Chapter 4

Linear algebra II

We introduce inner products as extra structure on real and complex vector spaces. With this inner product the Schwarz inequality holds. We are also able to build orthonormal bases following the Gram-Schmidt procedure and to define orthogonal projectors associated with the direct sum decomposition $V = U \oplus U_\perp$ *of a vector space* V *into orthogonal subspaces. The inner product allows the definition of a new linear operator: the adjoint* $T^\dagger$ *of a linear operator* T*. In an orthonormal basis the matrix for* $T^\dagger$ *is found by complex conjugation and transposition of the matrix for* T*. An operator is Hermitian if it is equal to its adjoint. Unitary operators are invertible operators that preserve the norm of vectors. They satisfy* $UU^\dagger = U^\dagger U = I$*.*

4.1 Inner products

We have been able to go a long way without introducing any additional structure on the vector spaces. We have been able to consider 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. With inner products we can introduce orthonormal bases for vector spaces and the concept of orthogonal projectors. Inner products will also allow us to define the *adjoint* of an operator. With adjoints available, we can define self-adjoint operators, usually called Hermitian operators in physics. We can also define unitary operators.

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, two vectors, and yields a number in $\mathbb{F}$. In order to motivate the definition of an inner product we first discuss the case of real vector spaces and begin by recalling the way in which we associate a length to a vector.

The length of a vector, or **norm** of a vector, is a real non-negative number, equal to 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

$$|a| \;=\; \sqrt{a_1^2 + \ldots a_n^2} \tag{4.1.1}$$

Squaring this we view $|a|^2$ as the *dot product* of a with a:

$$|a|^2 = a \cdot a = a_1^2 + \ldots a_n^2 \tag{4.1.2}$$

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

$$a \cdot b \;=\; a_1 b_1 + \;\ldots\; + a_n b_n\,. \tag{4.1.3}$$

In order to generalize this dot product we require the following properties:

1. $a \cdot a \;\geq 0$, for all vectors a.
2. $a \cdot a = 0$ if and only if $a = 0$.
3. $a \cdot (b_1 + b_2) \;=\; a \cdot b_1 + \, a \cdot b_2$.
4. $a \cdot (\alpha\, b) \;=\; \alpha\, a \cdot b$, with $\alpha \in \mathbb{R}$ 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

$$|a|^2 \;=\; a \cdot a\,. \tag{4.1.4}$$

The third property above is additivity on the second entry. Because of the fifth property, commutativity, additivity also holds for the second entry.
These axioms are satisfied by the definition (4.1.3) 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! (Which axiom goes wrong if we take some c_n equal to zero?). It follows that whatever we can prove with these axioms holds true not only for the conventional dot product but for many others.

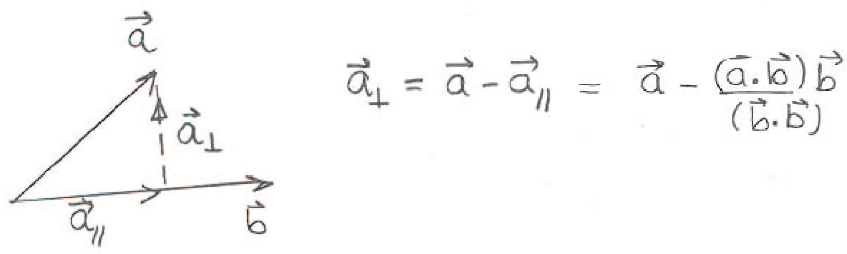
The above axioms guarantee a fundamental result called the Schwarz inequality:

$$|a \cdot b\,| \;\leq\; |\,a\,|\,|\,b\,|\,. \tag{4.1.5}$$

Note that on the left-hand side the bars denote absolute value while on the right-hand side they denote norm. If any of the two vectors is zero the inequality is trivially satisfied. To prove the inequality thus consider two nonzero vectors a and b and then examine the shortest vector joining a point on the line defined by the direction of b to the end of a (see the figure below). This is the vector $a_\perp$, given by

$$a_\perp \;\equiv\; a \;-\; \frac{a \cdot b}{b \cdot b}\; b\,. \tag{4.1.6}$$

The subscript $\perp$ indicates that the vector is perpendicular to b: $a_\perp \cdot b = 0$, 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 $b \to cb$ with c a constant; 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 for the next step, we could have just said "consider the following vector $a_\perp$" and written (4.1.6).



The next step is to apply axiom (1) to the vector $a_\perp$, so that $a_\perp \cdot a_\perp \geq 0$. Using the explicit expression for $a_\perp$ a short computation gives

$$a_\perp \cdot a_\perp \;=\; a \cdot a - \frac{(a\cdot b)^2}{b \cdot b} \;\geq\; 0\,. \tag{4.1.7}$$

Since b is not the zero vector we then have

$$(a \cdot b)^2 \;\leq\; (a\cdot a)(b \cdot b)\,. \tag{4.1.8}$$

Taking the square root of this relation we obtain the Schwarz inequality (4.1.5). 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

$$|z| \;=\; \sqrt{z_1^* z_1 + \ldots + z_n^* z_n}\,. \tag{4.1.9}$$

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

$$|z|^2 \;=\; z_1^* z_1 + \ldots + z_n^* z_n\,. \tag{4.1.10}$$

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

$$\langle w, z\rangle \;=\; w_1^* z_1 + \ldots + w_n^* z_n\,. \tag{4.1.11}$$

Note that we are not treating the two vectors in a symmetric 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, the new inner product is the complex conjugate of the original one. The order of vectors matters for the inner product in complex vector spaces. We can, however, define an inner product in a way that applies both to complex and real vector spaces. Let us do this now!

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.

The **norm** $|v|$ of a vector $v \in V$ is defined by relation

$$|v|^2 \;=\; \langle v\,, v\rangle\,. \tag{4.1.12}$$

Comparing with the dot product axioms for real vector spaces, the only major difference is in item five: the inner product in complex vector spaces is not symmetric. 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.

Let us make a few remarks. One can use (3) with $v_2 = 0$ to show that for all $u \in V$:

$$\langle u, 0\rangle = 0 \quad \rightarrow \quad \langle 0, u\rangle = 0, \tag{4.1.13}$$

where the second equation follows by axiom 5. 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:

$$\begin{aligned} \langle u_1 + u_2, v\rangle \;&= \langle v, u_1 + u_2\rangle^* \\ &= (\langle v, u_1\rangle + \langle v, u_2\rangle)^* \\ &= \langle v, u_1\rangle^* + \langle v, u_2\rangle^* \\ &= \langle u_1, v\rangle + \langle u_2, v\rangle\,. \end{aligned} \tag{4.1.14}$$

Homogeneity works differently on the first entry, however,

$$\begin{aligned} \langle \alpha\, u\,,\, v\rangle \;&= \langle v\,, \alpha\, u\rangle^* \\ &= (\alpha\langle v\,, u\rangle)^* \\ &= \alpha^*\,\langle u\,,\, v\rangle\,. \end{aligned} \tag{4.1.15}$$

In summary, we get **conjugate homogeneity** on the first entry:

$$\boxed{\begin{aligned} \langle u_1 + u_2, v\rangle &= \langle u_1, v\rangle + \langle u_2, v\rangle\,, \\ \langle \alpha\, u\,,\, v\rangle &= \alpha^*\,\langle u\,,\, v\rangle\,. \end{aligned}} \tag{4.1.16}$$

For a real vector space conjugate homogeneity is 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).

The inner product we have defined is **non-degenerate**: 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.

The **Pythagorean** identity holds for the norm-squared of orthogonal vectors in an inner-product vector space. As you can quickly verify,

$$|u + v|^2 \;=\; |u|^2 + |v|^2\,, \quad \text{for } u, v \in V, \text{ orthogonal vectors: } \langle u, v\rangle = 0\,. \tag{4.1.17}$$

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

$$\boxed{\text{Schwarz Inequality:} \qquad |\langle u\,, v\rangle| \;\leq\; |u|\,|v|\,.} \tag{4.1.18}$$

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**:

$$|u + v| \leq |u| + |v|\,, \tag{4.1.19}$$

which is saturated when $u = cv$ for c a real, positive constant. Our definition (4.1.12) 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 finite dimensional complex vector space with an inner product is a **Hilbert space**. An infinite dimensional complex vector space with an inner product is a Hilbert space if an additional *completeness* property holds: 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 $n, m > N$. Our study of quantum mechanics will often involve infinite-dimensional Hilbert spaces. For these familiar spaces the extra condition holds and we will not have to concern ourselves with it.

4.2 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. If $(e_1, \ldots, e_n)$ is a list of orthonormal vectors in V then

$$\langle e_i, e_j \rangle \;=\; \delta_{ij}\,. \tag{4.2.20}$$

We also have a simple expression for the norm of $a_1e_1 + \ldots + a_ne_n$, with $a_i \in \mathbb{F}$:

$$\begin{aligned} |a_1e_1 + \ldots + a_ne_n|^2 \;&= \big\langle a_1e_1 + \ldots + a_ne_n\,,\, a_1e_1 + \ldots + a_ne_n \big\rangle \\ &= \langle a_1e_1\,, a_1e_1\rangle + \ldots + \langle a_ne_n\,, a_ne_n\rangle \\ &= |a_1|^2 + \ldots + |a_n|^2\,. \end{aligned} \tag{4.2.21}$$

This result implies the somewhat nontrivial fact that *the vectors in any orthonormal list are linearly independent*. Indeed if $a_1e_1 + \ldots + a_ne_n = 0$ 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

$$v \;=\; a_1e_1 + \ldots + a_ne_n\,, \tag{4.2.22}$$

for some constants a_i that can be calculated as follows

$$a_i \;=\; \langle e_i, v\rangle\;. \tag{4.2.23}$$

Indeed:

$$\langle e_i, v\rangle = \sum_j \langle e_i, a_je_j\rangle = \sum_j a_j\langle e_i, e_j\rangle = \sum_j a_j\delta_{ij} = a_i\,. \tag{4.2.24}$$

Therefore any vector v can be written as

$$v \;=\; \langle e_1, v\rangle\, e_1 + \;\ldots\; + \langle e_n\,, v\rangle\, e_n \;=\; \sum_i \langle e_i\,, v\rangle\, e_i\,. \tag{4.2.25}$$

To find an orthonormal basis on an inner product space V we can start with any basis and then follow a certain procedure. A little more generally, we have the **Gram-Schmidt** procedure: *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 procedure goes as follows. You take e_1 to be v_1, normalized to have unit norm:

$$e_1 = \frac{v_1}{|v_1|}\,. \tag{4.2.26}$$

Then take

$$f_2 \equiv v_2 + \alpha e_1\,, \tag{4.2.27}$$

and fix the constant α so that this vector is orthogonal to e_1. The answer is clearly

$$f_2 \;=\; v_2 - \langle e_1, v_2\rangle e_1\,. \tag{4.2.28}$$

This vector, normalized by dividing it by its norm, is set equal to e_2, the second vector in our orthonormal list:

$$e_2 \;=\; \frac{v_2 - \langle e_1, v_2\rangle e_1}{|\, v_2 - \langle e_1, v_2\rangle e_1\,|}\,. \tag{4.2.29}$$

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:

$$e_j \;=\; \frac{v_j - \langle e_1, v_j\rangle e_1 - \ldots - \langle e_{j-1}, v_j\rangle e_{j-1}}{|v_j - \langle e_1, v_j\rangle e_1 - \ldots - \langle e_{j-1}, v_j\rangle e_{j-1}|}\,. \tag{4.2.30}$$

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

An inner product can help us construct interesting subspaces of a vector space V. Consider any *subset* U of vectors in V. 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:

$$U^\perp \;=\; \{v \in V\,|\langle v, u\rangle = 0, \text{ for all } u \in U\}\,. \tag{4.2.31}$$

This is clearly a subspace of V. When the set U is itself 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

$$v \;=\; \underbrace{(\langle e_1, v\rangle e_1 + \ldots + \langle e_n, v\rangle e_n\,)}_{\in U} \;+\; \underbrace{(\,v - \langle e_1, v\rangle e_1 - \ldots - \langle e_n, v\rangle e_n\,)}_{\in U_\perp}\,. \tag{4.2.32}$$

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 (3.2.12)). 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

$$v = u + w\,, \qquad \text{with } u \in U \text{ and } w \in U^\perp\,. \tag{4.2.33}$$

One can define a linear operator P_U, called the **orthogonal projection** of V onto U, that acting on v above gives the vector u: $P_U v = u$. It is clear from this definition that:

1. range $P_U = U$. Thus P_U is not surjective.

2. null $P_U = U_\perp$. Thus P_U is not invertible.

3. P_U acting on U is the identity operator. Thus if we act twice with P_U on a vector, the second action has no effect as it is acting on a vector in U. Thus

$$P_U P_U = I\, P_U = P_U \quad \rightarrow \quad \boxed{P_U^2 \;=\; P_U\,.} \tag{4.2.34}$$

4. With $(e_1, \ldots, e_n)$ an orthonormal basis for U then from (4.2.32)

$$P_U v \;=\; \langle e_1, v\rangle e_1 + \ldots + \langle e_n, v\rangle e_n\,. \tag{4.2.35}$$

5. $|P_U v| \leq |v|$. The action of P_U cannot increase the length of a vector. This follows from the decomposition (4.2.33) and the Pythagorean theorem:

$$|v|^2 = |u+w|^2 = |u|^2 + |w|^2 \geq |u|^2 = |P_U v|^2\,, \tag{4.2.36}$$

and taking the square root.

Alternatively, we can prove the following characterization of an orthogonal projector: *P is an orthogonal projection on V if*

$$V = \text{null}\, P \oplus \text{range}\, P\,, \tag{4.2.37}$$

and for all $u \in \text{range}\, P$ and $w \in \text{null}\, P$, we have $\langle w, u\rangle = 0$.
Indeed, if we define $U \equiv \text{range}\, P$ then all we have to show that is that $U_\perp = \text{null}\, P$. It is clear, by assumption, that $\text{null}\, P \subset U_\perp$. Now assume that there is a vector $x \in U_\perp$ that is not contained in null P. Then by the decomposition assumption (4.2.37) we can write $x = x' + x''$ with $x' \in \text{null}\, P$ and $x'' \neq 0$ and $x'' \in \text{range}\, P$. Since $x \in U_\perp$ we have $\langle x, u\rangle = 0$, for all $u \in \text{range}\, P$. Since $x' \in \text{null}\, P$, this implies $\langle x'', u\rangle = 0$, for all $u \in \text{range}\, P$. Given that $x'' \in \text{range}\, P$ this implies $\langle x'', x''\rangle = 0$ and $x'' = 0$, in contradiction with the assumption that there are vectors in $U_\perp$ that are not in null P. This completes the proof.

Comment: For any linear operator $T \in \mathcal{L}(V)$ the dimensions null T and range T add up to the dimension of V. But for an arbitrary T, the spaces null T and range T typically have a non-zero intersection and thus do not provide a direct sum decomposition of V. The T operator in (3.4.35) is an example where the null space and the range are in fact the same!

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

A matrix representation of orthogonal projectors is helpful to understand how the projector encodes the dimension of the space U is projects into. 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 $(e_1, \dots, e_n)$ be an orthonormal basis for U and $(f_1, \dots f_k)$ an orthonormal basis for $U^\perp$. We then see that the list of vectors

$$(e_1, \dots, e_n, f_1, \dots f_k) \text{ is an orthonormal basis for } V. \tag{4.2.38}$$

Since $P_U e_i = e_i$, for $i = 1, \dots n$ and $P_U f_i = 0$, for $i = 1, \dots, k$, a little thought shows that in this basis the projector operator is represented by the diagonal matrix:

$$P_U = \text{diag}(\underbrace{1, \dots 1}_{n \text{ entries}}, \underbrace{0, \dots, 0}_{k \text{ entries}}). \tag{4.2.39}$$

As expected from its non-invertibility, $\det(P_U) = 0$. More interestingly, the trace of the matrix P_U is n. Therefore

$$\boxed{\text{tr}\, P_U = \dim U .} \tag{4.2.40}$$

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 act of measuring an observable projects the physical state vector instantaneously to some invariant subspace of the observable.

4.3 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 generally 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 $a \in \mathbb{F}$.

Example: Consider the real vector space $\mathbb{R}^3$ with inner product equal to the familiar dot product. For any $v = (v_1, v_2, v_3) \in \mathbb{R}^3$, we take

$$\phi(v) = 3v_1 + 2v_2 - 4v_3 . \tag{4.3.1}$$

Linearity follows because the components v_1, v_2, and v_3 appear linearly on the right-hand side. We can actually use the vector $u = (3, 2, -4)$ to write the linear functional as an inner product. Indeed, one can readily see that for any $v \in \mathbb{R}^3$:

$$\phi(v) \;=\; \langle u, v\rangle. \tag{4.3.2}$$

This is no accident; we now prove that any linear functional $\phi(v)$ on a vector space 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

$$v = \langle e_1, v\rangle e_1 + \ldots + \langle e_n, v\rangle e_n\,. \tag{4.3.3}$$

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

$$\begin{aligned}\phi(v) &= \phi\big(\langle e_1, v\rangle e_1 + \ldots + \langle e_n, v\rangle e_n\big)\\ &= \langle e_1, v\rangle \phi(e_1) + \ldots + \langle e_n, v\rangle \phi(e_n)\\ &= \langle \phi(e_1)^* e_1, v\rangle + \ldots + \langle \phi(e_n)^* e_n\,, v\rangle\\ &= \big\langle \phi(e_1)^* e_1 + \ldots + \phi(e_n)^* e_n\,, v\big\rangle\,.\end{aligned} \tag{4.3.4}$$

We have thus shown that, as claimed

$$\phi(v) = \langle u, v\rangle \quad \text{with} \quad u = \phi(e_1)^* e_1 + \ldots + \phi(e_n)^* e_n\,. \tag{4.3.5}$$

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] $\square$

We can make our notation more explicit by writing

$$\phi_u(v) \;\equiv \langle u, v\rangle\,, \tag{4.3.6}$$

where the left-hand side makes it clear that this is a linear functional built from the vector u and the inner product.

We can now address the construction of the adjoint. Consider:

$$\phi(v) = \langle u\,,\, Tv\rangle\,, \tag{4.3.7}$$

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

$$\langle u, \,Tv\rangle \;=\; \langle \#\,, v\rangle\,, \tag{4.3.8}$$

[1] This theorem holds for infinite dimensional Hilbert spaces if we use what are called *continuous* linear functionals.

Of course, the vector # must depend on the vector u that appears on the left-hand side. Moreover, it must have something to do with the operator T, which does not appear on the right-hand side. We can think of # as a function of the vector u and thus write $\# = T^\dagger u$, where $T^\dagger$ denotes a function (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

$$\langle u\,, Tv\rangle \;=\; \langle T^\dagger u\,, v\rangle\,. \tag{4.3.9}$$

Our next step is to show that, in fact, $T^\dagger$ is a linear operator on V. Consider

$$\langle u_1 + u_2, Tv\rangle \;=\; \langle T^\dagger(u_1+u_2), v\rangle\,, \tag{4.3.10}$$

and expand the left-hand side to get

$$\begin{aligned}\langle u_1 + u_2, Tv\rangle \;&=\; \langle u_1, Tv\rangle + \langle u_2, Tv\rangle\\ &=\; \langle T^\dagger u_1, v\rangle + \langle T^\dagger u_2, v\rangle\\ &=\; \big\langle T^\dagger u_1 + T^\dagger u_2\,, v\big\rangle\,.\end{aligned} \tag{4.3.11}$$

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

$$T^\dagger(u_1+u_2) \;=\; T^\dagger u_1 + T^\dagger u_2\,. \tag{4.3.12}$$

Having established linearity now we establish homogeneity. Consider

$$\langle\, au,\, Tv\rangle \;=\; \langle T^\dagger(au)\,,\, v\rangle\,. \tag{4.3.13}$$

The left hand side is

$$\langle\, au, Tv\rangle \;=\; a^*\langle u, Tv\rangle \;=\; a^*\langle T^\dagger u, v\rangle \;=\; \langle\, aT^\dagger u, v\rangle\,. \tag{4.3.14}$$

This time we conclude that

$$T^\dagger(au) \;=\; aT^\dagger u\,. \tag{4.3.15}$$

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

The operator $T^\dagger \in \mathcal{L}(V)$ is called the **adjoint** of T. Its operational definition is the relation

$$\boxed{\langle u\,, Tv\rangle \;=\; \langle T^\dagger u\,, v\rangle\,.} \tag{4.3.16}$$

A couple of important properties are readily proven. The first is

$$(ST)^\dagger = T^\dagger S^\dagger\,. \tag{4.3.17}$$

Proof: By applying (4.3.16) twice $\langle u, STv\rangle = \langle S^\dagger u, Tv\rangle = \langle T^\dagger S^\dagger u, v\rangle$. By definition $\langle u, STv\rangle = \langle (ST)^\dagger u, v\rangle$. Comparison leads to the claim. $\square$

The other property is that the adjoint of the adjoint of an operator is the original operator

$$(S^\dagger)^\dagger = S\,. \tag{4.3.18}$$

Proof. 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.

It is a simple consequence of the above result that $\langle Su, v\rangle = \langle u, S^\dagger v\rangle$, as you can see by reading this equation from right to left. In summary, an operator can be moved from the left input to the right one, and vice versa, by adding a dagger.

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:

$$T(v_1, v_2, v_3) \;=\; (\,2v_2 + iv_3\,,\, v_1 - iv_2\,,\, 3iv_1 + v_2 + 7v_3\,)\,. \tag{4.3.19}$$

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 one on $\mathbb{C}^3$.

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

$$\langle\, Tu, v\,\rangle \;=\; (2u_2 + iu_3)^* v_1 \;+\; (u_1 - iu_2)^* v_2 \;+\; (3iu_1 + u_2 + 7u_3)^* v_3\,. \tag{4.3.20}$$

To identify this with $\langle u, T^\dagger v\rangle$ we rewrite the right-hand side factoring the various u_i^*'s

$$\langle u, T^\dagger v\rangle \;=\; u_1^*(v_2 - 3iv_3) + u_2^*(2v_1 + iv_2 + v_3) + u_3^*(-iv_1 + 7v_3) \tag{4.3.21}$$

We can therefore read the desired action of $T^\dagger$:

$$T^\dagger(v_1, v_2, v_3) \;=\; (\,v_2 - 3iv_3\,,\, 2v_1 + iv_2 + v_3\,,\, -iv_1 + 7v_3\,)\,. \tag{4.3.22}$$

To find the matrix representation we begin with T. Using basis vectors, (4.3.19) gives

$$Te_1 = T(1,0,0) = (0,1,3i) = e_2 + 3ie_3 = T_{11}e_1 + T_{21}e_2 + T_{31}e_3\,, \tag{4.3.23}$$

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 on the right-hand side of (4.3.19) fit into the i-th column of the matrix. Thus, we have

$$T = \begin{pmatrix} 0 & 2 & i \\ 1 & -i & 0 \\ 3i & 1 & 7 \end{pmatrix} \qquad \text{and} \qquad T^\dagger = \begin{pmatrix} 0 & 1 & -3i \\ 2 & i & 1 \\ -i & 0 & 7 \end{pmatrix}\,. \tag{4.3.24}$$

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

Let us calculate adjoints 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 (with repeated indices summed) as

$$\begin{aligned} \langle T^\dagger e_i, e_j\rangle &= \langle e_i, Te_j\rangle \\ \langle T^\dagger_{ki} e_k, e_j\rangle &= \langle e_i, T_{kj} e_k\rangle \\ (T^\dagger_{ki})^* \delta_{kj} &= T_{jk}\delta_{ik} \\ (T^\dagger)^*_{ji} &= T_{ij} \end{aligned} \tag{4.3.25}$$

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

$$\boxed{\text{In an orthonormal basis:} \quad (T^\dagger)_{ij} = (T_{ji})^*\,.} \tag{4.3.26}$$

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

4.4 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. Surprising, when you compare with 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

$$\langle\, v, Tv\rangle = 0 \quad \text{for all } v \in V. \tag{4.4.27}$$

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 *any such operator* T *necessarily vanishes.* This is a theorem:

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

$$\boxed{\text{If } \langle v\,, Tv\rangle = 0 \text{ for all } v \in V, \text{ then } T = 0.} \tag{4.4.28}$$

Proof: Any proof must be such that it fails to work for a 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

$$\langle u+v, T(u+v)\rangle \;-\; \langle u-v, T(u-v)\rangle \;=\; 2\langle u, Tv\rangle \;+\; 2\langle v, Tu\rangle\,. \tag{4.4.29}$$

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,

$$\langle u+iv, T(u+iv)\rangle \;-\; \langle u-iv, T(u-iv)\rangle \;=\; 2i\langle u, Tv\rangle \;-\; 2i\,\langle v, Tu\rangle\,. \tag{4.4.30}$$

It follows from the last two relations that

$$\langle u\,, Tv\rangle \;=\; \tfrac{1}{4}\Big(\langle u{+}v, T(u{+}v)\rangle - \langle u{-}v, T(u{-}v)\rangle + \frac{1}{i}\langle u{+}iv, T(u{+}iv)\rangle - \frac{1}{i}\langle u{-}iv, T(u{-}iv)\rangle\Big)\,. \tag{4.4.31}$$

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. You may know 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:

$$\boxed{\;T = T^\dagger \;\text{ if and only if }\; \langle v, Tv\rangle \in \mathbb{R} \text{ for all}\, v\,.\;} \tag{4.4.32}$$

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

$$\langle v, Tv\rangle \;=\; \langle T^\dagger v, v\rangle \;=\; \langle Tv, v\rangle = \langle v, Tv\rangle^*\,, \tag{4.4.33}$$

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

$$\langle v, Tv\rangle \;=\; \langle Tv, v\rangle \;=\; \langle v, T^\dagger v\rangle\,, \tag{4.4.34}$$

where the last equality, read from right to left, 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 quickly prove two additional results of Hermitian operators. 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.

Proof: 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

$$\langle v, Tv\rangle = \langle v, \lambda v\rangle = \lambda\langle v, v\rangle \,. \tag{4.4.35}$$

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

$$\langle v, Tv\rangle = \langle Tv, v\rangle = \langle \lambda v, v\rangle = \lambda^* \langle v, v\rangle \,. \tag{4.4.36}$$

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 thatλ is real. □

Theorem 2: Eigenvectors of a Hermitian operator associated with different eigenvalues are orthogonal.

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

$$Tv_1 = \lambda_1 v_1, \qquad Tv_2 = \lambda_2 v_2 \,, \tag{4.4.37}$$

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 by following the direction of the arrows emanating from the central term.

$$\lambda_2\langle v_2, v_1\rangle = \langle \lambda_2 v_2, v_1\rangle = \langle Tv_2, v_1\rangle \underset{\longleftarrow}{=} \langle v_2, Tv_1\rangle \underset{\longrightarrow}{=} \langle v_2, \lambda_1 v_1\rangle = \lambda_1\langle v_2, v_1\rangle \,, \tag{4.4.38}$$

Going left we used hermicity. Equating the left-most and right-most terms we conclude that

$$(\lambda_1 - \lambda_2)\langle v_1, v_2\rangle = 0 \,. \tag{4.4.39}$$

The assumption $\lambda_1 \neq \lambda_2$ leads to $\langle v_1, v_2\rangle = 0$, as desired. □

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:

$$|Uu| = |u| \,, \text{ for all } u \in V \,. \tag{4.4.40}$$

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.**

Example: The operator λI with λ a complex number of unit-norm is unitary. Indeed, with $|\lambda| = 1$ we have $|\lambda I u| = |\lambda u| = |\lambda||u| = |u|$ for all u. Moreover, the operator is clearly surjective since for any $v \in V$ we have $v = (\lambda I)\frac{1}{\lambda}v$.

For another useful characterization of unitary operators we begin by squaring (4.4.40)

$$\langle Uu, Uu\rangle = \langle u, u\rangle \tag{4.4.41}$$

By the definition of adjoint

$$\langle u, U^\dagger U\, u\rangle = \langle u, u\rangle \quad \to \quad \langle u\,,\, (U^\dagger U - I)u\rangle \;=\; 0 \;\text{ for all }\; u\,. \tag{4.4.42}$$

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 $UU^\dagger = I$:

$$\boxed{\;U^\dagger U \;=\; UU^\dagger \;=\; I\,.\;} \tag{4.4.43}$$

Any operator U that obeys these identities is unitary. Unitary operators *preserve inner products* in the following sense

$$\langle\, Uu\,,\, Uv\rangle \;=\; \langle\, u\,, v\rangle\,. \tag{4.4.44}$$

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 another 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:

$$f_i \;=\; U\, e_i\,. \tag{4.4.45}$$

This also means that $e_i = U^\dagger f_i$. The new vectors are also orthonormal:

$$\langle f_i\,, f_j\rangle \;=\; \langle\, Ue_i\,,\, Ue_j\rangle \;=\; \langle e_i\,,\, e_j\rangle \;=\; \delta_{ij}\,. \tag{4.4.46}$$

They are linearly independent because any list of orthonormal vectors is linearly independent. They span V because $\dim V$ linearly independent vectors span V. Thus the f_i's are an orthonormal *basis*. This is an important result: *the action of a unitary operator on an orthonormal basis gives us another orthonormal basis.*

The matrix elements of U in the e-basis are easily calculated:

$$\langle e_k\,, Ue_i\rangle \;=\; \langle e_k\,, \sum_p U_{pi} e_p\rangle \;=\; \sum_p U_{pi}\langle e_k\,, e_p\rangle \;=\; U_{ki}\,. \tag{4.4.47}$$

This is a nice formula to remember, valid for the matrix elements of *any* operator $T \in \mathcal{L}(V)$ in an orthonormal basis:

$$\boxed{\;T_{ki} \;=\; \langle e_k,\, Te_i\rangle\,.\;} \tag{4.4.48}$$

We can compute the matrix elements U'_{ki} of U in the f-basis using this formula

$$U'_{ki} \;=\; \langle f_k\,, Uf_i\rangle \;=\; \langle Ue_k\,, Uf_i\rangle \;=\; \langle e_k\,, f_i\rangle \;=\; \langle e_k\,, Ue_i\rangle \;=\; U_{ki}\,. \tag{4.4.49}$$

The matrix elements are the same! Do you remember when we saw this first?