Inner Product Spaces

Definition. Let V be a vector space over F, where $F = \mathbb{R}$ or $\mathbb{C}$. An **inner product** on V is a function $\langle \cdot, \cdot \rangle : V \times V \to F$ which satisfies:

(a) (Linearity) $\langle ax + y, z \rangle = a \langle x, z \rangle + \langle y, z \rangle$, for $x, y, z \in V$, $a \in F$.

(b) (Symmetry) $\langle x, y \rangle = \overline{\langle y, x \rangle}$, for $x, y \in V$. ("$\overline{x}$" denotes the complex conjugate of x.)

(c) (Positive-definiteness) If $x \neq 0$, then $\langle x, x \rangle \in \mathbb{R}$ and $\langle x, x \rangle > 0$.

A vector space with an inner product is an **inner product space**. If $F = \mathbb{R}$, V is a **real inner product space**; if $F = \mathbb{C}$, V is a **complex inner product space**.

Notation. There are various notations for inner products. You may see "(x, y)" or "$(x \mid y)$" or "$\langle x \mid y \rangle$", for instance. Some specific inner products come with established notation. For example, the **dot product**, which I'll discuss below, is denoted "$x \cdot y$".

Proposition. Let V be an inner product space over F, where $F = \mathbb{R}$ or $\mathbb{C}$. Let $x, y, z \in V$, and let $a \in F$.

(a) $\left\langle x, \vec{0} \right\rangle = 0$ and $\left\langle \vec{0}, x \right\rangle = 0$.

(b) $\langle x, ay + z \rangle = \overline{a} \langle x, y \rangle + \langle x, z \rangle$.
In particular, if $a \in \mathbb{R}$, then $\langle x, ay + z \rangle = a \langle x, y \rangle + \langle x, z \rangle$.

(c) If $F = \mathbb{R}$, then $\langle x, y \rangle = \langle y, x \rangle$.

(a)
$$\left\langle \vec{0}, x \right\rangle = \left\langle \vec{0} + \vec{0}, x \right\rangle = \left\langle \vec{0}, x \right\rangle + \left\langle \vec{0}, x \right\rangle, \quad \text{so} \quad 0 = \left\langle \vec{0}, x \right\rangle.$$

By symmetry, $\left\langle x, \vec{0} \right\rangle = 0$ as well.

(b)
$$\langle x, ay + z \rangle = \overline{\langle ay + z, x \rangle} = \overline{a \langle y, x \rangle + \langle z, x \rangle} = \overline{a} \overline{\langle y, x \rangle} + \overline{\langle z, x \rangle} = \overline{a} \langle x, y \rangle + \langle x, z \rangle.$$

If $a \in \mathbb{R}$, then $\overline{a} = a$, and so $\langle x, ay + z \rangle = a \langle x, y \rangle + \langle x, z \rangle$.

(c) As in the proof of (b), I have $\overline{\langle y, x \rangle} = \langle y, x \rangle$, so

$$\langle x, y \rangle = \overline{\langle y, x \rangle} = \langle y, x \rangle. \quad \square$$

Remarks. Why include complex conjugation in the symmetry axiom? Suppose the symmetry axiom had read

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

Then

$$0 < \langle ix, ix \rangle = i \langle x, ix \rangle = i \langle ix, x \rangle = i \cdot i \langle x, x \rangle = - \langle x, x \rangle.$$

This contradicts $\langle x, x \rangle > 0$. That is, I can't have *both* pure symmetry and positive definiteness. $\square$

Example. Suppose u, v, and w are vectors in a real inner product space V. Suppose

$$\langle u, v \rangle = 5, \quad \langle v, w \rangle = -2, \quad \langle u, w \rangle = 6,$$

$$\langle v, v \rangle = 4, \quad \langle w, w \rangle = 3.$$

(a) Compute $\langle u + 3w, v - 5w \rangle$.

(b) Compute $\langle v + w, v - w \rangle$.

(a) Using the linearity and symmetry properties, I have

$$\langle u + 3w, v - 5w \rangle = \langle u, v - 5w \rangle + \langle 3w, v - 5w \rangle = \langle u, v \rangle - \langle u, 5w \rangle + \langle 3w, v \rangle - \langle 3w, 5w \rangle =$$

$$\langle u, v \rangle - 5 \langle u, w \rangle + \langle w, v \rangle - 15 \langle w, w \rangle = 5 - 5 \cdot 6 + (-2) - 15 \cdot 3 = -72.$$

Notice that this "looks like" the polynomial multiplication you learned in basic algebra:

$$(x + 3z)(y - 5z) = xy - 5xz + 3zy - 15z^2.$$

(b)
$$\langle v + w, v - w \rangle = \langle v, v \rangle - \langle v, w \rangle + \langle w, v \rangle - \langle w, w \rangle = \langle v, v \rangle - \langle w, w \rangle = 4 - 3 = 1. \quad \square$$

Example. Let $(a_1, \ldots, a_n), (b_1, \ldots, b_n) \in \mathbb{R}^n$. The **dot product** on $\mathbb{R}^n$ is given by

$$(a_1, \ldots, a_n) \cdot (b_1, \ldots, b_n) = a_1 b_1 + \cdots + a_n b_n.$$

It's easy to verify that the axioms for an inner product hold. For example, suppose $(a_1, \ldots, a_n) \neq \vec{0}$. Then at least one of $a_1, \ldots, a_n$ is nonzero, so

$$(a_1, \ldots, a_n) \cdot (a_1, \ldots, a_n) = a_1^2 + \cdot + a_n^2 > 0.$$

This proves that the dot product is positive-definite. $\quad \square$

I can use an inner product to define **lengths** and **angles**. Thus, an inner product introduces (metric) geometry into vector spaces.

Definition. Let V be an inner product space, and let $x, y \in V$.

(a) The **length** of x is $\|x\| = \langle x, x \rangle^{1/2}$.

(b) The **distance** between x and y is $\|x - y\|$.

(c) The **angle** between x and y is the smallest positive real number θ satisfying

$$\cos \theta = \frac{\langle x, y \rangle}{\|x\| \|y\|}.$$

Remark. The definition of the **angle** between x and y wouldn't make sense if the expression $\dfrac{\langle x, y \rangle}{\|x\| \|y\|}$ was greater than 1 or less than -1, since I'm asserting that it's the cosine of an angle.

In fact, the **Cauchy-Schwarz inequality** (which I'll prove below) will show that

$$-1 \leq \frac{\langle x, y \rangle}{\|x\| \|y\|} \leq 1. \quad \square$$

Proposition. Let V be a real inner product space, $a \in \mathbb{R}$, $x, y \in V$.

(a) $\|x\|^2 = \langle x, x \rangle$.

(b) $\|ax\| = |a|\|x\|$. ("$|a|$" denotes the absolute value of a.)

(c) $x \neq 0$ if and only if $\|x\| > 0$.

(d) (**Cauchy- Schwarz inequality**) $|\langle x, y \rangle| \leq \|x\|\|y\|$.

(e) (**Triangle inequality**) $\|x + y\| \leq \|x\| + \|y\|$.

Proof. (a) Squaring $\|x\| = \langle x, x \rangle^{1/2}$ gives $\|x\|^2 = \langle x, x \rangle$.

(b) Since $\langle ax, ax \rangle = a^2 \langle x, x \rangle$,

$$\|ax\| = \sqrt{a^2}\|x\| = |a|\|x\|.$$

(c) $x \neq 0$ implies $\langle x, x \rangle > 0$, and hence $\|x\| > 0$. Conversely, if $x = 0$, then $\langle 0, 0 \rangle = 0$, so $\|x\| = 0$.

(d) If $x = \vec{0}$, then

$$|\langle x, y \rangle| = \left|\left\langle \vec{0}, y \right\rangle\right| = |0| = 0 \quad \text{and} \quad \|\vec{0}\|\|y\| = 0 \cdot \|y\| = 0.$$

Hence, $|\langle x, y \rangle| \leq \|x\|\|y\|$. The same is true if $y = \vec{0}$.

Thus, I may assume that $x \neq \vec{0}$ and $y \neq \vec{0}$.

The major part of the proof comes next, and it involves a trick. Don't feel bad if you wouldn't have thought of this yourself: Try to follow along and understand the steps.

If $a, b \in \mathbb{R}$, then by positive-definiteness and linearity,

$$0 \leq \langle ax - by, ax - by \rangle = a^2 \langle x, x \rangle - 2ab \langle x, y \rangle + b^2 \langle y, y \rangle = a^2\|x\|^2 - 2ab \langle x, y \rangle + b^2\|y\|^2.$$

The trick is to pick "nice" values for a and b. I will set $a = \|y\|$ and $b = \|x\|$. (A rationale for this is that I want the expression $\|x\|\|y\|$ to appear in the inequality.)

I get

$$\|y\|^2\|x\|^2 - 2\|y\|\|x\| \langle x, y \rangle + \|x\|^2\|y\|^2 \geq 0$$
$$2\|x\|^2\|y\|^2 - 2\|x\|\|y\| \langle x, y \rangle \geq 0$$
$$2\|x\|^2\|y\|^2 \geq 2\|x\|\|y\| \langle x, y \rangle$$
$$\|x\|^2\|y\|^2 \geq \|x\|\|y\| \langle x, y \rangle$$

Since $x \neq \vec{0}$ and $y \neq \vec{0}$, I have $\|x\| \neq 0$ and $\|y\| \neq 0$. So I can divide the inequality by $\|x\|\|y\|$ to obtain

$$\|x\|\|y\| \geq \langle x, y \rangle.$$

In the last inequality, x and y are arbitrary vectors. So the inequality is still true if x is replaced by $-x$. If I replace x with $-x$, then $\|-x\| = \|x\|$ and $\langle -x, y \rangle = - \langle x, y \rangle$, and the inequality becomes

$$\|x\|\|y\| \geq - \langle x, y \rangle.$$

Since $\|x\|\|y\|$ is greater than or equal to both $\langle x, y \rangle$ and $- \langle x, y \rangle$, I have

$$\|x\|\|y\| \geq |\langle x, y \rangle|.$$

(e)

$$\|x + y\|^2 = \langle x + y, x + y \rangle = \|x\|^2 + 2 \langle x, y \rangle + \|y\|^2 \leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2.$$

Hence, $\|x + y\| \leq \|x\| + \|y\|$. $\square$

Example. $\mathbb{R}^3$ is an inner product space using the standard dot product of vectors. The cosine of the angle between $(2, -2, 1)$ and $(6, -8, 24)$ is

$$\cos\theta = \frac{(2, -2, 1) \cdot (6, -8, 24)}{\|(2, -2, 1)\|\|(6, -8, 24)\|} = \frac{12 + 16 + 24}{(3)(26)} = \frac{2}{3}. \quad \square$$

Example. Let $C[0, 1]$ denote the real vector space of continuous functions on the interval $[0, 1]$. Define an inner product on $C[0, 1]$ by

$$\langle f, g \rangle = \int_0^1 f(x)g(x)\, dx.$$

Note that $f(x)g(x)$ is integrable, since it's continuous on a closed interval.

The verification that this gives an inner product relies on standard properties of Riemann integrals. For example, if $f \neq 0$,

$$\langle f, f \rangle = \int_0^1 f(x)^2\, dx > 0.$$

Given that this is a real inner product, I may apply the preceding proposition to produce some useful results. For example, the Cauchy-Schwarz inequality says that

$$\left(\int_0^1 f(x)^2\, dx \right)^{1/2} \left(\int_0^1 g(x)^2\, dx \right)^{1/2} \geq \left| \int_0^1 f(x)g(x)\, dx \right|. \quad \square$$

Definition. A set of vectors S in an inner product space V is **orthogonal** if $\langle v_i, v_j \rangle = 0$ for $v_i, v_j \in S$, $v_i \neq v_j$.

An orthogonal set S is **orthonormal** if $\|v_i\| = 1$ for all $v_i \in S$.

If you've seen dot products in a multivariable calculus course, you know that vectors in $\mathbb{R}^n$ whose dot product is 0 are perpendicular. With this interpretation, the vectors in an orthogonal set are mutually perpendicular. The vectors in an orthonormal set are mutually perpendicular *unit* vectors.

Notation. If I is an index set, the **Kronecker delta** δ_{ij} (or $\delta(i, j)$) is defined by

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

With this notation, a set $S = \{v_i\}$ is orthonormal if

$$\langle v_i, v_j \rangle = \delta_{ij}.$$

Note that the $n \times n$ matrix whose (i, j)-th component is δ_{ij} is the $n \times n$ identity matrix.

Example. The standard basis for $\mathbb{R}^n$ is
$$e_1 = (1, 0, 0, \ldots, 0)$$
$$e_2 = (0, 1, 0, \ldots, 0)$$
$$\vdots$$
$$e_n = (0, 0, 0, \ldots, 1)$$

It's clear that relative to the dot product on $\mathbb{R}^n$, each of these vectors as length 1, and each pair of the vectors has dot product 0. Hence, the standard basis is an orthonormal set relative to the dot product on $\mathbb{R}^n$. $\square$

4

Example. Consider the following set of vectors in $\mathbb{R}^2$:

$$\left\{ \left(\frac{3}{5}, \frac{4}{5} \right), \left(-\frac{4}{5}, \frac{3}{5} \right) \right\}.$$

I have

$$\left(\frac{3}{5}, \frac{4}{5} \right) \cdot \left(\frac{3}{5}, \frac{4}{5} \right) = \frac{9}{25} + \frac{16}{25} = 1,$$

$$\left(\frac{3}{5}, \frac{4}{5} \right) \cdot \left(-\frac{4}{5}, \frac{3}{5} \right) = -\frac{12}{25} + \frac{12}{25} = 0,$$

$$\left(-\frac{4}{5}, \frac{3}{5} \right) \cdot \left(-\frac{4}{5}, \frac{3}{5} \right) = \frac{16}{25} + \frac{9}{25} = 1.$$

It follows that the set is orthonormal relative to the dot product on $\mathbb{R}^2$. $\square$

Example. Let $C[0, 2\pi]$ denote the *complex-valued* continuous functions on $[0, 2\pi]$. Define an inner product by

$$\langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(x)\overline{g(x)} \, dx.$$

Let $m, n \in \mathbb{Z}$. Then

$$\frac{1}{2\pi} \int_0^{2\pi} e^{imx} e^{-inx} \, dx = \delta_{mn}.$$

It follows that the following set is orthonormal in $C[0, 2\pi]$ relative to this inner product:

$$\left\{ \frac{1}{\sqrt{2\pi}} e^{imx} \mid m = \dots, -1, 0, 1, \dots \right\} \quad \square$$

Proposition. Let $\{v_i\}$ be an orthogonal set of vectors, $v_i \neq 0$ for all i. Then $\{v_i\}$ is independent.

Proof. Suppose

$$a_1 v_{i_1} + a_2 v_{i_2} + \cdots + a_n v_{i_n} = \vec{0}.$$

Take the inner product of both sides with v_{i_1}:

$$a_1 \langle v_{i_1}, v_{i_1} \rangle + a_2 \langle v_{i_1}, v_{i_2} \rangle + \cdots + a_n \langle v_{i_1}, v_{i_n} \rangle = 0.$$

Since $\{v_i\}$ is orthogonal,

$$\langle v_{i_1}, v_{i_2} \rangle = \cdots = \langle v_{i_1}, v_{i_n} \rangle = 0.$$

The equation becomes

$$a_1 \langle v_{i_1}, v_{i_1} \rangle = 0.$$

But $\langle v_{i_1}, v_{i_1} \rangle > 0$ by positive-definiteness, since $v_{i_1} \neq \vec{0}$. Therefore, $a_1 = 0$.

Similarly, taking the inner product of both sides of the original equation with $v_{i_2}, \dots v_{i_n}$ shows $a_j = 0$ for all j. Therefore, $\{v_i\}$ is independent. $\square$

An orthonormal set consists of vectors of length 1, so the vectors are obviously nonzero. Hence, an orthonormal set is independent, and forms a basis for the subspace it spans. A basis which is an orthonormal set is called an **orthonormal basis**.

It is very easy to find the components of a vector relative to an orthonormal basis.

Proposition. Let $\{v_i\}$ be an orthonormal basis for V, and let $v \in V$. Then

$$v = \sum_i \langle v, v_i \rangle \, v_i.$$

Note: In fact, the sum above is a finite sum — that is, only finitely many terms are nonzero.

Proof. Since $\{v_i\}$ is a basis, there are elements $a_j \in F$ and $v_{i_v} \in \{v_i\}$ such that

$$v = a_1 v_{i_1} + a_2 v_{i_2} + \cdots + a_n v_{i_n}.$$

Take the inner product of both sides with v_{i_1}. Then

$$\langle v, v_{i_1} \rangle = a_1 \langle v_{i_1}, v_{i_1} \rangle + a_2 \langle v_{i_2}, v_{i_1} \rangle + \cdots + a_n \langle v_{i_n}, v_{i_1} \rangle .$$

As in the proof of the last proposition, all the inner product terms on the right vanish, except that $\langle v_{i_1}, v_{i_1} \rangle = 1$ by orthonormality. Thus,

$$\langle v, v_{i_1} \rangle = a_1.$$

Taking the inner product of both sides of the original equation with $v_{i_2}, \ldots v_{i_n}$ shows

$$\langle v, v_{i_2} \rangle = a_2, \ldots \langle v, v_{i_n} \rangle = a_n. \quad \Box$$

Example. Here is an orthonormal basis for $\mathbb{R}^2$:

$$\left\{ \begin{bmatrix} \dfrac{3}{5} \\ \dfrac{4}{5} \end{bmatrix}, \begin{bmatrix} -\dfrac{4}{5} \\ \dfrac{3}{5} \end{bmatrix} \right\}$$

To express $(-7, 6)$ in terms of this basis, take the dot product of the vector with each element of the basis:

$$(-7, 6) \cdot \left(\frac{3}{5}, \frac{4}{5} \right) = \frac{3}{5},$$

$$(-7, 6) \cdot \left(-\frac{4}{5}, \frac{3}{5} \right) = \frac{46}{5}.$$

Hence,

$$\begin{bmatrix} -7 \\ 6 \end{bmatrix} = \frac{3}{5} \cdot \begin{bmatrix} \dfrac{3}{5} \\ \dfrac{4}{5} \end{bmatrix} + \frac{46}{5} \cdot \begin{bmatrix} -\dfrac{4}{5} \\ \dfrac{3}{5} \end{bmatrix}. \quad \Box$$

Example. Let $C[0, 2\pi]$ denote the complex inner product space of complex-valued continuous functions on $[0, 2\pi]$, where the inner product is defined by

$$\langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(x) \overline{g(x)} \, dx.$$

I noted earlier that the following set is orthonormal:

$$S = \left\{ \frac{1}{\sqrt{2\pi}} e^{mix} \mid m = \ldots, -1, 0, 1, \ldots \right\}.$$

Suppose I try to compute the "components" of $f(x) = x$ relative to this orthonormal set by taking inner products — that is, using the approach of the preceding example.

For $m = 0$,

$$\frac{1}{\sqrt{2\pi}} \int_0^{2\pi} x \, dx = \pi\sqrt{2\pi}.$$

Suppose $m \neq 0$. Then

$$\frac{1}{\sqrt{2\pi}} \int_0^{2\pi} x e^{-mix} \, dx = \frac{1}{\sqrt{2\pi}} \left[\frac{i}{m} x e^{-mix} - \frac{1}{m^2} e^{-mix} \right]_0^{2\pi} = \frac{i\sqrt{2\pi}}{m}.$$

There are infinitely many nonzero components! Of course, the reason this does not contradict the earlier result is that $f(x) = x$ may not lie in the span of S. S is orthonormal, hence independent, but it is not a basis for $C[0, 2\pi]$.

In fact, since $e^{mix} = \cos mx + i \sin mx$, a finite linear combination of elements of S must be periodic.

It is still reasonable to ask whether (or in what sense) $f(x) = x$ can be represented by the the infinite sum

$$\pi\sqrt{2\pi} + \sum_{m=1}^{\infty} \left(\frac{i\sqrt{2\pi}}{m} e^{mix} - \frac{i\sqrt{2\pi}}{m} e^{-mix} \right).$$

For example, it is reasonable to ask whether the series converges uniformly to f at each point of $[0, 2\pi]$. The answers to these kinds of questions would require an excursion into the theory of **Fourier series**. □

Since it's so easy to find the components of a vector relative to an orthonormal basis, it's of interest to have an algorithm which converts a given basis to an orthonormal one.

The **Gram-Schmidt algorithm** converts a basis to an orthonormal basis by "straightening out" the vectors one by one.
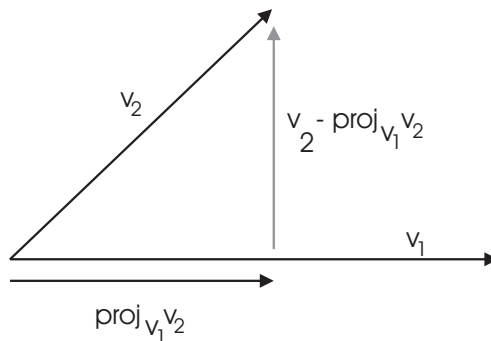
The picture shows the first step in the straightening process. Given vectors v_1 and v_2, I want to replace v_2 with a vector perpendicular to v_1. I can do this by taking the component of v_2 perpendicular to v_1, which is

$$v_2 - \frac{\langle v_1, v_2 \rangle}{\langle v_1, v_1 \rangle} v_1.$$

Lemma. (**Gram-Schmidt algorithm**) Let $\{v_1, \ldots, v_k\}$ be a set of nonzero vectors in an inner product space V. Suppose v_1, $\ldots$, v_{k-1} are pairwise orthogonal. Let

$$v_k' = v_k - \sum_{i=1}^{k-1} \frac{\langle v_i, v_k \rangle}{\langle v_i, v_i \rangle} v_i.$$

Then v_k' is orthogonal to v_1, $\ldots$, v_{k-1}.

7

Proof. Let $j \in \{1, \dots, k-1\}$. Then

$$\langle v_j, v'_k \rangle = \langle v_j, v_k \rangle - \sum_{i=1}^{k-1} \frac{\langle v_i, v_k \rangle}{\langle v_i, v_i \rangle} \langle v_j, v_i \rangle .$$

Now $\langle v_j, v_i \rangle = 0$ for $i \neq j$ because the set is orthogonal. Hence, the right side collapses to

$$\langle v_j, v_k \rangle - \frac{\langle v_j, v_k \rangle}{\langle v_j, v_j \rangle} \langle v_j, v_j \rangle = \langle v_j, v_k \rangle - \langle v_j, v_k \rangle = 0. \quad \square$$

Suppose that I start with an independent set $\{v_1, \dots, v_n\}$. Apply the Gram-Schmidt procedure to the set, beginning with $v'_1 = v_1$. This produces an orthogonal set $\{v'_1, \dots, v'_n\}$. In fact, $\{v'_1, \dots, v'_n\}$ is a *nonzero* orthogonal set, so it is independent as well.

To see that each v'_k is nonzero, suppose

$$0 = v'_k = v_k - \sum_{i=1}^{k-1} \frac{\langle v_i, v_k \rangle}{\langle v_i, v_i \rangle} v_i.$$

Then

$$v_k = \sum_{i=1}^{k-1} \frac{\langle v_i, v_k \rangle}{\langle v_i, v_i \rangle} v_i.$$

This contradicts the independence of $\{v_i\}$, because v_k is expressed as the linear combination of v_1, $\dots v_{k-1}$.

In general, if the algorithm is applied iteratively to a set of vectors, the span is preserved at each state. That is,

$$\langle v_1, \dots, v_k \rangle = \langle v'_1, \dots, v'_k \rangle.$$

This is true at the start, since $v_1 = v'_1$. Assume inductively that

$$\langle v_1, \dots, v_{k-1} \rangle = \langle v'_1, \dots, v'_{k-1} \rangle.$$

Consider the equation

$$v'_k = v_k - \sum_{i=1}^{k-1} \frac{\langle v_i, v_k \rangle}{\langle v_i, v_i \rangle} v_i.$$

It expresses v'_k as a linear combination of $\{v_1, \dots, v_k\}$. Hence, $\langle v'_1, \dots, v'_k \rangle \subset \langle v_1, \dots, v_k \rangle$.
Conversely,

$$v_k = v'_k + \sum_{i=1}^{k-1} \frac{\langle v_i, v_k \rangle}{\langle v_i, v_i \rangle} v_i \subset \langle v'_k \rangle + \langle v_1, \dots, v_{k-1} \rangle =$$

$$\langle v'_k \rangle + \langle v'_1, \dots, v'_{k-1} \rangle = \langle v'_1, \dots, v'_k \rangle.$$

It follows that $\langle v_1, \dots, v_k \rangle \subset \langle v'_1, \dots, v'_k \rangle$, so $\langle v_1, \dots, v_k \rangle = \langle v'_1, \dots, v'_k \rangle$, by induction.

To summarize: If you apply Gram-Schmidt to a set of vectors, the algorithm produces a new set of vectors with the same span as the old set. If the original set was independent, the new set is independent (and orthogonal) as well.

So, for example, if Gram-Schmidt is applied to a *basis* for an inner product space, it will produce an *orthogonal basis* for the space.

Finally, you can always produce *orthonormal* set from a orthogonal set (of nonzero vectors) — merely divide each vector in the orthogonal set by its length. $\quad \square$

Example. (**Gram-Schmidt**) Apply Gram-Schmidt to the following set of vectors in $\mathbb{R}^3$ (relative to the usual dot product):

$$\left\{ \begin{bmatrix} 3 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 7 \end{bmatrix}, \begin{bmatrix} 2 \\ 9 \\ 11 \end{bmatrix} \right\}.$$

$$v_1' = v_1 = (3, 0, 4),$$

$$v_2' = (-1, 0, 7) - \frac{(-1, 0, 7) \cdot (3, 0, 4)}{(3, 0, 4) \cdot (3, 0, 4)} (3, 0, 4) = (-4, 0, 3).$$

$$v_3' = (2, 9, 11) - \frac{(2, 9, 11) \cdot (3, 0, 4)}{(3, 0, 4) \cdot (3, 0, 4)} (3, 0, 4) - \frac{(2, 9, 11) \cdot (-4, 0, 3)}{(-4, 0, 3) \cdot (-4, 0, 3)} (-4, 0, 3) = (0, 9, 0).$$

(A common mistake here is to project onto v_1, v_2, ... I need to project onto the vectors that have *already been orthogonalized*. That is why I projected onto $(3, 0, 4)$ and $(-4, 0, 3)$ rather than $(3, 0, 4)$ and $(-1, 0, 7)$.)

The algorithm has produced the following orthogonal set:

$$\left\{ \begin{bmatrix} 3 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} -4 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 9 \\ 0 \end{bmatrix} \right\}.$$

The lengths of these vectors are 5, 5, and 9. For example

$$\|(3, 0, 4)\| = \sqrt{3^2 + 0^2 + 4^2} = 5.$$

The correponding orthonormal set is

$$\left\{ \frac{1}{5} \begin{bmatrix} 3 \\ 0 \\ 4 \end{bmatrix}, \frac{1}{5} \begin{bmatrix} -4 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}. \quad \square$$

Example. (**Gram-Schmidt**) Find an orthonormal basis (relative to the usual dot product) for the subspace of $\mathbb{R}^4$ spanned by the vectors

$$v_1 = (1, 0, 2, 2), \quad v_2 = (10, 1, 0, 4), \quad v_3 = (1, 1, 0, 13).$$

I'll use v_1', v_2', v_3' to denote the orthonormal basis.

To simplify the computations, you should fix the vectors so they're mutually perpendicular *first*. Then you can divide each by its length to get vectors of length 1.

First,

$$v_1' = v_1 = \langle 1, 0, 2, 2 \rangle.$$

Next,

$$v_2' = v_2 - \frac{v_2 \cdot v_1'}{v_1' \cdot v_1'} v_1' = (10, 1, 0, 4) - \frac{(10, 1, 0, 4) \cdot (1, 0, 2, 2)}{(1, 0, 2, 2) \cdot (1, 0, 2, 2)} (1, 0, 2, 2) =$$

$$(10, 1, 0, 4) - 2 \cdot (1, 0, 2, 2) = (10, 1, 0, 4) - (2, 0, 4, 4) = (8, 1, -4, 0).$$

You can check that $v_2' \cdot v_1' = 0$, so the first two are perpendicular.

Finally,

$$v_3' = v_3 - \frac{v_3 \cdot v_1'}{v_1' \cdot v_1'} v_1' - \frac{v_3 \cdot v_2'}{v_2' \cdot v_2'} v_2' =$$

$$(1,1,0,13) - \frac{(1,1,0,13) \cdot (1,0,2,2)}{(1,0,2,2) \cdot (1,0,2,2)}(1,0,2,2) - \frac{(1,1,0,13) \cdot (8,1,-4,0)}{(8,1,-4,0) \cdot (8,1,-4,0)}(8,1,-4,0) =$$

$$(1,1,0,13) - \frac{27}{9} \cdot (1,0,2,2) - \frac{9}{81} \cdot (8,1,-4,0) = (1,1,0,13) - (3,0,6,6) - \left(\frac{8}{9}, \frac{1}{9}, -\frac{4}{9}, 0\right) =$$

$$\left(-\frac{26}{9}, \frac{8}{9}, -\frac{50}{9}, 7\right).$$

If at any point you wind up with a vector with fractions, it's a good idea to clear the fractions before continuing. Since multiplying a vector by a number doesn't change its direction, it remains perpendicular to the vectors already constructed.

Thus, I'll multiply the last vector by 9 and use

$$v_3' = (-26, 8, -50, 63).$$

Thus, the *orthogonal* set is

$$v_1' = (1,0,2,2), \quad v_2' = (8,1,-4,0), \quad v_3' = (-26,8,-50,63).$$

The lengths of these vectors are 3, 9, and $\sqrt{7209} = 9\sqrt{89}$. Dividing the vectors by their lengths gives and orthonormal basis:

$$\left(\frac{1}{3}, 0, \frac{2}{3}, \frac{2}{3}\right), \left(\frac{8}{9}, \frac{1}{9}, -\frac{4}{9}, 0\right), \left(-\frac{26}{9\sqrt{89}}, \frac{8}{9\sqrt{89}}, -\frac{50}{9\sqrt{89}}, \frac{7}{\sqrt{89}}\right). \quad \Box$$

Recall that when an n-dimensional vector is interpreted as a matrix, it is taken to be an $n \times 1$ matrix: that is, an n-dimensional column vector

$$(v_1, v_2, \ldots v_n) = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

If I need an n-dimensional *row* vector, I'll take the transpose. Thus,

$$(v_1, v_2, \ldots v_n)^T = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}.$$

Lemma. Let A be an invertible $n \times n$ matrix with entries in $\mathbb{R}$. Let

$$\langle x, y \rangle = x^T A^T A y.$$

Then $\langle , \rangle$ defines an inner product on $\mathbb{R}^n$.

Proof. I have to check linearity, symmetry, and positive definiteness.

First, if $a \in \mathbb{R}$, then

$$\langle ax_1 + x_2, y \rangle = (ax_1 + x_2)A^T A y = a(x_1 A^T A y) + x_2 A^T A y = a \langle x_1, y \rangle + \langle x_2, y \rangle.$$

This proves that the function is linear in the first slot.

Next,

$$\langle x, y \rangle = x^T A^T A y = (y^T A^T A x)^T = y^T A^T A x = \langle y, x \rangle.$$

The second equality comes from the fact that $(BC)^T = C^T B^T$ for matrices. The third inequality comes from the fact that $y^T A^T A x$ is a 1×1 matrix, so it equals its transpose.

This proves that the function is symmetric.

Finally,
$$\langle x, x \rangle = x^T A^T A x = (Ax)^T (Ax).$$

Now Ax is an $n \times 1$ vector — I'll label its components this way:
$$Ax = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}.$$

Then
$$\langle x, x \rangle = (Ax)^T (Ax) = [\,u_1 \quad u_2 \quad \cdots \quad u_n\,] \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = u_1^2 + u_2^2 + \cdots + u_n^2 \geq 0.$$

That is, the inner product of a vector with itself is a nonnegative number. All that remains is to show that if the inner product of a vector with itself is 0, them the vector is $\vec{0}$.

Using the notation above, suppose
$$0 = (x, x) = u_1^2 + u_2^2 + \cdots + u_n^2.$$

Then $u_1 = u_2 = \cdots = u_n = 0$, because a nonzero u would produce a positive number on the right side of the equation.

So
$$Ax = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \vec{0}.$$

Finally, I'll use the fact that A is invertible:
$$A^{-1} A x = A^{-1}\vec{0}, \quad x = \vec{0}.$$

This proves that the function is positive definite, so it's an inner product. $\quad\square$

Example. The previous lemma provides lots of examples of inner products on $\mathbb{R}^n$ besides the usual dot product. All I have to do is take an invertible matrix A and form $A^T A$, defining the inner product as above.

For example, this 2×2 real matrix is invertible:
$$A = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix}.$$

Now
$$A^T A = \begin{bmatrix} 29 & 12 \\ 12 & 5 \end{bmatrix}.$$

(Notice that $A^T A$ will always be symmetric.) The inner product defined by this matrix is
$$\langle (x_1, x_2), y_1, y_2) \rangle = [\,x_1 \quad x_2\,] \begin{bmatrix} 29 & 12 \\ 12 & 5 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 29x_1 y_1 + 12 x_2 y_1 + 12 x_1 y_2 + 5 x_2 y_2.$$

For example, under this inner product,
$$\langle (1, 2), (-8, 3) \rangle = -358, \quad \text{and} \quad \|(5, -2))\| = \sqrt{505}. \quad\square$$

11

Definition. A matrix A in $M(n, \mathbb{R})$ is **orthogonal** if $AA^T = I$.

Proposition. Let A be an orthogonal matrix.

(a) $\det(A) = \pm 1$.

(b) $AA^T = I = A^T A$ — in other words, $A^T = A^{-1}$.

(c) The rows of A form an orthonormal set. The columns of A form an orthonormal set.

(d) A preserves dot products — and hence, lengths and angles — in the sense that

$$(Ax) \cdot (Ay) = x \cdot y.$$

Proof. (a) If A is orthogonal,

$$\det(AA^T) = \det(I) = 1$$
$$\det(A)det(A^T) = 1$$
$$(\det(A))^2 = 1$$

Therefore, $\det(A) = \pm 1$.

(b) Since $\det A = \pm 1$, the determinant is certainly nonzero, so A is invertible. Hence,

$$AA^T = I$$
$$A^{-1}AA^T = A^{-1}I$$
$$IA^T = A^{-1}I$$
$$A^T = A^{-1}$$

But $A^{-1}A = I$, so $A^T A = I$ as well.

(c) The equation $AA^T = I$ implies that the rows of A form an orthonormal set of vector. Likewise, $A^T A = I$ shows that the same is true for the columns of A.

(d) The ordinary dot product of vectors $x = (x_1, x_2, \ldots x_n)$ and $y = (y_1, y_2, \ldots y_n)$ can be written as a matrix multiplication:

$$x \cdot y = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = x^T y.$$

(Remember the convention that vectors are *column* vectors.)
Suppose A is orthogonal. Then

$$(Ax) \cdot (Ay) = (Ax)^T (Ay) = x^T A^T Ay = x^T Iy = x^T y = x \cdot y.$$

In other words, orthogonal matrices preserve dot products. It follows that orthogonal matrices will also preserve *lengths of vectors* and *angles between vectors*, because these are defined in terms of dot products.
□

Example. Find real numbers a and b such that the following matrix is orthogonal:

$$A = \begin{bmatrix} a & 0.6 \\ b & 0.8 \end{bmatrix}.$$

Since the columns of A must form an orthonormal set, I must have

$$(a, b) \cdot (0.6, 0.8) = 0 \quad \text{and} \quad \|(a, b)\| = 1.$$

(Note that $\|(0.6, 0.8)\| = 1$ already.) The first equation gives

$$0.6a + 0.8b = 0.$$

The easy way to get a solution is to swap 0.6 and 0.8 and negate one of them; thus, $a = -0.8$ and $b = 0.6$.

Since $\|(-0.8, 0.6)\| = 1$, I'm done. (If the a and b I chose had made $\|(a, b)\| \neq 1$, then I'd simply divide (a, b) by its length.) □

Example. Orthogonal 2×2 matrices represent **rotations of the plane about the origin** or **reflections across a line through the origin**.

Rotations are represented by matrices

$$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}.$$

You can check that this works by considering the effect of multiplying the standard basis vectors $(1, 0)$ and $(0, 1)$ by this matrix.

Multiplying a vector by the following matrix product reflects the vector across the line L that makes an angle θ with the x-axis:

$$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}.$$

Reading from right to left, the first matrix rotates everything by $-\theta$ radians, so L coincides with the x-axis. The second matrix reflects everything across the x-axis. The third matrix rotates everything by θ radians. Hence, a given vector is rotated by $-\theta$ and reflected across the x-axis, after which the reflected vector is rotated by θ. The net effect is to reflect across L.

Many transformation problems can be easily accomplished by doing transformations to reduce a general problem to a special case. □