LINEAR ALGEBRA: VECTOR SPACES AND OPERATORS

B. Zwiebach

October 21, 2013

Contents

1 Vector spaces and dimensionality

In quantum mechanics the state of a physical system is a *vector* in a *complex* vector space. Observables are linear operators, in fact, Hermitian operators acting on this complex vector space. The purpose of this chapter is to learn the basics of vector spaces, the structures that can be built on those spaces, and the operators that act on them.

Complex vector spaces are somewhat different from the more familiar real vector spaces. I would say they have more powerful properties. In order to understand more generally complex vector spaces it is useful to compare them often to their real dimensional friends. We will follow here the discussion of the book *Linear algebra done right*, by Sheldon Axler.

In a vector space one has vectors and numbers. We can add vectors to get vectors and we can multiply vectors by numbers to get vectors. If the numbers we use are real, we have a real vector space. If the numbers we use are complex, we have a complex vector space. More generally, the numbers we use belong to what is called in mathematics a 'field' and denoted by the letter $\mathbb{F}$. We will discuss just two cases, $\mathbb{F} = \mathbb{R}$, meaning that the numbers are real, and $\mathbb{F} = \mathbb{C}$, meaning that the numbers are complex.

The definition of a vector space is the same for $\mathbb{F}$ being $\mathbb{R}$ or $\mathbb{C}$. A vector space V is a set of vectors with an operation of **addition** $(+)$ that assigns an element $u + v \in V$ to each $u, v \in V$. This means that V is closed under addition. There is also a **scalar multiplication** by elements of $\mathbb{F}$, with $av \in V$

for any $a \in \mathbb{F}$ and $v \in V$. This means the space V is closed under multiplication by numbers. These operations must satisfy the following additional properties:

1. $u + v = v + u \in V$ for all $u, v \in V$ (addition is commutative).

2. $u + (v + w) = (u + v) + w$ and $(ab)u = a(bu)$ for any $u, v, w \in V$ and $a, b \in \mathbb{F}$ (associativity).

3. There is a vector $0 \in V$ such that $0 + u = u$ for all $u \in V$ (additive identity).

4. For each $v \in V$ there is a $u \in V$ such that $v + u = 0$ (additive inverse).

5. The element $1 \in \mathbb{F}$ satisfies $1v = v$ for all $v \in V$ (multiplicative identity).

6. $a(u+v) = au+av$ and $(a+b)v = av+bv$ for every $u, v \in V$ and $a, b \in \mathbb{F}$ (distributive property).

This definition is very efficient. Several familiar properties follow from it by short proofs (which we will not give, but are not complicated and you may try to produce):

- The additive identity is unique: any vector $0'$ that acts like 0 is actually equal to 0.

- $0v = 0$, for any $v \in V$, where the first zero is a number and the second one is a vector. This means that the number zero acts as expected when multiplying a vector.

- $a0 = 0$, for any $a \in \mathbb{F}$. Here both zeroes are vectors. This means that the zero vector multiplied by any number is still the zero vector.

- The additive inverse of any vector $v \in V$ is unique. It is denoted by $-v$ and in fact $-v = (-1)v$.

We must emphasize that while the numbers, in $\mathbb{F}$ are sometimes real or complex, we never speak of the vectors themselves as real or complex. A vector multiplied by a complex number is not said to be a complex vector, for example! The vectors in a real vector space are not themselves real, nor are the vectors in a complex vector space complex. We have the following examples of vector spaces:

1. The set of N-component vectors
$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix}, \quad a_i \in \mathbb{R}\,, \quad i = 1, 2, \ldots N\,. \tag{1.1}$$
form a real vector space.

2. The set of $M \times N$ matrices with complex entries
$$\begin{pmatrix} a_{11} & \ldots & a_{1N} \\ a_{21} & \ldots & a_{2N} \\ \vdots & \vdots & \vdots \\ a_{M1} & \ldots & a_{MN} \end{pmatrix}, \qquad a_{ij} \in \mathbb{C}\,, \tag{1.2}$$

is a complex vector space. In here multiplication by a constant multiplies each entry of the matrix by the constant.

3. We can have matrices with complex entries that naturally form a real vector space. The space of two-by-two *hermitian* matrices define a *real* vector space. They do not form a complex vector space since multiplication of a hermitian matrix by a complex number ruins the hermiticity.

4. The set $\mathcal{P}(\mathbb{F})$ of polynomials $p(z)$. Here the variable $z \in \mathbb{F}$ and $p(z) \in \mathbb{F}$. Each polynomial $p(z)$ has coefficients $a_0, a_1, \ldots a_n$ also in $\mathbb{F}$:

$$p(z) = a_0 + a_1 z + a_2 z^2 + \ldots + a_n z^n \,. \tag{1.3}$$

By definition, the integer n is finite but it can take any nonnegative value. Addition of polynomials works as expected and multiplication by a constant is also the obvious multiplication. The space $\mathcal{P}(\mathbb{F})$ of all polynomials so defined form a vector space over $\mathbb{F}$.

5. The set $\mathbb{F}^\infty$ of infinite sequences $(x_1, x_2, \ldots)$ of elements $x_i \in \mathbb{F}$. Here

$$\begin{aligned} (x_1, x_2, \ldots) + (y_1, y_2, \ldots) \; &= (x_1 + y_1, x_2 + y_2, \ldots) \\ a(x_1, x_2, \ldots) \; &= \; (ax_1, ax_2, \ldots) \quad a \in \mathbb{F} \,. \end{aligned} \tag{1.4}$$

This is a vector space over $\mathbb{F}$.

6. The set of complex functions on an interval $x \in [0, L]$, form a vector space over $\mathbb{C}$.

To better understand a vector space one can try to figure out its possible subspaces. A **subspace** of a vector space V is a subset of V that is also a vector space. To verify that a subset U of V is a subspace you must check that U contains the vector 0, and that U is closed under addition and scalar multiplication.

Sometimes a vector space V can be described clearly in terms of collection $U_1, U_2, \ldots U_m$ of subspaces of V. We say that the space V is the **direct sum** of the subspaces $U_1, U_2, \ldots U_m$ and we write

$$V \;=\; U_1 \oplus U_2 \oplus \cdots \oplus U_m \tag{1.5}$$

if any vector in V can be written *uniquely* as the sum $u_1 + u_2 + \ldots + u_m$, where $u_i \in U_i$. To check uniqueness one can, alternatively, verify that the only way to write 0 as a sum $u_1 + u_2 + \ldots + u_m$ with $u_i \in U_i$ is by taking all u_i's equal to zero. For the case of two subspaces $V = U \oplus W$, it suffices to prove that any vector can be written as $u + w$ with $u \in U$ and $w \in W$ and that $U \cap W = 0$.

Given a vector space we can produce lists of vectors. A **list** $(v_1, v_2, \ldots, v_n)$ of vectors in V contains, by definition, a finite number of vectors. The number of vectors in the list is the length of the list. The **span** of a list of vectors $(v_1, v_2, \cdots v_n)$ in V, denoted as $\text{span}(v_1, v_2, \cdots, v_n)$, is the set of all linear combinations of these vectors

$$a_1 v_1 + a_2 v_2 + \ldots a_n v_n \,, \qquad a_i \in \mathbb{F} \tag{1.6}$$

A vector space V is spanned by a list $(v_1, v_2, \cdots v_n)$ if $V = \text{span}(v_1, v_2, \cdots v_n)$.

Now comes a very natural definition: A vector space V is said to be **finite dimensional** if it is spanned by some list of vectors in V. If V is not finite dimensional, it is **infinite dimensional**. In such case, no list of vectors from V can span V.

Let us show that the vector space of all polynomials $p(z)$ considered in Example 4 is an infinite dimensional vector space. Indeed, consider any list of polynomials. In this list there is a polynomial of maximum degree (recall the list is finite). Thus polynomials of higher degree are not in the span of the list. Since no list can span the space, it is infinite dimensional.

For example 1, consider the list of vectors $(e_1, e_2, \ldots e_N)$ with

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \; e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots \quad e_N = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \tag{1.7}$$

This list spans the space (the vector displayed is $a_1 e_1 + a_2 e_2 + \ldots a_N e_N$). This vector space is finite dimensional.

A list of vectors $(v_1, v_2, \ldots, v_n)$, with $v_i \in V$ is said to be **linearly independent** if the equation

$$a_1 v_1 + a_2 v_2 + \ldots + a_n v_n \;=\; 0\,, \tag{1.8}$$

only has the solution $a_1 = a_2 = \cdots = a_n = 0$. One can show that the length of any linearly independent list is shorter or equal to the length of any spanning list. This is reasonable, because spanning lists can be arbitrarily long (adding vectors to a spanning list gives still a spanning list), but a linearly independent list cannot be enlarged beyond a certain point.

Finally, we get to the concept of a basis for a vector space. A **basis** of V is a list of vectors in V that both spans V and it is linearly independent. Mathematicians easily prove that any finite dimensional vector space has a basis. Moreover, all bases of a finite dimensional vector space have the same length. The **dimension** of a finite-dimensional vector space is given by the length of any list of basis vectors. One can also show that for a finite dimensional vector space a list of vectors of length $\dim V$ is a basis if it is linearly independent list or if it is a spanning list.

For example 1 we see that the list $(e_1, e_2, \ldots e_N)$ in (1.7) is not only a spanning list but a linearly independent list (prove it!). Thus the dimensionality of this space is N.

For example 3, recall that the most general hermitian two-by-two matrix takes the form

$$\begin{pmatrix} a_0 + a_3 & a_1 - i a_2 \\ a_1 + i a_2 & a_0 - a_3 \end{pmatrix}, \qquad a_0, a_1, a_2, a_3 \in \mathbb{R}. \tag{1.9}$$

Now consider the following list of four 'vectors' $(\mathbf{1}, \sigma_1, \sigma_2, \sigma_3)$. All entries in this list are hermitian matrices, so this is a list of vectors in the space. Moreover they span the space since the most general hermitian matrix, as shown above, is simply $a_0 \mathbf{1} + a_1 \sigma_1 + a_2 \sigma_2 + a_3 \sigma_3$. The list is linearly independent as $a_0\mathbf{1} + a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3 = 0$ implies that

$$\begin{pmatrix} a_0 + a_3 & a_1 - ia_2 \\ a_1 + ia_2 & a_0 - a_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \tag{1.10}$$

and you can quickly see that this implies a_0, a_1, a_2, and a_3 are zero. So the list is a basis and the space in question is a four-dimensional real vector space.

Exercise. Explain why the vector space in example 2 has dimension $M \cdot N$.

It seems pretty obvious that the vector space in example 5 is infinite dimensional, but it actually takes a bit of work to prove it.

2 Linear operators and matrices

A linear map refers in general to a certain kind of function from one vector space V to another vector space W. When the linear map takes the vector space V to itself, we call the linear map a linear operator. We will focus our attention on those operators. Let us then define a linear operator.

A **linear operator** T on a vector space V is a function that takes V to V with the properties:

1. $T(u+v) = Tu + Tv$, for all $u, v \in V$.
2. $T(au) = aTu$, for all $a \in \mathbb{F}$ and $u \in V$.

We call $\mathcal{L}(V)$ the set of all linear operators that act on V. This can be a very interesting set, as we will see below. Let us consider a few examples of linear operators.

1. Let V denote the space of real polynomials $p(x)$ of a real variable x with real coefficients. Here are two linear operators:

   - Let T denote differentiation: $Tp = p'$. This operator is linear because $(p_1 + p_2)' = p_1' + p_2'$ and $(ap)' = ap'$.
   - Let S denote multiplication by x: $Sp = xp$. S is also a linear operator.

2. In the space $\mathbb{F}^\infty$ of infinite sequences define the left-shift operator L by

   $$L(x_1, x_2, x_3, \ldots) = (x_2, x_3, \ldots). \tag{2.11}$$

   We lose the first entry, but that is perfectly consistent with linearity. We also have the right-shift operator R that acts as follows:

   $$R(x_1, x_2, \ldots) = (0, x_1, x_2, \ldots). \tag{2.12}$$

   Note that the first entry in the result is zero. It could not be any other number because the zero element (a sequence of all zeroes) should be mapped to itself (by linearity).

3. For any V, the zero map 0 such that $0v = 0$. This map is linear and maps all elements of V to the zero element.

4. For any V, the identity map I for which $Iv = v$ for all $v \in V$. This map leaves all vectors invariant.

Since operators on V can be added and can also be multiplied by numbers, the set $\mathcal{L}(V)$ introduced above is itself a vector space (the vectors being the operators!). Indeed for any two operators $T, S \in \mathcal{L}(V)$ we have the natural definition

$$\begin{aligned} (S+T)v &= Sv + Tv\,, \\ (aS)v &= a(Sv)\,. \end{aligned} \tag{2.13}$$

The additive identity in the vector space $\mathcal{L}(V)$ is the zero map of example 3.

In this vector space there is a surprising new structure: the vectors (the operators!) can be multiplied. There is a multiplication of linear operators that gives a linear operator. We just let one operator act first and the second later. So given $S, T \in \mathcal{L}(V)$ we define the operator ST as

$$(ST)v \equiv S(Tv) \tag{2.14}$$

You should convince yourself that ST is a linear operator. This product structure in the space of linear operators is associative: $S(TU) = (ST)U$, for S, T, U, linear operators. Moreover it has an identity element: the identity map of example 4. Most crucially this multiplication is, in general, *noncommutative*. We can check this using the two operators T and S of example 1 acting on the polynomial $p = x^n$. Since T differentiates and S multiplies by x we get

$$(TS)x^n = T(Sx^n) = T(x^{n+1}) = (n+1)x^n\,, \quad \text{while} \quad (ST)x^n = S(Tx^n) = S(nx^{n-1}) = nx^n\,. \tag{2.15}$$

We can quantify this failure of commutativity by writing the difference

$$(TS - ST)x^n = (n+1)x^n - nx^n = x^n = I\,x^n \tag{2.16}$$

where we inserted the identity operator at the last step. Since this relation is true for any x^n, it would also hold acting on any polynomial, namely on any element of the vector space. So we write

$$[\,T\,, S\,] \;=\; I\,. \tag{2.17}$$

where we introduced the commutator $[\cdot,\cdot]$ of two operators X, Y, defined as $[X, Y] \equiv XY - YX$.

The most basic features of an operator are captured by two simple concepts: its null space and its range. Given some linear operator T on V it is of interest to consider those elements of V that are mapped to the zero element. The **null space** (or kernel) of $T \in \mathcal{L}(V)$ is the subset of vectors in V that are mapped to zero by T:

$$\text{null}\, T \;=\; \{v \in V;\; Tv = 0\}\,. \tag{2.18}$$

Actually $\text{null}\, T$ is a *subspace* of V (The only nontrivial part of this proof is to show that $T(0) = 0$. This follows from $T(0) = T(0+0) = T(0) + T(0)$ and then adding to both sides of this equation the additive inverse to $T(0)$).

A linear operator $T : V \to V$ is said to be **injective** if $Tu = Tv$, with $u, v \in V$, implies $u = v$. An injective map is called a *one-to-one* map, because not two different elements can be mapped to the same one. In fact, physicist Sean Carroll has suggested that a better name would be *two-to-two* as injectivity really means that two different elements are mapped by T to two different elements! We leave for you as an exercise to prove the following important characterization of injective maps:

Exercise. Show that T is injective if and only if $\text{null}\, T = \{0\}$.

Given a linear operator T on V it is also of interest to consider the elements of V of the form Tv. The linear operator may not produce by its action all of the elements of V. We define the **range** of T as the image of V under the map T:

$$\text{range}\ T \;=\; \{Tv;\; v \in V\}\,. \tag{2.19}$$

Actually $\text{range}\, T$ is a *subspace* of V (can you prove it?). The linear operator T is said to be **surjective** if $\text{range}\, T = V$. That is, if the image of V under T is the complete V.

Since both the null space and the range of a linear operator $T : V \to V$ are subspaces of V, one can assign a dimension to them, and the following theorem is nontrivial:

$$\dim V \;=\; \dim(\text{null}\ T) \;+\; \dim(\text{range}\ T)\,. \tag{2.20}$$

Example. Describe the null space and range of the operator

$$T = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{2.21}$$

Let us now consider invertible linear operators. A linear operator $T \in \mathcal{L}(V)$ is **invertible** if there exists another linear operator $S \in \mathcal{L}(V)$ such that ST and TS are identity maps (written as I). The linear operator S is called the **inverse** of T. The inverse is actually unique. Say S and S' are inverses of T. Then we have

$$S \;=\; SI \;=\; S(TS') \;=\; (ST)S' \;=\; IS' \;=\; S'\,. \tag{2.22}$$

Note that we required the inverse S to be an inverse acting from the left and acting from the right. This is useful for infinite dimensional vector spaces. For finite-dimensional vector spaces one suffices; one can then show that $ST = I$ if and only if $TS = I$.

It is useful to have a good characterization of invertible linear operators. For a finite-dimensional vector space V the following three statements are equivalent!

$$\text{Finite dimension:} \quad \boxed{T \text{ is invertible}} \;\longleftrightarrow\; \boxed{T \text{ is injective}} \;\longleftrightarrow\; \boxed{T \text{ is surjective}} \tag{2.23}$$

For infinite dimensional vector spaces injectivity and surjectivity are not equivalent (each can fail independently). In that case invertibility is equivalent to injectivity plus surjectivity:

$$\text{Infinite dimension:} \quad \boxed{T \text{ is invertible}} \longleftrightarrow \boxed{T \text{ is injective and surjective}} \tag{2.24}$$

The left shift operator L is not injective (maps $(x_1, 0, \ldots)$ to zero) but it is surjective. The right shift operator is not surjective although it is injective.

Now we consider the **matrix associated to a linear operator** T that acts on a vector space V. This matrix will depend on the basis we choose for V. Let us declare that our basis is the list $(v_1, v_2, \ldots v_n)$. It is clear that the full knowledge of the action of T on V is encoded in the action of T on the basis vectors, that is on the values $(Tv_1, Tv_2, \ldots, Tv_n)$. Since Tv_j is in V, it can be written as a linear combination of basis vectors. We then have

$$\boxed{Tv_j \;=\; T_{1\,j}\, v_1 \,+\, T_{2\,j}\, v_2 \,+\, \ldots \,+ T_{n\,j}\, v_n\,,} \tag{2.25}$$

where we introduced the constants $T_{i,j}$ that are known if the operator T is known. As we will see, these are the entries form the matrix representation of the operator T in the chosen basis. The above relation can be written more briefly as

$$\boxed{Tv_j \;=\; \sum_{i=1}^{n} T_{ij}\, v_i\,.} \tag{2.26}$$

When we deal with different bases it can be useful to use notation where we replace

$$T_{ij} \quad \to \quad T_{ij}(\{v\})\,, \tag{2.27}$$

so that it makes clear that T is being represented using the v basis $(v_1, \ldots, v_n)$.

I want to make clear why (2.25) is reasonable before we show that it makes for a consistent association between operator multiplication and matrix multiplication. The left-hand side, where we have the action of the matrix for T on the j-th basis vector, can be viewed concretely as

$$Tv_j \quad \longleftrightarrow \quad \begin{pmatrix} T_{11} & \cdots & T_{1\,j} & \cdots & T_{1n} \\ T_{21} & \cdots & T_{2\,j} & \cdots & T_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ T_{n1} & \cdots & T_{n\,j} & \cdots & T_{nn} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \; j\text{-th position} \tag{2.28}$$

where the column vector has zeroes everywhere except on the j-th entry. The product, by the usual rule of matrix multiplication is the column vector

$$\begin{pmatrix} T_{1\,j} \\ T_{2\,j} \\ \vdots \\ T_{n\,j} \end{pmatrix} = T_{1\,j} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + T_{2\,j} \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \ \ldots \ + T_{n\,j} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \quad \longleftrightarrow \quad T_{1\,j} v_1 + \ldots T_{n\,j} v_n\,. \tag{2.29}$$

which we identify with the right-hand side of (2.25). So (2.25) is reasonable.

Exercise. Verify that the matrix representation of the identity operator is a diagonal matrix with an entry of one at each element of the diagonal. This is true for any basis.

Let us now examine the product of two operators and their matrix representation. Consider the operator TS acting on v_j:

$$(TS)v_j \;=\; T(Sv_j) \;=\; T\sum_p S_{pj}v_p \;=\; \sum_p S_{pj}\,Tv_p \;=\; \sum_p S_{pj}\,\sum_i T_{ip}v_i \tag{2.30}$$

so that changing the order of the sums we find

$$(TS)v_j \;=\; \sum_i\Big(\sum_p T_{ip}S_{pj}\Big)\,v_i\,. \tag{2.31}$$

Using the identification implicit in (2.26) we see that the object in parenthesis is the i,j matrix element of the matrix that represents TS. Therefore we found

$$(TS)_{ij} \;=\; \sum_p T_{ip}S_{pj}\,, \tag{2.32}$$

which is precisely the right formula for matrix multiplication. In other words, the matrix that represents TS is the product of the matrix that represents T with the matrix that represents S, in that order.

Changing basis

While matrix representations are very useful for concrete visualization, they are basis dependent. It is a good idea to try to figure out if there are quantities that can be calculated using a matrix representation that are, nevertheless, guaranteed to be basis independent. One such quantity is the **trace** of the matrix representation of a linear operator. The trace is the sum of the matrix elements in the diagonal. Remarkably, that sum is the same independent of the basis used. Consider a linear operator T in $\mathcal{L}(V)$ and two sets of basis vectors $(v_1,\ldots,v_n)$ and $(u_1,\ldots,u_n)$ for V. Using the explicit notation (2.27) for the matrix representation we state this property as

$$\mathrm{tr}\,T(\{v\}) \;=\; \mathrm{tr}\,T(\{u\})\,. \tag{2.33}$$

We will establish this result below. On the other hand, if this trace is actually basis independent, there should be a way to define the trace of the linear operator T *without* using its matrix representation. This is actually possible, as we will see. Another basis independent quantity is the determinant of the matrix representation of T.

Let us then consider the effect of a change of basis on the matrix representation of an operator. Consider a vector space V and a change of basis from $(v_1,\ldots v_n)$ to $(u_1,\ldots u_n)$ defined by the linear operator A as follows:

$$A:\; v_k\to u_k,\;\text{ for } k=1,\ldots,n\,. \tag{2.34}$$

This can also be written as

$$A v_k = u_k \tag{2.35}$$

Since we know how A acts on every element of the basis we know, by linearity how it acts on any vector. The operator A is clearly *invertible* because, letting $B : u_k \to v_k$ or

$$B u_k = v_k \,, \tag{2.36}$$

we have

$$\begin{aligned} BAv_k = B(Av_k) = Bu_k \;\; &= v_k \\ ABu_k = A(Bu_k) = Av_k \;\; &= u_k \,, \end{aligned} \tag{2.37}$$

showing that $BA = I$ and $AB = I$. Thus B is the inverse of A. Using the definition of matrix representation, the right-hand sides of the relations $u_k = Av_k$ and $v_k = Bu_k$ can be written so that the equations take the form

$$u_k = A_{jk}\, v_j \,, \qquad v_k = B_{jk}\, u_j \,, \tag{2.38}$$

where we used the convention that repeated indices are summed over. A_{ij} are the elements of the matrix representation of A in the v basis and B_{ij} are the elements of the matrix representation of B in the u basis. Replacing the second relation on the first, and then replacing the first on the second we get

$$\begin{aligned} u_k &= A_{jk}\, B_{ij}\, u_i = B_{ij} A_{jk}\, u_i \\ v_k &= B_{jk}\, A_{ij}\, v_i = A_{ij} B_{jk}\, v_i \end{aligned} \tag{2.39}$$

Since the u's and v's are basis vectors we must have

$$B_{ij} A_{jk} = \delta_{ik} \quad \text{and} \quad A_{ij} B_{jk} = \delta_{ik} \tag{2.40}$$

which means that the B matrix is the inverse of the A matrix. We have thus learned that

$$v_k = (A^{-1})_{jk}\, u_j \,. \tag{2.41}$$

We can now apply these preparatory results to the matrix representations of the operator T. We have, by definition,

$$T v_k = T_{ik}(\{v\})\, v_i \,. \tag{2.42}$$

We now want to calculate T on u_k so that we can read the formula for the matrix T on the u basis:

$$T u_k = T_{ik}(\{u\})\, u_i \,. \tag{2.43}$$

Computing the left-hand side, using the linearity of the operator T, we have

$$T u_k = T(A_{jk} v_j) = A_{jk} T v_j = A_{jk} T_{pj}(\{v\})\, v_p \tag{2.44}$$

and using (2.41) we get

$$Tu_k = A_{jk}T_{pj}(\{v\})\,(A^{-1})_{ip}\,u_i \;=\; \Big((A^{-1})_{ip}\,T_{pj}(\{v\})\,A_{jk}\Big)u_i \;=\; \big(A^{-1}T(\{v\})A\big)_{ik}u_i\,. \tag{2.45}$$

Comparing with (2.43) we get

$$T_{ij}(\{u\}) \;=\; \big(A^{-1}T(\{v\})A\big)_{ij} \quad\rightarrow\quad \boxed{\;T(\{u\}) \;=\; A^{-1}T(\{v\})A\,.\;} \tag{2.46}$$

This is the result we wanted to obtain.

The trace of a matrix T_{ij} is given by T_{ii}, where sum over i is understood. To show that the trace of T is basis independent we write

$$\begin{aligned} \mathrm{tr}(T(\{u\})) \;&=\; T_{ii}(\{u\}) \;=\; (A^{-1})_{ij}T_{jk}(\{v\})A_{ki} \\ &= A_{ki}(A^{-1})_{ij}T_{jk}(\{v\}) \\ &= \delta_{kj}T_{jk}(\{v\}) \;=\; T_{jj}(\{v\}) \;=\; \mathrm{tr}(T(\{v\}))\,. \end{aligned} \tag{2.47}$$

For the determinant we recall that $\det(AB) = (\det A)(\det B)$. Therefore $\det(A)\det(A^{-1}) = 1$. From (2.46) we then get

$$\det T(\{u\}) \;=\; \det(A^{-1})\,\det T(\{v\})\det A \;=\; \det T(\{v\})\,. \tag{2.48}$$

Thus the determinant of the matrix that represents a linear operator is independent of the basis used.

3 Eigenvalues and eigenvectors

In quantum mechanics we need to consider eigenvalues and eigenstates of hermitian operators acting on complex vector spaces. These operators are called observables and their eigenvalues represent possible results of a measurement. In order to acquire a better perspective on these matters, we consider the eigenvalue/eigenvector problem in more generality.

One way to understand the action of an operator $T \in \mathcal{L}(V)$ on a vector space V is to understand how it acts on subspaces of V, as those are smaller than V and thus possibly simpler to deal with. Let U denote a subspace of V. In general, the action of T may take elements of U outside U. We have a noteworthy situation if T acting on any element of U gives an element of U. In this case U is said to be **invariant** under T, and T is then a well-defined linear operator on U. A very interesting situation arises if a suitable list of invariant subspaces give the space V as a direct sum.

Of all subspaces, one-dimensional ones are the simplest. Given some vector $u \in V$ one can consider the one-dimensional subspace U spanned by u:

$$U = \{cu : \; c \in \mathbb{F}\}\,. \tag{3.49}$$

We can ask if the one-dimensional subspace U is left invariant by the operator T. For this Tu must be equal to a number times u, as this guarantees that $Tu \in U$. Calling the number λ, we write

$$T u = \lambda u . \tag{3.50}$$

This equation is so ubiquitous that names have been invented to label the objects involved. The number $\lambda \in \mathbb{F}$ is called an **eigenvalue** of the linear operator T if there is a **nonzero** vector $u \in V$ such that the equation above is satisfied. Suppose we find for some specific λ a nonzero vector u satisfying this equation. Then it follows that cu, for any $c \in \mathbb{F}$ also satisfies equation (3.50), so that the solution space of the equation includes the subspace U, which is now said to be an invariant subspace under T. It is convenient to call any vector that satisfies (3.50) for a given λ an **eigenvector** of T corresponding to λ. In doing so we are including the zero vector as a solution and thus as an eigenvector. It can often happen that for a given λ there are several linearly independent eigenvectors. In this case the invariant subspace associated with the eigenvalue λ is higher dimensional. The set of eigenvalues of T is called the **spectrum** of T.

Our equation above is equivalent to

$$(T - \lambda I)\, u = 0 , \tag{3.51}$$

for some nonzero u. It is therefore the case that

$$\boxed{\lambda \ \text{ is an eigenvalue}} \quad \longleftrightarrow \quad \boxed{(T - \lambda I) \quad \text{not injective.}} \tag{3.52}$$

Using (2.23) we conclude that λ is an eigenvalue also means that $(T - \lambda I)$ is **not invertible**, and not surjective. We also note that

$$\text{Set of eigenvectors of } T \text{ corresponding to } \lambda \ = \ \text{null}\,(T - \lambda I) . \tag{3.53}$$

It should be emphasized that the eigenvalues of T and the invariant subspaces (or eigenvectors associated with fixed eigenvalues) are basis independent objects. Nowhere in our discussion we had to invoke the use of a basis, nor we had to use any matrix representation. Below, we will discuss the familiar calculation of eigenvalues and eigenvectors using a matrix representation of the operator T in some particular basis.

Let us consider some examples. Take a real three-dimensional vector space V (our space to great accuracy!). Consider the rotation operator T that rotates all vectors by a fixed angle small about the z axis. To find eigenvalues and eigenvectors we just think of the invariant subspaces. We must ask which are the vectors for which this rotation doesn't change their direction and effectively just multiplies them by a number? Only the vectors along the z-direction do not change direction upon this rotation. So the vector space spanned by $\mathbf{e}_z$ is the invariant subspace, or the space of eigenvectors. The eigenvectors are associated with the eigenvalue of one, as the vectors are not altered at all by the rotation.

Consider now the case where T is a rotation by ninety degrees on a two-dimensional *real* vector space V. Are there one-dimensional subspaces left invariant by T? No, **all** vectors are rotated, none remains pointing in the same direction. Thus there are **no eigenvalues**, nor, of course, eigenvectors. If you tried calculating the eigenvalues by the usual recipe, you will find complex numbers. A complex eigenvalue is meaningless in a real vector space.

Although we will not prove the following result, it follows from the facts we have introduced and no extra machinery. It is of interest being completely general and valid for both real and complex vector spaces:

Theorem: Let $T \in \mathcal{L}(V)$ and assume $\lambda_1, \ldots \lambda_n$ are distinct eigenvalues of T and $u_1, \ldots u_n$ are corresponding nonzero eigenvectors. Then $(u_1, \ldots u_n)$ are linearly independent.

Note that we cannot ask if the eigenvectors are orthogonal to each other as we have not yet introduced an inner product on the vector space V. In this theorem there may be more than one linearly independent eigenvector associated with some eigenvalues. In that case any one eigenvector will do. Since an n-dimensional vector space V does not have more than n linearly independent vectors, no linear operator on V can have more than n distinct eigenvalues.

We saw that some linear operators in real vector spaces can fail to have eigenvalues. Complex vector spaces are nicer. In fact, *every linear operator on a finite-dimensional complex vector space has at least one eigenvalue.* This is a fundamental result. It can be proven without using determinants with an elegant argument, but the proof using determinants is quite short.

When λ is an eigenvalue, we have seen that $T - \lambda I$ is not an invertible operator. This also means that using any basis, the matrix representative of $T - \lambda I$ is non-invertible. The condition of non-invertibility of a matrix is identical to the condition that its determinant vanish:

$$\det(T - \lambda \mathbf{1}) \;=\; 0\,. \tag{3.54}$$

This condition, in an N-dimensional vector space looks like

$$\det\begin{pmatrix} T_{11}-\lambda & T_{12} & \ldots & T_{1N} \\ T_{21} & T_{22}-\lambda & \ldots & T_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ T_{N1} & T_{N2} & \ldots & T_{NN}-\lambda \end{pmatrix} \;=\; 0\,. \tag{3.55}$$

The left-hand side is a polynomial $f(\lambda)$ in λ of degree N called the *characteristic polynomial*:

$$f(\lambda) \;=\; \det(T - \lambda\mathbf{1}) = (-\lambda)^N + b_{N-1}\lambda^{N-1} + \ldots b_1\lambda + b_0\,, \tag{3.56}$$

where the b_i are constants. We are interested in the equation $f(\lambda) = 0$, as this determines all possible eigenvalues. If we are working on real vector spaces, the constants b_i are real but there is no guarantee of real roots for $f(\lambda) = 0$. With complex vector spaces, the constants b_i will be complex, but a complex solution for $f(\lambda) = 0$ always exists. Indeed, over the complex numbers we can factor the polynomial $f(\lambda)$ as follows

$$f(\lambda) \;=\; (-1)^N(\lambda - \lambda_1)(\lambda - \lambda_2)\ldots(\lambda - \lambda_N)\,, \tag{3.57}$$

where the notation does not preclude the possibility that some of the λ_i's may be equal. The λ_i's are the eigenvalues, since they lead to $f(\lambda) = 0$ for $\lambda = \lambda_i$. If all eigenvalues of T are different the spectrum of T is said to be *non-degenerate*. If an eigenvalue appears k times it is said to be a degenerate eigenvalue with of multiplicity k. Even in the most degenerate case we must have at least one eigenvalue. The eigenvectors exist because $(T - \lambda I)$ non-invertible means it is not injective, and therefore there are nonzero vectors that are mapped to zero by this operator.

4 Inner products

We have been able to go a long way without introducing extra structure on the vector spaces. We have considered linear operators, matrix representations, traces, invariant subspaces, eigenvalues and eigenvectors. It is now time to put some additional structure on the vector spaces. In this section we consider a function called an *inner product* that allows us to construct numbers from vectors. A vector space equipped with an inner product is called an inner-product space.

An **inner product** on a vector space V over $\mathbb{F}$ is a machine that takes an *ordered* pair of elements of V, that is, a first vector and a second vector, and yields a number in $\mathbb{F}$. In order to motivate the definition of an inner product we first discuss the familiar way in which we associate a length to a vector.

The length of a vector, or **norm** of a vector is a real number that is positive or zero, if the vector is the zero vector. In $\mathbb{R}^n$ a vector $a = (a_1, \ldots a_n)$ has norm $|a|$ defined by

$$|a| \;=\; \sqrt{a_1^2 + \ldots a_n^2} \tag{4.58}$$

Squaring this one may think of $|a|^2$ as the *dot product* of a with a:

$$|a|^2 = a \cdot a = a_1^2 + \ldots a_n^2 \tag{4.59}$$

Based on this the dot product of any two vectors a and b is defined by

$$a \cdot b \;=\; a_1 b_1 + \;\ldots\; + a_n b_n \,. \tag{4.60}$$

If we try to generalize this dot product we may require as needed properties the following

1. $a \cdot a \geq 0$, for all vectors a.

2. $a \cdot a = 0$ if and only if $a = 0$.

3. $a \cdot (b_1 + b_2) \;=\; a \cdot b_1 + \,a \cdot b_2$. Additivity in the second entry.

4. $a \cdot (\alpha\, b) \;=\; \alpha\, a \cdot b$, with α a number.

5. $a \cdot b \;=\; b \cdot a$.

Along with these axioms, the length $|a|$ of a vector a is the positive or zero number defined by relation

$$|a|^2 \;=\; a \cdot a\,. \tag{4.61}$$

These axioms are satisfied by the definition (4.60) but do not require it. A new dot product defined by $a \cdot b = c_1 a_1 b_1 + \ldots + c_n a_n b_n$, with $c_1, \ldots c_n$ positive constants, would do equally well! So whatever can be proven with these axioms holds true not only for the conventional dot product.

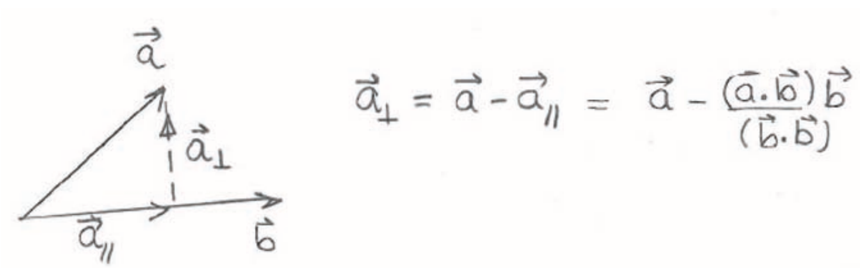
The above axioms guarantee that the Schwarz inequality holds:

$$|a \cdot b\,| \;\leq\; |\,a\,|\,|\,b\,|\,. \tag{4.62}$$

To prove this consider two (nonzero) vectors a and b and then consider the shortest vector joining the tip of a to the line defined by the direction of b (see the figure below). This is the vector $a_\perp$, given by

$$a_\perp \;\equiv\; a \;-\; \frac{a\cdot b}{b\cdot b}\; b\,. \tag{4.63}$$

The subscript $\perp$ is there because the vector is perpendicular to b, namely $a_\perp \cdot b = 0$, as you can quickly see. To write the above vector we subtracted from a the component of a parallel to b. Note that the vector $a_\perp$ is not changed as $b \to cb$; it does not depend on the overall length of b. Moreover, as it should, the vector $a_\perp$ is zero if and only if the vectors a and b are parallel. All this is only motivation, we could have just said "consider the following vector $a_\perp$".



Given axiom (1) we have that $a_\perp \cdot a_\perp \geq 0$ and therefore using (4.63)

$$a_\perp \cdot a_\perp \;=\; a\cdot a - \frac{(a\cdot b)^2}{b\cdot b} \;\geq\; 0\,. \tag{4.64}$$

Since b is not the zero vector we then have

$$(a\cdot b)^2 \;\leq\; (a\cdot a)(b\cdot b)\,. \tag{4.65}$$

Taking the square root of this relation we obtain the Schwarz inequality (4.62). The inequality becomes an equality only if $a_\perp = 0$ or, as discussed above, when $a = cb$ with c a real constant.

For complex vector spaces some modification is necessary. Recall that the length $|\gamma|$ of a complex number γ is given by $|\gamma| = \sqrt{\gamma^*\gamma}$, where the asterisk superscript denotes complex conjugation. It is not hard to generalize this a bit. Let $z = (z_1, \ldots, z_n)$ be a vector in $\mathbb{C}^n$. Then the length of the vector $|z|$ is a real number greater than zero given by

$$|z| \; = \; \sqrt{z_1^* z_1 + \ldots + z_n^* z_n}\,. \tag{4.66}$$

We must use complex conjugates, denoted by the asterisk superscript, to produce a real number greater than or equal to zero. Squaring this we have

$$|z|^2 \; = \; z_1^* z_1 + \ldots + z_n^* z_n\,. \tag{4.67}$$

This suggests that for vectors $z = (z_1, \ldots, z_n)$ and $w = (w_1, \ldots, w_n)$ an inner product could be given by

$$w_1^* z_1 + \ldots + w_n^* z_n\,, \tag{4.68}$$

and we see that we are not treating the two vectors in an equivalent way. There is the first vector, in this case w whose components are conjugated and a second vector z whose components are not conjugated. If the order of vectors is reversed, we get for the inner product the complex conjugate of the original value. As it was mentioned at the beginning of the section, the inner product requires an ordered pair of vectors. It certainly does for complex vector spaces. Moreover, one can define an inner product in general in a way that applies both to complex and real vector spaces.

An **inner product** on a vector space V over $\mathbb{F}$ is a map from an ordered pair (u, v) of vectors in V to a number $\langle u, v\rangle$ in $\mathbb{F}$. The axioms for $\langle u, v\rangle$ are inspired by the axioms we listed for the dot product.

1. $\langle v\,, v\rangle \geq 0$, for all vectors $v \in V$.
2. $\langle v, v\rangle = 0$ if and only if $v = 0$.
3. $\langle u\,, v_1 + v_2\rangle \; = \; \langle u\,, v_1\rangle + \langle u\,, v_2\rangle$. Additivity in the second entry.
4. $\langle u\,, \alpha\, v\rangle \; = \; \alpha\,\langle u\,, v\rangle$, with $\alpha \in \mathbb{F}$. Homogeneity in the second entry.
5. $\langle u\,, v\rangle \; = \; \langle v\,, u\rangle^*$. Conjugate exchange symmetry.

This time the **norm** $|v|$ of a vector $v \in V$ is the positive or zero number defined by relation

$$|v|^2 \; = \; \langle v\,, v\rangle\,. \tag{4.69}$$

From the axioms above, the only major difference is in number five, where we find that the inner product is not symmetric. We know what complex conjugation is in $\mathbb{C}$. For the above axioms to apply to vector spaces over $\mathbb{R}$ we just define the obvious: complex conjugation of a real number is a real number. In a real vector space the $*$ conjugation does nothing and the inner product is strictly symmetric in its inputs.

A few comments. One can use (3) with $v_2 = 0$ to show that $\langle u, 0\rangle = 0$ for all $u \in V$, and thus, by (5) also $\langle 0, u\rangle = 0$. Properties (3) and (4) amount to full linearity in the second entry. It is important to note that additivity holds for the first entry as well:

$$\begin{aligned}\langle u_1 + u_2, v\rangle &= \langle v, u_1 + u_2\rangle^* \\ &= (\langle v, u_1\rangle + \langle v, u_2\rangle)^* \\ &= \langle v, u_1\rangle^* + \langle v, u_2\rangle^* \\ &= \langle u_1, v\rangle + \langle u_2, v\rangle\,.\end{aligned} \tag{4.70}$$

Homogeneity works differently on the first entry, however,

$$\begin{aligned}\langle \alpha\, u\,,\, v\rangle &= \langle v\,, \alpha\, u\rangle^* \\ &= (\alpha\langle v\,, u\rangle)^* \\ &= \alpha^*\,\langle u\,,\, v\rangle\,.\end{aligned} \tag{4.71}$$

Thus we get **conjugate homogeneity** on the first entry. This is a very important fact. Of course, for a real vector space conjugate homogeneity is the same as just plain homogeneity.

Two vectors $u, v \in V$ are said to be **orthogonal** if $\langle u, v\rangle = 0$. This, of course, means that $\langle v, u\rangle = 0$ as well. The zero vector is orthogonal to all vectors (including itself). Any vector orthogonal to all vectors in the vector space must be equal to zero. Indeed, if $x \in V$ is such that $\langle x, v\rangle = 0$ for all v, pick $v = x$, so that $\langle x, x\rangle = 0$ implies $x = 0$ by axiom 2. This property is sometimes stated as the **non-degeneracy** of the inner product. The "Pythagorean" identity holds for the norm-squared of orthogonal vectors in an inner-product vector space. As you can quickly verify,

$$|u + v|^2 \;=\; |u|^2 + |v|^2\,, \quad \text{for } u, v \in V\text{, orthogonal vectors.} \tag{4.72}$$

The Schwarz inequality can be proven by an argument fairly analogous to the one we gave above for dot products. The result now reads

$$\boxed{\text{Schwarz Inequality:} \qquad |\langle u\,, v\rangle| \;\leq\; |u|\,|v|\,.} \tag{4.73}$$

The inequality is saturated if and only if one vector is a multiple of the other. Note that in the left-hand side |...| denotes the norm of a complex number and on the right-hand side each |...| denotes the norm of a vector. You will prove this identity in a slightly different way in the homework. You will also consider there the *triangle inequality*

$$|u + v| \leq |u| + |v|\,, \tag{4.74}$$

which is saturated when $u = cv$ for c a real, positive constant. Our definition (4.69) of norm on a vector space V is mathematically sound: a norm is required to satisfy the triangle inequality. Other properties are required: (i) $|v| \geq 0$ for all v, (ii) $|v| = 0$ if and only if $v = 0$, and (iii) $|cv| = |c||a|$ for c some constant. Our norm satisfies all of them.

A complex vector space with an inner product as we have defined is a *Hilbert space* if it is finite dimensional. If the vector space is infinite dimensional, an extra *completeness* requirement must be satisfied for the space to be a Hilbert space: all Cauchy sequences of vectors must converge to vectors in the space. An infinite sequence of vectors v_i, with $i = 1, 2, \ldots, \infty$ is a Cauchy sequence if for any $\epsilon > 0$ there is an N such that $|v_n - v_m| < \epsilon$ whenever $n, m > N$.

5 Orthonormal basis and orthogonal projectors

In an inner-product space we can demand that basis vectors have special properties. A list of vectors is said to be **orthonormal** if all vectors have norm one and are pairwise orthogonal. Consider a list $(e_1, \ldots, e_n)$ of orthonormal vectors in V. Orthonormality means that

$$\langle e_i, e_j \rangle = \delta_{ij} . \tag{5.75}$$

We also have a simple expression for the norm of $a_1 e_1 + \ldots + a_n e_n$, with $a_i \in \mathbb{F}$:

$$\begin{aligned} |a_1 e_1 + \ldots + a_n e_n|^2 &= \big\langle a_1 e_1 + \ldots + a_n e_n \,,\, a_1 e_1 + \ldots + a_n e_n \big\rangle \\ &= \langle a_1 e_1 \,, a_1 e_1 \rangle + \ldots + \langle a_n e_n \,, a_n e_n \rangle \\ &= |a_1|^2 + \ldots + |a_n|^2 . \end{aligned} \tag{5.76}$$

This result implies the somewhat nontrivial fact that *the vectors in any orthonormal list are linearly independent*. Indeed if $a_1 e_1 + \ldots + a_n e_n = 0$ then its norm is zero and so is $|a_1|^2 + \ldots + |a_n|^2$. This implies all $a_i = 0$, thus proving the claim.

An **orthonormal basis** of V is a list of orthonormal vectors that is also a basis for V. Let $(e_1, \ldots, e_n)$ denote an orthonormal basis. Then any vector v can be written as

$$v = a_1 e_1 + \ldots + a_n e_n , \tag{5.77}$$

for some constants a_i that can be calculated as follows

$$\langle e_i, v \rangle = \langle e_i \,, a_i e_i \rangle = a_i , \quad (\,i \text{ not summed}). \tag{5.78}$$

Therefore any vector v can be written as

$$v = \langle e_1, v \rangle \, e_1 + \ldots + \langle e_n \,, v \rangle = \langle e_i \,, v \rangle \, e_i . \tag{5.79}$$

To find an orthonormal basis on an inner product space V we just need to start with a basis and then use an algorithm to turn it into an orthogonal basis. In fact, a little more generally:

Gram-Schmidt: Given a list $(v_1, \ldots, v_n)$ of linearly independent vectors in V one can construct a list $(e_1, \ldots, e_n)$ of orthonormal vectors such that both lists span the same subspace of V.

The Gram-Schmidt algorithm goes as follows. You take e_1 to be v_1, normalized to have unit norm: $e_1 = v_1/|v_1|$. Then take $v_2 + \alpha e_1$ and fix the constant α so that this vector is orthogonal to e_1. The answer is clearly $v_2 - \langle e_1, v_2\rangle e_1$. This vector, normalized by dividing it by its norm, is set equal to e_2. In fact we can write the general vector in a recursive fashion. If we know $e_1, e_2, \ldots, e_{j-1}$, we can write e_j as follows:

$$e_j \;=\; \frac{v_j - \langle e_1, v_j\rangle e_1 - \ldots - \langle e_{j-1}, v_j\rangle e_{j-1}}{|v_j - \langle e_1, v_j\rangle e_1 - \ldots - \langle e_{j-1}, v_j\rangle e_{j-1}|} \tag{5.80}$$

It should be clear to you by inspection that this vector is orthogonal to the vectors e_i with $i < j$ and has unit norm. The Gram-Schmidt procedure is quite practical.

With an inner product we can construct interesting subspaces of a vector space V. Consider a subset U of vectors in V (not necessarily a subspace). Then we can define a subspace $U^\perp$, called the **orthogonal complement** of U as the set of all vectors orthogonal to the vectors in U:

$$U^\perp \;=\; \{v \in V\,|\langle v, u\rangle = 0, \text{ for all } u \in U\}\,. \tag{5.81}$$

This is clearly a subspace of V. When U is a subspace, then U and $U^\perp$ actually give a direct sum decomposition of the full space:

Theorem: If U is a subspace of V, then $V = U \oplus U^\perp$.

Proof: This is a fundamental result and is not hard to prove. Let $(e_1, \ldots e_n)$ be an orthonormal basis for U. We can clearly write any vector v in V as

$$v \;=\; (\langle e_1, v\rangle e_1 + \ldots + \langle e_n, v\rangle e_n\,) \;+\; (\,v - \langle e_1, v\rangle e_1 - \ldots - \langle e_n, v\rangle e_n\,)\,. \tag{5.82}$$

On the right-hand side the first vector in parenthesis is clearly in U as it is written as a linear combination of U basis vectors. The second vector is clearly in $U^\perp$ as one can see that it is orthogonal to any vector in U. To complete the proof one must show that there is no vector except the zero vector in the intersection $U \cap U^\perp$ (recall the comments below (1.5)). Let $v \in U \cap U^\perp$. Then v is in U and in $U^\perp$ so it should satisfy $\langle v, v\rangle = 0$. But then $v = 0$, completing the proof.

Given this decomposition any vector $v \in V$ can be written uniquely as $v = u + w$ where $u \in U$ and $w \in U^\perp$. One can define a linear operator P_U, called the **orthogonal projection** of V onto U, that and that acting on v above gives the vector u. It is clear from this definition that: (i) the range of P_U is U. (ii) the null space of P_U is $U^\perp$, (iii) that P_U is not invertible and, (iv) acting on U, the operator P_U is the identity operator. The formula for the vector u can be read from (5.82)

$$P_U v \;=\; \langle e_1, v\rangle e_1 + \ldots + \langle e_n, v\rangle e_n\,. \tag{5.83}$$

It is a straightforward but a good exercise to verify that this formula is consistent with the fact that acting on U, the operator P_U is the identity operator. Thus if we act twice in succession with P_U on a vector, the second action has no effect as it is already acting on a vector in U. It follows from this that

$$P_U P_U = I\, P_U = P_U \quad \rightarrow \quad \boxed{P_U^2 \;=\; P_U\,.} \tag{5.84}$$

The eigenvalues and eigenvectors of P_U are easy to describe. Since all vectors in U are left invariant by the action of P_U, an orthonormal basis of U provides a set of orthonormal eigenvectors of P all with eigenvalue one. If we choose on $U^\perp$ an orthonormal basis, that basis provides orthonormal eigenvectors of P all with eigenvalue zero.

In fact equation (5.84) implies that the eigenvalues of P_U can only be one or zero. T he eigenvalues of an operator satisfy whatever equation the operator satisfies (as shown by letting the equation act on a presumed eigenvector) thus $\lambda^2 = \lambda$ is needed, and this gives $\lambda(\lambda - 1) = 0$, and $\lambda = 0, 1$, as the only possibilities.

Consider a vector space $V = U \oplus U^\perp$ that is $(n+k)$-dimensional, where U is n-dimensional and $U^\perp$ is k-dimensional. Let $(e_1, \ldots, e_n)$ be an orthonormal basis for U and $(f_1, \ldots f_k)$ an orthonormal basis for $U^\perp$. We then see that the list of vectors $(g_1, \ldots g_{n+k})$ defined by

$$(g_1\,, \ldots, g_{n+k}) \;=\; (e_1, \ldots\,, e_n,\; f_1, \ldots f_k) \;\; \text{is an orthonormal basis for } V. \tag{5.85}$$

Exercise: Use $P_U e_i = e_i$, for $i = 1, \ldots n$ and $P_U f_i = 0$, for $i = 1, \ldots, k$, to show that in the above basis the projector operator is represented by the diagonal matrix:

$$P_U \;=\; \text{diag}\big(\, \underbrace{1, \ldots 1}_{n \text{ entries}}\,,\; \underbrace{0, \ldots, 0}_{k \text{ entries}}\,\big)\,. \tag{5.86}$$

We see that, as expected from its non-invertibility, $\det(P_U) = 0$. But more interestingly we see that the trace of the matrix P_U is n. Therefore

$$\text{tr}\, P_U \;=\; \dim U\,. \tag{5.87}$$

The dimension of U is the **rank** of the projector P_U. Rank one projectors are the most common projectors. They project to one-dimensional subspaces of the vector space.

Projection operators are useful in quantum mechanics, where observables are described by operators. The effect of measuring an observable on a physical state vector is to turn this original vector instantaneously into another vector. This resulting vector is the orthogonal projection of the original vector down to some eigenspace of the operator associated with the observable.

6 Linear functionals and adjoint operators

When we consider a linear operator T on a vector space V that has an inner product, we can construct a related linear operator $T^\dagger$ on V called the **adjoint** of T. This is a very useful operator and is typically different from T. When the adjoint $T^\dagger$ happens to be equal to T, the operator is said to be *Hermitian.* To understand adjoints, we first need to develop the concept of a linear functional.

A **linear functional** ϕ on the vector space V is a linear map from V to the numbers $\mathbb{F}$: for $v \in V$, $\phi(v) \in \mathbb{F}$. A linear functional has the following two properties:

1. $\phi(v_1 + v_2) = \phi(v_1) + \phi(v_2)\,$, with $v_1, v_2 \in V$.

2. $\phi(av) = a\phi(v)$ for $v \in V$ and $a \in \mathbb{F}$.

As an example, consider the three-dimensional real vector space $\mathbb{R}^3$ with inner product equal to the familiar dot product. Writing a vector v as the triplet $v = (v_1, v_2, v_3)$, we take

$$\phi(v) \;=\; 3v_1 + 2v_2 - 4v_3\,. \tag{6.1}$$

Linearity is clear as the right-hand side features the components v_1, v_2, v_3 appearing linearly. We can use a vector $u = (3, 2, -4)$ to write the linear functional as an inner product. Indeed, one can readily see that

$$\phi(v) \;=\; \langle u, v\rangle. \tag{6.2}$$

This is no accident, in fact. We can prove that any linear functional $\phi(v)$ admits such representation with some suitable choice of vector u.

Theorem: Let ϕ be a linear functional on V. There is a unique vector $u \in V$ such that $\phi(v) = \langle u, v\rangle$ for all $v \in V$.
Proof: Consider an orthonormal basis, $(e_1, \ldots, e_n)$ and write the vector v as

$$v = \langle e_1, v\rangle e_1 + \ldots + \langle e_n, v\rangle e_n\,. \tag{6.3}$$

When ϕ acts on v we find, first by linearity and then by conjugate homogeneity

$$\begin{aligned}
\phi(v) &= \phi\big(\,\langle e_1, v\rangle e_1 + \ldots + \langle e_n, v\rangle e_n\big) \\
&= \langle e_1, v\rangle \phi(e_1) + \ldots + \langle e_n, v\rangle \phi(e_n) \\
&= \langle \phi(e_1)^* e_1, v\rangle + \ldots + \langle \phi(e_n)^* e_n\,, v\rangle \\
&= \big\langle \phi(e_1)^* e_1 + \ldots + \phi(e_n)^* e_n\,, v\big\rangle\,.
\end{aligned} \tag{6.4}$$

We have thus shown that, as claimed

$$\phi(v) = \langle u, v\rangle \quad \text{with} \quad u = \phi(e_1)^* e_1 + \ldots + \phi(e_n)^* e_n\,. \tag{6.5}$$

Next, we prove that this u is unique. If there exists another vector, u', that also gives the correct result for all v, then $\langle u', v\rangle = \langle u, v\rangle$, which implies $\langle u - u', v\rangle = 0$ for all v. Taking $v = u' - u$, we see that this shows $u' - u = 0$ or $u' = u$, proving uniqueness.[1]

We can modify a bit the notation when needed, to write

$$\phi_u(v) \;\equiv \langle u, v\rangle\,, \tag{6.6}$$

where the left-hand side makes it clear that this is a functional acting on v that depends on u.

We can now address the construction of the adjoint. Consider: $\phi(v) = \langle u, T\,v\rangle$, which is clearly a linear functional, whatever the operator T is. Since any linear functional can be written as $\langle w, v\rangle$, with some suitable vector w, we write

$$\langle u,\, Tv\rangle \;=\; \langle w\,, v\rangle\,, \tag{6.7}$$

[1] This theorem holds for infinite dimensional Hilbert spaces, for *continuous* linear functionals.

Of course, the vector w must depend on the vector u that appears on the left-hand side. Moreover, it must have something to do with the operator T, who does not appear anymore on the right-hand side. So we must look for some good notation here. We can think of w as a function of the vector u and thus write $w = T^\dagger u$ where $T^\dagger$ denotes a map (not obviously linear) from V to V. So, we think of $T^\dagger u$ as the vector obtained by acting with some function $T^\dagger$ on u. The above equation is written as

$$\langle u\,, Tv\rangle \;=\; \langle T^\dagger u\,, v\rangle\,, \tag{6.8}$$

Our next step is to show that, in fact, $T^\dagger$ is a linear operator on V. The operator $T^\dagger$ is called the **adjoint** of T. Consider

$$\langle u_1 + u_2, Tv\rangle \;=\; \langle T^\dagger(u_1+u_2), v\rangle\,, \tag{6.9}$$

and work on the left-hand side to get

$$\begin{aligned}\langle u_1 + u_2, Tv\rangle \;&=\; \langle u_1, Tv\rangle + \langle u_2, Tv\rangle \\ &=\; \langle T^\dagger u_1, v\rangle + \langle T^\dagger u_2, v\rangle \\ &=\; \big\langle T^\dagger u_1 + T^\dagger u_2\,, v\big\rangle\,.\end{aligned} \tag{6.10}$$

Comparing the right-hand sides of the last two equations we get the desired:

$$T^\dagger(u_1+u_2) \;=\; T^\dagger u_1 + T^\dagger u_2\,. \tag{6.11}$$

Having established linearity now we establish homogeneity. Consider

$$\langle\, au,\, Tv\rangle \;=\; \langle T^\dagger(au)\,,\, v\rangle\,. \tag{6.12}$$

The left hand side is

$$\langle\, au, Tv\rangle \;=\; a^\star\langle u, Tv\rangle \;=\; a^\star\langle T^\dagger u, v\rangle \;=\; \langle\, aT^\dagger u, v\rangle\,. \tag{6.13}$$

This time we conclude that

$$T^\dagger(au) \;=\; aT^\dagger u\,. \tag{6.14}$$

This concludes the proof that $T^\dagger$, so defined is a linear operator on V.

A couple of important properties are readily proven:

Claim: $(ST)^\dagger = T^\dagger S^\dagger$. We can show this as follows: $\langle u, STv\rangle = \langle S^\dagger u, Tv\rangle = \langle T^\dagger S^\dagger u, v\rangle$.

Claim: The adjoint of the adjoint is the original operator: $(S^\dagger)^\dagger = S$. We can show this as follows: $\langle u, S^\dagger v\rangle = \langle (S^\dagger)^\dagger u, v\rangle$. Now, additionally $\langle u, S^\dagger v\rangle = \langle S^\dagger v, u\rangle^* = \langle v, Su\rangle^* = \langle Su, v\rangle$. Comparing with the first result, we have shown that $(S^\dagger)^\dagger u = Su$, for any u, which proves the claim

Example: Let $v = (v_1, v_2, v_3)$, with $v_i \in \mathbb{C}$ denote a vector in the three-dimensional complex vector space, $\mathbb{C}^3$. Define a linear operator T that acts on v as follows:

$$T(v_1, v_2, v_3) \;=\; (\,0v_1 + 2v_2 + iv_3\,,\, v_1 - iv_2 + 0v_3\,,\, 3iv_1 + v_2 + 7v_3\,)\,. \tag{6.15}$$

Calculate the action of $T^\dagger$ on a vector. Give the matrix representations of T and $T^\dagger$ using the orthonormal basis $e_1 = (1,0,0), e_2 = (0,1,0), e_3 = (0,0,1)$. Assume the inner product is the standard on on $\mathbb{C}^3$.

Solution: We introduce a vector $u = (u_1, u_2, u_3)$ and will use the basic identity $\langle u, Tv\rangle = \langle T^\dagger u, v\rangle$. The left-hand side of the identity gives:

$$\langle\, u, Tv\,\rangle \;=\; u_1^*(2v_2 + iv_3) \;+\; u_2^*(v_1 - iv_2) \;+\; u_3^*(3iv_1 + v_2 + 7v_3)\,. \tag{6.16}$$

This is now rewritten by factoring the various v_i's

$$\langle\, u, Tv\,\rangle \;=\; (u_2^* + 3iu_3^*)v_1 + (2u_1^* - iu_2^* + u_3^*)v_2 + (iu_1^* + 7u_3^*)v_3\,. \tag{6.17}$$

Identifying the right-hand side with $\langle T^\dagger u, v\rangle$ we now deduce that

$$T^\dagger(u_1, u_2, u_3) \;=\; (\,u_2 - 3iu_3\,,\; 2u_1 + iu_2 + u_3\,,\; -iu_1 + 7u_3\,)\,. \tag{6.18}$$

This gives the action of $T^\dagger$. To find the matrix representation we begin with T. Using basis vectors, we have from (6.15)

$$Te_1 = T(1,0,0) = (0,1,3i) = e_2 + 3ie_3 = T_{11}e_1 + T_{21}e_2 + T_{31}e_3\,, \tag{6.19}$$

and deduce that $T_{11} = 0$, $T_{21} = 1$, $T_{31} = 3i$. This can be repeated, and the rule becomes clear quickly: the coefficients of v_i read left to right fit into the i-th column of the matrix. Thus, we have

$$T = \begin{pmatrix} 0 & 2 & i \\ 1 & -i & 0 \\ 3i & 1 & 7 \end{pmatrix} \qquad \text{and} \qquad T^\dagger = \begin{pmatrix} 0 & 1 & -3i \\ 2 & i & 1 \\ -i & 0 & 7 \end{pmatrix}. \tag{6.20}$$

These matrices are related: one is the transpose and complex conjugate of the other! This is not an accident.

Let us reframe this using matrix notation. Let $u = e_i$ and $v = e_j$ where e_i and e_j are orthonormal basis vectors. Then the definition $\langle u, Tv\rangle = \langle T^\dagger u, v\rangle$ can be written as

$$\begin{aligned} \langle T^\dagger e_i, e_j\rangle &= \langle e_i, Te_j\rangle \\ \langle T^\dagger_{ki} e_k, e_j\rangle &= \langle e_i, T_{kj}e_k\rangle \\ (T^\dagger_{ki})^*\delta_{kj} &= T_{jk}\delta_{ik} \\ (T^\dagger)^*_{ji} &= T_{ij} \end{aligned} \tag{6.21}$$

Relabeling i and j and taking the complex conjugate we find the familiar relation between a matrix and its adjoint:

$$(T^\dagger)_{ij} = (T_{ji})^*\,. \tag{6.22}$$

The adjoint matrix is the transpose and complex conjugate matrix only if we use an orthonormal basis. If we did not, in the equation above the use of $\langle e_i, e_j\rangle = \delta_{ij}$ would be replaced by $\langle e_i, e_j\rangle = g_{ij}$, where g_{ij} is some constant matrix that would appear in the rule for the construction of the adjoint matrix.

7 Hermitian and Unitary operators

Before we begin looking at special kinds of operators let us consider a very surprising fact about operators on complex vector spaces, as opposed to operators on real vector spaces.

Suppose we have an operator T that is such that for any vector $v \in V$ the following inner product vanishes

$$\langle v, Tv \rangle = 0 \quad \text{for all } v \in V. \tag{7.23}$$

What can we say about the operator T? The condition states that T is an operator that starting from a vector gives a vector orthogonal to the original one. In a two-dimensional real vector space, this is simply the operator that rotates any vector by ninety degrees! It is quite surprising and important that for *complex* vector spaces the result is very strong: any such operator T necessarily vanishes. This is a theorem:

Theorem: Let T be a linear operator in a **complex vector space** V**:**

$$\boxed{\text{If } \langle v\,, Tv \rangle = 0 \text{ for all } v \in V, \text{ then } T = 0.} \tag{7.24}$$

Proof: Any proof must be such that it fails to work for real vector space. Note that the result follows if we could prove that $\langle u, Tv \rangle = 0$, for all $u, v \in V$. Indeed, if this holds, then take $u = Tv$, then $\langle Tv, Tv \rangle = 0$ for all v implies that $Tv = 0$ for all v and therefore $T = 0$.

We will thus try to show that $\langle u\,, Tv \rangle = 0$ for all $u, v \in V$. All we know is that objects of the form $\langle \#, T\# \rangle$ vanish, whatever # is. So we must aim to form linear combinations of such terms in order to reproduce $\langle u\,, Tv \rangle$. We begin by trying the following

$$\langle u + v, T(u + v) \rangle \; - \; \langle u - v, T(u - v) \rangle \; = \; 2\langle u, Tv \rangle \; + \; 2\langle v, Tu \rangle\,. \tag{7.25}$$

We see that the "diagonal" term vanished, but instead of getting just $\langle u\,, Tv \rangle$ we also got $\langle v\,, Tu \rangle$. Here is where complex numbers help, we can get the same two terms but with opposite signs by trying,

$$\langle u + iv, T(u + iv) \rangle \; - \; \langle u - iv, T(u - iv) \rangle \; = \; 2i\langle u, Tv \rangle \; - \; 2i\,\langle v, Tu \rangle\,. \tag{7.26}$$

It follows from the last two relations that

$$\langle u\,, Tv \rangle \; = \; \frac{1}{4}\Big(\langle u{+}v, T(u{+}v) \rangle - \langle u{-}v, T(u{-}v) \rangle + \frac{1}{i}\langle u{+}iv, T(u{+}iv) \rangle - \frac{1}{i}\langle u{-}iv, T(u{-}iv) \rangle \Big)\,. \tag{7.27}$$

The condition $\langle v, Tv \rangle = 0$ for all v, implies that each term of the above right-hand side vanishes, thus showing that $\langle u\,, Tv \rangle = 0$ for all $u, v \in V$. As explained above this proves the result.

An operator T is said to be **Hermitian** if $T^\dagger = T$. Hermitian operators are pervasive in quantum mechanics. The above theorem in fact helps us discover Hermitian operators. It is familiar that the expectation value of a Hermitian operator, on any state, is real. It is also true, however, that any operator whose expectation value is real for all states must be Hermitian:

$$\boxed{T = T^\dagger \;\text{ if and only if }\; \langle v, Tv\rangle \in \mathbb{R} \text{ for all } v\,.} \tag{7.28}$$

To prove this first go from left to right. If $T = T^\dagger$

$$\langle v, Tv\rangle \;=\; \langle T^\dagger v, v\rangle \;=\; \langle Tv, v\rangle = \langle v, Tv\rangle^*\,, \tag{7.29}$$

showing that $\langle v, Tv\rangle$ is real. To go from right to left first note that the reality condition means that

$$\langle v, Tv\rangle \;=\; \langle Tv, v\rangle \;=\; \langle v, T^\dagger v\rangle\,, \tag{7.30}$$

where the last equality follows because $(T^\dagger)^\dagger = T$. Now the leftmost and rightmost terms can be combined to give $\langle v, (T - T^\dagger)v\rangle = 0$, which holding for all v implies, by the theorem, that $T = T^\dagger$.

We can prove two additional results of Hermitian operators rather easily. We have discussed earlier the fact that on a complex vector space any linear operator has at least one eigenvalue. Here we learn that the eigenvalues of a hermitian operator are real numbers. Moreover, while we have noted that eigenvectors corresponding to different eigenvalues are linearly independent, for Hermitian operators they are guaranteed to be orthogonal. Thus we have the following theorems

Theorem 1: The eigenvalues of Hermitian operators are real.

Theorem 2: Different eigenvalues of a Hermitian operator correspond to orthogonal eigenfunctions.

Proof 1: Let v be a nonzero eigenvector of the Hermitian operator T with eigenvalue λ: $Tv = \lambda v$. Taking the inner product with v we have that

$$\langle v, Tv\rangle = \langle v, \lambda v\rangle = \lambda\langle v, v\rangle\,. \tag{7.31}$$

Since T is hermitian, we can also evaluate $\langle v, Tv\rangle$ as follows

$$\langle v, Tv\rangle = \langle Tv, v\rangle = \langle \lambda v, v\rangle = \lambda^*\,\langle v, v\rangle\,. \tag{7.32}$$

The above equations give $(\lambda - \lambda^*)\langle v, v\rangle = 0$ and since v is not the zero vector, we conclude that $\lambda^* = \lambda$, showing the reality of λ.

Proof 2: Let v_1 and v_2 be eigenvectors of the operator T:

$$Tv_1 = \lambda_1 v_1, \qquad Tv_2 = \lambda_2 v_2\,, \tag{7.33}$$

with λ_1 and λ_2 real (previous theorem) and different from each other. Consider the inner product $\langle v_2, Tv_1\rangle$ and evaluate it in two different ways. First

$$\langle v_2, Tv_1\rangle = \langle v_2, \lambda_1 v_1\rangle = \lambda_1\langle v_2, v_1\rangle\,, \tag{7.34}$$

and second, using hermiticity of T,

$$\langle v_2, Tv_1\rangle = \langle Tv_2, v_1\rangle = \langle \lambda_2 v_2, v_1\rangle = \lambda_2\langle v_2, v_1\rangle\,. \tag{7.35}$$

From these two evaluations we conclude that

$$(\lambda_1 - \lambda_2)\langle v_1, v_2\rangle = 0 \tag{7.36}$$

and the assumption $\lambda_1 \neq \lambda_2$, leads to $\langle v_1, v_2\rangle = 0$, showing the orthogonality of the eigenvectors.

Let us now consider another important class of linear operators on a complex vector space, the so-called unitary operators. An operator $U \in \mathcal{L}(V)$ in a complex vector space V is said to be a **unitary operator** if it is surjective and does not change the magnitude of the vector it acts upon:

$$|Uu| = |u|\,,\ \text{for all } u \in V\,. \tag{7.37}$$

We tailored the definition to be useful even for infinite dimensional spaces. Note that U can only kill vectors of zero length, and since the only such vector is the zero vector, $\text{null}\,U = 0$, and U is injective. Since U is also assumed to be surjective, **a unitary operator U is always invertible.**

A simple example of a unitary operator is the operator λI with λ a complex number of unit-norm: $|\lambda| = 1$. Indeed $|\lambda I u| = |\lambda u| = |\lambda||u| = |u|$ for all u. Moreover, the operator is clearly surjective.

For another useful characterization of unitary operators we begin by squaring (7.37)

$$\langle Uu, Uu\rangle = \langle u, u\rangle \tag{7.38}$$

By the definition of adjoint

$$\langle u, U^\dagger U\, u\rangle = \langle u, u\rangle \quad \to \quad \langle u\,,\, (U^\dagger U - I)u\rangle \;=\; 0 \ \text{ for all } \ u\,. \tag{7.39}$$

So by our theorem $U^\dagger U = I$, and since U is invertible this means $U^\dagger$ is the inverse of U and we also have $UU^\dagger = I$:

$$\boxed{\;U^\dagger U \;=\; UU^\dagger \;=\; I\,.\;} \tag{7.40}$$

Unitary operators *preserve inner products* in the following sense

$$\langle\, Uu\,,\, Uv\rangle \;=\; \langle u\,, v\rangle\,. \tag{7.41}$$

This follows immediately by moving the second U to act on the first input and using $U^\dagger U = I$.

Assume the vector space V is finite dimensional and has an orthonormal basis $(e_1, \ldots e_n)$. Consider the new set of vectors $(f_1, \ldots, f_n)$ where the f's are obtained from the e's by the action of a unitary operator U:

$$f_i \;=\; U\, e_i\,. \tag{7.42}$$

This also means that $e_i = U^\dagger f_i$. We readily see that the f's are also a basis, because they are linearly independent: Acting on $a_1 f_1 + \ldots + a_n f_n = 0$ with $U^\dagger$ we find $a_1 e_1 + \ldots + a_n e_n = 0$, and thus $a_i = 0$. We now see that the new basis is also orthonormal:

$$\langle f_i \,, f_j \rangle \;=\; \langle U e_i \,, U e_j \rangle \;=\; \langle e_i \,, e_j \rangle \;=\; \delta_{ij} \,. \tag{7.43}$$

The matrix elements of U in the e-basis are

$$U_{ki} \;=\; \langle e_k \,, U e_i \rangle \,. \tag{7.44}$$

Let us compute the matrix elements U'_{ki} of U in the f-basis

$$U'_{ki} \;=\; \langle f_k \,, U f_i \rangle \;=\; \langle U e_k \,, U f_i \rangle \;=\; \langle e_k \,, f_i \rangle \;=\; \langle e_k \,, U e_i \rangle \;=\; U_{ki} \tag{7.45}$$

The matrix elements are the same! Can you find an explanation for this result?

8.05 Quantum Physics II
Fall 2013