Inner Products and Norms

We would like to have a better understanding of the geometry of a given vector space; in particular, we would like to have some notion of "length" of a vector and "angle" between a pair of vectors. Of course, we already have a way to talk about lengths and angles in Euclidean space using the dot product. In this section, we will generalize the dot product to arbitrary vector spaces and use the generalization to discuss the inherent geometry of those spaces.

Review of the Dot Product

Definition 6.2. Given

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},\ y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n,$$

the *dot product* $x \cdot y$ is the number

$$x \cdot y = x_1y_1 + x_2y_2 + \ldots + x_ny_n.$$

On first inspection, the dot product seems a bit useless; however, it does actually give us tools to measure length and angles in $\mathbb{R}^n$.

Indeed, the length $||x||$ of vector $x \in \mathbb{R}^n$ is precisely

$$\begin{aligned} ||x|| &= \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2} \\ &= \sqrt{x \cdot x}. \end{aligned}$$

In addition, you probably remember from calculus that

$$\cos\theta = \frac{x \cdot y}{||x||\ ||y||},$$

where θ is the angle between x and y.

Thus the dot product really is a helpful geometric tool–it provides us with data about lengths and angles in $\mathbb{R}^n$.

Inherent in the utility of the dot product are its many properties. It is easy to see that

1. $||x|| \geq 0$;
2. $||x|| = 0$ if and only if $x = \mathbf{0}$;
3. $x \cdot y = y \cdot x$.

Generalization of the Dot Product

The following definition for the *inner product* is a generalization of the dot product. We will see shortly that this definition will give us the same geometric tools as does the dot product:

Definition 6.3. An *inner product* on a vector space V over $\mathbb{F}$ is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ (that is, $\langle u,\ v \rangle \in \mathbb{F}$) satisfiying the following properties:

1. $\langle v,\ v \rangle \geq 0$ for all $v \in V$;
2. $\langle v,\ v \rangle = 0$ if and only if $v = \mathbf{0}$;
3. $\langle u + v,\ w \rangle = \langle u,\ w \rangle + \langle v,\ w \rangle$ for all $u, v, w \in W$;
4. $\langle \lambda u,\ v \rangle = \lambda \langle u,\ v \rangle$ for all $u, v \in V$;
5. $\langle u,\ v \rangle = \overline{\langle v,\ u \rangle}$ for all $u, v \in V$.

Example. In $\mathbb{R}^n$, the dot product

$$\langle x, y \rangle = x \cdot y$$

is an example of an inner product (as we would expect, given that the dot product was our motivation for defining the inner product). We will not check the conditions explicitly, but it is easy to see that the dot product obeys all of the conditions from the theorem.

Example. We can build an inner product on $\mathbb{C}^n$ in a similar manner: given

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},\ y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{C}^n,$$

set

$$\langle x,\ y \rangle = x_1\overline{y_1} + x_2\overline{y_2} + \ldots + x_n\overline{y_n}.$$

Notice that the definition is a bit more complicated than in $\mathbb{R}^n$; indeed, we have to take the *complex conjugate* of each entry of the second vector in order to guarantee that property 5 from the definition is satisfied.

Again, we will not prove that the function defined above is an inner product on $\mathbb{C}^n$; it is relatively easy to see that our definition satisfies all of the requisite properties.

Example. Given $A,\ B \in \mathcal{M}_n(\mathbb{F})$, we define an inner product by

$$\langle A,\ B\rangle = \operatorname{tr}(B^* A),$$

where $A^* = \overline{A^\top}$.

Again, it is easy to see that this definition satisfies the requirements to be an inner product on $\in \mathcal{M}_n(\mathbb{F})$; let's check that property 5 holds:

$$\begin{aligned} \langle B, A\rangle &= \operatorname{tr}(A^* B) \\ &= \overline{\operatorname{tr}((A^* B)^*)} \\ &= \overline{(\operatorname{tr}(B^* A))} \\ &= \overline{\langle B, A\rangle}, \end{aligned}$$

as required.

Example. The set $\mathcal{C}[a, b]$ of all continuous real-valued functions on an interval $[a, b]$ is a vector space under the usual notion of function addition and scalar products. For example,

$$f(x) = \tan x \text{ and } g(x) = \frac{1}{x - \pi/2}$$

are both vectors in $\mathcal{C}[-\pi/4, \pi/4]$.

A function that is continuous on an $[a, b]$ is also *integrable* on $[a, b]$; we can use this fact to define an inner product on $\mathcal{C}[a, b]$ by

$$\langle f, g\rangle = \int_a^b f(x)g(x)\ \mathrm{d}x.$$

For example, $f(x) = 1$ and $g(x) = x^2$ are both vectors in $C[-1, 1]$. The inner product $\langle f, g\rangle$ is given by

$$\begin{aligned} \langle f, g\rangle &= \int_{-1}^{1} 1 \cdot x^2\ \mathrm{d}x \\ &= \int_{-1}^{1} x^2\ \mathrm{d}x \\ &= \frac{1}{3}x^3\ \Big|_{-1}^{1} \\ &= \frac{1}{3} + \frac{1}{3} \\ &= \frac{2}{3}. \end{aligned}$$

Again, it is easy to see that this function does satisfy the definition, and is thus an inner product on $\mathcal{C}[a, b]$.

Example. Given vectors

$$f(x) = a_0 + a_1x + \ldots + a_nx^n \text{ and } g(x) = b_0 + b_1x + \ldots + b_nx^n$$

in $\mathcal{P}_n(\mathbb{R})$, the standard inner product is defined by

$$\langle f, g \rangle = a_0b_0 + a_1b_1 + \ldots + a_nb_n.$$

For example, the inner product $\langle f, g \rangle$ of vectors $f(x) = 1 - x^3 + 2x^5$ and $g(x) = 2x - x^2 + x^3$ is given by

$$\begin{aligned} \langle f, g \rangle &= 1 \cdot 0 + (0 \cdot 2) + (0 \cdot -1) + (-1 \cdot 1) + (2 \cdot 0) \\ &= -1. \end{aligned}$$

Remark. The standard inner products defined above are not the only examples of inner products on the spaces; indeed, a particular vector space may have many different inner products, as indicated by the next example.

Example. Given $w_1, w_2, \ldots, w_n \in \mathbb{R}$ with each $w_i \geq 0$, the *weighted inner product* with weights $w_1, w_2, \ldots, w_n$ is defined on $\mathbb{C}^n$ by

$$\langle x,\ y \rangle = w_1x_1\overline{y_1} + w_2x_2\overline{y_2} + \ldots + w_nx_n\overline{y_n}.$$

Once more, the function is clearly an inner product; in particular, it is easy to see that there are *infinitely many* such inner products on $\mathbb{C}^n$.

Inner Product Spaces

An inner product endows a vector space with extra structure; thus when we wish to understand a vector space in the light of this extra structure, we indicate that fact with new terminology:

Definition 6.5. An *inner product space* is a vector space V along with an inner product on V.

The basic definition of an inner product actually gives us a great deal more structure than that indicated in the definition for an inner product. Indeed, the following theorem follows immediately from the definition:

Theorem 6.7. Let V be an inner product space with inner product $\langle \cdot, \cdot \rangle$.

(a) Given fixed $u \in V$, the function $\langle \cdot, u \rangle : V \to \mathbb{F}$ that sends $v \in V$ to $\langle v, u \rangle$ is a linear transformation.

(b) $\langle \mathbf{0}, u \rangle = 0$ for all $u \in V$.

(c) $\langle u, \mathbf{0} \rangle = 0$ for all $u \in V$.

(d) $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$ for all $u,\ v,\ w \in \mathbb{F}$.

(e) $\langle u, \lambda v \rangle = \overline{\lambda} \langle u, v \rangle$ for all $u,\ v \in V,\ \lambda \in \mathbb{F}$.

Proof. (a) Follows immediately from parts 3 and 4 of the definition of an inner product.

(b) Since

$$\begin{aligned} \langle \mathbf{0}, u \rangle &= \langle \mathbf{0} + \mathbf{0}, u \rangle \\ &= \langle \mathbf{0}, u \rangle + \langle \mathbf{0}, u \rangle, \end{aligned}$$

so that

$$\langle \mathbf{0}, u \rangle = \langle \mathbf{0}, u \rangle + \langle \mathbf{0}, u \rangle;$$

thus

$$\langle \mathbf{0}, u \rangle = 0.$$

(c) Follows from the proof of (b) and part 5 of the definition.

(d) Again, this follows from the definition, since

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

(e) Clearly

$$\begin{aligned} \langle u, \lambda v \rangle &= \overline{\langle \lambda v, u \rangle} \\ &= \overline{\lambda}\ \overline{\langle v, u \rangle} \\ &= \overline{\lambda} \langle u, v \rangle. \end{aligned}$$

Inner Products and Geometry

As promised, an inner product on a vector space gives us a tool for talking about lengths and angles in the space. With the formula for the length of a vector in $\mathbb{R}^n$ in mind, we introduce the idea of a *norm*:

Definitions 6.8. Given $v \in V$, where V is an inner product space over $\mathbb{F}$ equipped with $\langle \cdot, \cdot \rangle$, the *norm* of v, denoted $||v||$, is given by

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

Again, the norm is our analogue for the length of a vector in an arbitrary space.

Example. In $\mathbb{R}^n$ with the standard inner product

$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n,$$

the norm is given by

$$\begin{aligned} ||x|| &= \sqrt{\langle x, x \rangle} \\ &= \sqrt{x_1^2 + x_1^2 + \ldots + x_n^2}; \end{aligned}$$

thus the norm correpsonds to the standard notion of length of a vector in $\mathbb{R}^n$.

Example. The inner product defined on $\mathcal{M}_n(\mathbb{F})$ by

$$\langle A,\ B \rangle = \mathrm{tr}\,(B^* A)$$

yields norm

$$||A|| = \sqrt{\mathrm{tr}\,(A^* A)}.$$

This norm is known as the *Frobenius norm.*

As a specific example, set

$$A = \begin{pmatrix} i & 1 \\ 0 & 1 \end{pmatrix};$$

then $||A||$ is given by

$$\begin{aligned} ||A|| &= \sqrt{\mathrm{tr}\,(A^* A)} \\ &= \sqrt{\mathrm{tr}\left(\begin{pmatrix} -i & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} i & 1 \\ 0 & 1 \end{pmatrix}\right)} \\ &= \sqrt{\mathrm{tr}\begin{pmatrix} 1 & -i \\ i & 2 \end{pmatrix}} \\ &= \sqrt{3}. \end{aligned}$$

Example. The standard inner product in $\mathcal{P}_n(\mathbb{R})$, defined by

$$\langle f, g\rangle = a_0b_0 + a_1b_1 + \ldots + a_nb_n,$$

yields norm

$$||f|| = \sqrt{a_0^2 + a_2^2 + \ldots + a_n^2}.$$

Given $f = f(x) = 1 - x^3 + 2x^5$, we see that

$$||f|| = \sqrt{1+1+4} = \sqrt{6}.$$

Before moving on to the notion of angle, we record a few more helpful facts on norms:

Theorem 6.10. Let V be an inner product space, $v \in V$ and $\lambda \in \mathbb{F}$. Then

(a) $||v|| = 0$ if and only if $v = \mathbf{0}$;

(b) $||\lambda v|| = |\lambda|\ ||v||$, where $|\lambda|$ indicates the modulus $|\lambda| = \sqrt{a^2+b^2}$ for $\lambda = a + bi$, a, $b \in \mathbb{R}$.

Proof. (a) By the definition of the inner product, $\langle v, v\rangle = 0$ if and only if $v = \mathbf{0}$. The result follows immediately.

(b) Clear since

$$\begin{aligned} ||\lambda v|| &= \sqrt{\langle \lambda v, \lambda v\rangle} \\ &= \sqrt{\lambda\overline{\lambda}\langle v, v\rangle} \\ &= |\lambda|\sqrt{\langle v, v\rangle}. \end{aligned}$$

Orthogonality

Recall that, in $\mathbb{R}^n$, the angle θ between a pair x and y of vectors may be calculated using the formula

$$\cos\theta = \frac{x \cdot y}{||x||\ ||y||}.$$

In particular, if x and y are orthogonal (nonzero) vectors, so that $\theta = \pi/2$, we see that

$$\frac{x \cdot y}{||x||\ ||y||} = \cos\frac{\pi}{2} = 0;$$

in particular, this implies that

$$x \cdot y = 0.$$

We generalize this idea to arbitrary inner product spaces with the following definition:

Definition 6.11. Vectors u, $v \in V$ are *orthogonal* if $\langle u, v\rangle = 0$.

With the example of $\mathbb{R}^n$ in mind, you should think of orthogonal vectors as being "perpendicular".

Example. All of the vectors in the standard basis for $\mathbb{F}^n$ are mutually orthogonal under the standard inner product: indeed, given e_i, e_j with $i \neq j$, we have

$$\langle e_i,\ e_j\rangle = 0 \cdot 1 + 1 \cdot 0 = 0.$$

Example. Find a vector in $\mathcal{M}_n(\mathbb{F})$ orthogonal to

$$M_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

under the standard inner product

$$\langle X, Y\rangle = \operatorname{tr}(Y^* X).$$

In order to find vector X_θ orthogonal to M_θ, we need to guarantee that the trace of $X_\theta^* M_\theta$ is 0. The most obvious way to do this is to force each of the diagonal entries of the product to be 0. So a first guess for X_θ is the rotation matrix

$$X_\theta = \begin{pmatrix} \cos(\pi/2+\theta) & -\sin(\pi/2+\theta) \\ \sin(\pi/2+\theta) & \cos(\pi/2+\theta) \end{pmatrix} = \begin{pmatrix} -\sin\theta & -\cos\theta \\ \cos\theta & -\sin\theta \end{pmatrix}.$$

Let's evaluate the inner product–recall that $X^* = X^\top$ since X has all real entries:

$$\begin{aligned}
\langle M_\theta, X_\theta\rangle &= \operatorname{tr}(X_\theta^* M_\theta) \\
&= \operatorname{tr}\left(\begin{pmatrix} -\sin\theta & \cos\theta \\ -\cos\theta & -\sin\theta \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\right) \\
&= \operatorname{tr}\left(\begin{pmatrix} -\cos\theta\sin\theta + \cos\theta\sin\theta & \sin^2\theta + \cos^2\theta \\ -\sin^2\theta - \cos^2\theta & -\cos\theta\sin\theta + \cos\theta\sin\theta \end{pmatrix}\right) \\
&= \operatorname{tr}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \\
&= 0.
\end{aligned}$$

Thus for any $\theta \in \mathbb{R}$, matrices

$$M_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \text{ and } X_\theta = \begin{pmatrix} -\sin\theta & -\cos\theta \\ \cos\theta & -\sin\theta \end{pmatrix}$$

are orthogonal.

One of the most important inequalities in all of mathematics, the Cauchy-Schwarz inequality, relates the inner product of a pair of vectors with their associated norms. The inequality is a bit hard to interpret initially, but it immediately implies that a version of the *triangle inequality* holds in vector spaces. We will return to that interpretation momentarily.

Cauchy-Schwarz Inequality. Let V be an inner product space, and let u, $v \in V$. Then

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

Equality holds if and only if u and v are scalar multiples.

Example. Given the inner product on $\mathcal{M}_n(\mathbb{F})$ defined by by

$$\langle A,\ B\rangle = \mathrm{tr}\,(B^*A),$$

let us consider the Cauchy-Schwarz inequality for

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \text{ and } B = \begin{pmatrix} -\sqrt{3}/2 & 1/2 \\ -1/2 & -\sqrt{3}/2 \end{pmatrix}.$$

We calculate the norms of A and B as follows:

$$\begin{aligned} ||A|| &= \sqrt{\mathrm{tr}\,(A^*A)} \\ &= \sqrt{\mathrm{tr}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \\ &= \sqrt{2}, \\ ||B|| &= \sqrt{\mathrm{tr}\,(B^*B)} \\ &= \sqrt{\mathrm{tr}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \\ &= \sqrt{2}. \end{aligned}$$

Next, we find

$$\begin{aligned} \langle A, B\rangle &= \mathrm{tr}\,(B^*A) \\ &= \mathrm{tr}\begin{pmatrix} -1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix} \\ &= -1. \end{aligned}$$

Thus

$$|\langle A, B\rangle| = 1 < \sqrt{2}\sqrt{2} = ||A||\ ||B||,$$

as implied by the theorem.

Again, the Cauchy-Schwarz inequality seems a bit abstract on the surface, but it has an immediate geometric consequence–the Triangle Inequality:

Triangle Inequality. For any u, v in an inner product space V,

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

Proof. The proof is a straightforward application of the definition of norm and the Cauchy-Schwarz inequality:

$$\begin{aligned}
||u+v||^2 &= \langle u+v, u+v\rangle \\
&= \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle \\
&= \langle u,u\rangle + \langle u,v\rangle + \overline{\langle u,v\rangle} + \langle v,v\rangle \\
&= ||u||^2 + 2\text{Re}\langle u,v\rangle + ||v||^2 \\
&\leq ||u||^2 + 2|\langle u,v\rangle| + ||v||^2 \\
&\leq ||u||^2 + 2||u||\ ||v|| + ||v||^2 \\
&= (||u|| + ||v||)^2.
\end{aligned}$$

Remark. The Triangle inequality is really a generalization of the idea that the shortest path between a pair of points is a straight line connecting them. Indeed, the graph in $\mathbb{R}^2$ is as follows: