Math 350 Fall 2011
Notes about inner product spaces

In this notes we state and prove some important properties of inner product spaces.

First, recall the dot product on $\mathbb{R}^n$: if $x, y \in \mathbb{R}^n$, say $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$, then their dot product is given by

$$x \cdot y = x_1 y_1 + \cdots + x_n y_n.$$

The dot product is very useful in understanding geometry in $\mathbb{R}^n$. For instance, one can compute the length of a vector, the distance between two points, and the angle between two vectors using the dot product. What we will do now can be thought of as an effort to generalize the dot product to an abstract vector space. The generalization will be called an *inner product*.

1. Real inner product spaces

Suppose V is a vector space over $\mathbb{R}$. A (real) *inner product* on V is a map $\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{R}$ (i.e. an association of a number $\langle v, w \rangle \in \mathbb{R}$ for every $v, w \in V$) that satisfies the following axioms:

(i) $\langle v, v \rangle \geq 0$ for all $v \in V$;
(ii) $\langle v, v \rangle = 0$ if and only if $v = 0$;
(iii) $\langle v_1 + v_2, w \rangle = \langle v_1, w \rangle + \langle v_2, w \rangle$ for all $v_1, v_2, w \in V$, and $\langle cv, w \rangle = c\langle v, w \rangle$ for all $v, w \in V$ and all $c \in \mathbb{R}$;
(iv) $\langle v, w \rangle = \langle w, v \rangle$ for all $v, w \in V$.

A vector space over $\mathbb{R}$, with a real inner product, is called a *real inner product space*.

For instance, the dot product on $\mathbb{R}^n$ is a real inner product on $\mathbb{R}^n$. (Check!) In fact one should think of an inner product as a generalization of the dot product.

Natural examples of real inner products arise in the study of abstract vector spaces over $\mathbb{R}$. For instance, if $V = C[a, b]$, where $C[a, b]$ is the vector space over $\mathbb{R}$ of all real-valued continuous functions defined on an interval $[a, b]$, then a real inner product on V is defined by

$$\langle f, g \rangle = \int_a^b f(x) g(x) dx \quad \text{for all } f, g \in V.$$

More generally, if h is a (fixed) positive continuous function on $[a, b]$, then one can define a real inner product on $C[a, b]$ by

$$\langle f, g \rangle = \int_a^b f(x) g(x) h(x) dx \quad \text{for all } f, g \in C[a, b].$$

(So there are in general many possible inner products on a given vector space; one should always specify what inner product one uses unless there is a standard one and that it is understood that the standard one is being used.)

Also, if $V = M_{n\times n}(\mathbb{R})$, the vector space of all $n \times n$ matrices with real entries, then a real inner product on V can be defined by

$$\langle A, B\rangle = \text{Trace}(B^t A).$$

One should check that the above are examples of real inner products by using the axioms.

We mention some properties of real inner products. Suppose now V is a real inner product space, with inner product $\langle \cdot, \cdot\rangle$.

1. If $w \in V$ is fixed, and one defines a map $T: V \to \mathbb{R}$ such that $T(v) = \langle v, w\rangle$, then T is a linear map. This is evident from Axiom (iii) above.

2. We have

$$\langle w, v_1 + v_2\rangle = \langle w, v_1\rangle + \langle w, v_2\rangle \quad \text{for all } v_1, v_2, w \in V,$$

and

$$\langle w, cv\rangle = c\langle w, v\rangle \quad \text{for all } v, w \in V \text{ and } c \in \mathbb{R}.$$

This is because $\langle w, v_1+v_2\rangle = \langle v_1+v_2, w\rangle$ by Axiom (iv); then one applies Axiom (iii), and use Axiom (iv) again to conclude that $\langle w, v_1+v_2\rangle = \langle v_1, w\rangle+\langle v_2, w\rangle = \langle w, v_1\rangle+\langle w, v_2\rangle$ as desired. Similarly one can prove the assertion about $\langle w, cv\rangle$. It follows that if $w \in V$ is fixed, then the map $S: V \to \mathbb{R}$, defined by $S(v) = \langle w, v\rangle$, is also a linear map.

3. It follows from the above two properties that $\langle 0, w\rangle = \langle w, 0\rangle = 0$ for all $w \in V$. One way of seeing this is to quote the theorem that says every linear map applied to the zero vector yields zero; another way is to repeat that proof, and observe for instance that $\langle 0, w\rangle = \langle 0 + 0, w\rangle = \langle 0, w\rangle + \langle 0, w\rangle = 2\langle 0, w\rangle$, which implies $\langle 0, w\rangle = 0$.

4. If one wants to show that a vector $v \in V$ is zero, one just need to show that $\langle v, v\rangle = 0$ by Axiom (ii).

5. Alternatively, we have the following little, but remarkably useful, lemma:

   Lemma 1. *Suppose $v \in V$ is such that $\langle v, w\rangle = 0$ for all $w \in V$. Then $v = 0$.*

   The proof is one line: just take $w = v$. Then $\langle v, v\rangle = 0$, which implies $v = 0$ as we have just observed.

6. Recall that two vectors in $\mathbb{R}^n$ are said to be orthogonal to each other if their dot product is zero. This suggests the following terminology in our setting: Two vectors $v, w \in V$ are said to be *orthogonal* to each other if and only if $\langle v, w\rangle = 0$. A way of remembering the above lemma is then to say that the zero vector is the only vector that is orthogonal to every vector in V.

2. Complex inner product spaces

Next, we look at the situation where the scalars are complex numbers rather than real numbers. The definition of an inner product in this setting is very similar to the one we had when the vector spaces are over $\mathbb{R}$. However, as we will point out below, there will be some subtle (but important) differences, and one should always be careful about these.

First, let's review the standard 'dot product' in $\mathbb{C}^n$. If we denote this 'dot product' by $(\cdot,\cdot)$, then

$$(z,w) = z_1\overline{w_1} + \dots + z_n\overline{w_n}$$

if $z = (z_1, \dots, z_n)$ and $w = (w_1, \dots, w_n)$. Note that one has the complex conjugates over the w_j's. The complex conjugates were there because one would like to have $(z,z) = \|z\|^2 \geq 0$ for all $z \in \mathbb{C}^n$, where $\|z\|$ is the Euclidean norm of z, defined by

$$\|z\| = (|z_1|^2 + \dots + |z_n|)^{1/2};$$

if the complex conjugates were absent from the definition of (z,w), then it would no longer be true that $(z,z) = \|z\|^2$. On the other hand, the presence of the complex conjugates over the w_j's leads to the following fact: if $z, w \in \mathbb{C}^n$ and if $c \in \mathbb{C}$, then

$$(z, cw) = \overline{c}(z,w).$$

In other words, one cannot simply pull a scalar out of the second argument of this 'dot product'; one has to put a complex conjugate over the scalar if one pulls it out from the second argument of the 'dot product'. On the other hand, it is still true that

$$(cz, w) = c(z,w) \quad \text{for all } z, w \in \mathbb{C}^n \text{ and all } c \in \mathbb{C}.$$

Thus the first and second argument of the 'dot product' in $\mathbb{C}^n$ behaves slightly differently. In general, if V is a vector space over $\mathbb{C}$, when we define a complex inner product on V, one would like to build this difference into our definition. Thus we are led to the following.

Suppose now V is a vector space over $\mathbb{C}$. A (complex) inner product on V is a map $\langle\cdot,\cdot\rangle \colon V \times V \to \mathbb{C}$ (i.e. an association of a number $\langle v, w\rangle \in \mathbb{C}$ for every $v, w \in V$) that satisfies the following axioms:

(i) $\langle v, v\rangle \geq 0$ for all $v \in V$ (in particular $\langle v, v\rangle$ is real for every $v \in V$);
(ii) $\langle v, v\rangle = 0$ if and only if $v = 0$;
(iii) $\langle v_1 + v_2, w\rangle = \langle v_1, w\rangle + \langle v_2, w\rangle$ for all $v_1, v_2, w \in V$, and $\langle cv, w\rangle = c\langle v, w\rangle$ for all $v, w \in V$ and all $c \in \mathbb{C}$;
(iv) $\langle v, w\rangle = \overline{\langle w, v\rangle}$ for all $v, w \in V$.

A vector space over $\mathbb{C}$, with a complex inner product, is called a *complex inner product space*.

For instance, the 'dot product' on $\mathbb{C}^n$ is a complex inner product on $\mathbb{C}^n$. (Check! For instance, Axiom (iv) can be checked by beginning with the right hand side of the identity, and observing that

$$\overline{(w,z)} = \overline{w_1\overline{z_1} + \dots + w_n\overline{z_n}} = \overline{w_1}z_1 + \dots + \overline{w_n}z_n = z_1\overline{w_1} + \dots + z_n\overline{w_n} = (z,w)$$

for all $z, w \in \mathbb{C}^n$.)

Before we turn to more examples of complex inner products, let's make some observations first. Suppose now V is a complex inner product space, with inner product $\langle \cdot, \cdot \rangle$.

1. If $w \in V$ is fixed, and one defines a map $T \colon V \to \mathbb{C}$ such that $T(v) = \langle v, w \rangle$, then T is a linear map. This is evident from Axiom (iii) above.

2. We have
$$\langle w, v_1 + v_2 \rangle = \langle w, v_1 \rangle + \langle w, v_2 \rangle \quad \text{for all } v_1, v_2, w \in V,$$
and
$$\langle w, cv \rangle = \overline{c} \langle w, v \rangle \quad \text{for all } v, w \in V \text{ and } c \in \mathbb{C}.$$
These statements are very similar to the corresponding statements we had about real inner products, except for the complex conjugate over the c in the second identity. In fact, one can prove the first statement about $\langle w, v_1 + v_2 \rangle$ in exactly the same way we did for the real inner product; we leave that to the reader. On the other hand, suppose $v, w \in V$ and $c \in \mathbb{C}$. Then we have
$$\langle w, cv \rangle = \overline{\langle cv, w \rangle} = \overline{c \langle v, w \rangle} = \overline{c} \overline{\langle v, w \rangle} = \overline{c} \langle w, v \rangle,$$
proving the second identity. It follows that if $w \in V$ is fixed, then the map $S \colon V \to \mathbb{C}$, defined by $S(v) = \langle w, v \rangle$, is NOT a linear map, contrary to the case of real inner products. (It is sometimes called a *conjugate linear* map, because $S(v_1 + v_2) = S(v_1) + S(v_2)$ and $S(cv) = \overline{c} S(v)$, but we will not need this piece of terminology.)

3. It follows from the above two properties that $\langle 0, w \rangle = \langle w, 0 \rangle = 0$ for all $w \in V$. One way of seeing this is to quote the theorem that says every linear map applied to the zero vector yields zero; thus $\langle 0, w \rangle = 0$, from which it follows that $\langle w, 0 \rangle = \overline{\langle 0, w \rangle} = \overline{0} = 0$. Another way is to do this directly, and observe for instance that $\langle w, 0 \rangle = \langle w, 0 + 0 \rangle = \langle w, 0 \rangle + \langle w, 0 \rangle = 2 \langle w, 0 \rangle$, which implies $\langle w, 0 \rangle = 0$.

The next three properties works exactly the same way they did in a real inner product space:

4. If one wants to show that a vector $v \in V$ is zero, one just need to show that $\langle v, v \rangle = 0$ by Axiom (ii).

5. Alternatively, again we have the following little, but remarkably useful, lemma:

   Lemma 2. *Suppose $v \in V$ is such that $\langle v, w \rangle = 0$ for all $w \in V$. Then $v = 0$.*

   The proof is the same as in the case of real inner product spaces: just take $w = v$. Then $\langle v, v \rangle = 0$, which implies $v = 0$ as we have just observed.

6. Recall that two vectors in $\mathbb{C}^n$ are said to be orthogonal to each other if their 'dot product' is zero. This suggests the following terminology in our setting: Two vectors $v, w \in V$ are said to be *orthogonal* to each other if and only if $\langle v, w \rangle = 0$.

A way of remembering the above lemma is then to say that the zero vector is the only vector that is orthogonal to every vector in V.

Now we turn to some natural examples of complex inner product spaces. The first example is the following. Suppose V is the vector space over $\mathbb{C}$ of all complex-valued continuous functions defined on an interval $[a,b]$, then a complex inner product on V is defined by

$$\langle f, g\rangle = \int_a^b f(x)\overline{g(x)}dx \quad \text{for all } f, g \in V.$$

More generally, if h is a (fixed) positive continuous function on $[a,b]$, then one can define a complex inner product on V by

$$\langle f, g\rangle = \int_a^b f(x)\overline{g(x)}h(x)dx \quad \text{for all } f, g \in V.$$

Also, if $V = M_{n\times n}(\mathbb{C})$, the vector space of all $n \times n$ matrices with complex entries, then a complex inner product on V can be defined by

$$\langle A, B\rangle = \text{Trace}(B^*A).$$

Here B^* is the conjugate transpose of B; in other words, B^* is what one obtains if one first take the transpose of B, and then put complex conjugates over every entry entry of the resulting matrix.

One should check that the above are examples of complex inner products by using the axioms. Note that the complex conjugate is always somehow involved in the second argument of the complex inner product in these examples.

A piece of notation: Many definitions and results below hold for both real and complex inner product spaces. If we do not want to emphasize whether it is a real or complex inner product space, we just say inner product space. If there is a difference between real and complex inner product spaces, we will say so explicitly. All inner product spaces we have will be over $\mathbb{R}$ or $\mathbb{C}$, and on $\mathbb{R}^n$ and $\mathbb{C}^n$, we will always use the dot product as our (standard) inner product unless stated otherwise.

3. Norms, Distances and Angles

If V is a inner product space, with inner product $\langle\cdot,\cdot\rangle$, then the *length* of a vector v in V is by definition

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

For instance, if $V = \mathbb{R}^n$ with the standard inner product, and if $v \in \mathbb{R}^n$, then $\|v\|$ is the usual Euclidean norm of the vector v. Similarly if $V = \mathbb{C}^n$ with the standard inner product.

In general, if V is a inner product space, with inner product $\langle\cdot,\cdot\rangle$, and if $\|\cdot\|$ is the norm associated to the inner product, then we have:

(i) $\|v\| \geq 0$ for all $v \in V$;

(ii) $\|v\| = 0$ if and only if $v = 0$;
(iii) $\|cv\| = |c|\|v\|$ for all $v \in V$ and for all scalars c;
(iv) For any $v, w \in V$, we have

$$\|v + w\| \leq \|v\| + \|w\|. \tag{1}$$

The first three properties are very easy to verify using the definition of $\|\cdot\|$, and we leave it to the reader. The last one says the length of the sum of two vectors is less than or equal to the sum of the lengths of the two vectors. In $\mathbb{R}^n$ or $\mathbb{C}^n$ this is commonly known as the *triangle inequality*, because it asserts that the length of any side of a triangle cannot exceed the sum of the lengths of the other two sides. We will adopt the same terminology, and call (iv) above the *triangle inequality*.

The proof of the triangle inequality makes use of the following *Cauchy-Schwarz* inequality:

$$|\langle v, w\rangle| \leq \|v\|\|w\| \quad \text{for all } v, w \in V. \tag{2}$$

We defer the proofs of these until a point where we would have a better conceptual understanding of the proof of these inequalities. (c.f. Section 8.)

A function $\|\cdot\| \colon V \to \mathbb{R}$ satisfying Properties (i) to (iv) above are usually called a *norm* on V. It is not a linear function since in general $\|v + w\| \neq \|v\| + \|w\|$. The above says that if V is an inner product space, then the inner product on V defines a natural norm on V. The study of a general real or complex vector space with a norm is an interesting topic in itself, but we will not go into that.

Now given two points $v, w \in V$, we define the *distance* between v and w to be

$$d(v, w) = \|v - w\|.$$

This definition makes sense whenever one has a norm on V, but the reader may feel free to assume that the norm arises from an inner product on V just for the sake of concreteness. It follows from the Properties (i) to (iv) of the norm above that

(i) $d(v, w) \geq 0$ for all $v, w \in V$;
(ii) $d(v, w) = 0$ if and only if $v = w$;
(iii) $d(v, w) = d(w, v)$ for all $v, w \in V$;
(iv) $d(u, w) \leq d(u, v) + d(v, w)$ for all $u, v, w \in V$.

A function $d(\cdot,\cdot) \colon V \times V \to \mathbb{R}$ satisfying these properties is usually called a *metric* on V. The above says that a norm on a vector space V will naturally define a metric on V.

Example 1. *Suppose $V = C[0, 1]$, the vector space of all continuous real-valued functions on $[0, 1]$. Define an inner product on V by*

$$\langle f, g\rangle = \int_0^1 f(x)g(x)dx$$

for all $f, g \in V$. Define the norm $\|\cdot\|$ and the distance using the inner product. Compute $\langle x, e^{-2x}\rangle$, $\|x\|$, $\|e^{-2x}\|$ and check that $|\langle x, e^{-2x}\rangle| \leq \|x\|\|e^{-2x}\|$ in this case. Also, compute the distance between x and e^{-2x}.

Recall that on $\mathbb{R}^n$, if $\theta \in [0, \pi]$ is the angle between two non-zero vectors v and w, then

$$\langle v, w \rangle = \|v\| \|w\| \cos \theta.$$

It follows that

$$\theta = \cos^{-1} \left(\frac{\langle v, w \rangle}{\|v\| \|w\|} \right).$$

Motivated by this, we make the following definition.

Suppose V is a real inner product space, with inner product $\langle \cdot, \cdot \rangle$. Suppose $\| \cdot \|$ is the norm associated to the given inner product. For any non-zero vectors $v, w \in V$, we define the *angle* between v and w to be

$$\cos^{-1} \left(\frac{\langle v, w \rangle}{\|v\| \|w\|} \right).$$

This defines an angle in $[0, \pi]$ since we are taking the arc-cosine of a real number between -1 and 1, thanks to the Cauchy-Schwarz inequality (2). We then recover the identity $\langle v, w \rangle = \|v\| \|w\| \cos(\theta)$ in any real inner product space, if v, w are any non-zero vectors in V and θ is the angle between them.

Example 2. *Following the notations of Example 1 above, compute the angle between x and e^{-2x}.*

Recall that two vectors v, w in an inner product space are said to be orthogonal to each other if and only if $\langle v, w \rangle = 0$. It follows that if v, w are non-zero vectors in a real inner product space, then they are orthogonal to each other if and only if the angle between them is $\pi/2$. This is in turn the motivation for the definition of orthogonality.

From now on, $\| \cdot \|$ will always denote the norm defined by an inner product.

4. Orthogonal and orthonormal sets

Suppose now V is an inner product space (real or complex). A set of vectors $\{v_1, \dots, v_k\}$ in V is said to be *orthogonal* if $\langle v_i, v_j \rangle = 0$ for all $i \neq j$. It follows that $\{v_1, \dots, v_k\}$ is orthogonal, if and only if any two distinct vectors from the set is orthogonal to each other.

Furthermore, a set of vectors $\{v_1, \dots, v_k\}$ in V is said to be *orthonormal*, if first it is orthogonal, and second the length of every vector in the set is equal to 1.

For example, the standard basis in $\mathbb{R}^n$ or $\mathbb{C}^n$ are both orthogonal and orthonormal; the set $\{(1, 1), (1, -1)\}$ in $\mathbb{R}^2$ is orthogonal but not orthonormal.

Orthogonal sets of vectors enjoy a lot of good properties. For example, one has the Pythagoras theorem:

Proposition 1. *If $v, w \in V$ and $\{v, w\}$ is orthogonal, then*

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2.$$

In fact, to prove this, one simply observes that

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

and the last expression in this identity is $\|v\|^2 + \|w\|^2$ since $\langle v, w\rangle = 0 = \langle w, v\rangle$. More generally, we have:

Proposition 2. *If the set $\{v_1, \dots, v_k\}$ is orthogonal, then*

$$\|v_1 + \cdots + v_k\|^2 = \|v_1\|^2 + \cdots + \|v_k\|^2.$$

In particular, if the set $\{v_1, \dots, v_k\}$ is orthonormal, and if $v = a_1 v_1 + \cdots + a_k v_k$, then

$$\|v\|^2 = |a_1|^2 + \cdots + |a_k|^2.$$

Next, we have:

Proposition 3. *Suppose $S = \{v_1, \dots, v_k\}$ is an orthonormal set of vectors. Suppose v is in the span of S. We claim that*

$$v = \langle v, v_1\rangle v_1 + \cdots + \langle v, v_k\rangle v_k$$

and

$$\|v\|^2 = |\langle v, v_1\rangle|^2 + \cdots + |\langle v, v_k\rangle|^2.$$

In fact, any v in the span of S can be written as $v = a_1 v_1 + \cdots + a_k v_k$ for some scalars $a_1, \dots, a_k$. By taking the inner product of this equation with v_j, we get

$$\langle v, v_j\rangle = \langle a_1 v_1 + \cdots + a_k v_k, v_j\rangle = a_1\langle v_1, v_j\rangle + \cdots + a_k\langle v_k, v_j\rangle.$$

Since $\langle v_i, v_j\rangle = 0$ whenever $i \neq j$, and $\langle v_j, v_j\rangle = \|v_j\|^2 = 1^2 = 1$, we get $\langle v, v_j\rangle = a_j$, and by plugging this back into the equation $v = a_1 v_1 + \cdots + a_k v_k$, we obtain the first conclusion of the proposition.

The second conclusion in the proposition follows from the first one and the previous proposition. The second conclusion is sometimes called *Parseval's identity.*

Corollary 1. *Suppose V is an inner product space, and $\{v_1, \dots, v_n\}$ is an orthonormal set that forms a basis of V. Then for any $v \in V$, we have*

$$v = \langle v, v_1\rangle v_1 + \cdots + \langle v, v_n\rangle v_n$$

and

$$\|v\|^2 = |\langle v, v_1\rangle|^2 + \cdots + |\langle v, v_n\rangle|^2.$$

This is because then any $v \in V$ is in the span of $\{v_1, \dots, v_n\}$, and we can apply the previous proposition.

A basis of V that forms an orthonormal set is called an *orthonormal basis* of V. Thus by this corollary, in an inner product space, it is particularly easy to express a vector in terms an orthonormal basis.

Since it is advantageous to work with orthonormal basis, a natural question is then the following: when is an orthonormal set in V a basis of V? We have the following proposition:

Proposition 4. *If* $\{v_1, \dots, v_k\}$ *is orthonormal, then it is linearly independent.*

In fact, suppose $\{v_1, \dots, v_k\}$ is orthonormal. If $a_1, \dots, a_k$ are scalars such that

$$a_1 v_1 + \cdots + a_k v_k = 0,$$

then taking inner product of this with v_1, we get $a_1 = 0$; similarly, taking inner product with v_2, we get $a_2 = 0$. In fact, by taking inner product with v_j, we get $a_j = 0$ for all $j = 1, \dots, k$. Thus $\{v_1, \dots, v_k\}$ is linearly independent.

We then have the following corollaries:

Corollary 2. *If* $\{v_1, \dots, v_k\}$ *is orthonormal and spans* V*, then* $\{v_1, \dots, v_k\}$ *is a basis of* V*.*

Corollary 3. *If* $\{v_1, \dots, v_k\}$ *is orthonormal and* $\dim(V) = k$*, then* $\{v_1, \dots, v_k\}$ *is a basis of* V*.*

Example 3. *(a) Show that* $\{(1, 0, 1, 0), (0, -1, 0, -1), (1, 0, -1, 0)\}$ *is an orthogonal set of vectors in* $\mathbb{R}^4$*.*
(b) Let W *be the span of these vectors. Find an orthonormal basis of* W*.*

Example 4. *Suppose* $W = P_2(\mathbb{R})$*, the vector space of all polynomials of degree* ≤ 2 *with real coefficients. Define an inner product on* W *by*

$$\langle p, q \rangle = \frac{1}{2} \int_{-1}^{1} p(x)q(x)dx.$$

Let $u_1 = 1$*,* $u_2 = \sqrt{3}x$*, and* $u_3 = \sqrt{5}(3x^2 - 1)/2$*. Show that* $\{u_1, u_2, u_3\}$ *is an orthonormal basis of* W*.*

It will be shown that every finite dimensional inner product space admits an orthonormal basis. To do that, we need to first study orthogonal projections.

5. Orthogonal projections

Suppose V is an inner product space over a field F, where $F = \mathbb{R}$ or $\mathbb{C}$. Suppose $v, w \in V$ and $w \neq 0$. We define a function from F to $\mathbb{R}$ by

$$c \mapsto \|v - cw\|,$$

and ask when this function achieves its minimum as c varies over F. The answer is given by the following proposition:

Proposition 5. *Suppose* $v, w \in V$ *and* $w \neq 0$*. Then the function* $c \mapsto \|v - cw\|$ *achieves a global minimum at only one point, namely at*

$$c = \frac{\langle v, w \rangle}{\langle w, w \rangle}. \tag{3}$$

Proof. Since the minimum of $\|v - cw\|$ is achieved precisely when the minimum of $\|v - cw\|^2$ is achieved, we look at the latter function instead. Now suppose first $F = \mathbb{R}$. Then

$$\|v - cw\|^2 = \langle v - cw, v - cw \rangle = \langle v, v \rangle + 2c\langle v, w \rangle + c^2\langle w, w \rangle$$

is a quadratic polynomial in a real variable c, and the coefficient of c^2 is positive. Thus $\|v - cw\|^2$ achieves a global minimum; in fact

$$\frac{d}{dc}\|v - cw\|^2 = \frac{d}{dc}(\langle v, v\rangle - 2c\langle v, w\rangle + c^2\langle w, w\rangle) = -2\langle v, w\rangle + 2c\langle w, w\rangle,$$

so setting this equal zero, we see that the minimum of $\|v - cw\|^2$ occurs precisely when

$$-2\langle v, w\rangle + 2c\langle w, w\rangle = 0,$$

i.e. when

$$c = \frac{\langle v, w\rangle}{\langle w, w\rangle}.$$

The case when $F = \mathbb{C}$ can be dealt with similarly, except now one has a minimization problem involving one complex variable (or two real variables). We leave the details to the reader. □

Next, we observe that (3) holds if and only if $\langle v - cw, w\rangle = 0$, i.e. if and only if $v - cw$ is orthogonal to w. Motivated by this, we call cw the *orthogonal projection* of v onto w, if and only if

$$c = \frac{\langle v, w\rangle}{\langle w, w\rangle}.$$

The above proposition then says the orthogonal projection of v onto w is the closest point to v on the span of w.

More generally, suppose V is an inner product space (real or complex), and suppose W is a finite dimensional subspace of V.

Proposition 6. *For any $v \in V$, there exists one and only one vector $\pi(v) \in W$ such that $\langle v - \pi(v), w\rangle = 0$ for all $w \in W$. If $\{w_1, \dots, w_k\}$ is orthogonal and form a basis of W, then $\pi(v)$ is given by*

$$\pi(v) = c_1 w_1 + \cdots + c_k w_k$$

where $c_1, \dots, c_k$ are scalars such that

$$c_1 = \frac{\langle v, w_1\rangle}{\langle w_1, w_1\rangle}, \quad \dots, \quad c_k = \frac{\langle v, w_k\rangle}{\langle w_k, w_k\rangle}.$$

In addition, the function from W to $\mathbb{R}$, defined by $w \mapsto \|v - w\|$, achieves a global minimum at only one point, namely when $w = \pi(v)$.

We will call this $\pi(v)$ the *orthogonal projection* of v onto W. In particular, if $\{w_1, \dots, w_k\}$ is an orthonormal basis of W, then the orthogonal projection of v onto W is given by

$$\langle v, w_1\rangle w_1 + \cdots + \langle v, w_k\rangle w_k.$$

Proof of Proposition 6. Suppose $v \in V$ is given, and $\{w_1, \dots, w_k\}$ is an orthogonal set that forms a basis of W. To prove the uniqueness assertion in the proposition, suppose v_0 is in W and $\langle v - v_0, w\rangle = 0$ for all $w \in W$. We will determine v_0 explicitly. Since $\{w_1, \dots, w_k\}$ is a basis of W, we can write $v_0 = c_1 w_1 + \cdots + c_k w_k$

for some scalars $c_1, \dots, c_k$. Now by assumption, $\langle v - v_0, w\rangle = 0$ for all $w \in W$. This implies

$$\langle v - (c_1 w_1 + \cdots + c_k w_k), w_j\rangle = 0 \quad \text{for all } j = 1, \dots, k.$$

It follows that

$$\langle v, w_j\rangle - c_j\langle w_j, w_j\rangle = 0 \quad \text{for all } j = 1, \dots, k.$$

Thus

$$c_j = \frac{\langle v, w_j\rangle}{\langle w_j, w_j\rangle} \quad \text{for all } j = 1, \dots, k.$$

This proves that v_0 is uniquely determined, and proves the uniqueness assertion in the proposition.

On the other hand, if $v \in V$ is given, and if one defines $\pi(v)$ to be $c_1 w_1 + \cdots + c_k w_k$, where

$$c_j = \frac{\langle v, w_j\rangle}{\langle w_j, w_j\rangle} \quad \text{for all } j = 1, \dots, k,$$

then it is easy to show that $\langle v - \pi(v), w_j\rangle = 0$ for all $j = 1, \dots, k$. Since $\{w_1, \dots, w_k\}$ is a basis of W, this shows that $\langle v - \pi(v), w\rangle = 0$ for all $w \in W$. This proves the existence part of the proposition, and proves the given formula for $\pi(v)$.

A multivariable minimization gives the last part of the proposition, which we omit. □

Example 5. *Recall that in Example 3, W is the subspace of $\mathbb{R}^4$ spanned by the orthogonal set $\{(1,0,1,0), (0,-1,0,-1), (1,0,-1,0)\}$. Compute the orthogonal projection of $(1,2,3,4)$ onto W.*

Example 6. *Let $V = P_4(\mathbb{R})$, the vector space of all polynomials of degree ≤ 4 with real coefficients. Define an inner product on V by*

$$\langle p, q\rangle = \frac{1}{2}\int_{-1}^{1} p(x)q(x)dx.$$

Recall that in Example 4, if $W = P_2(\mathbb{R})$ is a subspace of V, then $u_1 = 1$, $u_2 = \sqrt{3}x$, and $u_3 = \sqrt{5}(3x^2 - 1)/2$ form an orthonormal basis of W. If $p(x) = x^4 \in V$, find the orthogonal projection of $p(x)$ onto W.

Corollary 4 (Bessel's inequality)**.** *Suppose V is an inner product space (real or complex), and that $\{v_1, \dots, v_k\}$ is an orthonormal set in V. Then for any $v \in V$, we have*

$$\|v\|^2 \geq |\langle v, v_1\rangle|^2 + \cdots + |\langle v, v_k\rangle|^2,$$

with equality holding if and only if v is in the span of $\{v_1, \dots, v_k\}$.

Proof. In fact, if W is the span of $\{v_1, \dots, v_k\}$, and w is the orthogonal projection of v onto W, then

$$\langle v - w, w\rangle = 0,$$

which implies by the Pythagoras theorem (Proposition 1) that

$$\|v\|^2 = \|v - w\|^2 + \|w\|^2 \geq \|w\|^2.$$

But

$$w = \langle v, v_1\rangle v_1 + \cdots + \langle v, v_k\rangle v_k,$$

and thus by Pythagoras theorem again (c.f. Proposition 2),

$$\|w\|^2 = |\langle v, v_1\rangle|^2 + \cdots + |\langle v, v_k\rangle|^2.$$

Together this proves the desired inequality.

The equality holds in the inequality if and only if $\|v - w\|^2 = 0$, i.e. if and only if $v = w$. This is equivalent to saying that v is in W, i.e. that v is the span of $\{v_1, \dots, v_k\}$. □

6. Orthogonal complements

Suppose V is an inner product space. If W is a subspace of V, the orthogonal complement of W in V is the set of all vectors in V that are orthogonal to every vector in W, and is usually denoted as $W^\perp$. In other words,

$$W^\perp = \{v \in V \colon \langle v, w\rangle = 0 \text{ for all } w \in W.\}$$

Example 7. *Find the orthogonal complement of the span of* $\{(1,0,i,1),(0,1,1,-i)\}$ *in* $\mathbb{C}^4$.

Example 8. *Suppose* $V = C[a,b]$ *is equipped with the inner product* $\langle f, g\rangle = \int_a^b f(x)g(x)dx$ *as in Section 1. If* W *is the vector subspace of all constant functions in* V*, show that the orthogonal complement of* W *consists of all functions* $f \in C[a,b]$ *such that* $\int_a^b f(x)dx = 0$.

Suppose now W is a finite dimensional subspace of V. It then follows from the previous section that if $v \in V$ and $\pi(v)$ is the orthogonal projection of v onto W, then $v - \pi(v) \in W^\perp$. In particular, every vector $v \in V$ can be written as the sum of a vector in W and a vector in $W^\perp$; in fact,

$$v = \pi(v) + (v - \pi(v)),$$

where $\pi(v) \in W$ and $v - \pi(v) \in W^\perp$. The intersection of W with $W^\perp$ is $\{0\}$. Thus V is the direct sum of W and $W^\perp$. In particular, if V is finite dimensional, then

$$\dim(V) = \dim(W) + \dim(W^\perp).$$

Example 9. *Suppose* V *is a finite dimensional inner product space (real or complex), and* W *is a subspace of* V*. For every* $v \in V$*, let* $\pi_1(v)$ *be the orthogonal projection of* v *onto* W*, and let* $\pi_2(v)$ *be the orthogonal projection of* v *onto* $W^\perp$*. If* Id *denotes the identity map on* V*, show that*

$$Id = \pi_1 + \pi_2, \quad \textit{and} \quad \pi_1 \circ \pi_2 = 0 = \pi_2 \circ \pi_1.$$

Hence show that

$$\pi_1^2 = \pi_1 \quad \textit{and} \quad \pi_2^2 = \pi_2.$$

Conclude that both π_1 *and* π_2 *are diagonalizable.*

7. The construction of orthogonal sets and the Gram-Schmidt process

We now turn to the construction of orthogonal sets in an inner product space. Suppose $\{v_1, \dots, v_k\}$ is a subset of an inner product space V. Define

$$u_1 = v_1,$$

$$u_2 = v_2 - \text{orthogonal projection of } v_2 \text{ onto the span of } \{u_1\},$$

$$u_3 = v_3 - \text{orthogonal projection of } v_2 \text{ onto the span of } \{u_1, u_2\},$$

etc. In general, for $j = 2, \dots, k$, define u_j to be

$$u_j = v_j - \text{orthogonal projection of } v_j \text{ onto the span of } \{u_1, \dots, u_{j-1}\}.$$

Then it is easy to see that $\langle u_i, u_j \rangle = 0$ for all $i > j$, and that the span of $\{u_1, \dots, u_k\}$ is the same as the span of $\{v_1, \dots, v_k\}$. It follows that

$$u_1 = v_1,$$

$$u_2 = v_2 - \frac{\langle v_2, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1,$$

$$u_3 = v_3 - \frac{\langle v_3, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1 - \frac{\langle v_3, u_2 \rangle}{\langle u_2, u_2 \rangle} u_2,$$

etc, and in general

$$u_j = v_j - \frac{\langle v_j, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1 - \dots - \frac{\langle v_j, u_{j-1} \rangle}{\langle u_{j-1}, u_{j-1} \rangle} u_{j-1}$$

for $j = 2, \dots, k$. The process of constructing $\{u_1, \dots, u_k\}$ from $\{v_1, \dots, v_k\}$ is called the Gram-Schmidt process. It constructs for us an orthogonal set of vectors $\{u_1, \dots, u_k\}$ that has the same span as the given set $\{v_1, \dots, v_k\}$.

Notice that if $\{v_1, \dots, v_k\}$ are linearly independent to begin with, then so is $\{u_1, \dots, u_k\}$. In particular, all of the u_j's are then not equal to zero. By replacing u_j by $u_j / \|u_j\|$, we can obtain an orthonormal set of vectors that has the same span as the given $\{v_1, \dots, v_k\}$. In particular, if V is a finite dimensional inner product space, then V has an orthonormal basis. (Just begin with an arbitrary basis of V and apply Gram-Schmidt.)

Example 10. *Suppose $V = P_4(\mathbb{R})$, the vector space of all real polynomials of degree ≤ 4, equipped with inner product*

$$\langle p(x), q(x) \rangle = \int_0^1 p(x)q(x)dx.$$

Let W be the subspace of V spanned by $\{1, x, x^2\}$. Find an orthonormal basis of W. Also find an orthonormal basis of $W^{\perp}$, the orthogonal complement of W in V.

8. Proof of the Cauchy-Schwarz and Triangle inequality

We finally have the conceptual framework that helps us understand the proof of the Cauchy-Schwarz inequality (2) and the triangle inequality (1).

To prove the Cauchy-Schwarz inequality, suppose V is an inner product space over $\mathbb{C}$. Suppose $v, w \in V$ are given. If $w = 0$, the Cauchy-Schwarz inequality is trivial. So suppose $w \neq 0$. Then we minimize $\|v - cw\|^2$ as c varies over $\mathbb{C}$. The minimum is achieved when

$$c = \frac{\langle v, w\rangle}{\langle w, w\rangle},$$

as we have seen in the section about orthogonal projections; on the other hand, the minimum value of $\|v - cw\|^2$ as c varies is certainly non-negative. Thus we have

$$\begin{aligned}
0 &\leq \|v - cw\|^2 \\
&= \langle v, v\rangle - c\langle w, v\rangle - \overline{c}\langle v, w\rangle + |c|^2\langle w, w\rangle \\
&= \|v\|^2 - 2\mathrm{Re}\,(c\overline{\langle v, w\rangle}) + |c|^2\|w\|^2 \\
&= \|v\|^2 - 2\frac{|\langle v, w\rangle|^2}{\langle w, w\rangle} + \left|\frac{\langle v, w\rangle}{\langle w, w\rangle}\right|^2 \|w\|^2 \\
&= \|v\|^2 - 2\frac{|\langle v, w\rangle|^2}{\|w\|^2} + \frac{|\langle v, w\rangle|^2}{\|w\|^2} \\
&= \|v\|^2 - \frac{|\langle v, w\rangle|^2}{\|w\|^2}.
\end{aligned}$$

The Cauchy-Schwarz inequality (2) then follows. A similar argument works when V is an inner product space over $\mathbb{R}$.

Next, to prove the triangle inequality, observe that

$$\|v+w\|^2 = \langle v+w, v+w\rangle = \langle v, v\rangle + \langle v, w\rangle + \langle w, v\rangle + \langle w, w\rangle = \|v\|^2 + 2\mathrm{Re}\,\langle v, w\rangle + \|w\|^2.$$

Now we invoke Cauchy-Schwarz:

$$2\mathrm{Re}\,\langle v, w\rangle \leq 2|\langle v, w\rangle| \leq 2\|v\|\|w\|.$$

So

$$\|v + w\|^2 \leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 = (\|v\| + \|w\|)^2.$$

The triangle inequality (1) follows.

9. The transpose and the adjoint of a matrix

Next we turn to another set of concepts revolving around the adjoint of a linear operator. To begin with, recall that the *transpose* of a matrix is obtained by replacing its rows by its columns. If A is a matrix, its transpose is usually written as A^t. The *conjugate transpose* of a matrix is what one obtains by transposing the matrix and putting complex conjugates over every entry; if A is a matrix, its conjugate transpose is usually written A^*. The conjugate transpose of a matrix is also sometimes called the *adjoint* of a matrix.

Note that if A is a real (resp. complex) $m \times n$ matrix, one can think of A as a linear map from $\mathbb{R}^n$ (resp. $\mathbb{C}^n$) to $\mathbb{R}^m$ (resp. $\mathbb{C}^m$), and one can think of A^t (resp. A^*) as a linear map from $\mathbb{R}^m$ (resp. $\mathbb{C}^m$) into $\mathbb{R}^n$ (resp. $\mathbb{C}^n$). Also, if A, B are two matrices of the correct sizes, then $(AB)^t = B^t A^t$ and $(AB)^* = B^* A^*$.

If we think of a vector in $\mathbb{R}^n$ as a column vector, then the dot product of x and y in $\mathbb{R}^n$ can be identified with $y^t x$, where $y^t x$ is the matrix product of y^t with x. Similarly, if we think of a vector in $\mathbb{C}^n$ as a column vector, then for $z, w \in \mathbb{C}^n$, the 'dot product' (z, w) can be identified with the matrix product of $w^* z$.

It follows that if A is an $m \times n$ matrix with real entries, then for all $x \in \mathbb{R}^n$ and all $y \in \mathbb{R}^m$, we have

$$\langle Ax, y \rangle = \langle x, A^t y \rangle;$$

here $\langle \cdot, \cdot \rangle$ denote the standard inner products on $\mathbb{R}^m$ and $\mathbb{R}^n$. This can be seen by just noticing that

$$\langle Ax, y \rangle = y^t A x = y^t (A^t)^t x = (A^t y)^t x = \langle x, A^t y \rangle.$$

One also has $\langle y, Ax \rangle = \langle A^t y, x \rangle$ for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$.

Similarly, if A is an $m \times n$ matrix with complex entries, then for all $z \in \mathbb{C}^n$ and all $w \in \mathbb{C}^m$, we have

$$\langle Az, w \rangle = \langle z, A^* w \rangle,$$

as well as $\langle w, Az \rangle = \langle A^* w, z \rangle$.

These are important identities, and should in fact be thought of as *defining* A^t and A^* when A is a given matrix. To illustrate the usefulness of these identities, we prove:

Proposition 7. *If A is a real $m \times n$ matrix, then the kernel of A is the orthogonal complement of the range of A^t in $\mathbb{R}^n$.*

Proof. Simply note that

$$\begin{aligned}
&y \text{ is in the orthogonal complement of the range of } A^t \\
\Leftrightarrow &\langle y, A^t x \rangle = 0 \quad \text{for all } x \in \mathbb{R}^n \\
\Leftrightarrow &\langle Ay, x \rangle = 0 \quad \text{for all } x \in \mathbb{R}^n \\
\Leftrightarrow &Ay = 0 \\
\Leftrightarrow &y \text{ is in the kernel of } A.
\end{aligned}$$

□

Similarly, one can prove that if A is a real $m \times n$ matrix, then the kernel of A^t is the orthogonal complement of the range of A in $\mathbb{R}^m$.

It follows that

1. The range of A is the orthogonal complement of the kernel of A^t in $\mathbb{R}^m$;
2. The range of A^t is the orthogonal complement of the kernel of A in $\mathbb{R}^n$;
3. $\mathbb{R}^n$ is the direct sum of the kernel of A and the range of A^t;

4. $\mathbb{R}^m$ is the direct sum of the range of A and the kernel of A^t.

In particular, the equation $Ax = b$ is solvable, if and only if b is orthogonal to the kernel of A^t.

The reader should formulate the complex version of these statements, when A is a complex $m \times n$ matrix and A^* is its conjugate transpose.

These generalizes to the setting where we have adjoints of linear maps. But before we can define that, we need the following Riesz representation theorem.

10. Riesz representation theorem

Suppose V is a vector space over a field F. Suppose $f\colon V \to F$ is a linear map. Then f is called a *linear functional* on V, and the set of all linear functionals on V is usually denoted as V^*. V^* is commonly known as the *dual space* of V. It is also a vector space, and in fact if V is finite dimensional, then so is V^*, and in fact $\dim(V^*) = \dim(V)$.

Suppose now V is an inner product space over F, where $F = \mathbb{R}$ or $\mathbb{C}$. If $w \in V$, then the map $f\colon V \to F$, defined by

$$f(v) = \langle v, w \rangle,$$

is a linear functional on V. The following Riesz representation theorem says that these are all the examples we have if V is finite dimensional; in fact different w defines different linear maps from V to F.

Theorem 5. *Suppose V is a finite dimensional inner product space over F, where $F = \mathbb{R}$ or $\mathbb{C}$. If $f\colon V \to F$ is a linear map, then there exists an unique $w \in V$ such that*

$$f(v) = \langle v, w \rangle$$

for all $v \in V$.

Proof. We carry out the proof when $F = \mathbb{C}$. The proof when $F = \mathbb{R}$ is an easy modification of this proof and left to the reader.

Suppose $f\colon V \to \mathbb{C}$ is a linear map. We first show the existence of w such that $f(v) = \langle v, w \rangle$ for all $v \in V$. If f is identically zero, then $f(v) = \langle v, 0 \rangle$ for all $v \in V$, so it suffices to take $w = 0$. Otherwise the rank of f is strictly bigger than zero, but since the rank of f is also at most 1 (since f maps V into $\mathbb{C}$ and $\dim(\mathbb{C}) = 1$), we have the rank of $f = 1$. This shows that the kernel of f is an $n-1$ dimensional subspace of V. Let W be the kernel of f. Then the orthogonal complement $W^\perp$ of W in V is one-dimensional. Let w_0 be an unit vector in $W^\perp$, and let $c = \overline{f(w_0)}$. We claim that

$$f(v) = \langle v, cw_0 \rangle$$

for all $v \in V$. In fact, it suffices to show that this is true for all $v \in W$ and for $v = w_0$, since V is the direct sum of W and the subspace spanned by $\{w_0\}$. But if

$v \in W$, then

$$f(v) = 0 = \langle v, cw_0 \rangle$$

by our construction of W and w_0. Also, if $v = w_0$, then

$$f(v) = f(w_0) = \overline{c},$$

and also

$$\langle v, cw_0 \rangle = \langle w_0, cw_0 \rangle = \overline{c}.$$

Thus we have $f(v) = \langle v, cw_0 \rangle$ for all $v \in V$, as desired.

Next, suppose there exists $w_1, w_2 \in V$ such that both $f(v) = \langle v, w_1 \rangle$ and $f(v) = \langle v, w_2 \rangle$ for all $v \in V$. Then

$$\langle v, w_1 \rangle = \langle v, w_2 \rangle$$

for all $v \in V$, i.e.

$$\langle v, w_1 - w_2 \rangle = 0$$

for all $v \in V$. This shows, by Lemma 2, that $w_1 - w_2 = 0$, i.e. $w_1 = w_2$. Thus we have the uniqueness assertion of the theorem. $\square$

11. The transpose and the adjoint of a linear map

We are now in the position to define the transpose and the adjoint of a general linear map. Suppose V, W are finite dimensional inner product spaces over $\mathbb{R}$. If $T\colon V \to W$ is a linear map, one can find a linear map $T^t\colon W \to V$ such that

$$\langle Tv, w \rangle = \langle v, T^t w \rangle \quad \text{for all } v \in V,\ w \in W.$$

This is because if $w \in W$ is given, then the map $v \mapsto \langle Tv, w \rangle$ defines a linear functional on V. Thus by the Riesz representation theorem, there exists a unique element $v_0 \in V$ such that $\langle Tv, w \rangle = \langle v, v_0 \rangle$ for all $v \in V$. This v_0 certainly depends on w, and we set $T^t w = v_0$. This defines a map from W to V, and one can show that the map T^t such defined is linear. We call T^t the *transpose* of the linear map T.

Similarly, if V, W are finite dimensional inner product spaces over $\mathbb{C}$, then for every linear map $T\colon V \to W$, there exists a linear map $T^*\colon W \to V$ such that

$$\langle Tv, w \rangle = \langle v, T^* w \rangle \quad \text{for all } v \in V,\ w \in W.$$

This linear map T^* is called the *adjoint* of the linear map T.

Example 11. *Show that if $T\colon \mathbb{C}^n \to \mathbb{C}^m$ is a linear map given by*

$$T(z) = Az,$$

where A is an $m \times n$ matrix with complex entries and one thinks of $z \in \mathbb{C}^n$ as a column vector, then $T^*\colon \mathbb{C}^m \to \mathbb{C}^n$ is given by

$$T^*(w) = A^* w$$

for all $w \in \mathbb{C}^m$, where A^* is the conjugate transpose of A and we think of w in $\mathbb{C}^m$ as a column vector.

Example 12. *Suppose V is the vector space over $\mathbb{R}$ consisting of the set of all linear combinations of $\sin x$, $\sin(2x)$, $\sin(3x)$ and $\sin(4x)$. We define an inner product on V by*

$$\langle f(x), g(x)\rangle = \int_0^{2\pi} f(x)g(x)dx.$$

Let $T\colon V \to V$ be the linear map defined by $T(f(x)) = f''(x)$ for all $f \in V$. Show that $T^t = T$.

12. The spectral theorem for matrices

The rest of the article is devoted to proving the spectral theorem in the setting of a finite dimensional real or complex inner product space. For the sake of exposition, we will first state the results for matrices, and then state the corresponding results for linear operators.

From now on all matrices will be square matrices (say $n \times n$), and all linear operators will be from a finite dimensional inner product space V into itself. First we recall some definitions.

First we recall some definitions.

An $n \times n$ real matrix P is said to be *orthogonal* if and only if $PP^t = P^tP = I$; here P^t is the transpose of the matrix P. It is known that if P is an $n \times n$ real matrix, then the following statements are equivalent:

(i) P is orthogonal;
(ii) $P^tP = I$;
(iii) the columns of P form an orthonormal basis of $\mathbb{R}^n$;
(iv) $PP^t = I$;
(v) the rows of P form an orthonormal basis of $\mathbb{R}^n$.

An $n \times n$ complex matrix U is said to be *unitary* if and only if $U^*U = UU^* = I$; here U^* is the conjugate transpose (or adjoint) of the matrix U. It is known that if U is an $n \times n$ complex matrix, then the following statements are equivalent:

(i) U is unitary;
(ii) $U^*U = I$;
(iii) the columns of U form an orthogonal basis of $\mathbb{C}^n$;
(iv) $UU^* = I$;
(v) the rows of U form an orthogonal basis of $\mathbb{C}^n$.

An $n \times n$ real matrix A is said to be *symmetric* if and only if $A^t = A$. An $n \times n$ complex matrix A is said to be *Hermitian* (or *self-adjoint*) if and only if $A^* = A$. An $n \times n$ complex matrix A is said to be *normal* if and only if $AA^* = A^*A = I$.

The first theorem is a characterization of (real) symmetric matrices:

Theorem 6. *Suppose A is an $n \times n$ real matrix. Then the following are equivalent:*

(a) A is symmetric;
(b) $\mathbb{R}^n$ has an orthonormal basis that consists of eigenvectors of A;
(c) there exists an orthogonal matrix P, and a diagonal matrix D, such that $A = PDP^{-1}$.

Note that it follows from the last statement of the above theorem that any symmetric matrix is diagonalizable over $\mathbb{R}$. Also, one should observe that in last statement of the above theorem, one could also have written

$$A = PDP^t,$$

since $P^t = P^{-1}$ when P is an orthogonal matrix.

Similarly, we have the following characterization of Hermitian matrices:

Theorem 7. *Suppose A is an $n \times n$ complex matrix. Then the following are equivalent:*

(a) A is Hermitian;
(b) $\mathbb{C}^n$ has an orthogonal basis that consists of eigenvectors of A, and all eigenvalues of A are real;
(c) there exists an unitary matrix U, and a diagonal matrix D with real entries, such that $A = UDU^{-1}$.

Finally we have the following characterization of normal matrices:

Theorem 8. *Suppose A is an $n \times n$ complex matrix. Then the following are equivalent:*

(a) A is normal;
(b) $\mathbb{C}^n$ has an orthogonal basis that consists of eigenvectors of A;
(c) there exists an unitary matrix U, and a diagonal matrix D, such that $A = UDU^{-1}$.

Again, it follows from the last statement of the two theorems above that any self-adjoint matrix is diagonalizable over $\mathbb{C}$, and so is any normal matrix. Also, one should observe that in last statement of the two theorems above, one could also have written

$$A = UDU^*,$$

since $U^* = U^{-1}$ when U is a unitary matrix.

In each of the three theorems above, the equivalence of (b) and (c) is an easy exercise. So the key is to prove the equivalence between (a) and (b). Before we do that, to put these theorems in perspective, we note the role played by orthogonality here. For instance, Theorem 8 can be seen as a characterization of complex $n \times n$ matrices A for which $\mathbb{C}^n$ admits an *orthonormal* basis consisting of eigenvectors of A. The class of such linear operators turns out to be quite small, as one sees from the theorem; in fact the set of normal matrices is nowhere dense in the set of complex $n \times n$ matrices. On the other hand, if we drop the assumption of orthogonality on the basis of $\mathbb{C}^n$, i.e. if we are interested instead in the set of all complex matrices for which $\mathbb{C}^n$ admits a basis (not necessarily orthonormal) consisting of eigenvectors of A, then we will get the set of all diagonalizable complex $n \times n$ matrices, and this set turns out to be dense in the set of complex $n \times n$ matrices.

13. The spectral theorem for linear operators

It turns out that the above three theorems have a counterpart that is stated in terms of linear operators. First we make the following definitions.

If V is a real inner product space, and $T: V \to V$ is linear, then T is said to be *symmetric* if and only if $T^t = T$.

If V is a complex inner product space, and $T: V \to V$ is linear, then T is said to be *self-adjoint* (or *Hermitian*) if and only if $T^* = T$, and T is said to be *normal* if and only if $TT^* = T^*T$.

One can show that every symmetric operator on a finite dimensional real inner product space extends to a self-adjoint operator on the complexification of the real inner product space, and every self-adjoint operator on a complex inner product space is normal.

We now have a characterization of symmetric operators on real inner product spaces:

Theorem 9. *Suppose V is a finite dimensional real inner product space, and $T: V \to V$ is linear. Then the following are equivalent:*

(a) T is symmetric;
(b) V has an orthonormal basis that consists of eigenvectors of T;
(c) V has an ordered orthonormal basis such that the matrix representation of T with respect to this basis is diagonal.

Similarly, we have the following characterization of self-adjoint operators on complex inner product spaces:

Theorem 10. *Suppose V is a finite dimensional complex inner product space, and $T: V \to V$ is linear. Then the following are equivalent:*

(a) T is self-adjoint;
(b) V has an orthonormal basis that consists of eigenvectors of T, and all eigenvalues of T are real;
(c) V has an ordered orthonormal basis such that the matrix representation of T with respect to this basis is diagonal with real entries.

Finally we have the following characterization of normal operators on complex inner product spaces:

Theorem 11. *Suppose V is a finite dimensional complex inner product space, and $T: V \to V$ is linear. Then the following are equivalent:*

(a) T is normal;
(b) V has an orthonormal basis that consists of eigenvectors of T;
(c) V has an ordered orthonormal basis such that the matrix representation of T with respect to this basis is diagonal.

Again the equivalence of (b) and (c) in the three theorems above are very easy. We leave them to the reader.

Observe that if one uses the standard inner products on $\mathbb{R}^n$ and $\mathbb{C}^n$, then a real $n \times n$ matrix A is symmetric if and only if the linear operator $x \mapsto Ax$ is a symmetric operator on $\mathbb{R}^n$, and a complex $n \times n$ matrix A is self-adjoint (resp. normal) if and only if the linear operator $z \mapsto Az$ is a self-adjoint (resp. normal) operator on $\mathbb{C}^n$. Thus to prove the equivalence of (a) and (b) in Theorems 6, 7 and 8, we only need to prove the equivalence of (a) and (b) in Theorems 9, 10 and 11; in fact it is beneficial to take this abstract point of view even if one is only interested in proving the results about matrices.

Since it is relatively easy to prove that (b) implies (a) in Theorems 9, 10 and 11, we leave them to the reader. In the next three sections, we prove that (a) implies (b) in Theorems 9, 10 and 11.

14. Proof of Theorem 10

In this section we prove that (a) implies (b) in Theorem 10. Suppose $T\colon V \to V$ is a self-adjoint linear operator on a finite dimensional complex inner product space V. First we prove the following proposition:

Proposition 8. *All the eigenvalues of T are real.*

Proof. Suppose $\lambda \in \mathbb{C}$ is an eigenvalue of T, and $v \neq 0$ is the corresponding eigenvector of T. Then $Tv = \lambda v$, so

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

Since $\|v\| \neq 0$, we get

$$\lambda = \overline{\lambda},$$

and thus λ is real. □

Next, we want to prove that V has an orthonormal basis consisting of eigenvectors of T. The proof here is by induction on the dimension of V. It comes in two steps. First, given T as above, we find an eigenvector v_1 of T. Without loss of generality we may assume that $\|v_1\| = 1$. Let W be the orthogonal complement of v_1 in V. We will then show[1] that W is a T-invariant subspace of V, and that the restriction of T to W is still self-adjoint. One can then invoke our induction hypothesis, and conclude that W has an orthonormal basis $\{v_2, \dots, v_n\}$ that consists

[1] This is a crucial place where orthogonality comes in; it is not just any complement of the span of v_1 in V that is T-invariant, but the orthogonal complement of v_1 that is T-invariant.

of eigenvectors of T. Then one can check that $\{v_1, \dots, v_n\}$ is an orthonormal basis of V consisting of eigenvectors of T.

As a result, we only need to prove the following:

(i) Every self-adjoint linear operator $T\colon V \to V$ on a finite dimensional complex inner product space V has an eigenvector;
(ii) If v_1 is an eigenvector of T and W is the orthogonal complement of v_1 in V, then W is T-invariant, and that $T|_W : W \to W$ is self-adjoint.

We begin with (ii). If v_1 is an eigenvector of T, say with eigenvalue λ ($\in \mathbb{R}$ by the previous proposition), and if W is the orthogonal complement, then for any $w \in W$, we have

$$\langle Tw, v_1 \rangle = \langle w, T^* v_1 \rangle = \langle w, T v_1 \rangle = \lambda \langle w, v_1 \rangle = 0.$$

Thus Tw is in W for all $w \in W$. This proves that W is T-invariant.

Next, since $T^* = T$, W is also T^*-invariant. Thus if U is the restriction of T to W, then, U^* is the restriction of T^* to W; in fact if $w_1, w_2 \in W$, then

$$\langle w_1, U^* w_2 \rangle = \langle U w_1, w_2 \rangle = \langle T w_1, w_2 \rangle = \langle w_1, T w_2 \rangle = \langle w_1, U w_2 \rangle,$$

which shows $U^* w_2 = U w_2$ for all $w_2 \in W$. (Here one uses $U w_2 \in W$.) It follows that $U^* = U$, and thus the restriction of T to W is self-adjoint.

We next prove (i). The easiest proof is to invoke the fundamental theorem of algebra, which says that every complex polynomial has a root in the complex plane. In particular, since V is finite dimensional, there exists some $\lambda_1 \in \mathbb{C}$ such that $\det(T - \lambda_1 I) = 0$. This says T has an eigenvalue λ_1, and thus T has an eigenvector v_1 corresponding to λ_1.

15. Proof of Theorem 11

In this section, suppose V is a finite dimensional complex inner product space, and $T\colon V \to V$ is a normal linear operator. We will prove that V has a basis that consists of eigenvectors of T, thereby proving that (a) implies (b) in Theorem 11.

Again the idea is to prove the following:

(i) T has an eigenvector in V;
(ii) If v_1 is an eigenvector of T and W is the orthogonal complement of v_1 in V, then W is T-invariant, and that $T|_W : W \to W$ is normal.

Once these are proved, the proof that (a) implies (b) in Theorem 11 can be completed as in the case of Theorems 10. We leave that to the reader.

First we prove an important property about eigenvectors of normal operators.

Proposition 9. *Suppose $T\colon V \to V$ is normal. If v is an eigenvector of T with eigenvalue λ, then v is also an eigenvector of T^* with eigenvalue $\overline{\lambda}$, and conversely.*

Proof. By normality of T, we know that $T-\lambda I$ and $T^*-\overline{\lambda}I$ commutes for all $\lambda\in\mathbb{C}$. Thus for any $v\in V$ and any $\lambda\in\mathbb{C}$, we have

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

But this says

$$(T-\lambda I)v=0 \Leftrightarrow (T^*-\overline{\lambda}I)v=0.$$

So we have the desired assertion. □

We are now ready to prove statement (ii) above. Suppose again that $T\colon V\to V$ is a complex normal operator. If v_1 is an eigenvector of T, say with eigenvalue λ, let W be the orthogonal complement of v_1. Now if $w\in W$, we want to show that $Tw\in W$. In other words, we want to show $\langle Tw,v_1\rangle=0$. But

$$\langle Tw,v_1\rangle=\langle w,T^*v_1\rangle=\langle w,\overline{\lambda}v_1\rangle=\lambda\langle w,v_1\rangle=0,$$

where the second equality follows from the previous proposition. Thus W is T-invariant. Similarly, one can show that W is T^*-invariant. Hence, if U is the restriction of T to W, then for all $w_1,w_2\in W$, we have

$$\langle w_1,U^*w_2\rangle=\langle Uw_1,w_2\rangle=\langle Tw_1,w_2\rangle=\langle w_1,T^*w_2\rangle,$$

which shows $U^*w_2=T^*w_2$ for all $w_2\in W$. (Here one uses $T^*w_2\in W$.) From normality of T, it then follows that $U^*U=UU^*$, and thus the restriction of T to W is still a normal operator.

Next, we turn to a proof of (i). Again the most direct one is to use the fundamental theorem of algebra, as in the first proof of (i) in Theorem 10. We leave the details to the reader.

16. Proof of Theorem 9

We now give the modifications of the proof of Theorem 10 that are necessary to prove that (a) implies (b) in Theorem 9. Thus in this section, unless otherwise stated, we assume that V is a real inner product space, and $T\colon V\to V$ is a symmetric linear operator. As in the proof of Theorem 10, we only need to prove the following:

(i) T has an eigenvector in V;
(ii) If v_1 is an eigenvector of T and W is the orthogonal complement of v_1 in V, then W is T-invariant, and that $T|_W : W\to W$ is symmetric.

The proof of (ii) in our current context is the same as that of the corresponding statement in the proof of Theorem 10. One just needs to keep in mind that if V is a real inner product space, then any eigenvalue of a symmetric linear operator $T\colon V\to V$ is real (by definition). One then replaces any occurrence of T^* in the previous proof by T^t. We leave the details to the reader.

To prove (i) in our current context, we use a *variational method*. Let S be the unit sphere in V, i.e. $S=\{v\in V\colon \|v\|=1\}$. It is a compact subset of V by the finite dimensionality of V. Now let $F(v)=\langle Tv,v\rangle$. This defines a smooth function from S to $\mathbb{R}$. Thus F achieves its maximum at some $v_1 \in S$. We can then show that v_1 is an eigenvector of T, pretty much in the same way that we had when we gave the second proof of (i) in the previous section. In fact, we then have

$$\left.\frac{d}{dt}\right|_{t=0} F\left(\frac{v_1 + tw}{\|v_1 + tw\|}\right) = 0.$$

But the left hand side of this equation is equal to

$$\left.\frac{d}{dt}\right|_{t=0} \frac{1}{\|v_1 + tw\|^2}\langle T(v_1 + tw), v_1 + tw\rangle$$

which in turn is given by

$$-\frac{\langle v_1, w\rangle + \langle w, v_1\rangle}{\|v_1\|^4}\langle Tv_1, v_1\rangle + \frac{1}{\|v_1\|^2}(\langle Tv_1, w\rangle + \langle Tw, v_1\rangle),$$

i.e.

$$-(\langle v_1, w\rangle + \langle w, v_1\rangle)\langle Tv_1, v_1\rangle + \langle Tv_1, w\rangle + \langle Tw, v_1\rangle.$$

Since we are working on a real inner product space, and since $T^t = T$, this implies

$$-2\langle v_1, w\rangle\langle Tv_1, v_1\rangle + 2\langle Tv_1, w\rangle = 0.$$

Thus letting $\lambda_1 = \langle Tv_1, v_1\rangle$, we get

$$\langle Tv_1 - \lambda_1 v_1, w\rangle = 0$$

for all $w \in V$. Thus $Tv_1 = \lambda_1 v_1$, and T has an eigenvector v_1. This completes our proof that (a) implies (b) in Theorem 9.