq \mathcal{B} \subseteq S$.*

Proof Let $\mathcal{A}$ denote the collection of all linearly independent subsets of V containing I and contained in S. Since $I \in \mathcal{A}$ this is non-empty. Now if $C = \{I_\alpha\}_{\alpha \in J}$ is a chain in $\mathcal{A}$, then the union $U = \bigcup_{\alpha \in J} I_\alpha$ is linearly independent and satisfies $I \subseteq U \subseteq S$. Indeed any linear dependence on U would simplify to a linear dependence on an element of C. Hence every chain in $\mathcal{A}$ has an upper bound in $\mathcal{A}$, so by Zorn's lemma $\mathcal{A}$ must contain a maximal element $\mathcal{B}$ which is linearly independent.

This $\mathcal{B}$ is a basis for $\langle S \rangle = S$, since if any $s \in S$ is not in span($\mathcal{B}$, then $\mathcal{B} \cup \{s\} \subseteq S$ would be a larger linearly independent set, contradicting maximality of $\mathcal{B}$. □

In the case of finite sets we perform an inductive argument. We enumerate $S_1 = S\backslash I = \{v_1, ..., v_n\}$. Then if $v_1 \notin \mathrm{span}(I)$ we form $I_2 = I \cup \{v_1\}$ which is linearly independent and set $S_2 = S_1\backslash\{v_1\}$, and otherwise we simply remove v_1. We repeat this process until we arrive at $S_{n+1} = \emptyset$, at which point I_{n+1} is a linearly independent set with equivalent span to S.

Theorem 1.4 *If $v_1, ..., v_n$ are linearly independent in V and $s_1, ..., s_m$ span V then $n \leq m$.*

Proof First we list the vectors as $s_1, ..., s_m; v_1, ..., v_n$. Since $s_1, ..., s_m$ span V, v_1 is a linear combination of the s_i's. This implies that there exists s_i which by re-indexing we can choose to be s_1 such that $\mathrm{span}(v_1, s_2, ..., s_m) = \mathrm{span}(s_1, s_2, ..., s_m)$. Then $v_1, s_2, ..., s_m; v_2, ..., v_n$ still represents a spanning and linearly independent set. Now if $m < n$ we can repeat this process to obtain $v_1, ..., v_m; v_{m+1}, ..., v_n$. However, $v_1, ..., v_m$ must span V but since $\{v_1, ..., v_n\}$ is linearly independent $v_n \notin \mathrm{span}(v_1, ..., v_m)$, which is a contradiction. Hence, $n \leq m$. □

Corollary 1.1 *If V has a finite spanning set, then any two bases of V have the same size.*

For arbitrary vector spaces we have the following theorem.

Theorem 1.5 *If $V \in \mathbf{Vect}_k$, then any two bases for V have the same cardinality.*

Proof We have already covered the finite case, so we may assume all bases of V are infinite. Let $\mathcal{B}$ and C be bases. Then any vector in C can be written as a linear combination of vectors in $\mathcal{B}$, and every vector in $\mathcal{B}$ must appear in at least one linear combination as otherwise a proper subset of $\mathcal{B}$ would span V. Thus $|\mathcal{B}| \leq \aleph_0|C| = |C|$ since the bases are infinite, and reversing the roles we obtain the reverse inequality so $|\mathcal{B}| = |C|$ by the Schröder-Bernstein theorem. □

In the finite dimensional case we find that being a basis of V is equivalent to being a spanning set with size $\dim V$ or a linearly independent set with size $\dim V$.

Theorem 1.6 *If $S \leq V$, then there exists $T \leq V$ such that $V = S \oplus T$.*

Proof Let $S \leq V$ and let $\mathcal{B}$ be a basis for S. As $\mathcal{B}$ is linearly independent in V we can extend it to a basis $\mathcal{B} \cup C$ of V. Let $T = \mathrm{span}(C)$. By linear independent of $\mathcal{B} \cup C$ we must have that $S \cap T = \{0\}$, and $V = S + T$ by construction, so $V = S \oplus T$. □

Theorem 1.7 *Let $S, T \leq V \in \mathbf{Vect}_k$. Then*

$$\dim(S) + \dim(T) = \dim(S + T) + \dim(S \cup T)$$

Proof Suppose $\mathcal{B}$ is a basis for $S \cap T$. Extend this to a basis $\mathcal{A} \cup \mathcal{B}$ for S and a basis $\mathcal{B} \cup C$ for T. We claim $\mathcal{A} \cup \mathcal{B} \cup C$ is a basis for $S + T$. Evidently it spans $S + T$. To see linear independence suppose to the contrary that we have a linear dependence $\alpha_1 v_1 + \cdots \alpha_n v_n$ with $\alpha_i \neq 0$ for all i. As $\mathcal{A} \cup \mathcal{B}$ and $\mathcal{B} \cup C$ are linearly independent, this expression must involve elements of $\mathcal{A}$ and C. Isolating the components of $\mathcal{A}$ on one side we have a nonzero vector $x \in \langle\mathcal{A}\rangle \cap \langle\mathcal{B} \cup C\rangle$. But then $x \in S \cap T$, so $x \in \langle\mathcal{A}\rangle \cap \langle\mathcal{B}\rangle$, which implies $x = 0$, a contradiction. Hence they indeed form a basis for $S + T$, and

$$\dim(S) + \dim(T) = |\mathcal{A} \cup \mathcal{B}| + |\mathcal{B} \cup C| = \dim(S + T) + \dim(S \cap T)$$

1.4 Ordered Bases and Coordinates

Now that we have a notion of minimal spanning sets which uniquely represent the vectors in a vector space, we can investigate how to represent vector spaces in a canonical fashion.

Definition 1.9 Let $V \in \mathbf{Vect}_k$ with $\dim V = n \in \mathbb{N}$. An **ordered basis** for V is an ordered n-tuple $(v_1, \ldots, v_n)$ of vectors for which the set $\{v_1, \ldots, v_n\}$ is a basis for V.

If $\mathcal{B} = (v_1, \ldots, v_n)$ is an ordered basis for V, then for each $v \in V$ we have a unique ordered n-tuple $(r_1, \ldots, r_n)$ of scalars for which $v = r_1 v_1 + \cdots + r_n v_n$. This defines a bijection known as the **coordinate map**: $\varphi_{\mathcal{B}} : V \to k^n$, with $\varphi_{\mathcal{B}}(v) = [v]_{\mathcal{B}} = \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}$.

Appendices: Constructions

Chapter 2

Linear Maps

As with all algebraic structures, one of the most important concepts is the structure preserving maps between the algebraic objects of interest. In this chapter we study these maps and their properties on abstract vector spaces.

2.1 Basic Linear Maps

Definition 2.1 Let $V, W \in \mathbf{Vect}_k$. A function $\tau : V \to W$ is a **linear transformation** if

$$\tau(u+v) = \tau(u) + \tau(v),\ \ \tau(ru) = r\tau(u)$$

for all scalars $r \in k$ and vectors $u, v \in V$. The set of all linear transformations from V to W forms a vector space $\mathcal{L}(V, W)$.

Linear maps of the form $\tau : V \to V$ are known as **linear operators** and their vector space is denoted $\mathcal{L}(V)$. Additionally, linear maps of the form $\tau : V \to k$ are known as **linear functionals**, and their vector space is denoted V^* and called the **dual space** of V.

Example 2.1 A few examples of important and well known linear transformations are as follows:

- The derivative $D : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ is a linear operator.
- The integral operator $\tau : \mathbb{R}[x] \to \mathbb{R}[x]$ defined by $\tau f = \int_0^x f(t)dt$ is a linear operator on $k[x]$.
- If $A \in M_{m,n}(k)$, the function $\tau_A : k^n \to k^m$ defined by $\tau_A v = Av$, where all vectors are written as column vectors, is a linear transformation from k^n to k^m.
- The coordinate map $\varphi : V \to k^n$ of an n-dimensional vector space is a linear transformation from V to k^n.

We note that in the case of $\mathcal{L}(V)$, the space is actually an associative algebra with unity, where multiplication is simply composition of operators. Now, recall that every vector space has a basis by an application of Zorn's Lemma. This provides vector spaces with an important universal property associated with maps on free structures.

Theorem 2.1 *Let V and W be vector spaces and let $\mathcal{B} = \{v_i | i \in I\}$ be a basis for V. Then there is a unique linear transformation $\tau \in \mathcal{L}(V, W)$ for which $\tau(v_i) = w_i \in W$.*

Proof Since each vector in V is a unique linear combination of vectors in the basis, defining $\tau(a_1 v_1 + \cdots + a_n v_n) = a_1 \tau v_1 + \cdots + a_n \tau v_n = a_1 w_1 + \cdots + a_n w_n$ for an arbitrary vector in V is a well-defined operation giving a linear map. Additionally this formula uniquely specifies the action of the map on any vector based solely on its action on a basis, so the map is unique. □

We briefly note that the category $\mathbf{Vect}_k$ is abelian, or in other words its arrows are enriched over $\mathbf{Ab}$. That is $S \circ (T_1 + T_2) = S \circ T_1 + S \circ T_2$ and $(T_1 + T_2) \circ S = T_1 \circ S + T_2 \circ S$. Now, for each linear map we have two important subspaces.

Definition 2.2 Let $V, W \in \mathbf{Vect}_k$ and $T \in \mathcal{L}(V, W)$. The subset $\ker(T) = \{v \in V | T(v) = 0\}$ of V is called the **kernel** or **null space** of T, and the subset $\mathrm{Im}(T) = \{T(v) | v \in V\}$ of W is called the **image** or **range** of T.

These are subspaces of V and W, respectively. We call $\dim(\ker(T))$ the **nullity** of T and $\dim(\mathrm{Im}(T))$ the **rank** of T, respectively. Note that by linearity the map $\tau \in \mathcal{L}(V, W)$ being injective is equivalent to $\ker(\tau) = \{0\}$.

Definition 2.3 A bijective linear transformation $\tau : V \to W$ is called a **linear isomorphism** from V to W. If a linear isomorphism from V to W exists we say they are **isomorphic** and write $V \cong W$.

The most common and useful example of a linear isomorphism in the case of finite dimensional vector spaces comes in the form of the coordinate map for a choice of ordered basis. An important point to note is that linear isomorphisms preserve the linear structure of a space entirely, so linearly independent sets are sent to linearly independent sets, spanning sets are sent to spanning sets, and consequently bases are sent to bases. In particular, a linear map is an isomorphism if and only if it sends some basis of the domain to some basis of the codomain.

Theorem 2.2 *Let $V, W \in \mathbf{Vect}_k$. Then $V \cong W$ if and only if* $\dim(V) = \dim(W)$.

If two vector spaces are isomorphic we know that we bijectively map bases to bases so the dimensions are equivalent. Conversely, if they have bases of cardinality κ, for B a set of cardinality κ we have $V \cong (k^B)_0 \cong W$ by a choice of basis.

2.2 Quotients and Dimension Theory

One of the most important results in linear algebra is the dimension theorem. It relies on the central algebraic notion of the first isomorphism theorem.

Definition 2.4 Let $V \in \mathbf{Vect}_k$ and let $W \leq V$. Then the coset space $V/W = \{v + W : v \in V\}$ forms a vector space under the natural induced operations from V such that the projection $\pi : V \to V/W$ is a linear map with kernel $\ker \pi = W$.

Then we have the first isomorphism theorem.

Theorem 2.3 *Let $\tau \in \mathcal{L}(V, W)$. Then there exists a unique map $\overline{\tau} : V/W \to W$ such that $\overline{\tau} \circ \pi = \tau$, and $\overline{\tau}$ is an isomorphism onto $\tau(W)$.*

As a corollary we obtain the dimension theorem.

Corollary 2.1 (Dimension Theorem) *Let $\tau \in \mathcal{L}(V, W)$. Then* $\dim V = \dim W + \dim V/W = \dim \ker(\tau) + \dim \mathrm{Im}(\tau)$.

2.3 Matrix Representations

First, we observe that for any linear transformation $\tau \in \mathcal{L}(k^n, k^m)$, by construction of matrix multiplication τ is simply given by matrix multiplication in the sense that $[\tau e_1 \ \cdots \tau e_n] e_i = \tau e_i$, and so $\tau = \tau_A$ for $A = [\tau e_1 \ \cdots \ \tau e_n$. This association is a linear isomorphism, so $\mathcal{L}(k^n, k^m) \cong M_{m,n}(k)$. In particular, $\tau_A : k^n \to k^m$ is injective if and only if $\mathbf{rank}(A) = n$ and surjective if and only if $\mathbf{rank}(A) = m$, using the dimension theorem.

In general, for linear operator $\tau \in \mathcal{L}(V, W)$ between finite dimensional vector spaces, a choice of bases $\beta = \{v_1, ..., v_n\}$ and $\gamma = \{w_1, ..., w_m\}$ specifies a linear isomorphism

$$[\cdot]_\beta^\gamma : \mathcal{L}(V, W) \to M_{m,n}(k)$$

sending τ to the matrix

$$[\tau]_\beta^\gamma = \big[[\tau v_1]_\gamma \ \cdots \ [\tau v_n]_\gamma \big]$$

In particular, by definition of matrix multiplication we have that $[\tau v]_\gamma = [\tau]_\beta^\gamma [v]_\beta$. Not only is the map a linear isomorphism, it also satisfies the property that if $\sigma \in \mathcal{L}(W, U)$ with basis ρ for U, then

$$[\sigma\tau]_\beta^\rho = [\sigma]_\gamma^\rho [\tau]_\beta^\gamma$$

Consequently we find that isomorphisms are sent to invertible matrices, with the inverse of the matrix being the representation of the inverse map in those bases.

Now, how would one go about moving between basis representations? The intuition comes from the fact that going from bases β to β' of V should be the matrix representation $[\mathrm{Id}_V]_\beta^{\beta'}$. In particular, this is simply the matrix representation of the map $\phi_{\beta'} \circ \phi_\beta^{-1} : k^n \to k^n$ in the canonical coordinates on k^n.We write $P_{\beta' \leftarrow \beta} := [\mathrm{Id}_V]_\beta^{\beta'}$. Then for any map $\tau \in \mathcal{L}(V, W)$ and any bases β, β' of V and γ, γ' of W,

$$[\tau]_{\beta'}^{\gamma'} = P_{\gamma' \leftarrow \gamma} [\tau]_\beta^\gamma P_{\beta \leftarrow \beta'}$$

Note that $P_{\beta' \leftarrow \beta} = P_{\beta \leftarrow \beta'}^{-1}$, as they correspond with the linear maps $\phi_{\beta'} \circ \phi_\beta^{-1}$ and $\phi_\beta \circ \phi_{\beta'}^{-1}$, respectively.

2.4 Dual Space

As defined previously, for a vector space $V \in \mathbf{Vect}_k$ we define its dual space to be $V^* = \mathcal{L}(V, k)$, the space of linear functionals on V.

Theorem 2.4 *Suppose $V \in \mathbf{Vect}_k$ with basis $\mathcal{B} = \{v_i | i \in I\}$. For each $i \in I$ we can define a linear functional $v_i^* \in V^*$ by the orthogonality condition $v_i^*(v_j) = \delta_{i,j}$ and extend by linearity. Then the set $\mathcal{B}^* = \{v_i^* | i \in I\}$ is a linearly independent subset.*

This result follows simply since for any relation $0 = a_{i_1} v_{i_1}^* + \cdots + a_{i_n} v_{i_n}^*$, if we evaluate it at v_{i_j}, for $1 \leq j \leq n$, we find $a_{i_j} = 0$ so the relation is trivial. In th case that V is finite dimensional this is a basis for V^*, called the **dual basis** of $\mathcal{B}$. Indeed, if $f \in V^*$ and $\mathcal{B} = \{v_1, ..., v_n\}$, we have

$$f(v_i) = \sum_{j=1}^n f(v_j) v_j^*(v_i)$$

so $f = \sum_{j=1}^n f(v_j) v_j^*$. We also have for finite dimensional vector spaces a natural isomorphism from V to the double dual V^{**}, given by the evaluation map $v \mapsto \mathrm{ev}_v$.

An important concept to dual spaces is the notion of annihilators.

Definition 2.5 Let $V \in \mathbf{Vect}_k$ and let $M \subseteq V$ be non-empty. The **annihilator** M^0 of M is

$$M^0 = \{f \in V^* | f(M) = \{0\}\}$$

Observe M^0 is a subspace of V^*. Additionally, if $M \subseteq N$, $N^0 \subseteq M^0$ so the operation is order reversing.

Theorem 2.5 *Let $S, T \leq V \in \mathbf{Vect}_k$ and $M \subseteq V$. Then*

- *if* $\dim V < \infty$*, then the natural map $\tau : \mathrm{span}(M) \to M^{00}$ is an isomorphism.*
- *For S, T, $(S \cap T)^0 = S^0 + T^0$ and $(S + T)^0 = S^0 \cap T^0$.*

To be continued

2.4.1 Operator Adjoints

For $\tau \in \mathcal{L}(V, W)$, we can define a map $\tau^t : W^* \to V^*$ by

$$\tau^t(f) = f \circ \tau = f\tau$$

for all $f \in W^*$. This map is called the **operator adjoint** of τ.

Theorem 2.6 *For $\tau, \sigma \in \mathcal{L}(V, W)$ and $a, b \in k$ we have $(a\tau + b\sigma)^t = a\tau^t + b\sigma^t$, and if $\rho \in \mathcal{L}(W, U)$, then $(\rho\sigma)^t = \sigma^t \rho^t$. Finally for invertible $\tau \in \mathcal{L}(V)$, $(\tau^{-1})^t = (\tau^t)^{-1}$.*

In relation to annihilators we have the following result.

Theorem 2.7 *Let $\tau \in \mathcal{L}(V, W)$, then*

- $\ker(\tau^t) = \mathrm{Im}(\tau)^0$
- $\mathrm{Im}(\tau^t) = \ker(\tau)^0$

Finally we can consider the matrix representation of the operator adjoint with respect to the dual bases for finite dimensional spaces.

Theorem 2.8 *Let $\tau \in \mathcal{L}(V, W)$ where V and W have bases $\mathcal{B} = \{v_1, \ldots, v_n\}$ and $\mathcal{C} = \{w_1, \ldots, w_m\}$, respectively, and corresponding dual bases $\mathcal{B}^*$ and $\mathcal{C}^*$. Then it follows that*

$$[\tau^t]_{\mathcal{C}^*}^{\mathcal{B}^*} = ([\tau]_{\mathcal{B}}^{\mathcal{C}})^T$$

Proof First we note that $\tau^t(w_j^*)(v_i)$ gives the i, j entry of its matrix with respect to the dual bases. Then we have

$$\tau^t(w_j^*)(v_i) = w_j^*(\tau(v_i))$$

which is the j, i entry of $[\tau]_{\mathcal{B}}^{\mathcal{C}}$. Hence

$$[\tau^t]_{\mathcal{C}^*}^{\mathcal{B}^*} = ([\tau]_{\mathcal{B}}^{\mathcal{C}})^T$$

Appendices A: Multilinear Maps

In this section we investigate the generalization of linear maps, known as multilinear maps.

Definition 2.6 Let $V, W \in \mathbf{Vect}_k$. A function $\phi : V^m \to W$ is called an m-linear form on V if it is linear in each component, where $V^m = \bigoplus_{i=1}^m V$.

We shall first consider the special case of multilinear forms, or multilinear maps of the form $\varphi : V \times \ldots \times V \to k$.

Example 2.2 Let k be a field and $A \in M_{m,m}(k)$. Then we define $\varphi : M_{m,n}(k) \times M_{m,n}(k) \to k$ by

$$\varphi(X, Y) = \mathrm{tr}(X^T AY), \quad X, Y \in M_{m,n}(k)$$

is a bilinear form on $M_{m,n}(k)$.

Another important example of a bilinear form are inner products on real vector spaces, which has been well-studied in the spectral theorems of this document. We also have the following important and recurring example.

Example 2.3 For $A \in M_{n,n}(k)$ fixed, consider the function $\varphi : k^n \times k^n \to k$ defined by $\varphi(x, y) = x^T Ay$.

We shall denote the m-linear maps from V to W by $\mathcal{L}_m(V, W)$, which is a vector space over k. An important construction of vector spaces which we shall justify more fully in our part on module theory is the notion of a tensor product of vector spaces. Here we shall simply state how a tensor space operates and its universal property without proof.

Definition 2.7 Let $V, W \in \mathbf{Vect}_k$. Then the tensor product over k is the vector space $V \otimes_k W$ generated by simple tensors $v \otimes w \in V \otimes_k W$ for $v \in V$ and $w \in W$, which are linear in each term. That is an arbitrary element is of the form

$$\sum_{i=1}^n a_i(v_i \otimes w_i)$$

The tensor product satisfies the following universal property.

Theorem 2.9 *Let $V, W, U \in \mathbf{Vect}_k$. Let $\iota : V \oplus W \to V \otimes W$ be the canonical map. Then for any multilinear map $\varphi : V \oplus W \to U$ there exists a unique linear map $\psi : V \otimes W \to U$ such that $\psi \circ \iota = \varphi$. This is characterized by the following commuting triangle.*

$$\begin{tikzcd} V \oplus W \arrow[r, hook, "\iota"] \arrow[dr, "\varphi"'] & V \otimes W \arrow[d, dashed, "\exists! \psi"] \\ & U \end{tikzcd}$$

This property implies that $\otimes : \mathbf{Vect}_k \times \mathbf{Vect}_k \to \mathbf{Vect}_k$ is a bi-functor of abelian categories. Now, the space $\mathcal{L}_m(V, k)$ is un-naturally isomorphic to $\bigotimes_{i=1}^m V^*$ for finite-dimensional V.

Theorem 2.10 *Let $V \in \mathbf{Vect}_k$ be of dimension n with basis $\mathcal{B} = \{v_1, ..., v_n\}$. Let $\mathcal{B}^* = \{v_1^*, ..., v_n^*\}$ denote the dual basis of V^*. Then the map $\Psi : \bigotimes_{i=1}^m V^* \to \mathcal{L}_m(V, k)$ defined on the basis of $\bigotimes_{i=1}^m V^*$ by*

$$\Psi(v_{i_1}^* \otimes \cdots \otimes v_{i_m}^*)(u_1, ..., u_m) = v_{i_1}^*(u_1)...v_{i_m}^*(u_m)$$

and extending by linearity is a linear isomorphism. In particular under this identification we denote $\{v_{i_1}^* \otimes \cdots \otimes v_{i_m}^* : 1 \leq i_j \leq n, 1 \leq j \leq m\}$ as the dual basis for $\mathcal{L}_m(V, k)$, so $\dim \mathcal{L}_m(V, k) = n^m$.

In particular in the case of bilinear forms we have a basis $\{v_i^* \otimes v_j^* : 1 \leq i, j \leq n\}$.

2.4.2 Matrix Representations of Bilinear Forms

Throughout let V be an n-dimensional vector space over k, and let $\mathcal{B} = \{v_1, ..., v_n\}$ be an ordered basis of V. If $\varphi \in \mathcal{L}_2(V, k)$, and $x, y \in V$ with $[x]_{\mathcal{B}} = [x_1 \ \cdots \ x_n]^T, [y]_{\mathcal{B}} = [y_1 \ \cdots \ y_n]^T$ the coordinate vectors of x and y, we have

$$\varphi(x, y) = \varphi\left(\sum_{i=1}^n x_i v_i, \sum_{j=1}^n y_j v_j\right) = \sum_{i=1}^n \sum_{j=1}^n x_i x_j \varphi(v_i, v_j) = [x]_{\mathcal{B}}^T \begin{bmatrix} \varphi(v_1, v_1) & \varphi(v_1, v_2) & \cdots & \varphi(v_1, v_n) \\ \varphi(v_2, v_1) & \varphi(v_2, v_2) & \cdots & \varphi(v_2, v_n) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi(v_n, v_1) & \varphi(v_n, v_2) & \cdots & \varphi(v_n, v_n) \end{bmatrix} [y]_{\mathcal{B}}$$

This suggests how we define a matrix representation of a bilinear form, being

$$[\varphi]_{\mathcal{B}} = [\varphi(v_i, v_j)]_{1 \leq i, j \leq n}$$

Now we want to know how to change bases for bilinear forms as we did for linear maps.

Theorem 2.11 *Let $\mathcal{A} = \{u_1, ..., u_n\}$ and $\mathcal{B} = \{v_1, ..., v_n\}$ be ordered bases of V and $P_{\mathcal{A} \leftarrow \mathcal{B}} = [[v_1]_{\mathcal{A}} \ \cdots \ [v_n]_{\mathcal{A}}]$ the change of bases matrix from $\mathcal{B}$ to $\mathcal{A}$. Then if $\varphi \in \mathcal{L}_2(V, k)$,*

$$[\varphi]_{\mathcal{B}} = P_{\mathcal{A} \leftarrow \mathcal{B}}^T [\varphi]_{\mathcal{A}} P_{\mathcal{A} \leftarrow \mathcal{B}}$$

Since change of variable matrices are invertible, both matrix representations of φ have the same rank.

Proof We observe that for $x, y \in V$,

$$\begin{aligned} [x]_{\mathcal{B}}^T [\varphi]_{\mathcal{B}} [y]_{\mathcal{B}} = \varphi(x, y) &= [x]_{\mathcal{A}}^T [\varphi]_{\mathcal{A}} [y]_{\mathcal{A}} \\ &= (P_{\mathcal{A} \leftarrow \mathcal{B}} [x]_{\mathcal{B}})^T [\varphi]_{\mathcal{A}} (P_{\mathcal{A} \leftarrow \mathcal{B}} [y]_{\mathcal{B}}) \\ &= [x]_{\mathcal{B}}^T (P_{\mathcal{A} \leftarrow \mathcal{B}}^T [\varphi]_{\mathcal{A}} P_{\mathcal{A} \leftarrow \mathcal{B}}) [y]_{\mathcal{B}} \end{aligned}$$

As this holds for all $x, y \in V$, in particular it holds for v_i and v_j, which gives $(P_{\mathcal{A} \leftarrow \mathcal{B}}^T [\varphi]_{\mathcal{A}} P_{\mathcal{A} \leftarrow \mathcal{B}})_{i,j} = ([\varphi]_{\mathcal{B}})_{i,j}$. Hence the change of basis formula holds. □

We then define the **rank** of a bilinear form to be the rank of any of its matrix representations. If $\operatorname{rank}(\varphi) = \dim V$, we say that φ is **nondegenerate** or **nonsingular**. This is equivalent to the statement that for all $u \in V$, $u \neq 0$, there exists $v \in V$ such that $\varphi(u, v) \neq 0$, and for all $v \in V$, $v \neq 0$, there exists $u \in V$ such that $\varphi(u, v) \neq 0$.

We now explore special multilinear forms.

Definition 2.8 Let $V \in \mathbf{Vect}_k$. We say a multilinear form $\varphi \in \mathcal{L}_m(V, k)$ is **symmetric** if $\varphi(u_1, ..., u_m) = \varphi(u_{\sigma(1)}, ..., u_{\sigma(m)})$ for any permutation $\sigma \in S_m$. If instead we have $\varphi(u_1, ..., u_m) = \operatorname{sign}(\sigma)\varphi(u_{\sigma(1)}, ..., u_{\sigma(m)})$ for any $\sigma \in S_m$, we say that φ is **anti-symmetric**.

In the case of a bilinear form, symmetric forms correspond to symmetric matrix representations, and anti-symmetric forms correspond to skew-symmetric matrix representations. Now we explore when a bilinear form has a diagonal matrix representation.

Definition 2.9 Let V be a finite dimensional vector space over a field k. A bilinear form $\varphi \in \mathcal{L}_2(V, k)$ is called **diagonalizable** if there exists an ordered basis $\mathcal{B} = \{v_1, ..., v_n\}$ of V such that the matrix $[\varphi]_{\mathcal{B}}$ of φ, relative to $\mathcal{B}$, is a **diagonal** matrix.

Notice that if the matrix representation is diagonal it is in particular symmetric. Consequently in order for a bilinear form to be diagonalizable it must be symmetric. In the case of fields of characteristic not equal to 2 this condition is also sufficient. First we prove a preliminary result.

Lemma 2.1 *Let $V \in \mathbf{Vect}_k$ with char$(k) \neq 2$. If $\varphi \in \mathcal{L}_2(V, k)$ is **symmetric** and not identically zero, then there exists $u \in V$, $u \neq 0$, such that $\varphi(u, u) \neq 0$.*

Proof Since $\varphi \neq 0$, there exists $x, y \in V$ such that $\varphi(x, y) \neq 0$. If either $\varphi(x, x) \neq 0$ or $\varphi(y, y) \neq 0$ we are done. Otherwise if $\varphi(x, x) = \varphi(y, y) = 0$, we have

$$\varphi(x + y, x + y) = \varphi(x, y) + \varphi(y, x) = 2\varphi(x, y) \neq 0$$

as desired. □

Now we have our main result.

Theorem 2.12 *Let $V \in \mathbf{Vect}_k$, with char$(k) \neq 2$ and $\dim V = n \in \mathbb{N}$. Let $\varphi : V \times V \to k$ be a symmetric bilinear form on V. Then there exists an ordered basis $\mathcal{B} = \{v_1, ..., v_n\}$ of V such that $[\varphi]_{\mathcal{B}}$ is a diagonal matrix.*

Proof If $\varphi = 0$ the claim immediately holds. Hence suppose $\varphi \neq 0$. We proceed by induction on n. If $n = 1$ any choice of basis diagonalizes φ. Suppose inductively that the claim holds for dimension $< n$. Now since $\varphi \neq 0$ we have $v_1 \in V$ such that $\varphi(v_1, v_1) \neq 0$ by the previous lemma. Consider the map $g : V \to k$ defined by $g(u) = \varphi(u, v_1)$. Then $g \neq 0$, so $\operatorname{rank}(g) = 1$, do $\dim \ker(g) = n - 1$. Let $\{v_2, ..., v_n\}$ be a basis of $\ker(g)$ which diagonalizes $\varphi|_{\ker(g)}$ by the induction hypothesis. Observe $\varphi(v_i, v_1) = 0$ for all $2 \leq i \leq n$, and by our induction hypothesis $\varphi(v_i, v_j) = 0$ for all $2 \leq i \neq j \leq n$. Thus we have that $\varphi(v_i, v_j) = 0$ for all $1 \leq i \neq j \leq n$, so $[\varphi]_{\mathcal{B}}$ is diagonal for $\mathcal{B} = \{v_1, ..., v_n\}$. □

Corollary 2.2 *If k is a field of characteristic different from 2, and $A \in M_{n,n}(k)$ is a symmetric matrix, then there exists an invertible matrix P over k such that $P^T AP$ is diagonal.*

Appendices B: Quadratic Forms

We now study the notion of quadratic forms, which relate in a particular manner to symmetric bilinear forms.

Definition 2.10 Let $V \in \mathbf{Vect}_k$. A **quadratic form** is a map $Q : V \to k$ such that $Q(\alpha v) = \alpha^2 Q(v)$ for all $\alpha \in k$ and $v \in V$, and $\omega_Q(x, y) = Q(x + y) - Q(x) - Q(y)$ is a symmetric bilinear form.

For finite dimensional vector spaces the matrix of a quadratic form is the matrix of its associated symmetric form. In particular, the matrix is always symmetric and for fields of characteristic different from 2,

$$Q(x) = \frac{1}{2}\omega_Q(x, x)$$

In particular from our results on symmetric bilinear forms all quadratic forms are diagonalizable. In the case of $k = \mathbb{R}$ we have by the real spectral theory that for an orthonormal basis $\{v_1, ..., v_n\}$ of eigenvectors with eigenvalues $\{\lambda_1, ..., \lambda_n\}$, then for any vector $x = x_1 v_1 + \cdots + x_n v_n$ we have

$$Q(x) = \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2$$

Then we can easily determine the sign of $Q(x)$ with it being always positive if $\lambda_i > 0$ for all i, and always negative if $\lambda_i < 0$ for all i.

Definition 2.11 Let $A \in \mathrm{Sym}_n(\mathbb{R})$. We say that A is **positive definite** if $x^T Ax > 0$ for all $x \neq 0$, **positive semi-definite** if $x^T Ax \geq 0$ for all $x \neq 0$, **negative definite** if $x^T Ax < 0$ for all $x \neq 0$, **negative semi-definite** if $x^T Ax \leq 0$ for all $x \neq 0$, and **indefinite** if $x^T Ax$ takes on positive and negative values.

Theorem 2.13 *If A is positive definite, then A is invertible.*

Indeed singular matrices have 0 as an eigenvalue.

Theorem 2.14 *Let $A \in Sym_n(\mathbb{R})$. The statements below are all equivalent.*

- *A is positive definite*
- *All eigenvalues of A are positive*
- *All the principal submatrices of A are positive definite*

- *All the principal minors of A are positive*

The kth principal submatrix of A is defined as $A_k = (A_{i,j})_{1 \leq i,j \leq k}$, where $A = (A_{i,j})_{1 \leq i,j \leq n}$. Then the kth principal minor of A is $\det A_k = \Delta_k$.

Corollary 2.3 *Let $A \in Sym_n(\mathbb{R})$. Then*

- *A is **negative definite** if and only if $\Delta_1 < 0, \Delta_2 > 0, \Delta_3 < 0, \ldots$*
- *A is **indefinite** if and only if either A has a negative principal minor of even order or A has two principal minors of odd order with opposite signs.*

Chapter 3

Matrix Algebra

Chapter 4

Inner Product Spaces

In this chapter we equip vector spaces over $k = \mathbb{R}$ or $\mathbb{C}$ with a structure of length, distance, and angle in the form of a operation known as an **inner product**.

4.1 Inner Product Construction

First we define what properties characterize an inner product on a vector space over k.

Definition 4.1 Let $V \in \mathbf{Vect}_k$. An **inner product** on V is a function $\langle , \rangle : V \times V \to k$ with the following properties.

- **Positive definiteness**: for all $v \in V$,

$$\langle v, v \rangle \geq 0 \text{ and } \langle v, v \rangle = 0 \iff v = 0$$

- **Conjugate symmetry**: for all $u, v \in V$,

$$\langle u, v \rangle = \overline{\langle v, u \rangle}$$

- **Linearity in the first coordinate**: for all $u, v, w \in V$ and $r, s \in k$,

$$\langle ru + sv, w \rangle = r\langle u, w \rangle + s\langle v, w \rangle$$

The pair $(V, \langle , \rangle)$ is refered to as a **real** (or **complex**) **inner product space**.

As a consequence of the definition of an inner product, it is conjugate linear in its second term. For convenience we define the notation for $X, Y \subseteq V$, $\langle X, Y \rangle = \{\langle x, y \rangle | x \in X, y \in Y\}$ and the special case $\langle v, X \rangle = \{\langle v, x \rangle | x \in X\}$. Note any subspace of an inner product space is again an inner product space under the restriction of the inner product. A complex inner product is often referred to as a **sesquilinear form**. There are a number of standard examples of inner products, a few of which are as follows.

Example 4.1 Observe the following are inner product spaces.

- The vector space $\mathbb{R}^n$ is an inner product space under the standard Euclidean inner product, or dot product,
$$\langle (r_1, ..., r_n), (s_1, ..., s_n) \rangle = \sum_{i=1}^{n} r_i s_i$$
We refer to $(\mathbb{R}^n, \langle , \rangle)$ as the n-dimensional Euclidean space.

- The vector space $\mathbb{C}^n$ is an inner product space under the standard inner product
$$\langle (r_1, ..., r_n), (s_1, ..., s_n) \rangle = \sum_{i=1}^{n} r_i \bar{s}_i$$
and it is refered to as the n-dimensional unitary space.

- The space $C[a, b]$ of all continuous complex-valued functions on $[a, b]$ is a complex inner product space under the inner product
$$\langle f, g \rangle = \int_a^b f(x)\overline{g(x)}dx$$

- The space ℓ^2 is an inner product space of k-sequences $(s_n)_{n\in\mathbb{N}}$ with inner product
$$\langle (s_n)_{n\in\mathbb{N}}, (t_n)_{n\in\mathbb{N}} \rangle = \sum_{n\in\mathbb{N}} s_n \bar{t}_n$$

- On $M_{m,n}(\mathbb{C})$ we have the inner product defined by $\langle A, B \rangle = \mathrm{tr}(A \cdot B^*)$, where B^* is the conjugate transpose of B. We refer to this as the **Frobenius inner product**.

We have the following important result:

Lemma 4.1 *If V is an inner product space and $\langle u, x \rangle = \langle v, x \rangle$ for all $x \in V$, then $u = v$.*

Now, the following result points out a key difference between real and complex inner product spaces.

Theorem 4.1 *Let $V \in$ **Vect**$_k$ be an inner product space and $\tau \in \mathcal{L}(V)$. Then if $\langle \tau v, w \rangle = 0$ for all $v, w \in V$, then $\tau = 0$. If in addition $k = \mathbb{C}$, then it is sufficient to have $\langle \tau v, v \rangle = 0$ for all $v \in V$.*

Proof The first claim follows immediately from the lemma. For the second claim let $v = rx + y$ for $x, y \in V$ and $r \in k$. Then

$$\begin{aligned} 0 &= \langle \tau(rx+y), rx+y \rangle \rangle \\ &= |r|^2 \rangle \tau x, x \rangle + \langle \tau y, y \rangle + r\langle \tau x, y \rangle + \overline{r}\langle \tau y, x \rangle \\ &= r\langle \tau x, y \rangle + \overline{r}\langle \tau y, x \rangle \end{aligned}$$

Setting $r = 1$ we obtain $\langle \tau x, y \rangle + \langle \tau y, x \rangle = 0$, and setting $r = i$ we can derive $\langle \tau x, y \rangle - \langle \tau y, x \rangle = 0$. Together these imply $\langle \tau x, y \rangle = 0$ for all $x, y \in V$. □

Now we can induce a **norm** on $(V, \langle , \rangle)$ by defining $||v|| = \sqrt{\langle v, v \rangle}$. We say a vector $v \in V$ is a **unit vector** if it is of unit norm, $||v|| = 1$.

Definition 4.2 A **norm** on a vector space $V \in \mathbf{Vect}_k$ is a map $|| \cdot || : V \longrightarrow \mathbb{R}$ satisfying the following properties.

- **Positive definiteness**: $||v|| \geq 0$ for all $v \in V$ and $||v|| = 0$ if and only if $v = 0$.
- **Absolute homogeneity**: for all $r \in k$ and $v \in V$, $||rv|| = |r| \, ||v||$.
- **Triangle inequality**: for all $v, w \in V$, $||v + w|| \leq ||v|| + ||w||$.

Another important property of norms induced by inner products is the following inequality.

Theorem 4.2 *For all $u, v \in V$ we have the* ***Cauchy-Schwarz inequality***

$$|\langle u, v \rangle| \leq ||u|| \, ||v||$$

with equality if and only if $v \in \operatorname{span}(u)$ or $u \in \operatorname{span}(v)$.

Proof If either u or v are zero the result follows, so assume $u, v \neq 0$. Then for any scalar $r \in k$,

$$0 \leq ||u - rv||^2 = \langle u, u \rangle - \overline{r}\langle u, v \rangle - r[\langle v, u \rangle - \overline{r}\langle v, v \rangle]$$

Choosing $r = \overline{\langle v, u \rangle}/\langle v, v \rangle$ makes the value in the square brackets zero, and so

$$0 \leq \langle u, u \rangle - \frac{\langle v, u \rangle \langle u, v \rangle}{\langle v, v \rangle} = ||u||^2 - \frac{|\langle u, v \rangle|^2}{||v||^2}$$

which is equivalent to the Cauchy-Schwarz inequality. Furthermore equality holds if and only if $||u - rv||^2 = 0$, which is to say $u - rv = 0$. □

We say that a pair $(V, ||\cdot||)$ is a **normed linear space**, or NLS for short. We also have the following two properties useful identities for norms and norms induces by inner products.

Proposition 4.1 *Let $x, y \in V$ for $(V, \langle , \rangle)$ an inner product space with induced norm $|| \cdot ||$. Then we have the* ***Parallelogram law***

$$||u + v||^2 + ||u - v||^2 = 2(||u||^2 + ||v||^2)$$

as well as the identity

$$||u+v||^2 - ||u-v||^2 = 2(\langle u,v\rangle + \langle v,u\rangle)$$

*If $k = \mathbb{R}$ we have the the **polarization identity***

$$4\langle u,v\rangle = ||u+v||^2 - ||u-v||^2$$

and if $k = \mathbb{C}$ we have the complex equivalent

$$4\langle u,v\rangle = ||u+v||^2 - ||u-v||^2 + i||u+iv||^2 - i||u-iv||^2$$

These identities can be proved through straight-forward computations.

4.2 Orthogonality

As stated previously we can use the inner product to define notions of angle, in particular a notion of orthogonality will be incredibly useful.

Definition 4.3 Let $(V, \langle,\rangle)$ be an inner product space. We say $u, v \in V$ are **orthogonal** and write $u \perp v$ if $\langle u,v\rangle = 0$. In general for two subsets $X, Y \subseteq V$ we say they are orthogonal and write $X \perp Y$ if $\langle X, Y\rangle = 0$.

We also define the **orthogonal complement** of a subset $X \subseteq V$ to be the set

$$X^\perp = \{v \in V | v \perp X\}$$

We have the following simple properties for orthogonal complements.

Theorem 4.3 *Let $(V, \langle,\rangle)$ be an inner product space. The orthogonal complement $X^\perp$ of any subset $X \subseteq V$ is a subspace of V. For any subspace S of V, $S \cap S^\perp = \{0\}$.*

Definition 4.4 A nonempty subset $O \subseteq V$ of an inner product space is said to be an **orthogonal set** if $u \perp v$ for all $u, v \in O$, $u \neq v$, and $0 \notin O$. If in addition each vector of O is of unit norm, we say the set is **orthonormal**.

We have the important property of orthogonal sets by Pythagoras.

Theorem 4.4 *Let $(V, \langle,\rangle)$ be an inner product space and let O be an orthogonal set. Then for any $v_1, \ldots, v_n \in O$,*

$$||v_1 + \cdots + v_n||^2 = ||v_1||^2 + \cdots + ||v_n||^2$$

Theorem 4.5 *Let $(V, \langle,\rangle)$ be an inner product space and O an orthogonal subset of V. Then O is linearly independent.*

Proof Suppose $\sum_{i=1}^{n} a_i v_i = 0$ is a linear dependence on O. Then

$$0 = \left\langle \sum_{i=1}^{n} a_i v_i, v_j \right\rangle = \sum_{i=1}^{n} a_i \langle v_i, v_j \rangle = a_j ||v_j||^2$$

As the elements of O are non-zero, $a_j = 0$ for each j so the relation is trivial and the set is linearly independent. □

Orthogonal bases are extremely convenient as they provide a simple way to represent vectors in the space.

Theorem 4.6 *Let $(V, \langle , \rangle)$ be an inner product space and $O = \{u_1, ..., u_n\}$ an orthogonal set. If $v \in \operatorname{span} O$ then we can write*

$$v = \sum_{i=1}^{n} \frac{\langle v, u_i \rangle}{||u_i||^2} u_i$$

Proof As $v \in \operatorname{span} O$ there exist $a_1, ..., a_n \in k$ such that $v = \sum_{i=1}^{n} a_i u_i$. Then for each j

$$\langle v, u_j \rangle = \sum_{i=1} a_i \langle u_i, u_j \rangle = a_j ||u_j||^2$$

which implies $a_j = \frac{\langle v, u_j \rangle}{||u_j||^2}$. □

To build such sets we can use the Gram-Schmidt procedure. First a preliminary result.

Theorem 4.7 *Let $(V, \langle , \rangle)$ be an inner product space and let $O = \{u_1, ..., u_n\}$ be an orthogonal set. If $v \notin \langle u_1, ..., u_n \rangle$, then there is a nonzero $u \in V$ for which $\{u_1, ..., u_n, u\}$ is an orthgonal set and $\langle u_1, ..., u_n, u \rangle = \langle u_1, ..., u_n, v \rangle$. In particular,*

$$u = v - \sum_{i=1}^{n} \frac{\langle v, u_i \rangle}{\langle u_i, u_i \rangle} u_i$$

Proof Set u as in the theorem. Then $u \neq 0$ as $v \notin \operatorname{span} O$, and

$$\langle u, u_i \rangle = \langle v, u_i \rangle - \frac{\langle v, u_i}{\langle u_i, u_i \rangle} \langle u_i, u_i \rangle = 0$$

as desired. □

We can apply this result to obtain the procedure.

Theorem 4.8 (The Gram-Schmidt Orthogonalization Procedure) *Let $\mathcal{B} = (v_1, v_2, ...)$ be a sequence of linearly independent vectors in an inner product space V. Define a sequence $O = (u_1, u_2, ...)$ by repeated application of the previous result, so*

$$u_k = v_k - \sum_{i=1}^{k-1} \frac{\langle v_k, u_i \rangle}{\langle u_i, u_i \rangle} u_i$$

Then O is an orthogonal sequence in v with the property that $\langle u_1, ..., u_k \rangle = \langle v_1, ..., v_k \rangle$ for all $k > 0$.

Proof The result holds immediately if $k = 1$. Assume it holds for $k - 1$, with $u_1, ..., u_{k-1}$ non-zero. As $v_k \notin \langle v_1, ..., v_{k-1} \rangle$, $u_k \neq 0$ and

$$\langle u_1, ..., u_k \rangle = \langle v_1, ..., v_{k-1}, u_k \rangle = \langle v_1, ..., v_{k-1}, v_k \rangle$$

Corollary 4.1 *Every finite dimensional inner product space has an orthonormal basis.*

4.2.1 Orthogonal Projections

Definition 4.5 Let $(V, \langle , \rangle)$ be an inner product space and let $W \leq V$. Let $\mathcal{B} = \{w_1, ..., w_r\}$ be an orthogonal basis of W. If $v \in V$, its **orthogonal projection** on W is the vector

$$\text{proj}_W(v) = \sum_{i=1}^{n} \frac{\langle v, w_i \rangle}{\langle w_i, w_i \rangle} w_i$$

The $\frac{\langle v, w_i \rangle}{\langle w_i, w_i \rangle}$ are called the **Fourier coefficients** of the projection. Projections give us best approximations of vectors in a subspace.

Theorem 4.9 *Let $(V, \langle , \rangle)$ be an inner product space and $O = \{u_1, ..., u_k\}$ an orthgonal subset and let $S = \operatorname{span} O$. Let $\hat{v}$ denote the projection of $v \in V$ onto S. Then $\hat{v}$ is the unique vector in S for which $(v - \hat{v}) \perp S$. Also $\hat{v}$ is the* ***best approximation*** *to v from within S, that is, $\hat{v}$ is the unique vector closest to $v$in the sense $||v - \hat{v}|| < ||v - s||$ for all $s \in S \backslash \{\hat{v}\}$. We also have* ***Bessel's inequality*** *for all $v \in V$, which states*

$$||\hat{v}|| \leq ||v||$$

Proof For part (1), since $\langle v - \hat{v}, u_i \rangle = \langle v, u_i \rangle - \langle \hat{v}, u_i \rangle = 0$, it follows $v - \hat{v} \in S^\perp$. Also if $v - s \in S^\perp$ for $s \in S$, then $s - \hat{v} \in S$ and

$$s - \hat{v} = (v - \hat{v}) - (v - s) \in S^\perp$$

so $s = \hat{v}$ as $S \cap S^\perp = \{0\}$. For the second statement, if $s \in S$, then $v - \hat{v} \in S^\perp$ implies that $(v - \hat{v}) \perp (\hat{v} - s)$, so

$$||v - s||^2 = ||v - \hat{v}||^2 + ||\hat{v} - s||^2$$

Hence $||v - s|| \geq ||\hat{v} - v||$, with equality if and only if $s = \hat{v}$. □

The following then follows.

Theorem 4.10 *If S is a finite dimensional subspace of an inner product space V, then $V = S \oplus S^\perp$, so* $\dim(V) = \dim(S) + \dim(S^\perp)$.

The result follows from the fact that if S is finite dimensional the projection of any vector in V is well defined, so $v = \text{proj}_S(v) + (v - \text{proj}_S(v)) \in S \oplus S^\perp$.

4.2.2 Riesz Representation Theory

Theorem 4.11 (The Riesz Representation Theorem) *Let $(V, \langle,\rangle)$ be a finite-dimensional inner product space.*

- *The map $\tau : V \longrightarrow V^*$ defined by $\tau x = \langle \cdot, x\rangle$ is a conjugate isomorphism. In particular, for each $f \in V^*$ there exists a unique vector $R_f \in V$ for which $f = \langle \cdot, R_f\rangle$ called the Riesz vector.*
- *If $O = \{u_1, \ldots, u_n\}$ is an orthonormal basis of V, then for $f \in V^*$, $R_f = \sum_{i=1}^n \overline{g(u_i)}u_i$.*
- *The map $R : V^* \longrightarrow V$ defined by $Rf = R_f$ is also a conjugate isomorphism and the inverse of τ.*

Proof We prove the first two points together. As inner products are positive definite, τ is an injection. Next, if $f \in V^*$ we observe that for any $w \in V$,

$$\langle w, R_f\rangle = \left\langle w, \sum_{i=1}^n \overline{f(u_i)}u_i\right\rangle = \sum_{i=1}^n \langle w, u_i\rangle f(u_i) = f(w)$$

Thus $\langle \cdot, R_f\rangle = f$, so τ is surjective. Further, if $\omega \in (\ker f)^\perp \backslash \{0\}$ for $f \neq 0$, then $R_f = \frac{\overline{f(\omega)}}{||\omega||^2}\omega$. Hence τ is a conjugate isomorphism. R being the inverse of a conjugate isomorphism is itself a conjugate isomorphism. Explicitly

$$\begin{aligned}\langle v, R_{rf+sg}\rangle &= (rf + sg)(v)\\ &= rf(v) + sg(v)\\ &= \langle v, \overline{r}R_f\rangle + \langle v, \overline{s}R_g\rangle\\ &= \langle v, \overline{r}R_f + \overline{s}R_g\rangle\end{aligned}$$

for al $v \in V$. □

4.3 Adjoint of a Linear Operator

For finite dimensional inner product spaces we have the existence of a unique map $\tau^* : W \longrightarrow V$ for each map $\tau : V \longrightarrow W$ defined by the condition $\langle \tau v, w\rangle_W = \langle v, \tau^* w\rangle_V$ for all $v \in V$ and $w \in W$. This is called the **adjoint** of τ.

Proof If τ^* exists it is unique. Indeed, if σ is another map satisfying the defining property of the adjoint, $\langle v, \sigma w\rangle_V = \langle v, \tau^* w\rangle_V$ for all $v \in V$ and $w \in W$, so $\sigma = \tau^*$.

If V and W are finite dimensional we have the Riesz representation theorem, and can use it to define τ^*. Specifically, for each $w \in W$ the linear functional $f_w \in V^*$ defined by $f_w v = \langle \tau v, w\rangle_W$ has the form $f_w v = \langle v, R_{f_w}\rangle_V$ where $R_{f_w} \in V$ is the Riesz vector. If we define τ^* by $\tau^* w = R_{f_w} = R(f_w)$, where R is the Riesz map, then

$$\langle v, \tau^* w\rangle_V = \langle v, R_{f_w}\rangle_V = f_w v = \langle \tau v, w\rangle$$

Finally, since $\tau^* = R \circ f$ is the composition of the Riesz map R and the map $f : w \mapsto f_w$, and since both are conjugate linear, their composition is linear. □

We have the following basic properties of the adjoint operation.

Theorem 4.12 *Let V and W be finite-dimensinoal inner product spaces. For every $\sigma, \tau \in \mathcal{L}(V, W)$ and $r \in k$,*

- $(\sigma + \tau)^* = \sigma^* + \tau^*$
- $(r\tau)^* = \overline{r}\tau^*$
- $\tau^{**} = \tau$
- *If $V = W$, then $(\sigma\tau)^* = \tau^*\sigma^*$*
- *If τ is invertible, then $(\tau^{-1})^* = (\tau^*)^{-1}$*
- *If $V = W$ and $p(x) \in \mathbb{R}[x]$, then $p(\tau)^* = p(\tau^*)$*

Moreover, if $\tau \in \mathcal{L}(V0$ and S is a subspace of V, then

- *S is τ-invariant if and only if $S^\perp$ is τ^*-invariant*
- *$(S, S^\perp)$ reduces τ if and only if S is both τ-invariant and τ^*-invariant, in which case $(\tau|_S)^* = (\tau^*)|_S$.*

Proof We prove the fifth statement and leave the rest to the reader. If τ is invertible, τ^{-1} is a linear operator and for any $v, u \in V$ we have

$$\langle u, v\rangle = \langle \tau\tau^{-1}u, v\rangle = \langle u, (\tau^{-1})^*\tau^* v\rangle$$

so $(\tau^{-1})^*\tau^* = \mathrm{id}_V$, and similarly $\tau^*(\tau^{-1})^* = \mathrm{id}_W$. □

We also have the following relations between τ and τ^*:

Theorem 4.13 *Let $\tau \in \mathcal{L}(V, W)$, where V and W are finite-dimensional inner product spaces.*

- $\ker(\tau^*) = \mathrm{Im}(\tau)^\perp$ *and* $\mathrm{Im}(\tau^*) = \ker(\tau)^\perp$

- $\ker(\tau^*\tau) = \ker(\tau)$ *and* $\ker(\tau\tau^*) = \ker(\tau^*)$
- $\mathrm{Im}(\tau^*\tau) = \mathrm{Im}(\tau^*)$ *and* $\mathrm{Im}(\tau\tau^*) = \mathrm{Im}(\tau)$

If we look at the matrix representation of the adjoint over an orthonormal basis it takes on a special simple form.

Theorem 4.14 *Let $\tau \in \mathcal{L}(V, W)$ where V and W are finite-dimensional inner product spaces. If $\mathcal{B}$ and $\mathcal{C}$ are ordered orthonormal bases for V and W, respectively, then*

$$[\tau^*]_{\mathcal{C}}^{\mathcal{B}} = ([\tau]_{\mathcal{B}}^{\mathcal{C}})^*$$

This follows from the result that for such orthonormal bases, $([\tau]_{\mathcal{B}}^{\mathcal{C}})_{i,j} = \langle \tau b_j, c_i \rangle$, so

$$([\tau^*]_{\mathcal{C}}^{\mathcal{B}})_{i,j} = \langle \tau^* c_j, b_i \rangle = \langle c_j, \tau b_i \rangle = \overline{\langle \tau b_i, c_j \rangle} = \overline{([\tau]_{\mathcal{B}}^{\mathcal{C}})_{j,i}}$$

4.4 Spectral Theory

We now use the properties in the relationship between τ and τ^* to determine a spectral theory for operators.

Lemma 4.2 *Let V be a finite dimensional inner product space and $\tau \in \mathcal{L}(V)$. If τ has an eigenvector, then so does τ^*.*

Proof Let u be an eigenvector fo τ with eigenvalue λ. Then for any $v \in V$ we can write

$$0 = \langle (\tau - \lambda \mathrm{id}_V)u, v \rangle = \langle u, (\tau^* - \overline{\lambda} \mathrm{id}_V)v \rangle$$

Thus $u \in \mathrm{Im}(\tau^* - \overline{\lambda}\mathrm{id}_V)^\perp$ with $u \neq 0$, so $\dim \mathrm{Im}(\tau^* - \overline{\lambda}\mathrm{id}_V) < \dim V$ which implies there exists $w \in \ker(\tau^* - \overline{\lambda}\mathrm{id}_V)$ with $w \neq 0$. Thus w is an eigenvector of τ^* of eigenvalue $\overline{\lambda}$. □

An important result in the spectral theory of vector space is Schur's Lemma.

Lemma 4.3 (Schur's Lemma) *Let V be a finite dimensional inner product space over k and $\tau \in \mathcal{L}(V)$. Suppose the characteristic polynomial of τ splits over k. Then there exists an orthonormal basis $\mathcal{B} = \{v_1, ..., v_n\}$ of V such that the matrix $[\tau]_{\mathcal{B}}^{\mathcal{B}}$ is upper triangular.*

Proof Let $p(t)$ be the characteristic polynomial of τ. We proceed by induction on $\dim V = n$. If $n = 1$ the result holds for any normalized basis of V. Suppose the result holds for any space of dimension $< n$. Then, as $p(t)$ splits over k it has a root $\lambda \in k$. It follows that $\dim E_\lambda \geq 1$. Let $U = \mathrm{Im}(\tau - \lambda\, \mathrm{Id}_V)$. Then $\dim U < n$ and U is τ invariant. Let $\{v_1, ..., v_m\}$ be an orthonormal basis of U for which $\tau v_i \in \mathrm{span}(v_1, ..., v_i)$ by the inductive hypothesis. Then extend it to an orthonormal basis $\{v_1, ..., v_n\}$ of V. It follows that

$$\tau v_j = (\tau - \lambda \operatorname{Id}_V)v_j + \lambda v_j \in \operatorname{span}(v_1, ..., v_m, v_j) \subseteq \operatorname{span}(v_1, ..., v_j)$$

for all $m < j \leq n$, so this is an orthonormal basis with respect to which the matrix of τ is upper triangular. □

We now look for the conditions needed to find an orthonormal basis $\mathcal{B}$ of eigenvectors which diagonalize an operator $\tau \in \mathcal{L}(V)$. In this case $([\tau]_{\mathcal{B}}^{\mathcal{B}})^* = [\tau^*]_{\mathcal{B}}^{\mathcal{B}}$, so τ and τ^* are simultaneously diagonalized, with conjugate eigenvalues. If we represent these matrices by D and D^*, we observe that $DD^* = D^*D$, so by the linear isomorphism given by matrix representations of operators we find that $\tau\tau^* = \tau^*\tau$ is a necessary condition for finding an orthonormal basis of eigenvectors. We shall shortly show that if $k = \mathbb{C}$ this condition is also sufficient.

Definition 4.6 Let V be an inner product space. $\tau \in \mathcal{L}(V)$ is called **normal** if $\tau\tau^* = \tau^*\tau$.

Here are some notable properties of normal operators.

Theorem 4.15 *Let V be an inner product space and $\tau \in \mathcal{L}(V)$ be normal. Then*

- $||\tau^* x|| = ||\tau x||$ *for all $x \in V$*
- $\tau - \alpha \operatorname{Id}_V$ *is normal for all $\alpha \in k$*
- *If u is an eigenvector of τ, then it is an eigenvector of τ^* as well.*
- *If $\lambda_1 \neq \lambda_2$ are eigenvalues of τ, then $\langle E_{\lambda_1}, E_{\lambda_2} \rangle = \{0\}$.*

Proof We prove the third and fourth claims. If u is an eigenvector of τ with eigenvalue λ, then $(\tau - \lambda \operatorname{Id}_V)u = 0$, so $||(\tau^* - \overline{\lambda} \operatorname{Id}_V)u|| = ||(\tau - \lambda \operatorname{Id}_V)u|| = 0$, so u must be an eigenvector of τ^* with eigenvalue $\overline{\lambda}$.

For the fourth claim, we observe that

$$\lambda_1 \langle u_1, u_2 \rangle = \langle \tau u_1, u_2 \rangle = \langle u_1, \tau^* u_2 \rangle = \langle u_1, \overline{\lambda}_2 u_2 \rangle = \lambda_2 \langle u_1, u_2 \rangle$$

for all $u_1 \in E_{\lambda_1}$ and $u_2 \in E_{\lambda_2}$. Thus as $\lambda_1 \neq \lambda_2$ we must have $\langle u_1, u_2 \rangle = 0$. □

Now we can finally state and prove the spectral theory of operators on complex vector spaces.

Theorem 4.16 ((Complex) Spectral Theorem) *Let V be an n-dimensional inner product space over $\mathbb{C}$ and let $\tau \in \mathcal{L}(V)$. Then the following statements are equivalent.*

- *τ is normal*
- *There exists an orthonormal basis of V that consists of eigenvectors of τ.*

Proof We have already showed that the second statement implies the first. Now, suppose τ is normal. By Schur's lemma we have an orthonormal basis $\mathcal{B}$ of V for which $A = [\tau]_{\mathcal{B}}^{\mathcal{B}}$ is upper triangular.

We proceed by induction on n to show A is in fact diagonal. If $n = 1$ A is trivially diagonal. Suppose the claim holds for any dimension $< n$. Now, we have that $||\tau v_1||^2 = |\langle \tau v_1, v_1\rangle|^2 = ||\tau^* v_1||^2 = \sum_{i=1}^n |\langle \tau^* v_1, v_i\rangle|^2 = \sum_{i=1}^n |\langle \tau v_i, v_1\rangle|^2$. Thus we have that $\langle \tau v_i, v_1\rangle = 0$ for $1 < i \leq n$. Thus $\tau v_i \in \mathrm{span}(v_2, ..., v_n)$ for $i > 1$, so $U = \mathrm{span}(v_2, ..., v_n)$ is a τ-invariant subspace of V, and $\mathcal{B}_1 = \{v_2, ..., v_n\}$ is an orthonormal basis for which the matrix of $\tau|_U : U \to U$ is upper triangular, so the matrix is diagonal. Hence for all $i \neq j > 1$, $\langle \tau v_i, v_j\rangle = 0$. Thus we have that $[\tau]_{\mathcal{B}}^{\mathcal{B}}$ is diagonal. Therefore $\mathcal{B}$ is a basis of eigenvectors of τ. □

For the case of the real spectral theory we require a stronger condition than normality.

Definition 4.7 Let V be an inner product space over k. Then $\tau \in \mathcal{L}(V)$ is called **self-adjoint** if $\tau^* = \tau$. We also say $A \in M_{n,n}(k)$ is **self-adjoint** if $A^* = A$. If $k = \mathbb{C}$ we also call A **Hermitian**, and if $k = \mathbb{R}$ we say that A is **symmetric**.

Theorem 4.17 *Let V be an inner product space over k. If $\tau \in \mathcal{L}(V)$ is self-adjoint, then the eigenvalues of T are real.*

Proof Let λ be an eigenvalue of T with eigenvector v. Observe that $\lambda\langle v, v\rangle = \langle \tau v, v\rangle = \langle v, \tau v\rangle = \overline{\lambda}\langle v, v\rangle$, so $\lambda = \overline{\lambda}$. Hence $\lambda \in \mathbb{R}$. □

We now can state and prove the real spectral theorem.

Theorem 4.18 ((Real) Spectral Theorem) *Let $\tau \in \mathcal{L}(V)$ for an n-dimensional real inner product space V. Then the following statements are equivalent.*

- *τ is self-adjoint*
- *There exists an orthonormal basis of V that consists of eigenvectors of τ.*

Proof If τ is self-adjoint all of its eigenvalues are real so its characteristic polynomial splits over $\mathbb{R}$. Consequently, as $\tau^*\tau = \tau^2 = \tau\tau^*$, we have by the proof of the complex spectral theory that there exists an orthonormal basis of eigenvectors of τ.

Conversely, if such an orthonormal basis exists, then $[\tau^*]_{\mathcal{B}}^{\mathcal{B}} = ([\tau]_{\mathcal{B}}^{\mathcal{B}})^*$. But as all the eigenvalues are real this corresponds simply with a transposition, so as the matrix is diagonal we have that $[\tau^*]_{\mathcal{B}}^{\mathcal{B}} = [\tau]_{\mathcal{B}}^{\mathcal{B}}$. By the linear matrix representation isomorphism for choice of bases, we have $\tau = \tau^*$ so τ is self-adjoint. □

Now we describe the change of bases matrices which appear in these spectral theories.

Definition 4.8 Let V be an inner product space over k. $\tau \in \mathcal{L}(V)$ is called **unitary** if $\tau^*\tau = \tau\tau^* = \mathrm{Id}_V$.

Unitary operators are characterized by the following properties.

Theorem 4.19 *Let $\tau \in \mathcal{L}(V)$ where V is a finite dimensional inner product space. The following statements are equivalent.*

- $\tau^*\tau = \tau\tau^* = \mathrm{Id}_V$
- $\langle \tau x, \tau y\rangle = \langle x, y\rangle$ *for all* $x, y \in V$
- *If* $\mathcal{B} = \{v_1, ..., v_n\}$ *is an orthonormal basis of* V*, then* $\{\tau v_1, ..., \tau v_n\}$ *is an orthonormal basis of* V*.*
- $||\tau x|| = ||x||$ *for all* $x \in V$

Proof The first three statements are easily seen to be equivalent, and the fourth is easily seen to be implied by the other three. For four implies one, we have that τ is injective and hence a linear isomorphism seen V is finite dimensional. Then $\langle x, (\tau^*\tau - \mathrm{Id}_V)x\rangle = 0$ for all $x \in V$. Notive that $(\tau^*\tau - \mathrm{Id}_V)^* = \tau^*\tau - \mathrm{Id}_V$ so the operator is self-adjoint. Therefore we have an orthonormal basis of eigenvectors of $\tau^*\tau - \mathrm{Id}_V$, so $0 = \langle u_i, (\tau^*\tau - \mathrm{Id}_V)u_i\rangle = \overline{\lambda}_i\langle u_i, u_i\rangle = \overline{\lambda}_i$. Thus all eigenvalues of $\tau^*\tau - \mathrm{Id}_V$ are 0, so with respect to its eigenbasis it is the zero matrix. By injectivity of the matrix representation we find $\tau^*\tau = \mathrm{Id}_V$. Similarly, $\tau\tau^* = \mathrm{Id}_V$. □

Unitary operators on complex inner product spaces are also characterized as follows.

Theorem 4.20 *Let* $\tau \in \mathcal{L}(V)$*, where* V *is an n-dimensional complex inner product space. The following statements are equivalent.*

- τ *is unitary*
- V *has an orthonormal basis of eigenvectors* $\mathcal{B} = \{v_1, ..., v_n\}$ *such that* $\tau v_i = \lambda_i v_i$ *and* $|\lambda_i| = 1$ *for* $i = 1, 2, ..., n$*.*

This follows by the complex spectral theorem and the last result. We also have a characterization in the case of real inner product spaces.

Theorem 4.21 *Let* $\tau \in \mathcal{L}(V)$*, where* V *is an n-dimensional real inner product space. The following statements are equivalent.*

- τ *is both self-adjoint and orthogonal*
- V *has an orthonormal basis of eigenvectors* $\mathcal{B} = \{v_1, ..., v_n\}$ *such that* $\tau v_i = \lambda_i v_i$ *and* $|\lambda_i| = 1$*, for* $i = 1, 2, ..., n$

This theorem follows from the real spectral theorem and our previous characterization.

Chapter 5

Spectral Theory

Chapter 6

Canonical Forms

This chapter aims to investigate the canonical forms of linear operators on finite dimensional vector spaces. A canonical form for linear operators $\mathcal{L}(V, W)$ is a collection of representatives for some equivalence relation on $\mathcal{L}(V, W)$, such that we only have one representative for each equivalence class. Some of the constructions in this chapter follow from results on module theory over PIDs which is discussed in Chapter 9. We begin by investigating the general form for operators in $\mathcal{L}(V, W)$ known as the Singular Value Decomposition.

6.1 Singular Value Decomposition

Theorem 6.1 (Singular Value Decomposition) *Let V and W be finite dimensional inner product spaces over k (= $\mathbb{C}$ or $\mathbb{R}$) and let $\tau \in \mathcal{L}(V, W)$ have rank r. Then there are ordered orthonormal bases $\mathcal{B}$ and C for V and W, respectively, for which $\mathcal{B} = (u_1, .., u_r, u_{r+1}, ..., u_n)$ where up to r is an ONB of* $\mathrm{Im}(\tau^*)$ *and from $r + 1$ to n is an ONB for* $\ker(\tau)$*, and where $C = (v_1, ..., v_r, v_{r+1}, ..., v_m)$ with the vectors up to r being an ONB for* $\mathrm{Im}(\tau)$ *and the vectors from $r + 1$ to m being an ONB for* $\ker(\tau^*)$*. Moreover, for $1 \le k \le r$,*

$$\tau u_i = s_i v_i, \quad \tau^* v_i = s_i u_i$$

*where $s_i > 0$ are called the **singular values** of τ and defined by $\tau^*\tau u_i = s_i^2 u_i$, $s_i > 0$ for $i \le r$.*

Proof First observe that $\tau^*\tau$ is self-adjoint and positive definite. Additionally $r = \mathrm{rank}(\tau) = \mathrm{rank}(\tau^*\tau)$, and by the spectral theorems V has an ordered orthonormal basis $\mathcal{B} = (u_1, ..., u_r, u_{r+1}, ..., r_n)$ of eigenvectors for $\tau^*\tau$, where the corresponding eigenvalues can be arranged so that $\lambda_1 \ge \cdots \ge \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_n$. The set $(u_{r+1}, ..., u_n)$ is an ONB for $\ker(\tau^*\tau) = \ker(\tau)$, and so $(u_1, ..., u_r)$ is an ONB for $\ker(\tau)^\perp = \mathrm{Im}(\tau^*)$. For $i = 1, ..., r$, the positive numbers $s_i = \sqrt{\lambda_i}$ are called the **singular values** of τ. If we set $s_i = 0$ for $i > r$, then $\tau^*\tau u_i = s_i^2 u_i$ for $i = 1, ..., n$. Set $v_i = (1/s_i)\tau u_i$ for each $i \le r$, so $\tau u_i = s_i v_i$ for $i \le r$, and $\tau^* v_i = s_i u_i$ for $i \le r$. The vectors $v_1, ..., v_r$ are orthonormal, since if $i, j \le r$, then

$$\langle v_i, v_j \rangle = \frac{1}{s_i s_j}\langle \tau u_i, \tau u_j \rangle = \frac{1}{s_i s_j}\langle \tau^*\tau u_i, u_j \rangle = \frac{s_i}{s_j}\langle u_i, u_j \rangle = \delta_{i,j}$$

Hence $(v_1, ..., v_r)$ is an orthonormal basis for $\mathrm{Im}(\tau) = \ker(\tau^*)^\perp$, which can be extended to an orthonormal basis $C = (v_1, ..., v_m)$ for v, with the extension $(v_{r+1}, ..., v_m)$ being an orthonormal basis for $\ker(\tau^*)$. Moreover, since $\tau\tau^* v_i = s_i \tau u_i = s_i^2 v_i$, the vectors $v_1, ..., v_r$ are eigenvectors for $\tau\tau^*$ with the same eigenvalues $\lambda_i = s_i^2$ as for $\tau^*\tau$. This completes the proof. □

Writing this result in the language of matrices gives the following decomposition.

Corollary 6.1 (Singular Value Decomposition) *If $A \in M_{m,n}(k)$ of rank r then there exists an $n \times n$ unitary matrix $Q = [q_1 \; q_2 \; \cdots \; q_n]$ where the columns of Q are eigenvectors of A^*A ordered in decreasing eigenvalue $\lambda_1 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_n$, where $p_1 = \frac{1}{\sqrt{\lambda_1}} A q_1, ..., p_r = \frac{1}{\sqrt{\lambda_r}} A q_r$ forms an orthonormal basis for $\mathrm{Im}(A)$, and extending it to an orthonormal basis of k^m, $p_1, ..., p_m$, the matrix $P = [p_1 \; \cdots \; p_m]$ is unitary. Finally, we have that*

$$P^*AQ = \sum = \left[\begin{array}{ccc|c} \sqrt{\lambda_1} & & & \\ & \ddots & & \mathbf{0} \\ & & \sqrt{\lambda_r} & \\ \hline & \mathbf{0} & & \mathbf{0} \end{array}\right]$$

6.2 Eigenspaces and Diagonalization

Next we investigate a special form for square matrices. However, this form is not canonical as it is too restrictive, so not every equivalence class has a diagonal representative.

To discuss the notion of diagaonalization we note that having a diagonal basis is equivalent to having a basis for which our operator acts by scaling on each vector.

Definition 6.1 Let $V \in \mathbf{Vect}_k$ and $\tau \in \mathcal{L}(V)$. A scalar $\lambda \in k$ is called an **eigenvalue** of τ if there exists a nonzero vector $v \in V$ for which $\tau v = \lambda v$. In this case v is called an **eigenvector** of τ with associated eigenvalue λ. The collection of all eigenvectors associated with λ, together with 0, forms a subspace of V called the **eigenspace** of λ and denoted by E_λ. The collection of all eigenvalues of τ is called its **spectrum**.

We have the following result.

Theorem 6.2 *Let $\tau \in \mathcal{L}(V)$. A scalar $\lambda \in k$ is an eigenvalue of τ if and only if $\tau - \lambda I_V$ is singular.*

Thus we can find the eigenvalues of an operator by finding the roots of the **characteristic polynomial** given by $c_\tau(\lambda) = \det(\tau - \lambda I_V)$, where we take any matrix representation of the transformation.

Theorem 6.3 *Suppose $\lambda_1, ..., \lambda_k \in Spec(\tau)$ are distinct eigenvalues of a linear operator $\tau \in \mathcal{L}(V)$. Then if $v_i \in E_{\lambda_i}$ are non-zero eigenvectors, $\{v_1, ..., v_k\}$ are linearly independent.*

Proof Suppose we have a linear relation of minimal size $r_1v_1 + \cdots + r_jv_j = 0$. Applying τ to both sides $r_1\lambda_1v_1 + \cdots + r_j\lambda_jv_j = 0$. Subtracting λ_1 times the first equation from the second, $r_2(\lambda_2 - \lambda_1)v_2 + \cdots + r_j(\lambda_j - \lambda_1)v_j = 0$. This is a smaller linear relation, contradicting minimality, so we must have that $\{v_1, ..., v_k\}$ is linearly independent. □

Since the eigenvalues of τ are given by the roots of a polynomial of order $\dim V$, τ has at most $\dim V$ eigenvalues. This can also be seen using the previous theorem and the uniqueness of the dimension of a space.

Corollary 6.2 *Let $V \in \mathbf{Vect}_k$. If $\lambda_1, ..., \lambda_k$ are pairwise distinct eigenvalues of $\tau \in \mathcal{L}(V)$, then the sum $E_{\lambda_1} + \cdots + E_{\lambda_k}$ is direct.*

We can now characterize diagonalizability of operators.

Theorem 6.4 *Let $V \in \mathbf{Vect}_k$, $\dim V = n \in \mathbb{N}$, and let $\tau \in \mathcal{L}(V)$. Then T is diagonalizable if and only if $V = \bigoplus_{\lambda \in Spec(\tau)} E_\lambda$.*

Definition 6.2 Let $V \in \mathbf{Vect}_k, \dim V = n \in \mathbb{N}$. Let $\tau \in \mathcal{L}(V)$ and λ a root of $c_\tau(t)$. Then the **algebraic multiplicity of** λ is the largest positive integer k such that $c_\tau(t) = (t-\lambda)^k q(t)$ with $q(\lambda) \neq 0$.

Theorem 6.5 *Let $V \in \mathbf{Vect}_k, \dim V = n \in \mathbb{N}$ and let $\tau \in \mathcal{L}(V)$. If $\lambda \in Spec(\tau)$ with algebraic multiplicity k, then $1 \leq \dim(E_\lambda) \leq k$.*

Proof Since λ is an eigenvalue of τ, $\dim(E_\lambda) \geq 1$. Now let $\{v_1, ..., v_p\}$ be a basis of E_λ and extend it into a basis $\mathcal{B}$ of V. Then

$$[\tau]_{\mathcal{B}}^{\mathcal{B}} = \left[\begin{array}{ccc|c} \lambda & & & \\ & \ddots & & \mathbf{A} \\ & & \lambda & \\ & \mathbf{0} & & \mathbf{B} \end{array}\right]$$

so $c_\tau(t) = (t-\lambda)^p c_{\mathbf{B}}(t)$. Thus we must have that $p \leq k$, so $\dim(E_\lambda) \leq k$. □

Definition 6.3 Let $\tau \in \mathcal{L}(V)$ and $\lambda \in \mathrm{Spec}(\tau)$. Then $\dim(E_\lambda)$ is called the **geometric multiplicity** of λ.

We then have the following result.

Theorem 6.6 *Let $V \in \mathbf{Vect}_k, \dim V = n \in \mathbb{N}$, and let $\tau \in \mathcal{L}(V)$. Then τ is diagonalizable if and only if for all $\lambda \in Spec(\tau)$, $\dim(E_\lambda) =$ the algebraic multiplicity of λ.*

6.2.1 Invariant Subspaces

Definition 6.4 Let $V \in \mathbf{Vect}_k$ and $\tau \in \mathcal{L}(V)$. A subspace $U \leq V$ is said to be **invariant under** τ or τ-**invariant** if $\tau(U) \subseteq U$.

This means that τ restricts to a linear operator on U, or $\tau|_U \in \mathcal{L}(U)$.

Theorem 6.7 *Let $V \in \mathbf{Vect}_k, \dim V = n \in \mathbb{N}$ and let $\tau \in \mathcal{L}(V)$. If $U \leq V$ is τ-invariant, then the characteristic polynomial of $\tau|_U$ divides the characteristic polynomial of τ.*

Proof Let $\beta = \{u_1, ..., u_m\}$ be a basis of U and extend to a basis $\mathcal{B} = \{u_1, ..., u_m, v_{m+1}, ..., v_n\}$ of V. Then

$$[\tau]_{\mathcal{B}} = \begin{bmatrix} [\tau|_U]_\beta & \mathbf{A} \\ \mathbf{0} & \mathbf{B} \end{bmatrix}$$

so $c_\tau(t) = c_{\tau|_U}(t)c_{\mathbf{B}}(t)$, as desired. □

Corollary 6.3 *Let $V \in \mathbf{Vect}_k, \dim V = n \in \mathbb{N}$ and $\tau \in \mathcal{L}(V)$. If $V = U_1 \oplus \cdots \oplus U_k$ for τ-invariant subspaces U_i, then we have that*

$$c_\tau(t) = \prod_{i=1}^{k} c_{\tau|_{U_i}}(t)$$

We have the following important result which shall be used in deriving our next canonical form.

Theorem 6.8 *Let $V \in \mathbf{Vect}_k, \dim V = n \in \mathbb{N}$ and let $\tau \in \mathcal{L}(V)$. For $u \in V\backslash\{0\}$ let $U = \operatorname{span}(u, \tau u, \tau^2 u, ...)$. Then $\dim(U) = k \leq n$, and we have*

- *$\{u, \tau u, ..., \tau^{k-1}u\}$ is a basis of U*
- *U is τ-invariant*
- *U is the smallest invariant subspace of V that contains u*
- *If $a_0, a_1, ..., a_{k-1} \in k$ such that $a_0 u + a_1 \tau u + \cdots + a_{k-1}\tau^{k-1}u + \tau^k u = 0$, then $c_{\tau|_U}(t) = a_0 + a_1 t + \cdots + a_{k-1}t^{k-1} + t^k$.*

Proof First let $k \in \mathbb{N}$ be the largest integer for which $\{u, \tau u, ..., \tau^{k-1}u\}$ is linearly independent. Since $\dim V = n$ k exists and is $\leq n$, and as $u \neq 0$, $k \geq 1$. It follows that $\tau^k u \in \operatorname{span}(u, \tau u, ..., \tau^{k-1}u)$, and so if $\tau^{k+l}u = \sum_{i=0}^{k-1} a_i \tau^i u$ for some $l \geq 0$, $\tau^{k+l+1}u = \sum_{i=0}^{k-1} a_i \tau^{i+1}u \in \operatorname{span}(u, \tau u, ..., \tau^{k-1}u)$. Hence $U \subseteq \operatorname{span}(u, \tau u, ..., \tau^{k-1}u) \subseteq U$, so $\mathcal{B} = \{u, \tau u, ..., \tau^{k-1}u\}$ is a basis for U. By construction U is τ invariant and the smallest subspace of V that contains u and is τ-invariant. As $\tau^k u \in \operatorname{span}(u, \tau u, ..., \tau^{k-1}u)$, there exist unique $a_0, ..., a_{k-1} \in k$ such that $a_0 u + a_1 \tau u + \cdots + a_{k-1}\tau^{k-1}u + \tau^k u = 0$. Now we have the matrix representation

$$[\tau|_U]_{\mathcal{B}} = \begin{bmatrix} 0 & 0 & 0 & \cdots & -a_1 \\ 1 & 0 & 0 & \cdots & -a_2 \\ 0 & 1 & 0 & \cdots & -a_3 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{k-1} \end{bmatrix}$$

so $c_{\tau|_U}(t) = a_0 + a_1 t + \cdots + a_{k-1}t^{k-1} + t^k$. $\square$

We then obtain the following well known result.

Theorem 6.9 (Cayley-Hamilton) *Let $V \in \mathbf{Vect}_k$, $\dim V = n \in \mathbb{N}$, and let $\tau \in \mathcal{L}(V)$. Then $c_\tau(\tau) = 0 \in \mathcal{L}(V)$.*

This result follows from the previous theorem and our result that the characteristic polynomial of inveriant subspaces divides the characteristic polynomial of the operator.

6.3 Jordan Canonical Form

Now we lessen the condition of diagonalization to obtain a true canonical form. In order to derive this form we introudce the notion of a generalized eigenvector.

Definition 6.5 Let $V \in \mathbf{Vect}_k$ and $\tau \in \mathcal{L}(V)$ with $\lambda \in \operatorname{Spec}(\tau)$. A nonzero vector $v \in V$ is said to be a **generalized eigenvector** of τ associated with λ if there exists $k \geq 1$ such that $(\tau - \lambda \operatorname{Id}_V)^k v = 0$ and $(\tau - \lambda \operatorname{Id}_V)^{k-1} v \neq 0$. k is called the **index** of the generalized eigenvector v.

Note eigenvectors are generalized eigenvectors of index 1, and in general $(\tau - \lambda I)^{k-1}v$ is an eigenvector of τ associated with λ. To show a vector $v \in V$, $v \neq 0$, is a generalized eigenvector of τ associated with λ, it sufficies to show that $(\tau - \lambda \operatorname{Id}_V)^m v = 0$ for some $m \geq 1$. Now, from our previous result we know $U = \operatorname{span}(u, (\tau - \lambda \operatorname{Id}_V)v, \ldots)$ is a τ-invariant subspace with basis $u, (\tau - \lambda \operatorname{Id}_V)u, \ldots, (\tau - \lambda \operatorname{Id}_V)^{k-1}u$. We now define the generalized eigenspace.

Definition 6.6 Let $\tau \in \mathcal{L}(V)$ and $\lambda \in \operatorname{Spec}(\tau)$. We define the **generalized eigenspace** of λ to be G_λ which consists of all generalized eigenvectors as well as $v = 0$.

If r is the largest index of all generalized eigenvectors in G_λ, $G_\lambda = \ker(\tau - \lambda \operatorname{Id}_V)^r$. As $r \leq n$ by our previous result and the fact that $\dim V = n$, we have always that $G_\lambda = \ker(\tau - \lambda \operatorname{Id}_V)^n$.

Definition 6.7 The largest index r of all generalized eigenvectors in G_λ is called the **index** of G_λ.

Although eigenspaces often were of insufficient dimension to give a full basis of the space, this issue doesn't arise anymore for generalized eigenspaces. We show this after showing we can construct bases of a particular type.

Theorem 6.10 *Let $V \in \mathbf{Vect}_k, \dim V = n \in \mathbb{N}$ and $N \in \mathcal{L}(V)$ a nilpotent operator. Then there exists a basis $\{N^{k_1}v_1, ..., Nv_1, v_1, ..., N^{k_j}v_j, ..., Nv_j, v_j\}$ of V.*

Proof We proceed by induction on the dimension n of V. If $n = 1$ then taking any nonzero v_1 gives a basis $\{v_1\}$ of V. Suppose the claim holds for all dimensions $< n$. Then as N is nilpotent $\dim(\operatorname{Im} N) < n$, so there exists a basis $\{N^{k_1}v_1, ..., Nv_1, v_1, ..., N^{k_j}v_j, ..., Nv_j, v_j\}$ of $\operatorname{Im} N$. Then there exists $u_1, ..., u_j$ such that $Nu_i = v_i$, so $\mathcal{A} = \{N^{k_1+1}u_1, ..., Nu_1, u_1, ..., N^{k_j+1}u_j, ..., Nu_j, u_j\}$ is a linearly independent set in V. We extend to a basis of V by adding $w_{j+1}, ..., w_m$. As $Nw_i \in \operatorname{Im} N$ there exists $v_i \in \operatorname{span}(\mathcal{A})$ such that $Nw_i = Nv_i$. Letting $u_i = w_i - v_i \notin \operatorname{span}(\mathcal{A})$ as $w_i \notin \operatorname{span}(\mathcal{A})$, we have our desired basis

$$\{N^{k_1+1}u_1, ..., Nu_1, u_1, ..., N^{k_j+1}u_j, ..., Nu_j, u_j, u_{j+1}, ..., u_m\}$$

Theorem 6.11 *Let $V \in \mathbf{Vect}_k, \dim V = n \in \mathbb{N}$ and $\tau \in \mathcal{L}(V)$. Let $\lambda \in Spec(\tau)$ with algebraic multiplicity m. Then $\dim(G_\lambda) = m$.*

Proof First note that $\tau|_{G_\lambda}$ is a nilpotent mapping, so by the previous theorem we have a basis $\{(\tau - \lambda \operatorname{Id}_V)^{k_1}v_1, ..., (\tau - \lambda \operatorname{Id}_V)v_1, v_1, ..., (\tau - \lambda \operatorname{Id}_V)^{k_j}v_j, ..., (\tau - \lambda \operatorname{Id}_V)v_j, v_j\}$ of G_λ. Now, as both are terminal in their respective sequences, $\ker(\tau - \lambda \operatorname{Id}_V)^n \cap \operatorname{Im}(\tau - \lambda \operatorname{Id}_V)^n = \{0\}$, so by the dimension theorem $V = G_\lambda \oplus \operatorname{Im}(\tau - \lambda \operatorname{Id}_V)^n$, both of which are τ-invariant. Then we can extend this to a basis $\mathcal{B}$ of V with a basis of $\operatorname{Im}(\tau - \lambda \operatorname{Id}_V)^n$. Then the associated matrix representation is

$$[\tau]_{\mathcal{B}} = \left[\begin{array}{c|c} \operatorname{diag}(J_{k_1}(\lambda), ..., J_{k_j}(\lambda)) & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{array}\right]$$

where $J_k(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda & 1 \\ 0 & \cdots & 0 & 0 & \lambda \end{bmatrix}$ is the **Jordan block of size** k. Then $c_\tau(t) = (t - \lambda)^{\dim G_\lambda} c_{\mathbf{A}}(t)$. As $G_\lambda \cap \operatorname{Im}(\tau - \lambda \operatorname{Id}_V)^n = \{0\}$ and all λ eigenvectors are in G_λ, we must have that $c_{\mathbf{A}}(\lambda) \neq 0$. Hence, $\dim G_\lambda = m$ must be the algebraic multiplicity of λ. □

We now have a type of spectral theorem for generalized eigenspaces.

Theorem 6.12 *Let $V \in \mathbf{Vect}_k, \dim V = n \in \mathbb{N}$ be a vector space over an algebraically closed field k. Let $\tau \in \mathcal{L}(V)$ and let $\lambda_1, ..., \lambda_k$ be the distinct eigenvalues of T with multiplicities $n_1, ..., n_k$. Then $V = G_{\lambda_1} \oplus \cdots \oplus G_{\lambda_k}$.*

Proof By the previous theorem it is sufficient to show that for $v_1 \in G_{\lambda_1}, ..., v_k \in G_{\lambda_k}$, not all zero, $v_1 + \cdots + v_k \neq 0$. Towards a contradiction suppose $v_1 + \cdots + v_k = 0$. Without loss of generality we may suppose $v_1 \neq 0$, and that v_1 has index k_1, and in general v_i has index k_i. Then if $(\tau - \lambda_1 \operatorname{Id}_V)^{k_1-1}v_1 = w \neq 0$, $\tau w = \lambda_1 w$, so $\prod_{i=2}^{k}(\tau - \lambda_i \operatorname{Id}_V)^{k_i} w = \prod_{i=1}^{2}(\lambda_1 - \lambda_i)^{k_i} w \neq 0$. However, $0 = (\tau - \lambda_1 \operatorname{Id}_V)^{k_1-1} \prod_{i=2}^{k}(\tau - \lambda_i \operatorname{Id}_V)^{k_i}(v_1 + \cdots + v_k) = \prod_{i=1}^{2}(\lambda_1 - \lambda_i)^{k_i} w$, which is a contradiction. Hence the generalized eigenspaces must be linearly independent. Hence, as $\dim(G_{\lambda_i}) = n_i$ from the previous result, and $n_1 + \cdots + n_k = n$, we have that

$$V = \bigoplus_{i=1}^{k} G_{\lambda_i}$$

The bases with Jordan blocks we used in the previous proofs is known as the **Jordan canonical basis** of τ, and gives the Jordan canonical form when the operator is represented in that basis.

Definition 6.8 Let $V \in \mathbf{Vect}_k$ and $\tau \in \mathcal{L}(V)$. Let $v \in G_\lambda \backslash \{0\}$ for $\lambda \in \mathrm{Spec}(\tau)$. Let $k \geq 1$ be the index of v. Then we call the ordered linearly independent set $\{(\tau - \lambda \, \mathrm{Id}_V)^{k-1} v, ..., (\tau - \lambda \, \mathrm{Id}_V) v, v\}$ is called a **cycle** of generalized eigenvectors of τ associated with the eigenvalue λ. $(\tau - \lambda \, \mathrm{Id}_V)^{k-1} v$ is called the **initial vector** of the cycle and v is called the **end vector** of the cycle. k is called the **length** of the cycle.

Note that the number of Jordan blocks total is given by the number of linearly independent eigenvectors, since all cycles must end in an eigenvector. The next result gives how to determine the number of blocks of any size.

Theorem 6.13 *Let* $\tau \in \mathcal{L}(V)$ *for* $\dim V = n$. *Let* $k_j = \dim(\ker(\tau - \lambda \, \mathrm{Id}_V)^j)$. *Then the number of Jordan blocks of eigenvalue* λ *of size* j *is given by*

$$m_j = 2k_j - k_{j+1} - k_{j-1}$$

Proof First I claim that there are $k_{j+1} - k_j$ Jordan blocks of size $> j$. As each Jordan block of size $> j$ will contribute +1 to the dimension of k_{j+1} versus k_j, $k_{j+1} - k_j$ must be the number of Jordan blocks of size $> j$. Thus, the number of Jordan blocks of size exactly j is $k_j - k_{j-1} - (k_{j+1} - k_j) = 2k_j - k_{j-1} - k_{j+1}$. □

Then we have the following algorithm for determining the Jordan canonical basis of an operator $\tau \in \mathcal{L}(V)$:

(1)Find the value minimal k for which $\ker(\tau - \lambda \, \mathrm{Id}_V)^k = G_\lambda$

(2)Find a basis $(v_k^1, ..., v_k^{m_k})$ of the subspace of $\ker(\tau - \lambda \, \mathrm{Id}_V)^k$ not including $\ker(\tau - \lambda \, \mathrm{Id}_V)^{k-1}$, where $m_k = k_k - k_{k-1}$

(3)If $k = 1$, stop, Else, apply $(\tau - \lambda \, \mathrm{Id}_V)$ to the vectors from the previous part. Extend $((\tau - \lambda \, \mathrm{Id}_V) v_k^1, ..., (\tau - \lambda \, \mathrm{Id}_V) v_k^{m_k})$ to a basis of $\ker(\tau - \lambda \, \mathrm{Id}_V)^{k-1}$ not including $\ker(\tau - \lambda \, \mathrm{Id}_V)^{k-2}$.

(4)Repeat step (3) replacing k with $k - 1$ until we hit $k = 1$.

Part II

Module Theory

Chapter 7

General Theory

Chapter 8

Universal Constructions

Chapter 9

Modules Over PIDs

Chapter 10

Tensor Products

Index