Linear Algebra

Karan Elangovan

April 4, 2021

Contents

1   Vector Spaces

1.1   Vector Space Axioms

Definition 1 (Vector Space). A vector space over a field k is an ordered triple $(V, +, f)$, with $+ : V^2 \to V$ and $f : k \times V \to V$ satisfying

1. $(V, +)$ is an abelian group

2. $1_k v = v$ for all $v \in V$

3. $a(bv) = (ab)v$ for all $a, b \in k$ and $v \in V$

4. $(a + b)v = av + bv$ for all $a, b \in k$ and $v \in V$

5. $a(v_1 + v_2) = av_1 + v_2$ for all $a \in k$ and $v_1, v_2 \in V$

Where we use adjacency to denote f.

   We call the elements of V vectors, the elements of k scalars and f scalar multiplication.

Definition 2 (Subspace). A subspace of a vector space V over k is a subset $U \subset V$ which is also a vector space over k under the induced operations of addition and scalar multiplication.

1.2 Linear Independence, Spanning Sets and Bases

Definition 3 (Linear Combination). Let $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ be a finite[1] set of vectors in V. Then we define a linear combination of $\mathcal{V}$ as a function $f : \mathcal{V} \to k$. To each linear combination corresponds a vector $f(v_1)v_1 + f(v_2)v_2 + \ldots + f(v_n)v_n$ as a linear combination. Whether we are speaking of the actual vector corresponding to a linear combination or the choice of scalars is usually easily inferred from context.

Definition 4 (Span). We define the span of a finite set of vectors $\mathcal{V}$ as the set of all of its linear combinations. If the span is V then we say $\mathcal{V}$ is a spanning set.

Theorem 1. Let $\mathcal{V}$ be a finite set of vectors. Then the following are all equivalent

1. The only linear combination of $\mathcal{V}$ equal to 0 has all of its scalars equal to 0.

2. For every v in the span of $\mathcal{V}$, the choice of scalars is unique.

3. For every $v \in \mathcal{V}$, v is not in the span of the remaining vectors in $\mathcal{V}$.

Proof. (2) is a stronger form of (1), hence (2) implies (1). Now assume (1). Say two linear combinations $a_1v_1 + a_2v_2 + \ldots + a_nv_n$ and $b_1v_1 + b_2v_2 + \ldots + b_nv_n$ are both equal to v, then we have $(a_1 - b_1)v_1 + (a_2 - b_2)v_2 + \ldots + (a_n - b_n)v_n = 0$, which by (1) yields $(a_i - b_i) = 0$, giving $a_i = b_i$, hence the choice of scalars is unique, and so (1) implies (2).

Now assume (1). Say (3) is not true, then we may assume WLOG by re-indexing that $v_n = a_1v_1 + a_2v_2 + \ldots a_{n-1}v_{n-1}$, yielding a non-trivial linear

---

[1] The definition may be extended for sets of infinite cardinality as well, with the restriction that all but finitely many of the choices made by f are 0, allowing us to sidestep issues of infinite sums. However most useful applications and results are with regards to the finite case, so we restrict our definitions and study to this case.

combination that is 0, contradicting (1), hence (1) implies (3). Now assume (3), and say that (1) is not true. Then we have $a_1 v_1 + a_2 v_2 + \ldots + a_n v_n = 0$ with some $a_i \neq 0$. So we may rearrange to yield $a_i = (-\frac{a_1}{a_i})v_1 + \ldots + (-\frac{a_{i-1}}{a_i})v_{i-1} + (-\frac{a_{i+1}}{a_i})v_{i+1} + \ldots + (-\frac{a_n}{a_i})v_n$, contradicting (3). Hence (3) implies (1). $\qquad \square$

Definition 5 (Linear Independence). We say a finite set of vectors $\mathcal{V}$ is linearly independent if it satisfies any of the equivalent conditions in theorem 1.

Theorem 2. The cardinality of any linearly independent set cannot exceed the cardinality of any spanning set.

Proof. Let $\mathcal{W} = \{w_1, w_2, \ldots w_m\}$ be an arbitrary linearly independent set and $\mathcal{V} = \{v_1, v_2 \ldots, v_n\}$ an arbitrary spanning set.

As $\mathcal{V}$ is spanning, w_1 belongs to its span. We claim that we may we may replace m of the v_i with distinct elements of $\mathcal{W}$ to yield a spanning set. We proceed by induction. Assume that the set

$$S = \{w_1, w_2, \ldots w_j, u_1, u_2, \ldots u_{n-j}\}.$$

where $j < m$ and u_i are distinct elements of $\mathcal{V}$ is spanning. Then w_{j+1} must belong to the span of S. If $w_{j+1} \in S$, then it must equal one of the u_i, and so we may replace that u_i with w_{j+1}. Otherwise consider

$$S' = \{w_{j+1}, w_1, w_2, \ldots, w_j, u_1, u_2, \ldots u_{n-j}\}.$$

which must be linearly dependent as w_{j+1} belongs to the span of the remaining vectors. We know that $\{w_{j+1}, w_1, w_2, \ldots w_j\}$ is linearly independent, and so if we keep adding u_i we eventually reach the first u_i in the span of

$$\{w_{j+1}, w_1, w_2, \ldots w_j, u_1, u_2, \ldots u_{i-1}\}.$$

which we know exists as otherwise S' would be linearly independent. Hence we may discard this u_i to leave the desired spanning set. The base case follows from $\mathcal{V}$ being spanning. Hence we have an injection from $\mathcal{W} \to \mathcal{V}$ by which $v \in \mathcal{V}$ is replaced by $w \in \mathcal{W}$ in the construction. Hence $|\mathcal{W}| \leq |\mathcal{V}|$. $\qquad \square$

As a trivial consequence of this, we have that any two bases have the same cardinality as their cardinalities bound eachother above and below as a basis is simultaeneously a spanning set and a linearly independent set. So we may call this invariant quantity the "dimension" of the vector space. However there is a caveat, namely that this assumes that there is a basis at all. When we extend our defintions to inlcude the idea of a linear combination of an infinite set as mentioned in footnote 1 it may be shown that every vector space has a basis (this is a trivial consequence of Zorn's lemma when we view a basis as a maximal linearly independent set)

Definition 6 (Finite-Dimensional Vector Spaces)**.** The vector space V is finite dimensional if it has a basis. If V is finite dimensional, then we call the common cardinality of all of its bases the dimension of V, $\dim V$.

2 Linear Maps

2.1 Definitions

The structure of a vector space is given by the operations of addition and scalar multiplication, so those maps between vector spaces which preserve this structure will be of interest. In the context of linear algebra, we call a homomorphism a linear map.

Definition 7 (Linear Map)**.** Let V and W be k-vector spaces. The function $T : V \to W$ is a linear map if it satisfies

1. $T(u + v) = T(u) + T(v)$ for all $u, v \in V$.

2. $T(\lambda v) = \lambda T(v)$ for all $\lambda \in k$ and $v \in V$.

We denote the set of all linear maps from V to W by $\mathcal{L}(V, W)$.

We call linear maps whose range is k itself treated as a vector space linear functionals.

Theorem 3. Let $\mathcal{V} = \{v_1, v_2, \ldots v_n\}$ be a basis of V. Then for any choice of $w_1, w_2, \ldots w_n \in W$, there is a unique linear map $T \in \mathcal{L}(V, W)$ that maps each v_i to w_i.

Proof. Consider the mapping from $V \to W$ defined by

$$T(a_1 v_1 + a_2 v_2 + \ldots + a_n v_n) = a_1 w_1 + a_2 w_2 + \ldots + a_n w_n.$$

This is well-defined as $\mathcal{V}$ is a basis and hence each vector has a unique representation as a linear combination. Homogeneity and additivity are trivially verifiable. Hence there does exist such a map. The uniqueness follows from expressing each vector as a linear combination of $\mathcal{V}$, from which the requirements of a linear map and our hypothesis leave only one possibile choice for the image. $\square$

A trivial but noteworthy corrolary is that if V and W are finite-dimensional and of equal dimension they must have a bijective linear map bewteen them, meaning they are isomorphic. This means that for finite dimensional vector spaces, the dimension totally determines the structure. So we may treat k^n as the canonical vector space of dimension n.

Definition 8 (Kernel)**.** Let $T \in \mathcal{L}(V, W)$ be a linear map. Then the kernel of T, $\ker T$ is the pre-image of 0.

Similarly to group homomorphisms, the kernel of a linear map is a subspace, and the triviality of the kernel is equivalent to injectivity, and there is also as one would expect an analogue to the isomorphism theorem for groups which we will be able to state and prove once having defined the notion of the quotient of a vector space and its subspace.

2.2 Quotient Spaces

Consider a vector space V and a subspace U. We wish to define a vector space V/U that is conceptually the vector space V, but in which we consider vectors which differ only by some vector in U as identical. More precisely V/U is the set of equivalence classes under the equivalence relation $v \sim u \iff (v - u) \in U$, with addition and scalar multiplication defined on coset representatives. These operations may be shown to be well-defined and to yield a vector space, we omit the proofs as they are essentially identical in logic to showing the cosets of a normal subgroup form a group.

Definition 9 (Quotient Space)**.** The quotient space V/U is the set of cosets of U in V, with the operations of addition and scalar multiplication defined on coset representatives.

Theorem 4 (Fundemental Theorem of Linear Maps)**.** Let $T \in \mathcal{L}(V, W)$ be a surjective linear map. Then $V/\ker T \cong W$.

Proof. We define the map $T' : V/\ker T \to W$ by

$$T'(\ker T + v) = T(v).$$

T' is well-defined as no matter which coset representative we choose, their difference will be in the kernel of T, and so will have no effect on the image by the additivity of T. T' is also clearly linear as T is. If $\ker T + v \in \ker(T')$, then

$T(v) = 0$, and so $\ker v + v = \ker v + 0$, hence the kernel of T' is trivial and so T' is injective. The surjectivity of T' follows from the surjectivity of T. Hence $V/\ker T$ and W are isomorphic by T'. □

Theorem 5. Let V be finite-dimensional, and U a subspace of V. Then $\dim(V/U) = \dim V - \dim U$.

Proof. Let $u_1, u_2, \ldots, u_n$ be a basis of U. Let us extend it into a basis for V to $u_1, u_2, \ldots, u_n, v_1, v_2, \ldots, v_m$. Hence $\dim(V) - \dim(U) = (n + m) - (n) = m$.

We claim that $\mathcal{V} = \{U + v_1, U + v_2, \ldots U + v_m\}$ is a basis of V/U.

$$\{U + u_1, U + u_2, \ldots, U + u_n, U + v_1, U + v_2, \ldots U + v_m\}$$
$$= \{U + 0, U + v_1, U + v_2, \ldots, U + v_m\}$$

spans V/U, and hence so does $\mathcal{V}$, as the 0 vector doesn't affect the span. Say we have a 0 linear combination of $\mathcal{V}$, $a_1(U+v_1)+a_2(U+v_2)+\ldots+a_m(U+v_m) = 0$, so we have $a_1 v_1 + a_2 v_2 + \ldots + a_m v_m \in U$, so we have $b_1, b_2, \ldots b_n \in k$ such that $a_1 v_1 + \ldots + a_m v_m = b_1 u_1 + \ldots + b_m u_m$ as $\{u_1, \ldots, u_m\}$ spans U. Hence by the linear independence of the entire basis of V, we have $a_1 = \ldots = a_m = b_1 = \ldots = b_n = 0$. Hence there are no non-trivial 0 linear combinations of $\mathcal{V}$. Hence $\mathcal{V}$ is a basis as we claimed, and so $\dim(V/U) = m = \dim(V) - \dim(U)$. □

2.3 Matrices

As we have established, a linear map $T \in \mathcal{L}(V, W)$ is uniquely identified by where it maps any given basis $\mathcal{V}$ of V, and to every possible choice of images of $\mathcal{V}$ there corresponds a linear map. This motivates us to define a matrix as a way to represent linear maps.

Definition 10 (Matrix). Let $T \in \mathcal{L}(V, W)$, and $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ and $\mathcal{W} = \{w_1, w_2, \ldots, w_m\}$ be bases of V and W, respectively. Then we define the matrix of T with respect to $\mathcal{V}$ and $\mathcal{W}$, $\mathcal{M}(T, \mathcal{V}, \mathcal{W})$ as an array with m rows and n columns, such that column j gives the scalars of the linear combination of $\mathcal{W}$ which corresponds to $T(v_j)$.

We define addition, multiplication (when the corresponding linear maps are composable) and scalar multiplication on matrices such that the resulting matrix corresponds to the resulting linear map when the corresponding linear maps of the original matrices are combined. Symbolically, $\mathcal{M}(S) + \mathcal{M}(T) = \mathcal{M}(T + S)$, $\lambda \mathcal{M}(T) = \mathcal{M}(\lambda T)$, and $\mathcal{M}(S)\mathcal{M}(T) = \mathcal{M}(S \circ T)$.

2.4 Operators

We consider the special case of linear maps whose domain and range are identical, that is, the endomorphisms of a vector space. The primary characteristic that leads to their richer theory is that any operator may be composed with itself, so we may consider powers of an operator, and hence polynomials of an operator.

Definition 11 (Operator). An operator is a linear map whose domain is the same as its range. We denote the operators of V as $\mathcal{L}(V)$.

Theorem 6. For an operator T on a finite-dimensional space V, injectivity is equivalent to surjectivity.

Proof. Let V have dimension n. We have by the fundemental theorem of linear maps, $V/\ker T \cong T(V)$.

First assume that T is injective, then the kernel is trivial and hence we have $V \cong T(V)$, and so $\dim T(V) = n$, and as $T(V)$ is a subspace of V, we have that $T(V) = V$, meaning T is surjective.

Now assume that T is surjective, then we have $V/\ker T \cong V$, and so $\dim \ker T = 0$, hence the kernel is trivial and so T is injective. $\qquad\square$

3 Eigenvalues and Eigenvectors

3.1 Invariant Subspaces

We may better better understand an operator on V by instead considering how it acts on each subspace in some direct sum decomposition of $V = U_1 \oplus \ldots \oplus U_n$. However T restricted to U_i, $T|_{U_i}$ may not be an operator. It is only an operator when it maps the subspace into itself. This motivates our definition of invariant subspaces.

Definition 12 (Invariant Subspace)**.** We call the subspace U invariant under $T \in \mathcal{L}(V)$ iff $T(U) \subset U$, or equivalently $T|_U$ is an operator.

The simplest case is when the subspace has dimension 1, in which case it is the span of some non-zero vector v, and so it must be the case that $T(v) = \lambda v$ for some $\lambda \in k$. When this is the case we call v an eigenvector and λ its corresponding eigenvalue. Whilst every eigenvector may of course only correspond to a single eigenvalue, the converse does not hold, indeed the eigenvectors of a specfic eigenvalue (which, alongside 0, trivially form a subspace) need not even have dimension 1.

Definition 13. Let $T \in \mathcal{V}$. Then $\lambda \in k$ is an eigenvalue if any of the following equivalent criteria hold

1. We have some non-zero $v \in V$ such that $T(v) = \lambda v$

2. $T - \lambda I$ is not injective.

For any $\lambda \in k$, we define the eigenspace $E(\lambda, T)$ as the kernel of $T - \lambda I$.

Theorem 7. If $v_1, v_2, \ldots, v_n$ are eigenvectors with distinct corresponding eigenvalues, then $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ is linearly independent.

Proof. Assume the contrary, and so we have some j such that $\{v_1, v_2, \ldots, v_j\}$ is linearly independent and $\{v_1, v_2, \ldots, v_{j+1}\}$ is linearly dependent. Hence we have for some choice of $a_1, a_2, \ldots, a_j \in k$ that $v_{j+1} = a_1 v_1 + a_2 v_2 + \ldots + a_j v_j$. We have by scalar multiplication by λ_{j+1} and by application of T that

$$\lambda_{j+1} v_{j+1} = \lambda_{j+1} a_1 v_1 + \lambda_{j+1} a_2 v_2 + \ldots + \lambda_{j+1} a_j v_j$$
$$\lambda_{j+1} v_{j+1} = \lambda_1 a_1 v_1 + \lambda_2 a_2 v_2 + \ldots + \lambda_j a_j v_j$$

Hence we have

$$a_1(\lambda_{j+1} - \lambda_1)v_1 + a_2(\lambda_{j+1} - \lambda_2)v_2 + \ldots + a_j(\lambda_{j+1} - \lambda_j)v_j = 0$$

The eigenvalues are all pairwise distinct, and hence by the linear independence of $\{v_1, v_2, \ldots, v_j\}$ we have that $a_1 = a_2 = \ldots = a_j = 0$. However this gives $v_{j+1} = 0$, which cannot be. Hence $\mathcal{V}$ is linearly independent. $\square$

As these eigenvalues are helpful in understanding the behaviour of an operator, we should ask for which operators there even exists an eigenvalue. In the most general case, if the operator is non-injective, then 0 will be an eigenvalue. Otherwise we restrict our attention to finite dimensional $\mathbb{R}$ and $\mathbb{C}$ vector spaces. An arbitrary operator on a finite dimensional $\mathbb{R}$ vector space need not have any eigenvalues (for example an order 4 turn about the origin), however as the next theorem shows, the answer is in the affirmative for finite dimensional $\mathbb{C}$ vector spaces.

Theorem 8. Let T be an operator over a finite dimensional, non-trivial, $\mathbb{C}$ vector space, V. Then T has an eigenvalue.

Proof. Let $\dim V = n$. Choose an arbitrary non-zero $v \in V$. Then $\mathcal{V} = \{v, T(v), \ldots, T^n(v)\}$ is linearly dependent, so we have some non-trivial zero linear combination of $\mathcal{V}$. Let the scalars be $a_0, a_1, \ldots a_n$ and j be the last non-zero index. j must exist as the combination is non-trivial, and $j \geq 1$, else we'd have that $v = 0$. So we have

$$(a_0 I + a_1 T + a_2 T^2 + \ldots + a_j T^j)v = 0$$

$j \geq 1$, hence by the Fundemental Theorem of Algebra, we may factorise the polynomial into a product of linear factors.

$$(T - \lambda_1 I)(T - \lambda_2 I) \ldots (T - \lambda_j I)v = 0$$

Hence not all of the linear factors may be bijective, and so at least one, $(T - \lambda_j I)$, is non-injective, injectivity is equivalent to bijectivity over finite dimensional vector spaces. This means that λ_j is the desired eigenvalue. $\qquad\square$

3.2 Upper Triangular Matrices

Definition 14 (Upper Triangular Matrix)**.** We say a square matrix is upper triangular if all entries under the diagonal are 0.

If we consider an operator $T \in \mathcal{L}(V)$, we may express the condition of upper-triangularity of $\mathcal{M}(T, \mathcal{V})$ in terms of the operator. Say $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$, then it is trivial to see that upper triangularity is equivalent to each $T(v_j)$ belonging to the span of $\{v_1, v_2, \ldots, v_j\}$.

Theorem 9. Let T be an operator over a finite-dimensional $\mathbb{C}$ vector space. Then there exists a basis $\mathcal{V}$ such that $\mathcal{M}(T, \mathcal{V})$ is upper-triangular.

Proof. Assume by induction that we have a linearly independent set $\mathcal{V} = \{v_1, v_2, \ldots, v_j\}$ such that each $T(v_i) \in \mathrm{span}\{v_1, v_2, \ldots, v_i\}$, with $j < n$.

Let $U = \mathrm{span}\,\mathcal{V}$. So V/U is a non-trivial $\mathbb{C}$ vector space, and U is an invariant subspace under T. So we have an eigenvector $U + v_{j+1}$ of the operator T/U, so $(T/U)(U + v_{j+1}) = U + \lambda v_{j+1}$.

So if we let v_{j+1} be a coset representative of $U + v_{j+1}$ then we have $T(v_{j+1}) = a_1 v_1 + a_2 v_2 + \ldots a_j v_j + \lambda v_{j+1} \in \mathrm{span}\{v_1, v_2, \ldots v_{j+1}\}$ and so we have $\mathcal{V} \cup \{v_{j+1}\}$ as a set of $j+1$ vectors that satisfies the property. So by induction we have the desired basis. $\qquad\square$

Theorem 10. Let V be a finite dimensional vector space and $T \in \mathcal{L}(V)$ have an upper triangular matrix with respect to some basis $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$. Then T is bijective iff all the entries on the diagonal of $\mathcal{M}(T, \mathcal{V})$ are non-zero.

Proof. First we assume the entries on the diagonal are non-zero. Let them be $\lambda_1, \lambda_2, \ldots, \lambda_n$ from left to right. We claim that $v_1, v_2, \ldots, v_n \in T(V)$. We proceed by induction, assume $v_1, v_2, \ldots, v_j \in T(V)$.

We have that $T(\frac{v_{j+1}}{\lambda_{j+1}}) = v_{j+1} + b_j v_j + \ldots + b_2 v_2 + b_1 v_1$ for some scalars $b_1, b_2, \ldots, b_j$, so we have $v_{j+1} \in T(V)$. So we have $v_1, v_2, \ldots, v_n \in T(V)$ as claimed, and so $V \subset T(V)$ as $\mathcal{V}$ spans V, so T is surjective. As V is finite dimensional, this means that T is bijective.

Now assume that T is bijective. Say that an entry $\lambda_j = 0$, then $T(v_1), T(v_2), \ldots, T(v_j) \in \mathrm{span}\{v_1, v_2, \ldots, v_j\}$. As the span has dimension j, the images must be linearly dependent, contradicting that bijective linear maps preserve linear independence. Hence all the entries on the diagonal are non-zero. $\qquad\square$

Theorem 11. Let V be a finite dimensional vector space and $T \in \mathcal{L}(V)$ have an upper triangular matrix with respect to some basis $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$. Then the eigenvalues of T are precisely the entries on the diagonals of $\mathcal{M}(T, \mathcal{V})$.

Proof. As V is finite dimensional, λ is an eigenvalue precisely when $T - \lambda I$ is non-bijective, which is precisely when λ is one of the entries on the diagonal. $\square$

4 Inner Product Spaces

4.1 Inner Products

A key aspect of the structure of the classical examples of vector spaces such as Euclidean space which we have not generalised is the notion of the dot product, whose key aspect is that it allows fora notion of angles between vectors, and most importantly the idea that vectors are orthogonal. We may generalise this idea into the inner product of a vector space, however we will only do so for $\mathbb{R}$ and $\mathbb{C}$ vector spaces.

Definition 15 (Inner Product)**.** Let V be an $\mathbb{R}$ or $\mathbb{C}$ vector space. Then we say a mapping $< .,. >: V^2 \to k$ is an inner product if it satisfies all of the following

1. $\langle v, v \rangle \geq 0$ for all $v \in V$ (non-negativity)

2. $\langle v, v \rangle = 0$ iff $v = 0$ (definiteness)

3. $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ for all $u, v, w \in V$ (additivity of the first argument)

4. $\langle \lambda u, v \rangle = \lambda \langle u, v \rangle$ for all $\lambda \in k$ and $u, v \in V$ (homogeneity of the first argument)

5. $\langle u, v \rangle = \overline{\langle v, u \rangle}$ (conjugate symmetry)

We call V equipped with an inner product an inner product space.

Definition 16 (Orthogonality)**.** Let V be an inner product space. We say $u, v \in V$ are orthogonal if $\langle u, v \rangle = 0$.

Definition 17 (Norm)**.** Let V be an inner product space. Then we define the norm of a vector $v \in V$ as

$$\|v\| = \sqrt{\langle v, v \rangle}$$

From these axioms we may trivially show that there is conjugate homogeneity in the second argument, additivity in the second argument and that the inner prodcut of any vector and 0 is 0 in either order.

The usual dot product we are familiar with over k^n is referred to as the Euclidean inner product. It is trivially verified that this is indeed an inner product. This may be thought of as the inne r product which arises when we require that the elements of the standard basis have norm 1 and are pairwise orthogonal.

Theorem 12 (Pythagorean Theorem)**.** If $u, v \in V$ are orthogonal, then

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2$$

Proof.

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

$\square$

It is often useful to be able to express a vector v as a sum of a scaled u and a vector orthogonal to u, as in Figure 3.
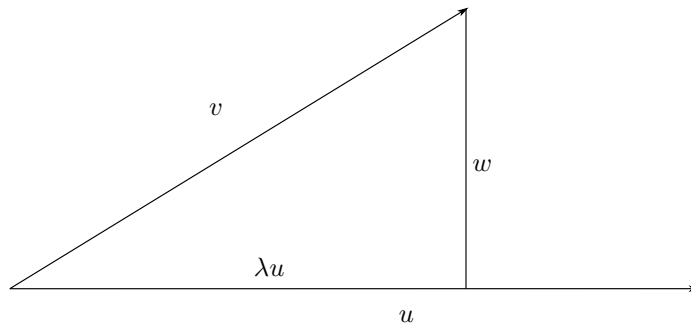
Figure 1: Illustration of v being broken down into orthogonal components.

We call this expression an orthogonal decomposition of v with respect to u. A trivial manipulaton yields that $v = \lambda u + (v - \lambda u)$, where $\lambda = \frac{\langle v, u \rangle}{\|u\|^2}$ is the desired sum. Of course this fails exactly when $u = 0$.

Theorem 13 (Cauchy-Scwarz Inequality). Let $u, v \in V$. Then $\|u\|\|v\| \geq |\langle u, v \rangle|$, with equality precisely when u and v are scalar multiplies of eachother.

Proof. If $u = 0$ then the result is trivial. Otherwise we may orthogonally decompose v to yield

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

16

We may now apply Pythagoras' theorem to yield

$$\|v\|^2 = \frac{|\langle v, u \rangle|^2}{\|u\|^4}\|u\|^2 + \|v - \frac{\langle v, u \rangle}{\|u\|^2}\|^2 \geq \frac{|\langle v, u \rangle|^2}{\|u\|^2}$$

Rearrangement yields the desired inequality. Equality arises precisely when the orthogonal component is zero, that is when u and v are scalar multiplies of eachother. $\square$

Theorem 14 (Triangle Inequality). Let $u, v \in V$. Then $\|u + v\| \leq \|u\| + \|v\|$, with equality precisely when u and v are nonnegative multiples of eachother.

Proof.

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

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

$$\leq \|u\|^2 + \|v\|^2 + 2|\langle u, v \rangle|$$

$$\leq \|u\|^2 + \|v\|^2 + 2\|u\|\|v\|$$

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

Equality occurs only when u and v are scalar multplies of eachother and $\Re\langle u, v \rangle = |\langle u, v \rangle|$, which is precisely when u and v are nonnegative multiples of eachother. $\square$

4.2   Orthonormal Bases

Definition 18 (Orthonormal Set). Let V be an inner product space, and $\mathcal{V}$ a finite set of vectors in V. Then we say $\mathcal{V}$ is orthonormal if all of its elements have norm 1 and are pairwise orthogonal.

Orthonormal sets behave quite nicely. This is due mainly to the fact that the norm of a linear combination is the sum of squares of the absolute values

of the scalars by repeated applications of Pythagoras and that we may "extract" any scalar in a linear combination by taking the inner product with the corresponding vector.

Theorem 15. Any orthonormal set, $\mathcal{V}$, is linearly independent.

Proof. Let $\mathcal{V} = \{e_1, e_2, \ldots, e_n\}$. Say we have a 0 linear combination

$$a_1 e_1 + a_2 e_2 + \ldots + a_n e_n = 0$$

Then taking the square of the norms of both sides yields

$$|a_1|^2 + |a_2|^2 + \ldots + |a_n|^2 = 0$$

This is a sum of real squares, hence every scalar is 0 and so the combination is trivial. $\qquad\square$

Theorem 16. Let $\mathcal{V} = \{e_1, e_2, \ldots, e_n\}$ be an orthonormal basis of V. Then for any $v \in V$, we have that

$$v = \langle v, e_1 \rangle e_1 + \langle v, e_2 \rangle e_2 + \ldots + \langle v, e_n \rangle e_n$$

Proof. $\mathcal{V}$ is a basis, so we for for some unique $a_1, a_2, \ldots, a_n \in k$ that

$$v = a_1 e_1 + a_2 e_2 + \ldots + a_n e_n$$

Taking the left inner product with e_j gives $\langle v, e_j \rangle = a_j$. $\qquad\square$

Theorem 17 (Gram-Schmidt Procedure). Let $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ be a linearly independent set of vectors in V. Define the vectors $\mathcal{V}' = \{e_1, e_2, \ldots, e_n\}$
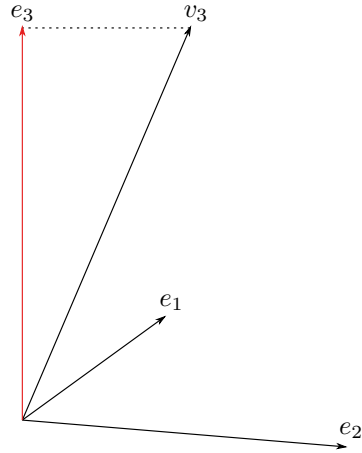
Figure 2: Illustration of how the Gram-Schmidt procedure removes the components of a vector parralel to the previous ones to form an orthogonal vector.

inductively by

$$e_{j+1} = \frac{v_{j+1} - (\langle v_{j+1}, e_1 \rangle e_1 + \langle v_{j+1}, e_2 \rangle e_2 + \ldots + \langle v_{j+1}, e_j \rangle e_j)}{\|v_{j+1} - (\langle v_{j+1}, e_1 \rangle e_1 + \langle v_{j+1}, e_2 \rangle e_2 + \ldots + \langle v_{j+1}, e_j \rangle e_j)\|}$$

Then $\mathcal{V}'$ is orthonormal and span $\mathcal{V} = \text{span}\,\mathcal{V}'$.

Proof. We will show that for every $j \leq n$, that

$$\text{span}\{v_1, v_2, \ldots, v_j\} = \text{span}\{e_1, e_2, \ldots, e_j\}$$

and $\{e_1, e_2, \ldots, e_j\}$ is orthonormal. We proceed by induction. First we show

that $\{e_1, e_2, \ldots, e_{j+1}\}$ is orthonormal. We have

$$\langle e_{j+1}, e_i \rangle = \lambda \langle v_{j+1} - (\langle v_{j+1}, e_1 \rangle e_1 + \langle v_{j+1}, e_2 \rangle e_2 + \ldots + \langle v_{j+1}, e_j \rangle e_j), e_i \rangle$$

$$= \lambda (\langle v_{j+1}, e_i \rangle - \langle v_{j+1}, e_i \rangle \langle e_i, e_i \rangle)$$

$$= 0$$

Hence e_{j+1} is orthogonal to the other vectors, and by induction the remaining vectors are all pairwise orthogonal, hence the entire set is pairwise orthogonal. The vectors are all normal by construction, so the set is orthonormal.

By induction, we have $\text{span}\{v_1, v_2, \ldots, v_j\} = \text{span}\{e_1, e_2, \ldots, e_j\}$, and hence $(\langle v_{j+1}, e_1 \rangle e_1 + \langle v_{j+1}, e_2 \rangle e_2 + \ldots + \langle v_{j+1}, e_j \rangle e_j) \in \text{span}\{v_1, v_2, \ldots, v_j\}$, so we have $e_{j+1} \in \text{span}\{v_1, v_2, \ldots, v_{j+1}\}$. Hence $\text{span}\{e_1, e_2, \ldots, e_{j+1}\} \subset \text{span}\{v_1, v_2, \ldots, v_{j+1}\}$. $\{e_1, e_2, \ldots, e_{j+1}\}$ is orthonormal and hence linearly independent, so $\text{span}\{e_1, e_2, \ldots, e_{j+1}\} = \text{span}\{v_1, v_2, \ldots, v_{j+1}\}$ as they have equal dimensions and one is contained in the other. $\qquad\square$

An extremely useful and immediate consequence of this procedure is that it means we may produce an orthonormal basis from a basis, and so every finite dimensional inner product space has an orthonormal basis.

Indeed, as the Gram-Schmidt procedure doesn't change the spans of any of the subsets of the form $\{v_1, v_2, \ldots, v_j\}$ this means that if a matrix is upper-triangular with respect to a basis, then the it will also be upper triangular with respect to the orthonormal basis produced by Gram Schmidt.

This means that all operators on non-trivial $\mathbb{C}$ vector spaces have an orthonormal basis with respect to which their matrix is upper-triangular. This result is known as Schurs' Theorem.

Theorem 18 (Riesz Representation Theorem)**.** Let φ be a linear functional on a finite dimensional inner product space V. Then for some unique $u \in V$, we

have

$$\varphi(v) = \langle v, u \rangle$$

Proof. First we show existence. As V is a finite dimensional inner product space, by the Gram Schmidt procedure we have some orthnormal basis $\mathcal{V} = \{e_1, e_2, \ldots, e_n\}$. So we have

$$\varphi(v) = \varphi(\langle v, e_1 \rangle e_1 + \langle v, e_2 \rangle e_2 + \ldots + \langle v, e_n \rangle e_n)$$
$$= \langle v, e_1 \rangle \varphi(e_1) + \langle v, e_2 \rangle \varphi(e_2) + \ldots + \langle v, e_n \rangle \varphi(e_n)$$
$$= \langle v, \overline{\varphi(e_1)} e_1 + \overline{\varphi(e_2)} e_2 + \ldots + \overline{\varphi(e_n)} e_n \rangle$$

Hence $u = \overline{\varphi(e_1)} e_1 + \overline{\varphi(e_2)} e_2 + \ldots + \overline{\varphi(e_n)} e_n$ is a valid choice.

Say that u_1 and u_2 satisfy the criteria, so $\langle v, u_1 \rangle = \langle v, u_2 \rangle$ for all $v \in V$. This yields $\langle v, u_1 - u_2 \rangle = 0$ for all $v \in V$, and letting $v = u_1 - u_2$ shows that $u_1 - u_2 = 0$, or $u_1 = u_2$, so the choice of u is unique. $\qquad \square$

4.3 Orthogonal Complements

Definition 19 (Orthogonal Complement). Let V be an inner product space and $U \subset V$ be a subset of V. Then we define the orthogonal complement of U, $U^{\perp}$ as the set of all vectors in V orthogonal to every element of U.

We have defined the orthogonal complement for an arbitrary subset, however due to the linearity and homogeneity of the inner product, $U^{\perp} = \text{span}(U)^{\perp}$. So whilst we will be mainly interested in the orthogonal complement of a subspace, in practice it is easiest to compute it by considering the orthogonal complement of a basis of the subspace.

When we use our geometric intuition, we see that it is plausible that V may be expressed as the direct sum of a subspace and its orthogonal complement.

In the finite dimensional case this turns out to be true.

Theorem 19. Let V be an inner product space and $U \subset V$ a finite-dimensional subspace of V. Then $V = U \oplus U^{\perp}$.

Proof. Let $\mathcal{U} = \{e_1, e_2, \ldots, e_n\}$ be an orthonormal basis of U. Consider an arbitrary $v \in V$, we may write v as $v = w + (v - w)$, where

$$w = \langle v, e_1 \rangle e_1 + \langle v, e_2 \rangle e_2 + \ldots + \langle v, e_n \rangle e_n$$

Intuitively we may regard w as the components of v parralel to U, so then in a manner similar to the Gram-Schmidt procedure, we have $v - w \in U^{\perp}$, so $v \in U + U^{\perp}$, and hence $V = U + U^{\perp}$.

The intersection of U and $U^{\perp}$ must be $\{0\}$, as 0 is the only vector orthogonal to itself by definiteness, and so the sum is direct. $\square$

Theorem 20. Let V be an inner product space. Then over the finite-dimensional subspaces of V, taking the orthogonal complement is an involution.

Proof. Let U be an arbitrary finite-dimensional subspace of V. Say that $u \in U$, then for all $w \in U^{\perp}$ we have that $u \perp w$, so $u \in (U^{\perp})^{\perp}$, and so $U \subset (U^{\perp})^{\perp}$.

As U is finite-dimensional, we have $(U^{\perp})^{\perp} \subset V = U \oplus U^{\perp}$, so for any $v \in (U^{\perp})^{\perp}$, we have $u \in U$ and $w \in U^{\perp}$ such that $v = u + w$.

Taking the left inner product with w yields $\langle w, w \rangle = 0$, so $w = 0$, and so $v = u \in U$. Hence $(U^{\perp})^{\perp} \subset U$. So we have $U = (U^{\perp})^{\perp}$. $\square$

In light of the idea of orthogonal complements, we may define the notion of the projection of a vector onto a subspace.

Definition 20 (Orthogonal Projection)**.** Let U be a finite dimensional subspace of V, then we define the orthogonal projection operator P_U of v as u, where $v = u + w$ is the unique sum of v as vectors from U and $U^{\perp}$.
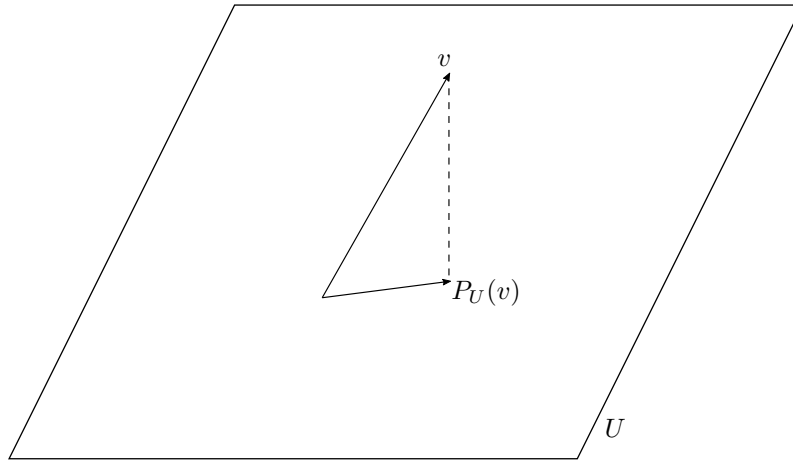
Figure 3: Illustration of orthogonal projection in $\mathbb{R}^3$ onto a plane

We have already used this idea in the Gram-Schmidt procedure, where the vector we subtract from v_{j+1} is really just its orthogonal projection onto the subspace spanned by $\{e_1, e_2, \ldots, e_j\}$. However it is more elegant to think in terms of orthogonal projections, as its properties reflect most of the intuitive notions of vectors we have as arrows in space.

One striking illustration of the power of being able to reason abstractly from geometric intuitions is that the orthogonal projection of v onto a finite dimensional subspace U is the element of U with minimal distance to v. Whilst this is trivial thinking of vectors as points in space, this result can apply to arbitrary vector spaces, so for example it allows us to find the polynomial of given degree that minimises the distance to a specific continuous function (we can use the integral of the difference as a measure of "distance" in this case, and so define the inner product appropriately)

Theorem 21. Let U be a finite-dimensional subspace of V, and $v \in V$. Then for all $u \in U$ we have

$$\|v - P_U(v)\| \le \|v - u\|$$

With equality precisely when $u = P_U(v)$.

Proof.

$$\|v - P_U(v)\|^2 \le \|v - P_U(v)\|^2 + \|u - P_U(v)\|^2$$
$$= \|v - u\|^2$$

There is equality precisely when $\|u - P_U(v)\|^2 = 0$, which is exactly when $u = P_U(v)$. $\qquad\square$

5 Operators on Inner Product Spaces

5.1 Self-Adjoint and Normal Operators

Definition 21 (Adjoint). Let T be a linear map between finite dimensional inner product spaces V and W, then we define the adjoint $T^* : W \to V$ as the unique mapping satisfying

$$\langle T(v), w \rangle = \langle v, T^*(w) \rangle$$

For all $v \in V$ and $w \in W$.

This definition makes sense due to Riesz representation theorem. Indeed it is trivially shown that $T*$ is actually a linear map, and various other properties such as $(S + T)^* = S^* + T^*$, $(T^*)^*$, etc, whose proofs we omit due to their triviality.

Theorem 22. Let T be a linear map between finite dimensional inner product spaces V and W. Then we have $\ker T^* = (T(V))^\perp$.

Proof.

$$w \in \ker T^*$$

$$\iff T^*(w) = 0$$

$$\iff \forall v \in V(\langle v, T^*(w) \rangle = 0)$$

$$\iff \forall v \in V(\langle T(v), w \rangle = 0)$$

$$\iff w \in (T(V))^\perp.$$

$\square$

Whilst our definition of the adjoint seems rather abstract and complicated, it has a very simple and concrete meaning with regards to the matrix representation, at least when we are using orthonormal bases.

Theorem 23 (Matrix of the Adjoint)**.** Let V and W be finite-dimensional inner product spaces, with $\mathcal{V} = \{e_1, e_2, \ldots, e_n\}$ and $\mathcal{W} = \{f_1, f_2, \ldots, f_m\}$ orthonormal bases of V and W respectively. If $T \in \mathcal{L}(V, W)$ is a linear map, then we have that $\mathcal{M}(T, \mathcal{V}, \mathcal{W})$ and $\mathcal{M}(T*, \mathcal{W}, \mathcal{V})$ are each others conjugate transpose.

Proof. Since $\mathcal{W}$ is orthonormal, we have that

$$T(e_j) = \langle T(e_j), f_1 \rangle f_1 + \langle T(e_j), f_2 \rangle f_2 + \ldots + \langle T(e_j), f_m \rangle f_m$$

$$= \langle e_j, T^*(f_1) \rangle f_1 + \langle e_j, T^*(f_2) \rangle f_2 + \ldots + \langle e_j, T^*(f_m) \rangle f_m$$

$$= \overline{\langle T^*(f_1), e_j \rangle} f_1 + \overline{\langle T^*(f_2), e_j \rangle} f_2 + \ldots + \overline{\langle T^*(f_m), e_j \rangle} f_m$$

So we have that $\mathcal{M}(T)_{ij} = \overline{\langle T^*(f_i), e_j \rangle}$

As $\mathcal{V}$ is orthonormal, we have that

$$T^*(f_j) = \langle T^*(f_j), e_1 \rangle e_1 + \langle T^*(f_j), e_2 \rangle e_2 + \ldots + \langle T^*(f_j), e_n \rangle e_n$$

So we have that $\mathcal{M}(T^*)_{ij} = \langle T^*(f_j), e_i \rangle$

Hence $\mathcal{M}(T)$ and $\mathcal{M}(T^*)$ are the conjugate transposes of each other. □

From here we restrict our attention to operators, as it is in the context of operators that the domain and range coincide. So a mapping may equal its own transpose, and we may consider the product of an operator with its transpose. This allows us to consider the notion of normal and self-adjoint operators.

Definition 22 (Self-Adjoint). Let V be a finite-dimensional inner product space. We say the operator $T \in \mathcal{L}(V)$ is self-adjoint if $T = T^*$.

Definition 23 (Normal Operator). Let V be a finite-dimensional inner product space. We say the operator $T \in \mathcal{L}(V)$ is normal if $TT^* = T^*T$.

Of course all self-adjoint operators are normal, so normality is a weaker condition. Whilst the criteria for normality and self-adjointedness may seem arbitrary and of little interest, they turn out to lead to extremely nice properties.

Theorem 24. Every eigenvalue of a self-adjoint operator is real.

Proof. Let λ be an eigenvalue of the self adjoint operator $T \in \mathcal{L}(V)$. Then for some non-zero $v \in V$, we have $T(v) = \lambda v$. By self-adjointedness, we have

$$\langle T(v), v \rangle = \langle v, T(v) \rangle$$
$$\langle \lambda v, v \rangle = \langle v, \lambda v \rangle$$
$$\lambda = \overline{\lambda}$$

Hence λ is real. □

This result is of course only of significance over a $\mathbb{C}$ vector space. As is often the case, there is an alternative proof of the result by consideration of the matrix of the operator.

Since the result is trivial in $\mathbb{R}$ spaces, we assume V is a complex inner product space, and so we have a basis in which the matrix is upper triangular. In this case, due to self-adjointedness, the matrix is its own conjugate transpose, meaning all the entries on the diagonals, that is, the eigenvalues, are their own complex conjugates.

Theorem 25. Let V be a complex inner product space. Then $T = 0$ is the only operator such that every vector is orthogonal to its own image under T.

Proof. Let $T \in \mathcal{L}(V)$ be such an operator. Let u and v be arbitrary vectors in V. Then we have

$$\langle T(u), v \rangle = \frac{\langle T(u+v), u+v \rangle - \langle T(u-v), u-v \rangle + i\langle T(u+iv), u+iv \rangle - i\langle T(u-iv), u-iv \rangle}{4}$$

$$= 0$$

So if we let $v = T(u)$, then $T(u) = 0$ for all $u \in V$, hence $T = 0$. $\qquad\square$

The hypothesis that the inner product space is complex is essential, as there are non-zero operators over real inner product spaces which satisfy this propery. For example an order 4 turn about the origin in $\mathbb{R}^2$.

Theorem 26. Let V be a complex inner product space. Then $T \in \mathcal{L}(V)$ is self-adjoint iff $\langle T(v), v \rangle \in \mathbb{R}$ for all $v \in V$.

Proof.

$$\forall v \in V : \langle T(v), v \rangle \in \mathbb{R}$$

$$\iff \forall v \in V : \langle T(v), v \rangle - \overline{\langle T(v), v \rangle} = 0$$

$$\iff \forall v :\in V \langle (T - T^*)v, v \rangle = 0$$

As V is a complex inner product space, this occurs precisely when $T - T^* = 0$, that is, when T is self-adjoint. $\square$

Theorem 27. Let V be an inner product space. Then $T = 0$ is the only self-adjoint operator such that every vector is orthogonal to its own image under T.

Proof. We have already shown this for complex spaces, so we assume that V is a real inner product space. Let u and v be arbitrary vectors in V. Then we have that

$$\langle T(u), v \rangle = \frac{\langle T(u + v), u + v \rangle - \langle T(u - v), u - v \rangle}{4}$$

$$= 0$$

So if we let $v = T(u)$, then $T(u) = 0$ for all $u \in V$, hence $T = 0$. $\square$

Theorem 28. Let V be an inner product space, and $T \in \mathcal{L}(V)$ an operator on V. Then T is normal iff $\|T(v)\| = \|T^*(v)\|$ for all $v \in V$.

Proof.

$$\forall v \in V : \|T(v)\| = \|T^*(v)\|$$

$$\iff \forall v \in V : \langle T(v), T(v) \rangle = \langle T^*(v), T^*(v) \rangle$$

$$\iff \forall v \in V : \langle (T^*T - TT^*)v, v \rangle = 0$$

28

The operator $TT^* - T^*T$ is self-adjoint, hence this is true precisely when $TT^* - T^*T = 0$, that is, when T is normal. $\qquad\square$

Theorem 29. Let $T \in \mathcal{L}(V)$ be a normal operator, and $v \in V$ an eigenvector of T with eigenvalue λ. Then v is also an eigenvector of T, with eigenvalue $\overline{\lambda}$.

Proof. By the normality of T, we have for any $v \in V$, that $\|(T - \lambda I)v\| = \|(T^* - \overline{\lambda}I)v|$. So $T - \lambda I$ is injective precisely when $T^* - \overline{\lambda}I$ is. $\qquad\square$

This is actually just a stronger statement of Theorem 24, as in the case of a self-adjoint operator, it is its own adjoint and so every eigenvalue is its own complex conjugate.

Theorem 30. Let $T \in \mathcal{L}(V)$ be a normal operator. Then the eigenvectors of T corresponding to distinct eigenvalues are orthogonal.

Proof. Let v_1 and v_2 be eigenvectors with distinct corresponding eigenvalues λ_1 and λ_2. T is normal, so $T^*(v_2) = \overline{\lambda}_2 v_2$.

$$
\begin{aligned}
(\lambda_1 - \lambda_2)\langle v_1, v_2 \rangle &= \langle \lambda_1 v_1, v_2 \rangle - \langle v_1, \overline{\lambda}_2 v_2 \rangle \\
&= \langle T(v_1), v_2 \rangle - \langle v_1, T^*(v_2) \rangle \\
&= 0
\end{aligned}
$$

As λ_1 and λ_2 are distinct, we have that $\langle v_1, v_2 \rangle = 0$, hence the eigenvectors are orthogonal. $\qquad\square$

5.2 The Spectral Theorem

Theorem 31 (Complex Spectral Theorem)**.** Let V be a complex inner product space and $T \in \mathcal{L}(V)$ an operator. Then T is normal precisely when there is an orthonormal basis of V consisting of eigenvectors of V.

Proof. First assume there is an orthonormal basis of eigenvectors. Then the matrices of T and T^* are diagonal with respect to that basis, and so they commute.

Now assume that T is normal. By Schur's theorem we have an orthonormal basis $\mathcal{V} = \{e_1, e_2, \ldots, e_n\}$ with respect to which $\mathcal{M}(T)$ is upper-triangular. We claim that $\mathcal{M}(T)$ is actually diagonal. We show non-diagonal entry on every row is 0. Assume by induction that this holds for the first j rows. We have, by the normality of T, that

$$\|T(e_j)\|^2 = \|T^*(e_j)\|^2$$

$$\|\mathcal{M}_{jj}\|^2 = \|\mathcal{M}_{jj}\|^2 + \|\mathcal{M}_{j(j+1)}\|^2 + \ldots + \|\mathcal{M}_{jn}\|^2$$

$$`0 = \|\mathcal{M}_{j(j+1)}\|^2 + \|\mathcal{M}_{j(j+2)}\|^2 + \ldots + \|\mathcal{M}_{jn}\|^2$$

Hence all the non-diagonal entries on row $j+1$ are zero. So by induction the matrix is diagonal, and so the orthonormal basis consists entirely of eigenvectors.

$\square$

Theorem 32. Let $T \in \mathcal{L}(V)$ be a self-adjoint operator, and b and c be reals such that $b^2 - 4c < 0$, then the operator $T^2 + bT + cI$ is bijective.

Proof. Consider an arbitrary non-zero $v \in V$. Then we have

$$\langle (T^2 + bT + cI)v, v \rangle$$

$$= \langle T(v), T(v) \rangle + b\langle T(v), v \rangle + c\langle v, v \rangle$$

$$\geq \|T(v)\|^2 - b\|T(v)\|\|v\| + c\|v\|^2$$

$$= (\|T(v)\| - \frac{b}{2}\|v\|)^2 + (c - \frac{b^2}{4})\|v\|^2$$

$$> 0$$

Hence the kernel of $T^2 + bT + cI$ is trivial and so it is bijective. $\square$

Theorem 33. Let V be a finite inner product space, and $T \in \mathcal{L}(V)$ be a self-adjoint operator. Then T has an eigenvalue.

Proof. If V is a complex inner product space, then we are done, as all operators on finite dimensional complex inner product spaces have an eigenvalue. So assume that V is a real inner product space, and let its dimension be n. Let v be a non-zero vector in V.

Then the set $\mathcal{V} = \{v, T(v), T^2(v), \ldots, T^n(v)\}$ is linearly dependent, so we have a choice of scalars $a_0, a_1, \ldots, a_n$ not all equal to 0 such that the corresponding linear combination of $\mathcal{V}$ is 0. Let the last non-zero scalar index be j, so $j \geq 1$ as $v \neq 0$. So we have

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

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

By the fundemental theorem of algebra, we may factorise the above expression into a product of linear and irreducible quadratic terms. As we have shown in Theorem 32, all of the irreducible quadratic factors are bijective and hence we may apply their inverses to yield

$$(T - \lambda_1 I)(T - \lambda_2 I)\ldots(T - \lambda_m I)v = 0$$

For some scalars $\lambda_1, \lambda_2, \ldots, \lambda_m$. Hence at least one of the factors is non-injective, and so one of λ_i is an eigenvalue. $\square$

Theorem 34. Let $T \in \mathcal{L}(V)$ be a self-adjoint operator, and U an invariant subspace under T. Then $U^\perp$ is also invariant under T.

Proof. Let w be an arbitrary vector in $U^\perp$. Then for all $u \in U$, we have $\langle w, T(u) \rangle = 0$, as U is an invariant subspace.

So we have by self-adjointedness that for all $u \in U$ that $\langle T(w), u \rangle = 0$, which means $T(w) \in U^\perp$, so $U^\perp$ is invariant. $\square$

Theorem 35 (Real Spectral Theorem)**.** Let V be a real inner product space, and $T \in \mathcal{L}(V)$ an operator over V. Then T is self-adjoint precisely when V has an orthonormal basis of eigenvectors of T.

Proof. First assume that there is an orthonormal basis of eigenvectors of V. Then the matrix of T is diagonal with respect to this basis, and as all the entries are real, the matrix is its own conjugate transpose, and so T is self-adjoint.

Now assume T is self-adjoint. We proceed by induction on the dimension of V. As T is self-adjoint, it has an eigenvector v. So $U = \text{span}\{v\}$ is an invariant subspace, and so, by self-adjointedness, so is $U^\perp$.

By induction, we have an orthonormal basis of eigenvectors, $\mathcal{V}$, of $U^\perp$. Hence $\mathcal{V} \cup \{\frac{v}{\|v\|}\}$ is the desired orthonormal basis of eigenvectors. $\square$

5.3 Positive Operators and Isometries

Definition 24 (Positive Operators)**.** We say an operator $T \in \mathcal{L}(V)$ is positive if T is self-adjoint and $\langle T(v), v \rangle \geq 0$ for all $v \in V$.

Definition 25 (Square Root)**.** We say an operator $R \in \mathcal{L}(V)$ is a square root of the operator $T \in \mathcal{L}(V)$ if $R^2 = T$.

Theorem 36. Let $T \in \mathcal{L}(V)$ be an operator. Then the following are all equivalent:

1. T is positive

2. T is self-adjoint and all of its eigenvalues are nonnegative

3. T has a positive square root

4. T has a self-adjoint square root

5. We have an operator $R \in \mathcal{L}(V)$ such that $T = R^*R$.

Proof. First assume (1), then by definition T must be self-adjoint. Say λ is an eigenvalue of T, then for some non-zero $v \in V$, we have $T(v) = \lambda v$, so by positivity we have $\lambda \|v\|^2 \geq 0$, so $\lambda \geq 0$. Now assume (2), and so let an orthonormal basis of eigenvalues be $\mathcal{V} = \{e_1, e_2, \ldots, e_n\}$ and the corresponding eigenvalues be $\lambda_1, \lambda_2, \ldots, \lambda_n$ which are all nonnegative. Hence we have that

$$\langle T(a_1 e_1 + a_2 e_2 + \ldots + a_n e_n), a_1 e_1 + a_2 e_2 + \ldots + a_n e_n \rangle$$
$$= \sum_{i=1}^{n} \lambda_i a_i \sum_{j=1}^{n} \overline{a_j} \langle e_i, e_j \rangle$$
$$= \sum_{i=1}^{n} \lambda_i a_i \overline{a_i}$$
$$= \sum_{i=1}^{n} \lambda_i |a_i|^2$$
$$\geq 0$$

Now assume (2). Then we have an orthonormal basis of eigenvectors of T, $\mathcal{V} = \{e_1, e_2, \ldots, e_n\}$, with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then the operator R defined by $R(e_j) = \sqrt{\lambda_j} e_j$ is the desired positive square root. Now if we assume (3), then T must be self-adjoint, and its eigenvalues must be the squares of the nonnegative eigenvalues of the positive square root.

Now assume (3), then (4) follows trivially as self-adjointness is a weaker condition than positivity. Now assume (4), so T has a self-adjoint square root, R. By the spectral theorem, we have an orthonormal basis of eigenvectors of T, $\mathcal{V} = \{e_1, e_2, \ldots, e_n\}$. Let the eigenvalues be $\lambda_1, \lambda_2, \ldots, \lambda_n$, which must be real as T is self-adjoint. So $T(v_j) = \lambda_j^2 v_j$, and hence $R'(v_j) = |\lambda_j| v_j$ is the desired positive square root.

Now assume (4), so there is a self-adjoint square root $R^2 = T$, so $R^*R =$

33

T. Assume (5), then T is self-adjoint, so we have an orthonormal basis of eigenvectors of T, $\mathcal{V} = \{e_1, e_2, \ldots, e_n\}$, with corresponding real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. So we have

$$R^* R(v_j) = \lambda_j v_j$$

$$\langle R^* R(v_j), v_j \rangle = \lambda_j \langle v_j, v_j \rangle$$

$$\|R(v_j)\|^2 = \lambda_j \|v_j\|^2$$

So each λ_j is nonnegative, and so T has a self-adjoint square root. $\qquad\square$

Theorem 37. Let V be a finite dimensional inner product space and $T \in \mathcal{L}(V)$ a positive operator. Then T has a unique positive square root.

Proof. We already have that T must have a positive square root, so we need only show uniqueness. Let R be a positive square root of T, and v an arbitrary eigenvector of T with eigenvalue λ.

By the positivity of R, we have an orthonormal basis $\mathcal{V} = \{e_1, e_2, \ldots, e_n\}$ of V consisting of eigenvectors of R. Let the corresponding eigenvalues by $\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_n}$, so $R(e_j) = \sqrt{\lambda_j} e_j$.

As $\mathcal{V}$ is a basis, we have for some choice of scalars a_i, that

$$v = a_1 e_1 + a_2 e_2 + \ldots + a_n e_n$$

$$R^2(v) = a_1 \lambda_1 e_1 + a_2 \lambda_2 e_2 + \ldots + a_n \lambda_n e_n$$

$$= \lambda v$$

$$a_1(\lambda_1 - \lambda)e_1 + a_2(\lambda_2 - \lambda)e_2 + \ldots + a_n(\lambda_n - \lambda)e_n = 0$$

So by the linear independence of $\mathcal{V}$, we have that for every j, $a_j = 0$ or $\lambda_j = \lambda$.

Hence we have

$$v = \sum_{\lambda_j = \lambda} a_j e_j$$

$$R(v) = \sqrt{\lambda} v$$

Since the square root is determined on the eigenvectors of T and by self-adjointedness, there is an orthonormal basis of V consisting of eigenvectors T, hence R is unique. $\qquad\square$

Definition 26 (Isometry). Let V be an inner product space, and $T \in \mathcal{L}(V)$ be an operator on V. Then T is an isometry if it preserves norms.

Theorem 38. Let $S \in \mathcal{L}(V)$ be an operator in V. Then the following are equivalent:

S is an isometry

$\langle S(u), S(v) \rangle = \langle u, v \rangle$ for all $u, v \in V$.

$\{S(e_1), S(e_2), \ldots, S(e_n)\}$ is orthnormal for all orthonormal $\{e_1, e_2, \ldots, e_n\}$

$\{S(e_1), S(e_2), \ldots, S(e_n)\}$ is orthnormal for some orthonormal orthonormal $\{e_1, e_2, \ldots, e_n\}$

$S^*S = SS^* = I$.

S^* is an isometry.

Proof. First assume (1). If V is a real inner product space then

$$\langle S(u), S(v) \rangle = \frac{\|u+v\|^2 - \|u-v\|^2}{4}.$$

If V is a complex inner product space then

$$\langle S(u), S(v) \rangle = \frac{\|u+v\|^2 - \|u-v\|^2 + i\|u+iv\|^2 - i\|u-iv\|^2}{4}.$$

In either case, we see the inner product may be computed from norms, and hence $\langle S(u), S(v)\rangle = \langle u, v\rangle$ for all $u, v \in V$. Of course (2) implies (1).

Now assume (2), then (3) trivially holds as orthnormality is based entirely on the inner product. Now assume (3), and let $\{e_1, e_2, \ldots, e_n\}$ be an orthonormal basis, so $\{S(e_1), S(e_2), \ldots, S(e_n)\}$ is as well. Consider an arbitrary $v \in V$, then

$$\|S(v)\|^2 = \|S(\langle v, e_1\rangle e_1 + \langle v, e_2\rangle e_2 + \ldots + \langle v, e_n\rangle e_n\|^2$$
$$= \|\langle v, e_1\rangle\|^2 + \|\langle v, e_2\rangle\|^2 + \ldots + \|\langle v, e_n\rangle\|^2$$
$$= \|v\|^2$$

Hence S preserves norms, and so it preserves the inner product, hence we have (2).

(3) and (4) are trivially equivalent.

Assume (4), and let $\{e_1, e_2, \ldots, e_n\}$ be an orthonormal basis, so $\{S(e_1), S(e_2), \ldots, S(e_n)\}$ is as well. Consider an arbitrary $v \in V$, then we have

$$v = \langle v, S(e_1)\rangle S(e_1) + \langle v, S(e_2)\rangle S(e_2) + \ldots + \langle v, S(e_n)\rangle S(e_n)$$
$$= S(\langle v, S(e_1)\rangle e_1 + \langle v, S(e_2)\rangle e_2 + \ldots + \langle v, S(e_n)\rangle e_n)$$
$$= S(\langle S^*(v), e_1\rangle e_1 + \langle S^*(v), e_2\rangle e_2 + \ldots + \langle S^*(v), e_n\rangle e_n)$$
$$= SS^*(v)$$

This implies S is bijective, and so S^* is its inverse, and so $SS^* = S^*S = I$. If

we assume (5), then for an arbitrary vector $v \in V$, we have

$$\|S(v)\|^2 = \langle S(v), S(v) \rangle$$
$$= \langle S^* S(v), v \rangle$$
$$= \langle v, v \rangle$$
$$= \|v\|^2.$$

So S preserves norms, and so we have (4).

Now assume (5), so $S^* = S^{-1}$, so S^* is an isometry as the inverse of an isometry is an isometry. Now assume (6), so $(S^*)^* = S$ is an isometry, so we have (5). $\qquad \square$

Theorem 39. Let V be a finite dimensional complex inner product space and $S \in \mathcal{L}(V)$ an operator. Then S is an isometry precisely when there is an orthonormal basis of V of eigenvectors of S with corresponding eigenvectors of magnitude 1.

Proof. First assume that S is an isometry. Then S and S^* are isometries, so for an arbitrary $v \in V$, we have $\|S(v)\| = \|v\| = \|S^*(v)\|$, so S is normal. So by the spectral theorem, we have an orthonormal basis $\mathcal{V} = \{e_1, e_2, \ldots, e_n\}$ of V consisting of eigenvectors of S. Consider an arbitrary eigenvalue λ of S, so we have for some non-zero $v \in V$, we have $T(v) = \lambda v$, which yields $|\lambda| = 1$ as T is an isometry.

Now assume that there is an orthonormal basis $\mathcal{V} = \{e_1, e_2, \ldots, e_n\}$ of V consisting of eigenvectors of S with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$

with magnitude 1. So for an arbitrary vector $v \in V$, we have

$$\|S(v)\|^2 = \|S(\langle v, e_1 \rangle e_1 + \langle v, e_2 \rangle e_2 + \ldots + \langle v, e_n \rangle e_n)\|^2$$

$$= \|\lambda_1 \langle v, e_1 \rangle e_1 + \lambda_2 \langle v, e_2 \rangle e_2 + \ldots + \lambda_n \langle v, e_n \rangle e_n\|^2$$

$$= \|\lambda_1 \langle v, e_1 \rangle\|^2 + \|\lambda_2 \langle v, e_2 \rangle\|^2 + \ldots + \|\lambda_n \langle v, e_n \rangle\|^2$$

$$= \|\langle v, e_1 \rangle\|^2 + \|\langle v, e_2 \rangle\|^2 + \ldots + \|\langle v, e_n \rangle\|^2$$

$$= \|v\|^2$$

Hence S is an isometry. $\qquad\square$

5.4 Polar Decomposition and Singular Value Decomposition

Definition 27 (Principal Square Root). Let V be a finite dimensional inner product space and $T \in \mathcal{L}(V)$ be a positive operator. Then we denote by $\sqrt{T}$ the unique positive square root of T.

Theorem 40 (Polar Decomposition). Let V be a finite-dimensional vector space, and $T \in \mathcal{L}(V)$ an operator. Then for we have for some isometry S

$$T = S\sqrt{T^*T}$$

Proof. We will construct the desired isometry by defining it on $\sqrt{T^*T}(V)$ and $(\sqrt{T^*T}(V))^\perp$, which will determine it on V as $V = \sqrt{T^*T}(V) \oplus (\sqrt{T^*T}(V))^\perp$.

We define S_1 on $\sqrt{T^*T}(V)$ by

$$S_1(\sqrt{T^*T}(v)) = T(v)$$

It trivially follows that S_1 is a well-defined linear map that preserves norms.

Clearly S_1 is bijective, hence we have $\dim(\sqrt{T^*T}(V))^\perp = \dim(T(V))^\perp = n$. So let $\{e_1, e_2, \ldots, e_n\}$ and $\{f_1, f_2, \ldots, f_n\}$ be orthnormal bases of $(\sqrt{T^*T}(V))^\perp$ and $(T(V))^\perp$, respectively. We may now define S_2 on $(\sqrt{T^*T}(V))^\perp$ by

$$S_2(e_j) = f_j$$

This is also trivally an isometry.

We now define S as the unique operator that coheres with S_1 and S_2 on their respective domains. For an arbitrary $v \in V$, we may write $v = u + w$ for $u \in \sqrt{T^*T}(V)$ and $w \in (\sqrt{T^*T}(V))^\perp$, so we have

$$
\begin{aligned}
\|S(v)\|^2 &= \|S(u) + S(w)\|^2 \\
&= \|S(u)\|^2 + \|S(w)\|^2 \\
&= \|u\|^2 + \|w\|^2 \\
&= \|u + w\|^2 \\
&= \|v\|^2
\end{aligned}
$$

$\square$

Definition 28 (Singular Value)**.** Let $T \in \mathcal{L}(V)$ be an operator on V. Then the singular values of T are the eigenvalues of $\sqrt{T^*T}$, repeated $\dim E(\lambda, \sqrt{T^*T})$ times.

The singular values are useful over the eigenvalues in particular as they always correspond to an orthnormal basis of eigenvectors, as $\sqrt{T^*T}$ is self-adjoint, whilst T is not necessarily even normal.

Theorem 41 (Singular Value Decomposition)**.** Let $T \in \mathcal{L}(V)$ be an operator with singular values $s_1, s_2, \ldots, s_n$. Then there are orthnormal bases $\mathcal{V}_1 =$

$\{e_1, e_2, \ldots, e_n\}$ and $\mathcal{V}_2 = \{f_1, f_2, \ldots, f_n\}$ such that for all $v \in V$ we have

$$T(v) = s_1 \langle v, e_1 \rangle f_1 + s_2 \langle v, e_2 \rangle f_2 + \ldots + s_n \langle v, e_n \rangle f_n.$$

Proof. By polar decomposition, we have an isometry S such that $T = S\sqrt{T^*T}$.

Let $\mathcal{V}_1$ be an orthonormal basis of eigenvectors $\sqrt{T^*T}$. Then $\mathcal{V}_2 = S(\mathcal{V}_1)$ is an orthonormal basis, as S is an isometry. So we have for an arbitrary $v \in V$ that

$$v = \langle v, e_1 \rangle e_1 + \langle v, e_2 \rangle e_2 + \ldots + \langle v, e_n \rangle e_n$$

$$\sqrt{T^*T}(v) = s_1 \langle v, e_1 \rangle e_1 + s_2 \langle v, e_2 \rangle e_2 + \ldots + s_n \langle v, e_n \rangle e_n$$

$$S\sqrt{T^*T}(v) = s_1 \langle v, e_1 \rangle S(e_1) + s_2 \langle v, e_2 \rangle S(e_2) + \ldots + s_n \langle v, e_n \rangle S(e_n)$$

$$T(v) = s_1 \langle v, e_1 \rangle f_1 + s_2 \langle v, e_2 \rangle f_2 + \ldots + s_n \langle v, e_n \rangle f_n$$

$\square$

The significance of this result is that we may always have a diagonal matrix with respect to some orthonormal bases for an operator, but given that the bases are not necessarily identical.

6 Operators on Complex Vector Spaces

6.1 Generalised Eigenvectors

Theorem 42. Let $T \in \mathcal{L}(V)$ be an operator. Then we have that

$$\ker(T^0) \subset \ker(T^1) \subset \ker(T^2) \subset \ldots$$

Proof. For any nonnegative integer k, we have

$$T^{k+1}(\ker(T^k)) = T(T^k(\ker(T^k)))$$
$$= T(\{0\})$$
$$= \{0\}$$

$\square$

Theorem 43. If $\ker(T^k) = \ker(T^{k+1})$, then we have

$$\ker(T^k) = \ker(T^{k+1}) = \ker(T^{k+2}) = \ldots$$

Proof. Consider an arbitrary $v \in \ker(T^{k+2})$, then we have

$$T^{k+1}T(v) = 0$$
$$T(v) \in \ker(T^{k+1})$$
$$T(v) \in \ker(T^k)$$
$$T^{k+1}(v) = 0$$
$$v \in \ker(T^{k+1})$$

So we have $\ker(T^{k+2}) \subset \ker(T^{k+1})$, and also the reverse subset relation. Hence we have that the kernels are equal, and so by induction the desired chain of equalities. $\square$

Theorem 44. Let n be the dimension of the vector space V, and $T \in \mathcal{L}(V)$ an operator on V. Then we have

$$\ker(T^n) = \ker(T^{n+1)}) = \ker(T^{n+2}) = \ldots$$

Proof. Assume the contrary, then we have that

$$\{0\} = \ker(T^0) \subsetneq \ker(T^1) \subsetneq \ldots \subsetneq \ker(T^n) \subsetneq \ker(T^{n+1})$$

Then this would imply that $\dim \ker(T^{n+1}) \geq n+1$, which is impossible as its a subspace of V. $\qquad\square$

As a consequence of the Fundemental Theorem of Linear Maps, we have that $\dim V = \dim \ker(T) + \dim T(V)$ for any operator $T \in \mathcal{L}(V)$. So if the sum were direct, then it would be a direct sum decomposition of V. However this is not necessarily the case, for example the operator $T(x,y,z) = (y,z,0)$ on $\mathbb{R}^3$. However as the next result shows, we may do so for a power of an operator.

Theorem 45. Let n be the dimension of the vector space V, and $T \in \mathcal{L}(V)$ an operator on V. Then we have

$$V = \ker(T^n) \oplus T^n(V)$$

Proof. We claim the terms are disjoint. Consider an arbitrary vector $v \in V$ in their intersection. Then we have for some $u \in V$ that $T^n(u) = v$, and $T^n(v) = 0$. So we have $T^{2n}(u) = 0$, which implies $v = T^n(u) = 0$, as the kernels eventually stop growing. So they are indeed disjoint, as so the sum is direct.

By the fundamental theorem of linear maps, we have that $n = \dim \ker(T^n) + \dim T^n(V)$, so the direct sum is V. $\qquad\square$

Definition 29 (Generalised Eigenvector). Let $T \in \mathcal{L}(V)$ be an operator. Then we say a non-zero vector $v \in V$ is a generalised eigenvector of T with corresponding eigenvalue λ if v belongs to the kernel of $(T - \lambda I)^j$ for some nonnegative integer j.

We let the generalised eigenvectors with corresponding eigenvalue λ along with 0 be the generalised eigenspace, $G(\lambda, T)$

Theorem 46. Let $T \in \mathcal{L}(V)$ be an operator and $\mathcal{V} = \{v_1, v_2, \ldots, v_m\}$ generalised eigenvectors with pairwise distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$. Then $\mathcal{V}$ is linearly independent.

Proof. Let j be the largest integer such that $(T - \lambda_1 I)^j(v_1) = w$ is non-zero. So we have that w is an eigenvector of T with eigenvalue λ_1.

Now consider an arbitrary zero linear combination of $\mathcal{V}$

$$a_1 v_1 + a_2 v_2 + \ldots + a_n v_n = 0$$

Applying the operator $(T - \lambda_1 I)^j (T - \lambda_2 I)^n (T - \lambda_3 I)^n \ldots (T - \lambda_m I)^n$ then yields

$$a_1 (\lambda_1 - \lambda_2)^n (\lambda_1 - \lambda_3)^n \ldots (\lambda_1 - \lambda_m)^n w = 0$$

As all the eigenvalues are pairwise distinct and $w \neq 0$, we have that $a_1 = 0$. By similar logic we have all the other scalars are 0. $\square$

As already established, the kernels of powers of an operator form an ascending chain under the $\subset$ relation which, in the case of finite-dimensional vector spaces, eventually becomes constant.

So it is interesting to ask what space does this chain terminate on, the two obvious interesting special cases are when the chain is eventually $\{0\}$ and V.

In the former case this is characteristic of bijective operators, however the latter case corresponds to an operator that is 0 when raised to a sufficiently high power. We will call these operators nilpotent.

Definition 30 (Nilpotent Operator). We say the operator $T \in \mathcal{L}(V)$ is nilpotent if $T^j = 0$ for some nonnegative integer j.

Theorem 47. Let $N \in \mathcal{L}(V)$ be a nilpotent operator. Then there is a basis of V with respect to which the matrix of N is upper triangular with all diagonal entries 0.

Proof. The kernel of powers of an operator form an ascending chain under the $\subset$ relation, and if there is ever equality then all proceeding terms are equal. Since the sequence corresponding to a nilpotent operator eventually has dimension n, we must have that $\dim \ker(N^j) \geq j$ for all $0 \leq j \leq n$.

This means we may choose a basis $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ such that each $v_j \in \ker(N^j)$. For any j, we have $N(v_j) \in \ker(N^{j-1})$, which is spanned by $v_1, v_2, \ldots, v_{j-1}$. So the matrix has the desired form. $\qquad\square$

In the case of a complex vector space, this can also be shown using the fact that every operator has a basis such that the matrix is upper-triangular, and the diagonal entries are eigenvalues, and in the case of a nilpotent operator all eigenvalues must be 0. However this would not be valid for spaces over arbitrary fields as not every operator over a general field need have an eigenvalue.

6.2   Decomposition of an Operator

Theorem 48. Let V be a complex vector space and $T \in \mathcal{L}(V)$ be an operator on V, whose distinct eigenvalues are $\lambda_1, \lambda_2, \ldots, \lambda_m$. Then we have

$$V = G(\lambda_1, T) \oplus G(\lambda_2, T) \oplus \ldots \oplus G(\lambda_m, T)$$

Proof. We show this by induction on the dimension of V. As V is a complex vector space, we have that $m \geq 1$, so $V = \ker(T - \lambda_1 I)^n \oplus (T - \lambda_1 I)^n(V)$. Hence we have

$$V = G(\lambda_1, T) \oplus U$$

Where $U = (T - \lambda_1 I)^n(V)$. By induction, we have that

$$U = G(\lambda_2, T|_U) \oplus \ldots \oplus G(\lambda_m, T|_U)$$

It is trivial that $G(\lambda_j, T|_U) \subset G(\lambda_j, T)$. Now consider an arbitrary $v \in G(\lambda_j, T) \subset V$. Then we have for unique $v_1 \in G(\lambda_1, T), v_2 \in G(\lambda_2, T|_U) \subset G(\lambda_2, T), \ldots, v_m \in G(\lambda_m, T|_U) \subset G(\lambda_m, T)$ that

$$v = v_1 + v_2 + \ldots + v_m$$

$$v_1 + v_2 + \ldots + (v_j - v) + \ldots + v_m = 0$$

By linear independence of generalised eigenvectors with distinct eigenvalues, every term is 0. In particular, we have $v = v_j \in G(\lambda_j, T|_U$, so we have $G(\lambda_j, T) \subset G(\lambda_j, T|_U)$, so they are equal.

Hence we have that $U = G(\lambda_2, T) \oplus G(\lambda_3, T) \oplus \ldots \oplus G(\lambda_m, T)$, yielding the desired direct sum decomposition. $\square$

In particular this means that for any operator we have a basis consiting of generalised eigenvectors.

Definition 31 (Multiplicity of an Eigenvalue)**.** Let λ be an eigenvalue of an operator $T \in \mathcal{L}(V)$. Then we call the dimension of the generalised eigenspace the algebraic multplicity and the dimension of the eigenspace the geometric multplicity.

In these notes when we use the term multiplicity unqualified we refer to the algebraic multiplicity.

Definition 32 (Block Diagonal Matrix)**.** We say a square matrix is block diagonal if all of its entries are zero except square matrices along the diagonal.