Chapter 6

Inner product spaces

Ordinary Euclidean space is a 3-dimensional vector space over $\mathbb{R}$, but it is more than that: the extra geometric structure (lengths, angles, etc.) can all be derived from a special kind of bilinear form on the space known as an inner product. We examine inner product spaces and their linear maps in this chapter.

One can also define inner products for complex vector spaces, but some adjustments need to be made. To avoid overloading the module, we will concentrate on real inner product spaces, and mention the adjustments required for complex spaces only briefly.

6.1 Inner products and orthonormal bases

Definition 6.1. An *inner product* on a real vector space V is a function that takes each pair of vectors $v, w \in V$ to a real number $v \cdot w$ satisfying the following conditions:

- The inner product is *symmetric*, that is, $v \cdot w = w \cdot v$ for all $v, w \in V$.
- The inner product is *bilinear*, that is, linear in the first variable when the second is kept constant and *vice versa*. (Symbolically, $(v + v') \cdot w = v \cdot w + v' \cdot w$ and $(av) \cdot w = a(v \cdot w)$ for all $v, v', w \in V$ and $a \in \mathbb{R}$.)
- The inner product is *positive definite*, that is, $v \cdot v \geq 0$ for all $v \in V$, and $v \cdot v = 0$ if and only if $v = \mathbf{0}$.

An inner product is sometimes called a *dot product* because of this notation.

Note that we don't need to insist that $v \cdot (w + w') = v \cdot w + v \cdot w'$ and $v \cdot (aw) = a(v \cdot w)$, for all $v, w, w' \in V$ and $a \in \mathbb{R}$, since these facts follow by symmetry from linearity in the first variable.

Geometrically, in a real vector space, we might define $v \cdot w = |v|\,|w| \cos\theta$, where $|v|$ and $|w|$ are the lengths of v and w, and θ is the angle between v and w. But we can easily reverse the order of doing things and define lengths and angles in terms of the inner product. Given an inner product on V, we define the *length* of any vector $v \in V$ to be

$$|v| = \sqrt{v \cdot v},$$

and, for any vectors $v, w \in V \setminus \{\mathbf{0}\}$, we define the *angle* between v and w to be θ, where

$$\cos\theta = \frac{v \cdot w}{|v|\,|w|}.$$

For this definition to make sense, we need to know that

$$-|v|\,|w| \leq v \cdot w \leq |v|\,|w|$$

for any vectors v, w (since $\cos\theta$ lies between -1 and 1). This is the content of the *Cauchy-Schwarz inequality*:

Theorem 6.2. *If v, w are vectors in an inner product space then*

$$(v \cdot w)^2 \leq (v \cdot v)(w \cdot w).$$

Proof. By definition, we have $(v + aw) \cdot (v + aw) \geq 0$ for any real number a. Expanding, we obtain

$$(w \cdot w)a^2 + 2(v \cdot w)a + (v \cdot v) \geq 0.$$

This is a quadratic function in a. Since it is non-negative for all real a, either it has no real roots, or it has two equal real roots; thus its discriminant is non-positive, that is,

$$(v \cdot w)^2 - (v \cdot v)(w \cdot w) \leq 0,$$

as required. □

Certain bases in an inner product space are particularly convenient to use.

Definition 6.3. A basis $(v_1, \ldots, v_n)$ for an inner product space is called *orthonormal* if $v_i \cdot v_j = \delta_{i,j}$ (the Kronecker delta) for $1 \leq i, j \leq n$.

Remark 6.4. If vectors $v_1, \ldots, v_n$ satisfy $v_i \cdot v_j = \delta_{i,j}$, then they are necessarily linearly independent. For suppose that $c_1 v_1 + \cdots + c_n v_n = \mathbf{0}$. Taking the inner product of this equation with v_i, we find that $c_i = 0$, for all i.

Theorem 6.5. *Let $\cdot$ be an inner product on a real vector space V. Then there is an orthonormal basis $\mathcal{B} = (v_1, \ldots, v_n)$ for V.*

Proof. The proof involves a constructive method for finding an orthonormal basis, known as the *Gram-Schmidt process*.

Let $w_1, \ldots, w_n$ be any basis for V. The Gram-Schmidt process works as follows.

Step 1 Since $w_1 \neq \mathbf{0}$, we have $w_1 \cdot w_1 > 0$, that is, $|w_1| > 0$. Let $v_1 = w_1/|w_1|$. [The effect of this step is to ensure that $v_1 \cdot v_1 = 1$, i.e., $|v_1| = 1$.]

Step 2 For $i = 2, \ldots, n$, let $w_i' = w_i - (v_1 \cdot w_i)v_1$. [The effect of this step is to ensure that

$$v_1 \cdot w_i' = v_1 \cdot w_i - (v_1 \cdot w_i)(v_1 \cdot v_1) = 0$$

for $2 \leq i \leq 2$, i.e., that v_1 is orthogonal to all of $w_2', \ldots, w_n'$.]

Step 3 Now apply the Gram-Schmidt process recursively to $w_2', \ldots, w_n'$ to obtain $v_2, \ldots, v_n$. [After this step, the vectors $v_1, \ldots, v_n$ form an orthonormal basis for V.]

Correctness of the Gram-Schmidt process can be seen by induction on n, the size of the initial basis. Correctness for $n = 1$ is clear. So suppose that $n \geq 2$ and that the process works for inputs with fewer than n basis vectors. We make some key observations.

- After Step 1, $v_1 \cdot v_1 = 1$.

- After Step 2, each of the original vectors w_i is a linear combination of v_1 and $w_2', \dots, w_n'$, so that $V = \langle v_1, w_2', \dots, w_n' \rangle$. It follows that $w_2', \dots, w_n'$ is a basis for some subspace of V (and hence is a valid input to Step 3 of the process).

- After Step 3, by the inductive hypothesis, $\langle v_2, \dots, v_n \rangle = \langle w_2', \dots, w_n' \rangle$, and $v_2, \dots, v_n$ are orthonormal.

- Also after Step 3, since v_1 is orthogonal to all of $w_2', \dots, w_n'$, it follows that v_1 is orthogonal to all of $v_2, \dots, v_n$.

Summarising, $v_i \cdot v_j = \delta_{i,j}$, for all $1 \le i, j \le n$, i.e., $v_1, \dots, v_n$ is an orthonormal basis of V. This completes the inductive step and the proof. □

There is essentially only one kind of inner product on a real vector space.

Proposition 6.6. *Suppose $\mathcal{B}$ is an orthonormal basis for the inner product space V of dimension n. If we represent vectors in coordinates with respect to $\mathcal{B}$, say $[v]_{\mathcal{B}} = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}^\top$ and $[w]_{\mathcal{B}} = \begin{bmatrix} b_1 & b_2 & \dots & b_n \end{bmatrix}^\top$, then*

$$v \cdot w = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n.$$

Proof. If $v = a_1 v_1 + \cdots + a_n v_n$ and $w = b_1 v_1 + \cdots + b_n v_n$, then

$$v \cdot w = (a_1 v_1 + \cdots + a_n v_n) \cdot (b_1 v_1 + \cdots + b_n v_n) = a_1 b_1 + \cdots + a_n b_n,$$

since all the cross terms are zero. □

Definition 6.7. The inner product on $\mathbb{R}^n$ for which the standard basis is orthonormal (that is, the one given in the above proposition) is called the *standard inner product* on $\mathbb{R}^n$.

Example 6.8. In $\mathbb{R}^3$ (with the standard inner product), apply the Gram-Schmidt process to the vectors $w_1 = \begin{bmatrix} 1 & 2 & 2 \end{bmatrix}^\top$, $w_2 = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}^\top$, $w_3 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^\top$.

We have $w_1 \cdot w_1 = 9$, so in the first step we put

$$v_1 = \tfrac{1}{3} w_1 = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}.$$

Now $v_1 \cdot w_2 = 1$ and $v_1 \cdot w_3 = \frac{1}{3}$, so in the second step we find

$$w_2' = w_2 - v_1 = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \\ -\frac{2}{3} \end{bmatrix} \quad \text{and} \quad w_3' = w_3 - \tfrac{1}{3} v_1 = \begin{bmatrix} \frac{8}{9} \\ -\frac{2}{9} \\ -\frac{2}{9} \end{bmatrix}.$$

Now we apply Gram-Schmidt recursively to w_2' and w_3'. We have $w_2' \cdot w_2' = 1$, so

$$v_2 = w_2' = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \\ -\frac{2}{3} \end{bmatrix}.$$

Then $v_2 \cdot w_3' = \frac{2}{3}$, so

$$w_3'' = w_3' - \tfrac{2}{3}v_2 = \begin{bmatrix} \frac{4}{9} \\ -\frac{4}{9} \\ \frac{2}{9} \end{bmatrix}.$$

Finally, $w_3'' \cdot w_3'' = \frac{4}{9}$, so

$$v_3 = \tfrac{3}{2}w_3'' = \begin{bmatrix} \frac{2}{3} \\ -\frac{2}{3} \\ \frac{1}{3} \end{bmatrix}.$$

You may check that the three vectors v_1, v_2 and v_3 we have found really do form an orthonormal basis.

Remark 6.9. If we reflect for a moment, we see that some changes must be made to deal with inner product spaces over $\mathbb{C}$. If V is a vector space over $\mathbb{C}$ and $v \in V$ any non-zero vector, then bilinearity of the inner product would imply $(iv) \cdot (iv) = i^2(v \cdot v) = -(v \cdot v)$. But then it cannot be the case that both $(iv) \cdot (iv)$ and $v \cdot v$ are positive real numbers, violating the requirement that the inner product should be positive definite. The fix is demand that the inner product satisfies *conjugate symmetry*, i.e., $v \cdot w = \overline{w \cdot v}$ rather than symmetry. (Overline here denotes complex conjugation.) A knock-on effect is that if we demand $(av) \cdot w = a(v \cdot w)$ then, necessarily, we must have $v \cdot (aw) = \bar{a}(v \cdot w)$. This follows from the chain of equalities

$$v \cdot (aw) = \overline{(aw) \cdot v} = \overline{a(w \cdot v)} = \bar{a}(\overline{w \cdot v}) = \bar{a}(v \cdot w).$$

We describe this situation by saying that the inner product is "sesquilinear". Note that these changes solve the problem we identified earlier, since now $(iv) \cdot (iv) = i(-i)v \cdot v = v \cdot v$. Note that the inner product of a vector with itself is certainly real, since $v \cdot v = \overline{v \cdot v}$.

6.2 Adjoints and orthogonal linear maps

Definition 6.10. Let V be an inner product space, and $\alpha : V \to V$ a linear map. Then the *adjoint* of α is the linear map $\alpha^* : V \to V$ defined by

$$v \cdot \alpha^*(w) = \alpha(v) \cdot w, \quad \text{for all } v \in V. \tag{6.1}$$

Given w, why should there exist a vector $\alpha^*(w)$ satisfying (6.1), and why should it be unique? If we fix w then the right hand side of (6.1) is a *linear functional* of v, that is, a function $V \to \mathbb{R}$ that is linear in its argument. There is a result, the *Riesz Representation Theorem*, that states that every linear functional in v has a unique representation as $v \cdot u$ for some $u \in V$. Thus $\alpha^*(w)$ exists and is unique. The Riesz Representation Theorem is beyond the scope of the course (though it is not so hard to prove).

Having seen that α^* is well defined, we need to show that α^* is linear. This is a short exercise on the final assignment, where you are also invited to show also that $\alpha^{**} = (\alpha^*)^*$ satisfies $\alpha^{**} = \alpha$.

Proposition 6.11. *If α is represented by the matrix A relative to an orthonormal basis of V, then α^* is represented by the transposed matrix $A^\top$.*

Proof. Denote the orthonormal basis by $\mathcal{B}$, and let the coordinate representations of vectors v and w in the basis $\mathcal{B}$ be $[v]_{\mathcal{B}} = \begin{bmatrix} b_1, \dots, b_n \end{bmatrix}^\top$ and $[w]_{\mathcal{B}} = \begin{bmatrix} c_1, \dots, c_n \end{bmatrix}^\top$; also let $A = (a_{i,j})$ and $A^* = (a^*_{i,j})$ be the representations of α and α^* in the basis $\mathcal{B}$. Then (6.1) expressed in the basis $\mathcal{B}$ becomes $[v]_{\mathcal{B}} \cdot (A^* [w]_{\mathcal{B}}) = (A [v]_{\mathcal{B}}) \cdot [w]_{\mathcal{B}}$, i.e.,

$$\sum_{i=1}^{n} b_i \sum_{j=1}^{n} a^*_{i,j} c_j = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} a_{j,i} b_i \right) c_j,$$

i.e.,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} b_i c_j a^*_{i,j} = \sum_{i=1}^{n} \sum_{j=1}^{n} b_i c_j a_{j,i}, \tag{6.2}$$

Since (6.1) holds for all v and w, and hence (6.2) holds for all $[v]_{\mathcal{B}}$ and $[w]_{\mathcal{B}}$, we can set $b_i = 1$ and $b_k = 0$ for all $k \neq i$, and also $c_j = 1$ and $c_k = 0$ for all $k \neq j$, to deduce from the above equation that $a^*_{i,j} = a_{j,i}$. Thus A^* is the transpose of A. □

Example 6.12. The matrix

$$A = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

represents a clockwise (looking down on the x, y-plane) rotation by $\pi/4$ about the z-axis in $\mathbb{R}^3$. Its adjoint is

$$A^* = A^\top = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and represents a rotation by $\pi/4$ anticlockwise.

Now we define two important classes of linear maps on V.

Definition 6.13. Let α be a linear map on an inner product space V.

(a) α is *self-adjoint* if $\alpha^* = \alpha$.

(b) α is *orthogonal* if it is invertible and $\alpha^* = \alpha^{-1}$.

There are several ways to look at orthogonal maps.

Theorem 6.14. *The following are equivalent for a linear map α on an inner product space V:*

(a) α is orthogonal;

(b) α preserves the inner product, that is, $\alpha(v) \cdot \alpha(w) = v \cdot w$;

(c) α maps any orthonormal basis of V to an orthonormal basis.

Proof. (a) ⇒ (b). Suppose α is orthogonal, so that $\alpha^*\alpha$ is the identity map. By the definition of adjoint,

$$\alpha(v) \cdot \alpha(w) = v \cdot \alpha^*(\alpha(w)) = v \cdot w.$$

(b) $\Rightarrow$ (c). Suppose that $(v_1, \dots, v_n)$ is an orthonormal basis, that is, $v_i \cdot v_j = \delta_{i,j}$ for all i, j. If (b) holds, then $\alpha(v_i) \cdot \alpha(v_j) = v_i \cdot v_j = \delta_{i,j}$, so that $(\alpha(v_1), \dots, \alpha(v_n))$ is an orthonormal basis, and (c) holds.

(c) $\Rightarrow$ (a). Suppose that α maps orthonormal basis $(v_1, \dots, v_n)$ to some other orthonormal basis $(\alpha(v_1), \dots, \alpha(v_n))$. Then, for all i, j, we have $\alpha(v_i) \cdot \alpha(v_j) = \delta_{i,j}$ and hence, by definition of adjoint, $v_i \cdot \alpha^*\alpha(v_j) = \delta_{i,j}$, for all i, j. Consider the effect of applying $\alpha^*\alpha$ to an arbitrary vector $v = c_1 v_1 + \cdots + c_n v_n$ expressed in terms of the basis vectors:

$$\alpha^*\alpha(v) = \alpha^*\alpha(c_1 v_1 + \cdots + c_n v_n) = c_1 \alpha^*\alpha(v_1) + \cdots + c_n \alpha^*\alpha(v_n).$$

The result can of course be expressed in terms of the basis vectors:

$$c_1 \alpha^*\alpha(v_1) + \cdots + c_n \alpha^*\alpha(v_n) = c_1' v_1 + \cdots + c_n' v_n,$$

for some scalars $c_1', \dots, c_n'$. Taking the inner product with v_i on both sides we obtain $c_i = c_i'$, since all but one term on each side is annihilated. So $\alpha^*\alpha$ is the identity map, and α is orthogonal. □

What do self-adjoint and orthogonal linear maps look like from the matrix perspective? The following is immediate from Proposition 6.11.

Corollary 6.15. *If α is represented by a matrix A (relative to an orthonormal basis), then*

(a) α is self-adjoint if and only if A is symmetric;

(b) α is orthogonal if and only if $A^\top A = I$.

Example 6.16. Returning to Example 6.12, the matrix $A^\top$ there is the inverse of the matrix A: the former is an anticlockwise rotation that undoes the clockwise rotation of the latter. Thus the matrix A represents an orthogonal linear map. Note that A preserves lengths and angles (see Theorem 6.14(b)) and maps orthonormal bases to orthonormal bases (see Theorem 6.14(c)).

A convenient characterisation of matrices representing orthogonal maps is the following.

Corollary 6.17. *Suppose the linear map $\alpha : V \to V$ is represented by the matrix A with respect to some orthonormal basis. Then α is orthogonal if and only if the columns of A form an orthonormal basis for V.*

Proof. Let A be the representation of linear map α with respect to some orthonormal basis. Let the columns of A be $v_1, \dots, v_n$, so that

$$A = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \quad \text{and} \quad A^\top = \begin{bmatrix} v_1^\top \\ v_2^\top \\ \vdots \\ v_n^\top \end{bmatrix}$$

and

$$A^\top A = \begin{bmatrix} v_1^\top v_1 & \dots & v_1^\top v_n \\ \vdots & & \vdots \\ v_n^\top v_1 & \dots & v_n^\top v_n \end{bmatrix} = \begin{bmatrix} v_1 \cdot v_1 & \dots & v_1 \cdot v_n \\ \vdots & & \vdots \\ v_n \cdot v_1 & \dots & v_n \cdot v_n \end{bmatrix}.$$

It is clear that $A^\top A = I$ if and only if $v_i \cdot v_j = \delta_{i,j}$, for $1 \leq i, j \leq n$, i.e., if and only if the vectors $v_1, \ldots, v_n$ are orthonormal. $\square$

Example 6.18. Returning again to Exercise 6.12, it can be checked that the columns of A (and hence the rows of A) are an orthonormal basis (relative to the standard orthonormal basis).

The above definitions suggest an equivalence relation on real matrices:

Definition 6.19. Two real $n \times n$ matrices are called *orthogonally similar* if and only if there is an orthogonal matrix P such that $A' = P^{-1}AP = P^\top AP$.

Here $P^{-1} = P^\top$ because P is orthogonal. Note that orthogonal similarity is a refinement of similarity. From Theorem 6.14 we know that orthogonal maps take orthonormal bases to orthonormal bases; in other words, orthogonal matrices are transition matrices between orthonormal bases. Thus, two real matrices A and A' are orthogonally similar if they represent the same linear map with respect to different orthonormal bases.

It is natural to ask when it is the case that a matrix is orthogonally similar to a diagonal matrix, and this is the question we turn to in the final chapter.