Chapter 4

Orthogonality, Inner Products and Projections

4.1 Inner Product Spaces

4.1.1 The ordinary scalar product on $\mathbb{R}^2$

In R^2, the scalar (or dot) product of the vectors $x = \binom{x_1}{x_2}$ and $y = \binom{y_1}{y_2}$ is given by

$$x \cdot y = x_1 y_1 + x_2 y_2 = x^T y = y^T x = y \cdot x.$$

We can interpret the length $||x||$ of the vector x as the length of the directed line segment from the origin to (x_1, x_2), which by the Theorem of Pythagoras is $\sqrt{x_1^2 + x_2^2}$ or $\sqrt{x \cdot x}$. Once we have a concept of length of a vector, we can define the distance $d(x, y)$ between two vectors x and y as the length of ther difference: $d(x, y) = ||x - y||$.

Similarly, from the Cosine Rule we can observe that $x \cdot y = ||x|| \, ||y|| \cos \theta$, where θ is the angle between the directed line segments representing x and y. In particular, x is orthogonal to y (or $x \perp y$) if and only if $x \cdot y = 0$.

So the scalar product encodes much of the geometry of $\mathbb{R}^2$, and it also provides a mechanism for defining concepts of length, distance and orthogonality on real vector spaces that do not necessarily have an obvious geometric structure.

4.1.2 Real Inner Products

Let V be a vector space over $\mathbb{R}$. An inner product on V is a function from $V \times V$ to $\mathbb{R}$ that assigns an element of $\mathbb{R}$ to every order pair of elements of V, and has the following properties. We write $\langle x, y \rangle$ for the inner product of x and y, and write the function as $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$.

1. Symmetry: $\langle x, y \rangle = \langle y, x \rangle$ for all $x, y \in V$
2. Linearity in both slots (bilinearity): For all $x, y, z \in V$ and all $a, b \in \mathbb{R}$, we have $\langle ax + by, z \rangle = a\langle x, z \rangle + b\langle y, z \rangle$ and $\langle x, ay + bz \rangle = a\langle x, y \rangle + b\langle x, z \rangle$.
3. Non-negativity: $\langle x, x \rangle \geqslant 0$ for all $x \in V$, and $\langle x, x \rangle = 0$ only if $x = 0_V$.

EXAMPLES We can check that each of the following satisfies the requirements to be an inner product.

1. The ordinary scalar product on $\mathbb{R}^n$.
2. Let C be the vector space of all continuous real-valued functions on the interval $[0, 1]$. The analogue of the ordinary scalar product on C is the inner product given by

$$\langle f, g \rangle = \int_0^1 f(x)g(x)\, dx, \text{ for } f, g \in C.$$

3. On the space $M_{m\times n}(\mathbb{R})$, the Frobenius inner product or trace inner product is defined by $\langle A, B\rangle = \text{trace}(A^T B)$. Note that trace ATB is the sum over all positions (i, j) of the products $A_{ij}B_{ij}$. So this is closely related to the ordinary scalar product, if the matrices A and B were regarded as vectors with mn entries over $\mathbb{R}$.

It is possible for a single vector space to have many different inner products defined on it, and if there is any risk of ambiguity we need to specify which one we are considering.

Given a real vector space and equipped with an inner product $\langle\cdot,\cdot\rangle$, we make the following two definitions.

Definition 4.1.1. *We define the length or norm of any vector v by*

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

and we define the distance between the vectors u and v by

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

Definition 4.1.2. *We say that vectors u and v are orthogonal (with respect to $\langle\cdot,\cdot\rangle$) if $\langle u, v\rangle = 0$.*

These definitions are consistent with "typical" geometrically motivated concepts of distance and orthogonality. First, we observe that the triangle inequality holds for the distance function defined on V. The triangle inequality captures the idea that if we want to travel from point A to point B, then travelling from A to a third point C, and then from C to B, should never amount to a shorter journey than travelling from A to B directly. It can have the same distance if C happens to be on a shortest path from A to B, but it can't be shorter.

The triangle inequality for an inner product space V is the statement that for any elements u, v, w of V,

$$d(u, v) \leqslant d(u, w) + d(w, v).$$

To prove the triangle inequality, we need to show that $\|u - v\| \leqslant \|u - w\| + \|w - v\|$ for all $u, v, w \in V$. since $u - v = (u - w) + (w - v)$, this will follow if we can show that $\|x + y\| \leqslant \|x\| + \|y\|$ for all $x, y \in V$. Now

$$\begin{aligned}
\|x + y\| \leqslant \|x\| + \|y\| \quad &\iff \sqrt{\langle x + y, x + y\rangle} \leqslant \sqrt{\langle x, x\rangle} + \sqrt{\langle y, y\rangle} \\
&\iff \sqrt{\langle x, x\rangle + 2\langle x, y\rangle +, \langle y, y\rangle} \leqslant \sqrt{\langle x, x\rangle} + \sqrt{\langle y, y\rangle} \\
&\iff \langle x, x\rangle + 2\langle x, y\rangle +, \langle y, y\rangle \leqslant \langle x, x + \langle y, y\rangle + 2\|x\|\|y\| \\
&\iff \langle x, y\rangle \leqslant \|x\|\|y\| \\
&\iff \langle x, y\rangle^2 \leqslant \langle x, x\rangle\langle y, y\rangle.
\end{aligned}$$

The last inequality here is known as the Cauchy-Schwarz Inequality and it is satisfied for all vectors x, y in V.

An element v of V is referred to as a unit vector if $\|v\| = 1$. The norm of elements of V has the property that $\|cv\| = |c|z, \|v\|$ for any vector v and real scalar c. To see this we can note that

$$\|cv\| = \sqrt{\langle cv, cv\rangle} = \sqrt{c^2\langle v, v\rangle} = c\|v\|.$$

So we can adjust the norm of any element of V, while preserving its direction, by multplying it by a positive scalar.

Definition 4.1.3. *If v is a non-zero vector in an inner product space V, then*

$$\hat{v} := \frac{1}{\|v\|}v$$

is a unit vector in the same direction as v, referred to as the normalization of v.

4.1.3 Orthogonal Projection

The following lemma says that any vector in an inner product space can be written as the sum of two orthogonal vectors, one in a pre-determined one-dimensional subspace and one orthogonal to that.

Lemma 4.1.4. *Let u and v be non-zero vectors in an inner product space V. Then it is possible to write $v = au + v'$, where a is scalar and v' is orthogonal to u.*

Proof. If v is orthogonal to u, or if v is a scalar multiple of u, there is nothing to do. In the first case, $a = 0$ and $v = v'$, and in the second case v' is the zero vector. Otherwise, to find a solution (for a) in the equation $v = au + v'$ (with $u \perp v'$), take an inner product with u on both sides. Then

$$\langle u, v\rangle = a\langle u, u\rangle + 0 \Longrightarrow a = \frac{\langle u, v\rangle}{||u||^2}.$$

We conclude that $v = \frac{\langle u,v\rangle}{||u||^2}u + (v - \frac{\langle u,v\rangle}{||u||^2}u)$, and it can be directly verified that the two components in this expression are orthogonal to each other with respect to the inner product. □

Example 4.1.5. *In $\mathbb{R}^2$, write $u = \binom{2}{1}$ and $v = \binom{6}{-2}$. Then*

$$||u||^2 = 5,\ u \cdot v = 10,\ \frac{u \cdot v}{||u||^2}u = \begin{pmatrix} 4 \\ 2 \end{pmatrix} \text{ and } v = \begin{pmatrix} 4 \\ 2 \end{pmatrix} + \begin{pmatrix} 2 \\ -4 \end{pmatrix}.$$

This is the unique expression for v as the sum of a scalar mutliple of u and a vector orthogonal to u.

Definition 4.1.6. *For non-zero vectors u and v in an inner product space V, the vector $\frac{\langle u, v\rangle}{||u||^2}u$ is called the projection of v on the 1-dimensional space spanned by u. It is denoted by $\text{proj}_u(v)$ and it has the property that $v - \text{proj}_u(v)$ is orthogonal to u.*

Lemma 4.1.7. $\text{proj}_u(v)$ *is the unique element of $\langle u\rangle$ whose distance from v is minimal.*

Proof. Let au be a scalar multiple of u. Then

$$d(au, v)^2 = \langle au - v, au - v\rangle = a^2\langle u, u\rangle - 2a\langle u, v\rangle + \langle v, v\rangle$$

Regarded as a quadratic function of a, this has a minimum when its derivative is 0, i.e. when $2a\langle u, u\rangle - 2\langle u, v\rangle = 0$, when $a = \frac{\langle u, v\rangle}{||u||^2}$. □

The concept and construction of orthogonal projection apply more generally than this. In general, if U is a subspace of an inner product space V, and v is a non-zero element of V, we can define $\text{proj}_U(v)$ to be the unique element u of U for which $(v - u)$ is orthogonal to every element of U. To see why such an element exists, we first consider orthogonal bases.

Let V be a finite dimensional inner product space with a basis $\mathcal{B} = \{b_1, \dots, b_n\}$. A basis of V is called orthogonal if its elements are all orthogonal to each other with respect to the inner product (it is called orthonormal if they are all unit vectors in addition).

We can adjust $\mathcal{B}$ to an orthogonal (or orthonormal) basis, using the following procedure, known as the Gram-Schmidt process.

1. Write $v_1 = b_1$.
2. Write $v_2 = b_2 - \text{proj}_{v_1}(b_2) = b_2 - \frac{\langle b_2, v_1\rangle}{\langle v_1, v_1\rangle}v_1$. Then $\langle v_2, v_1\rangle = 0$, and the linear independence of $\{v_1, v_2\}$ follows from that of $\{b_1, b_2\}$.
3. Write $v_3 = b_3 - \text{proj}_{v_1}(b_3) - \text{proj}_{v_2}(b_3) = b_3 - \frac{\langle b_3, v_1\rangle}{\langle v_1, v_1\rangle}v_1 - \frac{\langle b_3, v_2\rangle}{\langle v_2, v_2\rangle}v_2$.
   Then $\langle v_3, v_1\rangle = 0$, $\langle v_3, v_2\rangle = 0$, and v_3 is not a linear combination of v_1 and v_2 (since b_3 is not).
4. Continuing in this manner to adjust b_i to v_i, we arrive at an orthogonal basis. This can be adapted to an orthonormal basis, by scaling each basis element to a unit vector.

The Gram-Schmidt process is an algorithm for producing an orthogonal basis in an inner product space, from any basis as a starting point. It is also an justification for the assertion that every finite dimensional inner product space has an orthogonal (or orthonormal) basis. This second point can be used to define the concept of the ortogonal projection of a vector, not only onto a 1-dimensional subspace as we have seen, but on to any subspace.

Theorem 4.1.8. *Let $v \in V$, where V is an inner product space of dimension n, and let W be any subspace of V. Then there exist unique elements $u \in W$ and v' in V, for which $v = u + v'$, and v' s orthogonal to every element of W (we write $v' \in W^\perp$).*

Note: In this situation, u is called the orthogonal projection of v on W, and we write $u = \text{proj}_W(v)$.

Proof. Write k for the dimension of W. Using the Gram-Schidt process, we may construct an orthogonal basis $\{b_1, b_2, \ldots, b_k\}$ of W, and then (using Gram-Schmidt again), we may extend this to an orthogonal basis $\mathcal{B} = \{b_1, \ldots, b_k, c_{k+1}, \ldots, c_n\}$ of V. If we write v as a linear combination of the basis elements of V, we have

$$v = \sum_{i=0^k} r_i b_i + \sum_{j=k+1}^{n} s_j c_j,$$

for scalars r_i and s_j. Writing $u = \sum_{i=0^k} r_i b_i$ and $v' = \sum_{j=k+1}^{n} s_j c_j$ gives an expression of the required type.

For the uniqueness, suppose that $v = u + v'$ and also that $v = w + v''$, where both u and w belong to W, and both v' and v'' to $W^\perp$. Then since, $u - w = v' - v''$, it follows that $u - w$ belongs to both W and $W^\perp$. Since the zero vector is the only vector to be orthogonal to itself, it must be that $u = w$ and $v' = v''$. □

Example 4.1.9. *In $\mathbb{R}^3$, let W be the 2-dimensional space spanned by $u = (1, 2, 1)$ and $w = (4, -2, 0)$, and let $v = (5, 5, 2)$. Find $\text{proj}_W(v)$.*

First note that $u \perp w$, so that $\{u, w\}$ is an orthogonal basis of W. In this situation, $\text{proj}_W v$ is given by

$$\begin{aligned} \text{proj}_W(v) = \text{proj}_u(v) + \text{proj}_w(v) &= \frac{u \cdot v}{u \cdot u} u + \frac{w \cdot v}{w \cdot w} w \\ &= \frac{17}{6} u + \frac{10}{20} w \\ &= \frac{1}{6}(29, 28, 17). \end{aligned}$$

We can check directly that $v - \text{proj}_W(v) = \frac{1}{6}(1, 2, -5)$ is orthogonal to both u and w, hence to every element of W.

Lemma 4.1.10. *Let V be an inner product space with a subspace W, and let $v \in V$. Then $\text{proj}_W(v)$ is the nearest element of W to v, in terms of the distance determined by the inner product.*

Proof. Let $u = \text{proj}_W(v)$. For any element w of W, we have

$$\begin{aligned} d(v, w)^2 &= \langle v - w, v - w \rangle \\ &= \langle (v - u) + (u - w), (v - u) + (u - w) \rangle \\ &= \langle v - u, v - u \rangle + \underline{2\langle v - u, u - w \rangle} + \langle u - w, u - w \rangle \\ &\geqslant \langle v - u, v - u \rangle = d(v, u)^2, \end{aligned}$$

with equality only if $w = \text{proj}_W(v)$. □

Application: least squares for overdetermined systems

Example Consider the following overdetermined linear system.

$$\begin{array}{rcrcr} 2x & + & y & = & 3 \\ x & - & y & = & 0 \\ x & - & 3y & = & -4 \end{array} \qquad \begin{bmatrix} 2 & 1 \\ 1 & -1 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ -4 \end{bmatrix}$$

This system has three equations and only two variables. It is inconsistent and overdetermined - each pair of equations has a simultaneous solution, but all three do not.

Overdetermined systems arise quite often in applications, from observed data. Even if they do not have exact solutions, approximate solutions are of interest.

For a vector $b \in \mathbb{R}^3$, the system

$$\underbrace{\begin{bmatrix} 2 & 1 \\ 1 & -1 \\ 1 & -3 \end{bmatrix}}_{A} \begin{bmatrix} x \\ y \end{bmatrix} = b$$

has a solution if and only if b belongs to the 2-dimensional linear span W of the columns of the coefficient matrix A: $v_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ and $v_2 = \begin{bmatrix} 1 \\ -1 \\ -3 \end{bmatrix}$.

If not, then the nearest element of W to B is $b' = \text{proj}_W(b)$, and our approximate solutions for x and y are the entries of the vector c in $\mathbb{R}^2$ for which $Ac = b'$. We know that $b' - b$ is orthogonal to v_1 and v_2, which are the rows of A^T. Hence

$$A^T(b' - b) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Longrightarrow A^T b' = A^T A c = A^T b \Longrightarrow c = (A^T A)^{-1} A^T b$$

In our example,

$$\begin{bmatrix} 2 & 1 \\ 1 & -1 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ -4 \end{bmatrix}$$

$$A = \begin{bmatrix} 2 & 1 \\ 1 & -1 \\ 1 & -3 \end{bmatrix}, \; A^T = \begin{bmatrix} 2 & 1 & 1 \\ 1 & -1 & -3 \end{bmatrix}, \; A^T A = \begin{bmatrix} 6 & -2 \\ -2 & 11 \end{bmatrix}, \; A^T b = \begin{bmatrix} 2 \\ 15 \end{bmatrix}.$$

The least squares solution is given by

$$\begin{bmatrix} x \\ y \end{bmatrix} = c = (A^T A)^{-1} A^T b = \frac{1}{62} \begin{bmatrix} 11 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 2 \\ 15 \end{bmatrix} = \begin{bmatrix} \frac{26}{31} \\ \\ \frac{47}{31} \end{bmatrix}.$$