CHAPTER 6

Inner-Product Spaces

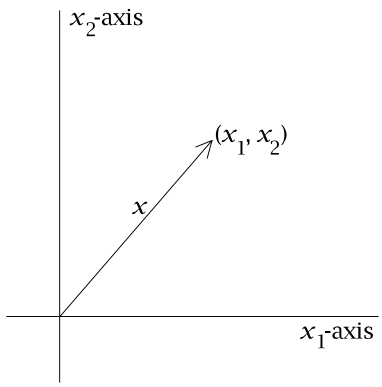
In making the definition of a vector space, we generalized the linear structure (addition and scalar multiplication) of $\mathbf{R}^2$ and $\mathbf{R}^3$. We ignored other important features, such as the notions of length and angle. These ideas are embedded in the concept we now investigate, inner products.

Recall that $\mathbf{F}$ denotes $\mathbf{R}$ or $\mathbf{C}$.
Also, V is a finite-dimensional, nonzero vector space over $\mathbf{F}$.

Inner Products

To motivate the concept of inner product, let's think of vectors in $\mathbf{R}^2$ and $\mathbf{R}^3$ as arrows with initial point at the origin. The length of a vector x in $\mathbf{R}^2$ or $\mathbf{R}^3$ is called the ***norm*** of x, denoted $\|x\|$. Thus for $x = (x_1, x_2) \in \mathbf{R}^2$, we have $\|x\| = \sqrt{x_1{}^2 + x_2{}^2}$.

If we think of vectors as points instead of arrows, then $\|x\|$ should be interpreted as the distance from the point x to the origin.



The length of this vector x is $\sqrt{x_1{}^2 + x_2{}^2}$.

Similarly, for $x = (x_1, x_2, x_3) \in \mathbf{R}^3$, we have $\|x\| = \sqrt{x_1{}^2 + x_2{}^2 + x_3{}^2}$. Even though we cannot draw pictures in higher dimensions, the generalization to $\mathbf{R}^n$ is obvious: we define the norm of $x = (x_1, \dots, x_n) \in \mathbf{R}^n$ by

$$\|x\| = \sqrt{x_1{}^2 + \cdots + x_n{}^2}.$$

The norm is not linear on $\mathbf{R}^n$. To inject linearity into the discussion, we introduce the dot product. For $x, y \in \mathbf{R}^n$, the ***dot product*** of x and y, denoted $x \cdot y$, is defined by

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

where $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$. Note that the dot product of two vectors in $\mathbf{R}^n$ is a number, not a vector. Obviously $x \cdot x = \|x\|^2$ for all $x \in \mathbf{R}^n$. In particular, $x \cdot x \geq 0$ for all $x \in \mathbf{R}^n$, with equality if and only if $x = 0$. Also, if $y \in \mathbf{R}^n$ is fixed, then clearly the map from $\mathbf{R}^n$ to $\mathbf{R}$ that sends $x \in \mathbf{R}^n$ to $x \cdot y$ is linear. Furthermore, $x \cdot y = y \cdot x$ for all $x, y \in \mathbf{R}^n$.

An inner product is a generalization of the dot product. At this point you should be tempted to guess that an inner product is defined by abstracting the properties of the dot product discussed in the paragraph above. For real vector spaces, that guess is correct. However, so that we can make a definition that will be useful for both real and complex vector spaces, we need to examine the complex case before making the definition.

Recall that if $\lambda = a + bi$, where $a, b \in \mathbf{R}$, then the absolute value of λ is defined by

$$|\lambda| = \sqrt{a^2 + b^2},$$

the complex conjugate of λ is defined by

$$\bar{\lambda} = a - bi,$$

and the equation

$$|\lambda|^2 = \lambda\bar{\lambda}$$

connects these two concepts (see page 69 for the definitions and the basic properties of the absolute value and complex conjugate). For $z = (z_1, \dots, z_n) \in \mathbf{C}^n$, we define the norm of z by

$$\|z\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2}.$$

The absolute values are needed because we want $\|z\|$ to be a nonnegative number. Note that

$$\|z\|^2 = z_1\overline{z_1} + \cdots + z_n\overline{z_n}.$$

We want to think of $\|z\|^2$ as the inner product of z with itself, as we did in $\mathbf{R}^n$. The equation above thus suggests that the inner product of $w = (w_1, \dots, w_n) \in \mathbf{C}^n$ with z should equal

$$w_1\overline{z_1} + \cdots + w_n\overline{z_n}.$$

If the roles of the w and z were interchanged, the expression above would be replaced with its complex conjugate. In other words, we should expect that the inner product of w with z equals the complex conjugate of the inner product of z with w. With that motivation, we are now ready to define an inner product on V, which may be a real or a complex vector space.

An ***inner product*** on V is a function that takes each ordered pair (u, v) of elements of V to a number $\langle u, v\rangle \in \mathbf{F}$ and has the following properties:

positivity
$\langle v, v\rangle \geq 0$ for all $v \in V$;

If z is a complex number, then the statement $z \geq 0$ means that z is real and nonnegative.

definiteness
$\langle v, v\rangle = 0$ if and only if $v = 0$;

additivity in first slot
$\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ for all $u, v, w \in V$;

homogeneity in first slot
$\langle av, w\rangle = a\langle v, w\rangle$ for all $a \in \mathbf{F}$ and all $v, w \in V$;

conjugate symmetry
$\langle v, w\rangle = \overline{\langle w, v\rangle}$ for all $v, w \in V$.

Recall that every real number equals its complex conjugate. Thus if we are dealing with a real vector space, then in the last condition above we can dispense with the complex conjugate and simply state that $\langle v, w\rangle = \langle w, v\rangle$ for all $v, w \in V$.

An ***inner-product space*** is a vector space V along with an inner product on V.

The most important example of an inner-product space is $\mathbf{F}^n$. We can define an inner product on $\mathbf{F}^n$ by

If we are dealing with $\mathbf{R}^n$ rather than $\mathbf{C}^n$, then again the complex conjugate can be ignored.

6.1 $$\langle (w_1, \dots, w_n), (z_1, \dots, z_n)\rangle = w_1\overline{z_1} + \cdots + w_n\overline{z_n},$$

as you should verify. This inner product, which provided our motivation for the definition of an inner product, is called the ***Euclidean inner product*** on $\mathbf{F}^n$. When $\mathbf{F}^n$ is referred to as an inner-product space, you should assume that the inner product is the Euclidean inner product unless explicitly told otherwise.

There are other inner products on $\mathbf{F}^n$ in addition to the Euclidean inner product. For example, if $c_1, \dots, c_n$ are positive numbers, then we can define an inner product on $\mathbf{F}^n$ by

$$\langle (w_1, \dots, w_n), (z_1, \dots, z_n)\rangle = c_1 w_1\overline{z_1} + \cdots + c_n w_n\overline{z_n},$$

as you should verify. Of course, if all the c's equal 1, then we get the Euclidean inner product.

As another example of an inner-product space, consider the vector space $\mathcal{P}_m(\mathbf{F})$ of all polynomials with coefficients in $\mathbf{F}$ and degree at most m. We can define an inner product on $\mathcal{P}_m(\mathbf{F})$ by

6.2
$$\langle p, q\rangle = \int_0^1 p(x)\overline{q(x)}\,dx,$$

as you should verify. Once again, if $\mathbf{F} = \mathbf{R}$, then the complex conjugate is not needed.

Let's agree for the rest of this chapter that V is a finite-dimensional inner-product space over $\mathbf{F}$.

In the definition of an inner product, the conditions of additivity and homogeneity in the first slot can be combined into a requirement of linearity in the first slot. More precisely, for each fixed $w \in V$, the function that takes v to $\langle v, w\rangle$ is a linear map from V to $\mathbf{F}$. Because every linear map takes 0 to 0, we must have

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

for every $w \in V$. Thus we also have

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

for every $w \in V$ (by the conjugate symmetry property).

In an inner-product space, we have additivity in the second slot as well as the first slot. Proof:

$$\begin{aligned}\langle u, v + w\rangle &= \overline{\langle v + w, u\rangle}\\ &= \overline{\langle v, u\rangle + \langle w, u\rangle}\\ &= \overline{\langle v, u\rangle} + \overline{\langle w, u\rangle}\\ &= \langle u, v\rangle + \langle u, w\rangle;\end{aligned}$$

here $u, v, w \in V$.

In an inner-product space, we have conjugate homogeneity in the second slot, meaning that $\langle u, av\rangle = \bar{a}\langle u, v\rangle$ for all scalars $a \in \mathbf{F}$. Proof:

$$\begin{aligned}\langle u, av\rangle &= \overline{\langle av, u\rangle}\\ &= \overline{a\langle v, u\rangle}\\ &= \bar{a}\overline{\langle v, u\rangle}\\ &= \bar{a}\langle u, v\rangle;\end{aligned}$$

here $a \in \mathbf{F}$ and $u, v \in V$. Note that in a real vector space, conjugate homogeneity is the same as homogeneity.

Norms

For $v \in V$, we define the ***norm*** of v, denoted $\|v\|$, by

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

For example, if $(z_1, \dots, z_n) \in \mathbf{F}^n$ (with the Euclidean inner product), then

$$\|(z_1, \dots, z_n)\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2}.$$

As another example, if $p \in \mathcal{P}_m(\mathbf{F})$ (with inner product given by 6.2), then

$$\|p\| = \sqrt{\int_0^1 |p(x)|^2 \, dx}.$$

Note that $\|v\| = 0$ if and only if $v = 0$ (because $\langle v, v\rangle = 0$ if and only if $v = 0$). Another easy property of the norm is that $\|av\| = |a| \, \|v\|$ for all $a \in \mathbf{F}$ and all $v \in V$. Here's the proof:

$$\begin{aligned}
\|av\|^2 &= \langle av, av\rangle \\
&= a\langle v, av\rangle \\
&= a\bar{a}\langle v, v\rangle \\
&= |a|^2 \|v\|^2;
\end{aligned}$$

taking square roots now gives the desired equality. This proof illustrates a general principle: working with norms squared is usually easier than working directly with norms.

*Some mathematicians use the term **perpendicular**, which means the same as orthogonal.*

Two vectors $u, v \in V$ are said to be ***orthogonal*** if $\langle u, v\rangle = 0$. Note that the order of the vectors does not matter because $\langle u, v\rangle = 0$ if and only if $\langle v, u\rangle = 0$. Instead of saying that u and v are orthogonal, sometimes we say that u is orthogonal to v. Clearly 0 is orthogonal to every vector. Furthermore, 0 is the only vector that is orthogonal to itself.

*The word **orthogonal** comes from the Greek word **orthogonios**, which means right-angled.*

For the special case where $V = \mathbf{R}^2$, the next theorem is over 2,500 years old.

6.3 Pythagorean Theorem: *If u, v are orthogonal vectors in V, then*

6.4
$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

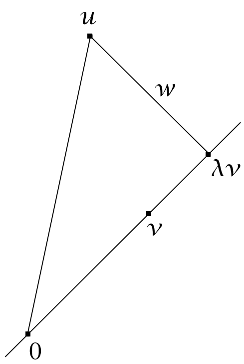
PROOF: Suppose that u, v are orthogonal vectors in V. Then

$$\begin{aligned}\|u+v\|^2 &= \langle u+v, u+v\rangle \\ &= \|u\|^2 + \|v\|^2 + \langle u,v\rangle + \langle v,u\rangle \\ &= \|u\|^2 + \|v\|^2,\end{aligned}$$

as desired. ■

The proof of the Pythagorean theorem shows that 6.4 holds if and only if $\langle u,v\rangle + \langle v,u\rangle$, which equals $2\,\mathrm{Re}\langle u,v\rangle$, is 0. Thus the converse of the Pythagorean theorem holds in real inner-product spaces.

Suppose $u, v \in V$. We would like to write u as a scalar multiple of v plus a vector w orthogonal to v, as suggested in the next picture.



An orthogonal decomposition

To discover how to write u as a scalar multiple of v plus a vector orthogonal to v, let $a \in \mathbf{F}$ denote a scalar. Then

$$u = av + (u - av).$$

Thus we need to choose a so that v is orthogonal to $(u - av)$. In other words, we want

$$0 = \langle u - av, v\rangle = \langle u, v\rangle - a\|v\|^2.$$

The equation above shows that we should choose a to be $\langle u,v\rangle/\|v\|^2$ (assume that $v \neq 0$ to avoid division by 0). Making this choice of a, we can write

6.5 $$u = \frac{\langle u,v\rangle}{\|v\|^2}v + \Big(u - \frac{\langle u,v\rangle}{\|v\|^2}v\Big).$$

As you should verify, if $v \neq 0$ then the equation above writes u as a scalar multiple of v plus a vector orthogonal to v.

The equation above will be used in the proof of the next theorem, which gives one of the most important inequalities in mathematics.

In 1821 the French mathematician Augustin-Louis Cauchy showed that this inequality holds for the inner product defined by 6.1. In 1886 the German mathematician Herman Schwarz showed that this inequality holds for the inner product defined by 6.2.

6.6 Cauchy-Schwarz Inequality: *If $u, v \in V$, then*

6.7
$$|\langle u, v\rangle| \leq \|u\|\,\|v\|.$$

This inequality is an equality if and only if one of u, v is a scalar multiple of the other.

PROOF: Let $u, v \in V$. If $v = 0$, then both sides of 6.7 equal 0 and the desired inequality holds. Thus we can assume that $v \neq 0$. Consider the orthogonal decomposition

$$u = \frac{\langle u, v\rangle}{\|v\|^2}v + w,$$

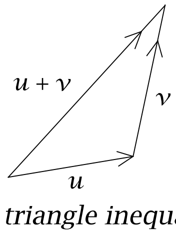
where w is orthogonal to v (here w equals the second term on the right side of 6.5). By the Pythagorean theorem,

$$\begin{aligned}\|u\|^2 &= \left\|\frac{\langle u, v\rangle}{\|v\|^2}v\right\|^2 + \|w\|^2\\ &= \frac{|\langle u, v\rangle|^2}{\|v\|^2} + \|w\|^2\\ &\geq \frac{|\langle u, v\rangle|^2}{\|v\|^2}. \qquad \textbf{6.8}\end{aligned}$$

Multiplying both sides of this inequality by $\|v\|^2$ and then taking square roots gives the Cauchy-Schwarz inequality 6.7.

Looking at the proof of the Cauchy-Schwarz inequality, note that 6.7 is an equality if and only if 6.8 is an equality. Obviously this happens if and only if $w = 0$. But $w = 0$ if and only if u is a multiple of v (see 6.5). Thus the Cauchy-Schwarz inequality is an equality if and only if u is a scalar multiple of v or v is a scalar multiple of u (or both; the phrasing has been chosen to cover cases in which either u or v equals 0). ■

The next result is called the triangle inequality because of its geometric interpretation that the length of any side of a triangle is less than the sum of the lengths of the other two sides.



The triangle inequality

6.9 Triangle Inequality: *If* $u, v \in V$*, then*

6.10 $$\|u + v\| \leq \|u\| + \|v\|.$$

This inequality is an equality if and only if one of u, v *is a nonnegative multiple of the other.*

The triangle inequality can be used to show that the shortest path between two points is a straight line segment.

PROOF: Let $u, v \in V$. Then

$$\begin{aligned}
\|u + v\|^2 &= \langle u + v, u + v \rangle \\
&= \langle u, u \rangle + \langle v, v \rangle + \langle u, v \rangle + \langle v, u \rangle \\
&= \langle u, u \rangle + \langle v, v \rangle + \langle u, v \rangle + \overline{\langle u, v \rangle} \\
&= \|u\|^2 + \|v\|^2 + 2\,\mathrm{Re}\langle u, v \rangle \\
&\leq \|u\|^2 + \|v\|^2 + 2|\langle u, v \rangle| && \textbf{6.11} \\
&\leq \|u\|^2 + \|v\|^2 + 2\|u\|\,\|v\| && \textbf{6.12} \\
&= (\|u\| + \|v\|)^2,
\end{aligned}$$

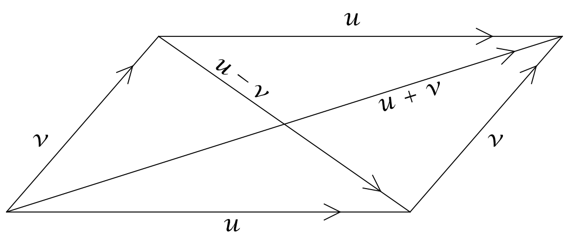
where 6.12 follows from the Cauchy-Schwarz inequality (6.6). Taking square roots of both sides of the inequality above gives the triangle inequality 6.10.

The proof above shows that the triangle inequality 6.10 is an equality if and only if we have equality in 6.11 and 6.12. Thus we have equality in the triangle inequality 6.10 if and only if

6.13 $$\langle u, v \rangle = \|u\|\|v\|.$$

If one of u, v is a nonnegative multiple of the other, then 6.13 holds, as you should verify. Conversely, suppose 6.13 holds. Then the condition for equality in the Cauchy-Schwarz inequality (6.6) implies that one of u, v must be a scalar multiple of the other. Clearly 6.13 forces the scalar in question to be nonnegative, as desired. ■

The next result is called the parallelogram equality because of its geometric interpretation: in any parallelogram, the sum of the squares of the lengths of the diagonals equals the sum of the squares of the lengths of the four sides.



The parallelogram equality

6.14 Parallelogram Equality: *If* $u, v \in V$, *then*

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

PROOF: Let $u, v \in V$. Then

$$\begin{aligned}\|u+v\|^2 + \|u-v\|^2 &= \langle u+v, u+v\rangle + \langle u-v, u-v\rangle \\ &= \|u\|^2 + \|v\|^2 + \langle u, v\rangle + \langle v, u\rangle \\ &\quad + \|u\|^2 + \|v\|^2 - \langle u, v\rangle - \langle v, u\rangle \\ &= 2(\|u\|^2 + \|v\|^2),\end{aligned}$$

as desired. ∎

Orthonormal Bases

A list of vectors is called ***orthonormal*** if the vectors in it are pairwise orthogonal and each vector has norm 1. In other words, a list $(e_1, \dots, e_m)$ of vectors in V is orthonormal if $\langle e_j, e_k\rangle$ equals 0 when $j \neq k$ and equals 1 when $j = k$ (for $j, k = 1, \dots, m$). For example, the standard basis in $\mathbf{F}^n$ is orthonormal. Orthonormal lists are particularly easy to work with, as illustrated by the next proposition.

6.15 Proposition: *If* $(e_1, \dots, e_m)$ *is an orthonormal list of vectors in* V, *then*

$$\|a_1e_1 + \cdots + a_me_m\|^2 = |a_1|^2 + \cdots + |a_m|^2$$

for all $a_1, \dots, a_m \in \mathbf{F}$.

PROOF: Because each e_j has norm 1, this follows easily from repeated applications of the Pythagorean theorem (6.3). ∎

Now we have the following easy but important corollary.

6.16 Corollary: *Every orthonormal list of vectors is linearly independent.*

PROOF: Suppose $(e_1, \dots, e_m)$ is an orthonormal list of vectors in V and $a_1, \dots, a_m \in \mathbf{F}$ are such that

$$a_1 e_1 + \cdots + a_m e_m = 0.$$

Then $|a_1|^2 + \cdots + |a_m|^2 = 0$ (by 6.15), which means that all the a_j's are 0, as desired. ■

An ***orthonormal basis*** of V is an orthonormal list of vectors in V that is also a basis of V. For example, the standard basis is an orthonormal basis of $\mathbf{F}^n$. Every orthonormal list of vectors in V with length $\dim V$ is automatically an orthonormal basis of V (proof: by the previous corollary, any such list must be linearly independent; because it has the right length, it must be a basis—see 2.17). To illustrate this principle, consider the following list of four vectors in $\mathbf{R}^4$:

$$\left(\left(\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}\right), \left(\tfrac{1}{2}, \tfrac{1}{2}, -\tfrac{1}{2}, -\tfrac{1}{2}\right), \left(\tfrac{1}{2}, -\tfrac{1}{2}, -\tfrac{1}{2}, \tfrac{1}{2}\right), \left(-\tfrac{1}{2}, \tfrac{1}{2}, -\tfrac{1}{2}, \tfrac{1}{2}\right)\right).$$

The verification that this list is orthonormal is easy (do it!); because we have an orthonormal list of length four in a four-dimensional vector space, it must be an orthonormal basis.

In general, given a basis $(e_1, \dots, e_n)$ of V and a vector $v \in V$, we know that there is some choice of scalars $a_1, \dots, a_m$ such that

$$v = a_1 e_1 + \cdots + a_n e_n,$$

but finding the a_j's can be difficult. The next theorem shows, however, that this is easy for an orthonormal basis.

The importance of orthonormal bases stems mainly from this theorem.

6.17 Theorem: *Suppose $(e_1, \dots, e_n)$ is an orthonormal basis of V. Then*

6.18 $$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n$$

and

6.19 $$\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2$$

for every $v \in V$.

PROOF: Let $v \in V$. Because $(e_1, \dots, e_n)$ is a basis of V, there exist scalars $a_1, \dots, a_n$ such that

$$v = a_1 e_1 + \cdots + a_n e_n.$$

Take the inner product of both sides of this equation with e_j, getting $\langle v, e_j \rangle = a_j$. Thus 6.18 holds. Clearly 6.19 follows from 6.18 and 6.15. ■

Now that we understand the usefulness of orthonormal bases, how do we go about finding them? For example, does $\mathcal{P}_m(\mathbf{F})$, with inner product given by integration on $[0, 1]$ (see 6.2), have an orthonormal basis? As we will see, the next result will lead to answers to these questions. The algorithm used in the next proof is called the ***Gram-Schmidt procedure***. It gives a method for turning a linearly independent list into an orthonormal list with the same span as the original list.

The Danish mathematician Jorgen Gram (1850–1916) and the German mathematician Erhard Schmidt (1876–1959) popularized this algorithm for constructing orthonormal lists.

6.20 Gram-Schmidt: *If $(v_1, \dots, v_m)$ is a linearly independent list of vectors in V, then there exists an orthonormal list $(e_1, \dots, e_m)$ of vectors in V such that*

6.21 $$\operatorname{span}(v_1, \dots, v_j) = \operatorname{span}(e_1, \dots, e_j)$$

for $j = 1, \dots, m$.

PROOF: Suppose $(v_1, \dots, v_m)$ is a linearly independent list of vectors in V. To construct the e's, start by setting $e_1 = v_1 / \|v_1\|$. This satisfies 6.21 for $j = 1$. We will choose $e_2, \dots, e_m$ inductively, as follows. Suppose $j > 1$ and an orthornormal list $(e_1, \dots, e_{j-1})$ has been chosen so that

6.22 $$\operatorname{span}(v_1, \dots, v_{j-1}) = \operatorname{span}(e_1, \dots, e_{j-1}).$$

Let

6.23 $$e_j = \frac{v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}}{\|v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}\|}.$$

Note that $v_j \notin \operatorname{span}(v_1, \dots, v_{j-1})$ (because $(v_1, \dots, v_m)$ is linearly independent) and thus $v_j \notin \operatorname{span}(e_1, \dots, e_{j-1})$. Hence we are not dividing by 0 in the equation above, and so e_j is well defined. Dividing a vector by its norm produces a new vector with norm 1; thus $\|e_j\| = 1$.

Let $1 \le k < j$. Then

$$\begin{aligned}\langle e_j, e_k\rangle &= \left\langle \frac{v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}}{\|v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}\|}, e_k \right\rangle \\ &= \frac{\langle v_j, e_k\rangle - \langle v_j, e_k\rangle}{\|v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}\|} \\ &= 0.\end{aligned}$$

Thus $(e_1, \ldots, e_j)$ is an orthonormal list.

From 6.23, we see that $v_j \in \operatorname{span}(e_1, \ldots, e_j)$. Combining this information with 6.22 shows that

$$\operatorname{span}(v_1, \ldots, v_j) \subset \operatorname{span}(e_1, \ldots, e_j).$$

Both lists above are linearly independent (the v's by hypothesis, the e's by orthonormality and 6.16). Thus both subspaces above have dimension j, and hence they must be equal, completing the proof. ■

Now we can settle the question of the existence of orthonormal bases.

6.24 Corollary: *Every finite-dimensional inner-product space has an orthonormal basis.*

Until this corollary, nothing we had done with inner-product spaces required our standing assumption that V is finite dimensional.

PROOF: Choose a basis of V. Apply the Gram-Schmidt procedure (6.20) to it, producing an orthonormal list. This orthonormal list is linearly independent (by 6.16) and its span equals V. Thus it is an orthonormal basis of V. ■

As we will soon see, sometimes we need to know not only that an orthonormal basis exists, but also that any orthonormal list can be extended to an orthonormal basis. In the next corollary, the Gram-Schmidt procedure shows that such an extension is always possible.

6.25 Corollary: *Every orthonormal list of vectors in V can be extended to an orthonormal basis of V.*

PROOF: Suppose $(e_1, \ldots, e_m)$ is an orthonormal list of vectors in V. Then $(e_1, \ldots, e_m)$ is linearly independent (by 6.16), and hence it can be extended to a basis $(e_1, \ldots, e_m, v_1, \ldots, v_n)$ of V (see 2.12). Now apply the Gram-Schmidt procedure (6.20) to $(e_1, \dots, e_m, v_1, \dots, v_n)$, producing an orthonormal list

6.26 $$(e_1, \dots, e_m, f_1, \dots, f_n);$$

here the Gram-Schmidt procedure leaves the first m vectors unchanged because they are already orthonormal. Clearly 6.26 is an orthonormal basis of V because it is linearly independent (by 6.16) and its span equals V. Hence we have our extension of $(e_1, \dots, e_m)$ to an orthonormal basis of V. ■

Recall that a matrix is called upper triangular if all entries below the diagonal equal 0. In other words, an upper-triangular matrix looks like this:

$$\begin{bmatrix} * & & * \\ & \ddots & \\ 0 & & * \end{bmatrix}.$$

In the last chapter we showed that if V is a complex vector space, then for each operator on V there is a basis with respect to which the matrix of the operator is upper triangular (see 5.13). Now that we are dealing with inner-product spaces, we would like to know when there exists an *orthonormal* basis with respect to which we have an upper-triangular matrix. The next corollary shows that the existence of any basis with respect to which T has an upper-triangular matrix implies the existence of an orthonormal basis with this property. This result is true on both real and complex vector spaces (though on a real vector space, the hypothesis holds only for some operators).

6.27 Corollary: *Suppose $T \in \mathcal{L}(V)$. If T has an upper-triangular matrix with respect to some basis of V, then T has an upper-triangular matrix with respect to some orthonormal basis of V.*

PROOF: Suppose T has an upper-triangular matrix with respect to some basis $(v_1, \dots, v_n)$ of V. Thus $\operatorname{span}(v_1, \dots, v_j)$ is invariant under T for each $j = 1, \dots, n$ (see 5.12).

Apply the Gram-Schmidt procedure to $(v_1, \dots, v_n)$, producing an orthonormal basis $(e_1, \dots, e_n)$ of V. Because

$$\operatorname{span}(e_1, \dots, e_j) = \operatorname{span}(v_1, \dots, v_j)$$

for each j (see 6.21), we conclude that $\operatorname{span}(e_1, \dots, e_j)$ is invariant under T for each $j = 1, \dots, n$. Thus, by 5.12, T has an upper-triangular matrix with respect to the orthonormal basis $(e_1, \dots, e_n)$. ■

The next result is an important application of the corollary above.

6.28 Corollary: *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some orthonormal basis of V.*

This result is sometimes called Schur's theorem. The German mathematician Issai Schur published the first proof of this result in 1909.

PROOF: This follows immediately from 5.13 and 6.27. ■

Orthogonal Projections and Minimization Problems

If U is a subset of V, then the ***orthogonal complement*** of U, denoted $U^\perp$, is the set of all vectors in V that are orthogonal to every vector in U:

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

You should verify that $U^\perp$ is always a subspace of V, that $V^\perp = \{0\}$, and that $\{0\}^\perp = V$. Also note that if $U_1 \subset U_2$, then $U_1^\perp \supset U_2^\perp$.

Recall that if U_1, U_2 are subspaces of V, then V is the direct sum of U_1 and U_2 (written $V = U_1 \oplus U_2$) if each element of V can be written in exactly one way as a vector in U_1 plus a vector in U_2. The next theorem shows that every subspace of an inner-product space leads to a natural direct sum decomposition of the whole space.

6.29 Theorem: *If U is a subspace of V, then*

$$V = U \oplus U^\perp.$$

PROOF: Suppose that U is a subspace of V. First we will show that

6.30 $$V = U + U^\perp.$$

To do this, suppose $v \in V$. Let $(e_1, \dots, e_m)$ be an orthonormal basis of U. Obviously

6.31
$$v = \underbrace{\langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m}_{u} + \underbrace{v - \langle v, e_1 \rangle e_1 - \cdots - \langle v, e_m \rangle e_m}_{w}.$$

Clearly $u \in U$. Because $(e_1, \ldots, e_m)$ is an orthonormal list, for each j we have

$$\begin{aligned}\langle w, e_j \rangle &= \langle v, e_j \rangle - \langle v, e_j \rangle \\ &= 0.\end{aligned}$$

Thus w is orthogonal to every vector in $\operatorname{span}(e_1, \ldots, e_m)$. In other words, $w \in U^\perp$. Thus we have written $v = u + w$, where $u \in U$ and $w \in U^\perp$, completing the proof of 6.30.

If $v \in U \cap U^\perp$, then v (which is in U) is orthogonal to every vector in U (including v itself), which implies that $\langle v, v \rangle = 0$, which implies that $v = 0$. Thus

6.32
$$U \cap U^\perp = \{0\}.$$

Now 6.30 and 6.32 imply that $V = U \oplus U^\perp$ (see 1.9). ∎

The next corollary is an important consequence of the last theorem.

6.33 Corollary: *If U is a subspace of V, then*

$$U = (U^\perp)^\perp.$$

PROOF: Suppose that U is a subspace of V. First we will show that

6.34
$$U \subset (U^\perp)^\perp.$$

To do this, suppose that $u \in U$. Then $\langle u, v \rangle = 0$ for every $v \in U^\perp$ (by the definition of $U^\perp$). Because u is orthogonal to every vector in $U^\perp$, we have $u \in (U^\perp)^\perp$, completing the proof of 6.34.

To prove the inclusion in the other direction, suppose $v \in (U^\perp)^\perp$. By 6.29, we can write $v = u + w$, where $u \in U$ and $w \in U^\perp$. We have $v - u = w \in U^\perp$. Because $v \in (U^\perp)^\perp$ and $u \in (U^\perp)^\perp$ (from 6.34), we have $v - u \in (U^\perp)^\perp$. Thus $v - u \in U^\perp \cap (U^\perp)^\perp$, which implies that $v - u$ is orthogonal to itself, which implies that $v - u = 0$, which implies that $v = u$, which implies that $v \in U$. Thus $(U^\perp)^\perp \subset U$, which along with 6.34 completes the proof. ∎

Suppose U is a subspace of V. The decomposition $V = U \oplus U^\perp$ given by 6.29 means that each vector $v \in V$ can be written uniquely in the form

$$v = u + w,$$

where $u \in U$ and $w \in U^\perp$. We use this decomposition to define an operator on V, denoted P_U, called the ***orthogonal projection*** of V onto U. For $v \in V$, we define $P_U v$ to be the vector u in the decomposition above. In the notation introduced in the last chapter, we have $P_U = P_{U,U^\perp}$. You should verify that $P_U \in \mathcal{L}(V)$ and that it has the following properties:

- range $P_U = U$;
- null $P_U = U^\perp$;
- $v - P_U v \in U^\perp$ for every $v \in V$;
- $P_U{}^2 = P_U$;
- $\|P_U v\| \leq \|v\|$ for every $v \in V$.

Furthermore, from the decomposition 6.31 used in the proof of 6.29 we see that if $(e_1, \dots, e_m)$ is an orthonormal basis of U, then

6.35 $$P_U v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m$$

for every $v \in V$.

The following problem often arises: given a subspace U of V and a point $v \in V$, find a point $u \in U$ such that $\|v - u\|$ is as small as possible. The next proposition shows that this minimization problem is solved by taking $u = P_U v$.

6.36 Proposition: *Suppose U is a subspace of V and $v \in V$. Then*

$$\|v - P_U v\| \leq \|v - u\|$$

for every $u \in U$. Furthermore, if $u \in U$ and the inequality above is an equality, then $u = P_U v$.

The remarkable simplicity of the solution to this minimization problem has led to many applications of inner-product spaces outside of pure mathematics.

PROOF: Suppose $u \in U$. Then

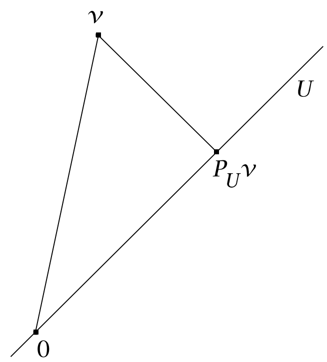
6.37 $$\|v - P_U v\|^2 \leq \|v - P_U v\|^2 + \|P_U v - u\|^2$$

6.38 $$= \|(v - P_U v) + (P_U v - u)\|^2$$

$$= \|v - u\|^2,$$

where 6.38 comes from the Pythagorean theorem (6.3), which applies because $v - P_U v \in U^\perp$ and $P_U v - u \in U$. Taking square roots gives the desired inequality.

Our inequality is an equality if and only if 6.37 is an equality, which happens if and only if $\|P_U v - u\| = 0$, which happens if and only if $u = P_U v$. ■



$P_U v$ is the closest point in U to v.

The last proposition is often combined with the formula 6.35 to compute explicit solutions to minimization problems. As an illustration of this procedure, consider the problem of finding a polynomial u with real coefficients and degree at most 5 that on the interval $[-\pi, \pi]$ approximates $\sin x$ as well as possible, in the sense that

$$\int_{-\pi}^{\pi} |\sin x - u(x)|^2 \, dx$$

is as small as possible. To solve this problem, let $C[-\pi, \pi]$ denote the real vector space of continuous real-valued functions on $[-\pi, \pi]$ with inner product

6.39 $$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x) g(x) \, dx.$$

Let $v \in C[-\pi, \pi]$ be the function defined by $v(x) = \sin x$. Let U denote the subspace of $C[-\pi, \pi]$ consisting of the polynomials with real coefficients and degree at most 5. Our problem can now be reformulated as follows: find $u \in U$ such that $\|v - u\|$ is as small as possible.

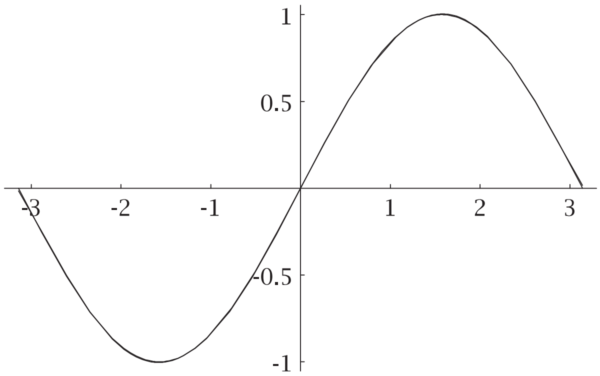
To compute the solution to our approximation problem, first apply the Gram-Schmidt procedure (using the inner product given by 6.39) to the basis $(1, x, x^2, x^3, x^4, x^5)$ of U, producing an orthonormal basis $(e_1, e_2, e_3, e_4, e_5, e_6)$ of U. Then, again using the inner product given by 6.39, compute $P_U v$ using 6.35 (with $m = 6$). Doing this computation shows that $P_U v$ is the function

A machine that can perform integrations is useful here.

6.40 $$0.987862x - 0.155271x^3 + 0.00564312x^5,$$

where the π's that appear in the exact answer have been replaced with a good decimal approximation.

By 6.36, the polynomial above should be about as good an approximation to $\sin x$ on $[-\pi, \pi]$ as is possible using polynomials of degree at most 5. To see how good this approximation is, the picture below shows the graphs of both $\sin x$ and our approximation 6.40 over the interval $[-\pi, \pi]$.



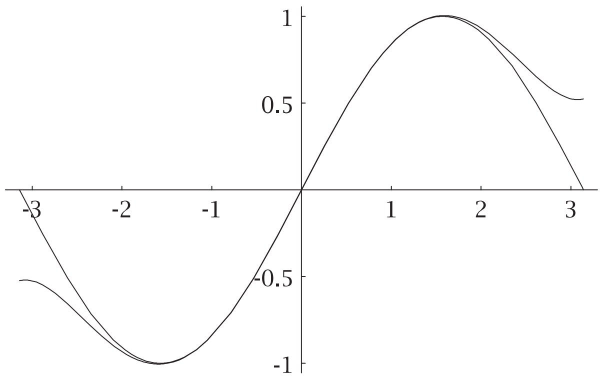
Graphs of $\sin x$ *and its approximation 6.40*

Our approximation 6.40 is so accurate that the two graphs are almost identical—our eyes may see only one graph!

Another well-known approximation to $\sin x$ by a polynomial of degree 5 is given by the Taylor polynomial

6.41 $$x - \frac{x^3}{3!} + \frac{x^5}{5!}.$$

To see how good this approximation is, the next picture shows the graphs of both $\sin x$ and the Taylor polynomial 6.41 over the interval $[-\pi, \pi]$.



Graphs of $\sin x$ *and the Taylor polynomial 6.41*

The Taylor polynomial is an excellent approximation to $\sin x$ for x near 0. But the picture above shows that for $|x| > 2$, the Taylor polynomial is not so accurate, especially compared to 6.40. For example, taking $x = 3$, our approximation 6.40 estimates $\sin 3$ with an error of about 0.001, but the Taylor series 6.41 estimates $\sin 3$ with an error of about 0.4. Thus at $x = 3$, the error in the Taylor series is hundreds of times larger than the error given by 6.40. Linear algebra has helped us discover an approximation to $\sin x$ that improves upon what we learned in calculus!

We derived our approximation 6.40 by using 6.35 and 6.36. Our standing assumption that V is finite dimensional fails when V equals $C[-\pi, \pi]$, so we need to justify our use of those results in this case. First, reread the proof of 6.29, which states that if U is a subspace of V, then

6.42 $$V = U \oplus U^{\perp}.$$

If we allow V to be infinite dimensional and allow U to be an infinite-dimensional subspace of V, then 6.42 is not necessarily true without additional hypotheses.

Note that the proof uses the finite dimensionality of U (to get a basis of U) but that it works fine regardless of whether or not V is finite dimensional. Second, note that the definition and properties of P_U (including 6.35) require only 6.29 and thus require only that U (but not necessarily V) be finite dimensional. Finally, note that the proof of 6.36 does not require the finite dimensionality of V. Conclusion: for $v \in V$ and U a subspace of V, the procedure discussed above for finding the vector $u \in U$ that makes $\|v - u\|$ as small as possible works if U is finite dimensional, regardless of whether or not V is finite dimensional. In the example above U was indeed finite dimensional (we had $\dim U = 6$), so everything works as expected.

Linear Functionals and Adjoints

A ***linear functional*** on V is a linear map from V to the scalars $\mathbf{F}$. For example, the function $\varphi\colon \mathbf{F}^3 \to \mathbf{F}$ defined by

6.43 $$\varphi(z_1, z_2, z_3) = 2z_1 - 5z_2 + z_3$$

is a linear functional on $\mathbf{F}^3$. As another example, consider the inner-product space $\mathcal{P}_6(\mathbf{R})$ (here the inner product is multiplication followed by integration on $[0, 1]$; see 6.2). The function $\varphi\colon \mathcal{P}_6(\mathbf{R}) \to \mathbf{R}$ defined by

6.44 $$\varphi(p) = \int_0^1 p(x)(\cos x)\,dx$$

is a linear functional on $\mathcal{P}_6(\mathbf{R})$.

If $v \in V$, then the map that sends u to $\langle u, v\rangle$ is a linear functional on V. The next result shows that every linear functional on V is of this form. To illustrate this theorem, note that for the linear functional φ defined by 6.43, we can take $v = (2, -5, 1) \in \mathbf{F}^3$. The linear functional φ defined by 6.44 better illustrates the power of the theorem below because for this linear functional, there is no obvious candidate for v (the function $\cos x$ is not eligible because it is not an element of $\mathcal{P}_6(\mathbf{R})$).

6.45 Theorem: *Suppose φ is a linear functional on V. Then there is a unique vector $v \in V$ such that*

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

for every $u \in V$.

Proof: First we show that there exists a vector $v \in V$ such that $\varphi(u) = \langle u, v\rangle$ for every $u \in V$. Let $(e_1, \dots, e_n)$ be an orthonormal basis of V. Then

$$\begin{aligned}\varphi(u) &= \varphi(\langle u, e_1\rangle e_1 + \cdots + \langle u, e_n\rangle e_n)\\ &= \langle u, e_1\rangle \varphi(e_1) + \cdots + \langle u, e_n\rangle \varphi(e_n)\\ &= \langle u, \overline{\varphi(e_1)}e_1 + \cdots + \overline{\varphi(e_n)}e_n\rangle\end{aligned}$$

for every $u \in V$, where the first equality comes from 6.17. Thus setting $v = \overline{\varphi(e_1)}e_1 + \cdots + \overline{\varphi(e_n)}e_n$, we have $\varphi(u) = \langle u, v\rangle$ for every $u \in V$, as desired.

Now we prove that only one vector $v \in V$ has the desired behavior. Suppose $v_1, v_2 \in V$ are such that

$$\varphi(u) = \langle u, v_1\rangle = \langle u, v_2\rangle$$

for every $u \in V$. Then

$$0 = \langle u, v_1\rangle - \langle u, v_2\rangle = \langle u, v_1 - v_2\rangle$$

for every $u \in V$. Taking $u = v_1 - v_2$ shows that $v_1 - v_2 = 0$. In other words, $v_1 = v_2$, completing the proof of the uniqueness part of the theorem. ■

In addition to V, we need another finite-dimensional inner-product space.

$$\boxed{\begin{array}{c}\text{Let's agree that for the rest of this chapter}\\ W \text{ is a finite-dimensional, nonzero, inner-product space over } \mathbf{F}.\end{array}}$$

*The word **adjoint** has another meaning in linear algebra. We will not need the second meaning, related to inverses, in this book. Just in case you encountered the second meaning for adjoint elsewhere, be warned that the two meanings for adjoint are unrelated to one another.*

Let $T \in \mathcal{L}(V, W)$. The ***adjoint*** of T, denoted T^*, is the function from W to V defined as follows. Fix $w \in W$. Consider the linear functional on V that maps $v \in V$ to $\langle Tv, w\rangle$. Let T^*w be the unique vector in V such that this linear functional is given by taking inner products with T^*w (6.45 guarantees the existence and uniqueness of a vector in V with this property). In other words, T^*w is the unique vector in V such that

$$\langle Tv, w\rangle = \langle v, T^*w\rangle$$

for all $v \in V$.

Let's work out an example of how the adjoint is computed. Define $T\colon \mathbf{R}^3 \to \mathbf{R}^2$ by

$$T(x_1, x_2, x_3) = (x_2 + 3x_3, 2x_1).$$

Thus T^* will be a function from $\mathbf{R}^2$ to $\mathbf{R}^3$. To compute T^*, fix a point $(y_1, y_2) \in \mathbf{R}^2$. Then

$$\begin{aligned}\langle (x_1, x_2, x_3), T^*(y_1, y_2)\rangle &= \langle T(x_1, x_2, x_3), (y_1, y_2)\rangle\\ &= \langle (x_2 + 3x_3, 2x_1), (y_1, y_2)\rangle\\ &= x_2y_1 + 3x_3y_1 + 2x_1y_2\\ &= \langle (x_1, x_2, x_3), (2y_2, y_1, 3y_1)\rangle\end{aligned}$$

for all $(x_1, x_2, x_3) \in \mathbf{R}^3$. This shows that

$$T^*(y_1, y_2) = (2y_2, y_1, 3y_1).$$

Note that in the example above, T^* turned out to be not just a function from $\mathbf{R}^2$ to $\mathbf{R}^3$, but a linear map. That is true in general. Specifically, if $T \in \mathcal{L}(V, W)$, then $T^* \in \mathcal{L}(W, V)$. To prove this, suppose $T \in \mathcal{L}(V, W)$. Let's begin by checking additivity. Fix $w_1, w_2 \in W$. Then

Adjoints play a crucial role in the important results in the next chapter.

$$\begin{aligned}
\langle Tv, w_1 + w_2 \rangle &= \langle Tv, w_1 \rangle + \langle Tv, w_2 \rangle \\
&= \langle v, T^* w_1 \rangle + \langle v, T^* w_2 \rangle \\
&= \langle v, T^* w_1 + T^* w_2 \rangle,
\end{aligned}$$

which shows that $T^* w_1 + T^* w_2$ plays the role required of $T^*(w_1 + w_2)$. Because only one vector can behave that way, we must have

$$T^* w_1 + T^* w_2 = T^*(w_1 + w_2).$$

Now let's check the homogeneity of T^*. If $a \in \mathbf{F}$, then

$$\begin{aligned}
\langle Tv, aw \rangle &= \bar{a} \langle Tv, w \rangle \\
&= \bar{a} \langle v, T^* w \rangle \\
&= \langle v, aT^* w \rangle,
\end{aligned}$$

which shows that $aT^* w$ plays the role required of $T^*(aw)$. Because only one vector can behave that way, we must have

$$aT^* w = T^*(aw).$$

Thus T^* is a linear map, as claimed.

You should verify that the function $T \mapsto T^*$ has the following properties:

additivity
$(S + T)^* = S^* + T^*$ for all $S, T \in \mathcal{L}(V, W)$;

conjugate homogeneity
$(aT)^* = \bar{a} T^*$ for all $a \in \mathbf{F}$ and $T \in \mathcal{L}(V, W)$;

adjoint of adjoint
$(T^*)^* = T$ for all $T \in \mathcal{L}(V, W)$;

identity
$I^* = I$, where I is the identity operator on V;

products

$(ST)^* = T^*S^*$ for all $T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, U)$ (here U is an inner-product space over $\mathbf{F}$).

The next result shows the relationship between the null space and the range of a linear map and its adjoint. The symbol $\iff$ means "if and only if"; this symbol could also be read to mean "is equivalent to".

6.46 Proposition: *Suppose $T \in \mathcal{L}(V, W)$. Then*

(a) $\operatorname{null} T^* = (\operatorname{range} T)^\perp$;

(b) $\operatorname{range} T^* = (\operatorname{null} T)^\perp$;

(c) $\operatorname{null} T = (\operatorname{range} T^*)^\perp$;

(d) $\operatorname{range} T = (\operatorname{null} T^*)^\perp$.

PROOF: Let's begin by proving (a). Let $w \in W$. Then

$$\begin{aligned} w \in \operatorname{null} T^* &\iff T^*w = 0 \\ &\iff \langle v, T^*w \rangle = 0 \text{ for all } v \in V \\ &\iff \langle Tv, w \rangle = 0 \text{ for all } v \in V \\ &\iff w \in (\operatorname{range} T)^\perp. \end{aligned}$$

Thus $\operatorname{null} T^* = (\operatorname{range} T)^\perp$, proving (a).

If we take the orthogonal complement of both sides of (a), we get (d), where we have used 6.33. Finally, replacing T with T^* in (a) and (d) gives (c) and (b). ■

If $\mathbf{F} = \mathbf{R}$, then the conjugate transpose of a matrix is the same as its ***transpose****, which is the matrix obtained by interchanging the rows and columns.*

The ***conjugate transpose*** of an m-by-n matrix is the n-by-m matrix obtained by interchanging the rows and columns and then taking the complex conjugate of each entry. For example, the conjugate transpose of

$$\begin{bmatrix} 2 & 3+4i & 7 \\ 6 & 5 & 8i \end{bmatrix}$$

is the matrix

$$\begin{bmatrix} 2 & 6 \\ 3-4i & 5 \\ 7 & -8i \end{bmatrix}.$$

The next proposition shows how to compute the matrix of T^* from the matrix of T. Caution: the proposition below applies only when we are dealing with orthonormal bases—with respect to nonorthonormal bases, the matrix of T^* does not necessarily equal the conjugate transpose of the matrix of T.

6.47 Proposition: *Suppose $T \in \mathcal{L}(V, W)$. If $(e_1, \dots, e_n)$ is an orthonormal basis of V and $(f_1, \dots, f_m)$ is an orthonormal basis of W, then*

$$\mathcal{M}(T^*, (f_1, \dots, f_m), (e_1, \dots, e_n))$$

is the conjugate transpose of

$$\mathcal{M}(T, (e_1, \dots, e_n), (f_1, \dots, f_m)).$$

The adjoint of a linear map does not depend on a choice of basis. This explains why we will emphasize adjoints of linear maps instead of conjugate transposes of matrices.

PROOF: Suppose that $(e_1, \dots, e_n)$ is an orthonormal basis of V and $(f_1, \dots, f_m)$ is an orthonormal basis of W. We write $\mathcal{M}(T)$ instead of the longer expression $\mathcal{M}(T, (e_1, \dots, e_n), (f_1, \dots, f_m))$; we also write $\mathcal{M}(T^*)$ instead of $\mathcal{M}(T^*, (f_1, \dots, f_m), (e_1, \dots, e_n))$.

Recall that we obtain the k^{th} column of $\mathcal{M}(T)$ by writing Te_k as a linear combination of the f_j's; the scalars used in this linear combination then become the k^{th} column of $\mathcal{M}(T)$. Because $(f_1, \dots, f_m)$ is an orthonormal basis of W, we know how to write Te_k as a linear combination of the f_j's (see 6.17):

$$Te_k = \langle Te_k, f_1 \rangle f_1 + \cdots + \langle Te_k, f_m \rangle f_m.$$

Thus the entry in row j, column k, of $\mathcal{M}(T)$ is $\langle Te_k, f_j \rangle$. Replacing T with T^* and interchanging the roles played by the e's and f's, we see that the entry in row j, column k, of $\mathcal{M}(T^*)$ is $\langle T^* f_k, e_j \rangle$, which equals $\langle f_k, Te_j \rangle$, which equals $\overline{\langle Te_j, f_k \rangle}$, which equals the complex conjugate of the entry in row k, column j, of $\mathcal{M}(T)$. In other words, $\mathcal{M}(T^*)$ equals the conjugate transpose of $\mathcal{M}(T)$. ∎

Exercises

1. Prove that if x, y are nonzero vectors in $\mathbf{R}^2$, then

$$\langle x, y \rangle = \|x\| \|y\| \cos\theta,$$

where θ is the angle between x and y (thinking of x and y as arrows with initial point at the origin). *Hint:* draw the triangle formed by x, y, and $x - y$; then use the law of cosines.

2. Suppose $u, v \in V$. Prove that $\langle u, v \rangle = 0$ if and only if

$$\|u\| \le \|u + av\|$$

for all $a \in \mathbf{F}$.

3. Prove that

$$\Big(\sum_{j=1}^{n} a_j b_j\Big)^2 \le \Big(\sum_{j=1}^{n} j {a_j}^2\Big)\Big(\sum_{j=1}^{n} \frac{{b_j}^2}{j}\Big)$$

for all real numbers $a_1, \dots, a_n$ and $b_1, \dots, b_n$.

4. Suppose $u, v \in V$ are such that

$$\|u\| = 3, \quad \|u + v\| = 4, \quad \|u - v\| = 6.$$

What number must $\|v\|$ equal?

5. Prove or disprove: there is an inner product on $\mathbf{R}^2$ such that the associated norm is given by

$$\|(x_1, x_2)\| = |x_1| + |x_2|$$

for all $(x_1, x_2) \in \mathbf{R}^2$.

6. Prove that if V is a real inner-product space, then

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

for all $u, v \in V$.

7. Prove that if V is a complex inner-product space, then

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

for all $u, v \in V$.

8. A norm on a vector space U is a function $\| \ \|: U \to [0, \infty)$ such that $\|u\| = 0$ if and only if $u = 0$, $\|\alpha u\| = |\alpha| \|u\|$ for all $\alpha \in \mathbf{F}$ and all $u \in U$, and $\|u + v\| \le \|u\| + \|v\|$ for all $u, v \in U$. Prove that a norm satisfying the parallelogram equality comes from an inner product (in other words, show that if $\| \ \|$ is a norm on U satisfying the parallelogram equality, then there is an inner product $\langle \ , \ \rangle$ on U such that $\|u\| = \langle u, u \rangle^{1/2}$ for all $u \in U$).

9. Suppose n is a positive integer. Prove that

$$\Big(\frac{1}{\sqrt{2\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \dots, \frac{\sin nx}{\sqrt{\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \dots, \frac{\cos nx}{\sqrt{\pi}}\Big)$$

is an orthonormal list of vectors in $C[-\pi, \pi]$, the vector space of continuous real-valued functions on $[-\pi, \pi]$ with inner product

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

This orthonormal list is often used for modeling periodic phenomena such as tides.

10. On $\mathcal{P}_2(\mathbf{R})$, consider the inner product given by

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

Apply the Gram-Schmidt procedure to the basis $(1, x, x^2)$ to produce an orthonormal basis of $\mathcal{P}_2(\mathbf{R})$.

11. What happens if the Gram-Schmidt procedure is applied to a list of vectors that is not linearly independent?

12. Suppose V is a real inner-product space and $(v_1, \dots, v_m)$ is a linearly independent list of vectors in V. Prove that there exist exactly 2^m orthonormal lists $(e_1, \dots, e_m)$ of vectors in V such that

$$\operatorname{span}(v_1, \dots, v_j) = \operatorname{span}(e_1, \dots, e_j)$$

for all $j \in \{1, \dots, m\}$.

13. Suppose $(e_1, \dots, e_m)$ is an orthonormal list of vectors in V. Let $v \in V$. Prove that

$$\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_m \rangle|^2$$

if and only if $v \in \operatorname{span}(e_1, \dots, e_m)$.

14. Find an orthonormal basis of $\mathcal{P}_2(\mathbf{R})$ (with inner product as in Exercise 10) such that the differentiation operator (the operator that takes p to p') on $\mathcal{P}_2(\mathbf{R})$ has an upper-triangular matrix with respect to this basis.

15. Suppose U is a subspace of V. Prove that
$$\dim U^{\perp} = \dim V - \dim U.$$

16. Suppose U is a subspace of V. Prove that $U^{\perp} = \{0\}$ if and only if $U = V$.

17. Prove that if $P \in \mathcal{L}(V)$ is such that $P^2 = P$ and every vector in $\operatorname{null} P$ is orthogonal to every vector in $\operatorname{range} P$, then P is an orthogonal projection.

18. Prove that if $P \in \mathcal{L}(V)$ is such that $P^2 = P$ and
$$\|Pv\| \leq \|v\|$$
for every $v \in V$, then P is an orthogonal projection.

19. Suppose $T \in \mathcal{L}(V)$ and U is a subspace of V. Prove that U is invariant under T if and only if $P_U T P_U = T P_U$.

20. Suppose $T \in \mathcal{L}(V)$ and U is a subspace of V. Prove that U and $U^{\perp}$ are both invariant under T if and only if $P_U T = T P_U$.

21. In $\mathbf{R}^4$, let
$$U = \operatorname{span}\big((1,1,0,0), (1,1,1,2)\big).$$
Find $u \in U$ such that $\|u - (1,2,3,4)\|$ is as small as possible.

22. Find $p \in \mathcal{P}_3(\mathbf{R})$ such that $p(0) = 0$, $p'(0) = 0$, and
$$\int_0^1 |2 + 3x - p(x)|^2 \, dx$$
is as small as possible.

23. Find $p \in \mathcal{P}_5(\mathbf{R})$ that makes
$$\int_{-\pi}^{\pi} |\sin x - p(x)|^2 \, dx$$
as small as possible. (The polynomial 6.40 is an excellent approximation to the answer to this exercise, but here you are asked to find the exact solution, which involves powers of π. A computer that can perform symbolic integration will be useful.)

24. Find a polynomial $q \in \mathcal{P}_2(\mathbf{R})$ such that

$$p(\frac{1}{2}) = \int_0^1 p(x)q(x)\,dx$$

for every $p \in \mathcal{P}_2(\mathbf{R})$.

25. Find a polynomial $q \in \mathcal{P}_2(\mathbf{R})$ such that

$$\int_0^1 p(x)(\cos \pi x)\,dx = \int_0^1 p(x)q(x)\,dx$$

for every $p \in \mathcal{P}_2(\mathbf{R})$.

26. Fix a vector $v \in V$ and define $T \in \mathcal{L}(V, \mathbf{F})$ by $Tu = \langle u, v \rangle$. For $a \in \mathbf{F}$, find a formula for T^*a.

27. Suppose n is a positive integer. Define $T \in \mathcal{L}(\mathbf{F}^n)$ by

$$T(z_1, \dots, z_n) = (0, z_1, \dots, z_{n-1}).$$

Find a formula for $T^*(z_1, \dots, z_n)$.

28. Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. Prove that λ is an eigenvalue of T if and only if $\bar{\lambda}$ is an eigenvalue of T^*.

29. Suppose $T \in \mathcal{L}(V)$ and U is a subspace of V. Prove that U is invariant under T if and only if $U^\perp$ is invariant under T^*.

30. Suppose $T \in \mathcal{L}(V, W)$. Prove that

(a) T is injective if and only if T^* is surjective;

(b) T is surjective if and only if T^* is injective.

31. Prove that

$$\dim \operatorname{null} T^* = \dim \operatorname{null} T + \dim W - \dim V$$

and

$$\dim \operatorname{range} T^* = \dim \operatorname{range} T$$

for every $T \in \mathcal{L}(V, W)$.

32. Suppose A is an m-by-n matrix of real numbers. Prove that the dimension of the span of the columns of A (in $\mathbf{R}^m$) equals the dimension of the span of the rows of A (in $\mathbf{R}^n$).