Inner Products and Orthogonality

This week we will learn about:

- Inner products (and the dot product again),
- The norm induced by the inner product,
- The Cauchy–Schwarz and triangle inequalities, and
- Orthogonality.

Extra reading and watching:

- Sections 1.3.4 and 1.4.1 in the textbook
- Lecture videos 17, 18, 19, 20, 21, and 22 on YouTube
- Inner product space at Wikipedia
- Cauchy–Schwarz inequality at Wikipedia
- Gram–Schmidt process at Wikipedia

Extra textbook problems:

⋆ 1.3.3, 1.3.4, 1.4.1

⋆⋆ 1.3.9, 1.3.10, 1.3.12, 1.3.13, 1.4.2, 1.4.5(a,d)

⋆⋆⋆ 1.3.11, 1.3.14, 1.3.15, 1.3.25, 1.4.16

☠ 1.3.18

There are many times when we would like to be able to talk about the angle between vectors in a vector space $\mathcal{V}$, and in particular orthogonality of vectors, just like we did in $\mathbb{R}^n$ in the previous course. This requires us to have a generalization of the dot product to arbitrary vector spaces.

Definition 5.1 — Inner Product

Suppose that $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $\mathcal{V}$ is a vector space over $\mathbb{F}$. Then an **inner product** on $\mathcal{V}$ is a function $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \to \mathbb{F}$ such that the following three properties hold for all $c \in \mathbb{F}$ and all $\mathbf{v}, \mathbf{w}, \mathbf{x} \in \mathcal{V}$:

a) $\langle \mathbf{v}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{v} \rangle}$ (conjugate symmetry)

b) $\langle \mathbf{v}, \mathbf{w} + c\mathbf{x} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle + c\langle \mathbf{v}, \mathbf{x} \rangle$ (linearity in 2nd entry)

c) $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$, with equality if and only if $\mathbf{v} = \mathbf{0}$. (positive definiteness)

- Why those three properties?

- Inner products are *not* linear in their first argument...

- OK, so why does property (a) have that weird complex conjugation in it?

- For this reason, they are sometimes called “sesquilinear”, which means...

Example. *Show that the following function is an inner product on* $\mathbb{C}^n$*:*

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^* \mathbf{w} = \sum_{i=1}^{n} \overline{v_i} w_i \quad \textit{for all} \quad \mathbf{v}, \mathbf{w} \in \mathbb{C}^n.$$

Example. *Let* $a < b$ *be real numbers and let* $\mathcal{C}[a,b]$ *be the vector space of continuous functions on the interval* $[a,b]$*. Show that the following function is an inner product on* $\mathcal{C}[a,b]$*:*

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

The previous examples are the "standard" inner products on those vector spaces. However, inner products can also be much uglier. The following example illustrates how the same vector space can have multiple different inner products, and at first glance they might look nothing like the standard inner products.

Example. *Show that the following function is an inner product on* $\mathbb{R}^2$*:*

$$\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + 2v_1 w_2 + 2v_2 w_1 + 5v_2 w_2 \quad \text{for all} \quad \mathbf{v}, \mathbf{w} \in \mathbb{R}^2.$$

There is also a "standard" inner product on $\mathcal{M}_n$, but before being able to explain it, we need to introduce the following helper function:

Definition 5.2 — Trace

Let $A \in \mathcal{M}_n$ be a square matrix. Then the **trace** of A, denoted by $\mathrm{tr}(A)$, is the sum of its diagonal entries:

$$\mathrm{tr}(A) \stackrel{\text{def}}{=} a_{1,1} + a_{2,2} + \cdots + a_{n,n}.$$

Example. *Compute the following matrix traces:*

The reason why the trace is such a wonderful function is that it makes matrix multiplication "kind of" commutative:

Theorem 5.1 — Commutativity of the Trace

Let $A \in \mathcal{M}_{m,n}$ and $B \in \mathcal{M}_{n,m}$ be matrices. Then

$$\mathrm{tr}(AB) = \mathrm{tr}(BA).$$

Proof. Just directly compute the diagonal entries of AB and BA:

■

The trace also has some other nice properties that are easier to see:

With the trace in hand, we can now introduce the standard inner product on the vector space of matrices:

Example. *Show that the following function is an inner product on* $\mathcal{M}_{m,n}$*:*

$$\langle A, B\rangle = \mathrm{tr}(A^*B) \quad \textit{for all} \quad A, B \in \mathcal{M}_{m,n}.$$

The above inner product is typically called the **Frobenius inner product** or **Hilbert–Schmidt inner product**. Also, a vector space together with a particular inner product is called an **inner product space**.

Norm Induced by the Inner Product

Now that we have inner products, we can define the length of a vector in a manner completely analogous to how we did it with the dot product in $\mathbb{R}^n$. However, in this more general setting, we are a bit beyond the point of being able to draw a geometric picture of what length means (for example, what is the "length" of a continuous function?), so we change terminology slightly and instead call this function a "norm."

Definition 5.3 — Norm Induced by the Inner Product

Suppose that $\mathcal{V}$ is an inner product space. Then the **norm induced by the inner product** is the function $\|\cdot\| : \mathcal{V} \to \mathbb{R}$ defined by

$$\|\mathbf{v}\| \stackrel{\text{def}}{=} \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} \quad \text{for all} \quad \mathbf{v} \in \mathcal{V}.$$

Example. *What is the norm induced by the standard inner product on $\mathbb{C}^n$?*

Example. *What is the norm induced by the standard inner product on $\mathcal{C}[a, b]$?*

Example. *What is the norm induced by the standard (Frobenius) inner product on $\mathcal{M}_{m,n}$?*

Perhaps not surprisingly, the norm induced by an inner product satisfies the same basic properties as the length of a vector in $\mathbb{R}^n$. These properties are summarized in the following theorem.

Theorem 5.2 — Properties of the Norm Induced by the I.P.

Suppose that $\mathcal{V}$ is an inner product space, $\mathbf{v} \in \mathcal{V}$ is a vector, and $c \in \mathbb{F}$ is a scalar. Then the following properties of the norm induced by the inner product hold:

a) $\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$, and

b) $\|\mathbf{v}\| \geq 0$, with equality if and only if $\mathbf{v} = \mathbf{0}$.

The two other main theorems that we proved for the length in $\mathbb{R}^n$ were the Cauchy–Schwarz inequality and the triangle inequality. We now show that these same properties hold for the norm induced by any inner product.

Theorem 5.3 — Cauchy–Schwarz Inequality

Suppose that $\mathcal{V}$ is an inner product space and $\mathbf{v}, \mathbf{w} \in \mathcal{V}$. Then

$$|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\|\|\mathbf{w}\|.$$

Furthermore, equality holds if and only if $\{\mathbf{v}, \mathbf{w}\}$ is a linearly dependent set.

Proof. Let $c, d \in \mathbb{F}$ be arbitrary scalars, and expand $\|c\mathbf{v}+d\mathbf{w}\|^2$ in terms of the inner product:

■

For example, if we apply the Cauchy–Schwarz inequality to the Frobenius inner product on $\mathcal{M}_{m,n}$, it tells us that

and if we apply it to the standard inner product on $\mathcal{C}[a, b]$ then it says that

Neither of the above inequalities are particularly pleasant to prove directly.

Just as was the case in $\mathbb{R}^n$, the triangle inequality now follows very quickly from the Cauchy–Schwarz inequality.

Theorem 5.4 — The Triangle Inequality

Suppose that $\mathcal{V}$ is an inner product space and $\mathbf{v}, \mathbf{w} \in \mathcal{V}$. Then

$$\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|.$$

Furthermore, equality holds if and only if $\mathbf{v}$ and $\mathbf{w}$ point in the same direction (i.e., $\mathbf{v} = \mathbf{0}$ or $\mathbf{w} = c\mathbf{v}$ for some $0 \leq c \in \mathbb{R}$).

Proof. Start by expanding $\|\mathbf{v} + \mathbf{w}\|^2$ in terms of the inner product:

■

Orthogonality

The most useful thing that we can do with an inner product is re-introduce orthogonality in this more general setting:

Definition 5.4 — Orthogonality

Suppose $\mathcal{V}$ is an inner product space. Then two vectors $\mathbf{v}, \mathbf{w} \in \mathcal{V}$ are called **orthogonal** if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$.

In $\mathbb{R}^n$, we could think of "orthogonal" as a synonym for "perpendicular", since two vectors were orthogonal if and only if the angle between them was $\pi/2$. In general inner product spaces this geometric picture makes much less sense (for example, what does it mean for the angle between two polynomials to be $\pi/2$?), so it is perhaps better to think of orthogonal vectors as ones that are "as linearly independent as possible."'

With this intuition in mind, it is useful to extend orthogonality to *sets* of vectors, rather than just pairs of vectors:

Definition 5.5 — Orthonormal Bases

A basis B of an inner product space $\mathcal{V}$ is called an **orthonormal basis** of $\mathcal{V}$ if

a) $\langle \mathbf{v}, \mathbf{w} \rangle = 0$ for all $\mathbf{v} \neq \mathbf{w} \in B$, and (mutual orthogonality)

b) $\|\mathbf{v}\| = 1$ for all $\mathbf{v} \in B$. (normalization)

Example. *Examples of orthonormal bases in our "standard" vector spaces include...*

Orthogonal and orthonormal bases often greatly simplify calculations. For example, the following theorem shows us that linear independence comes for free when we know that a set of vectors are mutually orthogonal.

Theorem 5.5 — Orthogonality Implies Linear Independence

Let $\mathcal{V}$ be an inner product space and suppose that the set $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \subset \mathcal{V}$ consists of non-zero mutually orthogonal vectors (i.e., $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ whenever $i \neq j$). Then B is linearly independent.

Proof. Suppose $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$. Then...

∎

A fairly quick consequence of the previous theorem is the fact that if a set of non-zero vectors is mutually orthogonal, and their number matches the dimension of the vector space, then...

Example. *Show that the set of Pauli matrices*

$$B = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \right\}$$

is an orthogonal basis of $\mathcal{M}_2(\mathbb{C})$. How could you turn it into an orthonormal basis?

We already learned that all finite-dimensional vector spaces are isomorphic (i.e., "essentially the same") to $\mathbb{F}^n$. It thus seems natural to ask the corresponding question about inner products—do all inner products on $\mathbb{F}^n$ look like the usual dot product on $\mathbb{F}^n$ in some basis? Orthonormal bases let us show that the answer is "yes."

Theorem 5.6 — All Inner Products Look Like the Dot Product

Suppose that B is an orthonormal basis of a finite-dimensional inner product space $\mathcal{V}$. Then

$$\langle \mathbf{v}, \mathbf{w} \rangle = [\mathbf{v}]_B \cdot [\mathbf{w}]_B \quad \text{for all} \quad \mathbf{v}, \mathbf{w} \in \mathcal{V}.$$

Proof. Write $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_n\}$. Since B is a basis of $\mathcal{V}$, we can write $\mathbf{v} = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n$ and $\mathbf{w} = d_1\mathbf{u}_1 + \cdots + d_n\mathbf{u}_n$. Then...

■

If we specialize even further to $\mathbb{C}^n$ rather than to an arbitrary finite-dimensional vector space $\mathcal{V}$, then we can say even more. Specifically, recall that if $\mathbf{v}, \mathbf{w} \in \mathbb{C}^n$, E is the standard basis of $\mathbb{C}^n$, and B is any basis of $\mathbb{C}^n$, then

By plugging this fact into the above characterization of finite-dimensional inner product spaces (and assuming that B is orthonormal), we see that every inner product on $\mathbb{C}^n$ has the form

We state this fact in a slightly cleaner form below:

Corollary 5.7 — Invertible Matrices Make Inner Products

A function $\langle \cdot, \cdot \rangle : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ is an inner product if and only if there exists an invertible matrix $P \in \mathcal{M}_n(\mathbb{F})$ such that

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^*(P^*P)\mathbf{w} \quad \text{for all} \quad \mathbf{v}, \mathbf{w} \in \mathbb{F}^n.$$

For example, the usual inner product (i.e., the dot product) on $\mathbb{C}^n$ arises when $P = I$. Similarly, the weird inner product on $\mathbb{R}^2$ from a few pages ago, defined by

$$\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + 2v_1 w_2 + 2v_2 w_1 + 5v_2 w_2 \quad \text{for all} \quad \mathbf{v}, \mathbf{w} \in \mathbb{R}^2,$$

is what we get if we choose $P = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$. To see this, we verify that

Orthogonalization

We already showed how to determine whether or not a particular set *is* an orthonormal basis, so let's turn to the question of how to *construct* an orthonormal basis. While this is reasonably intuitive in familiar inner product spaces like $\mathbb{R}^n$ or $\mathcal{M}_{m,n}(\mathbb{C})$, it becomes a bit more delicate when working in stranger inner products.

The process works one vector at a time to turn the vectors from some (not necessarily orthonormal) basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ into an orthonormal basis $C = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$. We start by simply defining

To construct the next member of our orthonormal basis, we define

In words, we are subtracting the portion of $\mathbf{v}_2$ that points in the direction of $\mathbf{u}_1$, leaving behind only the piece of it that is orthogonal to $\mathbf{u}_1$, as illustrated on the next page.

In higher dimensions, we would then continue in this way, adjusting each vector in the basis so that it is orthogonal to each of the previous vectors, and then normalizing it. The following theorem makes this precise and tells us that the result is indeed always an orthonormal basis.

Theorem 5.8 — Gram–Schmidt Process

Suppose $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis of an inner product space $\mathcal{V}$. Define

Then $C = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is an orthonormal basis of $\mathcal{V}$.

Proof. We actually prove that, not only is C an orthonormal basis of $\mathcal{V}$, but also that

for all $1 \leq k \leq n$.

We prove this result by induction on k. For the base case of $k = 1$, ...

$\blacksquare$

Since finite-dimensional inner product spaces (by definition) have a basis consisting of finitely many vectors, and the Gram–Schmidt process tells us how to convert that basis into an orthonormal basis, we now know that every finite-dimensional inner product space has an orthonormal basis:

Corollary 5.9 — Existence of Orthonormal Bases

Every finite-dimensional inner product space has an orthonormal basis.

Example. *Find an orthonormal basis for $\mathcal{P}^2[-1,1]$ with respect to the inner product*

$$\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\ dx.$$